International Journal of Research in Business, Management & Accounting
ISSN (Online): 2455-6114
Volume 02 Issue 01 January 2016
Available on-line at: https://gnpublication.org/index.php/bma
DOI: *https://doi.org/10.53555/bma.v3i3.1718* Received: *02-Jan-2016* | Accepted: *12-Jan-2016*

THE FARRAR-GLAUBAR APPROACH IN TESTING FOR MULTICOLLINEARITY IN ECONOMIC DATA

AKINNIYI, ALABA JOSEPH (PhD)
Department of Business Administration & Management Rufus Giwa Polytechnic, Owo Ondo State, Nigeria
SANNI, ENEJI ADEMOH
Department of Mathematics & Statistics Rufus Giwa Polytechnic, Owo Ondo State, Nigeria

Abstract:

This research aims at determining the presence of Multicollinearity in a function using farrar-glaubar test approach. In most economic data, there is the presence of Multicollinearity but the severity varies. The degree of this multicollinearity may vary from function to function. However, Farrar-Glaubar test is used to detect the presence and severity of Multicollinearity, location of Multicollinearity, and the pattern of Multicollinearity in a function. How to correct the effect of Multicollinearity was also covered this research. After analyses were done on the collected data, we realized that, Multicollinearity is most pronounced in Economic data.

Keywords:

multicollinearity, farrar-glaubar, economic data, variables.

*Correspondence Author:

Email: -------@yahoo.com (AKINNIYI, ALABA JOSEPH)

INTRODUCTION

Data is simply scientific term for facts, figures, information and measurements. Data, therefore, include the number of people who gain admission into universities each year, number of yam produced by each farmer in a year, etc. In other words, data can come from sector such as Agriculture, Business, Industrial etc. Economic data, on the other hand, are the data obtained from business transactions which can be in form of buying and selling of goods and services.

Econometrics is the application of mathematics and statistical method to the analysis of economic data as mathematical models help us to structure our perception about the forces generating the data we want to analyze, while statistical method helps to summarize the data, estimate the parameters of our models and interpret the strength of the evidence for various hypothesis that we wish to examine. The provided data affect our idea about the appropriateness of the original model and may result in significant revisions of such models.

There is, thus, a continuous interplay in economics between mathematical theoretical modelling of economic behavior, data collection, data summarizing, model fitting and model evaluation. Theory suggests data to be sought and examined; data availability suggests new theoretical questions and stimulates the development of new statistical method. The examination of data in the light of theory lend often to new interpretations and sometimes to question about its quality or relevance and to attempt to collect new and different data.

Collinearity refers to the existence of a single linear relationship. In other words, multicollinearity is simply the existence of multiple or several relationships in a linear relationship. Multicollinearity is not a condition that either exists or does not exist in economic function, but rather a phenomenon inherent in most relationship due to the nature of economic magnitudes. There is conclusive evidence concerning the degree of collinearity which, if present, will affect seriously the parameter estimate intuitively, when any two explanatory variables are changing in nearly the same way, it becomes extremely difficult to establish the influence of each one regressor on y separately (Armstong, 2012).

Aim

This research aims at determining the presence of multicollinearity in a function using farrar- glauber test approach and the specific objectives are:

i. Test for the location of Multicollinearity.
ii. Test for the pattern of Multicollinearity.

GENERAL LITERATURE

Correlation and Regression Analysis

There are various techniques of measuring the existence correlation between two variables. The most used techniques are correlation and regression analysis.

Correlation is the degree of relationships existing between two or more variables; the degree of relationship between two variables is called simple correlation. The primary objective investigating the correlation between two variables is to determine whether there is any causal connection between them (Becker, 1998).

Correlation Coefficient

The parameter ρi is called the population correlation coefficient and it measures the strength of the linear relationship between x and y. The statistic r measures the strength of relationship between the sample observations of two variables, it is also called sample estimate (Waegeman, 2009).

The sample correlation coefficient is defined by the formula:

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][\sum y^2 - (\sum y)^2]}}$$

Where Y is the dependent variable and X is the independent variable.
The value of r is always between -1 and +1 no matter the unit of X and Y. A value of r near or equal to zero implies little or no linear relationship. The closer r is to 1 or -1. The stronger the linear relationship between Y and X.
Test of significance for the sample coefficient

$\rho=0$ i.e. $r \sim N(0, \sigma_r)$

where $\sigma r = \sqrt{\frac{1-r^2}{n-2}}$

Test statistics t is estimated by $t = r\sqrt{\frac{n-2}{1-r^2}}$

With n-2 degree of freedom and this is compared with the appropriate theoretical value of t base on a level of significance.

Simple Regression Analysis

Least Square Method
If two variables X and Y are linearly related, their relationship can be expressed by the following simple linear. $Y = \alpha + \beta x + ei$
Where α and β are parameters called the regression constant and the regression coefficient respectively *ei* is a random variable with mean of zero and variance S^2 (Tofellis, 2009).

