International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 18 No. 4 Dec. 2016, pp. 990-996

http://www.ijias.issr-journals.org/

# Modeling Regression with Time Series Errors of Gross Domestic Product on Government Expenditure

***Emmanuel Alphonsus Akpan*[1]*, Imoh Udo Moffat*[2]*, and Ntiedo Bassey Ekpo*[3]**

[1]Department of Mathematics and Statistics, University of Uyo, Nigeria

[2]Department of Mathematics and Statistics, University of Uyo, Nigeria

[3]Department of Banking and Finance, University of Uyo, Nigeria



**ABSTRACT:** The study examined the relationship between Gross Domestic Product (GDP) and Government Expenditure between 1981 and 2012. The motivation was, in fitting regression model to time series data, autocorrelation in the error terms should be expected. Utilizing data from the Central Bank of Nigeria Statistical Bulletin, we found that regression model could capture the linear relationship between the dependent variable (GDP) and the independent (Government Expenditure). However, the error terms of the regression model were found to be autocorrelated and could be corrected by ARIMA(1,0,1) model. Moreover, regression model with an ARIMA(1,0,1) error was able to capture the linear relationship between GDP and the Government Expenditure alongside the autocorrelated errors. Evidence from the model revealed that Gross Domestic Product is a linear function of Government Expenditure at present and immediate previous year. The policy implication of this study is that if Government Expenditure is kept constant from immediate previous year to the present year, then, the GDP would tend to decrease, as such; Government should vary its expenditures in order to improve the GDP.

**KEYWORDS:** ARIMA model; autocorrelated errors; Government Expenditure; Gross Domestic Product; regression model.

## 1 INTRODUCTION

When regression is applied to time series data, the error terms are often autocorrelated (Wei, 2006). According to Box, Jenkins and Reinsel (2008), when fitting a regression model to time series data, one should always consider the possibility of autocorrelation in the error terms. If the error terms are found to be autocorrelated, then, the assumption that the errors of the observations from regression model are independent and identically distributed is clearly violated. The implication would be that ordinary least squares would no longer be considered the best in computing coefficients as it would tend to ignore time-relationship in data. Also, standard errors of coefficients would be incorrect, most likely too small, consequently, the validity and predictability potential of test results would be doubtful for decision making purposes; and even the information criteria of the fitted models would no longer be good guides as to which is the best model for forecasting. In most cases, the p-value associated with the coefficients will be too small, and so some predictor variables would appear even significant when they are not (Rawlins, Pantula and Dickey, 1998).

Brockwell and Davis (2002) argued that it is more appropriate to assume that the errors are observations of stationary process. However, if the errors are autocorrelated, Autoregressive Integrated Moving Average (ARIMA) model can be used to model the information they contain. The resulting model is the combination of regression model and an ARIMA model in the error terms. This combined model would enhance the possibility of obtaining more reliable estimates for the effect of the independent variables on the dependent variable. Thus, when analyzing management related time series data, prior studies preferred using regression models which are deficient in extracting the embedded information in the error terms thereby undermining the obvious advantages of time series models (ARIMA models). Therefore, this study contributes towards bridging the gap by combining both regression and ARIMA models in analyzing the relationship between Gross Domestic Product (GDP) and Government Expenditure (GEXP). That is to say, the combined model would be used to capture both the variation in GDP (explained by the regression model) and the unexplained variation (which is the extra information embedded in the error terms) in GDP (explained by the time series model).

## 2 METHODOLOGY

### REGRESSION MODEL

Rawlings, Pantula and Dickey (1998) defined a standard regression model as

$$Y_t = \beta_0 + \beta_{1,}X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \varepsilon_t \quad (2.1)$$

where $Y_t$ = dependent variable

$\beta_i$ = regression parameters, i = 1,…, k

$X_{it}$ = independent variables, i = 1,…, k

$\varepsilon_t$ = error term assumed to be i.i.d. N(0, $\sigma_t^2$)

Thus, the dependent variable for a time series regression model with independent variables is a linear combination of independent variables measured in the same time frame as the dependent variable. Estimates of the parameters of the model in (2.1) can be obtained by least squares estimation method. See Drasper and Smith (1998), Rawlings, Pantula and Dickey (1998) for more details on least squares estimation method.

