International Journal of Modern Mathematical Sciences, 2019, 17(1): 49-56

Modern Scientific Press | *International Journal of Modern Mathematical Sciences* | ISSN: 2166-286X

Journal homepage: www.ModernScientificPress.com/Journals/ijmms.aspx | *Florida, USA*

Article

Performance of Subset Autoregressive Integrated Moving Average Polynomial Distributed Lag Model

[*]Ayoola F. J. and Ojo J. F.

Department of Statistics, University of Ibadan, Ibadan, Nigeria

*Author to whom correspondence should be addressed; Email: fj.ayoola@ui.edu.ng, jfunminiyiojo@yahoo.co.uk, ayoolafemi@yahoo.com

Article history: Received 1 February2019, Revised 14 March 2019, Accepted 15 March 2019, Published 5 April 2019.

Abstract: This study considered three types of distributed lag models namely: Autoregressive Integrated Polynomial Distributed Lag (ARIPDL), Autoregressive Integrated Moving Average Polynomial Distributed Lag (ARIMAPDL) and Subset Autoregressive Integrated Moving Average Polynomial Distributed Lag (SARIMAPDL) models. Less attention have been given to these models in literature, hence the reason for this study. These models were further compared with the existing polynomial distributed lag models taking into consideration the stationary of the endogenous variable. The estimation technique was illustrated with respect to two time series. The optimal model were fitted using appropriate criteria and the forecast attached to these models were evaluated using appropriate measures. Among the three models, SARIMAPDL performed best when we studied the residual variance and SARIMAPDL outperformed ARIPDL and ARIMAPDL when the forecast performance were studied. In all, SARIMAPDL outperformed all the models studied in this work. We recommend SARIMAPDL especially when we are focusing on the stationary of the endogenous variable.

Keywords: Subset ARIMAPDL model, ARIMAPDL model, nominal gross domestic product, forecast.

---

1. Introduction

When a time series is stationary, the mean, variance and covariance of the series are all constant over time (Durkar and Pastorekova, 2012). When there is the need to fit a model to such series, having studied the autocorrelation and partial autocorrelation function, the recommended models to be fitted are autoregressive (AR), moving average (MA) or autoregressive moving average (ARMA). These models have been extensively studied by various authors, among whom are (Walker, 1952; Chatfield, 1980)
In a situation whereby the time series under study is not stationary, the series are always integrated by differencing once or more to achieve stationarity. The model that can be fitted to such series are autoregressive integrated (ARI), integrated moving average (IMA) or autoregressive integrated moving average (ARIMA). These models have been extensively studied by various authors, among whom are (Box and Jenkins, 1970; Anderson, 1971, 1977). Achieving stationarity for the models that are fitted prevent the system from explosion.

The reasons for lag in a model could be due to psychological, technological, institutional, political, business and economic decisions Ojo (2013). Due to this underlining fact, distributed lag models have been applied in various fields in the past few decades and a remarkable success in its application have been seen which help in the diverse areas of the economy (Kocky (1954); Almon (1965); Zvi (1961); Robert and Richard (1968); Frank (1972); Dwight (1971); Krinsten (1981); Wilfried (1991)).

In autoregressive distributed lag model, the regressors may include lagged values of the dependent variable and current and lagged values of one or more explanatory variables. This model allows us to determine what the effects are of a change in a policy variable, Chen (2010). It is imperative to see that adding an instrumental variable such as Moving Average (MA) to Autoregressive Polynomial Distributed Lag (ARPDL) model there is the likelihood of having a better model (Ojo and Ayiebutaju, 2015).

When a model that has taken care of stationarity form a part of any particular model such as distributed lag model, the result from such models will be a dependable one. Therefore, this study shall focus on autoregressive integrated moving average polynomial distributed lag (ARIMAPDL) model. In addition to ARIMAPDL we shall consider autoregressive integrated polynomial distributed lag (ARIPDL) and subset autoregressive integrated moving average polynomial distributed lag (SARIMAPDL) models.

The subset approach will remove redundant parameters from ARIMAPDL. Subset approach and algorithm for fitting subset models have been extensively studied by various authors among whom are (Haggan and Oyetunji 1980, Ojo 2007, 2008, 2013, Ojo and Ayoola 2013). The three distributed lag models shall be compared with the existing distributed lag models.

The traditional ARDL and the ARDL approach to cointegration for the analysis of short-run dynamic and long run relationship when series are difference stationary (series can be integrated of different orders). The two models were used to estimate the short-run dynamics and the long run relationships between selected Nigeria's macroeconomic series. The results compares favorably with the theory that the ARDL is equivalent to the short-run dynamics of the error correction model (the resultant model from the ARDL approach to cointegration).(Shittu, Yemitan and Yaya, 2012).

