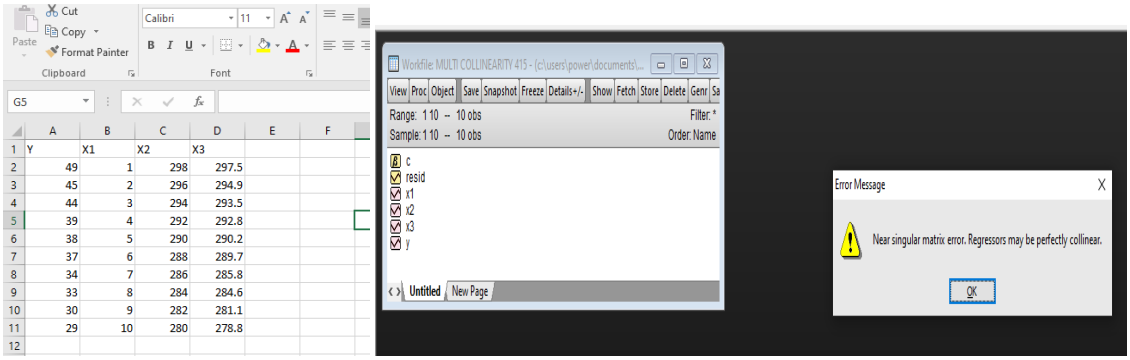
UNIT 1 MULTICOLLINEARITY

Meaning of Multicollinearity

Multicollinearity is a statistical concept where several independent variables in a model are correlated. Two variables are considered perfectly collinear if their correlation coefficient is +/- 1.0. Multicollinearity among independent variables will result in less reliable statistical inferences.

Consequences of Multicollinearity

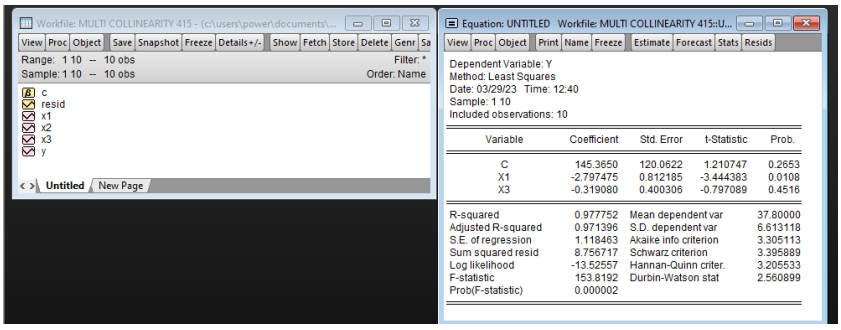
If we have the following data set , after importing it in EViews , we estimate regression equation (y_c_X1_X2_X3) in EViews by going to quick or selecting the variables , then the following will be the consequences if multicollinearity is present :



	A	B	C	D	E	F
1	Y	X1	X2	X3		
2	49	1	298	297.5		
3	45	2	296	294.9		
4	44	3	294	293.5		
5	39	4	292	292.8		
6	38	5	290	290.2		
7	37	6	288	289.7		
8	34	7	286	285.8		
9	33	8	284	284.6		
10	30	9	282	281.1		
11	29	10	280	278.8		
12						

The first consequence seen is that we can't estimate the model with perfect multicollinearity.

Now if we estimate y_c_X1_X3 , by dropping variable X2 , we get the following results:



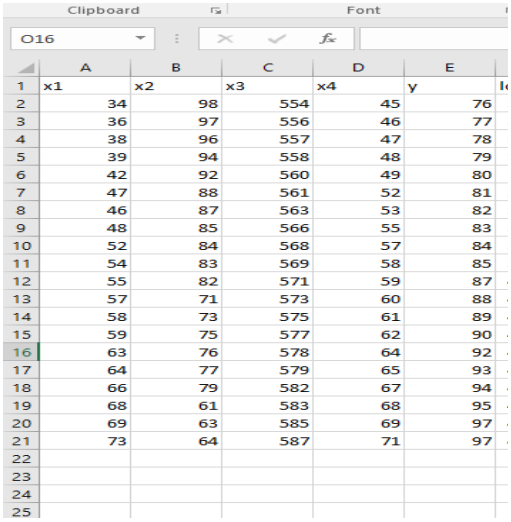
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	145.3650	120.0622	1.210747	0.2653
X1	-2.797475	0.812185	-3.444383	0.0108
X3	-0.319080	0.400306	-0.797089	0.4516

R-squared	0.977752	Mean dependent var	37.80000
Adjusted R-squared	0.971396	S.D. dependent var	6.613118
S.E. of regression	1.118463	Akaike info criterion	3.305113
Sum squared resid	8.756717	Schwarz criterion	3.395889
Log likelihood	-13.52557	Hannan-Quinn criter.	3.205533
F-statistic	153.8192	Durbin-Watson stat	2.560899
Prob(F-statistic)	0.000002		

Here because of multicollinearity , the value of standard errors becomes very large , due to this the value of t- statistic becomes small , leading to insignificant value of X3 as probability value is more than 5 percent.
NOTE- The OLS estimators and their standard errors become very sensitive to even small changes in the dataset due to presence of multicollinearity.

Detection of Multicollinearity

Multicollinearity can be detected by several ways. Some of them are listed below which will be applied on the following datasheet:


Clipboard Font

O16

	A	B	C	D	E
1	x1	x2	x3	x4	y
2	34	98	554	45	76
3	36	97	556	46	77
4	38	96	557	47	78
5	39	94	558	48	79
6	42	92	560	49	80
7	47	88	561	52	81
8	46	87	563	53	82
9	48	85	566	55	83
10	52	84	568	57	84
11	54	83	569	58	85
12	55	82	571	59	87
13	57	71	573	60	88
14	58	73	575	61	89
15	59	75	577	62	90
16	63	76	578	64	92
17	64	77	579	65	93
18	66	79	582	67	94
19	68	61	583	68	95
20	69	63	585	69	97
21	73	64	587	71	97
22					
23					
24					
25					

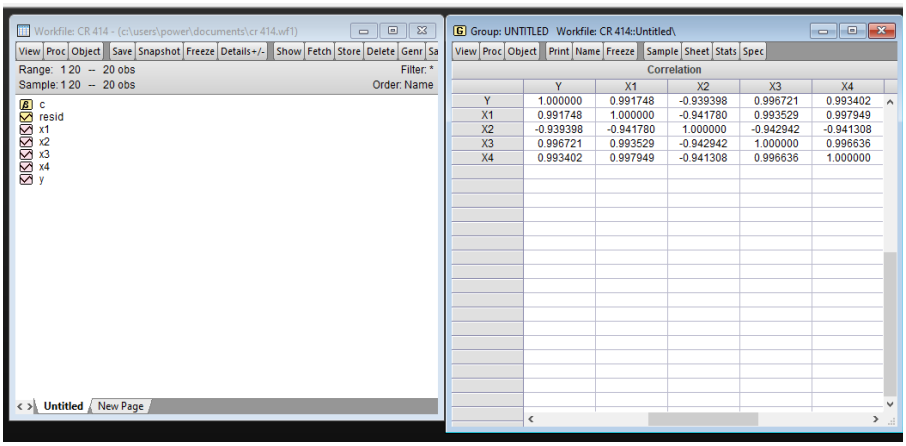
1.Pair wise Correlation

High pair wise correlation between independent variables represents presence of multicollinearity among them.

To calculate pair wise correlation coefficient in EViews the following steps should be taken:

a) Go to quick
b) Click on group statistics
c) Select the correlation option
d) A dialog box will appear on screen
e) Type y_X1_X2_X3_X4
f) Click on ok

The results in EViews will be shown as :



INTERPRETATION

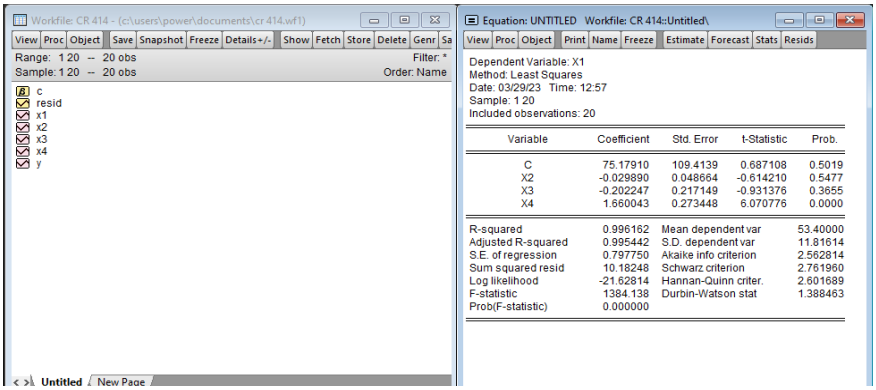
In the above correlation matrix , all the combinations of independent variables show a high or near to perfect multicollinearity as all the values are near to 1 . So , we can say that there is presence of multicollinearity among them.

