International Journal of Applied Research 2019; 5(6): 335-342

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2019; 5(6): 335-342
www.allresearchjournal.com
Received: 14-03-2019
Accepted: 18-04-2019

Ashis Kr. Mukherjee
Department of Economics,
Nistarini College, Purulia,
West Bengal, India

Moumi Laha
Department of Economics,
J.K.College, Purulia,
West Bengal, India

Correspondence
Ashis Kr. Mukherjee
Department of Economics,
Nistarini College, Purulia,
West Bengal, India

Estimation and interpretation of different types of regression model

Ashis Kr. Mukherjee and Moumi Laha

Abstract
The functional relationship between variables defines the dependence of dependent variable upon the independent variables in a specific form. The functional relation may be linear or log linear or semi log or lin-log or reciprocal type. Different models have different implications such as when we want to calculate growth rate we generally use semi log model, again to calculate elasticity of some economic variables we use log linear model etc.
In this paper main focus given on the estimation procedure and interpretation of the estimated equation. In this article we use Straight Line Model, Log-Linear Model, Semi-Log Model, Lin-Log Model. At first we explain the important features of a good regression model, and then we explain the estimation procedure and interpretation of different types of model.

Keywords: Regression, Straight Line, Log-Linear, Semi-Log, Lin-Log

Introduction

Features of a Best Regression Model

There are some important features of a best regression model. If these features are hold then we can use the data for forecasting purpose. The features are.

High R^2 Value

The regression model has a high R^2 value. That is the value of the coefficient of determination is large. In statistics, R^2 is the degree of association between variables. So a high value of R^2 means the regression line fits the data well. It also use as a guideline to measure the accuracy of the model.

No Serial Correlation

Classical linear regression model assumes that the successive values of disturbance term are temporarily independent. In other words, when observations are taken over time the effect of disturbance term occurring at one period does not carry over into another period. But if this assumption violated then the problem known as serial correlation. In this case the value of disturbance term in any particular period is correlated with its own preceding values, and parameters estimates are biased. The residuals in the regression model have no serial correlation. Serial correlation is defined as the correlation of a variable with itself over successive observations.

No Heteroscedasticity

In the classical linear model we assume that the variance of the disturbance term is constant. If this assumption is violated then a problem arises known as heteroscedasticity. In the presence of heteroscedasticity the OLS estimators are not best. A best regression model have no heteroscedasticity problem in the residuals terms. In statistics heteroscedasticity happens when the standard errors of a variable monitored over a specific amount of time are non-constant.

Residuals are Normally Distributed

To test that the residuals are normally distributed we use Jarque-Bera test. The test statistic is
$JB=n[S_2/6 + (K-3)^2/24]$

JB=n[$S_2/6 + (K-3)^2/24$]
Where S=Skewness and K= Kurtosis
In case of normal distribution Skewness is equal to zero and kurtosis is equal to 3. So (K-3) represent excess kurtosis. The statistic follows chi-square distribution with degree of freedom 2. If the p-value is sufficiently low then we can reject the hypothesis that the residuals are normally distributed. The residuals in the regression model are normally distributed.
When all the above features are fulfilled then we can forecast the model correctly.
To explain the above four features we take an example. Suppose the values of the three variables are shown in the following table.

Table 1: Values of the variables Y_i, (Dependent) X_{2i} and X_{3i} (independent)

Y_i	X_{2i}	X_{3i}
5.92	4.9	4.78
4.3	5.9	3.84
3.3	5.6	3.13
6.23	4.9	3.44
10.97	5.6	6.84
9.14	8.5	9.47
5.77	7.7	6.51
6.45	7.1	5.92
7.6	6.1	6.08
11.47	5.8	8.09
13.46	7.1	10.01
10.24	7.6	10.81
5.99	9.7	8

Considering Y_i as dependent variable and X_{2i}, X_{3i} as explanatory variables, at first we regress Y on X_{2i} and X_{3i} using OLS technique. The result of the regression analysis is:

$Y = 7.1933 - 1.3924X_{2i} + 1.4700X_{3i}$.......................... (1)

