

Residual Analysis for Auto-Correlated Econometric Model

Habib Ahmed Elsayir

Department of Mathematics, Al Qunfudha University College, Umm Al Qura University, Al Qunfudha, KSA

Email: Habibsayiroi@Yahoo.com

How to cite this paper: Elsayir, H.A. (2019) Residual Analysis for Auto-Correlated Econometric Model. *Open Journal of Statistics*, 9, 48-61.

<https://doi.org/10.4236/ojs.2019.91005>

Received: November 19, 2018

Accepted: January 22, 2019

Published: January 25, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The aim of this article is to provide residual analysis for a time series data of Gross Domestic Product (GDP) of the Sudan. An econometric time series model with macroeconomic variables is conducted to examine the goodness of fit using residual. Many statistical tests are used in time series models in order to make it a stationary series. After applying these tests, the time series became stationary and integrated; thus, Box-Jenkins procedure is used for the determination of ARIMA, AR (0,1,0) in this study. This identified technique is useful for analyzing this study.

Keywords

ARIMA Model, Autocorrelation, GDP, Residual Analysis

1. Introduction

When dealing with validation of data, usually there exists a portion not explained by the model, which combines two parts, the whiteness and the independence test. According to the first part (whiteness test criteria), a good model has the residual autocorrelation function within the confidence interval of the corresponding estimates, indicating that the residuals are uncorrelated; as according to the second criteria of independence test, residuals of a good model should be uncorrelated with previous elements. In fact, there is a need to check residuals in order to validate model performance in regression analysis, which is the main purpose of this empirical study.

Several procedures are used to analyze data within these domains. A useful common technique is the Box-Jenkins ARIMA method [1], which can be used for univariate or multivariate data set analyses. The ARIMA technique involves using Moving Averages (MA), smoothing and regression techniques to detect

and manipulate data autocorrelation problems. Error modeling approach has clearly been demonstrated by Firmino *et al.* [2] and Ikughur, A. J., *et al.* [3].

Diagnostic investigating and checking ARIMA time series models and residual analysis techniques have been used by McLeod and Li [4]. Lu [5] has introduced forecasting for China GDP of a time series model. Andreii and Bugudui [6] have presented an econometric modeling of GDP time series in the US economy. Other similar forecasting models using residual analysis were found in Okyere *et al.* [7], Boshnakov [8] and Lavrenz *et al.* [9]. Martin *et al.* [10] showed checking regression assumptions (errors normality, constant variance, residual analysis plots) in regression analysis fitting experimental data via residual plots.

Residuals Analysis

Consider the regression model:

$$y = a + bx + e \quad (1)$$

The gap between the original value of the dependent variable (y) and the estimated value (\hat{y}) is known as the residual (e). It is the amount of variability in dependent variable that is (left over) after accounting for the variation explained by the predictors in regression analysis. Each data point has one residual:

$$e = y - \hat{y},$$

where $\Sigma e = 0$ and $e = 0$.

Seen as a powerful diagnostic tool, checking the residuals of a regression is a way of checking whether a regression has achieved its goal to explain as much variation as possible in a dependent variable.

Ideally all residuals should be small and unstructured. Most problems that were initially overlooked when diagnosing the variables in the model or were impossible to see, will, turn up in the residuals, for instance:

- Big residuals shown due to outliers
- Certain structure will appear in residuals, will appear in the nonlinearity of the residuals.
- Heteroscedasticity problem, a type of violation of the assumption of non-constant model variation of the residuals.
- Examining residuals plots to check appropriateness of regression model for the data [11].
- If the points in a residual plot indicates appropriateness of a linear or non-linear regression model for the data, Lavrenz *et al.* [9].

The residuals plot shows a certain pattern. Random pattern might be an indication of goodness of fit to the data of a linear model. A form of an autoregressive process time series is as:

$$x_t = b_0 + b_1 x_{t-1} + \epsilon_t \quad (2)$$

For the basic analysis of residuals you will use the usual descriptive tools and scatter plots. A Q-Q Plot can be used to test for residual normality, besides Plotting the residuals to see if there appears any particular pattern (random cloud). A researcher may need to decide whether to adopt linear or log-linear trend

models after answering some questions related to the estimated relationship around the trend line and the correlation of the error terms.

