Tutorial 10

keywords: grouped data, variance, error, heteroskedasticity, homoskedasticity, residual plots, Breusch-Pagan test, inference, efficiency, Weighted Least Squares estimator

estimated reading time: 32 minutes

Quang Bui

May 7, 2018

Question 1

EViews workfile: *Tobin.wf1*

Tobin.wf1 contains data from the US 1941 Family Budget Survey data set. We have information about 37 family groups. Each group is based on number of people in the family (1,2,3,4,5+) and their income range (0-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-4000, 4000+). Information about each family group is held in the following variables:

$$\begin{aligned} food &- average\ expenditure\ on\ food\ for\ a\ particular\ group\ (\$) \\ inc &- average\ income\ for\ a\ particular\ group\ (\$) \\ nf &- average\ number\ of\ people\ in\ family\ for\ a\ particular\ group \\ ng &- number\ of\ families\ in\ each\ group \end{aligned}$$

	FOOD	INC	NF	NG
1	210	421	1	59
2	301	824	1	71
3	369	1287	1	40
4	433	1703	1	18
5	506	2150	1	11
6	621	2655	1	6
7	239	520	2	29
8	319	869	2	54
9	454	1379	2	67
10	517	1846	2	58

Table 1: Data on *average food expenditure*, *average income*, *average family size* and *number of families in the group* for the first 10 family groups in our data set of 37 family groups.

- Family group number 1 contains 59 families ($ng = 59$), spends \$210 on food on average ($food = 210$), has an average income of \$421 ($inc = 421$) and an average family size of 1 member ($nf = 1$). This family group represents familes with 1 family member with an income range of \$0-500.
- Family group number 2 contains 71 families ($ng = 71$), spends \$301 on food on average ($food = 301$), has an average income of \$824 ($inc = 824$) and an average family size of 1 member ($nf = 1$). This family group represents familes with 1 family member with an income range of \$500-1000.
- etc.

(a) Suppose that food consumption is related to family income and number of people in each family according to the following model.

$$food_i = \beta_0 + \beta_1 inc_i + \beta_2 nf_i + u_i$$

Argue that even if the above model is a fair description of each family's demand for food and all Gauss-Markov assumptions are satisfied for this model, we may expect heteroskedastic errors when we estimate the model using **group averaged data**.

If we estimate the model with data on **individual families** such that,

$$\begin{aligned} food_i &- i^{th}\ family's\ food\ expenditure \\ inc_i &- i^{th}\ family's\ income \\ nf_i &- i^{th}\ family's\ number\ of\ family\ members \end{aligned}$$

then we will assume that the errors are homoskedastic (although this assumption is not reasonable, we make this assumption to validate today's exercise). If instead, we estimate the model using **group averaged data** then it would be very unreasonable to assume homoskedastic errors. Why?

Background

Homoskedasticity

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.*

Homoskedasticity errors - the variance of the error is constant i.e. the variance is fixed across all observations (if the error does not vary differently across different observations then it is constant regardless of the value of x).

$$Var(u_i|inc_i, nf_i) = \sigma^2$$

Heteroskedasticity errors - the variance of the error is not constant.

$$Var(u_i|inc_i, nf_i) \neq \sigma^2$$

Since β_0, β_1 and β_2 are constants and inc_i and nf_i are known (we have conditioned on inc_i and nf_i),

$$\begin{aligned} Var(food_i|inc_i, nf_i) &= Var(\beta_0 + \beta_1 inc_i + \beta_2 nf_i + u_i|inc_i, nf_i) \\ &= Var(u_i|inc_i, nf_i) \end{aligned}$$

We expect the variability of a family group's average food expenditure,

$$Var(food_i|inc_i, nf_i)$$

to depend on the number of families in the group,

$$Var(food_i|inc_i, nf_i) = f(ng_i)$$

i.e. we expect average food expenditure to vary less across family groups with many families than family groups with few families,

$$Var(food_i|inc_i, nf_i) = \frac{\sigma^2}{ng_i} \neq \sigma^2$$

an inverse relationship.

Since,

$$Var(food_i|inc_i, nf_i) = Var(u_i|inc_i, nf_i)$$

when the variance of the dependent variable is not constant, the variance of the error will not be constant i.e. the error is heteroskedastic.

