Stat Papers
https://doi.org/10.1007/s00362-018-0992-2

REGULAR ARTICLE

Markov switching asymmetric GARCH model: stability and forecasting

N. Alemohammad[1] · **S. Rezakhah**[2] · **S. H. Alizadeh**[3]

Received: 28 October 2016 / Revised: 16 February 2018

Abstract A new Markov switching asymmetric GARCH model is proposed where each state follows the smooth transition GARCH model, represented by Lubrano (Recherches Economiques de Louvain 67:257–287, 2001), that follows a logistic smooth transition structure between effects of positive and negative shocks. This consideration provides better forecasts than GARCH, Markov switching GARCH and smooth transition GARCH models, in many financial time series. The asymptotic finiteness of the second moment is investigated. The parameters of the model are estimated by applying MCMC methods through Gibbs and griddy Gibbs sampling. Applying the log return of some part of $S\&P$ 500 indices, we show the competing performance of in sample fit and out of sample forecast volatility and value at risk of the proposed model. The Diebold–Mariano test shows that the presented model outperforms all competing models in forecast volatility.

Keywords Markov switching · Leverage effect · Smooth transition · DIC · Bayesian inference · Griddy Gibbs sampling

Mathematics Subject Classification 60J10 · 62M10 · 62F15

✉ S. Rezakhah
rezakhah@aut.ac.ir

1 Department of Mathematics and Computer Science, Shahed University, Tehran, Iran

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

3 Department of Computer Engineering and IT, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Published online: 10 March 2018

1 Introduction

Volatility modeling in financial time series has been widely studied over past few decades. The ARCH and GARCH models, introduced by Engle (1982) and Bollerslev (1986), are surely the most popular classes of volatility models. Hamilton and Susmel (1994) introduced the Markov-Switching GARCH (MS-GARCH) by merging GARCH model with a hidden Markov chain, where each state allows a different GARCH behavior. Such structure improves forecasting of volatilities. Gray (1996), Klaassen (2002) and Haas et al. (2004) proposed some different variants of MS-GARCH models. For further studies on MS-GARCH models, see Abramson and Cohen (2007), Ardia (2009), Alemohammad et al. (2016) and Bauwens et al. (2010).

One restriction of the GARCH model is its symmetry to the sign of past shocks. This is improved by letting the conditional variance to be a function of size and sign of the preceding observation. Such studies started by Black (1976), who investigated asymmetric effects of positive and negative shocks on volatilities. This consideration is important in financial markets as there exists higher volatility in response to bad news (negative shocks) Gonzalez-Rivera (1998). Study of the asymmetric GARCH started by Engle (1990) and continued as the Exponential GARCH (EGARCH) model by Nelson (1991), GJR-GARCH model by Glosten et al. (1993) and Threshold GARCH (TGARCH) model by Zakoian (1994). The other asymmetric structures are smooth transition models introduced by Gonzalez-Rivera (1998), Ardia (2009), Medeiros and Veiga (2009), and Haas et al. (2013).

In this paper, we study some Markov switching GARCH model where the volatility in each regime is coupled with the smooth transition between the effects of negative and positive shocks. More precisely, the presented model considers different smooth transition structure by states where each state describes some time dependent convex combination between asymmetric effects of positive and negative shocks. The new model obviates the absence of asymmetric property in the Markov switching GARCH model and switching between different levels of volatility in the smooth transition GARCH model (STGARCH), presented by Lubrano (2001).

Ardia (2009) considered some MS-GARCH model where the asymmetric effects of volatilities are considered by applying indicator functions of some predefined non-positive thresholds in states. This causes a sudden shift of volatility structure at corresponding threshold in each state. Alemohammad et al. (2016) considered a Markov switching GARCH model where the smooth transition between structures for high and low volatilities are in effect of size of the preceding return. In this paper we study the case where the volatility structure in each state follows some smooth transition between the effects of positive and negative shocks based on the preceding log return. So it is expected to provide much better fitting, especially when smooth transitions between such effects are evident. As such model employs all past observations, we reduce the volume of calculations by proposing a dynamic programming algorithm. We also derive sufficient condition for stability of the model by applying the method of Abramson and Cohen (2007) and Medeiros and Veiga (2009).

We present a simulation example to show the competitive performance of our model in compare to GARCH, MS-GARCH and STGARCH models. Using $S\&P$ 500 indices from 3/01/2005 to 3/11/2014 we show that our model has much better fitting by providing less forecasting error, better performance based on Diebold-Mariano test and value at risk of out-of-sample forecasting of one day ahead volatility, in compare with GARCH, MS-GARCH, EGARCH, GJR-GARCH and STGARCH models. We also show that our model outperforms the competing models for in-sample fit by using the Deviance information criterion.

The Markov switching smooth transition GARCH model is presented in Sect. 2. Section 3 is devoted to the statistical properties of the model. Estimation of the parameters of the model are studied in Sect. 4. Following simulation studies the competing performance of the presented model is shown in Sect. 5. Section 6 is dedicated to the analysis of the efficiency of the proposed model by applying the model to the $S\&P$ 500 indices for 3/01/2005 to 3/11/2014 . Section 7 concludes.

2 Markov switching asymmetric GARCH model

We consider the Markov switching smooth transition GARCH model, in summary MS-STGARCH as

$$y_t = \varepsilon_t \sqrt{H_{Z_t,t}}, \tag{2.1}$$

where $\{\varepsilon_t\}$ is an iid sequence of standard normal variables, and $\{Z_t\}$ an irreducible, aperiodic Markov chain on state space $E = \{1, 2, \ldots, K\}$ with transition probability matrix $P = ||p_{ij}||_{K\times K}$, where $p_{ij} = p(Z_t = j|Z_{t-1} = i)$, $i, j \in \{1, \ldots, K\}$, and stationary probability measure $\pi = (\pi_1, \ldots, \pi_K)'$. Also given that $Z_t = j$, $H_{j,t}$ (the conditional variance of regime j) is defined as

$$H_{j,t} = a_{0j} + y_{t-1}^2(d_{j,t}) + b_j H_{j,t-1}, \tag{2.2}$$

where

$$d_{j,t} = a_{1j}(1 - w_{j,t-1}) + a_{2j} w_{j,t-1}, \tag{2.3}$$

and the weights $(w_{j,t})$ are logistic function of the past observation as

$$w_{j,t-1} = \frac{1}{1 + \exp(-\gamma_j y_{t-1})} \qquad \gamma_j > 0, \quad j = 1, \ldots, K, \tag{2.4}$$

which are monotonically increasing with respect to previous observation and are bounded, $0 < w_{j,t-1} < 1$. The parameter γ_j is called the slope parameter. The weight function $w_{j,t-1}$ goes to one when $y_{t-1} \to +\infty$ and so $d_{j,t}$ tends to a_{2j}. Also it goes to zero when $y_{t-1} \to -\infty$ and so $d_{j,t}$ tends to a_{1j}. Therefore the effect of negative shocks are mainly described by a_{1j} and of positive shocks by a_{2j}, $j = 1, \ldots, K$. As often negative shocks have greater effect on volatilities than positive ones, one could assume that $a_{1j} > a_{2j}$ in each regime. To impose the idea in model building, we recommend to consider a higher prior for a_{1j} in each state and any estimation procedure. Logistic weight functions in states have the potential to describe different speed for smooth transitions which are in effect of γ_j, $j = 1, \ldots, K$ and also different effect limits as a_{1j} and a_{2j}. This enables one to provide a flexible model for describing such