Multiple Regression Estimation

Multiple regression analysis is a process whereby a relationship is established between two or more variables in term of an equation so that, given the value of one variable, the value of the other variable can be predicted (Oloyede 2012).
It is an attempt to determine approximately the value of the population parameters in the model of Awel (2014). Multiple regressions are concerned with obtaining a mathematical equation which describes the relationship among three or more variables. Then, the equation obtained can be used for comparism or purpose of estimation. Dependent variable is a variable that occur as a result of consequence of other variable called Independent variable.
The multicollinearity effect is observed in a function when all or some of the explanatory variable high correlated with each other than they are related to the dependent variable.

Multicollinearity

Cressie, (1996) viewed that; a crucial condition for the application of least squares is that the explanatory variables are not perfectly linearly correlated($rx_i rx_j \neq 1$). The term multicollinearity is used to denote the presence of linear relationship (or near linear
relationship) among explanatory variables. If the explanatory variables are perfectly linearly corrected, that is if the correlation coefficient for these variables is equal to unity, the parameters become indeterminate; it is

impossible to obtain numerical values for each parameter separately and the method of least square breaks down. At the other extreme, if the explanatory variables are not inter-correlated at all (that is if the correlation coefficient for these variables is not equal to zero), the variables are called orthogonal (variables whose covariance is zero: $\sum x_i x_j / n = 0$)and there are no problems concerning the estimates of the coefficients, at least so far as multicollinearity is concerned. Actually, in the case of orthogonal x's, there is no need to perform a multiple regression analysis. Each parameter can be estimated by a simple regression of y on the corresponding regressor: Y=f(x).

In practice, neither of the above extreme cases (of orthogonal X's or perfect collinear x's) is often met. In most cases, there is some degree of inter-correlation among the explanatory variables, due to the interdependence of many economic magnitudes over time. In this event, the simple correlation coefficient for each pair of explanatory variables will have a value between zero and unit, the multicollinearity problems may impair the accuracy and stability of the parameters estimates but the exact effects of colinearity have not yet been theoretically established.

Multicollinearity is not a condition that either exists or does not exist in economic function, but rather a phenomenon inherent in most relationship due to the nature of economic magnitudes. There is no conclusive evidence concerning the degree of co linearity which, if present, will affect seriously the parameter estimate. Intuitively, when any two explanatory variables are changing in nearly the same way, it becomes extremely difficult to establish the influence of each one regressor on y separately. For example, assume that the consumption expenditure of an individual depends on his income and liquid assets. If over a period of time, income and the liquid assets change by the same proportion, the influence on consumption of one of these variables may be erroneously attributed to the other. The effects of these variables on consumption cannot be sensibly investigated, due to their high inter-correlation (Aldrich, 2005).

The Nature of Multicollinearity

The term multicollinearity is due to Ragnar (2013) originally, it means the existence of a "perfect", or exact, linear relationship among some or all the explanatory variables of a regression model. Strictly speaking, multicollinearity refers to the existence of more than one exact linear relationship, and co linearity refers to the existence of a single linear relationship involving explanatory variable X_1, X_2 X_k(where X1=1 for all observations to allow for
the intercept term), an exact linear relationship is said to exist if the following condition is
satisfied:

$\lambda_1 X_1 + \lambda_2 X_2 + \quad + \lambda_k X_k = 0$

Where $\lambda 1 \lambda 2 \ldots \lambda$k are constants such that not all of them are zero simultaneously.

Today, however, the term multicollinearity is used in a broader sense to include the case of perfect multicollinearity. As it is, in the case where the X variables are inter- correlated but not perfectly so, as follows:

$\lambda_1 X_1 + \lambda_2 X_2 + \quad + \lambda_k X_k = 0$

Where Vi is a stochastic error term.

The preceding algebraic approach to multicollinearity can be portrayed succinctly by the Ballentine (David, 2005). Three variables say Y1 X2 and X3, represent the variations in Y (the dependent variable) and X2 and X3 (the explanatory variables). The degree of colinearity can be measured by the extent of the overlap (shaded area) of the X2 and X3 circles. In the diagram below − (a), there is no overlap between X2 and X3, and hence no colinearity. In (b) through (c), there is a "low' to "high" degree of colinearity - the greater the overlap between X2 and X3 (i.e. the larger the shaded area), the higher the degree of colinearity. In the extreme, if X2 and X3 were to overlap completely (or if X2 were completely inside X3, or vice versa), colinearity would be perfect.

Plausibility of the Assumption

Strictly speaking, the assumption concerning multicollinearity unit is that the variables are not perfectly linearly correlated and it is easily met in practice, because it is very rare for any two variables to exactly inter-correlated in a linear form. However, the estimates of least squares may be seriously affected with a less than perfect intercorrelation between the explanatory variable (Fotheringham, 2002).

