Estimates and Forecasts of GARCH Model under Misspecified Probability Distributions: A Monte Carlo Simulation Approach

OlaOluwa S. Yaya
University of Ibadan, Ibadan, Nigeria, os.yaya@ui.edu.ng

Olusanya E. Olubusoye
University of Ibadan, Ibadan, Nigeria, busoye2001@yahoo.com

Oluwadare O. Ojo
Federal University of Technology Akure, Akure, Nigeria, daruu208075@yahoo.com

Estimates and Forecasts of GARCH Model under Misspecified Probability Distributions: A Monte Carlo Simulation Approach

OlaOluwa S. Yaya
University of Ibadan
Ibadan, Nigeria

Olusanya E. Olubusoye
University of Ibadan
Ibadan, Nigeria

Oluwadare O. Ojo
Federal University of
Technology Akure
Akure, Nigeria

The effect of misspecification of correct sampling probability distribution of Generalized Autoregressive Conditionally Heteroscedastic (GARCH) processes is considered. The three assumed distributions are the normal, Student t, and generalized error distributions. The GARCH process is sampled using one of the distributions and the model is estimated based on the three distributions in each sample. Parameter estimates and forecast performance are used to judge the estimated model for performance. The AR-GARCH-GED performed better on the three assumed distributions; even, when Student t distribution is assumed, AR-GARCH-Student t does not perform as the best model.

Keywords: Generalized Error Distribution, forecasts, GARCH, misspecification, specification

Introduction

Since the introduction of Generalized Autoregressive Conditional Heteroscedastic (GARCH) model of Bollerslev (1986), thousands of articles have been published applying the model on financial series. The model captures volatility in the market, and its distributional specification makes it special among other nonlinear time series models. The GARCH process exists on the assumption of Normal, Student t, and Generalized Error Distributions (GED). The Normal distribution is the usual assumption in any time series estimation, but due to the fact that the distribution of GARCH process is leptokurtic, Normal distribution was found to be in appropriate

in capturing the tail behavior of the series. Bollerslev (1987) therefore proposed Student t distribution to capture the long tail behavior of the process. Nelson (1991) proposed the GED distribution.

Apart from the real applications of GARCH models on financial series, there is need to study the effect of misspecifying the GARCH distributional assumptions during estimation. Articles are very scarce along this line of thought. Wang (2002) affirms that spurious and inefficient inference is expected when pure GARCH models are misspecified. This as well may affect the Quasi Maximum Likelihood Estimates (QMLEs) of the misspecified model. The QMLE of pure GARCH(1,1) models indicates that the ARCH parameter is small, GARCH parameter is close to unity and the sum of both parameters approaches unity as the sampling frequency increases (Engle and Bollerslev, 1986; Bollerslev and Engle, 1993; Baillie, Bollerslev and Mikkelsen, 1996; Ding and Granger, 1996; Andersen and Bollerslev, 1997, and Engle and Patton, 2001.) This fact is reflected in the Integrated GARCH (IGARCH) of Engle and Bollerslev (1986). A more recent paper by Jensen and Lange (2010) shows that in a GARCH (1,1) model, the estimates of $\hat{\alpha}_1$ and $\hat{\beta}_1$ tend to zero and unity respectively as the sampling frequency increased, which is an IGARCH effect. This IGARCH effect is known for pure-GARCH processes. In a linear AR-GARCH or nonlinear AR-GARCH processes, IGARCH effect is not plausible. The present work considers AR-GARCH process, and therefore IGARCH effect may not be expected.

As tail distribution of the GARCH model is captured using the three distributions, and parameters estimated adjust accordingly, forecasts performances of the model are affected. Extensive Monte Carlo simulation was performed on the GARCH model using the three distributions.