Sampling variability of sample mean

Let X be a random variable with variance equal to σ^2,

$$Var(X) = \sigma^2$$

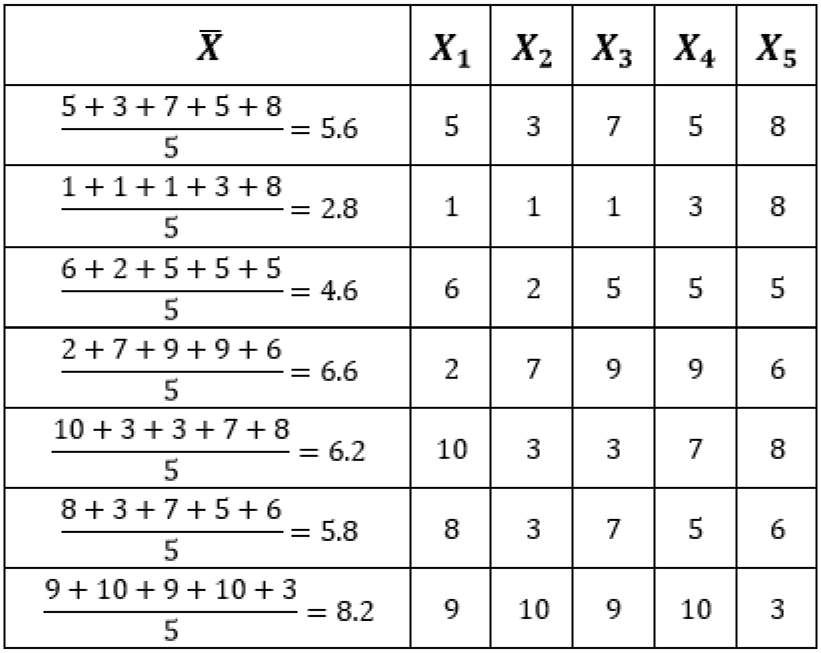
For n random variables $X_1, X_2, \ldots, X_n$, that are identically distributed to X and independent of each other, a draw from each random variable $X_1, X_2, \dots, X_n$ gives n independent observations i.e. a random sample of size n. Since each random variable is identical in distribution, the variance of each random variable equal to σ^2,

$$Var(X_1) = Var(X_2) = \cdots = Var(X_n) = \sigma^2$$

Since,

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

if follows that $\bar{X}$ is a random variable that depends on X_1, X_2 ... X_n and sample size n and is *subject to sampling variability.* For example, using a sample size of 5 to apply seven repeated sampling, we obtain the following seven $\bar{X}$,



$\bar{X}$	X_1	X_2	X_3	X_4	X_5
$\frac{5+3+7+5+8}{5} = 5.6$	5	3	7	5	8
$\frac{1+1+1+3+8}{5} = 2.8$	1	1	1	3	8
$\frac{6+2+5+5+5}{5} = 4.6$	6	2	5	5	5
$\frac{2+7+9+9+6}{5} = 6.6$	2	7	9	9	6
$\frac{10+3+3+7+8}{5} = 6.2$	10	3	3	7	8
$\frac{8+3+7+5+6}{5} = 5.8$	8	3	7	5	6
$\frac{9+10+9+10+3}{5} = 8.2$	9	10	9	10	3

(*can you guess the distribution of* X?)

(*what about the distribution of* $\bar{X}$?)

The sampling variability of $\bar{X}$ is given by,

$$\begin{aligned} Var(\bar{X}) &= Var\Big(\frac{X_1 + X_2 + \cdots + X_n}{n}\Big) \\ &= \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) \\ &= \frac{1}{n^2}\Big(Var(X_1) + Var(X_2) + \cdots + Var(X_n)\Big) + 0 + 0 + \cdots + 0 \end{aligned}$$

$$\begin{aligned}
&(Covariance\ between\ independent\ random\ variables\ equals\ to\ 0)\\
&= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots + \sigma^2)\\
&= \frac{1}{n^2}(n \times \sigma^2)\\
&= \frac{\sigma^2}{n}
\end{aligned}$$

which tells us that the sampling variability of $\bar{X}$ equals to the variance of X divided by the sample size. The greater the sample size used to obtain $\bar{X}$, the less sampling variability $\bar{X}$ will have.

