Modeling S&P 500 STOCK INDEX using ARMA-ASYMMETRIC POWER ARCH models

Jia Zhou

Supervisor: Changli He

HÖGSKOLAN
Dalarna

Master thesis in statistics

School of Economics and Social Science

HÖGSKOLAN Dalarna, Sweden

June, 2009

Contents

Abstract

In this paper, the S&P 500 stock index is studied for its time varying volatility and stylized facts. The ARMA mean equation with asymmetric power ARCH errors is used to model the series correlations and the conditional heteroscadesticity in the asset returns. The conditional distributions of the standardized residuals are assumed to be the normal distribution, the t distribution or the skew-t distribution. Furthermore, to capture the asymmetry and fat tail of the returns, an ARMA (0, 2)-APARCH (1, 1) model with the skew-t distribution is found to fit the data better than other models discussed in this paper. Finally, we use ARMA (0, 2)-APARCH (1, 1) model with the skew-t distribution to do the 10-step-ahead forecasting compared with the ARMA (0, 2)-GARCH (1, 1) model with the normal distribution and get the empirical conclusion that the ARMA (0, 2)-APARCH (1, 1) with the skew-t distribution also gives a better result in the forecasting.

Keywords: fat tail; asymmetry; GARCH; APARCH; non-Gaussian distribution

1. Introduction

In finance, researchers always put a lot of interests in modeling and forecasting volatility of asset returns. The reason is that the volatility of asset returns can be seen as a measurement of the risk for investment and provides essential information for the investors to make the correct decisions. While the asset returns themselves are uncorrelated or nearly uncorrelated, however, there exists high order dependence within the return series.

To model the time varying volatility (conditional heteroscedasticity), Autoregressive conditional heteroscedasticity (ARCH) introduced by Engle (1982) is used to model the series correlations in squared returns by allowing the conditional variance as a function of past errors and changing over time. Then generalized autoregressive conditional heteroscedasticity (GARCH) model introduced by Bollerslev (1986) extended the ARCH model to have longer memory and more flexible lag structure by adding lagged conditional variance to the model as well. Since then, GARCH model has been studied widely and proved a lot in the literature to be a competent model in fitting the financial time series, sometimes specify the mean equation with a low order of ARMA (p, q) process to capture the autocorrelation of the financial time series.

The empirical probability distributions for financial asset returns always exhibit some characteristics which are called stylized facts (Tavares, et al., 2008): Firstly, the well-known volatility clustering or persistence: large changes tend to followed by large changes, and small changes tend to followed by small changes are observed in the asset returns. Secondly, fat tail exits in the probability distribution of the assets returns that the kurtosis exceeds the value of the normal distribution which is 3. This fat tail phenomenon is called excess kurtosis, and the returns time series which exhibit fat tail are often called leptokurtic. From (Bai, et al., 2003), we know the excess kurtosis can be resulted from two aspects: one is the volatility clustering, which is a type of heteroscedasticity, accounts for some of the excess kurtosis; the other is the presence of non-Gaussian asset returns distribution can also results in the fat tail.

Thirdly, asymmetric distribution of the assets returns, which means volatility increase more when the change is negative than the change is positive, which is also called leverage effect (Black, 1976). That's why we always see the bad news give a greater impact on the volatility of the stock market than the good news.

In the initial assumption in ARCH model and GARCH model, the conditional distribution of the innovations is Gaussian. But in most case, the unconditional distribution of the high frequency financial time series seems to have fat tail than the Gaussian distribution. Assuming the fourth order moment exists, Bollerslev (1986) showed that the kurtosis implied by a GARCH (1, 1) process with normal errors is greater than 3, so the unconditional distribution of error which follows GARCH (1, 1) process is leptokurtic (fat tail). However, it's not adequate to capture the fairly high kurtosis present in the financial time series, sometimes GARCH model with a non-Gaussian errors are required to capture the observed fat-tailed behavior in asset returns. Then GARCH model with t distribution as conditional distribution was first introduced by Bollerslev (1987).

Further more, due to the skewness observed in the asset returns, the GARCH model with normal conditional distribution cannot capture the asymmetric characteristic of the returns. To overcome this drawback in the model, there are two ways: one is allowing for asymmetric conditional distribution, i.e. skewed t distribution which takes leptokurtic into account as well, while another is modeling the asymmetric directly in the conditional variance equation as nonlinear GARCH model: The APARCH (p, q) model- Asymmetric power ARCH model introduced by Ding, et al. (1993). It changes the second order of the error term into a more flexible varying exponent with an asymmetric coefficient which takes the leverage effect into account. The APARCH (p, q) model is a general class of model which includes special cases as ARCH, GARCH, TS-GARCH (Taylor, 1986 cited in Ding et al., 1993), GJR-GARCH (Glosten et al.,1993 cited in Ding et al., 1993) and TARCH (Zakoian, 1991 cited by Ding et al., 1993) by given the different definition of the model parameters.

In this paper, ARMA process with asymmetric power ARCH errors (including

GARCH, TS-GARCH and GJR-GARCH) is used to model the S&P 500 stock index returns with the t distribution and the skew-t distribution in order to see how well it captures the asymmetric and fat tail of the asset returns compared with the normal distribution. To measure the goodness of fit, we use maximum log-likelihood value, Akaike Information Criterion (AIC), Bayesian information criterion (BIC) to choose the better fitted model. Finally, we choose the ARMA (0, 2)-APARCH (1, 1) with skew-t distribution to be the best fitted model and do the 10-step-ahead forecasting.

This paper is organized as follows: section 2 will introduce the basic model which contains the ARMA (p, q) as the mean equation with innovations following GARCH and APARCH process. Normal distribution, student t distribution and skew-t distribution as the conditional distribution of the standardized residuals will be also specified. Section 3 will discuss about the statistical properties of the S&P 500 stock index returns, test for unit root and ARCH effect and give some empirical analysis. Section 4 is going to give the estimation result of the ARMA-GARCH/APARCH model with t and skew-t distribution compared with Gaussian distribution. Section 5 is going to do the forecasting based on the estimation results and the selected model and compare it with the ARMA (0, 2)-GARCH (1, 1) model with normal distribution. The final section will draw conclusions and give further discussion about this paper.