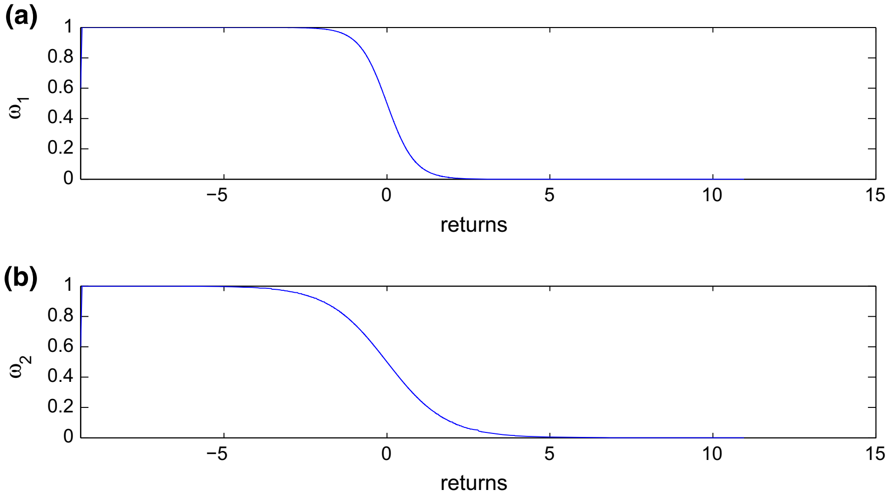

Fig. 1 Logistic function ($w_{j,t}$) for $S\&P$ returns **a** the weight of first regime, **b** the weight of second regime

different transitions. Plots of such logistic weight functions for the returns of $S\&P$ 500 indices, which are studied later in this paper, are presented in Fig. 1. Indeed in each regime the coefficient of y_{t-1}^2 is time dependent that causes the volatility structure being under the influence size and sign of the observations and it makes distinct from GARCH model.

As $\gamma_j \rightarrow \infty$, the logistic weight function considers a step function for positive and negative shocks. When γ_j approaches to zero, $w_{j,t-1}$ goes to $1/2$ and the MS-STGARCH model tends to the MS-GARCH model. In the case of single regime, our model is the STGARCH model that is introduced by Lubrano (2001).

As $\{\varepsilon_t\}$ and $\{Z_t\}$ are assumed independent, sufficient conditions which guarantees the conditional variance (2.2) to be strictly positive are that a_{1j}, a_{2j} and b_j be nonnegative and a_{0j} positive.

3 Statistical properties of the model

In this section, the statistical properties of the MS-STGARCH model are investigated and the conditional density and variance of the process is obtained. As the evaluation of the asymptotic behavior of the second moment in our model isn't so easy to follow, we apply the method of Abramson and Cohen (2007) and Medeiros and Veiga (2009) to obtain an appropriate upper bound for the asymptotic value of the unconditional variance to show its stability.

3.1 Conditional density and variance

Let $\mathcal{I}_t$ be the information up to time t. Following the method of Alemohammad et al. (2016), the conditional density function of y_t given past information can be written as

$$f(y_t|\mathcal{I}_{t-1}) = \sum_{j=1}^{K} \alpha_j^{(t)} \phi\left(\frac{y_t}{\sqrt{H_{j,t}}}\right), \tag{3.1}$$

where $\phi(.)$ is the probability density function of the standard normal distribution and $\alpha_j^{(t)} = p(Z_t = j|\mathcal{I}_{t-1})$, that is obtained in the following lemma.

Lemma 3.1 *The value of* $\alpha_j^{(t)}$ *is obtained recursively by*

$$\alpha_j^{(t)} = \frac{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})\alpha_m^{(t-1)} p_{m,j}}{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})\alpha_m^{(t-1)}}, \tag{3.2}$$

where $p_{mj} = p(Z_t = j|Z_{t-1} = m), m, j = 1, \ldots, K$ *are the transition probabilities and*

$$f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2}) = \phi\left(\frac{y_{t-1}}{\sqrt{H_{m,t-1}}}\right).$$

Proof 3.1 See Appendix A. □

The conditional variance of the MS-STGARCH model is given by

$$Var(Y_t|\mathcal{I}_{t-1}) = \sum_{j=1}^{K} \alpha_j^{(t)} H_{j,t}, \tag{3.3}$$

as $H_{j,t}$ is the conditional variance of jth state. This relation shows that the conditional variance of this model is affected by changes in regime and conditional variance of each state.

As using all past observations for forecasting could increase the complexity of the model, we reduce the volume of calculations by proposing a dynamic programming algorithm. At each time t, $\alpha_i^{(t)}$ (in Eqs. (3.1), (3.3)) can be obtained from a dynamic programming method based on the forward recursion algorithm, proposed in Lemma (3.1).

3.2 Stability

To show the asymptotic wide sense stationarity of the MS-GARCH model, Abramson and Cohen (2007) evaluated unconditional second moment by conditioning on past observations and providing a linear equation between present and past volatilities. They presented a necessary and sufficient condition for the asymptotic wide sense stationarity of the model. For the MS-STGARCH, this method fails as the logistic weights cause that the evaluation of unconditional variance of observations to be so complicated. Lubrano (2001) obtained a recursive relation for the volatility of STGARCH model and by imposing some conditions they showed the stationarity and persistence of the volatility of STGARCH Model. As we have different volatility structure in states of the MS-STGARCH model, their method is not applicable. We show the stability of the model by showing that the second order moments are asymptotically bounded. In this subsection, we investigate the stability of second moment of the MS-STGARCH model. So it would be enough to find an upper bound of the second moment of the process, see Alemohammad et al. (2016). Let M be a positive constant and

$$\boldsymbol{\Omega} = [a_{01} + |a_{21} - a_{11}|M^2, \ldots, a_{0K} + |a_{2K} - a_{1K}|M^2)]', \tag{3.4}$$

be a vector with K component, C denotes a K^2-by-K^2 block matrix as

$$C = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{K1} \\ C_{12} & C_{22} & \cdots & C_{K2} \\ \vdots & & & \vdots \\ C_{1K} & C_{2K} & \cdots & C_{KK} \end{pmatrix} \tag{3.5}$$

where

$$C_{jk} = p(Z_{t-1} = j | Z_t = k)(ue'_j + v), \qquad j, k = 1, \ldots, K, \tag{3.6}$$

$u = [a_{11} + (\delta + \frac{1}{2})|a_{21} - a_{11}|, \ldots, a_{1K} + (\delta + \frac{1}{2})|a_{2K} - a_{1K}|]'$, e_j is a K-by-1 vector that the jth element of it is one and other components are zero, and $v = \| v_{ij} \|_{i,j=1}^{K}$ is a diagonal matrix that $v_{jj} = b_j$ for $j = 1, \ldots, K$.

Let $\Pi = [\pi_1 e'_1, \ldots, \pi_K e'_K]$ and $\rho(A)$ denotes that spectral radius of matrix A. Now we present the following theorem regarding the stability condition of the MS-STGARCH model.

Theorem 3.1 *Let $\{Y_t\}_{t=0}^{\infty}$ follows the MS-STGARCH model, defined by (2.1)–(2.4), the process is asymptotically stable in variance and $\lim_{t\to\infty} E(Y_t^2) \le \Pi'(I - C)^{-1}\dot{\Omega}$, if $\rho(C) < 1$.*

Proof 3.2 See Appendix B. □

4 Estimation

For the estimation of parameters, we apply the Bayesian MCMC method which is extensively used in literature, see Bauwens and Storti (2009), Bauwens et al. (2010), Lubrano (2001) and Ardia (2009). The nonlinearity of Markov switching GARCH models cause the likelihood functions become tricky to maximize, see Lubrano (2001), but MCMC methods avoid the common problem of local maxima encountered in the Maximum likelihood (ML) estimation, see Ardia (2008) (Sect. 7.7) and Miazhynskia and Dorffner (2006). We quote from Ardia (2008) that the Bayesian approach is an adequate framework with less uncertainty in VaR estimates compared to other VaR methods. So by repeating the estimation for each draw in the posterior sample, we obtain an estimation for the VaR density. In the Bayesian framework, we can either integrate out the parameter uncertainty or choose a Bayes point estimate within the VaR density. We follow estimation of parameters by applying MCMC methods through Gibbs and Griddy-Gibbs sampling.