Show that it is logical to expect that variance of errors are inversely proportional to the number of familes in each group i.e. $Var(u_i|inc_i, nf_i) = \dfrac{\sigma^2}{ng_i}$.

$$(See\ above)$$

Assume that, if the model were estimated using **individual family data**, the errors are homoskedastic so that the variance of family food expenditure equals to σ^2,

$$Var(food_i^*|inc_i^*, nf_i^*) = Var(u_i^*|inc_i^*, nf_i^*) = \sigma^2$$

Since $food_i$ represents the average food expenditure for the i^{th} family group,

$$food_i = \overline{food_i^*}$$

the variance of the average food expenditure for the i^{th} family group is based on the number of families in the i^{th} group i.e. ng_i families. More specifically, the variance of the i^{th} group's average food expenditure equals to the variance of family food expenditure divided by ng_i,

$$\begin{aligned}
Var(food_i|inc_i, nf_i) &= \frac{\sigma^2}{ng_i}\\
&= Var(u_i|inc_i, nf_i)
\end{aligned}$$

Since the number of families in each group is not constant i.e. ng_i is not constant, the error is heteroskedastic i.e. the variance of the error when the model is estimated with **group average data** depends on the number of families in each group.

(b) If errors are heteroskedastic, would OLS be unbiased? Explain.

The condition for homoskedastic errors is not required for the OLS estimator to be unbiased, therefore, as long as the assumptions for unbiasedness holds, the OLS estimator will still be unbiased even when errors are heteroskedastic.

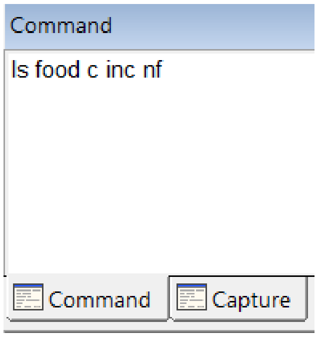
(c) Estimate the model using OLS. Then explore visually if the error variance is likely to be inversely proportional to the number of families in each group (ng), and then formally test the hypothesis of no heteroskedasticity against the alternative that,

$$Var(u_i|inc_i, nf_i) = \frac{\sigma^2}{ng_i}$$

Hint: Breusch-Pagan auxiliary regression is a regression of $\hat{u}^2$ on a constant and $\frac{1}{ng}$.

From the Command window:

$$ls\ food\ c\ inc\ nf$$



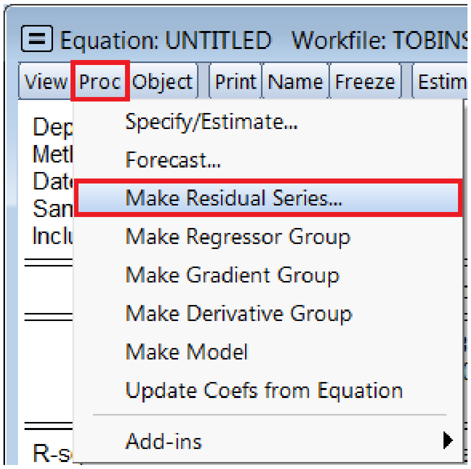
```
ls food c inc nf
```

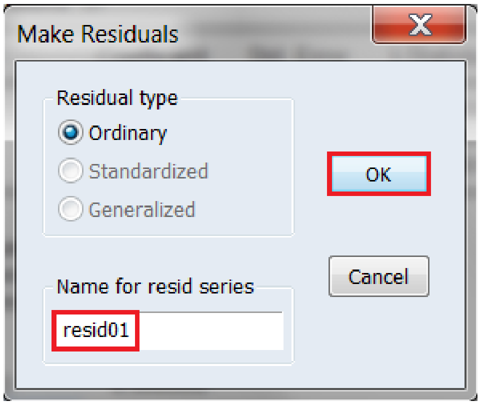
```
Dependent Variable: FOOD
Method: Least Squares
Sample: 1 37
Included observations: 37

        Variable          Coefficient   Std. Error   t-Statistic      Prob.

           C                 73.14384     48.74665      1.500490     0.1427
          INC                0.164541     0.012276      13.40395     0.0000
           NF                59.62959     11.42616      5.218691     0.0000

R-squared                    0.872122   Mean dependent var         662.0270
Adjusted R-squared           0.864600   S.D. dependent var         330.0085
S.E. of regression           121.4324   Akaike info criterion      12.51420
Sum squared resid            501358.4   Schwarz criterion          12.64481
Log likelihood              −228.5127   Hannan-Quinn criter.       12.56025
F-statistic                  115.9393   Durbin-Watson stat         1.749737
Prob(F-statistic)            0.000000
```

Table 2: Regression output of *average food expenditure of a family group* on a constant, *average income of a family group* and *number of families of a family group*.

