Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Markov switch smooth transition HYGARCH model: Stability and estimation

Ferdous Mohammadi Basatini & Saeid Rezakhah

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2019.1576884

Markov switch smooth transition HYGARCH model: Stability and estimation

Ferdous Mohammadi Basatini and Saeid Rezakhah

Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran

ABSTRACT
HYGARCH model is basically used to model long-range dependence in volatility. We propose Markov switch smooth-transition HYGARCH model, where the volatility in each state is a time-dependent convex combination of GARCH and FIGARCH. This model provides a flexible structure to capture different levels of volatilities and also short and long memory effects. The necessary and sufficient condition for the asymptotic stability is derived. Forecast of conditional variance is studied by using all past information through a parsimonious way. Bayesian estimations based on Gibbs sampling are provided. A simulation study has been given to evaluate the estimations and model stability. The competitive performance of the proposed model is shown by comparing it with the HYGARCH and smooth-transition HYGARCH models for some period of the *S&P*500 and Dow Jones industrial average indices based on volatility and value-at-risk forecasts.

ARTICLE HISTORY
Received 21 October 2017
Accepted 18 January 2019

KEYWORDS
HYGARCH; long rang dependence; Markov switching; smooth transition; Griddy Gibbs sampling

MATHEMATICS SUBJECT CLASSIFICATION
60J10; 91B84; 62F15

1. Introduction

ARCH and GARCH models introduced by Engle (1982) and Bollerslev (1986), respectively, are used to capture volatility of returns. In many financial time series, there exist some shocks that have long memory impacts on future volatilities with positive correlations which decay slowly to zero (Baillie, Bollerslev, and Mikkelsen 1996; Wang, Wu, and Wei 2011; Kwan, Li, and Li 2011). On the other hand, the autocorrelation functions (ACFs) of the ARCH and GARCH models decay exponentially, and cannot produce long-range dependence. Baillie, Bollerslev, and Mikkelsen (1996) proposed FIGARCH model to capture long-range dependence that possesses hyperbolic decay of ACF but has infinite variance which limits its application. Davidson (2004) proposed HYGARCH model, where the conditional variance is a convex combination of the conditional variances of GARCH and FIGARCH models. The ACF of HYGARCH model decays hyperbolically. So HYGARCH models capture long-range dependence but have finite variance under some conditions and have shown good performance in modeling long-range dependence in financial time series (Davidson 2004; Tang and Shieh 2006; Niguez and Rubia 2006).

CONTACT Saeid Rezakhah rezakhah@aut.ac.ir Department of Statistics, Faculty of Mathematics and Computer Science, Amirkabir University of Technology, 424 Hafez Avenue, Tehran, 15914, Iran.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsta.

In many financial time series there exist some time-varying structures of volatility which change over time. Markov switching (MS) models allow sudden changes in the volatility. Different variants of the MS models were proposed for GARCH models (see Cai 1994; Hamilton and Susmel 1994; Gray 1996; Klaassen 2002; Haas, Mittink, and Paollella 2004; Marcucci 2005). The stationary conditions for some of these models were investigated in the work of Abramson and Cohen (2007). Bauwens, Preminger, and Rombouts (2010) presented sufficient conditions for the geometric ergodicity and existence of moments of MS-GARCH model.

Smooth-transition (ST) models allow a continuum of changes between two extreme regimes which are associated with the extreme values of transition function. The ST weights are continuous functions that are bonded between two limits. Transition between regimes is imposed by the preceding observations. Logistic transition functions are the most popular ones in these studies. For a review on ST models, refer to Granger and Teräsvirta (1993), Teräsvirta (1998), Gonzales-Rivera (1998), Lubrano (2001) and Amado and Teräsvirta (2008).

Value-at-risk (VaR) is a useful measure for quantifying the risk and is used as a regulatory tool. The observed VaR of models must neither overestimate nor underestimate the true VaR (for more details, see Jorion 1996; Dowd 1998; Brooks and Persand 2000). As VaR depends directly on the volatility, the forecasts from various volatility models are evaluated and compared on the basis of how well they forecast VaR. Hence, some statistical hypothesis testing is performed to test whether the VaR forecasts by competing models display the required theoretical properties (Zhang and Nadarajah 2017).

Current authors (Mohammadi and Rezakhah 2017) studied smooth transition HYGARCH (ST-HYGARCH) model, where the volatility is stated as some smooth transition of GARCH and FIGARCH. Their model creates a convex combination with time-varying logistic weights between short memory GARCH and long memory FIGARCH. In this paper, the Markov switch smooth transition HYGARCH (MSST-HYGARCH) model is introduced. The new proposed model allows a Markov switch behavior of ST-HYGARCH model between different levels of volatility. Each state consist of a ST-HYGARCH model with time-dependent logistic weight function. This model has the potential to switch between different levels of volatility and creates dynamic memory in each state to react to different shocks. The ST-HYGARCH model is a special case when there exists just one state. We derive a necessary and sufficient asymptotic stability condition. A dynamic time-dependent relation for forecasting conditional variance is obtained. Due to the recursive structure of conditional variance in MSST-HYGARCH models, the path-dependence problem occurs. This means the maximum likelihood (ML) estimation needs to integrate all hidden states, which is infeasible. So Bayesian estimation according to the Markov chain Monte Carlo (MCMC) algorithm is implemented to overcome the estimation problem. The advantages of the Bayesian estimation can be stated as follows: the local maximum is prevented, the information of the uncertain parameters can be achieved via joint posterior distribution, and required constraints on the model parameters can be imposed on prior distributions. Some statistical hypothesis testings are provided to evaluate the VaR accuracy for proposed model. The theoretical results are examined via simulation. We consider some periods of $S\&P500$ and indices as real data to show the competitive behavior of MSST-HYGARCH model in comparison to HYGARCH and ST-HYGARCH based on volatility and VaR forecasting.

The paper is organized as follows. The MSST-HYGARCH model is defined in Section 2. Section 3 is devoted to the investigation of model stability. In Section 4, we obtain the forecasting conditional variance. Estimation of the parameters is followed in Section 5. The VaR and its statistical accuracy are provided in Section 6. Section 7 is dedicated to simulation studies. The performance of the model for the empirical data of $S\&P500$ and Dow Jones industrial average indices is reported in Section 8.

2. The model

2.1. HYGARCH model

Let $\{y_t\}$ follows a HYGARCH model as

$$\begin{aligned} y_t &= \epsilon_t \sqrt{h_t} \\ h_t &= \frac{\gamma}{1-\lambda B} + \left\{ 1 - \frac{1-\delta B}{1-\lambda B} \left[1 - w + w(1-B)^d \right] \right\} y_t^2, \end{aligned} \tag{2.1}$$

where B is the back-shift operator, $\gamma > 0, \lambda, \delta, w \geq 0$, and the sequence $\{\epsilon_t\}$ consist of *iid* random variables with mean 0 and variance 1. Also $(1-B)^d = 1 - \Sigma_{i=1}^{\infty} g_i B^i$ where $g_i = \frac{d\Gamma(i-d)}{\Gamma(1-d)\Gamma(i+1)}$ in which $0 < d < 1$. Let Υ_{t-1} be the information up to t-1 then $Var(y_t|\Upsilon_{t-1}) = h_t$. One can easily verify that h_t might be written as

$$\begin{aligned} h_t &= (1-w+w)\frac{\gamma}{1-\lambda B} + \left\{ (1-w+w) - \left[\left(\frac{1-\delta B}{1-\lambda B} \right)(1-w) + \left(\frac{1-\delta B}{1-\lambda B} \right) w(1-B)^d \right] \right\} y_t^2 \\ &= (1-w)\left[\frac{\gamma}{1-\lambda B} + \left(1 - \frac{1-\delta B}{1-\lambda B} \right) y_t^2 \right] + w \left[\frac{\gamma}{1-\lambda B} + \left(1 - \frac{1-\delta B}{1-\lambda B} \right)(1-B)^d y_t^2 \right] \end{aligned}$$

and so we have

$$h_t = (1-w)h_{1,t} + wh_{2,t} \tag{2.2}$$

where

$$h_{1,t} = a_0 + a_1 h_{1,t-1} + a_2 y_{t-1}^2 \tag{2.3}$$

is the conditional variance of the GARCH(1,1) and

$$h_{2,t} = b_0 + b_1 h_{2,t-1} + \left[1 - b_1 B - (1-b_2 B)(1-B)^d \right] y_t^2. \tag{2.4}$$

is the one of the FIGARCH(1,d,1), where $a_0 = \gamma, a_1 = \lambda, a_2 = (\delta - \lambda), b_0 = \gamma, b_1 = \lambda$ and $b_2 = \delta$. In this model the conditional variance, h_t, is a convex combination of $h_{1,t}$ and $h_{2,t}$ with fixed weights; by allowing the weights and parameters to be time-dependent we provide a more flexible model for describing the volatilities.

2.2. Smooth transition HYGARCH model

Let $\{y_t\}$ follows the ST-HYGARCH model as

$$\begin{aligned} y_t &= \sqrt{h_t}\epsilon_t \\ h_t &= (1-w_t)h_{1,t} + w_t h_{2,t} \end{aligned} \tag{2.5}$$

and

$$w_t = \frac{\exp(-\gamma y_{t-1})}{1 + \exp(-\gamma y_{t-1})} \tag{2.6}$$

where $h_{1,t}$ and $h_{2,t}$ are defined in (2.3) and (2.4) respectively. Also $\{\epsilon_t\}$ are *iid* random variables with mean 0 and variance 1, $a_0, a_1, a_2, b_0, \gamma > 0, 0 < b_2 \leq b_1 \leq d < 1$ and $(1-B)^d$ is defined as in (2.1).

2.3. The Markov switch smooth transition HYGARCH model

Let $\{y_t\}$ follows the MSST-HYGARCH model as

$$y_t = \sqrt{h_{t,Z_t}}\epsilon_t \tag{2.7}$$

where$\{\epsilon_t\}$ are *iid* random variables with mean 0 and variance 1 and are independent of $\{Z_t\}$. The $\{Z_t\}$ is a Markov chain which identify the state at time t as $z_t = 1, 2, \ldots, m$. Also the transition probability matrix $P = ||p_{rs}||_{m\times m}$ where $p_{rs} = p(Z_t = s|Z_{t-1} = r)$ $r, s = 1, 2, \ldots, m$, with stationary probabilities $\Pi = [\pi_1, \ldots, \pi_m]'$. The conditional variance in state j, $h_{t,j}, j = 1, 2, \ldots, m$ is given with

$$h_{t,j} = (1-w_{t,j})h_{1,t,j} + w_{t,j}h_{2,t,j}, \tag{2.8}$$

where

$$h_{1,t,j} = a_{0j} + a_{1j}h_{1,t-1,j} + a_{2j}y_{t-1}^2 \tag{2.9}$$

$$h_{2,t,j} = b_{0j} + b_{1j}h_{2,t-1,j} + \left[1-b_{1j}B-\left(1-b_{2j}B\right)\left(1-B\right)^{d_j}\right]y_t^2 \tag{2.10}$$

$$w_{t,j} = \frac{\exp(-\gamma_j y_{t-1})}{1 + \exp(-\gamma_j y_{t-1})} \tag{2.11}$$

and $a_{0j}, a_{1j}, a_{2j}, b_{0j} > 0, 0 < b_{2j} \leq b_{1j} \leq d_j < 1$ cause the conditional variance to be strictly positive. The $(1-B)^{d_j}$ for $j = 1, 2 \ldots m$ are defined as in (2.1). The parameters $\gamma_j > 0, j = 1, 2, \ldots, m$ are called smoothing parameters; they ensure a smooth transition from short to long memory and vice versa. The logistic weight functions $w_{t,j}, j = 1, 2, \ldots, m$ decrease monotonically and are bounded between 0 and 1.

