Tutorial 9

keywords: variance, error, heteroskedasticity, homoskedasticity, residual plots, Breusch-Pagan test, White test, WLS

estimated reading time: 36 minutes

Quang Bui

September 18, 2018

Question 1

EViews workfile: *profits.wf1*

(This is a continuation of question 2 in Part A) The workfile *profits.wf1* includes data on profits and assets of 88 firms (some firms having missing data). The variables in the data set are,

$$\begin{aligned} profits &- firm's\ profit\ in\ million\ dollars \\ assets &- firm's\ assets\ in\ million\ dollars \\ mno &- a\ dummy\ variable\ which\ equals\ 1\ if\ the\ CEO\ of\ the\ firm\ is\ not\ the\ owner, \\ &\quad 0\ otherwise \end{aligned}$$

	PROFITS	MNO	ASSETS
1	66.00	0.00	1918.30
2	28.40	0.00	789.20
3	18.40	0.00	574.30
4	5.80	0.00	353.00
5	54.50	0.00	568.40
6	3.80	0.00	280.00
7	7.40	0.00	161.40
8	24.20	0.00	274.40
9	4.80	0.00	184.60
10	12.80	0.00	192.40

Our objective is to test the hypothesis that the relationship between profits and assets is the same for owner-managed and non-owner-managed firms. The general model is:

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 (mno_i \times assets_i) + u_i$$

(a) Formulate the null hypothesis that the nature of ownership of the firm (i.e. whether the firm is managed by its owner or not) does not affect the relationship between profits and assets in a firm and the alternative that it does.

$$\begin{aligned} E(profits|mno = 0, assets) &= \beta_0 + \delta_0 \times 0 + \beta_1 assets + \delta_1 \times 0 \times assets \\ &= \beta_0 + \beta_1 assets \end{aligned}$$

$$\begin{aligned} E(profits|mno = 1, assets) &= \beta_0 + \delta_0 \times 1 + \beta_1 assets + \delta_1 \times 1 \times assets \\ &= (\beta_0 + \delta_0) + (\beta_1 + \delta_1) assets \end{aligned}$$

δ_0 and δ_1 represent the different in intercept and slope for the relationship between profits and assets for non-owner managed and owner managed firms.

If there is no difference:

$$\delta_0 = \delta_1 = 0$$

and if there is a difference:

$$\delta_0 \neq 0 \; and/or \; \delta_1 \neq 0$$

(b) Estimate the model using OLS and test whether the errors are homoskedastic using the following tests (in each case, answer the question as if it were a question in the final exam, i.e. write down the null and alternative hypothesis, the test statistics and its distribution under the null, the auxiliary regression you should estimate to compute the test statistic, then compute the test statistic and compare it with the critical value you get from statistical tables):

i. Breusch-Pagan test when the alternative hypothesis is $Var(u_i|mno_i, assets_i) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i$

ii. White test

iii. The special form of the White test that uses the predicted value of $profits$ and its square as predictors of variance.

Background

Homoskedasticity and the OLS estimator

For a model that is linear in parameters without perfect collinearity, if the assumption for unbiasedness (zero conditional mean) holds,

$$E(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{0}$$

the OLS estimator becomes an unbiased linear estimator. If, in addition to the unbiasedness assumption, the errors are serially uncorrelated and homoskedastic,

$$\begin{aligned} Var(\boldsymbol{u}|\boldsymbol{X}) &= \sigma^2 \boldsymbol{I}_n \\ &= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \vdots & 0 \\ \vdots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} \end{aligned}$$

$$(off\ diagonal\ elements = 0 \implies unserially\ uncorrelated\ errors)$$

$$(diagonal\ elements = \sigma^2 \implies homoskedastic\ errors)$$

then the OLS estimator becomes the *most efficient linear unbiased estimator.*

Consequences of heteroskedastic errors

- If all Gauss-Markov assumptions hold, except the assumption of homoskedastic errors, then the OLS estimator is no longer BLUE i.e. the OLS estimator is not the best linear unbiased estimator.

- A more several consequence is that the t and F test statistics not longer have a t and F distribution, so the t and F test are no longer valid.

Understanding the variance of the error: $Var(\boldsymbol{u}|\boldsymbol{X}) = Var(\boldsymbol{y}|\boldsymbol{X})$

In our model of a firm's $profit$,

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 mno_i \times assets_i + u_i$$

the error term is homoskedastic if the variance of the error is constant i.e. the variance is fixed across all i (if the error does not vary differently across different i then it is constant regardless of the value of x),

$$Var(u_i|mno_i, assets_i) = \sigma^2$$

If the error term is heteroskedastic, then the variance of the error is not constant,

$$Var(u_i|mno_i, assets_i) \neq \sigma^2$$

Since β_0, β_1, δ_0, and δ_1 are constants and mno_i and $assets_i$ are known (we have conditioned on mno_i and $assets_i$),

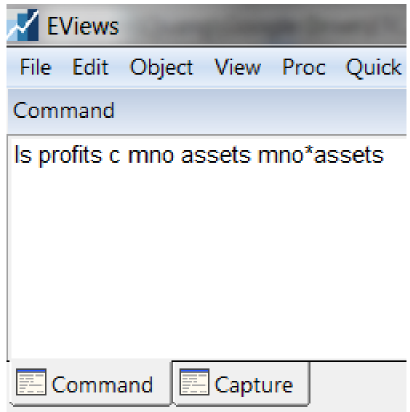
$$\begin{aligned} Var(profit_i|mno_i, assets_i) &= Var(\beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 mno_i \times assets_i \\ &\quad + u_i|mno_i, assets_i) \\ &= Var(u_i|mno_i, assets_i) \end{aligned}$$

This tells us that if the variance of the dependent variable is not constant, then the variance of the error will not be constant i.e. the error is heteroskedastic.

Estimate the model using OLS, then save the residuals (the 3 test for heteroskedasticity requires running a regression with the squared residual as the dependent variable).