Economic analysis suggests that there is a long run relationship between variables under consideration as stipulated by theory. This means that the long run relationship properties are intact. In other words, the means and variances are constant and not depending on time. However, most empirical researches have shown that the constancy of the means and variances are not satisfied in analyzing time series variables. In the event of resolving this problem most cointegration techniques are wrongly applied, estimated, and interpreted. One of these techniques is the Autoregressive Distributed Lag (ARDL) cointegration technique or bound cointegration technique. Hence, this study reviews the issues surrounding the way cointegration techniques are applied, estimated and interpreted within the context of ARDL cointegration framework (Nkoro and Uko, 2016).

2. Materials and Methods

2.1. Time Series Model

Time series models use the past movements of variables in order to predict their future values. The time series model proposed by Box-Jenkins has been widely used in literature because of its performance and simplicity. Most time series can be described by Autoregressive Moving Average (ARMA) model. If the series is difference-stationary, the integrated autoregressive moving average (ARIMA) model is implemented Jing *et al.* (2009).

2.2. Autoregressive Integrated Moving Average Model

In statistics, an autoregressive integrated moving average (ARIMA) model is a generalisation of an autoregressive moving average or (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series. The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively.

Given a time series of data X_t where t is an integer index and the X_t are real numbers, then an ARIMA(p,d, q) model is given by

$$Y_t - \psi_1 Y_{t-1} - \ldots\ldots - \psi_{p+d} Y_{t-p-d} = e_t + \theta_1 e_{t-1} + \ldots + \theta_q e_{t-q} \quad (2.1)$$

ψ_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and e_t are error terms. The error terms e_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean. The specification and estimation technique of equation (2.1) can be found in (Ojo, 2008).

In order to fit ARIMAPDL and SARIMAPDL models in this study, we first fit ARIMA and SARIMA models. The parameters of these models were used as the initial parameters in the estimation technique. Augmented Dickey-Fuller test (Dickey and Fuller, 1979) was employed to our Nigeria nominal GDP series, the dependent variable (Y_t). It was found out that the series has a unit root. Then, for ARIPDL, ARIMAPDL and SARIMAPDL models, we fixed the order of integration (d) at one.

2.3. Autoregressive Integrated Moving Average Polynomial Distributed Lag Model

An autoregressive integrated moving average polynomial distributed lag (ARPDL) model is one that contains lagged y_t's , x_t's and e_t's and is defined as follows:

$$Y_t - \psi_1 Y_{t-1} - ......- \psi_{p+d} Y_{t-p-d} = \theta_1 e_{t-1} + ...+ \theta_q e_{t-q} + \beta_0 x_t + \beta_1 x_{t-1} + ...+ \beta_j x_{t-j} + \upsilon_t$$

where ψ_i are the parameters of the integrated autoregressive part of the model, the θ_i are the parameters of the moving average part and e_t are error terms. and p and q are the lag length of the autoregressive component, $\beta_0, ..., \beta_j$ are the parameters of the polynomial distributed lag component, q is the lag length of the polynomial distributed lag component.

The subset approach and algorithm for fitting the best subset model have been discussed extensively in literature. See Hocking and Leslie (1967) Furnival (1971), Mallows (1973), Haggan and Oyetunji (1980), Ojo (2007, 2008 and 2013).

Estimation of parameters of Autoregressive Integrated Moving Average Polynomial Distributed Lag Model

$$Y_t = \psi_1 Y_{t-1} + ...+ \psi_{p+d} Y_{t-p-d} + e_t - \theta_1 e_{t-1} - ...- \theta_q e_{t-q} + \beta_0 x_t + \beta_1 x_{t-1} + ...+ \beta_j x_{t-j} + \upsilon_t$$

We consider Newton Raphson iterative method to estimate the parameters of the model. Representing the mean response as f_i the error term becomes $\upsilon_t = Y_t - f_i$

The least square estimator of $\hat{G}$ of G which minimizes the sum of the square of residual is

$$S(G) = \sum_{t=1}^{n} (v_t)^2 \ .$$

We differentiate S(G) with respect to the parameter $G(\psi_1, \psi_2, ..\psi_{p+d}, \theta_1, ..., \theta_q, \beta_0, \beta_1, ..., \beta_j)$. See Ojo and Ayiebutaju (2015), the approach are similar. For the selection of optimal lag for our models Akaike Information and Bayesian Information criteria were employed. The performance indicator for the forecast can be found in Ojo and Rufai (2016)

3. Results and Discussion

3.1. Numerical Example

To present the application of these models we will use a real time series dataset, monthly nominal gross domestic product and exchange rate series between 2002 and 2016 obtained from central bank of Nigeria. Nominal gross domestic product series is the endogenous variable (Y_t) while exchange rate represent the lag variables.