2. Through Auxiliary Equations

Auxiliary equations are such equations in which one independent variable serves as the role of dependent variable and other independent variables remaining the same . Here number of auxiliary equations equals to number of independent variables. In the above data set there are 4 independent variables , so there will be 4 auxiliary equations .

Estimating the auxiliary equation with X1 as the independent variable:

Estimate the regression equation in EViews as X1_c_X2_X3_X4. The below are the results given by Eviews:



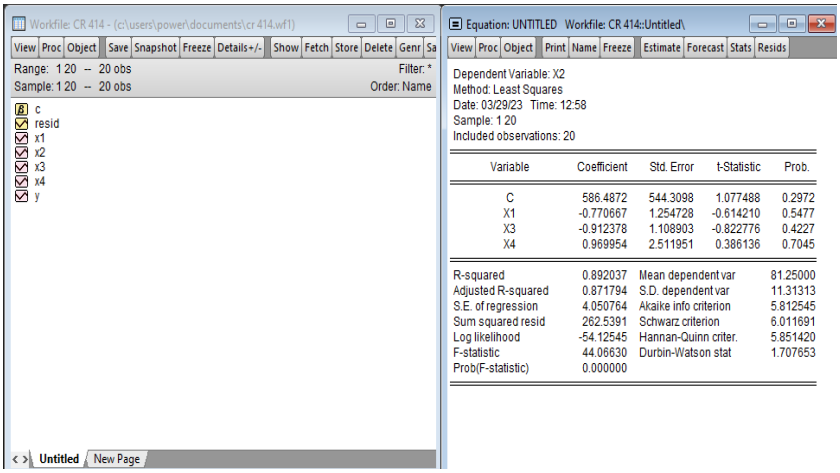
Since by above example the value of R^2_1 = 0.996, this means that 99.65% variation in X1 is explained by X2 and X3 and X4 and remaining is due to other factors which are unknown.

Formation of null hypothesis :
H_0 : b=c=d=0 , the impact of X2, X3 and X4 on X1 is not significant , $R^2_1 = 0$
H_1 : b≠c≠d≠0 , the impact of X2, X3 and X4 on X1 is significant , $R^2_1 \neq 0$

By the above results from EViews , the value of F- statistic = 1384.138 and the related probability value = 0.000 . Since p value is 0.000, which is less than 0.05 , so we reject our null hypothesis, that means we accept our alternate hypothesis . So , the value of R^2_1 is significant .

<u>Estimating the auxiliary equation with X2 as the independent variable:</u>

Estimate the regression equation in EViews as X2_c_X1_X3_X4. The below are the results given by Eviews:

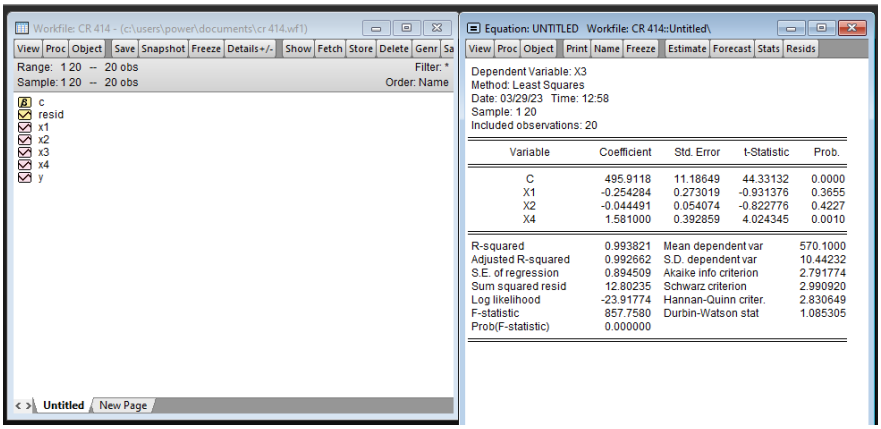


Since by above example the value of R^2_2 = 0.892 , this means that 89.20% variation in X2 is explained by X1 and X3 and X4 and remaining is due to other factors which are unknown.
Formation of null hypothesis :
H_0 : b=c=d=0 , the impact of X1, X3 and X4 on X2 is not significant , $R^2_2 = 0$
H_1 : b≠c≠d≠0 , the impact of X1, X3 and X4 on X2 is significant , $R^2_2 \neq 0$

By the above results from EViews , the value of F- statistic = 44.066 and the related probability value = 0.000 . Since p value is 0.000, which is less than 0.05 , so we reject our null hypothesis, that means we accept our alternate hypothesis . So , the value of R^2 is significant .

<u>Estimating the auxiliary equation with X3 as the independent variable:</u>

Estimate the regression equation in EViews as X3_c_X1_X2_X4. The below are the results given by EViews:



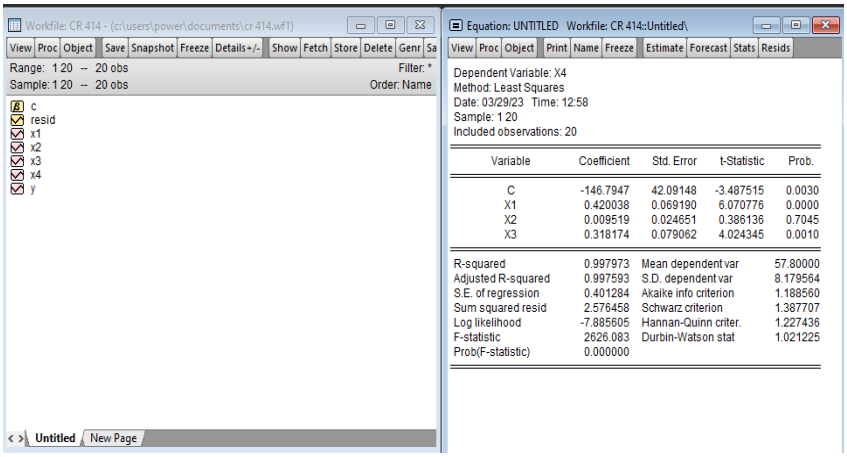
Since by above example the value of R^2_3 = 0.9938 , this means that 99.38 % variation in X3 is explained by X1 and X2 and X4 and remaining is due to other factors which are unknown.

Formation of null hypothesis :
H_0 : b=c=d=0 , the impact of X1, X2 and X4 on X3 is not significant , R^2_3 = 0
H_1 : b≠c≠d≠0 , the impact of X1, X2 and X4 on X3 is significant , R^2_3 ≠ 0

By the above results from EViews , the value of F- statistic = 857.76 and the related probability value = 0.000 . Since p value is 0.000, which is less than 0.05 , so we reject our null hypothesis, that means we accept our alternate hypothesis . So , the value of R^2_3 is significant .

Estimating the auxiliary equation with X4 as the independent variable:

Estimate the regression equation in EViews as X4_c_X1_X2_X3. The below are the results given by EViews:



Since by above example the value of R^2_4 = 0.9979 , this means that 99.79 % variation in X4 is explained by X1 and X2 and X3 and remaining is due to other factors which are unknown.

Formation of null hypothesis :
H_0 : b=c=d=0 , the impact of X1, X2 and X3 on X4 is not significant , R^2_4 = 0
H_1 : b≠c≠d≠0 , the impact of X1, X2 and X3 on X4 is significant , R^2_4 ≠ 0

By the above results from EViews , the value of F- statistic = 2626.083 and the related probability value = 0.000 . Since p value is 0.000, which is less than 0.05 , so we reject our null hypothesis, that means we accept our alternate hypothesis . So , the value of R^2_4 is significant .

