Tang, H, Chiu KC and Lei Xu, (2003), Proc. of 3rd International Workshop on Computational Intelligence in Economics and Finance (CIEF'2003), North Carolina, USA, September 26-30, 2003, pp.1112-1119.

Finite Mixture of ARMA-GARCH Model for Stock Price Prediction

Him Tang, Kai-Chun Chiu and Lei Xu *
Department of Computer Science and Engineering, The Chinese University of Hong Kong
Shatin, New Territories, Hong Kong, P. R. China
{htang,kcchiu,lxu}@cse.cuhk.edu.hk

Abstract

In the literature, the finite mixture of autoregressive (AR), finite mixture of autoregressive moving average (ARMA) and finite mixture of autoregressive generalized autoregressive conditional heteroscedasticity (AR-GARCH) models have been respectively adopted for finance exchange rate prediction. In this paper, we consider to extend the mixture of AR-GARCH model (W.C. Wong, F. Yip and L. Xu, 1998) to the mixture of ARMA-GARCH model and investigate its application in stock price prediction. A generalized expectation-maximization (GEM) algorithm is proposed to learn the mixture model. Experimental simulations show that the mixture of ARMA-GARCH model yields better prediction results than either the mixture of AR, or the mixture of AR-GARCH models.

Introduction

Financial time series prediction deals with the task of modelling the underlying data generation process using past observations and using the model to extrapolate the time series into the future. Due to its intrinsic difficulty and widespread applications, much effort was devoted in the past few decades to the development and refining of financial forecasting models.

In the literature, two major classes of models were studied by econometricians for the purpose of forecasting. They are the statistical time series models and structural econometric models. Linear time series models such as the Box-Jenkins autoregressive integrated moving average (ARIMA) models were among the first to be developed and subsequently widely studied. Despite its simplicity and versatility in modelling several types of linear relationship such as pure autoregressive, pure moving average and autoregressive moving average (ARMA) series, such type of models was constrained by its linear scope.

However, real-world systems are seldom linear. To tackle the problem of nonlinear modelling, several finite mixture models have been studied. For example, the finite mixture of AR model was used for forecasting financial time series with satisfactory results obtained. Further empirical comparisons between the mixture of AR and mixture of ARMA models for predicting financial exchange data were also conducted (Kwok et al., 1998). Recently, a theoretical study on the mixture of AR model was shown in (Wong et al., 2000), which highlights its advantages over conventional AR model.

It is not uncommon that many financial time series is still conditionally heteroscedastic after stationarity transformation. To specifically deal with this intricacy, the well-known generalized autoregressive conditional heteroscedasticity (GARCH) model (Bollerslev, 1986) was adopted in (Wong et al., 1998) to derive the so-called mixture of AR-GARCH model used in exchange rate prediction. Motivated by fact that financial time series can demonstrate behaviors of both ARMA and GARCH, this paper explores the mixture of ARMA-GARCH model for stock price prediction. A generalized expectation-maximization (GEM) algorithm is proposed for model parameter learning.

The rest of the paper is organized in the following way. Section introduces the finite mixture of ARMA-GARCH model. Section derives the GEM algorithm for implementation. Experimental simulations of the model for stock price prediction are given in section . Section concludes the paper.

The Finite Mixture of ARMA-GARCH Model

The mixture of ARMA-GARCH model is similar to the mixture of AR-GARCH model proposed in (Wong et al., 1998). Specifically, each component of the mixture model can be denoted as a normal ARMA series

$$y_{t,j} = \sum_{r=1}^{R} b_{rj} y_{t-r,j} + \sum_{s=1}^{S} a_{sj}\epsilon_{t-s,j} + \epsilon_{t,j}, \qquad (1)$$

Furthermore, each residual term $\epsilon_{t,j}$ is assumed gaussian white noise with variance denoted by the GARCH model

$$\sigma^2_{t,j} = \delta_{0j} + \sum_{q=1}^{Q} \delta_{qj}\epsilon^2_{t-q,j} + \sum_{p=1}^{P} \beta_{pj}\sigma^2_{t-p,j}, \qquad (2)$$

where $\delta_{qj} > 0$ for $q = 1, \ldots, Q$ and $\beta_{pj} > 0$ for $p = 1, \ldots, P$.

