International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 20 (2016) pp. 10125-10130

Validation of Combine White Noise using Simulated Data

Ayodele Abraham Agboluaje[1,2], Suzilah Ismail[1] and Chee Yin Yip[3]

[1]School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara, 06010 UUMSintok, Malaysia;
[2]Department of Mathematics and Computer Science, Faculty of Natural Sciences,
Ibrahim Badamasi Babangida University, Lapai, Nigeria;
[3]Department of Economics, Faculty of Business and Finance, Universiti Tuanku Abdul Rahman, Malaysia.

Abstract
Recent studies reveal that the data that exhibits heteroscedasticity are modelled by Exponential Generalized Autoregressive Conditional Heteroscedastic (EGARCH). Nevertheless, EGARCH model estimation is not efficient when the heteroscedasticity data have leverage effect. In this study, an algorithm is developed which is called Combine White Noise (CWN). The standardized residuals of EGARCH errors (heteroscedastic variance) are decomposed into equal variances (white noise series). The white noise series are transformed into Combine White Noise Model (CWN). The assessments of the model are based on data simulation. The simulated data of 200 and 300 sample sizes of EGARCH are generated. The generated EGARCH data are based on low, moderate and high values of leverage and skewness. Each of these generated EGARCH data is used for the estimation of EGARCH and Moving Average (MA). The same generated EGARCH data are transformed to obtain CWN data and VAR data for the estimation of CWN and VAR. Each CWN results outperformed every result of the existing models. These results confirm that CWN is the appropriate model for estimation. The CWN model fit best in the transformed 200 sample sizes of EGARCH generated data with moderate leverage and moderate skewness. While the best forecast is in the transformed 200 sample sizes of EGARCH generated data with high leverage and moderate skewness. 200 sample sizes of EGARCH generated data with right values of leverage and skewness are better than using 300 sample sizes to have reliable output.

Keywords: Heteroscedasticity data, Leverage effect, Combine white noise, model fit, Best forecast

INTRODUCTION
In recent research, the data that exhibits heteroscedasticity with or without leverage effect have been modelled by series of GARCH family models and these models have not given a better estimation output [1, 2, 3, 4, 5]. In stochastic time series, the error term exhibit different behaviour in the model is according to the characteristics of the data that are employed.

In the early time, stochastic time series data and even now, the error terms in some data manipulations are white noise errors which can easily be modelled with promising estimation. A white noise process is a time series that has mean zero, constant variance and all autocorrelations equal to zero [6].

(7) introduces VAR that offer a realistic and credible procedure to data description, forecasting, structural inference and policy assessment, white noise error is the error term of the VAR model. When the error term becomes heteroscedastic (unequal variances) in nature, VAR can no longer estimate this type of error term effectively [8].

(9) introduces Autoregressive Conditional Heteroscedasticity (ARCH) model to uplift the VAR weaknesses because of time varying volatility. ARCH models are able to grab the set of errors and economic forecaster can survive any changes that are made. ARCH cannot handle the large lag structure. (10) introduces generalized ARCH (GARCH) which is flexible to overcome the weaknesses of ARCH model. (11, 12) argue that there are excess kurtosis and volatility persistence in GARCH. This is GARCH weaknesses; therefore, series of GARCH family address the challenges as follows.

The integrated GARCH model explains the resemblance with ARIMA (0, 1, 1) model as the definition of an ACF of squared sample sizes .In case the sample sizes are stationary in first difference then the model is known as IGARCH. Exponential GARCH overcomes the problem of conditional variance persistent [6].

Threshold GARCH and exponential GARCH capture the asymmetric effects of positive and negative shocks of the same lengths on conditional volatility in different dimensions. Leverage is a particular case of asymmetry. Positivity restriction on the parameters of the model makes EGARCH to capture the asymmetry, but not the leverage effect [1, 2, 3, 4].

This study focus on the importance of Combine White Noise (CWN) model to have a reliable estimation for accurate forecasting and to improve the economy, this alleviates the weaknesses of asymmetry in conditional variance of EGARCH model.

