Econometric Analysis of Consumer Price Index on Some Major Economic Indicators

Yusuff, A. Q. & Ajayi, S. O.
Federal College of Forestry,
Ibadan, Nigeria
yusadeq9@gmail.com; +23480 3663 5343

Akanbi, O. A.
Ibadan City Polytechnic,
Ibadan, Nigeria

Amusa, S. O.
Yaba College of Technology,
Yaba, Lagos, Nigeria

Abstract
The study aimed and estimated and/or predicted the mean or average value of the Consumer Price Index on basis of Total Monetary Liabilities, Gross Domestic Product, Imports, Exports and Government Expenditure and also to estimate the strength of linear relationship between pairs of the economic indicators. The data used covered 31years was first reduced by 100,000 to facilitate easy computations. The analysis began with the estimation of the parameters for the linear model, and for heteroscedasticity, multicollinearity and autocorrelation was carried out, with every single one of theme present. The stationarity or otherwise of each of the time was tested for and each turned out to be non-stationery. However, test for cointegration revealed that these time series are cointegrated. Test for the presence of insignificant variable(s) necessitated the removal of GDP from the model which now left the regression with four explanatory variables. The analysis on the resulting model gave similar and even better results (e.g. variance) to that of model including GDP. This led credence to the fact that Gross Domestic Products is a superfluous variable. The new model was also plagued by heteroscedasticity, autocorrelation and multicollinearity. The time series are, however, still cointegrated. The log transformation adopted for the removal of heteroscedasticity gave a different result regards to the partial regression coefficient, especially that of Government Expenditure. This log linear model is however homoscedastic. The first difference method was used to overcome multicollinearity. Although effective, this left a regression model with no intercept term (regression through the origin).The model was subsequently tested for and appropriate corrections were made. The necessary recommendations were also made and we concluded that the Consumer Price Index is influenced by the economic indicators.

Keywords: *Heteroscedasticity, Autocorrelation, Multicollinearity, GDP, Price Index*

INTRODUCTION

The economy of any nation is a prime concern for government, business and the entire citizenry. It determines the social well-being and standard of living of the populace. It is, therefore, important that various economic indicators are available to inform on the progress and state of the economy (Iwedi and Igbanibo, 2017).

One of such indicators is the Consumer Price Index which is an index that shows the price of goods over a certain period of time. In other words, it is used to measure key changes in selling prices. The percentage changes in the Consumer Price Index for two different periods gives the rate of inflation from the previous period to the current period. This shows that any empirical study involving the Consumer Price Index invariably involves the inflation rate (Jennifer, *et al*; 2016).

It is no longer news that every government strives with all the resources at its disposal to keep the rate of inflation as low as possible (Iwedi and Igbanibo, 2017). This, however, is not possible without a corresponding stability in the Consumer Price Index from which the "almighty" inflation rate is calculated.

Econometrics can be defined as the social science in which tools of economic theory, mathematics and statistical inference are applied to the analysis of economic phenomena. Econometrics gives a quantitative or numerical expression to economic theory. This is achieved by starting economic relationships in mathematical form and applying the methods of statistical inference to the measurement of these economic relations as the verification of economic theorems (Koutsoyiannis, 2003; Robert and Daniel, 2006; Gujarati and Duncan, 2010). Econometrics is a product of three different disciplines; Economics, Mathematics and Statistics. As a result, the basic prerequisites for the understanding of econometrics are economic theory, mathematics and Statistics (Chong and Henry, 1986; Koutsoyiannis, 2003).

Econometric models are mathematical expressions of economic theory (Hendry, 1995). It should however be noted that a useful econometric model would emerge from a theoretical economic proposition only when the theoretical proposition can be cast in mathematical form (Koutsoyiannis, 2003).