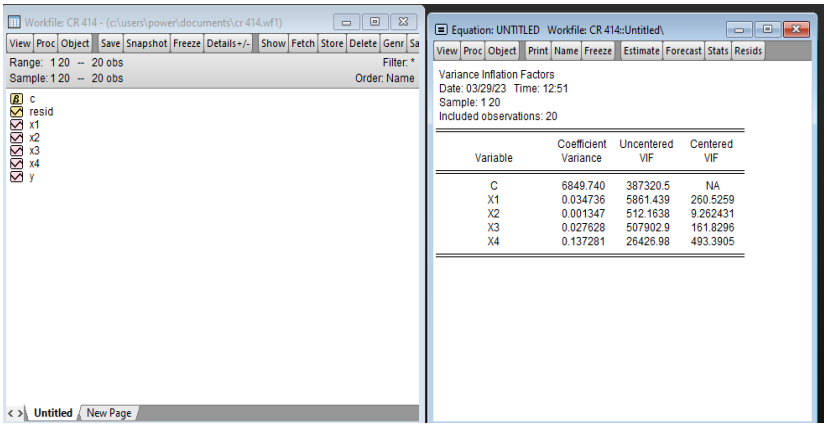
From the above results the value of R^2_2 is minimum . so in comparison with other cases this is not the X2 with multicollinearity and all other variables are suffering from multicollinearity.

3. Variance Inflation Factor Test

This is also a test used to detect multicollinearity in Eviews. The steps to test are as follows:

a) Estimate the equation as Y_c_X1_X2_X3_X4
b) Go to view from estimated results
c) Click on coefficient diagnostics
d) Select Variance Inflation Factor \
e) Click on OK

The results for the same in Eviews are:



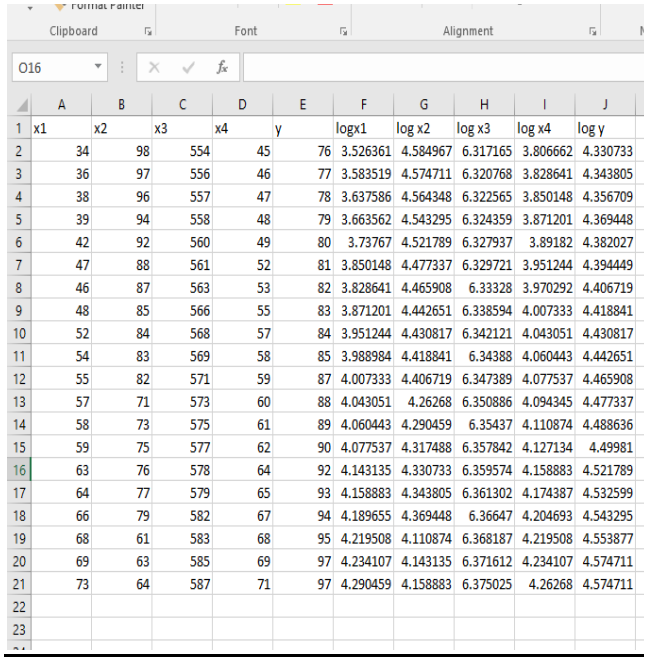
RULE OF THUMB- If the value of VIF (centred) is equal to or greater than 10, it will show the presence of multicollinearity in that variable.

INTERPRETATION

Since the values of centred VIF's of variables X1, X3 and X4 are more than 10 , there is presence of multicollinearity among them. X2 has value less than 10 , so as interpreted by auxiliary equations also , it doesn't have multicollinearity .

Remedial Measures

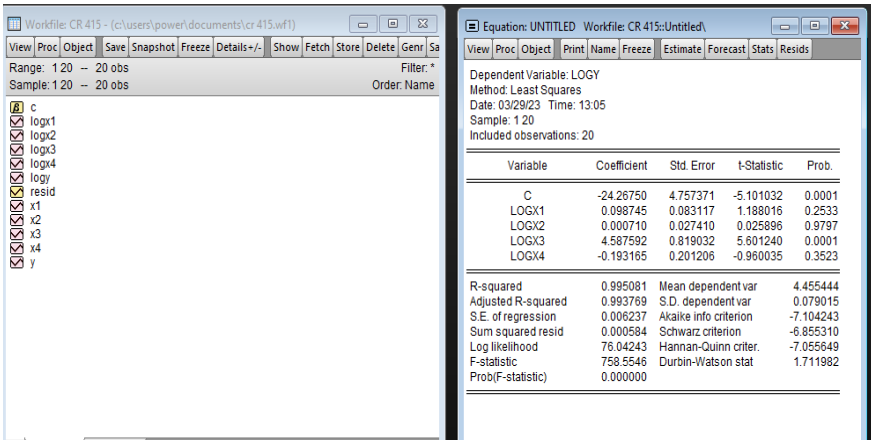
Multicollinearity is the phenomenon which makes our OLS estimators defective inspite of being BLUE.
So, this problem must be corrected . The below is a method used to correct the problem of multicollinearity . The dataset to be worked upon is:



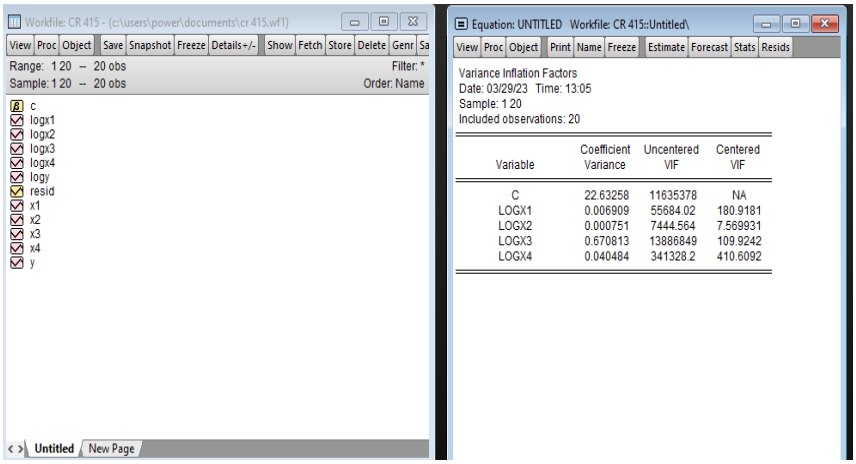
	x1	x2	x3	x4	y	logx1	log x2	log x3	log x4	log y
2	34	98	554	45	76	3.526361	4.584967	6.317165	3.806662	4.330733
3	36	97	556	46	77	3.583519	4.574711	6.320768	3.828641	4.343805
4	38	96	557	47	78	3.637586	4.564348	6.322565	3.850148	4.356709
5	39	94	558	48	79	3.663562	4.543295	6.324359	3.871201	4.369448
6	42	92	560	49	80	3.73767	4.521789	6.327937	3.89182	4.382027
7	47	88	561	52	81	3.850148	4.477337	6.329721	3.951244	4.394449
8	46	87	563	53	82	3.828641	4.465908	6.33328	3.970292	4.406719
9	48	85	566	55	83	3.871201	4.442651	6.338594	4.007333	4.418841
10	52	84	568	57	84	3.951244	4.430817	6.342121	4.043051	4.430817
11	54	83	569	58	85	3.988984	4.418841	6.34388	4.060443	4.442651
12	55	82	571	59	87	4.007333	4.406719	6.347389	4.077537	4.465908
13	57	71	573	60	88	4.043051	4.26268	6.350886	4.094345	4.477337
14	58	73	575	61	89	4.060443	4.290459	6.35437	4.110874	4.488636
15	59	75	577	62	90	4.077537	4.317488	6.357842	4.127134	4.49981
16	63	76	578	64	92	4.143135	4.330733	6.359574	4.158883	4.521789
17	64	77	579	65	93	4.158883	4.343805	6.361302	4.174387	4.532599
18	66	79	582	67	94	4.189655	4.369448	6.36647	4.204693	4.543295
19	68	61	583	68	95	4.219508	4.110874	6.368187	4.219508	4.553877
20	69	63	585	69	97	4.234107	4.143135	6.371612	4.234107	4.574711
21	73	64	587	71	97	4.290459	4.158883	6.375025	4.26268	4.574711

Rethink the Model

Rethinking the model means changing the functional form . The way dependent variable depends on independent variable is explained by the functional form . There are many types of functional form like log linear, semi log, cubic, quadratic , etc. The most commonly used form for correction of multicollinearity is log linear model. So , for the above data series if we take log of all variables and estimate the equation in EViews the following results will be shown:



From the above results , not so difference is seen even after changing the functional form of the model . Now, to test multicollinearity , we apply VIF test on modified model. The results in EViews for same are:



INTERPRETATION

Even after changing the functional form , results are not much changed for VIF test also . the variables X1 , X3 and X4 still possess multicollinearity as the value of centred VIF is more than 10 and the variable X2 doesn't have multicollinearity in it as it has its corresponding value less than 10.