2. Methodology

2.1Conditional Mean Equation

We can describe the mean equation of a financial time series y_t as

$$y_t = E(y_t|\psi_{t-1}) + \varepsilon_t \tag{1}$$

Where $E(y_t|\psi_{t-1})$ is the conditional mean of y_t given ψ_{t-1}, ψ_{t-1} is the information set at time $t-1$.

Sometimes in order to model the serial dependence and get the conditional mean equation, ARMA (p, q) model is used to fit the data to remove this linear dependence and get the residual ε_t which is uncorrelated (but not independent).

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \varphi_j \varepsilon_{t-j} + \varepsilon_t \tag{2}$$

This ARMA (p, q) process is stationary when all the roots of $\phi(z) = 1 - \phi_1 z - \phi_2 z - .... - \phi_p z = 0$ lie outside of the unit circle.

To specify the order of the ARMA process, use Akaike information criterion (AIC) and the Bayesian Schwarz criterion (BIC) to choose the ARMA term which minimize the corresponding value of the criterions.

2.2 Variance Equation

2.2.1 GARCH

The error term ε_t in the ARMA mean equation is defined by Engle(1982) as autoregressive conditional heteroscedastic process which can be decomposed as follows:

$$\varepsilon_t = \sigma_t z_t \qquad z_t \sim iid(0,1) \tag{3}$$

where $\sigma^2{}_t = E(\varepsilon_t^2 | \psi_{t-1})$ is the conditional variance of the error and $E(\varepsilon_t | \psi_{t-1}) = 0$.

The ε_t is uncorrelated but its conditional variance σ_t is changing over time as the function of the past errors defined in Engle(1982), and then generalized by Bollerslev (1986) who extended the ARCH model to have longer memory and more flexible lag structure by adding lagged conditional variance to the model:

$$\varepsilon_t | \psi_{t-1} \sim N(0, \sigma^2{}_t) \tag{4}$$

$$\sigma^2{}_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma^2{}_{t-i} = \alpha_0 + A(L)\varepsilon_t^2 + B(L)\sigma^2{}_t \tag{5}$$

Where $p \geq 0, q \geq 0,\ \alpha_0 > 0, \alpha_i \geq 0, i = 1,....,q,\ \beta_i \geq 0, i = 1,...., p$

When $p = 0$, the GARCH model is reduced to ARCH model. Bollerslev (1986) has proved that the GARCH (p, q) process as defined in (4) and (5) is wild-sense stationary with $E(\varepsilon_t) = 0$, $\operatorname{var}(\varepsilon_t) = \alpha_0 (1 - A(1) - B(1))^{-1}$ and $\operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0$ for $t \neq s$ if and only if $A(1) + B(1) < 1$.

The GARCH (1, 1) model is parsimonious but the most effective to model the conditional volatility. It has been studied a lot in the literature and yields abundant results modeling the conditional heteroscedasticity of the financial time series successfully.

$$\varepsilon_t | \psi_{t-1} \sim N(0, \sigma_t^{\ 2})$$

$$\sigma^2{}_t = \varpi + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{6}$$

Where $\varpi > 0, \alpha \geq 0, \beta \geq 0$, ψ_{t-1} is the information set available at time $t-1$, ε_t is wide-stationary if and only if $\alpha + \beta > 1$.

2.2.2 APARCH

The APARCH (p, q)- Asymmetric power ARCH model was introduced by Ding, et al., (1993). It changes the second order of the error term into a more flexible varying exponent with an asymmetric coefficient takes the leverage effect into account.

The variance equation of APARCH (p, q) can be written as

$$\varepsilon_t = z_t \sigma_t,\ \ z_t \sim N(0,1)$$

$$\sigma_t^{\ \delta} = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^{\delta},$$

$$\omega > 0, \delta > 0,\ \alpha_i \geq 0, -1 < \gamma_i < 1,\ i = 1,...., p,\ \beta_j \geq 0,\ j = 1,...., q$$

The APARCH(p,q) process is stationary if

$\omega > 0, \sum_i \alpha_i k_i + \sum_j \beta_j < 1, k_i = E(|z| - \gamma_i z)^\delta$

The APARCH (p, q) model is a general class of model which includes special cases as ARCH by Engle (1982), GARCH by Bollerslev(1986), TS-GARCH by Taylor and Schwert (1986 cited in Ding et al., 1993), GJR-GARCH by Glosten et al. (1993 cited in Ding et al., 1993), TARCH by Zakoian (1991 cited in Ding et al., 1993) and other two models by giving the different definition of the model parameter. In this paper, I only talk about TS-GARCH model, GJR-GARCH model and the complete APARCH model.

a) TS (Taylor-Schwert) GARCH

when $\delta = 1, \gamma_i = 0$

$$\sigma_t = \omega + \sum_{i=1}^{p} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^{q} \beta_j \sigma_{t-j} \tag{7}$$

$\omega > 0 \;\; \alpha_i \geq 0 \;\; i = 1,....,p \;\; \beta_j \geq 0, \; j = 1,....,q$

b) GJR-GARCH

when $\delta = 2$

$$\sigma_t^{\,2} = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{8}$$

$\omega > 0, \;\; \alpha_i \geq 0, -1 < \gamma_i < 1, \;\; i = 1,....,p, \;\; \beta_j \geq 0, \;\; j = 1,....,q$

c) DGE (Ding, Granger and Engle) GARCH

$$\sigma_t^{\,\delta} = \omega + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^\delta \tag{9}$$

$\omega > 0, \delta > 0, \;\; \alpha_i \geq 0, -1 < \gamma_i < 1, \;\; i = 1,....,p, \;\; \beta_j \geq 0, \;\; j = 1,....,q$

2.3 Conditional Distributions

2.3.1 Normal Distribution

The standard GARCH (p, q) model introduced by Tim Bollerslev(1986) is with normal distributed error $\varepsilon_t = \sigma_t z_t$, $z_t \sim iid(0,1)$. Use maximum log-likelihood method to estimate the parameter in the standard GARCH model, given the error following the Gaussian and we can get the log-likelihood function:

$$L(\varepsilon_t|\theta) = \ln\prod_t \frac{1}{\sqrt{2\pi\sigma^2{}_t}} e^{-\frac{\varepsilon_t^2}{2\sigma^2{}_t}} = \ln\prod_t \frac{1}{\sqrt{2\pi\sigma^2{}_t}} e^{-\frac{z_t^2}{2}} = -\frac{1}{2}\sum_t [\log(2\pi) + \log(\sigma^2{}_t) + z_t^2]$$

where θ is vector of the estimates parameters.