METHODOLOGY
Consider the autoregression model

$$y_t = \phi y_{t-1} + \varepsilon_t, \qquad (2.1)$$

Permit the stochastic approach of a real-valued time to be ε_t, and the complete information through t time is I. The GARCH model is

$$\varepsilon_t \mid \mathrm{I}_{t-1} \sim \mathrm{N}(0, h_t), \qquad (2.2)$$

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$
$$= \omega + A(L)\varepsilon_t^2 + B(L)h_t \qquad (2.3)$$

The EGARCH specification is

$$\log h_t = \alpha + \beta \mid z_{t-1} \mid + \delta z_{t-1} + \gamma \log h_{t-1}, \quad \mid \gamma \mid < 1 \qquad (2.4)$$

where $z_t = \varepsilon_t / \sqrt{h_t}$ is the standardized shocks, $z_t \sim iid(0,\alpha)$. $|\gamma|<1$ is when there is stability. The impact is asymmetric if $\delta \neq 0$, although, there is existence of leverage if $\delta < 0$ *and* $\delta < \beta < -\delta$. While both β and δ must be positive which the variances of two stochastic processes are, then, modeling leverage effect is not possible [3, 4].

The unequal variances (heteroscedastic errors) behaviors in the process of estimation being exhibited by GARCH models can be simplified into Combine White Noise model. The standardized residuals of GARCH errors which are unequal variances are decomposed into equal variances (white noise) in series to deal with the heteroscedasticity. The regression model is employed to transform each equal variances series to model.

Moving average process is employed for the estimation of these white noise series which is called Combine White Noise.

$$Y_1 = \varepsilon_{1t} + \theta_{11}\varepsilon_{1,t-1} + \theta_{12}\varepsilon_{1,t-2} + ..\theta_{jq}\varepsilon_{j,t-q}$$

$$Y_2 = \varepsilon_{2t} + \Phi_{21}\varepsilon_{2,t-1} + \Phi_{22}\varepsilon_{2,t-2} + ..\Phi_{jq}\varepsilon_{j,t-q}$$

$$Yj = \varepsilon_{jt} + \phi_{j1}\varepsilon_{j,t-1} + \phi_{j2}\varepsilon_{j,t-2} + ..\phi_{jq}\varepsilon_{j,t-q}$$

$$Y_{jt} = \sum_{j=1}^{q}\theta_j\varepsilon_{j,t-q} + \sum_{j=1}^{q}\Phi_j\varepsilon_{j,t-q} + \ldots\sum_{j=1}^{q}\phi_j\varepsilon_{j,t-q} \quad (2.5)$$

$$= A(L)\varepsilon_t + B(L)\varepsilon_t + \ldots$$

$$= \varepsilon_t[A(L) + B(L) + \ldots] \quad (2.6)$$

$$= Q\varepsilon_t \quad (2.7)$$

$$= U_t,$$

It can be written as

$$Y_t = U_t, \qquad (U_t \sim \mathrm{N}(0,\sigma_c^2)) \quad (2.8)$$

where A(L) + B(L) + … = Q which are the matrix polynomial, U_t is the error term of combine white noise model and σ_c^2 is the combination of equal variances.

The combine variances of the combine white noise is

$$\sigma_c^2 = \sigma_1^2 + \sigma_2^2 + \ldots \quad (2.9)$$

Considering the best two variances in the best two models produced by the Bayesian model averaging output. The combine variance follows:

$$\sigma_c^2 = \sigma_1^2 + \sigma_2^2 \quad (3.0)$$

The variance of errors, σ_c^2 in the combine white noise can be written:

$$\sigma_c^2 = W^2\sigma_1^2 + (1-W)^2\sigma_2^2 + 2\rho W\sigma_1(1-W)\sigma_2$$
(3.1)

where the balanced weight specified for the model is *W*. The least of σ_c^2 appearing, when the equation is differentiated with respect to *W* and equate to zero, obtaining:

$$W = \frac{\sigma_c^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (3.2)$$

Where ρ is the correlation; intra-class correlation coefficient is used for a reliable measurement.