In each state, the conditional variance is a time-dependent convex combination of GARCH(1,1) and FIGARCH $(1,d_j,1)$ conditional variances. The states can be considered for different levels of volatilities. For two states, one can be considered for low and the other for high volatility. As y_{t-1} tends to $-\infty$, w_t approaches one. So, at time t, the MSST-HYGARCH model tends to the MS-FIGARCH model. Also, as y_{t-1} tends to ∞, w_t approaches zero and the MSST-HYGARCH model tends to the MS-GARCH model at time t.

3. Stability

Stability of the model which refers to the asymptotic finiteness of the variance of the series can be imposed by considering some conditions to guarantee the asymptotic boundedness of unconditional second moment. Following Abramson and Cohen (2007) the unconditional second-order moment is calculated as

$$E\left(y_t^2\right) = E\left(h_{t,Z_t}\epsilon_t^2\right) = E\left(h_{t,Z_t}\right)E\left(\epsilon_t^2\right) = E\left(h_{t,Z_t}\right)$$

and so

$$E\left(h_{t,Z_t}\right) = E_{Z_t}\left[E_{t-1}\left(h_{t,Z_t}|Z_t = z_t\right)\right] = \sum_{z_t=1}^{m} \pi_{z_t} E_{t-1}\left(h_{t,z_t}|Z_t = z_t\right). \tag{3.1}$$

where $E_t(.)$ denotes the conditional expectation with respect to the information up to time t. we denote $E(.|Z_t = z_t)$ and $p(.|Z_t = z_t)$ by $E(.|z_t)$ and $p(.|z_t)$ respectively. By rewriting (2.10) as:

$$h_{2,t,j} = b_{0j} + b_{1j}h_{2,t-1,j} + \left(b_{2j} - b_{1j} + g_{1j}\right)y_{t-1}^2 + \sum_{i=0}^{\infty}\left(g_{i+2j} - b_{2j}g_{i+1j}\right)B^i y_{t-2}^2, \tag{3.2}$$

we have that

$$\begin{aligned}
E_{t-1}\left(h_{t,j}|z_t\right) &= E_{t-1}\left((1-w_{t,j})h_{1,t,j} + w_{t,j}h_{2,t,j}|z_t\right)\\
&= a_{0j} + \underbrace{\left(b_{0j}-a_{0j}\right)E_{t-1}\left(w_{t,j}|z_t\right)}_{I} + a_{1j}\underbrace{E_{t-1}\left((1-w_{t,j})h_{1,t-1,j}|z_t\right)}_{II} + b_{1j}\underbrace{E_{t-1}\left(w_{t,j}h_{2,t-1,j}|z_t\right)}_{III}\\
&\quad \underbrace{+\left(b_{2j}-b_{1j} + g_{1j} - a_{2j}\right)E_{t-1}\left(w_{t,j}y_{t-1}^2|z_t\right)}_{IV} + \sum_{i=0}^{\infty}\underbrace{\left(g_{i+2j} - b_2 g_{i+1j}\right)E_{t-1}\left(w_{t,j}y_{t-2-i}^2|z_t\right)}_{V}\\
&\quad + a_{2j}E_{t-1}\left(y_{t-1}^2|z_t\right)
\end{aligned} \tag{3.3}$$

and using the fact that $0 < w_{t,j} < 1$ we have the following bounds for terms $(I)-(V)$ in (3.3):

$$\begin{aligned}
&I \Rightarrow \left(b_{0j}-a_{0j}\right)E_{t-1}\left(w_{t,j}|z_t\right) \le |b_{0j}-a_{0j}|\\
&II \Rightarrow E_{t-1}\left((1-w_{t,j})h_{1,t-1,j}|z_t\right) \le E_{t-1}\left(h_{1,t,j}|z_t\right)\\
&III \Rightarrow E_{t-1}\left(w_{t,j}h_{2,t-1,j}|z_t\right) \le E\left(h_{2,t-1,j}|z_t\right)\\
&IV \Rightarrow \left(b_{2j}-b_{1j} + g_{1,j} - a_{2j}\right)E_{t-1}\left(w_{t,j}y_{t-1}^2\right) \le |b_{2j}-b_{1j} + g_{1j} - a_{2j}|E_{t-1}\left(y_{t-1}^2|z_t\right)\\
&V \Rightarrow \left(g_{i+2j}-b_2 g_{i+1j}\right)E_{t-1}\left(w_{t,j}y_{t-2-i}^2|z_t\right) \le |g_{i+2j}-b_2 g_{i+1j}|E_{t-1}\left(y_{t-2-i}^2|z_t\right).
\end{aligned} \tag{3.4}$$

The term $E_{t-1}(y_{t-i}^2|z_t)$ for $i = 1, 2, \ldots$ can be evaluated as:

$$\begin{aligned}
E_{t-1}\left(y_{t-i}^2|z_t\right) &= \sum_{z_{t-i}=1}^{m}\int_{\Upsilon_{t-1}} y_{t-i}^2 p(\Upsilon_{t-1}|z_t, z_{t-i})p(z_{t-i}|z_t)d\Upsilon_{t-1}\\
&= \sum_{z_{t-i}=1}^{m} p(z_{t-i}|z_t)E_{t-1}\left[y_{t-i}^2|z_{t-i}, z_t\right] \qquad i = 1, 2, \ldots \ .
\end{aligned} \tag{3.5}$$

Using the fact that given the current active state, the expected value of y_{t-i}^2 is independent of any future states (Abramson and Cohen 2007), we get

$$E_{t-1}\left[y_{t-i}^2|z_{t-i}, z_t\right] = E_{t-1}\left[y_{t-i}^2|z_{t-i}\right] \tag{3.6}$$

$$\begin{aligned} &= \int_{\Upsilon_{t-i-1}}\int_{y_{t-i}} y_{t-i}^2 p\left(y_{t-i}|\Upsilon_{t-i-1}, z_{t-i}\right)p(\Upsilon_{t-i-1}|z_{t-i})dy_{t-i}d\Upsilon_{t-i-1} \\ &= E_{t-i-1}\left[E\left(y_{t-i}^2|\Upsilon_{t-i-1}, z_{t-i}\right)|z_{t-i}\right] \\ &= E_{t-i-1}\left[h_{t-i,z_{t-i}}|z_{t-i}\right]. \end{aligned} \tag{3.7}$$

Remark 1. Relations (3.6) and (3.7) indicate that given the current state the returns and volatilities are independent of successive states, which means there is not leverage effect (McAleer 2014). Also

$$\begin{aligned} E_{t-1}\left[h_{k,t-1,j}|z_t\right] &= \sum_{z_{t-1}=1}^{m}\int_{\Upsilon_{t-1}} h_{k,t-1,j}p(\Upsilon_{t-1}|z_t, z_{t-1})p(z_{t-1}|z_t)d\Upsilon_{t-1} \\ &= \sum_{z_{t-1}=1}^{m} p(z_{t-1}|z_t)E_{t-2}\left[h_{k,t-1,j}|z_{t-1}\right] \qquad k = 1, 2. \end{aligned} \tag{3.8}$$

By replacing the results obtained in (3.4)–(3.8) in (3.3) we obtain the following bound for $E_{t-1}[h_{t,j}|z_t]$:

$$\begin{aligned} E_{t-1}\left[h_{t,j}|z_t\right] \le{} & \left(a_{0j} + |b_{0j} - a_{0j}|\right) + a_{1j}\sum_{z_{t-1}=1}^{m} p(z_{t-1}|z_t)E_{t-2}\left(h_{1,t-1,j}|z_{t-1}\right) \\ & + b_{1j}\sum_{z_{t-1}=1}^{m} p(z_{t-1}|z_t)E_{t-2}\left(h_{2,t-1,j}|z_{t-1}\right) \\ & + \left(|b_{2j} - b_{1j} + g_{1j} - a_{2j}|\right)\sum_{z_{t-1}=1}^{m} p(z_{t-1}|z_t)E_{t-2}\left(h_{t-1,z_{t-1}}|z_{t-1}\right) \\ & + \sum_{i=0}^{\infty}(g_{i+2j} - b_{2j}g_{i+1j})\sum_{z_{t-2-i}=1}^{m} p(z_{t-2-i}|z_t)E_{t-2-i-1}\left(h_{t-2-i,z_{t-2-i}}|z_{t-2-i}\right). \end{aligned} \tag{3.9}$$

Using Bayes rule

$$p(z_{t-i}|z_t) = \frac{\pi_{z_{t-i}}}{\pi_{z_t}}P^i_{z_{t-i},z_t}$$

where P^i is i-th power of the transition probability matrix.

Let $H_t = [E(h_{t,1}|z_t = 1), \ldots, E(h_{t,m}|z_t = m)]', H_{kt} = [E(h_{k,t,1}|z_t = 1), \ldots, E(h_{k,t,m}|z_t = m)]'$ for $k=1, 2$, $\tilde{H}_t = [H_t', H_{1t}', H_{2t}', H_{t-1}']'$ and

$$\Lambda = [v_1, \ldots, v_m, a_{01}, \ldots, a_{0m}, b_{01}, \ldots, b_{0m}, 0, \ldots, 0]'$$

where $v_i = a_{0i} + |b_{0i} - a_{0i}|$ for $i = 1, 2, \ldots, m$ be a vector of size 4 m. Also let $\nu_i = (b_{2i} - b_{1i} + g_{1i} - a_{2i})$ for $i = 1, 2, \ldots, m$, define the diagonal matrices $\delta = \text{diag}(\nu_1, \ldots, \nu_m)$, $\mathrm{a}_1 = \text{diag}(a_{11}, \ldots, a_{1m})$, $\mathrm{a}_2 = \text{diag}(a_{21}, \ldots a_{2m})$, $\mathrm{b}_1 = \text{diag}(b_{11}, \ldots, b_{1m})$, $\boldsymbol{c} = \text{diag}(\tau_1, \ldots, \tau_m)$ where

$\tau_i = (b_{2i} - b_{1i} + g_{1i}), i = 1, 2, ..., m$. Suppose $f = [f_{ij}], ij = 1, 2, ..., m$ be a square matrix with elements $f_{rj} = \sum_{i=0}^{\infty}(g_{i+2j} - b_{2j}g_{i+1j})p(z_{t-2-i} = r|z_t = j)B^i$. Let

$$Q = \begin{bmatrix} \delta\dot{p} & a_1\dot{p} & b_1\dot{p} & f \\ a_2\dot{p} & a_1\dot{p} & 0_m & 0_m \\ c\dot{p} & 0_m & b_1\dot{p} & f \\ I & 0_m & 0_m & 0_m \end{bmatrix}$$

be a $4m-by-4m$ block matrix where 0_m is a square matrix of zeros. Also I_m and $\dot{p}$ represent respectively the identity matrix and the transpose of the transition matrix. Then a recursive vector form of (3.9) is obtained as:

$$\tilde{H}_t \leq \Lambda + Q\tilde{H}_{t-1}, \quad t \geq 0 \tag{3.10}$$

with some initial conditions $\tilde{H}_{-1}$.

Let $\Pi = [\pi_1, ..., \pi_m]'$, if $\vartheta(.)$ denote the spectral radius of a matrix, then the next theorem expresses the stability condition of the MSST-HYGARCH model.