Let $Y_t = (y_1, \ldots, y_t)$ and $Z_t = (z_1, \ldots, z_t)$ be the samples of observations and hidden variables respectively. We consider two states for the model with parameters $\theta = (\theta_1, \theta_2)$, where $\theta_k = (a_{0k}, a_{1k}, a_{2k}, b_k, \gamma_k)$ for $k = 1, 2$ and transition probabilities $\eta = (\eta_{11}, \eta_{12}, \eta_{21}, \eta_{22})$ where $\eta_{ij} = p(z_{t+1} = j | z_t = i)$. The posterior density can be represented as

$$p(\theta, \eta, Z|Y) \propto p(\theta, \eta) p(Z|\theta, \eta) f(Y|\theta, \eta, Z), \tag{4.1}$$

where $Y = (y_1, \ldots, y_T)$, $Z = (z_1, \ldots, z_T)$, T is the total number of samples and $p(\theta, \eta)$ is the prior density. By assuming that the value of z_1 is known, conditional probability mass function of Z given the (θ, η) is independent of θ, so

$$\begin{aligned} p(Z|\theta, \eta) &= p(Z|\eta_{11}, \eta_{22}) \\ &= \prod_{t=1}^{T} p(z_{t+1}|z_t, \eta_{11}, \eta_{22}) \\ &= \eta_{11}^{n_{11}} (1 - \eta_{11})^{n_{12}} \eta_{22}^{n_{22}} (1 - \eta_{22})^{n_{21}}, \end{aligned} \tag{4.2}$$

where $n_{ij} = \#\{z_t = j | z_{t-1} = i\}$ (the number of transitions from regime i to regime j). The conditional density function of Y given the realization of Z and the parameters is factorized in the following way:

$$f(Y|\eta, \theta, Z) = \prod_{t=1}^{T} f(y_t|\theta, z_t = k, Y_{t-1}), \quad k = 1, 2, \tag{4.3}$$

where the one step ahead predictive densities are:

$$f(y_t|\theta, z_t = k, Y_{t-1}) = \frac{1}{\sqrt{2\pi H_{k,t}}} \exp\left(-\frac{y_t^2}{2H_{k,t}}\right). \tag{4.4}$$

Since the straight sampling from the posterior density (4.1) is not possible, we apply the Gibbs sampling algorithm for three blocks: θ, η and Z.
In implementing Gibbs algorithm, we consider the superscript (r) on a parameter to denote its value at the rth iteration of the algorithm. At any iteration of the algorithm, three steps are considered:

(i) Draw the random sample of the state variable $Z^{(r)}$ given , $\eta^{(r-1)}$, $\theta^{(r-1)}$.
(ii) Draw the random sample of the transition probabilities $\eta^{(r)}$ given $Z^{(r)}$.
(iii) Draw the random sample of the $\theta^{(r)}$ given $Z^{(r)}$ and $\eta^{(r)}$.

These steps are repeated until the convergence is obtained. In what follows the sampling of each block are explained.

4.1 Sampling z_t

This step is devoted to the sampling of the conditional probability $p(z_t|\eta, \theta, Y_t)$ which considered by Chib (1996), see also Kaufman and Fruhwirth-Schnatter (2002). Suppose $p(z_1|\eta, \theta, Y_0,)$ be the stationary distribution of the chain, then

$$p(z_t|\eta, \theta, Y_t) \propto f(y_t|\theta, z_t = k, Y_{t-1})p(z_t|\eta, \theta, Y_{t-1}), \tag{4.5}$$

where the predictive density $f(y_t|\theta, z_t = k, Y_{t-1})$ is calculated by (4.4) and by the law of total probability, $p(z_t|\eta, \theta, Y_{t-1})$ is given by

$$p(z_t|\eta, \theta, Y_{t-1}) = \sum_{z_{t-1}=1}^{K} p(z_{t-1}|\eta, \theta, Y_{t-1})\eta_{z_{t-1}z_t}, \tag{4.6}$$

where K is the number of states. Given the probabilities $(p(z_t|\eta, \theta, Y_t))$, we run a backward algorithm, starting from $t = T$, z_T is derived from $p(z_T|\eta, \theta, Y)$. For $t = T-1, \ldots, 0$, the corresponding samples are derived from $p(z_t|z_{t+1}, \ldots, z_T, \theta, \eta, Y)$, which satisfies

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

Derive z_t from $p(z_t|.) = p_{z_t}$ by the following procedure: first evaluate $q_j = p(Z_t = j|Z_t \geq j, .)$ by

$$p(Z_t = j|Z_t \geq j, .) = \frac{p_j}{\sum_{l=j}^{K} p_l},$$

then generate a number u from the standard uniform distribution (U(0,1)). If $u \leq q_j$ then put $z_t = j$ otherwise increasing j to $j + 1$ and generate another u from U(0,1) and repeat this step by comparing this with q_{j+1}.

4.2 Sampling η

This stage is devoted to sample $\eta = (\eta_{11}, \eta_{22})$ from the posterior probability $p(\eta|\theta, Y_t, Z_t)$ that is independent of Y_t, θ. We consider independent beta prior density for each of η_{11} and η_{22}. So,

$$p(\eta_{11}|Z_t) \propto p(\eta_{11})p(Z_t|\eta_{11}) = \eta_{11}^{c_{11}+n_{11}-1}(1 - \eta_{11})^{c_{12}+n_{12}-1},$$

where c_{11} and c_{12} are the parameters of beta prior, n_{ij} is the number of transition from $z_{t-1} = i$ to $z_t = j$. In the same way the sample of η_{22} is obtained.

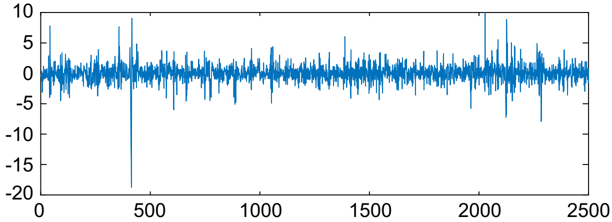

Fig. 2 Simulated data

4.3 Sampling θ

The posterior density of θ given the prior $p(\theta)$ is given by:

$$p(\theta|Y,Z) \propto p(\theta)\prod_{t=1}^{T} f(y_t|\theta, z_t = k, Y_{t-1}) = p(\theta)\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi H_{k,t}}}\exp\left(-\frac{y_t^2}{2H_{k,t}}\right), \tag{4.7}$$

which is independent of η. To sample from the $p(\theta|Y,Z)$ we use the Griddy Gibbs algorithm that introduced by Ritter and Tanner (1992). This method has had wide application in literature, see Bauwens and Lubrano (1998) , Bauwens and Storti (2009) and Bauwens et al. (2010).

Given samples at iteration r the Griddy Gibbs at iteration $r+1$ proceeds as follows:

1. Select a grid of points, such as $a_{0i}^1, a_{0i}^2, \ldots, a_{0i}^G$. Use (4.7) to evaluate the kernel of conditional posterior density function of a_{0i} given all the values of Z, Y and θ except a_{0i}, $k(a_{0i|Z_t,Y_t,\theta_{-a_{0i}}})$ over the grid points to obtain the vector $G_k = (k_1, \ldots, k_G)$.
2. By a deterministic integration rule using the G points, compute $G_\Phi = (0, \Phi_2, \ldots, \Phi_G)$ with

$$\Phi_j = \int_{a_{0i}^1}^{a_{0i}^j} k(a_{01}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t)da_{0i}, \quad i = 2, \ldots, G. \tag{4.8}$$

3. Simulate $u \sim U(0, \Phi_G)$ and invert $\Phi(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t)$ by numerical interpolation to obtain a sample $a_{0i}^{(r+1)}$ from $p(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t)$.
4. Repeat steps 1–3 for other parameters.

Prior densities of elements of θ can be considered as independent uniform densities over finite intervals.