2.3.2 Student t Distribution

Known fat tail in financial time series, it may be more appropriate to use a distribution which has fatter tail than the normal distribution. Bollerslev(1987) suggested fitting GARCH model using student t distribution for the standardized error to better capture the observed fat tails in the return series.

If z_t here has student t distribution with v degree of freedom, and its density function is given by

$$f(z_t|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}(1+\frac{z_t^2}{v-2})^{-\frac{v+1}{2}}$$

$$f(\varepsilon_t|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})}(1+\frac{\frac{\varepsilon_t}{\sigma_t}}{v-2})^{-\frac{v+1}{2}} \cdot (-\frac{1}{\sigma_t^2})$$

Where $v > 2$ is shape parameter, then the log-likelihood function of GARCH model with t distribution error can be built based on this probability density function which is: (Lambert and Laurent, 2002)

$$L(\varepsilon_t|\theta) = T\{\ln\Gamma(\frac{v+1}{2}) - \ln\Gamma(\frac{v}{2}) - \frac{1}{2}\ln[\pi(v-2)]\} - \frac{1}{2}\sum_{t=1}^{T}[\ln(\sigma_t^2) + (1+v)\ln(1+\frac{z_t^2}{v-2})]$$

2.3.3 Skew-t Distribution

Fernandez and Steel (1998 cited in Alberg et al., 2008) extended the student t distribution by adding a skew parameter then Lambert and Laurent (2001) applied it to the GARCH and the log-likelihood function for the skew t distribution is given as:

$$L(\varepsilon_t|\theta) = T(\ln\Gamma(\frac{v+1}{2}) - \ln(\frac{v}{2}) - \frac{1}{2}\ln(\pi(v-2)) + \ln(\frac{2}{\xi + (1/\xi)}) + \ln(s))$$

$$-\frac{1}{2}\sum_{t=1}^{T}(\ln(\sigma_t^2) + (1+v)\ln(1+\frac{(sz_t+m)^2}{v-2}\xi^{-2I_t}))$$

ξ is the skew parameter, v is degree of the freedom which is also called the shape parameter of the model.

$$I_t = \begin{cases} 1 & if \quad z_t \geq -\frac{m}{s} \\ -1 & if \quad z_t < -\frac{m}{s} \end{cases}$$

$$m = \frac{\Gamma(\frac{v+1}{2})\sqrt{v-2}}{\sqrt{\pi}\Gamma(v/2)}(\xi - 1/\xi)$$

$$s = \sqrt{(\xi^2 + \frac{1}{\xi^2} - 1) - m^2}$$

3. Data

3.1 Original Data and Pretreatment

The data used in this paper is standard & poor (S&P)'s 500 closing price over the period Jan 3, 1950 through Apr 14, 2009.Let p_t denotes the successive closing price observation at time t, corresponding transform the price series $\{p_t\}$ into a daily return series $\{r_t\}$ using $r_t = \log \frac{p_{t+1}}{p_t} = \log(p_{t+1}) - \log(p_t)$