So , this method is not useful in above taken dataset to correct multicollinearity.

Unit 2 Heteroscedasticity

Meaning Of Heteroscedasticity

Heteroscedasticity refers to situations where the variance of the residuals is unequal over a range of measured values. When running a regression analysis, heteroscedasticity results in an unequal scatter of the residuals (also known as the error term).

Detection of Heteroscedasticity

The detection of heteroscedasticity can be done by various tests in EViews . All the tests give almost same results .So, let's check them for the given data one by one:

Clipboard

K7

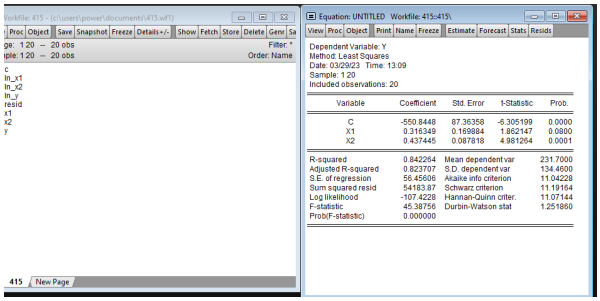
	A	B	C
1	y	x1	x2
2	121	546	1012
3	125	560	1013
4	120	612	1010
5	130	640	990
6	140	650	1120
7	141	700	1215
8	142	710	1180
9	150	780	1075
10	170	800	1200
11	180	780	1180
12	200	750	1080
13	201	740	1300
14	210	736	1310
15	250	740	1380
16	300	736	1375
17	310	730	1450
18	314	710	1479
19	320	680	1500
20	450	977	1678
21	660	999	1690
22			
23			

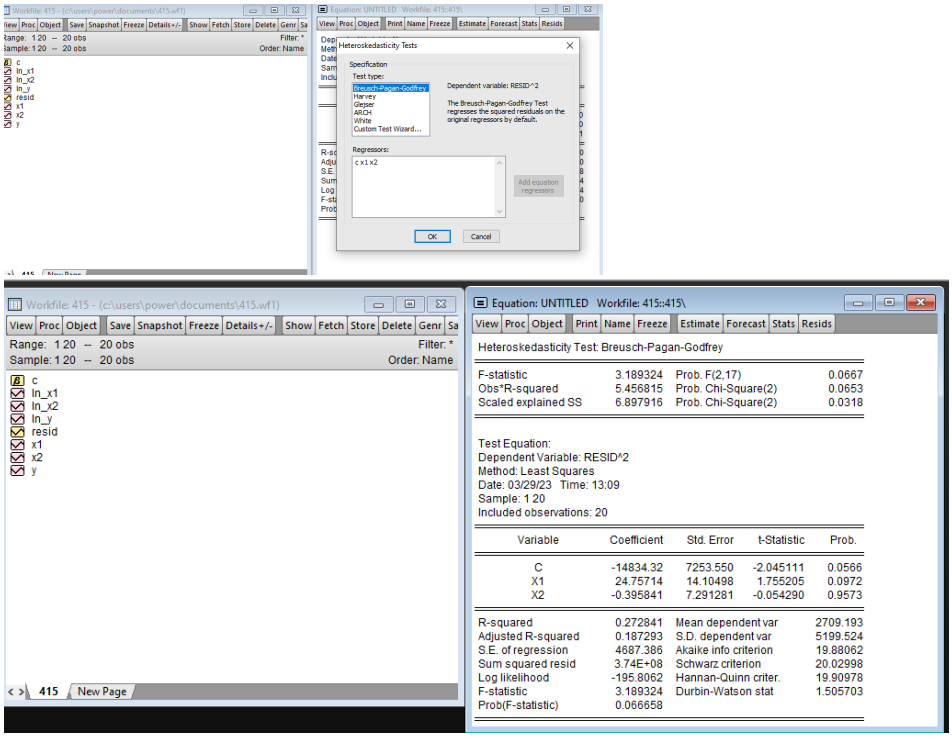
1.B-P-G Test

This is the most commonly used test to detect heteroscedasticity in the model. In EViews , the following are the steps to test this :

a) Estimate the regression model (Y_C_X1_X2)
b) Go to view from estimated results
c) Select residual diagnostics
d) Click on Heteroscedasticity tests
e) Select from test type – B.P.G test
f) Click ok

The results shown by EViews will be:





To check the presence of heteroscedasticity , we test the hypothesis of chi –square distribution (observations*R-squared).

INTERPRETATION

Hypothesis framing:
H_0: no heteroscedasticity or homoscedasticity present
H_1: presence of heteroscedasticity

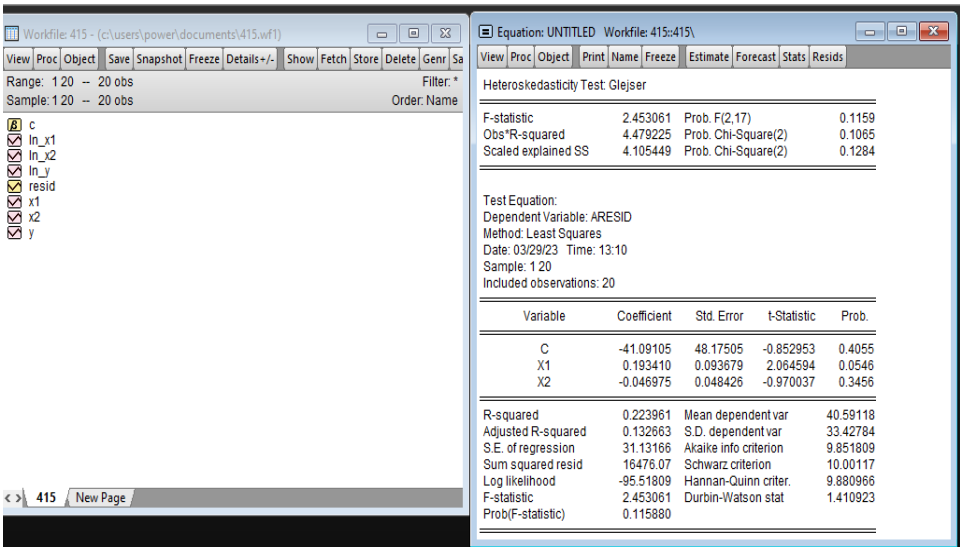
By the above results from EViews , the value of Obs*R-squared- statistic = 5.456 and the related probability value = 0.0653 . Since p value is 0.0653 , which is more than 0.05 , so we accept our null hypothesis, that means we reject our alternate hypothesis . So , there is no heteroscedasticity present in our model.

2. Glejser Test

This test is based on absolute vale of error . In EViews , the following are the steps to test Glejser Test :

a) Estimate the regression model (Y_C_X1_X2)
b) Go to view from estimated results
c) Select residual diagnostics
d) Click on Heteroscedasticity tests
e) Select from test type – Glejser test
f) Click ok

The results shown by EViews will be:



To check the presence of heteroscedasticity , we test the hypothesis of chi –square distribution (observations*R-squared).

INTERPRETATION

Hypothesis framing:
H_0: no heteroscedasticity or homoscedasticity present
H_1: presence of heteroscedasticity

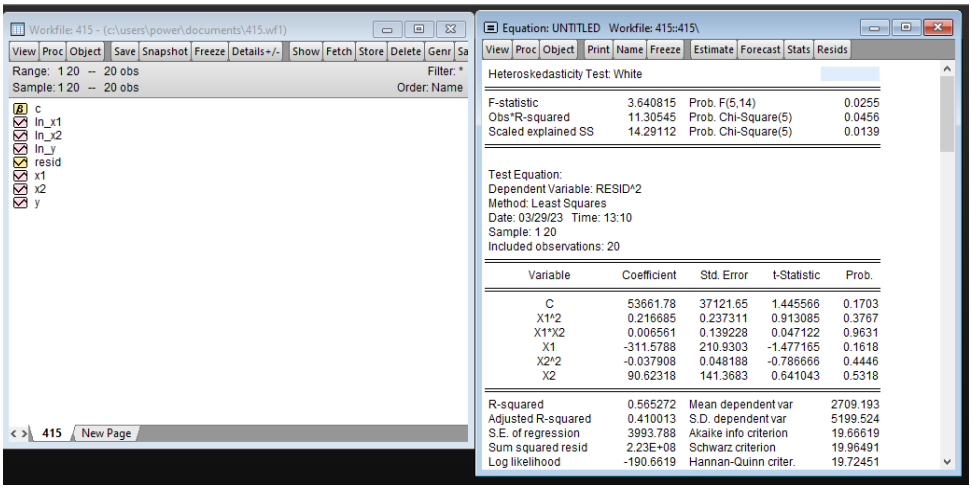
By the above results from EViews , the value of Obs*R-squared- statistic = 4.479 and the related probability value = 0.1065 . Since p value is 0.1065 , which is more than 0.05 , so we accept our null hypothesis, that means we reject our alternate hypothesis . So , there is no heteroscedasticity present in our model.

