Available online at www.sciencedirect.com
SCIENCE DIRECT®
International Journal of Forecasting 22 (2006) 73–89
international journal of forecasting
www.elsevier.com/locate/ijforecast

On a threshold heteroscedastic model

Cathy W.S. Chen [a], Mike K.P. So [b,*]

[a] Graduate Institute of Statistics and Actuarial Science, Feng Chia University, Taiwan
[b] Department of Information and Systems Management, School of Business and Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

Abstract

This paper proposes a threshold heteroscedastic model which integrates threshold nonlinearity and GARCH-type conditional variance for modeling mean and volatility asymmetries in financial markets. The main feature of this model is that the threshold variable for regime switching is formulated as a weighted average of important auxiliary variables. Estimation and diagnostic checks are performed using Markov chain Monte Carlo methods. Forecasts of volatility and value at risk can also be generated from predictive distributions. The proposed methodology is illustrated using both simulated and actual international market index data. Empirical results show higher average volatility and more persistent volatility when bad news arrives. While the domestic return is the major determinant of the regimes, both the SP 500 and Nikkei 225 indices also impact the dynamic structure of domestic market returns.

Keywords: Asymmetry; Auxiliary variables; GARCH model; Markov chain Monte Carlo; Model diagnostics; Stock returns

1. Introduction

It is well known that financial market volatility changes over time. A common way to capture this phenomenon was suggested by Engle (1982), who introduced the ARCH model for modeling time-varying conditional variance of financial returns. Bollerslev (1986) proposed extending this the GARCH model, which became a widely accepted model to describe the time series properties of financial market returns. However, if we model a financial series such as the returns of a stock index, the simple GARCH model cannot capture some stylized facts. For example, there is growing evidence that stock returns respond differently to positive and negative shocks. While the GARCH model assumes that positive and negative shocks have the same effect on future volatility, several specifications like the exponential GARCH model of Nelson (1991), the GJR model of Glosten, Jagannathan, and Runkle (1993) and the threshold model with ARCH errors of Li and Lam (1995) have been proposed to introduce asymmetry in the mean or volatility equations. In this paper, we develop a time series model that permits nonlinearity in the conditional mean and conditional variance.

* Corresponding author. Tel.: +852 23587726; fax: +852 23581946.
E-mail address: immkpso@ust.hk (M.K.P. So).

0169-2070/$ - see front matter © 2005 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2005.08.001

A popular model for explaining asymmetry in the mean is the threshold autoregressive model of Tong (1978); see also Tong (1983), Tong and Lim (1980) and Tsay (1989). To describe mean asymmetry and changing volatility, Tong (1990, p.116) advocates the use of a threshold model with ARCH errors. Li and Lam (1995) present some evidence of mean asymmetry in Hong Kong stock market returns using the threshold ARCH. Li and Li (1996) seem to be the first to attempt to model mean and volatility asymmetry together. They propose a double threshold ARCH model for financial market returns, applying an iteratively weighted least squares approach to obtain the maximum likelihood estimates, and using standardized residual autocorrelations and squared residual autocorrelations for checking model adequacy. Brooks (2001) further generalizes this to a double threshold GARCH model. However, both the models of Li and Li (1996) and of Brooks (2001) suffer from the weakness that the threshold variable is subjectively fixed. This model can be improved upon because the term structure of a financial return can be determined by many factors, such as its past values, international market movements, economic indices and interest rates. It would be too restrictive if we assume that the threshold variable, like security returns, undergoes regime shifts determined only by its own past values. In this paper, we introduce a threshold heteroscedastic model which has the threshold variable z_t defined by a weighted average of auxiliary variables:

$$z_t = w_1 z_{1t} + \ldots + w_m z_{mt}.$$

These auxiliary variables are believed to be highly relevant to regime switching. The threshold variable is only partially observable because it depends on the unknown parameters w_i. Differing from Li and Li (1996) and Brooks (2001), we let the data determine the most appropriate z_t by estimating w_i. This helps us understand the relative importance of z_{it}, $i=1,\ldots,m$, to the dynamic structure of security returns. In our model formulation, we also allow exogenous variables and a flexible error distribution which is essential for computing value at risk (VaR). Bayesian estimation is carried out by Markov chain Monte Carlo (MCMC) methods which are found to be very effective in simple threshold models (Chen & Lee, 1995). While the classical estimation method in Li and Li (1996) requires a fixed threshold value and a delay parameter in advance, we can jointly estimate the parameters in the mean and variance equations, the w_i, the threshold value and the time delay. We also perform diagnostic checking using the method in Gerlach, Carter, and Kohn (1999), which was demonstrated to be very effective in time series regression (Chen & Wen, 2001).

This paper is organized as follows. Section 2 defines the proposed threshold heteroscedastic model. Section 3 describes the Bayesian setup and detailed procedures for carrying out the Bayesian inference. Section 4 introduces a Bayesian forecasting method for volatility and VaR under the threshold heteroscedastic model. Model adequacy checking and model selection methods are presented in Section 5. Simulation results showing the performance of our Bayesian methods are given in Section 6. Applications of the model to financial markets are shown in Section 7. Section 8 contains concluding remarks.

2. A threshold heteroscedastic model

Motivated by the TAR model of Tong (1978, 1983) and the GARCH model of Bollerslev (1986), we introduce the following threshold heteroscedastic model:

$$y_t = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{l=1}^{q_j} \psi_l^{(j)} x_{lt} + a_t,$$

$$\text{if} \quad r_{j-1} \le z_{t-d} < r_j,$$

$$a_t = \sqrt{h_t}\epsilon_t, \qquad \epsilon_t \sim D(0,1),$$

$$h_t = \alpha_0^{(j)} + \sum_{i=1}^{d_j} \alpha_i^{(j)} a_{t-i}^2 + \sum_{l=1}^{c_j} \beta_l^{(j)} h_{t-l}, \tag{1}$$

where

$$z_t = w_1 z_{1t} + \ldots + w_m z_{mt}, \qquad 0 \le w_i \le 1,$$

$$\sum_{i=1}^{m} w_i = 1, \tag{2}$$

for $j=1,\ldots,g$, d is a positive integer and $D(0,1)$ is an error distribution with mean 0 and variance 1. The number of regimes is assumed to be g. Unknown parameters are also admitted in this error distribution.

The threshold values r_j satisfy $-\infty = r_0 < r_1 < \ldots < r_g = \infty$, and so the intervals $[r_{j-1}, r_j)$, $j=1,\ldots,g$, form a partition of the space of z_{t-d}. The positive integer d is commonly referred to as the delay (or threshold lag) of the model; it determines the time lag for which the threshold variable z_t has a large impact on the time series structure of y_t. To avoid the same process having more than one representation, we assume that $\phi_{Pj}^{(j)} \neq 0$ and that the g autoregressive vectors $(\phi_0^{(j)},\ldots,\phi_{Pj}^{(j)})$, $j=1,\ldots,g$, are distinct. The main feature of our proposed threshold heteroscedastic model in Eqs. (1) and (2) is the construction of the threshold variable z_t as a linear function of m auxiliary variables z_{it}, which are believed to affect the dynamic structure of y_t. In general, z_{it} can be any function of exogenous variables and $y_t,\ldots,y_1$. For example, we can choose $z_{1t}=y_t$ and $z_{it}=x_{it}$ for $i>1$, where x_{it} are exogenous variables. When $w_1=1$ and $w_i=0$ for $i>1$, our model reduces to the double threshold models of Li and Li (1996) and Brooks (2001). This choice of auxiliary variables is used in our application to real data. Another possibility is to equate z_{1t} with a_t instead of with y_t, as

$$a_t = \left(y_t - \phi_0^{(j)} - \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} - \sum_{l=1}^{q_j} \psi_l^{(j)} x_{lt}\right) \times I\left(r_{j-1} \leq z_{t-d} < r_j\right)$$

which is also a function of exogenous variables and y_ts. This particular z_{1t} enables us to account for the leverage effect; the mean and variance dynamics of a financial time series y_t respond differently to positive and negative price variations. A particular case belonging to this formulation of $z_{1t}=a_t$ is the GJR–GARCH model of Glosten et al. (1993):

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \delta a_{t-1}^2 I(a_{t-1}>0) + \beta_1 h_{t-1},$$

which corresponds to $w_1=1$ and $w_i=0$ for $i>1$ in our model.

To suitably define z_t, and to satisfy stationarity conditions, we apply constraints on the model parameters. Regarding the threshold variable, the equation $z_t = \sum_{i=1}^m w_i z_{it}$ and the inequalities $r_{j-1} \leq z_{t-d} < r_j$ cannot uniquely determine the threshold values or w_i. The reason for this can be understood by observing that $\sum_{i=1}^m w_i z_{it} < r_j$ if and only if $\sum_{i=1}^m (cw_i) z_{it} < cr_j$ for any positive constant c. Therefore, if w_i and r_j are the true parameter values, then cw_i and cr_j are also valid. To uniquely identify the threshold values r_j and the weights w_i, we impose positivity and sum-to-one constraints in Eq. (2), i.e., $0 \leq w_i \leq 1$ and $\sum_{i=1}^m w_i = 1$. Consequently, z_t is a weighted average of the auxiliary variables. In the case when z_{it} are returns of different markets, $z_t = \sum_{i=1}^m w_i z_{it}$ can be interpreted as a kind of portfolio return. The model specification in Eq. (1) implies that the conditional mean of y_t is given by

$$\mu_t = E\left[y_t | \boldsymbol{y}^{1,t-1}, \boldsymbol{x}_t\right] = \phi_0^{(j)} + \sum_{i=1}^{p_j} \phi_i^{(j)} y_{t-i} + \sum_{i=1}^{q_j} \psi_l^{(j)} x_{lt},$$

where $\boldsymbol{y}^{s,t} = (y_s,\ldots,y_t)'$, $\boldsymbol{x}_t = (x_{1t},\ldots,x_{qt})'$ and $q = \max\{q_1,\ldots,q_g\}$. This resembles a typical AR(p_j) process with q_j exogenous variables x_{lt}. We restrict the autoregressive parameters by

$$\sum_{i=1}^{p} \max_j |\phi_i^{(j)}| < 1,$$

where $p = \max\{p_1,\ldots,p_g\}$ and $\phi_i^{(j)}=0$ for $i>p_j$. Combining this with the standard restrictions on the variance parameters:

$$\alpha_0^{(j)}>0,\ \alpha_i^{(j)},\ \beta_l^{(j)} \geq 0 \quad \text{and} \quad \sum_{i=1}^{d_j} \alpha_i^{(j)} + \sum_{i=1}^{c_j} \beta_l^{(j)} < 1, \tag{3}$$

the results of Ling (1999) then guarantee the stationarity of our threshold heteroscedastic model.

