IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

MODELING THE ERROR TERM OF REGRESSION BY COMBINE WHITE NOISE

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

[1]School of Quantitative Sciences, College of Arts and Sciences, 06010UUM Sintok, Universiti Utara (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

ABSTRACT

This paper examines the utilization of combination model technique to model the standardized residual exponential generalized autoregressive conditional heteroscedastic (EGARCH) errors. The technique combine white noise (CWN) is found to be more efficient and overcome EGARCH weaknesses. The estimation results using Combine White Noise model satisfies stability condition and passes stationary, serial correlation, and the ARCH effect tests. It fails the histogram-Normality tests but passes the Levene's test of equal variances. Combine White Noise has minimum values of information criteria. From the results of the dynamic evaluation forecast errors, Combine White Noise has the minimum forecast errors which are indications of better results when compare with the EGARCH model dynamic evaluation forecast errors. Combine White Noise processes show the best fit with forecast accuracy.

Keywords: Combine white noise, More efficient, Minimum information criteria, Minimum forecast error, Best fit, Forecast accuracy

I. INTRODUCTION

This paper presents an approach to model the standardized residual GARCH errors with combination model procedures to overcome of the weaknesses in GARCH error term for more efficient results. This will meet the needs of the economy, for the policy makers to have reliable planning and accurate forecasting [1, 2, 3, 4, 5, 6]. Researchers have been so much worried about the error term that conceal some vital information, which supposed to have been modeled, of which may make the model ineffective and inefficient in a way, if it is not correctly griped .

This error term is an unobservable random variable in the empirical model [7]. The error term is made up of mostly missing variables, error in variables and simultaneous causality, and this error term components have made it difficult to find an accurate model at a given time. Researchers have been developing several models at different

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

time to overcome this challenge, which actually have efficient estimation for some time. The large data size and high data frequency determine the suitable model that will produce a better result at a particular time, and when the data size increases then the model will not produce a good result.

The belief of the econometricians are, whatever is left out from the theoretical regression models is complemented by the error terms of the estimated models. The theories provide incomplete explanations of economic systems. Accordingly, the econometrics extensive tradition has seen the dynamic evolution of the economy as a controlling force, having relationship directly with the theory [7].

The conviction that errors exclusively signify random shocks that is responsible for the production of business cycles, but fails to recognize that the properties of regression residuals are generated by the empirical model, sample data, and estimation process. Alternatively, in relation to any economic theory, the "innovational residuals" model design criterion can result in errors that cannot be interpreted. In the history of time series econometrics, the econometricians acknowledge these clarifications and incorporate the interpretations into a frame work that gives a sensible consistent manner of approach to handle the errors [7].

[8] argue that insufficient information about data collection procedure and the proxy of latent variables in a regression model can create measurement errors. Measurement errors are inevitable when the number of parameters in a model cannot take care of the number of unknowns in the model.

[9] recommends that reporting a separated variant of the data is required to a regression model in which the data gathering organisation observes data with serially associated errors. Sargent demonstrates that more steps are required to estimate the parameters, when the information is not error free.

The above are the ways the stochastic models integrate error terms. The error term can also be assumed as white noise. White noise series is when the time series is stationary, having a sequence of unrelated variables with constant mean and constant variance. It is also called innovations or shocks in economic time series. The White noise series being modeled are referring to as combine white noise model.

Therefore, the Combine White Noise uplifted the EGARCH models weaknesses to model the error terms for appropriate estimation and has reasonable outputs.

II. METHODOLOGY

Considering the autoregression model:

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

Allow stochastic procedure of a real-valued time to be ε_t, and the entire information through t time is I. The GARCH model is:

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

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

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

$$= \omega + A(L)\varepsilon_t^2 + B(L)h_t \tag{2.3}$$

The GARCH models which have been exhibiting unequal variances (heteroscedastic errors) behaviors in the process of estimation, its error term can be simplified. The standardized residuals of GARCH errors which are unequal variances are decomposed into equal variances (white noise) in series to deal with the heteroscedasticity. Each equal variances series will be modeled using regression model.