The econometric analysis of time series focuses on the statistical aspects of model building, with an emphasis on providing an understanding the main ideas and concepts in econometrics rather than presenting a series rigorous proofs (Hendry, 1995; Clements and Hendry,1998a; 2001). It explores the way in which recent advances in time series analysis have affected the development of a theory of dynamic econometrics, set out an integrated approach to the problems of estimation and testing based on the method of maximum likelihood and presents a coherent strategy model selection (Clements and Hendry, 1988a).

This study apart from being a time series data was also used in forecasting non-stationary economic time series data. Economies evolve and are subject to sudden shifts precipitated by legislative changes, economic policy, major discoveries, and political turmoil (Diebold and Mariano, 1995).

Macro econometric models a very imperfect tool for forecasting this highly complicated and changing process. Ignoring these factors leads to a wide discrepancy between theory and practice (Chong and Hendry, 1986).

In the second book on economic forecasting, Michael P. Clements and David F. Hendry (2001) asked why some practices seemed to work empirically despite lack of formal support from theory. After reviewing the conventional approach to economic forecasting, they looked at the implications for causal modelling, present a taxonomy of forecast errors, and delineate the sources of forecast failure. They showed that forecast-period shifts in deterministic factors interacting with model mis-specification, collinearity and inconsistent estimation are the dominant source of systematic failure. They, then, considered various approaches for avoiding systematic forecasting errors, including intercept collections, differencing, co-breaking, and modeling regime shifts; they emphasize the distinction between equilibrium collection (based on cointegration) and error correction (automatically offsetting past errors). Finally, they present three applications to test the implications of their framework. Their results on forecasting have wider implications for the conduct of empirical econometric research, model formulation, the testing of economic hypotheses, and model-based policy analyses.

Total Monetary Liabilities, Gross Domestic Product, Exports, Imports, and Total Government Expenditure are tagged with a particular given price on amount; they are associated with the

Consumer Price Index which will determine its variability's. Therefore, this study examined the effects of some other economic factors on the Consumer Price Index (vis-á-vis inflation rate).

Methodology

The data available for the study is a secondary data which covers a period of thirty-one years. A researcher has to keep in mind always that the results of research are only as good as the quality of the data. Hence, this study has some limitations such as: the non-experimental nature of the data used but there is often no choice but to depend on the available data; there is possibility of observational errors, either of omission or commission; the data used is secondary in nature, there is little to say on the method of collection and economic data are generally available at a highly aggregate level (Micheal *et al*, 2005).

Model

The model considered is:

$$CPI : \beta_0 + \beta_1 TML_i + \beta_2\, IMP_i + \beta_3 GDP_i \; + \; \beta_4 EXP_i + \beta_5 GOVT.EXP_i + \mu_i$$

Where

Consumer Price Index (CPI) = Y
Total Monetary Liabilities (TML) = X_1
Imports (IMP) = X_2
Gross Domestic Product (GDP) = X_3
Export (EXP) = X_4
Government Expenditure (GOVT. EXP) = X_5
Error/Disturbance Term = μ_i

It is true that we seldom can identify or measure all the explanatory variables relevant to a study, but it is imperative to think hard to identify and measure as many of the important ones as possible (David, 2017). When two or more explanatory variables are involved we move from the concept of simple regression to that of multiple regression, in which case there are three or more variables models to handle (Koutsoyiannis, 2003; Micheal *et al*, 2005; Robert and Daniel, 2006; Schroeder and Daniel, 2006).
The two variables regression model often represents a gross oversimplification of real world situations in econometrics studies.
Most often, one dependent variable would be inadequate in explaining or predicting changes in a dependent variable.
The simplest form of multiple regressions is the three variable regression models, which includes one dependent variable and two explanatory variables (Schroederr and Panda, 1986).
There is also the K-variable linear regression model in which case K > 2.

