OPTIMUM CONDITIONAL DISTRIBUTION FOR ESTIMATING THE VOLATILITY OF STOCK INDICES

Akinlawon, O.J., Asiribo, O.E. and Adebanji, A.O.
Department of Statistics, College of Natural Sciences, University of Agriculture,
P.M.B. 2240, Abeokuta, 11001, Ogun State, Nigeria.
e-mail: dedot_001@yahoo.com

Abstract
This study aims at minimizing the excess kurtosis observed in the financial data especially in stock price by obtaining an optimum conditional distribution for estimating the linear GARCH (p, q) model. The study used the daily stock prices of First Bank Nigeria Plc. traded in the Nigeria Stock Exchange from January 2001 to July 2008 and the results from the statistical properties of the daily returns and that of the Jarque-Bera test showed that the returns series is leptokurtic. The selected GARCH model was compared for estimation based on the Normal, Student-t distribution and Generalized Error distribution (GED). The optimum distribution was selected using the Bayesian information criterion (BIC) and the Likelihood values and the results indicated that the kurtosis displayed was reduced when GED is used in the estimation. For stationarity, the parameters sum is less than one and this implies that the parameters satisfy the second order stationary (weakly stationary) conditions.

Keywords: GARCH Models, Model Selection, Optimum Conditional Distribution, GED.

Introduction

Financial time series often exhibit some well-known characteristics. First, large changes tend to be followed by large changes and small changes tend to be followed by small changes. Secondly financial time series often exhibit leptokurtosis (Kurtosis > 3), which means that the distribution of their returns is fat-tailed (i.e. relative high probability for extreme values).The GARCH model successfully captures the first property described above (Chen et al, 2006; Hansen and Lunde, 2005; Caiado, 2007; Hourvouliades 2007; Jafari et. al., 2008), but sometimes fails to capture the fat-tail property of financial data. This has led to the use of non-normal distributions to better model the fat-tailed characteristic. Not only the unconditional distribution but also the conditional distribution of daily asset returns is known to be leptokurtic. Thus, some authors have suggested using ARCH-type models with leptokurtic distributions such as the t-distribution, normal-lognormal mixture distribution, normal-Poisson distribution, Power exponential distribution for the conditional distribution and they conclude that student-t distribution is suited for capturing the excess kurtosis of conditional distribution for daily stock returns (Baillie and DeGennaro, 1990; Hsieh, 1989; Bollerslev and Mikkelsen, 1996).

The purpose of this paper is to examine what distribution is best fit for the conditional distribution of daily Nigeria stock returns by comparing two fat-tailed distributions viz: Student-t and Generalized error distributions in order to minimize the excess kurtosis of conditional distribution for Nigeria stock returns using Bayesian Information Criterion (BIC).

PRELIMINARIES

Financial time series analysis is directed to the understanding of the mechanism that drives a given time series of data, or in other words: financial time series analysis focuses on "the truth behind the data" so that one can find physical models that explain the empirically observed features of real life data. With such models one can make distributional forecasts for future values in time series.
Today there exist many different types of financial data but if one focuses on share prices, stock indices and foreign exchange rates (which we denote P_t, t=1,2,..., where t can be minutes, hours, days, etc.), they behave very similar after the transformation:

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \log P_t - \log P_{t-1} \quad \ldots\ldots(1)$$

The series Y_t is referred to as the log return series. These series have the advantage that they are free of unit, and can therefore be compared with each other.

2.1 Kurtosis

The observations of the time series Y_t have a distribution, which often is assumed to be normal (Gaussian). However, empirical studies of practically any financial time series show that this is not quite correct (Mandelbrot, 1963 and Fama (1963, 1965)) are the first studies to report this feature). One way to quantify this property is to look at the kurtosis of the distributions.
Kurtosis is a measure of the extent to which observed data fall near the centre of a distribution or in the tails i.e.is the degree of peakedness or flatning of a distribution usually taken relative to a normal distribution and is calculated as:

$$K = \frac{E\left(y_t^4\right)}{\sigma_t^4} \text{.........}(2)$$

- For a normal distribution, the kurtosis is 3 and is referred to as mesokurtic
- A platykurtic distribution has a kurtosis value less than that of a standard, normal distribution. This type of distribution has a fat mid-range on either side of the mean and a low peak.
- A leptokurtic distribution has a kurtosis value greater than that of a standard, normal distribution which gives the distribution a high peak, a thin midrange, and fat (heavy) tails.

