1-1: 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.[1](Subhendu Dutta , Introductory- Economics,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, labour, 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. .[2] (Estimation of Sudan's GDP Function, Mayson A. Raheem-2012).

We 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.

1-2: Problem of the Study:

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

1-3: Important of the Study:

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.

1-4: Objectives of the Study:

1- Illustrate the general trend of GDP.

2- Choose the Best Model for estimating the GDP applying the PRESS method.

3- Comparing the result between PRESS model and traditional model.

1-5: Hypotheses of the Study:

1- Aspects of spending in the economic sectors (government sector, consumer sector, investment sector and the external sector) affect on GDP.
2- Best Model can be obtained by applying PRESS method.

1-6: Methodology of the Study:

The study is a descriptive analytical type of studies, to realize its objectives the study depend on secondary data to analyze the components of GDP and estimate components by estimating Models of multiple regression and choose the best Model using PRESS method.

1-7: Study data:

The study deals with GDP data issued by the Central Bureau of Statistics in the period of 1982-2008.

1-8: Previous studies:

1). Study that addressed the GDP of scientific research done by Aqeel Abdul Hakim (2006) on the impact of exports GDP in the Republic of Yemen.[3].. Study handled the evolution of GDP and its composition, and the average growth of real GDP during the study period. It also showed how Yemeni exports depend mainly on oil exports, and in addition to, he addressed the ratio of the contribution of exports to GDP growth through the standard model represented by the agent simple linear regression demonstrated a strong positive relationship between exports and changes in GDP. He estimates the rate of change in GDP when the change of the unity of (one dollar).

2). Another study prepared by Mayson A/Raheem (2012), estimated the function of GDP by preparing Models of record represent the estimate and predict the values of GDP The study found that expenditure contributes a high proportion in the interpretation of GDP .[2].

In this research through the conducting the best model we will find out the impact of external spending (Net Exports) on of GDP in addition to studying the impact of government spending and consumer spending (domestic) and investment spending on GDP in Sudan using

No study on the statistical estimation methods of GDP tested by PRESS method , and this will done by this research.

1-9 : study structure :

This study consist of five chapters, the first one is introduction to the study which include the problem, importance, objective, hypotheses and the methodology of the study also it included the source of the data and the pervious study related to this topic, and the structure of the study.

The second one is literature review about the Gross Domestic Product, the importance of its account, method of calculating (production, income and expenditure)approach, component of GDP by expenditure , background of Sudan economy

The third one is statistical methodology which about regression analysis, linear regression model, ordinary least square, model selection cross validation and PRESS

The fourth one is analysis and discussion of the findings .The last one is conclusion and recommendation , this follow by references and appendices.

2-1: Gross Domestic Product:

It is the money value of all goods and services produced in a country during one year, by the population of the country regardless of this citizenship which means GDP is a geographical concept determined by a geographical area.

2-1-1: Importance of GDP accounting:

The best way to understand a country's economy is by looking at its Gross Domestic Product (GDP). This economic indicator measures the country's total output. This includes everything produced by all the people and all the companies in the country, so it helps to:

1- Follow economic fluctuations (periodic and non-periodic) short, medium and long-term index reveals where GDP for prosperity or recession in the economy determine the reality of the economy, compared to other economies (follow-up economic growth)

2- Development of policies on population, especially when determining the per capita income of Sudan.

3-Any deficit or surplus in the national accounts is reflected in the growth rates of GDP, as well as on the nominal and real value placed and determine economic policies in the light of his calculations.

2-1-2: Methods for calculating GDP:

GDP can be determined in three ways, all of which should, in principle, give the same result. They are the production (or output) approach, the income approach, or the expenditure approach.

The most direct of the three is the production approach, which sums the outputs of every class of enterprise to arrive at the total. The expenditure approach works on the principle that all of the product must be bought by somebody, therefore the value of the total product must be equal to people's total expenditures in buying things. The income approach works on the principle that the incomes of the productive factors ("producers," colloquially) must be equal to the value of their product, and determines GDP by finding the sum of all producers' incomes.

Production approach:

It is the sum total of market value of final goods and services produced in a country during 1 year by :

1. Estimate the gross value of domestic output out of the many various economic activities;

2. Determine the intermediate consumption, i.e., the cost of material, supplies and services used to produce final goods or services.

3. Deduct intermediate consumption from gross value to obtain the gross value added.

Gross value added = gross value of output – value of intermediate consumption.

Value of output = value of the total sales of goods and services plus value of changes in the inventories.

The sum of the gross value added in the various economic activities is known as "GDP at factor cost".

GDP at factor cost plus indirect taxes less subsidies on products = "GDP at producer price".

For measuring output of domestic product, economic activities (i.e. industries) are classified into various sectors. After classifying economic activities, the output of each sector is calculated by any of the following two methods:

1. By multiplying the output of each sector by their respective market price and adding them together .

2. By collecting data on gross sales and inventories from the records of companies and adding them together.

The gross value of all sectors is then added to get the gross value added (GVA) at factor cost. Subtracting each sector's intermediate consumption from gross output gives the GDP at factor cost. Adding indirect tax minus subsidies in GDP at factor cost gives the "GDP at producer prices".

Income approach:

It is Sum total of incomes of individuals living in a country during 1 year. Another way of measuring GDP is to measure total income. If GDP is calculated this way it is sometimes called gross domestic income (GDI), or GDP (I). GDI should provide the same amount as the expenditure method described below. (By definition, GDI = GDP. In practice, however, measurement errors will make the two figures slightly off when reported by national statistical agencies.)

This method measures GDP by adding incomes that firms pay households for factors of production they hire- wages for labour, interest for capital, rent for land and profits for entrepreneurship.

Expenditure approach:

It defined as all expenditure incurred by individuals during 1 year. In economics, most things produced are produced for sale and then sold. Therefore, measuring the total expenditure of money used to buy things is a way of measuring production. This is known as the expenditure method of calculating GDP. Note that if you knit yourself a sweater, it is production but does not get counted as GDP because it is never sold. Sweater-knitting is a small part of the economy, but if one counts some major activities such as child-rearing (generally unpaid) as production, GDP ceases to be an accurate indicator of production. Similarly, if there is a long term shift from non-market provision of services (for example cooking, cleaning, child rearing, do-it yourself repairs) to market provision of services, then this trend toward increased market provision of services may mask a dramatic decrease in actual domestic production, resulting in overly optimistic and inflated reported GDP. This is particularly a problem for economies which have shifted from production economies to service economies.

2-1-3: Components of GDP by expenditure

GDP (Y) is the sum of government spending (G), consumption (C), investment (I), and net exports (X – M).

$$Y = G + C + I + (X - M) \qquad (2\text{-}1)$$

The description of each GDP component:

G: (government spending) is the sum of government expenditures on final goods and services. It includes salaries of public servants, purchases of weapons for the military and any investment expenditure by a government. It does not include any transfer payments, such as social security or unemployment benefits.

C: (consumption) is normally the largest GDP component in the economy, consisting of private (household final consumption expenditure) in the economy. These personal expenditures fall under one of the following categories: durable goods, non-durable goods, and services. Examples include food, rent, jewelry, gasoline, and medical expenses but does not include the purchase of new housing.

I: (investment) includes, for instance, business investment in equipment, but does not include exchanges of existing assets. Examples include construction of a new mine, purchase of software, or purchase of machinery and equipment for a factory. Spending by households (not government) on new houses is also included in investment. In contrast to its colloquial meaning, "investment" in GDP does not mean purchases of financial products. Buying financial products is classed as 'saving', as opposed to investment. This avoids double-counting: if one buys shares in a company, and the company uses the money received to buy plant, equipment, etc., the amount will be counted toward GDP when the company spends the money on those things; to also count it when one gives it to the company would be to count two times an amount that only corresponds to one group of products. Buying bonds or stocks is a swapping of deeds, a transfer of claims on future production, not directly an expenditure on products.

X: (exports) represents gross exports. GDP captures the amount a country produces, including goods and services produced for other nations' consumption, therefore exports are added.

M: (imports) represents gross imports. Imports are subtracted since imported goods will be included in the terms G, I, or C, and must be deducted to avoid counting foreign supply as domestic.

A fully equivalent definition is that GDP (Y) is the sum of final consumption expenditure (FCE), gross capital formation (GCF), and net exports (X – M).

$$Y = FCE + GCF + (X - M) \qquad (2\text{-}2)$$

2-2: Background of Sudan Economy:

Since the last strategic plan (2007-11) Sudan has seen achieved many developments. General census was conducted in 2008 to update the country's population data. The country had elections in 2010 and a referendum was held in January, 2011, when the Southerners voted to secede, changing the political landscape of Sudan to form a new country, the Southern Sudan. The economy continued to grow, but following the separation of Southern Sudan, the oil revenue was shrinking. In this background, the following is the picture of Sudan, seen from different aspects, (FMOH, strategic plan, 2012).

2-2-1: Geography and ecology:

With land area of 1.8 million square kilometres, traversed by the Nile and its tributaries, Sudan shares its borders with Southern Sudan, Central African Republic, Chad, Libya, Egypt, Eritrea and Ethiopia. It has access to the Red Sea with 853 kilometres long coastline. Its terrain is generally flat, featureless plain, mountains in northeast, and west, while desert dominates the north.

Sudan's climate is tropical in south, arid desert in north, and the rainy season vary by region from April to November. With dust storms and periodic persistent droughts and flooding, the country faces soil erosion and desertification, inadequate supplies of potable water and wildlife is threatened by excessive hunting. Its geography and ecology contribute to the shaping of health, nutrition and population situation. Vast distances and poor roads and transport infrastructure affect coverage of health services, while natural and manmade disasters cause humanitarian emergencies and ecological factors expose population to infectious and parasitic diseases.

