Markov Switching Component ARCH Model: Stability and Forecasting

N. Alemohammad, S. Rezakhah, S. H. Alizadeh *

March 25, 2013

Abstract

This paper introduces an extension of the Markov switching ARCH model where the volatility in each state is a convex combination of two different ARCH components with time varying weights with different volatilities. The asymptotic behavior of the second moment is investigated and an appropriate upper bound for it is evaluated. The estimation of the parameters by using the Bayesian method via Gibbs sampling algorithm is studied. We propose a dynamic programming algorithm for the forecasting. Finally we illustrate the efficiency of the model by simulation and forecasting the volatility. We show that this model provide much better forecast of the volatility than the Markov switching ARCH model.

Keywords: ARCH models, Markov process, Stability, Component GARCH models, Forecasting, Bayesian inference, Griddy Gibbs sampling.

Mathematics Subject Classification: 60J10, 62M10, 62F15.

1 Introduction

In the past three decades, there has been a growing interest in using non linear time series models in finance and economy. For financial time series, the ARCH model and GARCH model , introduced by Engle [11] and Bollerslev [7], are surely the most popular class of volatility models. Although these models have been applied extensively in the modeling of financial time series, but the dynamic streucture of volatility can not be captured passably by such models. For more consistent volatility modelling, the models by time varying parameters are introduced. One

*Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran. Email: n-alemohammad@aut.ac.ir, rezakhah@aut.ac.ir, sasan.alizadeh@gmail.com

class of such models is that of smooth transition GARCH models that presented by Gonzalez-Rivera [15], (see also Hagerud [20] and Medeiros and Veiga [26]). Another class is that of Markov switching models. These models are obtained by by Merging (G)ARCH model with a Markov process, where each state of the Markov model allows a different (G)ARCH behavior. These models are introduced by Cai [8] and Hamilton and Susmel [19]. This feature extends the dynamic formulation of the model and potentially enables improving forecasts of the volatility [1]. Gray [16], Klaassen [22], Haas, Mittnik and Paolella [18] proposed different variants of Markov-Switching GARCH models. See also further studies, Abramson and Cohen [1], Alexander and Lazar [2] and Bauwens et al. [6]. The component GARCH models, introduced first by Ding and Granger [10], are also a generalization of constant parameter GARCH model. These models have been widely applied in modeling the financial time series (e.g. [12], [25] and [13]). In the structure of component GARCH model ([10]), two different ARCH component contribute to the overall conditional variance at time t. One component has the high volatility (integrated variance component) and the other component has the low volatility. A generalization of the component GARCH model of Ding and Granger is the weighted GARCH model that is peoposed by Bauwens and Storti [5]. In this model the weights of GARCH components are the function of lagged values of the conditional standard deviation or squared past observations.

In this paper we consider a Markov switching model that the volatility of each state is a convex combination of two ARCH regimes with time varying coefficients which is in effect of previous observation. This model has the potential to model the effect of more complicated resources which are in effect of some volatility components and the share of these components could change in time. We consider different weight functions for each state that allow volatility in each state to react differently to the shocks of equal size. 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. We derive necessary and sufficient conditions for stability and obtain an upper bound for the limit of the second moment by using the method of Abramson and Cohen [1] and Medeiros [26]. For the estimation of the parameters we use the Bayesian inference via the Gibbs sampling. We compare the performance of our model to Markov switching ARCH model. The Markov switching component ARCH model can forecast the conditional variance much better than MS-ARCH model.

The paper is organized as follows: in section 2 we introduce the smooth transition Markov switching ARCH model. Section 3 investigates the statistical properties of the model. Section 4 is devoted to estimation of the parameters of the model.

Section 5 dedicated to the analyzing of the efficiency of the proposed model through simulation and the comparison of the forecast errors with the MS-ARCH model. Section 6 concludes.

