Available online at http://www.mecs-press.net/ijem

On a GARCH Model with Normal Scale Mixture Innovations

Feng Feng [*]

School of Management, Fuzhou University, Fuzhou, Fujian Province, China

Abstract

Recently, there has been a lot of interest in modeling real data with a heavy tailed distribution. A popular candidate is the so-called generalized autoregressive conditional heteroscedastic (GARCH) model. Unfortunately, the tails of normal GARCH models are not thick enough in some applications. In this paper, we propose a GARCH model with normal scale mixture innovations, the parameters estimation procedure using EM algorithm is also provided. It is shown that GARCH models with normal scale mixture innovations have tails thicker than those of normal GARCH models. Therefore, the GARCH models with normal scale mixture innovations are more capable of capturing the heavy-tailed features in real data. Shanghai Stock Market Index as a real example illustrates the results.

Index Terms: GARCH model; Normal scale mixture; EM algorithm; Tail behavior; Volatility Clustering

1. Introduction

The volatility of returns has important theoretical and practicality meaning in the applications of risk management and asset pricing. Volatility clusters (i.e. Volatility may be high for certain periods and low for other periods) are commonly seen in financial time series. Also, financial time series, which marginal distribution have higher excess kurtosis, typically exhibit the feature of heavy tail.

In the past two decades, various models have been proposed in order to describe these features. Among them, the generalized autoregressive conditional heteroscedastic (GARCH) model of Bollerslev (1986) has been proven to be a powerful one in capturing the empirical features. The GARCH(p,q) model is defined as

$$\begin{cases} X_t = \varepsilon_t \sqrt{h_t}, \\ h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, t \in N, \end{cases} \tag{1}$$

* Corresponding author.
E-mail address: ffeng@126.com

Where $p, q \geq 0$ are integers; $\alpha_0 > 0$, $\alpha_j, \beta_i \geq 0$ are constants , to ensure positivity of h_t, for all t; ε_t are i.i.d. with zero mean and unit variance; and $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$, to ensure covariance stationarity. The random variables ε_t are called the innovations and usually assumed to be standard-normal. When $p = 0$, the model reduces to the ARCH model of Engle (1982).

Unfortunately, the normal GARCH models seem to be much thinner than the tails apparent in the data, that is to say, the normal GARCH model can't interpret all the excess kurtosis. A popular candidate is the GARCH-t model (Bollerslev, 1987), which used the t distribution innovations assumption. However, the discussion of GARCH process excess kurtosis makes some senses only if the freedom of t innovation distribution greater than 4, otherwise, the ability of capture the tails behavior for the model will be cut down. Other papers dealing with alternative mixture GARCH models are Wong and Li (2001), Zhang (2006).

In the statistics literature, mixtures of distributions have been widely used in modeling of heavy-tailed distributions, and noticed the fact that we can obtain heavier tail marginal distribution than the innovations' distribution by the GARCH process, we considered a natural extension of Bollerslev (1986) by using NSM innovations instead of normal innovations in this paper.

The rest of this paper is organized as follows. Section 2 presents the GARCH model with mixture normal innovations and illustrates its flexibility in capturing the characteristics in financial time series. Section 3 describes the parameters estimation procedure using EM algorithm for the model. Section 4 illustrates our procedure for the log return series of the Shanghai Market Index which is a clear example of a series large kurtosis and extreme returns.

2. The GARCH model with normal scale mixture innovations

2.1. Normal Scale Mixture Distribution

Mixture Models are a type of density model which comprise a number of component functions, usually Gaussian. The two component normal scale mixture (NSM) distribution under considered is defined by

$$f(y; \sigma^2, \lambda, \rho) = \rho f_1(y) + (1 - \rho) f_2(y) \tag{2}$$

Where $f_1 \sim N(0, \sigma^2)$, $f_2 \sim N(0, \sigma^2/\lambda)$ and $0 < \rho, \lambda < 1$.

An alternative definition of $Y \sim NSM(\lambda, \rho)$ as

$$Y \sim \begin{cases} N(0, \sigma^2) & \text{with probability } \rho, \\ N(0, \sigma^2/\lambda) & \text{with probability } 1 - \rho. \end{cases} \tag{3}$$

This family was introduced to model a population which follows a normal distribution except on those few occasions where a grossly atypical observation is recorded.