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

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

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

$$\vdots \qquad .$$

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

$$Y_{jt} = \sum_{j=1}^{q}\theta_j \varepsilon_{j,t-q} + \sum_{j=1}^{q}\Phi_j \varepsilon_{j,t-q} + \ldots \sum_{j=1}^{q}\phi_j \varepsilon_{j,t-q} \tag{2.4}$$

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

$$= \varepsilon_t[A(L) + B(L) + \ldots] \tag{2.5}$$

$$= Q\varepsilon_t$$

$$= U_t \tag{2.6}$$

It can be written as:

$$Y_t = U_t\,, \qquad U_t \sim \mathrm{N}(0, \sigma_c^2) \tag{2.7}$$

Equation (2.6) can be inverted if the coefficient of absolute values of error term is less than one:

$$Q^{-1}Y_{jt} = \varepsilon_t \qquad \textbf{for} \quad |Q^{-1}| < 1$$

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

where A(L) + B(L) + … = Q which are the matrix polynomial, $Q\varepsilon_t$ = U_t which 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 are:

$$\sigma_c^2 = \sigma_1^2 + \sigma_2^2 + \ldots \tag{2.9}$$

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

$$\sigma_c^2 = \sigma_1^2 + \sigma_2^2 \tag{3.0}$$

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

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

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:

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

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

III. DATA APPLICATION

The France Gross Domestic Product (GDP) quarterly data from 1960Q3 to 2015Q2 is used for this study. It is retrieved from the DataStream of Universiti Utara Malaysia library. The data show an upward trend which is a behavior of non-stationary.

The data is transformed in returns series to observe the volatility clustering, long tail skewness and excess kurtosis which are the characteristics of heteroscedasticity. The graph exhibits irregular variances that indicate volatility.

The table 1 reports that, there are right tail skewness, excess kurtosis, and Jarque-Bera test is highly significant, that is an indication of non-normality. Standard deviation is less than one.

The table 1 above shows the ARCH LM tests for the effect of heteroscedasticity in the data series; F-Statistic and Obs*R-squared are highly significant which indications ARCH presence in the data.

Table 1: Histogram-Normality and ARCH Tests

Normal test	Coefficient/ value	probability
Standard deviation	0.920518	
Skewness	0.763937	
Kurtosis	22.97828	
Jarque-Bera	3663.369	0.0000
ARCH Tests		
F-Statistic		0.0000
Obs*R-squared		0.0000

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The table 2 below shows that the AIC, BIC and HQ minimum information criteria with log-likelihood that are used to select the appropriate model between ARCH and GARCH estimation. EGARCH model is choosing because it has minimum values of AIC, BIC and HQ with high log-likelihood values.

Combine White Noise (CWN) estimation has the minimum information criteria with high log likelihood. The CWN estimation gives better results with minimum information criteria and high log likelihood when compared with the GARCH estimation.

Table 2 .France data ARCH, EGARCH and CWN models coefficients, information criteria and log likelihood values

	α	β	δ	γ	AIC	BIC	LL
ARCH	0.4745 (0.0003)	0.7089 (0.0000)			1.9446	2.0067	-207.9622
EGARCH	0.0.5839 (0.0000)	0.4966 (0.0018)	-0.1223 (0.1715)	0.6544 (0.0000)	1.7018	1.8104	-178.4925
CWN					-4.6571	-4.564	515.95

Note: α is the coefficient of the mean equation, β and δ are the coefficients of the variance equations, while γ is the coefficient of the log of variance equation. In the parentheses are the probability values (PV).

In GARCH modeling, the leverage is not possible because, any restriction imposed is positivity restriction which has no leverage effect [5, 6]. To avoid the above challenges of leverage effect in EGARCH, the standardized residuals graph of the GARCH model (GARCH errors) with unequal variances and zero mean are decomposed into equal variances series (white noise series). The graphs of equal variances with mean zero being obtained from graphs of GARCH errors are white noise series. These white noise series are fit into regression model to make each a model.

The implementation of Bayesian model averaging produces two best models [10]. For confirmations, fitting linear regression with autoregressive errors; 220 is the number of observation, with zero mean and variance one [11].Therefore, the best two models are white noise models.

Table 3 shows that independent samples test for testing whether data set of the two white noise models have equal variances or not. The test in Table 3 reveals that the variability in the distribution of the data is no significantly different value which is greater than the p-value 0.05. Thus, the model had equal variances [12].