3. White's Test

This test is not based on normal distribution assumption. In EViews , the following are the steps to test White's Test :

a) Estimate the regression model (Y_C_X1_X2)
b) Go to view from estimated results
c) Select residual diagnostics
d) Click on Heteroscedasticity tests
e) Select from test type – White's test
f) Click ok

The results shown by Eviews will be:



To check the presence of heteroscedasticity , we test the hypothesis of chi –square distribution (observations*R-squared).

INTERPRETATION

Hypothesis framing:
H_0: no heteroscedasticity or homoscedasticity present
H_1: presence of heteroscedasticity

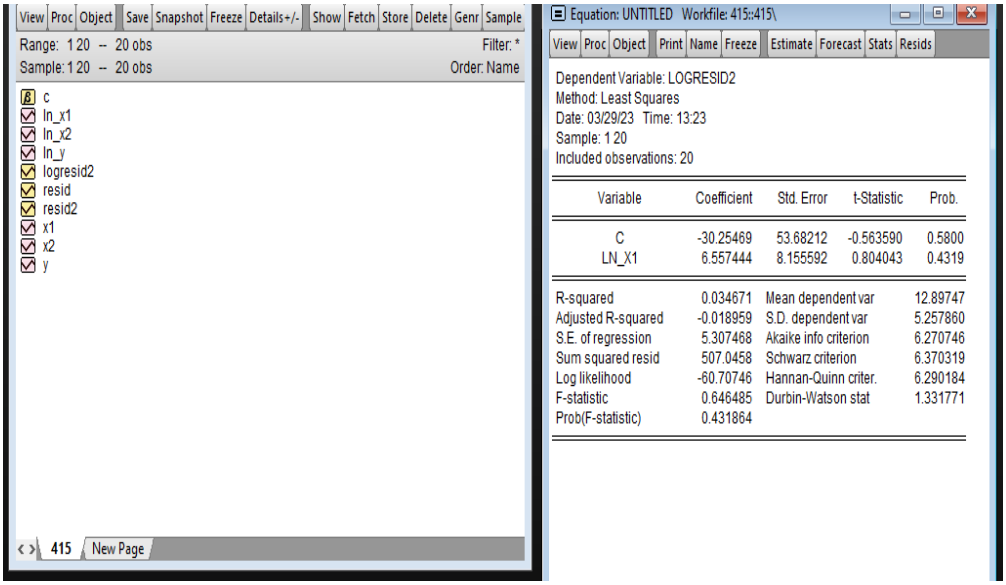
By the above results from EViews , the value of Obs*R-squared- statistic = 11.3054 and the related probability value = 0.0456 . Since p value is 0.0456 , which is less than 0.05 , so we reject our null hypothesis, that means we accept our alternate hypothesis . So , there is heteroscedasticity present in our model according to this test.

4. Park Test

In this test , error variance is regressed with independent variable . To test this in EViews , the following steps are involved :

a) After exporting of data from excel to EViews , with log forms of independent variables
b) From dialog box click on genr
c) Firstly square the resid term by – (resid2=resid^2)
d) Now take the log of same by – [logresid2=log(resid2)]
e) Now estimate the equation by either of two methods as – log(resid2)_c_lnX1
f) Click ok

The below is the result in EViews :



In this case , the presence of heteroscedasticity is tested by hypothesis testing of the slope coefficient of estimated model.

INTERPRETATION

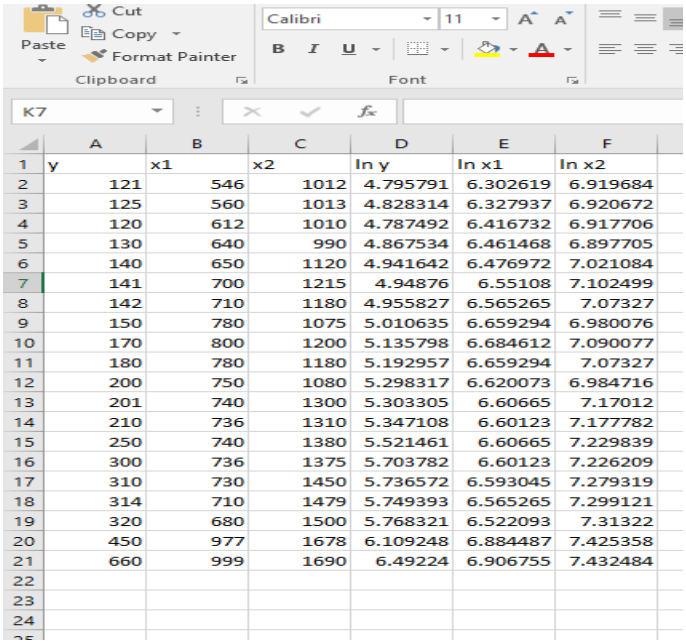
Formation of hypothesis:

H_0: No heteroscedasticity or homoscedasticity present
H_1: Heteroscedasticity present

By the above results from EViews, the value of slope coefficient = 6.5574 and the related probability value = 0.4319 . Since p value is 0.4319, which is more than 0.05 , so we accept our null hypothesis, that means we reject our alternate hypothesis . So , there is no heteroscedasticity present in our model according to this test.

Remedial Measures

Like multicollinearity, there are methods to correct heteroscedasticity too. One such method is changing the functional form as done in multicollinearity.
To do so take the log forms of all variable for converting it into log linear model and import in EViews .