The GARCH (1,1) model

The GARCH (1,1) model proposed in Bollerslev (1986) is

$$\sigma_t^2 = w + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{1}$$

where ε_t are the returns series of the financial asset; σ_t is the volatility at time t and z_t gives the assumed distribution. The parameters, α_1 and β_1 are conditioned as $w > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ in order to ensure stationarity of the whole process (Bollerslev, 1986). This condition is establish by defining

$k_t = \varepsilon_t^2 - \sigma_t^2 = \left(z_t^2 - 1\right)\sigma_t^2$ where $z_t \approx N(0,1)$. Using this in (1) results in the Autoregressive Moving Average (ARMA) representation

$$\varepsilon_t^2 = w + \left(\alpha_1 + \beta_1\right)\varepsilon_{t-1}^2 - \beta_1 k_{t-1} + k_t \tag{2}$$

where k_t is serially uncorrelated with mean zero. Stationarity of the process is then ensured when the roots of $1 - \alpha(1) - \beta(1) = 1 - (\alpha_1 + \beta_1) = 0$ lie outside the unit circle and this is not conditioned on time t as it is measured directly from the parameters of the model. Hence it is expected that

$$\left(\alpha_1 + \beta_1\right) < 1 \tag{3}$$

for existence of covariance stationary process. For the stationary process, the finite unconditional variance of ε_t is given by

$$\sigma^2 = \frac{w}{1 - \alpha_1 - \beta_1} \tag{4}$$

Kurtosis of GARCH (1,1) model

For any GARCH (p,q) process, $E(z_t) = 0$ and $Var(z_t) = 1$. $E\left(z_t^3\right)$ is the skewness and $E\left(z_t^4\right)$ gives the measure of skewness. Because the emphasis is on tail behaviour of GARCH residuals, the expression for the unconditional kurtosis is next derived.

Assuming that $E\left(\sigma_t^2\right)$ and $E\left(\sigma_t^4\right)$ exist, then it suffices to write

$E\left(z_t^4\right) = k_z + 3;$

because $E\left(\varepsilon_t\right) = 0;$

$E\left(\varepsilon_t^4\right) = E\left(\sigma_t^4\right)E\left(z_t^4\right)$ because $\varepsilon_t = \sigma_t z_t$ and ε_t and z_t are independent

Then, squaring GARCH (1,1) model,

$$\sigma_t^2 = w + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

gives

$$\sigma_t^2 = w^2 + \alpha_1^2 \varepsilon_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2w\alpha_1 \varepsilon_{t-1}^2 + 2w\beta_1 \sigma_{t-1}^2 + 2\alpha_1 \beta_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2.$$

Taking expectation of the resulting expansion, as well as applying the properties outlined above

$$E\left(\sigma_t^2\right) = \frac{w^2\left(1 + \alpha_1 + \beta_1\right)}{\left[1 - \left(\alpha_1 + \beta_1\right)\right]\left[1 - \alpha_1^2\left(k_z + 2\right) - \left(\alpha_1 + \beta_1\right)^2\right]}$$

Using the relation $E\left(\varepsilon_t^4\right) = \left(k_z + 3\right)E\left(\sigma_t^4\right)$ where k_z is the excess kurtosis of z_t, then

$$E\left(\varepsilon_t^4\right) = \frac{w^2\left(1 + \alpha_1 + \beta_1\right)\left(k_z + 3\right)}{\left[1 - \left(\alpha_1 + \beta_1\right)\right]\left[1 - \alpha_1^2\left(k_z + 2\right) - \left(\alpha_1 + \beta_1\right)^2\right]}$$

Using the formula $k_\varepsilon = \dfrac{E\left(\varepsilon_t^4\right)}{\left[E\left(\varepsilon_t^2\right)\right]^2} - 3$ for excess kurtosis and with the fact that

$E\left(\varepsilon_t^2\right) = \dfrac{w}{1 - \alpha_1 - \beta_1}$ from the properties above,

$$
\begin{aligned}
k_\varepsilon &= \frac{w^2\left(1 + \alpha_1 + \beta_1\right)\left(k_z + 3\right)\left[1 - \left(\alpha_1 + \beta_1\right)\right]^2}{w^2\left[1 - \left(\alpha_1 + \beta_1\right)\right]\left[1 - \alpha_1^2\left(k_z + 2\right) - \left(\alpha_1 + \beta_1\right)^2\right]} - 3 \\[2mm]
&= \frac{\left(k_z + 3\right)\left[1 - \left(\alpha_1 + \beta_1\right)^2\right]}{\left[1 - 2\alpha_1^2 - \left(\alpha_1 + \beta_1\right)^2 - k_z\left(\alpha_1^2\right)\right]} - 3
\end{aligned}
$$

