ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 5 No. 1b June 2019*

Forecasting of Per Capita Income using the Dynamic Model (A Case Study of Nigeria)

Shamsu Shehu
Department of mathematical Sciences
Al-qalam University Katsina.
Email: shamsushehu@yahoo.com

Abstract
Many Time Series Analysts or Econometricians studied and forecasted the per capita income using multivariate regression techniques. This study investigated the PCI of Nigeria using a different approach known as Univariate Time Series technique to forecast the series for the period of 2009 to 2030 using ARIMA model. The results revealed that there is an upward trend and the 2nd difference of the series was stationary, meaning that the series was I(2). Base on the AIC and BIC selection criteria, the best model that explains the series was found to be ARIMA (1, 2, 0). Residual diagnosis of the model confirmed that the error was white noise, no correlation found and forecast of 22 years terms was made which indicates that the series continue to increase with these forecasted time period.

Keyword: Per capita income, Autocorrelation function, Partial Autocorrelation function, AIC, BIC, ARIMA

INTRODUCTION
Generally speaking, per capital income (PCI) or income per person can be defined as the mean income of the people in an economic unit such as city or country. It is determined for a country by dividing the gross domestic product (GDP) of such country by its population. The importance of PCI can never be neglected because it counts each man, woman and child, even newborn babies as a member of the population. Therefore, it stands in contrast to other common measures of an area's prosperity such as household income, which counts all people residing under one roof as a household, and family income, which counts a family those related by birth, marriage or adoption who live under the same roof. The most important uses of PCI include:

- Access an area's affordability
- To ascertain an area's wealth or lack of wealth e.g. PCI is one metric the United States Bureau of Economic Analysis (BEA) uses to rank the wealthiest counties in the US, the other being median household income.
- It can be used in conjunction with data on real estate prices

Dynamic econometric models are regression models that take time lags into account (Gujarati, 2004). They are also referred to as lagged regression models. There are two types of dynamic models namely: the autoregressive and distributed-lag models. According to the literature, Box and Jenkins (1976) applied autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models to find the best fit of a time series model using past values of a time series. This can also be seen in William and Jane (1978).

Many researchers have used multiple linear regression approaches to study and forecast the PCI. For instance, Faggohun and Adekoya (2016) investigated the long-run per-capita income growth in Nigeria within the period of 1970 to 2014. They regressed capital (k), government expenditure (gexp), trade openness (top) and enrolment (entl) variables on the real GDP per-capita. His empirical founding revealed that only trade openness (top) has positive and significant impact on growth rate of per-capita income of Nigeria.

Also, Bello (2010) claimed that only the Gross Domestic Product (GDP), inflation rate and exchange rate have significant effect on the PCI using multiple linear regression models. He discovered that 96.3% of the total variation in per capita income can be accounted for by these three economic indicators.

Charles *et al.* (2010) worked on convergence of real per capita GDP within Common Market for Eastern and Southern Africa (COMESA) using a panel unit root evidence technique. Their studies showed no evidence of convergence process for the income in the COMESA. This implies that economies of COMESA are locked into sustain poverty trap process.

In a study, Bipasha and Bani (2012) fitted a very simple tentative ARIMA (1, 2, 2) model to the GDP of India. Their results suggested that only one period of autoregressive and moving average terms are statistically significant. Further results showed that absolute values of forecasted GDP indicated an increasing trend and its respective growth rates revealed an opposite trend in future.

The ARMA model was used differently by Cejun and Chou-Lin (2004) to forecast the crash fatalities during summer holiday in the United States. They were able to show that fatalities are higher during holiday than during non-holiday periods.

The PCI in this study will be investigated through the use of univariate time series approach. That is, the past records of the variable will be used to study its future. The objectives are three. Firstly, the pattern of PCI of Nigeria will be examined. Secondly, a stochastic ARIMA model that best fit the PCI will be developed. Thirdly, we will use the developed model to forecast future values of the PCI for another twenty-two years.

