ON THE EFFECTS OF NON-ROBUSTNESS IN THE SPURIOUS REGRESSION MODEL*

D.K. SHANGODOYIN[1], A.W. ADEOGUN[2] and T.O. OLATAYO[3]

1. Department of Statistics, University of Ibadan, Ibadan, Nigeria.
2. Department of Statistics, Yaba College of Technology, Lagos State, Nigeria.
3. Department of Mathematical Sciences, Olabisi Onabanjo University, Ago-Iwoye , Nigeria.

(*Received*: August 2007; *Accepted*: March 2008)

Abstract

The effects of violation of basic assumptions on spurious regression with Time Series data was carried out. The requirements of establishing non-robustness of a Time Series regression model, identification of spurious regression through formal process were illustrated with foreign exchange of Nigeria, United state of America and Great Britain.

It was found that violation of these assumptions play an important role in determining if a spurious regression emanates from the statistically related model for reliable predictive purposes.

Key words: Spurious regression, non robustness and foreign exchange.

1. Introduction

It is well known that, if auto correlated errors in time series regression equations are ignored, problems arises involving inefficient parameter estimates, invalid significance tests and sub-optimal results when the fitted equations are used to derive forecasts,

Newbold and Davies (1978). Spurious regression have a long history in statistics, dating back at least to Yule (1926). Yet another is econometric example of alchemy reported by Hendry (1980) between the price level and cumulative rainfall in the U.K.

The latter "relation" proved resilient to many econometric diagnostic tests and was humorously advanced by its author as a new "theory" of inflation.

Therefore, spurious relationship, referring to a correlation induced between two variables that are casually related but both dependent on other common variables. This is accomplished either by improving the function of time as a repressors or by subtracting a function of time from all series used.

Granger and Newbold (1974) shows that this phenomenon occurs when independent random walks are regressed on one another and warned that spurious relationship may be formed between the levels of trending time series that are actually independent. Philip (1998) and Wayne (2003) developed tools for understanding and analyzing spurious regression in modeling and forecasting financial theories.

Nevertheless, the above named authors and many others had employed ordinary least square (OLS) method to estimate the parameters of the equation:

$$Y_t = \alpha + \beta X_t + \varepsilon_t \qquad \text{eqn (1)}$$

Where $\mathring{a}_t$ is a white noise, and t =1, 2,.................,n.

The regression model in eqn (1) is predictive; as such it is vital to measure the error inherent in the analysis processes as it affect the variable involved.

In this wise the error terms are indeed very basic in the stochastic assumptions made about specified regression model.

Among such assumptions are that the error terms ($\mathring{a}_t$);

i. Are random variables
ii. Are normally distributed with mean zero i.e. $\mathring{a}_t \approx N(0, \ddot{a}^2)$
iii. Posses same variance (homoscedastic property) for all X_t
iv. Exhibit no serial correlation (absence of autocorrelation).

+ *corresponding author* (email:)

* Presented in part at the First Faculty of Science Conference, Obafemi Awolowo University, Ile-Ife, July 3-5, 2007.

When assumptions are violated, estimated regression coefficient suffers statistically in one form or the other. There is difficulties in estimating their variance and intervals in the case of homoscedasticity property violation.

When normality assumption is not satisfied it leads to the inability to perform test of significance of estimates which prevents an assessment of statistical reliability of the estimates. Violation of nonserial correlation of errors assumption affects the values and standard errors of estimate parameters; they tend not to posses statistical bias and any forecast made from the OLS estimates are generally inefficient.

A non robust model brakes down completely, when any of the assumptions made about its adoption is violated. In regression analysis, linear models eqn. (1), especially are know to exhibit the properties of non robustness when either of the stochastic or non stochastic assumptions are violated.

In relation to this, one of the main objectives of performing a regression analysis apart from forecasting purpose is to explain the nature of occurrence of a dependant variable, by explanatory variables. When the variables are time series generated processes, it become more involving in that a high degree measure of "goodness of fit" does not really imply that all is well.

The purpose of this paper is to develop some alternative methods of analyzing and estimating the parameters of coefficient in the regression equation, for efficient and reliable predictive purposes.

The methods in comparison with OLS were analyzed empirically with exchange rate data of U.S dollar to Nigerian Naira and U.S dollar to Great Britain pound sterling and therefore the analysis of model when the assumptions are violated could lead to spurious regression and any forecast from such model could be mis-leading.