2 Markov switching Component ARCH model

The Markov switching component ARCH model, MS-CARCH, for time series $\{y_t\}$ is defined as

$$y_t = \varepsilon_t \sqrt{H_{t,Z_t}}, \qquad H_{t,Z_t} = w_{t,Z_t} h_{1,t,Z_t} + (1 - w_{t,Z_t}) h_{2,t,Z_t}, \tag{2.1}$$

where $\{\varepsilon_t\}$ are iid standard normal variables, $\{Z_t\}$ is an irreducible and aperiodic Markov chain on finite state space $E = \{1, 2, \cdots, K\}$ with transition probability matrix $P = ||p_{ij}||_{K \times K}$, where $p_{ij} = p(Z_t = j | Z_{t-1} = i)$, $i, j \in \{1, \cdots, K\}$, and stationary probability measure $\pi = (\pi_1, \cdots, \pi_K)'$. Also

$$h_{1,t,Z_t} = a_{0,Z_t} + a_{1,Z_t} y_{t-1}^2, \qquad h_{2,t,Z_t,} = b_{0,Z_t} + b_{1,Z_t} y_{t-1}^2, \tag{2.2}$$

and each of the weights $(w_{t,i}, i = 1, \cdots, K)$ is a function of past observation as

$$w_{t,i} = \frac{1 - \exp(-\gamma_i |y_{t-1}|)}{1 + \exp(-\gamma_i |y_{t-1}|)} \qquad \gamma_i > 0, \tag{2.3}$$

which is bounded , $0 < w_{i,t-1} < 1$. The parameter γ_i is called the slope parameters, that explains the speed of transition from one component to the other one: the higher γ_i, the faster the transition. Since $\gamma_i > 0$, when the absolute value of y_{t-1} increases, the impact of $h_{1,i,t}$ is increases and consequently the effect of $h_{2,i,t}$ decreses and vice versa. For this reason we consider the first ARCH component in each state with the high volatility and the second component with low volatility. So when γ tending to zero or infinity and the MS-CARCH model tends to MS-ARCH model.

It is assumed that $\{\varepsilon_t\}$ and $\{Z_t\}$ are independent. Sufficient conditions to guarantee strictly positive conditional variance are $a_{0,i}, b_{0,i}$ to be positive and $a_{1,i}, a_{2,i}, b_{1,i}, b_{2,i}$ being nonnegative.

Let $\mathcal{I}_t$ be the observation set up to time t. The conditional density function of y_t given past information is obtained as follows:

$$
\begin{aligned}
f(y_t|\mathcal{I}_{t-1}) &= \sum_{i=1}^{K} f(y_t, Z_t = i|\mathcal{I}_{t-1}) \\
&= \sum_{i=1}^{K} p(Z_t = i|\mathcal{I}_{t-1}) f(y_t|\mathcal{I}_{t-1}, Z_t = i) \\
&= \sum_{i=1}^{K} \alpha_i^{(t)} \phi(\frac{y_t}{\sqrt{H_{t,i}}}) \qquad (2.4)
\end{aligned}
$$

in which $\alpha_i^{(t)} = p(Z_t = i|\mathcal{I}_{t-1})$ (that is obtained in next section), and $\phi(.)$ is the probability density function of the standard normal distribution.

3 Statistical Properties of the model

In this section, the statistical properties of the MS-CARCH model are investigated and the conditional variance of the process is obtained. We show that the model, under some conditions on coefficients and transition probabilities , is asymptotically stable in the second moment. An appropriate upper bound for the limiting value of the second moment is obtained.

3.1 Forecasting

The forecasting volatility (conditional variance) of MS-CARCH model is given by

$$
Var(Y_t|\mathcal{I}_{t-1}) = \sum_{i=1}^{K} \alpha_i^{(t)} H_{t,i} = \sum_{i=1}^{K} \alpha_i^{(t)} (w_{t,i} h_{1,t,i} + (1 - w_{t,i}) h_{2,t,i}) \qquad (3.5)
$$

as $H_{k,t}$ is the conditional variance of k-th state. This relation shows that the conditional variance of this model is affected by the changes in states, the volatility of components and the weight functions in each state.
At each time t, $\alpha_i^{(t)}$ (in equation (2.4), (3.5)) can be obtained from a dynamic programming method based on forward recursion algorithm, proposed in remark (3.1).

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

$$
\alpha_j^{(t)} = \frac{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m|\mathcal{I}_{t-2}) p_{m,j}}{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m|\mathcal{I}_{t-2})}. \qquad (3.6)
$$