2-2-2: Demography:

With the total population 33,419,625 people, growing at a rate of 2.8% per annum, 88% of it is settled, including 49% living in urban areas, while 8% is nomadic. Almost 2% of the population is internally displaced, while another 1.4% resides in institutions and the remaining 0.6% lives in the cattle camps. There has been increasing urbanization, with natural disasters, civil conflicts and poor conditions in rural areas, contributing to this. The average household size is 5 – 6 persons, while fertility rate is 3.9; crude birth rate is 31.2; and crude death rate is 16.7 per 1,000 people (17.2 males, 16.3 females). About 43.2% of population is young under 15 years including 15% under 5 years, 49.6 male and 53.4% female are in age group 15-64 years, and 5.4% of male, 3.4% of female are 60 years and above, life expectancy at birth is 59 years (58 years for males and 61years for females),(CBS, 2008).

2-2-3: Economic picture:

Sudan is rich in natural resources, including oil, agriculture and animal resources. Its economy, with the export of crude oil in 1999, boomed due to increases in oil production, high oil prices, and significant inflows of foreign direct investment. Despite sanctions and additional safety guard policies of the west, it was one of the world's fastest growing economies until the second half of 2008. The gross domestic product (GDP) grew from US$9.9 billion in 1980 to US$57.9 billion in 2008, due to the effect of global financial crisis shranked to 52.2 billion in 2009, achieving GDP growth rate of 6.7%, and declined to 6.2% in 2011. While the oil sector has been the driving force behind growth and services and utilities playing an increasingly important role, agriculture remained important in the economy as it employs 80.2% of the work force and contributes a third of GDP,(UNDP,MDGs in Sudan,2010).

The economic growth has however benefitted mainly the capital cities, leading to the increasing disparities between rural and urban areas as well as between states and the geographical regions of the country. Poverty remains widespread with Sudan ranking 147th among 177 countries on human development index, (ibid). About 46.5% of the population live below poverty line with less than 1$ earning a day, while 8% live in extreme poverty. The hardest hit by the poverty are the rural dwellers, particularly women and internally displaced people. The unemployment rate is 11%, (CBS,2010, SNHS,2009).

2-2-4: Political landscape:

Sudan comprises 15 states each divided into localities, which in total are 144, but varies with time due to redrawing the boundaries of the existing ones. Sudan with multiparty system is a federated republic with powers devolved to states under Local Government Act (2003), often referred to as the Decentralization Act. However, precise legislative and organizational arrangements may vary from state to state.

2-2-5: Socio-cultural scene:

Sudan is a multicultural society with hundreds of ethnic and tribal divisions and languages. The adult literacy rate in Sudan is 69%. The primary enrolment is 80.5%, while 36.4% of the cohort completed primary school education. 58.7% of the population has access to improved drinking water source, while 55% enjoy improved sanitation, (SHHS, 2006).

3-1: Introduction:

Statistics refers to the collection, presentation, analysis and utilization of numerical data to make inferences and reach decisions in the face of uncertainty in economics, business and other social and physical sciences.

Statistics is subdivided into descriptive and inferential. Descriptive statistics is concerned with summarizing and describing a body of data. Inferential statistics is the process of reaching generalization about the whole (called population) by examining apportion (called the sample). In order for this to be valid, the sample must be representative of the population and the probability of error also must be specified.

Econometrics refers to the application of economic theory, mathematics, and statistical techniques for the purpose of testing hypotheses and estimating and forecasting economic phenomena. Econometrics has become strongly identified with regression analysis. This relates a dependent variable to one or more independent or explanatory variables. Since relationships among economic variables are generally inexact, a disturbance or error term (with well-defined probabilistic properties) must be included.

Econometric research, in general, involves the following four stages:

1. Specification of the model or maintained hypothesis in explicit stochastic equation form (regression equation $Y=b_0+b_1X+$ disturbance or error term) together with prior theoretical expectations about the sign and size of the parameters of the function.
2. Collection of data on the variables of the model and estimation of the coefficient of the functions with appropriate econometric techniques
3. Evaluate of the estimated coefficient of the function on the basis of economic, statistical, and econometric criteria.
4. Application and forecasting (we use estimated model to forecast the dependent value in the future). [4] (Salvator Dominick and Reagle Derrick, (2001)," Statistics and Econometrics", second edition, McGraw-Hill).

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. [5] (Jack Johnston, john dinardo, (1997), “Econometric Method” – fourth edition).

3-2: Regression analysis:

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 the independent variables, 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.

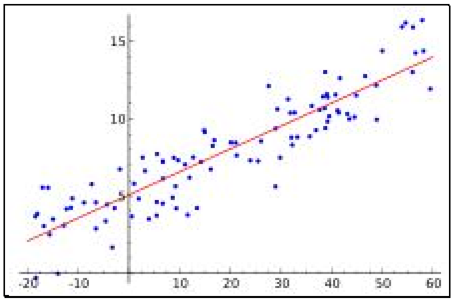
3-3: linear regression model:

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.

3-3-1: Simple linear regression analysis:

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[6](R. carter Hill, William E.Griffiths ," Principle of Econometrics".

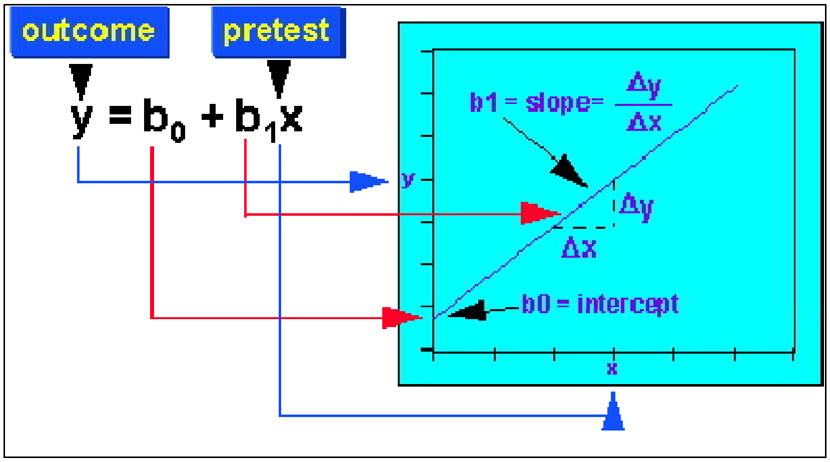
Fig. (3-1) : Scatter diagram:



Source: www.socialresearchmethods.net.

This figure show that there is approximate linear relationship between x &y. In other words, simple linear regression fits a straight line through the set of (n) points in such a way that makes the sum of squared residuals of the model (that is, vertical distances between the points of the data set and the fitted line) as small as possible,

Fig. (3-2): Equation for a straight line.



Source: www.socialresearchmethods.net

The equation is:

$$Y_i = \beta_0 + \beta_1 X_i \qquad (3\text{- }1)$$

Since the points are unlikely to fall precisely on the line, the exact linear relationship in Eq. (3- 1) must be modified to include a random disturbance, error, or stochastic term, u_i:

$$Y_i = \beta_0 + \beta_1 X_i + U_i \qquad (3\text{- }2)$$

The error term is assumed to be:

(1) Normally distributed,

(2) Zero expected value or mean $E(U_i) = 0 \;\forall\; i = 1,2,\ldots,n$

(3) Constant variance $V(U_i) = E[U_i - E(U_i)]^2 = E(U_i)^2 = \sigma_u^2 \;\forall\; i = 1,2,\ldots,n$

(4) The error terms are uncorrelated or unrelated to each other $E(U_iU_j) = 0 \;\forall\; i \neq j$

(5) the explanatory variable assumes fixed values in repeated sampling(so that X_i and u_i are also uncorrelated) $E(U_iX_i) = 0 \;\forall\; i = 1,2,\ldots,n$

Simple linear regression can be used to describe the strength of association between two continuous variables. The correlation between variables helps to develop a prediction equation that will allow to predict the value of one variable based on the value of the other.

Regression is a good way to predict outcomes and explain the relationship between variables. If the correlation between two continuous variables is perfect positive or negative, it will be able to get a perfect prediction about the outcome so the higher the correlation, the more precise about the predictions, so the requirement for simple linear regression the dependent and Independent variable are continuous variable, also to know the proportion of variance in the dependent variable that explained by independent variable

The hypothesis test for simple linear regression:

Null hypothesis (H_0): The regression coefficients = 0

Alternative hypothesis (H_1): The regression coefficients ≠ 0

The assumption for that is:
1- The sample must be representative of the population.
2- The variable of X and Y must have a normal distribution.
3- The variance is homoscedasticity.
4- The relationship between X and Y must be linear.
And the most important assumptions for residuals are:

1- Normal distribution of residuals

2- Homoscedasticity of residuals

Homoscedasticity means that the variance should be constant and the residuals are equal for all values of predicted variable that means :

1. $E(U_i) = 0 \ \forall \ i = 1,2,\dots,n$

2. $V(U_i) = E[U_i - E(U_i)]^2 = E(U_i)^2 = \sigma_u^2 \ \forall \ i = 1,2,\dots,n$

Regression Equation**:**

The regression equation for the straight line is written as:

$$Y_i = \beta_0 + \beta_1 X_i + U_i \qquad (3\text{-}3)$$

Y_i : the predicted value for dependent variable
β_0 : Constant, Intercept, the value of Y when X = 0
$\beta_1 X_i$: it measures the slope of regression line, it is the rate of change in Y with a unit change in X
U_i : it is a random error

3-3-2: Ordinary least square method:

The ordinary-least square method (OLS) is a technique for fitting the "best" straight line to the sample of XY observation. It involves minimizing the sum of the squared (vertical) deviations of points from the line:

$$\underset{i}{Min} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \qquad (3\text{-}4)$$

Where Yi refers to the actual observation, and $\hat{Y}_i$ refers to the corresponding fitted values, so that $e_i = Y_i - \hat{Y}_i$ the residual. This gives the following two normal equations:

$$\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \qquad (3\text{-}5)$$

$$\sum_{i=1}^{n} X_i Y_i - (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^{n} X_i - \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0 \qquad (3\text{-}6)$$

where (n) is the number of observations and $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimator of the true parameters β_0 and β_1.