Theoretically, spurious regression occurs when the variables being regressed are integrated variables of order one i.e. I (1), in which case they are not stationary, but stationary if difference once, Wen-jen Tsay (1999). The measure of "goodness of fit" tends to be statistically significant but economically insignificant through poor values estimated for the regression coefficient or vise versa.

Almost all economic variables are I (1), hence expected to lead to spurious regression.

Exchange range data of financial currency is suitable financial economic indicator and measure of the performance of the nation's economy. This variable is not entirely immune to spurious regression if not properly analyzed since the explanatory variable(s) may be statistically estimating the independent variable significantly, but the regression coefficient(s) remain insignificant, thus in this case rendering the model meaningless.

It is not all spurious regression that is meaningless or would run out to be false model.

When variables generated by an I (1).process are regressed, the regression spurious and the model specified may be true or false depending on further variations such as the analysis of ex-post forecast. A true model that lead to spurious regression, tend to fit a very good forecast, while a false model bring bad forecast. In addition since non stationary nature of the variables contribute significantly to spurious results; the first deference of the variables when regressed makes a true model to survive while a false model disintegrates.

2. Materials and Methods

Let the regression OLS model for estimating the exchange rate of US$ to Nigerian ₦ be specified as below: equivalent to eqn. (1):

$$\text{InDN} = \eta_0 + \eta_1 InDP_t + \varepsilon_t \qquad \text{eqn. (2)}$$

Where DN_t = the exchange rate of IUS$ to Nigerian ₦

DP_t = the exchange rate of IUS$ to Great Britain pounds

ε_t = residual or disturbance terms

The data collected for analysis is for daily exchange rate, of which 99 are used for analysis of various regression models specified, while 12 data points are used for expost forecast analysis.

OLS Estimation Process

Using SPSS application package, the OLS estimation of the regression coefficients gives the result below.

InDN=4.023-2.111InDP_t

From the results the following can be inferred:

Results and Deduction form OLS Analysis

1) The value η_1 = -2.111 indicates that economically the model is not significant it implies that for every 1 Dollar exchange for sterling pound, ₦ 2.111 is lost to a Dollar.
2) Also the adjust R^2 value of 0.236 shows that the exchanged for sterling rate of Dollar to sterling pound poorly contributes to the exchange of rate of dollar to Naira using the model.

3) Testing for Normality of Errors
   Using the Jarque-Bera Test

   $H_0 : e_t$ are $\approx N$

   $H_1 : e_t$ are not $\approx N$

   J-B value =24.96

   $X^2_{2,0.05} = 5.99$ Reject H_0 if J-B $\succ$ 5.99

   H_0 is rejected, hence the errors does not follows a normal distribution.

   From the Normal Probability Plot below, the deviation fom normality of the errors is confirmed.
4) Testing for Error Homoskedasticity
   Using the Breusch-Pagan test procedure, the hypothesis tested are:

   $H_0 : e_t$ are homoskedasticity

   $H_1 : e_t$ are not homoskedasticity

   B-P value =13.33

   $X^2_{1,0.05} = 3.84$ Reject H_0 if B-P $\succ$ 3.84

   H_0 is rejected, hence the errors are homoskedastic.
5) <u>Testing for Autocorrelation:</u>
   The hypothesis to be tested is

   H_0: There is no autocorrelation of residuals

   H_1: There is autocorrelation of residuals

The Durbin-Watson statistic is 0.442 from the procedure involved in the performance of the autocorrelation test, from the table at 5% significance level, H_0 is rejected, this implies that there is a serial correlation of errors problem.

From the above result this three basic stochastic assumptions are shown to be violated in the estimation process and the expected consequence are as numerated earlier. Next we shall explore the effects of these violations (if any) on the model, if it can be ascertained that the analysis of the model will indeed lead to spurious regression. The process requires examining the specified model with various relevant properties to identify a spurious regression situation.

Preliminary Analysis of Data Structure

The chart shows the time-plot of natural log transformation of the exchange rates data for each of the variables in the specified model.

The chart shows a near stationary trend movement for DN_t, while for DP_t a pronounced upward trend movement is indicated. The data series for the two variables can be said to be non-stationary though at varying degree.

Result and deduction from OLS model analysis

In addition to result, the following can be observed.

