ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

EViews Workshop I

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Agenda (1)

I. Introduction to EViews

- EViews Interface
- Types of Data based on EViews
  - *Unstructured/Undated Data*
  - *Dated –Regular frequency Data*
  - *Panel Balanced Data*

II. Empirical Example 1 : Unstructured/Undated Data

- Example of a test in Finance for 60 students
- Probability density function
- Basic Distribution parameters (mean, standard deviation, skewness, kurtosis)

Agenda (2)

III. Empirical Example 2 : Dated – Regular frequency Data

- Example of monthly returns for equity indices of G7 countries
- Distribution parameters (covariance, correlation)

IV. Empirical Example 3 : Transformations of data

- Example of simple and continuous compounding returns for ETE

V. Empirical Example 4: Correlation vs. Regression

- Example of simple returns of ETE and ATHEX

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Agenda (3)

VI. Classical Linear Regression Model Estimation

- **Empirical Example 5 :** CAPM model
  - ➢ Model Estimation
  - ➢ Hypothesis Testing
  - ➢ Wald Test
  - ➢ Multiple Hypothesis : the F -test

VII. Multiple Linear Regression Model Estimation

- **Empirical Example 6 :** APT Model
  - ➢ Model Estimation
  - ➢ Hypothesis Testing
  - ➢ Wald Test
  - ➢ Multiple Hypothesis : the F -test
  - ➢ Stepwise procedure equation estimation
  - ➢ R-squared & F -Statistic

Econometric Packages for modeling financial data

Econometric Package	Package Software Supplier
Eviews	QMS Software
GAUSS	Aptech Systems
LIMDEP	Econometric Software
MATLAB	The Mathworks
RATS	Estima
SAS	SAS Institute
SHAZAM	Northwest Econometrics
SPLUS	Insightful Corporation
SPSS	SPSS
TSP	TSP International

Source : Introductory Econometrics
for Finance, Chris Brooks

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

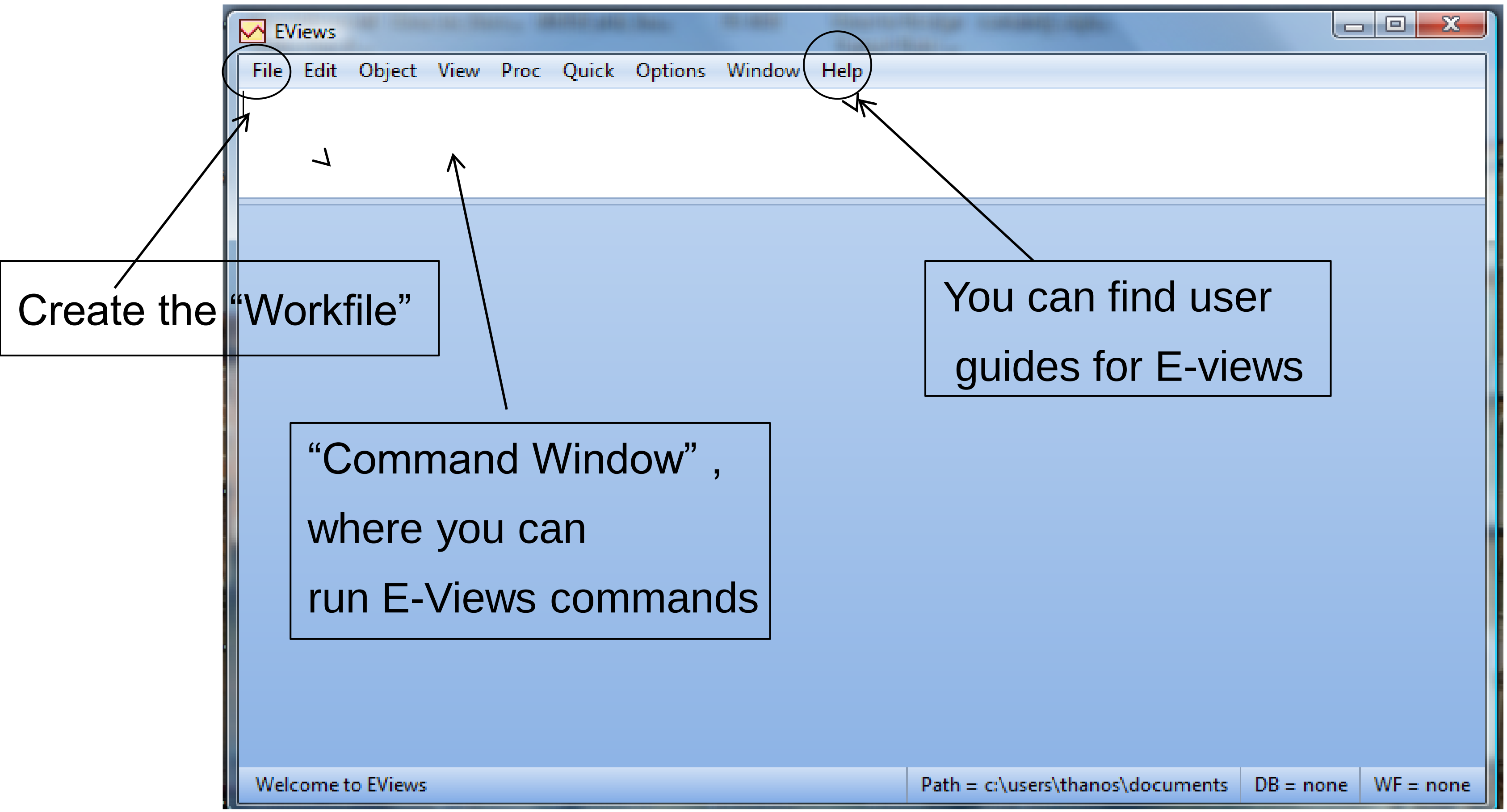
Introduction to EViews

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

E- views Interface

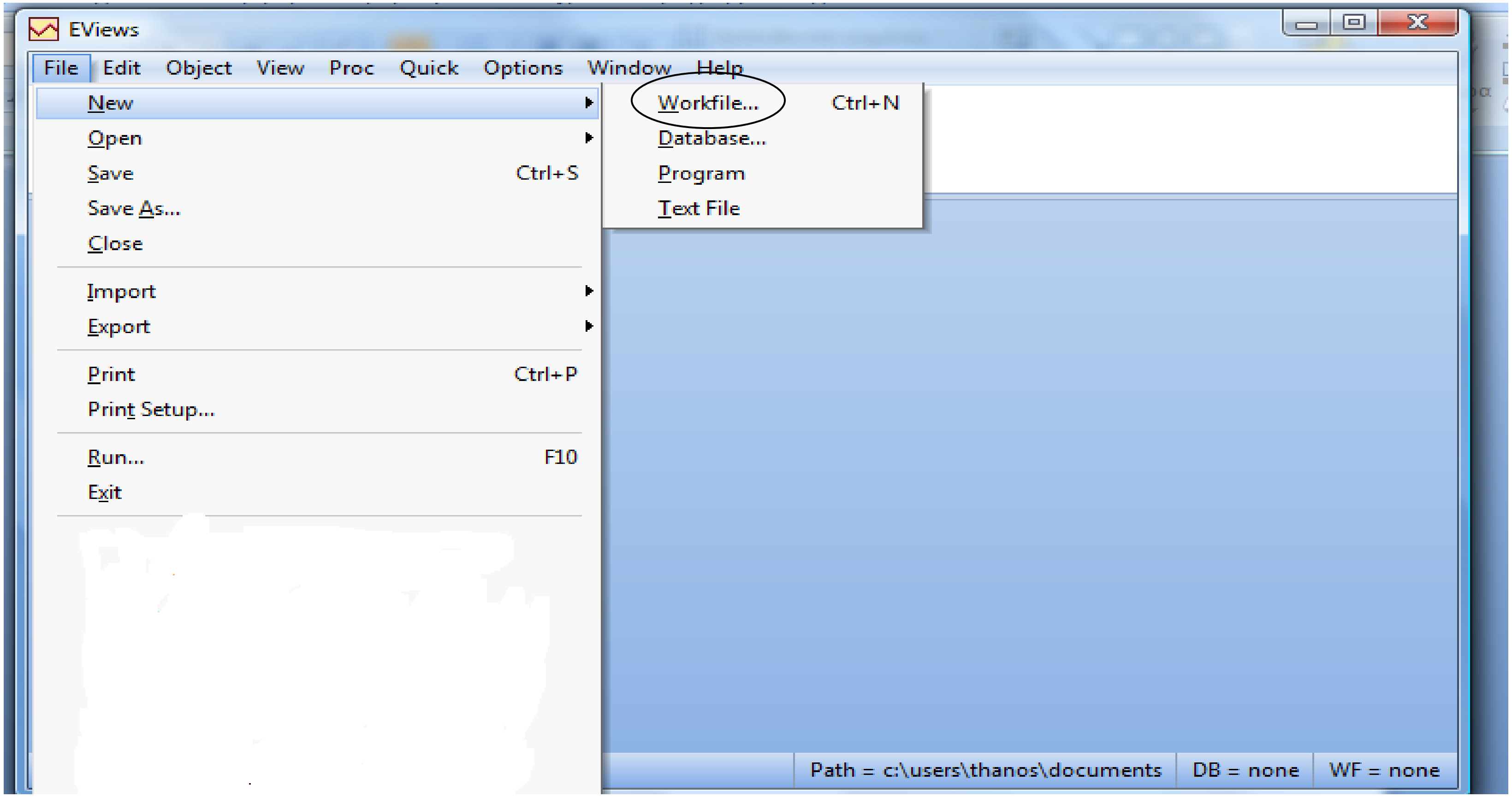
1.Open *EViews* from PC- lab

- Double Click on the EViews on the desktop of your pc



2. Go to File ⟶ New ⟶ **Workfile**

- **Workfile** constitutes the main workspace in EViews



Create a Workfile

3. *Workfile Create*

Types of Data

I. <u>Cross- sectional Data *(Unstructured/Undated)*</u>

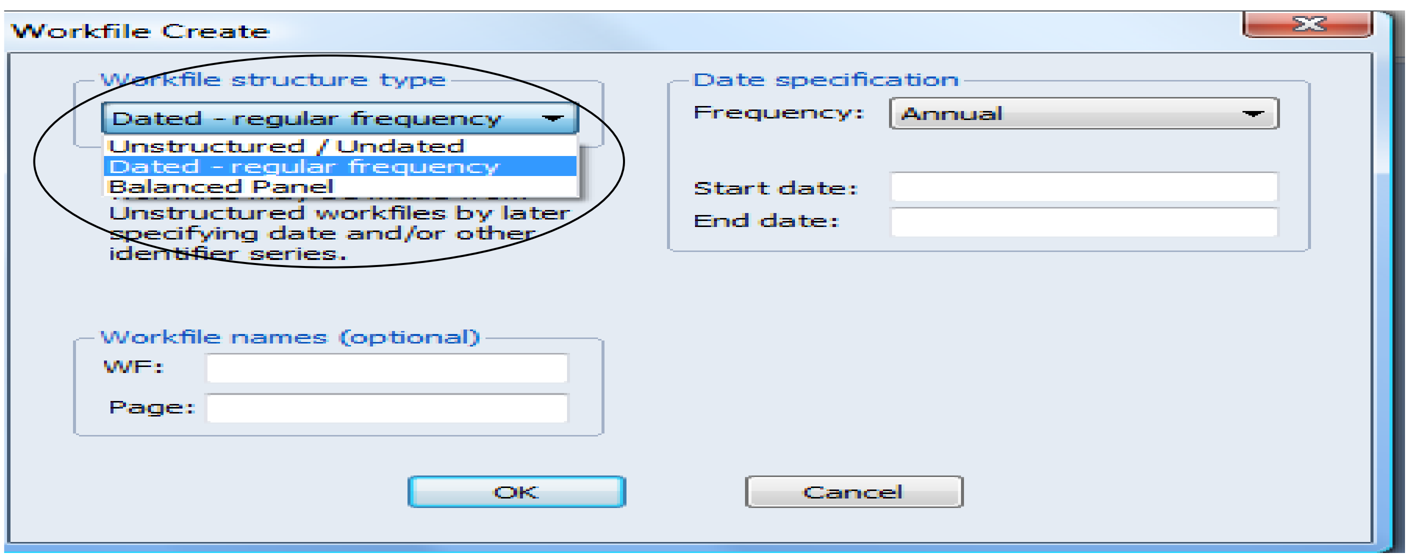
Data on one or more variables collected at a single point in time (i.e. Cross-section of stock returns *on the Athens Stock Exchange)*

II. <u>Time–Series Data *(Dated–regular frequency)*</u>

We need start & end date + frequency (i.e. Daily/ Weekly/Monthly/ Annually Returns on S&P500)

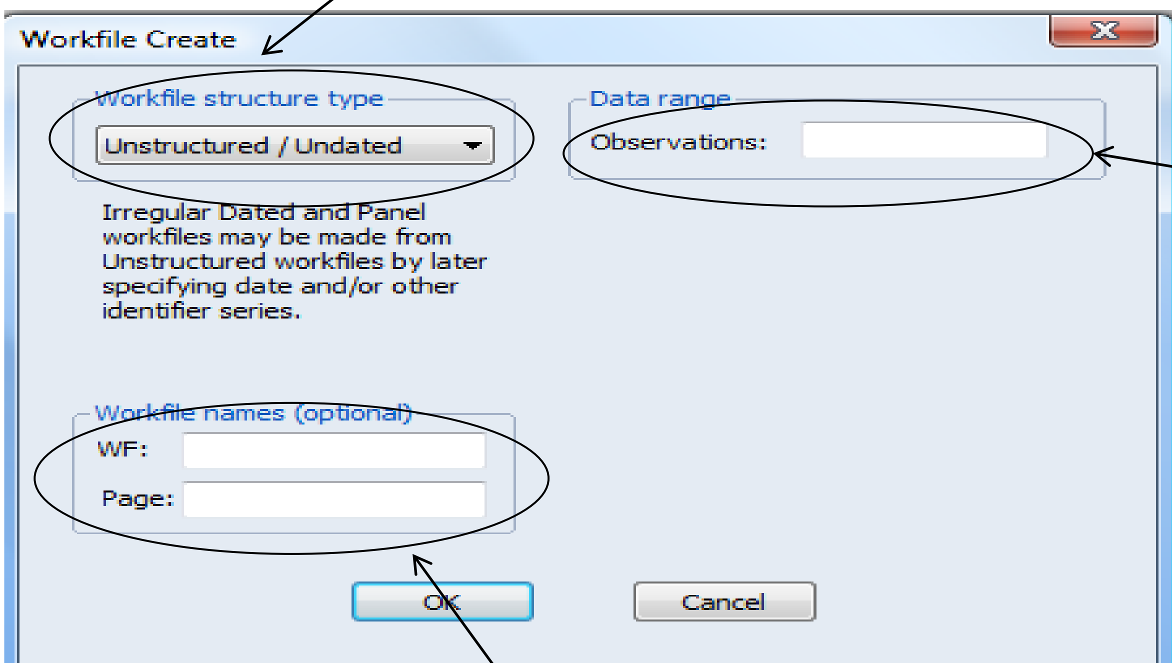
III. <u>Panel Data *(Balanced Panel)*</u>

Combination of I and II. (i.e. Monthly prices of energy stocks over five years)



3.1 *Unstructured/Undated Data*

Type of Data (*here* **Unstructured/Undated**)



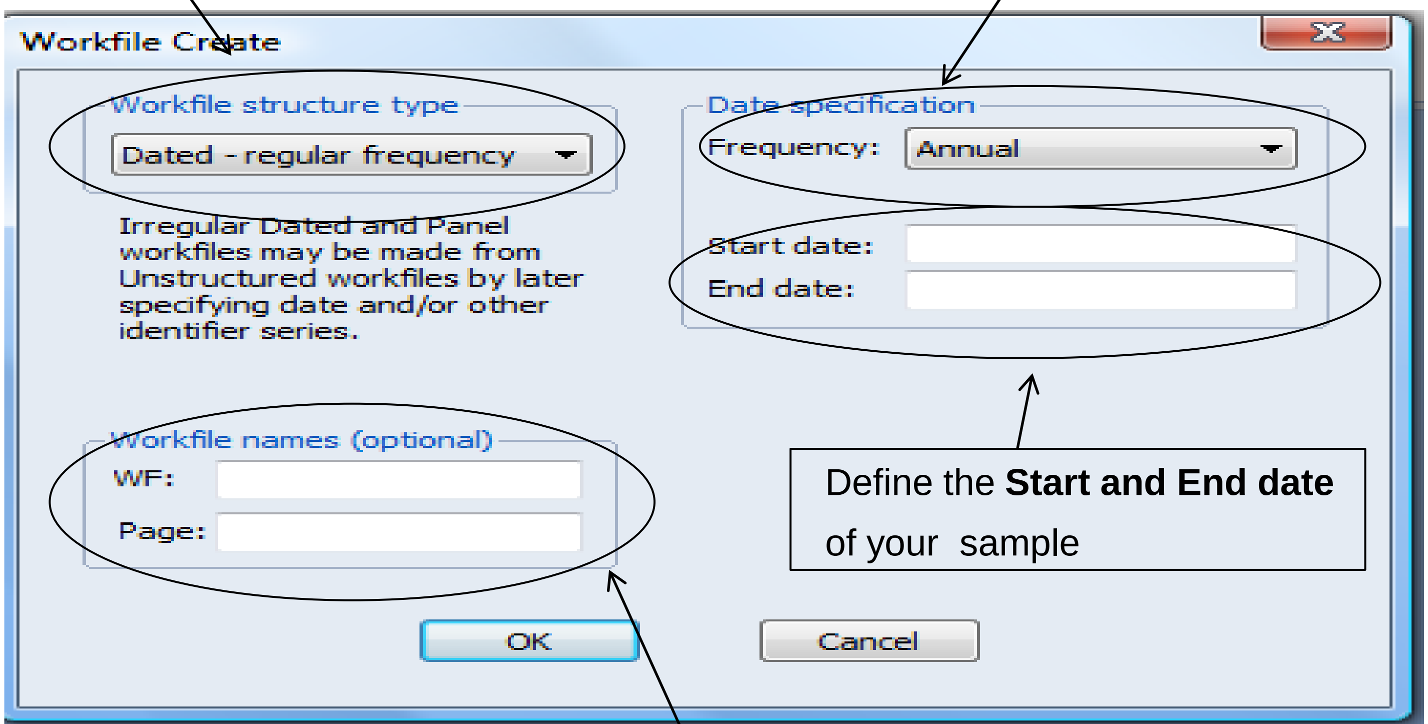
Define the **number of observations** of your sample

Define the name of your workfile

3.2 Dated – Regular frequency Data

Type of Data (*here* **Dated–Regular Frequency Data**)

Define the **frequency** of the data (i.e. annually, monthly, daily, etc.)



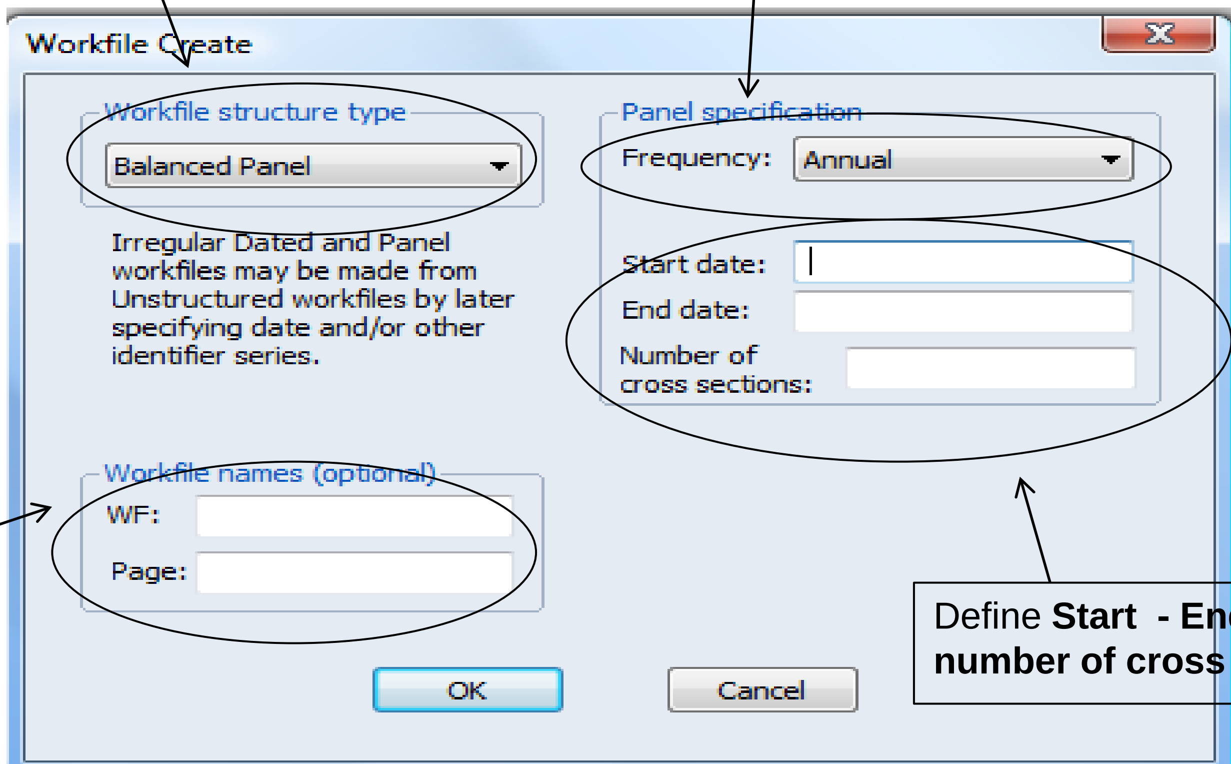
Define the **Start and End date** of your sample

Define the name of your workfile

3.3 Balanced Panel Data

Type of Data (*here* **Balanced Panel Data)**

Define the **frequency** of the data (i.e. annually, monthly, daily, etc.)



Define the name of your workfile

Define **Start - End date** and the **number of cross sections**

Empirical Example 1 :

Unstructured/Undated Data

Empirical Example 1 - *Unstructured/Undated Data (1)*

1. Go to folder *Empirical Examples* ⟶ *Example_1*

- Shows the results of a test in Finance for 60 students

(Source: "Econometrics for Financial Analysis", A. G. Merikas, A. A. Merika)

2. Open txt file: *example_1.txt*

3. Define the type of the data : *Unstructured/Undated Data*

4. Define the number of observations of the sample: *60*

5. Close txt file

Empirical Example 1 - *Unstructured/Undated Data (2)*

6 . Open *EViews* from PC - lab

7. *Go to File* → *New* → *Workfile*

8. Set *Workfile Structure Type* : Unstructured/Undated

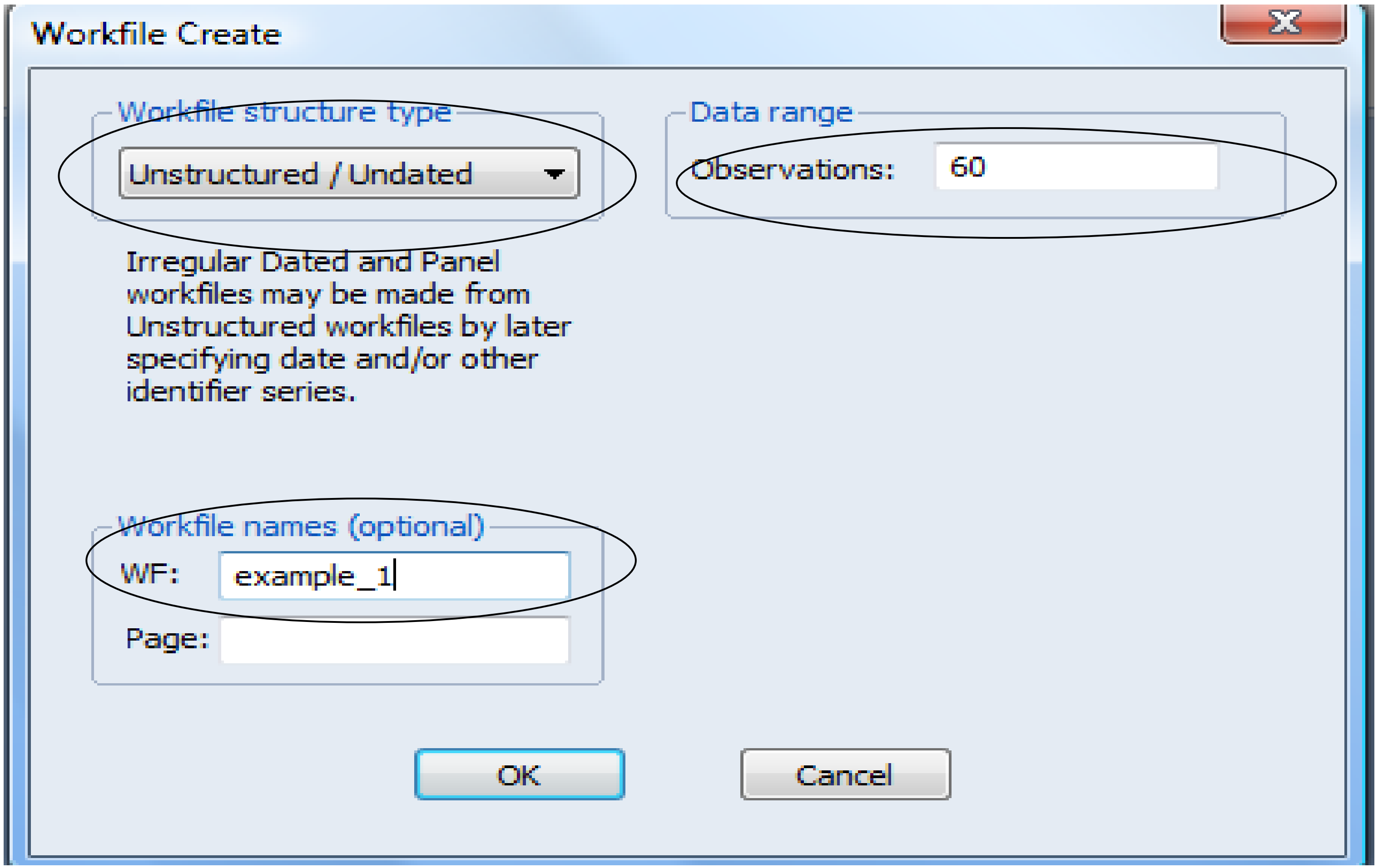
9. Set *Data Range / Observations* equal to 60

10. Set the name of your workfile *WF : Example_1*

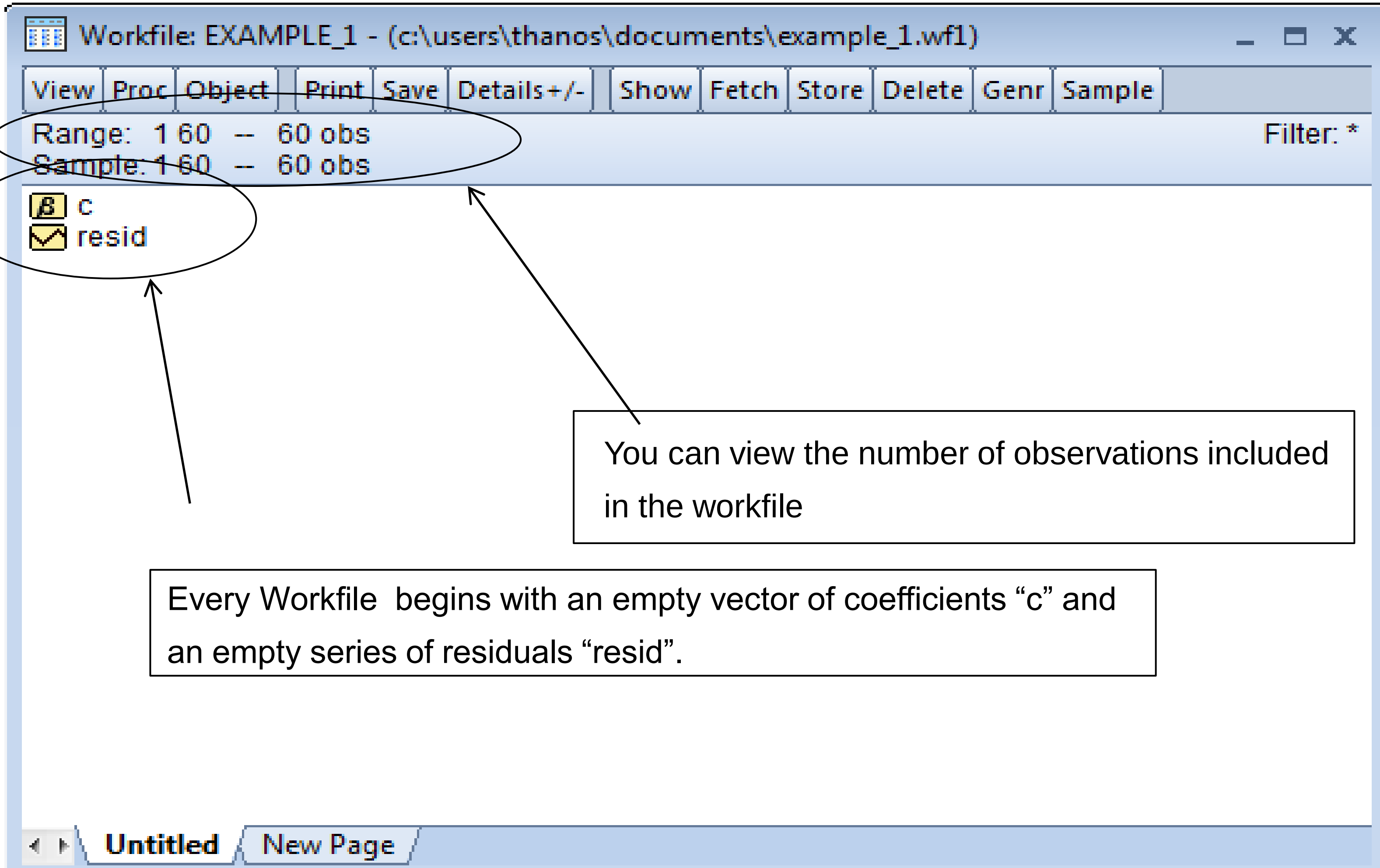
- Set the *Page Blank, as it is.*

11. Click *OK.*

Empirical Example 1 - ***Unstructured/Undated Data (3)***



Empirical Example 1 - *Unstructured/Undated Data (4)*

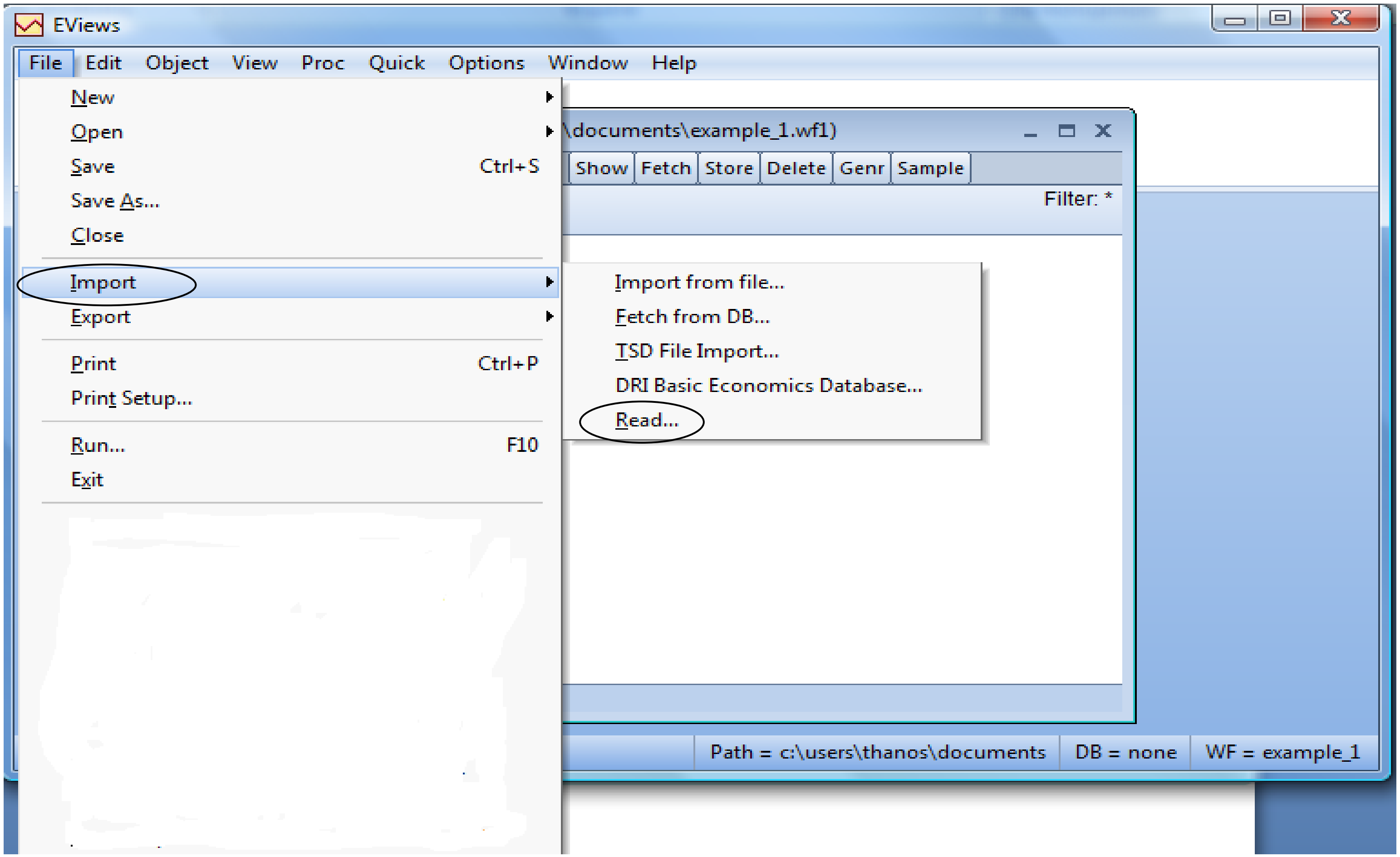


You can view the number of observations included in the workfile

Every Workfile begins with an empty vector of coefficients "c" and an empty series of residuals "resid".

