Sudan Journal of Science and Technology
Journal homepage:
http://jst.sustech.edu/

Determining the Best Method for Estimating in Sudan during the period (1982-2008)

Rabab M. Sorkatti [1], Dr. Afra H. Abdel-lateef [2], Dr.Khalid R. Khider[3]
Sudan University of Science and Technology.
Received: 2/1/2019 Accepted:29/4/2019

Abstract:
The paper aimed at determining the Best Method for Estimating in Sudan during the period (1982-2008).The paper used the descriptive – analytic approach (summarizing and describing a body of data and testing the significancy of the parameters) for the data was given by Central Bureau of Statistics (CBS) /Sudan also it use the data from reports and publications from other relevant sides like Central Bank of Sudan and Sudan household survey 2006. for estimating GDP by Prediction Residual Error Sum of Squares (PRESS) method after excluding the outliers values, the prediction sum of squares is a useful statistic for comparing different models. Moreover, the findings showed that PRESS method model bring estimated value for DGP more relevant to the actual value than traditional method which indicate that PRESS method is the best for estimating the future value. Based on these findings the paper calls for more application of the criteria for selection the best model.
Keywords: regression, outlier, P.R.E.S.S, G.D.P.

Introduction:
The study of the overall operation and performance of the economy can be done by looking to the information provided by a lot of macro variables; the main one is the national income. Accordingly, national income helps to assess and compare the progress achieved by a country over a period of time. National income is generally defined as the value of final goods and services produced in a country in an accounting year. However, it can be defined in terms of total output, total factor income and total expenditure **(**Subhendu, 2006**).**
In terms of total factor income, it is the sum of factor incomes (wage, rent, interest, profit) in a country in a year. Factors of production are land, labor, capital and organization/ entrepreneur earns reward as rent, wage, interest and profit respectively. The sum of these rewards is the national income in terms of income generated in the economy. National income, in terms of total expenditure, is the aggregate expenditure of a country in a year's time. Spending of households, private sector and government sector in a country adds up to national income by expenditure method. National income at current prices is the money value of all goods and services produced in a country estimated at the prevailing prices. National income at constant prices is the national income estimated at a base year, which is an earlier year to the current year. National income at constant prices is used for making comparisons of national income and related data. The important of National Income is to estimate economic development and how far development objectives were achieved, the contribution of various sectors to national income. The basic concepts of national income are Gross national product, Gross domestic product, private income, personal income and personal disposable income.

The Gross Domestic Product (GDP), is the sum of all the final goods and services produced in the country within a specified period of time (usually one year). The money value of all these goods and services taken together gives us the GDP. Various methods are used for computing GDP, such as value added method, expenditure methods

GDP is one of important economic indicators that reflect the nature of the economic activities, and it is a tool of evaluating the economic performance at the same time it helps in predicting some indicators such as inflation, unemployment.

When determining the value of the GDP some problems occur such as the distinction between current prices and the prices of the cost, and the distinction between intermediate goods and final goods as well as the implicit value problem and other problems that will be covered in the study. (Mayson, 2012).

It find that the level of output is determined based on the availability of economic resources and the optimal use of these resources and the suitability of economic and political policies in the country. In previous study they focus on conducting the model of the gross domestic product as dependent variable and the government expenditure, consumption expenditure, investment expenditure and the net between the export and import expenditure as independent variables, and not concerning about the best model to represent this relationship.

In this paper we study the best model for this relationship and using the PRESS method, and after that use the model for prediction.

As mentioned GDP used as an indicator to the economic performance and as a tool to predict these indicators, so it helps to find process to increase the output for increasing the per capita income.

The importance of the study comes from the importance of the GDP in improving the performance of the economy, so it is necessary to evaluate the model that used in measuring the GDP to find the best model that should be adopted.

Multiple linear regression model using different criteria has been configured to test the significant of the model, but PRESS method is the best in determining the best estimate Model for estimate.

The objectives of the Study, Illustrate the general trend of GDP, Choose the Best Model for estimating the GDP applying the PRESS method, comparing the result between PRESS model and traditional model.The hypotheses of the Study, Aspects of spending in the economic sectors (government sector, consumer sector, investment sector and the external sector) affect on GDP, Best Model can be obtained by applying PRESS method.