Unlike traditional threshold modeling that chooses the threshold variable subjectively, see for example Tong and Lim (1980), Tsay (1989), Li and Li (1996) and Brooks (2001), our z_t is only partially observed because of the unknown parameters w_i. Therefore, the above model generalizes the double threshold ARCH of Li and Li (1996). It also extends the double threshold GARCH model of Brooks (2001) by incorporating exogenous variables and flexible error distribution D(0,1). Typical empirical evidence in the literature tells us that ϵ_t is usually fat-tailed. Therefore, assuming normality as in Li and Li (1996) and Brooks (2001) may create substantial bias in the return percentile and volatility forecasts, especially when we are interested in extreme percentiles of the return distribution, as in VaR estimation. In the empirical applications of this paper, we assume that $D(0,1)$ is the standardized t-distribution which captures the usual conditional leptokurtosis in financial return data.

Some advantages of our way of formulating z_t are noteworthy. First, our specification in Eq. (2) allows the simultaneous influence of more than one auxiliary variable in determining the regime of the threshold heteroscedastic model. Second, we can let the data choose the most suitable z_t by estimating w_i. Third, the weights w_i can reflect the relative importance of each z_{it} in governing the time series behavior of y_t. Concerning the identification of z_t, we choose promising values of z_{it} closely related to the dynamics of y_t. In practice, the auxiliary variables can be y_t and other related factors. For instance, if y_t is a stock return, z_{it} can be taken as a market index return and the returns of major stocks in the same industry. Similarly, if y_t is a market index return, we can choose z_{it} to be y_t, index returns of major financial markets, interest rates and other economic variables. After fixing the auxiliary variables, we estimate the weights by the Bayesian procedures discussed in Section 3.

3. Bayesian inference

Parameter estimation in homoscedastic threshold models is usually performed in two steps; see for example Tong and Lim (1980), Tong (1990) and Tsay (1989, 1998). For fixed values of d and r_j, the other parameters are estimated first. Then estimates of d and r_j can be determined by minimizing the AIC (Tong, 1990; Tong & Lim, 1980), by minimizing a nonlinearity test statistic and using scatterplots (Tsay, 1989), or by minimizing a conditional least square (Tsay, 1998). In the above two-step procedure, the threshold variable z_t is assumed to be known. It is not clear whether existing asymptotic results, like that in Tsay (1998, p.1195), hold in our case with partially unknown z_t. Moreover, the classical methods in Li and Li (1996) and Brooks (2001) are subject to some weaknesses. First, the uncertainty concerning r_j and d cannot be taken into account for statistical inference if we follow Li and Li (1996) and Brooks (2001) and fix r_j and d to maximize the likelihood function. Second, their final choice of r_j and d is dependent on the criteria they use. To tackle the above weaknesses, we suggest in this paper a Bayesian approach to performing simultaneous inference for all parameters, including d, r_j and w_i. We generate samples from the joint posterior distribution of all parameters via Markov chain Monte Carlo methods. One advantage of our approach is that it allows r_j and d to be estimated simultaneously with other parameters and thus incorporates their variations into the statistical inference. Another advantage of the Bayesian approach is that we can estimate z_t via w_i rather than fixing it subjectively. Estimating z_t and the unknown parameters simultaneously enables us to take into account the uncertainty in z_t in estimating the parameters. In other words, the statistical inference, like model selection and diagnostic checking, can be done while accounting for possible the estimation error of z_t. The final advantage is that the prior information regarding the parameters can also be incorporated in their prior distributions.

Define $\boldsymbol{\phi}_j=(\phi_0^{(j)},\ldots,\phi_{p_j}^{(j)},\ \psi_1^{(j)},\ldots,\psi_{qj}^{(j)})'$, $\boldsymbol{\alpha}_j=(\alpha_0^{(j)},\ldots,\alpha_{d_j}^{(j)},\ \beta_1^{(j)},\ldots,\beta_{c_j}^{(j)})'$, $\boldsymbol{r}=(r_1,\ldots,r_{g-1})'$ and $\boldsymbol{w}=(w_1,\ldots,w_{m-1})'$ as the mean, variance, threshold and weight parameter vectors, respectively. The weight $\boldsymbol{w}$ has dimension $m-1$ instead of m because the last element w_m can be determined as $1-\boldsymbol{w}'\mathbf{1}$, where $\mathbf{1}$ is a column vector of 1s. Denote the unknown parameter $(\boldsymbol{\phi}_1',\ldots,\boldsymbol{\phi}_g',\boldsymbol{\alpha}_1',\ldots,\boldsymbol{\alpha}_g',\boldsymbol{r}',\boldsymbol{w}',d)'$ by $\boldsymbol{\theta}$. Let d_0 be the maximum delay and $s=\max\{p_1,\ldots,p_g,\ d_0\}$. The conditional likelihood function of the model is given by

$$p\left(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}\right)=\prod_{t=s+1}^{n}\left[\sum_{j=1}^{g}\frac{1}{\sqrt{h_t}}p_\epsilon\left(\frac{y_t-\mu_t}{\sqrt{h_t}}\right)I_{jt}\right],\qquad(4)$$

where $p_\epsilon(\cdot)$ is the probability density function of ϵ_t, and I_{jt} is the indicator variable $I(r_{j-1}\le z_{t-d}<r_j)$. The dependence of the conditional likelihood on $\boldsymbol{y}^{1,s}$ is ignored for notational simplicity.

3.1. The prior

With maximum delay d_0, we assume the uniform prior $p(d)=\frac{1}{d_0}$ for $d=1,\ldots,d_0$. To represent prior ignorance regarding $\boldsymbol{w}$, we take $p(\boldsymbol{w})=I(B)$, where B is the set of $\boldsymbol{w}$ having $0\le w_i\le 1$ and $\sum_{i=1}^{m}w_i=1$. Similarly, $p(\alpha_j)$ is the indicator $I(S_j)$, $j=1,\ldots,g$, where S_j is the set of $\boldsymbol{\alpha}_j$ that satisfy the restrictions in Eq. (3). In threshold modeling, it is important to set the minimum number of observations in each regime so that we can generate meaningful inference results. Since z_{t-d} is dependent on $\boldsymbol{w}$, rather than selecting an unconditional prior for $\boldsymbol{r}$ to fix the minimum number, we choose the conditional prior of $\boldsymbol{r}$ given $\boldsymbol{w}$, $p(\boldsymbol{r}|\boldsymbol{w})$, as $I(A)$ where A is the event that each regime contains at least h percent z_{t-d}. For example, in a 2-regime threshold model, i.e. $g=2$, with $z_{1t}=y_t$ and $d=1$, having at least 10% of the data in each regime ($h=10$) means that $p(\boldsymbol{r}|w_1=1)$, the conditional prior of $\boldsymbol{r}=r_1$ given $w_1=1$, is the uniform distribution from the 10th percentile of y_{t-1} to the 90th percentile of y_{t-1}. Overall, the following prior for $\boldsymbol{\theta}$ is adopted:

$$p(\boldsymbol{\theta}) \propto \left[\prod_{j=1}^{g} p(\phi_j)p(\boldsymbol{\alpha}_j)\right] p(d)p(\boldsymbol{r}|\boldsymbol{w})p(\boldsymbol{w}), \tag{5}$$

where $p(\phi_j) \sim N(\phi_{j0}, \boldsymbol{V}_j)$.

3.2. Sampling scheme

Let π_t be the time index in ascending order for the set of time points corresponding to regime j, i.e., let $\{t:t=s+1,\ldots,n,\ I_{jt}=1\}$ and $a=\sum_{t=s+1}^{n} I_{jt}$ be the number of observations in regime j. Also, let $\boldsymbol{\theta}_{-\gamma}$ be the parameter vector $\boldsymbol{\theta}$ excluding the element γ. The prior in Eq. (5) leads to the conditional posterior

$$p\left(\phi_j|\boldsymbol{y}^{s+1,n}, \theta_{-\phi_j}\right) \propto p\left(\boldsymbol{y}^{s+1,n}|\theta\right)p\left(\phi_j\right). \tag{6}$$

In the simplest case, that is, when ϵ_t is normal and h_t is constant, Eq. (6) reduces to

$$\begin{aligned} p(\phi_j|y^{s+1,n}, \boldsymbol{\theta}_{-\phi_j}) &= \prod_{t=s+1, I_{jt}=1}^{n} \left[\frac{1}{\sqrt{h_t}} p_\epsilon\left(\frac{y_t-\mu_t}{\sqrt{h_t}}\right)\right] p(\phi_j) \\ &\propto |\boldsymbol{H}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{Z}\phi_j\right)^T \boldsymbol{H}^{-1}\left(\boldsymbol{y}-\boldsymbol{Z}\phi_j\right) - \frac{1}{2}\left(\phi_j-\phi_{j0}\right)^T \boldsymbol{V}_j^{-1}\left(\phi_j-\phi_{j0}\right)\right\} \\ &\propto |\boldsymbol{H}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\phi_j-\phi^*\right)^T \boldsymbol{V}^{*-1}\left(\phi_j-\phi^*\right)\right\}, \end{aligned} \tag{7}$$

where

$$\phi^* = \boldsymbol{V}^*\left(\boldsymbol{Z}^T\boldsymbol{H}^{-1}\boldsymbol{y} + \boldsymbol{V}_j^{-1}\phi_{j0}\right),$$

$$\boldsymbol{V}^* = \left(\boldsymbol{Z}^T\boldsymbol{H}^{-1}\boldsymbol{Z} + \boldsymbol{V}_j^{-1}\right)^{-1},$$

$$\boldsymbol{Z} = \begin{bmatrix} 1 & y_{\pi_1-1} & \cdots & y_{\pi_1-p_j} & x_{1_{\pi_1}} & \cdots & x_{q_j\pi_1} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & y_{\pi_a-1} & \cdots & y_{\pi_a-p_j} & x_{1_{\pi_a}} & \cdots & x_{q_j\pi_a} \end{bmatrix},$$

$$\boldsymbol{y} = \begin{bmatrix} y_{\pi_1} \\ y_{\pi_2} \\ \vdots \\ y_{\pi_a} \end{bmatrix} \quad \text{and} \quad \boldsymbol{H} = \begin{bmatrix} h_{\pi_1} & 0 & \cdots & 0 \\ 0 & h_{\pi_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & h_{\pi_a} \end{bmatrix}.$$

The conditional posterior in Eq. (7) is a Gaussian distribution which provides us with a kernel for the drawing of ϕ_j, under general p_ϵ and heteroscedasticity, by the Metropolis–Hasting (MH) method. Let f denote the target density in Eq. (6) for notational convenience. Details of the MH steps for ϕ_j are as follows:

Step 1: At iteration i, generate a point ϕ_j^* from $\phi_j^*=\phi_j^{[i-1]}+\epsilon$, where $\phi_j^{[i-1]}$ is the $(i-1)$th iterate of ϕ_j and $\epsilon \sim N(0,a\Omega)$.
Step 2: Accept ϕ_j^* as $\phi_j^{[i]}$ with probability $p=\min\{1, f(\phi_j^*)/f(\phi_j^{[i-1]})\}$. Otherwise, set $\phi_j^{[i]}=\phi_j^{[i-1]}$.