12. Go to *File* ⟶ *Import* ⟶ *Read*

Find the txt file example_1.txt in the folder *"empirical examples"*

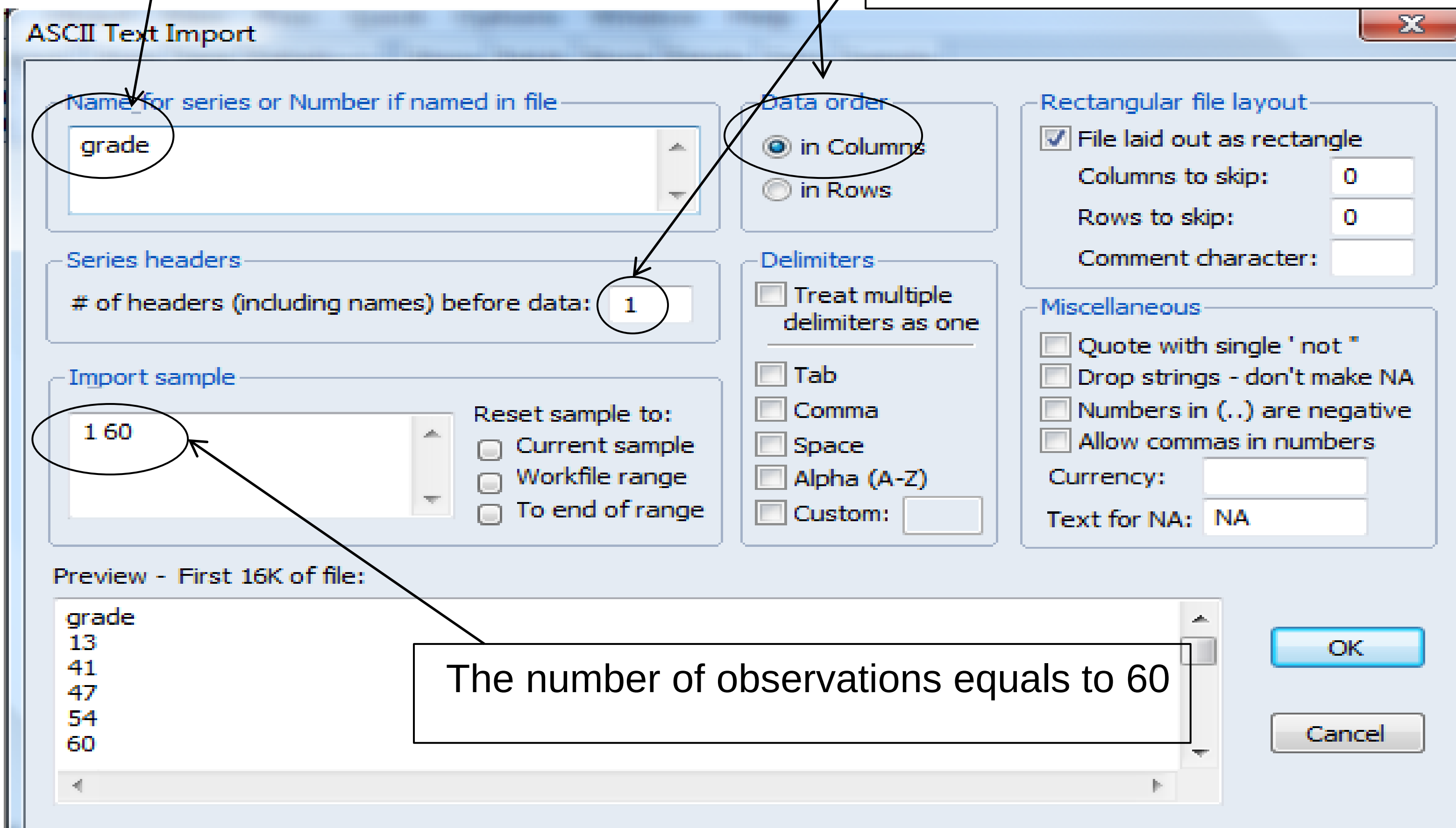


Empirical Example 1 - *Unstructured/Undated Data (6)*

You can either write “grade” or “1”, *(the number of the variable in this example equals to one)*

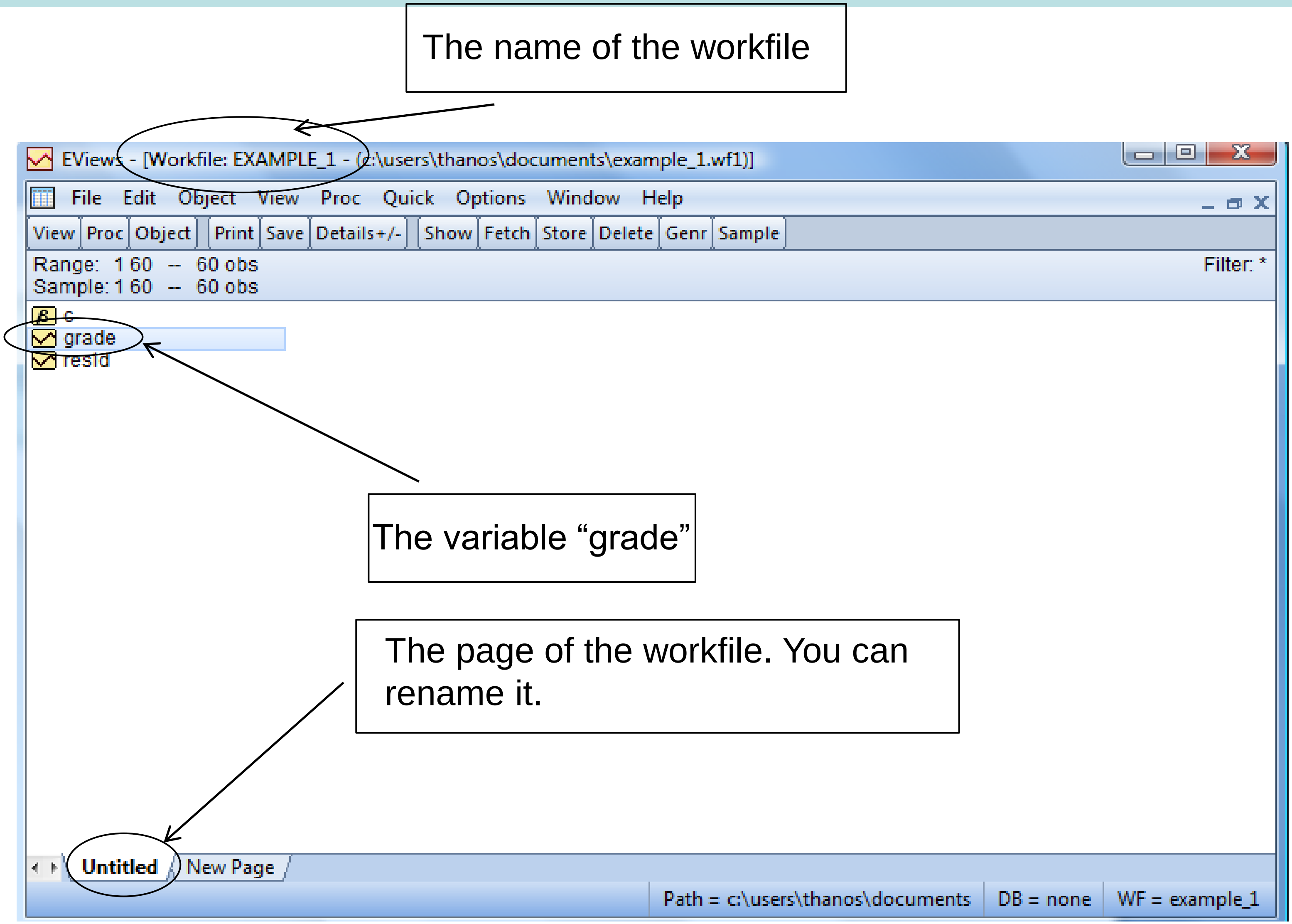
The order of the data are in a column setting

We set 1, since we have one row of headers



The number of observations equals to 60

13. Click OK


The name of the workfile
EViews - [Workfile: EXAMPLE_1 - (c:\users\thanos\documents\example_1.wf1)]
File Edit Object View Proc Quick Options Window Help
View Proc Object Print Save Details+/- Show Fetch Store Delete Genr Sample
Range: 1 60 -- 60 obs
Sample: 1 60 -- 60 obs
Filter: *
c
grade
resid
The variable "grade"
The page of the workfile. You can rename it.
Untitled
New Page
Path = c:\users\thanos\documents
DB = none
WF = example_1

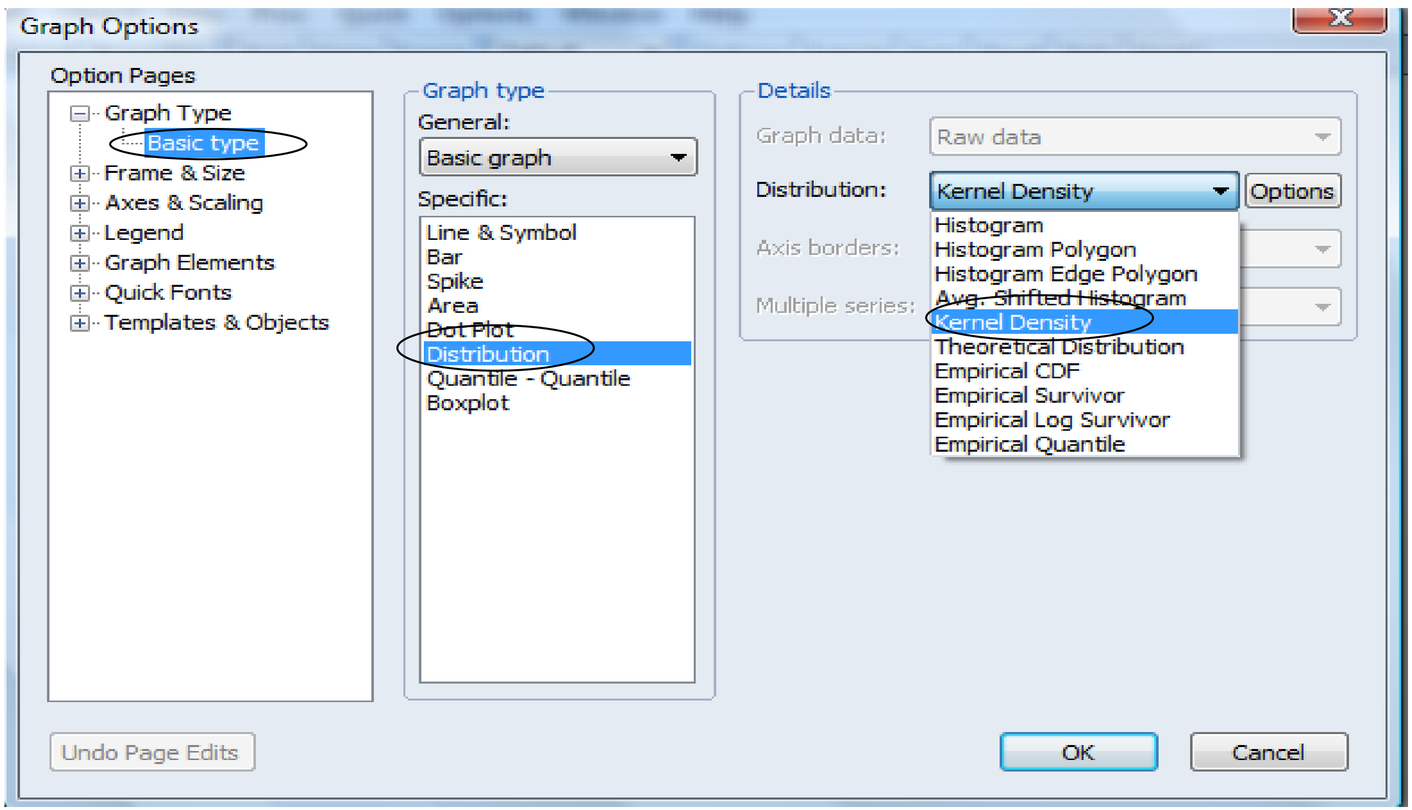
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Empirical Example 1 - *Unstructured/Undated Data (8)*

14. Double Click on the variable *“grade”*

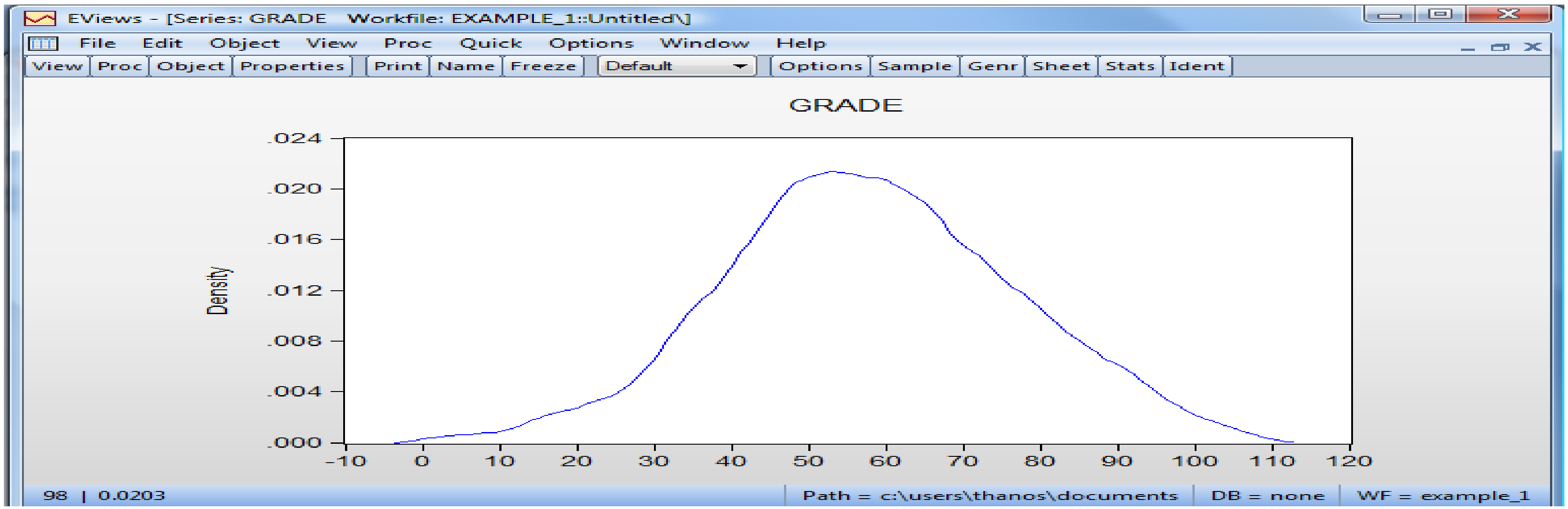
15. Go to View ⟶ Graph

- Select Basic type / Distribution / Kernel Density



We can visualize the shape of distribution

- Distinction between normal and non-normal distributions



The above diagram shows the probability density function (pdf) of the variable "grade". In a first view resembles the "normal distribution" with pdf function :

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

However, we have to derive the basic distribution parameters.

Empirical Example 1 - *Unstructured/Undated Data (10)*

We can now define the basic distribution parameters

- **Mean:** *The expected value* $E(X)=\frac{1}{n}\sum_{i=1}^{n}X_i$
- **Standard deviation:** *A measure of spread* $s=\sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n}(X-E(X))^2\right)}$
- **Skewness:** *Measures the extent to which the distribution is* ***not*** *symmetric*

$$\mu_3=E\left[X-E(X)\right]^3\equiv\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X-E(X)}{s\sqrt{(n-1)/n}}\right)^3$$

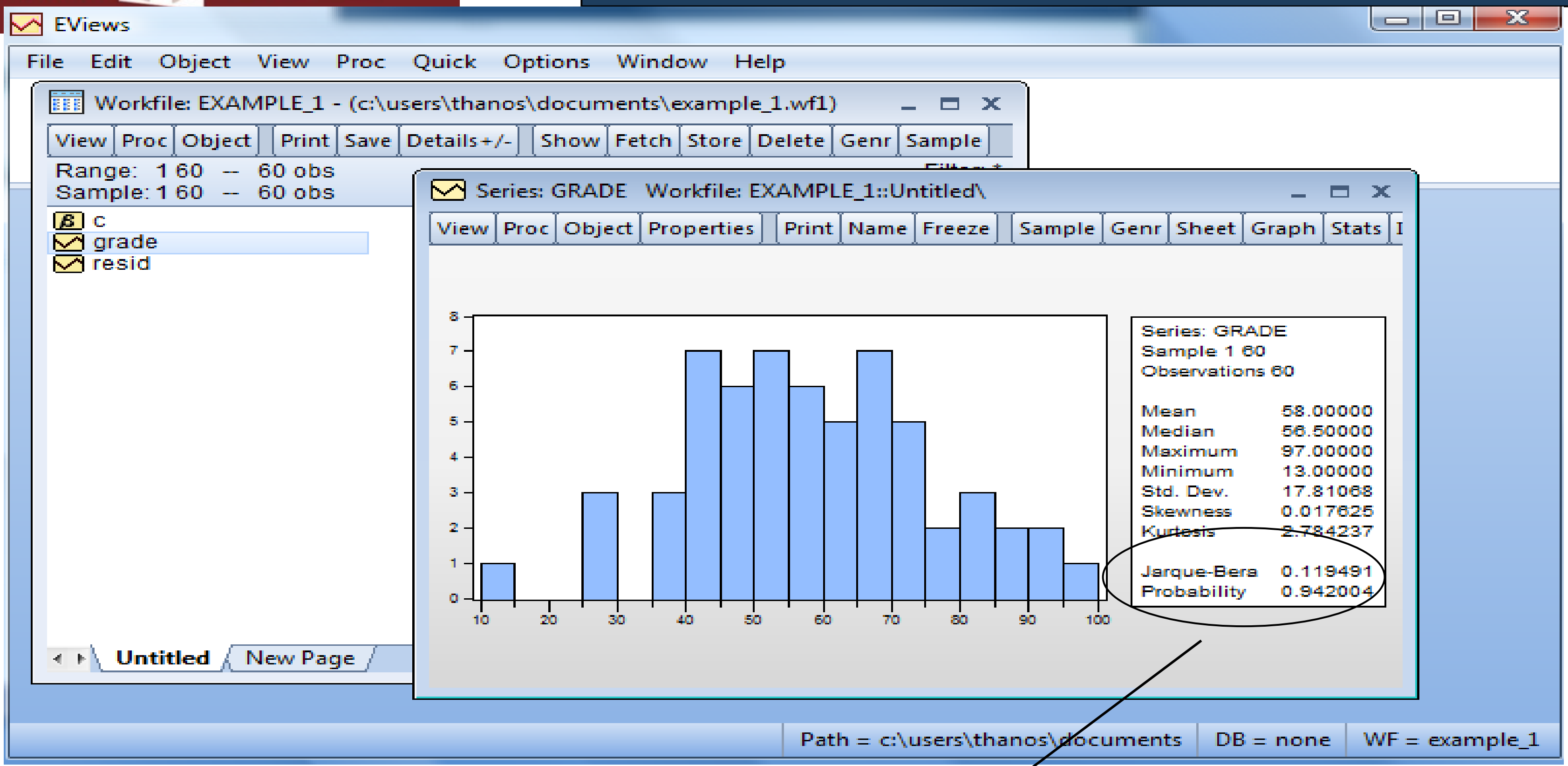
- If $\mu_3>0$ the right tail is the longer
- If $\mu_3<0$ the left tail is the longer
- If $\mu_3=0$ the distribution is normal

- **Kurtosis :** *Measures the "thickness of the tails"* $\mu_4=E\left[X-E(X)\right]^4\equiv\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X-E(X)}{s\sqrt{(n-1)/n}}\right)^4$

- If $\mu_4>3$ the distribution has long or thick tails (leptokurtic)
- If $\mu_4<3$ the distribution has short or thin tails (platykurtic)
- If $\mu_4=3$ the distribution is normal

16. Go to View $\longrightarrow$ Descriptive Statistics and Tests $\longrightarrow$ Histogram and Stats

Empirical Example 1 - *Unstructured/Undated Data (11)*



Jarque-Bera: is a test statistic **for testing whether the series are normally distributed**. **Probability** is the probability that Jarque-Bera statistic exceeds in absolute value the observed value under **the null hypothesis of a normal distribution.**

$$JB = \frac{n}{6}\left(\mu_3^2 + \frac{(\mu_4 - 3)^2}{4}\right)$$

17. Close the above window ⟶ Go to *File* ⟶ *Save as…*

Alternative ways of importing data

- So far we have shown how to import data from a txt file

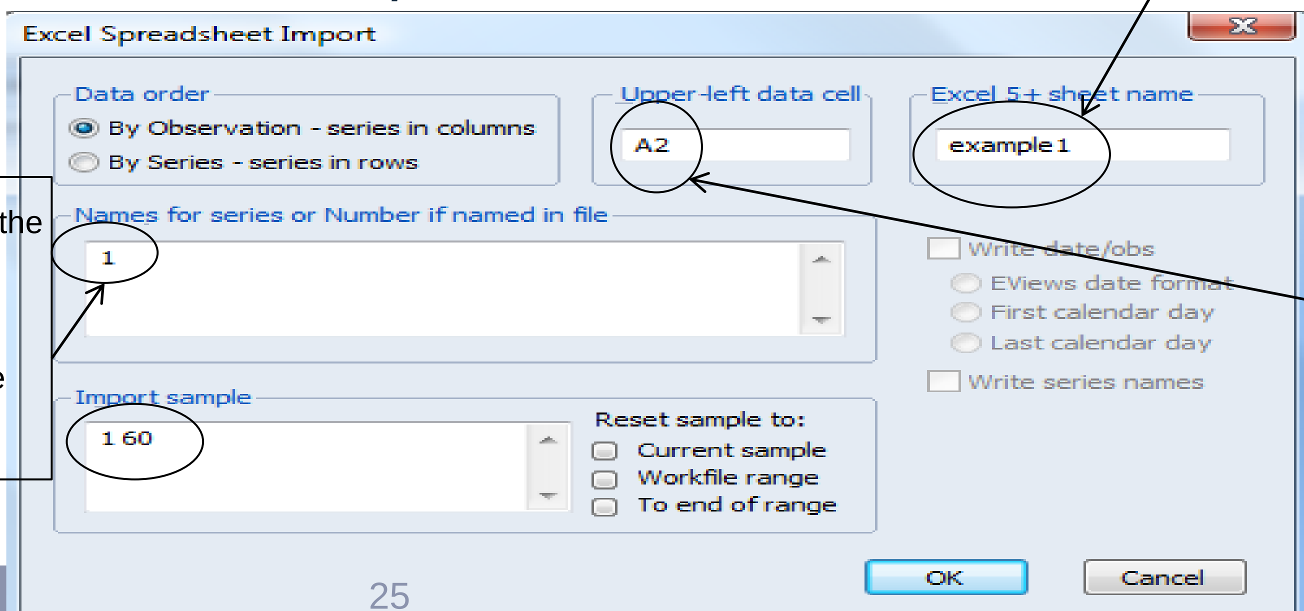
From xls file

1. Create the Workfile (Steps 1 to 11)
2. Go to ***File*** ⟶ ***Import*** ⟶ ***Read***
3. ***Select*** *example_1.xls*

The name of worksheet of the xls file

The number **or** the name of variables (here we set the number)

The cell of xls file that the data begin.



A general way to import any types of files:

1. Create the Workfile (Steps 1 to 11)
2. Go to ***File*** → ***Import*** → ***Import from file***
3. ***Select*** *example_1.xlsx/example_1.xls/example.csv /etc..*

Empirical Example 2 :

Dated – Regular frequency Data

Empirical Example 2- ***Dated – Regular frequency Data (1)***

1. Go to folder *Empirical Examples* ⟶ *Example_2*

- Shows the monthly total simple returns(capital + dividends) in $ of the equity indices of G7 countries from 31/01/1980 – 31/10/2012 .

(Source : DataStream)

2. Open txt. file: *example_2.txt*

3. Define the type of the data : *Dated - Regular frequency data*

4. Close txt. file

Empirical Example 2- *Dated – Regular frequency Data (2)*

5 . Open *EViews* from PC - lab

6. *Go to File* → *New* → *Workfile*

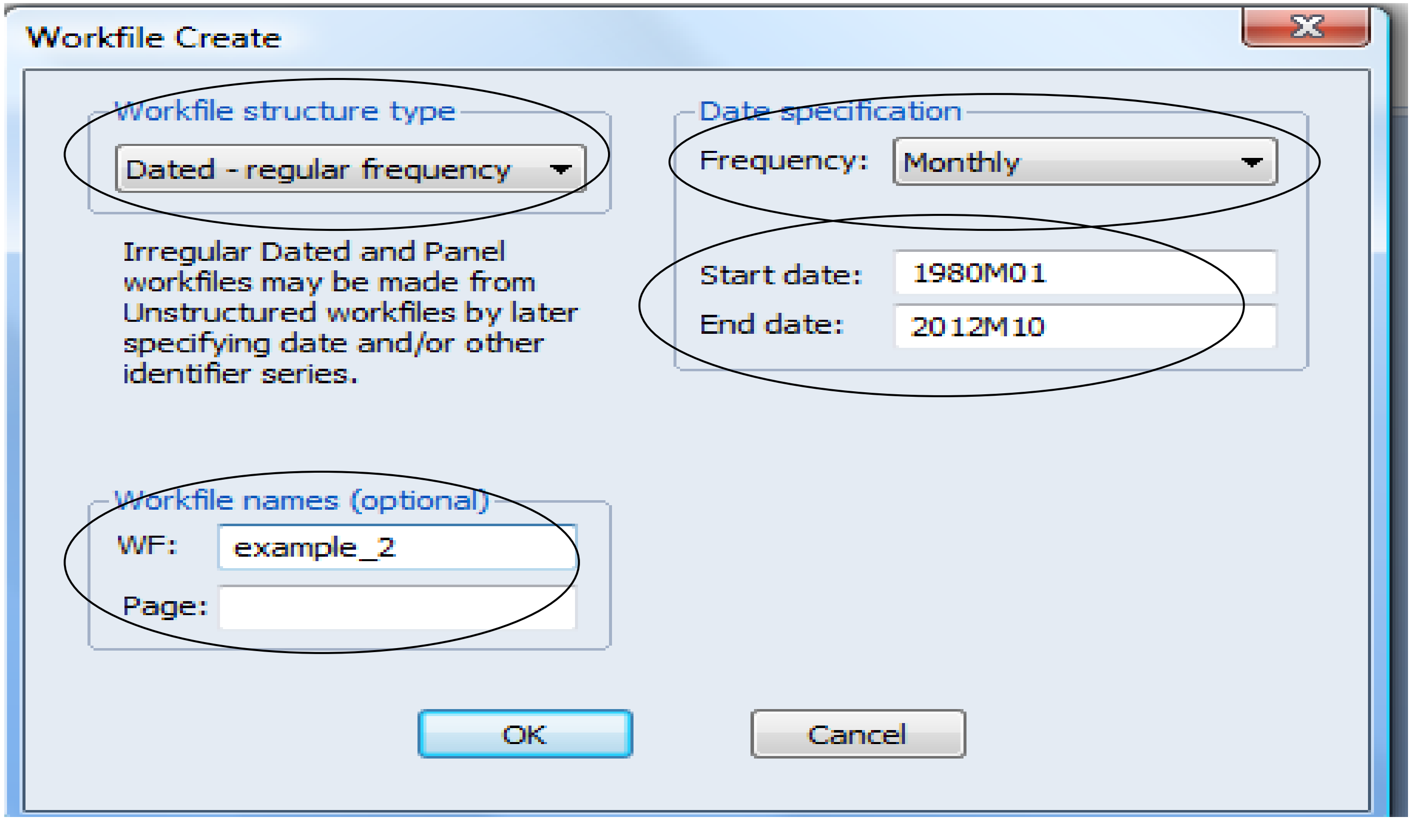
7. Set *Workfile Structure Type* : Dated - Regular frequency data

8. Set *Start Date : 1980M01 - End Date: 2012M10*

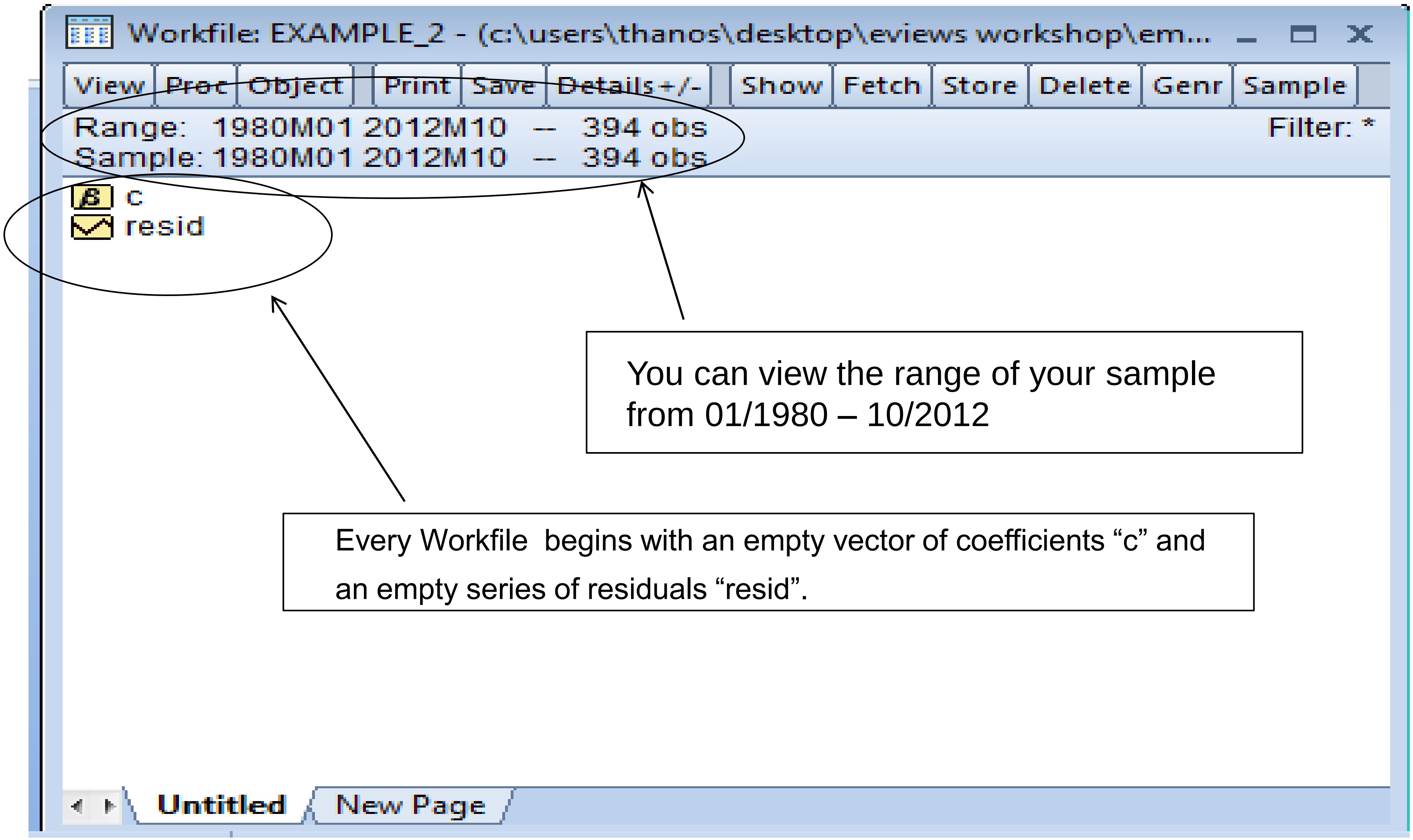
9. Set the name of your workfile *WF : Example_2*

- Set the *Page Blank, as it is.*

10. Click *OK.*


Workfile Create
Workfile structure type
Dated - regular frequency
Irregular Dated and Panel workfiles may be made from Unstructured workfiles by later specifying date and/or other identifier series.
Date specification
Frequency: Monthly
Start date: 1980M01
End date: 2012M10
Workfile names (optional)
WF: example_2
Page:
OK
Cancel

Empirical Example 2- *Dated – Regular frequency Data (4)*

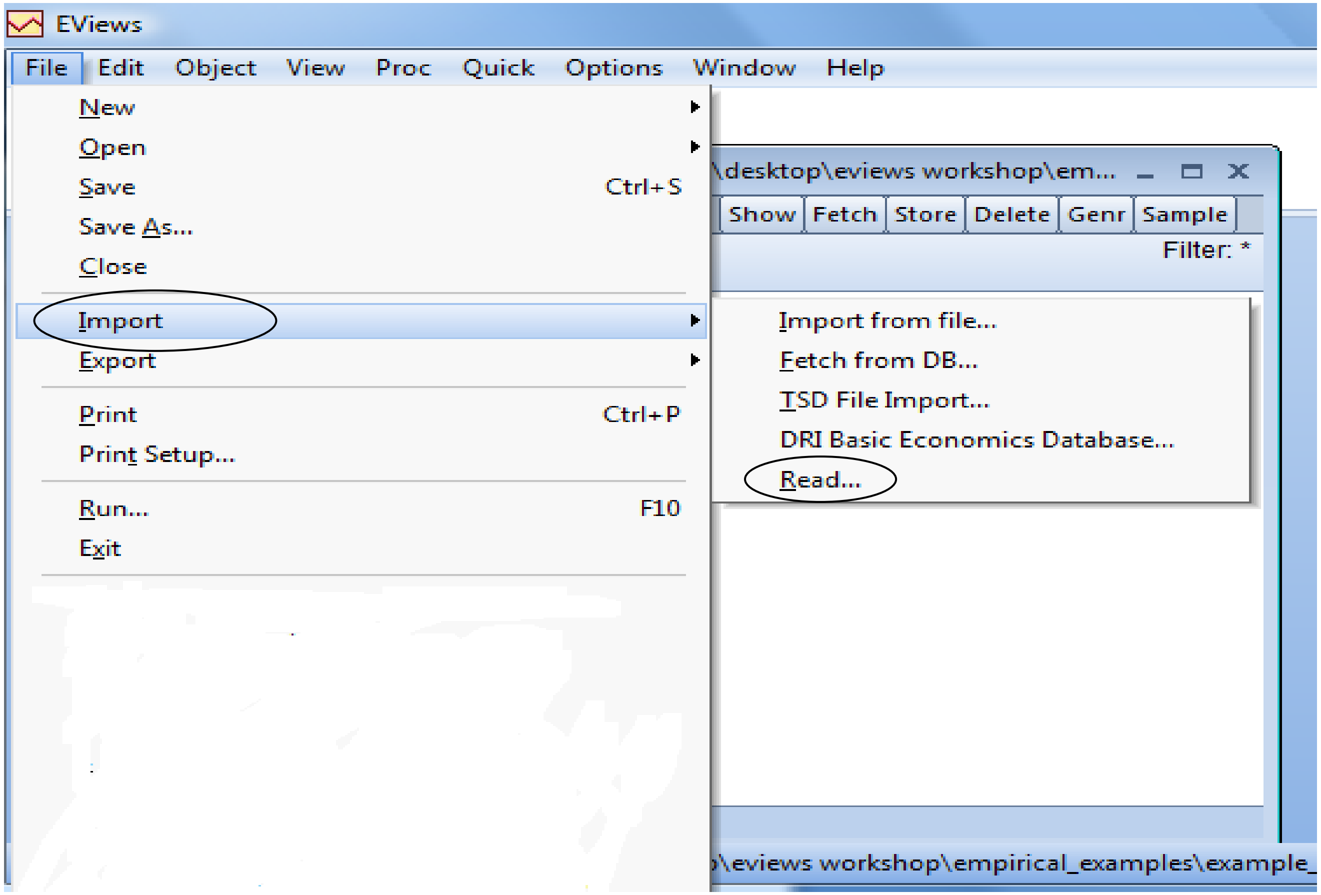


You can view the range of your sample from 01/1980 – 10/2012

Every Workfile begins with an empty vector of coefficients “c” and an empty series of residuals “resid”.