Theorem 3.1. *The time series $\{y_t\}$ defined in relations (2.7)–(2.11) is asymptotically stable in unconditional second moment, $lim_{t\to\infty}E(y_t^2) \leq \Pi'(I-Q)^{-1}\Lambda$, if and only if $\vartheta(Q) < 1$.*

Proof. Let the recursive inequality (3.10) be written as

$$\tilde{H}_t \leq \Lambda \sum_{i=0}^{t-1} Q^i + Q^t\tilde{H}_0.$$

Using the matrix convergence theorem (Lancaster and Tismenetsky 1985), if $\vartheta(Q) < 1$ then Q^t is convergent to zero as $t \to \infty$ and if matrix $(I-Q)$ is invertible then $\sum_{i=0}^{t-1} Q^i$ convergences to $(I-Q)^{-1}$. So if $\vartheta(Q) < 1$,

$$lim_{t\to\infty}\tilde{H}_t \leq (I-Q)^{-1}\Lambda.$$

The asymptotic behavior of the unconditional second moment is bounded with

$$lim_{t\to\infty}E\left(y_t^2\right) \leq \Pi'(I-Q)^{-1}\Lambda$$

otherwise, if $\vartheta(Q) \geq 1$, the unconditional second moment goes to infinity with the growth of the time index (Abramson and Cohen 2007). □

Remark 2. For ST-HYGARCH, the $\tilde{H}_t, \Lambda$ and Q can be defined as:

$$\tilde{H}_t = \left[E(h_t), E(h_{1,t}), E(h_{2,t}), E(h_{t-1})\right]', \quad \Lambda = [a_0 + |b_0 - a_0|, a_0, b_0, 0]'$$

and

$$Q = \begin{bmatrix} (|b_2 - b_1 + g_1 - a_2| + a_2) & a_1 & b_1 & \sum_{i=0}^{\infty}|g_{i+2} - b_2 g_{i+1}|B^i \\ a_2 & a_1 & 0 & 0 \\ (b_2 - b_1 + g_1) & 0 & b_1 & \sum_{i=0}^{\infty}|g_{i+2} - b_2 g_{i+1}|B^i \\ 1 & 0 & 0 & 0 \end{bmatrix},$$

then the stability of unconditional second moment is resulted from theorem 1.

Remark 3. For MS-HYGARCH, the stability of unconditional second moment is resulted from theorem 1 if $0 < w_j < 1$; and $\tilde{H}_t, \Lambda$ and Q are defined as in MSST-HYGARCH model.

Remark 4. For different variants of MS-GARCH model Abramson and Cohen (2007) derived the stability conditions of unconditional second moment.

4. Forecasting

In this section we calculate the forecasting conditional variance of MSST-HYGARCH model. The conditional density function of y_t given the Υ_{t-1} can be written as:

$$f(y_t|\Upsilon_{t-1}) = \sum_{j=1}^{m} p(Z_t = j|\Upsilon_{t-1}) f(y_t|Z_t = j, \Upsilon_{t-1}). \tag{4.1}$$

If $\{\epsilon_t\}$ are standard normal variables then $f(y_t|Z_t = j, \Upsilon_{t-1}) = \frac{1}{\sqrt{2\pi h_{t,j}}} exp(-\frac{y_t^2}{h_{t,j}})$ and $p(Z_t = j|\Upsilon_{t-1})$ can be obtained recursively by the same method as in Alemohammad, Rezakhah, and Alizadeh (2016):

$$\psi_j^{(t)} = p(Z_t = j|\Upsilon_{t-1}) = \frac{\sum_{k=1}^{m} f(y_{t-1}|Z_{t-1} = k, \Upsilon_{t-2}) p(Z_{t-1} = k|\Upsilon_{t-2}) p_{kj}}{\sum_{k=1}^{m} f(y_{t-1}|Z_{t-1} = k, \Upsilon_{t-2}) p(Z_{t-1} = k|\Upsilon_{t-2})}. \tag{4.2}$$

So the conditional variance can be evaluated as:

$$\begin{aligned} V(y_t|\Upsilon_{t-1}) &= \sum_{k=1}^{m} \psi_k^{(t)} h_{t,k} \\ &= \sum_{k=1}^{m} \psi_k^{(t)} \left((1-w_{t,k})h_{1,t,k} + w_{t,k} h_{2,t,k}\right). \end{aligned} \tag{4.3}$$

where $\psi_k^{(t)} = p(Z_t = k|\Upsilon_{t-1})$ is defined by (4.2).

5. Estimation

Markov switching models cause difficulties in the ML estimation since the conditional variance at time t depends on the whole state path up to t; since this path is hidden, the likelihood of the observations can be calculated by integrating all possible state paths. This integration grows exponentially with the size of the observations. So, it is infeasible numerically. Bayesian inference is a technique that tackles the estimation issue of the Markov switching models very well (Bauwens, Preminger, and Rombouts 2010; Ardia 2009; Alemohammad, Rezakhah, and Alizadeh 2016). In this framework, the latent states are treated as parameters of the model and will be estimated. In this section it is assumed that $\{\epsilon_t\}$ are *iid* standard normal variables. The parameters of the MSST-HYGARCH model are

$\theta = (a_{01}, a_{11}, a_{21}, b_{01}, b_{11}, b_{21}, d_1, \gamma_1, a_{02}, a_{12}, a_{22}, b_{02}, b_{12}, b_{22}, d_2, \gamma_2)$ and $\eta = (p_{11}, p_{22})$.

Denoting $Y = (y_1, ..., y_T), Z = (z_1, ..., z_T), Y_t = (y_1, ..., y_t)$ and $Z_t = (z_1, ..., z_t)$, where T the size of data.

We use the Gibbs sampling algorithm (Gelfand and Smith 1990) to estimate the parameters of the MSST-HYGARCH model. This process generates a Markov chain which after warm-up phase convergences to the posterior distribution under regularity conditions (Robert and Casella 2004). The idea of this algorithm is to sample from the posterior density $p(\theta, \eta, Z|Y)$. These samples then serve to estimate features of the posterior distribution, like means, standard deviations and marginal densities.

Since the posterior density $p(\theta, \eta, Z|Y)$ is not standard hence the Gibbs sampling is down using the lower dimensional distributions, called blocks. For the MSST-HYGARCH model the blocks are θ, η and Z.

Gibbs algorithm steps: Let $\theta^{(r)}, \eta^{(r)}$ and $Z^{(r)}$ denote the draws at r-th iteration of the algorithm.

1. At iteration 1, initial value $\theta^{(0)}, \eta^{(0)}, Z^{(0)}$ must be used.
2. Given the (r-1)-th sample, the next one is found as:
   (i) $Z^{(r)}$ is sampled from $p(Z|\theta^{(r-1)}, \eta^{(r-1)}, Y)$. In this step we use the method of Chib (1996).
   (ii) $\eta^{(r)}$ is sampled from $p(\eta|\theta^{(r-1)}, Z^{(r)}, Y)$ that is independent from Y and θ.
   (iii) $\theta^{(r)}$ is sampled from $p(\theta|Z^{(r)}, \eta^{(r)}, Y)$ which is independent from η. As the $p(\theta|Z^{(r)}, \eta^{(r)}, Y)$ dose not have a closed-form in this step we use the Griddy Gibbs algorithm introduced by Ritter and Tanner (1992).
3. Increase r.
4. Repeat 2–3 until convergence.

For more details, see Bauwens, Preminger, and Rombouts (2010), Ardia (2009) and Alemohammad, Rezakhah, and Alizadeh (2016). We will now explain the above-mentioned steps in detail.

5.1. Sampling z_t

To obtain a sample of z_t we use the method of Chib (1996). For $t=1$ to $t=T$ repeat the following steps.

Prediction step: by the law of total probability determines

$$p(z_t|\eta, \theta, Y_{t-1}) = \sum_{z_{t-1}} p(z_{t-1}|\eta, \theta, Y_{t-1}) p_{z_{t-1}z_t}.$$

Update step: by the Bayes theorem determines

$$p(z_t|\eta, \theta, Y_t) \propto f\left(y_t|\theta, z_t = j, Y_{t-1}\right) p(z_t|\eta, \theta, Y_{t-1})$$

where $f(y_t|\theta, z_t = j, Y_{t-1}) = \frac{1}{\sqrt{2\pi h_{t,j}}} exp(-\frac{y_t^2}{h_{t,j}})$.

For $p(z_1|\eta, \theta, Y_0)$ we use the stationary probability of the chain. Then z_T is sampled from $p(z_T|\eta, \theta, Y)$ and for $t = T-1, ..., 1$ we run a backward algorithm to sample from

$$p(z_t|z_{t+1}, ..., z_T, \eta, \theta, Y) \propto p(z_t|\eta, \theta, Y) p_{z_t z_{t+1}}.$$

where z_t is sampled from $p(z_t|.)$ like sampling from Bernoulli distribution.

5.2. Sampling η

The posterior probability $p(\eta|\theta, Z, Y)$ is independent from θ and Y. Hence given the states, Z

$$p(p_{11}|Z) \propto p(p_{11})p(Z|p_{11}) = p_{11}^{c_{11}+n_{11}-1}(1-p_{11})^{c_{12}+n_{12}-1}$$

and

$$p(p_{22}|Z) \propto p(p_{22})p(Z|p_{22}) = p_{22}^{c_{22}+n_{22}-1}(1-p_{22})^{c_{21}+n_{21}-1}$$

where $p(p_{11}), p(p_{22})$ are independent beta prior densities respectively for p_{11} and p_{22}. Also c_{11}, c_{12}, c_{21}, c_{22} are the parameters of the beta prior and n_{ij} is the number of transitions from $z_{t-1} = i$ to $z_t = j$.

5.3. Sampling θ

The posterior density of θ is independent from η. So if θ_i is an arbitrary element of the vector θ then

$$p(\theta_i|Z, Y, \theta_{-\theta_i}) \propto p(\theta_i) \prod_{t=1}^{T} f(y_t|\theta, z_t = j, Y_{t-1}) = p(\theta_i) \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi h_{t,j}}} exp\left(\frac{-y_t^2}{h_{t,j}}\right) \tag{5.1}$$

where $p(\theta_i)$ is the prior of θ_i, also $\theta_{-\theta_i}$ denotes vector θ excluding θ_i and $h_{t,j}$ is defined in relations (2.7)–(2.11). If $h_{t,j}$ were fixed (5.1) would be a normal density. But h_t is a function of both θ_i and $\theta_{-\theta_i}$ so the conditional posterior density of θ_i contains $h_{t,j}$ which is also a function of θ_i. Obviously $p(\theta_i|Z, Y, \theta_{-\theta_i})$ doesn't belong to normal or any other well known density; so we can't sample from it in straightforward manner. Griddy Gibbs algorithm can be used to handle such situations. Given the draws of iteration r for iteration $r+1$ Griddy Gibbs algorithm runs as follows:

1. Set $(\theta_i^{(1)}, \theta_i^{(2)}, \ldots, \theta_i^{(H)})$ as a grid of points for θi. Using (5.1) compute the kernel of posterior density function $k(\theta_i|Z, Y, \theta_{-\theta_i})$ and evaluate it over the grid points to compute the vector $G_k = (k_1, \ldots, k_H)$. H refers to the number of grid points.
2. Compute $G_\Phi = (0, \phi_2, \ldots, \phi_H)$ where ϕ_j is obtained by using deterministic integration rule as
$$\phi_j = \int_{\theta_i^1}^{\theta_i^j} k(\theta_i|Z^{(r)}, Y, \theta_{-\theta_i}^{(r)})d\theta_i, \quad j = 2, \ldots, H.$$
3. Draw $u \sim U(0, \phi_H)$ and invert $\phi(\theta_i|Z^{(r)}, Y, \theta_{-\theta_i}^{(r)})$ by numerical interpolation to get the sample $\theta_i^{(r+1)}$.
4. Repeat steps 1–3 for other parameters.

