Physica A 462 (2016) 56–66

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Binomial Markov-Switching Multifractal model with Skewed t innovations and applications to Chinese SSEC Index

Yufang Liu [a], Weiguo Zhang [a,*], Junhui Fu [b,c]

[a] School of business administration, South China University of Technology, Guangzhou, 510641, China
[b] School of Finance, Zhejiang University of Finance and Economics, Hangzhou, 310018, China
[c] China Academy of Financial Research, Zhejiang University of Finance and Economics, Hangzhou, 310018, China

HIGHLIGHTS

- A sliding window multifractal detrending moving average (MF-DMA) method is proposed.
- A Log-normal multifractal model is presented and the parameter is estimated.
- Both sliding window MF-DMA method and Log-normal multifractal model perform well.

ARTICLE INFO

Article history:
Received 30 October 2015
Received in revised form 27 March 2016
Available online 16 June 2016

Keywords:
BMSM models
Volatility forecasting
Skewed t innovations

ABSTRACT

This paper presents the Binomial Markov-switching Multifractal (BMSM) model of asset returns with Skewed t innovations (BMSM-Skewed t for short), which considers the fat tails, skewness and multifractality in asset returns simultaneously. The parameters of BMSM-Skewed t model can be estimated by Maximum Likelihood (ML) methods, and volatility forecasting can be accomplished via Bayesian updating. In order to evaluate the performance of BMSM-Skewed t model, BMSM model with Normal innovations (BMSM-N), BMSM model with Student-t innovations (BMSM-t) and GARCH(1,1) models (GARCH-N, GARCH-t and GARCH-Skewed t) are chosen for comparison. Through empirical studies on Shanghai Stock Exchange Composite Index (SSEC), we find that for sample estimation, BMSM models outperform the GARCH(1,1) models through BIC and AIC rules, and BMSM-Skewed t performs the best among all the models due to its fat tails, skewness and multifractality. In addition, BMSM-Skewed t model dominates other models at most forecasting horizons for out-of-sample volatility forecasts in terms of MSE, MAE and SPA test.

1. Introduction

Financial volatility modeling is one of the core contents in modern financial theory. Accurate modeling of volatility dynamics plays major roles in portfolio allocation, derivative pricing and risk management [1,2].

Among the existed literature on volatility modeling and forecasting, the most common used models are the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) class [3], which filter out the heteroscedasticity in financial returns. The conditional volatility in GARCH models is driven by a smooth autoregressive transition. Due to the reason that it has difficulty to capture the sudden changes in returns, Andersen [4,5] proposed the stochastic volatility (SV) model, where the

* Corresponding author.
E-mail addresses: liufang_0205@126.com (Y. Liu), wgzhang@scut.edu.cn (W. Zhang), fujh02@gmail.com (J. Fu).

http://dx.doi.org/10.1016/j.physa.2016.06.014

volatility is hit by independent shocks. However, the volatility state is not observable in SV model, parameter estimations of this model could not be realized by the maximum likelihood estimation. Although GARCH models perform well in forecasting short-term volatility, they have difficulties to capture the long-term volatility dynamics accurately. For this reason, Baillie et al. [6] extended the GARCH models to fractionally integrated GARCH (FIGARCH) by adding one parameter to characterize the long memory existed in financial returns. Also, Breidt et al. [7] gave the long-memory SV model. Besides, there has been another type of volatility models by applying regime-switching approach which was first introduced to finance by Hamilton [8–10]. The volatility dynamics were defined by the conditional variance of a Markov state variable. The disadvantage of these models is that the Markov states will increase exponentially as the number of components of Markov variable increases, which may cause the calculation great complexity.

Yet in this paper, we consider the volatility model named the Markov-Switching Multifractal (MSM) Model, which was first introduced by Calvet and Fisher [11]. Researches on this model have shown that it can forecast future volatility more accurately than traditional models, such as GARCH and Markov-Switching GARCH models [11–13]. The essential feature of MSM model is its hierarchical, multiplicative structure to characterize multifractality, which is a stylized fact in empirical finance [14–20]. Meanwhile, this model captures long memory and fat tails exhibited in returns. As a Markov-switching process, MSM model is a highly tractable volatility model that both parameter estimation and volatility forecasting can be realized by standard econometric methods.

There have been a few ongoing research since the MSM model appeared. For example, Calvet and Fisher [21] extended the univariate MSM models to multivariate MSM and developed a particle filter that permitted inference and forecasting using simulations for a large number of components of Markov state vector. As maximum likelihood (ML) estimation can only be used for discrete distributions with limited number of volatility components in MSM models, Lux [13] proposed an alternative GMM (Generalized method of moments) estimator together with linear forecasts which is applicable for any distribution with any number of volatility components. In consideration that MSM models with Student-t innovations might improve models with normal innovations, Lux and Morales–Arias [22] extended the MSM models of asset returns with normal innovations to Student-t innovations (MSM-t). Both ML and GMM methods were used for parameters estimation, and volatility forecasting was performed via Bayesian updating and best linear forecasts. They also observed that forecast combinations obtained by MSM and (FI)GARCH models could provide an improvement from single models.

Although research on MSM models increased the attention of financial economists, almost all the empirical studies existed were aimed at the financial markets of developed countries, there have been few investigations on developing countries including Chinese financial markets. Besides, most of the literature assumed that the innovations of returns were driven from normal distribution. Considering the fact that empirical financial returns exhibit non-Gaussian, the performance of volatility forecasting may be improved by adopting innovations' distribution with stylized features of returns, such as fat tails and skewness [23,24].