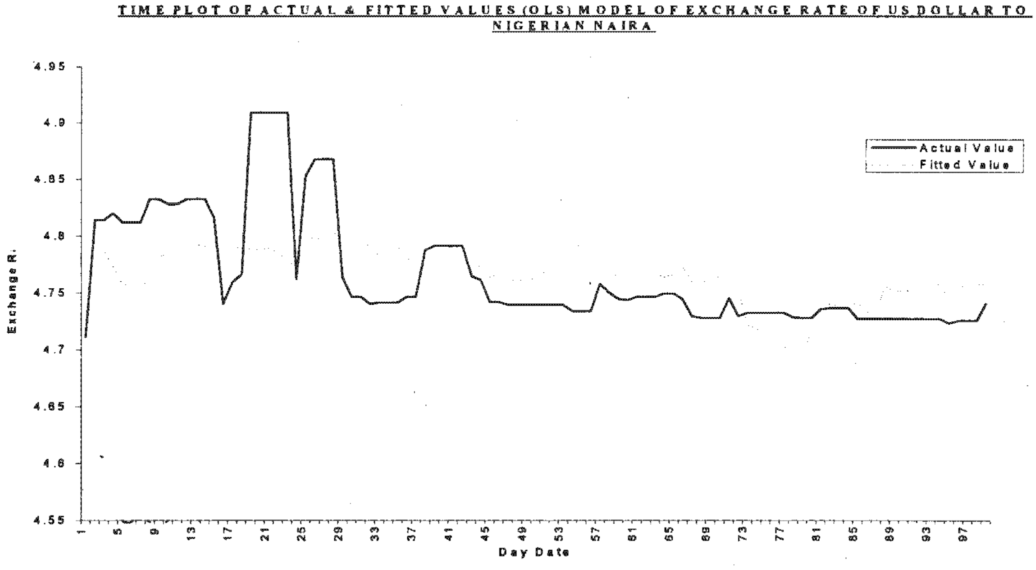
(1) The time-plot of actual and fitted (OLS) model reveals lack of good fit, (The time plot of residual values) actually appears to be serially correlated with large swing around the axis.

(2) The Ex-post Forecast
    The chart indicates in general a bad forecast resulting from estimations based on the specified model.

The identified problem of serial correlation of errors and bad Ex-post Forecasts points to the need for

a) The removal of serial correlation effect, which may have influenced the forecast result.
b) Further exploration of the data analysis technique to improve on its estimation result, at least economically.

The process of improvement on the analysis is done by specifying another model.



The AR (1) Model Specification

The attempt to solve the problem identified in the OLS estimation of the earlier regression model requires the specification of an AR (1) model,
Box and Jenkins (1970), as stated below:

$$\text{InDN}_t = \eta_0 + \eta_1 InDP_t + \rho\upsilon_{t-1} + \varepsilon_t$$

Where $u_t = \rho\upsilon_{t-1} + \varepsilon_t$ is the residual of the OLS model.

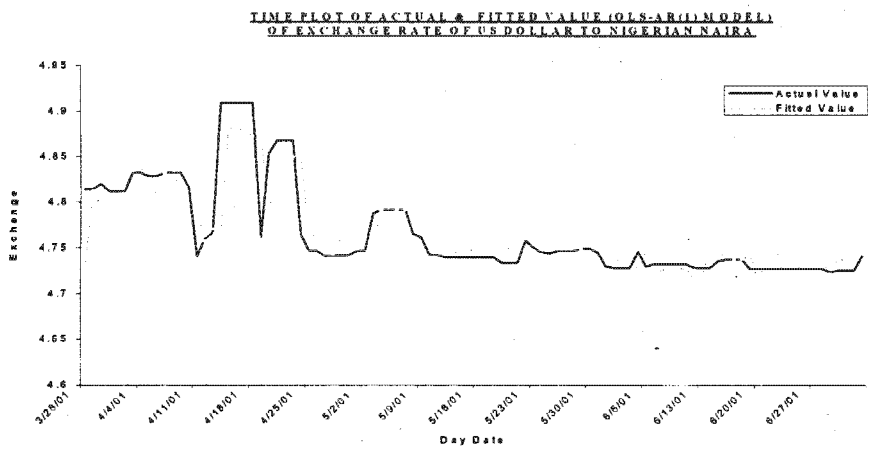
ε_t Is a white noise process.