2.2 Diagnostic Checking of Normality

There are several statistical tests used for the diagnostic checking of normality. In this study, Jarque Bera (JB) test was used for the diagnostic checking of residuals for normality and it is denoted by:

$$JB = \left(\frac{N}{6}\right)\left(S^2 + (K-3)^2/4\right)\text{........}(3)$$

Where S=Skewness, K=Kurtosis.

It follows a χ^2 distribution with 2 d.f. If the JB Statistic is greater than critical value of the χ^2 $(JB) > \chi_c^2$, we reject the null hypothesis of normality.

2.3 Diagnostic Checking of Serial Correlation

Testing for serial correlation is a fundamental problem in time series analysis. To determine whether a time series is independent, the autocorrelation function (ACF) of the series is examined. If the ACF is significantly different from zero, this implies that there is dependence between observations. Therefore, ACF is a powerful complementary tool for testing independence (Janacek and Swift, 1993; Fernando et al, 2000). The residual autocorrelation should be obtained to determine whether the residuals are white noise. In this study, the Ljung-Box Q statistic is used as alternative approaches for the diagnostic checking of residuals for serial correlation and is given as (Ljung and Box, 1978):

$$Q_k = n(n+2)\sum_{i=1}^{k}\frac{r_i}{(n-i)}\text{........}(4)$$

where n is the number of samples, k is the number of lags and r_i is the i^{th} autocorrelation. If Q_k is large then the probability that the process has uncorrelated data decreases. Q_k is asymptotically approximate as χ_k^2.

2.4 Model Order Selection

Since GARCH model can be treated as Autoregressive Moving Average (ARMA) model for squared residuals, traditional model selection criteria such as Akaike Information Criterion (AIC) and BIC can also be used for selecting models. When a model involving q independently adjusted parameters is fitted to data, the Bayesian information criterion (BIC) which is a modification of AIC is given as:

$$BIC(q) = -2\ln(\textit{Maximized likelihood}) + q\ln(n)\text{.......}(5)$$

where n is the number of observation.The model order with the minimum BIC would provide a good fit for analyzing the return series. This technique is also employed for the selection of the conditional distribution for the estimation of the parameters of the fitted model in order to minimize the kurtosis.

REVIEW OF RELEVANT MODELS

Most models for financial returns are of the form:

$$y_t = \sigma_t Z_t, \ldots\ldots\ldots \quad \ldots\ldots\text{(6)}$$

where Z_t is a sequence of i.i.d. symmetric random variables, and $\{\sigma_t\}$ is a nonnegative stochastic process such that Z_t and σ_t are independent for fixed t. There is strong empirical support for stochastic volatility in financial time series and the presence of stochastic volatility implies that returns are not necessarily independent over time. The standard assumption for the noise Z_t is that $\{Z_t\} \sim N(0,1)$ with $\{Z_t\}$ independent of the standard deviation process $\{\sigma_t\}$.

3.1 Autoregressive Conditional Heteroscedsatic (ARCH) Model

Volatility is a central part of most asset pricing models. In these models, one often assumes that the volatility is constant over time. However, it is well known that financial time series exhibit time-varying volatility. Engle (1982) proposed a model for σ_t.

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i y_{t-i}^2 \ldots\ldots\text{(7)}$$

This model is called the Autoregressive Conditional Heteroscedasticity (ARCH process) where the "autoregressive" property in principle means that old events leave waves behind a certain time after the actual time of the action. The process depends on its past. The terms "conditional heteroscedasticity" means that the variance (conditional on the available information) varies and depends on old values of the process.

3.2 General Autoregressive Conditional Heteroscedastic (GARCH) Model

Empirical evidence shows that high ARCH order has to be selected in order to catch the dynamic of the conditional variance. This leads to the Generalised ARCH model (GARCH) introduced by Bollerslev (1986).