This paper attempts to extend the MSM model with Skewed t innovations. Refer to Ref. [22], as the distribution of volatility components of MSM model is specified to Binomial distribution, such model is called Binomial MSM-Skewed t model (BMSM-Skewed t for short). The parameters of this model can be estimated by ML method and it is convenient to use the Bayesian updating to forecast the future volatility. In order to examine the superior performance of BMSM-Skewed t model, we consider the BMSM model with Normal and Student-t innovations, and GARCH(1,1) models along with Normal, Student-t and Skewed t innovations. Empirical analysis on Shanghai Stock Exchange Composite Index (SSEC) in China shows obvious improvement by adding fat tails and skewness in BMSM model. Since BMSM-Skewed t model performs the best for both parameter estimation and out-of-sample volatility forecasting among all the models considered, it is thought to be the most appropriate model for SSEC index.

The rest of this paper is organized as follows. Section 2 introduces the BMSM-Skewed t model and corresponding methods of parameter estimation and volatility forecasting. Also, we briefly reviews the benchmark model, GARCH(1,1) model. Section 3 presents the empirical findings of SSEC via different volatility models. Finally, Section 4 summarizes the conclusion.

2. Volatility models

Let $r_t,\ t = 1, 2, \ldots, T$ denote the log-return series of a financial asset. This paper considers the following specification of returns:

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t u_t, \tag{1}$$

where μ_t is the conditional mean of the return series, and is assumed to follow a first order autoregressive model. σ_t is the volatility process, and u_t is the independently and identically distributed (i.i.d.) innovations that can be drawn from various stationary distributions.

2.1. BMSM-Skewed t model

In this section, we introduce the construction of Binomial Markov-switching Multifractal volatility model with Skewed t innovations. In BMSM model, the stochastic volatility σ_t is determined by a first-order Markov state vector M_t with $\bar{k}$

volatility components $M_{1,t}, M_{2,t}, \ldots, M_{\bar{k},t}$ and a scale factor σ:

$$\sigma_t = \sigma \left(\prod_{k=1}^{\bar{k}} M_{k,t} \right)^{1/2}. \tag{2}$$

Assume that the volatility components have the same marginal distribution. For each $k \in \{1, \ldots, \bar{k}\}$, $M_{k,t}$ is drawn from a fixed distribution M with probability γ_k, and remains unchanged with probability $1 - \gamma_k$. That is,

$$\begin{array}{ll} M_{k,t} \text{ drawn from distribution } M & \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \text{with probability } 1 - \gamma_k. \end{array}$$

The transition probabilities $\gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_{\bar{k}})$ are specified as

$$\gamma_k = 1 - (1 - \gamma_1)^{b^{k-1}}, \tag{3}$$

where $\gamma_1 \in (0, 1)$ and $b \in (1, \infty)$. The same way with Ref. [11], we choose $(\gamma_{\bar{k}}, b)$ as the parameter vector to be estimated for transition probabilities.

The distribution M satisfies that $M \geq 0$ and $\mathbb{E}(M) = 1$. Here the distribution of M is specified as the Binomial distribution. That is, M takes only two values, m_0 or m_1. Then the Markov state vector M_t can take $d = 2^{\bar{k}}$ states, and denote each state as $m^i,\ i = 1, 2, \ldots, d$. For simplicity, we assume these two outcomes occur with the same probability and have $m_1 = 2 - m_0$ due to $\mathbb{E}M = 1$.

Based on the fact that empirical finance always exhibits skewness besides fat tails, it is natural and reasonable to consider the BMSM volatility model with Skewed t innovations. The skewed t distribution was first proposed by Hansen [25]. He added another parameter to describe the skewness of the distribution based on Student-t distribution. The probability density function of Skewed t distribution is as follows:

$$f(x; \nu, \mu) = \begin{cases} bc \left(1 + \dfrac{1}{\nu - 2} \left(\dfrac{bx + a}{1 - \mu} \right)^2 \right)^{-(\nu+1)/2}, & x < -a/b, \\ bc \left(1 + \dfrac{1}{\nu - 2} \left(\dfrac{bx + a}{1 + \mu} \right)^2 \right)^{-(\nu+1)/2}, & x \geq -a/b, \end{cases} \tag{4}$$

where $2 < \nu < \infty$, and $-1 < \mu < 1$. The constants a, b and c are given by

$$a = 4\mu \left(\frac{\nu - 2}{\nu - 1} \right),$$
$$b^2 = 1 + 3\mu^2 - a^2,$$

and

$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu - 2)}\Gamma\left(\frac{\nu}{2}\right)}.$$

The freedom degree ν controls the tails of distribution function. The distribution function has thinner tails than Normal distribution with $2 < \nu < \infty$. The lower ν, the fatter the tails. μ is the skewness parameter, the function is right skewed with $\mu > 0$ and left skewed with $\mu < 0$. When $\nu \to \infty$ and $\mu = 0$, the density function degenerates to Normal density function.

The BMSM parameter vector can be estimated by the Maximum Likelihood (ML) method, and the log-likelihood function is

$$\ln L(r_1, \ldots, r_T; \psi) = \sum_{t=1}^{T} \ln[\omega(r_t) \cdot (\Pi_{t-1} A)], \tag{5}$$

where $\omega(r_t)$ is the conditional density function of x_t, $\Pi_t = \left(\Pi_t^1, \ldots, \Pi_t^d\right)$ is the conditional probability vector with $\Pi_t^i \equiv \mathbb{P}\left(M_t = m^i | r_1, \ldots, r_t\right)$, and $A = (a_{i,j})_{1 \leq i,j \leq d}$ is the transition matrix with components $a_{i,j} = \mathbb{P}\left(M_{t+1} = m^j | M_t = m^i\right)$. The parameter vector of BMSM-Skewed t model is given by $\psi \equiv (b, m_0, \gamma_k, \mu, \nu, \sigma)$.