### AUTOREGRESSIVE MOVING AVERAGE (ARMA) PROCESS

A natural extension of pure autoregressive and pure moving average processes is the mixed autoregressive moving average ($ARMA$) processes, which includes the autoregressive and moving average as special cases (Wei, 2006).

A stochastic process $\{X_t\}$ is an $ARMA(p, q)$ process if $\{X_t\}$ is stationary and if for every $t$,

$$\varphi(B)X_t = \theta(B)a_t \quad (2.2)$$

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ is the autoregressive coefficient polynomial.

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the moving average coefficient polynomial.

### AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL

Box, Jenkins and Reinsel (2008) considered the extension of ARMA model in (2.2) to deal with homogenous non-stationary time series in which $X_t$, itself is non-stationary but its $d^{th}$ difference is a stationary ARMA model. Denoting the $d^{th}$ difference of $X_t$ by

$$\varphi(B) = \phi(B)\nabla^d X_t = \theta(B)a_t \quad (2.3)$$

where $\varphi(B)$ is the nonstationary autoregressive operator such that d of the roots of $\varphi(B) = 0$ are unity and the remainder lie outside the unit circle. $\phi(B)$ is a stationary autoregressive operator.

Therefore, (2.3) is called an autoregressive integrated moving average model and can be referred to as an $ARIMA(p, d, q)$ model.

### REGRESSION MODEL WITH AN ARIMA ERROR

The idea of regression model with an ARIMA error structure is to refine the ordinary regression estimates in that ARIMA structure exists in the residuals. According to Box, Jenkins and Reinsel (2008), it is known that the sample autocorrelations (ACF) and Partial autocorrelations (PACF) calculated from the residuals of the preliminary least squares fit are asymptotically equivalent to those obtained from the actual noise series and as such they can be used to identify an appropriate ARIMA model for the error term (see also Fuller, 1996).

The complete model considered is

$$Y_t = \beta_0 + \beta_{1,}X_{1,t} + \beta_2 X_{2,t} + \cdots + \beta_k X_{k,t} + \varepsilon_t,\ \varphi(B)(1-B)^d \varepsilon_t = \theta(B)a_t \quad (2.4)$$

t = 1,…, n,

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ is the autoregressive coefficient polynomial.

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is the moving average coefficient polynomial.

Estimates of all parameters in the model (2.4), $\beta_0, \ldots, \beta_k, \varphi_1, \ldots, \varphi_p, \theta_1, .., \theta_q, \sigma_a^2$ , can be obtained by maximum likelihood estimation methods (see Brockwell and Davis, 2002; Box, Jenkins and Reinsel, 2008; Wei, 2006; and, Cryer and Chan, 2008).

#### MODEL SELECTION CRITERIA

For a given data set, when there are multiple adequate models, the selection criterion is normally based on summary statistics from residuals of a fitted model (Wei, 2006).

There are several model selection criteria based on residuals (see Wei, 2006). For the purpose of this study, we consider the well-known Akaike's information criterion (AIC), (Akaike, 1973) defined as

$$\text{AIC} = -2\ ln(\text{likelihood}) + 2(\text{number of parameters})$$

where the likelihood function is evaluated at the maximum likelihood estimates. The optimal order of the model is chosen by the value of the number of parameters, so that AIC is minimum (Wei, 2006).

#### MODEL DIAGNOSTIC CHECKING

Box and Pierce (1970) proposed the Portmanteau statistics:

$$Q^*(\text{m}) = \text{T} \sum_{l=1}^{m} \hat{\rho}_l^2 \quad (2.5)$$

where T is the number of observations.

Ljung and Box (1978) modify the $Q^*$(m) statistic to increase the power of the test in finite samples as follows:

$$\text{Q(m)} = \text{T(T + 2)} \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l} . \quad (2.6)$$

where T is the number of observations.

The decision rule is to reject $H_0$ if Q(m) > $\chi_\alpha^2$, where $\chi_\alpha^2$ denotes the 100 (1 – $\alpha$)th percentile of a Chi-squared distribution with m – (p + q) degree of freedom (see for example Akpan, Moffat and Ekpo, 2016). But in this paper, we are using Chi-squared distribution with m degree of freedom as provided by R package for the analysis. The decision rule can also reject $H_0$ if the p-value is less than or equal to $\alpha$, the significance level.