The Autoregressive Time Series Models

Abbreviated as $AR(p)$ models, where p stands for the number of the lagged values of the dependent variable which known as the model “order”. The “order” of the $AR(p)$ models is the number of prior values used in the model. Thus:

$$AR(1): x_t = b_0 + b_1x_{t-1} + \epsilon_t \quad (3)$$

$$AR(2): x_t = b_0 + b_1x_{t-1} + b_2x_{t-2} + \epsilon_t \quad (4)$$

and so on.

When considering the autocorrelation of the residuals that are to be used to evaluate model fit, the testing procedure includes estimating the AR model and calculating of residuals (or error terms), estimating the autocorrelations for the error terms (residuals) and testing to see the inside structure of autocorrelations to see if statistically different from zero. For an $AR(1)$, the values will stay constant when:

$$x_t = b_0/(1-b_1)$$

And rise when:

$$x_t < b_0/(1-b_1)$$

The value falls if:

$$x_t > b_0/(1-b_1)$$

To determine whether a time series is an $AR(p)$ or a Moving average- $MA(q)$ -, examine the autocorrelations. Generally, AR model autocorrelations start at large values and then decline gradually, while the MA autocorrelations drop dramatically after reach of q lags. This behavior describes both the MA process and its order.

Types of Residual Plot

In first graph bellow, linear regression model is preferred since dots are randomly dispread, while in second and third plot, a non-linear regression method is suggested since dots are non-randomly dispread. The randomness property in the first plot shows indicates a good fit for a linear model, as the rest of the plot patterns indicate a non-random structure which expressed by (U and inverted U shaped), suggesting a non-linear model structure (**Figure 1**).

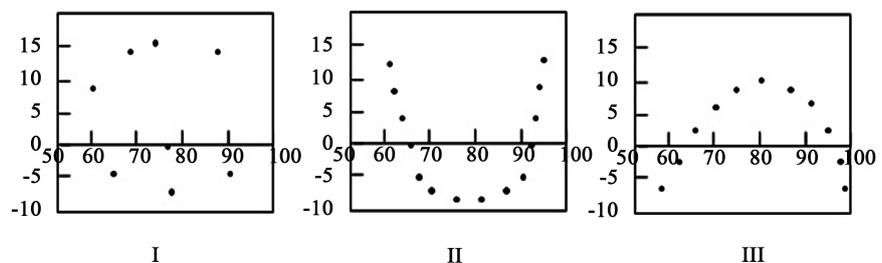


Figure 1. Types of residual plots.

Expected Mean Square (EMS)

Expected Mean Square (EMS) represent the values that we will get for any given mean square (MS) statistic under distribution, on average over repeated experiment.

Let:

$$\begin{aligned} \bar{\mu} &= \sum_{i=1}^k \mu_i / k \\ \lambda_i &= \mu_i - \bar{\mu}, \text{ and} \\ \sigma_A^2 &= \sum_{i=1}^k \lambda_i^2, \text{ (It is variance like)} \end{aligned}$$

Let σ_e^2 be the true error variance. Then the ANOVA is as **Table 1** and **Table 2**.

The Proposed Model

Let Y be the target predicting value and Y_t is the value of Y at time t , then we aimed at constructing a model of the type:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-n}) + e_t \tag{5}$$

where, Y_{t-1} is the previous observation value of Y , Y_{t-2} is the value two observations before, etc. and e_t (a random shock or the noise term). The values of underlying variables that occur prior to the current observation are called lag values. In a repeating pattern time series, the value of Y_t is expected to be highly correlated with $Y_{t-\text{cycle}}$. Thus, the goal of constructing a time series model is to build a model such that the error term to be as small as possible.