S.E (1.594) (0.3050) (0.1758)
t-Value (4.5105) (-4.5652) (8.3626)
p-Value (.0011) (0.0010) (0.000)

R^2= 0.8765
F-Value: 35.515 (0.000)

We also calculate the residuals. The residuals output shown in the following table

Table 2: Predicted values of y_i and residuals

Observation	Predicted Yi	Residuals
1	7.396997579	-1.477
2	4.622695292	-0.3227
3	3.99671413	-0.69671
4	5.427154558	0.802845
5	9.450533238	1.519467
6	9.278548559	-0.13855
7	6.041231216	-0.27123
8	6.009395539	0.440604
9	7.637072784	-0.03707
10	11.00957895	0.460421
11	12.02182685	1.438173
12	12.50161648	-2.26162
13	5.446634834	0.543365

With the help of the above example we can explain all the above features one by one.

High R^2 Value

In statistical analysis the coefficient of determination or the square of the correlation coefficient assesses how well a model explains and predicts future outcomes. It also use as a guideline to measure the accuracy of the model. If the value of R^2 is 0.50, then we say that approximately half (50%) of the observed variation can be explained by the model. Since the value of the value of R^2 lies between -1 and +1 the closer the value is to 1 the better the fit or relationship between two variables.
From the above regression result we can say that, the coefficient of determination, R^2= 0.8765 means that 87.65 percent of the variation in the Y_i is explained by the two variables X_{2i} and $X_{3i,}$ therefore our regression line fits the given data well. We see that F statistic is also significant because the p-value (shown in the bracket) is 0.000 which is less than 5%. In other words, this means that the explanatory variables X_{2i} and X_{3i} jointly can influence the dependent variable Y_i.

No Serial Correlation

To test that residuals in the regression model have no serial correlation we set two hypotheses
The null hypothesis is, H_0: (residuals are not serially correlated)
The alternative hypothesis is, H_1: (residuals are serially correlated).
To test that whether the residuals are serially correlated or not we use Breusch-Godfrey serial correlation LM test.

Table 3: Results of Breusch-Godfrey Serial Correlation LM Test

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	0.682521	Prob. F(2,8)		0.5325
Obs*R-squared	1.894871	Prob. Chi-Square(2)		0.3877
Test Equation: Dependent Variable: RESID Method: Least Squares Date: 04/27/19Time: 14:41 Sample: 1 13 Included observations: 13 Presample missing value lagged residuals set to zero.				
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
X3I	0.231121	0.275075	0.840209	0.4252
X2I	-0.084210	0.434665	-0.193735	0.8512
C	-0.944937	2.345057	-0.402948	0.6975
RESID(-1)	-0.550770	0.504996	-1.090642	0.3072

RESID(-2)	-0.477210	0.659536	-0.723554	0.4900
R-squared	0.145759	Mean dependent var		-6.15E-16
Adjusted R-squared	-0.281361	S.D. dependent var		1.068611
S.E. of regression	1.209638	Akaike info criterion		3.502242
Sum squared resid	11.70579	Schwarz criterion		3.719531
Log likelihood	-17.76458	Hannan-Quinn criter.		3.457580
F-statistic	0.341261	Durbin-Watson stat		1.614148
Prob(F-statistic)	0.842970			

From the above table we see that the value of observed R^2 is 1.894871 and corresponding probability is 0.3877. Since the value is greater than 5%, we cannot reject the null hypothesis, and conclude that the residuals are free from serially correlation.

No Heteroscedasticity

To test there is no heteroscedasticity problem we use Breuch-pagan-godfrey test. To test that residuals in the regression model have no heteroscedastic problem, we set two hypotheses

The null hypothesis is, H_0: (residuals are not heteroscedastic)

The alternative hypothesis is, H_1: (residuals are heteroscedastic).