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

$$\begin{aligned}
\alpha_j^{(t)} =& p(Z_t = j|\mathcal{I}_{t-1}) = \sum_{m=1}^{K} P(Z_t = j, Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \sum_{m=1}^{K} p(Z_t = j|Z_{t-1} = m, \mathcal{I}_{t-1}) p(Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \sum_{m=1}^{K} p(Z_t = j|Z_{t-1} = m) p(Z_{t-1} = m|\mathcal{I}_{t-1}) \\
&= \frac{\sum_{m=1}^{K} f(\mathcal{I}_{t-1}, Z_{t-1} = m) p_{m,j}}{\sum_{m=1}^{K} f(\mathcal{I}_{t-1}, Z_{t-1} = m)} \\
&= \frac{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m|\mathcal{I}_{t-2}) p_{m,j}}{\sum_{m=1}^{K} f(y_{t-1}|Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m|\mathcal{I}_{t-2})},
\end{aligned} \tag{3.7}$$

where

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

3.2 Stability

In this subsection, we investigate the stability of second moment of MS-CARCH model. Indeed we are looking for an upper bound for the second moment of our model. The second moment of the model can be calculated as:

$$E(y_t^2) = E(H_{t,Z_t}) = E_{Z_t}[E_{t-1}(H_{t,Z_t}|z_t)]$$

$$= \sum_{z_t=1}^{K} \pi_{z_t} E_{t-1}(H_{t,Z_t}|z_t). \tag{3.8}$$

$E_t(\cdot)$ denotes the expectation with respect to the information up to time t. Also for summarization, we shall use $E(\cdot|z_t)$ and $p(\cdot|z_t)$ to represent $E(\cdot|Z_t = z_t)$ and $P(\cdot|Z_t = z_t)$, respectively, where z_t is the realization of the state at time t. We investigate $E_{t-1}[H_{t,Z_t}|z_t]$ as follows:

$$E_{t-1}(H_{t,Z_t}|z_t) = E_{t-1}([w_{t,z_t}(a_{0,z_t} + a_{1,z_t}y^2_{t-1}) + (1 - w_{t,z_t})(b_{0,z_t} + b_{1,z_t}y^2_{t-1})]|z_t)$$
$$= \underbrace{b_{0,z_t}}_{I} + \underbrace{b_{1,z_t}E_{t-1}[y^2_{t-1}|z_t]}_{II} + \underbrace{(a_{0,z_t} - b_{0,z_t})E_{t-1}[w_{t,z_t}|z_t]}_{III}$$
$$+ \underbrace{(a_{1,z_t} - b_{1,z_t})E_{t-1}[w_{t,z_t}y^2_{t-1}|z_t]}_{IV}. \tag{3.9}$$

The relation (II) in (3.9) can be interpreted as follows:

$$E_{t-1}[y^2_{t-1}|z_t] = \sum_{z_{t-i}=1}^{K} \int_{S_{\mathcal{I}_{t-1}}} y^2_{t-1} p(\mathcal{I}_{t-1}|z_t, z_{t-1}) p(z_{t-1}|z_t) d\mathcal{I}_{t-1}$$
$$= \sum_{z_{t-1}=1}^{K} p(z_{t-1}|z_t) E_{t-1}[y^2_{t-1}|z_{t-1}, z_t], \tag{3.10}$$

where $S_{\mathcal{I}_{t-1}}$ is the support of $\mathcal{I}_{t-1} = (y_1, \cdots, y_{t-1})$. Since the expected value of y^2_{t-1} conditional on the present state is independent of any future state, so

$$E_{t-1}[y^2_{t-1}|z_{t-1}, z_t] = E_{t-1}[y^2_{t-1}|z_{t-1}]. \tag{3.11}$$

Also using the tower property of the conditional expectation, that is $E[E(Y|X,Z)|X] = E(Y|X)$ [see Grimmett and Stirzaker (2001, p. 69)], we have

$$E_{t-1}[y^2_{t-1}|z_{t-1}] = E_{t-2}[E_{t-1}(y^2_{t-1}|\mathcal{I}_{t-2}, z_{t-1})|z_{t-1}]$$
$$= E_{t-2}[H_{t-1,Z_{t-1}}|z_{t-1}]. \tag{3.12}$$

The calculation of $E_{t-1}[w_{t,z_t}|z_t]$ and $E_{t-1}[w_{t,z_t}y^2_{t-1}|z_t]$ is a problem that can not be easily done, For this reason we will try to find an upper bound for them.