6. Value-at-risk

VaR(ρ) is a value that with probability ρ the losses are equal to or exceed it at given trading period and with probability $(1-\rho)$ the losses are lower than it. VaR is obtained by calculating the ρ, the percentile of the predictive distribution (Ardia 2009). We use the relation

$$VaR_t(\rho) = F^{-1}(\rho)\sigma_t$$

to calculate value-at-risk in $(1-\rho)$ confidence level that $F^{-1}(\rho)$ is the inverse distribution of standardized observation (y_t/σ_t) where $\sigma_t = \sqrt{V(y_t|\Upsilon_{t-1})}$; $V(y_t|\Upsilon_{t-1})$ is computed in relation (4.3). Due to the importance of VaR in management risk, evaluating the accuracy of the VaR forecasts from different models is a substantial task. Here, we use some tests to examine the accuracy of the VaR forecasts.

6.1. Unconditional coverage test

A well-specified VaR model should produce VaR forecasts that cover the pre-specified probability. This means that 5% of time the losses should exceed the VaR(0.05). If the number of exceedances substantially differs from what is expected, then the model's accuracy is questionable. If the actual loss exceeds the VaR forecasts, this is termed an "exception," which can be presented by the indicator variable q_t as

$$q_t = \begin{cases} 1 & \text{if} \quad y_t < VaR_t(\rho) \\ 0 & \text{if} \quad y_t \geq VaR_t(\rho) \end{cases}$$

Obviously, q_t is a Bernoulli random variable with probability φ. The Kupiec test (Kupiec 1995), also known as the unconditional coverage (UC) test, is designed to test the number of exceptions based on the likelihood ratio (LR) test. The null hypothesis of the UC test is $H_0 : \rho = \varphi$. Then the LR test of the unconditional coverage (LR_{uc}) is defined as

$$LR_{UC} = -2\log\left(\frac{L^0_{UC}}{L^1_{UC}}\right) = -2\log\left(\frac{\rho^n(1-\rho)^{T-n}}{\widehat{\varphi}^n(1-\widehat{\varphi})^{T-n}}\right) \tag{6.1}$$

where L^0_{uc} and L^1_{uc} are the likelihood functions respectively under H_0 and H_1, T is the number of the forecasting samples, n is the number of the exceptions and $\widehat{\varphi} = \frac{n}{T}$ is the ML estimate of the φ under H_1. Under H_0, the LR_{UC} is asymptotically distributed as a χ^2 random variable with one degree of freedom.

6.2. Independent test

If the volatilities are low in some periods and high in others, the forecasts should respond to this clustering event. It means that the VaR should be small in times of low volatility and high in times of high volatility. So, the exceptions are spread over the entire sample period independently and do not appear in clusters (Sarma, Thomas, and Shah 2003). A model that cannot capture the clustering of volatilities will exhibit the symptom of clustering of the exceptions. Kupiec's test cannot check the clustering of the exceptions. Christoffersen (1998) designed an independent (IND) test based on the LR to test the clustering of the exceptions. The null hypothesis of the IND test assumes that the probability of an exception on a given day t is not influenced by what happened the day before. Formally, $H_0 : \varphi_{10} = \varphi_{00}$, where φ_{ij} denotes that the probability of an i event on day $t - 1$ must be followed by a j event on day t; $\varphi_{ij} = p(q_t = j|q_{t-1} = i)$, where $i, j = 0, 1$. The LR statistic of the IND test (LR_{IND}) can be obtained as

$$LR_{IND} = -2\log\left(\frac{L^0_{IND}}{L^1_{IND}}\right) = -2\log\left(\frac{\widehat{\varphi}^n(1-\widehat{\varphi})^{T\bar{n}}}{\widehat{\varphi}_{01}^{n_{01}}(1-\widehat{\varphi}_{01})^{n_{00}}\widehat{\varphi}_{11}{}^{n_{11}}(1-\widehat{\varphi}_{11})^{n_{10}}}\right). \tag{6.2}$$

Where n_{ij} is the number of observations with value i followed by value j $(i,j=0,1)$, $\widehat{\varphi}_{01}=\frac{n_{01}}{n_{00}+n_{01}}$ and $\widehat{\varphi}_{11}=\frac{n_{11}}{n_{10}+n_{11}}$. Under H_0, the LR_{UC} is asymptotically distributed as a χ^2 random variable with one degree of freedom.

6.3. Conditional coverage test

The IND test is not complete on its own. Hence, Christoffersen (1998) proposed a joint test: the conditional coverage (CC) test, which combines the properties of both the UC and IND tests. The null hypothesis of the CC test checks both the exception cluster and consistency of the exceptions with VaR confidence level. The null hypothesis of the test is $H_0:\varphi_{01}=\varphi_{11}=\rho$. The LR test statistic is obtained as

$$LR_{CC} = -2\log\left(\frac{L^0_{CC}}{L^1_{CC}}\right) = -2log\left(\frac{\rho^n(1-\rho)^{T-n}}{\widehat{\varphi}_{01}^{n_{01}}(1-\widehat{\varphi}_{01})^{n_{00}}\widehat{\varphi}_{11}{}^{n_{11}}(1-\widehat{\varphi}_{11})^{n_{10}}}\right), \tag{6.3}$$

Under H_0, LR_{CC} is asymptotically distributed as a χ^2 random variable with two degrees of freedom. It is a summation of two separate statistics, LR_{UC} and LR_{IND}, as given below:

$$\begin{aligned}
LR_{CC} &= -2\left[\log\left(L^0_{CC}\right) - \log\left(L^1_{CC}\right)\right] \\
&= -2\left[\log\left(L^0_{UC}\right) - \log\left(L^1_{IND}\right)\right] \\
&= -2\left[\log\left(L^0_{UC}\right) - \log\left(L^1_{UC}\right) + \log\left(L^0_{IND}\right) - \log\left(L^1_{IND}\right)\right]\Big)\Big] \\
&= -2\left[\log\left(L^0_{UC}\right) - \log\left(L^1_{UC}\right)\right] - 2\left[\log\left(L^0_{IND}\right) - \log\left(Ł^1_{IND}\right)\right] \\
&= LR_{UC} + LR_{IND}
\end{aligned} \tag{6.4}$$

6.4. Testing based on loss function

Diebold and Mariano (1995) proposed a hypothesis test of the efficiency of the models based on quadratic loss function. The quadratic loss function for model i is defined as:

$$l_{i,t} = \begin{cases} 1 + (y_t - Var_{i,t})^2 & \text{if} \quad y_t < VaR_{i,t} \\ 0 & \quad\quad otherwise \end{cases}$$

and the loss differential between model i and model j is defined as $o_t = l_{i,t} - l_{j,t}$. Negative values of o_t indicate a superiority of model i over model j. In this framework the null hypothesis, $H_0:\zeta=0$ is tested against the one-sided hypothesis $H_1:\zeta<0$ where ζ is the median of the distribution of o_t. Define an indicator variable

$$q_t = \begin{cases} 1 & \text{if} \quad o_t > 0 \\ 0 & \text{if} \quad o_t \le 0 \end{cases}$$

and the sign statistic $S_{i,j}=\sum_{t=1}^{T} q_t$ is the number of non-negative o_t's. If o_t are *iid*, then under H_0, $S_{i,j}$ is binomial $(T, 0.5)$. For large sample $S^a_{i,j}=\frac{S_{i,j}-0.5T}{0.25T}$ is asymptotically

Table 1. Descriptive statistics of the simulated observations from MSST-HYGARCH model.

Mean	Std.dev	Minimum	Maximum	Skewness	Kurtosis
0.080	1.491	−7.030	6.166	−0.070	1.982

Table 2. Estimation results of MSST-HYGARCH on simulated observations based on 10,000 Gibbs sampling iterations.

	True Value	Mean	Bias	Std.dev.
a_{01}	0.180	0.204	0.024	0.048
a_{11}	0.200	0.205	0.005	0.049
a_{21}	0.250	0.254	0.004	0.049
b_{01}	0.150	0.205	0.055	0.049
b_{11}	0.140	0.130	−0.010	0.037
d_1	0.400	0.404	0.004	0.049
γ_1	0.600	0.609	0.009	0.098
a_{02}	1.500	1.528	0.028	0.244
a_{12}	0.400	0.406	0.006	0.049
a_{22}	0.350	0.354	0.004	0.049
b_{02}	1.000	1.030	0.030	0.246
b_{12}	0.180	0.179	−0.001	0.037
d_2	0.850	0.805	0.045	0.049
γ_2	2.000	2.017	0.017	0.146
p_{11}	0.850	0.900	0.050	0.036
p_{22}	0.600	0.780	0.180	0.093

distributed as $N(0,\ 1)$ and H_0 is rejected at 5% significance level if $S^a_{i,j} < -1.645$. Rejection of H_0 would imply that model i is significantly better than model j.

7. Simulated data

This section conducts three simulation experiments to evaluate the performance of the MSST-HYGARCH model defined in (2.7)–(2.11). A two-state Markov chain was considered where the first state corresponds to low volatilities and the second corresponds to higher volatilities and it is assumed that $\{\epsilon_t\}$ are *iid* standard normal variables. In the first experiment, we simulated 1,000 samples based on the parameters that are presented in the second column of Table 2. To ensure simplicity in the calculations, it is assumed that $b_{21}, b_{22} = 0$.

Table 1 contains the descriptive statistics of the simulated data and Figure 1 shows the simulated series. The parameters are estimated using Gibbs algorithm, which was discussed in Section 4. We have used the uniform priors. The number of iterations for the Gibbs algorithm was set to 10,000. The initial 5,000 draws are considered as the warm-up phase and discarded.

Table 2 gives the posterior means and standard deviations based on the Gibbs sampling. The posterior means are considered as the estimates of the parameters while the standard deviation is a measure of the Gibbs sampling variability. We also computed the biases of the estimates. The reported results show that the biases and standard deviations are small in general. By changing the priors, one may get more or less bias and standard deviation. So, from the Bayesian viewpoint, the bias and standard deviation is not important (Bauwens, Preminger, and Rombouts 2010). The diagrams at the top of Figure 2 display the estimated posterior densities of p_{11} and p_{22}, while the lower diagram shows the estimated conditional transition probabilities to the second state (high-volatility state), $\psi_2^{(t)}$ computed by (4.2). Matrix Q is calculated as

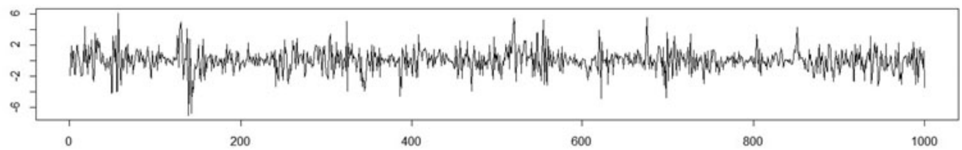

Figure 1. Simulated time series of MSST-HYGARCH model.