Solving simultaneously Eqs. (3-5) and (3-6), we get:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right) \Big/ n}{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n} \quad (3\text{-}7)$$

The estimated least-square regression (OLS) equation is then:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (3\text{-}8)$$

The meaning of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$\hat{\beta}_0$: The y intercept, or the value of dependent variable when independent variable is zero

$\hat{\beta}_1$: The slope of the estimated regression line , or the change in dependent variable when a unit change by independent variable.

3-3-3 : Coefficient of Determination R^2:

Is defined as the proportion of the total variation in Y explained by the regression of Y on X. $\mathbf{R^2}$ 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).

$$R^2 = 1- \sum e^2_i / \sum y^2_i \quad (3\text{-}9)$$

3-3-4: 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. The generic form of the linear regression model is:

$$Y = f(X_1, X_2, \ldots, X_K) + \varepsilon$$

$$Y= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdot \cdot \cdot + \beta_k X_K + u \quad (3\text{-}10)$$

Where Y is the dependent or explained variable and $X_1, \ldots, X_K$ are the independent or explanatory variables. This function is commonly called the regression equation of Y on and $X_1, \ldots X_K$. In this setting, *Y* is the regress and X_k, $k=1, \ldots, K$, are the regressors or covariates. The underlying theory will specify the dependent and independent variables in the model.

The term ε is a random disturbance, so named because it "disturbs" an otherwise stable relationship. The disturbance arises for several reasons, primarily because we cannot hope to capture every influence on an economic variable in a model, no matter how elaborate. The net effect, which can be positive or negative, of these omitted factors is captured in the disturbance. There are many other contributors to the disturbance in an empirical model. Probably the most significant is errors of measurement. It is easy to theorize about the relationships among precisely defined variables; it is quite another to obtain accurate measures of these variables. For example, the difficulty of obtaining reasonable measures of profits, interest rates, capital stocks, or, worse yet, flows of services from capital stocks is a recurrent theme in the empirical literature.

We assume that each observation in a sample $(Y_i , X_{i1}, X_{i2}, . . . , X_{iK})$, $i = 1, . . . , n$, is generated by an underlying process described by

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + U_i \qquad (3\text{-}11)$$

$$i = 1, . . . , n$$

and the form of the estimated model is:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdot\ \cdot\ \cdot + \hat{\beta}_k X_{Ki} + u \qquad (3\text{-}12)$$

The observed value of Y_i is the sum of two parts, a deterministic part and the random part, Ui. Our objective is to estimate the unknown parameters of the model, use the data to study the validity of the theoretical propositions, and use the model to predict the variable y. How we proceed from here depends crucially on what we assume about the stochastic process that has led to our observations of the data in hand.

Equation (3-11) contain n equations and (k+1) parameters so that:

$$\left.\begin{array}{l} i = 1 \Rightarrow Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + ... + \beta_k X_{k1} + U_1 \\ i = 2 \Rightarrow Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + ... + \beta_k X_{k2} + U_2 \\ \vdots \qquad\qquad\qquad \vdots \\ i = n \Rightarrow Y_n = \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + ... + \beta_k X_{kn} + U_n \end{array}\right\} \qquad (3\text{-}13)$$

We can write in matrix form as:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} , \ U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}_{n\times 1} , \ \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & ..... & X_{k1} \\ 1 & X_{12} & X_{22} & ..... & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & ..... & X_{kn} \end{bmatrix}_{n\times(k+1)}$$

(3-14)

We can write general liner model as:

$$Y = X\beta + U \qquad (3\text{-}15)$$

And this is the equation of GLM

Where as :

Y : vector of dependent variable

X : matrix of In dependent variables

β: vector of parameter of the model

U: vector of error term

The assumption for GLM:

In addition to the assumption for simple regression model we have :

1.U vector is normally distributed

2. $E(U_i)=0 \quad \forall\ i=1,2,\dots,n$ so $E(U)=\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}=0$ (3-16)

3. The variance covariance matrix is: $V-Cov(U)=E[U-E(U)][U-E(U)]'$

$$=E(UU)'=E\begin{bmatrix}U_1\\U_2\\\vdots\\U_n\end{bmatrix}[U_1 \quad U_2 \quad \dots\dots \quad U_n]$$

$$=E\begin{bmatrix}U_1^2 & U_1U_2 & \dots & U_1U_n\\U_2U_1 & U_2^2 & \dots & U_2U_n\\\vdots & \vdots & & \vdots\\U_nU_1 & U_nU_2 & \dots & U_n^2\end{bmatrix}=\begin{bmatrix}E(U_1^2) & E(U_1U_2) & \dots & E(U_1U_n)\\E(U_2U_1) & E(U_2^2) & \dots & E(U_2U_n)\\\vdots & \vdots & & \vdots\\E(U_nU_1) & E(U_nU_2) & \dots & E(U_n^2)\end{bmatrix}$$

From the assumption for simple regression model we have

$$E(U_i^2)=\sigma_u^2 \quad \forall\ i=1,2,\dots,n$$
$$E(U_iU_j)=0 \quad \forall\ i\neq j;\quad i,j=1,2,\dots,n$$

So we have

$$E(UU')=\begin{bmatrix}\sigma_u^2 & 0 & \cdots & 0\\ 0 & \sigma_u^2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \sigma_u^2\end{bmatrix}=\sigma_u^2\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\sigma_u^2 I_n \qquad (3\text{-}17)$$

4. From simple regression we have $U_i \sim N(0,\sigma_u^2)$

So here we have $U \sim N(0,\sigma_u^2 I_n)$ (3-18)

5. U is dependent from the explanatory variable :

$$Cov(X',U)=EX'[U-E(U)]$$

$$=E(X'U)=E\left\{\begin{bmatrix}1 & 1 & \cdots & 1\\ X_{11} & X_{12} & \cdots & X_{1n}\\ \vdots & \vdots & & \vdots\\ X_{k1} & X_{k2} & \cdots & X_{kn}\end{bmatrix}.\begin{bmatrix}U_1\\ U_2\\ \vdots\\ U_n\end{bmatrix}\right\}$$

$$=E\begin{bmatrix}U_1+U_2+\ldots+U_n\\ X_{11}U_1+X_{12}U_2+\ldots+X_{1n}U_n\\ \vdots\\ X_{k1}U_1+X_{k2}U_2+\ldots+X_{kn}U_n\end{bmatrix}=E\begin{bmatrix}\sum_{i=1}^{n}U_i\\ \sum_{i=1}^{n}X_{1i}U_i\\ \vdots\\ \sum_{i=1}^{n}X_{ki}U_i\end{bmatrix}$$

$$=\begin{bmatrix}\sum_{i=1}^{n}E(U_i)\\ \sum_{i=1}^{n}X_{1i}E(U_i)\\ \vdots\\ \sum_{i=1}^{n}X_{ki}E(U_i)\end{bmatrix} \quad \text{because} \quad E(U_i)=0 \quad \forall\ i=1,2,\ldots,n \quad \therefore E(X'U)=\begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix}=0 \qquad (3\text{-}19)$$

OLS Estimation :

Estimated model will be : $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki}$ (3-20)

And in matrix form $\hat{Y} = X.\hat{\beta}$

The estimated parameter $\frac{\partial \hat{Y}}{\partial X_j} = \hat{\beta}_j \qquad j = 1,2,\ldots,k$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2 - \ldots - \hat{\beta}_k \overline{X}_k \tag{3-21}$$

And the residual is $e = Y - \hat{Y}$ so

$$Q = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \tag{3-22}$$

$$\frac{\partial Q}{\partial \hat{\beta}} = \begin{bmatrix} \frac{\partial Q}{\partial \hat{\beta}_0} \\ \frac{\partial Q}{\partial \hat{\beta}_1} \\ \vdots \\ \vdots \\ \frac{\partial Q}{\partial \hat{\beta}_k} \end{bmatrix}$$ from this we get $\Rightarrow \hat{\beta} = (X'X)^{-1} X'Y$

$$X'X = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ X_{11} & X_{12} & \ldots & X_{1n} \\ \vdots & \vdots & & \vdots \\ X_{k1} & X_{k2} & \ldots & X_{kn} \end{bmatrix} . \begin{bmatrix} 1 & X_{11} & \ldots & X_{k1} \\ 1 & X_{12} & \ldots & X_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1n} & \ldots & X_{kn} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{2i} & \ldots & \sum_{i=1}^{n} X_{ki} \\ \sum_{i=1}^{n} X_{1i} & \sum_{i=1}^{n} X_{1i}^2 & \sum_{i=1}^{n} X_{1i} X_{2i} & \ldots & \sum_{i=1}^{n} X_{1i} X_{ki} \\ \sum_{i=1}^{n} X_{2i} & \sum_{i=1}^{n} X_{2i} X_{1i} & \sum_{i=1}^{n} X_{2i}^2 & \ldots & \sum_{i=1}^{n} X_{2i} X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ki} & \sum_{i=1}^{n} X_{ki} X_{1i} & \sum_{i=1}^{n} X_{ki} X_{2i} & \ldots & \sum_{i=1}^{n} X_{ki}^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ \vdots & \vdots & & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix} . \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{1i} Y_i \\ \vdots \\ \sum_{i=1}^{n} X_{ki} Y_i \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} n & \sum X_{1i} & \sum X_{2i} & \cdots & \sum X_{ki} \\ & \sum X_{1i}^2 & \sum X_{1i} X_{2i} & \cdots & \sum X_{1i} X_{ki} \\ & & & & \vdots \\ & & & & \vdots \\ & & & & \sum X_{ki}^2 \end{bmatrix}^{-1} . \begin{bmatrix} \sum Y_i \\ \sum X_{1i} Y_i \\ \vdots \\ \vdots \\ \sum X_{ki} Y_i \end{bmatrix} \qquad (3\text{-}23)$$

And this the equation for practical use for the model estimation .[7] (Johnston, J.(1984),” Econometric Methods ”, 3rd ed., McGraw-Hill Book Company, London.