Table 4: Results of Breusch-Pagan-Godfrey Test

Heteroskedasticity Test: Breusch-Pagan-Godfrey				
F-statistic	3.484218	Prob. F(2,10)		0.0711
Obs*R-squared	5.338717	Prob. Chi-Square(2)		0.0693
Scaled explained SS	2.944052	Prob. Chi-Square(2)		0.2295
Test Equation: Dependent Variable: RESID^2 Method: Least Squares Date: 04/27/19Time: 14:44 Sample: 1 13 Included observations: 13				
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
C	1.760610	1.716065	1.025958	0.3291
X3I	0.497796	0.189153	2.631705	0.0251
X2I	-0.606395	0.328213	-1.847567	0.0944
R-squared	0.410671	Mean dependent var		1.054089
Adjusted R-squared	0.292805	S.D. dependent var		1.497859
S.E. of regression	1.259623	Akaike info criterion		3.498677
Sum squared resid	15.86651	Schwarz criterion		3.629050
Log likelihood	-19.74140	Hannan-Quinn criter.		3.471879
F-statistic	3.484218	Durbin-Watson stat		1.513457
Prob(F-statistic)	0.071087			

From the table we see that the value of observed R^2 is 5.338717 and corresponding probability is 0.0693. Since the value is greater than 5%, we cannot reject the null hypothesis, and conclude that the residuals are free from serially correlation.

Residuals are Normally Distributed

Now to test whether the residuals are normally distributed or not we use Jarque-Bera test. To test this at first we set two hypothesis

The null hypothesis is, H_0: (residuals are normally distributed)

The alternative hypothesis is, H_1: (residuals are not normally distributed).

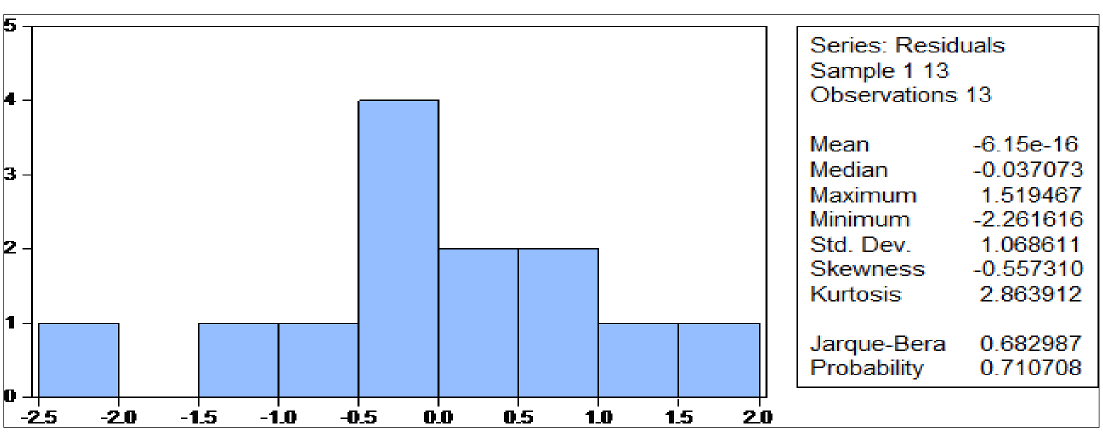

Fig 1: Results of Jarque Bera Test

From the table we see that the value of the Jarque-Bera statistic is 0.6830 and corresponding probability is 0.7107. Since the value is greater than 5% we cannot reject the null hypothesis and conclude that the residuals are normally distributed.
So from the above analysis we say that above data is best. Analyse the data we see that the value of R^2 is very high, there is no serial correlation, there is no heteroskedasticity problem and also the residuals are normally distributed. So we can use the data for forecasting purpose.

Estimation and interpretation of Linear Equation
The linear equation can be written in the form,

$Y_i = a + bX_i$

Where, Y is the dependent variable.
X is the independent or explanatory variable.
'a' is intercept term, 'b' is the slope.
Now adding disturbance term we can write,
$Y= a + bX + U$.
We want to explain how the equation of such type can be estimated and interpreted. To explain this we take the above example.
From the above regression result we can say that,

- The coefficient of determination, $R^2= 0.8765$ means that 87.65 percent of the variation in the Y is explained by the two variables X_{2i} and X_{3i}, therefore our regression line fits the given data well.
- Holding X_{3i} input constant, one unit increases in X_{2i} will led to, on the average, a 1.3924 unit decreases in the output. The coefficient -1.3924 (known as partial regression coefficient). Since the sign of the coefficient is negative we say that there is an inverse relationship between Y_i and X_{2i}.
- Holding X_{2i} input constant, one unit increases in X_{3i} will led to, on the average, a 1.47 percent increase in the output. The coefficient 1.27 (known as partial regression coefficient). Since the sign of the coefficient is positive we say that there is a positive relationship between Y_i and X_{3i}.