Save the estimated model's OLS residuals into a separate series,

$$Proc \rightarrow Make\ to\ Residual\ Series \rightarrow Name\ for\ resid : resid01 \rightarrow OK$$



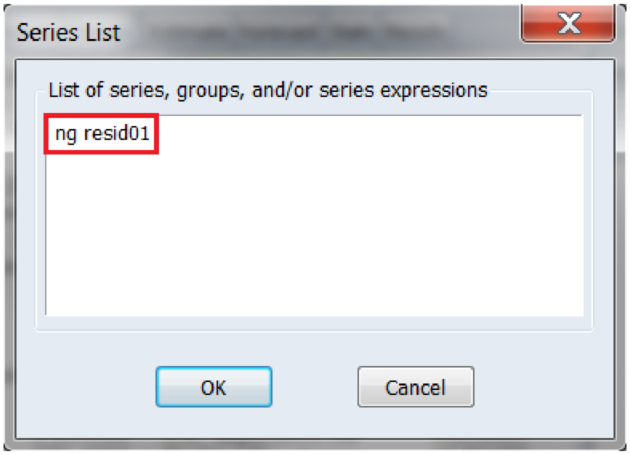

Make Residuals

Residual type
● Ordinary
○ Standardized
○ Generalized

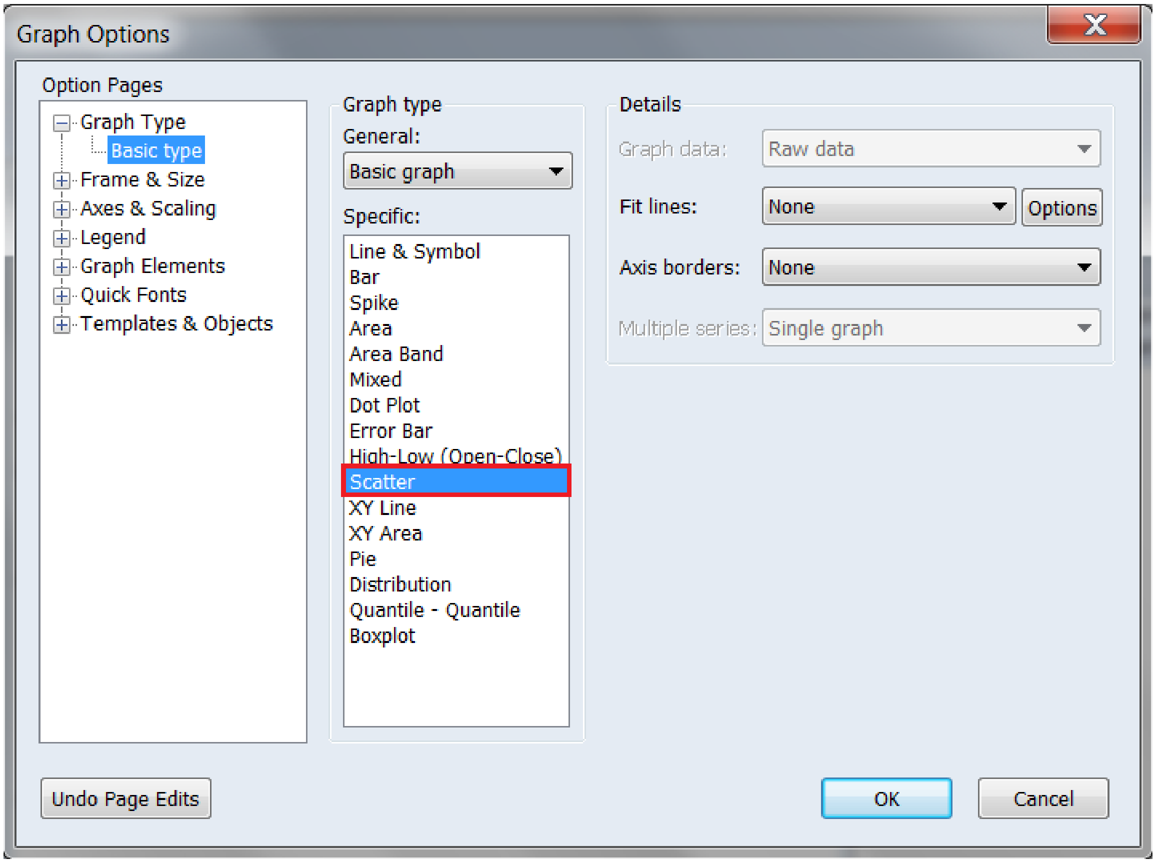
Name for resid series: resid01

To plot the OLS residuals against ng (residual plot),

$$Quick \rightarrow Graph \rightarrow ng\ resid01 \rightarrow Specific : Scatter \rightarrow OK$$