Based on the Bayesian method, we can forecast the future volatility with the estimated parameters. Given the conditional probability vector Π_t, the probability distribution of future volatility state is

$$\mathbb{P}\left(M_{t+l} = m^i | r_1, \ldots, r_t\right) = \left(\Pi_t A^l\right)_i, \quad 1 \leq i \leq d, \tag{6}$$

where l is the forecasting horizon. Then the future volatility is

$$\mathbb{E}\left(\sigma_{t+l}^2\right) = \sum_{i=1}^{d} \sigma_t^2\left(m^i\right) * \mathbb{P}\left(M_{t+l} = m^i | r_1, \ldots, r_t\right) = \sum_{i=1}^{d} \sigma_t^2\left(m^i\right) * \left(\Pi_t A^l\right)_i, \tag{7}$$

where $\sigma_t^2\left(m^i\right) = \sigma \prod_{k=1}^{\bar{k}} M_{k,t}^i$ with $M_{k,t}^i = m_0$ or m_1.

For comparisons of the performance of model estimation and volatility forecasting, we also study the BMSM models with standard Normal and Student-t innovations. The BMSM model with standard Normal innovations is just the model considered by Calvet and Fisher [11], and the BMSM model with Student-t innovations is the same with Lux and Leonardo [22].

Normal distribution. It is the most common distribution used in modeling financial asset returns. The density function of standard Normal distribution is:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{8}$$

Student-t distribution. Most empirical studies show that financial assets have fat tails. Student-t distribution has fatter tails than Normal distribution, so Student-t distribution is usually considered as an improvement for better fitting real financial data than normal distribution. The probability density function of Student-t distribution is:

$$f(x;\nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \quad x \in \mathbb{R}, \tag{9}$$

where $\Gamma(\cdot)$ is the Gamma function, and ν is the degree of freedom.

2.2. GARCH(1,1) model

Nowadays, the most common used approach to modeling volatility is the generalized autoregressive conditional heteroskedasticity (GARCH) model proposed by Bollerslev [3]. The standard GARCH(1,1) assumes that the conditional variance σ_t satisfies the following autoregressive process:

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{10}$$

where the parameters are restricted as $\omega > 0, \alpha, \beta \geq 0$ and $\alpha + \beta < 1$. Then the unconditional variance of r_t is given by $\sigma^2 = \omega/(1-\alpha-\beta)$.

GARCH(1,1) model can be estimated by ML method. The l-step ahead forecast is given by

$$\hat{\sigma}_{t+l} = \frac{\omega\left[1-(\alpha+\beta)^{l-1}\right]}{1-\alpha-\beta} + (\alpha+\beta)^{l-1}\hat{\sigma}_{t+1}. \tag{11}$$

3. Empirical analysis

3.1. Data

We choose the daily closing prices of SSEC from July 15, 1991 to June 8, 2015 (up to 5841 trading days) for our empirical study. Let $P_t, t = 0, 1, \ldots, T$ denote the closing price of stock index at time t, we analyze the daily log-returns using the equation,

$$r_t = 100 * (\ln P_t - \ln P_{t-1}), \quad t = 1, 2, \ldots, T. \tag{12}$$

The SSEC log-return series is illustrated in Fig. 1. We can see that the log-return series is highly volatile and shows apparent volatility clustering. China started to carry out the price limits policy on December 16, 1996, and stipulated the range of price change in any direction (amount of increase and amount of decrease) is limited to 10%. The SSEC log-return series fluctuated violently before the policy was implemented, even achieved 100 ∗ 71.92% on May 21, 1992. The skewness and kurtosis have great dependence on these extremes. In order to improve the out-of-sample forecasting ability where the forecast interval is after the implementation of the price limits policy, and reserve the historical information provided before the policy, we eliminate the outliers of log-return series which exceed 10.

Table 1 provides the descriptive statistics of SSEC log-return series after elimination. The skewnesses of log-return series is bigger than 0, and the kurtosises are bigger than 3 obviously, which state that the log-return series does not satisfy the law of normal distribution, and has leptokurtosis properties. This is also verified by the Jarque–Bera statistics. Thereby, it is not reasonable to use Efficient Market theory to characterize the SSEC log-returns. In addition, the statistics of the Augmented

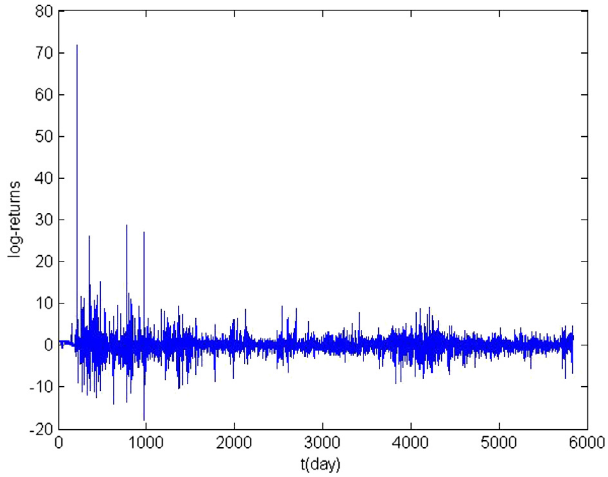


Fig. 1. Log-returns of SSEC.

Table 1
Descriptive statistics of SSEC log-returns.

Mean	Std. dev	Skewness	Kurtosis	JB	ADF	Size of data
0.026	1.926	−0.129	7.336	4566.044*	−11.977	5807

Notes: The Jarque–Bera statistic tests for the null hypothesis of normality in the sample. ADF is the statistic of the Augmented Dickey–Fuller unit root test based on the least AIC criterion.
* Refers to statistical significance at the 1% level.

Dickey–Fuller (ADF) unit root tests reject the null hypothesis of a unit root at 1% significance level indicating that the log-return series is stationary. The statistics results show the need for a model that can incorporate long memory, fat tails and skewness.

To detect the multifractality of SSEC log-return series, we use the MF-DMA (multifractal detrending moving average) algorithm proposed by Gu and Zhou [26]. For two time series, one can refer to Refs. [27,28]. Consider the SSEC log-return series $r(t),\ t = 1, 2, \ldots, T$, where T is the length of the time series.