3-4: Model selection:

In its most basic forms, model selection is one of the fundamental tasks of scientific inquiry. Determining the principle that explains a series of observations is often linked directly to a mathematical model predicting those observations. Of the countless number of possible mechanisms and processes that could have produced the data, how can one even begin to choose the best model? The mathematical approach commonly taken decides among a set of candidate models; this set must be chosen by the researcher. Burnham and Anderson (2002) emphasize throughout their book the importance of choosing models based on sound scientific principles, such as understanding of the phenomenological processes or mechanisms.

Once the set of candidate models has been chosen, the mathematical analysis allows us to select the best of these models. What is meant by best is controversial. A good model selection technique will balance goodness of fit with simplicity. More complex models will be better able to adapt their shape to fit the data but the additional parameters may not represent anything useful. The complexity is generally measured by counting the number of parameters in the model.

Model selection techniques can be considered as estimators of some physical quantity, such as the probability of the model producing the given data. The bias and variance are both important measures of the quality of this estimator; a standard example of model selection is that of curve fitting, we must select a curve that describes the function that generated the points.

In statistics, regression model validation is the process of deciding whether the numerical results quantifying hypothesized relationships between variables, obtained from regression analysis, are in fact acceptable as descriptions of the data. The validation process can involve analyzing the goodness of fit of the regression, analyzing whether the regression residuals are random, and checking whether the model's predictive performance deteriorates substantially when applied to data that were not used in model estimation.

3.5: Cross validation:

Cross-validation, sometimes called rotation estimation, is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. It is worth highlighting that in a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (testing dataset). The goal of cross validation is to define a dataset to "test" the model in the training phase (i.e., the validation dataset), in order to limit problems like overfitting, give an insight on how the model will generalize to an independent data set (i.e., an unknown dataset, for instance from a real problem.

One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, multiple rounds of cross-validation are performed using different partitions, and the validation results are averaged over the rounds.

3.5.1: Purpose of cross validation

Suppose we have a model with one or more unknown parameters, and a data set to which the model can be fit (the training data set). The fitting process optimizes the model parameters to make the model fit the training data as well as possible. If we then take an independent sample of validation data from the same population as the training data, it will generally turn out that the model does not fit the validation data as well as it fits the training data. This is called over fitting, and is particularly likely to happen when the size of the training data set is small, or when the number of parameters in the model is large. Cross-validation is a way to predict the fit of a model to a hypothetical validation set when an explicit validation set is not available.

Linear regression provides a simple illustration of overfitting. In linear regression we have real response values Y_1, ..., Y_n, and vector covariates X_1, ..., X_p. We can use least squares to fit a hyper plane $a + b_1X_1 + ... + b_pX_p$ between the Y and X data, and then assess the fit using the mean squared error (MSE):

$$\frac{\sum_i (Y_i - a - b_1 X_{1i} - \cdots - b_p X_{pi})^2}{n}$$

where X_{ji} is the value of variable X_j corresponding to the i^{th} response value Y_i.

It can be shown under mild assumptions that the expected value of the MSE for the training set is $(n - p - 1)/(n + p + 1) < 1$ times the expected value of the MSE for the validation set the expected value is taken over the distribution of training sets. Thus if we fit the model and compute the MSE on the training set, we will get an optimistically biased assessment of how well the model will fit an independent data set. This biased estimate is called the in-sample estimate of the fit, whereas the cross-validation estimate is an out-of-sample estimate.

Since in linear regression it is possible to directly compute the factor $(n - p - 1)/(n + p + 1)$ by which the training MSE underestimates the validation MSE, cross-validation is not practically useful in that setting. In most other regression procedures (e.g. logistic regression), there is no simple formula to make such an adjustment. Cross-validation is, thus, a generally applicable way to predict the performance of a model on a validation set using computation in place of mathematical analysis.

3.5.2. Common types of cross-validation:

1. K-fold cross-validation

In k-fold cross-validation, the original sample is randomly partitioned into k equal size subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k - 1$ sub samples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling (see below) is that all observations are used for both training and validation, and each observation is used for validation exactly once. 10-fold cross-validation is commonly used, but in general k remains an unfixed parameter.

In stratified k-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds. In the case of a dichotomous classification, this means that each fold contains roughly the same proportions of the two types of class labels.

2. 2-fold cross-validation

This is the simplest variation of k-fold cross-validation. Also, called holdout method for each fold, we randomly assign data points to two sets d_0 and d_1, so that both sets are equal size (this is usually implemented by shuffling the data array and then splitting it in two). We then train on d_0 and test on d_1, followed by training on d_1 and testing on d_0.

This has the advantage that our training and test sets are both large, and each data point is used for both training and validation on each fold.

3. Repeated random sub-sampling validation

This method randomly splits the dataset into training and validation data. For each such split, the model is fit to the training data, and predictive accuracy is assessed using the validation data. The results are then averaged over the splits. The advantage of this method (over k-fold cross validation) is that the proportion of the training/validation split is not dependent on the number of iterations (folds). The disadvantage of this method is that some observations may never be selected in the validation subsample, whereas others may be selected more than once. In other words, validation subsets may overlap. This method also exhibits Monte Carlo variation, meaning that the results will vary if the analysis is repeated with different random splits.

In a stratified variant of this approach, the random samples are generated in such a way that the mean response value (i.e. the dependent variable in the regression) is equal in the training and testing sets. This is particularly useful if the responses are dichotomous with an unbalanced representation of the two response values in the data.

4. Leave-one-out cross-validation

As the name suggests, leave-one-out cross-validation (LOOCV) involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. This is the same as a K-fold cross-validation with K being equal to the number of observations in the original sampling.

3.5.3: Computational issues:

Most forms of cross-validation are straightforward to implement as long as an implementation of the prediction method being studied is available. In particular, the prediction method need only be available as a "black box" – there is no need to have access to the internals of its implementation. If the prediction method is expensive to train, cross-validation can be very slow since the training must be carried out repeatedly. In some cases such as least squares and kernel regression, cross-validation can be sped up significantly by pre-computing certain values that are needed repeatedly in the training, or by using fast "updating rules" such as the Sherman–Morrison formula. However one must be careful to preserve the "total blinding" of the validation set from the training procedure, otherwise bias may result. An extreme example of accelerating cross-validation occurs in linear regression, where the results of cross-validation have a closed-form expression known as the prediction residual error sum of squares (PRESS).

3.5.4: Relationship to other forms of validation:

In "true validation," or "holdout validation," a subset of observations is chosen randomly from the initial sample to form a validation or testing set, and the remaining observations are retained as the training data. Normally, less than a third of the initial sample is used for validation data. This would generally not be considered to be cross-validation since only a single partition of the data into training and testing sets is used.

3.6: PRESS statistic

In statistics the predicted residual 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:

$$\mathrm{PRESS} = \sum_{i=1}^{n} (y_i - \hat{y}_{i,-i})^2$$

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.

What is prediction sum of squares (PRESS)?

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 [8], 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 splines . 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 mod.[8] (Allen, D.M.,(1974) "The relationship between variable selection and data augmentation and a method for prediction".

4.1: Introduction:

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 (table1 appendix), the researcher prepare (table 2 appendix) 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 from equ.(3-11) is:

$$GDP = \beta_0 + \beta_1 G + \beta_2 C + \beta_3 I + \beta_4 Net \qquad (4\text{-}1)$$

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 (external expenditure)

4.2: 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 (4-1) : KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.527
Bartlett's Test of Sphericity	Approx. Chi-Square	191.807
	df	10
	Sig.	.000

Source: by the researcher 2014, using SPSS

According to table (4-3) 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.

4.3: Descriptive of the study variables:

In this study we have five variables as follows:

4.3.1: Government expenditure:

G : the first independent variable (Government expenditure)

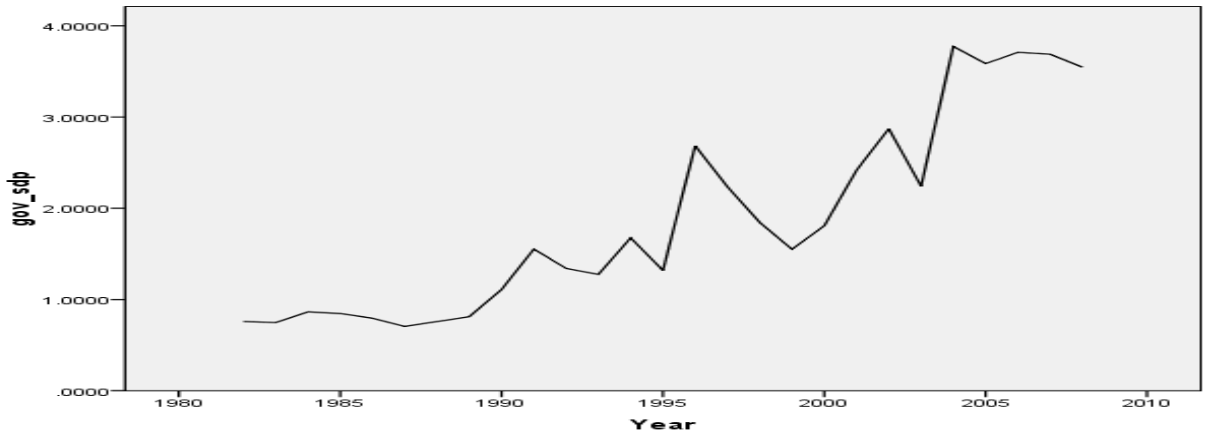
Table (4-2): Descriptive Statistics of Government expenditure

	N	Minimum	Maximum	Mean	Std. Deviation
gov	27	.7045	3.7740	1.870833	1.0650637

Source: by the researcher 2014, using SPSS

Table (4-2) provides some information about the mean of government expenditure which is (1.8708) million Sudanese pounds with standard deviation equal to (1.0651) & minimum value is (.7045) in year 1988 & maximum value is (3.774) in year 2004, and here we notice that the high number of government expenditure after the applying the policy of decentralization in Sudan which take place in 2003.