## 3 DATA ANALYSIS AND DISCUSSION

This study considers the Federal Government of Nigeria Expenditure (N' Billion) as the independent variable and the Gross Domestic Product by Expenditure (N' Billion) as the dependent variable. The data were obtained from the Central Bank of Nigeria Statistical Bulletin for a period spanning from 1981 to 2012. Each series consists of 32 observations.

#### REGRESSION MODEL

First, we fit a regression model of the Gross Domestic Product (GDP) on Government Expenditure (GEXP) with the trend component inclusive to avoid the need for differencing since it is not of our interest in this paper to address the effect nonstationary data can have on regression model.

The estimated regression model with trend component is presented in equation (3.1)

$$GDP_t = -2921.444 + 171.951 TREND + 27.790\ GEXP_t$$

| | | | | |
|---|---|---|---|---|
| s.e | (2349.442) | (222.461) | (6.001) | |
| t-value | $(-1.243)$ | (0.773) | (4.631) | (3.1) |
| p-value | (0.224) | (0.446) | $(7.07e^{-05})$ | |

[Excerpts from Table 1]

***Table 1. Output of Regression Model with Trend***

```
Call:
lm(formula = GDP ~ TREND + GEXP)

Residuals:
     Min       1Q   Median       3Q      Max
-10873.8  -2507.1    639.2   1852.9  14284.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2921.444   2349.442  -1.243    0.224
TREND         171.951    222.461   0.773    0.446
GEXP           27.790      6.001   4.631 7.07e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5143 on 29 degrees of freedom
Multiple R-squared:  0.8336,      Adjusted R-squared:  0.8221
F-statistic: 72.63 on 2 and 29 DF,  p-value: 5.096e-12
```

From the estimated model in equation (3.1), it is observed that the trend component is not significant since the p – value = 0.446 $>$ 0.05 level of significance, as such the trend component is dropped from the model. Although, the intercept is not also significant with p – value = 0.224 $>$ 0.05 level of significance, we still include it in the model since it is the value of the dependent variable if the coefficient of the independent variable appears to be zero. Therefore, equation (3.1) is refined and presented in equation (3.2),

$$GDP_t = -1372.321 + 31.950\, GEXP_t$$

| | | | |
|---|---|---|---|
| s.e | (1217.772) | (2.638) | |
| t-value | $(-1.127)$ | (12.109) | (3.2) |
| p-value | (0.269) | $(4.45e^{-13})$ | |

$R^2 = 0.8032$ [Excerpts from Table 2].

From the regression model in (3.2), it is observed that the inclusion of the independent variable (Government Expenditure) in the model is significant since the p – value = $4.45e^{-13} < 0.05$ level of significance, implying that there is very strong evidence to conclude that Government Expenditure has a significant contribution to Gross Domestic Product. The coefficient of determination ($R^2$) indicates that the Government Expenditure is able to explain about 80.32% of the total variation in Gross Domestic Product. Therefore, we can deduce that about 19.68% of the total variation embedded in the error terms could not be explained by the regression model and the 19.68% of the total variation is significant if the error term is found to be autocorrelated, hence the need for regression modeling with an ARIMA error.

***Table 2. Output of Regression Model***

```
Call:
lm(formula = GDP ~ GEXP)

Residuals:
     Min       1Q   Median       3Q      Max
-11227.5  -1812.2    840.7   1272.1  14598.9

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1372.321   1217.772  -1.127    0.269
GEXP           31.950      2.638  12.109 4.45e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5108 on 30 degrees of freedom
Multiple R-squared:  0.8302,      Adjusted R-squared:  0.8245
F-statistic: 146.6 on 1 and 30 DF,  p-value: 4.448e-13
```

In order to check if autocorrelations exist in the residuals obtained from the regression model in equation (3.2), we consider the ACF [Figure 1] and PACF [Figure 2] of the residuals from the regression model:

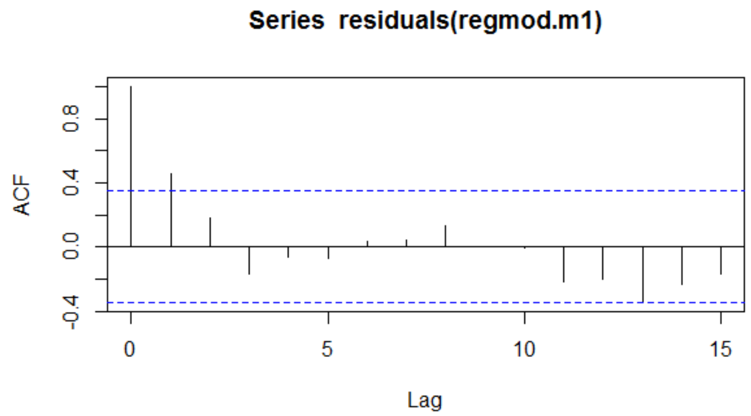


***Fig. 1. ACF of Residuals from Regression Model***

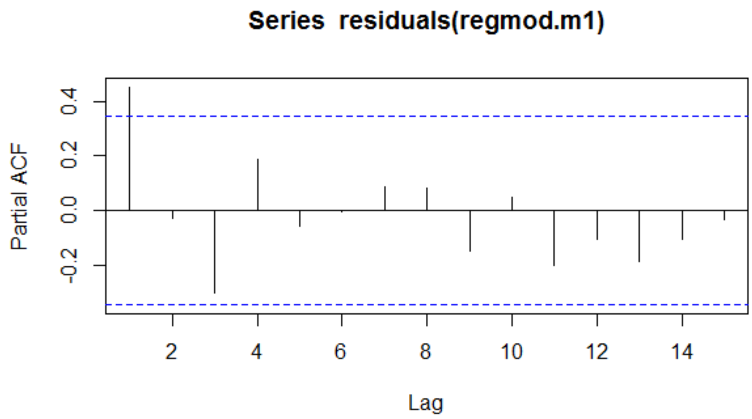


***Fig. 2. PACF of Residuals from Regression Model***

From the ACF [Figure 1], it is observed that the lag starts from zero which is always one. Both the ACF and PACF [Figure 2] indicate that there is a significant spike at lag1 which is more than 5% of the total lags while all other lags fall within the confidence bounds, hence, the residuals from the regression model appear to be autocorrelated. Also, evidence from Box – Ljung test confirms the presence of autocorrelations in the residuals from the regression model since $\chi^2$ = 25.304, df =14 with corresponding p – value = 0.03168 $<$ 0.05 level of significance [Excerpts from Table 3].

***Table 3. Box – Ljung Test for Residuals from Regression Model***

```
Box-Ljung test

data:  residuals(regmod.m1)
X-squared = 25.304, df = 14, p-value = 0.03168
```

Now, it is evident that autocorrelations exist in the residuals from the regression model in (3.2) and the fact that both the ACF and PACF [Figures 1 and 2] respectively, cut off at lag 1 is a clear indicate that there is an ARIMA(1,0,1) structure in the error terms . Therefore, it can be deduced that the unexplained 19.68% of the total variation of the regression model could be modeled by an ARIMA(1,0,1) model.

### REGRESSION MODEL WITH AN ARIMA ERROR

The errors of the regression model in (3.2) appear to be autocorrelated and seem to contain an ARIMA(1,0,1) structure, therefore, we build a regression model that takes into account the ARIMA(1,0,1) structure in the error terms. We also fit tentatively, regression models with ARIMA(1,0,0) and ARIMA(0,0,1) errors, [Tables 5 and 6] respectively. We then compare their information criteria and observe that regression model with ARIMA(1,0,1) error has the smallest information criterion and thus becomes our chosen model for this study. Therefore, the estimated regression model with an ARIMA(1,0,1) structure in the error is presented in equation (3.3)

$$GDP_t = 22317.92 - 6.4314\, GEXP_t\,,\ \ \varepsilon_t = 0.9920\varepsilon_{t-1} + a_t + 0.4595a_{t-1} \tag{3.3}$$

s.e (21755.78) (2.7686) (0.0108) (0.1321)

where $\varepsilon_t$ is the original regression error term while $a_t$ is the corrected (white noise) error term which contains no further information, that is $a_t \sim$ i.i.d. N(0, $\sigma^2$) [Excerpts from Table 4]

From the model in (3.3) it can be deduced that Gross Domestic Product (GDP) is a linear function of Government Expenditure (GEXP) at present year and the residuals at previous year. Also, a unit increase in the Government Expenditure would decrease the GDP by 6.4314 (N' Billion).