From the Command window:

$$ls\ profit\ c\ mno\ assets\ mno^{*}assets$$



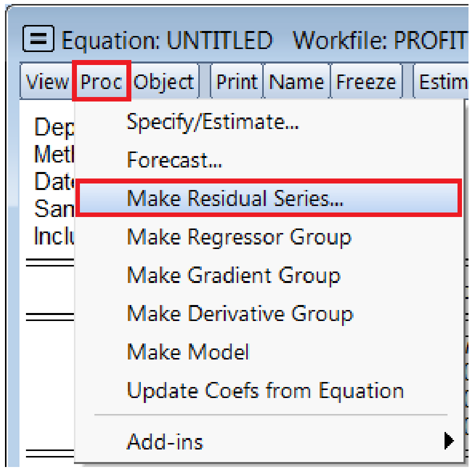
Dependent Variable: PROFITS
Method: Least Squares
Date: 09/15/18 Time: 17:13
Sample: 1 88
Included observations: 69

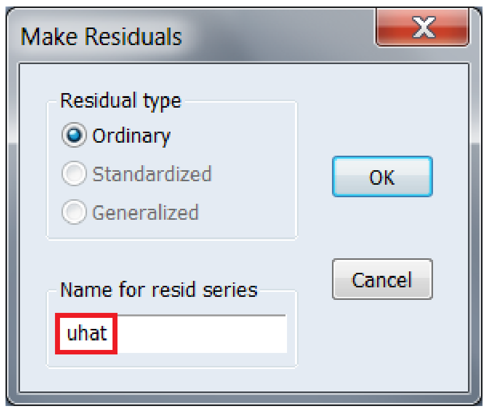
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.561369	2.780089	0.561626	0.5763
MNO	8.231865	4.087202	2.014059	0.0481
ASSETS	0.050647	0.006196	8.174306	0.0000
MNO*ASSETS	−0.052172	0.009294	−5.613231	0.0000

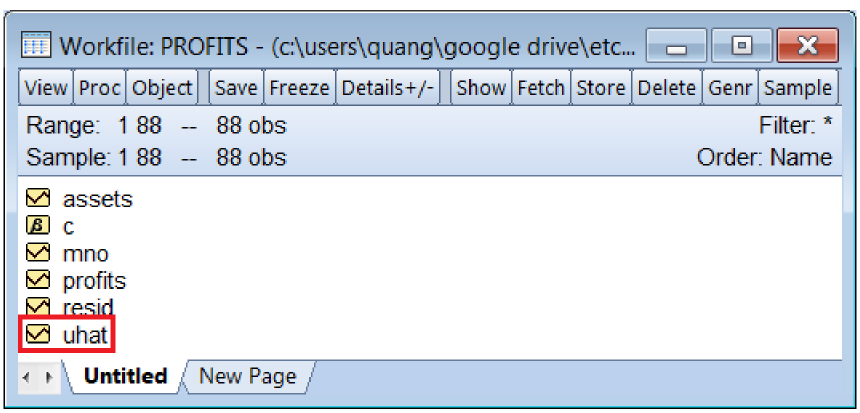
R-squared	0.521963	Mean dependent var	12.83188
Adjusted R-squared	0.499900	S.D. dependent var	18.62111
S.E. of regression	13.16843	Akaike info criterion	8.049745
Sum squared resid	11271.50	Schwarz criterion	8.179258
Log likelihood	−273.7162	Hannan-Quinn criter.	8.101127
F-statistic	23.65757	Durbin-Watson stat	2.410291
Prob(F-statistic)	0.000000		

Save the OLS residuals into a separate series (the series *resid* contains the residuals from the last estimated model),

$$Proc \rightarrow Make\ to\ Residual\ Series \rightarrow Name\ for\ resid : uhat \rightarrow OK$$


Equation: UNTITLED Workfile: PROFIT
View Proc Object Print Name Freeze Estim
Specify/Estimate...
Forecast...
Make Residual Series...
Make Regressor Group
Make Gradient Group
Make Derivative Group
Make Model
Update Coefs from Equation
Add-ins


Make Residuals
Residual type
Ordinary
Standardized
Generalized
OK
Cancel
Name for resid series
uhat


Workfile: PROFITS - (c:\users\quang\google drive\etc...
View Proc Object Save Freeze Details+/- Show Fetch Store Delete Genr Sample
Range: 1 88 -- 88 obs
Filter: *
Sample: 1 88 -- 88 obs
Order: Name
assets
c
mno
profits
resid
uhat
Untitled
New Page

Background

Breusch-Pagan test for heteroskedasticity

Since $E(u_i|mno_i, assets_i) = 0$,

$$\begin{aligned} Var(u_i|mno_i, assets_i) &= E(u_i^2|mno_i, assets_i) - (E(u_i|mno_i, assets_i))^2 \\ &= E(u_i^2|mno_i, assets_i) \end{aligned}$$

we can think about a possible 'model' for the $Var(u_i|mno_i, assets_i)$ i.e. one that contains variables that helps to explain $Var(u_i|mno_i, assets_i)$.

By considering a model of the squared error term from our model of firm's profits,

$$\begin{aligned} u_i^2 &= \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq} + v_i \\ &= E(u_i^2|z_{i1}, z_{i2}, \ldots, z_{iq}) + v_i \\ &= Var(u_i|z_{i1}, z_{i2}, \ldots, z_{iq}) + v_i \end{aligned}$$

we see that it is easy to perform a test to see if at least one of the z variables helps to explain the variance of the error.

Consider z to be a variable of any function of the x variables in our model **or** any other variable that we have data on which we think can help to explain $Var(profits_i|mno_i, assets_i)$ and $\therefore$ $Var(u_i|mno_i, assets_i)$ (so the z variables do not have to be *mno* and *assets*). That is,

$$Var(u_i|mno_i, assets_i) = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq}$$

Ideally, we would want to run a regression on,

$$\begin{aligned} u_i^2 &= \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq} + v_i \\ &= E(u_i^2|mno_i, assets_i) + v_i \\ &= Var(u_i|mno_i, assets_i) + v_i \end{aligned}$$

and test for heteroskedasticity by testing if at least one of the z variables helps to explain the variance of u and therefore the variance of $profits$ (any variable z that helps to explain u^2, is a variable that helps to explain the variance of u),

$$H_1 : \textit{at least one of } \delta_j \textit{ does not equal } 0 \quad \textit{for } j = 1, 2, \ldots, q$$

but this is not feasible because we do not observe u^2 so we cannot run the regression.

What we do is replace u with the observable OLS residuals $\hat{u}$,

$$\hat{u}_i^2 = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq} + v_i$$

run this *auxiliary regression*,

$$Quick \rightarrow Estimate\ Equation \rightarrow \ldots$$

then test if at least one of the z variables has explanatory power in explaining the variance of u,

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_q = 0$$
$$H_1 : at\ least\ one\ of\ the\ above\ \delta_i \neq 0\ for\ i = 1, 2, .., q$$

Concretely, this is equivalent to testing the null that u is homoskedastic against the alternative that it is heteroskedastic,

$$H_0 : E(u_i^2|mno_i, assets_i) = \sigma^2 \quad (\delta_1 = \delta_2 = \cdots = \delta_q = 0 \quad homoskedastic\ errors)$$
$$H_1 : E(u_i^2|mno_i, assets_i) \neq \sigma^2 \quad (at\ least\ one\ of\ the\ above\ \delta's \neq 0 \quad heteroskedastic\ errors)$$

(σ^2 and δ_0 are constants. Notation is arbitrary so we could have used δ_0 in our null and alternative hypothesis.)