We observe from Eq. (7) that $p(\phi_j|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{-\phi j})$ is a normal density with covariance matrix $\boldsymbol{V}^*$ under normality and constant conditional variance. Therefore, it is anticipated that $\boldsymbol{V}^*$ contains useful information about the conditional correlations of $\phi_i^{(j)}$ and $\psi_i^{(j)}$, even when p_ϵ is non-normal. Toward this goal, we choose Ω to be $\boldsymbol{V}^*$ with all h_t in $\boldsymbol{H}$ replaced by the sample variance of y_t. The positive scale a is fixed by controlling the acceptance rate (Gelman, Roberts & Gilks, 1996). A suitable value of a with good convergence properties can usually be selected by having an acceptance probability of 25% to 50%.

We again apply the random walk MH algorithm to the GARCH parameter $\boldsymbol{\alpha}_j$ for the first M iterations, with the target density f given by

$$p\left(\boldsymbol{\alpha}_j|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{-\alpha_j}\right) \propto p\left(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}\right)I\left(S_j\right).$$

We form the sample mean μ_α and sample covariance Ω_α using the first M iterates of $\boldsymbol{\alpha}_j$. Then, using the Gaussian proposal distribution with mean μ_α and covariance Ω_α, we apply the following independent kernel MH algorithm starting from the $(M+1)$th iteration to speed up the convergence:

Step 1: At iteration i, generate a point $\boldsymbol{\alpha}_j^*$ according to $\boldsymbol{\alpha}_j^* = \mu_\alpha + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(0, \Omega_\alpha)$.
Step 2: Accept $\boldsymbol{\alpha}_j^*$ as $\boldsymbol{\alpha}_j^{[i]}$ with probability

$$p = \min\left\{1, \frac{f\left(\boldsymbol{\alpha}_j^*\right)g\left(\boldsymbol{\alpha}_j^{[i-1]}\right)}{f\left(\boldsymbol{\alpha}_j^{[i-1]}\right)g\left(\boldsymbol{\alpha}_j^*\right)}\right\},$$

where $\boldsymbol{\alpha}_j^{[i]}$ is the ith iterate of $\boldsymbol{\alpha}_j$ and $g(\boldsymbol{\alpha})$ is the Gaussian proposal density with mean μ_α and covariance matrix Ω_α. Otherwise, set $\boldsymbol{\alpha}_j^{[i]} = \boldsymbol{\alpha}_j^{[i-1]}$.

To draw the threshold vector $\boldsymbol{r}$ and the weight vector $\boldsymbol{w}$, we also use the random walk MH algorithm with the target densities given by

$$p\left(\boldsymbol{r}|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{-\boldsymbol{r}}\right) \propto p\left(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}\right)p(\boldsymbol{r}|\boldsymbol{w})$$

and $\quad p\left(\boldsymbol{w}|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{-\boldsymbol{w}}\right) \propto p\left(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}\right)p(\boldsymbol{r}|\boldsymbol{w})p(\boldsymbol{w}).$

Drawing d is made easy by noting that it is discrete and has the posterior probabilities

$$p\left(d = j|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{-d}\right) = \frac{p(\boldsymbol{y}^{s+1,n}|d = j, \boldsymbol{\theta}_{-d})}{\sum_{i=1}^{d_0} p\left(\boldsymbol{y}^{s+1,n}|d = i, \boldsymbol{\theta}_{-d}\right)},$$

$j = 1, \ldots, d_0.$

4. Forecasting volatility and value at risk

In addition to estimation of unknown parameters, volatility and value at risk forecasting are also important in analyzing derivative pricing, yielding a good portfolio for investment and managing market risk in financial markets. This section derives a simulation-based approach for predicting future values of h_{n+k}, $k \geq 1$, and for generating VaR forecasts. Applying the sampling scheme in Section 3 allows us to generate a Monte Carlo sample $\boldsymbol{\theta}^{[i]}$, $i = 1, \ldots, N$, from the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}^{1,n})$. As the conditional variance h_t is a deterministic function of $\boldsymbol{\theta}$ and $\boldsymbol{y}^{1,n}$, smoothed values of h_t, $t = 1, \ldots, n$, that are compiled from $p(h_t|\boldsymbol{y}^{1,n})$ are readily obtained. Let $h_t = g_t(\boldsymbol{y}^{1,t-1}, \boldsymbol{\theta})$ where

$$g_t\left(\boldsymbol{y}^{1,t-1}, \theta\right) = \sum_{j=1}^{g}\left[\alpha_0^{(j)} + \sum_{i=1}^{d_j}\alpha_i^{(j)}a_{t-i}^2 + \sum_{l=1}^{c_j}\beta_l^{(j)}h_{t-l}\right]I_{jt}. \tag{8}$$

Then $h_t^{[i]} = g_t(\boldsymbol{y}^{1,t-1}, \boldsymbol{\theta}^{[i]})$, $i = 1, \ldots, N$, form a posterior sample of $p(h_t|\boldsymbol{y}^{1,n})$. A natural smoothed estimate of h_t is

$$\frac{1}{N}\sum_{i=1}^{N} h_t^{[i]}.$$

To yield volatility forecasts, we simulate h_{n+k} $(k \geq 1)$ from the predictive distribution $p(h_{n+k}|\boldsymbol{y}^{1,n})$. Because of the deterministic relationship in Eq. (8), it is sufficient to generate Monte Carlo samples from $p(y_{n+1}, \ldots, y_{n+k-1}, \boldsymbol{\theta}|\boldsymbol{y}^{1,n})$. In particular for $k = 1$, $h_{n+1}^{[i]} = g_{n+1}(\boldsymbol{y}^{1,n}, \boldsymbol{\theta}^{[i]})$ are draws from the predictive distribution $p(h_{n+k}|\boldsymbol{y}^{1,n})$ and so h_{n+1} can be estimated by the sample mean $\sum_{i=1}^{N} h_{n+1}^{[i]}/N$ or other location measures. For $k \geq 2$, we need to sample $y_{n+1}^{[i]}, \ldots, y_{n+k-1}^{[i]}$ from $p(y_{n+1}, \ldots, y_{n+k-1}, \boldsymbol{\theta}|\boldsymbol{y}^{1,n})$. This can be done by the method of composition and decomposition:

$$p\left(y_{n+1}, \ldots, y_{n+k-1}, \boldsymbol{\theta}|\boldsymbol{y}^{1,n}\right) = p\left(\boldsymbol{\theta}|\boldsymbol{y}^{1,n}\right)\prod_{j=1}^{k-1} p\left(y_{n+j}|\boldsymbol{y}^{1,n+j-1}, \boldsymbol{\theta}\right).$$

We draw $y_{n+1}, \ldots, y_{n+k-1}$ sequentially as follows:

1. With $h_{n+1}^{[i]}$, draw $y_{n+1}^{[i]}$ from $p(y_{n+1}|\boldsymbol{y}^{1,n}, \boldsymbol{\theta}^{[i]})$ and set $j = 2$.
2. Calculate $h_{n+j}^{[i]} = g_{n+j}(\boldsymbol{y}^{1,n}, y_{n+1}^{[i]}, \ldots, y_{n+j-1}^{[i]}, \boldsymbol{\theta}^{[i]})$.
3. Draw $y_{n+j}^{[i]}$ from $p(y_{n+j}|\boldsymbol{y}^{1,n}, y_{n+1}^{[i]}, \ldots, y_{n+j-1}^{[i]}, \boldsymbol{\theta}^{[i]})$.
4. Repeat steps 2 and 3 for $j = 3, \ldots, k-1$.

According to Eqs. (1) and (2), simulations in steps 1 and 3 are done via suitable t-distributions. Implementing the above iterative scheme creates a draw of $y_{n+1}^{[i]}, \ldots, y_{n+k-1}^{[i]}$ from $p(y_{n+1}, \ldots, y_{n+k-1}, \boldsymbol{\theta}|\boldsymbol{y}^{1,n})$. Obviously, k-step-ahead forecasts of volatility can be formed by using $h_{n+j}^{[i]}$ in step 2.

VaR is a common risk measure. It is usually defined as the loss of a portfolio that is exceeded with a predetermined probability over a time horizon. To forecast VaR, we need to estimate extreme pth percentiles of y_{n+k}. For multiple-period VaR estimation, $k > 1$, Bayesian forecasts are created with the percentiles of the predictive distribution $p(y_{n+k}|\boldsymbol{y}^{1,n})$. These 'predictive' percentiles can be estimated by the sample pth percentiles of $y_{n+k}^{[i]}$, $i = 1, \ldots, N$. The Bayesian VaR predictor derived using $p(y_{n+k}|\boldsymbol{y}^{1,n})$ under our threshold heteroscedastic model provides an alternative to existing methods like those in RiskMetrics and Wong and So (2003).