Upper bound to III. As $0 < w_{t,i} < 1$, so an upper bound for the relation III in (3.9) is obtained by

$$(a_{0,z_t} - b_{0,z_t})E_{t-1}[w_{t,z_t}|z_t] \leq (a_{0,z_t} - b_{0,z_t}) < \infty. \tag{3.13}$$

Upper bound to IV. Let $0 < M < \infty$ be a constant, so

$$E_{t-1}[w_{t,z_t}y^2_{t-1}|z_t] = E_{t-1}[w_{t,z_t}y^2_{t-1}I_{|y_{t-1}|<M}|z_t]$$
$$+ E_{t-1}[w_{t,z_t}y^2_{t-1}I_{|y_{t-1}|\geq M}|z_t]$$

in which

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

As by (2.3), $0 < w_{t,z_t} < 1$ and so

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

also

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

by (2.3),

$$\lim_{y_{t-1} \to +\infty} w_{t,z_t} = 0, \qquad \lim_{y_{t-1} \to -\infty} w_{t,z_t} = 0, \tag{3.14}$$

therefore according to the definition of limit at infinity, for a small number $\delta > 0$, there will exist a finite constant $M > 0$ such that if $y_{t-1} \geq M$, $|w_{t,z_t}| \leq \delta$ and if $y_{t-1} \leq -M$, $|w_{t,z_t}| \leq \delta$. Hence

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

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

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

and

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

Therefor

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

By replacing the obtained upper bounds and relations (3.10)-(3.12) in (3.9), the upper bound for $E_{t-1}(H_{t,Z_t}|z_t)$ is aquired as:

$$\begin{aligned} E_{t-1}(H_{t,Z_t}|z_t) &\leq a_{0,z_t} + (a_{1,z_t} - b_{1,z_t})M^2 \\ &+ \sum_{z_{t-1}=1}^{K} [b_{1,z_t} + (a_{1,z_t} - b_{1,z_t})\delta] p(z_{t-1}|z_t) E_{t-2}[H_{t,Z_{t-1}}|z_{t-1}], \end{aligned} \quad (3.15)$$

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

$$\mathbf{\Omega} = [a_{0,1} + (a_{1,1} - b_{1,1})M^2, \cdots, a_{0,K} + (a_{1,K} - b_{1,K})M^2)]', \quad (3.16)$$

be a vector with K component, $\mathbf{C}$ be a K-by-K matrix with elements

$$\{\mathbf{C}_{jk}\} = [b_{1,j} + (a_{1,j} - b_{1,j})\delta] \frac{\pi_k}{\pi_j} \{P_{kj}\}, \quad (3.17)$$

and

$$\mathbf{A}_t = [E_{t-1}(H_{t,1}|Z_t = 1), \cdots, E_{t-1}(H_{t,K}|Z_t = K)]', \quad (3.18)$$

be a K-by-1 vector.
Hence by (3.16)-(3.18) we have the following recursive inequality,

$$\mathbf{A}_t \leq \mathbf{\Omega} + \mathbf{C}\mathbf{A}_{t-1}, \quad t \geq 0. \quad (3.19)$$

with some initial conditions $\mathbf{A}_{-1}$.
Suppose $\rho(A)$ denotes the spectral radius of a matrix A, then we have the following theorem.

Theorem 3.1 *Let $\{Y_t\}_{t=0}^{\infty}$ follows the MS-CARCH model, defined by (2.1)-(2.3), the process is asymptotically stable in variance and* $\lim_{t\to\infty} E(Y_t^2) \leq \pi'(\mathbf{I} - \mathbf{C})\mathbf{\Omega}$ *if and only if* $\rho(\mathbf{C}) < 1$.

Proof 3.2 *[1], By recursive inequality (3.19),*

$$\boldsymbol{A}_t \leq \mathbf{\Omega} \sum_{i=0}^{t-1} \mathbf{C}^i + \mathbf{C}^t \boldsymbol{A}_0 := \boldsymbol{B}_t \quad (3.20)$$

By the matrix convergence theorem [23] , a necessary and sufficient condition for the convergence of $\boldsymbol{B}_t$ when $t \to \infty$ is $\rho(\mathbf{C}) < 1$ (the value of δ can be considered small enough to be negligible). Under this condition, $\mathbf{C}^t$ converges to zero as t goes to ∞ and $\sum_{i=0}^{t-1} \mathbf{C}^i$ converges to $(I-\mathbf{C})^{-1}$ provided that matrix $(I-\mathbf{C})$ is invertible. So if $\rho(\mathbf{C}) < 1$,

$$\lim_{t\to\infty} \boldsymbol{A}_t \leq (I - \mathbf{C})^{-1}\boldsymbol{\Omega}$$

and by (3.8) we attain the upper bound for the asymptotic behaviour of unconditional variance,

$$lim_{t\to\infty} E(y_t^2) \leq \pi'(\mathbf{I} - \mathbf{C})\boldsymbol{\Omega}.$$

4 Estimation

In this section we describe the estimation of the parameters of the MS-CARCH model. We consider Bayesian MCMC method using Gibbs algorithm by following methods of sampling of a hidden Markov process ([9] and [21]), MS-GARCH model and weighted GARCH model ([5] and [6]) for estimation of parameters.