Step 1. Construct the new sequence by the cumulative sums of $r(t)$:

$$y(t) = \sum_{i=1}^{t} r(i), \quad t = 1, 2, \ldots, T. \tag{13}$$

Step 2. Calculate the moving average function in a moving window,

$$\tilde{y} = \frac{1}{s} \sum_{k=-\lfloor (s-1)\theta \rfloor}^{\lceil (s-1)(1-\theta) \rceil} y(t-k), \tag{14}$$

where s is the window size, $\lfloor m \rfloor$ is the largest integer not larger than m, $\lceil m \rceil$ is the smallest integer not smaller than m, and θ is the position parameter varying from 0 to 1.

Step 3. Detrend the series by subtracting the moving average functions, and obtain the residual sequence $\varepsilon(i)$ through

$$\varepsilon(i) = y(i) - \tilde{y}(i), \tag{15}$$

where $s - \lfloor (s-1)\theta \rfloor \le i \le N - \lfloor (s-1)\theta \rfloor$.

Step 4. Divide the residual sequences into N_s non-overlapped segments with the equal length s, where $N_s = \lfloor N/s - 1 \rfloor$. Each segment can be written as $\varepsilon_v(i) = \varepsilon(l + i)$ for $1 \le i \le s$, respectively, where $l = (v-1)s$. Then determine the fluctuation variance

$$F_v^2(s) = \frac{1}{s} \sum_{i=1}^{s} \varepsilon_v^2(i). \tag{16}$$

Step 5. Average over all segments to obtain the qth order fluctuation function

$$F_q(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [F_v^2(s)]^{q/2} \right\}^{1/q}, \tag{17}$$

Table 2
ML estimation results of Binomial MSM-Skewed t model.

$\bar{k}$	1	2	3	4	5	6	7
ML estimation of BMSM-Skewed t							
b	–	15.5892 (7.3280)	5.0755 (1.1906)	3.3107 (0.7815)	2.6274 (0.5192)	2.3148 (0.4905)	2.0485 (0.3840)
$m0$	1.7262 (0.0125)	1.5866 (0.0137)	1.5102 (0.0150)	1.4815 (0.0175)	1.4145 (0.0196)	1.4028 (0.0205)	1.3559 (0.0217)
$\gamma_{\bar{k}}$	0.0238 (0.0041)	0.0598 (0.0114)	0.063 (0.0118)	0.0597 (0.0164)	0.0678 (0.0195)	0.063 (0.0199)	0.0946 (0.0360)
ν	5.336 (0.4316)	7.0897 (0.7911)	7.7841 (0.8667)	7.9462 (1.0178)	7.5555 (0.9803)	7.4645 (0.9165)	7.8421 (0.9995)
μ	−0.0551 (0.0164)	−0.054 (0.0182)	−0.0573 (0.0187)	−0.0567 (0.0188)	−0.0564 (0.0172)	−0.057 (0.0174)	−0.0568 (0.0192)
σ	2.1864 (0.07889)	2.1679 (1.1943)	2.2363 (1.2740)	1.9078 (1.04886)	2.1527 (2.2454)	1.8587 (1.5520)	1.9082 (3.5725)
$\ln L$	−10 901.83	−10 832.03	−10 799.76	−10 790.48	−10 790.90	−10 789.25	−10 791.67

where q takes any real value except zero. When $q = 0$, we have

$$\ln[F_0(s)] = \frac{1}{N_s}\sum_{v=1}^{N_s}\ln[F_v(s)] \tag{18}$$

according to l'Hôspital's rule. Repeat steps 2–5 for each time scale s.

Step 6. Determine the scaling behavior of the fluctuation functions by analyzing log — log plots $F_q(s)$ versus s for each value of q

$$F_q(s) \propto s^{h(q)}. \tag{19}$$

When $h(q)$ is a constant, the series is monofractal. When $h(q)$ is dependent on q, the series is multifractal. $h(q)$ is defined to be generalized Hurst exponent.

In the standard multifractal formalism based on partition function, the multifractal nature is characterized by the scaling exponent $\tau(q)$, the relation between $h(q)$ and $\tau(q)$ is:

$$\tau(q) = qh(q) - 1. \tag{20}$$

Multiscale spectrum function $f(\alpha)$, which is used to characterize multiscale time series, can be obtained through the Legendre transformation:

$$\begin{aligned} \alpha &= h(q) + qh'(q), \\ f(\alpha) &= q[\alpha - h(q)] + 1, \end{aligned} \tag{21}$$

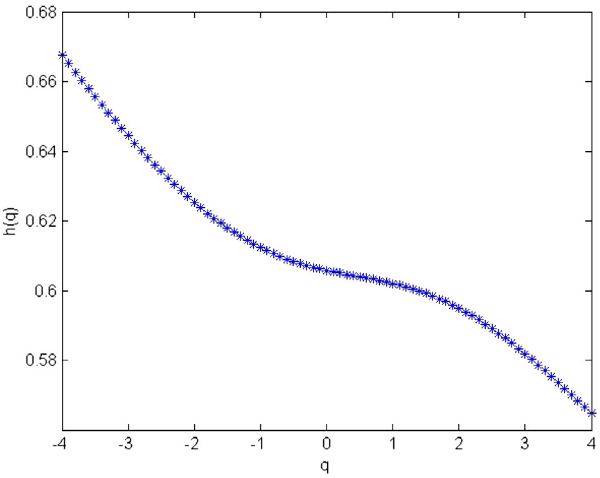
where α is the singular index that reflects the singularity degree during different intervals in a complex system. $f(\alpha)$ is called the multifractal spectrum, whose value reflects the fractal dimension with singular index α.