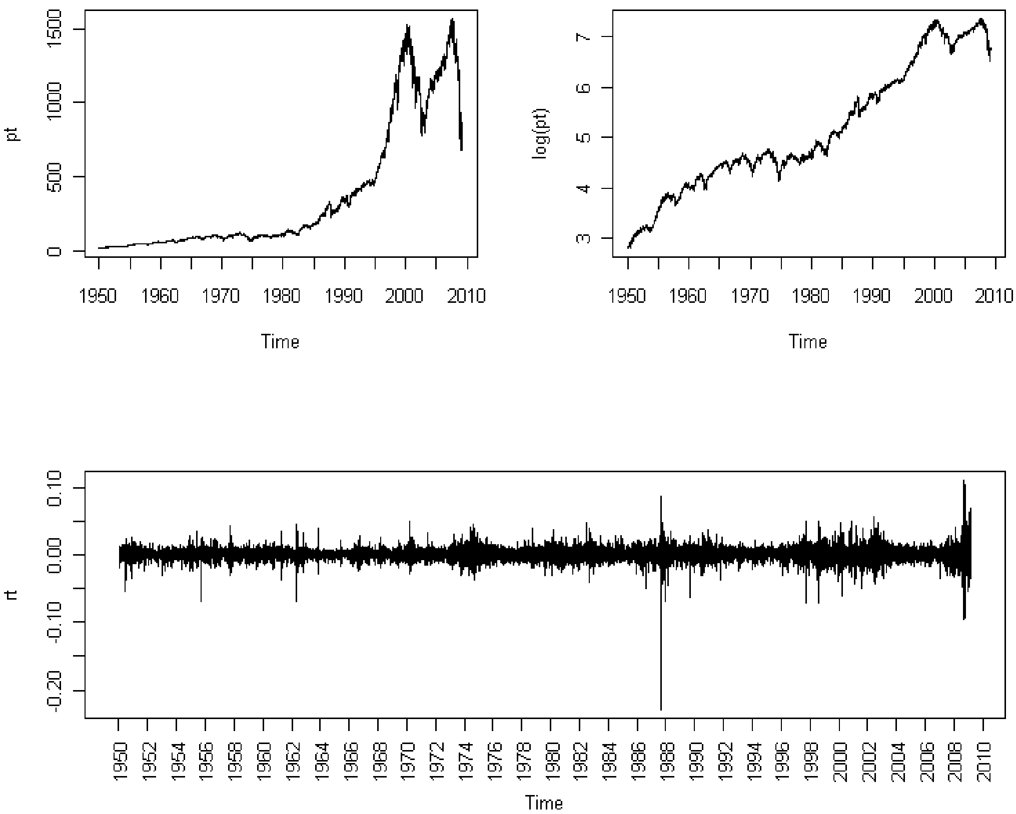

Figure 1: S&P 500 Daily Closing Price p_t, $\log(p_t)$ and Daily Returns r_t

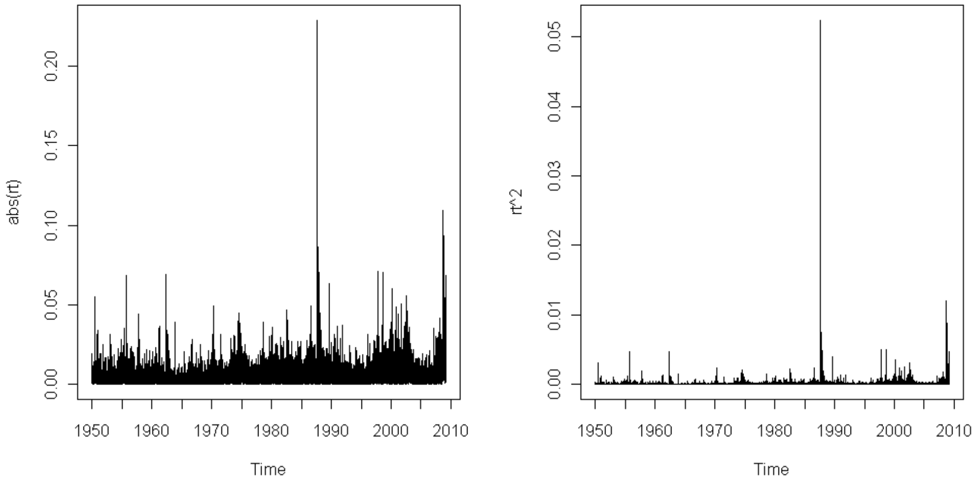

Figure 2: Absolute Returns and Squared Returns for the S&P 500 Stock Index

3.2 Stylized Fats of Asset Returns

Table 1: Summary Statistics for Daily Stock Returns.

Statistics of rt	values
Num of rt	14917
Minimum	-0.229
Maximum	0.109572
Mean	0.000265
Median	0.000444
Variance	0.000093
Stdev	0.009653
Skewness	-1.100235
Excess Kurtosis	30.702161
JB test	589061.9 (<2.2e-16)

Notes: Sample Period is 01/03/50 - 04/17/09 Giving 19417 Daily Observations.

In figure 2, obvious volatility clustering in the returns is indicated that low volatility is followed by low volatility and high volatility is followed by high volatility. It is also confirmed by the autocorrelation of squared returns in figure 3. As can be seen from table 1, the distribution of daily returns is clearly non-normal with negative skewness which means there is a long tail in the negative direction, and excess kurtosis is significantly high.

3.3 Tests for Normality and Unit Root

To test for the normality, use Jarque-Bera test calculated by $JB = \frac{n}{6}(\hat{S}^2 + \frac{(\hat{K}-3)^2}{4})$, in which n is the number of the observations, $\hat{S}$ denotes the sample skewness, $\hat{K}$ denotes the sample kurtosis and it has an asymptotic chi-square distribution with two degrees of freedom. The null hypothesis of Jarque-Bera test is a joint hypothesis of the skewness and the excess kurtosis being zero, since samples from a normal distribution have an expected skewness and excess kurtosis of zero. From the result shown in table2, we can get the conclusion that the Gaussian distribution hypothesis for the empirical returns distribution is clearly rejected.

Table 2: ADF Test and PP Test for the Daily Stock Return with P-Value

	ADF Test	P-value	PP Test	P-value
$\log(p_t)$	-1.7053	0.7034	-6.8999	0.7253
r_t	-24.4742	<0.01	-13589.04	<0.01

To test for stationary, we use ADF (augmented Dickey–Fuller) test and PP (Phillips–Perron) test and get the result from which we know that the S&P 500 index (in logarithm) is non-stationary while the returns are stationary and it won't cause the nonlinearity in the returns.

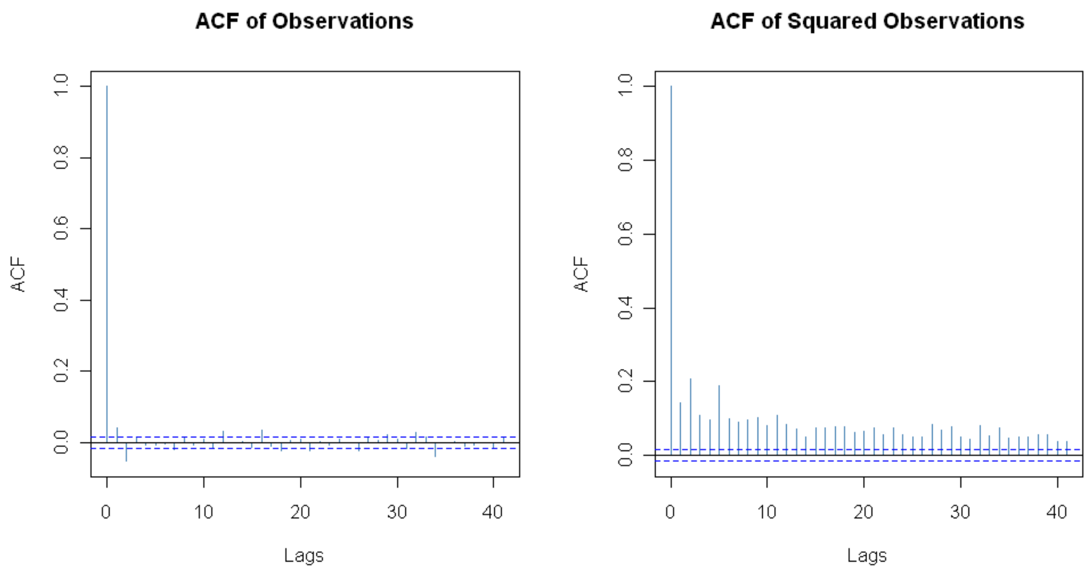

Figure 3: Autocorrelation Function of Asset Returns and Squared Asset Returns

3.4 Tests for Serial Dependence in Asset Returns and ARCH-effect of ε_t in the Fitted ARMA (0, 2) Model

Ljung-box (Ljung & Box, 1978) statistic $Q = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}^2{}_j}{n-j}$, where *n* is the sample size, $\hat{\rho}_j$ is the sample autocorrelation at lag *j*, and *h* is the number of lags being tested. For significance level α, the critical region for rejection of the hypothesis of randomness is rejected if $Q > \chi^2_{1-\alpha,h}$, where $\chi^2_{1-\alpha,h}$ is the α-quantile of the chi-square distribution with *h* degrees of freedom.