(5)

with normally distributed innovations z_t, $k_z = 0$ and

$$k_\varepsilon = \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} \tag{6}$$

with non-normally distributed innovations z_t, as in Student t and GED, $Var(z_t) = E(z_t^2) = 1$ and $k_z = \dfrac{E(z_t^4)}{\left[E(z_t^2)\right]^2} = E(z_t^4) \neq 0$, then

$$k_\varepsilon = \frac{k_z - k_z(\alpha_1 + \beta_1)^2 + 6\alpha_1^2 + 2k_z\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - k_z(\alpha_1^2)} \tag{7}$$

In these two cases, it is observed that α_1^2 is important in determining the tail behavior of ε_t, because once $\alpha_1^2 = 0$, $k_\varepsilon = 0$. Hence, $k_\varepsilon = k_z$ for the non-normally distributed case and it implies the similarity of the tail behaviors of both ε_t and z_t

Distributional Assumptions and Estimation

For GARCH models, the unconditional distributions are always non normal, and this gives fatter tails than the normal distribution. In practice, z_t is assumed to follow the normal distribution or non-normal distributions. These non-normal distributions have been proved to perform well in modeling the fatter tails (leptokurticity) observed in GARCH residuals. The non-normal distributions are the Student t distribution proposed in Bollerslev (1987) and Generalized Error Distribution (GED) by Nelson (1991)

The standardized Normal distribution is

$$f(z_t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_t^2\right), \quad -\infty < z_t < \infty \tag{8}$$

with the log likelihood function

$$L(z_t) = -\frac{1}{2}\left[N\log(2\pi) + \sum_{t=1}^{N} z_t^2\right] \tag{9}$$

where N is the sample size. The standardized Student t distribution proposed in Bollerslev (1987) is given as

$$f\left(z_t,v\right)=\Gamma\left((v+1)/2\right)\left[\Gamma\left(v/2\right)\right]^{-1}\left[\pi(v)\right]^{-1/2}\left(1+\frac{z_t^2}{v}\right)^{-(v+1)/2},\ -\infty<z_t<\infty \qquad (10)$$

This distribution is symmetric around zero as it is observed in its specification with $v>2$. At $v=1$, the Student t reduces to Cauchy distribution. At $2<v\le 4$, its conditional kurtosis is less than 3, which means that the resulting tail effect is normal. For $v>4$, the kurtosis becomes $3(v-1)(v-4)^{-1}$, which is greater than 3, hence the tail effect becomes non-normal distribution. As $v\to\infty$, the distribution converges to normal distribution. The log likelihood function of Student t distribution is then simplified as

$$L\left(z_t,v\right)=-\frac{1}{2}\left\{N\log\left(\frac{\pi(v-2)\Gamma(v/2)^2}{\Gamma\left((v+1)/2\right)^2}\right)+(v+1)\sum_{t=1}^{N}\log\left[1+\frac{z_t^2}{(v-1)}\right]\right\} \qquad (11)$$

The standardized GED proposed in Nelson (1991) is given as

$$f\left(z_t,v\right)=2^{-1}v\Gamma\left(\frac{3}{v}\right)^{1/2}\left[\Gamma\left(\frac{1}{v}\right)^{3/2}\right]^{-1}\exp\left[-\left|z_t\right|^v\left\{\Gamma\left(\frac{3}{v}\right)\left[\Gamma\left(\frac{1}{v}\right)\right]^{-1}\right\}^{v/2}\right] \qquad (12)$$