5. Testing model adequacy and model selection

5.1. Diagnostic checking

In the usual frequentist approach to parameter estimation, residual autocorrelation has become a standard device for checking time series model inadequacy. In threshold ARCH modeling, Li and Li (1996) use residual and squared residual autocorrelations to detect misspecification in the mean and variance equations. Based on our MCMC methods for carrying out the statistical inference, it is natural to consider a Bayesian way of performing diagnostic checking. In this paper, we choose the method of Gerlach et al. (1999), which is based upon the time series

$$u_t = F\left(y_t|\mathbf{y}^{1,t-1}\right), \qquad t = 1, \ldots, n, \tag{9}$$

where $F(\cdot)$ is the conditional distribution function of y_t given $\mathbf{y}^{1,t-1}$ under the tested model. It is not difficult to see that if the underlying tested model is correct, u_t will be independent and identically distributed (i.i.d.) uniform on [0,1]; see Rosenblatt (1952). Similarly, $v_t=\boldsymbol{\Phi}^{-1}(u_t)$; the inverse transformation of the standard normal distribution function has a standard normal distribution under the correct model specification. Smith (1985) uses these properties of u_t and v_t to build up diagnostic tests for time series models with known parameters. In the more realistic situation that $\boldsymbol{\theta}$ is unknown, Gerlach et al. (1999) propose a simulation-based method to estimate u_t and v_t. Suppose that we have a posterior sample of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_k^{[i]}$, $i=1,\ldots,N$, from $p(\boldsymbol{\theta}|\mathbf{y}^{1,k})$ for a given k. This sample can be obtained by our sampling scheme in Section 3. Gerlach et al. (1999) show that for $k \geq t$,

$$\hat{u}_t = \frac{\sum_{i=1}^{N} F\left(y_t|\mathbf{y}^{1,t-1}, \boldsymbol{\theta}_k^{[i]}\right)/p\left(\mathbf{y}^{t,k}|\mathbf{y}^{1,t-1}, \boldsymbol{\theta}_k^{[i]}\right)}{\sum_{i=1}^{N} 1/p\left(\mathbf{y}^{t,k}|\mathbf{y}^{1,t-1}, \boldsymbol{\theta}_k^{[i]}\right)} \tag{10}$$

converges to u_t as $N \rightarrow \infty$. Similarly, for $k<t$,

$$\hat{u}_t = \frac{\sum_{i=1}^{N} F\left(y_t|\mathbf{y}^{1,t-1}, \boldsymbol{\theta}_k^{[i]}\right) p\left(\mathbf{y}^{k+1,t-1}|\mathbf{y}^{1,k}, \boldsymbol{\theta}_k^{[i]}\right)}{\sum_{i=1}^{N} p\left(\mathbf{y}^{k+1,t-1}|\mathbf{y}^{1,k}, \boldsymbol{\theta}_k^{[i]}\right)} \tag{11}$$

converges to u_t as $N \rightarrow \infty$. Since the variance of $\hat{u}_t$ in Eq. (10) increases with $k-t$, we need to calculate $\hat{u}_t$ with t reasonably close to k. This can be achieved by increasing k sequentially, say to 100, 200,..., n, and evaluate $\hat{u}_t$ using Eq. (10) with $k-t$ not greater than the increments. We can improve the precision of the approximation by reducing the increments, but longer computational time is involved. A similar arrangement to control the size of $|k-t|$ is required if we use Eq. (11) to estimate u_t.

Based on the convergence properties of $\hat{u}_t$, we construct $\hat{v}_t=\boldsymbol{\Phi}^{-1}(\hat{u}_t)$, which is approximately i.i.d. $N(0,1)$ under the correct model. We then examine whether our threshold model formulation is adequate by applying two standard tests to $\hat{v}_t$. The first one is analogous to Li and Li (1996) and explores its time series structure. The usual Portmanteau test statistic is applied for testing independence:

$$Q = n(n+2) \sum_{j=1}^{J} (n-j)^{-1} \hat{\rho}_j^2,$$

where $\hat{\rho}_j^2$ is the sample autocorrelation of $\hat{v}_t$. In addition, we also perform the Studentized range test (D'Agostino & Stephens, 1986, p.392) to see whether $\hat{v}_t$ deviates significantly from a standard normal distribution. The test statistic is the sample range of $\hat{v}_t$. The critical values of this range test are determined by simulations.

5.2. Model selection

To investigate whether our proposed formulation of z_i is superior to the original threshold model in Li and Li (1996) and Brooks (2001), we need to set up a model selection method. Evaluating the performance of our formulation means comparing our model in Eq. (1) with those in Li and Li (1996) and Brooks (2001). To compare two models A and B in the Bayesian framework, say A is our model and B is Li and Li's

(1996) model, we compute the posterior odds ratio of A vs. B as

$$\mathrm{POR}_{AB} = \frac{p_A(\boldsymbol{y}^{s+1,n})}{p_B(\boldsymbol{y}^{s+1,n})} \times \text{ prior odds ratio}, \qquad (12)$$

where $p_A(\boldsymbol{y}^{s+1,n})$ and $p_B(\boldsymbol{y}^{s+1,n})$ are the marginal likelihoods under models A and B, respectively. Without any prior information on which model is better, we set the prior odds ratio equal to 1 so that POR_{AB} is simplified to the Bayes factor $p_A(\boldsymbol{y}^{s+1,n})/p_B(\boldsymbol{y}^{s+1,n})$. The larger the value of POR_{AB}, the more evidence we have in support of A over B; normally, when $\mathrm{POR}_{AB}>1$, we conclude that A is preferable to B. Based on the above discussion, it suffices to compute the marginal likelihoods for model selection.

We adopt the method in Gerlach et al. (1999) to estimate the marginal likelihood of a model. By replacing $F(y_t|\boldsymbol{y}^{1,t-1},\boldsymbol{\theta}_k^{[i]})$ with $p(y_t|\boldsymbol{y}^{1,t-1},\boldsymbol{\theta}_k^{[i]})$ in Eq. (10), we obtain an estimate for the predictive density $p(y_t|\boldsymbol{y}^{1,t-1})$ when $k \geq t$:

$$\hat{p}\left(y_t|\boldsymbol{y}^{1,t-1}\right) = \frac{\sum_{i=1}^{N} p\left(y_t|\boldsymbol{y}^{1,t-1},\boldsymbol{\theta}_k^{[i]}\right)/p\left(\boldsymbol{y}^{t,k}|\boldsymbol{y}^{1,t-1},\boldsymbol{\theta}_k^{[i]}\right)}{\sum_{i=1}^{N} 1/p\left(\boldsymbol{y}^{t,k}|\boldsymbol{y}^{1,t-1},\boldsymbol{\theta}_k^{[i]}\right)}.$$

Likewise, we can do the same substitution in Eq. (11) to estimate the predictive density when $k<t$. An estimate of the marginal likelihood $p(\boldsymbol{y}^{s+1,n})$ is then evaluated as

$$\prod_{t=s+1}^{n} \hat{p}\left(y_t|\boldsymbol{y}^{1,t-1}\right).$$

6. Simulation studies

We perform simulation studies to illustrate the proposed Bayesian estimation and diagnostic checking methods. In the first part, we simulate time series from two threshold models to investigate the finite sample properties of the posterior mean estimator. In the second part, the model checking method discussed in Section 5 is applied to time series generated by five different models to study the power of the diagnostic test.

6.1. Parameter estimation

The orders p_j, q_j, c_j and d_j of the two threshold models considered are set to 1. The first model is

Model 1:

$$y_t = \begin{cases} 0.2+0.1y_{t-1}+a_t, & z_{t-1}\leq 0.1, \\ 0.2-0.1y_{t-1}+a_t, & z_{t-1}>0.1, \end{cases}$$

$$a_t=\sqrt{h_t}\epsilon_t, \quad \epsilon_t \sim t_{10},$$

$$h_t = \begin{cases} 0.5+0.02a_{t-1}^2+0.9h_{t-1}, & z_{t-1}\leq 0.1, \\ 0.2+0.05a_{t-1}^2+0.6h_{t-1}, & z_{t-1}>0.1, \end{cases}$$

$$z_t = 0.6z_{1t}+0.2z_{2t}+0.2z_{3t},$$

which has two regimes ($g=2$). The threshold variable z_t is a linear combination of $m=3$ auxiliary variables. We set the variance equation the match the empirical evidence that the first regime has higher volatility persistence, with $\alpha_1^{(1)}+\beta_1^{(1)}$ closer to 1 than $\alpha_1^{(2)}+\beta_1^{(2)}$, and a higher average variance, i.e. $\alpha_0^{(1)}/(\alpha_1^{(1)}+\beta_1^{(1)})>\alpha_0^{(2)}/(\alpha_1^{(2)}+\beta_1^{(2)})$. We assume that y_{t-2} is in the model, though $\phi_2^{(1)}$ and $\phi_2^{(2)}$ are zero in Model 1. To formulate the threshold variable, the exogenous variable x_{1t} is generated from the GARCH(1,1) process:

$$x_{1t} = \sqrt{h_t}\xi_t, \qquad h_t = 0.01+0.2x_{1t-1}^2+0.6h_{t-1},$$
$$\xi_t \sim N(0,1).$$

The three auxiliary variables z_{1t}, z_{2t} and z_{3t} are defined as y_t, x_{1t} and N(0,1) white noise, respectively. Model 2, formulated below, has specifications similar to Model 1:

Model 2 :

$$y_t = \begin{cases} 0.2+0.1y_{t-1}+0.3x_{1t}+a_t, & z_{t-1}\leq 0.1, \\ 0.2-0.1y_{t-1}-0.2x_{1t}+a_t, & z_{t-1}>0.1, \end{cases}$$

$$a_t = \sqrt{h_t}\epsilon_t, \qquad \epsilon_t \sim t_6,$$

$$h_t = \begin{cases} 0.5+0.02a_{t-1}^2+0.9h_{t-1}, & z_{t-1}\leq 0.1, \\ 0.2+0.05a_{t-1}^2+0.6h_{t-1}, & z_{t-1}>0.1, \end{cases}$$

$$z_t = 0.4z_{1t}+0.3z_{2t}+0.3z_{3t}.$$

The major differences between Models 1 and 2 are in the w_i and in the degrees of freedom v; the latter has more balance weights and fewer degrees of freedom.

To allow ϵ_t to be t-distributed, that is,

$$p_\epsilon(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \frac{1}{\sqrt{\pi(v-2)}} \left[1 + \frac{x^2}{v-2}\right]^{-\left(\frac{v+1}{2}\right)}, \tag{13}$$

we have to draw v in the MCMC iterations. This is achieved via a random walk MH algorithm as used in simulating ϕ_j. We adopt the prior $p(v)=I(v>2)$ so that ϵ_t has finite variance. For drawing d, the maximum delay is fixed at $d_0=3$. We also choose the non-informative prior for ϕ_j that V_j^{-1} is set to 0. In fact, simulations show that using informative priors for ϕ_j has a negligible effect on its posterior. To retain a sufficient number of observations for Bayesian analysis, we choose $h=5$ for $n=2000$ and $h=2.5$ for $n=4000$ in the prior $p(\boldsymbol{r}|\boldsymbol{w})$ so that there are at least 100 observations in each regime. In this simulation experiment, two sample sizes $n=2000$ and $n=4000$ with 100 replications are used for demonstration.