Series List

List of series, groups, and/or series expressions: ng resid01



General: Basic graph

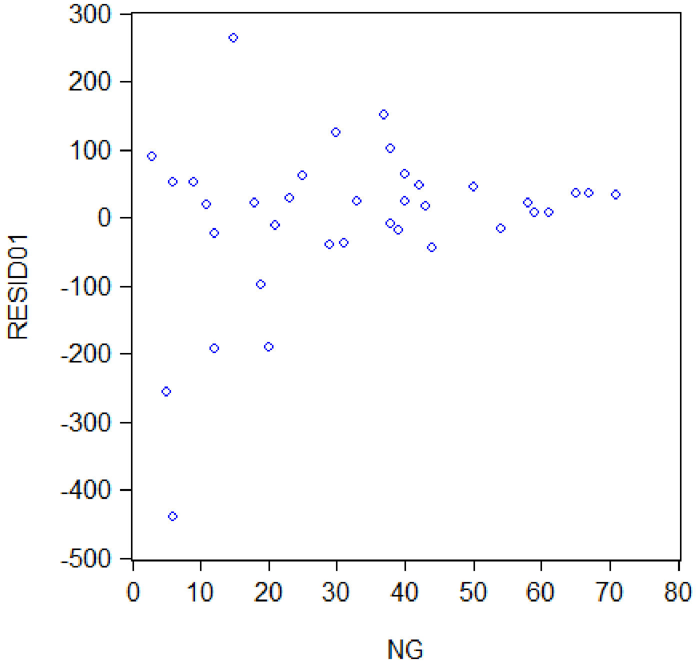
Specific: Scatter

Graph data: Raw data

Fit lines: None

Axis borders: None

Multiple series: Single graph



Background

Breusch-Pagan test for heteroskedasticity

Since $E(u_i|inc_i, nf_i) = 0$,

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

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

By considering a model of the squared error term from our model of average food expenditure $u_i = food_i - (\beta_0 + \beta_1 inc_i + \beta_2 nf_i)$,

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

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 $Var(food_i|inc_i, nf_i)$ i.e. $Var(u_i|inc_i, nf_i)$ (so the z variables do not have to be inc and nf). That is,

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

Ideally, we would want to run a regression on,

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

and test for heteroskedasticity by testing if at least one of the z variables has any explanatory power in explaining the variance of u (and therefore the variance of $food$),

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

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,

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

For the general case of q, z variables, so we would formulate the null and alternative hypothesis as follows,

$$Var(u_i|inc_i, nf_i) = E(u_i^2|inc_i, nf_i) = \delta_0 + \delta_1 z_{i1} + \delta_2 z_{i2} + \cdots + \delta_q z_{iq}$$

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

(σ^2 and δ are constants)

Using our intuition and visual aid, we suspect that ng_i has an inverse relationship with $Var(u_i|inc_i, nf_i)$, therefore, we let $\dfrac{1}{ng_i}$ be a variable in our **auxiliary regression**,

$$\hat{u}_i^2 = \delta_0 + \delta_1 \frac{1}{ng_i} + v_i$$

(If we suspect that there are other variables that affect the variance of the error and we have data on these variables, we should include them in the auxiliary regression when testing for heteroskedasticity in the error using Breusch-Pagan test.)

$H_0 : E(u_i^2|inc_i, nf_i) = \sigma^2 \qquad (\delta_1 = 0 \quad homoskedastic\ errors)$

$H_1 : E(u_i^2|inc_i, nf_i) \neq \sigma^2 \qquad (\delta_1 \neq 0 \quad heteroskedastic\ errors)$

The Breusch-Pagan test statistic,

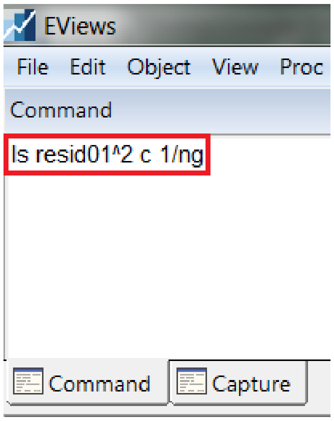
$$LM = N \times R_{aux}^2 \sim \chi_q^2$$

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

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

where q represents the number of variables which we believe impact the variance of the error and are include as regressors in the auxiliary regression of the Breusch-Pagan test.