Let Y_t denote a real-valued discrete-time stochastic process as defined in eqn. (6). The GARCH (p,q) process proposed by Bollerslev is then given by:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{p} a_i y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \ldots\ldots\text{(8)}$$

$p > 0,\ q \geq 0,\ a_0 > 0,\ a_i \geq 0,$ i=1,2,…,p, $b_j \geq 0,$ j=1,2,…,q.

We can write (5) in terms of lag-operators B, such that

$$\sigma_t^2 = a_0 + a(B)y_t^2 + b(B)\sigma_t^2 \ldots\ldots\text{(9)}$$

where

$$a(B) = a_1 B + a_2 B^2 + \ldots + a_p B^p \ldots\ldots\ldots\ldots \quad ..\text{(10)}$$
$$b(B) = b_1 B + b_2 B^2 + \ldots + b_q B^q$$

According to Bollerslev (1986), the necessary and sufficient condition for weak stationarity of the GARCH model is:

$$\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1 \ldots\ldots\text{(11)}$$

CONDITIONAL DISTRIBUTIONS

As discussed earlier, observations of the financial time series { Y_t } tend to be leptokurtic (fat tailed). It may be expected that excess kurtosis displayed by the residuals of conditional heteroscedasticity models will be reduced when a more appropriate distribution is used for the estimation of the parameters of the fitted model. We shall consider the normal and two non-guassian error distributions for the estimation in this study.
Here follows some further information about these distributions.

4.1 Normal Distribution

The normal (or Gaussian) distribution is a symmetric distribution with density function:

$$f_y(Y_t) = \left(2\pi\sigma_t^2\right)^{-\frac{1}{2}} \exp\left\{-\frac{(Y_t-\mu)^2}{2\sigma_t^2}\right\} \text{.....(12)}$$

where μ is the expectation value expected to be zero and σ_t^2 is the variance of the stochastic variable Y_t, thus $Y_t \sim$ N (0, σ_t^2). The so called standard normal distribution is given by taking μ =0 and σ_t^2 =1. For a normal distribution, the Skewness=0 and the Kurtosis=3.

4.2 Student-t Distribution

If a random variable Y_t has a student-t distribution with h degrees of freedom, the probability density function (PDF) of Y_t is given by:

$$f_y(y_t;h) = \frac{\Gamma[(h+1)/2]}{\sqrt{h\pi}\left(\Gamma\left(\frac{h}{2}\right)\right)\left(1+\frac{y_t^2}{h}\right)^{(h+1)/2}} \text{.....(13)}$$

where $\Gamma(\cdot)$ is the gamma function, with $y = \dfrac{z}{\sqrt{\dfrac{u}{h}}}$;

Also, z is a standard normal and u is a χ^2 distribution with h degree of freedom ($h > 2$). Like the normal distribution, the t distribution is symmetric for large sample size. The mean, variance and kurtosis of the distribution are:

1. $\mu = 0$ for $h \geq 2$
2. $\sigma^2 = \dfrac{h}{h-2}$ for $h \geq 3$
3. $K = \dfrac{6}{h-4}$ for $h \geq 5$.

The Student-t distribution with unit variance has the following density function:

$$f_y(Y_t;h) = \frac{\Gamma[(h+1)/2]}{\sqrt{h\pi}\left(\Gamma\left(\frac{h}{2}\right)\right)\left(1+\frac{y_t^2}{(h-2)}\right)^{(h+1)/2}} \text{.......(14)}$$

4.3 Generalized Error Distribution (GED)

Generalized error distribution (GED) proposed by Nelson (1991) is a symmetric distribution that can both be leptokurtic and platykurtic depending on the degree of freedom v ($v > 1$). The GED has the following density function:

$$f_y(Y_t,v)=\frac{v\exp\left[-(\tfrac{1}{2})/\tfrac{y_t}{\lambda}/^v\right]}{\lambda 2^{\frac{(v+1)}{v}}\Gamma(1/v)}......(15)$$

where

$$\lambda=\left[\frac{2^{-2/v}\Gamma(1/v)}{\Gamma(3/v)}\right]^{1/2}........(16)$$