Let us consider an X_p time series model, then the Autoregressive Moving Average (ARMA) model combines two components, an autoregressive (AR) component and a moving average (MA) part. Following Brockwell and Davis [12] and Pasavento [13], the AR(p) model can be written in the form:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t \tag{6}$$

where, ϕ_1, \dots, ϕ_p are the model parameters, c is a constant (which may be omitted for simplicity) and ε_t is an error term. The MA(q) notation stands for the moving average model of order q .

Table 1. The EMS for one-way ANOVA.

Source of variation (MS)	Mean Squares Squares (EMS)	Expected Mean
Factor A	MS_A	$\sigma_e^2 + n\sigma_A^2$
Error (Residual)	MS_{error}	σ_e^2

Table 2. The EMS for a two-way ANOV.

Source of variation	MS	EMS
Factor A	MS_A	$\sigma_e^2 + bn\sigma_A^2$
Factor B	MS_B	$\sigma_e^2 + an\sigma_B^2$
A * B interaction	MS_{A*B}	$\sigma_e^2 + n\sigma_{AB}^2$
Error (Residual)	MS_{error}	σ_e^2

$$Y_t = \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (7)$$

where, the $\theta_1, \dots, \theta_q$ are the parameters of the model and the $\varepsilon_t, \varepsilon_{t-1}, \dots$ is an error term.

The ARMA(p, q) notation represents the p autoregressive terms and q moving average terms of the underlying model. This model is given the notation AR(p) and MA(q) models, which is:

$$Y_t = \varepsilon_t + \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (7)$$

where the error terms ε_t are distributed randomly and assumed to be independent with mean zero and variance σ^2 that is $\varepsilon_t \sim N(0, \sigma^2)$.

The process (Y) _{t} is defined to be ARIMA(p, d, q) if:

$$(1-l)^d \varnothing^*(l) Y_t = c + \theta(l) \varepsilon_t \quad (9)$$

where, $\varnothing^*(l)$ is defined in $\varnothing(l) = (1-l)\varnothing^*(l)$, $\varnothing^*(z) \neq 0$ for all $|z| \leq 1$. And $\theta(l)$ is defined in $\theta(z) \neq 0$ for all $|z| \leq 1$.

The process (Y) _{t} is said to be stationary (*i.e.*, its mean, variance and autocorrelation should be approximately constant through time) if and only if $d = 0$ in which case it eliminates to ARMA(p, q) process:

$$\varnothing(l) Y_t = c + \theta(l) \varepsilon_t \quad (10)$$

where, $\varepsilon_t \sim wN(0, \sigma^2)$.

The Box-Jenkins methodology Box *et al.* [1] is a five-step procedure for analyzing and assessing time series data models. Adopting the ARIMA (auto-regressive, integrated, moving average) method iteratively, to best-fit time series data, then the (AR) component in ARIMA is structured as p , the integrated component (I) as d and moving average (MA) as q . The (AR) component represents the previous observations effects, where the (I) component represents model trends and the MA part holds effects of previous random shocks (or error). Then the order of the time series ARIMA fit can be selected assigning an integer value (0, 1, or 2) for each component.

2. Data and Methodology of Collection

The Sudan Central Bureau of Statistics (CBS), issue annual report include all National accounts, while the Central Bank of the Sudan [14] also issues its annual economic records. United Nations forms are used for annual gatherings of the official national accounts data presented to the United Nations Statistics Division by the countries, according to International Monetary fund [15]. If a full set of official data is not reported, estimation methods are used to obtain estimates for the entire time series. Then annual percentage growth indicators including annual rates of GDP based on market prices is presented on constant local currency, while the aggregates are based on constant U.S. dollars [16].

3. Data Analysis

In this section, the data of GDP statistics of Sudan, which include the current and constant prices in million US\$ for the period (1960-2015) will be investigated using SPSS Time Series Modeler. The created series for the GDP model with creating function has been made and presented in **Table 3** and result variables are displayed in **Table 4**. The result variables for the GDP model functions are identical. As in **Table 5**, applying the model specifications, it seems that it is not applicable for calculating standard errors of the partial autocorrelations, for independence (white noise).