Reasons for Multicollinearity

Firstly, there is a tendency of economic variable to move together over time. Economic magnitudes are influenced by the same factors and in consequence, once these determining factors become operative, the economic variables show the same broad pattern or behavior over time. For example, in periods of booms rapid economic growth, the basic economic magnitudes grow, although some tend to lag behind others. Thus income, consumption, savings, investment, prices, employment, tends to rise in periods of economic expansion and decrease in periods of recession. Growth and tend factors in some series are the most serious cause of multicollinearity.

Secondly, the use of lagged values of some explanatory variables as separate independent factors in the relationship. Models with distributed lags have given satisfactory results in many fields of applied econometrics, and their use is expanding fast. For example, in consumption function. It has become customary to include among the explanatory variables past as well as the present levels of income. Similarly, in investment function, distributed lags concerning past levels of economic activity are introduced as separate explanatory variables. Naturally, the successive values of certain variable are inter-correlated, for example, income in the current period is partly determined by its own value in the previous period, and so on. Thus, multicollinearity is almost certain to exist in distributed lag models.

Taking the above considerations into account, it is clear that some degree of co linearity is expected to appear in most economic relationships. It should be noted that although multicollinearity is usually connected with time series, it is quite frequent in cross section data as well. For example, in across section, sample of manufacturing firms, labour and capital inputs are almost always highly inter-correlated because large firms tend to have large quantities of both factors, while small firms usually have smaller quantities of both labour and capital. However, multicollinearity tends to be more common and more serious a problem in time series.

Effects of Multicollinearity

1. The estimates of the coefficient are indeterminate
2. The standard errors of these estimates become infinitely i.e. the matrix (X^1X) become singular.
3. The estimate regression coefficients individual statistically significant even though a definite statistical relation exists between the dependent variable and the set of independent variables.

Consequences of Multicollinearity

In the presence of multicollinearity, the estimate of one variable's impact on y while controlling for the other tends to be less precise than if predictors were uncorrelated with one another. The usual interpretation of a regression coefficient is that it provides an estimate of the effect of a one unit change in an independent variable, X1, holding the other variables constant. If X1 is highly correlated with another independent

variable, X2 in the given data set, then we only have observations for which X1 and X2 have particular relationship (either positive or negative). We don't have observations for which X1 changes Independently of X2, so we have an imprecise estimates of the effect of Independent changes in X1.
Kutner (2004) affirms that, in some sense, the collinear variables contain the same information about the dependent variable. If nominally "different" measures actually quantify the same phenomenon, then they are redundant. Alternatively, if the variables are accorded different names and perhaps employ different numeric measurement scales but are highly correlated with each other then they suffer from redundancy.
One of the features of multicollinearity is that the standard errors of the affected coefficients tend to be large. In that case, the test of the hypothesis that the coefficient is equal to zero against the alternative that it is not equal to zero leads to a failure to reject the null hypothesis. However, if a simple linear regression of the dependent variable on this explanatory variable is estimated, the coefficient will be found to be significant; specifically, the analysis will reject the hypothesis that the coefficient is insignificant. In the presence of multicollinearity, an analyst might falsely conclude that there is no linear relationship between an independent and dependent variable. A principal danger of such data redundancy is over fitting in regressions analysis models. The best regression models are those in which the predictor analysis models correlate highly with the dependent (outcome) variables but correlate at most only minimally with each other. Such a mode is often called "low noise" and
will be statistically robust (that is it will predict reliably across numerous samples of variables sets drawn from the same statistical population).

Solutions for the Incidence of Multicollinearity

The solutions which may be adopted if multicollinearity exists in a function vary, depending of the severity of multicollinarity, on availability of other sources of data (larger samples, cross-section samples etc.), on the importance of factors which are multicollinearity on the purpose for which the function is being estimated and other considerations.

1. Some writers have suggested that, if multicollinearity does not seriously affect the estimates of the coefficients, one may tolerate its presence in the function, although the integrity of the least estimates is to a certain extent impaired.

2. Others have suggested that if multicollinearity affects some unimportant factors, one may exclude these factors from the function. Again specification error may well be expected to undermine the BLU character of the ordinary least squares.

3. Multicollinearity may affect only a part of the b's, while other estimates may remain fairly stable and reliable. In this case:

a. The reliable b's may be used for any purpose for any purpose, fore case or policy formulation (which require reliable information about structural coefficients);

b. All the estimates may be used to exist in the forecast period.

Corrective Solutions

1. Increase of the Size of the Sample: It has been suggested that multicollinearity may be avoided or reduced if we increased the size of the sample by gathering more observations. Thus Christ says that by increasing the sample. High covariance among estimated parameters resulting from multicollinearity in an equation can be reduced, because this covariance's inversely

proportional to sample size. This is true only if multicollinearity is due to errors of measurements, as well as when intercorrelation happens to exist only in our original sample but not in the population of the X's. If the populations of the variables are multi-collinear, obviously an increase in the size of the sample will not help in the reduction of multi-collinear relations among the variables.