Fig (4-1): general trend of Government expenditure



Source: by the researcher 2014, using SPSS

From this fig. we notice that high fluctuation of the government expenditure between 1995 and 2008 and high expenditure in 2003.

4.3.2 : consumption expenditure:

C : the second independent variable (consumption expenditure)

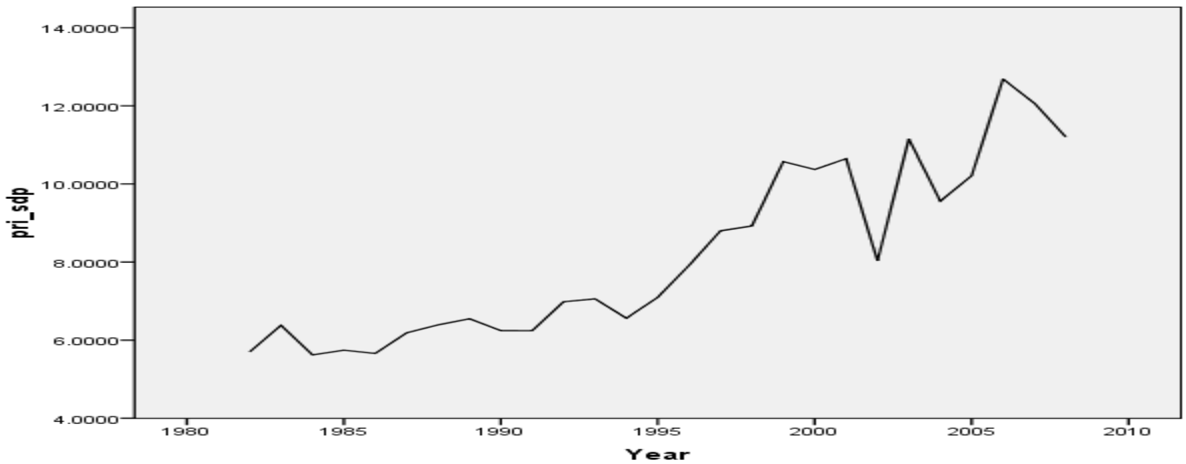
Table (4-3): Descriptive Statistics of consumption expenditure

	N	Minimum	Maximum	Mean	Std. Deviation
Pvi. con.	27	5.6230	12.6866	8.1691000	2.2280534

Source: by the researcher 2014, using SPSS

From table (4-3) we find the mean of consumption expenditure is (8.1691) million Sudanese pounds with standard deviation equal to (2.2281) & minimum value is (5.623) in year 1985 & maximum value is (12.6866) in year 2006.

Fig (4-2): general trend of consumption expenditure



Source: by the researcher 2014, using SPSS

This fig. show the general trend of the consumption expenditure which increases sharply after 1994 with fluctuation between 2002 and 2008. The expenditure after the applying the policy of decentralization in Sudan which take place in 2003 also affect the consumption sector.

4.3.3 Investment expenditure:

I : the third independent variable (investment expenditure)

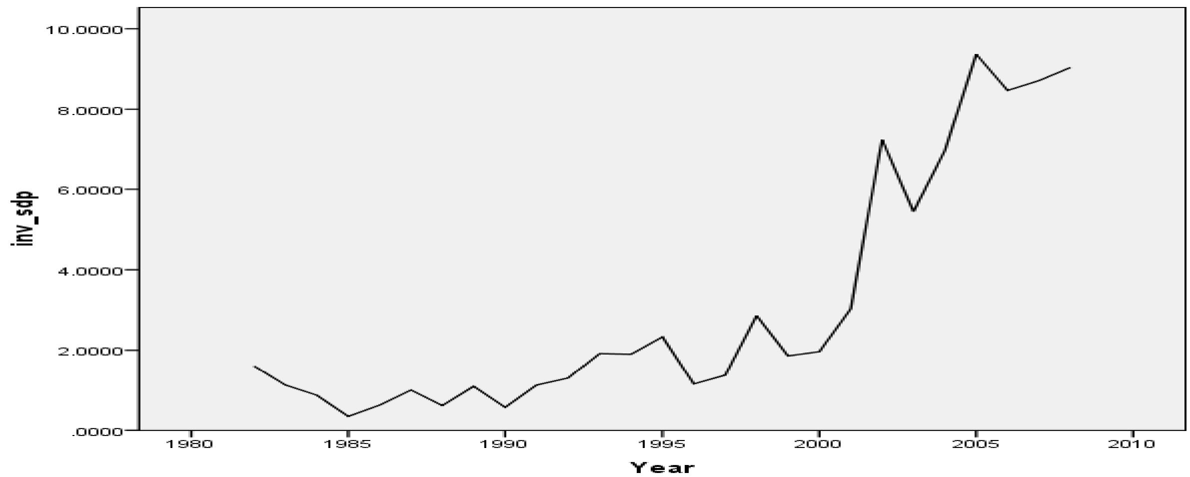
Table(4-4): Descriptive Statistics of investment expenditure

	N	Minimum	Maximum	Mean	Std. Deviation
inv_sdp	27	.3458	9.3659	3.107363	3.0273605

Source: by the researcher 2014, using SPSS

Table (4-4) show the mean of investment expenditure is (3.1074) million Sudanese pounds with standard deviation equal to (3.0274) & minimum value is (.3458) in year 1983 & maximum value is (9.3659) in year 2005 which looks high variation between 2000 and 2005 and this due to exporting the crude oil from Sudan which started in 1999 and many capital foreign came to Sudan for investment.

Fig (4-3): general trend of Investment expenditure



Source: by the researcher 2014, using SPSS

This fig. show the very high increase of the investment expenditure after 2000 and this due to export of the crude oil from Sudan as said before.

4.3.4 External expenditure:

Net: the fourth independent variable (external expenditure)

Which mean the difference net between export and import

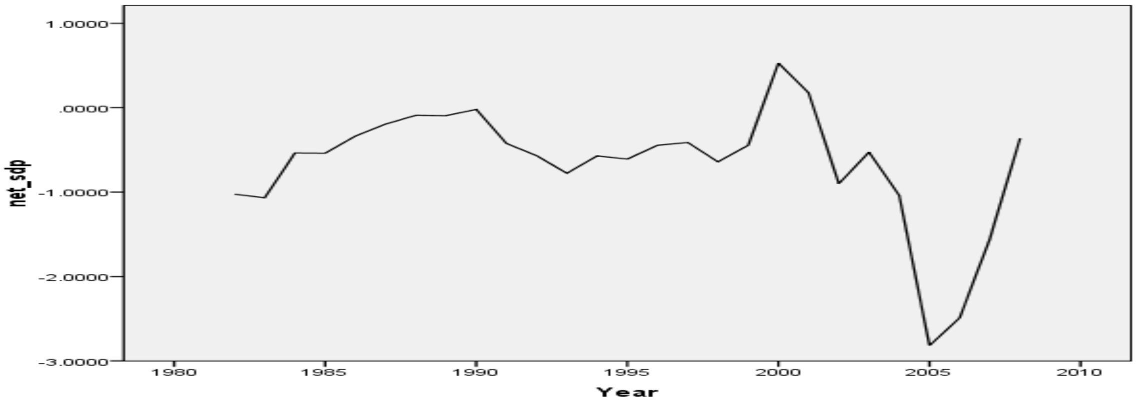
Table (4-5): Descriptive Statistics of external expenditure

	N	Minimum	Maximum	Mean	Std. Deviation
net_exp_imp	27	-2.8180E2	.528500	-.6584111	7.11490544

Source: by the researcher 2014, using SPSS

The mean of external expenditure is (- 6.5841E1) million Sudanese pounds with standard deviation equal to (7.11491), here we notice that the minimum value is negative (-.28180E2) in year 2005 which indicate that the imports is greater than exports & maximum value is (.5285) in year 2000 which looks high variation between 2000 and 2005 and this due to exporting the crude oil from Sudan.

Fig (4-4): general trend of external expenditure



Source: by the researcher 2014, using SPSS

From this fig. we notice that only in the period between 2000 and 2001 the net between export and import is positive and all the rest is negative and this due to excesses of the import more than export and very high decrease in export in the period between 2005 till 2007.

4.3.5 : Gross Domestic Product expenditure:

GDP: dependent variable (Gross Domestic Product)

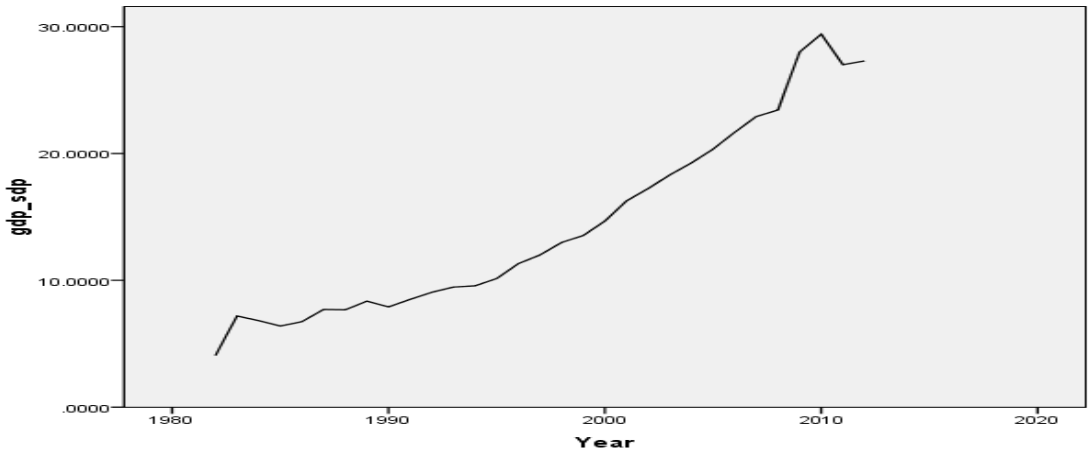
Table (4-6): Descriptive Statistics of GDP:

	N	Minimum	Maximum	Mean	Std. Deviation
gdp_sdp	27	4.0701	23.4244	14.36225	7.4890083

Source: by the researcher 2014, using SPSS

The mean of GDP expenditure is (14.36225) million Sudanese pounds with standard deviation equal to (7.4890) & minimum value is (4.0701) in year 1982 & maximum value is (23.4244) in year 2008 which represent growth in GDP and in economy as whole.