The GED with a unit variance has the following density function:

$$f_y(Y_t,v)=\frac{v\exp\left[-(\tfrac{1}{2})/\tfrac{y_t}{\lambda}/^v\right]}{\lambda.2^{\frac{(v+1)}{v}}\Gamma(1/v)\sqrt{\frac{v}{(v-2)}}},\quad \text{for } v>2........(17)$$

and v is a positive parameter governing the thickness of the tail behaviour of the distribution. In equation (15) above, when v=2, the PDF reduces to the standard normal PDF; when v<2, the density has thicker tails than the normal density; when v>2, the density has thinner tails than the normal density.

4.4 Parameter Estimation in the GARCH Model

To be able to minimize the kurtosis displayed by financial time series, one first has to fit the GARCH-model to the time series in question. This is done via estimation of the parameters in the model. As stated by Peters (2001), the most common method of this estimation is the maximum-likelihood estimation (MLE).

4.4.1 Gaussian Quasi Maximum-Likelihood Estimation

Suppose that Y_t follows a normal distribution, then we substitute conditional variance for unconditional variance in the normal likelihood and then maximize it with respect to the parameters. The approach is as follows:

The joint probability density for n observations $(Y_1, Y_2, ..., Y_n)$

$$f(y_n, y_{n-1}, ..., y_1)=f(y_1)\prod_{t=2}^{n} f(y_t/y_{t-1}).........(18)$$

Where $f(y_t/y_{t-1})$ is conditional distribution (Gaussian) and $f(y_1)$ is unconditional non-Gaussian distribution, and no closed form description. So, we used conditional MLE (MLE_c). Then the Log – likelihood function is written as:

$$LogL\ (Y;\psi)=\sum_{t=2}^{n}\log\ f(y_t/y_{t-1})......(19)$$

For GARCH (p,q) process, we drop the terms $f(y_t/y_{t-1})$ for $t \le p$.

$$\therefore\ MLE_c(Y;\psi)=\sum_{t=p+1}^{n}\log\ f(y_t/y_{t-1})$$

This implies that:

$$\sum_{t=p+1}^{n}\log\frac{1}{\sqrt{2\pi\sigma_t^2}}\exp-\tfrac{1}{2}\left(\frac{y_t^2}{\sigma_t^2}\right)........(20)$$

Where σ_t^2 is given by (8).

Hence maximizing (20) gives the required parameters.

4.4.2 Fat-Tailed Maximum Likelihood Estimation

An alternative way of dealing with non-Gaussian errors is to assume a distribution that reflects the features of the data better than the normal distribution, and estimate the parameters using this distribution in the likelihood function instead of the Gaussian. Thus, the problem with the calculation of choosing a distribution for the innovations, BIC can be very helpful. In this paper two distributions, apart from the Gaussian, are considered; the Student-t Distribution (t Distribution), the Generalised Error Distribution (GED).

The likelihood functions for the distributional assumptions are:

- the log-likelihood function for the Student-t distribution:

$$L_n = \ln\left[\Gamma\left(\frac{h+1}{2}\right)\right] - \ln\left[\Gamma\left(\frac{h}{2}\right)\right] - 0.5\ln[\pi(h-2)] - 0.5\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (1+h)\ln\left(1+\frac{Y_t^2}{h-2}\right)\right] \text{...(21)}$$

- the log-likelihood function for the GED:

$$\ln = \sum_{t=1}^{n}\{\log(\frac{v}{\lambda}) - 0.5|\frac{Y_t}{\sigma_t\lambda}|^v - (1+v^{-1})\log(2) - \log[\Gamma(\frac{1}{v})] - 0.5\log(\sigma_t^2)\} \text{....(22)}$$

Where $\Gamma(.)$ is the gamma function, and $\lambda = \left[\frac{2^{-2/v}\Gamma(1/v)}{\Gamma(3/v)}\right]^{1/2}$

These log-likelihood functions are maximised with respect to the unknown parameters (the same procedure as in the Gaussian quasi MLE case).