To estimate the auxiliary regression in EViews,

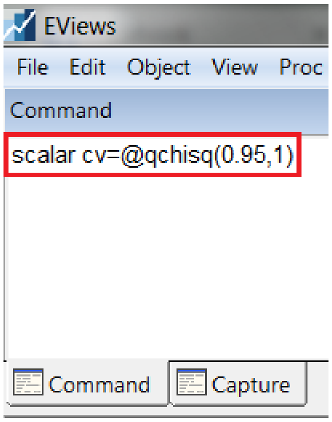


```
ls resid01^2 c 1/ng
```

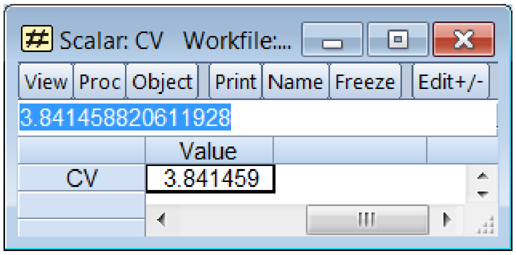
```
Dependent Variable: RESID01^2
Method: Least Squares
Date: 05/06/18 Time: 19:16
Sample: 1 37
Included observations: 37

        Variable          Coefficient   Std. Error   t-Statistic     Prob.

            C              1372.979     7123.362     0.192743      0.8483
          1/NG             212500.3     82855.09     2.564722      0.0148

R-squared                  0.158205   Mean dependent var       13550.23
Adjusted R-squared         0.134153   S.D. dependent var       34713.39
S.E. of regression         32301.11   Akaike info criterion    23.65613
Sum squared resid         3.65E+10    Schwarz criterion        23.74321
Log likelihood            -435.6384   Hannan-Quinn criter.     23.68683
F-statistic                6.577799   Durbin-Watson stat       2.379673
Prob(F-statistic)          0.014773
```

$$LM_{calc} = 37 \times 0.158205 = 5.8534$$

To obtain the critical value in EViews,



```
scalar cv=@qchisq(0.95,1)
```



3.841458820611928

	Value
CV	3.841459

Since $LM_{calc} = 5.8534 > \chi^2_{crit} = 3.8415$, 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.

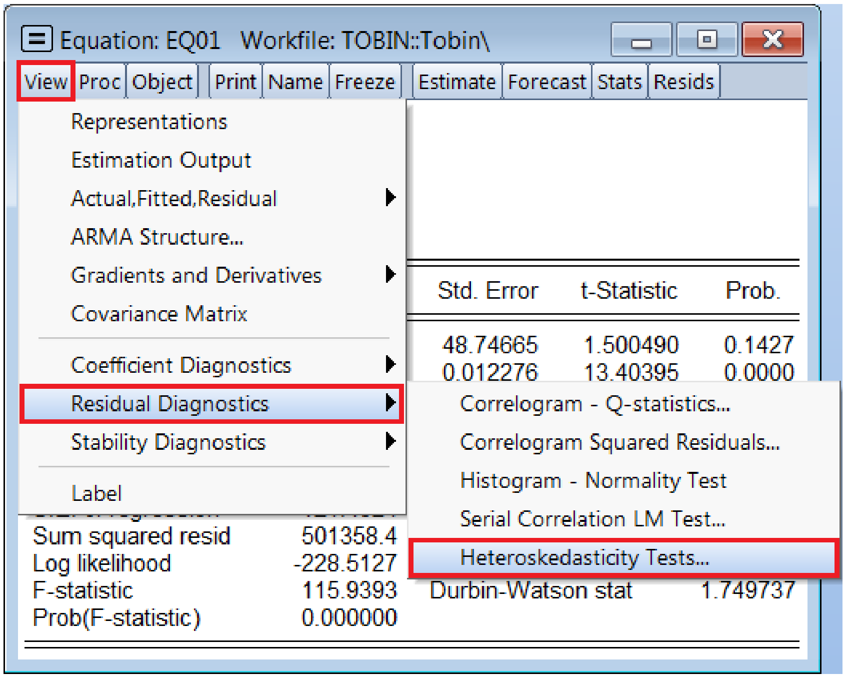
Note: This test is only valid for larger samples, so our conclusion should be taken with consideration to the sample size and used in conjunction with theoretical reasoning and our visual findings.