Fig (4-5): general trend of GDP expenditure :



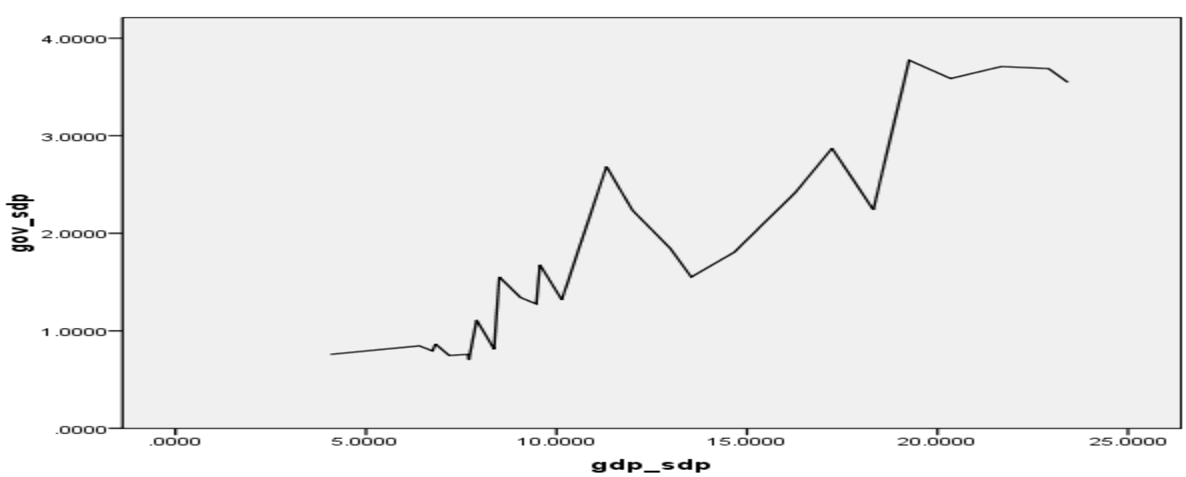
Source: by the researcher 2014, using SPSS

This fig. show the general trend of the GDP expenditure which increases with the time normally reflecting growth in economy of the Sudan between 2000 and 2010 this is healthy for the economy but with limitation, and that is, the increase should be in investment expenditure and net between export and import to reflect the well done of the economy more than increasing in government expenditure and consumption .

4.4: The relationship between GDP and the type of Expenditure

4.4.1 The relationship between GDP and government expenditure

Fig.(4-6): The relationship between GDP and government expenditure

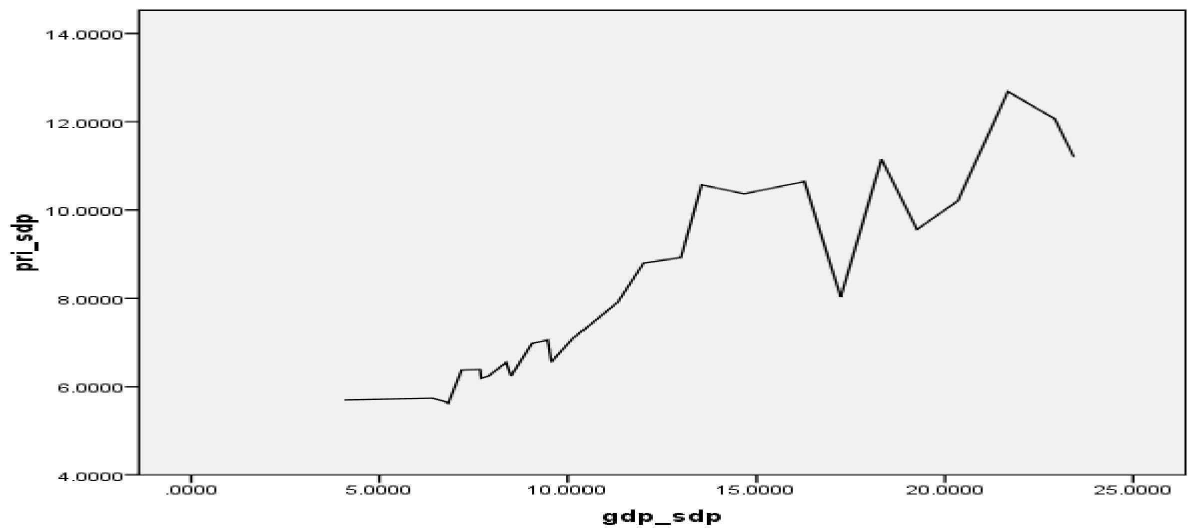


Source: by the researcher 2014, using SPSS

From this fig. we notice that there is positive (but not exact linear) relationship between GDP and government expenditure

4.4.2: The relationship between GDP and consumption expenditure

Fig .(4-7): The relationship between GDP and consumption expenditure

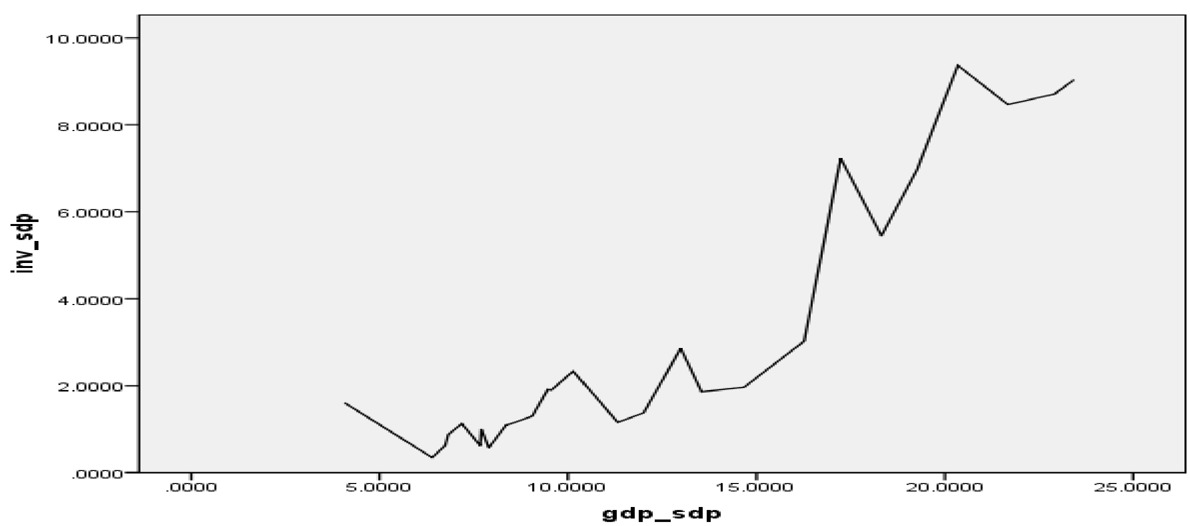


Source: by the researcher 2014, using SPSS

From this fig. we notice that there is positive (but not exact linear) relationship between GDP and consumption expenditure

4.4.3: The relationship between GDP and investment expenditure

Fig .4.8: The relationship between GDP and investment expenditure

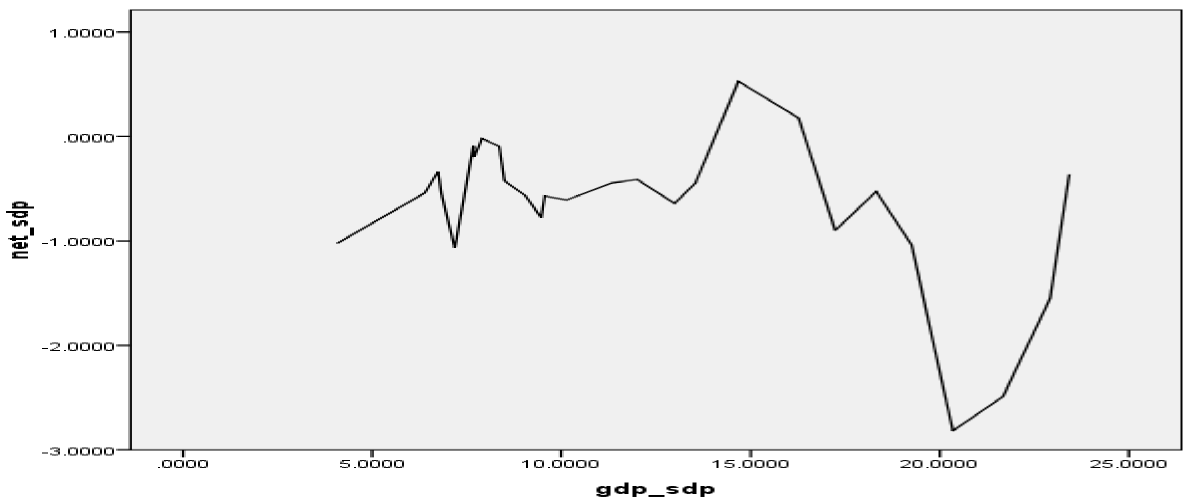


Source: by the researcher 2014, using SPSS

From this fig. we notice that there is no relationship between GDP and investment expenditure

4.4.4 : The relationship between GDP and external expenditure

Fig .4.9: The relationship between GDP and external expenditure



Source: by the researcher 2014, using SPSS

From this fig. we notice that there is no relationship between GDP and external expenditure

4.5 : Test the normality of the data

To test the normality of the data we test the following hypothesis:

H_0: the data follow the normal distribution

H_1: the data not follow the normal distribution

Table(4-7): Tests the Normality of the data

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
gov_sdp	.139	27	.192	.871	27	.003
Con pri_sdp	.203	27	.006	.891	27	.008
inv_sdp	.277	27	.000	.761	27	.000
net_sdp	.213	27	.003	.845	27	.001
gdp_sdp	.170	27	.043	.911	27	.024

Source: by the researcher 2014, using SPSS

From the table (4-7) 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 accept H_1 that the data not follow the normal distribution so we use the log transformation by taking the log of the actual data and the result is the following table.