We can deduce an important property of NSM distribution concerned on its kurtosis by calculating some expectations.

THEOREM 1.   Suppose $Y \sim NSM(\lambda, \rho)$, $0 < \rho, \lambda < 1$. Then the kurtosis of Y is given by

$$K_Y = 3 + K_e > 3 \text{ , where } K_e = \frac{3\rho(1-\rho)((1/\lambda)-1)^2}{[\rho + (1-\rho)/\lambda]^2}.$$

Proof   Since $E(Y^2) = \int_{-\infty}^{+\infty} y^2 f(y; \rho, \lambda) dy$

$$= \rho \int_{-\infty}^{+\infty} \frac{y^2}{\sigma} \varphi(\frac{y}{\sigma}) dy + (1-\rho) \int_{-\infty}^{+\infty} \frac{y^2}{\sigma/\sqrt{\lambda}} \varphi(\frac{y}{\sigma/\sqrt{\lambda}}) dy$$

$$= \rho \int_{-\infty}^{+\infty} t_1^2 \sigma^2 \frac{1}{\sigma} \phi(t_1) \sigma dt_1 + (1-\rho) \int_{-\infty}^{+\infty} t_2^2 \frac{\sigma^2}{\lambda} \frac{1}{\sigma/\sqrt{\lambda}} \phi(t_2) \frac{\sigma}{\sqrt{\lambda}} dt_2$$

$$= \rho \sigma^2 + (1-\rho) \frac{\sigma^2}{\lambda}, \text{ and } E(Y^4) = \int_{-\infty}^{+\infty} y^4 f(y; \rho, \lambda) dy$$

$$= \rho \int_{-\infty}^{+\infty} \frac{y^4}{\sigma} \phi(\frac{y}{\sigma}) dy + (1-\rho) \int_{-\infty}^{+\infty} \frac{y^4}{\sigma/\sqrt{\lambda}} \phi(\frac{y}{\sigma/\sqrt{\lambda}}) dy$$

$$= \rho \int_{-\infty}^{+\infty} t_3^4 \sigma^4 \frac{1}{\sigma} \phi(t_3) \sigma dt_3 + (1-\rho) \int_{-\infty}^{+\infty} t_4^4 \frac{\sigma^4}{\lambda^2} \frac{1}{\sigma/\sqrt{\lambda}} \phi(t_4) \frac{\sigma}{\sqrt{\lambda}} dt_4$$

$$= 3\rho(\sigma^2)^2 + 3(1-\rho)(\sigma^2/\lambda)^2, \text{ Thus we have } K_Y = \frac{E(Y^4)}{[E(Y^2)]^2} = \frac{3\rho + 3(1-\rho)(1/\lambda)^2}{[\rho + (1-\rho)/\lambda]^2}$$

$$= 3 + \left[\frac{3\rho + 3(1-\rho)(1/\lambda)^2}{[\rho + (1-\rho)/\lambda]^2} - 3 \right]$$

$$= 3 + \frac{3\rho(1-\rho)(1/\lambda - 1)^2}{[\rho + (1-\rho)/\lambda]^2} > 3.$$

2.2. The Definition of GARCH Model with Normal Scale Mixture Innovations

We consider the innovations in equation (1) follow normal scale mixture distribution, that is $\varepsilon_t \sim NSM(\lambda, \rho)$, so we have the GARCH(p,q)- $NSM(\lambda, \rho)$ model defined by

$$\begin{cases} X_t = \varepsilon_t \sqrt{h_t}, \\ h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \\ \varepsilon_t \ i.i.d. \sim NSM(\lambda, \rho), t \in N, \end{cases} \tag{4}$$

Where $p, q \geq 0$ are integers; $0 < \lambda, \rho < 1$; $\alpha = (\alpha_0, \ldots, \alpha_q)^T$ and $\beta = (\beta_1, \ldots, \beta_p)^T$ follow the restrictions, $\alpha_0 > 0$, $\alpha_j, \beta_i \geq 0$ are constants, $\sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1$; and $\sigma^2 = 1/(\rho + (1-\rho)/\lambda)$, so that $Var(\varepsilon_t) = 1$.