2. Substitution of Lagged variables for other Explanatory variables in Distribution Lag models.

3. Introduction of Additional Equations in the Models Application of the Principal Components Method.

METHODOLOGY

Research methodology is carried out in any sector of the economy in order to discover the facts that economic data are highly affected by the presence of multicollinearity. In summary, it is a process of finding answer to problem and deals with multicollinearity in economic data.

The Farrar-Glaubar Test for Multicollinearity

A statistical test for multicollinearity has been recently developed by Farrar and Glauber. It is really a set of three tests, that is, the authors use three statistics for testing for multicollinearity. The first test is a chi-square test for the detection of the existence and the severity of multicollinearity in a function including several explanatory variables. The second test is F-tests for locating which variables are multi-collinear. The third test is a t-Test for finding out the pattern of multicollinearity, which is for determining which variables are responsible for the appearance of multi-collinear variables.
Farrar Glaubar considers multicollinearity in a sample departure of the observed X's from orthogonality. Their approach emerged from the general ideals developed in the preceding paragraphs namely that if multicollinearity is perfect, then the coefficients becomes indeterminate, and that the inter-correlations among the various explanatory variables can be measured by multiple correlation coefficients and partial correlation coefficients. The Farrar Glauber test may be outlined as follows:

Steps in Carrying Out the Farrar-Glaubar Test

i. Conduct the Chi - Square test to detect the existence of severity of multicollinearity.

ii. Carry out F-test to locate the variables(s) inter-correlated, if Chi-Square test is positive.

iii. Conduct T-test to detect the variables(s) that are responsible for multicollinearity if the F-test is positive.

The Chi - Square Test

The following steps are taken in conducting the Chi-Square test.

1. The ideal behind multicollinearity may be considered as a departure from orthogonality. The stronger the departure from orthogonality, that is the closer the value of the determined to zero, the stronger the degree of multicollinearity, and vice-versa, starting from this fact, Farrar-Glauber suggested the following X^2 test for detecting the strength of multicollinearity over the whole set of explanatory variables. The basic hypothesis here is:

Ho: The X's are orthogonal

It tested against the alternative hypothesis H1: The X's are not orthogonal

2. Compute the matrix (r,) of the Simple correlated coefficients between x1. For instance, if the explanatory variables are three, then xi becomes x1, x2 and x3.

3.

$$\begin{bmatrix} 1 & rx_2x_3 & rx_2x_3 \\ rx_2x_3 & 1 & rx_2x_3 \\ rx_2x_3 & rx_2x_3 & 1 \end{bmatrix}$$

4. Calculate D = |rij| the determinant of the matrix (rij)

5. Calculate the Statistics (Farrar and Glauber have found that the quantity) χ^2= - [T - 1 - 1/6(2k + 5)] Loge (value of the std. determinant)

(Where χ^2=observed (compound from the sample). T size of the sample, and K = number of explanatory variable) has a χ^2 distributed with V = ½k (k-1) degrees of freedom.

5. Check $\chi\alpha^2$ with 1/2k (k-1) degrees of freedom (α is the level of significance).

It should be clear that the theoretical value of χ^2, is the value that defines the critical region of the test at the chosen level of significance and with the appropriate degrees of freedoms.

If the observed χ^2 is greater than the theoretical value of χ^2 with ½k (k-1) degrees of freedom, we reject the assumption of orthogonality, that is, we accept that there is multicollinearity in the function. The higher the observed χ^2, the more severe is the multicollinearity.

If the observed $\chi^2 < \chi^2$, we accept the assumption of orthogonal, that is we accept that there is no significant multicollinearity in the function.

3.2.2. An F-Test for the Location of Multicollinearity

If the chi-square test is positive, the multicollinearity exists.

The next step is to locate the factors which are multi-collinear. Farrar Glaubar computes the multiple correlation coefficients among the explanatory variables (Rx12….x2….x3xn1, $Rx^2\ldots_2x^2\ldots x._1\ldots..x_3$) and they test the statistical significance of these multiple correlation coefficients with an F-test.

Steps:

i. Write the xi which is suspected to be inter-correlated with other Xs' as a function of other Xs'. Thus, xi=f(x1, x2, x3 …xk). This can also be rewritten as (Given x2x3 as inter-correlated variable then),

Xi=f(x2x3)=> B2X2+B3X3+U

ii. Compute the parameter

$b = \binom{b2}{b3}$As b = $(x_1x)\text{-}_1X_1X$, where X = (x_2x_3)

iii. Compute $R^2 = \frac{b_2 \sum x_1 x_2 + b_3 \sum x_1 x_3}{\sum x_1^2}$

Where $x_i x_j = \frac{\sum x_i x_j - \sum x_i x_j}{T}$

iv. Compute the F – statistic

$$F = \frac{R_{i^2}/(K-1)}{(1 - R_{i2})/(T - K)}$$

Where k = 3 and T = 12

And check $F_{.05}$ (k-1, T-k) from the F-distribution table.

v. Define the hypothesis

H_o: x_i is not inter-correlated with x_2 and x_3

H_i: x_i is inter-correlated with x_2 and x_3

vi. If $F_{.05} < F_{cal}$, accept H_1 and conclude that x_1 is inter-correlated with x_2 and x_3

vii. Repeat the test for other xii suspected to be inter-correlated with others.