We carry out $N=20{,}000$ MCMC iterations and discard the first $M=8000$ burn-in iterates for each series. In all replicates, the posterior probabilities for $d=1$ are very close to 1, implying that the posterior mode of d accurately estimates the delay parameter. Tables 1 and 2 display the summary statistics from 100 replications for the posterior mean estimators of other parameters. The means of the estimators are close to the respective true values, indicating that the posterior mean obtained by our sampling scheme is a reliable estimator. We can also see that the bias is substantially reduced when the sample size is doubled. The only exceptions are for $\phi_2^{(2)}$ in Model 1 and v in Model 2 where a small bias is observed in both sample sizes. Concerning the variability of the posterior mean estimators, the standard property that their standard deviations diminish as the sample size increases is revealed. Even for the moderately large size of $n=2000$, we can conclude from the small standard deviations that the posterior mean is a precise estimator for all parameters.

Table 1
Summary statistics of the posterior mean estimator for Model 1 obtained from 100 replications

	True value	Mean	SD	Mean	SD
		$n=2000$		$n=4000$	
$\phi_0^{(1)}$	0.2	0.194	0.097	0.203	0.045
$\phi_1^{(1)}$	0.1	0.107	0.047	0.101	0.033
$\phi_2^{(1)}$	0	−0.005	0.043	−0.003	0.026
$\phi_0^{(2)}$	0.2	0.188	0.062	0.201	0.044
$\phi_1^{(2)}$	−0.1	−0.093	0.049	−0.103	0.034
$\phi_2^{(2)}$	0	−0.004	0.029	−0.005	0.019
$\alpha_0^{(1)}$	0.5	0.632	0.238	0.546	0.075
$\alpha_1^{(1)}$	0.02	0.040	0.022	0.028	0.013
$\beta_1^{(1)}$	0.9	0.826	0.087	0.868	0.046
$\alpha_0^{(2)}$	0.2	0.246	0.100	0.223	0.064
$\alpha_1^{(2)}$	0.05	0.056	0.024	0.054	0.016
$\beta_1^{(2)}$	0.6	0.566	0.064	0.581	0.051
r_1	0.1	0.088	0.115	0.105	0.015
v	10	11.547	4.404	10.333	1.952
w_1	0.6	0.615	0.061	0.609	0.039
w_2	0.2	0.184	0.073	0.189	0.051
w_3	0.2	0.200	0.032	0.202	0.017

Table 2
Summary statistics of the posterior mean estimator for Model 2 obtained from 100 replications

	True value	Mean	SD	Mean	SD
		$n=2000$		$n=4000$	
$\phi_0^{(1)}$	0.2	0.208	0.051	0.202	0.035
$\phi_1^{(1)}$	0.1	0.105	0.046	0.099	0.029
$\psi_1^{(1)}$	0.3	0.332	0.182	0.289	0.126
$\phi_0^{(2)}$	0.2	0.198	0.049	0.199	0.033
$\phi_1^{(2)}$	−0.1	−0.096	0.041	−0.098	0.026
$\psi_1^{(2)}$	−0.2	−0.211	0.143	−0.200	0.102
$\alpha_0^{(1)}$	0.5	0.637	0.148	0.547	0.090
$\alpha_1^{(1)}$	0.02	0.046	0.026	0.031	0.017
$\beta_1^{(1)}$	0.9	0.814	0.075	0.870	0.048
$\alpha_0^{(2)}$	0.2	0.261	0.138	0.224	0.071
$\alpha_1^{(2)}$	0.05	0.058	0.025	0.053	0.016
$\beta_1^{(2)}$	0.6	0.560	0.084	0.580	0.055
r_1	0.1	0.089	0.089	0.104	0.014
v	6	6.045	0.762	6.053	0.535
w_1	0.4	0.422	0.050	0.414	0.030
w_2	0.3	0.267	0.078	0.279	0.050
w_3	0.3	0.311	0.041	0.308	0.026

6.2. Diagnostic test performance

To study the power of the diagnostic test outlined in Section 5, we generate data series from the following five models:

Model 3 :

$$y_t = \begin{cases} 0.2 + 0.15y_{t-1} + 0.1y_{t-2} + a_t, & y_{t-1} \le -0.3, \\ -0.2 + 0.05y_{t-1} - 0.1y_{t-2} + a_t, & y_{t-1} > -0.3, \end{cases}$$

$$h_t = \begin{cases} 0.2 + 0.05a_{t-1}^2 + 0.85h_{t-1}, & y_{t-1} \le -0.3, \\ 0.05 + 0.1a_{t-1}^2 + 0.8h_{t-1}, & y_{t-1} > -0.3, \end{cases}$$

$$z_t = y_t \quad \text{and} \quad \epsilon_t \sim N(0,1),$$

Model 4 : Same as Model 3 but $\epsilon_t \sim t_5$.

Model 5 : Same as Model 3 but

$$y_t = \begin{cases} 0.2 + 0.15y_{t-1} + a_t, & y_{t-1} \le -0.3, \\ -0.2 + 0.05y_{t-1} - 0.2y_{t-2} + 0.4y_{t-3} + a_t, & y_{t-1} > -0.3. \end{cases}$$

Model 6 : Same as Model 3 but

$$y_t = \begin{cases} 0.2 + 0.15y_{t-1} + a_t, & y_{t-1} \le -0.3, \\ -0.2 + 0.05y_{t-1} - 0.2y_{t-2} + 0.4y_{t-3} + a_t, & y_{t-1} > -0.3. \end{cases}$$

Model 7: Same as Model 6 but $\epsilon_t \sim t_5$.

The fitted threshold model for the simulated series has $p_1 = p_2 = 2$, $q_1 = q_2 = 0$, $d_j = c_j = 1$, $z_t = y_t$ and $\epsilon_t \sim N(0,1)$, ensuring its adequacy for the data of Model 3. However, it is inadequate for the data of Model 4, as we assume that ϵ_t is normal in the model fitting. Likewise, the mean equation is misspecified for the time series of Models 5 and 6, which have correct order $p_2 = 3$. In Model 7, the error distribution and the autoregressive orders are both misspecified.

In this simulation experiment, we apply the MCMC method with $M = 8000$ and $N = 20\,000$ to each simulated series to estimate u_t, $t = 501, \ldots, 2000$, via Eq. (10). This is done by setting the value of k sequentially at 550, 600,..., 2000 and estimating u_t with $k - t < 50$ for each k. The Portmanteau test with $J = 10$ and the Studentized range test are used to check whether $\hat{v}_t$ violates the hypothesis that it is i.i.d. $N(0,1)$. Under the null hypothesis, the 90% confidence interval $(6.03, 7.55)$ and the 95% confidence interval $(5.93, 7.76)$ of the Studentized range test statistic for 1500 observations are obtained by simulations to conduct the diagnostic test. Table 3 reports the number of rejections out of 1000 replications. Since the data from Model 3 are correctly fitted, it is good to see that the empirical sizes of the Portmanteau test (4.7% and 10.0%) and the Studentized range test (5.2% and 10.6%) match the nominal values of 5% and 10% very closely. For Model 4, where there is misspecification in the error distribution, the powers of the Studentized range test, 99.7% and 100.0%, are very close to 100%, implying that the test is very powerful in detecting lack of fit in the distribution of ϵ_t. The Portmanteau test has reasonably high rejection rates when the autoregressive order p_2 is wrongly assumed to be 2 for Model 5 though it has low power for Model 4. When the discrepancy between the fitted model and the true model is expanded by increasing $\phi_3^{(2)}$ to 0.4 in Model 6, the powers of the Portmanteau test rise to over 85%. This observation provides clear evidence that the Portmanteau test is useful in detecting misspecification of the autoregressive orders; the larger the discrepancy, the higher is the power. In the case where there is misspecification in both the orders and the error distribution, as in Model 7, the power of both tests can be well above 90%. Overall, this experiment demonstrates that the Bayesian diagnostic test is a useful device for examining the suitability of the fitted threshold model. We can also develop the following decision rule according to the results of the two tests in the following table.

Table 3
Proportions, out of 1000 replications, for which the fitted model is found to be inadequate for Models 3 to 7

	Portmanteau		Studentized range	
	5%	10%	5%	10%
Model 3	4.7	10.0	5.2	10.6
Model 4	5.9	11.9	99.7	100.0
Model 5	50.6	61.8	5.3	11.4
Model 6	85.8	90.7	6.0	11.8
Model 7	94.3	96.5	99.6	99.8

The levels of significance are 5% and 10%.

Portmanteau	Studentized range	Implication
Significant	Not significant	Misspecification in the orders only
Not significant	Significant	Misspecification in the error distribution only
Significant	Significant	Misspecification in the orders and error distribution

7. Application to financial markets

7.1. Data and model description

The analysis undertaken in this article is based on daily closing prices of five Asian and two U.S. stock market indices: the Hang Seng Index of Hong Kong (HSI); the Straits Times Industrial Index of Singapore (STII); the Taiwan Stock Exchange Weighted Stock Index (TWSI); the Korea Composite Price Index (KCPI); the Nikkei 225 Index, Nasdaq Composite

Table 4
Summary statistics of the market index returns

	Mean	STD	Skewness	Kurtosis
HSI	0.048	1.781	−3.391	75.033
STII	0.023	1.447	−2.012	51.091
TWSI	0.044	1.995	−0.059	2.889
KCPI	0.038	1.799	−0.021	4.746
Nasdaq	0.046	1.388	−0.519	12.205
Nikkei 225	−0.004	1.371	−0.133	8.920
SP 500	0.043	1.048	−2.647	56.136

Index (Nasdaq); and Standard and Poor's 500 Index (SP 500). The data, obtained from Datastream International, run from January 4, 1985, to February 1, 2002, giving a total of 4455 return observations. All subsequent analysis is performed on the daily log returns, $y_t = (\log p_t - \log p_{t-1}) * 100$, where p_t is the price index at time t, with summary statistics in Table 4. All seven returns series exhibit the standard property of asset return data in that they have fat-tailed distributions, as indicated by excess kurtosis.