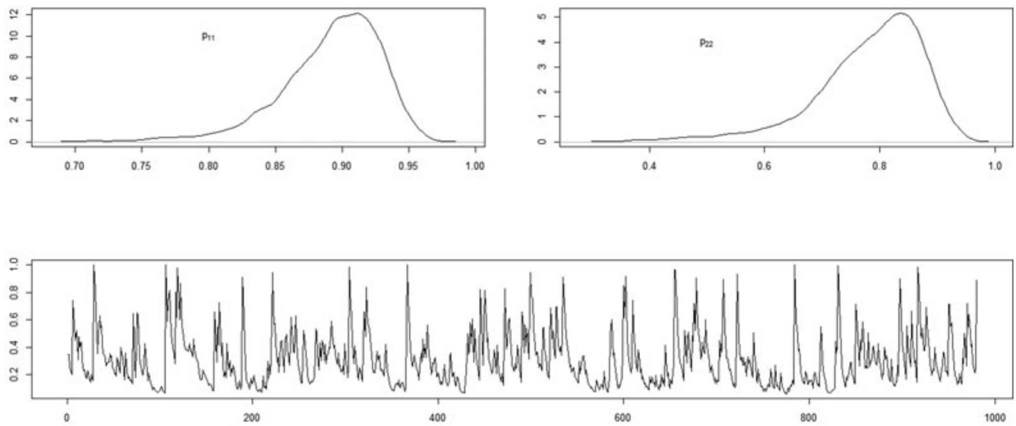

Figure 2. (Up): Estimated posterior density of the p_{11}(left) and p_{22}(right) for simulated data; (Bottom): The estimated conditional transition probabilities to the second state (high-volatility state), $\psi_2^{(t)}$ computed by (4.2).

$$\begin{pmatrix} 0.094 & 0.067 & 0.170 & 0.120 & 0.119 & 0.084 & 0.401 & 0.390 \\ 0.049 & 0.132 & 0.060 & 0.160 & 0.027 & 0.072 & 0.027 & 0.031 \\ 0.212 & 0.150 & 0.170 & 0.120 & 0 & 0 & 0 & 0 \\ 0.052 & 0.140 & 0.060 & 0.160 & 0 & 0 & 0 & 0 \\ 0.307 & 0.216 & 0 & 0 & 0.119 & 0.084 & 0.401 & 0.390 \\ 0.102 & 0.272 & 0 & 0 & 0.027 & 0.072 & 0.027 & 0.031 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\vartheta(Q) = 0.90$; so according to theorem 1, the model is stable. The second experiment is designed to study the numerical behavior of the second-order moment (M_2), third-order moment (M_3), forth-order moment (M_4) and also kurtosis (K)[1] of MSST-HYGARCH model in comparison to GARCH and HYGARCH models. Hence we have generated samples from MSST-HYGARCH model considered in the first experiment, HYGARCH model defined in (2.1)–(2.4) with $(a_0, a_1, a_2, b_0, b_1, b_2, d, \gamma)' = (.4, .35, .3, .5, .25, .2, .8, .7)'$ when $\{\epsilon_t\}$ are standard normal variables; and also GARCH model defined as

$$\begin{aligned} y_t &= \epsilon_t \sqrt{h_t} \\ h_t &= a_0 + a_1 h_{t-1} + a_2 y_{t-1}^2 \end{aligned}$$

where $\{\epsilon_t\}$ are *iid* standard normal variables and $(a_0, a_1, a_2)' = (1, .4, .35)'$. We have computed M_2, M_3, M_4 and K sequences for generated samples with different sizes from

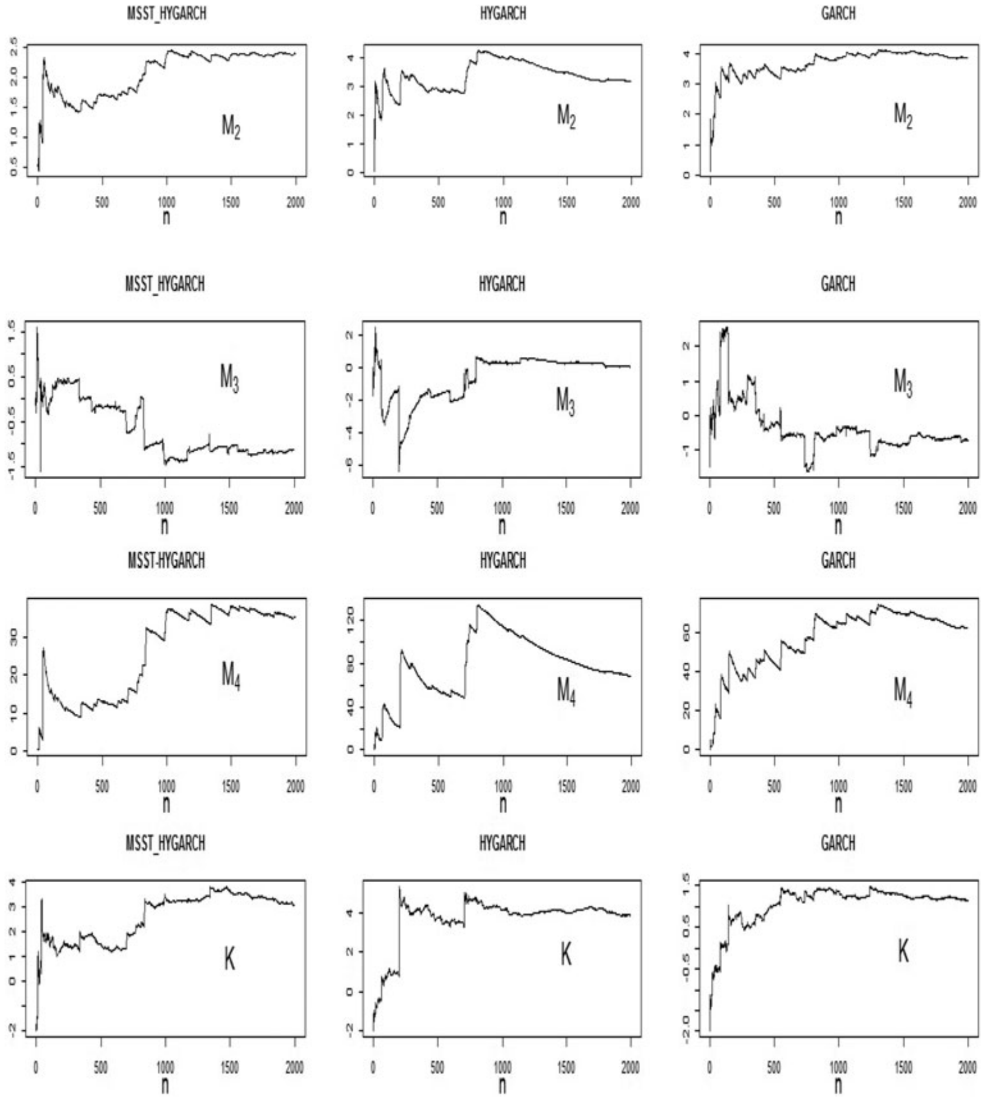

Figure 3. The M_2, M_3, M_4 and K sequences are plotted for MSST-HYGARCH, HYGARCH and GARCH models based on simulated samples with sizes from $1, \ldots, 2000$.

1 to 2000. The results are plotted in Figure 3. It is obvious that all of the sequences are converge but with different rates. M_2 sequence for GARCH model converge more quickly than MSST-HYGARCH and HYGARCH models. It seems M_3 sequence of HYGARCH model converge sooner than other models. Although the M_4 sequence of GARCH model converge sooner than MSST-HYGARCH but it vary in a larger rang. Finally the K sequence converge more quickly for GARCH model.

The third experiment is conducted to show that the MSST-HYGARCH as a long-memory process significantly improves forecasting performance in comparison to MSST-GARCH. For this, 1000 samples are simulated from long-memory FIGARCH model defined as

Table 3. Estimation results of fitting MSST-HYGARCH and MSST-GARCH models on simulated observations from FIGARCH model.

	MSST-HYGARCH	MSST-GARCH
a_{01}	0.206 (0.050)	0.323 (0.049)
a_{11}	0.255 (0.049)	0.303 (0.049)
a_{21}	0.204 (0.049)	0.304 (0.050)
b_{01}	0.206 (0.048)	0.305 (0.049)
b_{11}	0.162 (0.024)	0.254 (0.049)
b_{21}	0.063 (0.025)	0.255 (0.049)
d_1	0.656 (0.048)	—
γ_1	0.530 (0.247)	0.531 (0.241)
a_{02}	0.609 (0.098)	0.756 (0.098)
a_{12}	0.403 (0.049)	0.405 (0.048)
a_{22}	0.354 (0.049)	0.356 (0.049)
b_{02}	0.609 (0.099)	0.608 (0.074)
b_{12}	0.202 (0.025)	0.404 (0.049)
b_{22}	0.102 (0.025)	0.356 (0.049)
d_2	0.856 (0.049)	—
γ_2	1.526 (0.245)	1.525 (0.247)
p_{11}	0.645 (0.162)	0.769 (0.137)
p_{22}	0.975 (0.024)	0.973 (0.019)

Standard deviations are given in parentheses.

$$
\begin{aligned}
y_t &= \epsilon_t \sqrt{h_t} \\
h_t &= b_0 + b_1 h_{t-1} + \left[1 - b_1 B - (1 - b_2 B)(1 - B)^d\right] y_t^2,
\end{aligned}
$$

where $\{\epsilon_t\}$ are *iid* standard normal variables and $(b_0, b_1, b_2, d)' = (.4, .3, .2, .8)'$. Then we have fitted the MSST-HYGARCH, defined by (2.7)–(2.11), and MSST-GARCH defined as:

$$
y_t = \sqrt{h_{t,Z_t}} \epsilon_t
$$

where $\{Z_t\}$ is a Markov chain with values $\{1, 2, \ldots, m\}$ and $\{\epsilon_t\}$ are *iid* standard normal variables, independent of $\{Z_t\}$. The conditional variances $h_{t,j}, j = 1, 2, \ldots, m$ follow the recursive equations:

$$
h_{t,j} = (1 - w_{t,j}) h_{1,t,j} + w_{t,j} h_{2,t,j}
$$

where

$$
\begin{aligned}
h_{1,t,j} &= a_{0j} + a_{1j} h_{1,t-1,j} + a_{2j} y_{t-1}^2 \\
h_{2,t,j} &= b_{0j} + b_{1j} h_{2,t-1,j} + b_{2j} y_{t-1}^2 \\
w_{t,j} &= \frac{\exp\left(-\gamma_j y_{t-1}\right)}{1 + \exp\left(-\gamma_j y_{t-1}\right)}
\end{aligned}
$$

where $a_{0j}, b_{0j} > 0, 0 < a_{1,j} + a_{2j} < 1$ and $0 < b_{1,j} + b_{2j} < 1$. The estimated parameters are given in Table 3. In order to evaluate the goodness-of-fit of the models in forecasting performance we compare AIC of the fitted models. The MSST-HYGARCH model is favored with the AIC value 3713.2 to the MSST-GARCH with AIC value 3716.9.