where $-\infty<z_t<\infty$ and $v>0$. The GED reduces to the standard normal distribution at $v=4$. At $0<v<2$, the distribution has thicker tail than the normal distribution, for example, at $v=1$ the distribution becomes a double exponential (Laplace) distribution. At $v>2$, the distribution of z_t has thinner tails than the normal distribution, for example, as v tends to infinity, z_t reduces to a uniform distribution on the interval $\left(-\sqrt{3},\sqrt{3}\right)$. The log likelihood of this distribution is then expanded as

$$f\left(z_t,v\right)=2^{-1}v\Gamma\left(\frac{3}{v}\right)^{1/2}\left[\Gamma\left(\frac{1}{v}\right)^{3/2}\right]^{-1}\exp\left[-\left|z_t\right|^v\left\{\Gamma\left(\frac{3}{v}\right)\left[\Gamma\left(\frac{1}{v}\right)\right]^{-1}\right\}^{v/2}\right] \qquad (13)$$

These likelihood functions are then estimated using the numerical derivatives based on the fact that GARCH models lack closed form estimation. Berndt, Hall, Hall and Hausman (BHHH) algorithm of Berndt, et al (1974) is then used. This algorithm is termed Gauss-Newton in general Nonlinear Least Squares (NLS) and BHHH in MLE estimation. Unlike some other derivatives, it uses only first derivatives of the likelihood function and computes a set of parameter values as

$$\psi^{(i+1)} = \psi^{(i)} - \left(\sum_{t=1}^{N} \frac{\partial L\left(z_t,. \right)_t^{(i)}}{\partial \psi} \cdot \frac{\partial L\left(z_t,. \right)_t^{(i)}}{\partial \psi'} \right)^{-1} \frac{\partial L\left(z_t,. \right)_N^{(i)}}{\partial \psi} \tag{14}$$

where $L(z_t,.)$ is the likelihood function. The initial parameter set is given as $\psi^{(0)}$ and the parameter set which maximize the likelihood function is denoted as $\psi^{(i+1)}$. The estimation of GARCH (1, 1) model with Student t distribution and GED follow the usual Quasi Maximum Likelihood Estimation (QMLE) because normality assumption is violated in these cases.

Misspecification of distribution of GARCH model could lead to stationarity and explosion of the series in some points. Though standard errors will be consistent; the QML estimators $\tilde{\psi}^{(i+1)}$ are generally closed to the exact ML estimator $\hat{\psi}^{(i+1)}$ for symmetric GARCH distribution. For non-symmetric conditional distributions, both the asymptotic and finite sample loss in efficiency are quite large and parametric estimation approach are not applicable in this regard (Mills and Markellos, 2008).

Forecasts Evaluation

Forecast evaluation criteria considered are the Root Mean Squares Forecast Error (RMSFE), Mean Absolute Error (MAE), Mean Absolute Percentage Forecast Error (MAPFE) and Theil Inequality of Theil (1961;1966). The MSFE is defined as

$$MSFE = \frac{1}{m} \sum_{t=1}^{m} \left(\hat{\sigma}_t^2 - \sigma_t^2 \right)^2 \tag{15}$$

where $\hat{\sigma}_t^2$ is the predicted in-sample conditional variances, and this depends on the scale of the variance series, σ_t^2. The square root of MSFE is the RMSFE

485

$$RMSFE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}\left(\hat{\sigma}_t^2 - \sigma_t^2\right)^2} \tag{16}$$

The MAFE and MAPFE are obtained by taking the absolute differences of the predicted conditional volatilities and the observed volatilities as

$$MAFE = \frac{1}{m}\sum_{t=1}^{m}\left|\hat{\sigma}_t^2 - \sigma_t^2\right| \tag{17}$$

$$MAPFE = 100\sum_{t=1}^{m}\left|\frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}\right| \tag{18}$$

The Theil inequality is given as

$$TI = \frac{\sqrt{\frac{1}{m}\sum_{t=1}^{m}\left(\hat{\sigma}_t^2 - \sigma_t^2\right)^2}}{\sqrt{\frac{1}{m}\sum_{t=1}^{m}\hat{\sigma}_t^2} + \sqrt{\frac{1}{m}\sum_{t=1}^{m}\sigma_t^2}} \tag{19}$$

The inequality coefficient is time invariant and always lies between 0 and unity. The smaller these forecast evaluation criteria, the better the candidate model represent well the data.