The K-Variable linear Regression Model

The K-variable population regression model involving the dependent variable Y and K-1 explanatory variables X_2, X_3, … X_k may be written as:

$$Y_1 = \beta_1 + \beta_2\, X_{2i} + \beta_3\, X_{3t} + ... + \beta_k X_{ki} + U_i$$

Where

$\beta_i = the\ intercept,\ \beta_2\ to\ \beta_k\ =\ partial\ slope\ coefficients$

U = stochastic disturbance term and
i = i^{th} observation, n being t the size of the population (Koutsoyiannis, 2003)

The population regression model gives the mean or expected value of Y conditional upon the fixed (in repeated sampling) value of X_2, X_3, ….X_k that is $E(Y/X_{2i}, X_{3i}, \ldots X_{ki})$.

The above equation is a shorthand expression for the following set of n simultaneous equation a:

$$Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_k X_{ki} + U_1$$

$$Y_2 = \beta_2 + \beta_2 X_{22} + \beta_3 X_{33} + \ldots + \beta_k X_{ki} + U_2$$

…………………………………………..

$$Y_n = \beta_1 + \beta_2 X_{2n} + \beta_3 X_{3n} + \ldots + \beta_k X_{kn} + U_n$$

The above system of equations in an alternative but more illuminating way as follows:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 & X_{21} & X_{31} & X_{k1} \\ 1 & x_{22} & x_{32} & X_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} & X_{kn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}$$

$$\underset{n\times 1}{Y} \qquad \underset{n\times k}{X} \qquad \underset{k\times i}{\beta} \qquad + \underset{n\times 1}{U}$$

where y = n × 1 column vector of observation on the dependent variable Y;
X = n × k matrix giving n observation on K - 1 variables.
X_2 to X_k, the first column of l's representing the intercept term (This matrix is also known as DATA MATRIX)
β = k × 1 column vector of the unknown parameters $\beta_1, \beta_2, \ldots, \beta_k$
This system is known as the MATRIX REPRESENTATION OF THE GENERAL (K-VARIABLE) LINEAR REGRESSION MODEL.
It is written more compactly as:

$$\underset{n\times 1}{Y} \qquad \underset{n\times k}{X} \qquad \underset{k\times i}{\beta} \qquad + \underset{n\times 1}{U}$$

Or simply as

y = Xβ + U (Agunbiade, 2011)

Data Analysis

$$CPI : \beta_0 + \beta_1 TML_i + \beta_2 IMP_i + \beta_3 GDP_i + \beta_4 EXP_i + \beta_5 GOVT.EXP_i + \mu_i$$

Where CPI = Y
Total monetary liabilities = (TML)
Imports = (IMP)
Gross Domestic Product = (GDP)
Exports = (EXP.)
Government Expenditure = (Govt. Exp.)
To estimate the parameters in the model, we use

$$\hat{\beta} = (X'X)^{-1}\ X'Y$$

$$(X'X) = \begin{bmatrix} N & \sum X_1 & \sum X_2 & \sum X_3 & \sum X_4 & \sum X_5 \\ \sum X_1 & \sum X_1 & \sum X_1X_2 & \sum X_1X_3 & \sum X_1X_4 & \sum X_1X_5 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 & \sum X_2X_3 & \sum X_2X_4 & \sum X_2X_5 \\ \sum X_3 & \sum X_3X_2 & \sum X_3X_2 & \sum X_3^2 & \sum X_3X_4 & \sum X_3X_5 \\ \sum X_4 & \sum X_4X_1 & \sum X_4X_2 & \sum X_4X_3 & \sum X_4^2 & \sum X_4X_5 \\ \sum X_5 & \sum X_5X_1 & \sum X_5X_2 & \sum X_5X_3 & \sum X_5X_4 & \sum X_5^2 \end{bmatrix}$$

INTERPRETATION

From the analysis carried out on the first model, $\hat{\beta}$ yielded the following result; $\hat{\beta} = -1.775523$ intercept term);

$\hat{\beta}_1 = 159.55873(\times 1/100{,}000)$; $\hat{\beta}_2 = 88.92592(\times 1/100{,}000)$;

$\hat{\beta}_3 = 21913\ (\times 1/100{,}000)$; $\hat{\beta}_4 = 72.755195(\times 1/100{,}000)$;

$\hat{\beta}_5 = 129.15988(\times 1/100{,}000)$. These regression coefficients represent the change in Y (CPI) for every unit change in any of the explanation variables.