1.2.3. A T-Test for the Pattern of Multicollinearity

This is a T-test which aims at the detection of the variables which cause multicollinearity.

To find variables which are responsible for the multicollinearity, we compute the partial correlation coefficients among the explanatory variables and test their statistical significance with the t-statistic.

Recall that the partial correlation coefficient between any two variable xi and xj, shows the degree of correlation between these two variables, all others being kept constant. For the two – variable model, the partial correlation coefficients are given by the formulae.

$$r^2 x_1 x_{2\ldots} x_3 = \frac{(r_{12} - r_{13} r_{23}).^2}{(1 - r^2_{23})(1 - r^2_{3})}$$

$$r^2 x_1 x_{2\ldots} x_3 = \frac{(r_{13} - r_{13} r_{23}).^2}{(1 - r^2_{23})(1 - r^2_{12})}$$

$$r^2 x_1 x_{2\ldots} x_3 = \frac{(r_{2} - r_{21} r_{31}).^2}{(1 - r^2_{13})(1 - r^2_{12})}$$

The basic hypothesis here is

H_o: $r x_i x_j \ldots x_1 .. x_2 \ldots\ldots x_n = 0$ and is tested against the alternative hypothesis

H_1: $r x_i x_j \ldots x_1 \ldots x_2 \ldots\ldots x_n \neq 0$

Having estimated the partial correlation coefficients, we test their significance by computing for each of them the statistic.

$$t* = \frac{(r x_i x_j \ldots\ldots\ldots x_k)/\sqrt{T-K})}{\sqrt{(1 - r^2 x_i x_j \ldots\ldots x_1 \ldots x_2 \ldots\ldots x_k)}}$$

Where $r^2 x_i x_j \ldots x_1 \ldots x_2 \ldots\ldots x_k$ denotes the partial correlation coefficient between xi and x_j.

The observed value t* is compared with the theoretical t value with v = (T-K) degrees of freedom at the chosen level of significance). If t* > t, we accept that the partial correlation coefficient between the variable Xi and Xj is significant, that is, the variables Xi and Xj are responsible for multicollinearity in the function. If t* < t we accept that Xi and Xj are not the cause of multicollinearity, since their partial correlation coefficient is not statistically significant. With the above three statistics, we find the severity, the location and the pattern of multicollinearity.

ANALYSIS AND PRESENTATION OF DATA

Presentation of Data Y = Consumption, X1= Wage income

X2= Non-wage, non-farm income, and

X3 = Farm income

S/N	Y	X1	X2	X3
1	62.8	43.41	17.10	3.96
2	65.0	46.44	18.65	5.48
3	63.9	44.35	17.09	4.37
4	67.5	47.82	19.28	4.51
5	71.3	51.02	23.24	4.88
6	76.6	58.71	28.11	6.37
7	86.3	87.69	30.29	8.96
8	95.7	76.73	28.26	9.76
9	98.3	75.91	27.91	9.31
10	100.3	77.62	32.30	9.85
11	103.2	78.01	31.39	7.21
12	108.9	83.57	35.61	7.39
13	108.5	90.95	37.58	7.98
14	11.4	95.47	35.17	7.42

Source: CIA World fact book, 2011.

ANALYSIS:

$\sum Y = 1219.7$, $\sum X_1 = 957.34$, $\sum X_2 = 381.98$, $\sum X_3 = 97.45$,

$\sum X_1Y = 877.152$, $\sum X_2 Y = 31929.9640$, $\sum X_3 Y = 8876.111$

$\sum X_1X_2 = 27761.387$, $\sum X_1X_3 = 7076.0291$, $\sum X_2X_3 = 2799.5687$

$\sum Y2 = 110780.37$, $\sum X12 = 55701.43$, $\sum X22 = 215255.3$, $\sum X32 = 734.111$

T = 14

Hypothesis Statement:

Ho: The X's are orthogonal

And it tested against the alternative hypothesis H1: The X's are not orthogonal

$$X_{1^2} = \frac{1}{T-1}\left[\sum x_i^2 - \frac{(\sum X)^2}{T}\right]$$

$$= \frac{1}{14-1}\left[70099.4766 - \frac{957.34^2}{14}\right]$$

$$= 4{,}635.1998$$

In the same manner,

X1X2 = 1641.0406, X1X3 = 412.2589, $X2^2$ = 653.1965,

X2X3 = 140.715, $X3^2$ = 55.7895, X1Y = 4310.3716,

X2Y = 2144.89, X3Y = 386.1278

Compute the matrix (r_{ij}) using the formula:

$$r_{ij} = \frac{X_i X_j}{\sqrt{X_{i2} X_{j2}}}$$

$$r_{12} = \frac{1641.0406}{\sqrt{4635.1998 \times 653.1965}}$$

$$= 0.9431$$

$$r_{13} = \frac{412.2589}{\sqrt{14635.1998 \times 55.7895}}$$

$$= 0.8107$$

$$r_{23} = \frac{140.7751}{\sqrt{653.1965 \times 55.7895}}$$

$$= 0.7371$$

It is a known fact that $r_{ij}=1$, if $j=i$.

Therefore, $r_{11} r_{22} = r_{33} = 1$

$$r_{ij} = \begin{bmatrix} 1 & 0.9431 & 0.8107 \\ 0.9431 & 1 & 0.7371 \\ 0.8107 & 0.7371 & 1 \end{bmatrix}$$

$|D| = |r_{ij}| = 0.0370$

$\text{Log}_e D = \text{Log}_e\ 0.0370 = -3.2968$

$\chi^2_{cal} = -\ [T - 1 - 1/6(2k + 5)]\ \text{Log}_e$

$= -\ [23 - 1 - 1/6(2 \times 3 + 5)]\ (-3.2968)$

$= (12.833) \times 3.2968$

$= 42.3089$

$\chi^2_{0.05} V$, where

$V = ½\ k\ (k-1)$

$V = ½ \times 3(3-1)$

$V = ½ \times 3(2)$

$V = 3$

$\chi^2_{0.05}\ 3 = 7.82$

Decision Rule:

Since reject Ho if χ^2cal>χ^2tab, otherwise accept Ho

Conclusion

Since reject Ho if χ^2cal>χ^2tab, we therefore reject Ho and conclude that there is the presence of Multicollinearity in the function.

Test 2: An F Test for the Location of Multicollinearity

HO: xi is not inter-correlated with x2 and x3 Hi: xi is inter-correlated with x2 and x3.

Hi: xi is inter-correlated with x2 and x3.

Write X11 = F(x2, x3)

= x1 => B2x2 + B3x3 + u

Thus $b = \begin{bmatrix} b_2 \\ B_3 \end{bmatrix}$ $(X^1X^1)^{-1}X^1X$

Observed that

$$X^1X = \begin{bmatrix} X^{12}X_2X^{12}X_3 \\ X^{13}X_2X^{13}X_3 \end{bmatrix} = \begin{bmatrix} 653.196140.715 \\ 140.7151 \quad 55.7895 \end{bmatrix}$$

$$|X^1X| = 16{,}640.7668$$

$$(x^1x)^{-1} = \frac{1}{16{,}640.7668}\begin{bmatrix} 55.7895 & -140.7151 \\ -140.7151 & 653.1965 \end{bmatrix}$$

$$= \begin{bmatrix} 0.0034 & -0.0085 \\ -0.0085 & 0.0393 \end{bmatrix}$$

$$b = \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.0034 & -0.0085 \\ -0.0085 & 0.0393 \end{bmatrix}\begin{pmatrix} 1641.0406 \\ 412.2589 \end{pmatrix}$$

$$= \begin{vmatrix} 2.0753 \\ 2.2529 \end{vmatrix}$$

$b_2 = 2.0753$, $b_3 = 2.2529$

$$R_{1^2} = \frac{b_2 \sum x_1 x_2 + b_3 \sum x_1 x_3}{\sum x_1^{\,2}}$$

$$= \frac{2.0753 \times 1641.0406 + 2.2529 \times 412.2589}{4635.1998}$$

$$= 0.9351 \cong 0.94$$

To compute the F – Statistics

$$F_1 = \frac{R_{1^2}/(k-1)}{(1-R_{1^2})/(T-K)}$$

$$= \frac{0.9351/2}{0.0649/11}$$

$$= 79.2458$$

$F_{0.05}(k - 1, T - k) = F_{0.05}(2, 11) = 3.98$

Decision Rule:

If Ftab<Fcal, accept H1 otherwise reject H1.

Conclusion

Since Ftab<Fcal,i.e 3.98 < 79.2456, we thereby accept the alternative hypothesis and conclude that X1 is inter-correlated with X2 and X3.

In the same manner, F2 = Fcal = 15.3375 F3 = Fcal = 21.9863

Testing each Fcal against Ftable i.e comparing F calculation with F tabulated, we realized that each of the Fcal is greater than Ftab, we therefore conclude that X2 is inter- correlated with X1 and X3 in the case. And X3 is inter-correlated with X1 and X2 in the last case.