To test the intercept term (i.e.β_1)
The Null Hypothesis is H_0: ($\beta_1 = 0$)
Against the Alternative Hypothesis H_1: ($\beta_1 \neq 0$)
The test statistic is

$t = \beta_1/S.E(\beta_1)$
$= (7.1933)/ (1.594) = 4.5105$
Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.
The confidence interval of β_1 is

$\beta_1 \pm t_{0.025,\, n-2}\, SE(\beta_1)$
$= 7.1933 \pm 3.106(1.594)$
=12.144 (taking +ve sign) and, 2.2423(taking negative sign).
This means that value of β_1 vary from 2.2423 to 12.144 units.

To test the coefficient of X_{2i}(i.e.β_2)
The Null Hypothesis is H_0: ($\beta_2 = 0$)
Against the alternative hypothesis H_1: ($\beta_2 \neq 0$)
The test statistic is

$t = \beta_2/S.E(\beta_2)$
$= (-1.3924)/ (0.3050) = -4.5652$
Or, $|t| = 4.5652$
Degree of freedom =13-2 = 11
Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.
The confidence interval of β_2 is

$\beta_2 \pm t_{0.025,n-2}\, SE(\beta_2)$
$= -1.3924 \pm 3.106(0.3050)$
= -0.4451 (taking +ve sign) and, -2.34 (taking negative sign).

This means that value of β_2 vary from -0.4451 to -2.34 units.

To test the coefficient of X_{3i} (i.e. β_3)
The Null Hypothesis is H_0: ($\beta_3 = 0$)
Against the alternative hypothesis H_1: ($\beta_3 \neq 0$)
The test statistic is
$t = \beta_3/S.E(\beta_3)$
$= (1.4700)/ (0.1758) = 8.3626$
Degree of freedom =13-2 = 11
Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_3 is not equal to zero.
The confidence interval of β_3 is
$\beta_3 \pm t_{0.025,n-2}\, SE(\beta_3)$
$= 1.4700 \pm 3.106(0.1758)$
=2.02 (taking +ve sign)
and, 0.924 (taking negative sign).
This means that value of β_3 vary from 2.02 to 0.924 units.

Test for Stability: (CUSUM Test)
Now to check that the model is stable or not we use CUSUM (Cumulative Sum Control Charts) test. This chart is used to monitoring the mean of a process based on sample taken form a given time. The measurement of the sample at a given time constitute a subgroup. rather than examining the mean of each sub groups independently, the CUSUM, chrts shows that accumulation of information of current and previous sample. The cusum chart typically signals an out of control process by an upward or downward drift of the cumulative sum until it crossed the boundary. If it crossed the boundary we suspect that the process of out of control.