The partial autocorrelation for the built model is the autocorrelation of time series observations separated by a lag of 16 time units with the effects of the intervening observations eliminated. Autocorrelation presented in **Table 6** with autocorrelation function (**Figure 2**)—values seen above zero in (**Figure 4**) too—and partial autocorrelation tables (**Table 7**) with partial autocorrelation (**Figure 3**) are also provided for the residuals (errors) between the actual and predicted values of the time series. Examining the autocorrelation table shown in **Table 6**, we see that the highest autocorrelation is 0.875 (the first value in the

Table 3. Created series.

	Series Name	Case number of non-missing values			Creating Function
		First	Last	N of Valid Cases	
1	GDP_1	2	56	55	DIFF (GDP,1)
1	GDP_1	2	56	55	MA (GDP,1,1)
1	GDP_1	2	56	55	T4253 H (GDP)

Table 4. Result variables.

Result Variables						
	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	GDP_1	1	1	57	57	SMEAN (G DP)
Result Variables						
	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	GDP_1	1	1	57	57	SMEAN (G DP)
Result Variables						
	Result Variable	N of Replaced Missing Values	Case Number of Non-Missing Values		N of Valid Cases	Creating Function
			First	Last		
1	GDP_1	1	1	57	57	SMEAN (G DP)

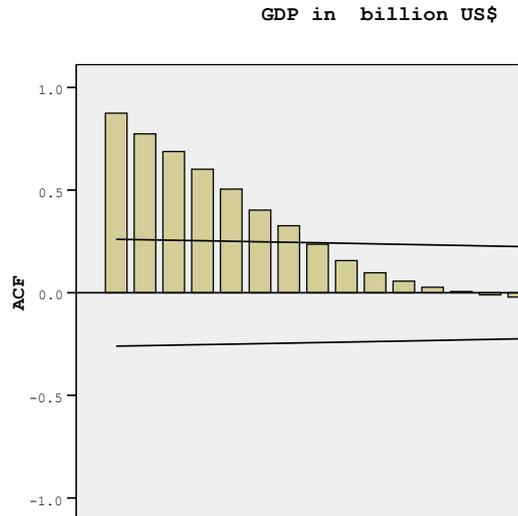


Figure 2. Autocorrelation function.

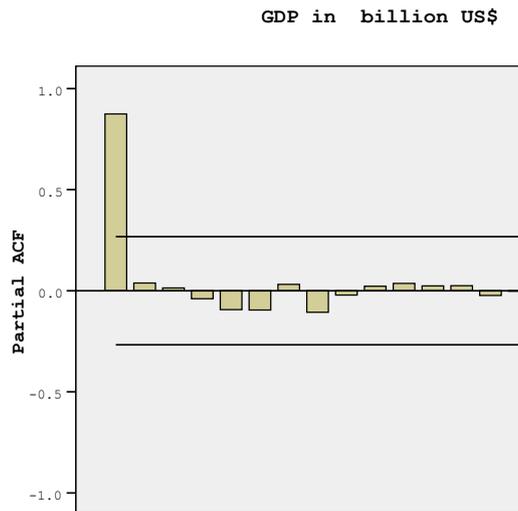


Figure 3. Partial autocorrelation function.

Table 5. Model description.

Model Description		
Model Name	MOD_1	
Series Name	1	GDP in billion US\$
Transformation	None	
Non-Seasonal Differencing	0	
Seasonal Differencing	0	
Length of Seasonal Period	No periodicity	
Maximum Number of Lags	16	
Process Assumed for Calculating the Standard Errors of the Autocorrelations	Independence (white noise) ^a	
Display and Plot	All lags	

Applying the model specifications from MOD_1. ^aNot applicable for calculating the standard errors of the partial autocorrelations.

Table 6. Model autocorrelations.