To test for serial dependence and ARCH-effect, use Ljung-box test from 1 up to twentieth order autocorrelation of returns r_t, squared returns r_t^2, residuals of the fitted ARMA (0, 2) model ε_t and squared residuals of the fitted ARMA (0, 2) model ε_t^2.

Figure 3 shows a small relevant first and second order autocorrelation for the asset returns. And these correlations can be confirmed from the ljung-box test in table 3, the series of returns are correlated over time, so I use ARMA (p,q) model to fit the asset returns which will be mentioned in the next part of this paper. The Ljung–Box statistic for up to twentieth-order serial correlation of squared returns is highly significant suggesting the presence of strong nonlinear dependence in the data. The volatility clustering shown in the figure 1 and 2 suggests the presence of time varying conditional volatility (conditional heteroskedasticity). And ε_t is uncorrelated according to the testing result of Ljung-Box, which conforms to the assumption of the GARCH-type model.

Table 3: Ljung-Box Test with P-Value for Asset Returns, Squared Asset Returns and Residuals of the Fitted ARMA (0, 2) Model and Squared Residuals of the Fitted ARMA (0, 2) Model

	Q(1)	Q(5)	Q(10)	Q(20)
r_t	23.7631	68.9155	79.1619	132.0471
	(1.089e-06)	(1.723e-13)	(7.327e-13)	(< 2.2e-16)
r_t^2	302.9294	1787.788	2453.268	3308.723
	(< 2.2e-16)	(< 2.2e-16)	(< 2.2e-16)	(< 2.2e-16)
ε_t	0.0055	3.3376	16.7926	73.5572
	(0.9411)	(0.6481)	(0.07908)	4.733e-08
ε_t^2	373.3585	1738.825	2435.722	3310.868
	(< 2.2e-16)	(< 2.2e-16)	(< 2.2e-16)	(< 2.2e-16)

4. Results

4.1 Order Selection of ARMA (p, q)

Use ARMA (p, q) model to fit the data and get the innovations series $\{\varepsilon_t\}$. From (Hannan, 1980) we know $\hat{p}$, $\hat{q}$ from AIC (p, q) are not weakly consistent, and if series { ε_t } are not independent but $E\{|\varepsilon_t|^r\} < \infty$, $\gamma < 4$, then the estimates via BIC (p, q) are strongly consistent. BIC always gives penalty for the additional parameters more than AIC does. So I choose ARMA (0, 2) as the mean equation mainly take account of the BIC.

Therefore, the conditional mean equation with error term following conditional heteroskedastic process is $y_t = u + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$,

Then, together with the GARCH (1, 1) and APARCH class of models, use R program to do the estimation and get the estimation result.

Table 4: Criterion for ARMA (p, q) Order Selection

	AIC	BIC
ARMA(0,0)	-96108.65	-96093.43
ARMA(1,0)	-96130.43	-96107.6
ARMA(2,0)	-96173.09	-96142.64
ARMA(3,0)	-96174.68	-96136.63
ARMA(0,1)	-96133.33	-96110.5
ARMA(0,2)*	-96175.9	**-96145.45***
ARMA(0,3)	-96175.7	-96137.65
ARMA(1,1)	-96164.66	-96134.22
ARMA(2,1)	-96175.78	-96137.73
ARMA(1,2)	-97176*	-96137.95
ARMA(2,2)	-96173.93	-96128.27

4.2 Estimation Results

4.2.1 ARMA (0, 2)-GARCH (1, 1)

In this paper I choose fit the data with a GARCH model with normal distribution and non-normal distribution for the standardized residual, so the complete model is

ARMA(0,2)-GARCH(1,1) with conditional distribution is t distribution and skewed t distribution compared with normal distribution.

In table 5, we can see the estimations of the model parameters are all significant for norm, t and skewed t distribution.

Compare the log-likelihood and information criterion in table 6 within the three conditional distributions, the model with conditional distribution of skewed t has larger log- likelihood and smaller information criterion statistics than estimated by normal and t distribution which means this model is better fitted.

Table 5: Estimation of the ARMA (0, 2)-GARCH (1, 1) with Different Conditional Distributions

	Norm	T	skewed t
Mu	4.57e-04 (7.06e-14)	5.28e-04 (< 2e-16)	4.50e-04 (2.46e-14)
ma1	1.11e-01 (< 2e-16)	1.06e-01 (< 2e-16)	1.03e-01 (< 2e-16)
ma2	-1.96e-02 (0.024)	-2.98e-02 (0.000273)	-3.43e-02 (3.26e-05)
Omega	7.12e-07 (1.82e-14)	5.52e-07 (2.81e-10)	5.36e-07 (4.43e-10)
alpha	8.04e-02 (< 2e-16)	7.27e-02 (< 2e-16)	7.20e-02 (< 2e-16)
beta	9.15e-01 (< 2e-16)	9.23e-01 (< 2e-16)	9.24e-01 (< 2e-16)
Shape	-----------	6.89e+00 (< 2e-16)	7.01e+00 (< 2e-16)
Skew	-----------	------------	9.51e-01 (< 2e-16)

Table 6: Analysis of Standardized Residual and Information of the Fitted Parameters in ARMA(0,2)-GARCH(1,1)