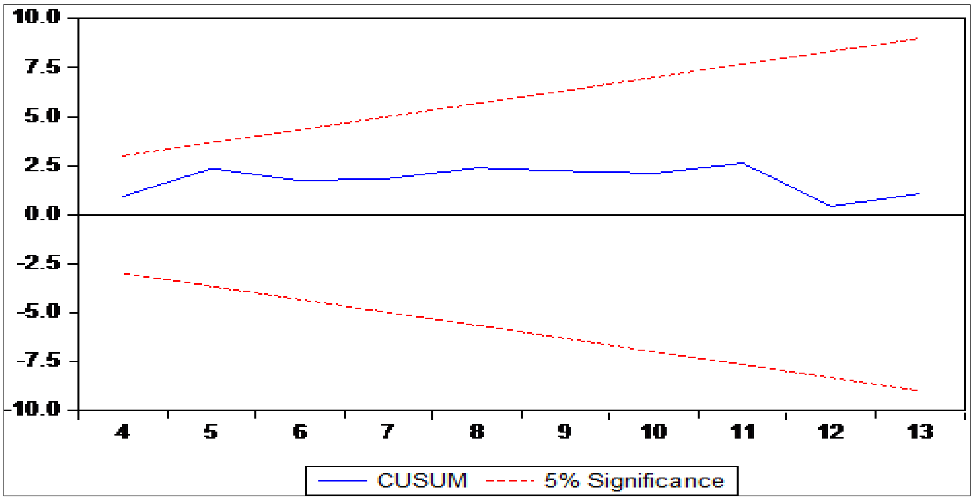

Fig 2: Results of CUSUM Test

In the graph the two red lines indicates the boundary. since the blue line is located between the two red lines meaning that our model is stable. In other words the dependent variable Yi has stability.

Estimation and interpretation of Log Linear Model

Let us consider an exponential regression model as

$Y_i = \beta_1 X_i^{\beta_2} e^{U_i}$

Where Y_i = dependent variable,

X_i = independent variable,

U_i=random error term,

β_1 and β_2 are regression parameters.

Taking Logarithm both side,

$\text{Log } Y_i = \text{Log } \beta_1 + \beta_2 \text{Log} X_i + U_i$

let $\text{Log } \beta_1 = \alpha$ then we can write

$\text{Log } Y_i = \alpha + \beta_2 \text{Log} X_i + U_i$

This model is linear in the parameter α and β and we can estimate the values of the parameters of this equation by using OLS technique. Such models are known as Log-Log or double Log or Log linear model. We can use the model when we measure elasticity. To explain the model let us take an example.

Suppose the values of the three variables are shown in the following table. We have to analyse the data using OLS technique. At first we calculate log Y, $\log X_{2i}$ and $\log X_{3i}$ which is shown in the following table:

Table 5: Logarithmic values of the variables

Y	X_{2i}	X_{3i}	Log Y	Log X_{2i}	Log X_{3i}
5.92	4.9	4.78	0.772322	0.690196	0.679428
4.3	5.9	3.84	0.633468	0.770852	0.584331
3.3	5.6	3.13	0.518514	0.748188	0.495544
6.23	4.9	3.44	0.794488	0.690196	0.536558
10.97	5.6	6.84	1.040207	0.748188	0.835056
9.14	8.5	9.47	0.960946	0.929419	0.97635
5.77	7.7	6.51	0.761176	0.886491	0.813581
6.45	7.1	5.92	0.80956	0.851258	0.772322
7.6	6.1	6.08	0.880814	0.78533	0.783904
11.47	5.8	8.09	1.059563	0.763428	0.907949
13.46	7.1	10.01	1.129045	0.851258	1.000434
10.24	7.6	10.81	1.0103	0.880814	1.033826
5.99	9.7	8	0.777427	0.986772	0.90309

Then we regress logY on $\log X_{2i}$ and $\log X_{3i}$ and get the following results,

$\log Y_i = 0.8979 - 1.291 \log X_{2i} + 1.273 \log X_{3i}$

S.E	(0.156)	(0.2524)	(0.1308)
t-Value	(5.7471)	(-5.117)	(9.735)
p-Value	(0.000)	(0.000)	(0.000)

R^2= 0.9080

From the above regression result we can say that

- The coefficient of determination, R^2= 0.9080 means that 90.80 percent of the variation in the (log of) Yis explained by the two variables (log of) X_{2i} and (log of) X_{3i}, therefore our regression line fits the given data well.
- Holding X_{3i} input constant, one percent increases in X_{2i} will led to, on the average, a 1.29 percent decreases in the output. The coefficient -1.29 (known as partial regression coefficient) is output elasticity of Y with respect to X_{2i}. Since the sign of the coefficient is negative we say that there is an inverse relationship between $\log Y_i$ and $\text{Log} X_{2i}$.
- Holding X_{2i} input constant, one percent increases in X_{3i} will led to, on the average, a 1.27 percent increase in the output. The coefficient 1.27 (known as partial regression coefficient) is known as output elasticity of Y with respect to X_{3i}. Since the sign of the coefficient is positive we say that there is a positive relationship between $\log Y_i$ and $\text{Log} X_{3i}$.
- The sum of the elasticities = -1.29+1.27 = -0.02 which gives us the value of the returns to scale. Since the sum of the elasticities is less than one, this means that the function is DRS (decreasing returns to scale) type.