Here we specify $s \in [10, T/10]$ and use the backward moving average with $\theta = 0$. The generalized Hurst exponent $h(q)$ and scaling exponent $\tau(q)$ are obtained from $q \in [-4, 4]$ with 0.1 step. Fig. 2 shows the multifractal results. The generalized Hurst exponent $h(q)$ decreases with q, the scaling function $\tau(q)$ is clearly nonlinear, and the width of the multifractal spectrum is not zero. For monofractal series, the multifractal spectrum degenerates into one point. All the results show clear evidence of the existence of multifractality in out data, which indicates that multifractal model better matches our data rather than monofractal models.

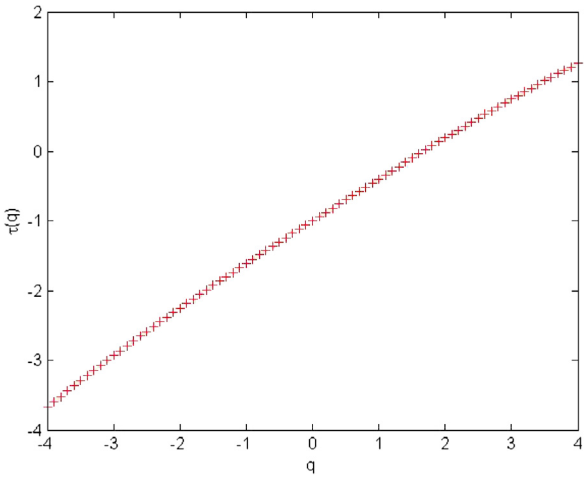
3.2. Parameter estimation

We use the ML method for parameter estimation. The number of volatility components $\bar{k}$ varies from 1 to 7. The parameter estimation results of the BMSM-Skewed t model are showed in Table 2. The multiplier m_0 is bigger than 1, which shows clear evidence of variation in volatility. The bigger m_0 is, the more outliers there are in log-return series. m_0 tends to decline with $\bar{k}$ increasing because more volatility components need less variability in M_t to match the fluctuations in log-return series. The transaction probability $\gamma_{\bar{k}}$ tends to increase as $\bar{k}$ increases, The estimates of freedom parameters ν in BMSM-Skewed t are between 5 and 8, which state that the log-return series has fatter tails than Normal distribution. The estimates of skewness parameter μ in BMSM-Skewed t are less than 0 and show left skewness, which is consistent with the real data. The value of log-likelihood function increases and reaches a maximum at $\bar{k} = 6$.

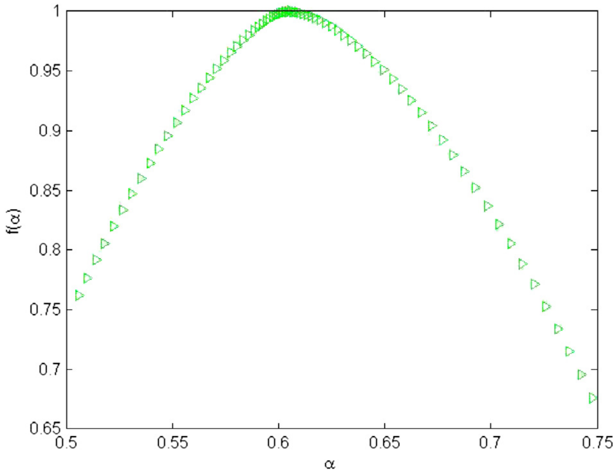
We also provide the ML estimates of BMSM-N and BMSM-t models in Table 3. The parameter estimates have similar properties to those of BMSM-Skewed t model. By comparison, we find that the estimates of m_0 and $\gamma_{\bar{k}}$ in BMSM-t and



(a) Generalized Hurst exponent.



(b) Scaling exponent.



(c) Multifractal spectrum.

Fig. 2. Multifractal results of SSEC log-returns.

BMSM-Skewed t are smaller than those of BMSM-N model. There exists a trade-off between the fat tails and long memory. As Skewed t distribution can model both the fat tails and skewness of log-return series, the BMSM-Skewed t model produces the largest values of log-likelihood function for fixed $\bar{k}$.

Results of GARCH(1,1) models are reported in Table 4. GARCH-Skewed t model, which incorporates fat tails and skewness, improves the results of model estimation in terms of $\ln L$.

To measure the performance of all models considered in this paper, BIC and AIC values are reported in Table 5. We find that for BMSM models, adding fat tails with the assumption of Student-t innovations, improves the original model with Normal innovations. BMSM-Skewed t model performs better than BMSM-t model due to the skewness besides fat tails. Results of GARCH(1,1) models are quite the same. On the other hand, all the BMSM models outperform the GARCH(1,1) models. The BMSM-Skewed t model has the smallest BIC and AIC values. The BIC and AIC values of BMSM-N model are slightly larger than other BMSM models, but still smaller than those of GARCH(1,1) models. This may be explained by that BMSM models can capture the properties of long memory and multifractal exhibited in real data, while GARCH(1,1) models are incapacity of this.

3.3. Out-of-sample volatility forecasting

This section studies the out-of-sample volatility forecasting performance of all the models considered by dynamic prediction over forecasting horizons $l = 1, 5, 10, 20, 50$ and 100 days. We divide the overall sample ($t = 1, 2, \ldots, T$) into two parts. One part containing $H = 5000$ observations is used for parameter estimation by ML method with $\bar{k} = 6$. The other part ($m = H + l, H + l + 1, \ldots, H + M, M = 807 - l$) is for out-of-sample volatility forecasting. First, choose

Table 3
ML estimation results of other Binomial MSM models.