In summary, the previous studies reported in the literature focused on the dependency of the PCI on other (explanatory) variables or per capita GDP was studied using a panel unit root approach; which excludes forecasting. Further, it is not clear why many studies have not forecasted the PCI using univariate time series analysis technique. Since the PCI serves as a measure of economic development of any nation, hence forecasting it will assist the Government through the policy makers to foresee the growth of economic development in Nigeria

SOURCE OF DATA

Data used for this study are annual figures covering the period 1991-2008 and the univariate variable is the per capita income (PCI). Data for the GDP were obtained from the Statistical Bulletin of the Central Bank of Nigeria (2007) while data for the population were obtained from National Bureau of Statistics (NBS).

MATERIALS AND METHODS

In order to calculate the PCI we used the formula: $\mathrm{PCI} = \frac{\mathrm{GDP_i}}{\mathrm{population_i}}$ Where; $\mathrm{GDP_i} =$ GDP for ith year and $\mathrm{population_i} =$ population for ith year

The stationarity of the Per Capita Income (PCI) series has to be determined first. This is done by taking the following two steps: (1) observing the time plot of the series whether there is trend or not. (2) confirming the results reported by time plot by conducting the unit root test using the Augmented-Dickey Fuller (ADF) test approach. The study used the ADF model with drift for its unit root tests.

Augmented-Dickey Fuller (ADF) test

The Dickey-Fuller (DF) test can be used to test any of the three possible models of the PCI data under three different null hypotheses as specified below:

$\Delta PCI_t = \delta PCI_{t-1} + u_t$ (None) (1)

$\Delta PCI_t = \beta_1 + \delta PCI_{t-1} + u_t$ (With drift) (2)

$\Delta PCI_t = \beta_1 + \beta_2 t + \delta PCI_{t-1} + u_t$ (With drift and trend) (3)

Where: $\delta = \rho - 1$, $\beta_1 =$ drift or intercept term, $t =$ trend or time, $u_t =$ white noise, PCI_t is a random walk, PCI_{t-1} is the lagged one period of the PCI variable.

In order to conduct the Dickey-Fuller (DF) test for three equations above, we assume that the error u_t is uncorrelated. Should the u_t are correlated, the ADF developed by Dickey and Fuller is used to correct the correlation by "augmenting" the three equations by adding the lagged values of the dependent variable ΔPCI_t

The following hypotheses have been formulated with respect to equation (2)

$H_o: \delta = 0$(PCI has a unit root) vsH_o: Not H_1(PCI does not have a unit root)

Test statistic:

The test statistic is given by: $t_\alpha = \frac{\widehat{\alpha}}{Se(\widehat{\alpha})}$

Decision rule: Reject the null hypothesis if t_α is greater than the 5% critical value. Otherwise, we do not reject the null hypothesis.

If the null hypothesis is accepted, that means the PCI series is not stationary at level. That is, it has to be differenced. And differencing a time series once or twice may make such series stationary (Box and Jenkins, 1976). Once stationarity is attained, we therefore proceed to modeling approach.

Modeling approach

This paper utilized one of the most popular methods of forecasting known as the ARIMA model to forecast the PCI from 2009 through 2028. The AR stands for the autoregressive i.e. regressing the dependent variables with linear combination of its lagged values, MA denotes moving average i.e. regressing the dependent error with linear combination of its past error or innovation and "I" denotes differencing order (i.e. number of differencing applied on the stochastic process to attain stationarity). The model is given by:

$$X_t = \mu + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + \varepsilon_t \quad (4)$$

In the general form of ARIMA (p, d, q), p denotes the number of periods in the past for AR, q denotes the number of periods in the past for MA, and d denotes the integrating order.

According to the Box and Jenkins (1976) methodology, there are four steps to be taken in order to achieve accurate forecasting. These steps are classified as: model identification, model estimation, diagnostic checking and forecasting.

1. **Model identification:** The appropriate values of p, d and q in ARIMA (p, d, q) model will be pre-determined using the Autocorrelation Function (A.C.F) and Partial Auto Correlation Function (P.A.C.F) of the correlogram.
2. **Model estimation:** Here, the parameters of the pre-determined or identified ARIMA (p, d, q) model is estimated using the iterative methods to further confirm the identified model under the criteria measurements of Alkaike Information Criteria (AIC) and Bayesian Information Criteria (BIC)
3. **Diagnostic checking:** We have to check if the estimated residual of the fitted model is white noise. This can be checked either by plotting the ACF and PACF of the residual or by performing the Ljung Box test on residuals. This paper used ACF and PACF plots for its diagnosis checking.
4. **Forecasting:** Once step 1-3 is properly taken care of, the next thing is forecasting.