RESULTS

The 200 and 300 sample sizes simulated data are used to estimate the parameters of EGARCH generated using the betategarch package in R software. The200 and 300 sample sizes generated EGARCH data are based on low, moderate and high values of leverage and skewness [13, 14].
The estimates of the parameters of 200 sample sizes generated EGARCH with the values of low, moderate and high leverage and skewness are reported in the Table 1. The estimated parameters are close to the EGARCH for the postulation model.

Table 1. The Estimated Parameters of the Simulated Data of EGRCH for Postulated (CWN) Model

200 sample sizes with low leverage and low skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	0.5
Estimates	-0.06	0.51	0.08	0.01	7.50	0.46
200 sample sizes with low leverage and moderate skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	0.7
Estimates	-0.03	0.69	0.04	-0.03	11.14	0.56
200 sample sizes with both low leverage and high skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	1.2
Estimates	0.06	0.51	0.05	0.04	7.57	1.21

200 sample sizes with moderate leverage and low skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	0.5
Estimates	-0.09	0.06	0.05	0.07	9.99	0.41

200 sample sizes with moderate leverage and moderate skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	0.7
Estimates	0.01	0.63	0.06	0.01	9.99	0.60

200 sample sizes with moderate leverage and high skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	1.2
Estimates	0.07	0.51	0.05	0.09	8.04	1.21

200 sample sizes with high leverage and low skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	0.5
Estimates	-0.13	0.21	0.05	0.11	9.98	0.40

200 sample sizes with moderate leverage and moderate skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	0.7
Estimates	0.01	0.46	0.07	0.05	9.38	0.60

200 sample sizes with high leverage and high skewness

Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	1.2
Estimates	0.08	0.52	0.05	0.13	8.78	1.21

The simulation of the generated 200 sample sizes of EGARCH with low, moderate and high leverage and skewness are used for the estimation of CWN, VAR, EGARCH and MA. Among the four models, CWN outperform the VAR, EGARCH and MA output.

Among all the CWN, the sample sizes of 200 generated EGARCH data with moderate leverage and moderate skewness output prove that CWN5 output have the minimum information criteria values of AIC equals -7.8471, BIC equals -7.7478, standard error value of 0.0239 and log likelihood highest value of 786.787. This makes the model to be the best fit among the different values of low, moderate and high leverage and skewness of the 200 generated EGARCH sample sizes. The best forecast for CWN output are in 200 generated EGARCH sample sizes with high leverage and moderate skewness of values 0.0417 RMSE, 0.0087 MAE and 1.0861MAPE [15] as reported in Table 2.

Table 2. 200 observations with leverage and skewness

	CWN1	CWN2	CWN3	CWN4	CWN5	CWN6	CWN7	CWN8	CWN9
Estimation									
Model fit									
Std Error	0.0716	0.1242	0.0498	0.1282	0.0239	0.2502	0.0718	0.0959	0.2120
Log L	162.5676	328.5	390.947	372.375	786.787	30.005	487.072	510.1128	92.8775
AIC	-1.5735	-3.2412	-3.8688	-3.6822	-7.8471	-0.2413	-4.8349	-5.0664	-0.8731
BIC	-1.4742	-3.1419	-3.7695	-3.5829	-7.7478	-0.1420	-4.7356	-4.9671	-0.7738
Dynamic Forecast Evaluation									
RMSE	0.1630	0.1247	0.0283	0.1504	0.0240	0.4098	0.0424	0.0417	0.1617
MAE	0.1058	0.0623	0.0080	0.0898	0.0116	0.3509	0.0121	0.0087	0.0349
MAPE	10.6726	6.2157	1.9950	9.1120	5.7104	1.3992	1.9950	1.0861	1.1621

CWN1 stand for CWN with low leverage and low skewness. CWN2 stand for CWN with low leverage and moderate skewness. CWN3 stand for CWN with low leverage and high skewness. CWN4 stand for CWN with moderate leverage and low skewness. CWN5 stand for CWN with *moderate leverage and moderate skewness. CWN6 stand for CWN with moderate leverage and high skewness. CWN7 stand for CWN with high leverage and low skewness. CWN8 stand for CWN with high leverage and moderate skewness. CWN9 stand for CWN with high leverage and high skewness.*

The estimates of the parameters of 300 sample sizes generated EGARCH with the values of low, moderate and high leverage and skewness are reported Table 3. The estimated parameters are close to the EGARCH for the postulation model.