An equivalent and more intuitive way to write the null and alternative hypothesis is,

$$H_0 : Var(u_i|mno_i, assets_i) = \sigma^2 \quad (homoskedastic\ errors)$$
$$H_1 : Var(u_i|mno_i, assets_i) \neq \sigma^2 \quad (heteroskedastic\ errors)$$

Breusch-Pagan test

Since we are performing a Breusch-Pagan test when the alternative hypothesis is $Var(u_i|mno_i, assets_i) = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i$, our z variables are mno_i and $assets_i$, which gives the following *auxiliary regression*

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i + v_i$$

$$H_0 : E(u_i^2|mno_i, assets_i) = \sigma^2 \qquad (\alpha_1 = \alpha_2 = 0 \quad homoskedastic\ errors)$$

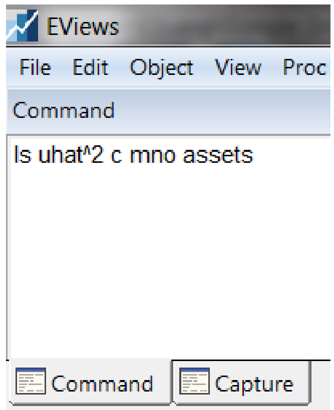
$$H_1 : E(u_i^2|mno_i, assets_i) \neq \sigma^2 \qquad (\alpha_1\ and/or\ \alpha_2 \neq 0 \quad heteroskedastic\ errors)$$

The Breusch-Pagan test statistic:

$$BP = n \times R^2_{aux} \sim \chi^2_{q=2} \quad under\ H_0$$

$$R^2_{aux} : R^2 \ from \ the \ auxiliary \ regression$$
$$q : number \ of \ regressors \ in \ the \ auxiliary \ regression$$

Estimate the auxiliary regression for the Breusch-Pagan test in EViews,



Dependent Variable: UHAT^2
Method: Least Squares
Date: 09/15/18 Time: 18:42
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.83989	119.7623	0.107211	0.9149
MNO	−157.4662	150.7769	−1.044365	0.3001
ASSETS	0.806402	0.219029	3.681709	0.0005

R-squared	0.182682	Mean dependent var	163.3550
Adjusted R-squared	0.157914	S.D. dependent var	680.5634
S.E. of regression	624.5205	Akaike info criterion	15.75435
Sum squared resid	25741705	Schwarz criterion	15.85149
Log likelihood	−540.5251	Hannan-Quinn criter.	15.79289
F-statistic	7.375942	Durbin-Watson stat	2.564701
Prob(F-statistic)	0.001285		

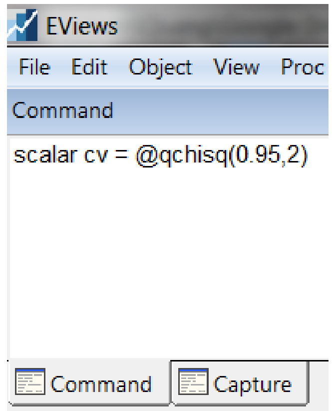
The estimated auxiliary regression:

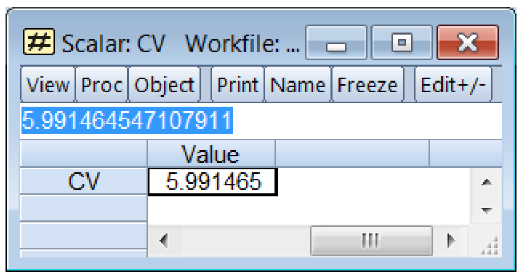
$$\widehat{u_i^2} = 12.8399 - 157.4662 mno_i + 0.8064 assets_i \quad R^2_{aux} = 0.1827$$

The calculated BP test statistic:

$$BP_{calc} = 69 \times 0.1827 = 12.6050$$

To obtain the critical value in EViews,





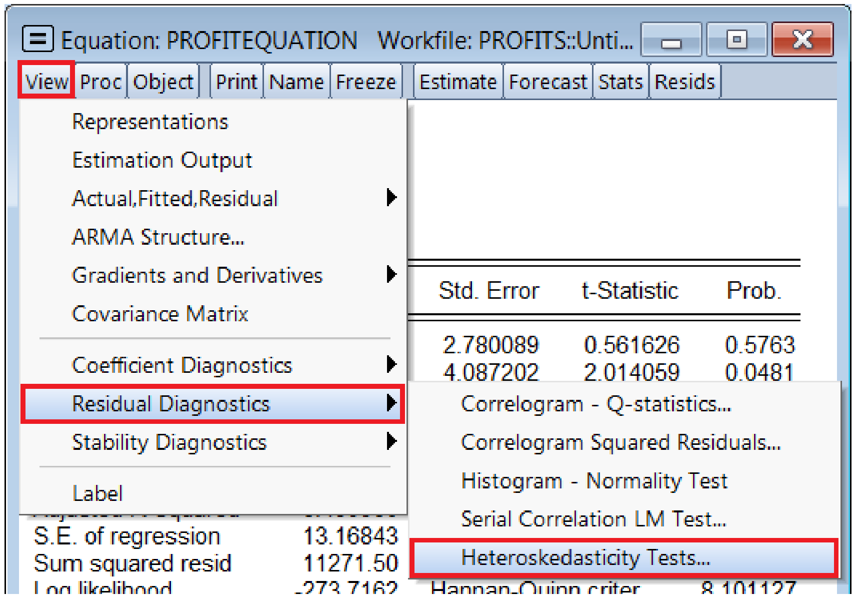
Since $BP_{calc} = 12.6050 > \chi^2_{crit} = 5.99$, we reject H_0 at the 5% significance level and conclude that there is sufficient evidence from our sample to suggest that the errors are heteroskedastic.