Let $Y_t = (y_1, \cdots, y_t)$ and $Z_t = (z_1, \cdots, z_t)$. For the case of two states, the transition probabilities are $\eta = (\eta_{11}, \eta_{12}, \eta_{21}, \eta_{22})$ and the parameters of the model are $\theta = (\theta_1, \theta_2)$, where $\theta_k = (a_{0k}, b_{0k}, a_{1k}, b_{1k}, \gamma_k)$ for $k = 1, 2$.

The purpose of Bayesian inference is to simulate from the distributions of the parameters and the state variables given the observations. As $Z = (z_1, \cdots, z_T)$ and $Y = (y_1, \cdots, y_T)$ the posterior density of our model is:

$$p(\theta, \eta, Z|Y) \propto p(\theta, \eta)p(Z|\theta, \eta)f(Y|\theta, \eta, Z), \tag{4.21}$$

in which $p(\theta, \eta)$ is the prior of the parameters. The conditional probability mass function of Z given the (θ, η) is independent of θ, so

$$\begin{aligned} p(Z|\theta, \eta) =& p(Z|\eta_{00}, \eta_{11}) \\ &= \prod_{t=1}^{T} p(z_{t+1}|z_t, \eta_{00}, \eta_{11}) \\ &= p_{00}^{n_{00}}(1 - p_{00})^{n_{01}} p_{11}^{n_{11}}(1 - p_{11})^{n_{10}}, \end{aligned} \tag{4.22}$$

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

$$f(Y|\eta, \theta, Z) = \prod_{t=1}^{T} f(y_t|\theta, z_t = k, Y_{t-1}), \quad k = 1, 2, \tag{4.23}$$

where the one step ahead predictive densities are:

$$f(y_t|\theta, z_t = k, Y_{t-1}) = \frac{1}{\sqrt{2\pi H_{t,k}}} \exp(-\frac{y_t^2}{H_{t,k}}). \tag{4.24}$$

Since the posterior density (4.21) is not standard we can not sample it in a straightforward manner. Gibbs sampling of Gelfand and Smith [14] is a repetitive algorithm to sample consecutively from the posterior distribution. Under regularity conditions, the simulated distribution converges to the posterior distribution, (see e.g Robert and Casella [27]). The blocks of parameters are θ, η and the realizations of Z.

A brief description of the Gibbs algorithm: Let use the superscript (r) on Z, θ and η to denote the estimators of Z_t), η, and θ at the r-th iteration of the algorithm. Each iteration of the algorithm consist of three steps:

(i) Drawing an estimator random sample of the state variable $Z^{(r)}$ given , $\eta^{(r-1)}$, $\theta^{(r-1)}$.

(ii) Drawing a random sample of the transition probabilities $\eta^{(r)}$ given $Z^{(r)}$.

(iii) Drawing a random sample of the $\theta^{(r)}$ given Z^r and $\eta^{(r)}$.

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

4.1 Sampling z_t

The purpose of this step is to obtain the sample of $p(z_t|\eta, \theta, Y_t)$ that is performed by Chib[9], (see also [21]). Suppose $p(z_1|\eta, \theta, Y_0,)$ be the stationary distribution of the chain,

$$p(z_t|\eta, \theta, Y_t) \propto f(y_t|\theta, z_t = k, Y_{t-1}) p(z_t|\eta, \theta, Y_{t-1}), \tag{4.25}$$

where the predictive density $f(y_t|\theta, z_t = k, Y_{t-1})$ is calculated by the relation (4.24) and by the law of total probability $p(z_t|\eta, \theta, Y_{t-1})$ is given by:

$$p(z_t|\eta, \theta, Y_{t-1}) = \sum_{z_{t-1}=1}^{K} p(z_{t-1}|\eta, \theta, Y_{t-1}) \eta_{z_{t-1} z_t}. \tag{4.26}$$

Given the filter probabilities ($p(z_t|\eta, \theta, Y_t)$), we run a backward algorithm, starting from $t = T$ that z_T is derived from $p(z_T|\eta, \theta, Y)$. For $t = T-1, \cdots, 0$ the sample is derived from $p(z_t|z_{t+1}, \cdots, z_T, \theta, \eta, Y)$,which is obtained by

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

To derive z_t from $p(z_t|\cdot) = p_{z_t}$ is by sampling from the conditional probabilities $q_j = p(Z_t = j|Z_t \geq j, .)$ which are given by

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

After generating a uniform (0,1) number U, if $U \leq q_j$ then $z_t = j$, otherwise increase j to $j+1$ and generate another uniform (0,1) and compare it by 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}. For example,

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

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.