Monte Carlo Simulations

The Monte Carlo experiment is set up using the AR(1)−GARCH(1,1) DGP

$$y_t = 0.15 + 0.5y_{t-1} + \varepsilon_t, \ \sigma_t^2 = 0.02 + 0.25\varepsilon_{t-1}^2 + 0.60\sigma_{t-1}^2, \tag{20}$$

with the error distribution $\varepsilon_t = \sigma_t z_t$ where z_t is assumed to follow Normal, Student t and GED distributions. The parameters of the AR(1) and GARCH(1,1) models are set within the stationary region in order to avoid problems data explosion. The sample sizes N are varied as 2000, 4000 and 6000 with in-sample forecasts generated as 25% of the data length. The results are then presented as Scenarios 1

to 3 in Tables 1−6 below. Each Scenario gives results for parameter estimation, volatility, excess kurtosis and forecasts evaluations criteria.

Scenario 1: When the true Distribution is Normal

Tables 1 and 2 present the results when the GARCH processes are simulated based on Normal distribution assumption, and these processes are used to estimate the GARCH process based on Student t, GED and the same Normal distribution. The results in Table 1 show that the AR-GARCH parameter estimates, measures of volatility and kurtosis are not consistent with sample sizes. Both the AR and GARCH parameter estimates computed for Student t distribution have larger biases in compared with that of Normal and GED distributions, even though excess kurtosis of the AR-GARCH-Student t model is the smallest. Volatility of the AR-GARCH-Student t model is also observed to be higher than that of the Normal and GED distributions. The excess kurtosis of the AR-GARCH-Normal model was expected to be the smallest because the series is sampled from Normal distribution but this was not the case.

Looking at the results of the in-sample forecasts realized from the AR-GARCH models as given in Table 2, the AR-GARCH-Normal and AR-GARCH-GED model perform better than AR-GARCH-Student t model on forecasts as given by the minimum values of the RMSPE and Theil inequality coefficients. The AR-GARCH-GED is expected to realize better forecasts than AR-GARCH-Normal model.

Table 1. Model Parameter, Volatility and Kurtosis when GARCH processes are simulated based on Normal distribution assumption

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	$\hat{w}$ (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1480	0.4839	0.0169	0.2110	0.6596	0.8706	0.1306	1.6427
	4000	0.1518	0.4724	0.0173	0.2049	0.6590	0.8639	0.1271	1.8430
	6000	0.1475	0.4750	0.0180	0.2052	0.6503	0.8555	0.1246	1.3275
Student t	2000	0.1462	0.4868	0.0820	0.1500	0.6000	0.7500	0.3280	0.2895
	4000	0.1471	0.4794	0.0794	0.1500	0.6000	0.7500	0.3176	1.8408
	6000	0.1576	0.4916	0.0849	0.1500	0.6000	0.7500	0.3396	0.4501
GED	2000	0.1499	0.4811	0.0160	0.2125	0.6831	0.8956	0.1533	2.3761
	4000	0.1547	0.4662	0.0172	0.2106	0.6743	0.8849	0.1494	2.1683
	6000	0.1495	0.4723	0.0185	0.2118	0.6612	0.8730	0.1457	1.7707

Table 2. Forecast evaluation estimates when GARCH processes are simulated based on Normal distribution assumption

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0003	0.0112	48.0967	0.0289
	4000	0.0002	0.0104	44.8820	0.0257
	6000	0.0001	0.0101	49.5855	0.0287
Student *t*	2000	0.0011	0.0464	500.3362	0.1270
	4000	0.0007	0.0437	529.5600	0.1242
	6000	0.0008	0.0602	503.3612	0.1363
GED	2000	0.0003	0.0110	50.1820	0.0276
	4000	0.0002	0.0102	47.7504	0.0246
	6000	0.0001	0.0099	53.1115	0.0276

Scenario 2: When the true Distribution is Student t

Tables 3 and 4 present the results when the true GARCH distribution follows Student *t*. Here, the distinctions in the GARCH estimates can only be made using the persistence and unconditional volatility measures. The AR-GARCH-Student *t* model still presents smallest persistence and highest volatility. The excess kurtosis of the AR-GARCH-Student *t* model is the smallest followed by that of AR-GARCH-Normal model.