Two-regime models are adopted to capture mean and volatility asymmetry in the market returns. As it is commonly observed that a GARCH(1,1) model is sufficient to explain the conditional heteroscedasticity (see Bollerslev, Chou, & Kroner, 1992), we take $c_1 = c_2 = d_1 = d_2 = 1$. In addition, we assume $p_1 = p_2 = 1$ and $q_1 = q_2 = 2$. Allowing y_{t-1} in the conditional mean helps account for possible autocorrelations in the market returns. We also suspect that individual stock market movement can be affected by external events that are related to the economy worldwide. To reflect the relationships among different financial markets, it is reasonable to postulate that today's market return is also driven by past returns in other markets. With this in mind, for HSI, STII, TWSI, KCPI and Nasdaq, we set x_{1t} to be the SP 500 return at time t and x_{2t} to be the Nikkei 225 return at time t. The above choices of the exogenous variables are based on the global economic scale of the US and Japan and their strong influence on other countries' economic growth. For SP 500, we set x_{1t} to be the FTSE 100 return of the UK and x_{2t} to be the Nikkei 225 return at time t. Similarly, for Nikkei 225, we set x_{1t} to be the SP 500 return and x_{2t} to be the FTSE 100 return at time t.

In threshold modeling, it is crucial to select a suitable threshold variable to determine the regimes. Due to the temporal dependence among international financial markets, the use of $z_t = y_t$ as in Li and Li (1996) and Brooks (2001) can be improved upon. In our empirical study, we choose $m = 3$ and the auxiliary

Table 5
Posterior mean, 2.5th percentile (Lowlim) and 97.5th percentile (Uplim) of the unknown parameters for the HSI of Hong Kong, STII of Singapore and TWSI of Taiwan

	Hong Kong (HSI)			Singapore (STII)			Taiwan (TWSI)		
	Mean	Lowlim	Uplim	Mean	Lowlim	Uplim	Mean	Lowlim	Uplim
$\phi_0^{(1)}$	−0.473	−0.721	−0.235	−0.234	−0.591	0.245	−0.326	−0.856	0.310
$\phi_1^{(1)}$	−0.230	−0.333	−0.124	−0.001	−0.140	0.164	−0.109	−0.267	0.063
$\psi_1^{(1)}$	0.535	0.423	0.645	0.514	0.403	0.640	0.356	0.200	0.533
$\psi_2^{(1)}$	−0.080	−0.156	−0.003	−0.083	−0.173	0.006	0.061	−0.060	0.181
$\phi_0^{(2)}$	0.052	0.015	0.088	−0.005	−0.032	0.021	0.030	−0.014	0.077
$\phi_1^{(2)}$	0.085	0.047	0.122	0.191	0.156	0.225	0.078	0.040	0.118
$\psi_1^{(2)}$	0.414	0.368	0.461	0.323	0.286	0.357	0.164	0.112	0.213
$\psi_2^{(2)}$	−0.009	−0.039	0.022	0.009	−0.016	0.034	0.050	0.011	0.088
$\alpha_0^{(1)}$	0.343	0.200	0.519	0.459	0.282	0.676	0.526	0.216	0.858
$\alpha_1^{(1)}$	0.124	0.078	0.183	0.212	0.140	0.301	0.077	0.037	0.122
$\beta_1^{(1)}$	0.848	0.764	0.911	0.745	0.621	0.840	0.914	0.864	0.955
$\alpha_0^{(2)}$	0.083	0.051	0.124	0.103	0.069	0.151	0.040	0.016	0.068
$\alpha_1^{(2)}$	0.074	0.049	0.102	0.141	0.103	0.186	0.078	0.058	0.102
$\beta_1^{(2)}$	0.846	0.804	0.883	0.742	0.675	0.795	0.896	0.869	0.921
r_1	−0.757	−0.840	−0.691	−0.836	−0.902	−0.733	−1.263	−1.942	−0.828
v	5.241	4.523	6.076	4.975	4.328	5.754	5.364	4.565	6.363
w_1	0.660	0.595	0.734	0.551	0.474	0.685	0.695	0.487	0.921
w_2	0.295	0.239	0.355	0.283	0.231	0.462	0.104	0.003	0.290
w_3	0.045	0.003	0.121	0.166	0.002	0.253	0.201	0.006	0.333

variables $z_{1t}=y_t$, $z_{2t}=x_{1t}$ and $z_{3t}=x_{2t}$ because we believe that the factors appearing in the mean equation can also have strong impacts on the dynamic structure of y_t. For example, if y_t is the HSI return at time t and $d=1$, then $z_{1t-1}=y_{t-1}$, $z_{2t-1}=x_{1t-1}$ and $z_{3t-1}=x_{2t-1}$, representing the HSI, SP 500 and Nikkei 225 returns at time $t-1$, respectively, fix the time series properties of y_t. Similarly for y_t denoting the SP 500 return at time t and $d=1$, $z_{1t-1}=y_{t-1}$, $z_{2t-1}=x_{1t-1}$ and $z_{3t-1}=x_{2t-1}$, representing the SP 500, FTSE 100 and Nikkei 225 returns at time $t-1$, together define the regime of y_t. The combined effect of three past market changes, which is quantified by the weighted average of the three market returns, determines the switching between regimes. It is interesting to see how the weights w_1, w_2 and w_3, are distributed so that we understand more of how the market return dynamic reacts to past market movements.

To complete the model specification, we assume ϵ_t to be t-distributed rather than standard normal as in Li and Li (1996) and Brooks (2001). This distributional assumption matches the empirical evidence in the GARCH literature better and allows us to produce more reliable volatility and VaR forecasts. In summary, the working model considered in this section is

$$y_t=\begin{cases}\phi_0^{(1)}+\phi_1^{(1)}y_{t-1}+\psi_1^{(1)}x_{1t-1}+\psi_2^{(1)}x_{2t-1}+a_t, & z_{t-d}\le r_1,\\ \phi_0^{(2)}+\phi_1^{(2)}y_{t-1}+\psi_1^{(2)}x_{1t-1}+\psi_2^{(2)}x_{2t-1}+a_t, & z_{t-d}>r_1,\end{cases}$$

$$h_t=\begin{cases}\alpha_0^{(1)}+\alpha_1^{(1)}a_{t-1}^2+\beta_1^{(1)}h_{t-1}, & z_{t-d}\le r_1,\\ \alpha_0^{(2)}+\alpha_1^{(2)}a_{t-1}^2+\beta_1^{(2)}h_{t-1}, & z_{t-d}>r_1,\end{cases}$$

$$z_t=w_1z_{1t}+w_2z_{2t}+w_3z_{3t},\quad z_{1t}=y_t,\quad z_{2t}=x_{1t},\quad z_{3t}=y_{2t}, \tag{14}$$

where $\epsilon_t \sim t_v$, the standardized t-distribution with mean 0, variance 1 and degrees of freedom v. The probability density function of ϵ_t is given in Eq. (13).

7.2. Estimation results

To implement our MCMC sampling scheme, we choose the same prior as in the simulation experiment except that $h=10$, that is, we keep at least 10% of the observations in each regime. A total of 50,000 iterations are performed with the first 10,000 warm-up iterates discarded. Based on the posterior mode of d, a time delay of one day is selected for the indices. Tables 5–7 present the posterior mean, 2.5th and 97.5th percentiles of the other parameters. Concerning the autoregressive coefficients, $\phi_1^{(1)}$ and $\phi_1^{(2)}$ are of opposite signs except in Nikkei 225. This result is in agreement with Li and Lam (1995). With SP 500 being the only exception, all values of $\psi_1^{(1)}$ and $\psi_1^{(2)}$ are positive, indicating that the previous day's SP 500 return has a positive effect on the direction of other markets' movements. According to the posterior 95% confidence interval of $\psi_1^{(1)}$ and $\psi_1^{(2)}$, the impact of FTSE 100 on the mean level of SP 500 is negligible. Similarly, from the sizes of $\psi_2^{(1)}$ and $\psi_2^{(2)}$, we observe that the Nikkei 225 return has only a minor influence on the mean returns. Therefore, the common dominant characteristic of the US market is reflected in the mean equations.

Comparing the variance equations in the two regimes, we can see that $\alpha_0^{(1)}$ is substantially greater than $\alpha_0^{(2)}$. In addition, the persistence parameter in the lower regime is higher than that in the upper regime, i.e. $\alpha_1^{(1)}+\beta_1^{(1)}>\alpha_1^{(2)}+\beta_1^{(2)}$, in most cases. The only exception is KCPI, which has persistence parameters

Table 6
Posterior mean, 2.5th percentile (Lowlim) and 97.5th percentile (Uplim) of the unknown parameters for the KCPI of Korea and Nasdaq of the US

	Korea (KCPI)			US (Nasdaq)		
	Mean	Lowlim	Uplim	Mean	Lowlim	Uplim
$\phi_0^{(1)}$	−0.174	−0.617	0.307	−0.127	−0.385	0.021
$\phi_1^{(1)}$	−0.070	−0.224	0.092	−0.124	−0.257	−0.011
$\psi_1^{(1)}$	0.247	0.088	0.417	0.141	0.044	0.237
$\psi_2^{(1)}$	−0.058	−0.179	0.061	−0.056	−0.131	−0.004
$\phi_0^{(2)}$	−0.018	−0.052	0.017	0.097	0.058	0.172
$\phi_1^{(2)}$	0.093	0.058	0.131	0.117	0.037	0.179
$\psi_1^{(2)}$	0.093	0.053	0.135	0.070	0.014	0.122
$\psi_2^{(2)}$	0.045	0.011	0.078	−0.025	−0.052	0.002
$\alpha_0^{(1)}$	0.558	0.333	0.738	0.052	0.015	0.096
$\alpha_1^{(1)}$	0.071	0.040	0.124	0.116	0.081	0.159
$\beta_1^{(1)}$	0.896	0.790	0.952	0.879	0.835	0.915
$\alpha_0^{(2)}$	0.015	0.006	0.030	0.010	0.003	0.019
$\alpha_1^{(2)}$	0.081	0.062	0.102	0.098	0.070	0.130
$\beta_1^{(2)}$	0.905	0.884	0.923	0.879	0.846	0.911
r_1	−1.422	−1.654	−0.823	−0.331	−0.629	0.071
v	5.628	4.735	6.812	5.492	4.700	6.470
w_1	0.857	0.520	0.978	0.669	0.181	0.935
w_2	0.065	0.001	0.453	0.182	0.003	0.615
w_3	0.078	0.003	0.371	0.148	0.007	0.346