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

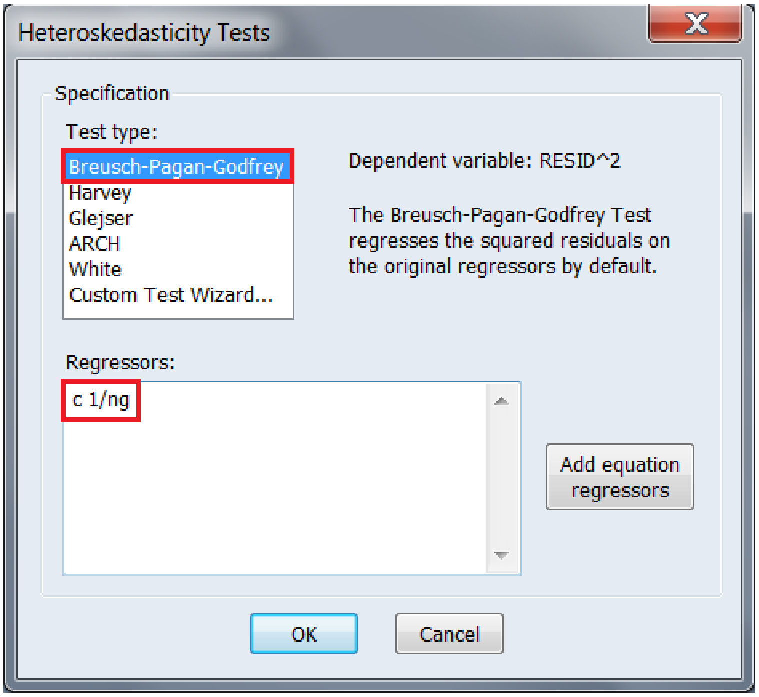
After you estimate your model with OLS

From the equation object :

View → Residual Diagnostics → Heteroskedasticity Tests...



```
Std. Error    t-Statistic    Prob.

48.74665      1.500490       0.1427
0.012276      13.40395       0.0000

Sum squared resid     501358.4
Log likelihood       -228.5127
F-statistic           115.9393    Durbin-Watson stat    1.749737
Prob(F-statistic)     0.000000
```



Dependent variable: RESID^2

The Breusch-Pagan-Godfrey Test regresses the squared residuals on the original regressors by default.

Regressors: c 1/ng

Test type : $Breusch - Pagan - Godfrey$

Regressors : $c \quad 1/ng$

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	6.577799	Prob. F(1,35)	0.0148
Obs*R-squared	5.853570	Prob. Chi-Square(1)	0.0155
Scaled explained SS	15.78144	Prob. Chi-Square(1)	0.0001

Test Equation:
Dependent Variable: RESIDˆ2
Method: Least Squares
Date: 05/08/18 Time: 18:09
Sample: 1 37
Included observations: 37

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1372.979	7123.362	0.192743	0.8483
1/NG	212500.3	82855.09	2.564722	0.0148

R-squared	0.158205	Mean dependent var	13550.23
Adjusted R-squared	0.134153	S.D. dependent var	34713.39
S.E. of regression	32301.11	Akaike info criterion	23.65613
Sum squared resid	$3.65E+10$	Schwarz criterion	23.74321
Log likelihood	-435.6384	Hannan-Quinn criter.	23.68683
F-statistic	6.577799	Durbin-Watson stat	2.379673
Prob(F-statistic)	0.014773		

$$LM_{calc} = 5.8536 \qquad p - value = 0.0155$$

(d) Suppose we reject homoskedasticity in favour of,

$$Var(u_i|inc_i, nf_i) = \frac{\sigma^2}{ng_i}$$

Describe how we should transform the variables and which regression we should run to obtain the best linear unbiased estimator for β_0, β_1 and β_2.

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,

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

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).