11. Go to *File* ⟶ *Import* ⟶ *Read*

Find the xlsx file example_2.xlsx in the folder *“empirical examples”*

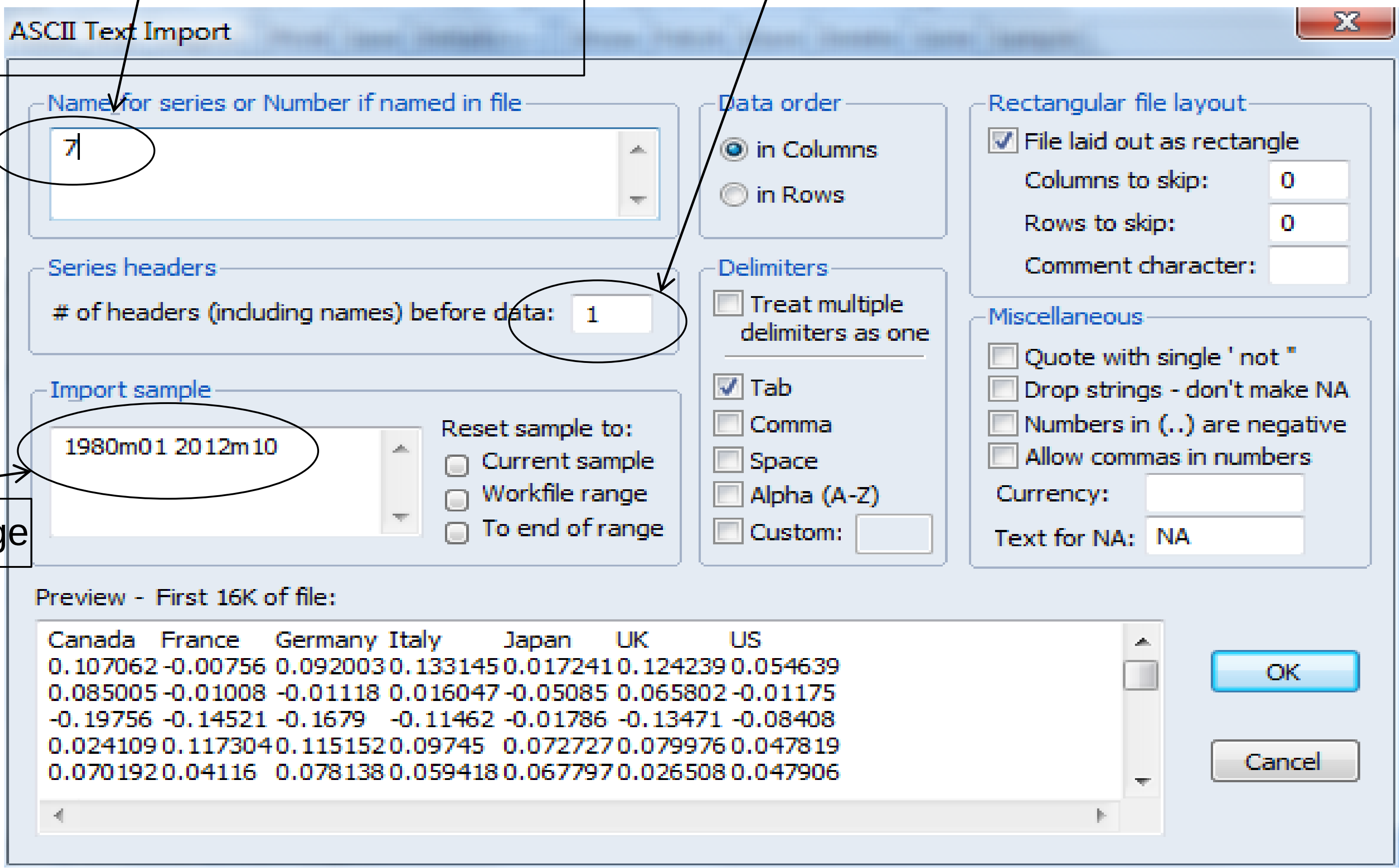


Empirical Example 2- ***Dated – Regular frequency Data (6)***

You can either write the name of countries or "7",*(the number of the variable in this example equals to seven)*

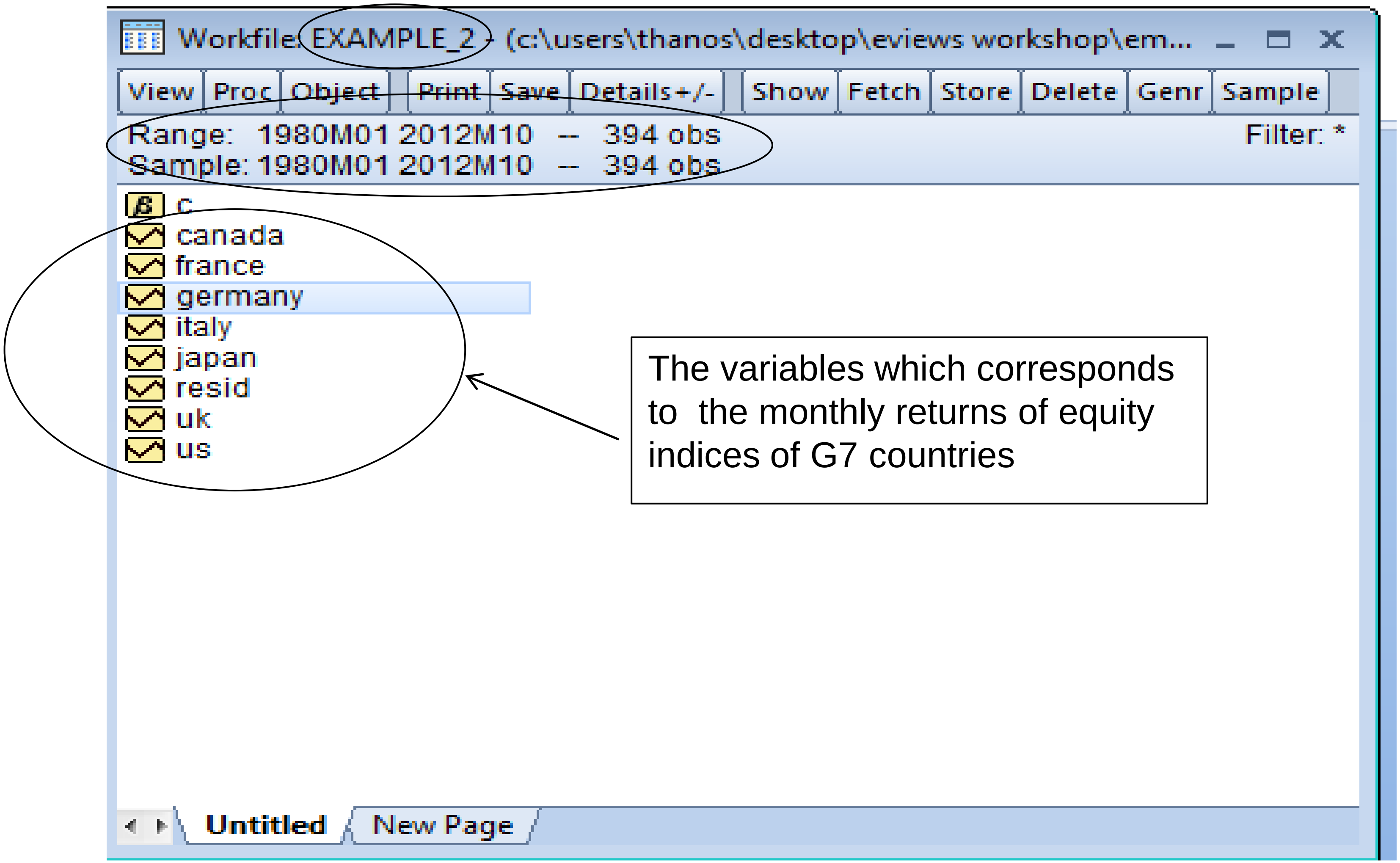
We set 1, since we have one row of headers

Date Range



Empirical Example 2- *Dated – Regular frequency Data (7)*

12. Click OK

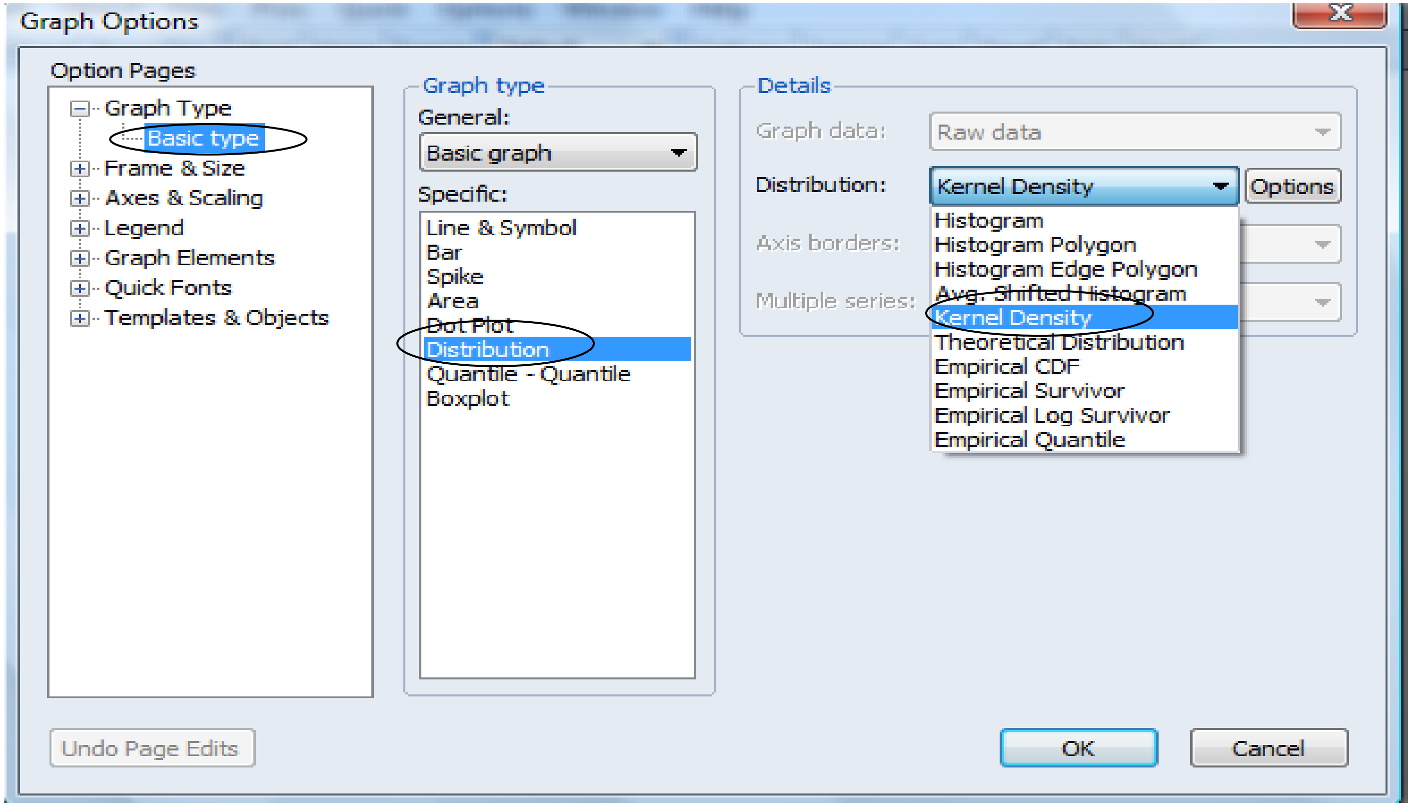


Empirical Example 2- *Dated – Regular frequency Data (8)*

13. Double Click on the variable *"us"*

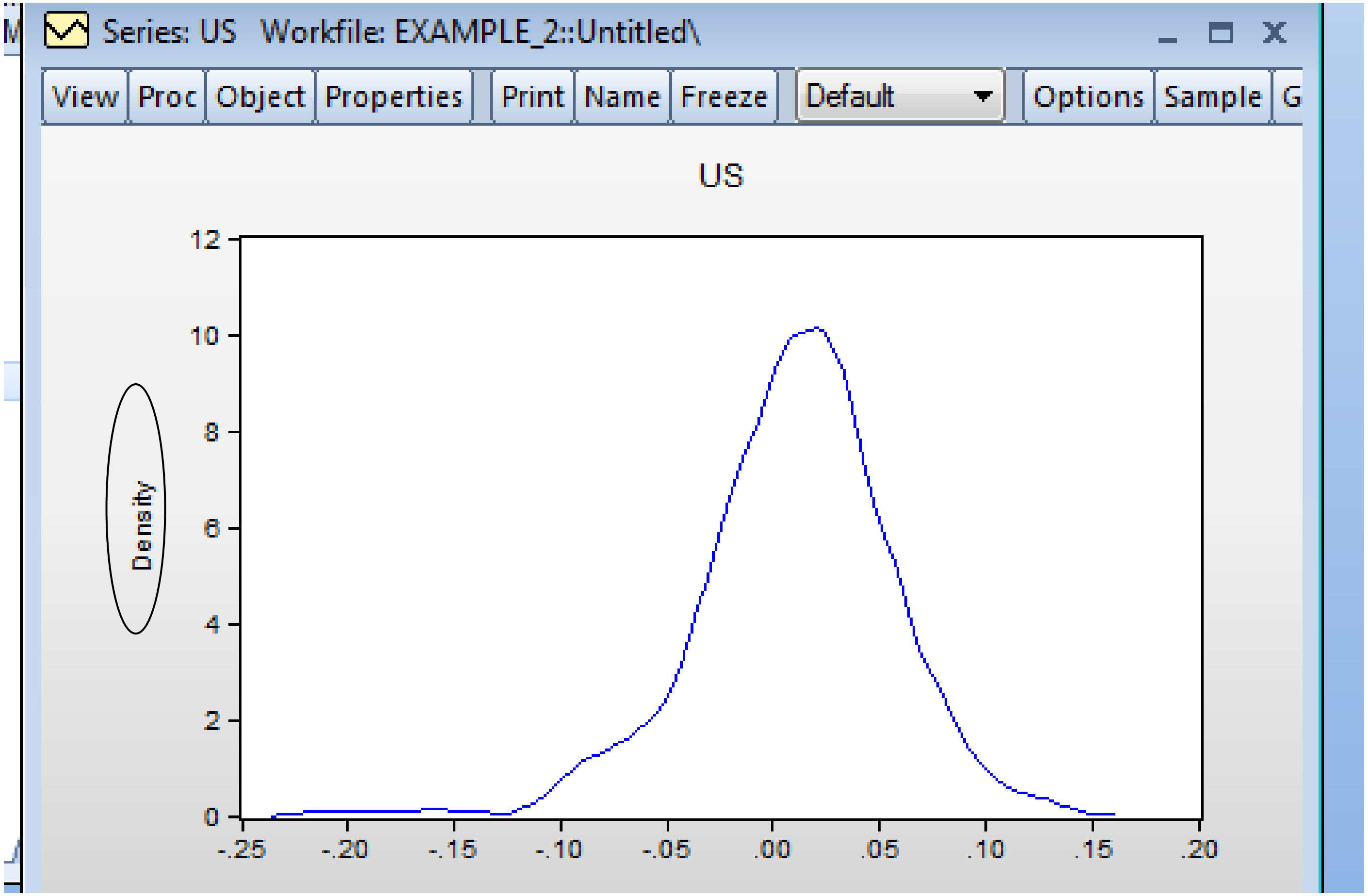
14. Go to View ⟶ Graph

- Select Basic type / Distribution / Kernel Density



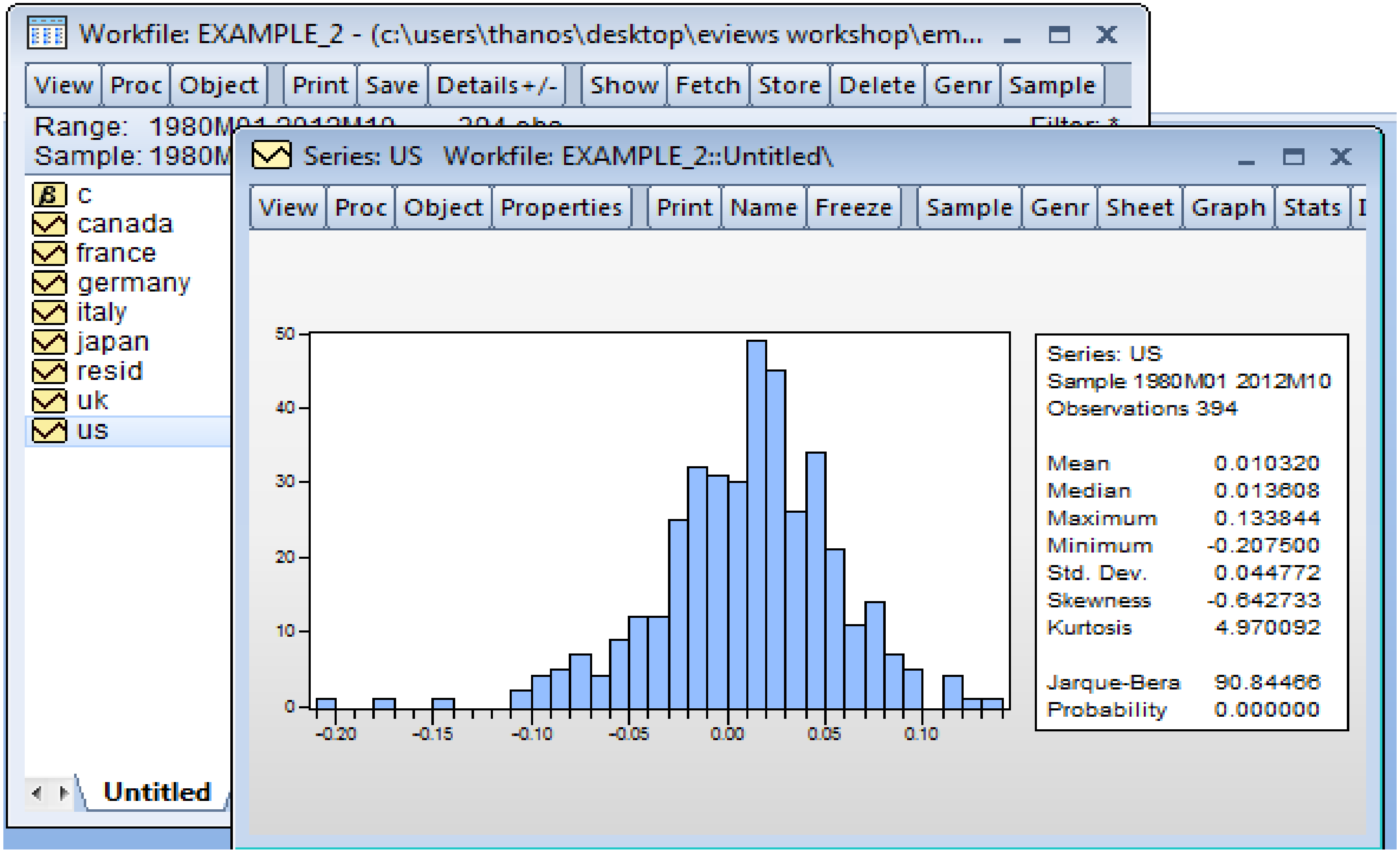
We can visualize the shape of distribution

- Distinction between normal and non-normal distributions



Empirical Example 2- *Dated – Regular frequency Data (10)*

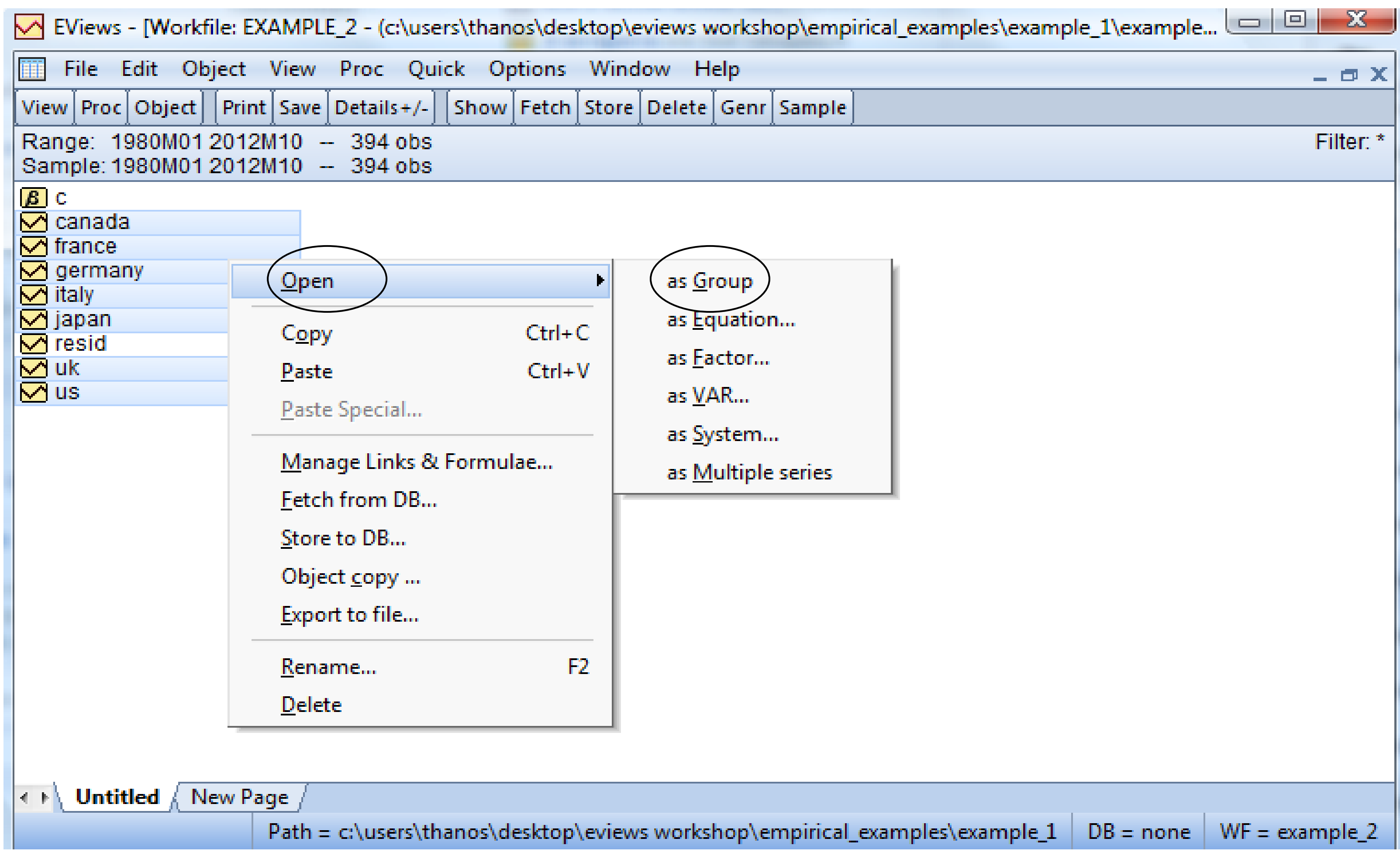
15. Go to View ⟶ Descriptive Statistics and Tests ⟶ Histogram and Stats



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

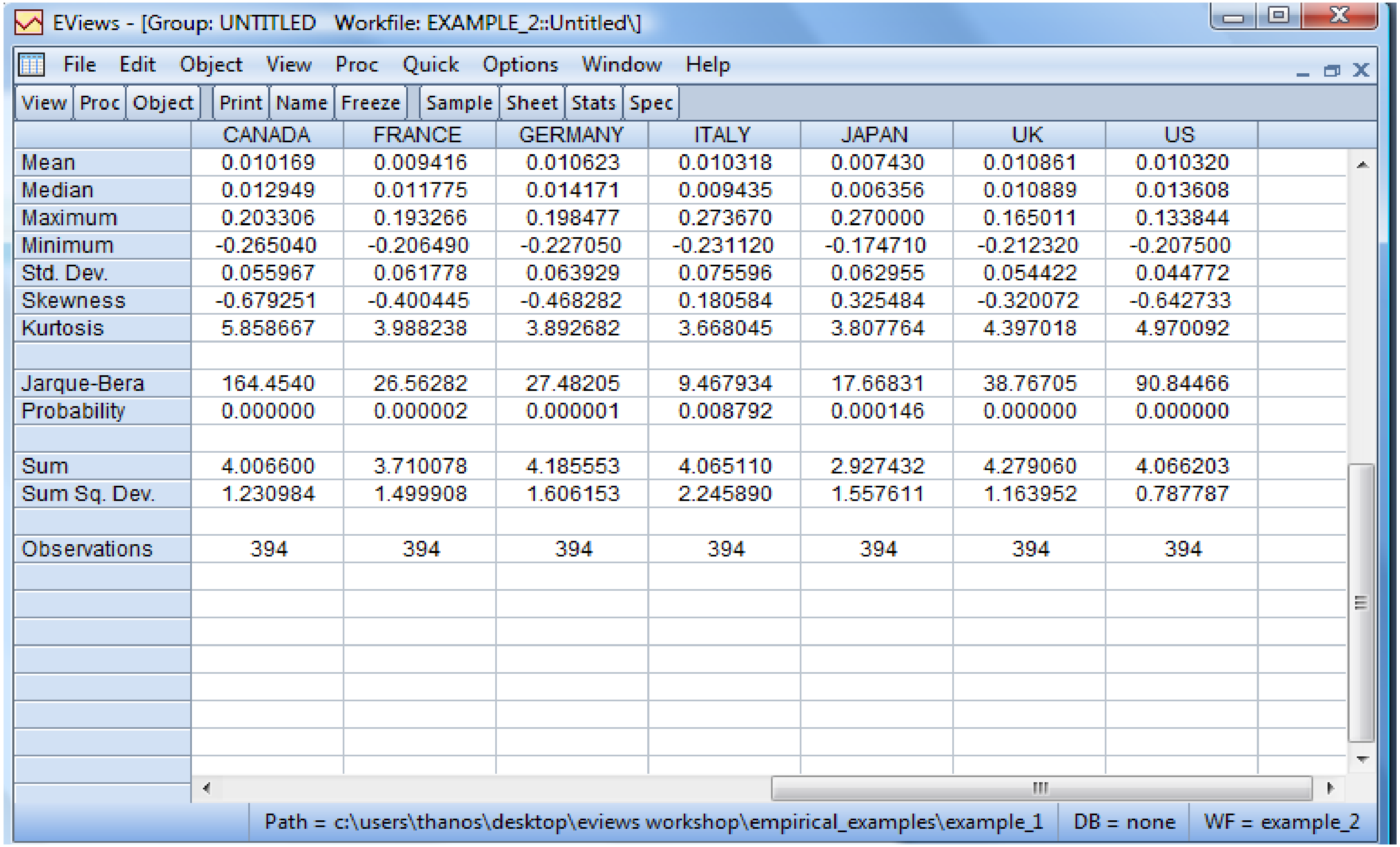
Empirical Example 2- *Dated – Regular frequency Data (11)*

16. Select G7 ⟶ Right Click ⟶ Open ⟶ as Group



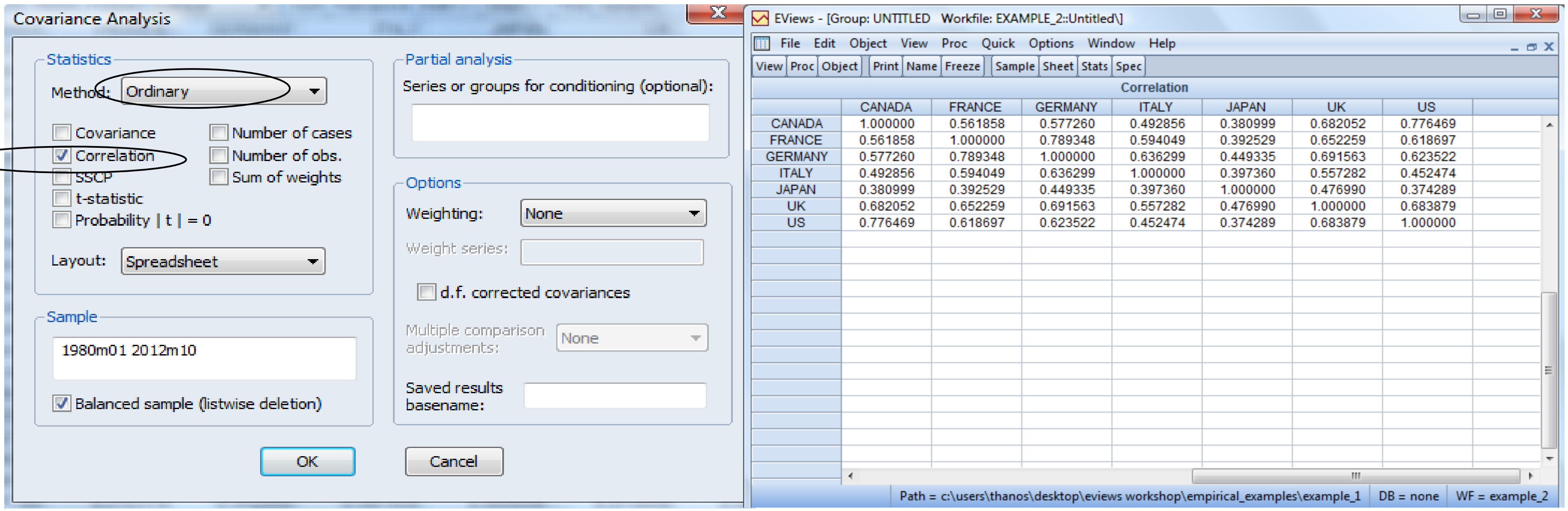
Empirical Example 2- *Dated – Regular frequency Data (12)*

17. Go to *View* ⟶ *Descriptive Stats* ⟶ *Common Sample*



	CANADA	FRANCE	GERMANY	ITALY	JAPAN	UK	US
Mean	0.010169	0.009416	0.010623	0.010318	0.007430	0.010861	0.010320
Median	0.012949	0.011775	0.014171	0.009435	0.006356	0.010889	0.013608
Maximum	0.203306	0.193266	0.198477	0.273670	0.270000	0.165011	0.133844
Minimum	-0.265040	-0.206490	-0.227050	-0.231120	-0.174710	-0.212320	-0.207500
Std. Dev.	0.055967	0.061778	0.063929	0.075596	0.062955	0.054422	0.044772
Skewness	-0.679251	-0.400445	-0.468282	0.180584	0.325484	-0.320072	-0.642733
Kurtosis	5.858667	3.988238	3.892682	3.668045	3.807764	4.397018	4.970092
Jarque-Bera	164.4540	26.56282	27.48205	9.467934	17.66831	38.76705	90.84466
Probability	0.000000	0.000002	0.000001	0.008792	0.000146	0.000000	0.000000
Sum	4.006600	3.710078	4.185553	4.065110	2.927432	4.279060	4.066203
Sum Sq. Dev.	1.230984	1.499908	1.606153	2.245890	1.557611	1.163952	0.787787
Observations	394	394	394	394	394	394	394

18. Go to *View* ⟶ *Covariance Analysis* ⟶ *Correlation*



Correlation

	CANADA	FRANCE	GERMANY	ITALY	JAPAN	UK	US
CANADA	1.000000	0.561858	0.577260	0.492856	0.380999	0.682052	0.776469
FRANCE	0.561858	1.000000	0.789348	0.594049	0.392529	0.652259	0.618697
GERMANY	0.577260	0.789348	1.000000	0.636299	0.449335	0.691563	0.623522
ITALY	0.492856	0.594049	0.636299	1.000000	0.397360	0.557282	0.452474
JAPAN	0.380999	0.392529	0.449335	0.397360	1.000000	0.476990	0.374289
UK	0.682052	0.652259	0.691563	0.557282	0.476990	1.000000	0.683879
US	0.776469	0.618697	0.623522	0.452474	0.374289	0.683879	1.000000