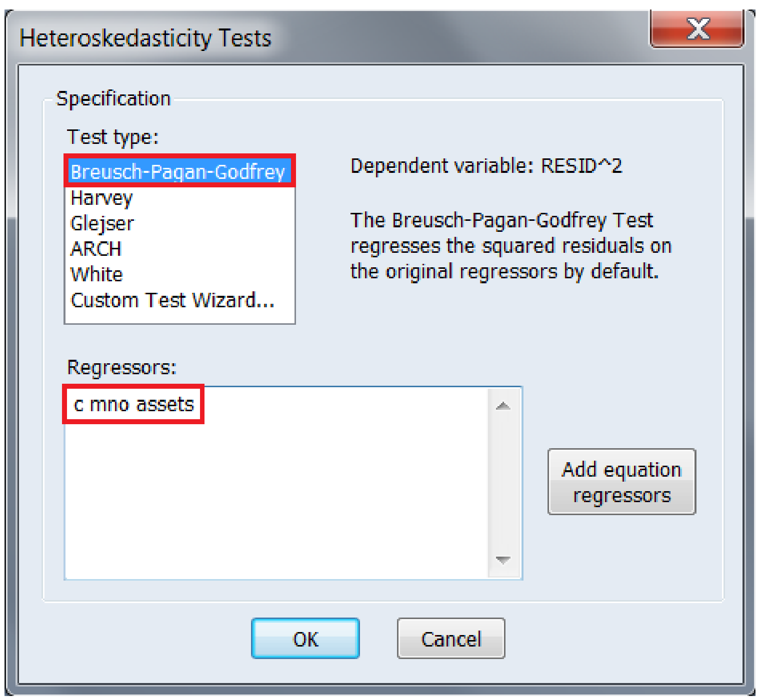
Note: EViews has an inbuilt Breusch-Pagan and White test for heteroskedasticity which you can use to verify your results,

$$\textit{After you estimate your model with OLS}$$

$$\textit{From the equation object}:$$

$$\textit{View} \rightarrow \textit{Residual Diagnostics} \rightarrow \textit{Heteroskedasticity Tests...}$$





$$\textbf{Test type}: Breusch - Pagan - Godfrey$$

$$\textbf{Regressors}: c \quad mno \quad assets$$

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	7.375942	Prob. F(2,66)	0.0013
Obs*R-squared	12.60503	Prob. Chi-Square(2)	0.0018
Scaled explained SS	95.66969	Prob. Chi-Square(2)	0.0000

Test Equation:
Dependent Variable: RESIDˆ2
Method: Least Squares
Date: 09/15/18 Time: 19:06
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.83989	119.7623	0.107211	0.9149
MNO	−157.4662	150.7769	−1.044365	0.3001
ASSETS	0.806402	0.219029	3.681709	0.0005

R-squared	0.182682	Mean dependent var	163.3550
Adjusted R-squared	0.157914	S.D. dependent var	680.5634
S.E. of regression	624.5205	Akaike info criterion	15.75435
Sum squared resid	25741705	Schwarz criterion	15.85149
Log likelihood	−540.5251	Hannan-Quinn criter.	15.79289
F-statistic	7.375942	Durbin-Watson stat	2.564701
Prob(F-statistic)	0.001285		

$$BP_{calc} = 12.6050 \qquad p - value = 0.0018$$

White test

Background

White test for heteroskedasticity

If we wish to test for heteroskedasticity without precise knowledge of the relevant variables, we could implement the White test. White's test for heteroskedasticity specifies that the variance of the error term is a function of *all regressors in the model of y, the squares of the regressors, and all cross-product combinations of the regressors* (obviously omitting any duplicate variables as this results in perfect collinearity).

Our model of $profits$

$$profits_i = \beta_0 + \delta_0 mno_i + \beta_1 assets_i + \delta_1 mno_i^* assets_i + u_i$$

contains the following regressors:

- mno
- $assets$
- $mno^* assets$

so the functional form of the error variance for White's test of heteroskedasticity is given by:

$$\begin{aligned} Var(u_i|mno_i, assets_i) = \alpha_0 &+ \alpha_1 mno_i + \alpha_2 assets_i + \alpha_3 mno_i^* assets_i + \alpha_4 assets_i^2 \\ &+ \alpha_5 mno_i^* assets_i^* assets_i \end{aligned}$$

As you can see, there are no duplicate variables in the error variance e.g. mno_i and mno_i^2 are the same variables so only need to include one of them (here I have included mno_i and omitted mno_i^2).

Ideally, we would estimate the regression model,

$$\begin{aligned} u_i^2 &= \alpha_0 + \alpha_1 mno_i + \alpha_2 assets_i + \alpha_3 mno_i^* assets_i + \alpha_4 assets_i^2 \\ &\quad + \alpha_5 mno_i^* assets_i^* assets_i + v_i \\ &= E(u_i^2|mno_i, assets_i) + v_i \\ &= Var(u_i|mno_i, assets_i) + v_i \end{aligned}$$

then test if at least one of the variables helps to explain the error variance but this is not feasible because u^2 is unobserved.

To perform's White's test, we replace u with the observed OLS residuals $\hat{u}$ to give us the following *auxiliary regression*:

$$\hat{u}_i^2 = \alpha_0+\alpha_1 mno_i+\alpha_2 assets_i+\alpha_3 mno_i^* assets_i+\alpha_4 assets_i^2+\alpha_5 mno_i^* assets_i^* assets_i+v_i$$

then test the null and alternative hypothesis:

$$H_0 : E(u_i^2|mno_i, assets_i) = \sigma^2 \quad (\alpha_1 = \alpha_2 = \cdots = \alpha_5 = 0 \quad homoskedastic\ errors)$$
$$H_1 : E(u_i^2|mno_i, assets_i) \neq \sigma^2 \quad (at\ least\ one\ of\ the\ above\ \delta's \neq 0)$$

(The alternative hypothesis is that the variance is a smooth unknown function of mno_i and $assets_i$. σ^2 and δ_0 are constants. Notation is arbitrary so we could have used δ_0 in our null and alternative hypothesis.)