Materials and Methods:

An econometric study begins with a set of propositions about some aspect of the economy. The theory specifies a set of precise, deterministic relationships among variables. Familiar examples are demand equations, production functions, and macroeconomic models. The empirical investigation provides estimates of unknown parameters in the model, such as elasticities or the effects of monetary policy, and usually attempts to measure the validity of the theory against the behavior of observable data. Once suitably constructed, the model might then be used for prediction or analysis of behavior. (Johnston, Dinardo, 1997).

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More

specifically, regression analysis helps one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given that the independent variables are fixed, that is the average value of the dependent variable when the independent variables are fixed. Many techniques for carrying out regression analysis have been developed. The familiar method is ordinary least squares to estimate the linear regression model. (Abdul Rahman, 1417 e). It is the single most useful tool in the econometrician's kit. Though to an increasing degree in the contemporary literature, it is often only the departure point for the full analysis, it remains the device used to begin almost all empirical research.

Simple linear regression is the least squares estimator of a linear regression model with a single explanatory variable. The two-variable linear model, or simple regression analysis, is used for testing hypothesis about the relationship between a dependent variable *Y* and an independent or explanatory variable *X* and for prediction. Simple linear regression analysis usually begins by plotting the set of *XY* values on a scatter diagram and determining by inspection if there exists an approximate linear relationship (Carter Hill, other 2011). It defined as the proportion of the total variation in Y explained by the regression of *Y* on *X*. $\boldsymbol{R^2}$ is unit free and its value ranged between 0 (when the estimated regression equation explained none of variation in *Y*) to 1 (when all points lie on the regression line).

General Linear Model:

The multiple regression analysis is used for testing hypotheses about the relationship between a dependent variable *Y* and two or more independent variables *X* and for prediction

Press Statistcs:

In statistics the predicted residual error sums of squares (PRESS) statistic is a form of cross validation used in regression analysis to provide a summary measure of the fit of a model to a sample of observations that were not themselves used to estimate the model. It is calculated as the sums of squares of the prediction residuals for those observations.

A fitted model having been produced, each observation in turn is removed and the model is refitted using the remaining observations. The out-of-sample predicted value is calculated for the omitted observation in each case, and the PRESS statistic is calculated as the sum of the squares of all the resulting prediction errors.

Given this procedure, the PRESS statistic can be calculated for a number of candidate model structures for the same dataset, with the lowest values of PRESS indicating the best structures. Models that are over-parameterized over fitted would tend to give small residuals for observations included in the model-fitting but large residuals for observations that are excluded.

The prediction sum of squares is a useful statistic for comparing different models. It is based on the principle of leave-one-out or ordinary cross-validation, whereby every measurement is considered in turn as a test set, for the model parameters trained on all but the held out measurement. As for linear least squares problems, there is a simple well-known non-iterative formula to compute the prediction sum of squares without having to refit the model as many times as the number of measurements. We extend this formula to cases where the problem has multiple parameter or measurement sets.

The prediction sum of squares (PRESS), similar to the sum of squares of the residual error (SSE), is the sum of squares of the prediction error. PRESS differs from the sum of squares of the residual error in that each fitted value, Y hat $_i$, for PRESS is obtained from the

remaining n – 1 observations, then using the fitted regression function to obtain the predicted value for the i^{th} observation.

Use PRESS to assess your model's predictive ability. Usually, the smaller the PRESS value, the better the model's predictive ability. PRESS is used to calculate the predicted R^2 which is usually more intuitive to interpret. Together, these statistics can help prevent over-fitting the model because these statistics are calculated using observations not included in model estimation. Over-fitting refers to models that seem to explain the relationship between the predictor and response variables for the data set used for model calculation but fail to provide valid predictions for new observations.

The prediction sum of squares (PRESS) is a statistic based on the leave-one-out technique. It was proposed by Allen in 1974, and is typically used to compare different models. It is equivalent to the sum of standardized residuals, and can be extended to select parameters such as the regularization weight in smoothing spines. The press is a statistic that depends on a chosen cost function, and is in a sense complementary to this cost function. The cost function often expresses the discrepancy between measurements and the values predicted by a parametric model. While minimizing the cost function allows one to find the model parameters, it is clear that the most complex model always has the lowest residual error. In other words, the 'best' model cannot be selected based on the residual error only. The PRESS statistic, however, does not depend on some particular model parameters, but on the model itself. As with techniques based on cross-validation, it expresses to which extent a particular model is able to generalize to new data. The PRESS should therefore be used as a measure of predictivity to compare and select the 'best' model, while minimizing the cost function gives the parameters of a particular model (David M 1974) "The relationship between variable selection and data augmentation and a method for prediction".