Results and Discussion

From Table 1, we see that the returns series is leptokurtic in the sense that the kurtosis exceeds positive three with positive non-zero skewness. The positive skewness shows that the upper tail of the distribution is thicker than the lower tail implying that the market increases occur more than the market declines. So the results showed that the returns have significant skewness, excess kurtosis and hence the assumption of normal distribution is not satisfied.

Further on conducting the normality test using the Jarque-Bera test for normality; it indicated non-normality of the distribution i.e. under the null hypothesis of normal distribution the p-value was essentially zero which is less than the 5 percent level of significance. The tests results are presented in table 2.

Different p and q values for the standard GARCH model were tested using the AIC and BIC techniques. So from Table 3, if the various BIC values are compared, we established that GARCH (1,2) is the one with the minimum BIC which is also one with the minimum AIC and highest likelihood value. Therefore, GARCH (1,2) is chosen as the optimum order of the linear GARCH model.

To evaluate which conditional distribution better describes the observed characteristics of daily stock price returns of the First Bank Nigeria Plc., three different distributions were compared for the identified GARCH (1,2) model using BIC and their likelihood values as shown in Table 4. According to the table, generalized exponential distribution (GED) has the smallest AIC and BIC values with highest likelihood value. This implies that the excess kurtosis and skewness displayed by the residuals of GARCH (1,2) is reduced with the use of GED. Therefore, the parameters of the GARCH (1,2) model are going to be estimated using the identified distribution.

However, Table 5 shows the values of the parameters for the identified model. Consequently, the following GARCH (1,2) process is fitted to the data in order to model the volatility:

$$\sigma_t^2 = 0.0002553 + 0.8291\, y_{t-1}^2 + 0.0009142\, \sigma_{t-1}^2 + 0.1463\, \sigma_{t-2}^2$$

When we check for persistency of our chosen model, we have that the sum

$$0.8291 + 0.0009142 + 0.1463 = 0.9763142 < 1$$

Table 1: The Descriptive Statistics

Mean	*Min.*	*Max.*	*Median*	*St.Dev.*	*Skewness*	*Kurtosis*
-0.0002735642	-4.60517	4.60517	0.0000	0.1817525	0.229075	512.5148

Table 2: The Test Results for Testing Normal Hypothesis

Test	*Test Statistic*	*Null Hypothesis*	*P-value*
Jarque-Bera	20587201	Normally distributed	0.0000

Table 3: The Various Fitted GARCH (p,q) Models

Order	*AIC*	*BIC*	*Likelihood*
0,1	-995.169	-978.557	500.6
0,2	-663.8	-641.6	335.9
1,0	-1319.41	-1302.798	662.7
1,1	-2189.236	-2167.086	1099
1,2	-2216.163	-2188.476	1113
2,0	8877	8900	-4435
2,1	-2202.176	-2174.489	1106
2,2	-2085.656	-2052.432	1049

Table 4: Conditional Distribution for The Identified GARCH (1,2) Model

Distribution	*AIC*	*BIC*	*Likelihood*
Normal	-2216	-2188	1113
Student-t	-9247	-9214	4629
Generalized Exponential	-9287	-9253	4649

Table 5: Estimation of GARCH (1,2) with Conditional GED Distribution with Estimated Parameter, V=0.7434626, and Standard Error 0.006131138.

Parameters	*Value*
Constant	2.553e-004
ARCH (1)	0.8291
GARCH (1)	0.0009142
GARCH (2)	0.1463

Table 6: Test Results for The Parameters of The Chosen Model

Parameter	*Value*	*Std. Error*	*t-value*	*Null Hypothesis*	*P-value*
Constant	2.553e-004	0.00001744	14.6404817	Zero	0.0000
ARCH(1)	0.8291	0.08250219	10.0493057	Zero	0.0000
GARCH(1)	0.0009142	0.00183175	0.4990830	Zero	0.3089
GARCH(2)	0.1463	0.02085932	7.0152251	Zero	1.598e-012