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

Table 3: Levene's test for equal variances, Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
B Equal variances assumed	1.414	.235	2.159	438	.031	.05909	.02737	.00530	.11288
Equal variances not assumed			2.159	255.236	.032	.05909	.02737	.00519	.11299

Table 4 reveals that CWN for France data estimation appeared to be more appropriate model for estimation and forecasting in comparison with EGARCH models. In France CWN have minimum values of root mean square error (RMSE) and mean absolute error (MAE) with high mean absolute percentage error (MAPE) when compared with EGARCH forecast errors.

Table 4 The summary of CWN and EGARCH models estimation and forecasting evaluation for France data set

	CWN	EGARCH
Stationary	Stationary	Stationary
Log Likelihood	515.95	-178.49
AIC	-4.6571	1.7018
BIC	-4.5642	1.8104
Correlogram Std Resid Squared/Lag Structure	Stable	Model Specified Correctly
Histogram-Normality Tests	Not Normal	Not Normal
ARCH Test	No ARCH effect	No ARCH effect
RMSE	0.0532	0.6684
MAE	0.0145	100.07
MAPE	1.8169	0.3192

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

GRMSE	0.0021	0.3192
Correlogram Std Resid Squared/Lag Structure	Stable	Model Specified Correctly
Histogram-Normality Tests	Not Normal	Not Normal
Heteroscedasticity Test	No ARCH effect	No ARCH effect

IV. CONCLUSION

The CWN estimation reveals that minimum information criteria and high log likelihood values in France data estimation are better than EGARCH information criteria and log likelihood values. The EGARCH estimation reveals that the data set contain leverage effect. CWN have the minimum forecast errors which are indications of better results when compared with the EGARCH model dynamic evaluation forecast errors in France data set forecast evaluation [13, 14]

Based on every result in the empirical analysis of the France GDP data set, CWN is the more appropriate model. For this reason, CWN is recommended for modeling the leverage effect in the data that exhibits heteroscedasticity.

REFERENCES

[1] H White, A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity, *Econometrica, 48(4),* 1980, 817-838

[2] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica, Journal Econometric Society, 50(4)* , 1982, 987-1007.

[3] T Bollerslev, Generalized autoregressive conditional heteroscedasticity, *Journal of econometrics, 31 (3),*1986, 307-327.

[4] D. B. Nelson, Conditional heteroscedasticity in asset returns: A new approach, *Econometrica: Journal of Econometric Society, 59(2),*1991, 347-370.

[5] M McAleer, Asymmetry and Leverage in Conditional Volatility Models, *Econometrics, 2 (3),* 2014, 145-150

[6] G. G. Martinet, and M. McAleer, On the Invertibility of EGARCH (p, q)*, *Econometric Reviews,* Published online: 08 Apr 2016

[7] D Qin, and C. L. Gilbert, The error term in the history of time series econometrics, *Econometric theory, 17(2*), 2001, 424-450

[8] I Komunjer, and S. Ng, Measurement errors in dynamic models. *Econometric Theory, 30(01),* 2014, 150-175

[9] T. J. Sargent, Two models of measurements and the investment accelerator. *The Journal of Political Economy, 97(2*), 1989. 251-287

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

[10] Z Asatryan, and L. P. Feld, Revisiting the link between growth and federalism, A Bayesian model averaging approach, *Journal of Comparative Economics, 43(3),* 2015, 772-781.

[11] M. L. Higgins, and A. K. Bera, A class of nonlinear ARCH models, *International Econometrics Review, 33(1),* 1992, 137-158

[12] D. D. Boos, and C. Brownie, Comparing variances and other measures of dispersion, *Statistical Science, 19(4),* 2004, 571-578

[13] S Ismail, and T. Z. Muda, *Comparing forecasting effectiveness through air travel data,* In Proceedings of Knowledge Management International Conference & Exhibition (KMICE) , 6 - 8 June 2006 Legend Hotel Kuala Lumpur. Faculty of Information Technology, Universiti Utara Malaysia, Sintok, 2006, 594-602

[14] R. Fildes, Y. Wei, and S. Ismail, Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures, *International Journal for Forecasting, 27(3),* 2011, 902-922