$\bar{k}$	1	2	3	4	5	6	7
ML estimation of BMSM-*N*							
b	–	15.1376 (4.3650)	7.4849 (2.0324)	9.9819 (2.4436)	6.0585 (1.3820)	6.2109 (0.9696)	4.2382 (0.5227)
$m0$	1.7639 (0.0086)	1.6370 (0.0126)	1.5440 (0.0133)	1.5029 (0.0154)	1.4620 (0.0150)	1.4496 (0.0124)	1.3992 (0.0131)
$\gamma_{\bar{k}}$	0.0495 (0.0061)	0.1609 (0.0310)	0.2128 (0.0766)	0.6636 (0.1384)	0.6976 (0.1759)	0.8731 (0.0882)	0.9185 (0.0785)
σ	2.2224 (0.0391)	2.0862 (0.0502)	2.1516 (0.100)	2.1857 (0.0950)	1.9714 (0.0950)	1.8173 (0.0722)	2.1203 (0.0995)
$\ln L$	−110 989.57	−10 871.62	−10 832.47	−10 817.15	−10 810.11	−10 805.71	−10 804.69
ML estimation of BMSM-*t*							
b	–	15.4418 (6.8277)	5.1882 (1.4440)	3.4414 (0.7978)	2.6473 (0.5092)	2.3571 (0.4524)	2.0755 (0.4014)
$m0$	1.7274 (0.0125)	1.5870 (0.0141)	1.5107 (0.0142)	1.4857 (0.0175)	1.4170 (0.0200)	1.4046 (0.0177)	1.3580 (0.0204)
$\gamma_{\bar{k}}$	0.0238 (0.0046)	0.0592 (0.0112)	0.0644 (0.0131)	0.0608 (0.0150)	0.0693 (0.0193)	0.0656 (0.0197)	0.0967 (0.0404)
ν	5.4288 (0.4495)	7.2059 (0.9245)	8.0113 (0.9112)	8.1446 (1.0657)	7.7856 (0.9495)	7.7008 (0.9405)	8.0726 (1.1276)
σ	2.1880 (0.8251)	2.1495 (1.2391)	2.2388 (1.2089)	1.9006 (1.1114)	2.1282 (2.1441)	1.8381 (1.4374)	1.9081 (3.7212)
$\ln L$	−10 907.46	−10 836.42	−10 804.55	−10 795.05	−10 795.47	−10 793.91	−10 796.2

Table 4
ML results of GARCH(1,1).

	ω	α	β	ν	μ	$\ln L$
GARCH-*N*	0.0392 (0.0059)	0.0912 (0.0081)	0.9015 (0.0080)	–	–	−11 059.4
GARCH-*t*	0.0377 (0.0090)	0.1057 (0.0121)	0.8929 (0.0113)	4.8279 (0.3207)	–	−10 827
GARCH-Skewed *t*	0.0398 (0.0091)	0.1092 (0.0122)	0.8895 (0.1125)	4.7937 (0.3166)	−0.0581 (0.0156)	−10 820

Table 5
BIC and AIC values.

	GARCH-*N*	GARCH-*t*	GARCH-Skewed *t*	BMSM-*N*	BMSM-*t*	BMSM-Skewed *t*
BIC	3.8135	3.7349	3.7340	3.7272	3.7250	3.7249
AIC	3.8100	3.7303	3.7283	3.7226	3.7200	3.7188

$\varepsilon_t,\ t = 1, 2, \ldots, H$ for in-sample estimation, and obtain the l-step ahead forecast $\hat{\sigma}^2_{H+l}$. Then we use $\varepsilon_t,\ t = 2, 3, \ldots, H+1$ for parameter estimation and get the l-step volatility forecast $\hat{\sigma}^2_{H+l+1}$. Repeat this procedure, we obtain all the out-of-sample volatility forecasts $\hat{\sigma}^2_{H+l}, \hat{\sigma}^2_{H+l+1}, \ldots, \hat{\sigma}^2_{H+M}$.

It is worth noting that there is no unified conclusion in which loss function is the most reasonable benchmark as the deviation measure. This paper considers two loss functions to compare the forecasts performance across models in empirical study: MSE (mean square error) and MAE (mean absolute error) their definitions are as follows:

$$MSE = M^{-1} \sum_{m=H+l}^{H+M} \left(\varepsilon_m^2 - \hat{\sigma}_m^2\right)^2; \tag{22}$$

$$MAE = M^{-1} \sum_{m=H+l}^{H+M} \left|\varepsilon_m^2 - \hat{\sigma}_m^2\right|, \tag{23}$$

Table 6
Future volatility forecasting results.

	MSE					
	1	5	10	20	50	100
BMSM-*N*	0.6742	0.6838	0.6849	0.7136	**0.7385**	0.7770
BMSM-*t*	0.6683	0.6768	0.6786	0.7085	0.7428	0.7775
BMSM-Skewed *t*	**0.6672**	**0.6758**	**0.6775**	**0.7068**	0.7416	**0.7750**
GARCH-*N*	0.6848	0.7014	0.6949	0.7389	0.8069	0.9014
GARCH-*t*	0.6869	0.7048	0.7008	0.7543	0.8748	1.1393
GARCH-Skewed *t*	0.6871	0.7051	0.7008	0.7548	0.8775	1.1485
	MAE					
	1	5	10	20	50	100
BMSM-*N*	0.7139	0.5225	0.5279	0.5571	0.6026	0.6632
BMSM-*t*	**0.7013**	0.5095	0.5122	0.5394	0.5823	0.6386
BMSM-Skewed *t*	0.7019	**0.5095**	**0.5119**	**0.5381**	**0.5802**	**0.6335**
GARCH-*N*	0.7273	0.5353	0.5423	0.5977	0.7149	0.8730
GARCH-*t*	0.7393	0.5489	0.5637	0.6357	0.8161	1.0943
GARCH-Skewed *t*	0.7392	0.5490	0.5639	0.6364	0.8187	1.0991

where squared returns ε_m^2 are a proxy of 'true' volatility used in the asset volatility literature to evaluate forecast errors. In order to reflect the improvements of volatility models versus historical volatility in forecasting, we display the relative MSEs and MAEs obtained in percentage of the MSEs and MAEs of a naive forecast using the in-sample variance.