***Table 4. Output of Regression Model with an ARIMA(1, 0, 1) Structure***

```
Call:
arima(x = gdp, order = c(1, 0, 1), xreg = gep, method = "ML")

Coefficients:
         ar1     ma1  intercept      gep
      0.9920  0.4595   22317.92  -6.4314
s.e.  0.0108  0.1321   21755.78   2.7686

sigma^2 estimated as 3957282:  log likelihood = -291.03,  aic = 592.05
```

***Table 5. Output of Regression Model with an ARIMA(1, 0, 0) Structure***

```
Call:
arima(x = gdp, order = c(1, 0, 0), xreg = gep, method = "ML")

Coefficients:
         ar1  intercept      gep
      0.9934   19895.27  -1.3703
s.e.  0.0090   19599.90   3.6717

sigma^2 estimated as 5524690:  log likelihood = -295.97,  aic = 599.93
```

***Table 6. Output of Regression Model with an ARIMA(0, 0, 1) Structure***

```
Call:
arima(x = gdp, order = c(0, 0, 1), xreg = gep)

Coefficients:
         ma1  intercept      gep
      0.3955  -741.5707  30.4416
s.e.  0.1325  1490.5666   3.2175

sigma^2 estimated as 19781628:  log likelihood = -314.3,  aic = 636.59
```

### MODEL DIAGNOSTIC CHECKING

The results from Box – Ljung test indicate that the residuals from the model in (3.3) are uncorrelated (white noise) since $\chi^2$ = 8.517, df =16 with corresponding p – value = 0.932 $>$ 0.05 level of significance [Excerpts from Table 7]

***Table 7. Box – Ljung Test for Residuals from Regression Model with an ARIMA(1, 0, 1) Structure***

```
Box-Ljung test

data:  residuals(regmod.7)
X-squared = 8.517, df = 16, p-value = 0.932
```

## 4 CONCLUSION

In this study, the linear relationship between the Gross Domestic Product (GDP) and Government Expenditure (GEXP) is performed by regression model using least squares estimation method and the extra information embedded in the error terms is found to follow an ARIMA(1,0,1) structure. The fact that the error term of the regression model follows an ARIMA(1,0,1) structure implies that the unexplained variance in the GDP which is embedded in the error term is being captured by ARIMA(1,0,1) model, that is to say, the error term of the regression model is autocorrelated and could be corrected by ARIMA(1,0,1). Moreover, a single model that takes into account both the linear relationship between GDP and Government Expenditure, and the extra information in the error term is put forward using maximum likelihood estimation method, that is, regression model with an ARIMA(1,0,1) error. The evidence from the model shows that GDP is a linear function of Government Expenditure at present and immediate previous year. This implies that if the Government Expenditure is kept constant from the immediate previous year to the present year, then, the GDP would tend to decrease. Therefore, Government should vary its expenditures so as to improve the GDP. Methodologically, this study should be extended to cover heteroscedasticity modeling of the nonconstant variances of the error terms in order to accommodate the varying Government Expenditures.

## REFERENCES


[1] H. Akaike, A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. 19, no 6, pp. 716 – 723, 1973.

[2] E. A. Akpan, I. U. Moffat, and N. B. Ekpo, Arma – Arch Modeling of the Returns of First Bank of Nigeria, *European Scientific Journal*, Vol.12, no.8, pp. 257 – 266, 2016.

[3] G.E.P. Box, G. M. Jenkins and G.C.Reinsel, *Time Series Analysis: Forecasting and Control.* 3rd Ed., New Jersey: Wiley and Sons, pp. 5-22, 2008.

[4] G. E. P.Box and D.Pierce, Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models, *Journal of the American Statistical Association,* 65, pp.1509-1526, 1970.

[5] P. J. Brockwell and R. A.Davis, *Introduction to Time Series and Forecasting, 2nd Ed,.* Springer, 2002.

[6] J. D. Cryer and K. Chan, *Time Series Analysis with Application in R,* 2nd Ed., Springer, 2008.

[7] N. R. Drasper and H. Smith, *Applied Regression Analysis,* 3rd Ed., New York, John Wiley and Sons, 1998.

[8] W. A. Fuller, *Introduction to Statistical Time Series,* 2nd Ed., New York ,John Wiley and Sons, 1996.

[9] G. Ljung and G. C. Box,. On a Measure of Lack of Fit in Time series models. *Biometrica, Vol.* 2 no.66, pp. 265-270, 1978.

[10] J. O Rawlings, S. G. Pantula and D. A. Dickey, *Applied Regression Analysis: A Research Tool,* 2nd Ed., Springer, 1998.

[11] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*, 2nd Ed., Adison Westley, 2006.