Results:

To study the influence of the type of expenditure on general Domestic product and depending on the data which had been taken from national accounts administration in the Central Bureau of Statistics, the researcher prepare to apply multiple regression model to represent the relationship between the GDP and the type of expenditure, and all these variable in fixed price for the year(1981-1982) in SDD million and this to remove the influence of the change in price which is due to inflation factor and this appear in GDP at current price. We use SPSS to estimate the model after testing the data for adequacy and normality, and we use Minitab for choosing the best model according to PRESS process. So, the relation according to multiple regression models: (Osama Rabie, 2008)

$$GDP = \beta_0 + \beta_1 G + \beta_2 C + \beta_3 I + \beta_4 Net$$

Where:

GDP: dependent variable (Gross Domestic Product)

G: the first independent variable (Government expenditure)

C: the second independent variable (consumption expenditure)

I: the third independent variable (investment expenditure)

Net: the fourth independent variable (net expenditure)

The sufficient of the sample:

To test the adequacy of the sample data, the researcher use the KMO and Bartlett's test of sphericity , and the sample seems to be adequacy when the value of the Kaiser statistic is between (.5 – 1) .

Table (1) : KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.527
Bartlett's Test of Sphericity	Approx. Chi-Square	191.807
	df	10
	Sig.	0.000

Source: by the researcher 2018, using SPSS

According to table (1) which provides the result of KMO test, it is clear that Kaiser Statistic is equal to 0.527, and this indicates that the sample is sufficient enough for the analysis.

Table (2): Descriptive Statistics of the study variables

	N	Minimum	Maximum	Mean	Std. Deviation
GDP	27	4.0701	23.4244	14.36225	7.4890083
G	27	.7045	3.7740	1.870833	1.0650637
C	27	5.6230	12.6866	8.1691000	2.2280534
I	27	.3458	9.3659	3.107363	3.0273605
Net	27	-2.8180E2	.528500	-.6584111	7.11490544

Source: by the researcher 2018, using SPSS

Table (3): Tests the Normality of the data

	Kolmogorov-Smirnova		
	Statistic	df	Sig.
G	0.139	27	0.192
C	0.203	27	0.006
I	0.277	27	0.000
Net	0.213	27	0.003
GDP	0.170	27	0.043

Source: by the researcher 2018, using SPSS

From the table (3) and according to Kolmogorov _ Smirnov to test the normality of the data, we found that four variables not normally distributed (consumption, investment, external and GDP expenditure) because their *P* value is less than 0.05, so we use the log transformation by taking the *log* of the actual data and the result is the following table.

Table (4): Tests of Normality by using the log transformation

	Kolmogorov-Smirnova		
	Statistic	df	Sig.
Log G	0.150	27	0.124
Log C	0.174	27	0.035
Log I	0.138	27	0.200
Log Net	0.135	27	0.200
Log GDP	0.112	27	0.200

Source: by the researcher 2018, using SPSS

From the table (4) only one variable not normally distributed (log consumption) because its *P* value is less than 0.05 but the other variable is normally distributed.

Table (5) : Descriptive statistics of the variable of the model

	Mean	Std. Deviation	N
Log GDP	1.0478	0.20078	27
Log G	0.2025	0.25304	27
Log C	0.8972	0.11511	27
Log I	0.3018	0.41462	27
Log Net	-0.2154	0.15264	27

Source: by the researcher 2018, using SPSS

From the table (5) find the mean and standard deviation of the dependent and independent variables of the model as above.

Table(6): Summary of the model

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.972	0.946	0.936	0.05094

Source: by the researcher 2018, using SPSS

From the table (6) find that R square and adjusted R square is very high which equal to 94.6% and 93.6 % consequently , and that means 93.6% change in dependent variable (GDP) have been explained by the change in the independent variables, and here the R square is very high .

Table (7): Analysis of variance table (ANOVA):

S.O.V	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.991	4	0.248	95.471	0.000
Residual	0.057	22	0.003		
Total	1.048	26			

Source: by the researcher 2018, using SPSS

From this ANOVA table (7) find that the whole model is significant because P value for the model is significant.

Table (8): Significant of the coefficients of the model:

Model	B	Std. Error	t	Sig.
(Constant)	0.394	0.171	2.306	0.031
Log G	0.336	0.089	3.790	0.001
Log C	0.662	0.198	3.346	0.003
Log I	0.090	0.050	1.803	0.085
Log Net	0.161	0.075	2.160	0.042

Source: by the researcher 2018, using SPSS

From this table find that all the coefficients of the parameters is significant because *t* value for them is significant, except for the investment not significant, that means investment is not statistically important for the model.