*The work described in this paper was fully supported by a grant from the Research Grant Council of the Hong Kong SAR (Project No: CUHK 4184/03E).
CIEF 2003 (http://aiecon.org/conference/cief2003.html).

Mathematically, the finite mixture of ARMA-GARCH model can be denoted as a K-component gaussian mixture model

$$p(y_t) = \sum_{j=1}^{K} \alpha_j G(y_t; \hat{y}_{t,j}, \sigma^2_{t,j}), \quad (3)$$

$$\hat{y}_{j,t} = \sum_{r=1}^{R} b_{rj} y_{t-r,j} + \sum_{s=1}^{S} a_{sj}\epsilon_{t-s,j}, \quad (4)$$

where $\alpha_j > 0$ and $\sum_{j=1}^{K}\alpha_j = 1$.

Once the model has been learned, one-step ahead prediction can be done via taking expectation of y_t

$$E(y_t) = \alpha_1\hat{y}_{t,1} + \ldots + \alpha_K\hat{y}_{t,K}. \quad (5)$$

Derivation of the Generalized Expectation-Maximization (GEM) Algorithm for Implementation

The E step

The joint probability of the finite mixture model is

$$p(\mathbf{y}) = \prod_{t=1}^{T}\sum_{j=1}^{K} \alpha_j G(y_t; \hat{y}_{t,j}, \sigma^2_{t,j}), \quad (6)$$

and we want to maximize the log-likelihood of (6)

$$\hat{\Theta} = \arg\max_{\Theta} \ln p(\mathbf{y}) \quad (7)$$

where $\Theta = \{\{a_{sj}\}_{s=1}^{S}, \{b_{rj}\}_{r=1}^{R}, \{\beta_{pj}\}_{p=1}^{P}, \{\delta_{qj}\}_{q=1}^{Q}, \alpha_j\}_{j=1}^{K}$, $\mathbf{y} = \{y_t\}_{t=1}^{T}$.

We use the EM algorithm (Dempster et al., 1977) to maximize the log-likelihood. Let the unobserved variables are $\mathbf{Z} = \{Z_t\}_{t=1}^{T}$. We define that if y_t is produced by the uth component, then $Z_t = u$. And so

$$p(y_t|Z_t = u) = G(y_t; \hat{y}_{t,u}, \sigma^2_{t,u}), \quad (8)$$

$$p(Z_t = u) = \alpha_u. \quad (9)$$

The complete data joint distribution is

$$p(\mathbf{y}, \mathbf{Z}) = \prod_{t=1}^{T} p(y_t|Z_t)p(Z_t), \quad (10)$$

and its log-likelihood is

$$\ln p(\mathbf{y}, \mathbf{Z}) = \sum_{t=1}^{T} \ln p(y_t|Z_t)p(Z_t), \quad (11)$$

And the Q function of the EM algorithm will be

$$\begin{aligned}
&Q(\Theta, \Theta^*) \\
&= E_{\mathbf{Z}|\mathbf{y},\Theta^*}\left\{\sum_{t=1}^{T}\ln[p(y_t|Z_t)p(Z_t)]\right\} \\
&= \sum_{Z_1=1}^{K}\cdots\sum_{Z_T=1}^{K}\left\{\sum_{t=1}^{T}\ln[p(y_t|Z_t)p(Z_t)]\prod_{l=1}^{T}p(Z_l|y_l,\Theta^*)\right\} \\
&= \sum_{Z_t=1}^{K}\sum_{t=1}^{T}\ln[p(y_t|Z_t)p(Z_t)]p(Z_t|y_t,\Theta^*) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{K}\ln[p(y_t|Z_t)p(Z_t=j)]p(Z_t=j|y_t,\Theta^*) \\
&= \sum_{t=1}^{T}\sum_{j=1}^{K}p(Z_t=j|y_t,\Theta^*)\ln\alpha_j G(y_t;\hat{y}_{t,j},\sigma^2_{t,j}). \quad (12)
\end{aligned}$$