DATA ANALYSIS

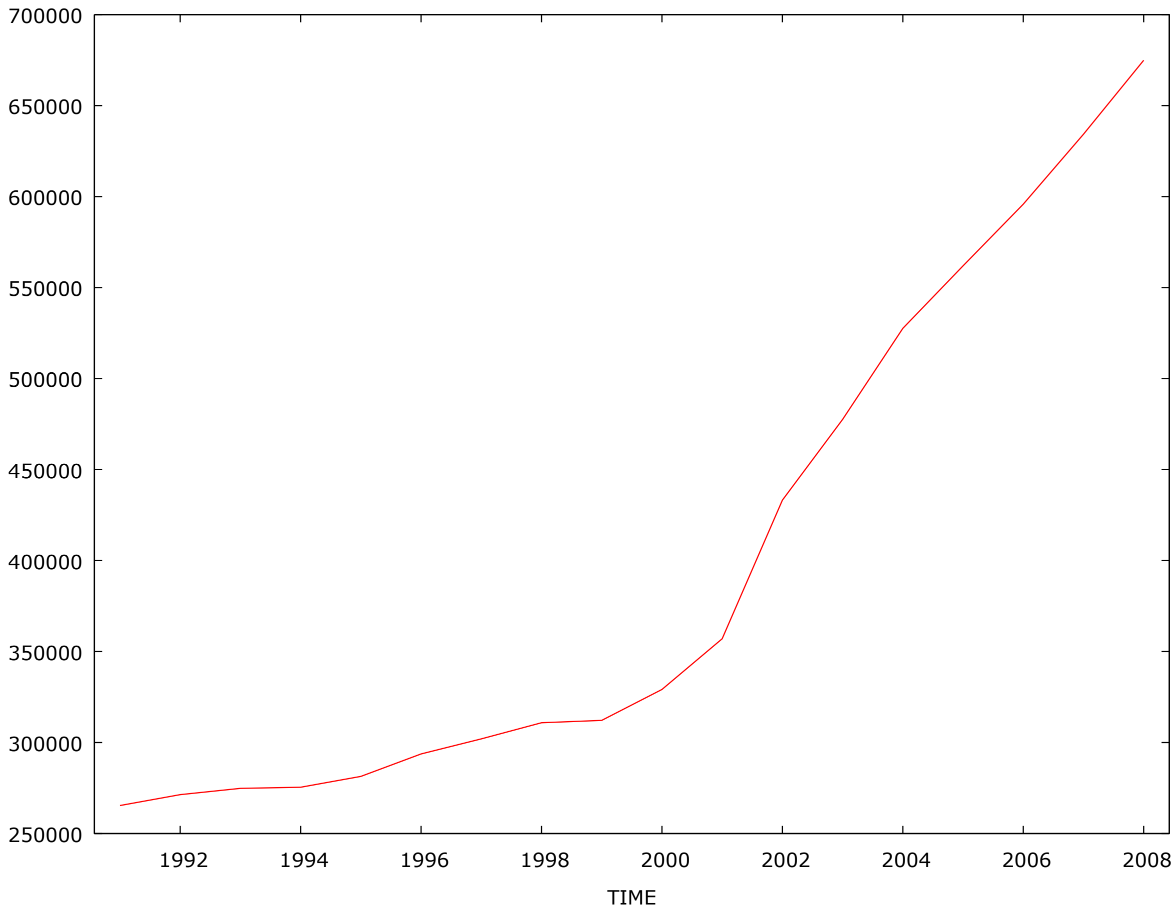

Figure1: Time plot of PCI over a period 1991-2008

The graphical representation of the data in Figure1 reveals that the PCI series exhibits an upward trend. This means that the series does not follow a stationary pattern. Hence, differencing the data once or twice may keep it stationary as suggested by Box and Jenkins (1976). The result reported by the time plot is further confirmed by subjecting the data to a unit root test, to see truly if the data is not stationary at level. This paper used the ADF test for its unit root analysis. Table 1 presents the results of the ADF tests under three different null hypotheses.