Table 3. The Estimated Parameters of the Simulated Data of EGRCH for Postulated (CWN) Model

300 sample sizes with low leverage and low skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	0.5
Estimates	0.04	0.46	0.10	-0.01	6.06	0.56
300 sample sizes with low leverage and moderate skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	0.7
Estimates	0.01	0.55	0.10	0.02	5.70	0.82
300 sample sizes with low leverage and high skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.01	10	1.2
Estimates	-0.04	0.61	0.08	0.02	6.61	1.32
300 sample sizes with moderate leverage and low skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	0.5
Estimates	0.04	0.46	0.10	0.05	5.86	0.57
300 sample sizes with moderate leverage and moderate skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	0.7
Estimates	0.01	0.56	0.10	0.07	5.63	0.82
300 sample sizes with moderate leverage and high skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.05	10	1.2
Estimates	-0.02	0.58	0.07	0.11	8.38	1.33
300 sample sizes with high leverage and low skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	0.5
Estimates	0.07	0.47	0.10	0.10	6.13	0.57
300 sample sizes with high leverage and low skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	0.7
Estimates	-0.01	0.56	0.09	0.11	5.6	0.82
300 sample sizes with high leverage and high skewness						
Parameters	ω	α	β	δ	df	γ
For Postulation	0.01	0.5	0.1	0.09	10	1.2
Estimates	-0.02	0.58	0.07	0.11	8.38	1.33

The simulation of the generated 300 sample sizes of EGARCH with low, moderate and high leverage and skewness are used for the estimation of CWN, VAR, EGARCH and MA. Among the four models, CWN outputs outperform the VAR, EGARCH and MA outputs.

Among all the CWN, 300 sample sizes of generated EGARCH data with low leverage and low skewness outputs prove that CWN1 output have the minimum information criteria values of AIC equals -6.0041, BIC equals -5.9298 and log likelihood highest value of 903.6062 that give the best output, but the lowest standard error is 0.0911 which is in low leverage and high skewness. This makes the model to be the best fit among the different values of low, moderate and high leverage and skewness of the 300 generated EGARCH sample sizes. The best forecast for CWN output are in 300 generated EGARCH sample sizes with high leverage and high skewness which are of values 0.0722 RMSE, 0.0306 MAE and 3.3643 [15] as reported in Table 4.

Table 4. 300 observations with leverage and skewness

	CWN1	CWN2	CWN3	CWN4	CWN5	CWN6	CWN7	CWN8	CWN9
Estimation									
Model fit									
Std Error	0.1249	0.1191	0.0911	0.1694	0.0978	0.1254	0.0249	0.0266	0.0492
Log L	903.606	426.739	434.348	807.345	493.960	563.634	827.110	867.438	759.876
AIC	-6.0041	-2.8143	-2.8652	-5.3602	-3.2639	-3.7300	-5.4924	-5.7621	-5.0426
BIC	-5.9298	-2.7401	-2.7909	-5.2859	-3.1897	-3.6557	-5.4181	-5.6879	-4.9684
Dynamic Forecast Evaluation									
RMSE	0.1253	0.1363	0.1356	0.1634	0.0789	0.1130	0.0221	0.0197	0.0722
MAE	0.0633	0.0743	0.0615	0.0890	0.0309	0.0510	0.0098	0.0077	0.0306
MAPE	6.2777	7.4340	5.1043	7.4251	3.8897	5.1075	4.8911	3.8896	3.3643

CWN1 stand for CWN with low leverage and low skewness. CWN2 stand for CWN with low leverage and moderate skewness. CWN3 stand for CWN with low leverage and high skewness. CWN4 stand for CWN with moderate leverage and low skewness. CWN5 stand for CWN with moderate leverage and moderate skewness. CWN6 stand for CWN with moderate leverage and high skewness. CWN7 stand for CWN with high leverage and low skewness. CWN8 stand for CWN with high leverage and moderate skewness. CWN9 stand for CWN with high leverage and high skewness.