Table (4-8) : Tests of Normality by using the log transformation

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
log_gov_sdp	.150	27	.124	.913	27	.027
log_pri_sdp	.174	27	.035	.906	27	.019
log_invi_sdp	.138	27	.200*	.935	27	.091
log_net	.135	27	.200*	.929	27	.065
log_gdp_sdp	.112	27	.200*	.957	27	.321

Source: by the researcher 2014, using SPSS

From the above table only one variable not normally distributed (log consumption) because its p value is less than 0.05 but the other variable is normally distributed so we accept H_0 that the data follow the normal distribution

And this prove our first hypothesis about the normality of the data

4.6: Multiple regression model and the significance test:

4.6.1 : Descriptive statistics of the variables of the model:

Table (4-9) : Descriptive statistics of the variable of the model

	Mean	Std. Deviation	N
log_gdp_sdp	1.0478	.20078	27
log_gov_sdp	.2025	.25304	27
log_pri_sdp	.8972	.11511	27
log_invi_sdp	.3018	.41462	27
log_net	-.2154-	.15264	27

Source: by the researcher 2014, using SPSS

From this table we find the mean and standard deviation of the dependent and independent variables of the model as above.

Table(4-10): Summary of the model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972a	.946	.936	.05094

Source: by the researcher 2014, using SPSS

From this table we 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 d (GDP) have been explained be the change in the independent variables, and here the R square is very high .

4.6.2: The significance test of the model:

To test the significance of the model we test the following hypothesis:

H_0: the regression model is not significant

H_1: the regression model is significant

Table (4-11): Analysis of variance table (ANOVA):

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.991	4	.248	95.471	.000a
	Residual	.057	22	.003		
	Total	1.048	26			

Source: by the researcher 2014, using SPSS

From this ANOVA table we find that the whole model is significant because P value for the model is significant < .05 so we reject H_0 and accept H_1which mean the model is significant at 5% & at 1% significance level and that means one parameter at least not equal to zero.

To test the significance of the coefficients for the model we test the following hypothesis:

1. H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

   H_1: not all the β_i value are 0

Table (4-12): Significant of the coefficients of the model:

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.394	.171		2.306	.031
	log_gov_sdp	.336	.089	.423	3.790	.001
	log_pri_sdp	.662	.198	.379	3.346	.003
	log_invi_sdp	.090	.050	.185	1.803	.085
	log_net	.161	.075	.123	2.160	.042

Source: by the researcher 2014, using SPSS

From this table we find that all the coefficients of the parameters is significant because t value for the them is significant < .05 except for the investment not significant because t value >.05 that means investment is not statistically important for the model.

From the above table, our model according to equ. (3-11) is:

$$\text{Log GDP} = 0.394+0.336 \text{ log gov} +0.662 \text{ log pri} + 0.161 \text{ log net}$$

Table (4-13): Statistics of the residual:

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	.7725	1.3582	1.0478	.19523	27
Residual	-.19842-	.04396	.00000	.04686	27
Std. Predicted Value	-1.410-	1.590	.000	1.000	27
Std. Residual	-3.895-	.863	.000	.920	27

Source: by the researcher 2014, using SPSS

This table shows minimum, maximum, mean and std. deviation of the predicted value of GDP and the residual from the prediction model and the sum of square of the residual ($\sum e^2$) is equal to (0.1147) and that means the summation of the square residual that between the actual value and the predicted value is equal to (0.1147)

4.7: Prediction residual sum of square (PRESS)

The best model according to PRESS method is:

log gdp sdp = 0.275 + 0.409 log gov sdp + 0.801 log pri sdp + 0.132 log net

4-7-1: Summary of PRESS :

Table(4-14):

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

Source: by the researcher 2014, using Minitab

We conclude that the model which done using the PRESS method is the best model because their residual sum of square which equal (0.0655) is less than the model with the same set of the variables done by the usual regression model which their residual sum of square equal to (0.1147) and when the residual sum of square is less that means the model is the best model.

4-8: Prediction:

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

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

From this table we notice 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.

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.
- We use PRESS method to assess the model's predictive ability for the future value for GDP and to calculate the predicted R^2.
- We 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

Recommendation

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

- 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 .
- Encourage the economy to go towards international market by more exports.
- 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
- Above all, the Statistical analysis for all economic procedure should be supported to hold a scientific approach for every policy.

References:

1. Dutta , Subhendu, (2004)“Introductory- Economics” .
2. A. Raheem, Mayson,(2012),” Estimation of Sudan's GDP Function”.
3. Abdul Hakim, Aqeel ,(2006),” the impact of exports on GDP in the Republic of Yemen” (Adan magazine).
4. Salvator Dominick and Reagle Derrick, (2001),” Statistics and Econometrics”, second edition, McGraw-Hill.
5. Jack Johnston, john dinardo, (1997), “Econometric Method” – fourth edition-
6. R. carter Hill, William E.Griffiths ,” Principle of Econometrics”.
7. Allen, D.M.,(1974) “The relationship between variable selection and data augmentation and a method for prediction”
8. Johnston, J.(1984),” Econometric Methods ”, 3rd ed., McGraw-Hill Book Company, London

المراجع العربيه:

1. عبد الرجمن، عبد المحمود محمد (1417هـ)، "مقدمه في الاقتصاد القياسي" ، مطابع جامعة الملك سعود ، الرياض
2. الصياد، جلال مصطفي ،(1993)، "الإستدلال الإحصائي"، دار المريخ للنشر الرياض.
3. العقيلي، صالح إرشيد- الشايب ، سامر محمد (1998)، "التحليل الإحصائي يإستخدام البرنامج SPSS، عمان دار الشروق .
4. أمين، اسامه ربيع،(2008)، "التحليل الإحصائي للمتغيرات المتعدده بإستخدام برنامج SPSS ، جامعة المنوفيه.

Appendices:

Table (1): GDP by type of expenditure at constant prices (1982-2008)

year	Gov	Con/pri	Inv	Exp	Emp	GDP
1982	0.7586	5.6994	1.6066	0.6849	1.7094	4.0701
1983	0.8461	5.7436	0.3458	0.4184	0.9573	6.3966
1984	0.7943	5.6595	0.6258	0.4827	0.8194	6.7429
1985	0.8642	5.623	0.8744	0.6525	1.1888	6.8253
1986	0.747	6.3798	1.1262	0.7497	1.8175	7.1852
1987	0.7584	6.3918	0.6163	0.5643	0.6545	7.6763
1988	0.7045	6.1885	1.0027	0.4747	0.6686	7.7018
1989	1.1093	6.2446	0.5708	0.441	0.4612	7.9045
1990	0.8123	6.5488	1.0953	0.6074	0.7019	8.3619
1991	1.5527	6.243	1.1256	0.3566	0.7797	8.4982
1992	1.3411	6.9847	1.3005	0.5284	1.0975	9.0572
1993	1.2745	7.058	1.9157	0.4929	1.2701	9.471
1994	1.6756	6.5614	1.8998	0.6383	1.2088	9.5663
1995	1.3182	7.098	2.3322	0.507	1.1154	10.14
1996	2.685	7.917	1.1556	0.4304	0.8756	11.3124
1997	2.235	8.8005	1.3741	0.4932	0.9052	11.9976
1998	1.845	8.9248	2.8583	0.4693	1.111	12.9864
1999	1.5509	10.5711	1.8598	0.6024	1.0478	13.5364
2000	1.8073	10.371	1.9646	1.441	0.9125	14.6714
2001	2.4169	10.646	3.0269	1.557	1.381	16.2658
2002	2.8706	8.0308	7.2308	1.799	2.699	17.2322

2003	2.241	11.1555	5.4465	1.833	2.36	18.316
2004	3.774	9.5603	6.9623	2.053	3.093	19.2566
2005	3.586	10.2104	9.3659	2.332	5.15	20.3443
2006	3.7093	12.6866	8.4662	2.995	5.484	21.6731
2007	3.6882	12.0655	8.7123	3.496	5.046	22.916
2008	3.5465	11.2021	9.0378	4.236	4.598	23.4244

Source: central Bureau of statistics – national accounts administration

Table (2): GDP by type of expenditure at constant prices (1982-2008) after transforming by log :

year	log gov sdp	log pri sdp	log inv sdp	log net	log gdp sdp
1982	-0.119987162	0.755829138	0.205907763	-0.397216533	0.60960508
1983	-0.072578305	0.759184187	-0.461175011	-0.359456383	0.805949194
1984	-0.100015437	0.752778064	-0.203564441	-0.229818661	0.828846719
1985	-0.063385738	0.749968084	-0.058289851	-0.26052828	0.834121745
1986	-0.126679398	0.804807064	0.051615523	-0.384586909	0.856438861
1987	-0.120101676	0.805623177	-0.210207832	-0.064399601	0.885151939
1988	-0.152119003	0.791585395	0.001171015	-0.148747141	0.886592237
1989	0.045049013	0.795504625	-0.243516036	-0.019450709	0.897874404
1990	-0.090283547	0.816161727	0.039533088	-0.062800455	0.92230497
1991	0.191087553	0.795393335	0.051384084	-0.339746195	0.929326948
1992	0.127461163	0.844147757	0.114110357	-0.31744172	0.956993958
1993	0.10533984	0.848681654	0.282327499	-0.411079098	0.976395837
1994	0.224170352	0.816996514	0.278707883	-0.277329607	0.980743996
1995	0.119981307	0.851135995	0.367765791	-0.342422681	1.006037955
1996	0.42894429	0.898560645	0.062807533	-0.30843349	1.053554753
1997	0.349277527	0.944507347	0.13801834	-0.263721477	1.079094378
1998	0.26599637	0.950598493	0.456107809	-0.374263505	1.113488776
1999	0.190583796	1.024120181	0.269466243	-0.240393431	1.131503179
2000	0.257030249	1.015820634	0.29327414	0.198431108	1.166471558
2001	0.383258682	1.027186462	0.480998073	0.052094934	1.211275428
2002	0.457972681	0.90475881	0.859186349	-0.176171721	1.236340726
2003	0.350441857	1.047489041	0.736117508	-0.109749538	1.262830635
2004	0.576801896	0.980471521	0.842752733	-0.177990971	1.284579609
2005	0.554610285	1.009042756	0.971549516	-0.344078683	1.308442751
2006	0.56929196	1.103345247	0.927688524	-0.262700619	1.335921035
2007	0.566814464	1.081545324	0.940132822	-0.159375822	1.360138814
2008	0.549799964	1.049299445	0.956062726	-0.035613015	1.369668476