	Norm	t	skewed t
loglikelihood	51056.68	51487.61	51497.23
Jarque-Bera Test	14571.51	15042.07	14941.50
	(0.00000)	(0.00000)	(0.00000)
Ljung-Box Test	12.64196	16.29922	19.60367
R Q(10)	(0.24438)	(0.09138)	(0.03323)
Ljung-Box Test	17.35922	20.90359	24.23974
R Q(15)	(0.29785)	(0.13994)	(0.06113)
Ljung-Box Test	22.97918	26.59462	29.88170
R Q(20)	(0.28982)	(0.14706)	(0.07179)
Ljung-Box Test	18.11756	21.49811	21.62693
R^2 Q(10)	(0.05301)	(0.01788)	(0.01712)
Ljung-Box Test	20.66037	23.93266	24.03359
R^2 Q(15)	(0.14804)	(0.06625)	(0.06453)
Ljung-Box Test	24.51930	27.99330	28.10664
R^2 Q(20)	(0.22044)	(0.10956)	(0.10690)
LM ARCH test	19.28457	22.50661	22.62559
	(0.08189)	(0.03222)	(0.03108)
AIC	-6.84455	-6.90227	-6.90343
BIC	-6.84149	-6.89870	-6.89935

4.2.2 ARMA (0, 2)-APARCH (1, 1)

Table 7 gives us the results of the parameter estimation of the TS-GARCH model, GJR-GARCH model and DGE-GARCH model. The parameters estimated in these three models are all significant except for the coefficient of the second term of moving average process under the normal distribution. And also according to the log-likelihood value and information criterion of the estimated models, we can get the comparison results: Firstly, compare with TS-GARCH model and GJR-GARCH model, the DGE-GARCH model has larger log-likelihood value and smaller information criterion. Secondly, compare within the DGE-GARCH models under normal distribution, t distribution and skewed t distribution, obviously the model with conditional distribution is skewed t distribution outperforms other two distributions which means this model is superior in modeling the S&P 500 data stock index returns with asymmetry and fat tail.

In this paper, the estimated power δ is 1.402 which is slightly different from the estimated result of Ding, Granger and Engle (1993)'s under the normal distribution which is 1.43. This may be caused by the time period of the data is different and then mean equation is also different to model the data. But δ in this paper is still significantly different from 1 (TS-GARCH) or 2 (GARCH). When the conditional distribution changes to t distribution and skew t distribution, δ is getting smaller to 1.15, however, using the same test as in Ding, Granger and Engle (1993)'s paper, let l_0 be the log-likelihood of value under the GARCH model which is set as the null hypothesis, while the alternative hypothesis is APARCH model with log-likelihood is l, then $2(l-l_0)$ have a χ^2 distribution with 2 degrees of freedom when H_0 is true. Then, in this paper, under the skew-t distribution, $2(l-l_0)=2(51615-51497)=216$ which means we can reject the null hypothesis that the data is generated from GARCH model. And also in the same way we can reject that the data is generated from TS-GARCH model.

Table 7: Estimation of the ARMA (0, 2)-APARCH (1, 1) Models with Different Conditional Distributions

	TS-GARCH			GJR-GARCH			DGE-GARCH		
conditional distribution	norm	t	skewd t	Norm	t	Skewed t	Norm	t	Skewed t
mu	0.00045	0.00052	0.00044	0.00028	0.00041	0.00033	0.00026	0.00037	0.00030
	(0.000)	(< 2e-16)	(0.000)	(0.00001)	(0.000)	(0.000)	(0.00004)	(0.000)	(0.000)
ma1	0.10240	0.10440	0.10070	0.11480	0.11000	0.10750	0.11120	0.10790	0.10590
	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)
ma2	-0.01434	-0.02843	-0.03284	-0.01057	-0.02285	-0.02579	-0.00895	-0.02155	-0.02427
	(0.07340)	(0.00024)	(0.00007)	(0.22300)	(0.00533)	(0.00171)	(0.30100)	(0.00888)	(0.00286)
omega	0.00010	0.00007	0.00007	0.00000	0.00000	0.00000	0.00002	0.00004	0.00004
	(< 2e-16)	(0.00000)	(0.00000)	(< 2e-16)	(0.00000)	(0.00000)	(< 2e-16)	(0.00000)	(0.00000)
alpha	0.09161	0.07671	0.07639	0.06740	0.06448	0.06441	0.07562	0.06922	0.06910
	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)
Gamma	0.00000	0.00000	0.00000	0.30670	0.34980	0.34810	0.39990	0.55000	0.54850
				(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)
Beta	0.91940	0.93390	0.93430	0.91770	0.92030	0.92090	0.92340	0.93400	0.93430
	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)	(< 2e-16)
Delta	1.00000	1.00000	1.00000	2.00000	2.00000	2.00000	1.40200	1.15000	1.15100
							(< 2e-16)	(< 2e-16)	(< 2e-16)
shape	-------------	6.75	6.861	--------------	7.372	7.473	---------------	7.366	7.46500
		(< 2e-16)	(< 2e-16)		(< 2e-16)	(< 2e-16)		(< 2e-16)	(< 2e-16)
Skew	--------------	---------------	0.95200	---------------	---------------	0.94890	---------------	---------------	0.94840
			(< 2e-16)			(< 2e-16)			(< 2e-16)

Table 8: Analysis of Standardized Residual and Information of the Fitted Parameters in ARMA (0, 2)-APARCH (1, 1) Models