CONCLUSION

The GARCH family models are the traditional ways of analyzing heteroscedasticity data, but it cannot model heteroscedasticity data that include leverage effect efficiently. This reveals that the combine white noise model is a flexible tool with assurance that improves the outcome of heteroscedasticity data, irrespective of the volatility clustering that includes the leverage effect. The combine white noise estimation outperforms the EGARCH estimation when the heteroscedasticity data have no leverage effect. The combine white noise data are obtained from the decomposition of standardized residual of EGARCH data.

The validation of the performance of combine white noise model with simulation is carried out with two different sample sizes in connection with the low, moderate and high leverage, and skewness in ordered form. Combine white noise performs well in validation process. The model fit best in transformed 200 sample sizes of EGARCH generated data with moderate leverage and moderate skewness. While the best forecast is in transformed 200 sample sizes of EGARCH generated data with high leverage and moderate skewness. The results show that 200 sample sizes of EGARCH generated data with right values of leverage and skewness are better than 300 sample sizes for a reliable output.

ACKNOWLEDGEMENT

The authors thank the Universiti Utara Malaysia for the financial support in carrying out this research. The authors thank the reviewers that take their time to perfect this article.

REFERENCES

[1] Nelson, D. B., 1991, "Conditional heteroscedasticity in asset returns: A new approach," Econometrica: Journal of the Econometric Society, pp. 347-370.

[2] Hentschel, L., 1995, "All in the family nesting symmetric and asymmetric GARCH models," Journal of Financial Economics, *39*(1), pp. 71-104.

[3] McAleer, M., 2014, "Asymmetry and Leverage in Conditional Volatility Models,"Econometrics, 2(3), pp. 145-150, doi: 10.3390/econometrics2030145.

[4] McAleer, M., and Hafner, C. M., 2014, "A one line derivation of EGARCH," Econometrics, 2(2), pp. 9297, doi: 10.3390/econometrics2020092.

[5] Moorthy, R., H. K. Sum, and Benny, G., 2015, "Power Asymmetry and Nuclear Option in India-Pakistan Security Relations," Asian J. Sci. Res., pp. 8:80-94, DOI:10.3923/ajsr.

[6] Harvey, A. C., 1993, Time series models 2nd edition, The MIT Press, Cambridge, Massachusetts, Harvester Wheatsheaf Publisher.
[7] Sims, C. A., 1980, "Macroeconomics and Reality," Econometrica," *48*(1), pp. 1–48
[8] White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," Econometrica, 48(4), pp. 817-838
[9] Engle, R. F., 1982, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," Econometrica: Journal of the Econometric Society, pp. 987-1007.
[10] Bollerslev, T., 1986, "Generalized autoregressive conditional heteroscedasticity," Journal of econometrics, *31*(3), pp. 307-327
[11] Vivian, A., and Wohar, M. E., 2012, "Commodity volatility breaks," Journal of International Financial Markets, Institutions and Money, *22*(2), pp. 395-422.
[12] Ewing, B. T., and Malik, F., 2013, "Volatility transmission between gold and oil futures under structural breaks," International Review of Economics & Finance, *25*, pp. 113-121, journal homepage: www.elsevier.com/locate/iref.
[13] Ng, H. S., and Lam, K. P., 2006, "How Does the Sample Size Affect GARCH Model?"
[14] Sucarrat, G., 2013, "betategarch: Simulation, estimation and forecasting of Beta-Skew-t-EGARCH Models," The R Journal, 5(2), pp. 137-147.
[15] Lazim, M. A., 2013, Introductory business forecasting A practical approach 3rd edition, Penerbit Press, University Technology Mara, Printed in Kuala Lumpur, Malaysia.