According to the definition, the innovations ε_t are generated from a normal distribution with variance σ^2 with probability ρ, or from a normal distribution with variance σ^2/λ with probability $1-\rho$. We also impose the condition that the probability ρ is restricted to the interval (0.5, 1) to ensure that the component with largest number of elements is the one with smallest variance.

The reason of using the NSM distribution to model the innovations instead of a Student's t distribution are as follows. If ε_t follows a Student's t distribution with v degrees of freedom, then $K_\varepsilon = 6/v - 4$. Note that

the second and fourth moments of X_t only exist if $v > 4$, implying that the excess kurtosis K_ε should be positive. However, in practice, the degrees of freedom parameter, v is either fixed to be larger than or equal to 5, in which case the implied kurtosis of the estimated model does not match the observed kurtosis, or it is estimated, in which case its estimate is usually smaller than 5, and the estimated excess kurtosis does not exist.

3. Estimation via EM algorithm

In this section, we discuss the estimation of the parameters of a GARCH model with normal scale mixture innovations using the EM algorithm (Dempster et al., 1977). Suppose that the observed time series $r = \{r_1, \ldots, r_T\}$ is generated from the following model (5).

$$
\begin{cases}
r_t = \mu_t + X_t, \ X_t = \varepsilon_t \sqrt{h_t}, \\
h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j X_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \\
\varepsilon_t \ i.i.d. \sim NSM(\lambda, \rho). \ t = 1, \ldots, T.
\end{cases}
\tag{5}
$$

where μ_t is a stationary mean process of $\{r_t\}$, other notations are same as (4), (5). Let $z = \{z_1, \ldots, z_T\}$ be the unobserved indicators, where z_t equals to 1 if $\varepsilon_t \sim N(0, \sigma^2)$ and equals to 0 if $\varepsilon_t \sim N(0, \sigma^2/\lambda)$. Then, conditional on these indicators, we have that,

$$
r_t | h_t, z_t \sim
\begin{cases}
N(\mu_t, \sigma^2 h_t) & \text{if } z_t = 0, \\
N(\mu_t, \sigma^2 h_t / \lambda) & \text{if } z_t = 1,
\end{cases}
\qquad t = 1, \ldots, T.
\tag{6}
$$

We call r the incomplete data with missing data z. Let $\theta = (\lambda, \rho, \alpha^T, \beta^T)^T$ be the parameter vector of model (5). The likelihood is given by

$$
L(\theta | r, z) = \prod_{t=c+1}^{T} \left\{ \left[\frac{\rho}{\sigma \sqrt{h_t}} \varphi\left(\frac{r_t - \mu_t}{\sigma \sqrt{h_t}} \right) \right]^{z_t} \times \left[\frac{(1-\rho)\sqrt{\lambda}}{\sigma \sqrt{h_t}} \varphi\left(\frac{(r_t - \mu_t)\sqrt{\lambda}}{\sigma \sqrt{h_t}} \right) \right]^{1-z_t} \right\}
\tag{7}
$$

where $c = \max(p, q)$; $\phi(\cdot)$ is standard normal density function, For simplicity, we denote that

$$
f_{1t} = \frac{1}{\sigma \sqrt{h_t}} \varphi\left(\frac{r_t - \mu_t}{\sigma \sqrt{h_t}} \right), f_{2t} = \frac{\sqrt{\lambda}}{\sigma \sqrt{h_t}} \varphi\left(\frac{(r_t - \mu_t)\sqrt{\lambda}}{\sigma \sqrt{h_t}} \right), f_t = \rho f_{1t} + (1 - \rho) f_{2t}, \text{ then the joint density of } r_t
$$

and z_t will be $g(r_t, z_t; \theta) = \left[\rho f_{1t} \right]^{z_t} \left[(1 - \rho) f_{2t} \right]^{1-z_t}$ and the conditional distribution of z_t conditional on r_t given by

$$
P(z_t = 1 | r_t; \theta) = \rho \frac{f_{1t}}{f_t}, \qquad P(z_t = 0 | r_t; \theta) = (1 - \rho) \frac{f_{2t}}{f_t}, \qquad t = 1, \ldots, T
\tag{8}
$$

According to (8), we can obtain the (conditional) log-likelihood as follows

$$
l = \sum_{t=c+1}^{T} \ln g(r_t, z_t; \theta)
\tag{9}
$$

The iterative EM procedure for estimating the parameters by maximizing the log-likelihood function (9) consists of an E-step and an M-step. These steps are described in the following.