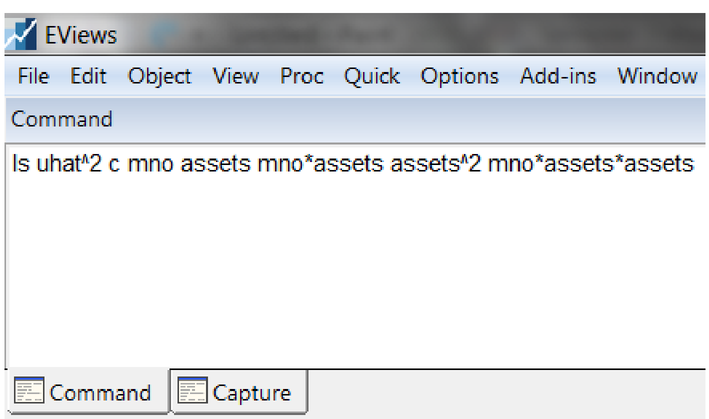
The White test statistic:

$$W = n \times R^2_{aux} \sim \chi^2_{q=5} \quad under\ H_0$$

$$R^2_{aux} : R^2\ from\ the\ auxiliary\ regression$$

$$q : number\ of\ regressors\ in\ the\ auxiliary\ regression$$

Estimate the auxiliary regression for White's test in EViews,



Dependent Variable: UHAT^2
Method: Least Squares
Date: 09/16/18 Time: 13:57
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−470.1101	170.4323	−2.758340	0.0076
MNO	536.1701	263.3993	2.035579	0.0460
ASSETS	3.330239	0.803775	4.143251	0.0001
MNO*ASSETS	−3.291956	1.244107	−2.646040	0.0103
ASSETS^2	−0.001130	0.000452	−2.498314	0.0151
MNO*ASSETS*ASSETS	0.001122	0.000663	1.692514	0.0955

R-squared	0.371760	Mean dependent var	163.3550
Adjusted R-squared	0.321900	S.D. dependent var	680.5634
S.E. of regression	560.4225	Akaike info criterion	15.57820
Sum squared resid	19786623	Schwarz criterion	15.77247
Log likelihood	−531.4479	Hannan-Quinn criter.	15.65527
F-statistic	7.456027	Durbin-Watson stat	2.531454
Prob(F-statistic)	0.000015		

The estimated auxiliary regression:

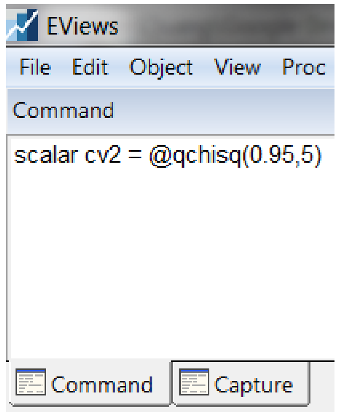
$$\begin{aligned}\widehat{\hat{u}_i^2} &= -470.1101 + 536.1701 mno_i + 3.3302 assets_i - 3.2920 mno_i^* assets_i \\ &- 0.0011 assets_i^2 + 0.0011 mno_i^* assets_i^* assets_i\end{aligned}$$

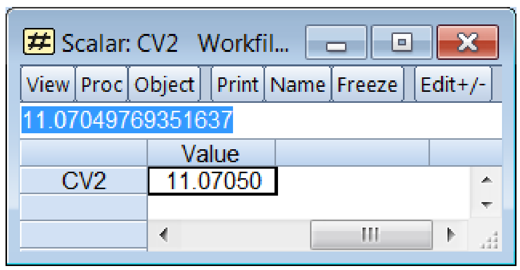
$$R^2_{aux} = 0.3718$$

The calculated White test statistics:

$$W_{calc} = 69 \times 0.3718 = 25.6542$$

To obtain the critical value in EViews,





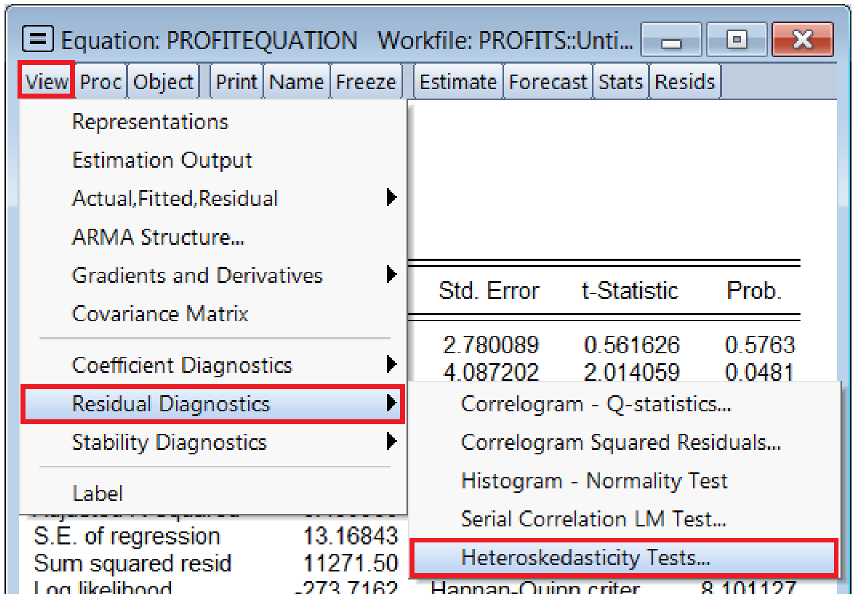
Since $W_{calc} = 25.6542 > W_{crit} = 11.0705$, we reject H_0 at the 5% significance level and conclude that there is sufficient evidence from our sample to suggest that the errors are heteroskedastic.

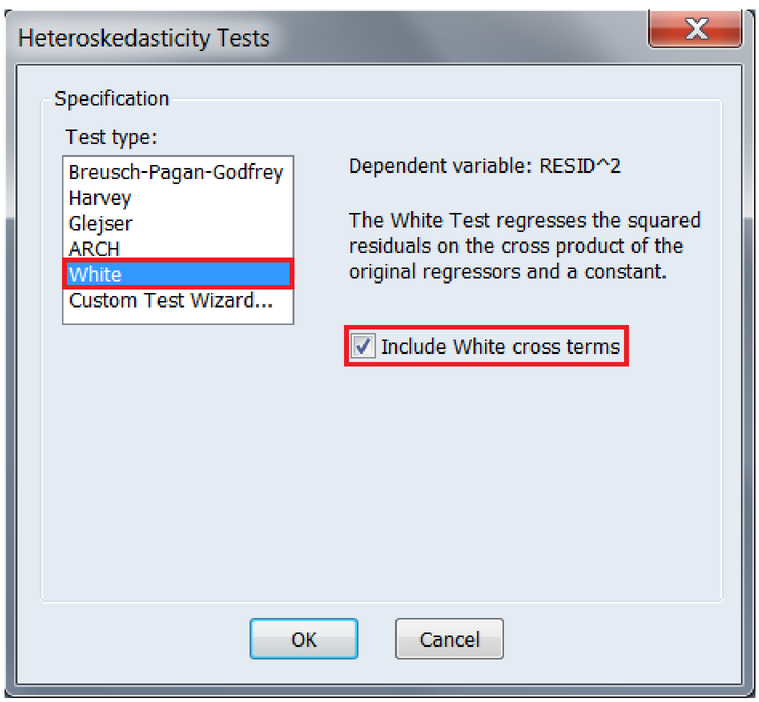
Using the in-built White test in EViews:

$$After\ you\ estimate\ your\ model\ with\ OLS$$

$$From\ the\ equation\ object:$$

$$View \rightarrow Residual\ Diagnostics \rightarrow Heteroskedasticity\ Tests...$$