Now we want to test the validity of the intercept term and the coefficients β_1, β_2 of the above regression result:

To test the intercept term (i.e.β_1)

The Null Hypothesis is H_0: ($\beta_1 = 0$)

Against the Alternative Hypothesis H_1: ($\beta_1 \neq 0$)

The test statistic is

$t = \beta_1 / S.E(\beta_1)$

$= (0.8979)/(.1562) = 5.7471$

Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106

Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{2i} (i.e.β_2)
The Null Hypothesis is H_0: ($\beta_2 = 0$)
Against the alternative hypothesis H_1: ($\beta_2 \neq 0$)
The test statistic is

t = β_2/S.E(β_2)
= (-1.291)/(0.2524) = -5.115
Or, |t| = 5.115
Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{3i} (i.e. β_3)
The Null Hypothesis is H_0: ($\beta_3 = 0$)
Against the alternative hypothesis H_1: ($\beta_3 \neq 0$)
The test statistic is

t = β_3/S.E(β_3)
= (1.273)/(0.1308) = 9.7324
Degree of freedom =13-2 = 11
Tabulated value at 1% level of significance = 3.106

Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_3 is not equal to zero.

Estimation and interpretation of Semi Log Model
To calculate the rate of growth of certain economic variables such as GNP, GDP, Money Supply, BOP deficit etc. we use semi-log model.
Consider the following model:

$Y_t = Y_0(1+r)^t$

Where r = Compound rate of growth of Y.
Y_0= Initial value of Y.t= time period.
Taking logarithm both side,

In Y_t=InY_0+t In(1+r)..............................(i)
Let β_1 = InY_0 and β_2=In (1+r)
Then from (i) we can write InY_t= $\beta_1 + \beta_2 t$

Adding the disturbance term we obtain,

InY_t= $\beta_1 + \beta_2 t + U_t$

This model is called semi-log model because only one variable appears in the logarithmic form. For descriptive purpose a model in which the regress sand is logarithmic will be called a log-lin model. The slope coefficient is the constant proportional or relative change in Y for a given absolute change in the value of the Regressor. That is,
β_2= (relative change in regress and)/ (absolute change in Regressor)

To explain the model let us take an example.

- Here we consider the production of a commodity in different years. We want to calculate the growth rate of the production of the commodity. So at first we calculate the logarithmic value of Y.

Table 6: Year Wise Production of a Commodity

Year	Production(Y)	X	Log Y	Year	Production(Y)	X	LogY
1992	1365	-19	3.135133	2002	746	1	2.872739
1993	2795	-17	3.446382	2003	1129	3	3.052694
1994	2525	-15	3.402261	2004	1256	5	3.09899
1995	958	-13	2.981366	2005	842	7	2.925312
1996	1480	-11	3.170262	2006	365	9	2.562293
1997	1998	-9	3.300595	2007	110	11	2.041393
1998	1850	-7	3.267172	2008	95	13	1.977724
1999	955	-5	2.980003	2009	40	15	1.60206
2000	1121	-3	3.049606	2010	120	17	2.079181
2001	1228	-1	3.089198	2011	70	19	1.845098

Regressing log Y on X we get the following result
Log Y= 2.7939 - 0.0409t

S.E (0.0667) (0.00578)
t-Value (41.88) (-7.0755)
p-Value (0.000) (0.000)R^2 = 0.7355

The slope coefficient gives us the growth rate of the commodity. From theresult we say that the production of the commodity decline at the rate of 4.09 percent.
We find that,

InY_0 = 2.7939
Or,Y_0 = antilog(2.7939) = 622.16

That is, at the beginning of the year estimated production was about 622.16 units. The slope coefficient of 0.0409 of the growth model gives the instantaneous rate of growth.
We can write,

In (1+r) = -0.0409
Or, (1+r) = antilog(0.0409)
Or, r = 1.0987-1 = 0.0987

That is over the study period the compound rate of growth was about 1.656 percent per year.