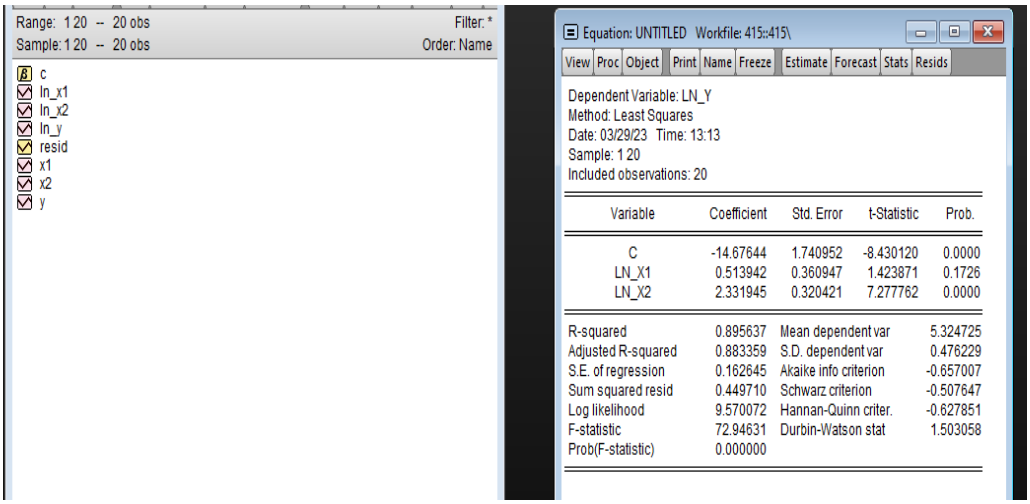


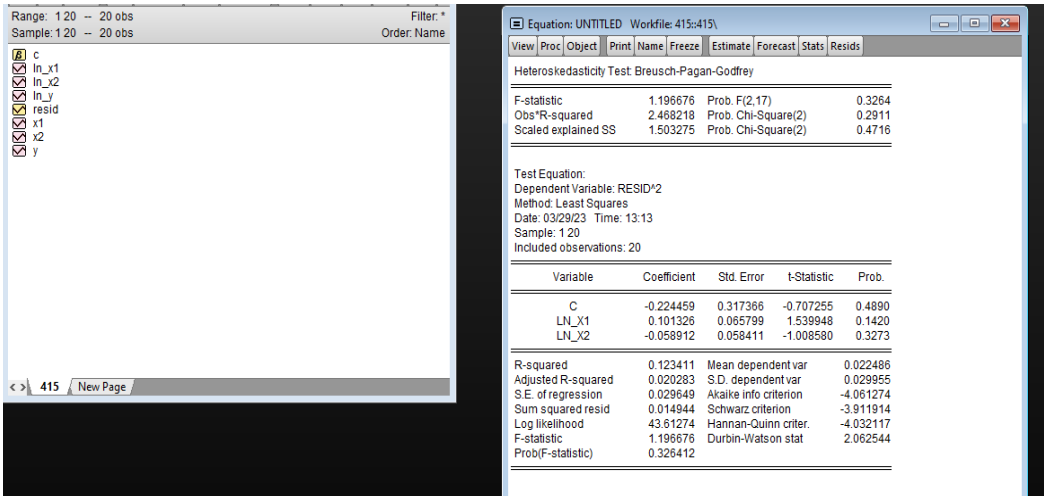
	A	B	C	D	E	F
1	y	x1	x2	ln y	ln x1	ln x2
2	121	546	1012	4.795791	6.302619	6.919684
3	125	560	1013	4.828314	6.327937	6.920672
4	120	612	1010	4.787492	6.416732	6.917706
5	130	640	990	4.867534	6.461468	6.897705
6	140	650	1120	4.941642	6.476972	7.021084
7	141	700	1215	4.94876	6.55108	7.102499
8	142	710	1180	4.955827	6.565265	7.07327
9	150	780	1075	5.010635	6.659294	6.980076
10	170	800	1200	5.135798	6.684612	7.090077
11	180	780	1180	5.192957	6.659294	7.07327
12	200	750	1080	5.298317	6.620073	6.984716
13	201	740	1300	5.303305	6.60665	7.17012
14	210	736	1310	5.347108	6.60123	7.177782
15	250	740	1380	5.521461	6.60665	7.229839
16	300	736	1375	5.703782	6.60123	7.226209
17	310	730	1450	5.736572	6.593045	7.279319
18	314	710	1479	5.749393	6.565265	7.299121
19	320	680	1500	5.768321	6.522093	7.31322
20	450	977	1678	6.109248	6.884487	7.425358
21	660	999	1690	6.49224	6.906755	7.432484

From the above results , we apply B-P-G test to test the presence or absence of heteroscedasticity after changing the functional form to log linear .
In EViews , the following are the steps to test this :

a) Estimate the regression model (lnY_C_lnX1_lnX2)
b) Go to view from estimated results
c) Select residual diagnostics
d) Click on Heteroscedasticity tests
e) Select from test type – B.P.G test
f) Click ok

The results shown by EViews will be:





INTERPRETATION

Hypothesis framing:
H_0: no heteroscedasticity or homoscedasticity present
H_1: presence of heteroscedasticity

By the above results from EViews, the value of Obs*R-squared- statistic = 2.468 and the related probability value = 0.2911 . Since p value is 0.2911, which is more than 0.05, so we accept our null hypothesis, that means we reject our alternate hypothesis . So, there is no heteroscedasticity present in our model in case of log linear also. Our prob. value increased in the case of log linear model than that of linear.

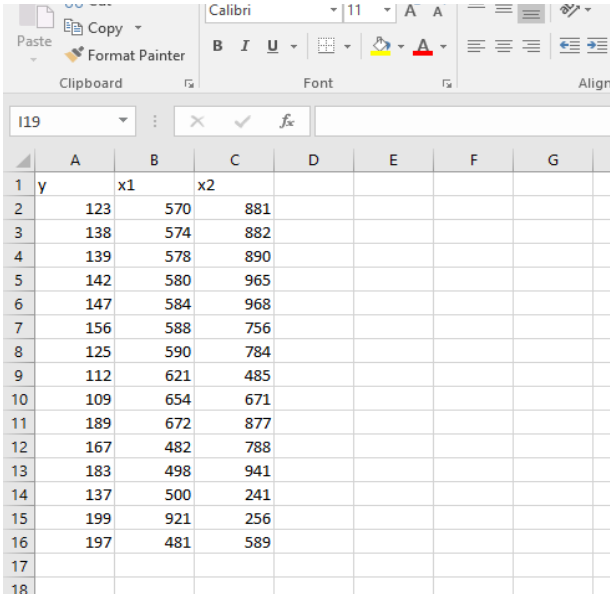
UNIT 3 AUTOCORRELATION

Meaning of Autocorrelation

Autocorrelation means the relationship between each value of errors in the equation. Or in the other hand, autocorrelation means the self relationship of errors. This assumption is popularly found in time-series data. It is the lagged relationship between errors.

Detection of Autocorrelation

The data series to be worked upon and some ways through which autocorrelation can be detected are given below :



	A	B	C	D	E	F	G
1	y	x1	x2				
2	123	570	881				
3	138	574	882				
4	139	578	890				
5	142	580	965				
6	147	584	968				
7	156	588	756				
8	125	590	784				
9	112	621	485				
10	109	654	671				
11	189	672	877				
12	167	482	788				
13	183	498	941				
14	137	500	241				
15	199	921	256				
16	197	481	589				
17							
18							

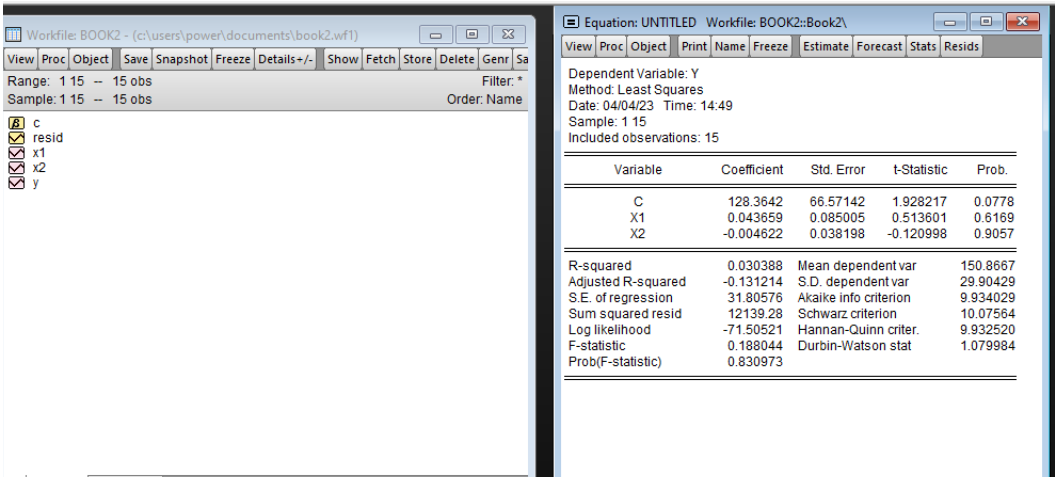
1.Durbin Watson- d Statistic

Steps to test autocorrelation under this test are:

a) Estimate equation by going to quick or selecting the variables after importing of data .
b) Estimate equation as Y_C_X1_X2
c) In results there will be value of Durbin Watson statistic too.

RULE OF THUMB – If the value of d is approximately 2 , then there is no autocorrelation present . If the value is less than 2 , there is positive autocorrelation and if the value is more than 2 , there is negative autocorrelation.

Below are the results shown by Eviews-



INTERPRETATION

Since by above results, the value of Durbin Watson d statistic is 1.079 , which is less than 2
So , there will exist positive autocorrelation in the model.

DRAWBACK

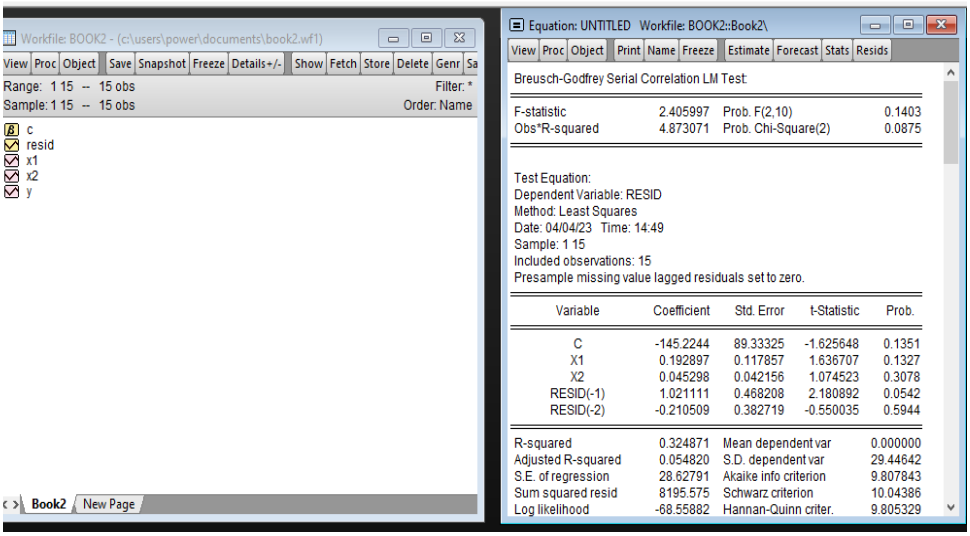
Durbin Watson d statistic doesn't provide the p value to test the significance level for hypothesis testing. Also , it doesn't talk about serial correlation i.e., autocorrelation present in time series data.