We can here define

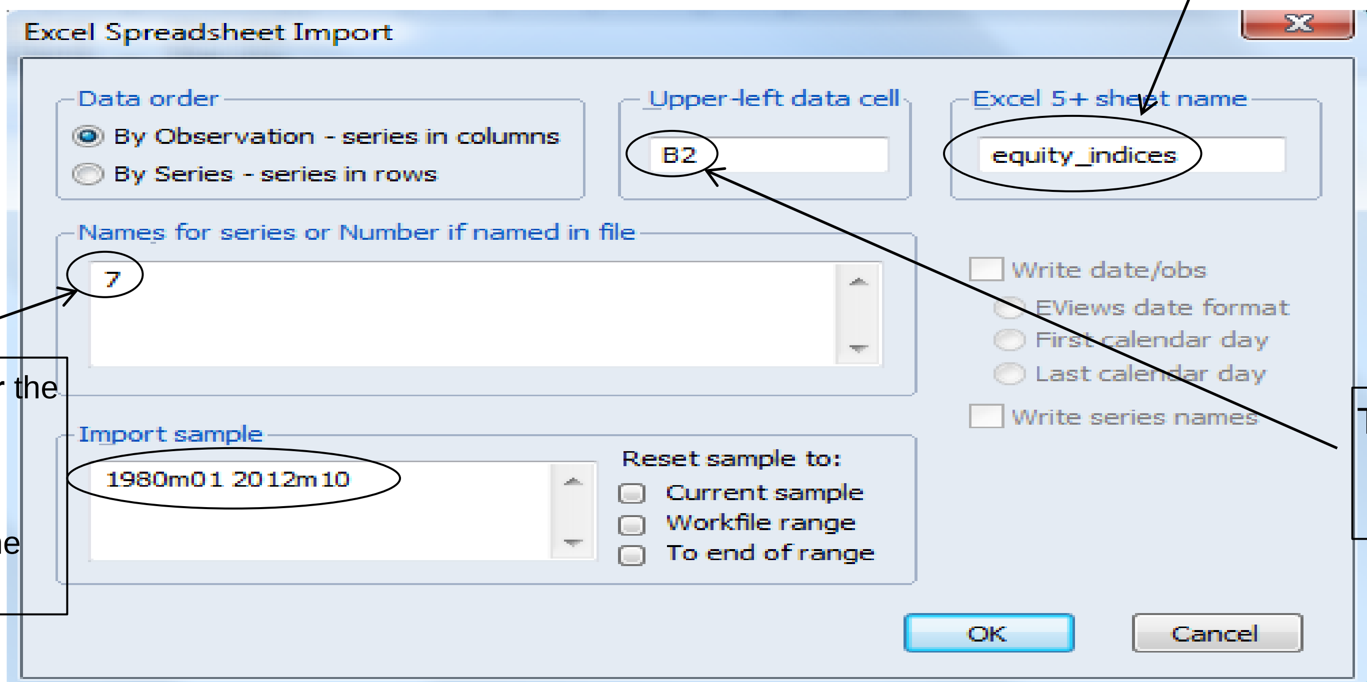
- Covariance between X and Y variables $Cov(X,Y)=E(XY)-E(X)E(Y)$
- Correlation between X and Y variables $\rho=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

19 . Close the above window ⟶ Go to *File* ⟶ *Save as…*

From xls file

1. Create the Workfile (Steps 1 to 11)
2. Go to ***File*** ⟶ ***Import*** ⟶ ***Read***
3. ***Select*** *example_2.xls*

The name of worksheet of the xls file



The number **or** the name of variables (here we set the number)

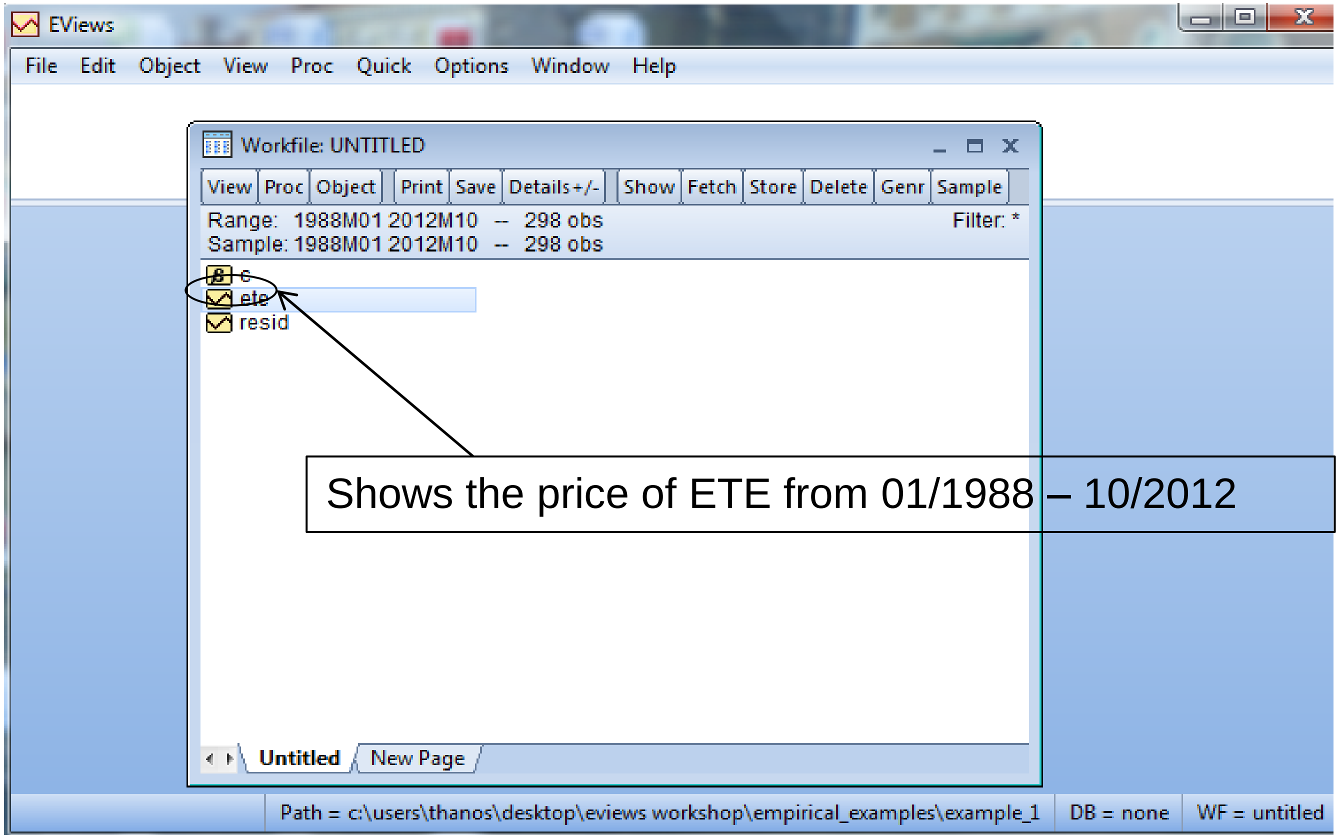
The cell of xls file that the data begin

Empirical Example 3 :

Transformations of Data

1.Go to folder *Example_3* → *Double click on Example_3.wf1 (eviews workfile)*

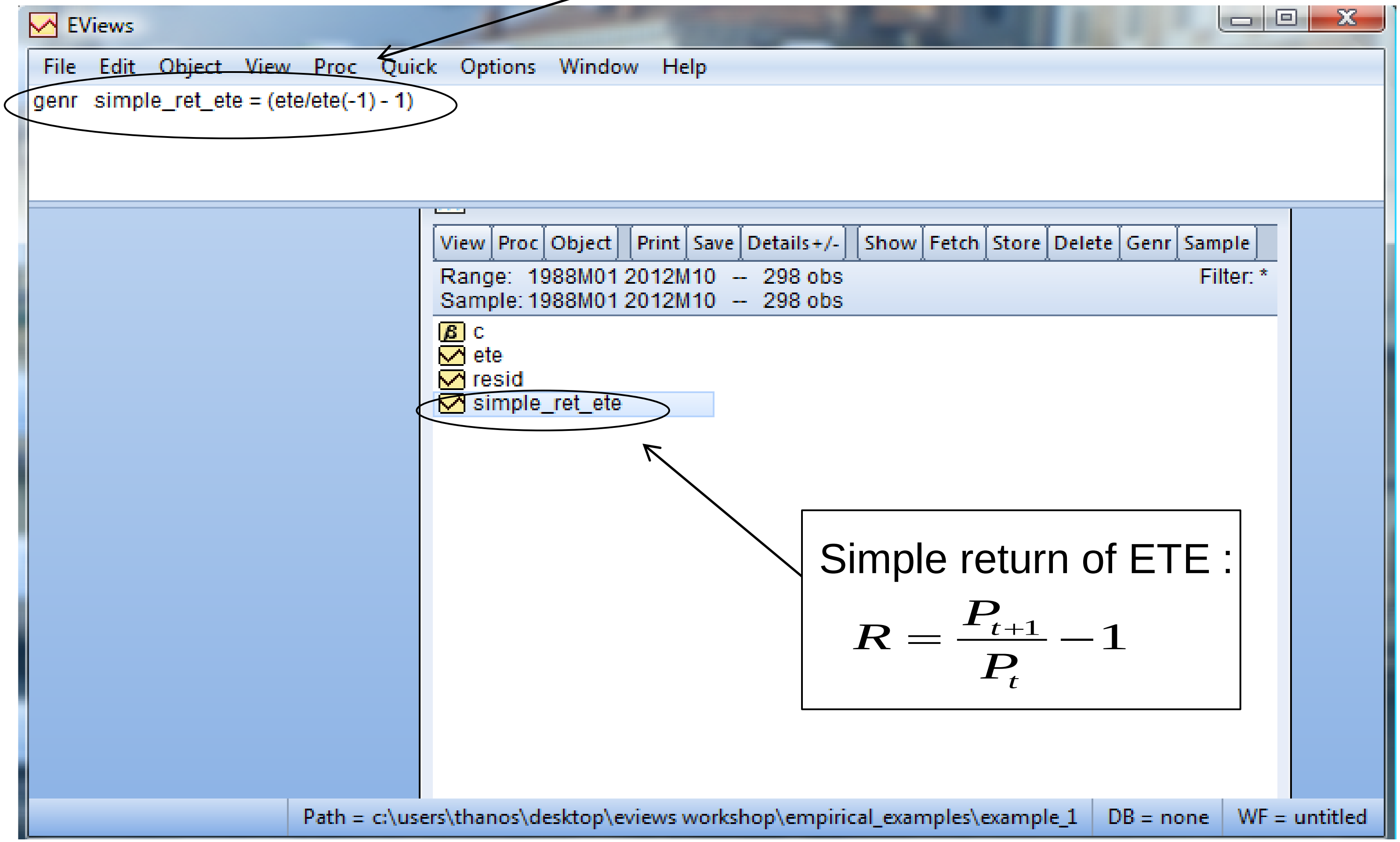
- We present the price of ETE from 01/1988 – 10/2012 *(Source : DataStream)*



Shows the price of ETE from 01/1988 – 10/2012

Empirical Example 3 : *Transformation of data (2)*

On the command window **type** *genr simple_ret_ete = (ete/ete(-1) - 1)*

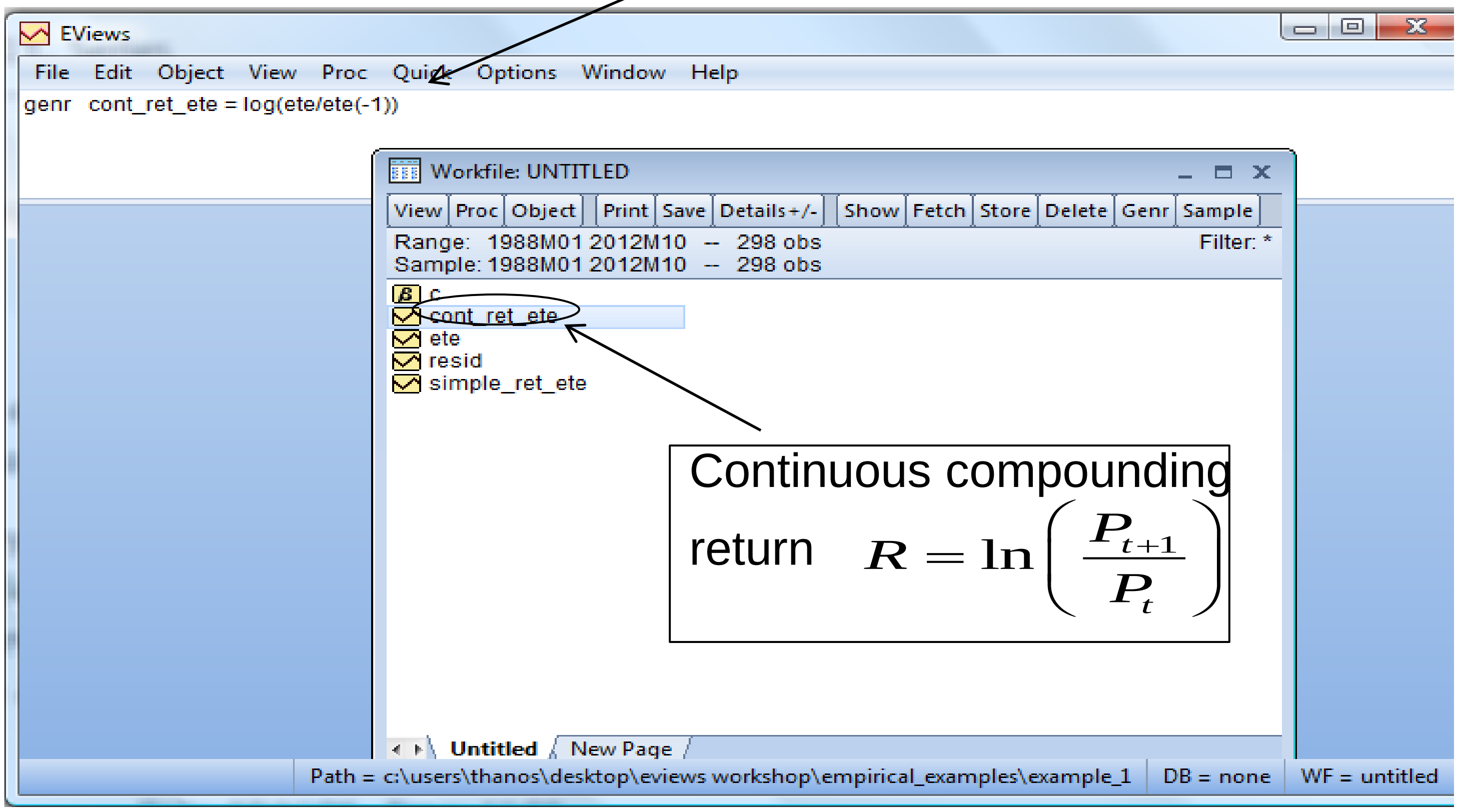


Simple return of ETE :

$$R = \frac{P_{t+1}}{P_t} - 1$$

Empirical Example 3 : *Transformation of data (3)*

On the command window **type** *genr cont_ret_ete = log(ete/ete(-1))*



Continuous compounding return $R = \ln\left(\frac{P_{t+1}}{P_t}\right)$

Continuous compounding or log- returns

Advantages

- Are time additive.
- Assets can be compared since the frequency of compounding return does not play any role.

Disadvantages

- In Investments , the simple portfolio return is a weighted average of the simple returns on the individual assets. $R_{pt} = \sum_{i=1}^{n} w_i R_{it}$
- **However,** this is not feasible for log returns since the log of a sum is not the same as the sum of a log.

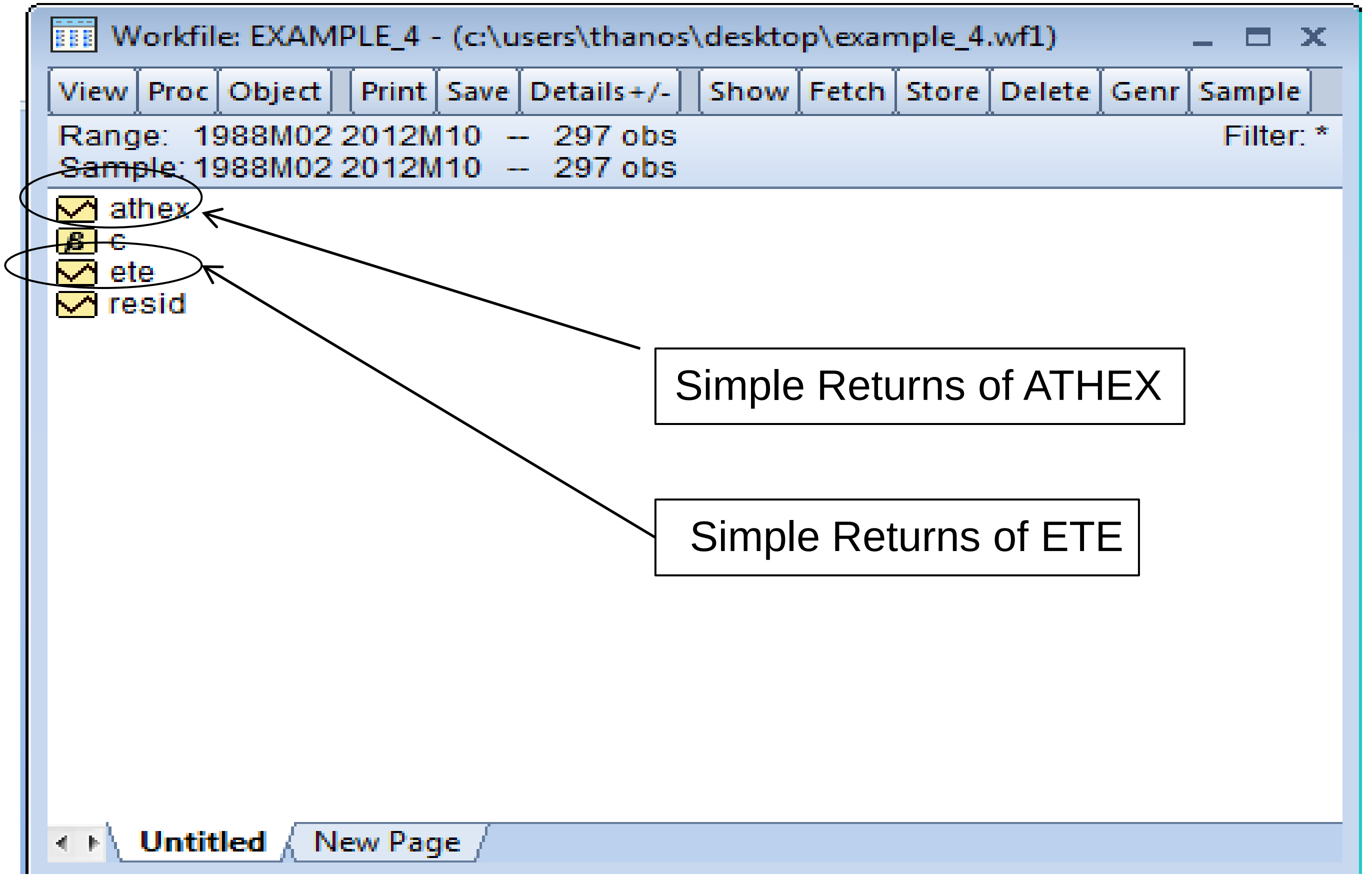
Empirical Example 4 :

Correlation vs. Regression

Empirical Example 4: Correlation vs. Regression(1)

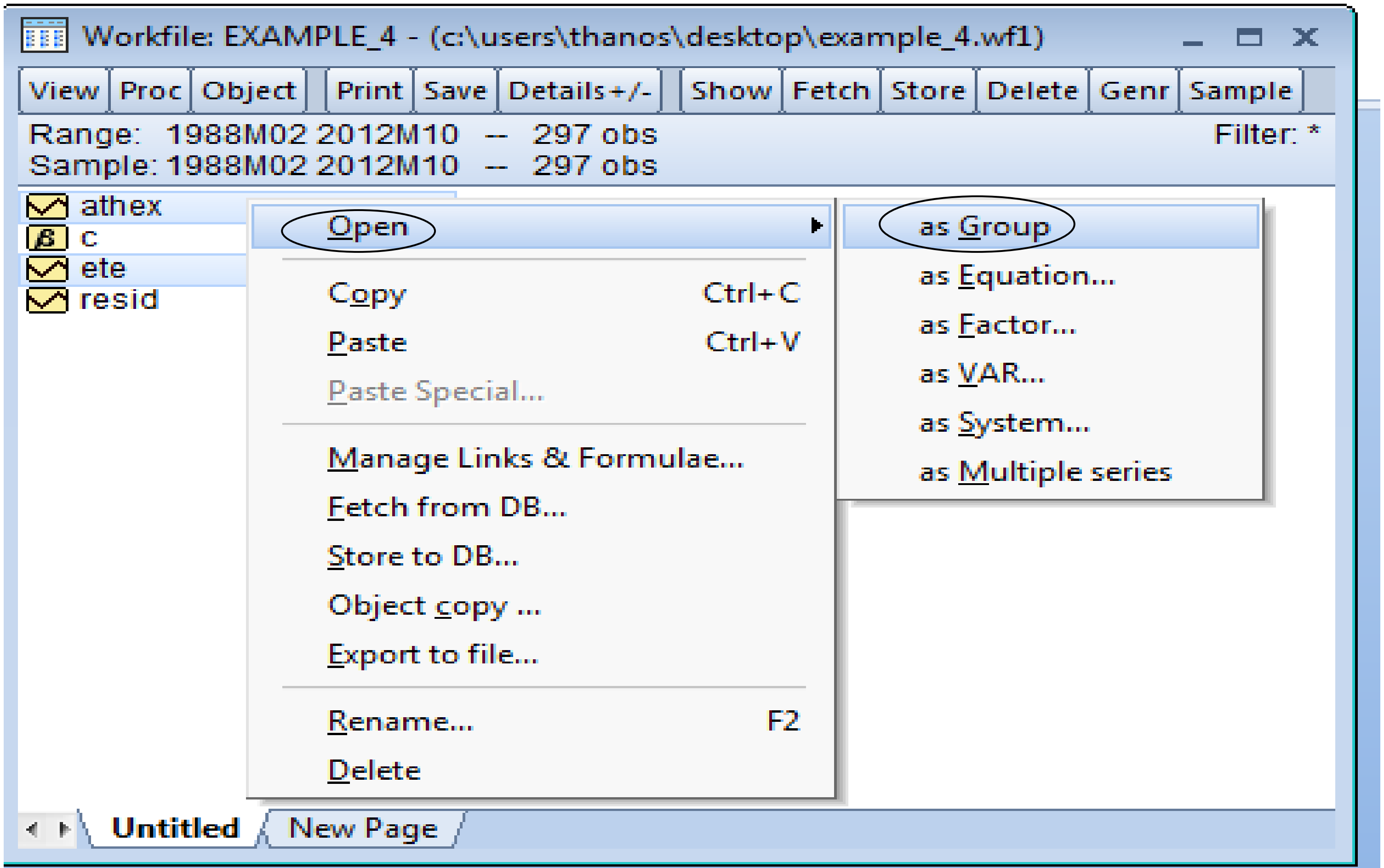
1.Go to folder *Example_4* → *Double click on Example_4.wf1 (eviews workfile)*

- We present the simple returns of ETE and ATHEX from 02/1988 – 10/2012 *(Source : DataStream)*



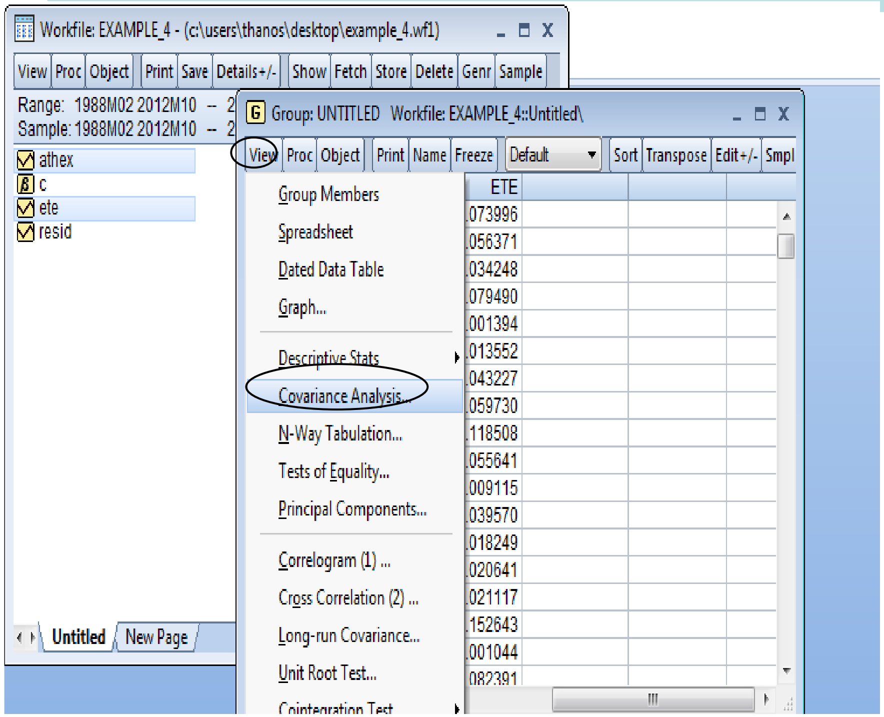
Empirical Example 4: Correlation vs. Regression(2)

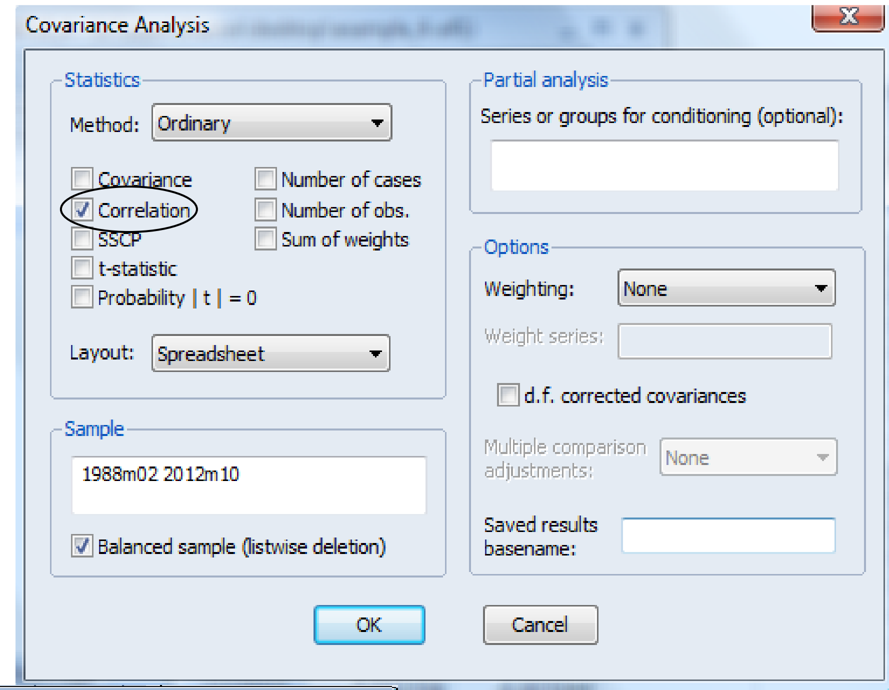
2.Select athex & ete → right click → open → as group

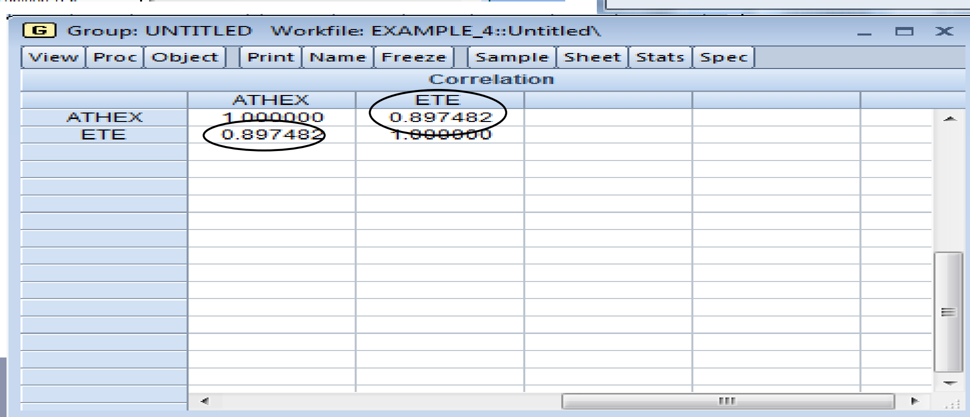


Empirical Example 4: Correlation vs. Regression(3)

3. Define the correlation : View ⟶ Covariance Analysis ⟶ Correlation







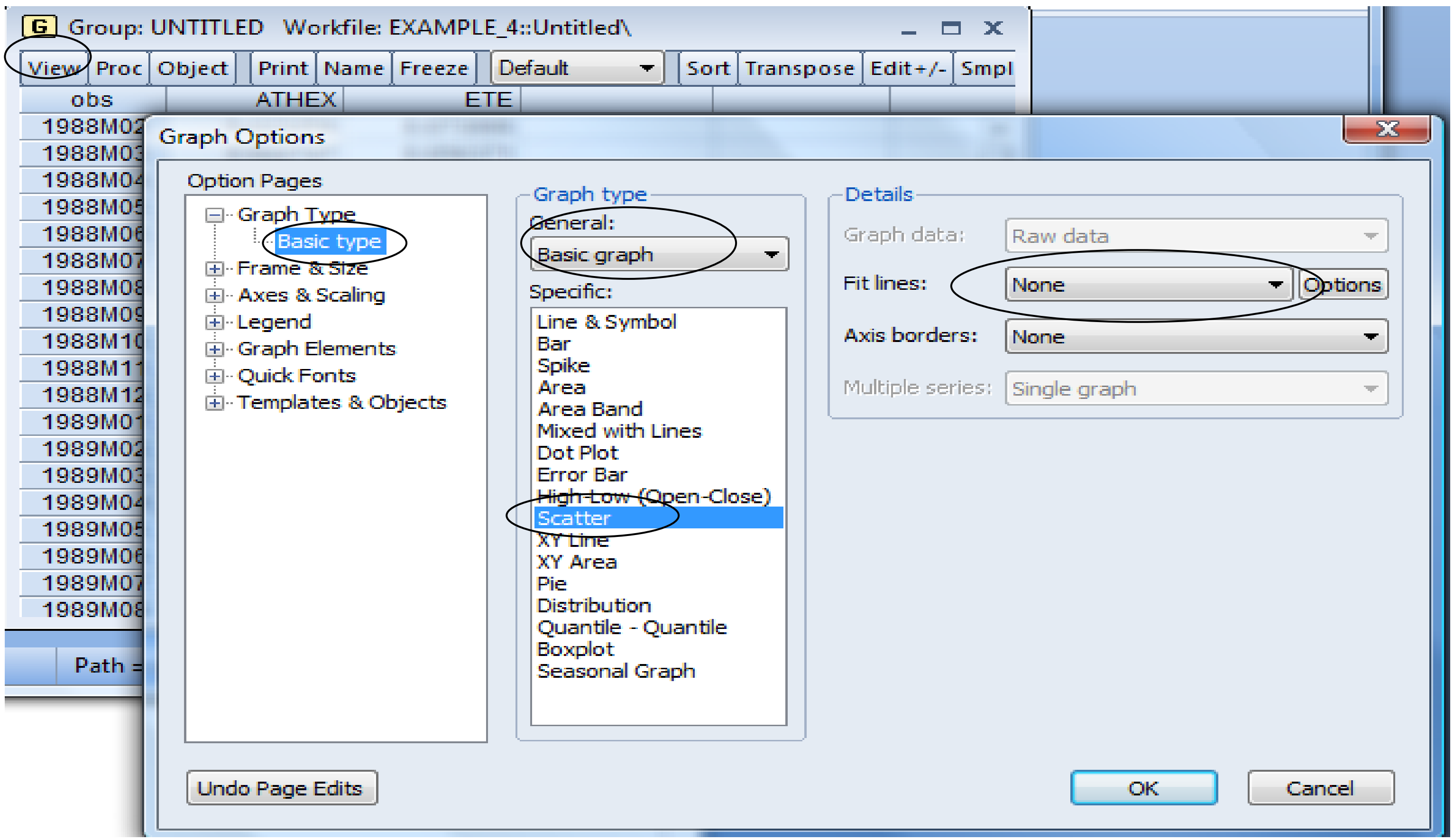
Empirical Example 4: Correlation vs. Regression(4)

- The ***correlation coefficient*** is a **measure of linear association between two variables**. Values of the correlation coefficient are always between -1 and +1.
- A correlation coefficient of **+1** indicates that two variables are **perfectly related** in a **positive** linear sense,
- A correlation coefficient of **-1** indicates that two variables are **perfectly related** in a **negative** linear sense,
- A correlation coefficient of **0** indicates that there is **no linear relationship** between the two variables.
- It is **not** implied that changes in one variable causes changes in the other variable and vice versa.
- Correlation shows an evidence for a linear relationship between two variables.

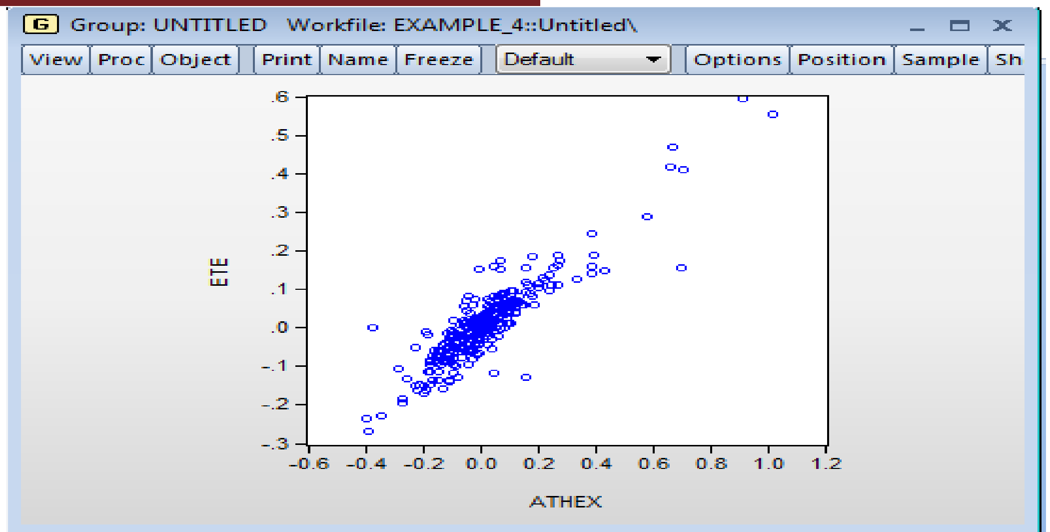
Empirical Example 4: Correlation vs. Regression(5)

4. Scatter plot : Follow step 2 ⟶ View ⟶ Scatter

Scatter plot shows the quantitative relationship between two sets of data.