Autocorrelations					
Series: GDP in billion US\$					
Lag	Autocorrelation	Std. Error ^a	Box-Ljung Statistic		
			Value	df	Sig. ^b
1	0.875	0.130	45.167	1	0.000
2	0.774	0.129	81.184	2	0.000
3	0.688	0.128	110.152	3	0.000
4	0.601	0.127	132.735	4	0.000
5	0.505	0.125	148.958	5	0.000
6	0.402	0.124	159.476	6	0.000
7	0.326	0.123	166.539	7	0.000
8	0.236	0.122	170.309	8	0.000
9	0.157	0.120	172.002	9	0.000
10	0.097	0.119	172.665	10	0.000
11	0.056	0.118	172.892	11	0.000
12	0.027	0.116	172.945	12	0.000
13	0.006	0.115	172.948	13	0.000
14	-0.011	0.114	172.957	14	0.000
15	-0.022	0.112	172.994	15	0.000
16	-0.027	0.111	173.054	16	0.000

^aThe underlying process assumed is independence (white noise). ^bBased on the asymptotic chi-square approximation.

Table 7. Partial autocorrelations.

Partial Autocorrelations		
Series: GDP in billion US\$		
Lag	Partial Autocorrelation	Std. Error
1	0.875	0.134
2	0.038	0.134
3	0.014	0.134
4	-0.039	0.134
5	-0.093	0.134
6	-0.095	0.134
7	0.032	0.134
8	-0.106	0.134
9	-0.022	0.134
10	0.022	0.134
11	0.036	0.134
12	0.024	0.134
13	0.025	0.134
14	-0.023	0.134
15	-0.002	0.134
16	-0.003	0.134

lags) which occurs with a lag of 15. Now we aim to be sure to include lag values up to 15 when constructing the model. Based on the assumption that the series are not cross correlated and that one of the series is white noise, the cross correlations are found in **Table 8**.

Using Time Series Modeler, the model specification was shown with range of lags from -7 to $+7$ as seen in **Table 9** and described in **Table 10** that is (ARIMA(0,1,0)) and the model fit summary is shown in **Table 11** and the residuals ACF summary is presented in **Table 12**.

Table 8. Model cross correlations.

Cross Correlations		
Series Pair: GDP in billion US\$ with year		
Lag	Cross Correlation	Std. Error ^a
-7	0.603	0.143
-6	0.631	0.141
-5	0.658	0.140
-4	0.685	0.139
-3	0.711	0.137
-2	0.737	0.136
-1	0.761	0.135
0	0.785	0.134
1	0.678	0.135
2	0.587	0.136
3	0.507	0.137
4	0.431	0.139
5	0.352	0.140
6	0.274	0.141
7	0.211	0.143

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 9. Model description.

Model Description		
Model Name		MOD_2
Series Name	1	GDP in billion US\$
	2	year
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Range of Lags	From	-7
	To	7
Display and Plot		All lags

Applying the model specifications from MOD_2.

Table 10. Model description.

Model Description	
Model Type	
Model ID GDP in billion US\$ Model_1	ARIMA (0,1,0)

Table 11. Model summary.

Fit Statistic	Model Fit										
	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Stationary R-squared	-2.2E-016		-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016	-2.2E-016
R-squared	0.982		0.982	0.982	0.982	0.982	0.982	0.982	0.982	0.982	0.982
RMSE	3.157		3.157	3.157	3.157	3.157	3.157	3.157	3.157	3.157	3.157
MAPE	13.577		13.577	13.577	13.577	13.577	13.577	13.577	13.577	13.577	13.577
MaxAPE	77.561		77.561	77.561	77.561	77.561	77.561	77.561	77.561	77.561	77.561
MAE	2.190		2.190	2.190	2.190	2.190	2.190	2.190	2.190	2.190	2.190
MaxAE	7.296		7.296	7.296	7.296	7.296	7.296	7.296	7.296	7.296	7.296
Normalized BIC	2.372		2.372	2.372	2.372	2.372	2.372	2.372	2.372	2.372	2.372

Table 12. Residual ACF summary.