Table 1: Results of unit root test using ADF model with drift

Variable	Various unit root tests	ADF-statistic	Critical values	p-value	Decision	Remarks
PCI with intercept	At level	0.7203	1%= -3.9204 5%= -3.0656 10%= -2.6735	0.9888	Accept H_o	Non-stationary
PCI with intercept & trend	First difference	-1.5737	1%= -3.9204 5%= -3.0656 10%= -2.6735	0.4722	Accept H_o	Non-stationary
PCI with none	Second difference	-5.0353	1%= -3.9591 5%= -3.0810 10%= -2.6813	**0.0014***	Reject H_o	Stationary

Note: the ADF test is only significant at second difference

Table 1 shows that the PCI series is not stationary at level. That is, the data have been differenced twice in order to attain stationarity and is only significant at second difference since the p-value (=0.0082) is less than level of significance (=0.05). Before further examination of the results, we have to choose the appropriate model for the ADF test. Table2 presents the estimated coefficients of the three different possible ADF equations.

Table 2: Summary of coefficients of the ADF equations under three different null hypotheses

Coefficients of PCI_{t-1}			
	Intercept	**Intercept & Trend**	**None**
At level	0.0335*(0.4841)*	-0.0720 *(0.3598)*	0.0313*(0.1052)*
First difference	-0.2965*(0.1379)*	-0.6405*(0.0286)*	-0.0827*(0.5413)*
Second difference	-1.3193***(0.0002*)***	-1.3199***(0.0004*)***	-1.2947 ***(0.0002*)***

Note: the p-value is reported in parenthesis ( ) as indicated in the table. The best results as reported by the p-value are asterisked (*)

Table 2 shows that all the three different possible models at their second difference are appropriate for the unit root test, since coefficients of the PCE_{t-1} in each case are not only negatives (i.e. -1.3193, -1.3199 and -1.2947) but also the absolute value of ρ in each case (i.e. $|\rho|$=0.3193 for the model with drift, $|\rho|$=0.3199 for the model with intercept & trend, $|\rho|$=0.2947 for the model with none) is surely less than one. Hence, the PCI is an integrated series of order 2 denoted by I (2). The ADF model with drift is thus given by:

$$\triangle \widehat{\mathrm{PCET}}_{t-1} = -775.2286 - 1.3193\mathrm{PCE}_{t-1}$$