E-step: Let $\theta^{(0)}$ be the value specified initially for θ, Then on the (m)th iteration of the EM algorithm, the E-step requires the computation of the conditional expectation of l given z, using $\theta^{(m-1)}$ for θ, which can be written as

$$Q(\theta, \theta^{(m-1)}) = \sum_{t=c+1}^{T} E_{(z_t|r_t;\theta^{(m-1)})} \left[\ln(g(r_t, z_t; \theta)) \right]$$ (10)

$$= \sum_{t=c+1}^{T} \left[\frac{\rho^{(m-1)} f_{1t}^{(m-1)}}{f_t^{(m-1)}} \ln(\rho f_{1t}) + \frac{(1-\rho^{(m-1)}) f_{2t}^{(m-1)}}{f_t^{(m-1)}} \ln((1-\rho) f_{2t}) \right],$$

where $\theta^{(m-1)} = (\lambda^{(m-1)}, \rho^{(m-1)}, \alpha^{(m-1)^T}, \beta^{(m-1^T)})^T$, $f_t^{(m-1)} = f_t(\theta^{(m-1)})$, $f_{1t}^{(m-1)} = f_{1t}(\theta^{(m-1)})$, $f_{2t}^{(m-1)} = f_{2t}(\theta^{(m-1)})$, $f_t = f_t(\theta)$, $f_{1t} = f_{1t}(\theta)$, $f_{2t} = f_{2t}(\theta)$;

M-step: The M-step on the (m)th iteration requires the global maximization of $Q(\theta, \theta^{(m-1)})$ with respect to θ over the parameter space Ω to give the updated estimated $\theta^{(m)}$, which can be obtained by solve the equations $\frac{\partial Q(\theta, \theta^{(m-1)})}{\partial \theta} = 0$

The estimates of the parameters θ are obtained by iterating the E and M steps until convergence.

4. An example

The word "data" is plural, not singular. In American Eng As an illustration, we select the Shanghai stock market daily closing index from December 16th, 1996, to April, 27th, 2007 (2500 observations). The computational results and the analysis of the real data example in this section have been carried out by means of various routines written by the author in MATLAB (The MathWorks, Inc.). The sample mean, variance and kurtosis coefficient of the log return series are 5.30×10^{-4}, 2.46×10^{-4}, 8.97 respectively. Note the large sample kurtosis of the returns.

We first fit a normal GARCH model to the log return series, in which innovations assumed to be standard-normal. It is also the best model in the sense that it minimizes the AIC and BIC over all the normal GARCH models. The estimated model is given in the following (11),

$$\begin{cases} r_t = 0.34745 r_{t-3} + X_t, X_t = \varepsilon_t \sqrt{h_t}, \\ \quad (1.7085) \\ h_t = 6.2585 \times 10^{-6} + 0.1162 X_{t-1}^2 + 0.8647 h_{t-1}, \\ \quad (7.4987) \qquad (16.4347) \qquad (144.4248) \\ \varepsilon_t \ i.i.d. \sim N(0,1). \ t = 1, 2, \ldots, 2499, \\ AIC = -1.4181 \times 10^4, BIC = -1.4158 \times 10^4. \end{cases}$$ (11)

where the numbers in brackets are T statistics values. According to the upper graph in Fig.1, in which displays a quantile-quantile plot of stand residuals versus standard normal distribution in model (11), it is not difficult to find that model (11) is poor in capturing the heavy-tailed features in the data.