- We consider another example. Suppose the values of the three variables are shown in the following table. We have to analyse the data using OLS technique. At first we calculate log Y, which is shown in the following table.

Table 7: Logarithmic values of the dependent variable Y_i

Y	X_{2i}	X_{3i}	Ln Y
5.92	4.9	4.78	0.772322
4.3	5.9	3.84	0.633468
3.3	5.6	3.13	0.518514
6.23	4.9	3.44	0.794488
10.97	5.6	6.84	1.040207
9.14	8.5	9.47	0.960946
5.77	7.7	6.51	0.761176
6.45	7.1	5.92	0.80956
7.6	6.1	6.08	0.880814
11.47	5.8	8.09	1.059563
13.46	7.1	10.01	1.129045
10.24	7.6	10.81	1.0103
5.99	9.7	8	0.777427

Then we regress log Y on X_{2i} and $X_{3i}$and get the following estimated equation as

Log Y_i= 0.7859 -0.0736X_{2i} + 0.0839X_{3i}
S.E (0.1001) (0.0193) (0.0111)
t-Value (7.8044) (-3.8195) (7.560)
p-Value (0.000) (0.0033) (0.000)R^2= 0.8550

From the above regression result we can say that

- Coefficient of determination or R^2= 0.8550 means that about 85.50 percent of the variation in the (log of) Y is explained by the the two variables X_{2i} and X_{3i}, and can therefore our regression line fits the given data well.
- Holding X_{3i} input constant, one unit increases in X_{2i} led on the average to about a 0.0736 percent decreases in the (log of) output. Since the sign of the coefficient is negative we say that there is an inverse relationship between X_{2i} and logY_i.
- Holding X_{2i} input constant, one percent increases in X_{3i} led on the average to about a 0.0839 percent increase in the output. Since the sign of the coefficient is positive we say that there is a positive relationship between X_{3i} and logY_i

Now we want to test the validity of the intercept term and the coefficients β_1, β_2 of the above regression result:

To test the intercept term (i.e.β_1)

The Null Hypothesis is H_0: ($\beta_1 = 0$)
Against the alternative hypothesis H_1: ($\beta_1 \neq 0$)
The test statistic is

t = β_1/S.E(β_1)
= (0.7859)/(0.1001) = 7.8511
Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{2i}(i.e.β_2)

The Null Hypothesis is H_0: ($\beta_2 = 0$)
Against the alternative hypothesis H_1: ($\beta_2 \neq 0$)
The test statistic is

t = β_2/S.E(β_2)
= (-0.0736)/(0.01926) = -3.8213
Or, |t| = 3.8213
Degree of freedom =13-2 = 11
Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{3i}(i.e. β_3)

The Null Hypothesis is H_0 : ($\beta_3 = 0$)
Against the alternative hypothesis H_1: ($\beta_3 \neq 0$)
The test statistic is

t = β_3/S.E(β_3)
= (0.0839)/(0.0111) = 7.5585
Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_3 is not equal to zero.

Estimation and interpretation of Lin-Log Model

Now we are interested in finding the absolute change in the dependent variable for a percent change in independent variable. In this case we use lin-log model. The model can be written in the form
$Y_t = \beta_1 + \beta_2 InX_t + U_{ta}$
This model is called lin-log model. For descriptive purpose a model in which the regressor is logarithmic will be called a lin-log model. The slope coefficient can be defined as
β_2= (absolute change in regress and)/ (relative change in regressor)
To explain the model let us take an example.

- Suppose the values of the three variables are shown in the following table. We have to analyse the data using OLS technique. At first we calculate logX_{2i} and logX_{3i} which is shown in the following table:

Table 8: Logarithmic values of the independent variables

Y	X_{2i}	X_{3i}	LnX_{2i}	LnX_{3i}
5.92	4.9	4.78	0.690196	0.679428
4.3	5.9	3.84	0.770852	0.584331
3.3	5.6	3.13	0.748188	0.495544
6.23	4.9	3.44	0.690196	0.536558
10.97	5.6	6.84	0.748188	0.835056
9.14	8.5	9.47	0.929419	0.97635
5.77	7.7	6.51	0.886491	0.813581
6.45	7.1	5.92	0.851258	0.772322
7.6	6.1	6.08	0.78533	0.783904
11.47	5.8	8.09	0.763428	0.907949
13.46	7.1	10.01	0.851258	1.000434
10.24	7.6	10.81	0.880814	1.033826
5.99	9.7	8	0.986772	0.90309