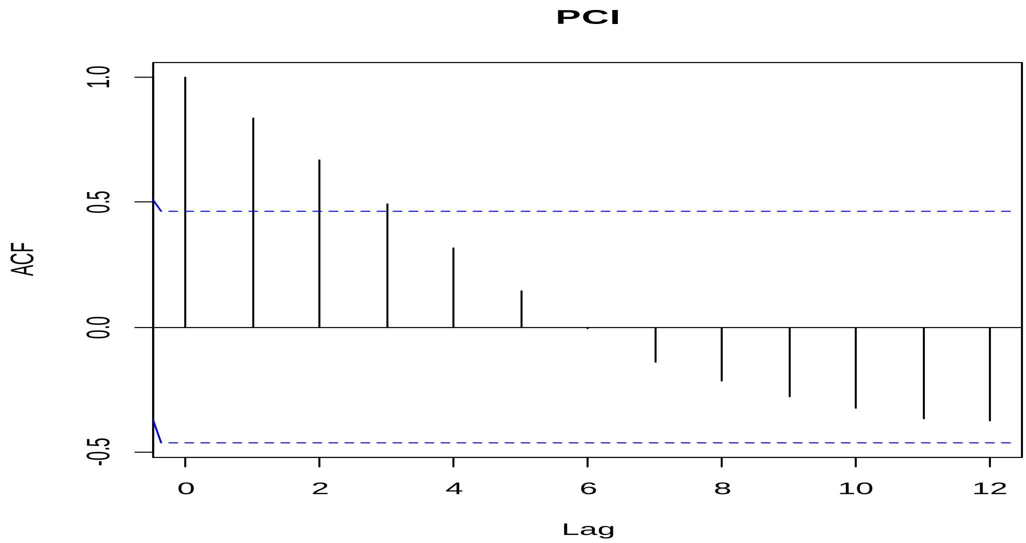

Figure 2a: Autocorrelation function for PCI at level form

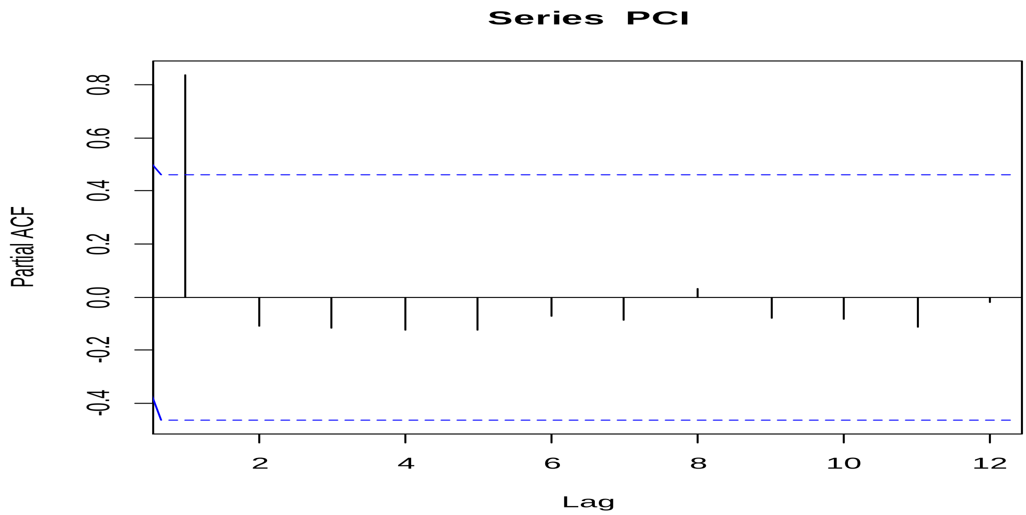

Figure2b: Partial autocorrelation function for PCI at level form

Figure 2a and Figure 2b revealed that the spikes are decaying exponentially for the autocorrelation function (ACF) while for the partial autocorrelation function (PACF), the spike is significant only at lag1.This means that the AR (1) model is suggested to fit the data. The suggested model is further subjected to confirmation by estimating various ARIMA model. Thus, Table 3 presents the results of various ARIMA model.

ESTIMATION OF PARAMETERS

Table 3: Summary of various estimated ARIMA (p, d, q) models

S/n	(p, d, q)	AIC	BIC	S.E	LogL
1	(0,2,1)	359.7635	362.0813	2676.82	-176.8818
2	(2,2,0)	361.3135	364.4039	3324.63	-176.6568
3	(1,2,0)	**359.6002***	**361.9179***	2958.51	-176.8001
4	(1,2,1)	360.8937	363.9840	3224.63	-176.4468
5	(2,2,2)	362.4995	367.1350	1142.23	-175.2497
6	(2,2,1)	362.8745	366.7374	3324.72	-176.4372
7	(1,2,2)	362.8148	366.6777	3578.43	-176.4074
8	(1,2,3)	360.8264	365.4619	1093.89	-174.4132
9	(3,2,1)	359.8567	364.4923	749.804	-173.9284
10	(1,2,4)	368.2231	362.8150	1106.62	-174.4075
11	(4,2,1)	361.6531	367.0612	708.571	-173.8266
12	(1,2,5)	364.7769	370.9576	1155.21	-174.3885

*Note: AIC= Akaike information criteria, BIC= Bayesian information criteria, S.E=Standard error and LogL= Log likelihood. The best results as reported by the criteria are asterisked *.*

From Table 3, the least values of the AIC and BIC estimation criteria occurs at ARIMA (1, 2, 0). Hence, ARIMA (1, 2, 0) has been confirmed to be the best model. Hence, Table 4 presents the model estimates

Table 4: Estimates of ARIMA (1, 2, 0) model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27426.55	14724.09	1.862699	0.0872
AR(1)	0.680805	0.207691	3.277978	0.0066

R-squared	0.472415	Mean dependent var	23175.44
Adjusted R-squared	0.428449	S.D. dependent var	22357.51
S.E. of regression	16902.49		
Sum squared resid	3.43E+09		
Log likelihood	-155.0791		
F-statistic	10.74514	Durbin-Watson stat	2.265664
Prob(F-statistic)	0.006605		
Inverted AR Roots	.68		