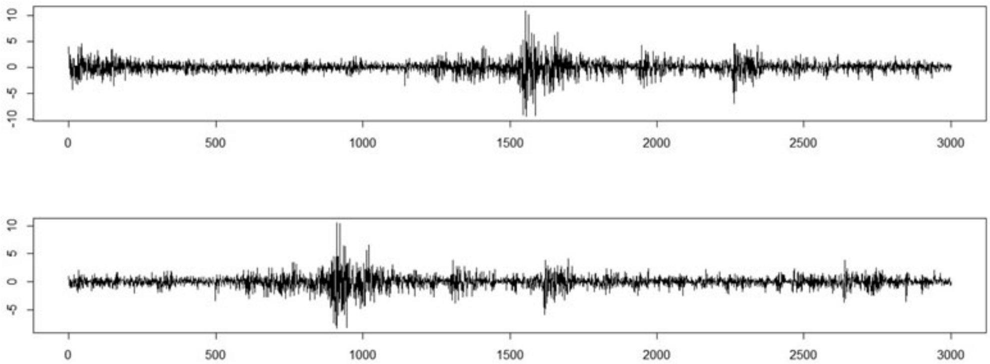

Figure 4. (Up): Percentage log returns of $S\&P500$ daily log returns. (Bottom): Percentage log returns of DJIA daily log returns.

8. Empirical data

In this section, we apply the MSST-HYGARCH model as well as the ST-HYGARCH and HYGARCH models on the daily percentage log returns of the $S\&P500$ indices from August 13, 2002 to July 15, 2014 (3000 observations) and also Dow Jones industrial average (DJIA) indices from March 3, 2005 to February 1, 2017 (3000 observations). Figure 4 presents the sample path of data, which show evidence of two states, where the first state is associated with low volatilities and the second state relates to high volatilities. Table 4 includes the descriptive statistics of the $S\&P500$ and DJIA indices. We observe the negative skewness and excess kurtosis of these returns. The whole sample is divided into two parts. The first part contains 2,000 observations and is used as in-sample data to conduct model estimation. The second part is used as out-of-sample data to evaluate model forecasting. Evidence in the literature shows returns tend to be leptokurtic. Thus we let $\{\epsilon_t\}$ follow the standardized t-distribution with κ degree of freedom. Three models are then applied to the first part of data. Using Section 4, the parameters of the models are estimated and the results are reported in Tables 5 and 6. The value of γ_1 shows the speed of transition from the short memory component to the long memory component in the low-volatility state to be smaller than the value of γ_2, which shows this specification in the high-volatility state. Also Tables 5 and 6 show that the Estimated transition probabilities (p_{11} and p_{22}) are close to one which indicate less switch between regimes. Figure 5 displays the estimated conditional transition probabilities to the second state (high-volatility state), $\psi_2^{(t)}$ computed by (4.2). To evaluate the performance of the different models in computing true conditional variances that are measured by squared returns, we calculated the root mean squared error (RMSE) and the log likelihood value (LLV) for in-sample and out-of-sample data. As out-of-sample performance, the one-day-ahead forecasts are computed using estimated models. The results are given in Tables 7 and 8; the results show that the MSST-HYGARCH has the best fitting and forecasting to data. For in-sample data of the $S\&P500$ the HYGARCH model has the worst performance. The MSST-HYGARCH model outperforms the ST-HYGARCH model, and has a lower RMSE and a higher LLV. In out-of-sample data the MSST-HYGARCH and HYGARCH models have a same RMSE but the MSST-HYGARCH model has the larger LLV. In the DJIA data MSST-HYGARCH model has the best performance in fitting and forecasting task based on LLV and RMSE.

Table 4. Descriptive statistics of *S&P*500 and DJIA daily log returns.

Series	Mean	Std.dev	Minimum	Maximum	Skewness	Kurtosis
S&P	0.027	1.267	−9.469	10.957	−0.275	10.045
DJIA	0.020	1.137	−8.200	10.508	−0.097	10.891

Table 5. Estimation results of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on *S&P*500 daily log returns.

	MSST-HYGARCH		ST-HYGARCH	HYGARCH
a_{01}	0.205 (0.049)	a_0	0.415 (0.124)	0.418 (0.149)
a_{11}	0.255 (0.048)	a_1	0.313 (0.098)	0.311 (0.098)
a_{21}	0.205 (0.049)	a_2	0.310 (0.099)	0.310 (0.098)
b_{01}	0.205 (0.050)	b_0	0.361 (0.123)	0.362 (0.123)
b_{11}	0.162 (0.025)	b_1	0.228 (0.037)	0.405 (0.049)
b_{21}	0.062 (0.024)	b_2	0.083 (0.035)	0.229 (0.037)
d_1	0.655 (0.048)	d	0.758 (0.074)	0.658 (0.074)
γ_1	0.266 (0.119)	γ	0.264 (0.123)	—
κ_1	6.110 (1.461)	κ	6.148 (1.483)	6.146 (1.484)
a_{02}	0.610 (0.098)	w	—	0.552 (0.235)
a_{12}	0.406 (0.049)		—	
a_{22}	0.353 (0.048)		—	
b_{02}	0.609 (0.098)		—	
b_{12}	0.203 (0.024)		—	
b_{22}	0.102 (0.024)		—	
d_2	0.855 (0.049)		—	
γ_2	2.029 (0.246)		—	
κ_2	7.201 (1.467)		—	
p_{11}	0.997 (0.003)		—	
p_{22}	0.970 (0.071)		—	

Standard deviations are given in parentheses.

Table 6. Estimation results of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA daily log returns.

	MSST-HYGARCH		ST-HYGARCH	HYGARCH
a_{01}	0.205 (0.049)	a_0	0.405 (0.124)	0.419 (0.149)
a_{11}	0.256 (0.049)	a_1	0.312 (0.099)	0.312 (0.098)
a_{21}	0.204 (0.049)	a_2	0.283 (0.086)	0.309 (0.099)
b_{01}	0.205 (0.049)	b_0	0.381 (0.122)	0.362 (0.123)
b_{11}	0.163 (0.024)	b_1	0.229 (0.037)	0.405 (0.049)
b_{21}	0.063 (0.024)	b_2	0.083 (0.034)	0.229 (0.037)
d_1	0.656 (0.049)	d	0.758 (0.074)	0.658 (0.074)
γ_1	0.527 (0.245)	γ	0.264 (0.124)	—
κ_1	6.157 (1.466)	κ	6.149 (1.484)	6.145 (1.484)
a_{02}	0.612 (0.049)	w	—	0.552 (0.235)
a_{12}	0.405 (0.049)		—	
a_{22}	0.356 (0.049)		—	
b_{02}	0.611 (0.049)		—	
b_{12}	0.202 (0.024)		—	
b_{22}	0.103 (0.024)		—	
d_2	0.854 (0.049)		—	
γ_2	1.521 (0.246)		—	
κ_2	7.119 (1.467)		—	
p_{11}	0.996 (0.003)		—	
p_{22}	0.948 (0.100)		—	

Standard deviations are given in parentheses.

The MSST-HYGARCH model has the potential to make a continuum of regimes in each state using the ST structure; Thus presenting two states would be enough as further transition can be evaluated by the smooth transition of GARCH and FIGARCH

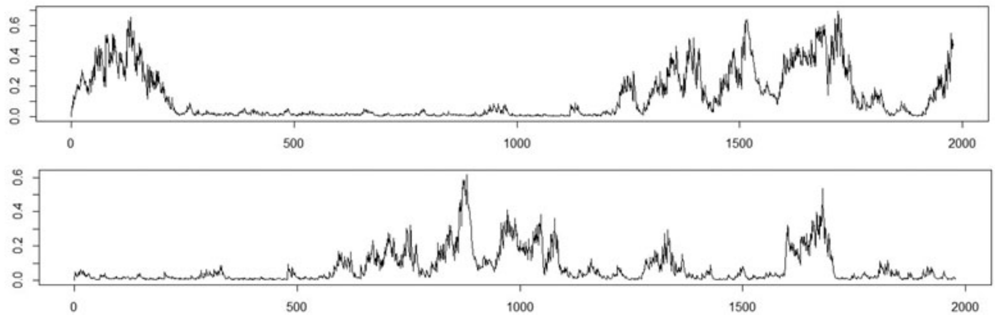

Figure 5. Estimated conditional transition probabilities to the second states (the high-volatility state) of the fitted MSST-HYGARCH, $\psi_2^{(t)}$. (Up): For the $S\&P500$ daily log returns. (Bottom): For the DJIA daily log returns.

volatiliies from time-dependent convex combination between extremes in each state. Hence there is no need to assume multiple states apriori which is subject to over or under-fitting the states. To clarify the out-performance of the MSST-HYGARCH model, we plot the forecasting conditional variances and true conditional variances for some of the data in Figures 6 and 7. When the level of the true conditional variances changes, the MSST-HYGARCH perceives this matter very well and switches from the low-volatility (high-volatility) state to the high-volatility (low-volatility) state. Hence, the MSST-HYGARCH model is more flexible than the HYGARCH and ST-HYGARCH models in accommodating different degrees of memory and different sizes of shocks. For $S\&P500$ data matrix Q is calculated as

$$\begin{pmatrix} 0.456 & 0.015 & 0.254 & 0.008 & 0.161 & 0.005 & 0.143 & 0.025 \\ 0.001 & 0.339 & 0.001 & 0.394 & 0.001 & 0.196 & 0.001 & 0.024 \\ 0.204 & 0.006 & 0.254 & 0.008 & 0 & 0 & 0 & 0 \\ 0.001 & 0.342 & 0.001 & 0.394 & 0 & 0 & 0 & 0 \\ 0.690 & 0.021 & 0 & 0 & 0.161 & 0.005 & 0.143 & 0.025 \\ 0.002 & 0.742 & 0 & 0 & 0.001 & 0.196 & 0.001 & 0.024 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\vartheta(Q) = 0.917$ and for DJIA data

$$\begin{pmatrix} 0.388 & 0.020 & 0.255 & 0.013 & 0.162 & 0.008 & 0.239 & 0.054 \\ 0.002 & 0.388 & 0.002 & 0.384 & 0.001 & 0.191 & 0.001 & 0.020 \\ 0.203 & 0.011 & 0.255 & 0.013 & 0 & 0 & 0 & 0 \\ 0.001 & 0.337 & 0.002 & 0.384 & 0 & 0 & 0 & 0 \\ 0.591 & 0.031 & 0 & 0 & 0.162 & 0.008 & 0.239 & 0.054 \\ 0.003 & 0.726 & 0 & 0 & 0 & 0.191 & 0.001 & 0.020 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\vartheta(Q) = 0.920$. So, according to Theorem 1, the estimated MSST-HYGARCH models are stable. Based on the out-of-sample data, one-day-ahead VaR forecasts at a level risk of $\rho = 0.05, 0.10$ for all models are calculated and the accuracy tests that are discussed in Section 6 are performed. The results are reported in Tables 9 and 10.

Table 7. Measures of performance of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on *S*&*P*500 daily log returns.

	In-sample		Out-of-sample	
Model	RMSE	LLV	RMSE	LLV
MSST-HYGARCH	2.969	−2949.895	2.505	−1244.690
ST-HYGARCH	3.343	−3017.876	2.541	−1301.528
HYGARCH	3.457	−3027.765	2.505	−1318.688

Table 8. Measures of performance of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA daily log returns.