	TS-GARCH			GJR-GARCH			DGE-GARCH		
Conditional Distribution	Norm	t	skewd t	Norm	t	Skewed t	Norm	t	Skewed t
Log likelihood	50930.16	51469.95	51477.79	51161.62	51577.67	51588.24	51164.49	51605.25	51615.36
Jarque-Bera	26377.41	44648.65	45435.88	17389.7	19709.81	19431.68	21411.98	40324.03	39628.14
Test	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	(0.34813)	0.00000	0.00000
Ljung-Box	11.06500	11.37540	13.75185	10.68769	11.62235	13.06543	11.12157	10.28695	10.85113
R Q(10)	(0.35248)	(0.32903)	(0.18461)	(0.38236)	(0.31113)	(0.22004)	(0.34813)	(0.41569)	(0.36923)
Ljung-Box	14.79436	14.70279	17.05722	15.22662	15.82967	17.34297	15.36056	13.85703	14.45316
R Q(15)	(0.46633)	(0.47303)	(0.31546)	(0.43522)	(0.39346)	(0.29878)	(0.42577)	(0.53640)	(0.49147)
Ljung-Box	20.87438	20.95681	23.29916	20.29701	20.84051	22.28721	20.58139	19.07978	19.62234
R Q(20)	(0.40456)	(0.39968)	(0.27435)	(0.43949)	(0.40657)	(0.32513)	(0.42213)	(0.51665)	(0.48177)
Ljung-Box	199.36710	544.53720	590.40960	14.49714	15.20908	15.25789	23.91320	90.55194	90.81995
R^2 Q(10)	0.00000	0.00000	0.00000	(0.15150)	(0.12462)	(0.12294)	(0.00783)	(0.00000)	(0.00000)
Ljung-Box	199.85280	545.73980	591.68590	16.51665	17.38286	17.39979	25.06601	91.07042	91.33419
R^2 Q(10)	0.00000	0.00000	0.00000	(0.34857)	(0.29650)	(0.29553)	(0.04906)	(0.00000)	(0.00000)
Ljung-Box	202.22930	547.43880	593.36660	19.16863	20.18055	20.20704	27.30961	92.39806	92.66480
R^2 Q(10)	0.00000	0.00000	0.00000	(0.51089)	(0.44669)	(0.44505)	(0.12678)	(0.00000)	(0.00000)
LM Arch	54.25578	80.61406	79.71624	15.56676	16.39303	16.43228	24.80044	43.31198	43.26175
Test	(0.00000)	(0.00000)	(0.00000)	(0.21189)	(0.17389)	(0.17223)	(0.01580)	(0.00002)	(0.00002)
AIC	-6.82767	-6.89991	-6.90082	-6.85857	-6.91421	-6.91550	-6.85882	-6.91778	-6.91900
BIC	-6.82461	-6.89634	-6.89674	-6.85500	-6.91013	-6.91091	-6.85474	-6.91319	-6.91390

5. Forecast

Comparing within the asymmetric power ARCH models, finally we get the result that the ARMA (0, 2)-APARCH (1, 1) model with skew-t distribution is the most fitted model to model the conditional heteroscedasticity. Then the goal is to use this model do the forecasting both for the future returns of S&P 500 stock index and its conditional volatility and compare it with ARMA (0, 2)-GARCH (1, 1) model.

5.1 Forecasting Conditional Mean

As we select the ARMA (0, 2) as the mean equation, then we do the forecasting based on this model to forecast the future returns of the S&P 500 stock index.

ARMA (0, 2) process is shown as follows:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

which can be rewritten as $y_t - \mu = (1 + +\theta_1 L + \theta_2 L^2)\varepsilon_t$

According to (Hamilton, 1994), the forecast becomes

$\hat{y}_{t+h|t} = \mu + (\theta_h + \theta_{h+1} L + ... + \theta_2 L^{2-h})\hat{\varepsilon}_t$, when $h = 1,2$ $\hat{\varepsilon}_t = (y_t - \mu) - \theta_1 \hat{\varepsilon}_{t-1} - \theta_2 \hat{\varepsilon}_{t-2}$

The forecast farther than 2 periods in the future will be just the unconditional mean u .

5.2 Forecasting Conditional Variance

The conditional variance forecasting is independently from the conditional mean. In this paper, we choose APARCH (1, 1) model with skew-t distribution as the final model.

The h-step-ahead forecasting for APARCH (1, 1) model is (Laurent, Lambert, 2002)

$$\hat{\sigma}^{\delta}_{t+h|t} = E(\hat{\sigma}^{\delta}_{t+h|t} |\psi_t) = E\left(\hat{\omega} + \hat{\alpha}\ (|\varepsilon_{t+h-1}| - \hat{\gamma}\varepsilon_{t+h-1})^{\hat{\delta}} + \hat{\beta}\ \sigma^{\hat{\delta}}_{t+h-1} |\psi_t\right)$$

$$= \hat{\omega} + \hat{\alpha}\ E[(\varepsilon_{t+h-1} - \hat{\gamma}\varepsilon_{t+h-1})^{\hat{\delta}} |\psi_t] + \hat{\beta}\ \sigma^{\hat{\delta}}_{t+h-1|t}$$

Where $E[(\varepsilon_{t+k} - \hat{\gamma}\varepsilon_{t+k})^{\hat{\delta}} | \psi_t] = \kappa_i \sigma_{t+k|t}^{\hat{\delta}}$, for k>1, and Lambert and Laurent (2001) show in the skew-t distribution that

$$\kappa_i = \left\{\xi^{-(1+\delta)}(1+\gamma)^{\delta} + \xi^{1+\delta}(1-\gamma)^{\delta}\right\} \frac{\Gamma(\frac{\delta+1}{2})\Gamma(\frac{v-\delta}{2})(v-2)^{\frac{1+\delta}{2}}}{(\xi+\frac{1}{\xi})\sqrt{(v-2)\pi}\,\Gamma(\frac{v}{2})}$$

Table 9 and Table 10 show the 10-step-ahead forecast for the S&P 500 stock index returns modeled by ARMA(0,2)-APARCH(1,1) model under the skew-t distribution and ARMA (0, 2)-GARCH (1, 1) model with normal distribution. We can see that after the first two-step forecasting, the conditional mean are all the same with the unconditional mean. The forecasting standard deviation in the table is the conditional standard deviation of the asset returns which is $\hat{\sigma}_{t+h|t}$. Comparing the forecasting result of the two models, ARMA (0, 2)-APARCH (1, 1) model with skew-t distribution gives us a better forecasting result with a smaller mean error.

