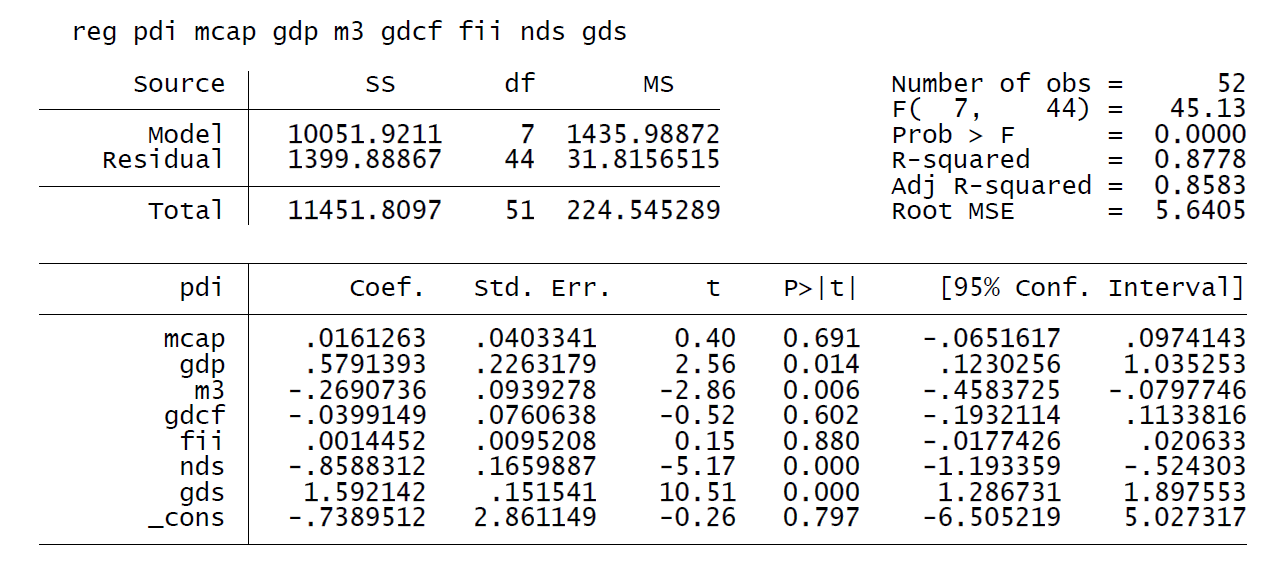
# ASSIGNMENT 1 – Simultaneous Equation Models (SEM)

1. **Dataset**

The dataset that we are going to use consists of mixture of variables such as gdp, market cap, nds as so on. There in all 8 variables.

1. The first step is to regress a variable over all the other variables to find out if there is presence of endogeneous variables.
2. Here we are going to regress “pdi” over all the other variables and look at some of the statistics. Running the below command in STATA gives the output shown in the following table.



1. ***F-statistic and Probability***

F statistic tells overall level of significance

Null hypothesis – is the model significant. As prob is less than 0.05 hence reject and the values ae statistically significant at all levels of confidence intervals. This means our regression model has some explanatory powers.

The R Square here says that, 87.7% of the pdi can be explained by its other parameters on which it is regressed.

The coefficients here show the nature of the relationship. When the values of the parameters is zero then the log pdi is -.7389.

The positive values of the coefficients show that they have a positive correlation with the pdi. Similarly negative values would indicate the negative correlation.