5 Simulation results

We have simulated 2500 sample from the proposed model (2.1)–(2.4) for two states, $j = 1, 2$. Figure 2 shows the plot of the simulated time series and Table 1 reports some descriptive statistics of the simulated data.

Table 1 Descriptive statistics for the simulated data

	Mean	SD	Skewness	Maximum	Minimum	Kurtosis
Simulated data	– 0.03	1.470	– 1.621	9.057	– 18.754	24.623

Table 2 Results of the Bayesian estimation of the simulated MS-STGARCH model

	True values	Mean	SD	MSE
a_{01}	0.300	0.314	0.031	0.0001
a_{11}	0.200	0.224	0.022	0.0005
a_{21}	0.050	0.12	0.014	0.0002
b_1	0.500	0.510	0.049	0.002
γ_1	1.500	1.517	0.179	0.032
a_{02}	1.900	1.750	0.09	0.008
a_{12}	0.700	0.619	0.060	0.004
a_{22}	0.100	0.094	0.016	0.0003
b_2	0.250	0.217	0.021	0.0004
γ_2	0.500	0.619	0.013	0.0002
η_{11}	0.970	0.982	0.093	0.009
η_{22}	0.850	0.853	0.097	0.009

Using the Bayesian inference, we estimate the parameters of the MS-STGARCH by applying the first 2000 samples. The prior density of each parameter is assumed to be uniform over a finite interval except for transition probabilities η_{11} and η_{22} which are drawn from some beta distribution. Table 2 demonstrates the true values of the parameters and also posterior means and standard deviations of the corresponding estimators over 10,000 iterations, which 5000 of them are discarded as burn-in samples. The results of this table shows that the mean square errors (MSE) of the estimated parameters are adequately small. Using simulated data, we compare the in-sample fit and out-of-sample forecasting performance of the presented model with EGARCH, GARCH, GJR-GARCH, STGARCH and MS-GARCH models in Sects. 5.1 and 5.2 respectively.

5.1 In sample performance analysis

In order to compare the goodness of fit of the proposed model with EGARCH, GARCH, GJR-GARCH, STGARCH and MS-GARCH, we apply the DIC introduced by Speigelhalter et al. (2002). DIC is a Bayesian version of the reputable Akaike information criterion (AIC) that designed specifically for Bayesian estimation and involves MCMC simulations, see Gelman et al. (2014) and StataCrop (2015) . The smallest DIC determines the best model. Berg et al. (2004) applied DIC for the family of stochastic volatility (SV) models and Ardia (2009) for the family of asymmetric GARCH models.

Table 3 Deviance information criterion (DIC) for simulated data

Model	$S\&P$ 500 returns
EGARCH	6114.5
GARCH	5896
GJR-GARCH	5886.8
ST-GARCH	6021.4
MS-GARCH	5887.8
MS-STGARCH	5858.6*

*The Deviance information criterion (DIC) of the MS-STGARCH model is less in compare to the DSI of other models

In the Markov-switching models, the likelihood is calculated by the following formula:

$$f(Y|\Theta) = \prod_{t=1}^{T} f(y_t|\mathcal{I}_{t-1}, \Theta),$$

in which Θ is the vector of all parameters in model and $f(y_t|\mathcal{I}_{t-1})$ is obtained from (3.1). The deviance information criterion is computed as:

$$DIC = 2\log(f(Y|\hat{\Theta})) - 4E_{Y|\Theta}[\log(f(Y|\Theta))], \tag{5.1}$$

where $\hat{\Theta}$ is the posterior means of the vector Θ. The results concerning DIC of the simulated data are reported in Table 3. It is apparent that the DIC of MS-STGARCH is the smallest value in the table. Thus our considered model has the best fit to the simulated data set among competing models.

5.2 Out-of-sample forecasting performance analysis

For appraising the performance of MS-STGARCH in forecasting, we survey the one-day-ahead value at risk (VaR) forecasts for the last 500 data of the simulated data. The one-day-ahead value at risk level $\alpha \in (0, 1)$, VaR(α) is obtained by calculating the $(1-\alpha)$th percentile of the one-day-ahead predictive distribution (4.4). To test the VaR at level α, we evaluate the sequence $\{V_t(\alpha)\}$ by

$$V_t(\alpha) = \begin{cases} I\{y_{t+1} < VaR(\alpha)\} & \text{if} \quad \alpha > 0.5 \\ I\{y_{t+1} > VaR(\alpha)\} & \text{if} \quad \alpha \leq 0.5. \end{cases}$$

The out-of-sample VaR at level α has good performance if the sequence $\{V_t(\alpha)\}$ are independent and obey the following distribution

$$V_t(\alpha) \sim \begin{cases} Bernoulli(1-\alpha) & \text{if} \quad \alpha > 0.5 \\ Bernoulli(\alpha) & \text{if} \quad \alpha \leq 0.5. \end{cases}$$

The three likelihood ratio statistics for unconditional coverage (LR_{uc}), independence (LR_{ind}) and conditional coverage (LR_{cc}) are as follows Christofferssen (1998):

1. LR statistic for the test of unconditional coverage,

$$LR_{uc} = -2\ln[\frac{\phi^{n_1}(1-\phi)^{n_0}}{\hat{\pi}^{n_1}(1-\hat{\pi})^{n_0}}] \sim \chi^2_{(1)},$$

where ϕ is the parameter of related Bernoulli distribution, which could be $1-\alpha$ or α, n_1 is the number of 1's and n_0 is the number of 0's in the $V_t(\alpha)$ series and $\hat{\pi} = \frac{n_1}{n_1+n_0}$.

2. LR statistic for the test of independence,

$$LR_{ind} = -2\ln[\frac{\hat{\pi}_*^{n_{00}+n_{10}}(1-\hat{\pi}_*)^{n_{11}+n_{01}}}{\hat{\pi}_1^{n_{00}}(1-\hat{\pi}_1)^{n_{01}}\hat{\pi}_2^{n_{11}}(1-\hat{\pi}_2)^{n_{10}}}] \sim \chi^2_{(1)},$$

where n_{ij} is the number of transition from i to j ($i, j = 0, 1$) in the $V_t(\alpha)$ series, $\hat{\pi}_1 = \frac{n_{00}}{n_{00}+n_{01}}$, $\hat{\pi}_2 = \frac{n_{11}}{n_{10}+n_{11}}$ and $\hat{\pi}_* = \frac{n_{00}+n_{10}}{n_{00}+n_{01}+n_{10}+n_{11}}$.

3. LR statistic for the test of conditional coverage,

$$LR_{cc} = LR_{ind} + LR_{uc},$$

LR_{cc} has χ^2 distribution with two degrees of freedom. When the value of LR_{cc} is less than the critical value of χ^2 distribution one infer that the conditional coverage is correct and there exist good VaR forecasts.
The results of the tests for simulation example are reported in Table 4. The second and third columns demonstrate the theoretical expected violations and the number of empirical violations respectively.

According to the results of Table 4, at the 5% significance levels, as $\chi^2_{1,0.95} = 3.841$, the LR_{uc} test is rejected five times for EGARCH, four times for GJR-GARCH, GARCH and STGARCH, two times for MS-GARCH and one time for MS-STGARCH models. For some risk levels the test of independence (IND test) is not applicable since no consecutive violations have been occurred. In such cases $n_{00} = 0$ and so the LR_{ind} statistic becomes infinity. The LR_{ind} statistic at 5% significance level is bigger than critical value for one case of GJR-GARCH, GARCH and MS-STGARCH and also two cases of EGARCH and STGARCH. The conditional coverage (CC) test is higher than critical value $\chi^2_{2,0.95} = 5.991$ with two degrees of freedom, five times for EGARCH, four times for GJR-GARCH and STGARCH, three times for the GARCH, two times for the MS-GARCH and one time for the MS-STGARCH.