4.3 Sampling θ

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

$$p(\theta|Y, Z, \eta) \propto p(\theta)\prod_{t=1}^{T} f(y_t|\theta, z_t = k, Y_{t-1}) = p(\theta)\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi H_{t,k}}}\exp(-\frac{y_t^2}{H_{t,k}}), \quad (4.27)$$

which is independent of η. Since the conditional distribution of θ does not have a closed-form (because for example $p(a_{0k}|Y_t, Z_t, \theta_{-a_{0k}})$, in which $\theta_{-a_{0k}}$ is the parameter vector without a_{0k}, contains $H_{k,i}$, which is also a function of a_{0k}. Therefor it isn't a normal density.) using the Gibbs sampling in this situation may be complicated. The Griddy Gibbs algorithm, that introduced by Ritter and Tanner (1992), can be a solution of this problem. This method is very applicable in researches (for example [4] , [5] and [6]).

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

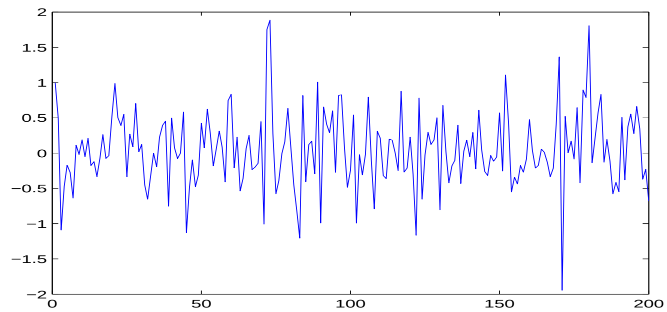

Figure 1: Simulated time series of MS-CARCH model.

Table 1: Descriptive statistics for the simulated data (sample size=200)

Mean	Standard deviation	Skewness	Maximum	Minimum	Kurtosis
.051	.632	.445	2.893	-1.833	5.777

1. Select a grid of points, such as $a_{0i}^1, a_{0i}^2, \cdots, a_{0i}^G$. Using (4.27), evaluate the conditional posterior density function $k(a_{0i|Z_t,Y_t,\theta_{-a_{0i}}})$ over the grid points to obtain the vector $G_k = (k_1, \cdots, k_G)$.
2. By a deterministic integration rule using the G points, compute $G_\Phi = (0, \Phi_2, \cdots, \Phi_G)$ with

$$\Phi_j = \int_{a_{0i}^1}^{a_{0i}^j} k(a_{01}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t) da_{0i}, \quad i = 2, \cdots, G. \tag{4.28}$$

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 $a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t$.
4. Repeat steps 1-3 for other parameters.

For the prior densities of all elements of θ, it can be can considered independent uniform densities over the finite intervals.

5 Simulation results

In this section we provide some simulation results of MS-CARCH model defined by equations (2.1)-(2.3) for two states. We simulate 200 sample from the following MS-CARCH model:

$$y_t = \varepsilon_t \sqrt{H_{t,Z_t}}, \tag{5.29}$$

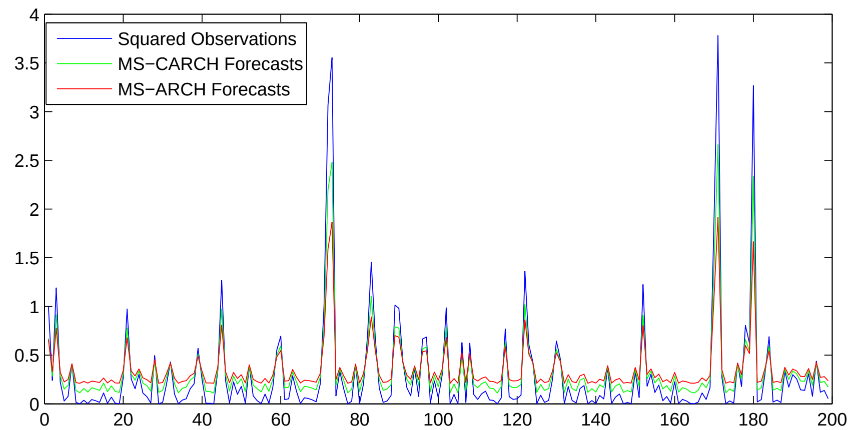

Figure 2: Squared observations of the simulated time series (blue), forecasts by MS-ARCH model (red) and forecasts by MS-CARCH model (green)