Test3: A T-Test for the Pattern of Mulitcollinearity Hypothesis Statement

H_o: $r_{X_iX_j.X_1.X_2 \ldots\ldots\ldots\ldots n} = 0$

H_1: $r_{X_iX_j.X_1.X_2 \ldots\ldots\ldots\ldots n} \neq 0$

$$r^2x_1x_2x_3 = \frac{(r_{12} - r_{13}r_{23})^2}{(1 - r^2_{23})(1 - r^2_{13})}$$

$$t_2 = \frac{\sqrt{0.2642}/\sqrt{14-3}}{\sqrt{1-0.2642}}$$

$$= 1.842$$

$$= \frac{[0.9431 - (0.8107 \times 0.7371)]^2}{(1-0.7371^2)(1-0.8107^2)}$$

$$= \frac{0.1194}{0.4567 \times 0.03428}$$

$$= 0.7627$$

$$r^2x_1x_3x_2 = \frac{(r_{13} - r_{12}r_{23})^2}{(1 - r^2_{23})(1 - r^2_{12})}$$

$$= \frac{[0.8107 - (0.9431 \times 0.7371)]^2}{(1-0.7371^2)(1-0.9431^2)}$$

$$= \frac{0.0133}{0.4567 \times 0.03428}$$

$$= 0.2642$$

$$r^2x_2x_3x_1 = \frac{(r_{23} - r_{21}r_{13})^2}{(1 - r^2_{13})(1 - r^2_{13})}$$

$$= \frac{[0.7371 - (0.9431 \times 0.0.8107)]^2}{(1-0.9431^2)(1-0.8107^2)}$$

$$= \frac{0.008}{0.0379}$$

$$= 0.0200$$

$$r^2x_1x_{2\ldots}x_3 = \frac{(r_{12} - r_{13}r_{23}).^2}{(1-r^2_{23})(1-r^2_{13})}$$

$$r^2x_1x_{2\ldots}x_3 = \frac{(r_{13} - r_{13}r_{23}).^2}{(1-r^2_{23})(1-r^2_{12})}$$

$$r^2x_1x_{2\ldots}x_3 = \frac{(r_2 - r_{21}r_{31}).^2}{(1-r^2_{13})(1-r^2_{12})}$$

The basic hypothesis here is

H_o: $r_{X_iX_j}$...x_1..x_2.....$x_n = 0$ and is tested against the alternative hypothesis

H_1: $r_{X_iX_j}$...x_1...x_2.....$x_n \neq 0$

Having estimated the partial correlation coefficients, we test their significance by computing for each of them the statistic.

$$t* = \frac{(r_{X_iX_j \ldots\ldots\ldots x_k})/\sqrt{T-K}}{\sqrt{(1-r^2_{X_iX_j \ldots\ldots x_1 \ldots x_2 \ldots\ldots x_k})}}$$

Where $r^2x_ix_j$...x_1...x_2.....x_k denotes the partial correlation coefficient between xi and x_j.

$= 1.842$

$t_3 = \frac{\sqrt{0.7627}/\sqrt{14-3}}{\sqrt{1-0.7627}}$

$= 0.734$

Table Value

$t_{tab} = t_\alpha V$

$\alpha = 0.05$

$V = T - K = 14 - 3 = 11$

$T_{0.05}\ 11 \quad = 1.796$

Decision Table

If $t_{cal} > t_{tab}$, accept H_o otherwise reject H_o

Conclusion

Since $t_{cal} > t_{tab}$ only in the third case, we thereby accept and conclude that the variable X_2 is responsible for the multicollinearity in the function.

Correcting the Effect of Multicollinearity

As we rightly know, the effect of multicollinearity can be corrected through several ways. In this case, we shall adopt the method which says that this can be minimized or corrected by increasing the sample size.

We, therefore, increase the sample size by eleven (11) and test for the presence of multicollinearity.

The following results were obtained from the analysis.