Result and deduction from AR (1) model analysis

The estimated model is:

$$InDN_t = 4.030 - 2.092 InDP_t + 0.765\upsilon_{t-1} + \varepsilon_t$$

1) The estimated values of the regression coefficient $\eta_0 and \eta_1$ are obtained with very little change. the regression coefficient of the AR residual term is significant enough to stabilize the estimation process.
2) The adjusted R^2 value improved considerably to 0.698, $DP_t and \upsilon_{t-i}$ Contribute significantly to the estimation of DN_t, hence The model is a better fit.
3) The Durbin – Watson value of 1.870; implies that H_0 is not rejected. The serial correlations of error term have been removed.
4) The chart reflects a better model fit. The residual time plot shows less swing around the time axis, further confirming removal of residual autocorrelation.



Fig(2)

(5) To further confirm the removal of residual autocorrelation, AR (1) diagnostic check is performed. Given that an estimated autocorrelation coefficient to be significantly different from zero, its absolute value should exceed 2/?N, where N is the number of observation utilized in the analysis. Evaluating to be 0.201 and none of the estimated autocorrelation co-efficient and partial autocorrelation coefficients is greater than 0.2 in absolute value and the correlograms shows less swing around to time axis.

(6) Ex-post Forecast analysis

I. The scatter plot is less dispersed . The time plot of the actual and forecasted values of the model shows better forecast patter and the forecast residuals are better spread to even out.

II. The improvement in the forecast is determined by performing the Mean Absolute percentage error (MAPE) analysis to compare between OLS and AR(1) model forecast estimates. The expression for determining the MAPE is given by

$$\text{MAPE}=\left[\frac{1}{k}\sum_{j=1}^{k}\frac{InDN_{(7/4/0)+J}-In\hat{DN}_{(7/4/0)+J}}{InDN_{(7/4/0)+J}}\right]\times 100\%.$$

Where k = no data point used for ex-post forecast

OLS=28.05% and AR(1)=15.35% Estimation improvement is 12.70%,this is quite encouraging. But the magnitude of 15.35% is still regarded very high, which is not yet satisfactory as a forecast error.

Summary of inference AR (1)

In general , the AR(1) estimation model provided a better goodness of fit measure , removed the serial correlation and some worth improves the forecast , however the MAPE measure of the AR(1) model, suggest that the estimation process can still be improved upon by specifying another model.

Partial Adjustment Model (PAM) Specification

Following the quest for estimation improvement, a simple partial adjustment model is specified as below:

$$InDN_T = \eta_0 + \varphi InDN_{T-1} + \eta_1 InDP_T + \varepsilon_T$$

Result and deduction from PAM Analysis

The estimated mode is:

$$InDN_T = 0.923 + 0.762DN_{T-1} - 0.599InDP_T$$

1) The estimated value of the regression coefficient η_1 improved although still economically insignificant.
2) The adjusted R^2 value further improved 0.703, DN_{T-1}, DP_T contributes significantly to the estimation of DN_T , Hence the model is a better fit.
3) The Durbin-Watson, jarque-Bera and Breusch-pagan test values gives positive estimation measures.
4) The residual time plot shows much tapered swing about the time axis, indicating residual autocorrelation.

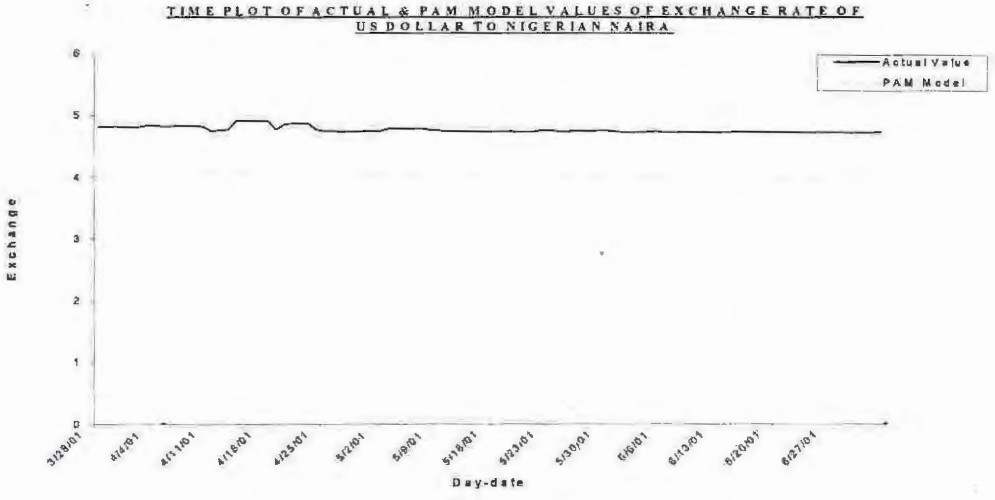


Fig. (3)

Ex- post forecast Analysis

I. The scatter plot is quite dispersed indicating non linear relationship. The time plots of the actual and forecasted values of the model shows a bad forecast pattern and the forecast residuals consistently remain negative.