The estimated model is $\mathrm{PCI_t} = 27426.55 + 0.680805\mathrm{PCI_{t-1}} + \varepsilon_t$
From Table 4, the coefficient of the ARIMA (1, 2, 0) model was valid and stationary condition was met and satisfied since the coefficient is less than one. It was also found to be significant since its p-value (= 0.0066) is less than level of significance (= 0.05). This is also justified by the

p-value of value (= 0.006605) was less than the exact probability (0.05) , these means that the probability of rejecting the overall significance of ARIMA (1, 2, 0) model is 0.6605% while the probability of accepting it is 99.3395%. Hence, AR (1) thus explains the series. The accuracy of the model is also reported by comparing the R^2 (= 0.472415)and Durbin Watson static (=2.265664), R^2tends to be lower than DW statistic, which in accordance of a good model.

Further model accuracy was presented in Figure3, where ACF and PACF of the error was presented.

MODEL EVALUATION

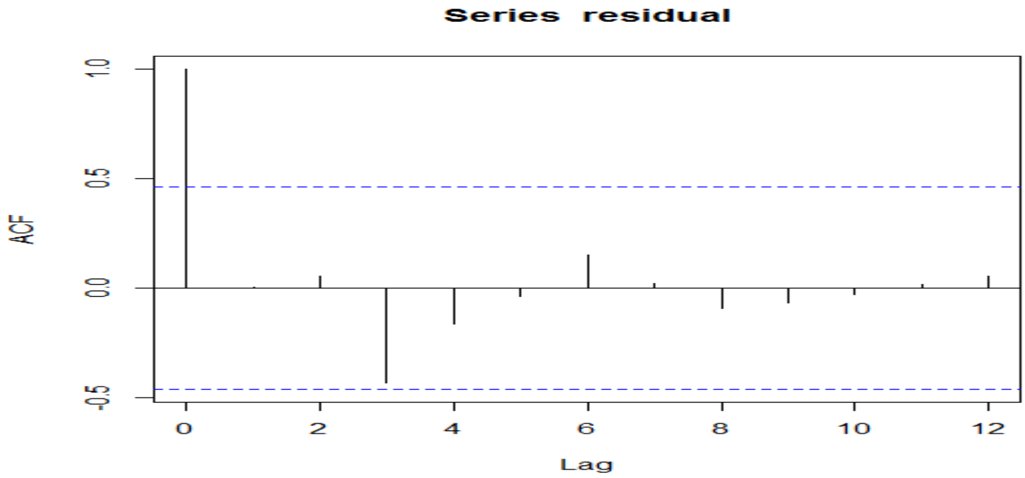

Figure 3a: Autocorrelation of the residuals obtained from ARIMA (1, 2, 0)