3.2. Fitted Optimal Autoregressive Integrated Moving Average Polynomial Distributed Lag (ARIMAPDL) Model

$$\hat{Y}_t = 0.903164Y_{t-1} - 0.002914Y_{t-2} - 0.588295Y_{t-3} + 0.438590Y_{t-4} - 0.215136e_{t-1}$$
$$+ 0.11334.85X_t - 3671.660X_{t-1}$$

3.3. Fitted Optimal Subset Autoregressive Integrated Moving Average Polynomial Distributed Lag (SARIMAPDL) Model

$$\hat{Y}_t = 0.902865Y_{t-1} - 0.588458Y_{t-3} + 0.438130Y_{t-4} - 0.212208e_{t-1} + 11368.68X_t$$
$$- 3771.795X_{t-1}$$

3.4. Fitted Optimal Autoregressive Integrated Polynomial Distributed Lag (ARIPDL) Model

$$\hat{Y}_t = 0.759801Y_{t-1} + 11351.22X_t - 4075.575X_{t-1}$$

3.5. Fitted Optimal Autoregressive Polynomial Distributed Lag (ARPDL) Model

$$\hat{Y}_t = 1.009496Y_{t-1} + 1673.742X_t - 1654.871X_{t-1}$$

3.6. Fitted Optimal Polynomial Distributed Lag (PDL) Model

$$\hat{Y}_t = 40798.96X_t - 11441.58X_{t-1}$$

3.7. Model Performance of ARIMAPDL and SARIMAPDL

Table 2.1: Performance Indicators of ARIMAPDL and SARIMAPDL Models

Model	R^2	$\bar{R}^2$	AIC	BIC	Residual variance
ARIMAPDL	0.966340	0.966146	28.89404	28.93007	$(448658.6)^2$
SARIMAPDL	0.966893	0.966702	28.87748	28.91351	$(444955.8)^2$
ARIPDL	0.968753	0.968577	28.83450	28.87011	$(435422)^2$
***ARPDL**	0.999540	0.999538	24.61581	24.651430	$(53116.54)^2$
PDL	0.470653	0.467663	31.69985	31.66423	$(1792077)^2$

The performance of the five distributed lag models were shown in table 2.1. The $\bar{R}^2$ showed the contribution of explanatory variables, lagged values of endogenous variable inclusive in those models except polynomial distributed lag model. Though autoregressive polynomial distributed lag model has the minimum variance but because of the non-stationary of the endogenous variable the root mean square forecast error in table 2.2 for ARPDL was the highest. This is an explosion. The result from ARPDL proved the focus of this study that is fitting a stationary model in addition to the lagged exogenous variable. The criteria for the order determination namely Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC) were given also in table (2.1). From these performance indicator at model level we could see that the values for SARIMAPDL was smaller than the other models since ARPDL could not capture stationary.

3.8. Forecast Performance of ARIMAPDL and SARIMAPDL

Table 2.2: Performance Indicators of ARIMAPDL and SARIMAPDL Models at the Level of Forecast

Model	RMSFE	MAFE	MAPFE
ARIMAPDL	1736908	1515932	64.29787
SARIMAPDL	1735623	1514953	64.23443
ARIPDL	1764508	1538307	67.55238
ARPDL	1802775	1487810	30.10018
PDL	1797674	1552013	70.37826

From table 2.2, we could see the performance of the five distributed lag models. The performance measures are Root Mean Square Forecast Error (RMSFE), Mean Absolute Forecast Error (MAFE) and Mean Absolute Percentage Forecast Error (MAPFE). From these performance indicators at forecast level we could see that the values for SARIMAPDL was smaller than the other models.

4. Conclusion

In this study, three types of distributed lag models were considered namely: Autoregressive Integrated Polynomial Distributed Lag (ARIPDL), Autoregressive Integrated Moving Average Polynomial Distributed Lag (ARIMAPDL) and Subset Autoregressive Integrated Moving Average Polynomial Distributed Lag (SARIMAPDL) models. These models were further compared with the existing polynomial distributed lag models taking into consideration the stationary of the endogenous variable. The estimation technique was illustrated with respect to two time series. The optimal model were fitted using appropriate criteria and the forecast attached to these models were evaluated using appropriate measures. Among the three models, SARIMAPDL performed best when we studied the residual variances and SARIMAPDL outperformed ARIPDL and ARIMAPDL when the forecast performance were studied. In all, SARIMAPDL outperformed all the models studied in this work. We recommend SARIMAPDL especially when we are focusing on the stationary of the endogenous variable.