Table 7: Test Results for Testing Various Hypotheses for The Chosen Model

Test	*Test Statistic*	*Null Hypothesis*	*P-value*
Jarque-Bera	139840117	Normally Distributed	0.0000
Ljung-Box(Std. Residuals)	0.6138	No Autocorrelation	1.0000
Ljung-Box(Squared Std. Resid.)	0.03133	No Autocorrelation	1.0000

This shows that the parameters satisfy the second order stationary (weakly stationary) conditions with high degree of persistence in the conditional variance. Various test statistics carried out to assess the performance of the GARCH (1,2) model are shown in Tables 6 and 7. From Table 6, all the parameters including the constant value except GARCH (1) parameter, are significantly different from zero at 5 percent level of significance. Also, from Table 7, the test for serial correlation structure showed that no autocorrelation left in the standardized residuals and the squared standardized residuals since we fail to reject both the Null hypotheses in the Ljung-Box test.

Conclusion

This study is concerned with obtaining the optimum model order of the linear GARCH model in modeling the First Bank returns series. The results from the descriptive statistics of the daily return series supported the claim that most of the financial series are leptokurtic. By comparing the orders of the models using BIC technique as well as their likelihood values, GARCH (1,2) model was identified to be the most appropriate for the time-varying volatility of the data.

In order to account for the fat tails (minimize the excess kurtosis), the chosen model was compared for estimation based on the Normal, Student-t and Generalized Exponential distributions and the parameters of GARCH (1,2) model were obtained using GED. Serial correlation structure of the residuals and squared residuals of the identified model was examined using the Ljung-Box statistic. The results from all these tests showed that the selected model fulfilled all the diagnostic checks assumptions.

References

Baillie, R.T. and DeGennaro, R.P. (1990). Stock returns and Volatility. Journal of Financial and Quantitative Analysis, 25: 203-214.

Bollerslev, T. and Mikkelsen, H.O.A. (1996). Modelling and Pricing Long-Memory in Stock Market Volatility. Journal of Econometrics. 73(1): 151-184.

Caiado, J. (2007). Modeling and forecasting the volatility of the Portuguese stock index PSI-20. MPRA paper No. 2077.

Chen, Z. Daigler, R. and Parhizgari, A. (2006). Persistence of volatility in future markets. Journal of futures markets. 26(6): 571-594.

Bollerslev, T.P. (1986). Generalized Autoregressive conditional heteroscedasticity. Journal of Econometrica. 31: 307-327.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 50(4): 987-1008.

Fama, E.F. (1963). Mandelbrot and the Stable Paretian Distribution. Journal of Business. 36: 420-429.

Fernando, T.S., Genest, C. and Hallin, M., (2000). Kendal's Tau for serial Dependence. The Canadian Journal of Statistics, 28: 587-604.

Hansen, P. and Lunde, A. (2005). A forecast comparison of Volatility models: Does anything beat a

GARCH (1,1)? Journal of Applied Econmetrics. 20: 873-889.

Hourvouliades, N.L. (2007). Volatility clustering in the Greek future market: Curse or Blessing. International Journal of Finance and Economics. 11: 1450-2887.

Hsieh, D.A. (1989). Modelling Heteroscedasticity in Daily Foreign Exchange Rates. Journal of Business and Economics Statistics, 7: 307-317.

Jafari, G.R., Bahraminasab, A. and Norouzzadeh, P. (2008). Why does the standard GARCH (1,1) model work well? Journal of Statistical Physics. ArXiv: Physics/050327V2.

Janacek, G. and Swift, C. (1993). Time Series Forecasting, Simulation and Application. Ellis Horwood, New York, USA, 1993.

Ljung, G.M. and Box, G.E.P. (1978). On a Measure of Lack of Fit in Time Series Models. Biometrica, 65: 297-303.

Mandelbrot, B. (1963). The variation of certain speculative prices. Journal of Bussiness. 36: 394-416.

Nelson, D.B. (1990). Stationary and Persistence in the GARCH (1,1) model, Economic Theory. 6: 318-334.

Nelson, D.B. (1991). Conditional heteroscedasticity in asset returns. A new approach, Econometrica. 59: 347-370.

Peters, J.P. (2001). Estimating and Forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities. Unpublished paper. Universite de Liege, Belgium.