Test type : *White*

Heteroskedasticity Test: White

F-statistic	7.456027	Prob. F(5,63)	0.0000
Obs*R-squared	25.65143	Prob. Chi-Square(5)	0.0001
Scaled explained SS	194.6893	Prob. Chi-Square(5)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/16/18 Time: 09:08
Sample: 1 88
Included observations: 69
Collinear test regressors dropped from specification

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−470.1101	170.4323	−2.758340	0.0076
MNO^2	536.1701	263.3993	2.035579	0.0460
MNO*ASSETS	−3.291956	1.244107	−2.646040	0.0103
ASSETS^2	−0.001130	0.000452	−2.498314	0.0151
ASSETS*MNO*ASSETS	0.001122	0.000663	1.692514	0.0955
ASSETS	3.330239	0.803775	4.143251	0.0001

R-squared	0.371760	Mean dependent var	163.3550
Adjusted R-squared	0.321900	S.D. dependent var	680.5634
S.E. of regression	560.4225	Akaike info criterion	15.57820
Sum squared resid	19786623	Schwarz criterion	15.77247
Log likelihood	−531.4479	Hannan-Quinn criter.	15.65527
F-statistic	7.456027	Durbin-Watson stat	2.531454
Prob(F-statistic)	0.000015		

Special form of the White test that uses the predicted value of $profits$ and its square as predictors of variance.

$$H_0 : Var(u_i|mno_i, assets_i) = \sigma^2$$
$$H_1 : Var(u_i|mno_i, assets_i) \neq \sigma^2$$

The alternative hypothesis is that the error variance is a smooth unknown function of mno_i and $assets_i$.

The functional form of the error variance for this special form of White test:

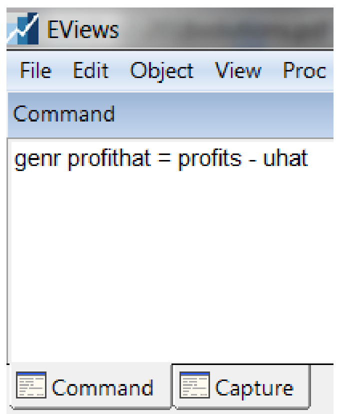
$$Var(u_i|mno_i, assets_i) = \alpha_0 + \alpha_1\widehat{profit}_i + \alpha_2\widehat{profit}_i^2$$

The auxiliary regression:

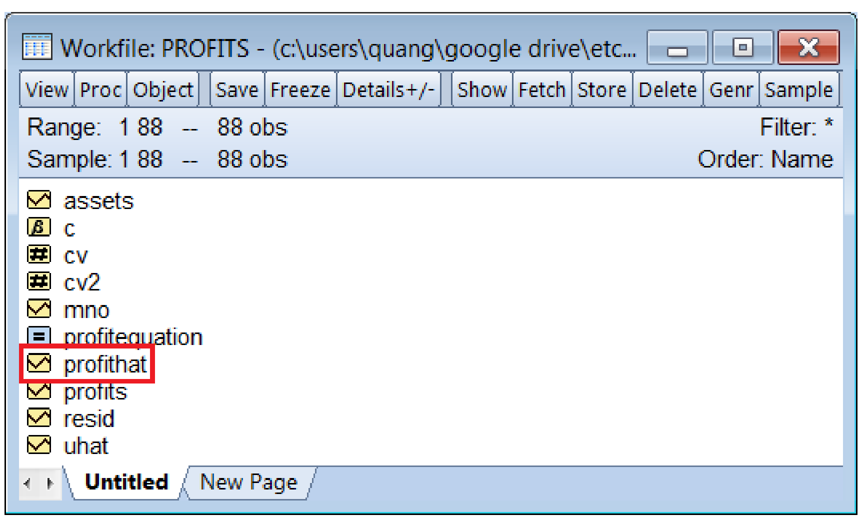
$$\hat{u}_i^2 = \alpha_0 + \alpha_1\widehat{profit}_i + \alpha_2\widehat{profit}_i^2 + v_i$$

Before estimating the auxiliary regression, we should save $\widehat{profit}_i$ as a variable in our EViews workfile. From the Command Window:

$$genr\ profithat = profits - uhat$$

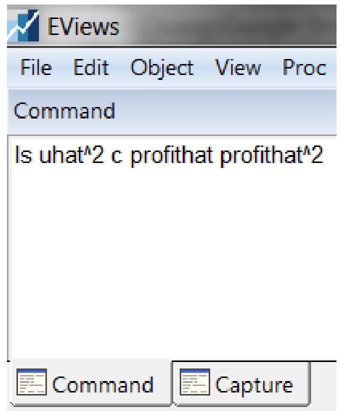


This generates a new object called $profithat$ which is equal to $profits$ minus the OLS residual $uhat$. Note, $uhat$ is the OLS residual (from the estimated model of $profits$) that we saved earlier.



To estimate the auxiliary regression from the Command Window in EViews:

$$ls\ uhat^2\ c\ profitshat\ profitshat^2$$



Dependent Variable: UHAT^2
Method: Least Squares
Date: 09/16/18 Time: 17:54
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	−529.7017	161.4841	−3.280210	0.0017
PROFITHAT	65.44824	15.73677	4.158937	0.0001
PROFITHAT^2	−0.427865	0.171206	−2.499129	0.0149

R-squared	0.368618	Mean dependent var	163.3550
Adjusted R-squared	0.349485	S.D. dependent var	680.5634
S.E. of regression	548.9051	Akaike info criterion	15.49623
Sum squared resid	19885591	Schwarz criterion	15.59337
Log likelihood	−531.6200	Hannan-Quinn criter.	15.53477
F-statistic	19.26627	Durbin-Watson stat	2.541133
Prob(F-statistic)	0.000000		

The White test statistic:

$$W = n \times R^2_{aux} \sim \chi^2_{q=2} \quad under\ H_0$$

$$R^2_{aux} : R^2\ from\ the\ auxiliary\ regression$$

$$q : number\ of\ regressors\ in\ the\ auxiliary\ regression$$

The calculated White test statistic:

$$W_{calc} = 69 \times 0.3686 = 25.4334$$

The critical value:

$$W_{crit} = \chi^2_{crit} = \chi^2_{2,0.95} = 5.99$$

Since $W_{calc} = 25.4334 > W_{crit} = 5.99$, we reject H_0 at the 5% significance level and conclude that there is sufficient evidence from our sample to suggest that the errors are heteroskedastic.