6 Empirical data set

By applying daily log returns of the $S\&P$ 500 for the period of 03/01/2005 to 03/11/2014 (2500 observations), we compare the performance of our model with the GARCH, EGARCH, GJR-GARCH, MS-GARCH and STGARCH ones. From the 2500 observations of $S\&P$ 500, the first 2000 observations are employed to estimate

Table 4 VaR results of simulated data

Model	α	$E(V_t(\alpha))$	N	UC	IND	CC
EGARCH	0.99	5	18	20.458	5.161	25.619
	0.95	25	43	11.331	4.744	16.075
	0.9	50	72	9.6	1.742	11.345
	0.1	50	61	2.53	0.037	2.568
	0.05	25	40	8.079	0.606	8.686
	0.01	5	19	23.129	1.505	24.634
GJR-GARCH	0.99	5	11	5.419	5.539	10.959
	0.95	25	38	6.181	1.525	7.706
	0.9	50	63	3.499	0.175	3.674
	0.1	50	53	0.196	0.091	0.287
	0.05	25	38	6.181	0.357	6.538
	0.01	5	15	13.162	0.537	13.699
GARCH	0.99	5	13	8.973	4.243	13.216
	0.95	25	42	10.195	1.756	11.95
	0.9	50	62	2.997	0.03	3.027
	0.1	50	50	0	0.2664	0.2664
	0.05	25	37	5.317	0.256	5.573
	0.01	5	12	7.111	1.15	8.263
STGARCH	0.99	5	12	7.111	4.854	11.964
	0.95	25	38	6.181	1.527	7.706
	0.9	50	63	3.499	0.005	3.504
	0.1	50	53	0.197	0.030	0.227
	0.05	25	39	7.102	0	7.103
	0.01	5	12	7.111	4.854	11.964
MS-GARCH	0.99	5	13	8.970	0.914	9.887
	0.95	25	34	3.080	2.81	5.892
	0.9	50	56	0.773	0.558	1.330
	0.1	50	50	0	0.266	0.266
	0.05	25	38	6.181	0.357	6.538
	0.01	5	10	3.914	1.75	5.665
MS-STGARCH	0.99	5	8	1.538	NA	NA
	0.95	25	36	4.510	3.987	8.498
	0.9	50	62	2.996	0.015	3.011
	0.1	50	54	0.347	0.274	0.622
	0.05	25	34	3.081	0.215	3.295
	0.01	5	9	2.613	2.126	4.739

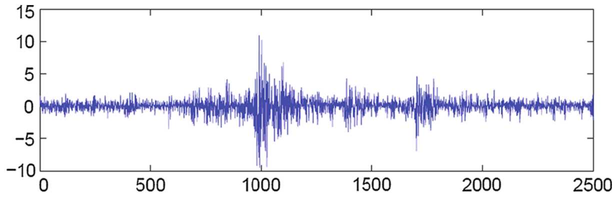

Fig. 3 Percentage daily log returns of S&P 500 data

Table 5 Descriptive statistics for the $S\&P$ 500 daily log returns

	Mean	SD	Skewness	Maximum	Minimum	Kurtosis
$S\&P$ 500	0.023	1.287	− 0.337	10.957	− 9.469	14.049

the parameters and the remaining 500 samples are used for forecasting analysis. Figure 3 plots the daily log returns in percentages[1] of the $S\&P$ 500 indices.

In Table 5, the descriptive statistics of the log returns in percentages are presented. This table shows that the means are close to zero and there are some slightly negative skewness and excess kurtosis for the data set.

6.1 Estimation of the parameters

Applying the S&P 500 set of samples and using the Bayesian MCMC method through Gibbs and griddy Gibbs sampling, we estimate the parameters GARCH EGARCH, GJR-GARCH, STGARCH, two-state MS-GARCH and MS-STGARCH to compare their performance. The prior density of transition probabilities η_{11} and η_{22} are drawn from the beta distribution and priors for the other parameters are assumed to be uniform over some finite intervals. We consider 10,000 iterations of Gibbs algorithm which half of them are burn-in-phase. The posterior means and standard deviations for the parameters of the models corresponding to S&P 500 data are reported in Tables 6 and 7, which show that the standard deviations are small enough in all cases, except for the slop parameter γ_1 which relates to the state with low volatility.

The single-regime STGARCH has potential to react differently to negative and positive shocks but does not consider shifting between different levels of volatilities. The estimation results show that the level of volatility of the second regimes in MS-STGARCH and MS-GARCH are higher. This is by the fact that the coefficient a_{i2} for $j = 0, 1, 2$ are respectively greater than a_{i1}. In Table 6, we see that the estimated parameters of the MS-STGARCH satisfies $a_{11} > a_{21}$ and $a_{12} > a_{22}$, so the negative shocks have more affect on volatility than the positive shocks as $w_{j,t-1}$, for $j = 1, 2$ for large negative shocks approaches to zero.

Table 6 shows that the posterior means of transition probabilities (η_{11} and η_{22}) are close to one which indicate less switch between regimes. Estimated conditional

[1] log return in percentage is defined as $r_t = 100 * \log\left(\frac{P_t}{P_{t-1}}\right)$, where P_t is the index level at time t.

Table 6 Posterior means and standard deviations (S&P 500 daily log returns)

	MS-STGARCH		MS-GARCH		STGARCH		GARCH	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
a_{01}	0.194	0.001	0.233	0.004	0.336	0.011	0.269	0.005
a_{11}	.276	0.008	0.278	0.012	0.421	0.019	0.120	0.007
a_{21}	0.085	0.006	0	0	0.121	0.016	0	0
b_1	0.289	0.003	0.320	0.009	0.364	0.014	0.439	0.004
γ_1	2.345	0.132	0	0	2.206	0.218	0	0
a_{02}	0.717	0.087	0.779	0.012	–	–	–	–
a_{12}	0.677	0.007	0.430	0.011	–	–	-	–
a_{22}	0.365	0.013	0	0	–	–	–	–
b_2	0.264	0.015	0.207	0.007	–	–	–	–
γ_2	1.097	0.017	0	0	–	–	–	–
η_{11}	0.986	0.004	0.994	0.003	–	–	–	–
η_{22}	0.985	0.005	0.991	0.004	–	–	–	–

Table 7 Parameter estimation of the EGARCH and GJR-GARCH models

	EGARCH		GJR-GARCH	
Parameters	Mean	SD	Mean	SD
Constant	0.362	0.009	0.447	0.011
arch effect	0.139	0.022	0.151	0.007
garch effect	0.301	0.013	0.293	0.006
leverage effect	0.433	0.010	0.207	0.026

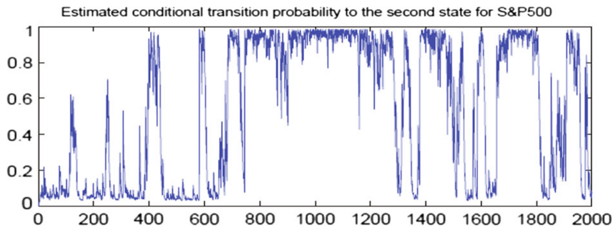

Fig. 4 Estimated conditional transition probabilities to the second states (the high volatility regime) of the fitted MS-STGARCH, $\alpha_2^{(t)}$ calculated by (3.6), for the S&P 500 log returns

transition probabilities to the second state (high volatility regime), $\alpha_2^{(t)}$ computed by (3.6) plotted in Fig. 4.

The MS-STGARCH has the potential to present better forecasting when different levels of volatilities are presented and there is different effect for negative and positive shocks. The results of Table 8 demonstrates that the MS-STGARCH has the best fitting to data. For appraising the performance of MS-STGARCH in forecasting, we survey the one-day-ahead VaR forecasts for the samples of S&P 500. Based on the last 500 returns (of S&P 500), the out of sample VaR forecasts are calculated.