Table 9: 10-Step-Ahead Forecast for the S&P 500 Stock Index Returns Modeled by ARMA (0, 2)-APARCH (1, 1) Model under the Skew-t Distribution

	Mean Forecast	Mean Error	Standard Deviation
1	0.0003601666	0.009651459	0.01921733
2	0.0002086897	0.009705407	0.01910896
3	0.0002964023	0.009708232	0.01900164
4	0.0002964023	0.009708232	0.01889538
5	0.0002964023	0.009708232	0.01879016
6	0.0002964023	0.009708232	0.01868598
7	0.0002964023	0.009708232	0.01858282
8	0.0002964023	0.009708232	0.01848067
9	0.0002964023	0.009708232	0.01837952
10	0.0002964023	0.009708232	0.01827937

Table 10: 10-Step-Ahead Forecast for the S&P 500 Stock Index Returns Modeled by ARMA (0, 2)-GARCH (1, 1) Model under the Normal Distribution

	Mean Forecast	Mean Error	Standard Deviation
1	0.0005730848	0.009657464	0.02140746
2	0.0003914337	0.009717193	0.02136904
3	0.0004567766	0.009719039	0.02133076
4	0.0004567766	0.009719039	0.02129260
5	0.0004567766	0.009719039	0.02125457
6	0.0004567766	0.009719039	0.02121667
7	0.0004567766	0.009719039	0.02117889
8	0.0004567766	0.009719039	0.02114124
9	0.0004567766	0.009719039	0.02110372
10	0.0004567766	0.009719039	0.02106633

6. Conclusions and Further Discussion

To modeling the financial time series data, we review of the autoregressive conditional heteroscedasticity (ARCH) and generalized autoregressive conditional heteroscedasticity (GARCH) model. Consider of the stylized facts of the asset return series, GARCH model with normal distribution sometimes fails to capture the fat tail and asymmetry in the observed return series.

To overcome these drawbacks, one is to relax the assumptions of the conditional distribution to non-Gaussian i.e. student t distribution proposed by Bollerslev (1987) which can capture the excess kurtosis in the return series. On the other hand, to fit the leverage effect, adding the skew parameter into conditional distribution and the variance equation are both effectively capture the skewness of the data. According to the log-likelihood value and AIC, BIC, we get the empirical results that the skew-t distribution along with asymmetry power ARCH model outperforms the standard

GARCH model with normal conditional distribution.

This paper is based on the S&P 500 stock index returns, use ARMA model to fit the series dependence and get the error term which is generated from the conditional heteroscedasticity process. Use GARCH (1, 1) model and a more flexible exponent and asymmetric power ARCH (1, 1) models including TGARCH and GJR-GARCH to fit the data with the t distribution and the skew-t distribution comparing with normal distribution and get the empirical results that the ARMA (0, 2)-APARCH (1, 1) model estimated using asymmetric leptokurtic distribution is superior to other counterparts with t distribution or normal distribution. And based on the estimated model, a 10-step-ahead forecasting is taken to forecast the future value of the stock index returns and the conditional volatility. And also we compare the forecasting result between the ARMA (0, 2)-APARCH (1, 1) model with skew-t distribution and the ARMA (0, 2)-GARCH (1, 1) model with normal distribution, get the result that the ARMA (0, 2)-APARCH (1, 1) model is also superior in the forecasting.

However, there are several limitations in the paper: Besides GARCH and APARCH model, there are still many other linear and non-linear model to model the conditional heteroscedasticity such as TGARCH, EGARCH; Secondly, except for the skew-t distribution take the fat tail and asymmetry into consideration, the generalized error distribution and skew-generalized error distribution, stable Paretian distribution for the innovation can also used to model the heavy tails and leverage effect; Thirdly, if modeling the S&P 500 stock index compared with other similar stock market may give us more comprehensive result.

Reference

[1] Alberg, D., Shalit, H., & Yosef, R., (2008). Estimating stock market volatility using asymmetric GARCH models. *Applied Financial Economics*, 18, pp.1201–1208.

[2] Bai, X., Russell, J.R. & Tiao, G.C., (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2), pp.349-360.

[3] Black, F., (1976). Studies in stock price volatility changes, Proceedings of the 1976 business meeting of the business and economics statistics section. *American Statistical Association*, pp.177-181.

[4] Bollerslev, T., (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), pp.307-327.

[5] Bollerslev, T., (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics,* 69(3), pp.542–547.

[6] Ding, Z., Engle, R.F., & Granger, C.W.J., (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1, pp.83-106.

[7] Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica*, 50, pp.987-1008.

[8] Engle, R.F. (2001). GARCH 101: The use of ARCH/GARCH model in applied economics. *Journal of Economic Perspectives*, 15(4), pp.157-168.

[9] Fernandez, C. and Steel, M., (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93, pp.359–371.

[10] Hamilton, J.D., (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

[11] Hannan, E. J., (1980).The estimation of the order of an ARMA process. *The Annals of Statistics*, 8(5), pp.1071-1081.

[12] Ljung, G.M., Box, G. E. P., (1978). On a Measure of a Lack of Fit in Time Series Models. *Biometrika,* 65, pp. 297-303.

[13] Lambert, P., Laurent, S., (2001). Modelling financial time series using GARCH-type models and a skewed student density. Mimeo, Universite de Liege.

[14] Laurent, S., Lambert, P., (2002). A tutorial for GARCH 2.3, a complete Ox package for estimating and forecasting ARCH models. *GARCH 2.3 Tutorial*, pp. 71.

[15] Taylor, S.J., (1986). *Modeling financial time series*, New York: John Wiley & son.

[16] Tavares, A., Curto, J.D. & Tavares, G.N., (2008). Modeling heavy tails and asymmetry using ARCH-type models with stable paretian distributions. *Nonlinear Dynamics*, *Springer*, 51(1), pp. 231-243.

[17] Wurtz, D., Chalabi, Y., Luksan, Y., (2002). Parameter Estimation of ARMA Models with GARCH/APARCH Errors. An R and SPlus Software Implementation, zob. *Journal of Statistical Software.*

[18] Zivot, E., Practical Issues in the Analysis of Univariate GARCH Models, Department of Economics, University of Washington, No UWEC-2008-03-FC, Working Papers from University of Washington, Department of Economics.