2. B.G. – L.M. Serial Correlation Test

To test serial correlation and to remove the drawback of D.W. test , B.G.- L.M. test is used .
The steps of performing this test in Eviews are :

a) Estimate the equation – Y_c_X1_X2
b) Go to view from the estimated results
c) Select residual diagnostics
d) Click on serial correlation LM test
e) Confirm lags as shown by the software
f) Press ok

The results for the same in EViews will be shown as –



INTERPRETATION

In this case test of significance approach based on Chi-Square is used .

Formation of hypothesis:

H_0 : No serial correlation present
H_1 : Serial correlation present

By the above results from EViews, the value of Obs*R-squared- statistic = 4.873 and the related probability value = 0.0675 . Since p value is 0.0675, which is more than 0.05, so we accept our null hypothesis, that means we reject our alternate hypothesis . So, there is no serial correlation present in our model.

UNIT 4 SPECIFICATION ERROR

MEANING OF SPECIFICATION ERROR

In regression analysis and related fields such as econometrics, specification is the process of converting a theory into a regression model.
There could be four types of specification errors; inclusion of an irrelevant variable, exclusion of a relevant variable, incorrect functional form and measurement error. When the econometric model is not specified correctly, the coefficient estimates, the confidence intervals, and the hypothesis tests are misleading and inconsistent.

Detection of Specification error

The data series to be worked upon and some ways through which specification error can be detected are given below :

	A	B	C	D	E
1	y	x1	x2	x3	
2	245	867	741	345	
3	256	753	369	465	
4	356	159	157	789	
5	566	534	322	236	
6	598	234	158	156	
7	845	487	741	735	
8	547	745	458	789	
9	512	411	822	732	
10	245	125	211	987	
11	589	689	265	923	
12	988	745	598	973	
13	974	448	574	845	
14	521	987	412	634	
15	211	852	256	267	
16	117	123	349	845	
17	145	456	943	825	
18	658	987	746	456	
19	956	741	158	734	
20	688	369	987	734	
21	999	125	913	999	
22					
23					
24					

1. Ramsey RESET (Regression Equation Specification Error Test)

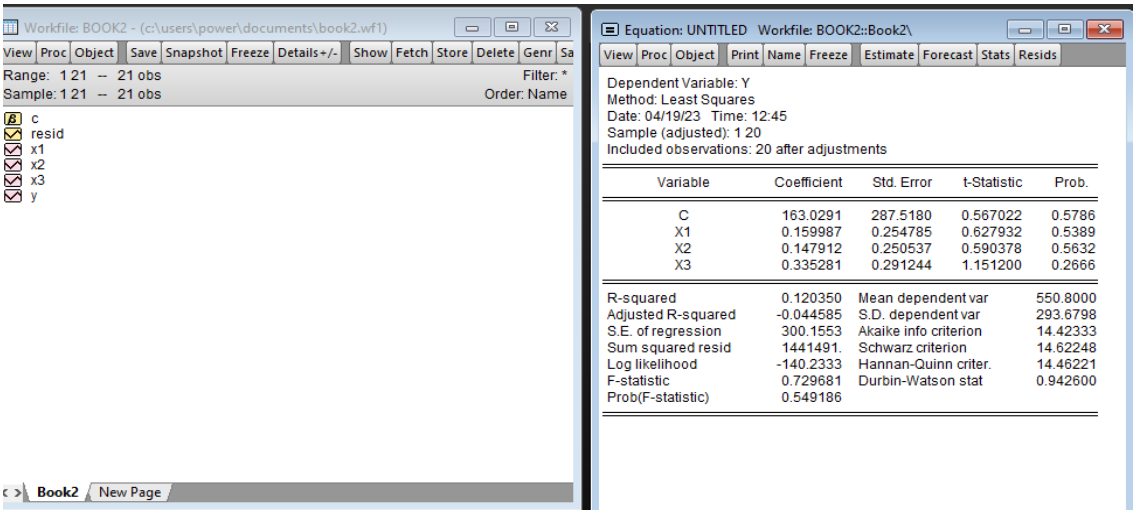
This is a general test which is used to test the specification error.

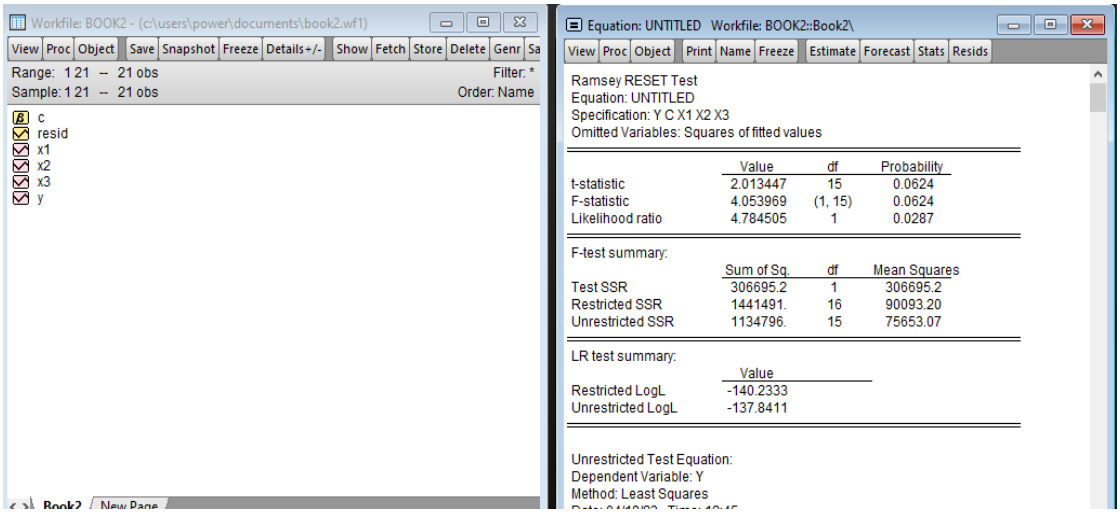
Steps to perform this test in EViews are :

a) Estimate the equation – Y_c_X1_X2_X3
b) Go to view from the estimated results
c) Select stability diagnostics
d) Click on Ramsey RESET
e) Confirm lags as shown by the software

f) Press ok

The results for the same in EViews will be shown as –





INTERPRETATION

In this case test of significance approach is based on t- statistic , F- statistic and likelihood ratio. Mainly F – statistic is used for the purpose.

Formation of hypothesis:

H_0 : No specification error present
H_1 : Specification error present

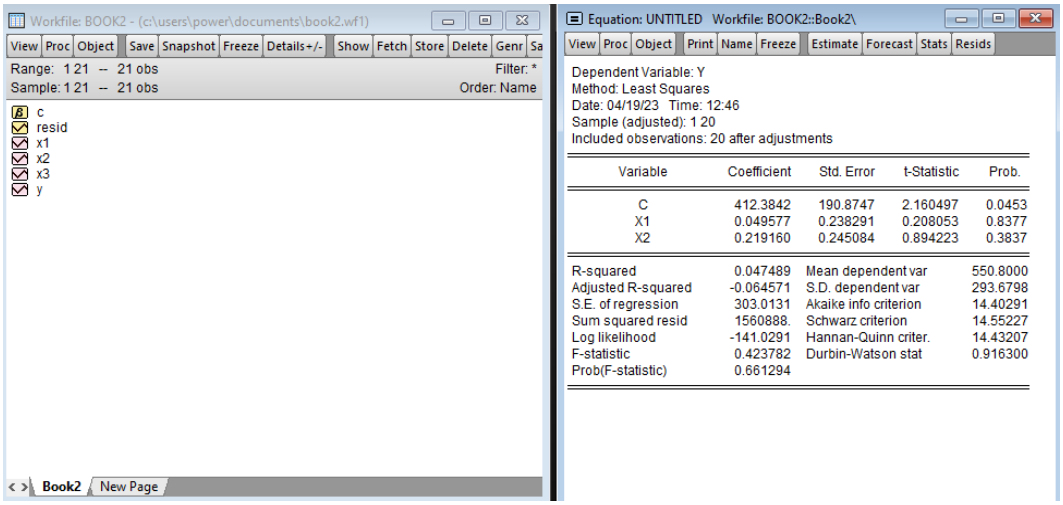
By the above results from EViews, the p value is more than 0.05 for t and F statistic , so we accept our null hypothesis, that means we reject our alternate hypothesis . So, there is no specification error present in our model but for likelihood ratio since the p value is less than 0.05 , the hypothesis is rejected.