1 unit increase in mcap causes .016 units of change in pdi

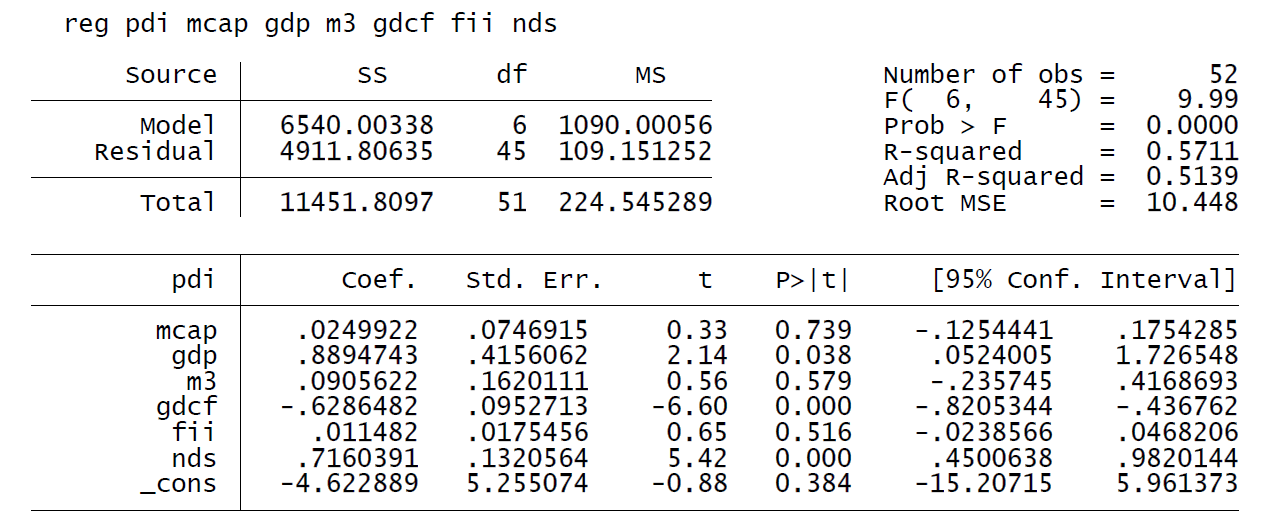
The t values indicate the important parameters that actually affect the pdi

1. **Test of endogeneity**

Here we will find out if the variables that we are dealing with are endogenous or not. Endogenous variables are those variables which are dependent on each other. For example, if you have represented any dependent variable say “y” in terms of independent variables say x1, x2, x3, for a simple regression equation in such case only one variable (y) is dependent of variables x1, x2, x3. But when there is presence of endogeneity, the dependence is both ways i.e “y” is dependent on x1, x2, or x3 also either of x1, x2, and x3 is also dependent on “y”.

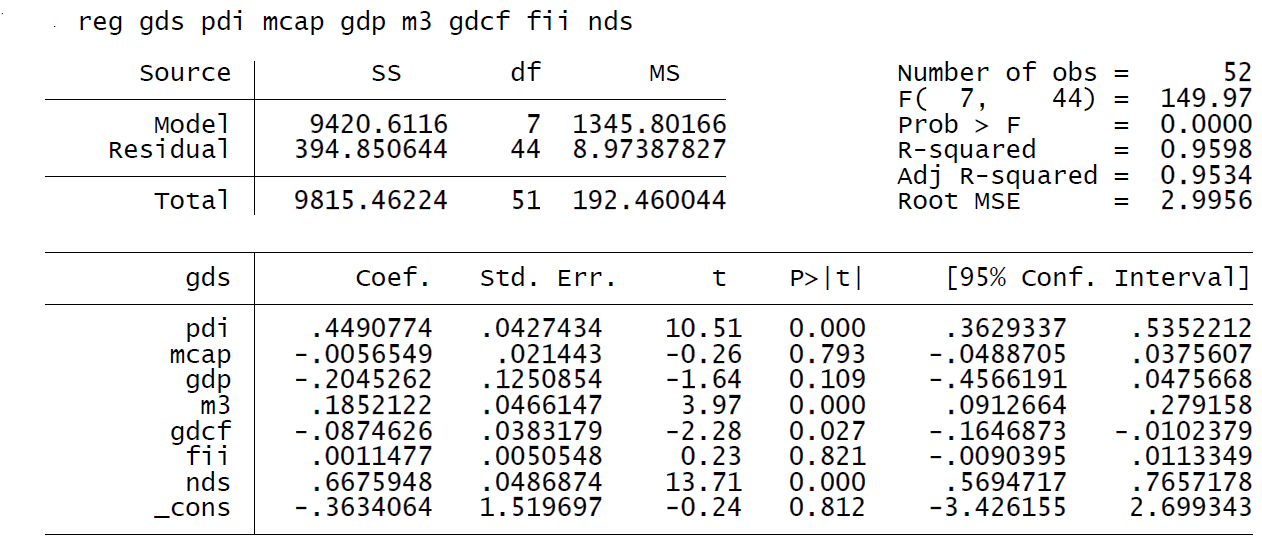
In our case, we don’t know if the endogeneity condition exists or not. So, let’s say that we have a doubt on “gds” meaning that it is endogenous. So, in order to check the endogeneity of the variable, we remove that particular variable and regress all the other variables together.

Run the following command in STATA – **reg pdi mcap gdp m3 gdcf fii nds**



We are now going to regress “gds” over all other variables. Run the following command in STATA.