Lag	Residual ACF Summary										
	Mean	SE	Minimum	Maximum	Percentile						
					5	10	25	50	75	90	95
Lag 1	0.056		0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056
Lag 2	0.060		0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060	0.060
Lag 3	-0.020		-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020	-0.020
Lag 4	0.141		0.141	0.141	0.141	0.141	0.141	0.141	0.141	0.141	0.141
Lag 5	-0.181		-0.181	-0.181	-0.181	-0.181	-0.181	-0.181	-0.181	-0.181	-0.181
Lag 6	-0.050		-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050	-0.050
Lag 7	-0.070		-0.070	-0.070	-0.070	-0.070	-0.070	-0.070	-0.070	-0.070	-0.070
Lag 8	0.174		0.174	0.174	0.174	0.174	0.174	0.174	0.174	0.174	0.174
Lag 9	-0.034		-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034	-0.034
Lag 10	0.102		0.102	0.102	0.102	0.102	0.102	0.102	0.102	0.102	0.102
Lag 11	-0.161		-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161	-0.161
Lag 12	0.025		0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025
Lag 13	0.042		0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042	0.042
Lag 14	-0.143		-0.143	-0.143	-0.143	-0.143	-0.143	-0.143	-0.143	-0.143	-0.143
Lag 15	-0.313		-0.313	-0.313	-0.313	-0.313	-0.313	-0.313	-0.313	-0.313	-0.313
Lag 16	-0.064		-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064
Lag 17	-0.103		-0.103	-0.103	-0.103	-0.103	-0.103	-0.103	-0.103	-0.103	-0.103
Lag 18	-0.083		-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083	-0.083
Lag 19	-0.064		-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064	-0.064
Lag 20	0.047		0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047	0.047
Lag 21	0.082		0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082
Lag 22	-0.102		-0.102	-0.102	-0.102	-0.102	-0.102	-0.102	-0.102	-0.102	-0.102
Lag 23	-0.069		-0.069	-0.069	-0.069	-0.069	-0.069	-0.069	-0.069	-0.069	-0.069
Lag 24	-0.008		-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008	-0.008

The autocorrelation ACF and partial autocorrelation PACF tables provide valuable information about the significance of the lag variables. An autocorrelation is the correlation between the target variable (GDP) and lag values for the same variable. It is known that correlation values lies between -1 and $+1$. A value of $+1$ indicates that the two variables move together perfectly; a value of -1 indicates that they move in opposite directions (see the results of **Table 14**. The third column of the autocorrelation table shows the standard error of the autocorrelation, this is followed by Box-ljung statistics based on the asymptotic chi-square approximation (all values are significant) in the fourth column. The autocorrelation bar chart is used to indicate positive or negative correlations up or down of the centerline. The dots shown in the chart mark the points two standard deviations from zero. If the autocorrelation bar is longer than the dot marker (that is, it covers it), then the autocorrelation should be considered significant. In this model, significant autocorrelations occurred for all lags except for lag 15. On the basis of the assumption that the series are not cross correlated and that one of the series is white noise, the cross correlations and range of lags (from -7 to $+7$ are displayed in **Table 13** and **Figure 4**. The figure shows confidence limit to be all above zero for the GDP.

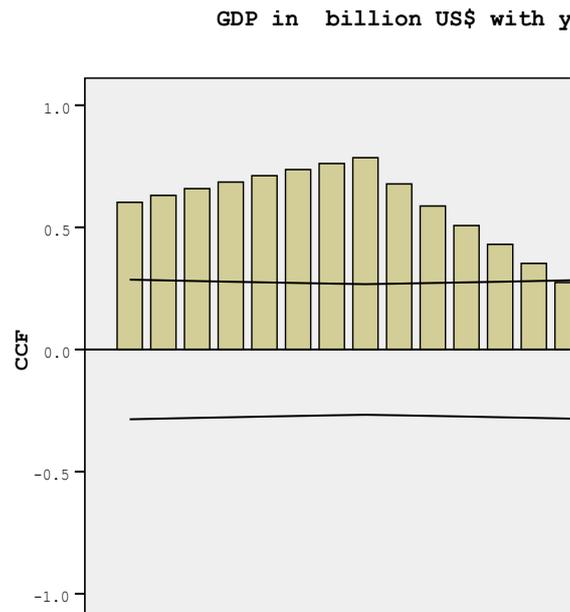


Figure 4. Cross correlation function.