Source: by the researcher 2014, using

PRESS RESULT:

The researcher use Minitab software for conducting PRESS and get the following result :

(1)Regression Analysis: log gdp sdp versus log gov sdp, log pri sdp, log inv sdp, log net

The regression equation is:

log gdp sdp = 0.394 + 0.336 log gov sdp + 0.662 log pri sdp + 0.0895 log inv sdp + 0.161 log net

Predictor	Coef	SE Coef	T	P
Constant	0.3937	0.1707	2.31	0.031
log gov sdp	0.33574	0.08858	3.79	0.001
log pri sdp	0.6618	0.1978	3.35	0.003
log inv sdp	0.08950	0.04965	1.80	0.085
log net	0.16122	0.07463	2.16	0.042

S = 0.0509419 R-Sq = 94.6% R-Sq(adj) = 93.6%

PRESS = 0.0980996 R-Sq(pred) = 90.64%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	0.99102	0.24775	95.47	0.000
Residual Error	22	0.05709	0.00260		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log pri sdp	1	0.08340
log inv sdp	1	0.00485
log net	1	0.01211

(2)Regression Analysis: log gdp sdp versus log gov sdp, log pri sdp, log inv sdp

The regression equation is

log gdp sdp = 0.196 + 0.307 log gov sdp + 0.858 log pri sdp + 0.0662 log inv sdp

Predictor	Coef	SE Coef	T	P
Constant	0.1958	0.1551	1.26	0.219
log gov sdp	0.30688	0.09428	3.25	0.003
log pri sdp	0.8580	0.1891	4.54	0.000
log inv sdp	0.06623	0.05219	1.27	0.217

S = 0.0548524 R-Sq = 93.4% R-Sq(adj) = 92.5%

PRESS = 0.112143 R-Sq(pred) = 89.30%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.97891	0.32630	108.45	0.000
Residual Error	23	0.06920	0.00301		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log pri sdp	1	0.08340
log inv sdp	1	0.00485

(3) Regression Analysis: log gdp sdp versus log gov sdp, log pri sdp, log net

The regression equation is

log gdp sdp = 0.275 + 0.409 log gov sdp + 0.801 log pri sdp + 0.132 log net

Predictor	Coef	SE Coef	T	P
Constant	0.2749	0.1650	1.67	0.109
log gov sdp	0.40917	0.08241	4.96	0.000
log pri sdp	0.8007	0.1908	4.20	0.000
log net	0.13204	0.07633	1.73	0.097

S = 0.0533746   R-Sq = 93.7%   R-Sq(adj) = 92.9%

PRESS = 0.0915093   R-Sq(pred) = 91.27%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.98259	0.32753	114.97	0.000
Residual Error	23	0.06552	0.00285		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log pri sdp	1	0.08340
log net	1	0.00852

(4)Regression Analysis: log gdp sdp versus log gov sdp, log inv sdp, log net

The regression equation is

log gdp sdp = 0.961 + 0.491 log gov sdp + 0.154 log inv sdp + 0.276 log net

Predictor	Coef	SE Coef	T	P
Constant	0.96112	0.02403	39.99	0.000
log gov sdp	0.49139	0.09056	5.43	0.000
log inv sdp	0.15429	0.05493	2.81	0.010
log net	0.27596	0.07964	3.46	0.002

S = 0.0612025 R-Sq = 91.8% R-Sq(adj) = 90.7%

PRESS = 0.131741 R-Sq(pred) = 87.43%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.96196	0.32065	85.60	0.000
Residual Error	23	0.08615	0.00375		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log inv sdp	1	0.02633
log net	1	0.04497

(5)Regression Analysis: log gdp sdp versus log pri sdp, log inv sdp, log net

The regression equation is

log gdp sdp = 0.073 + 1.06 log pri sdp + 0.176 log inv sdp + 0.119 log net

Predictor	Coef	SE Coef	T	P
Constant	0.0733	0.1865	0.39	0.698
log pri sdp	1.0554	0.2116	4.99	0.000
log inv sdp	0.17605	0.05544	3.18	0.004
log net	0 .11855	0.09277	1.28	0.214

S = 0.0640574 R-Sq = 91.0% R-Sq(adj) = 89.8%

PRESS = 0.141003 R-Sq(pred) = 86.55%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	0.95373	0.31791	77.48	0.000

Residual Error	23	0.09438	0.00410
Total	26	1.04811	

Source	DF	Seq SS
log pri sdp	1	0.91209
log inv sdp	1	0.03494
log net	1	0.00670

(6)Regression Analysis: log gdp sdp versus log gov sdp, log pri sdp

The regression equation is

log gdp sdp = 0.131 + 0.368 log gov sdp + 0.938 log pri sdp

Predictor	Coef	SE Coef	T	P
Constant	0.1313	0.1484	0.88	0.385
log gov sdp	0.36796	0.08210	4.48	0.000
log pri sdp	0.9384	0.1805	5.20	0.000

S = 0.0555456 R-Sq = 92.9% R-Sq(adj) = 92.3%

PRESS = 0.0946407 R-Sq(pred) = 90.97%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.97406	0.48703	157.85	0.000
Residual Error	24	0.07405	0.00309		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log pri sdp	1	0.08340

(7)Regression Analysis: log gdp sdp versus log gov sdp, log inv sdp

The regression equation is

log gdp sdp = 0.897 + 0.529 log gov sdp + 0.145 log inv sdp

Predictor	Coef	SE Coef	T	P
Constant	0.89672	0.01840	48.72	0.000
log gov sdp	0.5289	0.1086	4.87	0.000
log inv sdp	0.14549	0.06627	2.20	0.038

S = 0.0739150 R-Sq = 87.5% R-Sq(adj) = 86.4%

PRESS = 0.185746 R-Sq(pred) = 82.28%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.91699	0.45849	83.92	0.000
Residual Error	24	0.13112	0.00546		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log inv sdp	1	0.02633

(8)Regression Analysis: log gdp sdp versus log gov sdp, log net

The regression equation is

log gdp sdp = 0.962 + 0.707 log gov sdp + 0.266 log net

Predictor	Coef	SE Coef	T	P
Constant	0.96177	0.02726	35.27	0.000
log gov sdp	0.70710	0.05444	12.99	0.000
log net	0.26561	0.09026	2.94	0.007

S = 0.0694322 R-Sq = 89.0% R-Sq(adj) = 88.0%

PRESS = 0.145225 R-Sq(pred) = 86.14%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.93241	0.46620	96.71	0.000
Residual Error	24	0.11570	0.00482		
Total	26	1.04811			

Source	DF	Seq SS
log gov sdp	1	0.89066
log net	1	0.04175

(9) Regression Analysis: log gdp sdp versus log pri sdp, log inv sdp

The regression equation is

log gdp sdp = - 0.055 + 1.18 log pri sdp + 0.153 log inv sdp

Predictor	Coef	SE Coef	T	P
Constant	-0.0549	0.1593	-0.34	0.733
log pri sdp	1.1776	0.1913	6.16	0.000
log inv sdp	0.15295	0 .05310	2 .88	0.008

S = 0.0648966 R-Sq = 90.4% R-Sq(adj) = 89.6%

PRESS = 0.143575 R-Sq(pred) = 86.30%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.94703	0.47352	112.43	0.000
Residual Error	24	0.10108	0.00421		
Total	26	1.04811			

Source	DF	Seq SS
log pri sdp	1	0.91209
log inv sdp	1	0.03494

(10) Regression Analysis: log gdp sdp versus log pri sdp, log net

The regression equation is

log gdp sdp = - 0.398 + 1.62 log pri sdp + 0.022 log net

Predictor	Coef	SE Coef	T	P
Constant	-0.3980	0.1326	-3.00	0.006
log pri sdp	1.6168	0.1365	11.84	0.000
log net	0.0225	0.1030	0.22	0.829

S = 0.0752083 R-Sq = 87.0% R-Sq(adj) = 86.0%

PRESS = 0.171893 R-Sq(pred) = 83.60%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.91236	0.45618	80.65	0.000
Residual Error	24	0.13575	0.00566		
Total	26	1.04811			

Source	DF	Seq SS
log pri sdp	1	0.91209
log net	1	0.00027

(11) Regression Analysis: log gdp sdp versus log inv sdp, log net

The regression equation is

log gdp sdp = 0.995 + 0.407 log inv sdp + 0.328 log net

Predictor	Coef	SE Coef	T	P
Constant	0.99548	0.03427	29.05	0.000
log inv sdp	0.40706	0.04303	9.46	0.000
log net	0.3276	0.1169	2.80	0.010

S = 0.0904691   R-Sq = 81.3%   R-Sq(adj) = 79.7%

PRESS = 0.238398   R-Sq(pred) = 77.25%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.85168	0.42584	52.03	0.000
Residual Error	24	0.19643	0.00818		
Total	26	1.04811			

Source	DF	Seq SS
log inv sdp	1	0.78737
log net	1	0.06431