Table 8 Deviance information criterion (DIC)

Model	S&P 500 returns
EGARCH	8889.3
GARCH	8464.8
GJR-GARCH	7623.1
STGARCH	7513.5
MS-GARCH	7257.1
MS-STGARCH	7147.8*

*The Deviance information criterion (DIC) of the MS-STGARCH model is less in compare to the DSI of other models

According to the results of Table 9, at the 5 and 10% significance levels, the LR_{uc} test is rejected three times for EGARCH, GJR-GARCH and STGARCH, two times for MS-GARCH and is accepted at all risk levels α for the GARCH and MS-STGARCH models. The LR_{ind} statistic at 5% significance level is smaller than critical value $\chi^2_{0.95}$ with one degree of freedom for all cases that test is applied. Also excluding the cases of risk level 0.95 for the EGARCH and GJR-GARCH, the IND test is accepted at 10%. At the 5% significance level, the conditional coverage (CC) test is higher than critical value $\chi^2_{0.95}$ with two degrees of freedom two times for the EGARCH, GJR-GARCH and STGARCH while at the 10% significance level this test is rejected three times for the EGARCH, two times for the GJR-GARCH and STGARCH and one time for the MS-GARCH.

In order to appraise the ability of competing models to forecast volatility, we apply the Diebold Mariano test. Testing for equal forecast accuracy is an approach to evaluate the predictive capability of competitor models. Diebold and Mariano (1995) proposed a unified method for testing the null hypothesis of no difference in the forecasting accuracy of two competing models Xekalaki and Stavros (2010). Harvey et al. (1997) suggested a modified of Diebold Mariano (DM) test for small sample. The DM test or its HLN variant are applied widely in empirical forecasting research, see Diebold and Mariano (1995), Harvey et al. (1997) and Curto and Pinto (2012). Consider two forecast sequences as

$$\{\hat{y}_{it} : t = 1, \ldots, T\}, \qquad i = 1, 2;$$

and define

$$e_{it} = \hat{y}_{it} - y_t;$$

that $\{y_t, t = 1, \ldots, T\}$ are actual values. Let $g(e_{it}) = e_{it}^2$ and

$$d_t = g(e_{1t}) - g(e_{2t});$$

we would like to test the null hypothesis:

$$H_0 : E(d_t) = 0, \quad \forall t$$

Table 9 VaR results of S&P 500 daily log returns

Model	α	$E(V_t(\alpha))$	N	UC	IND	CC
EGARCH	0.99	5	9	2.596	0.330	2.926
	0.95	25	26	0.038	2.740	2.778
	0.9	50	42	1.531	0.070	1.601
	0.1	50	32	8.227	0.001	8.228
	0.05	25	12	8.790	1.155	9.946
	0.01	5	1	4.829	0.004	4.833
GJR-GARCH	0.99	5	10	3.891	0.408	4.299
	0.95	25	26	0.037	2.740	2.778
	0.9	50	42	1.531	0.070	1.601
	0.1	50	31	9.236	0.036	9.239
	0.05	25	16	3.925	2.775	6.700
	0.01	5	3	0.950	0.036	0.987
GARCH	0.99	5	9	2.596	NA	NA
	0.95	25	18	2.276	2.024	4.3
	0.9	50	44	0.830	0.252	1.082
	0.1	50	42	1.395	0.071	1.466
	0.05	25	27	2.276	0.142	2.418
	0.01	5	4	0.229	NA	NA
STGARCH	0.99	5	5	0	NA	NA
	0.95	25	20	1.147	NA	NA
	0.9	50	31	9.235	0.5317	9.767
	0.1	50	28	12.684	1.191	13.875
	0.05	25	8	16.441	NA	NA
	0.01	5	2	2.365	NA	N
MS-GARCH	0.99	5	8	1.526	NA	NA
	0.95	25	27	0.156	0.1787	0.335
	0.9	50	42	1.531	0.827	2.358
	0.1	50	37	4.112	0.252	4.364
	0.05	25	15	4.926	0.539	5.465
	0.01	5	2	2.365	NA	NA
MS-STGARCH	0.99	5	9	2.596	NA	NA
	0.95	25	27	0.156	0.1787	0.335
	0.9	50	45	0.595	0.301	0.897
	0.1	50	44	0.9	0.252	1.109
	0.05	25	18	2.3	0.275	2.575
	0.01	5	2	2.38	NA	NA

Table 10 DM test results

Comparison of MS-STGARCH with	Statistic value
EGARCH	– 3.89
GJR-GARCH	– 3.73
GARCH	– 3.84
STGARCH	– 4.08
MS-GARCH	– 2.08

versus the alternative hypothesis

$$H_1 : E(d_t) < 0,$$

under covariance stationarity of the process $\{d_t : 1, \ldots, T\}$, the DM statistic for testing the null hypothesis is given by:

$$\frac{\bar{d}}{\sqrt{V\hat{a}r((\bar{d})}},$$

and is approximately normally distributed for large samples. For evaluating the performance of our model in one-step ahead conditional variance forecast, we compute the DM statistic for pairwise comparison of MS-STGARCH model with GARCH, STGARCH, MS-GARCH, GJR and EGARCH models. According to the test results demonstrated in Table 10, the null hypothesis is rejected at the 5% significance level for all cases as all the statistics are less than $Z_{0.05}$. So our presented model has an improvement in the forecasting performance.

Also for specifying the out of sample forecast performance of the MS-STGARCH toward the competing models, We compare the forecasting volatility $E(Y_t^2|\mathcal{F}_{t-1})$, or conditional variance, of GARCH, STGARCh and MS-GARCH with the squared returns. In Fig. 5, the squared returns of $S\&P$ 500 and the forecasting values of some competing models are plotted. According to this figure the differences of forecast and real values (errors) in the MS-STGARCH always has been much less than other compared models. The results of Table 11 show that the least values of the MSE and MAE are related to the MS-STGARCH model that reveals the best forecast compared with the other reviewed models in this paper.

7 Conclusion

Applying Markov switching structure cause to have a better fitting while the existence of different levels of volatilities are evident. Also the asymmetry effects of negative and positive shocks in many cases are trivial and transition between these effects happen in some smooth ways and not sudden. So in many cases the use of MS-STGARCH has advantages to the other methods as GARCH, MS-GARCH, ST-GARCH, EGARCH and GJR-GARCH as we find this performance for $S\&P$ 500 indices where studied in

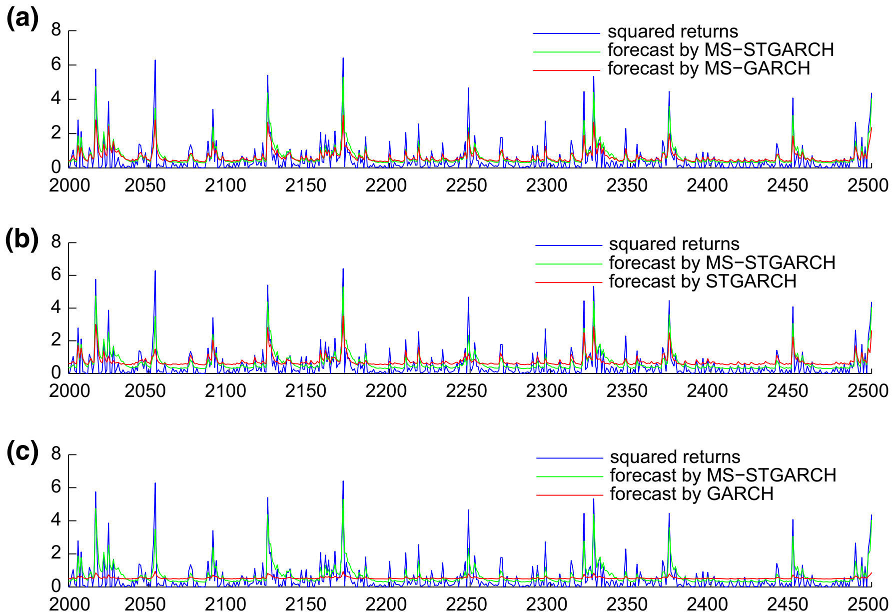

Fig. 5 Comparing forecasts of MS-STGARCH with forecasts of MS-GARCH, STGARCH and GARCH models by applying squared returns of S&P 500 (blue) (a): forecast by MS-STGARCH (green) and forecast by MS-GARCH (red), (b) forecast by MS-STGARCH (green) and forecast by STGARCH (red), (c) forecast by MS-STGARCH (green) and forecast by GARCH (red). (Color figure online)