Table 13. Model statistics.

Model Statistics									
Model	Number of Predictors	Model Fit statistics				Ljung-BoxQ (18)			Number of Outliers
		Stationary R-squared	R-squared	RMSE	MAE	Statistics	DF	Sig.	
GDP in billion US\$-Model_1	0	-2.22E-016	0.982	3.157	2.190	20.011	18	0.332	0

Proportion of variance explained by model is the best single measure of how well the estimated values match the original values. If the estimated values exactly match the original values, then the model would explain all the amount of variation (100%). In fact this is not always the case (here the model explains 98.2% of the variance due to the R square value), as seen in **Table 13**. The ARIMA model parameters using natural log show significant t value (0.001) as in **Table 14**. The residual ACF is displayed in **Table 15** and the residual PACF is presented in **Table 16**.

Finally, the model forecast has been shown in **Table 17**, since one of the strengths of time series model analysis is its ability of generation of future data forecasts by depending on observed past observations. Thus, if we rely on this information, we may conclude that we have a good model fit.

Table 14. ARIMA model parameters.

ARIMA Model Parameters				Estimate	SE	t	Sig.
GDP in billion US\$-Model_1	GDP in billion US\$	Natural Log	Constant Difference	0.0781	0.023	3.378	0.001

Table 15. Residual autocorrelation function.

Residual ACF																									
Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
GDP in billion 1	AC	0.056	0.060	-0.020	0.141	-0.181	-0.050	0.070	0.174	-0.034	0.102	-0.161	0.025	0.042	-0.143	-0.313	-0.064	-0.103	-0.083	-0.064	0.047	0.082	-0.102	-0.069	0.008
	SE	0.135	0.135	0.136	0.136	0.138	0.143	0.143	0.144	0.147	0.147	0.149	0.152	0.152	0.152	0.155	0.166	0.166	0.167	0.168	0.169	0.169	0.169	0.171	0.171

Table 16. Residuals partial autocorrelation function.

Residual PACF																									
Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
GDP in billion 1	PA	0.056	0.057	-0.027	0.141	-0.199	-0.042	0.039	0.168	-0.002	0.072	-0.193	0.026	0.133	-0.191	-0.236	-0.108	-0.105	-0.043	-0.045	0.074	0.030	-0.161	-0.056	0.072
	SE	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135	0.135

Table 17. Model forecasts.

Forecast		
Model		57
GDP in billion US\$-Model_1	Forecast	106.639
	UCL	148.340
	LCL	74.426

For each model, forecasts start after the last non-missing in the range of the requested estimation period, and end the last period for which non-missing values of all the predictors are available or at the end date of the request forecast period, whichever is earlier.

4. Discussion

A residual is the vertical distance between a data point and the regression line. Considering residual plots, it can be used to assess whether the observed error (residuals) is consistent with random (or stochastic) error. In a regression model, the residuals should not be either systematically high or low. So, they should be spread about zero throughout the range of fitted values. Further, in the context of OLS method, random errors are assumed to yield normally distributed residuals. This means that the residuals should be distributed in a symmetrical pattern and spread constantly throughout the range.

We evaluate Autoregressive Integrated Moving Average (ARIMA) model of the GDP series using Box-Jenkins methodology by using four different equations which are, linear, logarithmic, quadratic and exponential equations. Rely on the parameter estimates, it is found that the ARIMA (0,1,0) is the best model for the data. Comparing with other models, ARIMA model has been selected as the final model. The method for prediction and forecasting then has been provided based on data, which may be performable and useful to governmental and nongovernment institutions.

In time series regression analysis of data, caution must be taken about using this technique, because of autocorrelation and violation of the assumption of errors independence which leads to increase in type I error when autocorrelation is present. Furthermore, time series patterns must be accounted for within the analysis.