Table 7
Posterior mean, 2.5th percentile (Lowlim) and 97.5th percentile (Uplim) of the unknown parameters for Nikkei 225 of Japan and SP 500 of the US

	Japan (Nikkei 225)			US (SP 500)		
	Mean	Lowlim	Uplim	Mean	Lowlim	Uplim
$\phi_0^{(1)}$	−0.270	−0.457	−0.079	−0.144	−0.398	0.189
$\phi_1^{(1)}$	−0.213	−0.319	−0.107	−0.116	−0.250	0.004
$\psi_1^{(1)}$	0.345	0.260	0.430	−0.039	−0.130	0.055
$\psi_2^{(1)}$	0.116	0.020	0.209	−0.058	−0.120	0.000
$\phi_0^{(2)}$	0.026	−0.003	0.056	0.050	0.026	0.073
$\phi_1^{(2)}$	−0.027	−0.065	0.013	0.026	−0.011	0.063
$\psi_1^{(2)}$	0.221	0.184	0.257	0.016	−0.015	0.046
$\psi_2^{(2)}$	0.079	0.043	0.116	−0.008	−0.027	0.011
$\alpha_0^{(1)}$	0.265	0.191	0.351	0.143	0.071	0.212
$\alpha_1^{(1)}$	0.092	0.061	0.128	0.042	0.019	0.072
$\beta_1^{(1)}$	0.898	0.854	0.934	0.939	0.889	0.974
$\alpha_0^{(2)}$	0.020	0.013	0.028	0.010	0.004	0.017
$\alpha_1^{(2)}$	0.046	0.029	0.066	0.028	0.013	0.052
$\beta_1^{(2)}$	0.886	0.864	0.905	0.935	0.915	0.953
r_1	−0.629	−0.668	−0.595	−0.712	−0.855	−0.476
v	6.104	5.128	7.316	5.017	4.286	5.910
w_1	0.757	0.695	0.784	0.844	0.627	0.958
w_2	0.188	0.142	0.218	0.084	0.002	0.301
w_3	0.055	0.015	0.139	0.072	0.004	0.131

of similar magnitude. Based on the above results, higher average volatility is evident when bad news ($z_{t-d}<r_1$) arrives. As established from $\alpha_1^{(i)}+\beta_1^{(i)}$, $i=1,2$, it takes markets longer to digest bad news than good news. In other words, the effect of the current h_t on future volatility diminishes at a slower rate in the lower regime. It is interesting to note that in TWSI, Nasdaq and Nikkei 225, $\alpha_1^{(1)}+\beta_1^{(1)}>0.99$, signifying highly persistent volatility.

Regarding the distribution of ϵ_t, all degrees of freedom are less than 7. This suggests the existence of leptokurtosis and that it is more appropriate to use a fat-tailed error distribution than a normal. Also, it makes sense to have all threshold values negative. Most of the thresholds range from −0.9 to −0.6. While Nasdaq has the largest r_1, which is not significantly different from zero, TWSI and KCPI's values are comparatively small. In the upper regime, the model structure does not change until the threshold variable drops more than the borderline given by r_1. The negative threshold value accounts for the baseline market variation in that a small negative return in z_t is still regarded as normal. In contrast, abnormal situations that make z_t smaller than the negative cut-off r_1 confirm the 'bad news' regime.

The relative importance of z_{it} in governing the change in the regimes is studied through the weights. From the Bayesian estimation results, all w_1 are greater than 0.5, implying that the domestic return y_t is the most important determinant for the regimes. However, most of the weights deviate significantly from 1, demonstrating that fixing $z_t=y_t$ subjectively can lead to misleading results and inaccurate volatility forecasts. Comparing w_2 and w_3 in HSI, STII and Nasdaq, the SP 500 return is more important than the Nikkei 225 return in setting up the regimes whereas in TWSI, Nikkei 225 is more important. Similarly, SP 500 plays a more relevant role than FTSE 100 in defining the Nikkei 225 threshold variable. We also find that $w_1>0.8$ in both KCPI and SP 500 where the auxiliary variables z_{2t} and z_{3t} have only small effects on z_t. In these two indices, the threshold model with $z_t=y_t$ may provide a reasonable approximation to our general model.

7.3. Model diagnostics and selection results

To perform the diagnostic checking for the seven data series, we produce $\hat{u}_t$, $t=1389,\ldots,4388$ using Eq. (10) and compute $\hat{v}_t=\Phi^{-1}(\hat{u}_t)$. To apply the Studentized range test and the Portmanteau test, 3000 $\hat{v}_t$ were obtained. The 95% confidence interval (6.336, 8.111) of the Studentized range test is generated by simulations under the null hypothesis that $\hat{v}_t$ is i.i.d. $N(0,1)$. Table 8 displays the model checking results. We find no sign of inadequacy in HSI, STII, KCPI or Nikkei 225. Though the current models for TWSI, Nasdaq and SP 500 are not completely satisfactory, they are certainly useful in understanding the

Table 8
Studentized range test statistics and Portmanteau test statistics (with p-values in parentheses) obtained from modeling various market returns

	Studentized range test	Portmanteau test
HSI	7.100	13.91 (0.18)
STII	7.656	3.87 (0.95)
TWSI	6.836	34.37 (0.00)
KCPI	6.343	8.90 (0.54)
Nasdaq	5.935[a]	4.15 (0.94)
Nikkei225	6.526	11.94 (0.29)
SP500	6.195[a]	23.51 (0.01)

[a] Indicates that the Studentized range statistic lies outside the 95% confidence interval.

asymmetric behavior in these indices and their structural relationships with other indices.

To investigate how our proposed model performs relative to the threshold GARCH model in Li and Li (1996) and Brooks (2001), we perform a model comparison using the Bayesian method in Section 5.2. The model in Eq. (14) is named the full model F, and Li and Li's (1996) model is denoted L. T and W denote submodels of F with $w_1=1$ and $w_2=w_3=0$, and with normally distributed errors. The logarithm of the posterior odds ratio of F vs. T, F vs. W and F vs. L are given in Table 9. All $\mathrm{POR}_{\mathrm{FL}}$ have a magnitude of at least 10^{40}, implying that the superiority of F to L is obvious. Clearly, F is better than W for the same reason. As F and W differ only in the distribution of ϵ_t, very strong evidence is provided in support of the use of t-distributed innovations instead of normal innovations. Furthermore, we test whether the weighted average formulation $z_t = \sum_{i=1}^{3} w_i z_{it}$ is significantly better than $z_t=y_t$ provided that ϵ_t is t-distributed. The fact that all $\mathrm{POR}_{\mathrm{FT}}$ are greater than 30 except in KCPI signifies strong or even decisive evidence in favor of F over T. The exception of KCPI is reasonable as it has the largest value of w_1 among the seven indices. The change of regime in KCPI is mainly governed by its past returns. Overall, the value of auxiliary variables and t innovations is confirmed.

7.4. Forecast evaluation

Using the Bayesian prediction methods described in Section 4, we generate volatility and VaR forecasts for $t=n+1$, the first trading day of November 2001.

Table 9
Logarithm of the posterior odds ratio of F vs. T, F vs. W and F vs. L for the seven indices, where F stands for the full model with unknown w_i and t-distributed errors, T is the submodel of F with $w_1=1$ and $w_2=w_3=0$, W is the submodel of F with normally distributed errors and L stands for the model in Li and Li (1996)

	F vs. T	F vs. W	F vs. L
HSI	5.4	169.0	189.1
STII	14.4	226.9	335.4
TWSI	3.5	118.6	103.5
KCPI	−1.3	104.6	92.2
Nasdaq	4.3	150.5	156.2
Nikkei 225	16.3	114.3	141.1
SP 500	7.6	191.0	188.3

There are altogether 40 000 MCMC iterates of $\sqrt{h_{n+1}^{[i]}}$ and $y_{n+1}^{[i]}$ simulated from the predictive distributions $p(\sqrt{h_{n+1}}|\boldsymbol{y}^{1,n})$ and $p(h_{n+1}|\boldsymbol{y}^{1,n})$, respectively. Figs. 1 and 2 show sketches of the distributions constructed by standard kernel methods. In Fig. 1, we can observe that the predictive distributions of volatility are roughly symmetric, except in TWSI and KCPI, and little positive skewness is revealed. Point estimates of the volatility $\sqrt{h_{n+1}}$, obtained by the posterior mean of $p(\sqrt{h_{n+1}}|\boldsymbol{y}^{1,n})$, are 1.64, 1.03, 1.87, 1.59, 2.19, 1.49, and 1.35 for HSI, STII, TWSI, KCPI, Nasdaq, Nikkei 225 and SP 500. While we find Nasdaq the most volatile, STII has the smallest variation. Indeed, the order of the indices is STII < SP 500 < Nikkei 225 < KCPI < HSI < TWSI < Nasdaq with respect to the volatility $\sqrt{h_{n+1}}$. The predictive distributions of y_{n+1} in Fig. 2 demonstrate no evidence of asymmetry. From the simulated sample $y_{n+1}^{[i]}$, we forecast the first percentile of y_{n+1} to be −4.31, −2.82, −4.90, −4.08, −5.41, −3.78 and −3.47 for HSI, STII, TWSI, KCPI, Nasdaq, Nikkei 225 and SP 500. As these percentile estimates, or 1% VaR predicted values, are directly related to the variation in markets, it is not surprising to see that the order of the indices with respect to the magnitude of the percentiles is the same as that for volatility. In particular, the largest 1% VaR tells us that Nasdaq drops by more than 5.41% with a probability of 0.01.