The probability $p(Z_t = j|y_t, \Theta^*)$ (denoted as $h_j(t)$) can be obtained as follows,

$$h_j(t) = p(Z_t = j|y_t, \Theta^*) = \frac{\alpha_j G(y_t; \hat{y}_{t,j}, \sigma^2_{t,j})}{\sum_{i=1}^{K}\alpha_i G(y_t; \hat{y}_{t,j}, \sigma^2_{t,i})}. \quad (13)$$

This is the E step of the EM algorithm.

The M step

In the M step we maximize (12). Since β, δ and α must bigger than zero, so we replace them by:

$$\delta_{0j} = e^{\gamma_{0j}}, \quad (14)$$

$$\delta_{qj} = e^{\gamma_{qj}}, \quad \text{where } q = 1, \ldots, Q \quad (15)$$

$$\beta_{pj} = e^{\rho_{pj}}, \quad \text{where } p = 1, \ldots, P \quad (16)$$

To ensure $\{\alpha_j\}_{j=1}^{K}$ sum to unity and each α_j is larger than zero, we use $\alpha_j = \frac{e^{m_j}}{\sum_{i=1}^{K} e^{m_i}}$.

The parameters that we will need to adjust are m_j and $\Omega = \{a_{1j}, \ldots, a_{Sj}, b_{1j}, \ldots, b_{Rj}, \rho_{1j}, \ldots, \rho_{Pj}, \gamma_{0j}, \ldots, \gamma_{Qj}\}_{j=1}^{K}$.

The first derivatives of (12) with respect m_j is

$$\frac{\partial Q(\Theta, \Theta^*)}{\partial m_j} = h_j(t) - \alpha_j. \quad (17)$$

For all other parameters, the first derivative is

$$\frac{\partial Q(\Theta, \Theta^*)}{\partial \omega} = h_j(t)\left[\frac{1}{2\sigma^2_{t,j}}\left(\frac{\epsilon^2_{t,j}}{\sigma^2_{t,j}} - 1\right)\frac{\partial\sigma^2_{t,j}}{\partial\omega} - \frac{\epsilon_{t,j}}{\sigma^2_{t,j}}\frac{\partial\epsilon_{t,j}}{\partial\omega}\right] \quad (18)$$

Where $\omega \in \Omega$. To calculate (18), we also need the following first derivatives:

$$\begin{aligned}
\frac{\partial \sigma^2_{t,j}}{\partial a_{sj}} &= 2\sum_{c=1}^{Q} \epsilon_{t-c,j} e^{\gamma_{cj}} \frac{\partial \epsilon_{t-c,j}}{\partial a_{sj}} + \sum_{d=1}^{P} e^{\rho_{dj}} \frac{\partial \sigma^2_{t-d,j}}{\partial a_{sj}}, \\
\frac{\partial \epsilon_{t,j}}{\partial a_{sj}} &= -\epsilon_{t-s,j} - \sum_{c=1}^{S} a_{cj} \frac{\partial \epsilon_{t-c,j}}{\partial a_{sj}}, \\
\frac{\partial \sigma^2_{t,j}}{\partial b_{rj}} &= 2\sum_{c=1}^{Q} \epsilon_{t-c,j} e^{\gamma_{cj}} \frac{\partial \epsilon_{t-c,j}}{\partial b_{rj}} + \sum_{d=1}^{P} e^{\rho_{dj}} \frac{\partial \sigma^2_{t-d,j}}{\partial b_{rj}}, \\
\frac{\partial \epsilon_{t,j}}{\partial b_{rj}} &= -y_{t-r,j} - \sum_{c=1}^{S} a_{cj} \frac{\partial \epsilon_{t-c,j}}{\partial b_{rj}}, \\
\frac{\partial \sigma^2_{t,j}}{\partial \gamma_{0,j}} &= e^{\gamma_{j0}} + \sum_{c=1}^{P} e^{\rho_{cj}} \frac{\partial \sigma^2_{t-c,j}}{\partial \gamma_{0j}}, \\
\frac{\partial \sigma^2_{t,j}}{\partial \gamma_{qj}} &= \epsilon^2_{t-q,j} e^{\gamma_{qj}} + \sum_{c=1}^{P} e^{\rho_{cj}} \frac{\partial \sigma^2_{t-c,j}}{\partial \gamma_{qj}}, \\
\frac{\partial \sigma^2_{t,j}}{\partial \rho_{pj}} &= \sigma^2_{t-p,j} e^{\rho_{pj}} + \sum_{c=1}^{P} e^{\rho_{cj}} \frac{\partial \sigma^2_{t-c,j}}{\partial \rho_{pj}}, \\
\frac{\partial \epsilon_{t,j}}{\partial \gamma_{0j}} &= \frac{\partial \epsilon_{t,j}}{\partial \gamma_{qj}} = \frac{\partial \epsilon_{t,j}}{\partial \rho_{pj}} = 0.
\end{aligned}$$