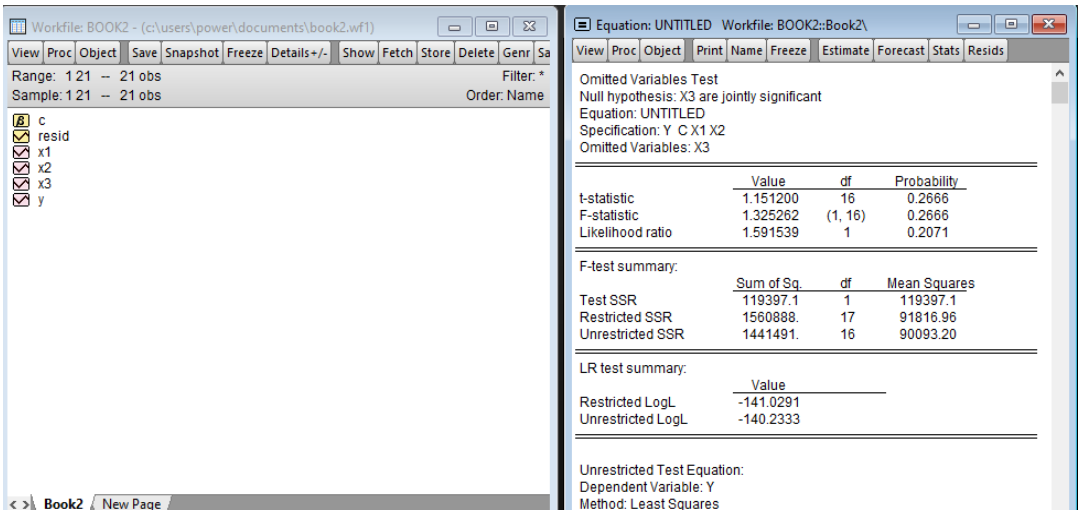
2. Omitted Variable Case

Steps to perform this test in Eviews are :

a) Estimate the equation – Y_c_X1_X2 (omitting X3)
b) Go to view from the estimated results
c) Select coefficient diagnostics
d) Click on Omitted variable test – Likelihood ratio
e) A new dialog box will appear
f) Type omitted variable as X3
g) Click ok

The results in EViews will be shown as:





INTERPRETATION

In this case test of significance approach is based on t- statistic , F- statistic and likelihood ratio.

Formation of hypothesis:

H_0 : No impact of omitted variable on model or X3 is irrelevant
H_1 : omitted variable has an impact on model

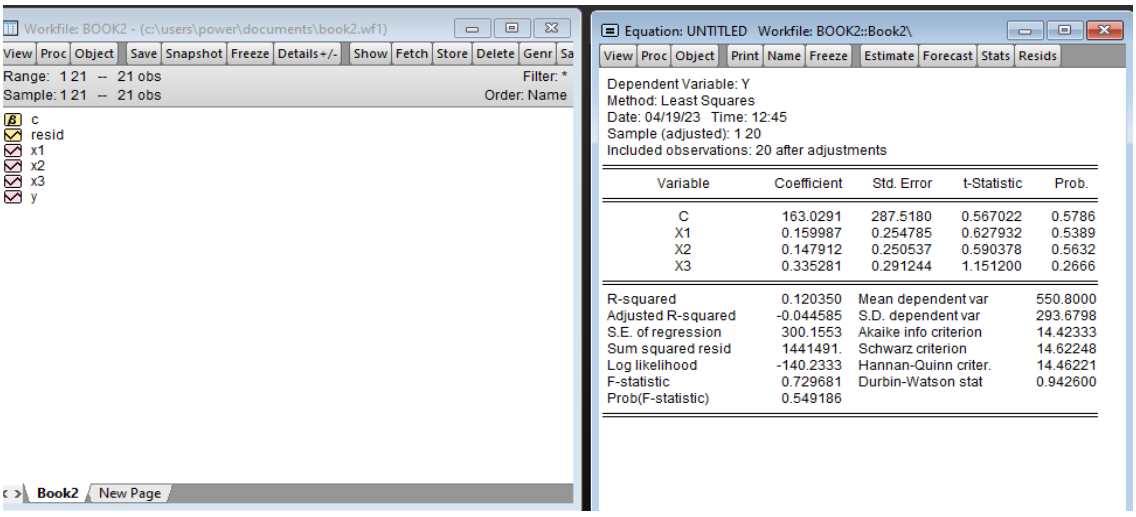
By the above results from EViews, the p value is more than 0.05 for t-statistic ,F statistic and likelihood ratio , so we accept our null hypothesis, that means we reject our alternate hypothesis . So, there is no impact of omitted variable on model or X3 is irrelevant.

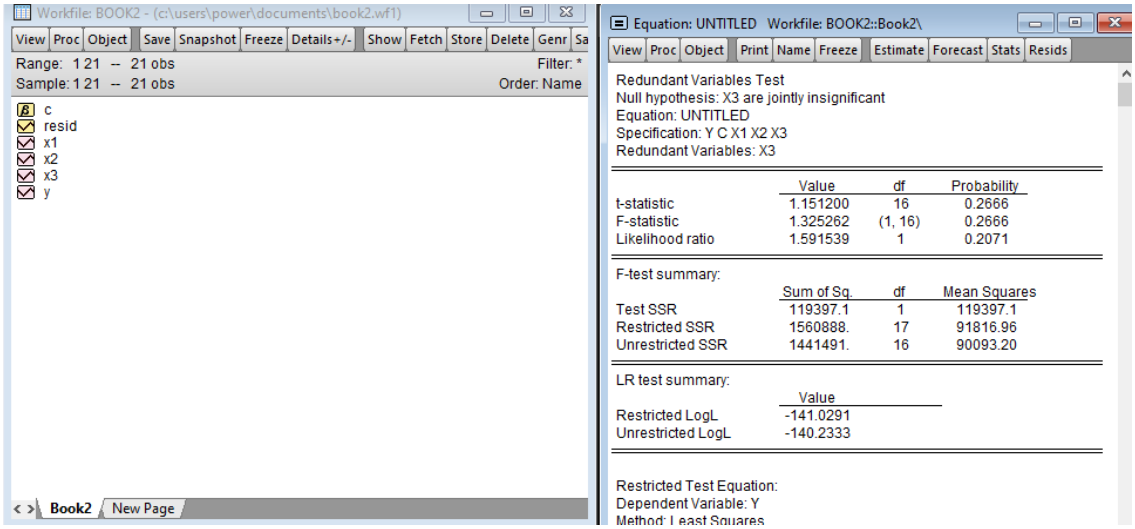
3. Redundant Variable Case(Inclusion of Irrelevant Variable)

Steps to perform this test in EViews are :

a) Estimate the equation – Y_c_X1_X2_X3 (inclusion of X3)
b) Go to view from the estimated results
c) Select coefficient diagnostics
d) Click on Redundant variable test – Likelihood ratio
e) A new dialog box will appear
f) Type included variable as X3
g) Click ok

The results in EViews will be shown as:





INTERPRETATION

In this case test of significance approach is based on t- statistic , F- statistic and likelihood ratio.

Formation of hypothesis:

H_0 : X3 is redundant variable or is irrelevant
H_1 : X3 is not redundant variable and is relevant

By the above results from EViews, the p value is more than 0.05 for t-statistic ,F statistic and likelihood ratio , so we accept our null hypothesis, that means we reject our alternate hypothesis . So, X3 is irrelevant or redundant variable.