II. The MAPE analysis to compare between OLS, AR (1) and PAM forecast estimate is given as OLS=28.05%,AR(1)=15.35% and PAM=27.52%.The estimation improvements are 12.70% and 0.52% respectively.

From the result, the PAM specification provides very little improvement in the forecast estimate with a value less than 1%.

Summary of Inference (PAM)

The PAM specification improved the statistical estimation process very well, relatively large values of t-statistics and adjusted R^2 with normally distributed error terms, equal variance and no problem of serial correlation, however the forecast were as bad as that of the OLS model specification.

Table 1: Result of the analyzed model specifications:

Xteristics/model	OLS	AR(1)	PAM
Goodness of fit	Bad	Good	Better
Serial correlation	Present	Solved	Absent
Forecast	Bad	Good	Bad

Table 2: Economic and statistical relevance of trade-off between the models analyzed.

Mode	OLS	AR(1)	PAM
Economic (model parameter)	Meaningless	Meaningless	Improved significantly
Statistical (Model forecast)	Poor	Improved significantly	Poor

Tables 1 and 2 are summary of result from the three specified models.

3. Summary and Conclusion

The two models AR (1) and PAM are specified to improve the OLS model, but they give contrasting economic and statistical improvements. The AR (1) provides improvement which supports statistical theory of regression analysis, while PAM improved the economic requirement; hence the research is put in a trade off condition. Should he favor improvement along statistical theory or economic theory?

Let us pause to consider the suitability or otherwise of variables in the models analyzed. So far the attempt is estimating exchange rate of US$ to Nigeria ₦ with that of exchange rate of US$ to Great Britain pound, or put in another form, we are trying to explain a major financial economic factor of Nigeria relative to that of two entirely different economies i.e. that of the US and Great Britain. This is an "antipodal" theory of financial economy. The explanation of the influence of US$ on Nigeria economy with that of the influence of US$ on Great Britain economy is indeed economically meaningless, though may be statistically significant.

In this context the standard statistical criteria for evaluating an estimation equation does not ensure meaningful relationship, although they may be statistically convincing, this is exhibited with the goodness of fit which is getting better with the different models, despite lack of true economic relationship between variables in the regression equations.

We can thus come to the conclusion that instead of trading off between statistical and economic theories, the true relationship between variables is not ascertain by the goodness of fit and goodness of forecast measures as had been demonstrated, this is an indication of spurious regressions. Therefore, the result presented in this paper suggest that inference and forecast in regressions involving economic time series can be greatly affected by error structure assumed. We would hope that economics analyst will consider a wider range of possible modeling structures than has been the case in the past.

REFERENCES

Box, G.E.P. and Jenkins, G.M., 1970. Time series Analysis, forecasting and control; San Francisco: Holden-Day.

Granger, C.W.J. and Newbold, P., 1974. Spurious Regression in Econometrics. Journal of Econometrics, 2, 111-120.

Ferson, W.E., Sarkissian, S. and Simin, T.T., 2003. Spurious Regressions in financial Economics? The Journal of Finance, 58(4), 1393-1412.

Hendry, D.F., 1980. Econometrics: Alchemy or science? Economica, 47, 387-406.

Newbold, P. and Davies, N., 1978. Error mis-specification and spurious Regressions. International Economic Review, 19(2), 513-519.

Newbold, P. and Granger, C.W.J., 1974. Experience with forecasting univariate time series and the combination of Forecasts. Journal of the Royal statistical society, A 137, 131-146.

Philip, P.C.B. 1998. New Tools for Understanding spurious Regressions. Econometrica, 66(6), 1299-1325.

Tsay, W., 1999. Spurious Regression between I(1) processes with Infinite Variance Error. Econometric theory, 15(4), 622-628.

Yule, G.U., 1926. Why do we sometimes get nonsense correlation between Time series? A study in sampling and the nature of Time series; journal of the Royal statistical society, 89,

1-69.