Where $r = 1, \ldots, R$, $s = 1, \ldots, S$, $q = 1, \ldots, Q$, $p = 1, \ldots, P$.

Since we use GEM algorithm, we only adjust each parameter more towards the maximum in each M step. Let θ be one of the parameters. To update θ we use:

$$\theta^{(n+1)} = \theta^{(n)} + \lambda_n \frac{\partial E(l_t)}{\partial \theta}, \qquad (19)$$

where $\theta \in \Theta$.

To ensure the estimated model will be stationary, the initial characteristic equations of every ARMA model

$$1 - b_{1,j} z - b_{2,j} z^2 - \cdots - b_{R,j} z^R = 0, \qquad (20)$$

and every GARCH model

$$1 - \beta_{1,j} z - \beta_{2,j} z^2 - \cdots - \beta_{P,j} z^P = 0, \qquad (21)$$

must have all their roots lie outside the unit circle (Greene, 2000). During our experiments, if the initial characteristic equations have all their roots lie outside the unit circle, the estimated results will also be stationary.

Experimental Simulations

We apply the mixture of ARMA(3,3)-GARCH(3,3) model to predict the daily stock prices of Cheung Kong Holding (CK HDG) and HSBC Holding (HSBC HDG) respectively. The period used to train the model is from Sep 15, 1995 to Jul 16, 1999, consisting of 1001 data points. Prediction is then effected on the following 120 days stock prices, which constitutes our test set. To enable comparisons, we also present results using conventional ARMA(3,3)-GARCH(3,3) model and the mixture of AR(3)-GARCH(3,3) model.

Results for CK HLD prices prediction using the conventional ARMA-GARCH model, mixture of AR-GARCH model and mixture of ARMA-GARCH model are shown in Figures 1, 2 and 3 respectively, with that of HSBC HDG shown in 4, 5 and 6 respectively.

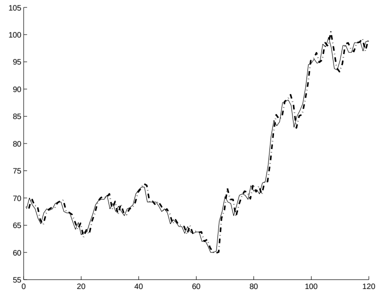

Figure 1: Prediction of CK prices with conventional ARMA-GARCH.

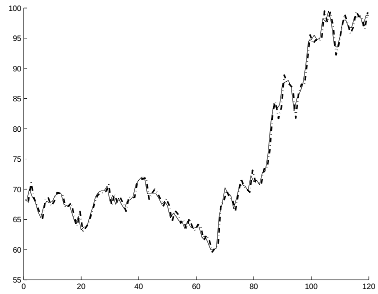

Figure 2: Prediction of CK prices with mixture of AR-GARCH.