	In-sample		Out-of-sample	
Model	RMSE	LLV	RMSE	LLV
MSST-HYGARCH	5.339	−2817.593	1.169	−1083.905
ST-HYGARCH	5.380	−2903.266	1.235	−1169.072
HYGARCH	5.413	−2922.748	1.257	−1194.636

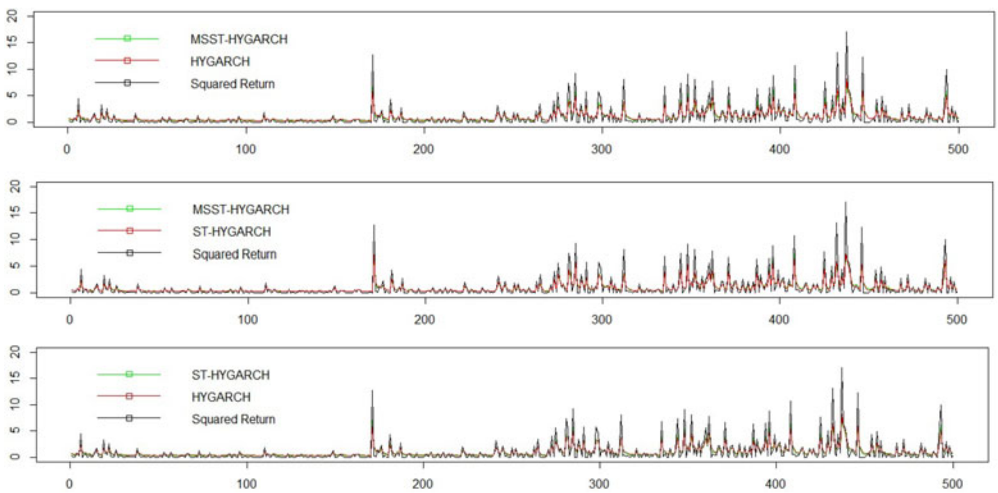

Figure 6. (Up): Forecasting conditional variances of the MSST-HYGARCH and HYGARCH models for some of the *S*&*P*500 daily log returns. (Middle): Forecasting conditional variances of the MSST-HYGARCH and ST-HYGARCH models for some of the *S*&*P*500 daily log returns. (Bottom): Forecasting conditional variances of the ST-HYGARCH and HYGARCH models for some of the *S*&*P*500 daily log returns.

Where the first and second rows show the number of expected exceptions (Ex.e) and empirical exceptions (Em.e) respectively. For the *S*&*P*500 data, it can be seen that at 5% and 10% significance levels, the LR_{UC} is rejected one time for MSST-HYGARCH and ST-HYGARCH models and two times for HYGARCH model. The LR_{IND} is rejected one time for ST-HYGARCH model and is accepted two times for MSST-HYGARCH and HYGARCH models. The LR_{CC} is rejected one time for MSST-HYGARCH and ST-HYGARCH models and two times for HYGARCH model. For the DJIA data, it can be seen that at 5% and 10% significance levels, the LR_{UC} is accepted two times for MSST-HYGARCH and rejected two times for ST-HYGARCH and HYGARCH models. The LR_{IND} is accepted two times for all models. The LR_{CC} is accepted two times for MSST-HYGARCH and rejected two times for ST-HYGARCH and HYGARCH models. For

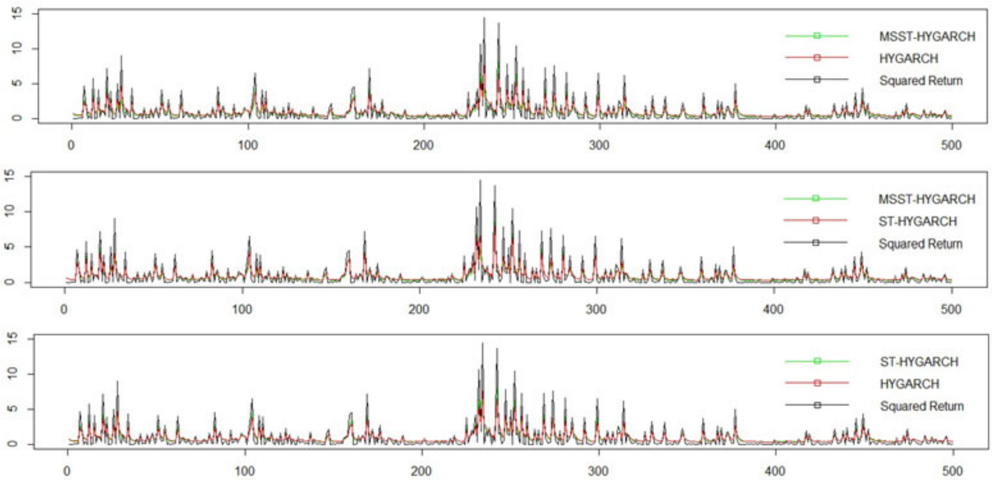

Figure 7. (Up): Forecasting conditional variances of the MSST-HYGARCH and HYGARCH models for some of the DJIA daily log returns. (Middle): Forecasting conditional variances of the MSST-HYGARCH and ST-HYGARCH models for some of the DJIA daily log returns. (Bottom): Forecasting conditional variances of the ST-HYGARCH and HYGARCH models for some of the DJIA daily log returns.

Table 9. VaR forecasting for MSST-HYGARCH, ST-HYGARCH and HYGARCH models on $S\&P500$ daily log returns at level $\rho = 0.05, 0.10$.

	MSST-HYGARCH		ST-HYGARCH		HYGARCH	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
Ex.e	50	100	50	100	50	100
Em.e	69	102	46	77	37	70
LR_{UC}	6.830	0.044*	0.346*	6.332	3.895	11.054
LR_{IND}	0.997*	2.870*	4.534	0.341*	2.922*	0.147*
LR_{CC}	7.827	2.914*	4.882*	6.674	6.818	11.202

Notes: At the 5% significance level the critical value of the LR_{UC} and LR_{IND} is 3.84 and for LR_{CC} is 5.99. "*" indicates that the model passes the test at 5% significance level.

evaluating the performance of MSST-HYGARCH, ST-HYGARCH and HYGARCH which we refer them by 1,2 and 3 respectively, the $S^a_{i,j}$ for $i, j = 1, 2, 3$ defined at subsection 6.4 are calculated. According to the results demonstrated in Table 11 the H_0 is rejected at the 5% significance level for all cases. So there is no significant different between competing models in terms of quadratic loss functions.

As investigate the moments of the fitted model, we let $\{\epsilon_t\}$ follow the *iid* standard normal variables. The daily percentage log returns of the $S\&P500$ indices from February 17, 2009 to February 6, 2013 (1000 observations) are considered. Table 12 includes the descriptive statistics of the data. The MSST-HYGARCH, ST-HYGARCH and HYGARCH models are fitted on data. The estimated noises of data from fitted models, $\widehat{\epsilon_t} = \frac{y_t}{\sqrt{\widehat{h_t}}}$ are calculated. Table 13 reports the M_2, M_3, M_4 and K for etimated noises from different models; the first row corresponds to the standard normal variable. By inspection of Table 13, the M_2 from MSST-HYGARCH is closer to standard normal variable. But the M_3, M_4 and K from HYGARCH are closer to standard normal variable.

Table 10. VaR forecasting for MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA daily log returns at level $\rho = 0.05, 0.10$.

	MSST-HYGARCH		ST-HYGARCH		HYGARCH	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
Ex.e	50	100	50	100	50	100
Em.e	52	90	34	57	28	50
LR_{UC}	0.083*	1.146*	6.043	23.940	12.036	33.413
LR_{IND}	0.141*	0.190*	0.931*	1.041*	0.114*	0.975*
LR_{CC}	0.224*	1.336*	6.136	24.982	12.150	34.370

Notes: At the 5% significance level the critical value of the LR_{UC} and LR_{IND} is 3.84 and for LR_{CC} is 5.99. "*" indicates that the model passes the test at 5% significance level.

Table 11. Diebold and Mariano statistic.

	*S&P*500		DJIA	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
$S^a_{1,2}$	−27.385	−25.484	−28.397	−26.057
$S^a_{1,3}$	−27.258	−25.298	−28.397	−26.057
$S^a_{2,3}$	−28.776	−27.006	−29.598	−28.271
$S^a_{2,1}$	−31.496	−31.243	−29.851	−29.093
$S^a_{3,1}$	−31.622	−31.433	−29.978	−29.409
$S^a_{3,2}$	−31.559	−31.370	−31.622	−31.559

Table 12. Descriptive statistics of 1000 observations from *S&P*500 daily log returns.

Mean	Std.dev	Minimum	Maximum	Skewness	Kurtosis
0.065	1.246	−6.896	6.837	−0.118	3.619

Table 13. Calculating M_2, M_3, M_4 and K for estimated noises from fitting MSST-HYGARCH, ST-HYGARCH and HYGARCH models on 1000 observations of *S&P*500 daily log returns when $\{\epsilon_t\}$ follow *iid* standard normal variables; the first row corresponds to the standard normal variable.

	M_2	M_3	M_4	K
Standard Normal	1	0	3	0
MSST-HYGARCH	0.994	−0.338	4.227	1.280
ST-HYGARCH	1.094	−0.405	5.483	1.573
HYGARCH	0.940	−0.300	3.7618	1.257

To survey the effect of skewed t-distribution, we consider daily percentage log returns of the DJIA indices from April 29, 2011 to April 24, 2015 (1000 observations) and US/Euro exchange rates from September 8, 2013 to September 6, 2017 (1000 observations). Table 14 gives the descriptive statistics for the data. The first 500 observations are used as in-sample data and the second 500 observations are used as out-of-sample data. Then, we let $\{\epsilon_t\}$ follow the skewed t-distribution (Hansen 1994) with κ degree of freedom and $\imath$ skewed parameter. The estimated parameters are reported in Tables 15 and 16. Table 17 shows in-sample fitting results. By inspection of Table 17, the MSST-HYGARCH model has the best performance. Using the constant rolling window procedure with window size 100, the ond-day-ahead forecasts are calculated as out-of-sample performance. Tables 18 and 19 give the VaR forecasts and related tests. Tables 18 and 19 demonstrate the MSST-HYGARCH model gives more accurate VaR forecasts. Diebold and Mariano statistic for different models is computed and the results are reported in Table 20; obviously from Table 20, the H_0 is rejected at the 5% significance level for all cases.

Table 14. Descriptive statistics of DJIA and US/Euro exchange rates daily log returns (1000 observations).

Series	Mean	Std.dev	Minimum	Maximum	Skewness	Kurtosis
DJIA	0.034	0.906	−5.740	4.119	−0.469	4.464
US/Euro	−0.010	0.558	−2.662	3.075	0.077	2.321

Table 15. Estimation results of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA daily log returns (500 observations).

	MSST-HYGARCH		ST-HYGARCH	HYGARCH
a_{01}	0.204 (0.050)	a_0	0.405 (0.123)	0.418 (0.148)
a_{11}	0.256 (0.048)	a_1	0.307 (0.099)	0.306 (0.099)
a_{21}	0.203 (0.049)	a_2	0.283 (0.083)	0.309 (0.096)
b_{01}	0.204 (0.049)	b_0	0.385 (0.122)	0.366 (0.122)
b_{11}	0.162 (0.025)	b_1	0.229 (0.037)	0.406 (0.049)
b_{21}	0.063 (0.024)	b_2	0.083 (0.034)	0.229 (0.036)
d_1	0.655 (0.049)	d	0.758 (0.073)	0.658 (0.072)
γ_1	0.541 (0.240)	γ	0.261 (0.122)	—
κ_1	2.056 (0.491)	κ	2.027 (0.491)	2.027 (0.492)
ι_1	−0.239 (0.120)	ι	−0.239 (0.120)	−0.238 (0.120)
a_{02}	0.607 (0.047)	w	—	0.549 (0.231)
a_{12}	0.406 (0.049)		—	
a_{22}	0.356 (0.049)		—	
b_{02}	0.609 (0.096)		—	
b_{12}	0.202 (0.024)		—	
b_{22}	0.102 (0.025)		—	
d_2	0.856 (0.049)		—	
γ_2	1.529 (0.243)		—	
κ_2	2.046 (0.480)		—	
ι_2	−0.236 (0.119)		—	
p_{11}	0.983 (0.023)		—	
p_{22}	0.779 (0.167)		—	

Standard deviations are given in parentheses.