The coefficient of multiple determination R^2 = (0.9930 15465) shows that the regression model explains almost all the variations in the Consumer Price Index (y).

Having gotten a very R^2 value, it would be expected that the individual regression coefficients would be significant. The t-test showed that Total monetary liabilities (X_1) and Gross Domestic Product (X_3) are not significant. This means that they are individually uninfluential with respect to the Consumer Price Index. The F-test revealed the extent to which the regression (the explanatory variables combined) influenced the Consumer Price Index (F-value is 717).

Tests on various assumptions carried out showed that the model violates a number of the assumption of the linear regression model. The simple correlation coefficient for the pairs of the explanatory variables are very high, confirming the presence of multicollinearity. The Spearman's rank correlation test revealed that all the explanatory variables have systematic relationship with the error, this signaling the existence of heteroscedasticity. The Durbin-Watson test showed that there is negative autocorrelation in the data. Stationarity test revealed that the time series are not individually nonstationary (i.e. their mean and variance are not time invariant). Hence, analysis especially predictions about them would be dubious. The good news, however, is the linear combination of these time series (CPI, TML, IMPORTS, GDP, EXPORT and GOVT. EXP.) is stationary. This means they are cointegrated and whatever result we have from the regression would not be spurious.

Test for the presence of unnecessary variable(s) revealed that the GDP is not relevant (at least for this data). With the removal of Gross Domestic Product, the resulting model was now run

$$\left(CPI = \beta_0 + \beta_1 TML + \beta_3 IMPORTS_i + \beta_4 EXPORTS_i + \beta_5 GOVT.EXPEND_i + U_i\right)$$

Running the regression model above gave similar results to that of the previous model (R^2 = 0.0993080205, R^2 = 0.992). However, the variance of the error is smaller than that of the original result, multicollinearity which was very severe in the first model is now moderate, condition index is 15.021 compare with CI for the first model 45.839. The F-test gave an even higher F-value, 932.834246 (compare with 717.0702264 obtained in the first model).

The tests of assumptions however revealed that this model violates the assumption of no multicollinearity, homoscedasticity and independent residuals. The CRWD test carried out showed however that these time series are cointegrated.

The log-linear model (log transformation) was adopted to overcome heteroscedasticity and the first difference method employed to combat multicollinearity. In both cases, the measures turned out to be adequate (this is evident from Spearman's correlation coefficient and simple correlation coefficients computed after these measures had been adopted).

It must be noted here that the log-linear model gave some differing results, the most notable being the negative sign of the partial regressing coefficient for Government Expenditure. The regression coefficient in the log-linear model represents the elasticity of with respect to the corresponding explanatory variables .This should not be confused with β in the linear model which represents the net change in Y with respect to a unit change in the explanatory variable.

First differencing of the log-linear model to eliminate multicollinearity led to regression through the origin (i.e. a regression with no intercept term). A notable feature of this kind of

regression is that the sum of residuals from this regression need not be zero and R^2 computed is not the conventional coefficient of determination (since the raw sum of squares is used, hence, the appellation raw R^2).
The model was then corrected for autocorrelation and the runs test made use of to confirm if the correction was effective. The model were left with is devoid of multicollinearity, heteroscedasticity and autocorrelation, but, this is only half the battle since from Raw R^2 it is clear the resultant model will have very little (or no) predictive power.