Then We regress Y on logX_{2i} and log$X_{3i}$and get the following estimated equation as

Y_i = 9.4576 - 23.1643LnX_{2i}+ 21.6059LnX_{3i}
S.E(3.128)(5.052)(2.618)
t-Value(3.0235)(-4.5842)(8.2526)
p-Value(0.000)(0.000)(0.000)R^2= 0.8749

From the above regression result we can say that

- Coefficient of determination or R^2= 0.8749 means that about 87.49 percent of the variation in theY is explained by the (log of) two variables X_{2i} and X_{3i}, and can therefore our regression line fits the given data well.

- Holding X_{3i} input constant, one percent increases in (log of) X_{2i} led on the average to about a 0.0736 unit decreases in the output. Since the sign of the coefficient is negative we say that there is an inverse relationship between LnX_{2i} and $Y_{i.}$
- Holding X_{2i} input constant, one percent increases in (log of) X_{3i} led on the average to about a 0.0839 unit increase in the output. Since the sign of the coefficient is positive we say that there is a positive relationship between X_{3i} and $\log Y_i$

Now we want to test the validity of the intercept term and the coefficients β_1, β_2 of the above regression result:

To test the intercept term (i.e.β_1)

The Null Hypothesis is H_0: ($\beta_1 = 0$)
Against the alternative hypothesis H_1: ($\beta_1 \neq 0$)
The test statistic is

$t = \beta_1/S.E(\beta_1)$
$= (9.4576)/(3.128) = 3.0235$
Degree of freedom =13-2 = 11

Tabulated value at 5% level of significance = 2.201
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{2i} (i.e.β_2)

The Null Hypothesis is H_0: ($\beta_2 = 0$)
Against the alternative hypothesis H_1: ($\beta_2 \neq 0$)
The test statistic is

$t = \beta_2/S.E(\beta_2)$
$= (-23.1643)/(5.05298) = -4.584$
Or, $|t| = 4.584$
Degree of freedom =13-2 = 11

Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_1 is not equal to zero.

To test the coefficient of X_{3i} (i.e. β_3)

The Null Hypothesis is H_0: ($\beta_3 = 0$)
Against the alternative hypothesis H_1: ($\beta_3 \neq 0$)
The test statistic is

$t = \beta_3/S.E(\beta_3)$
$= (21.6059)/(2.6180) = 8.2526$

Degree of freedom =13-2 = 11
Tabulated value at 1% level of significance = 3.106
Since, the estimated value is greater than the tabulated value we can reject the null hypothesis and conclude that the coefficient β_3 is not equal to zero.

Conclusion

- From the above analysis we say that a best regression model must have four features, a high R^2 value, residuals in the regression model have no serial correlation, no heteroscedasticity in the residuals terms, and residuals in the regression model are normally distributed.
- Linear equation shows the change in dependent variable when independent variable change by one unit. The slope coefficient is the marginal change
- Log linear model shows the rate of change of dependent variable when independent variable change by one percent. The slope coefficient is the output elasticity with respect to the independent variable.
- To calculate the rate of growth of certain economic variables such as GNP, GDP, Money Supply, BOP deficit etc. we use semi-log model.
- Lin log model shows the absolute change in the dependent variable for a percent change in independent variable.

References

1. Koutsoyiannis A. Theory of Econometrics, ELBS with Macmillan.
2. Madnani GMK. Introduction to Econometrics, Oxford & IBH Publishing co. Pvt. Ltd.
3. Damodar N. Gujarati, Basic Econometrics, McGraw Hill Education Private Limited.
4. Croxton and Cowden, Applied General Statistics Prentice Hall, Inc, 1964.
5. DasNG. Statistical Method, M. Das & Co.
6. Goon Gupta, Dasgupta. Fundamentals of Statistics, The World Press.