Table 16. Estimation results of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on US/Euro daily log returns (500 observations).

	MSST-HYGARCH		ST-HYGARCH	HYGARCH
a_{01}	0.207 (0.049)	a_0	0.343 (0.113)	0.369 (0.123)
a_{11}	0.254 (0.050)	a_1	0.308 (0.099)	0.255 (0.074)
a_{21}	0.204 (0.048)	a_2	0.283 (0.083)	0.310 (0.096)
b_{01}	0.207 (0.049)	b_0	0.342 (0.109)	0.316 (0.097)
b_{11}	0.162 (0.025)	b_1	0.283 (0.062)	0.365 (0.049)
b_{21}	0.061 (0.024)	b_2	0.083 (0.034)	0.229 (0.036)
d_1	0.760 (0.073)	d	0.658 (0.073)	0.658 (0.072)
γ_1	0.532 (0.241)	γ	0.314 (0.146)	—
κ_1	3.045 (0.496)	κ	2.528 (0.491)	2.528 (0.491)
ι_1	0.111 (0.046)	ι	0.108 (0.047)	0.108 (0.047)
a_{02}	0.458 (0.073)	w	—	0.551 (0.231)
a_{12}	0.406 (0.050)		—	
a_{22}	0.354 (0.048)		—	
b_{02}	0.459 (0.073)		—	
b_{12}	0.202 (0.024)		—	
b_{22}	0.102 (0.024)		—	
d_2	0.760 (0.073)		—	
γ_2	1.524 (0.246)		—	
κ_2	3.056 (0.494)		—	
ι_2	0.109 (0.046)		—	
p_{11}	0.994 (0.006)		—	
p_{22}	0.774 (0.156)		—	

Standard deviations are given in parentheses.

Table 17. Measures of performance of MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA and US/Euro daily log returns (500 observations).

	DJIA		US/Euro	
Model	RMSE	LLV	RMSE	LLV
MSST-HYGARCH	1.392	−666.358	0.628	−444.471
ST-HYGARCH	1.520	−686.317	0.713	−500.028
HYGARCH	1.311	−679.870	0.727	−502.478

Table 18. VaR forecasting for MSST-HYGARCH, ST-HYGARCH and HYGARCH models on DJIA daily log returns at level $\rho = 0.05, 0.10$ using constant rolling window proceduar (500 observations).

	MSST-HYGARCH		ST-HYGARCH		HYGARCH	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
Ex.e	25	50	25	50	25	50
Em.e	18	32	14	26	14	25
LR_{UC}	2.276*	8.147	6.018	15.254	6.017	16.706
LR_{IND}	1.420*	0.134*	0.865*	0.409*	0.865*	0.163*
LR_{CC}	3.697*	8.282	6.883	15.663	6.883	16.869

Notes: At the 5% significance level the critical value of the LR_{UC} and LR_{IND} is 3.84 and for LR_{CC} is 5.99. "*" indicates that the model passes the test at 5% significance level.

Table 19. VaR forecasting for MSST-HYGARCH, ST-HYGARCH and HYGARCH models on US/Euro daily log returns at level $\rho = 0.05, 0.10$ using constant rolling window proceduar (500 observations).

	MSST-HYGARCH		ST-HYGARCH		HYGARCH	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
Ex.e	25	50	25	50	25	50
Em.e	18	32	6	14	3	10
LR_{UC}	2.276*	8.147	21.642	39.163	32.281	51.265
LR_{IND}	1.420*	0.134*	0.170*	0.865*	0.048*	0.449*
LR_{CC}	3.697*	8.282	21.794	40.0283	32.330	51.715

Notes: At the 5% significance level the critical value of the LR_{UC} and LR_{IND} is 3.84 and for LR_{CC} is 5.99. "*" indicates that the model passes the test at 5% significance level.

Table 20. Diebold and Mariano statistic using constant rolling window proceduar (500 observations).

	DJIA		US/Euro	
	VaR (0.05)	VaR (0.10)	VaR (0.05)	VaR (0.10)
$S^a_{1,2}$	−21.108	−20.125	−21.198	−20.482
$S^a_{1,3}$	−21.108.	−20.125	−22.360	−21.108
$S^a_{2,3}$	−21.108	−20.035	−21.824	−21.108
$S^a_{2,1}$	−22.003	−21.734	−22.271	−22.271
$S^a_{3,1}$	−22.003	−21.734	−20.392	−22.361
$S^a_{3,2}$	−22.360	−22.360	−22.360	−22.360

9. Conclusion

MSST-HYGARCH has the potential to consider low volatility and high volatility as two clusters based on all past information and also determines the smooth transition weights between short and long memory based on the preceding observation. This model offers much better description of the dynamic volatilities, and exploits a smooth transition structure to create time-varying convex combination of the short memory GARCH and long memory FIGARCH in each state.The transition probabilities between states are calculated using all past observations. The necessary and sufficient asymptotically stability condition is derived. The simulation study showed that Gibbs sampling provides credible estimates of the parameters. The empirical example of some periods of $S\&P500$ and Dow Jones industrial average indices showed that the MSST-HYGARCH model gives better forecasting of volatilities and more accurate VaR than the ST-HYGARCH and HYGARCH.

Note

1. $K = \frac{M_4}{M_2^2} - 3$, where $M_r = E(y^r)$.

ORCID

Saeid Rezakhah http://orcid.org/0000-0003-4466-0492

References

Abramson, A., and I. Cohen. 2007. On the stationary of Markov-switching GARCH processes. *Econometric Theory* 23:485–500.

Alemohammad, N., S. Rezakhah, and S. H. Alizadeh. 2016. Markov switching component GARCH model theory and methods. *Communication in Statistics-Theory and Methods* 45 (15): 4332–48. doi:10.1080/03610926.2013.841934.

Amado, C., and T. Teräsvirta. 2008. Modelling conditional and unconditional heteroscedasticity with smoothly time-varying structure. CREATES; Research Paper. http://ssrn.com/abstract=1148141.

Ardia, D. 2009. Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with student t innovations. *Econometrics Journal* 12 (1):105–26. doi:10.1111/j.1368-423X.2008.00253.x.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen. 1996. Fractionally integrated generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 74 (1):3–30. doi:10.1016/S0304-4076(95)01749-6.

Bauwens, L., A. Preminger, and V. K. Rombouts. 2010. Theory and inference for a Markov switching GARCH model. *Econometrics Journal* 13 (2):218–44. doi:10.1111/j.1368-423X.2009.00307.x.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31:307–27.

Brooks, C., and G. Persand. 2000. Value at risk and market crashes. The Journal of Risk 2 (4): 5–26. doi:10.21314/JOR.2000.029.

Cai, J. 1994. A Markov model of switching-regime ARCH. *Journal of Business and Economic Statistics* 12:309–18. doi:10.2307/1392087.

Chib, S. 1996. Calculating posterior distributions and model estimate in Markov mixture models. *Journal of Econometrics* 75 (1):79–97. doi:10.1016/0304-4076(95)01770-4.

Christoffersen, P. 1998. Evaluating interval forecasts. *International Economic Review* 39 (4): 841–62. doi:10.2307/2527341.

Davidson, J. 2004. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business and Economic Statics* 22 (1):16–29. doi:10.1198/073500103288619359.

Diebold, F. X., and R. Mariano. 1995. Camparing predictive accuracy. *Journal of Business and Economic Statistics* 13 (3):253–63.

Dowd, K. 1998. *Beyond value at risk: The new science of risk management*. Chichester, UK: Wiley.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4):987–1007. doi:10.2307/1912773.

Gelfand, A., and A. Smith. 1990. Sampling based approaches to calculating marginal densities. *Journal of American Statistical Association* 85 (410):398–409. doi:10.2307/2289776.

Gonzales-Rivera, G. 1998. Smooth transition GARCH models. *Studies in Non-linear Dynamics and Econometrics* 3:61–78.

Granger, C. W. J., and T. Teräsvirta. 1993. *Modelling non-linear economic relationships*. Oxford, UK: Oxford University Press.

Gray, S. F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42 (1):27–62. doi:10.1016/0304-405X(96)00875-6.

Haas, M., S. Mittink, and M. S. Paollella. 2004. A new approach to Markov switching GARCH models. *Journal of Financial Econometrics* 2 (4):493–530. doi:10.1093/jjfinec/nbh020.

Hamilton, J. D., and R. Susmel. 1994. Autoregressive conditional heteroscedasticity and changes in regime. *Journal of Econometrics* 64 (1–2):307–33. doi:10.1016/0304-4076(94)90067-1.

Hansen, B. E. 1994. Autoregressive conditional density estimation. *International Economic Review* 35 (3):705–30. doi:10.2307/2527081.

Jorion, P. 1996. *Value at risk: The new benchmark for controlling market risk*. Chicago, IL: Irwin.

Klaassen, F. 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27 (2):363–94. doi:10.1007/s001810100100.

Kwan, W., W. K. Li, and G. Li. 2011. On the threshold hyperbolic GARCH models. *Statistics and Its Interface* 4 (2):159–66. doi:10.4310/SII.2011.v4.n2.a11.

Kupiec, P. 1995. Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives 3 (2):73–84. doi:10.3905/jod.1995.407942.

Lancaster, P., and M. Tismenetsky. 1985. *The theory of matrices*. 2nd ed. Cambridge, MA: Academic press.

Lubrano, M. 2001. Smooth transition GARCH models: A Bayesian approach. Researches. *Economiques Louvain* 67:257–87.

Marcucci, J. 2005. Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 9:1–53.

McAleer, M. 2014. Asymmetry and leverage in conditional volatility models. *Econometrics* 2 (3): 145–50. doi:10.3390/econometrics2030145.

Mohammadi, F., and S. Rezakhah. 2017. Smooth transition HYGARCH model: Stability and testing. http://arxiv.org/pdf/1701.05358.pdf.

Niguez, T., and A. Rubia. 2006. Forecasting the conditional covariance matrix of a portfolio under long run temporal dependence. *Journal of Forecasting* 25:439–58. doi:10.1002/for.997.

Ritter, C., and M. A. Tanner. 1992. Facilitating the Gibbs sampler: The Gibbs stopper and Griddy-Gibbs sampler. *Journal of the American Statistical Association* 87 (419):861–8. doi: 10.2307/2290225.

Robert, C., and G. Casella. 2004. *Monte Carlo statistical methods*. New York, NY: Springer.

Sarma, M., S. Thomas, and A. Shah. 2003. Selection of value-at-risk models. *Journal of Forecasting* 22 (4):337–58. doi:10.1002/for.868.

Tang, T. L., and S. J. Shieh. 2006. Long memory in stock index futures markets: A value-at-risk approach. *Physica A* 366:437–48. doi:10.1016/j.physa.2005.10.017.

Teräsvirta, T. 1998. Modelling economic relationships with smooth transition regressions. In *Handbook of applied economic statistics*. New York, NY: Marcel Dekker.

Wang, Y., C. Wu, and Y. Wei. 2011. Can GARCH-class models capture long memory in WTI crude oil markets? *Economic Modeling* 28 (3):921–7. doi:10.1016/j.econmod.2010.11.002.

Zhang, Y., and S. Nadarajah. 2017. A review of back-testing for value-at-risk. *Communications in Statistics-Theory and methods*, accepted. doi:10.1080/03610926.2017.1361984.