(c) Is it likely that a log-log formulation that uses $log(profits)$ and $log(assets)$ would solve the heteroskedasticity problem in this application? Explain.

Lec8 - page 21

If the population model has $log(y)$ as the dependent variable but we have used y, this kind of mis-specification can show up as heteroskedastic errors. So, if log-transformation is admissible (i.e. if y is positive), moving to a log model may solve the problem, and the OLS estimator on the log-transformed model will then be BLUE and standard errors will be useful. Of course, when we consider transforming y, we should think if a log-level or a log-log model makes better sense.

Since $profits$ can be negative, log transformation is not an option.

(d) In each of the following scenarios, determine the appropriate weight that solves the problem of heteroskedasticity when it multiplies on both sides of the equation of $profits$:

$$Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$$
$$Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i^2$$
$$Var(u_i|mno_i, assets_i) = \sigma^2 \times log(assets_i)$$

Background

Weighted Least Squares Estimator

It is helpful to consider the WLS estimator as a 2-step estimator:

- At step 1, apply some weighting/transformation to the original model to obtain the weighted model.
- At step 2, estimate the weighted model by OLS.

If the variance of the error has the following known functional form,

$$Var(u_i|x_{i1}, x_{i2}, ...) = \sigma^2 \times h_i$$

then weighing the original model by,

$$w_i = \frac{1}{\sqrt{h_i}}$$

produces the following weighted model,

$$w_i y_i = \beta_0 w_i + \beta_1 w_i x_{i1} + \beta_2 w_i x_{i2} + \cdots + w_i u_i$$

with a constant error variance (homoskedastic error),

$$\begin{aligned} Var(w_i u_i|x_{i1}, x_{i2}, \dots) &= w_i^2 Var(u_i|x_{i1}, x_{i2}, \dots) \\ &= \frac{1}{h_i} Var(u_i|x_{i1}, x_{i2}, \dots) \\ &= \frac{1}{h_i}\sigma^2 h_i \\ &= \sigma^2 \end{aligned}$$

The weight, w_i, is known and not treated a random variable.

The Weighted Least Squares (WLS) estimator is the OLS estimator used to estimate the weighted model of y (weighted so that the error has constant variance).

If the variance of the error takes the following form,

$$Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$$

then the weight that we need to apply to our original model to obtain a weighted model with constant error variance (homoskedastic error) is given by,

$$w_i = \frac{1}{\sqrt{assets_i}}$$

Multiplying w_i on both sides of the original model of $profits_i$ gives the following weighted model of $profits_i$,

$$\frac{profits_i}{\sqrt{assets_i}} = \frac{\beta_0}{\sqrt{assets_i}} + \frac{\delta_0 mno_i}{\sqrt{assets_i}} + \frac{\beta_1 assets_i}{\sqrt{assets_i}} + \frac{\delta_1 mno_i^* assets_i}{\sqrt{assets_i}} + \frac{u_i}{\sqrt{assets_i}}$$

and the error term in this weighted model has constant variance,

$$\begin{aligned} Var(\frac{u_i}{\sqrt{assets_i}}|mno_i, assets_i) &= \frac{1}{assets_i} Var(u_i|mno_i, assets_i) \\ &= \frac{1}{assets_i}\sigma^2 \times assets_i \\ &= \sigma^2 \end{aligned}$$

If the variance of the error takes the following form,

$$Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i^2$$

then the weight that we need to apply to our original model to obtain a weighted model with constant error variance (homoskedastic error) is given by,

$$w_i = \frac{1}{assets_i}$$

If the variance of the error takes the following form,

$$Var(u_i|mno_i, assets_i) = \sigma^2 \times log(assets_i)$$

then the weight that we need to apply to our original model to obtain a weighted model with constant error variance (homoskedastic error) is given by,

$$w_i = \frac{1}{\sqrt{log(assets_i)}}$$

(e) Suppose we know $Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$. Test the hypothesis that you formulated in part (a), i.e. that the nature of the ownership of a firm does not effect the relationship between its profits and its assets against the alternative that it does.

If $Var(u_i|mno_i, assets_i) = \sigma^2 \times assets_i$ then our weighted model of $profits$ is given by:

$$\frac{profits_i}{\sqrt{assets_i}} = \frac{\beta_0}{\sqrt{assets_i}} + \frac{\delta_0 mno_i}{\sqrt{assets_i}} + \frac{\beta_1 assets_i}{\sqrt{assets_i}} + \frac{\delta_1 mno_i^* assets_i}{\sqrt{assets_i}} + \frac{u_i}{\sqrt{assets_i}}$$

We wish to test the null hypothesis the nature of ownership of a firm does not affect the relationship between its profits and assets,

$$H_0 : \delta_0 = \delta_1 = 0$$

against the alternative hypothesis that it does,

$$H_1 : \delta_0 \neq 0 \ and/or \ \delta_1 \neq 0$$

Since we are testing multiple linear restrictions (2 restrictions), we require a F test.

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(69-3-1)} \sim F_{2,65} \quad under \ H_0$$

$$\begin{aligned} n &= sample\ size = 69 \\ k &= number\ of\ regressors\ in\ the\ unrestricted\ model = 3 \\ q &= number\ of\ restrictions\ \ = 2 \\ SSR_r &= sum\ of\ squared\ residuals\ from\ estimated\ restricted\ model \\ SSR_{ur} &= sum\ of\ squared\ residuals\ from\ estimated\ unrestricted\ model \end{aligned}$$

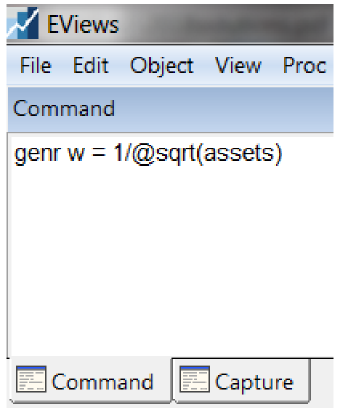
The unrestricted and restricted model:

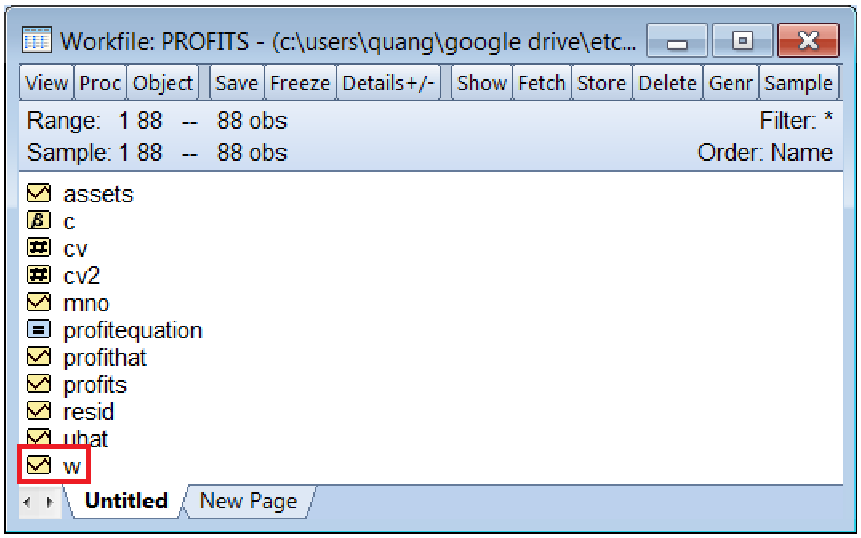
$$\frac{profits_i}{\sqrt{assets_i}} = \frac{\beta_0}{\sqrt{assets_i}} + \frac{\delta_0 mno_i}{\sqrt{assets_i}} + \frac{\beta_1 assets_i}{\sqrt{assets_i}} + \frac{\delta_1 mno_i^* assets_i}{\sqrt{assets_i}} + \frac{u_i}{\sqrt{assets_i}}$$

$$\frac{profits_i}{\sqrt{assets_i}} = \frac{\beta_0}{\sqrt{assets_i}} + \frac{\beta_1 assets_i}{\sqrt{assets_i}} + \frac{u_i}{\sqrt{assets_i}}$$

Before estimating the unrestricted and restricted, let us generate the variable $w = \frac{1}{\sqrt{assets}}$ in EViews. From the Command Window:

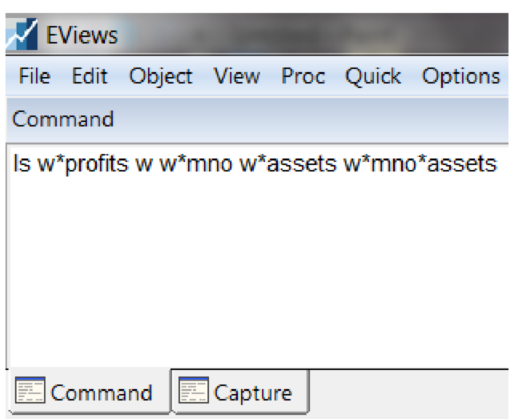
$$genr\ w = 1/@sqrt(assets)$$





To estimate the weighted unrestricted model from the Command Window,

$$ls\ w\ w^{*}mno\ w^{*}assets\ w^{*}mno^{*}assets$$



Dependent Variable: W*PROFITS
Method: Least Squares
Date: 09/17/18 Time: 19:47
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
W	0.123099	1.937385	0.063539	0.9495
W*MNO	2.500248	2.657450	0.940845	0.3503
W*ASSETS	0.055756	0.009207	6.056060	0.0000
W*MNO*ASSETS	−0.030944	0.013195	−2.345092	0.0221

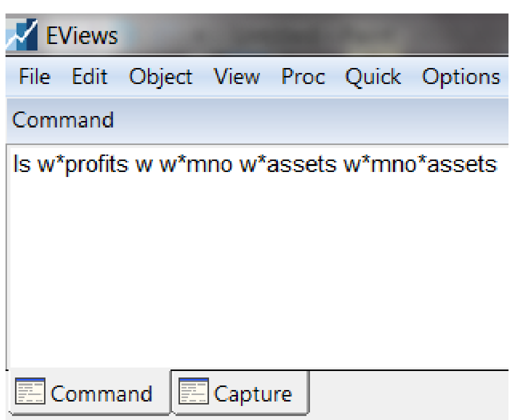
R-squared	0.175888	Mean dependent var	0.748290
Adjusted R-squared	0.137852	S.D. dependent var	0.672177
S.E. of regression	0.624129	Akaike info criterion	1.951304
Sum squared resid	25.31993	Schwarz criterion	2.080818
Log likelihood	−63.32000	Hannan-Quinn criter.	2.002687
Durbin-Watson stat	2.179079		

$$\widehat{w^{*}profits}_i = \underset{(1.9374)}{0.1231}w + \underset{(2.6575)}{2.5002}w^{*}mno + \underset{(0.0092)}{0.0448}w^{*}assets - \underset{(0.0131)}{0.0309}w^{*}mno^{*}assets$$

$$SSR_{ur} = 25.3199$$

To estimate the weighted model from the Command Window,

$$ls\ w\ w^{*}assets$$



Dependent Variable: W*PROFITS
Method: Least Squares
Date: 09/17/18 Time: 19:56
Sample: 1 88
Included observations: 69

Variable	Coefficient	Std. Error	t-Statistic	Prob.
W	1.395225	1.371934	1.016977	0.3128
W*ASSETS	0.041256	0.006804	6.063294	0.0000

R-squared	0.090491	Mean dependent var	0.748290
Adjusted R-squared	0.076916	S.D. dependent var	0.672177
S.E. of regression	0.645809	Akaike info criterion	1.991932
Sum squared resid	27.94365	Schwarz criterion	2.056688
Log likelihood	−66.72164	Hannan-Quinn criter.	2.017623
Durbin-Watson stat	2.223289		

$$\widehat{w^*profits}_i = \underset{(1.3719)}{1.3952}w + \underset{(0.0068)}{0.0413}w^*assets$$
$$SSR_r = 27.9437$$

$$F_{calc} = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(65)} = \frac{(27.94 - 25.32)/2}{25.32/(65)} = 3.36$$
$$F_{crit} = 3.15 \quad (5\%\ level.\ Stat\ Table)$$

Since $F_{calc} = 3.36 > F_{crit} = 3.15$, we reject the null at the 5% significance level and conclude that there is some difference in the relationship between profits and assets across owner managed and non-owner managed firms.

Note: This process of applying a weighted transformation to the model is a device to converting a model with heteroskedastic errors into one with homoskedasticity error. It is NOT something that changes the inherent meaning of the coefficients. As such, we still interpret the 'weighted' coefficients the same way as we would with the original model.