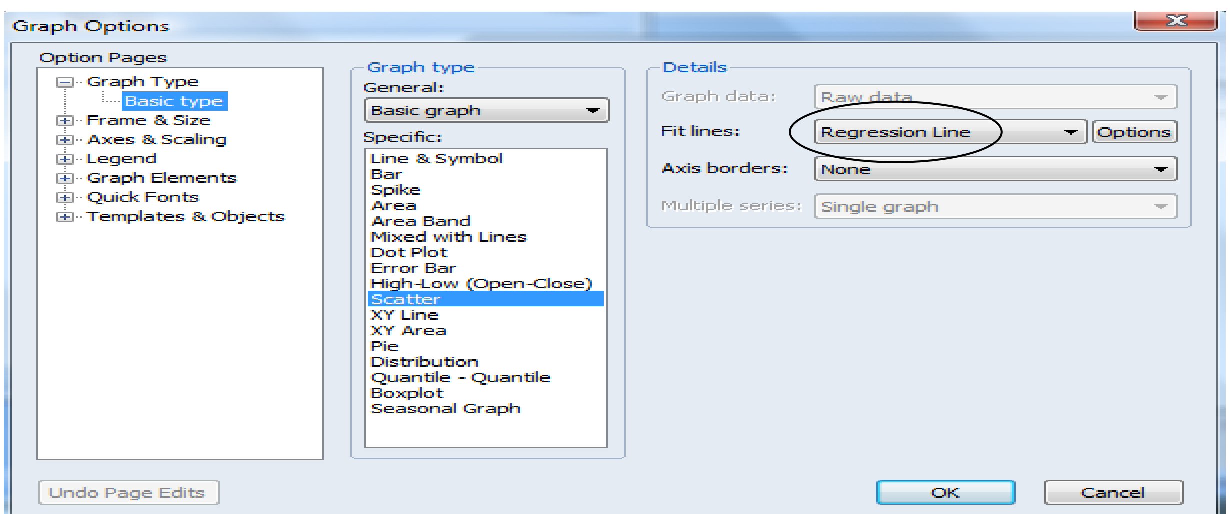


Empirical Example 4: Correlation vs. Regression(6)

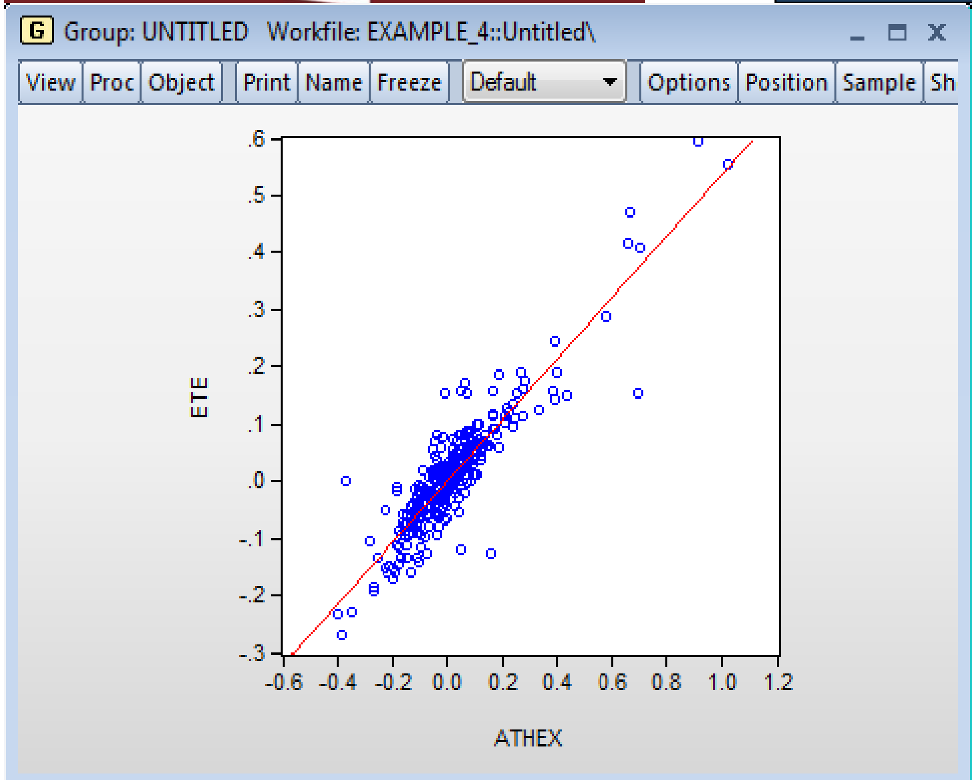


If a scatter diagram shows a linear relationship, as in our case, we would like to review the overall pattern by drawing a line on the scatter diagram.

5. Scatter plot : Go back to ***Graph*** and select ***Regression Line***



Empirical Example 4: Correlation vs. Regression(7)



We need a formal way to draw an optimal line that will be set as close to the points as possible .

Based on the least square method we can define the least squares regression line of Y (*here* ete) on X (*here* athex) :

$$Y = a + bX + error$$

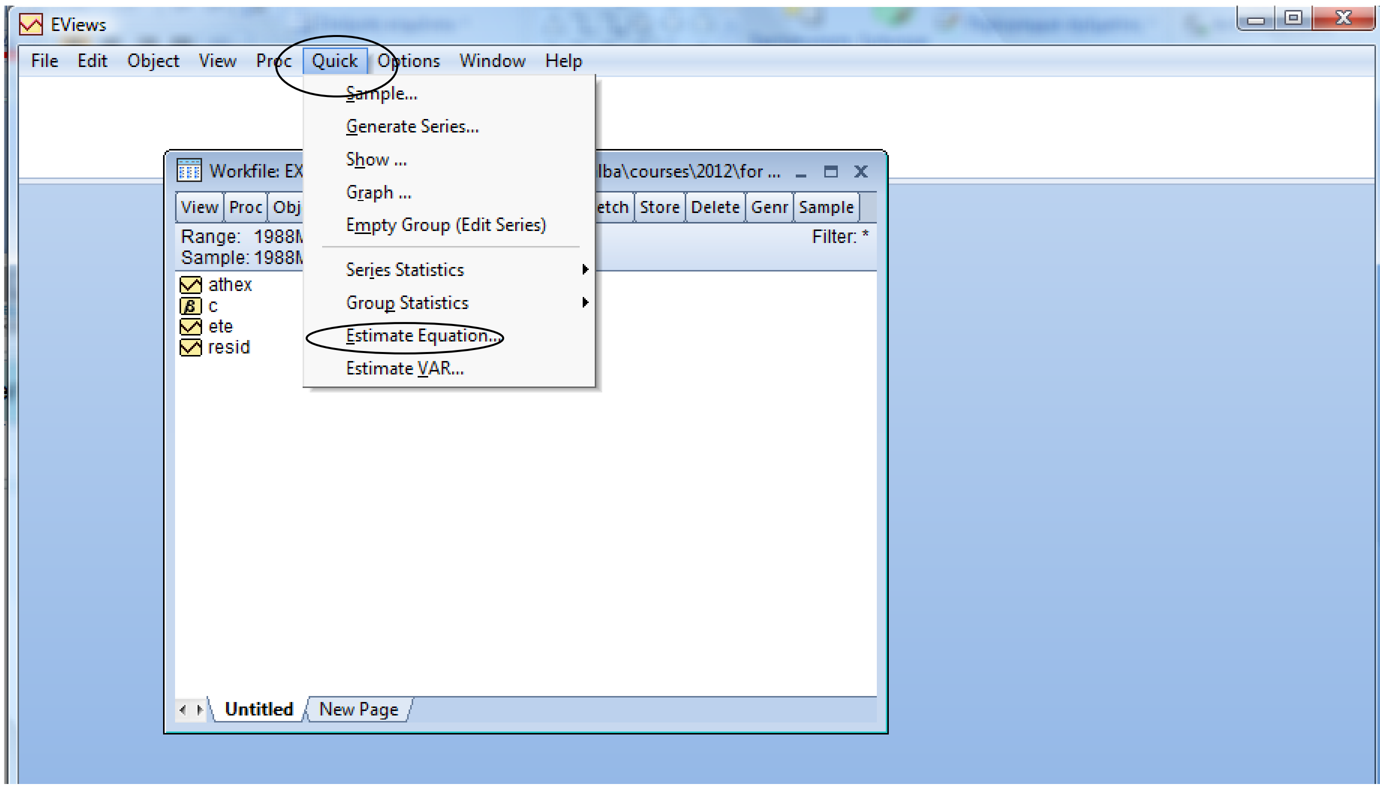
$$ETE = a + bATHEX + error$$

where a: the intercept and b :slope

- In regression, variations in ATHEX **cause changes** in ETE
- Regression describes the relationship between a given variable and one (simple regression) or more other variables

Regression as a tool is more flexible and powerful than correlation

6. How to run a regression : ***Quick*** ⟶ ***Estimate Equation***



Empirical Example 4: Correlation vs. Regression(9)

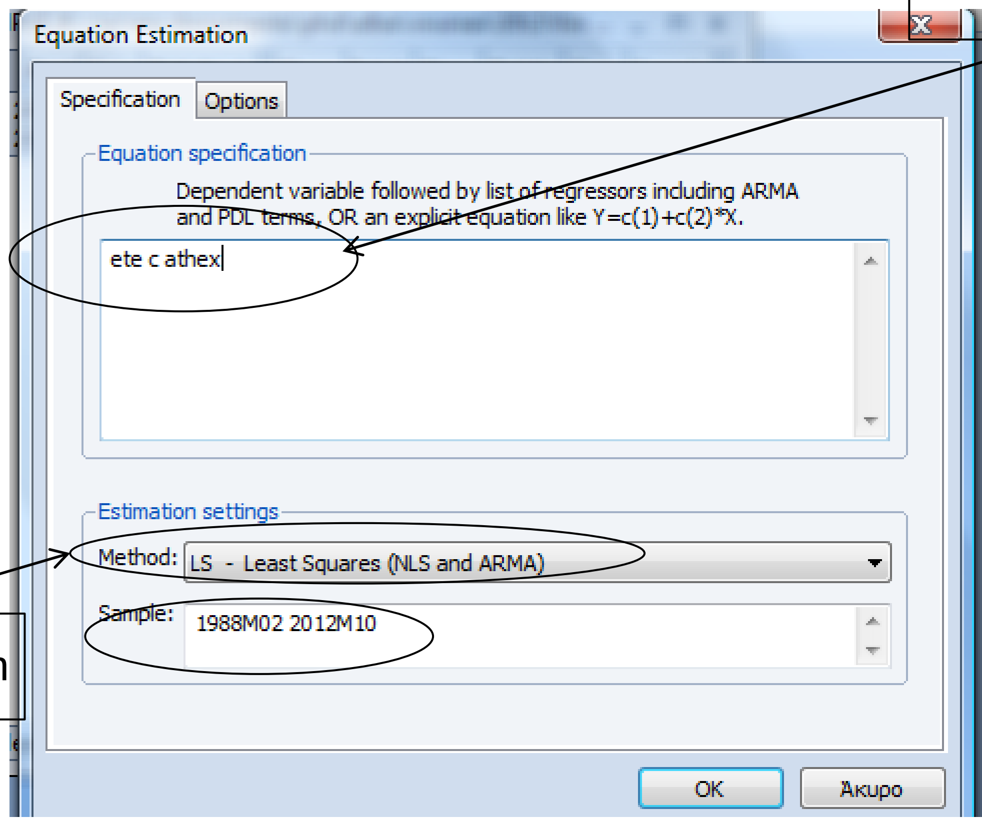
7. Equation Estimation

$$ETE = a + bATHEX + error$$

ete: dependent variable

c : constant term

athex :independent variable



OLS Regression

Empirical Example 4: Correlation vs. Regression(10)

8. Equation Estimation

$$ETE = a + b ATHEX + error$$

Equation: UNTITLED   Workfile: EXAMPLE_4::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: ETE
Method: Least Squares
Date: 11/08/12   Time: 13:51
Sample: 1988M02 2012M10
Included observations: 297

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001584	0.002642	0.599513	0.5493
ATHEX	0.536615	0.015354	34.95005	0.0000

R-squared	0.805474	Mean dependent var	0.009689
Adjusted R-squared	0.804814	S.D. dependent var	0.102678
S.E. of regression	0.045363	Akaike info criterion	-3.341527
Sum squared resid	0.607053	Schwarz criterion	-3.316653
Log likelihood	498.2168	Hannan-Quinn criter.	-3.331569
F-statistic	1221.506	Durbin-Watson stat	2.099498
Prob(F-statistic)	0.000000		

Hypothesis Testing

Critical value approach

p-value approach

Classical Linear Regression Model Estimation

Empirical Example 5:CLR

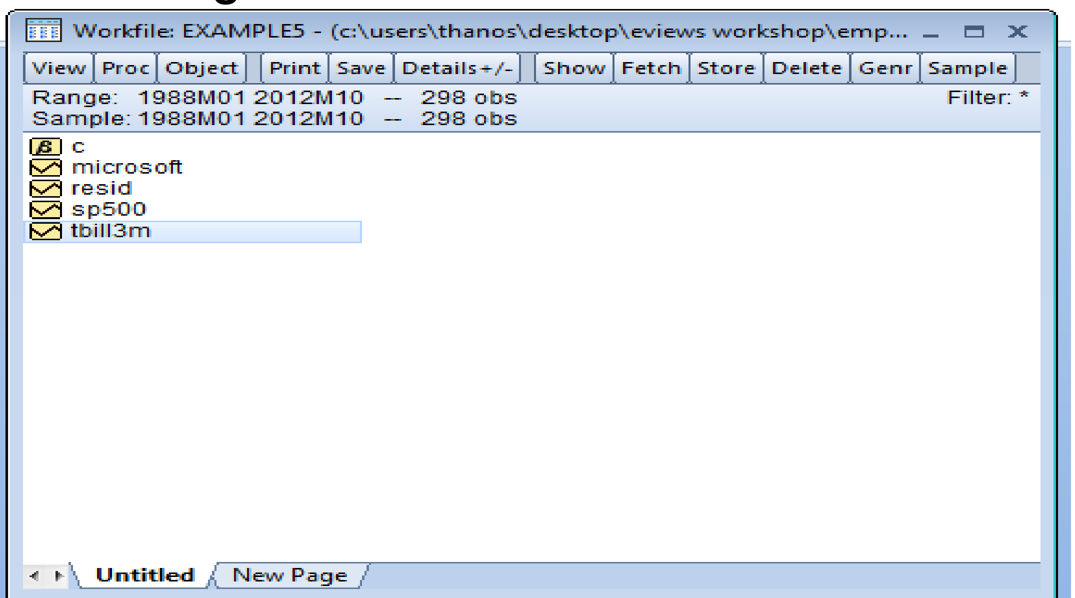
Microsoft stock and S&P 500 Index

1. Import the "empirical_example5.xls" file

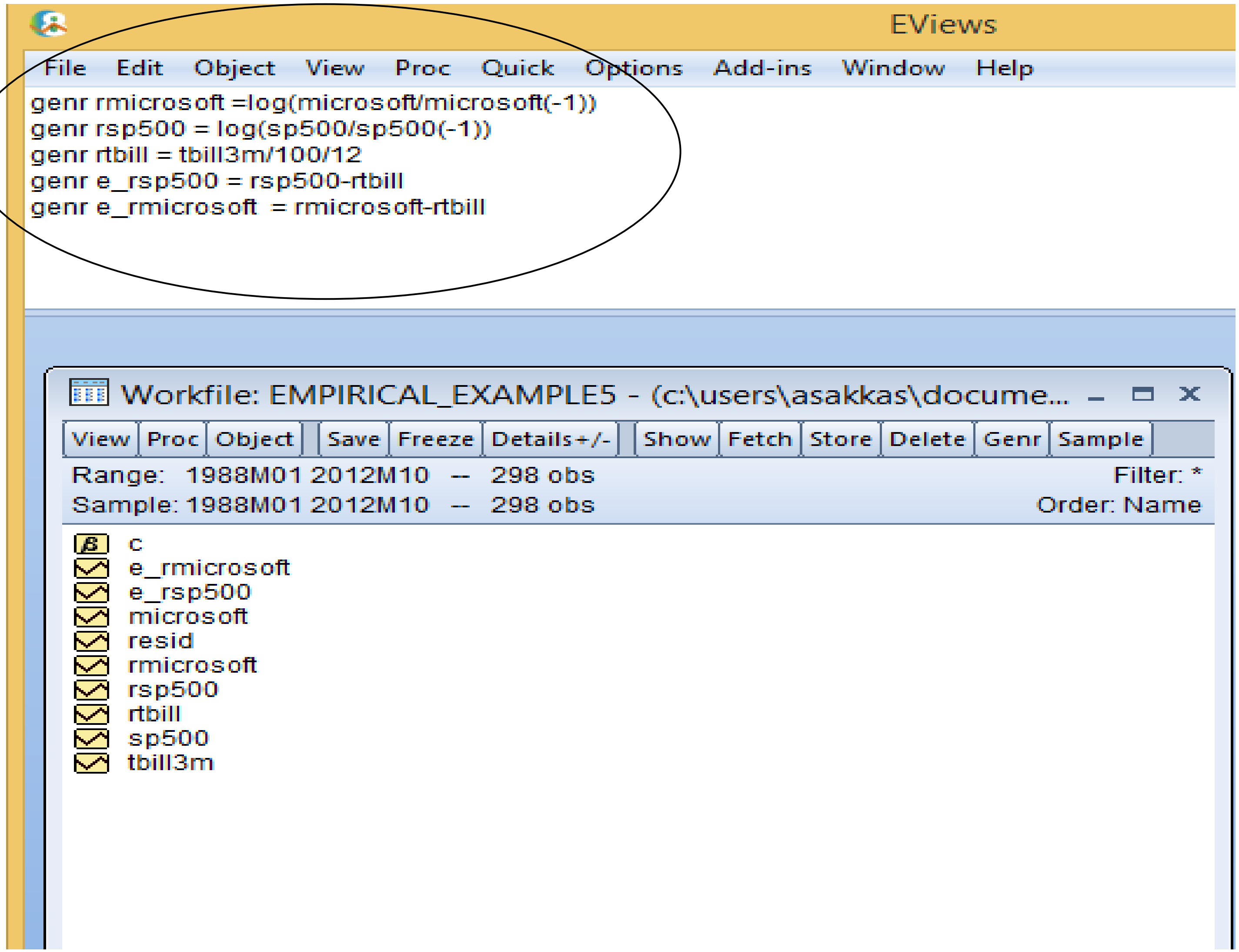
- Monthly RI for Microsoft and S&P500 and T-bill rate for 29/01/21988- 31/10/2012

2. Calculate the log returns for Microsoft and the S&P 500 and the monthly T-bill3m

•Use the **genr** in the command window

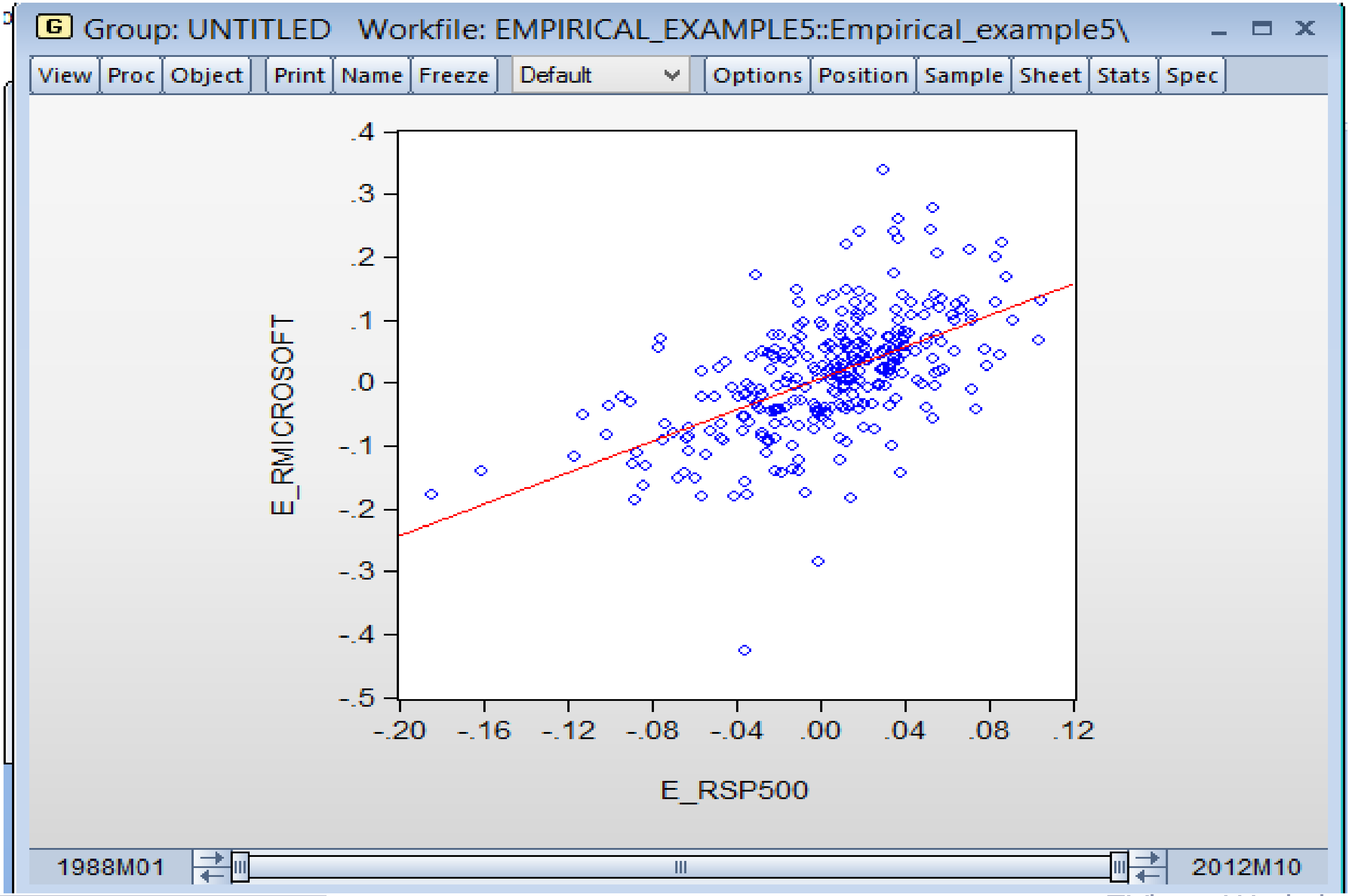


Empirical Example 5:CLR



Empirical Example 5:CLR

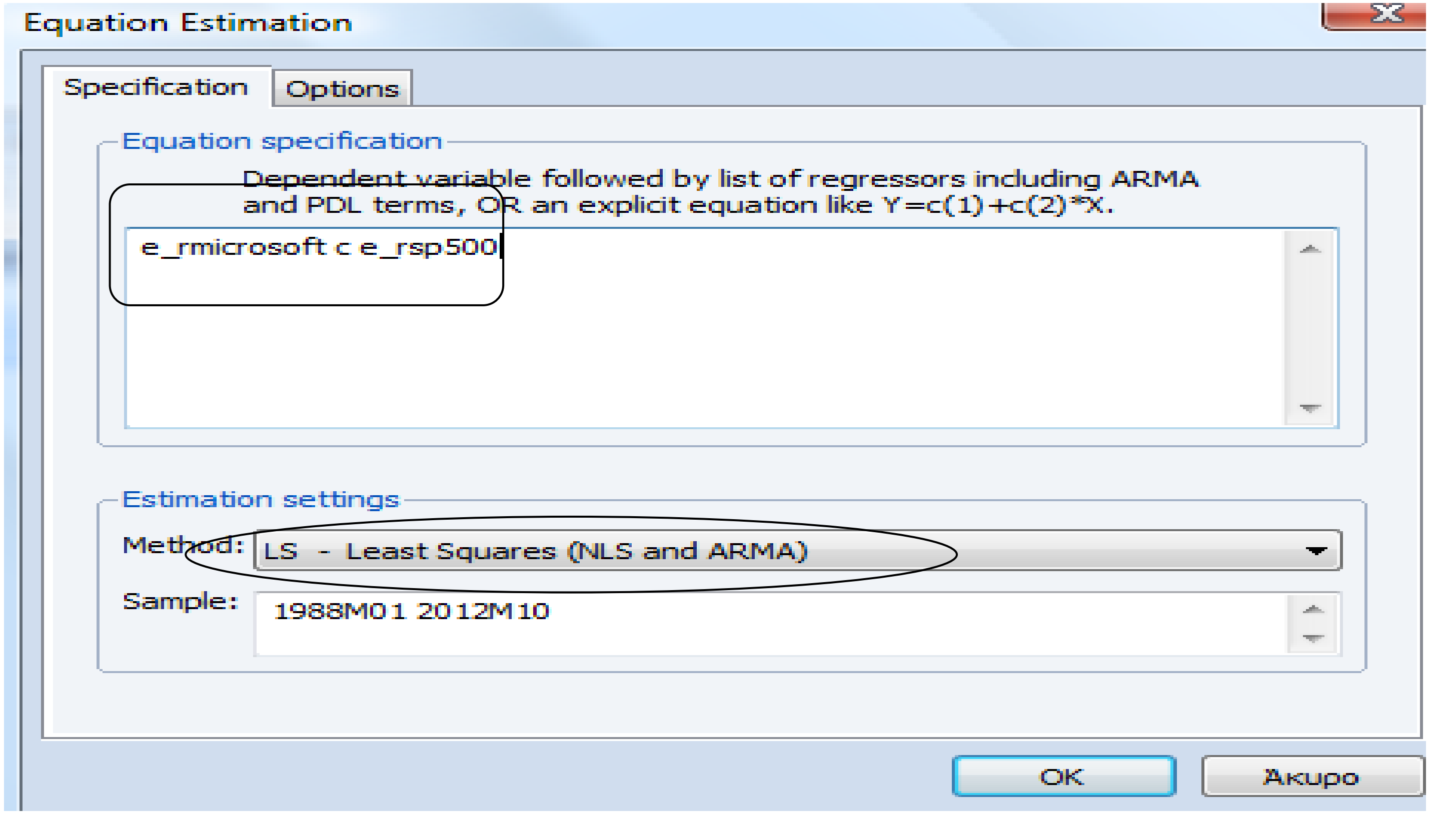
3. Do the Scatter Plot



Empirical Example 5:CLR

3. Estimate the CAPM equation : $\left(R_{Microsoft} - r_f\right)_t = \alpha + \beta\left(R_M - r_f\right)_t + u_t$

- Quick ⟶ Estimate Equation



Empirical Example 5:CLR

Assumptions underlying the CLR model

$E(u_t) = 0$ The errors have zero mean (**Mean Independence**)

$\mathrm{var}(u_t) = \sigma^2$ The variance of the errors is constant (**Homoskedasticity**)

$\mathrm{cov}(u_i, u_j) = 0$ The errors are linearly independent of one other

$\mathrm{cov}(u_t, x_t) = 0$ There is no relationship between the error and the corresponding variate x

$u_t \sim N(0, \sigma^2)$ The errors are normally distributed (**Normality**)

Empirical Example 5:CLR

Equation: UNTITLED   Workfile: EMPIRICAL_EXAMPLE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15   Time: 14:54
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
E_RSP500	1.251419	0.105816	11.82641	0.0000

R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

Any set of (X,Y) *are specific to the sample used in their estimation. It would be desirable to have an idea of how reliable/ precise these estimates are. Thus we need an estimate given by the* ***standard error.***

The standard error

$$SE\left(\hat{a}\right)$$

The standard error

$$SE\left(\hat{\beta}\right)$$

The beta coefficient (the slope coefficient) estimate is 1.25

$$\hat{\beta} = \frac{Cov\left(ExcR_{Microsoft}, ExcR_{SP500}\right)}{Var\left(ExcR_{SP500}\right)}$$

The a coefficient estimate is 0.006

$$\hat{\alpha} = E(Y) - \hat{\beta}\,E(X)$$

Standard error of regression (or Root Mean Square error)

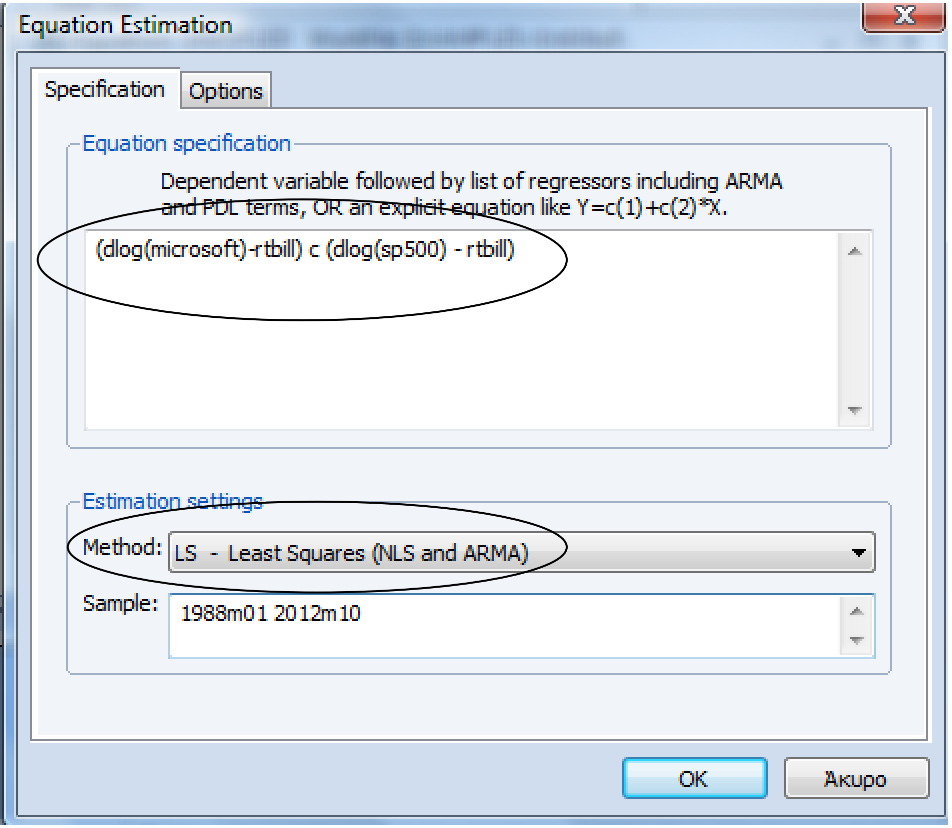
$$s = \sqrt{\frac{1}{n-2}\sum_i \hat{u}_i^2}$$

Measures the standard deviation of Y around the estimated regression line

Empirical Example 5:CLR

An other way…

- Quick ⟶ Estimate Equation



File Edit Object View Proc Quick Options Add-ins Window Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: DLOG(MICROSOFT)-RTBILL
Method: Least Squares
Date: 11/13/15 Time: 14:57
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
DLOG(SP500)-RTBILL	1.251419	0.105816	11.82641	0.0000

R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

Empirical Example 5:CLR

4.1. Hypothesis Testing – Critical value approach

Two –sided Test

$$H_0 : \alpha = 0$$
$$H_A : \alpha \neq 0$$

$$H_0 : \beta = 0$$
$$H_A : \beta \neq 0$$

Critical value approach

a = 5% significance level

$$test\ statistic = \frac{\hat{a} - a}{SE\left(\hat{a}\right)}$$

We do **not** reject the Null Hypothesis for a ; thus a is **insignificant**

$$test\ statistic = \frac{\hat{\beta} - \beta}{SE\left(\hat{\beta}\right)}$$

We reject the Null Hypothesis for b ; thus b is **significant**

Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15 Time: 14:54
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
E_RSP500	1.251419	0.105816	11.82641	0.0000

R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

Empirical Example 5:CLR

4.2. Hypothesis Testing – Confidence interval approach

Two –sided Test

$$H_0: \alpha = 0$$
$$H_A: \alpha \neq 0$$

Confidence interval approach →

a = 5% significance level

$$\hat{a} \pm t_{crit} SE(\hat{a})$$

(-0.002,0.014)
We do **not** reject the Null Hypothesis for a ; thus a is **Insignificant,** since 0 lies within confidence interval

$$H_0: \beta = 0$$
$$H_A: \beta \neq 0$$

$$\hat{\beta} \pm t_{crit} SE(\hat{\beta})$$

(1.05,1.45)
We reject the Null Hypothesis for b ; thus b is **significant,** since 0 does **not** lie within confidence interval

Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15 Time: 14:54
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
E_RSP500	1.251419	0.105816	11.82641	0.0000

R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

Empirical Example 5:CLR

4.3. Hypothesis Testing – p-value approach

Two –sided Test

$$H_0 : \alpha = 0$$
$$H_A : \alpha \neq 0$$

$$H_0 : \beta = 0$$
$$H_A : \beta \neq 0$$

P-value approach

a = 5% significance level

p-value is termed as the

“plausibility” of the Null Hypothesis;

the smaller the p-value, the less plausible is the null hypothesis.

Is the largest significance level at which we fail to reject

the null hypothesis.

Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15 Time: 14:54
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
E_RSP500	1.251419	0.105816	11.82641	0.0000

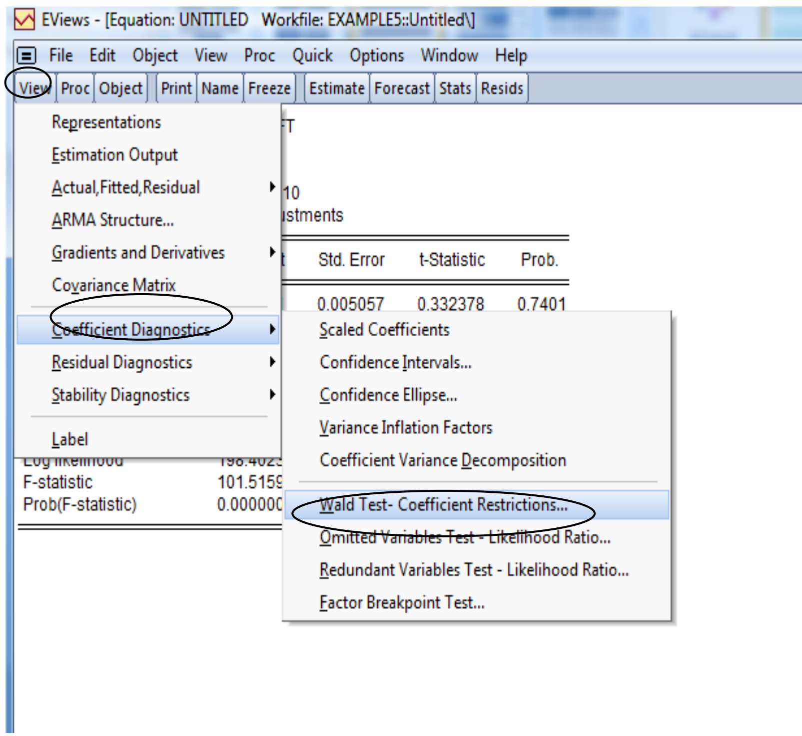
R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

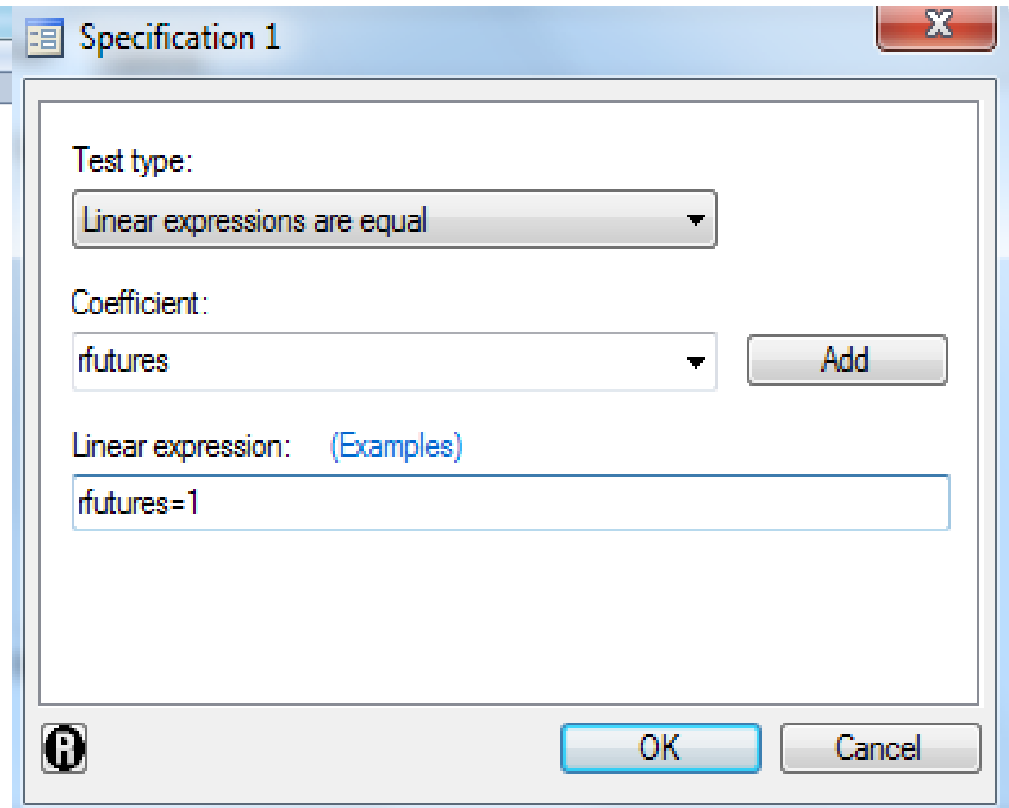
Empirical Example 5:CLR

5.Suppose now we want to test the null hypothesis that

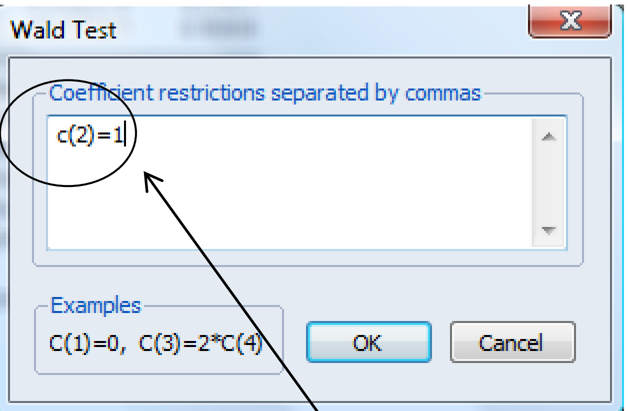
$$H_0 : \beta = 1$$
$$H_A : \beta \neq 1$$

5.1 Go to View ⟶ Coefficient Diagnostics ⟶ Wald Test Coefficient Restrictions

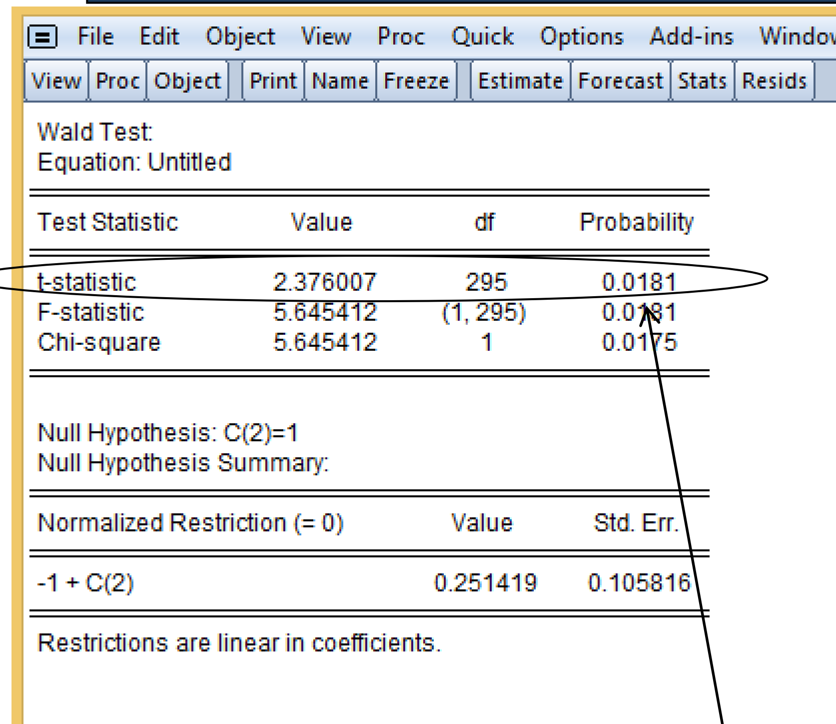




Empirical Example 5:CLR



c(2) denotes the beta coeff.



File Edit Object View Proc Quick Options Add-ins Window

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	2.376007	295	0.0181
F-statistic	5.645412	(1, 295)	0.0181
Chi-square	5.645412	1	0.0175

Null Hypothesis: C(2)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(2)	0.251419	0.105816

Restrictions are linear in coefficients.

We reject the Null Hypothesis

- F-statistic and X-square statistic are identical since we are testing a hypothesis about only one parameter

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ — ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Empirical Example 5:CLR

6. **Testing Multiple Hypothesis : The F- test**

The t-test was used to test single- hypothesis (one coefficient hypothesis)

For more than one parameter hypothesis we use F - statistic

$$t - statictic = \frac{RRSS - URSS}{URSS} \times \frac{T-k}{m} \sim F(m, T-k)$$

$$Z \sim t_{T-k}$$
$$Z^2 \sim {t^2}_{T-k} \sim F(1, T-k)$$

- URSS: Residual sum of squares from unrestricted regression
- RRSS : Residual sum of squares from restricted regression
- m : number of restrictions
- T : number of observations
- k : number of regressors in unrestricted regression

Reject the Null when $F \succ t_{crit}$

Empirical Example 5:CLR

Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15 Time: 14:54
Sample (adjusted): 1988M02 2012M10
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006634	0.004571	1.451431	0.1477
E_RSP500	1.251419	0.105816	11.82641	0.0000

R-squared	0.321627	Mean dependent var	0.012261
Adjusted R-squared	0.319327	S.D. dependent var	0.094962
S.E. of regression	0.078346	Akaike info criterion	-2.248650
Sum squared resid	1.810743	Schwarz criterion	-2.223776
Log likelihood	335.9245	Hannan-Quinn criter.	-2.238692
F-statistic	139.8639	Durbin-Watson stat	2.175777
Prob(F-statistic)	0.000000		

$$H_0 : \beta = 0$$
$$H_A : \beta \neq 0$$

We reject the Null Hypothesis for levels of significance 1%, 5% and 10%, since p-value of F – statistic is 0.000. **In this case F-test statistic is equal to the square of the slope t-stat.**

Wald Test

Coefficient restrictions separated by commas

C(1)=1,C(2)=1

Examples

C(1)=0, C(3)=2*C(4)

OK Cancel

File Edit Object View Proc Quick Options Add-ins W

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Re

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	23821.68	(2, 295)	0.0000
Chi-square	47643.37	2	0.0000

Null Hypothesis: C(1)=1,C(2)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(1)	-0.993366	0.004571
-1 + C(2)	0.251419	0.105816

Restrictions are linear in coefficients.

$$H_0 : \alpha = 1 \quad and \quad \beta = 1$$
$$H_A : \alpha \neq 1 \quad or \quad \beta \neq 1$$

We reject the joint Null Hypothesis for levels of significance 1%, 5% and 10%.

- The F-version is adjusted for small sample bias and should be used when the regression is estimated using a small sample

Multiple Linear Regression Model Estimation

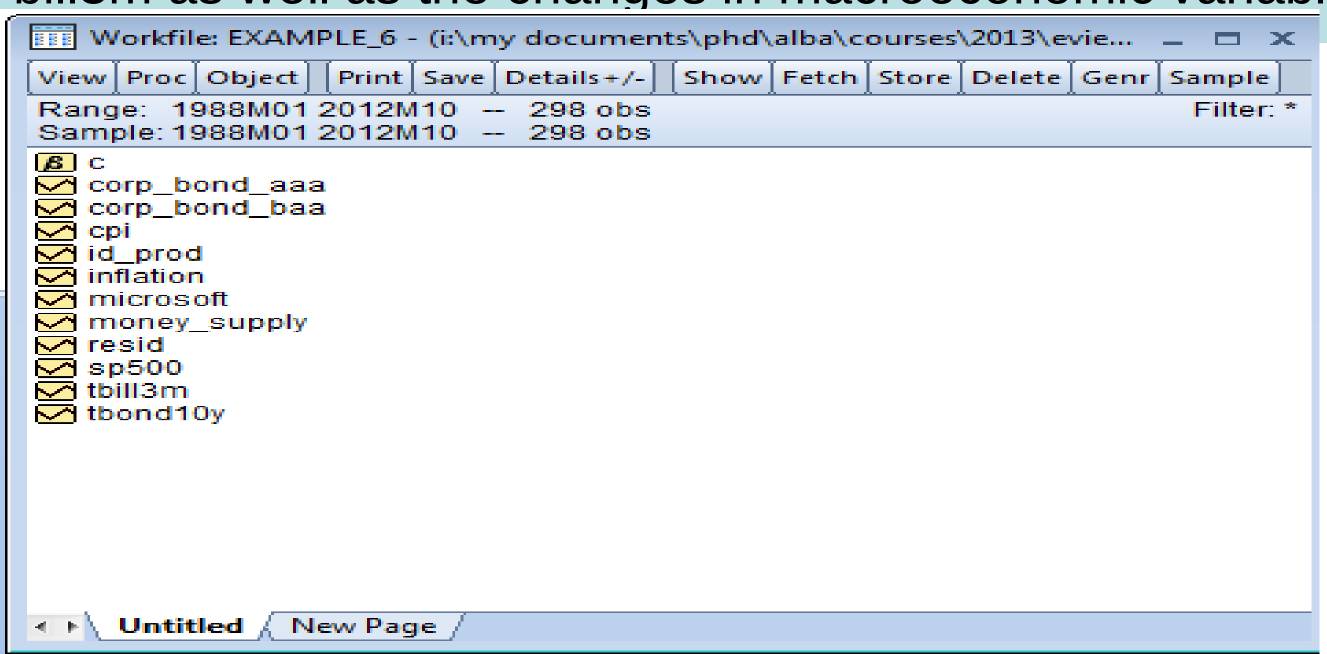
Empirical Example 6:MLR

APT model : Microsoft stock - Market + Macroeconomic and financial variables

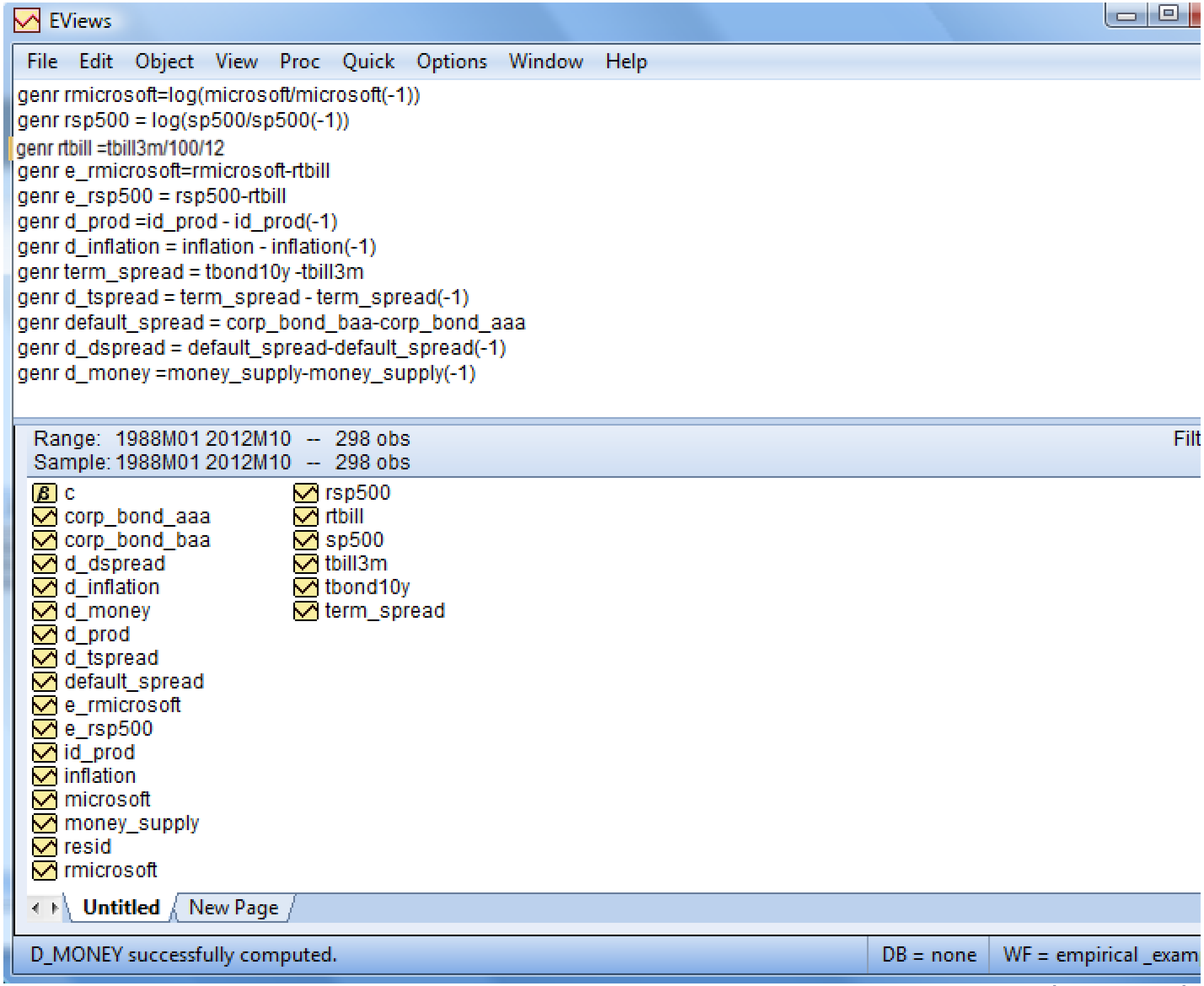
1. Open “empirical_example6” eviews workfile

- Monthly RI for Microsoft and S&P500,T-bill rate CPI, money supply, industrial production, corporate bond yields with rates AAA and BAA, yield of US Treasury Bond10y for 29/01/21988- 31/10/2012

2. Calculate the log returns for Microsoft and the S&P 500 and the monthly T-bill3m as well as the changes in macroeconomic variables



Empirical Example 6:MLR



Empirical Example 6:MLR

Quick ⟶ Estimate equation

$$\left(R_{Microsoft}-r_f\right)_t=\alpha+\beta_1\left(R_M-r_f\right)_t+\beta_2(d\_prod)+\beta_3(d\_money)+$$
$$\beta_4(d\_\text{inflation})+\beta_5(d\_tspread)+\beta_6(d\_dspread)+u_t$$

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_6::Untitl...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 11/13/15 Time: 15:21
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009011	0.005125	1.758323	0.0798
E_RSP500	1.294789	0.112600	11.49906	0.0000
D_PROD	-0.011834	0.008256	-1.433299	0.1529
D_MONEY	-0.000189	0.000335	-0.565633	0.5721
D_INFLATION	1.320730	1.675276	0.788366	0.4311
D_TSPREAD	0.013109	0.015505	0.845463	0.3986
D_DSPREAD	0.043376	0.039935	1.086151	0.2783

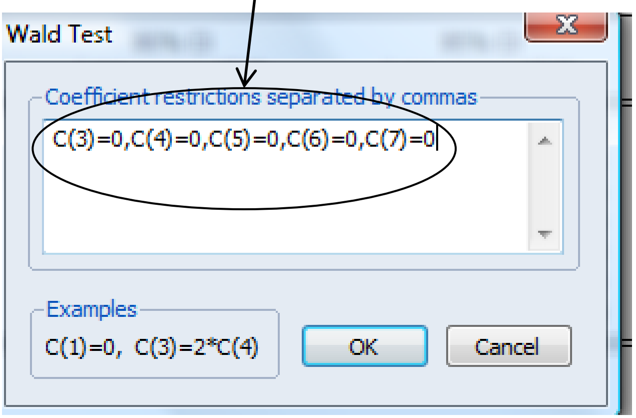
R-squared	0.332096	Mean dependent var	0.012106
Adjusted R-squared	0.318229	S.D. dependent var	0.095085
S.E. of regression	0.078511	Akaike info criterion	-2.227793
Sum squared resid	1.781386	Schwarz criterion	-2.140521
Log likelihood	336.7134	Hannan-Quinn criter.	-2.192851
F-statistic	23.94944	Durbin-Watson stat	2.196550
Prob(F-statistic)	0.000000		

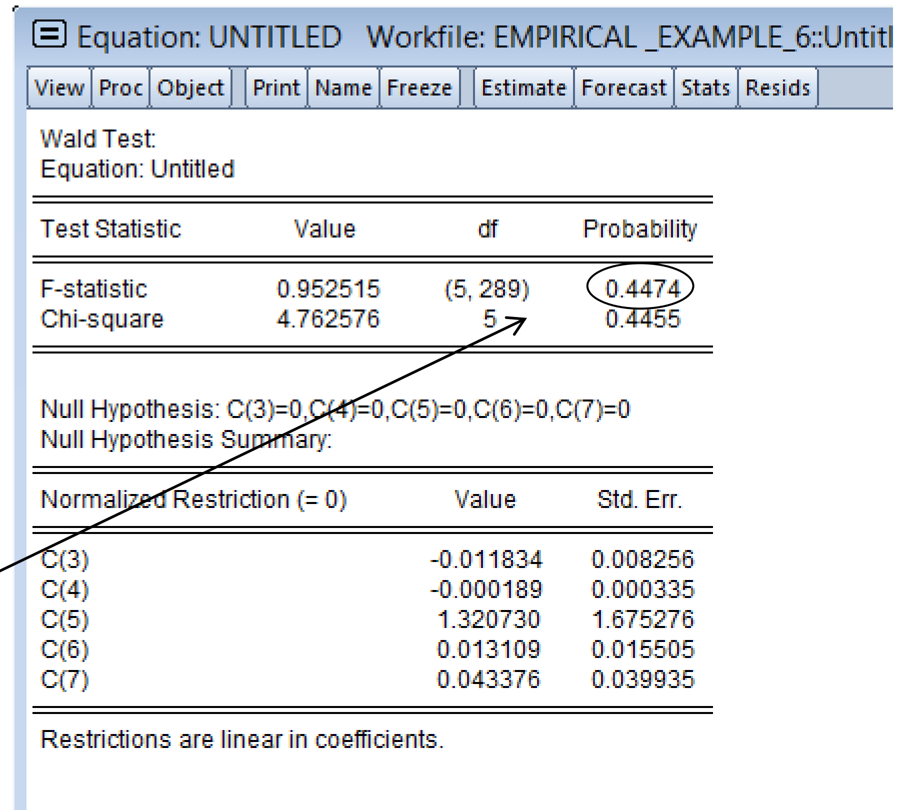
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ — ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Empirical Example 6:MLR

$$H_0: \beta_2 = 0 and \beta_3 = 0 and \beta_4 = 0 and \beta_5 = 0 and \beta_6 = 0$$

$$H_A: \beta_2 \neq 0 or \beta_3 \neq 0 or \beta_4 \neq 0 or \beta_5 \neq 0 or \beta_6 \neq 0$$





Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE_6::Untitl

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	0.952515	(5, 289)	0.4474
Chi-square	4.762576	5	0.4455

Null Hypothesis: C(3)=0,C(4)=0,C(5)=0,C(6)=0,C(7)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(3)	-0.011834	0.008256
C(4)	-0.000189	0.000335
C(5)	1.320730	1.675276
C(6)	0.013109	0.015505
C(7)	0.043376	0.039935

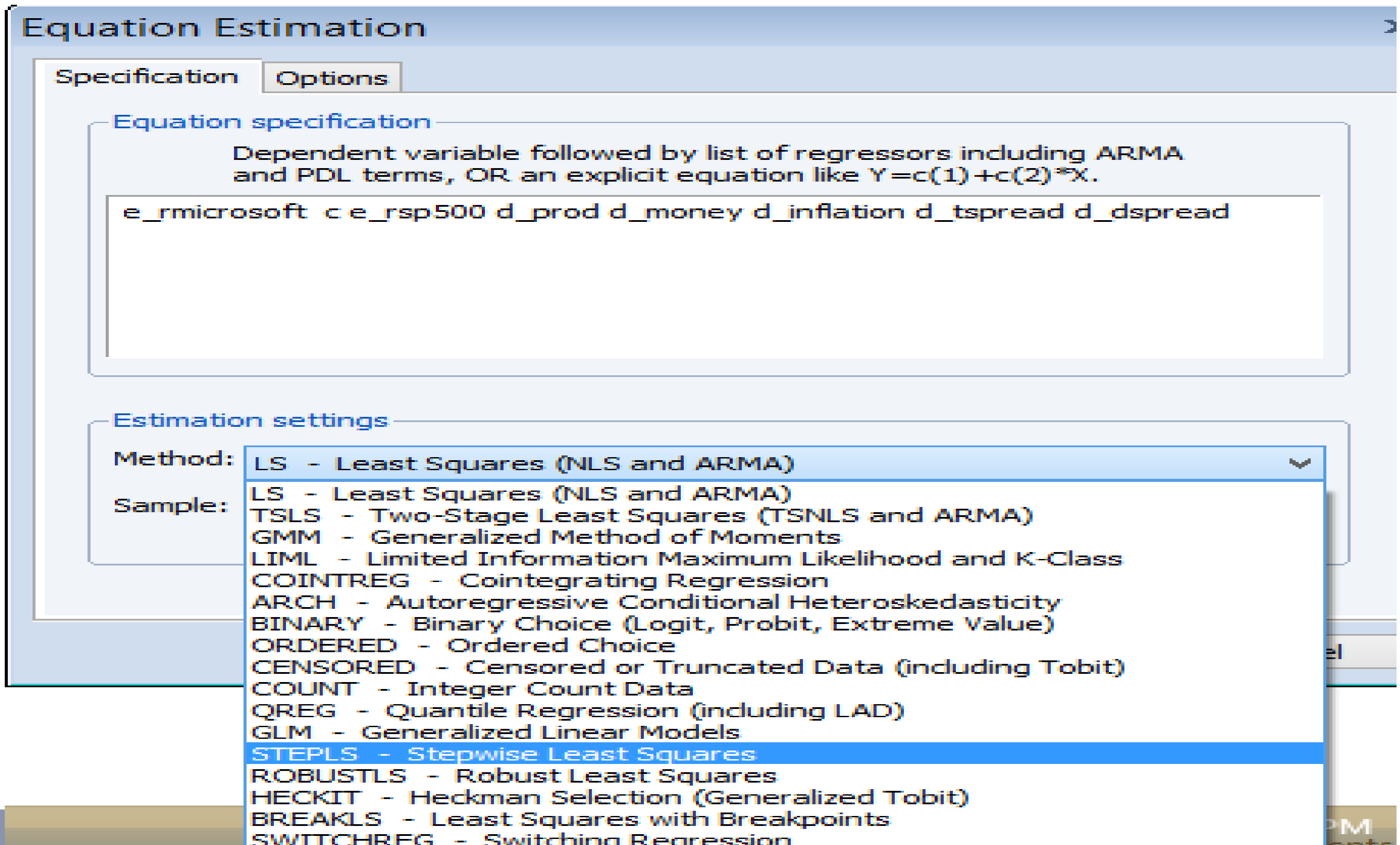
Restrictions are linear in coefficients.

Do not Reject the joint Null Hypothesis

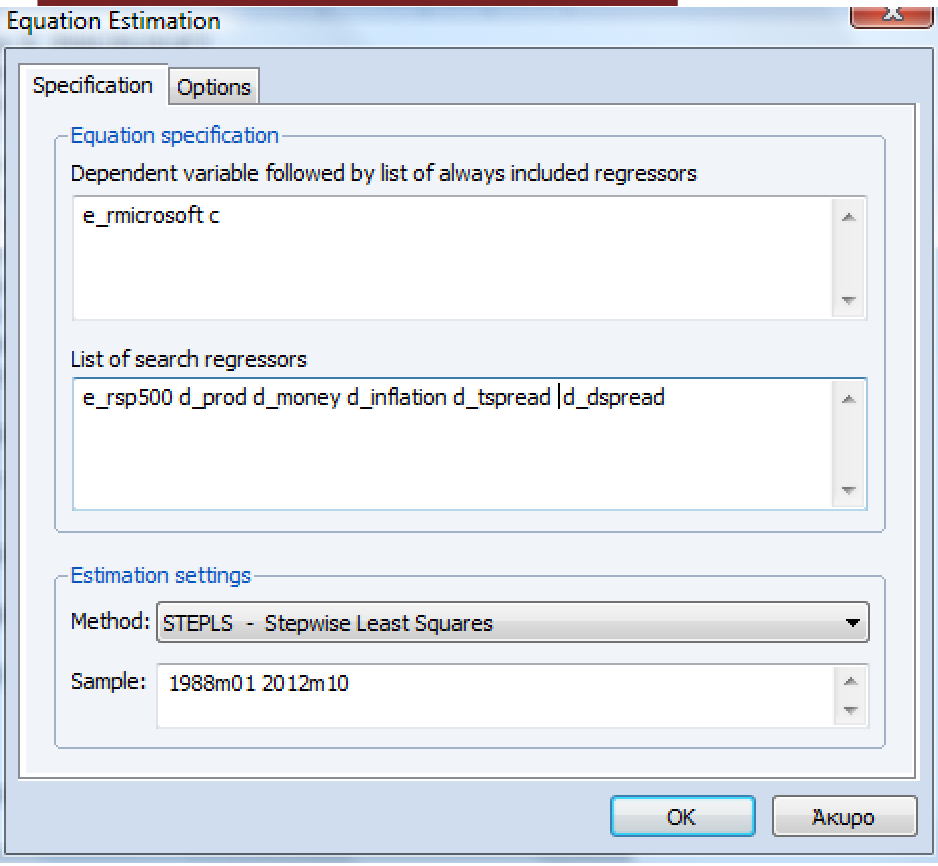
Empirical Example 6:MLR

Stepwise Regression

It constitutes an automatic variable selection procedure which chooses jointly the most important independent variables from a set of candidate variables



Empirical Example 6:MLR



Dependent Variable: E_RMICROSOFT
Method: Stepwise Regression
Date: 11/13/15 Time: 15:27
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
Number of always included regressors: 1
Number of search regressors: 6
Selection method: Stepwise forwards
Stopping criterion: p-value forwards/backwards = 0.5/0.5

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
C	0.007848	0.004689	1.673807	0.0952
E_RSP500	1.295735	0.112455	11.52225	0.0000
D_PROD	-0.011036	0.008125	-1.358237	0.1754
D_TSPREAD	0.013468	0.015474	0.870370	0.3848
D_DSPREAD	0.041593	0.039764	1.045985	0.2964
D_INFLATION	1.227519	1.665196	0.737162	0.4616

R-squared	0.331356	Mean dependent var	0.012106
Adjusted R-squared	0.319828	S.D. dependent var	0.095085
S.E. of regression	0.078419	Akaike info criterion	-2.233444
Sum squared resid	1.783358	Schwarz criterion	-2.158639
Log likelihood	336.5497	Hannan-Quinn criter.	-2.203493
F-statistic	28.74275	Durbin-Watson stat	2.194578
Prob(F-statistic)	0.000000		

Selection Summary

Added E_RSP500
Added D_PROD
Added D_TSPREAD
Added D_DSPREAD
Added D_INFLATION

Empirical Example 6:MLR

Goodness of fit measures

R^2 A measure of how well the regression model fits the data or
how well does the model containing the propose independent
variables which explain the variation in the dependent variable

$$R^2 adjusted = 1 - \left[\frac{T-1}{T-k} \left(1 - R^2\right) \right]$$

Decision making tool for determining whether a given variable should be included in a regression model : if R-squared adjusted rise include the variable

The relationship between regression F and R^2

$$F - stat = \frac{R^2 (T-k)}{1 - R^2 (k-1)}$$

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

EViews Workshop II

Agenda (1)

I. Testing for heteroskedasticity

- Wald Test
- Breusch-Pagan- Godfrey Test

II. Testing for serial correlation

- Durbin- Watson Test
- Cohrane - Orcutt Test
- Breusch-Godfrey Test

III. Testing for non normality

- Jarque – Bera Test
- Dummies

IV. Testing for multicollinearity

- Correlation Matrix
- Add/Remove of Explanatory variable

Agenda (2)

V. Testing for linear relationship between Y and X

- Ramsey RESET Test

VI. Testing for stability

- Chow Test

VII. Univariate Time Series Modelling of US Home Prices

- Autoregressive Process (AR)
- Moving Average Process (MA)
- ARMA model

VIII. Stationarity

Assumptions underlying the CLR model

$E(u_t)=0$ The errors have zero mean (Mean Independence)

$\mathrm{var}(u_t)=\sigma^2$ The variance of the errors is constant (Homoskedasticity)

$\mathrm{cov}(u_i,u_j)=0$ The errors are linearly independent of one other

$\mathrm{cov}(u_t,x_t)=0$ There is no relationship between the error and the corresponding variate x

$u_t \sim N(0,\sigma^2)$ The errors are normally distributed (Normality)

Violation of one of the above assumptions may lead to

1. Biased coefficient estimates
2. Biased standard errors
3. Inappropriate distributions

Thus, we need to test and solve for these violations

Statistical Distributions for diagnostic tests

The tests that detect any violation are based on the calculation of test statistic

LM test

- Chi-squared distribution
- df equal to the number of restrictions

Wald Test

- F-distribution
- df equal to $(m, T-k)$

$$\frac{\chi^2(m)}{m} \overset{A}{\sim} F(m, T-k)$$

1st Assumption: Mean Independence

$E(u_t) = 0$ The errors have zero mean (Mean Independence)

•If we include a constant term in the regression equation, this assumption **will never be** violated.

•If financial theory suggest a model **without** intercept then

a. R-squared may be negative (the sample average of y explains more of the variation in y than the explanatory variables x).

b. Severe biases in slope coefficients.

Testing for Heteroskedasticity

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

2nd Assumption: Homoskedasticity

$var(u_t)=\sigma^2$ The variance of the errors is constant (Homoskedasticity)

•You can plot the residuals with an explanatory variable; however, it is difficult to detect the presence or not of heteroskedasticity, since we do not know the form of the latter.