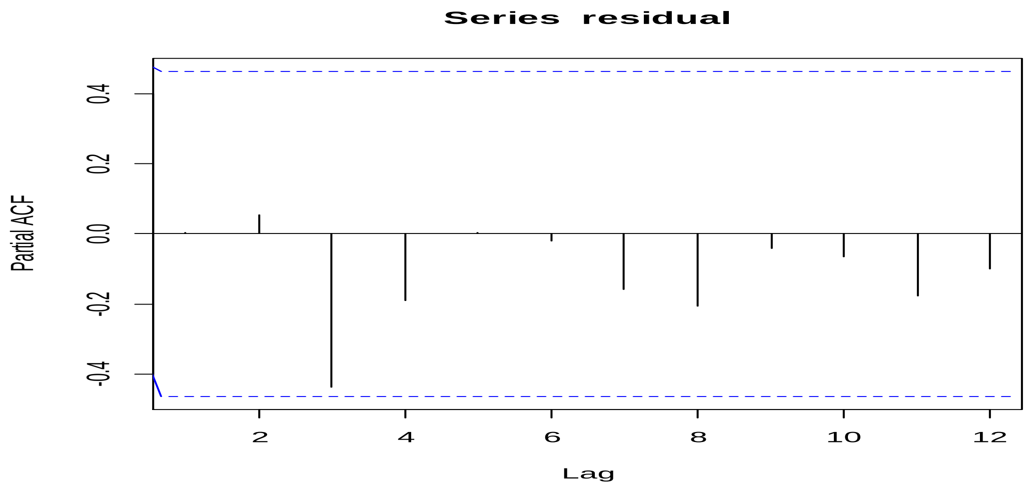

Figure 3b: Partial autocorrelation of the residuals obtained from ARIMA (1, 2, 0)
Figure 3a and Figure 3b showed that the spikes of both the ACF and PACF are within the 95% confidence interval. This means that the ARIMA (1, 2, 0) model is a reasonable fit to the PCI data.

Table 5: Twenty-two years forecast of the PCI series

Year	Forecast	Std. Error	95% Confidence interval
2009	717770.4	15185.12	688008.1 - 747532.7
2010	762884.6	29919.57	704243.3 - 821525.9
2011	810234.9	48178.33	715807.1 - 904662.7
2012	859820.3	68949.91	724681.0 - 994959.6
2013	911641.1	92069.82	731187.6 - 1092094.7
2014	965697.2	117290.93	735811.2 - 1195583.2
2015	1021988.7	144459.25	738853.7 - 1305123.6
2016	1080515.4	173442.12	740575.1 - 1420455.8
2017	1141277.5	204133.17	741183.9 - 1541371.2
2018	1204275.0	236442.11	740856.9 - 1667693.0
2019	1269507.7	270291.84	739745.4 - 1799270.0
2020	1336975.8	305615.26	737980.9 - 1935970.7
2021	1406679.2	342353.34	735679.0 - 2077679.4
2022	1478617.9	380453.65	732942.5 - 2224293.4
2023	1552792.0	419869.14	729863.6 - 2375720.4
2024	1629201.4	460557.38	726525.5 - 2531877.3
2025	1707846.1	502479.75	723003.9 - 2692688.3
2026	1788726.2	545600.97	719367.9 - 2858084.4
2027	1871841.5	589888.61	715681.1 - 3028002.0
2028	1957192.2	635312.74	712002.1 - 3202382.3
2029	2044778.3	681845.64	708385.4 - 3381171.2
2030	2134599.6	729461.49	704881.4 - 3564317.9