Table 3. Model Parameter, Volatility and Kurtosis when the true GARCH distribution follows Student *t*

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	$\hat{w}$ (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1408	0.5096	0.0232	0.2295	0.6105	0.8400	0.1450	2.5966
	4000	0.1478	0.4830	0.0237	0.2576	0.5762	0.8338	0.1426	4.0384
	6000	0.1497	0.4878	0.0219	0.2483	0.6032	0.8515	0.1475	3.7926
Student *t*	2000	0.1510	0.5165	0.0938	0.1500	0.6000	0.7500	0.3752	0.7336
	4000	0.1472	0.4829	0.0911	0.1500	0.6000	0.7500	0.3644	0.8526
	6000	0.1408	0.5114	0.0961	0.1500	0.6000	0.7500	0.3844	1.1013
GED	2000	0.1416	0.4967	0.0261	0.2476	0.5979	0.8455	0.1689	3.5293
	4000	0.1473	0.4792	0.0243	0.2620	0.5937	0.8557	0.1684	5.7359
	6000	0.1491	0.4890	0.0230	0.2585	0.6093	0.8678	0.1740	5.9655

Table 4. Forecast evaluation estimates when the true GARCH distribution follows Student *t*

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0004	0.0145	63.7385	0.0332
	4000	0.0002	0.0131	62.2680	0.0324
	6000	0.0003	0.0247	63.1032	0.0392
Student *t*	2000	0.0015	0.0627	635.1114	0.1552
	4000	0.0010	0.0574	611.0665	0.1471
	6000	0.0009	0.0678	645.0181	0.1582
GED	2000	0.0004	0.0156	76.3588	0.0355
	4000	0.0002	0.0132	68.0163	0.0321
	6000	0.0004	0.0253	70.6578	0.0396

In terms of forecasts, the AR-GARCH-Student *t* model is the worst, even though the DGP is realized from the same probability distribution. The forecast performances of AR-GARCH-Normal and AR-GARCH-GED seem not different from each other as indicated by the forecast evaluation estimates.

Scenario 3: When the true Distribution is GED

Table 5 and 6 present the results when the true GARCH distribution is GED. In Table 5, in the AR(1) estimates, the estimates for the constant $\hat{\phi}_0$ are all consistent with sample sizes when the three probability distributions are assumed. The autoregressive parameters $\hat{\phi}_1$ are not consistent with sample sizes. The GARCH parameter estimates computed for Student t distribution are the same to that of Table 1 and 3 while the AR(1) parameter are different in the two results. The Student *t* distribution assumption of GARCH model still presents model estimates with highest volatility but with lowest persistence of this volatility. Misspecifying GED for Student *t* distribution here also caused the excess kurtosis to be negative in AR-GARCH-Student *t* model and this is a very spurious case.

The forecast evaluation results of the model estimates follow in In Table 6. Starting with the AR(1)−GARCH(1,1)−Student *t* model, the model is the worst in terms of forecasts because it presents the highest RMSPE, MAPE, MAPFE and Theil inequality coefficient. The best model is AR(1)−GARCH(1,1)−GED model, and this is expected because the DGP assumed GED initially. The performance of

AR(1)−GARCH(1,1)−Normal in terms of forecast is very close to that of AR(1)−GARCH(1,1)−GED model.