From the above table, our model according to this is:

$$\log GDP = 0.394 + 0.336\log G + 0.662\log C + 0.161\log Net$$

Table (9): Statistics of the residual

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	0.7725	1.3582	1.0478	0.195	27
Residual	-0.19842	0.04396	0.000	0.047	27
Std. Predicted Value	-1.4100	1.590	0.000	1.000	27
Std. Residual	-3.8950	0.863	0.000	0.920	27

Source: by the researcher 2018, using SPSS

From the table (9) shows minimum, maximum, mean and std. deviation of the predicted value of GDP and the residual from the prediction model.

Prediction residual sum of square (PRESS)

The best model according to PRESS method is:

$$\log GDP = 0.275 + 0.409\log G + 0.801\log C + 0.132\log Net$$

Table (10): Summary of PRESS

Model	R^2	R^2_{pred}	SS_{res}	S^2	Press
G,C,I,Net	94.6	90.64	0.0571	0.0026	0.0981
G,C,I	93.4	89.3	0.0692	0.003	0.1121
G,C,Net	93.7	91.27	0.0655	0.0029	0.0915
G,I,Net	91.8	87.43	0.0862	0.0037	0.1317
C,I,Net	91	86.55	0.0944	0.0041	0.141
G,C	92.9	90.97	0.0741	0.0031	0.0946
G,I	87.5	82.28	0.1311	0.0055	0.1857
G,Net	89	86.14	0.1157	0.0048	0.1452
C,I	90.4	86.3	0.1011	0.0042	0.1435
C,Net	87	83.6	0.1358	0.0057	0.1719
I,Net	81.3	77.25	0.1964	0.0082	0.2384

Source: by the researcher 2018, using Minitab

Conclude that the model which done using the PRESS method is the best model, and when the residual sum of square is less that means the model is the best model.

Prediction:

Here predict GDP for 2009 and 2010 and compare between the value obtained from the traditional method and PRESS method as follow:

Table (11): Prediction

DGP	2009	2010
Actual value	1.3957	1.423
Traditional method	1.2712	1.2967
PRESS method	1.3415	1.3674

Source: by the researcher 2018, using Minitab

From this table (11) notice that PRESS method model bring estimated value for DGP more relevant to the actual value than traditional method which indicates that PRESS method is the best for estimating the future value.

Conclusion:

This study handled the impact of type of expenditure on GDP, Hence, there seems to emerge the various type of expenditure on GDP and their effect on GDP. Their implication on affecting of GDP need to be investigated.The study takes the four type of expenditure on GDP and test the significance of the model.The results of the data reflect the importance of three types of expenditure on influence on the GDP.Furthermore, the study finding in relation to the above the equation of the best model for GDP using PRESS Method.

The PRESS method to assess the model's predictive ability for the future value for GDP and to calculate the predicted R^2.The predict GDP for 2009 and 2010 and compare between the value obtained from the traditional method and PRESS method, we notice that PRESS method model bring estimated value for DGP more relevant to the actual value than traditional method. Finally, it could be predicting the GDP in future using this model

Recommendations:

According to the study finding and conclusion, this study recommended the following:

1.More applying the criteria of the best model selection in order to get the best model for representing the data while dealing with economic phenomena.

2.Encourage the economy to go towards international market by more exports.

3.The government should influence the internal market by more expenditure for establishing good environment for investment because this study notice that investment on GDP not significant

4.Above all, the Statistical analysis for all economic procedure should be supported to hold a scientific approach for every policy.

References:

1.A. Raheem, Mayson,(2012), Estimation of Sudan's GDP Function. Sudan University of Science and Technology. Sudan.

2.Allen, David M,(1974). The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. Technometrics. **16:** 125-127

3.Amin, Osama Rabie, (2008), Statistical Analysis of Multiple Variables Using SPSS Program, University of Menoufia

4.Dutta , Subhendu, (2006) .Introductory Economics (Micro and Macro): A Textbook for Class XII. New Age International. New Delhi.

5.Jack Johnston, John Dinardo.(1997), Econometric Methods, fourth edition, McGraw-Hill Book Company, London

6.R. Carter Hill, William E. Griffiths and Guay C. Lim (2011), Principle of Econometrics, Fourth *Edition.* John Wiley & Sons, Inc, USA