S/N	Y	X_1	X_2	X_3	Y^2	X_1^2	X_2^2	X_3^2
1	62.8	43.41	17.1	3.96	3943.84	1884.428	292.41	15.6816
2	65	46.44	18.65	5.48	4225	2156.674	347.8225	30.0304
3	63.9	44.35	17.09	4.37	4083.21	1966.923	292.0681	19.0969
4	67.5	47.82	19.28	4.51	4556.25	2286.752	371.7184	20.3401
5	71.3	51.02	23.24	4.88	5083.69	2603.04	540.0976	23.8144
6	76.6	58.71	28.11	6.37	5867.56	3446.864	790.1721	405769
7	86.3	87.69	30.29	8.96	7447.69	7689.536	917.4841	80.2816
8	95.7	76.73	28.26	9.76	9158.49	5887.493	798.6276	95.2576
9	98.3	75.91	27.91	9.31	9662.89	5762.328	778.9681	86.6761
10	100.3	77.62	32.3	9.85	10060.09	6024.864	1043.29	97.0225
11	103.2	78.01	31.39	7.21	10650.24	6085.56	985.3321	51.9841
12	108.9	83.57	35.61	7.39	11859.21	6983.945	1268.072	54.6121
13	108.5	90.59	37.58	7.98	11772.25	8206.548	1412.256	63.6804
14	11.4	95.47	35.17	7.42	129.96	9114.521	1236.929	55.0564
15	47.82	46.44	18.63	7.37	2286.752	2156.674	347.0769	54.3169
16	51.02	44.35	17.01	8.67	2603.04	1966.923	289.3401	75.1689
17	58.71	47.82	19.28	9.79	3446.864	2286.752	371.7184	95.8441
18	87.69	51.02	23.22	9.31	7689.536	2603.04	539.1684	86.6761
19	76.73	58.71	28.11	9.85	5887.493	3446.864	790.1721	97.0225
20	75.91	87.69	30.29	7.21	5762.328	7689.536	917.4841	51.9841
21	77.62	76.73	28.64	7.39	6024.864	5887.493	820.2496	54.6121
22	78.01	75.91	27.91	7.98	6085.56	5762.328	778.9681	63.6804
23	83.57	77.62	32.3	7.42	6983.945	6024.864	1043.29	55.0564
24	90.59	78.01	31.39	9.6	8206.548	6085.56	985.3321	92.16
25	37.58	63.9	35.61	10.5	1412.256	4083.21	1268.072	110.25

1884.95	1665.54	674.37	192.54	154889.6	118092.7	19226.12	1570.883	

Source: CIA World fact book, 2011.

Hypothesis Statement:

Ho: The X's are orthogonal

And it tested against the alternative hypothesis H1: The X's are not orthogonal

In the same manner,

χ^2cal = -[T – 1 – 1 1/6(2k + 5)] Loge

= 6.3089

$\chi^2$0.05V, where v = ½ k(k – 1)

v = ½ x 3(3 – 1) v = x 3(2)

v = 3

$\chi^2$0.05 3 = 7.82

Decision Rule:

Since reject Ho if χ^2cal>χ^2tab, otherwise accept Ho

Conclusion

Since χ^2cal<χ^2tab, we therefore fail to reject Ho and conclude that there is absent of multicollinearity in the function.

CONCLUSION AND RECOMMENDATIONS

Conclusion

It is obvious from the analysis that the presence and severity of multicollinearity, as well as the pattern and location of multicollinearity, in a function can be easy detected by Ferrar- Glauber. In this analysis, we realized that variable X2 is responsible for the multicollinearity in the function.

Recommendations

From this analysis and other analysis, it is obvious that multicollinearity is always present in Economic data but the severity differs. The following recommendations can be made from this analysis and from other submissions on multicollinearity.

1. If the multicollinearity does not seriously affect the estimates of the coefficients, one may tolerate its presence in the function. Although the integrity of the least estimates is to a certain extent impaired.
2. The use of Lagged variables for other explanatory variables in Distribution LagModels can reduce the presence of colinearity.
3. Introduction of Additional Equations in the Models.

References

Aldrich, John (2005). "Fisler and Regression" *Statistical Science*. 20(4): 401- 417.

Armstrong, J. Scot (2012). "Illusions in Regression Analysis". *International Journal of Forecasting*. **28** (3): 689.

Becker, A.J., and Hoover E., M. (1998) *Population Growth and Economic Development in Low-Income Countries*. Princeton: Princeton University Press, 610-619.

CIA World facts books, Vol. 5, No. 1, 2011: 89-95.

Cressie, N. (1996) "Change of Support and the Modiable Area Unit Problem" *Geographical Systems* 3: 169 – 180.

David A. Freedom (2005), *Statistical Models: Theory and Practice*, Cambridge University Press.

Fotheringham, AS; Wong, DWS (1st Jan, 2002). "The Modifiable Area Unit Problem in Multivariate Statistical Analysis" in *Environment and Planning*. A. 23(7): 1025 – 1044.

Kutner, M.H., C.J Nachtshein, and J. Neter (2004), *Applied Linear Regression Models*. 25(7):25 – 31.

Smith, J.B. (1997). Effects of eighth-grade transition programs on high school retention and experiences. *The Journal of Educational Research*, 90 (3): 144 – 152.

Thirwal, Dolf (2007) "Growth and Development with Special Reference to DevelopingEconomies", *University of Kent at Canterbury*, 5th Edition. **111** (8): 143 – 155.

Tofallis, C. (2009). "Least Squares Percentage Regression". *Journal of Modern AppliedStatistical Methods.* 7: 526 – 534.

Waegeman, Willem; De Baets, Bernanrd; Boullart, Luc (2009). "Roc analysis in Ordinal Regression Learning" Pattern Recognition Learning" Pattern Recognition Letters. 29:1-9.

Yang Jing Long (2009). "Human Age Estimation by Metric Learning for Regression Problems". 21(6): 74 – 82.