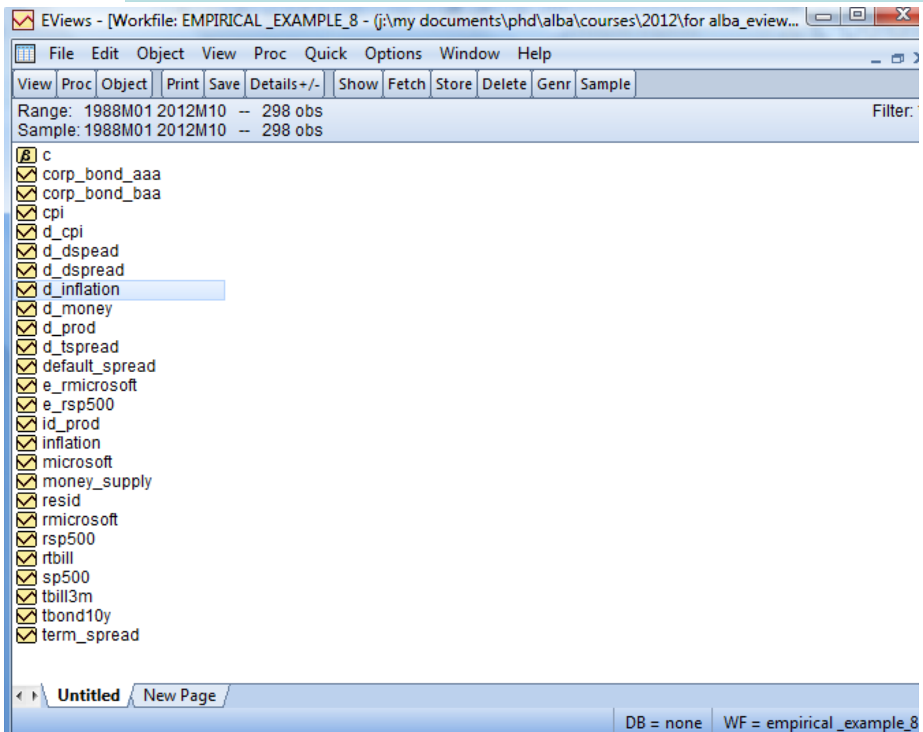
Thus, we use a number of tests that detect heteroskedasticity

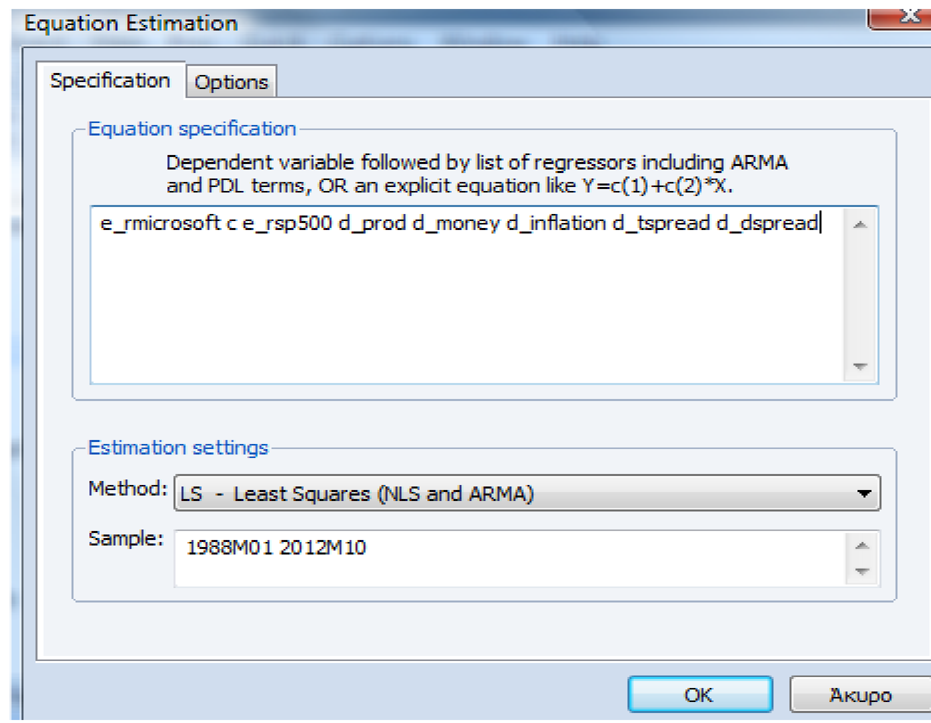
here in EViews: ***White Test*** and ***Breusch-Pagan- Godfrey Test***

Testing for heteroskedasticity

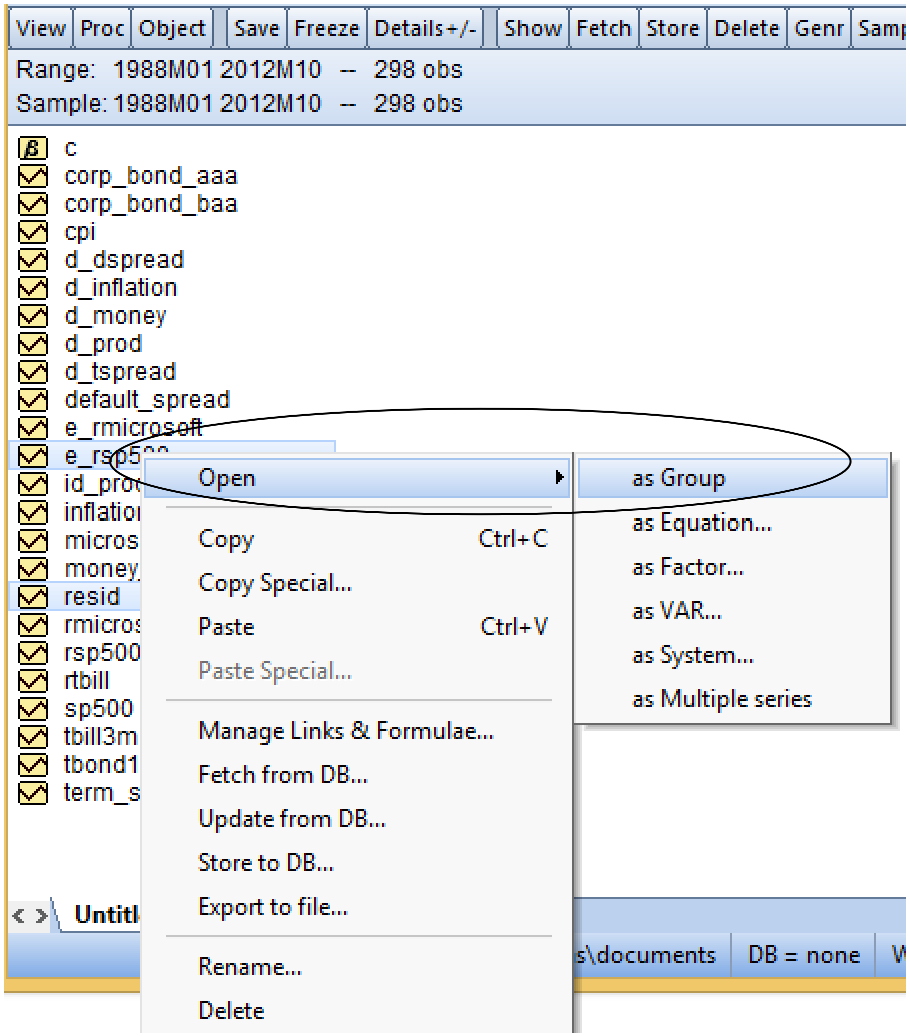
Open empirical_example_8.wf1 ⟶ Quick ⟶ Estimate equation

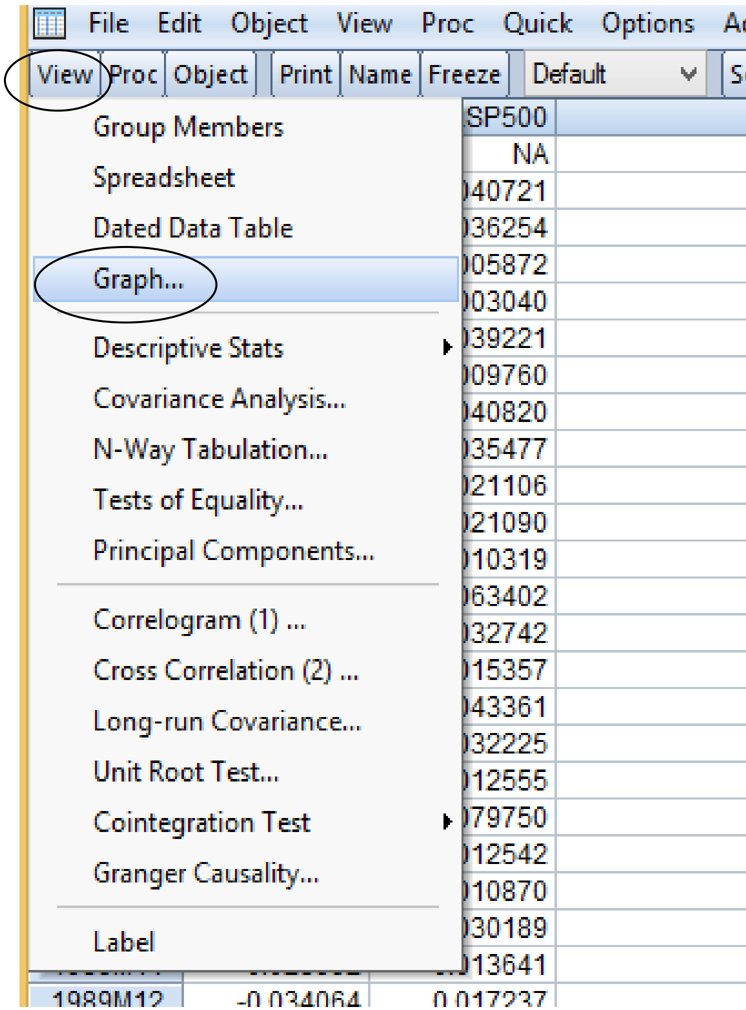
$$\left(R_{Microsoft}-r_f\right)_t=\alpha+\beta_1\left(R_M-r_f\right)_t+\beta_2(d\_prod)+\beta_3(d\_money)+$$
$$\beta_4(d\_\mathrm{inflation})+\beta_5(d\_tspread)+\beta_6(d\_dspread)+u_t$$



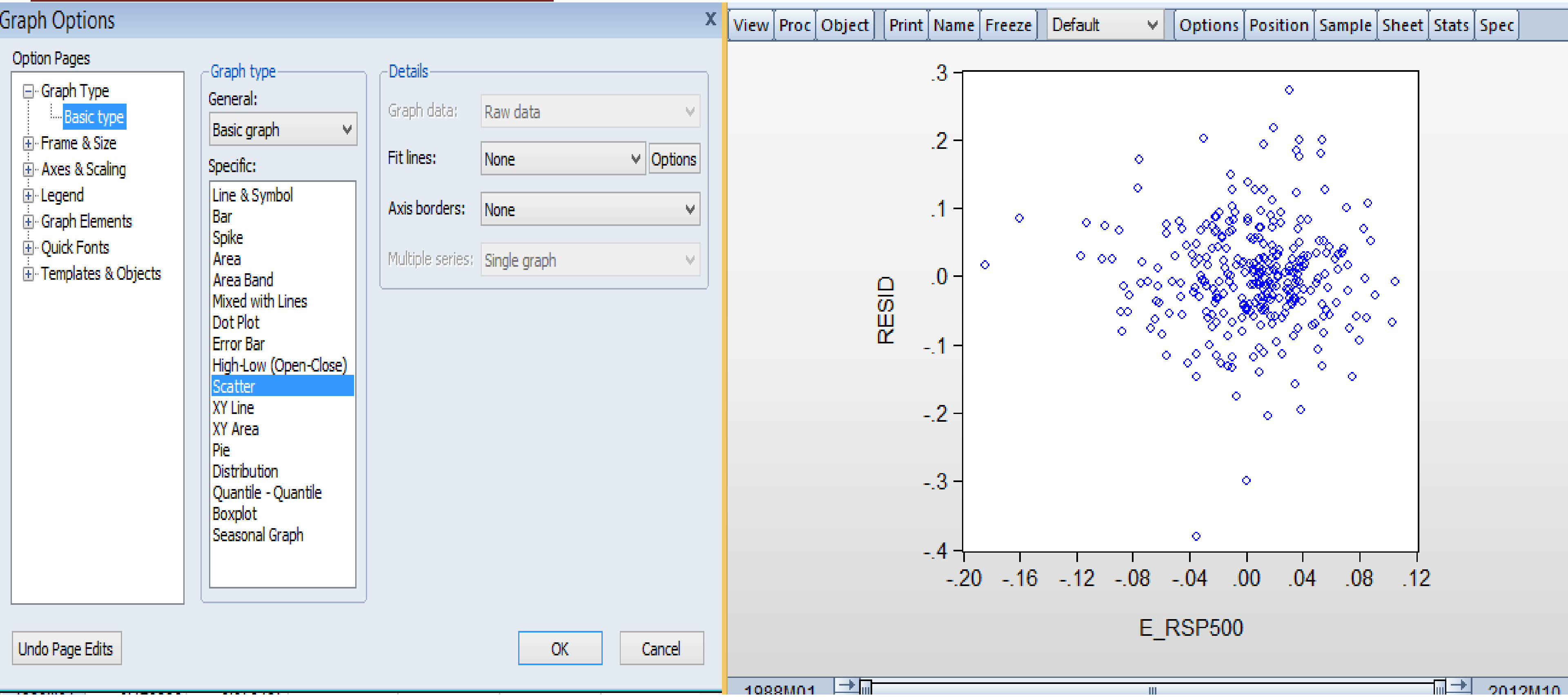


Testing for heteroskedasticity





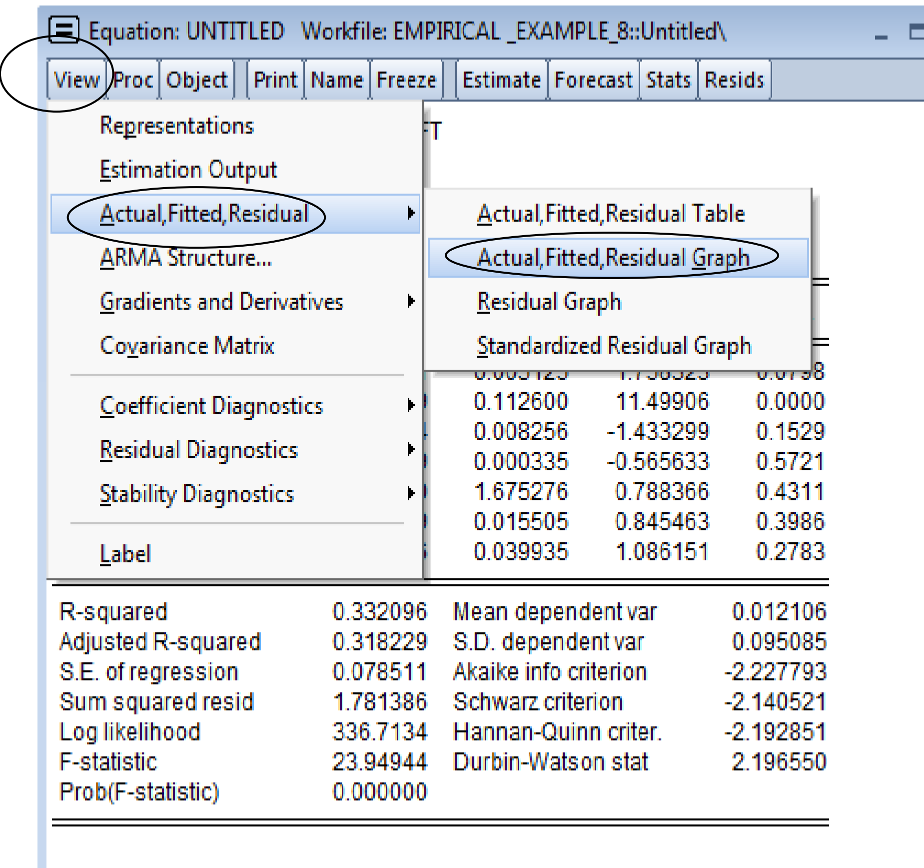
Testing for heteroskedasticity

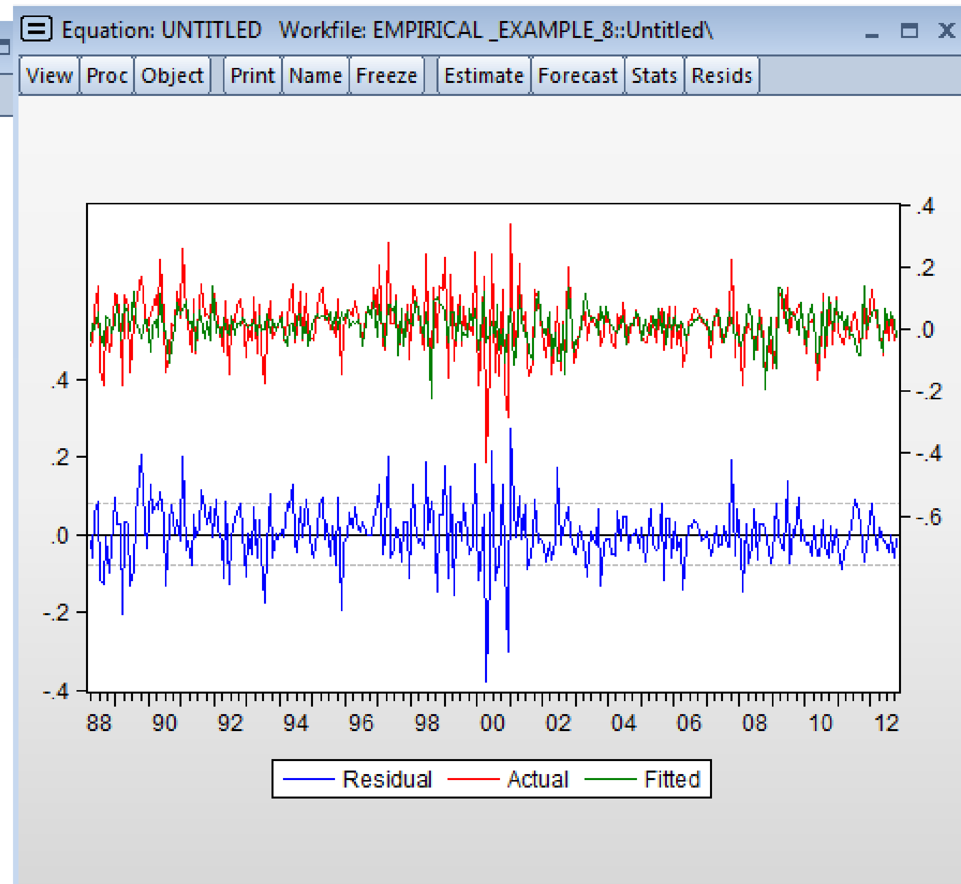


Testing for heteroskedasticity

View ⟶ Actual fitted Residual ⟶ Actual fitted Residual Graph

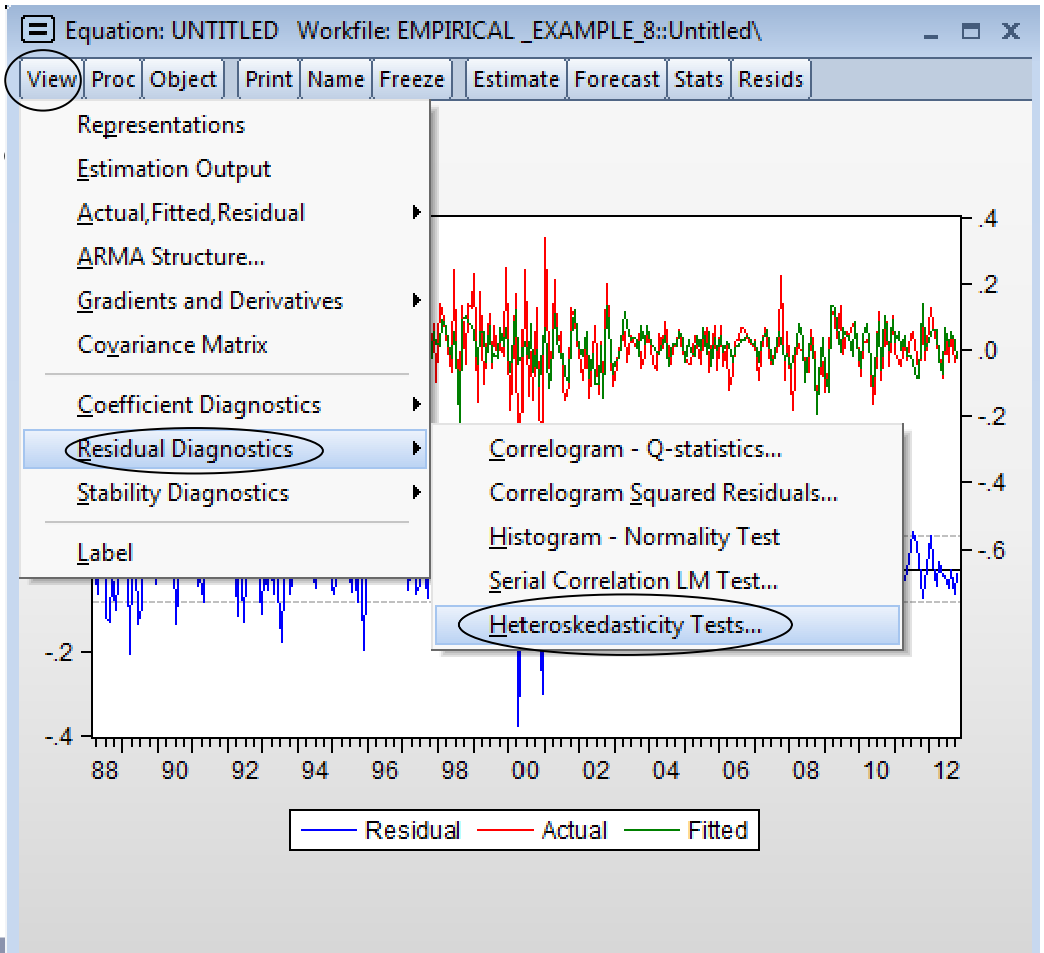
Graphical Illustration of possible heteroskedasticity

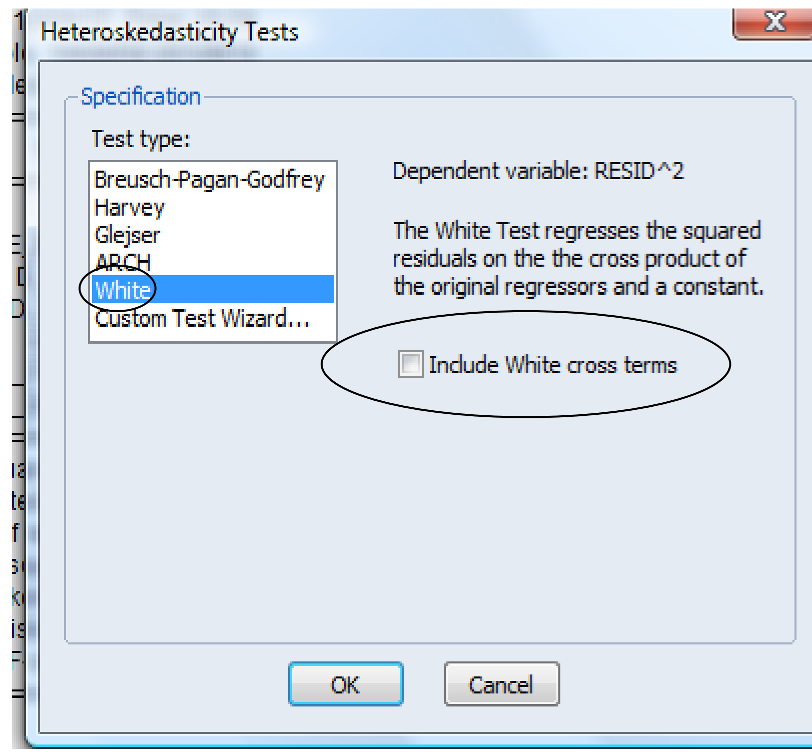




Testing for heteroskedasticity

How to detect it? : View ⟶ Residual Diagnostics ⟶ Heteroskedasticity Tests ⟶ Select ***White***





ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ATHENS UNIVERSITY OF ECONOMICS

Testing for heteroskedasticity

Output of White Test

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: White

F-statistic	0.602509	Prob. F(6,289)	0.7283
Obs*R-squared	3.656875	Prob. Chi-Square(6)	0.7230
Scaled explained SS	8.276685	Prob. Chi-Square(6)	0.2185

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/05/13 Time: 14:14
Sample: 1988M03 2012M10
Included observations: 296

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005711	0.001025	5.571103	0.0000
E_RSP500^2	-0.163700	0.287665	-0.569064	0.5698
D_PROD^2	-0.000662	0.000813	-0.813788	0.4164
D_MONEY^2	-6.92E-07	9.02E-07	-0.766947	0.4437
D_INFLATION^2	58.22221	47.92108	1.214960	0.2254
D_TSPREAD^2	0.006649	0.005908	1.125542	0.2613
D_DSPREAD^2	-0.004564	0.011703	-0.390010	0.6968

R-squared	0.012354	Mean dependent var	0.006018
Adjusted R-squared	-0.008150	S.D. dependent var	0.013137
S.E. of regression	0.013190	Akaike info criterion	-5.795346
Sum squared resid	0.050279	Schwarz criterion	-5.708074
Log likelihood	864.7112	Hannan-Quinn criter.	-5.760404
F-statistic	0.602509	Durbin-Watson stat	1.765153
Prob(F-statistic)	0.728289		

$$H_0 : Homoskedasticity$$

$$H_A : Heteroskedasticity$$

p-values are above 0.05 for **F and chi-squared versions** of the test statistic, so we do not find evidence for presence of heteroskedasticity.

p-value for the third version **" Scaled explained SS"** is 0.21 so we do not find evidence for the presence of heteroskedasticity.

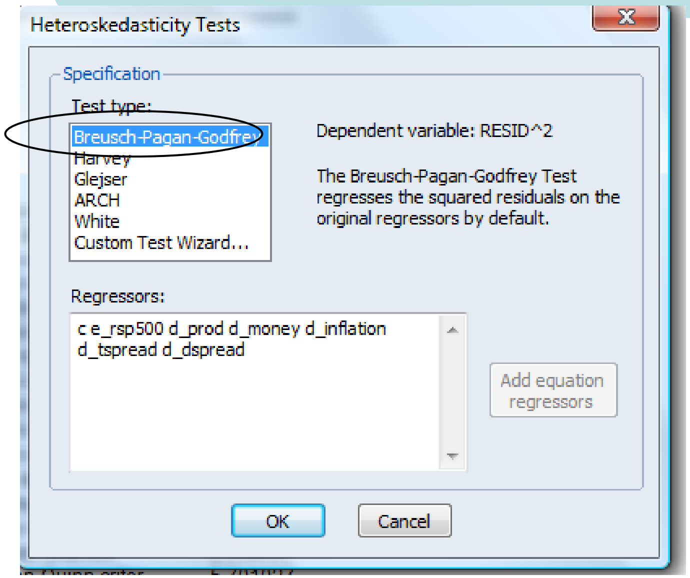
Conclusion: Based on the three versions of the test we do not find evidence for the presence of Heteroskedasticity.

Auxiliary Regression

$$\hat{u}_t^2 = \alpha + \beta_1 \left(R_M - r_f \right)_t^2 + \beta_2 (d\_prod)_t^2 + \beta_3 (d\_money)_t^2 + \beta_4 (d\_inflation)_t^2 + \beta_5 (d\_tspread)_t^2 + \beta_6 (d\_dspread)_t^2 + v_t$$

Testing for heteroskedasticity

How to detect it? Another Test : View ⟶ Residual Diagnostics ⟶ Heteroskedasticity Tests ⟶Select ***Breusch-Pagan-Godfrey***



Equation: UNTITLED Workfile: EMPIRICAL_EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.375560	Prob. F(6,289)	0.8943
Obs*R-squared	2.280082	Prob. Chi-Square(6)	0.8912
Scaled explained SS	5.183193	Prob. Chi-Square(6)	0.5205

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/05/13 Time: 14:19
Sample: 1988M03 2012M10
Included observations: 296

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006456	0.000863	7.480416	0.0000
E_RSP500	-0.005609	0.018961	-0.295791	0.7676
D_PROD	-0.001032	0.001390	-0.742402	0.4584
D_MONEY	-5.10E-05	5.63E-05	-0.905875	0.3658
D_INFLATION	-0.068130	0.282108	-0.241504	0.8093
D_TSPREAD	0.001465	0.002611	0.561234	0.5751
D_DSPREAD	-0.004953	0.006725	-0.736529	0.4620

R-squared	0.007737	Mean dependent var	0.006018
Adjusted R-squared	-0.012864	S.D. dependent var	0.013137
S.E. of regression	0.013221	Akaike info criterion	-5.790682
Sum squared resid	0.050514	Schwarz criterion	-5.703410
Log likelihood	864.0209	Hannan-Quinn criter.	-5.755740
F-statistic	0.375560	Durbin-Watson stat	1.763544
Prob(F-statistic)	0.894290		

p-values are above 0.05; so we do not find evidence for the presence of heteroskedasticity.

Auxiliary Regression

$$\hat{u}_t^2 = \alpha + \beta_1 \left(R_M - r_f\right)_t + \beta_2 (d\_prod)_t + \beta_3 (d\_money)_t + \beta_4 (d\_inflation)_t + \beta_5 (d\_tspread)_t + \beta_6 (d\_dspread)_t + v_t$$

Testing for heteroskedasticity

Change the estimation period ⟶ Detect Heteroskedasticity

1990M01- 2000M10

Equation: UNTITLED   Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: White

F-statistic	3.976778	Prob. F(6,123)	0.0011
Obs*R-squared	21.12129	Prob. Chi-Square(6)	0.0017
Scaled explained SS	32.29206	Prob. Chi-Square(6)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/05/13   Time: 15:06
Sample: 1990M01 2000M10
Included observations: 130

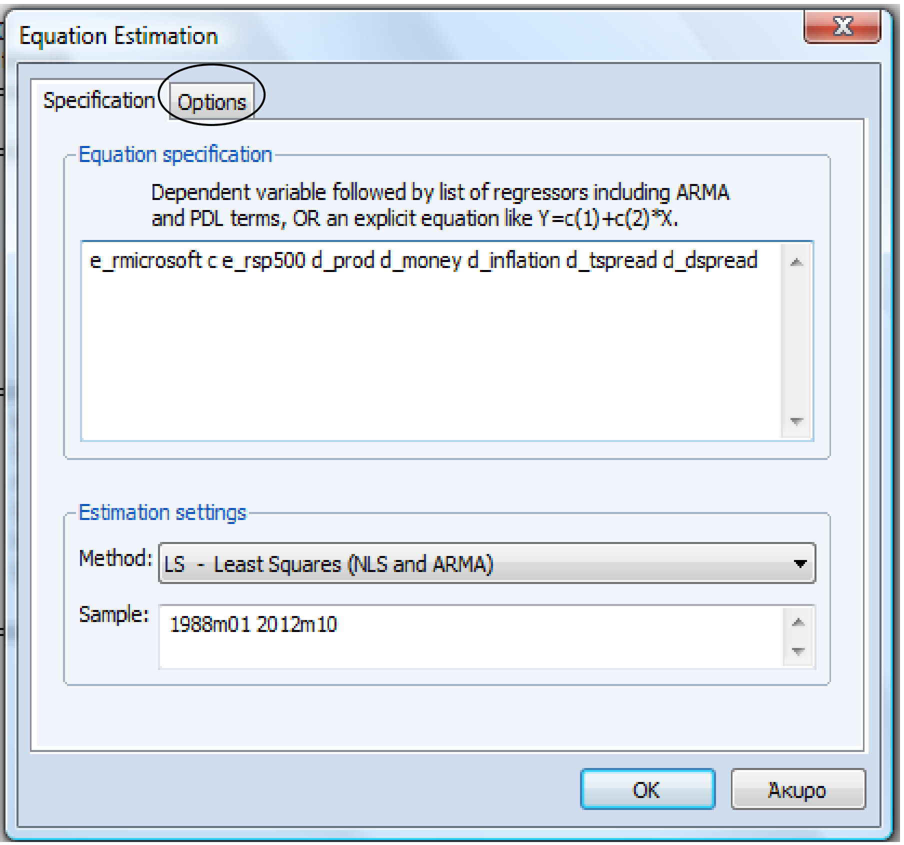
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004518	0.001878	2.405559	0.0176
E_RSP500^2	0.084755	0.487055	0.174015	0.8621
D_PROD^2	-0.000346	0.003782	-0.091596	0.9272
D_MONEY^2	-7.09E-06	2.07E-05	-0.342389	0.7326
D_INFLATION^2	653.7758	140.0555	4.667975	0.0000
D_TSPREAD^2	-0.006284	0.012581	-0.499506	0.6183
D_DSPREAD^2	0.118641	0.122018	0.972324	0.3328

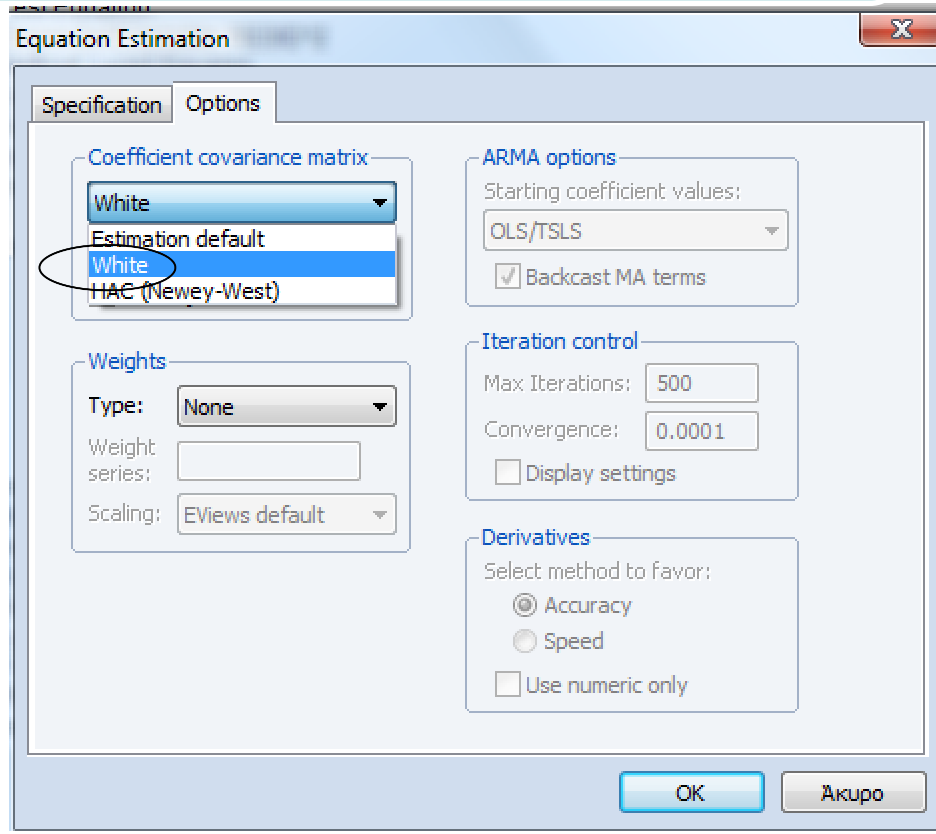
R-squared	0.162471	Mean dependent var	0.007274
Adjusted R-squared	0.121616	S.D. dependent var	0.013496
S.E. of regression	0.012649	Akaike info criterion	-5.850171
Sum squared resid	0.019679	Schwarz criterion	-5.695765
Log likelihood	387.2611	Hannan-Quinn criter.	-5.787431
F-statistic	3.976778	Durbin-Watson stat	2.029333
Prob(F-statistic)	0.001135		

Correcting for heteroskedasticity

Use White robust standard errors in the estimation

Estimate Equation → Options → Select ***White*** → Press Ok





Comparison

Comparing the results *before and after* using the heteroskedasticity robust standard errors.

After….

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 15:09
Sample: 1990M01 2000M10
Included observations: 130
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.021246	0.010146	2.094027	0.0383
E_RSP500	1.431976	0.244356	5.860210	0.0000
D_PROD	-0.026183	0.020297	-1.289981	0.1995
D_MONEY	-0.000125	0.001217	-0.102596	0.9185
D_INFLATION	5.429882	5.818300	0.933242	0.3525
D_TSPREAD	0.007676	0.032402	0.236894	0.8131
D_DSPREAD	0.021260	0.122796	0.173134	0.8628

R-squared	0.318175	Mean dependent var	0.026881
Adjusted R-squared	0.284916	S.D. dependent var	0.103690
S.E. of regression	0.087683	Akaike info criterion	-1.977844
Sum squared resid	0.945653	Schwarz criterion	-1.823439
Log likelihood	135.5599	Hannan-Quinn criter.	-1.915104
F-statistic	9.566384	Durbin-Watson stat	2.232866
Prob(F-statistic)	0.000000		

Before….