Reference

[1] Almon S. The Distributed lag between capital appropriation and expenditures. *Econometrics,* 30(1965): 407-423.

[2] Anderson, T. W. The Statistical Analysis of Time Series. *New York and London Wiley*. (1971)

[3] Anderson, T. W. Estimation for Autoregressive Moving Average Models in the Time and Frequency Domains. *The Ann. of Statis.* 5(5)(1977): 842-865

[4] Box, G. E. P. and Jenkins, G. M. Time Series, Forecasting and Control. *Holden-Day: San Francisco*. (1970)

[5] Chatfield, C. The Analysis of Time Series- An Introduction. *Chapman and Hall*. (1980)

[6] Chaleampong Kongcharoen and Tapanee Kruangpradit Autoregressive Integrated Moving Average with Explanatory variable (ARIMAX) Model for Thailand Export. *Presented at the 33rd International Symposium on forecasting, South Korea* (2013): 1-8

[7] Chen Yi-Yi, Autoregressive Distributed Lag (ADL) Model (2010). *http://mail.tku.edu.tw/chenyiyi/ADL.pdf*

[8] Koyck L.M. Distributed lag and investment analysis. *Amsterdam, North Holland Publishing Company*. (1954).

[9] Durka P. and Pastorekova S. ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series? *Proceedings of 30th International conference mathematical methods in Economics*. (2012): 136-140.

[10] Furnival, G. M. All possible Regressions with Less Computation. *Technometrics*, 13(1971): 403-408.

[11] Haggan, V. and Oyetunji, O. B. On the Selection of Subset Autoregressive Time Series Models. *UMIST Technical Report*, 124(1980).

[12] Hocking, R. R. and Leslie R. N. Selection of the Best Subset in Regression Analysis. *Technometrics*, 9(1967): 531-540

[13] Jing Fan, Rui Shan, Xiaoqin Cao and Peiliang Li. The Analysis to Tertiary-industry with ARIMAX Model. *Journal of Mathematics Research*. 1(2)(2009):156-162

[14] Nkoro, E. and Uko, A. K. Autoregressive Distributed Lag (ARDL) Cointegration Technique: Application And Interpretation. *Journal of Statistical and Econometric Methods*. 5(4)(2016): 63-91. doi: 10.1002/jae.616.

[15] Mallows, C. L. Choosing Subset Regression. *Technometrics*, 17(1973): 213-219.

[16] Ojo, J. F. and Ayoola, F. J. Listing of all Subsets and Selection of Best Subset in Finite Distributed Lag Models using Meteorological Data. *Continental Journal of Applied Sciences*, 8(2)(2013): 26-33.

[17] Ojo, J. F. Identification of Optimal models in Higher Order Integrated Autoregressive Models and Autoregressive Integrated Moving Average Models in the Presence of 2^k-1 Subsets. *Journal of Modern Mathematics and Statistics*, 2(1)(2008): 7-11.

[18] Ojo, J. F. On the Estimation and Performance of Subset Autoregressive Moving Average Models. *European Journal of Scientific Research.* 18(4)(2007): 700-706.

[19] Ojo J. F. and Aiyebutaju M. O. On the Performance of Autoregressive Moving Average Polynomial Distributed Lag Model. *Ife Journal of Science.* 17(2)(2015): 247-254.

[20] Ojo J.F On the performance of Subset Autoregressive Polynomial Distributed Lag Model. American *Journal of Scientific and Industrial Research*, 4(4)(2013): 399-403.

[21] Ojo, J. F. and Rufai O. On Subset Autoregressive Fractionally Integrated Moving Average Models. *International Journal of Physical Science*, 11(1)(2016): 15-20.

[22] Shittu, O. I., Yemitan, R. A. and Yaya, O. S. On Autoregressive Distributed Lag, Cointegration and Error Correction Model. [An Application to Some Nigeria Macroeconomic Variables]', *Australian Journal of Business and Management Research*, 2(8)(2012): 56-62.

[23] Walker, A. M. Some properties of Asymptotic Power Functions of Goodness of fit Tests for Linear Autoregressive Schemes. *J. Roy. Stat. Soc.*, 14(1952): 117-134