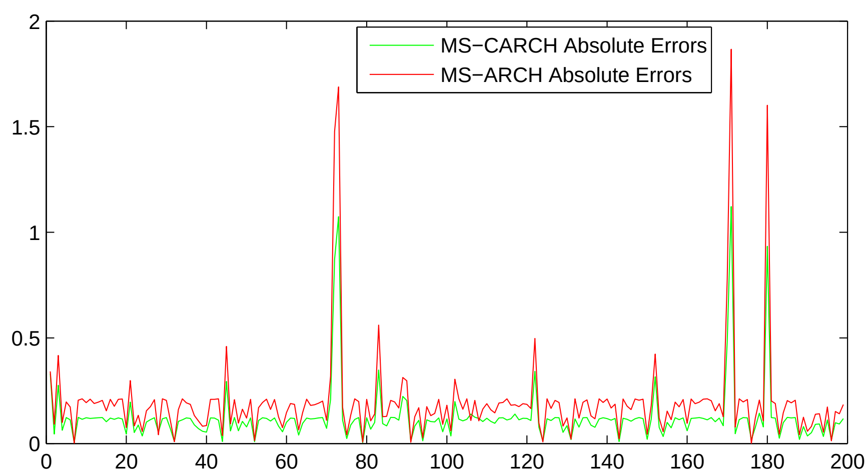

Figure 3: Forecast Errors of square of the observations in the MS-ARCH model (red) and in MS-CARCH model (green).

Table 2: Results of the Bayesian Estimation of the simulated MS-ARCH model

	True values	Mean	Std. dev.
a_{01}	0.10	0.097	0.049
a_{11}	0.40	0.397	0.164
b_{01}	0.20	0.199	0.053
b_{11}	0.90	0.845	0.190
a_{02}	0.10	0.158	0.081
a_{12}	0.05	0.124	0.067
b_{02}	0.30	0.298	0.110
b_{12}	0.20	0.200	0.108
γ_1	5	5.994	.272
γ_2	0.10	0.106	0.057
η_{11}	0.90	0.892	0.078
η_{22}	0.85	0.709	0.159

where $\{\varepsilon_t\}$ is an iid sequence of standard normal variables, $\{Z_t\}$ is a Markov chain on finite state space $E = \{1, 2\}$ with transition probability matrix

$$P = \begin{pmatrix} .90 & .10 \\ .05 & .85 \end{pmatrix},$$

and

$$\begin{aligned} H_{t,1} =& (1 - w_{t,1})(.1 + .4y_{t-1}^2) + w_{t,1}(.2 + .9y_{t-1}^2), \quad w_{t,1} = \frac{1 - \exp(-5|y_{t-1}|)}{1 + \exp(-5|y_{t-1}|)} \\ H_{t,2} =& (1 - w_{t,2})(.1 + .05y_{t-1}^2) + w_{t,2}(.3 + .2y_{t-1}^2), \quad w_{t,2} = \frac{1 - \exp(-.1|y_{t-1}|)}{1 + \exp(-.1|y_{t-1}|)} \end{aligned} \tag{5.30}$$

The first state implies a higher conditional variance than the second one and in each state, the first component has the lower volatility than another component. By theorem (1), the assumption for the exictence an upper bound for the second moment is checked.

In table 1, we report summery statistics for simulated data and figure 1 shows the plot of the simulated time series.

Using the Bayesian inference, we estimate the parameters of the MS-CARCH model. The prior density of each parameter is assumed to be uniform restricted over a finite interval (except for η_{11} and η_{22}, since they are drawed from the beta distribution). The number of iterations of the Gibbs sampler was 20000 and the initial 10000 draws were discarded. Table 2 demonstrates the performance of our estimation methods for the model. The results of these tables show that the standard deviation are small enough in most cases.

For clarifying the performane of MS-CARCH model toward MS-ARCH model, We compare the forecasting volatility ($E(Y_t^2|\mathcal{F}_{t-1})$) of each model to the squared observations. The forecast error (the difference between the forecasting volatility and the squared observations) of our model is more smaller than the MS-ARCH model especially in picked points.