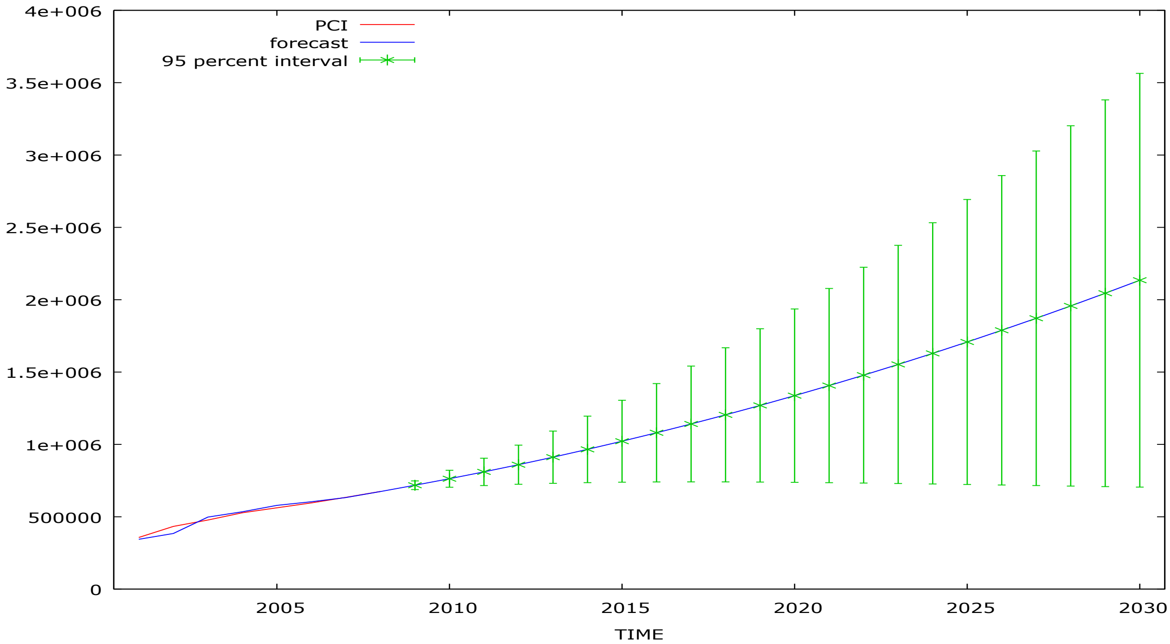

Figure 5: Forecast plot of PCI series from 2009-2030

Table 5 and Figure 5 showed that the forecasted values of PCI will actually go up in the next twenty-two years.

EMPIRICAL RESULTS

The Per Capita Income (PCI) series was first subjected to the unit root test using three different approaches. The approaches are: (1) plotting the time plot, (2) fitting the series to an autoregressive model one (i.e.AR(1)) with drift and check if the coefficient of the lagged one period of the PCI (i.e. PCI_{t-1}) is not only negative but also less than one. (3) conducting ADF tests on the series Figure1 showed that the series exhibited an upward trend. This means that the series is moving with time; which indicates that the series is not stationary at level. To justify the time plot, Table1 reports the AR (1) model used for the ADF test. Tables 2 reported the coefficients of each of the three possible ADF equations at level, first and second their second differences while Figure2 presents ACF and PACF of the PCI series at level form. Summary of various estimates of ARIMA (p, d, q) models were presented in Table3; in which the estimation criteria AIC and BIC were used to further confirm the results reported in Figure2.Table4 presented the estimates of the best model while Figure4 showed that the spikes of both the ACF and PACF of the residuals of the best model are within the 95% confidence interval; which means that the ARIMA (1, 2, 0) model is a reasonable fit to the PCI data. Table5 and Figure5 revealed that the PCI series will continue to increase for the specify period of time

CONCLUSION

The aim of this study is to model and forecast the Per Capita Income (PCI) of Nigeria for the period of 2009 to 2030 through the Box-Jenkins fundamental approach. The modeling cycle was in four stages, the first stage was model identification stage; where the series was not stationary at level form base on the result provided by ADF test and time plot. It was found that the series was stationary at second difference. Base on the selection criteria AIC and BIC, reports showed that the identified ARIMA (1, 2, 0) model was confirmed to reasonably fit the data. The second stage was the model estimation, where the coefficient of parameter of the lagged one period of PCI conforms to the stationary condition (less than one), and finally the third stage was model diagnosis; where the errors derived from the ARIMA (1, 2, 0) model was normally distributed, random (independent of time) and no presence of error serial correlation . An out sample forecast for period of 20 years term was made, and this shows that the PCI will continue to increase for these forecasted time period.Eviews7 and gretlw32 statistical software packages were used for computations in this study.

Further studies can still be conducted on forecasting the PCI alongside other variables like the interest rate, exchange etc. using the multivariate time series approach.

REFERENCES

Bipasha, M and Bani, C (2012). "Forecasting GDPgrowth rates of india". International Journal of Economics and Management Sciences. Vol.1 .No. 9.pp. 52-58

Bello, I.G. (2010). "Statistical analysis of the effect of some economic indicators on per capita income". Unpublished

Box, G.E and Jenkins, G.M. (1976). *"Time Series Analysis, Forecasting and Control";* San Francisco, Holden-Day Inc., USA.

Cejun, L and Chou – Lin, C (2004).Time Series Analysis and Forecast, "National Centre for Statistics and Analysis". DOT HS 809 736

Charles, A., Darne, O. and Hoaran, J.F. (2009). "Convergence of real per capita GDP within COMESA countries: A panel unit root evidence". Annals of Regional Science, Springer Verlag (Germany), 2012, 49 (1), pp.53-71

Fagbohun, A and Adekoya, O.M (2016). "Investment as a determinant per-capita income growth in Nigeria" European Journal of Business and Management. Vol 8. No. 11. Pp. 175-181

Gujarati, D.N. (2004): "Basic Econometrics, 4th Edition", McGraw-hill, New York

William, M and Janes E.R. (1978). Statistics for Management and Economics (Fourth edition). Duxbury Press, Reading, MA 01867