Results of the relative MSEs and MAEs of out-of-sample forecasts are reported in Table 6. We find supportive results for the forecast capabilities of BMSM models. With respect to the BMSM models, we find that they produce relative MSEs and MAEs that are lower than one, there exists an obvious improvement of BMSM models over historical volatility method in future forecasting. Moreover, BMSM models have lower relative MSEs and MAEs than GARCH(1,1) models at all forecasting horizons. GARCH models perform well at short forecast horizons, but not very satisfying for longer horizons. For BMSM models, adding fat tails and skewness improves the forecasting capability. BMSM-Skewed *t* model produces the lowest relative MSEs and MAEs at most horizons. However, results of GARCH models are different. There are no improvements in forecasting performance in terms of both relative MSEs and MAEs when Student-*t* or Skewed *t* innovations are adopted. Though the results are not entirely supportive for BMSM-Skewed *t* model, there is only one exception by the loss measurement of relative MSEs or relative MAEs. overall, the results are quite encouraging.

There exists the argument that the MSE and MAE obtained only by one empirical study and specific data sets are not robust. In order to solve this problem, Hansen [29] proposed the superior predictive ability (SPA) test. He confirmed that SPA test exhibited outstanding model discrimination ability and robustness. The procedure of SPA test is as follows [30]: Suppose there are $K + 1$ different volatility models M_k, $k = 0, 1, \ldots, K$. Let the first model M_0 be the benchmark model that is compared to M_k, $k = 1, 2, \ldots, K$. Each model leads to a sequence of losses, $L_{k,m} \equiv L(\hat{\sigma}_m^2, r_m^2)$, $m = 1, 2, \ldots, M$, and we define the relative loss function:

$$X_{k,m} = L_{0,m} - L_{k,m}, \quad k = 1, 2, \ldots, K, \ m = 1, 2, \ldots, M. \tag{24}$$

The null hypothesis is that the benchmark model is as good as any other model in terms of expected loss, which can be formulated as the hypothesis $H_0 : \lambda_k \equiv \mathbb{E}(X_{k,m}) \leq 0$ for all $k = 1, 2, \ldots, K$. Hansen and Lunde [30] proved that the SPA test is based on the test statistic,

$$T_M^{SPA} \equiv \max_{k=1,2,\ldots,K} \bar{X}_k / \hat{\omega}_{kk}, \tag{25}$$

where $\bar{X}_k \equiv M^{-1} \sum_{m=1}^{M} X_{k,m}$, and $\hat{\omega}_{kk}^2 = var(\sqrt{M}\bar{X}_k)$.

In order to obtain a precise distribution of T_M^{SPA} and the critical value for T_M^{SPA}, Hansen and Lunde proposed to adopt stationary bootstrap of Politis and Romano [31]. Repeat the bootstrap procedure B times and obtain B resamples with length M, denote as $X_{k,m}^i$, $i = 1, 2, \ldots, B$. First, calculate the sample averages $\bar{X}_k^i = \frac{1}{M} \sum_{m=1}^{M} X_{k,m}^i$, $i = 1, 2, \ldots, B$, and $\hat{\omega}_{kk}^2 = \frac{M}{B} \sum_{i=1}^{B} \left(\bar{X}_k^i - \bar{X}_k\right)^2$. Then, define $\bar{Z}_k^i = \bar{X}_k^i - \bar{X}_k \cdot \mathbf{1}_{\bar{X}_k > -A_k}$, where $A_k = \frac{1}{4} M^{-1/4} \hat{\omega}_{kk}$. This enables us to approximate the distribution of T_M^{SPA} by the empirical distribution of

$$T_b^{SPA*,i} = \max_{k=1,2,\ldots,K} \frac{M^{1/2} \bar{Z}_k^i}{\hat{\omega}_{kk}}, \quad i = 1, 2, \ldots, B. \tag{26}$$

The p-value $\hat{p}_{SPA} \equiv B^{-1} \sum_{b=1}^{B} \mathbf{1}_{\left\{T_b^{SPA*,i} > T_M^{SPA}\right\}}$, and the null hypothesis is rejected for small p-values.

Table 7 shows the SPA test results for our empirical study. The first column shows the two loss functions, the names of benchmark models are in the second column. The bigger the p-value is, the better performance the benchmark model M_0 is. From Table 7, we can see that BMSM models outperform GARCH(1,1) models, and BMSM-Skewed *t* model performs the best for most of the cases. These are very consistent with the results in Table 6.

Table 7
Results of SPA tests.

		1	5	10	20	50	100
MSE	BMSM-*N*	0.0295	0.0140	0.0245	0.0360	**1**	0.4535
	BMSM-*t*	0.1535	0.0565	0.0710	0.0305	0.0565	0.0215
	BMSM-Skewed *t*	**1**	**1**	**1**	**1**	0.3070	1
	GARCH-*N*	0.0470	0.0400	0.0885	0.0465	0	0
	GARCH-*t*	0.0185	0.0120	0.0085	0	0	0
	GARCH-Skewed *t*	0.0145	0.0175	0.0105	0.0015	0	0
MAE	BMSM-*N*	0.0005	0	0	0	0	0
	BMSM-*t*	**1**	0.4875	0.3290	0.0315	0.0140	0
	BMSM-Skewed *t*	0.215	**1**	**1**	**1**	**1**	**1**
	GARCH-*N*	0.0090	0.0050	0.0025	0	0	0
	GARCH-*t*	0	0	0	0	0	0
	GARCH-Skewed *t*	0	0	0	0	0	0

4. Conclusion

This paper has proposed the BMSM model with Skewed t innovations which incorporates both fat tails and skewness. The BMSM-Skewed t model has been estimated via ML method, and volatility forecasting has been performed by Bayesian updating rule.