Table 5. Model Parameter, Volatility and Kurtosis when the true GARCH distribution is GED

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	$\hat{w}$ (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1409	0.5120	0.0219	0.2260	0.6112	0.8372	0.1345	0.1984
	4000	0.1478	0.4848	0.0236	0.2576	0.5639	0.8215	0.1322	0.4768
	6000	0.1497	0.4880	0.0218	0.2465	0.5932	0.8397	0.1360	0.3859
Student t	2000	0.1373	0.5192	0.0877	0.1500	0.6000	0.7500	0.3508	-0.3883
	4000	0.1469	0.4863	0.0854	0.1500	0.6000	0.7500	0.3416	-0.3554
	6000	0.1500	0.4860	0.0875	0.1500	0.6000	0.7500	0.3500	-0.4099
GED	2000	0.1424	0.5024	0.0225	0.2284	0.6031	0.8315	0.1335	0.1874
	4000	0.1481	0.4834	0.0228	0.2493	0.5763	0.8256	0.1307	0.4139
	6000	0.1498	0.4900	0.0214	0.2430	0.5973	0.8403	0.1340	0.3518

Table 6. Forecast evaluation estimates when the true GARCH distribution is GED

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0002	0.0099	48.9918	0.0280
	4000	0.0002	0.0100	50.8079	0.0292
	6000	0.0002	0.0142	50.9182	0.0325
Student t	2000	0.0013	0.0529	521.4746	0.1396
	4000	0.0008	0.0496	499.5437	0.1329
	6000	0.0008	0.0547	494.8839	0.1379
GED	2000	0.0002	0.0100	49.5704	0.0284
	4000	0.0002	0.0097	48.9723	0.0284
	6000	0.0002	0.0140	49.9722	0.0322

Conclusion

The misspecification of GARCH probability distribution functions were considered. These are the Normal, Student t and Generalized Error Distributions (GED). The estimation convergence time varied based on the distribution and the set sample sizes. When a Normal distribution was assumed, the AR−GARCH−GED seemed to perform marginally better than AR−GARCH−Normal model in terms of forecasts as revealed in the estimates of the Theil inequality. Though, the AR−GARCH−Normal was the best model here in terms of parameter estimates, and this was expected because the DGP assumed Normal distribution initially. With the assumption of Student t distribution in the DGP, the forecast performance of the models computed with Normal distribution and GED reduced and these still presented better models than the corresponding AR−GARCH−Student t model. Similar results were obtained when the DGP assumed GED.

It was also observed that all the results obtained, particularly the parameter estimates were not consistent with sample sizes. These are expected because volatility came into play. In empirical modeling research like this, interest should either lie in the behavior of the volatility–assuming a probability distribution which will give us the best volatility measurement–or in the forecasts. The best GARCH model may not actually produce the best forecast estimates and probability distributions have effect on the tail distribution of the innovations. This work can be replicated using higher order of the model, and in that case, more sophisticated software is recommended for the simulation in order to avoid convergence problems.

References

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance, 4*, 115-158.

Baillie, R.T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractional integrated generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics, 74*, 3-30.

Berndt, E., Hall, B., Hall, R., & Hausman, J. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement, 3*, 653-665.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics, 69*, 542-547.

Bollerslev, T., & Engle, R. F. (1993). Common persistence in conditional variances. *Econometrica, 61*, 167-186.

Ding, Z., & Granger, C. W. J. (1996), Modelling volatility persistence of speculative returns: a new approach. *Journal of Econometrics, 73*, 185-215.

Engle, R. F., & Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews, 5*, 1–50.

Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, *1*, 237-245.

Jensen, A. T., & Lange, T. (2010). On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics, 2*(1), 1-31.

Mills, T. C., & Markellos, R. (2008). *The Econometric Modelling of Financial Time Series* (3rd ed.). Cambridge, UK: Cambridge University Press.

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

Theil, H. (1961). *Economic forecasts and policy*. Amsterdam: North-Holland.

Theil, H. (1966). *Applied Economic forecasting*. Amsterdam: North-Holland

Wang, Y. (2002), Asymptotic nonequivalence of GARCH models and diffusions. *Annals of Statistics*, *30*(3), 754-783.