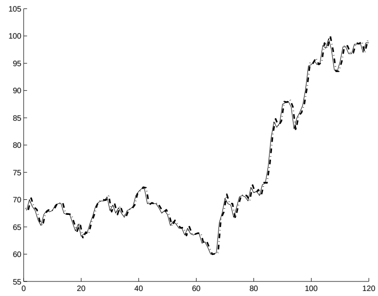

Figure 3: Prediction of CK prices with mixture of ARMA-GARCH.

Results Evaluation

By comparing the mean square error performance metrics shown in table 1, we can see that both the mixture of ARMA-GARCH and mixture of AR-GARCH models outperform the conventional ARMA-GARCH model by quite a significant margin. The reason is probably due to the capability of non-linear modelling by the mixture models. Another observation as revealed by the results

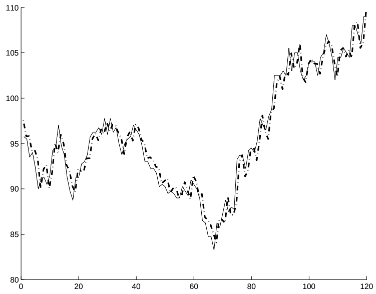

Figure 4: Prediction of HSBC prices with conventional ARMA-GARCH.

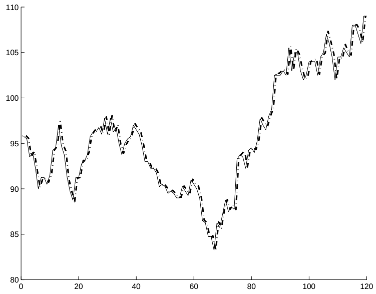

Figure 5: Prediction of HSBC prices with mixture of AR-GARCH.

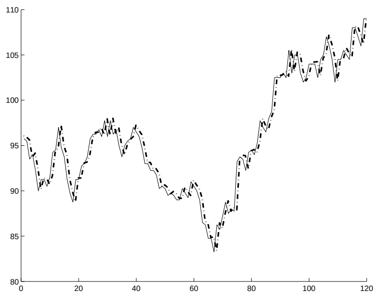

Figure 6: Prediction of HSBC prices with mixture of ARMA-GARCH.

Table 1: A summary of mean square errors for different approaches

	CK HDG	HSBC HDG
Conv. ARMA-GARCH	2.7571	2.3714
Mixture AR-GARCH	2.6037	2.2409
Mixture ARMA-GARCH	2.5609	2.1828

is that the mixture of ARMA-GARCH is better than the mixture of AR-GARCH model.

Conclusion

In this paper, we derive a GEM algorithm for the mixture of ARMA-GARCH model. Its relative empirical performance in stock price prediction against the conventional ARMA-GARCH and mixture of AR-GARCH model is investigated. Results reveal that both mixture models outperform the conventional ARMA-GARCH model, with the best results obtained by the mixture of ARMA-GARCH model.

Acknowledgements

The first and second authors would like to thank Mr. Zhi-Yong Liu for helpful discussions.

References

H. Y. Kwok, C. M. Chen, and L. Xu, "Comparison between mixture of ARMA and mixture of AR model with application to time series forecasting," in *Proceedings of Fifth International Conference on Neural Information Processing*, 1998, pp. 1049–1052.

C. S. Wong and W. K. Li, "On a mixture autoregressive model," *Journal of the Royal Statistical Society Series B – Statistical Methodology*, vol. 62, no. 1, pp. 95–115, 2000.

T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

W. C. Wong, F. Yip, and L. Xu, "Financial Prediction by Finite Mixture GARCH Model," in *Proceedings of Fifth International Conference on Neural Information Processing*, 1998, pp. 1351–1354.

A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of The Royal Statistical Society Series B – Statistical Methodology*, vol. 39, no. 1, pp. 1–38, 1977.

W. H. Greene, *Econometric Analysis*, Prentice Hall, New Jersey, fourth edition, 2000.