Recommendations

The followings are recommended based on the analysis and subsequent discoveries:

- Close attention should be paid to the Total Monetary Liabilities, Imports Exports, Government Expenditure since the analysis revealed they account for almost all the variations in the Consumer Price Index.
- Though, the Gross Domestic Product turned out to be irrelevant in this analysis, it (may) significantly influence each of the explanatory variables, this is evident from the reduction in the level of multicollinearity upon its removal.
- Extensive research still needs to be carried out on the Consumer Price Index since this index determines the cost of living vis-â-vis, the standard of living of the citizenry and also the inflation rate.
- And lastly, determine which of the models (linear or log-linear) is the most appropriate for the expression of the relationship between the Consumer Price Index and other economic indicators.

This is essential, though this study assumed that the relationship is linear. To choose between the models, use the MWD test, as follows:

(i) Estimate the linear model and obtain the estimated Y values call them Yf(i.e. $\hat{Y}$).
(ii) Estimate the log-linear model and obtain the estimated In Y values call then in F (i.e. ln Y).
(iii) Obtain Z_i, (In Yf – lnf)
(iv) Regress Y on X's and Z_1 obtained in step III, reject H_0 if the coefficient of Z_1 is statistically significant by the usual t-test.
(v) Obtain Z_Q = (antilog of Inf Yf).
(vi) Regress log of Y on the logs of X's and Z_2 in statistically significant by the usual test.

The hypothesis is as follows:
H_0: Linear model; Y is a linear function of regressions, the X's.
H_1: Log-linear model; In Y in a linear function of logs of regressors, the logs of X's.

Conclusion

Based on the analysis carried out, the Consumer Price Index is influenced greatly by the Economic indicators; Total Monetary Liabilities, Imports, Exports and Government Expenditure.

REFERENCES

Agunbiade, D.A. (2011): Parameters estimation of simultaneous equation Model in the Presence of Multicollinearity. Bull. Cal. Math. Soc. 103, (6) 489 – 500.
Clements, M.P. and Hendry, D.F. (2001): *Forecasting Non-Stationary Economic Time Series.* The MIT Press, London.

Clements, M.P. and Hendry, D.F. (1998a): Forecasting Economic Processes. International Journal of Forecasting, 14, pp 111-131.

Chong, Y.Y. and Hendry, D.F. (1986): Econometric Evaluation of Linear Macro-Economic Models. Review of Economic Studies, 53, pp 671-690.

David J. Olive (2017): *Linear Regression*. Springer International Publishing AG, Switzerland

Diebold, F.X. and Mariano, R. S. (1995): Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, pp 253-263

Gujarati N. Damodar and Dawn, C Porter (2010): *Essentials of Econometrics*. 4th Edition McGraw-Hill, Irwin.

Hendry, D.F. (1995): *Dynamic Econometrics*, Oxford. Oxford University Press.

Iwedi Marshal and Igbanibo Dumini Solomon (2017): Nigeria Economy and the Politics of Recession: A critique. Journal of Advances in Economics and Finances. Vol.2 (4)

Jennifer, L. Castle, Michael, P. Clements and David, F. Hendry (2016): An Overview of Forecasting Facing Breaks. Journals of Business Cycle Research, 12:3-23.

Koutsoyiannis, A. (2003): *Theory of Econometrics: An Introductory Exposition of Econometric Methods:* 2nd Ed. Palgrave Publications. N.Y. 10010.

Michael, H. Kutneri, Christopher, J. Nachtsheim, John Neter and Willian, Li (2005): *Applied Linear Statistical models.* 5th Edition. McGraw-Hill/Irwin, NY, 10020.

Robert, S.P. and Daniel, L.R. (2006): *Econometric Model and Economic Forecasts.* 4th Edition. Sage Publications U.K.

Schroederr, L.D. and Pauda, E.S. (1986): U*nderstanding Regression Analysis: An Introducing Guide*. Thousand Oaks, C.A. Sage Publications Series. U.K. Quantitative Applications in the Social Sciences. No. 57.