To evaluate the performance of BMSM-Skewed t model, an empirical study on Chinese SSEC index has been conducted. We have used the GARCH(1,1) model as the benchmark model and considered three types of innovations: Normal, Student-t and Skewed t. Besides, we have used the BMSM models with Normal and Student-t innovations for comparison. The empirical research has shown strong evidence of fat tails, skewness and multifractality in SSEC index. BMSM models have outperformed GARCH(1,1) models for out-of-sample volatility forecasting and produced relative MSEs and MAEs that are less than one. GARCH(1,1) models are less satisfying in forecasting future volatility over longer horizons, due to that they have difficulty in capturing the long memory.

Moreover, adding fat tails and skewness with Skewed t innovations has improved forecasts from BMSM models in terms of relative MSEs, MAEs and SPA test at most horizons. However, it is the opposite for GARCH(1,1) models. Though there are exceptions for superior of BMSM-Skewed t model in volatility forecasting, BMSM-Skewed t model dominates the most cases and the results are quite encouraging.

Acknowledgments

The authors would like to acknowledge the financial support of Major Project of the National Social Science Foundation of China (No. 11&ZD156), Guangzhou Financial Services Innovation Risk Management Research Base, National Natural Science Foundation of China (No. 71401151) and Zhejiang Provincial Natural Science Foundation of China (No. LQ13G010002).

References

[1] T. Andersen, T. Bollerslev, P. Christoffersen, F. Diebold, Volatility Forecasting, NBER Working Paper No. 11188, 2005, pp. 1–111. http://dx.doi.org/10.3386/w11188.
[2] S.H. Poon, C.W.J. Granger, Forecasting volatility in financial markets: A review, J. Econom. Lit. 41 (2003) 478–539.
[3] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, J. Econometrics 31 (1986) 307–327.
[4] T.G. Andersen, Stochastic autoregressive volatility: a framework for volatility modeling, Math. Finance 4 (1994) 75–102.
[5] T.G. Andersen, Return volatility and trading volume: An information flow interpretation of stochastic volatility, J. Financ. 51 (1996) 169–204.
[6] R.T. Baillie, T. Bollerslev, H.O. Mikkelsen, Fractionally integrated generalized autoregressive conditional heteroscedasticity, J. Econometrics 74 (1996) 3–30.
[7] F.J. Breidt, N. Crato, P.de. Lima, The detection and estimation of long memory in stochastic volatility, J. Econometrics 83 (1998) 325–348.
[8] J.D. Hamilton, Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates, J. Econom. Dynam. Control 12 (1988) 385–423.
[9] J.D. Hamilton, A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica 57 (1989) 357–384.
[10] J.D. Hamilton, Analysis of time series subject to changes in regime, J. Econometrics 45 (1990) 39–70.
[11] L. Calvet, A. Fisher, How to forecast long-run volatility: Regime switching and the estimation of multifractal processes, J. Financ. Econ. 2 (2004) 49–83.
[12] T. Lux, T. Kaizoji, Forecasting volatility and volume in the Tokyo stock market: Long memory, fractality and regime switching, J. Econom. Dynam. Control 31 (2007) 1808–1843.
[13] T. Lux, The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility, J. Bus. Econom. Statist. 26 (2008) 194–210.
[14] L. Calvet, A. Fisher, Multifractality in asset returns: Theory and evidence, Rev. Econ. Stat. 84 (2002) 381–406.
[15] K. Kim, S.M. Yoon, Multifractal features of financial markets, Physica A 344 (2004) 272–278.
[16] P. Norouzzadeh, G.R. Jafari, Application of multifractal measures to Tehran price index, Physica A 356 (2005) 609–627.
[17] P. Norouzzadeh, B. Rahmani, A multifractal detrended fluctuation description of Iranian rial-US dollar exchange rate, Physica A 367 (2006) 328–336.
[18] R.P. Liu, T.D. Matteo, T. Lux, True and apparent scaling: The proximity of the Markov-switching multifractal model to long-range dependence, Physica A 383 (2007) 35–42.
[19] D.O. Cajueiro, B.M. Tabak, Long-range dependence and multifractality in the term structure of LIBOR interest rates, Physica A 373 (2007) 603–614.
[20] P. Suárez-Garcia, D. Gómez-Ullate, Multifractality and long memory of a financial index, Physica A 394 (2014) 226–234.
[21] L.E. Calvet, A.J. Fisher, S.B. Thompson, Volatility comovement: a multifrequency approach, J. Econometrics 131 (2006) 179–215.

[22] T. Lux, L. Morales-Arias, Forecasting volatility under fractality, regime-switching, long memory and student-t innovations, Comput. Statist. Data Anal. 54 (2010) 2676–2692.
[23] B. Podobnik, D. Horvatic, A.M. Petersen, B. Urošević, H.E. Stanley, Bankruptcy risk model and empirical tests, Proc. Natl. Acad. Sci. USA 107 (2010) 18325–18330.
[24] B. Podobnik, A. Valentinčič, D. Horvatić, H.E. Stanley, Asymmetric Lévy flight in financial ratios, Proc. Natl. Acad. Sci. USA 108 (2011) 17883–17888.
[25] B.E. Hansen, Autoregressive conditional density estimation, Internat. Econom. Rev. 35 (1994) 705–730.
[26] G.F. Gu, W.X. Zhou, Detrending moving average algorithm for multifractals, Phys. Rev. E 82 (2010) 011136.
[27] B. Podobnik, H.E. Stanley, Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series, Phys. Rev. Lett. 100 (2008) 084102.
[28] Z.Q. Jiang, W.X. Zhou, Multifractal detrending moving-average cross-correlation analysis, Phys. Rev. E 84 (2011) 016106.
[29] P.R. Hansen, A test for superior predictive ability, J. Bus. Econom. Statist. 23 (2005) 365–380.
[30] P.R. Hansen, A. Lunde, A forecast comparison of volatility models: does anything beat a GARCH(1,1)? J. Appl. Econometrics 20 (2005) 873–889.
[31] D.N. Politis, J.P. Romano, The stationary bootstrap, J. Amer. Statist. Assoc. 89 (1994) 1303–1313.