To gauge the forecasting performance, we evaluate out-of-sample volatility forecasts in a testing period using a rolling sample approach. Parameter estimates are calculated in our learning period and a one-step-ahead forecast of h_t is produced. The next one-step-ahead forecast is then obtained by shifting the whole learning period forward by one day so that the number of observations for estimation is kept unchanged. The learning period starts with the first 4388 daily returns and our testing period contains the last 67 returns. The forecasted volatility at time t is denoted $\hat{h}_t$ using our heteroscedastic model. We calculate root mean square error (RMSE) and mean absolute error (MAE) for the forecast errors $(y_t^2-\hat{h}_t)$, $t=n+1,\ldots, n+67$. The two measures for evaluating the forecasting performance are presented in Table 10. The order of indices using the RMSE is SP 500 < HSI < Nikkei 225 < STII < Nasdaq < TWSI < KCPI, while that using the MAE is SP 500 < STII < HSI < Nasdaq < Nikkei 225 < TWSI < KCPI. In terms of the two measures,

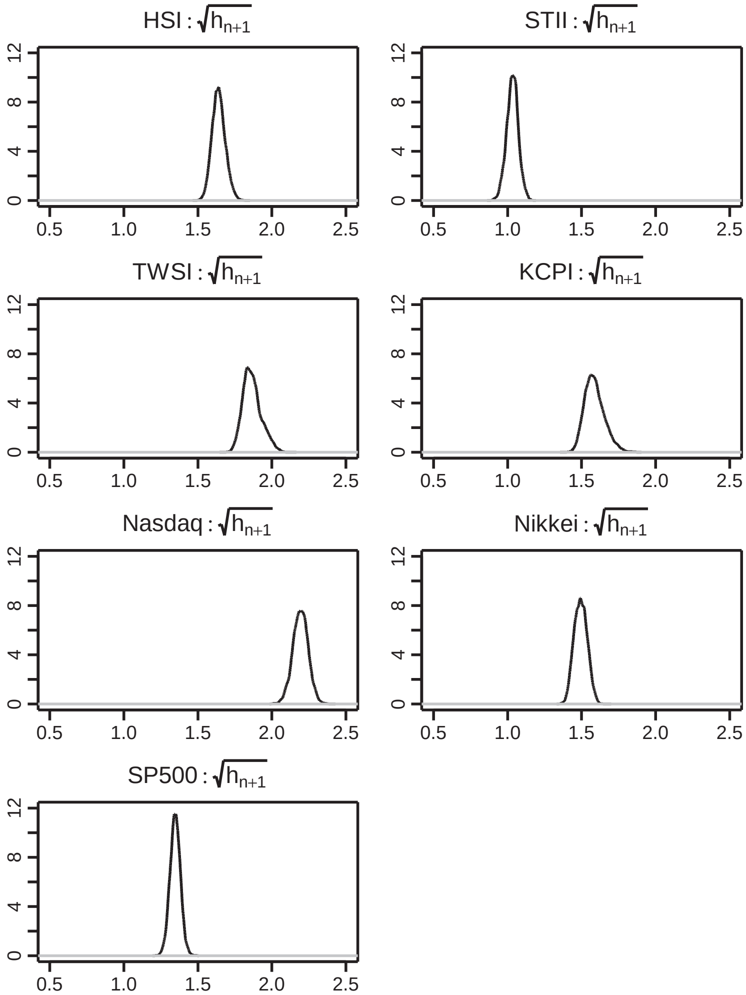

Fig. 1. The one-step-ahead predictive distribution of volatility, $p(\sqrt{h_{n+1}}|y^{1,n})$, for the seven market indices, where $t=n+1$ is the first trading day of November 2001.

the model forecasts the best in SP 500 and the worst in TWSI and KCPI. We observe in Table 4 that the standard deviations of TWSI and KCPI are the largest and the smallest is for SP 500. The relationship between forecast errors and variation in market returns is worth further exploration in future research.

8. Concluding remarks

A threshold heteroscedastic model is proposed to capture the mean and variance asymmetries in financial markets. The main feature of this model is that it allows the threshold variable to be formulated with auxiliary variables. This avoids subjectively choosing the threshold variable and enables the relative importance of the auxiliary variables to be examined after model fitting. Simultaneous estimation of time delay, threshold values and other parameters is feasible via our MCMC sampling methods. We also perform Bayesian diagnostic checking for the threshold heteroscedastic model. Simulations show that the Bayesian approach can provide accurate estimates for the unknown parameters and a reliable scrutiny of the fitted model.

Several observations from the real data application are worth while highlighting. First, there is

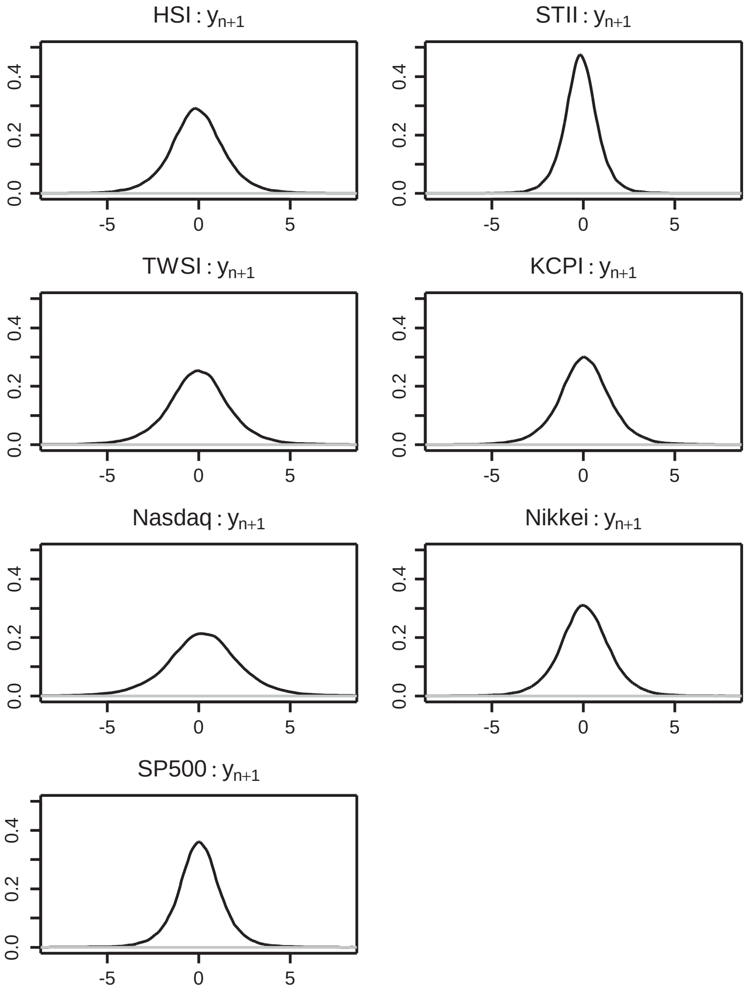

Fig. 2. The one-step-ahead predictive distribution of return, $p\left(y_{n+1}|y^{1,n}\right)$, for the seven market indices, where $t=n+1$ is the first trading day of November 2001.

strong evidence that the error distribution is leptokurtic. Modeling the threshold model with normal error as in Li and Li (1996) and Brooks (2001) can

Table 10
Forecast evaluation based on 67-day returns

	RMSE	MAE
HSI	2.841	2.032
STII	3.725	1.850
TWSI	6.444	4.392
KCPI	7.541	4.959
Nasdaq	4.058	2.669
Nikkei 225	3.667	2.754
SP 500	1.580	0.946

result in substantial bias in the VaR and volatility forecasts. Second, besides clear mean asymmetry, a dominant positive effect of the SP 500 return on the conditional mean is discovered. This supports the argument that international market index returns are driven by US market movements. In contrast, Nikkei 225 has minor influence on the mean equation. Third, higher average volatility and more persistent volatility are found in the lower regime. Therefore, the arrival of bad news in general makes the markets more volatile and bad news also has a more prolonged effect on return volatility than good news. Finally, we can see from the weights associated with the auxiliary variables that the domestic return is the major determinant of the regimes. However, both SP 500 and Nikkei 225 do impact the dynamic structure of the domestic market returns. Further research is necessary on how to select important auxiliary variables when forming the threshold variable.

Acknowledgements

The authors thank Professor Michael Clements, an associate editor, and two anonymous referees for their valuable suggestions and comments. C.W.S. Chen would like to thank Professor W.K. Li for valuable discussions during her visit to The University of Hong Kong. The research of C.W.S. Chen was supported by the National Science Council of Taiwan grant NSC93-2118-M-035-003. The research of Mike So was supported by the Hong Kong RGC Direct Allocation grants 01/02.BM38 and 02/03.BM29.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

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

Brooks, C. (2001). A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting*, *20*, 135–143.

Chen, C. W. S., & Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, *16*, 483–492.

Chen, C. W. S., & Wen, Y. W. (2001). On goodness of fit for time series regression models. *Journal of Statistical Computation and Simulation*, *6*, 239–256.

D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-fit techniques*. New York: Marcel Dekker.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, *50*, 987–1008.

Gelman, A., Roberts, G. O., & Gilks, W. R. (1996). Efficient Metropolis jumping rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 5* (pp. 599–607). Oxford: Oxford University Press.

Gerlach, R., Carter, C. K., & Kohn, R. (1999). Diagnostics for time series analysis. *Journal of Time Series Analysis*, *20*, 309–330.

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

Li, C. W., & Li, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics*, *11*, 253–274.

Li, W. K., & Lam, K. (1995). Modelling asymmetry in stock returns by a threshold ARCH model. *The Statistician*, *44*, 333–341.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability*, *36*, 688–705.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, *23*, 470–472.

Smith, J. Q. (1985). Diagnostic checks of non-standard time series. *Journal of Forecasting*, *4*, 283–291.

Tong, H. (1978). On a threshold model. In C. H. Chen (Ed.), *Pattern recognition and signal processing*. Amsterdam: Sijhoff & Noordhoff.

Tong, H. (1983). *Threshold models in non-linear time series analysis* (vol. 21 of Lecture Notes in Statistics). New York: Springer-Verlag.

Tong, H. (1990). *Nonlinear time series: A dynamical system approach*. Oxford: Oxford University Press.

Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *Journal of the Royal Statistical Society. Series B*, *42*, 245–292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association*, *84*, 231–240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, *93*, 1188–1202.

Wong, C. M., & So, M. K. P. (2003). On conditional moments of GARCH models, with applications to multiple period value at risk estimation. *Statistica Sinica*, *13*, 1015–1044.

Dr. Cathy W. S. Chen is professor of Statistics at Feng Chia University, Taiwan. Her research interests include financial time series modeling, econometric methods, MCMC methods, and statistical approach to epidemiology. She has published papers in the *Journal of Time Series Analysis, Journal of Forecasting*, *Statistica Sinica*, *Statistics in Medicine*, and *Aids*, etc. Her publication lists can be found at http://www.stat.fcu.edu.tw/faculty/Chen.html/.

Mike K. P. So is an associate professor of the Department of Information and Systems Management at The Hong Kong University of Science and Technology. His research interests include nonlinear time series analysis, financial time series modeling, market volatility studies, modeling the dynamic structure of economic data and value at risk estimation in financial markets.