The inadequate of model (11) suggest us consider a GARCH model with heavy-tailed innovations distribution for the data. A candidate can be GARCH model with normal scale mixture innovations which provided in (4). The GARCH model with normal scale mixture innovations can be estimated using the EM algorithm in part 2 for detail and given in the following (12),

$$
\begin{cases}
r_t = 0.06550 r_{t-3} + X_t, \; X_t = \varepsilon_t \sqrt{h_t}, \\
h_t = 1.01502 \times 10^{-6} + 0.14239 X_{t-1}^2 + 0.71540 h_{t-1}, \\
\varepsilon_t \; i.i.d. \sim \begin{cases} N(0, 0.77924^2) & , with\; prob.\; 0.81188, \\ N(0, 0.77924^2 / 0.22531), with\; prob.\; 1 - 0.81188, \end{cases} \\
t = 1, 2, \ldots, 2499, \\
AIC = -1.4237 \times 10^4, BIC = -1.4202 \times 10^4.
\end{cases}
\tag{12}
$$

For purposes of comparison the two candidate models, we plot the quantile-quantile plot of stand residuals versus innovations' distribution each other and put together in Fig 1. Obviously, the GARCH model with normal scale mixture innovations is much better than the normal GARCH model for fitting this series. On the other hand, the BIC and AIC model selection criteria also give the same choice.
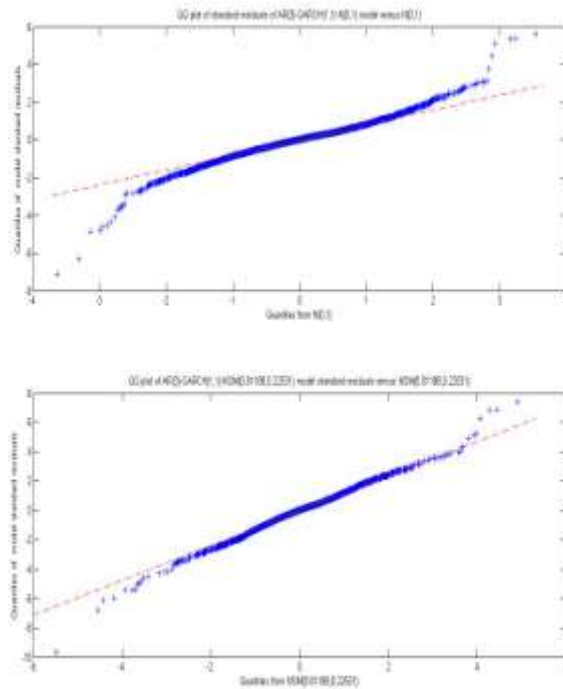

Fig. 1 PP plots of the two comparative models

5. Conclusion

The proposed GARCH model with normal scale mixture innovations provides an extension of the standard GARCH model. Its tail behavior can be thicker than that of the normal GARCH model. The GARCH Model with Normal Scale Mixture Innovations may be worth considering if accurate modeling of the tail is important in such applications like the estimation of value-at-risk.

Acknowledgements

References

[1] Bollerslev, "Generalised autoregressive conditional heteroscedasticity," J. Econometrics, vol. 51, pp. 307–327, 1986.

[2] Bollerslev, "A conditional heteroscedastic time series model for speculative prices and rates of return," Rev. Econom Statist, Vol. 69, pp.542–547,1987.

[3] Chun Shan Wong and Wai Keung Li, "On a mixture autoregressive model," Journal of Royal Statistical Society, vol. 62, pp.95–115,2000.

[4] Chun Shan Wong and Wai Keung Li, "On a mixture autoregressive conditional heteroscedastic model," Journal of American Statistical Association, vol. 96, pp.982–995,2001.

[5] Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of the UK inflation," Econometrica, vol. 50, pp.987–1008,1982.

[6] Mar´ıa Concepci´on Aus´ın Pedro Galeano, "Bayesian estimation of the gaussian mixture GARCH model," Computional Statistics & Data, vol. 51, pp.2636–2652,2007.

[7] Pi Liu-yi, Liu Zhong and Mao Shi-song, "Probability Distribution Analysis on Market Value, Amount and Varities of Stocks," Chinese Journal of Applied Probability and Statistics, vol. 14, pp.386–394,1998.

[8] Zhang Z., W.K. Li and K.C. Yuen, "On a mixture GARCH time series model," Journal of Time Series Analysis, vol. 27, pp.577–597,2006