Table 11 Measures of performance forecasting

Model	Mean square error (MSE)	Mean absolute error (MAE)
EGARCH	0.676	0.528
GJR-GARCH	0.665	0.517
GARCH	0.707	0.524
STGARCH	0.384	0.498
MS-GARCH	0.302	0.395
MS-STGARCH	0.227*	0.369*

*The Mean absolute error (MAE) of the MS-STGARCH model is less in compare to the MAE of other models

the paper. In such cases a much better fit to the data can be provided by the presented model which leads to the forecasts with much smaller error. The MS-STGARCH extends the MS-GARCH model by considering convex combination of time dependent logistic weights between the effects of negative and positive shocks in each regime. It also extends the STGARCH model by considering transition between different levels of volatilities. We show that the existence of a simple condition suffices for the existence of an asymptotic upper bound for the second moment of returns which causes the stability of the model.

By fitting the EGARCH, GJR-GARCH, GARCH, STGARCH, MS-GARCH and MS-STGARCH models to the S&P 500 log returns we find that our model has the best DIC and provides the best forecast volatilities by performing Diebold Mariano test. The one-day ahead VaR forecast of our model has better performance to the EGARCH, GJR-GARCH, STGARCH and MS-GARCH.

Further researches could be oriented to investigate the existence of the third and the fourth moments of the process and derive the necessary and sufficient conditions for stationarity and ergodicity of the process. For the sake of simplicity it was assumed that the process conditional mean is zero, this assumption could be relaxed by refining the structure of model to allow ARMA structure for conditional mean. Since Financial time series data are typically observed to have heavy tails Liu et al. (2008), it might be interesting to replace the Gaussian distribution with Student's t or stable Paretian distributions to investigate the ability for modeling heavy tailed property of financial time series such as Curto et al. (2009) work.

Acknowledgements This paper initiated during Professor Rezakhah sabbatical at Institute of Mathematics at the EPFL where benefited from the discussion of the paper with Professor Stephan Morgenthaler and careful written comments and suggestions of Professor Anthony Davison which caused to improve the quality of this paper.

Appendix A

Proof of Lemma 3.1 As the hidden variables $\{Z_t\}_{t\geq 1}$ have Markov structure in MS-STARCH model, so

$$\begin{aligned}
\alpha_j^{(t)} = p(Z_t = j|\mathcal{I}_{t-1}) &= \sum_{m=1}^{K} P(Z_t = j, Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \sum_{m=1}^{K} p(Z_t = j|Z_{t-1} = m, \mathcal{I}_{t-1}) p(Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \sum_{m=1}^{K} p(Z_t = j|Z_{t-1} = m) p(Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \frac{\sum_{m=1}^{K} f(\mathcal{I}_{t-1}, Z_{t-1} = m) p_{m,j}}{\sum_{m=1}^{K} f(\mathcal{I}_{t-1}, Z_{t-1} = m)} \\
&= \frac{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})\alpha_m^{(t-1)} p_{m,j}}{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2})\alpha_m^{(t-1)}}.
\end{aligned} \tag{7.1}$$

□

Appendix B

Proof of Theorem 3.1 Let $E_t(.)$ denotes the expectation with respect to the information up to time t. Thus the second moment of the model can be calculated as, see Abramson and Cohen (2007):

$$
\begin{aligned}
E(y_t^2) &= E(H_{Z_t,t}) = E_{Z_t}[E_{t-1}(H_{Z_t,t}|z_t)] \\
&= \sum_{z_t=1}^{K} \pi_{z_t} E_{t-1}(H_{Z_t,t}|z_t).
\end{aligned} \tag{7.2}
$$

Also let $E(.|z_t)$ and $p(.|z_t)$ denote $E(.|Z_t = z_t)$ and $P(.|Z_t = z_t)$, respectively, where z_t is the realization of the state at time t. Applying to the method of Medeiros and Veiga (2009), we find an upper bound of $E_{t-1}(H_{m,t}|z_t)$, for $m = 1, 2, \ldots, K$ by the following

$$
\begin{aligned}
E_{t-1}(H_{m,t}|z_t) &= E_{t-1}(a_{0m} + a_{1m}y_{t-1}^2(1 - w_{m,t-1}) + a_{2m}y_{t-1}^2 w_{m,t-1} + b_m H_{m,t-1}|z_t) \\
&= \underbrace{a_{0m}}_{I} + \underbrace{a_{1m}E_{t-1}[y_{t-1}^2|z_t]}_{II} + \underbrace{(a_{2m} - a_{1m})E_{t-1}[y_{t-1}^2 w_{m,t-1}|z_t]}_{III} \\
&\quad + \underbrace{b_m E_{t-1}[H_{m,t-1}|z_t]}_{IV}.
\end{aligned} \tag{7.3}
$$

The term (II) in (7.3) can be interpreted as follows:

$$
\begin{aligned}
E_{t-1}[y_{t-1}^2|z_t] &= \sum_{z_{t-1}=1}^{K} \int_{S_{\mathcal{I}_{t-1}}} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t, z_{t-1}) p(z_{t-1}|z_t) d\mathcal{I}_{t-1} \\
&= \sum_{z_{t-1}=1}^{K} p(z_{t-1}|z_t) E_{t-2}[H_{Z_{t-1},t-1}|z_{t-1}],
\end{aligned} \tag{7.4}
$$

where $S_{\mathcal{I}_{t-1}}$ is the support of $\mathcal{I}_{t-1} = (y_1, \ldots, y_{t-1})$. □

Upper bound for III in (7.3): Let $0 < M < \infty$ be a constant, so

$$
\begin{aligned}
E_{t-1}[y_{t-1}^2 w_{m,t-1}|z_t] =& E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}|<M}|z_t] \\
&+ E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}|\geq M}|z_t]
\end{aligned}
$$

in which

$$
I_{x<a} = \begin{cases} 1 & \text{if } x < a \\ 0 & \text{otherwise.} \end{cases}
$$

As by (2.4), $0 < w_{m,t-1} < 1$ and so

$$
E_{t-1}[y_{t-1}^2 w_{m,t-1}|z_t] \leq M^2 + E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}|\geq M}|z_t],
$$

also

$$E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}|\geq M}|z_t] = \int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\leq -M} y_{t-1}^2[w_{m,t-1}]p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} + \int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\geq M} y_{t-1}^2[w_{m,t-1}]p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1},$$

by (2.4),

$$\lim_{y_{t-1}\to+\infty} w_{m,t-1} = 1 \qquad \lim_{y_{t-1}\to-\infty} w_{m,t-1} = 0. \tag{7.5}$$

So for any fixed positive small number $\delta > 0$, we can consider $M > 0$ so large that for $y_{t-1} \geq M$, $|w_{m,t-1} - 1| \leq \delta$ and for $y_{t-1} \leq -M$, $|w_{m,t-1}| \leq \delta$. Hence

$$E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}|\geq M}|z_t] \leq \delta \int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\leq -M} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} + (\delta+1)\int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\geq M} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1}.$$

Since the distribution of the $\{\varepsilon_t\}$ is symmetric, then

$$\delta \int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\leq -M} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} \leq \delta \int_{S_{\mathcal{I}_{t-2}}, -\infty<y_{t-1}<0} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} = \delta \frac{E_{t-1}[y_{t-1}^2|z_t]}{2}$$

and

$$(\delta+1)\int_{S_{\mathcal{I}_{t-2}}, y_{t-1}\geq M} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} \leq (\delta+1)\int_{S_{\mathcal{I}_{t-2}}, 0<y_{t-1}<\infty} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t)d\mathcal{I}_{t-1} = (\delta+1)\frac{E_{t-1}[y_{t-1}^2|z_t]}{2}.$$