5. Conclusion

This article has discussed the technique of residual analysis for an economic GDP model. The procedures—which mainly depend on time series analysis and ARIMA method in particular—used here might be valuable only for a time series that is stationery and it is much preferred and more recommended when carried out for a model with at least 50 data observations (our model has 57 observations). The procedure of residual analysis in an economic time series model is outlined. The model has been investigated to describe the data to see how well the underlying model fits the GDP data which can help econometricians to understand the behavior and structure of the Sudanese economy. Thus, this analysis may provide practical utility of the procedure of residual analysis.

Acknowledgements

I am deeply indebted to the editorial board and the reviewers of the Journal of Mathematics and Statistics for their valuable comments and efforts in order to publish this paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis: Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliffs, NJ.
- [2] Firmino, P.R.A., de Mattos Neto, P.S.G. and Ferreira, T.A.E. (2015) Error Modeling Approach to Improve Time Series Forecasters. *Neurocomputing*, **153**, 242-254. <https://doi.org/10.1016/j.neucom.2014.11.030>
- [3] Ikughur, A.J., Uba, T. and Ogunmola, A.O. (2015) Application of Residual Analysis in Time Series Model Selection. *Journal of Statistical and Econometric Methods*, **4**, 41-53. http://www.scienpress.com/Upload/JSEM%2fVol%204_4_3.pdf
- [4] McLeod, A.I. and Li, W.K. (1983) Diagnostic Checking ARMA Time Series Models Using Squared-Residual Autocorrelations. *Journal of Time Series Analysis*, **4**, 269-273. <https://doi.org/10.1111/j.1467-9892.1983.tb00373.x>
- [5] Lu, Y. (2009) Modeling and Forecasting China's GDP Data with Time Series Models. D-Level Essay in Statistics. Department of Economics and Society, Hogskolan Dalarna, Sweden.
- [6] Andreii, E.A. and Bugudui, E. (2011) Econometric Modeling of GDP Time Series. *Theoretical and Applied Economics*, **18**, 91-98. <http://store.ectap.ro/articole/652.pdf>
- [7] Okyere, F., Mahama, F., Yemidi, S. and Krampa, E. (2015) An Econometric Model for Inflation Rates in the Volta Region of Ghana. *IOSR Journal of Economics and Finance*, **6**, 48-55.
- [8] Boshnakov, G.N. (2016) Introduction to Time Series Analysis and Forecasting. 2nd Edition, John Wiley and Sons, Hoboken.
- [9] Lavrenz, S.M., Vlahogianni, E.I., Gkritza, K. and Ke, Y. (2018) Time Series Modeling in Traffic Safety Research. *Accident Analysis & Prevention*, **117**, 368-380. <https://doi.org/10.1016/j.aap.2017.11.030>
- [10] Martin, J., de Adana, D.D.R. and Asuero, A.G. (2017) Fitting Models to Data: Residual Analysis, a Primer, Uncertainty Quantification and Model Calibration. Jan Peter Hessling, IntechOpen. <https://www.intechopen.com/books/uncertainty-quantification-and-model-calibration/fitting-models-to-data-residual-analysis-a-primer>
- [11] Frost, J. (2012) Why You Need to Check Your Residual Plots for Regression Analysis.
- [12] Brockwell, P.J. and Davis, R.A. (2002) Introduction to Time Series and Forecasting. 2nd Edition, Springer, New York. <https://doi.org/10.1007/b97391>
- [13] Pasavento, E. (2007) Residuals-Based Tests for the Null of No-Co-Integration: An Analytical Comparison. *Journal of Time Series Analysis*, **28**, 111-137. <https://doi.org/10.1111/j.1467-9892.2006.00501.x>
- [14] Central Bank of Sudan (2015) Sudan GDP and Economic Data. Country Report 2015.
- [15] International Monetary Fund (IMF) (2014) Global Finance Magazine. World Economic Outlook (WEO) Database. International Monetary Fund.
- [16] Sudan GDP Annual Growth Rate (2015). <http://www.tradingeconomics.com/sudan/gdp-growth-annual/forecast>