Since the variance of the error takes the following form,

$$Var(u_i|inc_i, nf_i) = \frac{\sigma^2}{ng_i}$$

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

$$w_i = \frac{1}{\sqrt{1/ng_i}} = \sqrt{ng_i}$$

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

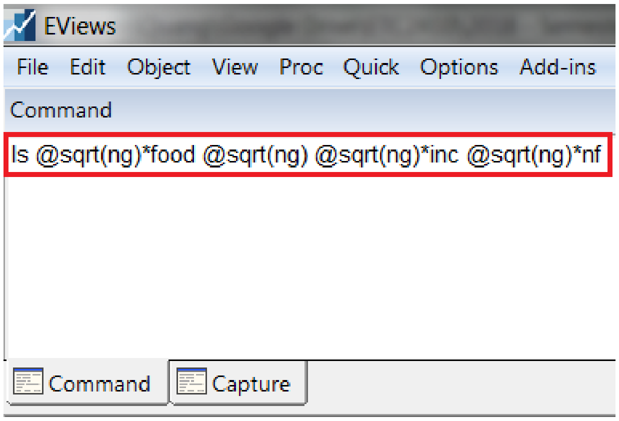
$$\sqrt{ng_i} food_i = \beta_0 \sqrt{ng_i} + \beta_1 \sqrt{ng_i} inc_i + \beta_2 \sqrt{ng_i} nf_i + \sqrt{ng_i} u_i$$

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

$$\begin{aligned} Var(\sqrt{ng_i} u_i|inc_i, nf_i) &= (\sqrt{ng_i})^2 Var(u_i|x_{i1}, x_{i2}, \dots) \\ &= ng_i Var(u_i|x_{i1}, x_{i2}, \dots) \\ &= ng_i \frac{\sigma^2}{ng_i} \\ &= \sigma^2 \end{aligned}$$

To estimate the weighted model from the **Command window**,

$$ls \quad @sqrt(ng)^* food \quad @sqrt(ng) \quad @sqrt(ng)^* inc \quad @sqrt(ng)^* nf$$



```
ls @sqrt(ng)*food @sqrt(ng) @sqrt(ng)*inc @sqrt(ng)*nf
```

Dependent Variable: @SQRT(NG)*FOOD
Method: Least Squares
Date: 05/06/18 Time: 19:45
Sample: 1 37
Included observations: 37

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SQRT(NG)	70.31890	28.51754	2.465813	0.0189
@SQRT(NG)*INC	0.177184	0.010268	17.25529	0.0000
@SQRT(NG)*NF	57.70058	8.056305	7.162164	0.0000

$$\widehat{food}_i = \underset{(28.5175)}{70.3189} + \underset{(0.0103)}{0.1772} inc_i + \underset{(8.0563)}{57.7006} nf_i$$

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

(e) Based on the weighted least squares estimates test the hypothesis that, other things staying the same, a 1 dollar increase in family income will result a 20 cent increase in family food consumption, against the alternative that it will result in a less than 20 cent increase in food consumption. Perform this test at the 5% level of significance.

Dependent Variable: @SQRT(NG)*FOOD
Method: Least Squares
Date: 05/06/18 Time: 19:45
Sample: 1 37
Included observations: 37

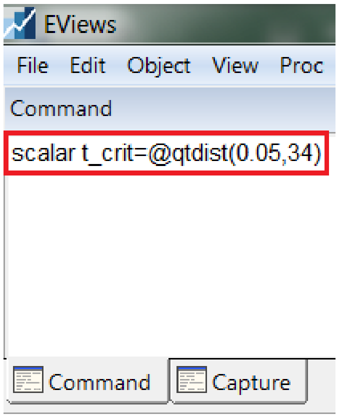
Variable	Coefficient	Std. Error	t-Statistic	Prob.
@SQRT(NG)	70.31890	28.51754	2.465813	0.0189
@SQRT(NG)*INC	0.177184	0.010268	17.25529	0.0000
@SQRT(NG)*NF	57.70058	8.056305	7.162164	0.0000

$$\widehat{food}_i = \underset{(28.5175)}{70.3189} + \underset{(0.0103)}{0.1772}inc_i + \underset{(8.0563)}{57.7006}nf_i$$

$$H_0 : \beta_1 = 0.2$$
$$H_1 : \beta_1 < 0.2$$

$$t = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - 0.2}{se(\hat{\beta}_1)} \sim t_{n-k-1} \; under \; H_0$$

$$t_{calc} = \frac{0.1772 - 0.2}{0.0103} = -2.2136$$



```
scalar t_crit=@qtdist(0.05,34)
```

$$-t_{crit} = t_{34,0.05} = -1.6909$$

Since $t_{calc} = -2.2136 > -t_{crit} = -1.6909$ we reject the null and conclude that there is sufficient evidence from our sample to suggest that a dollar increase in family income will result in a less than 20 cent increase in food consumption, holding nf constant.