6 Conclusion

In this paper a generalization of the MS-ARCH model has been presented where the conditional variance in each state is a convex combination of two different ARCH components with time varying coefficients, one of the component with higher volatility than other component. Our model can providemore better forecast of volatility toward MS-ARCH model. For the estimation of parameters we have applied the Bayesian estimation algorithm. We provide simple necessary and sufficient condition for the existence of an upper bound for second moment.

This work has the potential to be applied in the context of financial time series. The empirical distribution of daily returns doesn't generally have a Gaussian distribution. They have fat tails densities (they are called leptokurtic). One of the extending of this work is considering the fat tail densities instead of Gaussian distribution, that can cause better modeling of the financial time series. Also we can generalize this model by using of the GARCH structure instead of the ARCH structure for the better results.

References

[1] Abramson, A. Cohen, I. (2007). On the stationarity of Markov-Switching GARCH processes, *Econometric Theory*, 23, 485-500.

[2] Alexander, C. Lazar, E. (2008). Markov-Switching GARCH diffusion, *ICMA Center Discussion Paper in Finance 2008-01*, University of Reading.

[3] Alizadeh, S. H. Rezakhah, S. (2013). Hidden Markov mixture autoregressive models:stability and moments,*Communications in Statistics- Theory and Methods*, 42:6, 1087-1104.

[4] Bauwens, L. Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics journal*, 1, 23-46.

[5] Bauwens, L. Storti, G. (2009). A component GARCH model with time varying weights, *Studies in Nonlinear Dynamics and Econometrics*, 13 (2), Article 3.

[6] Bauwens, L. Preminger, A. Rombouts, V.K. (2010). Theory and inferencefor Markov switching GARCH model , *Econometrics journal*, 13, 218-244.

[7] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.

[8] Cai, J. (1994). A markonv model of switching -regime ARCH, *Journal of Business and Economic Statistics*, 12, 309-316.

[9] Chib, S. (1996). Calculating posterior distributions and model estimates in Markov mixture models, *Journal of Econometrics*, 75, 79-97.

[10] Ding, Z. Granger, C.W.J. (1996).Modeling volatility persistence of speculative returns: a new approach, *Journal of Econometrics*, 173, 185-215.

[11] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica*, 50, 987-1007.

[12] Engle, R. F. Lee, G.G.J. (1999). A long-run and short-run component model of stock return volatility, *Cointegration, Causality, and Forecasting (Engle R.F. and White H., eds)*, Oxford University press. 3, 103-106.

[13] Engle, R. F. Rangel, J. (2008). The Spline-GARCH model for low frequency volatility and its global macroeconomic causes , *Review of Financial Studies*, 21, 1187-1222.

[14] Gelfand, A. Smith, A. (1990). Sampling based approaches to calculating marginal densities, *Journal of American Statistical Assosiation*, 85, 398-409.

[15] Gonzalez-Rivera, G. (1998). Smooth transition GARCH models, Studies in Nonlinear Dynamics and Econometrics, 3, 61-78.

[16] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, 42, 27-62.

[17] Grimmett, G. Stirzaker, D. (2001). *Probability and random processes*, Oxford University press, New York.

[18] Haas, M. Mittnik, S. Paolella, M.S. (2004). A new approach to markov-switching GARCH models, *Journal of Financial Econometrics*, 2, 493-530.

[19] Hamilton, J. D. Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, 64, 307-333.

[20] Hagerud, G. (1997). A smooth transition ARCH model for asset returns, SSE/EFI, Working Paper Series in Economic and Finance, No.162.

[21] Kaufman, S. Fruhwirth-Schnatter, S. (2002). Bayesian analysis of switching ARCH models , *Journal of Time Series Analysis*, 23, 425-458.

[22] Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH, *Empirical Economics* 27, 363-394.

[23] Lancaster, P. Tismenetsky, M. (1985). *The theory of matrices, 2nd ed*, Academic press.

[24] lubrano, M. (2001). Smooth transition GARCH models: a Bayesian approach mixture models, *Recherches Economiques de Louvain*, 67, 257-287.

[25] Maheu, J. (2005). Can GARCH models capture long range dependence?, *Studies in Nonlinear Dynamics and Econometrics*, 9 (4), Article 1.

[26] Medeiros, M.C. Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1), *Journal of Econometric Theory*, 25, 117-161.

[27] Robert, C. Casella, G. (2004). *Monte carlo statistical methods*, New York: Springer.