Therefore

$$E_{t-1}[y_{t-1}^2 w_{m,t-1}|z_t] \leq M^2 + (\delta + \frac{1}{2})E_{t-1}[y_{t-1}^2|z_t].$$

Upper bound for IV in (7.3):

$$b_m E_{t-1}(H_{m,t-1}|z_t) = b_m \int_{S_{\mathcal{I}_{t-1}}} H_{m,t-1} p(\mathcal{I}_{t-1}|z_t) d\mathcal{I}_{t-1}$$
$$= b_m \sum_{z_{t-1}=1}^{K} p(z_{t-1}|z_t) E_{t-2}(H_{m,t-1}|z_{t-1}). \tag{7.6}$$

By replacing the obtained upper bounds and relations (7.4) in (7.3), the upper bound for $E_{t-1}(H_{m,t}|z_t)$ is obtained by:

$$E_{t-1}(H_{m,t}|z_t) \leq a_{0m} + |a_{2m} - a_{1m}|M^2$$
$$+ \sum_{z_{t-1}=1}^{K} [a_{1m} + |a_{2m} - a_{1m}|(\delta + \frac{1}{2})] p(z_{t-1}|z_t) E_{t-2}[H_{z_{t-1},t-1}|z_{t-1}]$$
$$+ \sum_{z_{t-1}=1}^{K} b_m p(z_{t-1}|z_t) E_{t-2}(H_{m,t-1}|z_{t-1}), \tag{7.7}$$

in which by Bayes' rule

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

where P is the transition probability matrix.

Let $A_t(j,k) = E_{t-1}[H_{j,t}|Z_t = k]$, $A_t = [A_t(1,1), A_t(2,1), \ldots, A_t(K,1), A_t(1,2), \ldots, A_t(K,K)]$ be a K^2-by-1 vector and consider $\dot{\Omega} = (\Omega', \ldots, \Omega')'$ be a vector that is made of K vector Ω. By (7.23)–(7.26), the following recursive inequality is attained,

$$\mathbf{A}_t \leq \dot{\Omega} + \mathbf{C}\mathbf{A}_{t-1}, \quad t \geq 0. \tag{7.8}$$

with some initial conditions $\mathbf{A}_{-1}$. The relation (7.8) implies that

$$A_t \leq \dot{\Omega} \sum_{i=0}^{t-1} C^i + C^t A_0 := B_t. \tag{7.9}$$

Following the matrix convergence theorem Lancaster and Tismenetsky (1985), the necessary condition for the convergence of B_t when $t \to \infty$ is that $\rho(C) < 1$. Under this condition, C^t converges to zero as t goes to infinity and $\sum_{i=0}^{t-1} C^i$ converges to $(I - C)^{-1}$ provided that matrix $(I - C)$ is invertible. So if $\rho(C) < 1$,

$$\lim_{t \to \infty} A_t \leq (I - C)^{-1} \dot{\Omega}.$$

By (7.2) the upper bound for the asymptotic behavior of unconditional variance is given by

$$\lim_{t \to \infty} E(y_t^2) \leq \Pi'(I - C)^{-1}\dot{\Omega}.$$

References

Abramson A, Cohen I (2007) On the stationarity of Markov-Switching GARCH processes. Econom Theory 23:485–500

Alemohammad N, Rezakhah S, Alizadeh SH (2016) Markov switching component GARCH model: stability and forecasting. Commun Stat Theory Methods 45(15):4332–4348

Ardia D (2008) Financial risk management with Bayesian estimation of GARCH models: theory and applications, vol 612. Lecture notes in economics and mathematical systems. Springer, Heidelberg

Ardia D (2009) Bayesian estimation of a Markov switching threshold asymmetric GARCH model with Student-t innovations. Econom J 12(2):105–126

Bauwens L, Lubrano M (1998) Bayesian inference on GARCH models using the Gibbs sampler. Econom J 1:23–46

Bauwens L, Storti G (2009) A component GARCH model with time varying weights. Stud Nonlinear Dyn Econom 13:Article 1

Bauwens L, Preminger A, Rombouts VK (2010) Theory and inference for a Markov switching GARCH model. Econom J 13:218–244

Berg A, Meyer R, Yu J (2004) Deviance information criterion for comparing stochastic volatility models. J Bus Econ Stat 22:107–120

Black F (1976) studies of stock price volatility changes. In: Proceedings of the 1976 meetings of the business and economic statistics section. American Statistical Association, Washington DC, pp 177–181

Bollerslev T (1986) Generalized autoregressive conditional heteroscedasticity. J Econom 31:307–327

Chib S (1996) Calculating posterior distributions and model estimates in Markov mixture models. J Econom 75:79–97

Christofferssen P (1998) Evaluating interval forecasting. Int Econ Rev 39:841–862

Curto JD, Pinto JC (2012) Predicting the financial crisis volatility. Econ Comput Econ Cybern Stud Res J 46:183–195

Curto JD, Pinto JC, Tavares GN (2009) Modeling stock markets volatility using GARCH models with Normal, Students t and stable Paretian distributions. Stat Pap 50:311–321

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. J Bus Econ Stat 13:253–263

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50:987–1007

Engle RF (1990) Discusion: stock volatility and the crash of ’87. Rev Financ Stud 3:103–106

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian. Stat Comput 24:997–1016

Glosten LR, Jagannathan R, Runkle D (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. J Financ 48:1779–1801

Gonzalez-Rivera G (1998) Smooth transition GARCH models. Stud Nonlinear Dyn Econom 3:61–78

Gray SF (1996) Modeling the conditional distribution of interest rates as a regime-switching process. J Financ Econ 42:27–62

Haas M, Mittnik J, Paolella MS (2004) A new approach to markov-switching GARCH models. J Financ Econom 2:493–530

Haas M, Krause J, Paolella MS, Steudc SC (2013) Time varying mixture GARCH models and asymmetric volatility. N A J Econ Financ 26:602–623

Hamilton JD, Susmel R (1994) Autoregressive conditional heteroskedasticity and changes in regime. J Econom 64:307–333

Harvey D, Leybourns S, Newbold P (1997) Testing the equality of prediction mean squared error. Int J Forecast 13:281–291

Kaufman S, Fruhwirth-Schnatter S (2002) Bayesian analysis of switching ARCH models. J Time Ser Anal 23:425–458

Klaassen F (2002) Improving GARCH volatility forecasts with regime-switching GARCH. Empir Econ 27:363–394
Lancaster P, Tismenetsky M (1985) The theory of matrices, 2nd edn. Academic Press, New York
Liu S, Heyde CC, Wong WK (2008) On estimation in conditional heteroskedastic time series models under non-normal distributions. Stat Pap 49:455–469
Lubrano M (2001) Smooth transition GARCH models: a Bayesian approach mixture models. Recherches Economiques de Louvain 67:257–287
Medeiros MC, Veiga A (2009) Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1). Econom Theory 25:117–161
Miazhynskia T, Dorffner G (2006) A comparison of Bayesian model selection based on MCMC with an application to GARCH-type models. Stat Pap 47:525–549
Nelson DB (1991) Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59:347–370
Ritter C, Tanner MA (1992) Facilitating the Gibbs sampler: the Gibbs Stopper and the Griddy-Gibbs Sampler. J Am Stat Assoc 87:861–868
Speigelhalter DJ, Best NG, Carlin BP, Van del lindle A (2002) Bayesian measures of model complexity and fit. J R Stat Soc Ser 64:583–639
StataCrop (2015) Stata: release 14 Statistical software. StataCrop LLC, College station
Xekalaki E, Stavros D (2010) ARCH nodels for financial applications. Wiley, New York
Zakoian JM (1994) Threshold heteroskedastic models. J Econ Dyn Control 27:577–597