**Command – reg gds pdi mcap gdp m3 gdcf fii nds**

****

Interpreting the above table, we can see that for “gds” as our independent variable, “pdi” is significant. So, when, “pdi” was used as an independent variable before “gds” was significant and now when “pdi” was replaced with “gds” as the independent variable, “pdi” was seen to be significant. So, this means that both “pdi” and “gds” are dependent on each other. This proves that, endogeneity exists between these variables. Such equations in which endogeneity exists are called as Simultaneous equations (SEM) and these equations cannot be solved by using Ordinary Least Square Methods (OLS).

**METHOD FOR SEM**

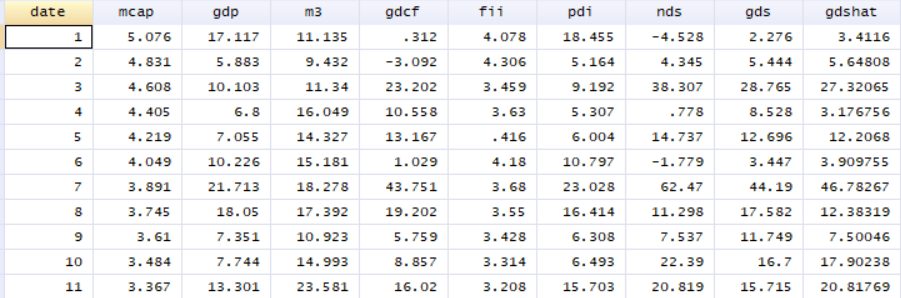
1. **Indirect Least Squares (ILS)**

For a just or exactly identified structural equation, the method of obtaining the estimates of the structural coefficients from the OLS estimates of the reduced-form coefficients is known as the **method of indirect least squares (ILS),** and the estimates thus obtained are known as the **indirect least squares estimates.** ILS involves the following three steps:

Here, we are going to apply ILS method. We are going to predict the residual value for “gds” called “gdshat”. Run the following command in SATA.

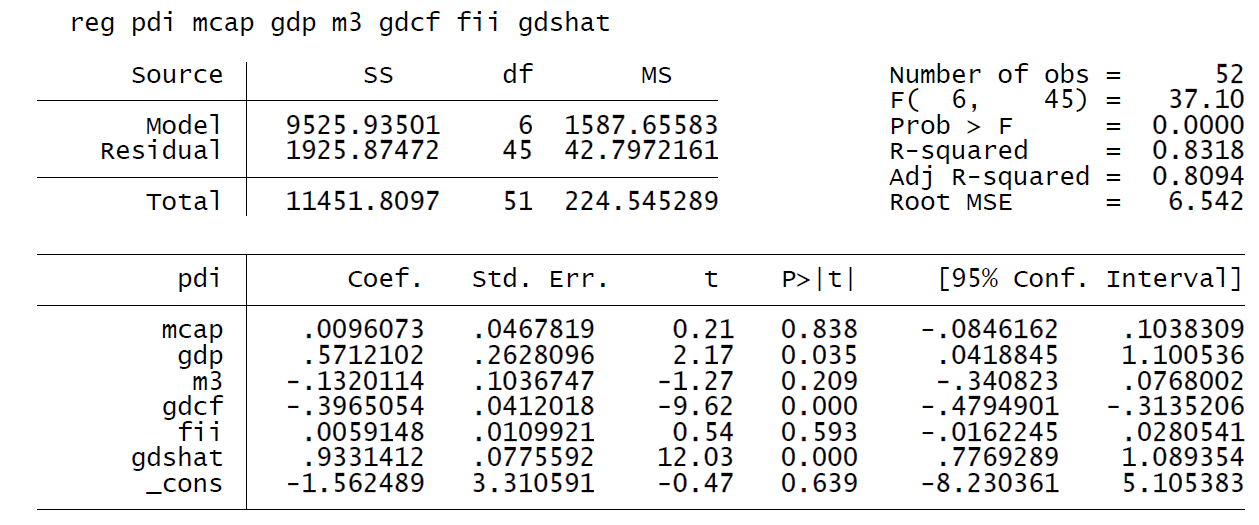
**Command – predict gdshat, xb**

Once running the above command, we should see a new column “gdshat” appear in our data as shown below. This column is nothing but predicted values of “gds”

****

Once we have predicted the “gdshat” values, we can now use “gdshat” in place of “gds” and see if we are overestimating the values of “gds”. Run the following command in STATA

**Command – reg pdi mcap gdp m3 gdcf fii gdshat**

****

Compare gdshat from the output table with gds from the first table. If coefficient of gdshat < gds. Say it gds coeff is 2 and gdshat coeff is 1.5 that means we are overestimating the value by 50% because of endogeneity.