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 15:03
Sample: 1990M01 2000M10
Included observations: 130

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.021246	0.010114	2.100670	0.0377
E_RSP500	1.431976	0.213663	6.702029	0.0000
D_PROD	-0.026183	0.020534	-1.275091	0.2047
D_MONEY	-0.000125	0.001397	-0.089384	0.9289
D_INFLATION	5.429882	3.802481	1.427984	0.1558
D_TSPREAD	0.007676	0.032508	0.236119	0.8137
D_DSPREAD	0.021260	0.114629	0.185471	0.8532

R-squared	0.318175	Mean dependent var	0.026881
Adjusted R-squared	0.284916	S.D. dependent var	0.103690
S.E. of regression	0.087683	Akaike info criterion	-1.977844
Sum squared resid	0.945653	Schwarz criterion	-1.823439
Log likelihood	135.5599	Hannan-Quinn criter.	-1.915104
F-statistic	9.566384	Durbin-Watson stat	2.232866
Prob(F-statistic)	0.000000		

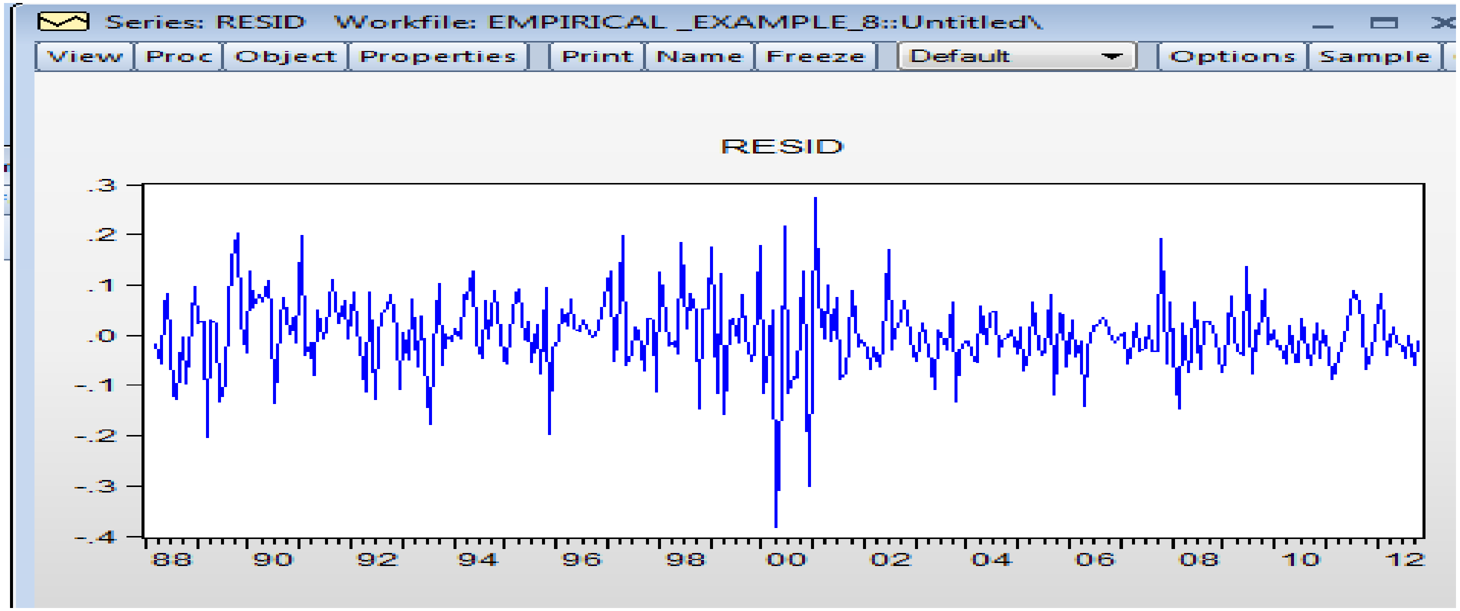
Testing for Serial Correlation/Autocorrelation

3nd Assumption: No serial autocorrelation

$cov(u_i, u_j) = 0$ The errors are linearly independent of one other

- Errors are uncorrelated with one another
- If errors are not uncorrelated with one another, it would be stated that they are ***autocorrelated or serially correlated.***

Plot of residuals over time for our model

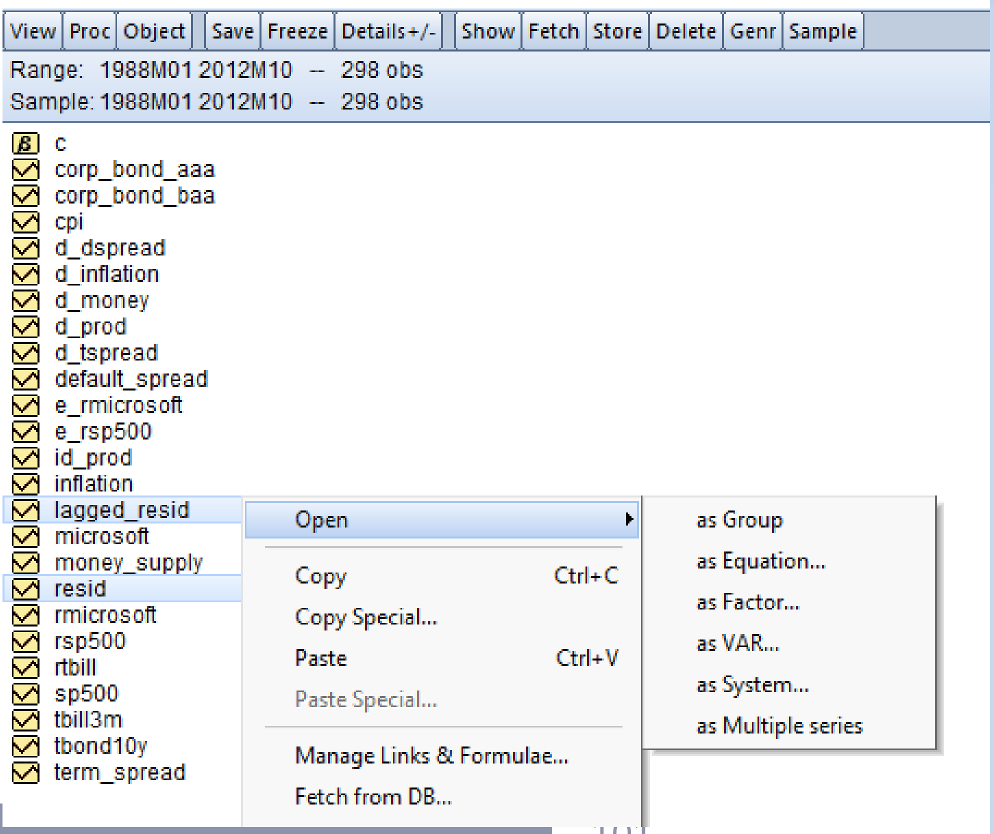


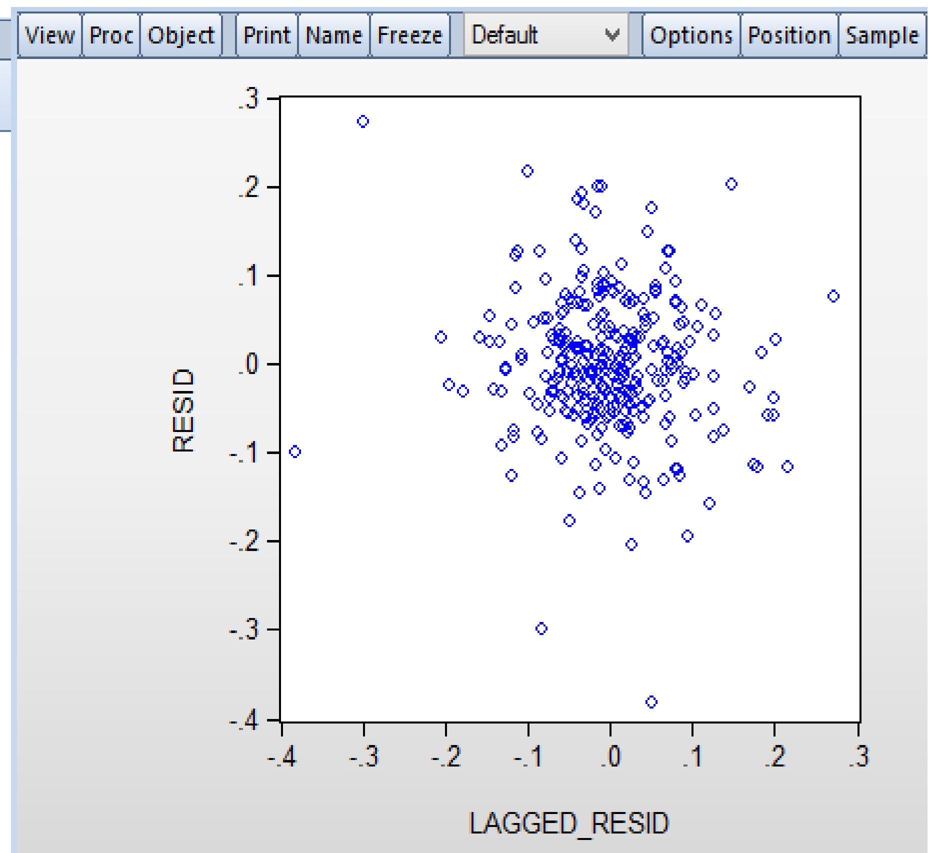
3rd Assumption: No serial autocorrelation

Plot of residuals against their lagged value

- In the command window type :

genr lagged_resid = resid(-1)





Testing for serial correlation

How detect autocorrelation??

From the estimation output a simple test is Durbin –Watson Test

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:27
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009011	0.005324	1.692479	0.0916
E_RSP500	1.294789	0.106406	12.16844	0.0000
D_PROD	-0.011834	0.007143	-1.656595	0.0987
D_MONEY	-0.000189	0.000292	-0.647642	0.5177
D_INFLATION	1.320730	1.757288	0.751573	0.4529
D_TSPREAD	0.013109	0.016608	0.789352	0.4306
D_DSPREAD	0.043376	0.033288	1.303038	0.1936

R-squared	0.332096	Mean dependent var	0.012106
Adjusted R-squared	0.318229	S.D. dependent var	0.095085
S.E. of regression	0.078511	Akaike info criterion	-2.227793
Sum squared resid	1.781386	Schwarz criterion	-2.140521
Log likelihood	336.7134	Hannan-Quinn criter.	-2.192851
F-statistic	23.94944	Durbin-Watson stat	2.196550
Prob(F-statistic)	0.000000		

The Durbin- Watson test statistic is 2.19, close to 2

Durbin – Watson(DW) is a test for **first order autocorrelation.**(tests the relationship between an error and its immediately previous value).

$$u_t = \rho u_{t-1} + v_t$$

$$H_0 : \rho = 0 (\text{No Autocorrelation})$$

$$H_A : \rho \neq 0 (\text{Autocorrelation})$$

Conditions for DW to be a valid Test

1. Existence of a constant term.
2. Non –stochastic regressors.
3. **No** lags of dependent variable.

$$DW \approx 2(1-\rho)$$

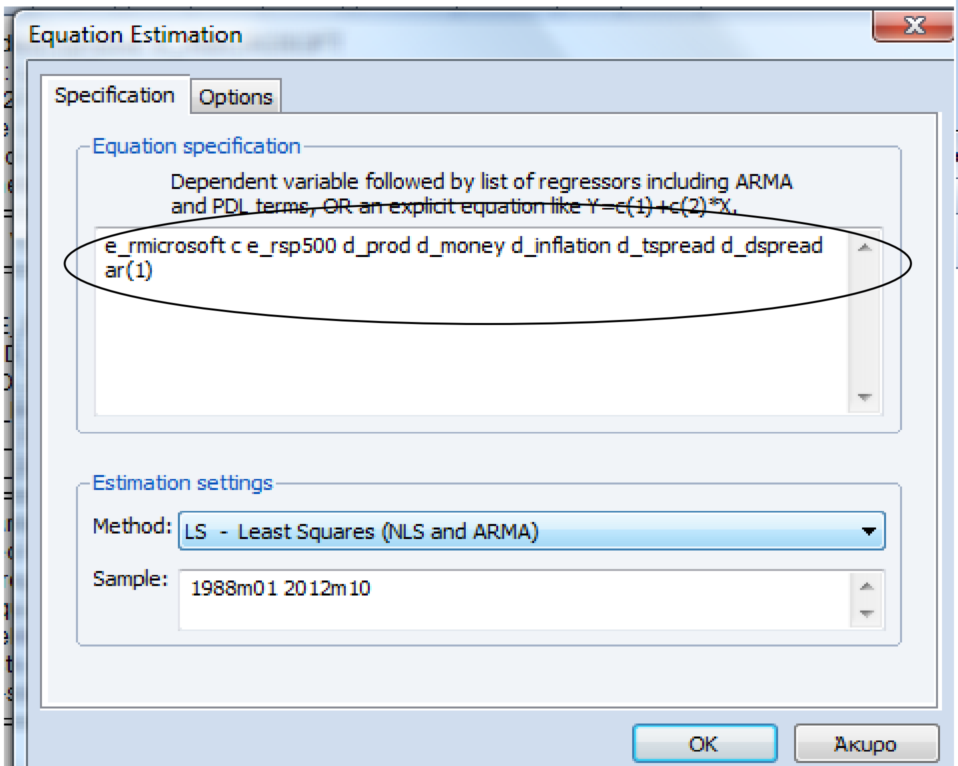
Testing for serial correlation

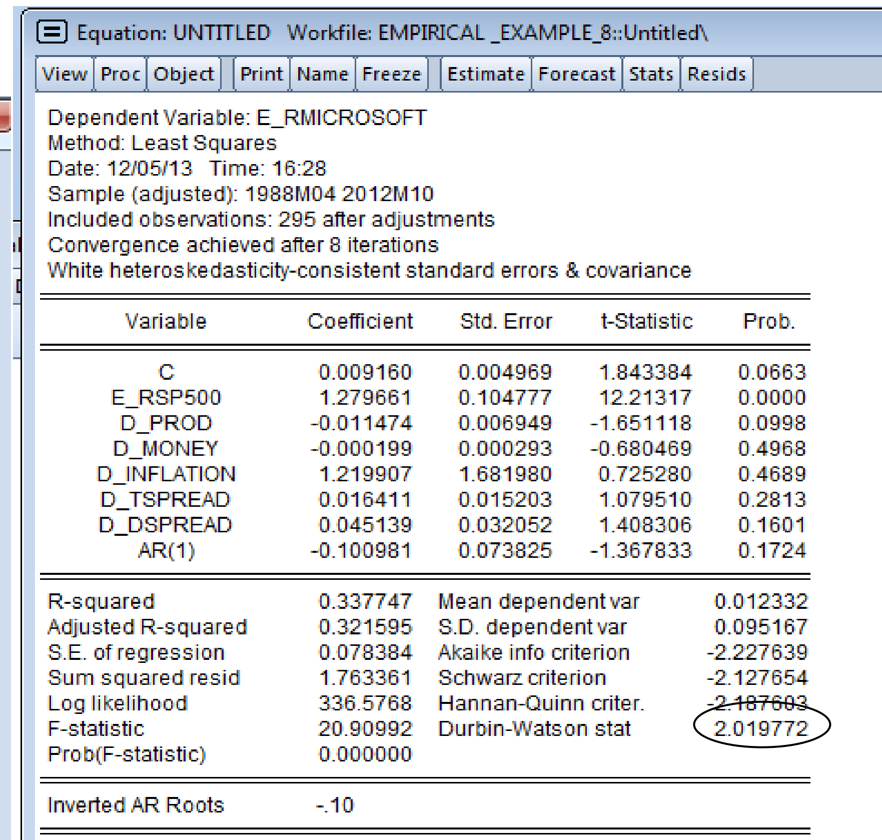
- **Another way**

Cohrane –Orcut Procedure (Recalculate the model assuming the error term follows a **first** order autoregressive process)

$$Y_t = a + b_1 x_{1t} + \ldots + b_6 x_{6t} + u_t$$

$$u_t = \rho u_{t-1} + v_t$$





Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:28
Sample (adjusted): 1988M04 2012M10
Included observations: 295 after adjustments
Convergence achieved after 8 iterations
White heteroskedasticity-consistent standard errors & covariance

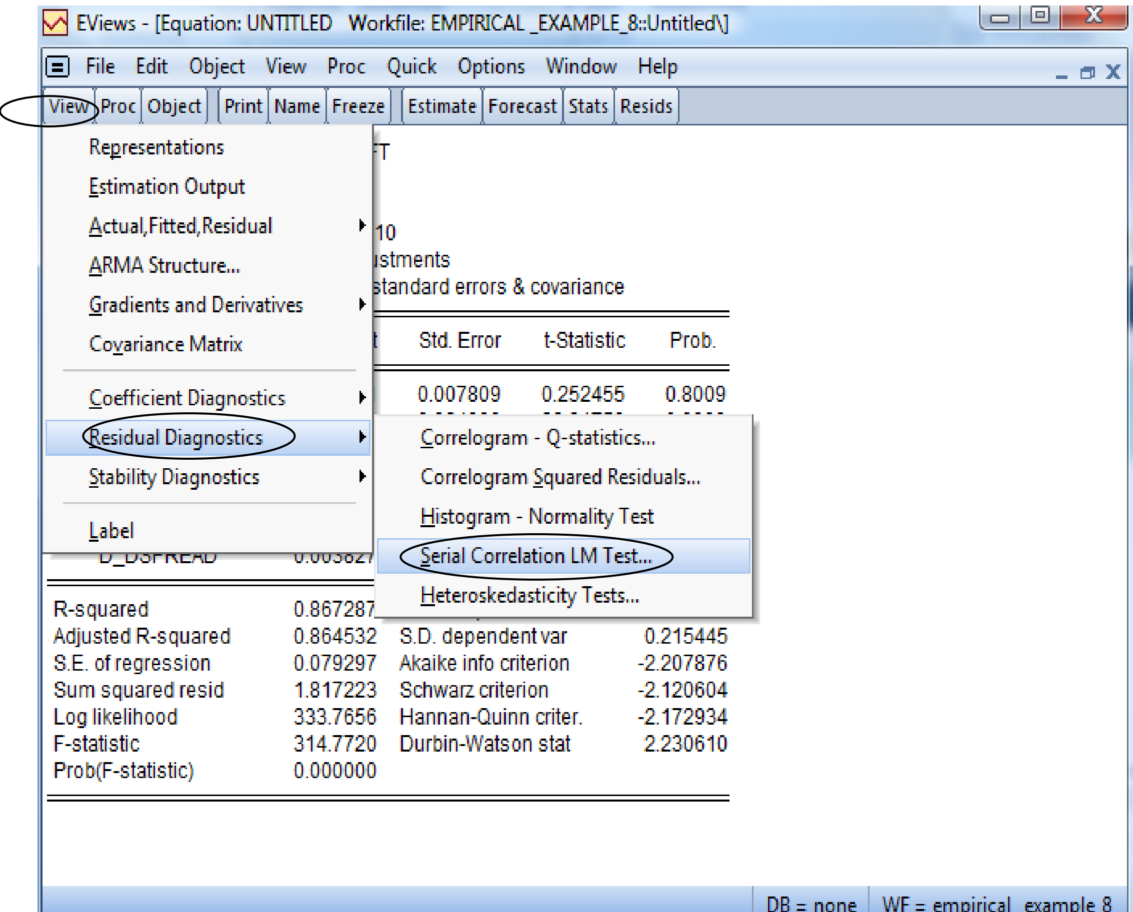
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009160	0.004969	1.843384	0.0663
E_RSP500	1.279661	0.104777	12.21317	0.0000
D_PROD	-0.011474	0.006949	-1.651118	0.0998
D_MONEY	-0.000199	0.000293	-0.680469	0.4968
D_INFLATION	1.219907	1.681980	0.725280	0.4689
D_TSPREAD	0.016411	0.015203	1.079510	0.2813
D_DSPREAD	0.045139	0.032052	1.408306	0.1601
AR(1)	-0.100981	0.073825	-1.367833	0.1724

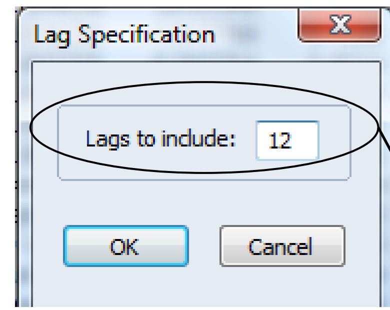
R-squared	0.337747	Mean dependent var	0.012332
Adjusted R-squared	0.321595	S.D. dependent var	0.095167
S.E. of regression	0.078384	Akaike info criterion	-2.227639
Sum squared resid	1.763361	Schwarz criterion	-2.127654
Log likelihood	336.5768	Hannan-Quinn criter.	-2.187603
F-statistic	20.90992	Durbin-Watson stat	2.019772
Prob(F-statistic)	0.000000		

Inverted AR Roots -.10

Testing for serial correlation

Another more robust test than DW is **Breush – Godfrey Test**





Specify the number of lags equal to12. There is no an obvious answer to this, you can experiment on a range of number. You can relate the number of lags with the frequency of your data. (for monthly data use 12, for quarterly data 4, etc)

Testing for serial correlation

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.169460	Prob. F(12,277)	0.3049
Obs*R-squared	14.27300	Prob. Chi-Square(12)	0.2836

p-values for both versions of the test **F and chi-squared** are in excess of 0.05; thus we cannot reject the Null Hypothesis of no autocorrelation / no serial correlation.

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 12/05/13 Time: 16:33
Sample: 1988M03 2012M10
Included observations: 296
Presample missing value lagged residuals set to zero.

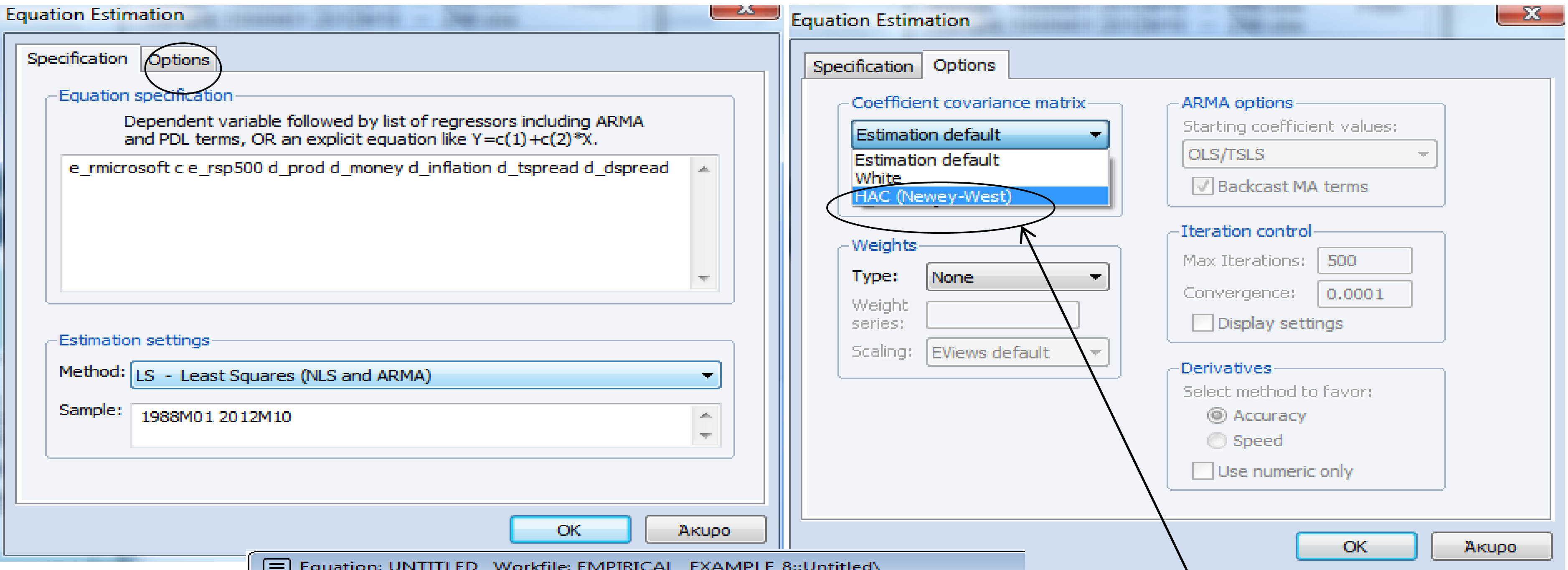
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.20E-05	0.005116	0.006257	0.9950
E_RSP500	-0.014864	0.114914	-0.129345	0.8972
D_PROD	-0.000382	0.008360	-0.045694	0.9636
D_MONEY	5.90E-06	0.000336	0.017543	0.9860
D_INFLATION	-0.968758	1.700562	-0.569669	0.5694
D_TSPREAD	0.000876	0.015749	0.055592	0.9557
D_DSPREAD	0.002545	0.040532	0.062796	0.9500
RESID(-1)	-0.104130	0.060425	-1.723308	0.0859
RESID(-2)	-0.085033	0.060916	-1.395916	0.1639
RESID(-3)	0.090240	0.061194	1.474662	0.1414
RESID(-4)	0.002533	0.061301	0.041321	0.9671
RESID(-5)	0.004603	0.060912	0.075573	0.9398
RESID(-6)	-0.024474	0.060441	-0.404918	0.6859
RESID(-7)	0.099058	0.061313	1.615602	0.1073
RESID(-8)	0.056159	0.060662	0.925772	0.3554
RESID(-9)	-0.052242	0.061241	-0.853065	0.3944
RESID(-10)	-0.025008	0.061504	-0.406602	0.6846
RESID(-11)	-0.030298	0.060947	-0.497129	0.6195
RESID(-12)	0.066893	0.060719	1.101688	0.2716

R-squared	0.048220	Mean dependent var	-3.90E-18
Adjusted R-squared	-0.013629	S.D. dependent var	0.077708
S.E. of regression	0.078236	Akaike info criterion	-2.196133
Sum squared resid	1.695488	Schwarz criterion	-1.959252
Log likelihood	344.0277	Hannan-Quinn criter.	-2.101291
F-statistic	0.779640	Durbin-Watson stat	1.990552
Prob(F-statistic)	0.723794		

$$H_0: \rho_1 = 0 \text{ and } \rho_2 = 0 \text{ and} \ldots \rho_{12} = 0$$

$$H_A: \rho_1 \neq 0 \text{ or } \rho_2 \neq 0 \text{ or} \ldots \rho_{12} \neq 0$$

Summarizing…



Equation: UNTITLED   Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:35
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009011	0.004833	1.864355	0.0633
E_RSP500	1.294789	0.111250	11.63852	0.0000
D_PROD	-0.011834	0.006198	-1.909107	0.0572
D_MONEY	-0.000189	0.000320	-0.591941	0.5544
D_INFLATION	1.320730	1.915675	0.689433	0.4911
D_TSPREAD	0.013109	0.017875	0.733379	0.4639
D_DSPREAD	0.043376	0.026522	1.635447	0.1030

R-squared	0.332096	Mean dependent var	0.012106
Adjusted R-squared	0.318229	S.D. dependent var	0.095085
S.E. of regression	0.078511	Akaike info criterion	-2.227793
Sum squared resid	1.781386	Schwarz criterion	-2.140521
Log likelihood	336.7134	Hannan-Quinn criter.	-2.192851
F-statistic	23.94944	Durbin-Watson stat	2.196550
Prob(F-statistic)	0.000000		

Newey & West for both ***heteroskedasticity and autocorrelation***

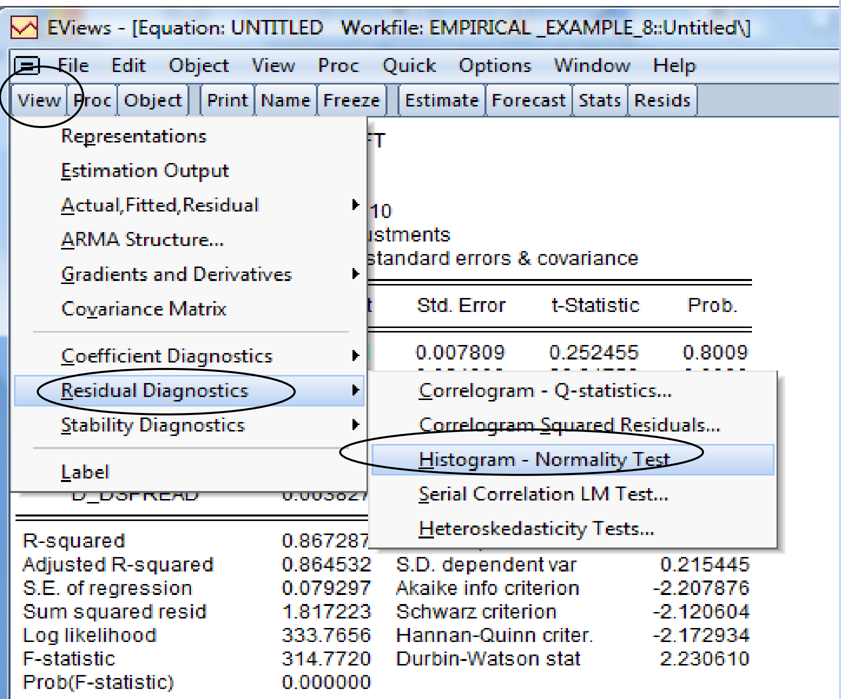
Testing for Non- Normality

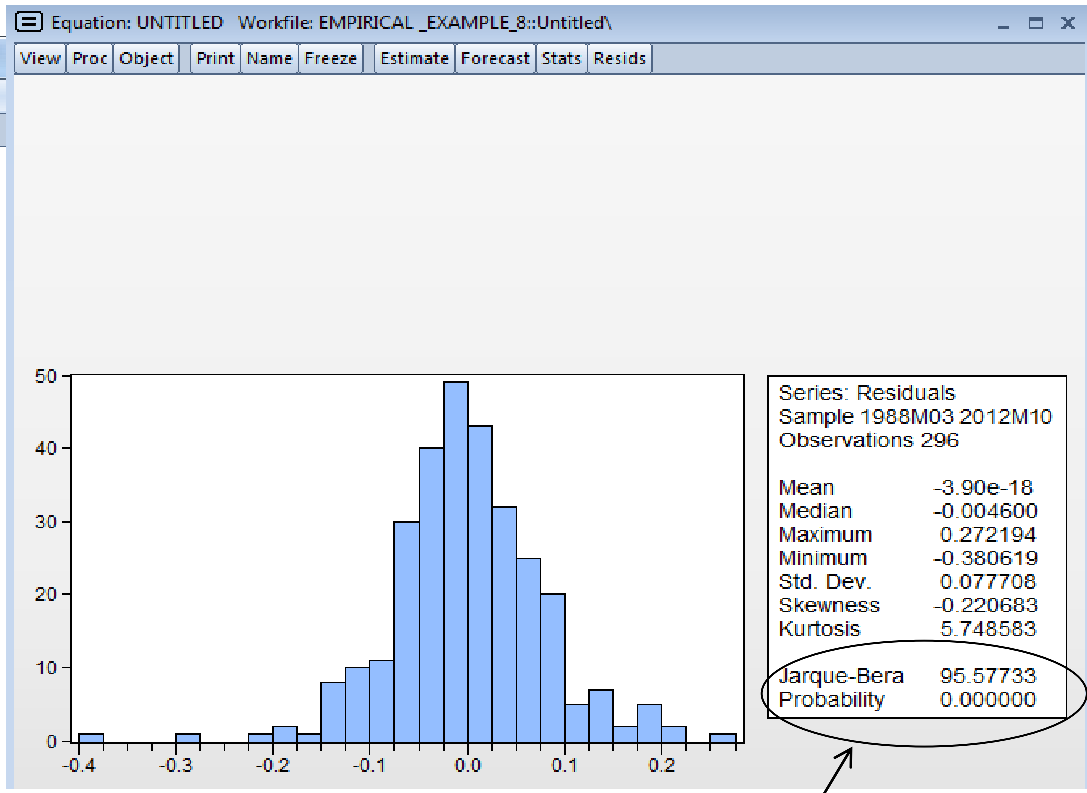
Testing for Non- Normality

$u_t \sim N(0,\sigma^2)$ The errors are normally distributed (Normality)

How to detect Non –Normality ?? Use Jarque – Bera Test :

View → Resid. Diagnostics → Histogram –Normality Test





H_0 : Normality

H_A : Non-Normality

p-value is 0.000, thus we reject the Null Hypothesis of residual normality. Residuals negatively skewed and are leptokurtic

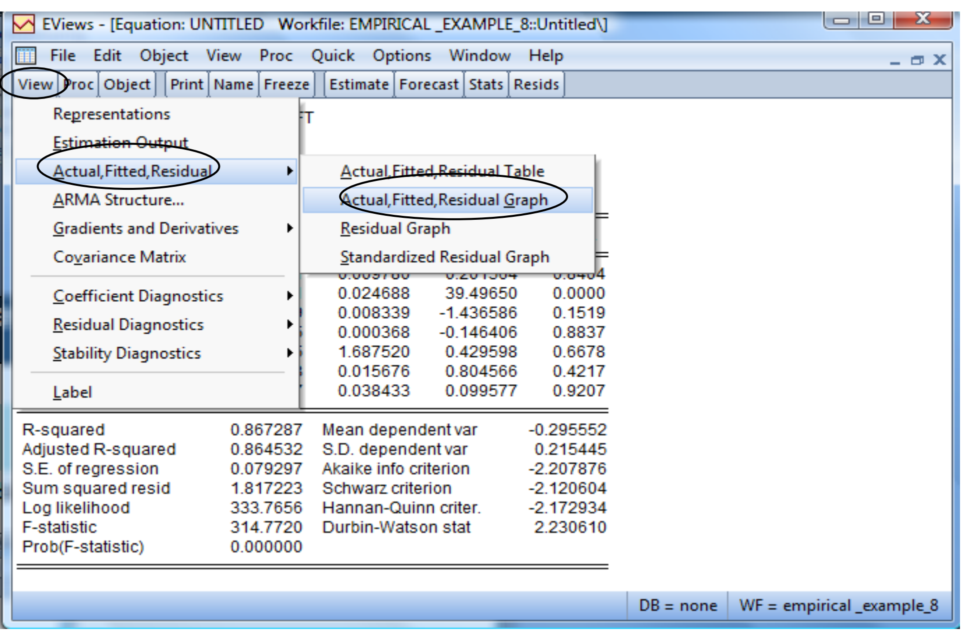
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

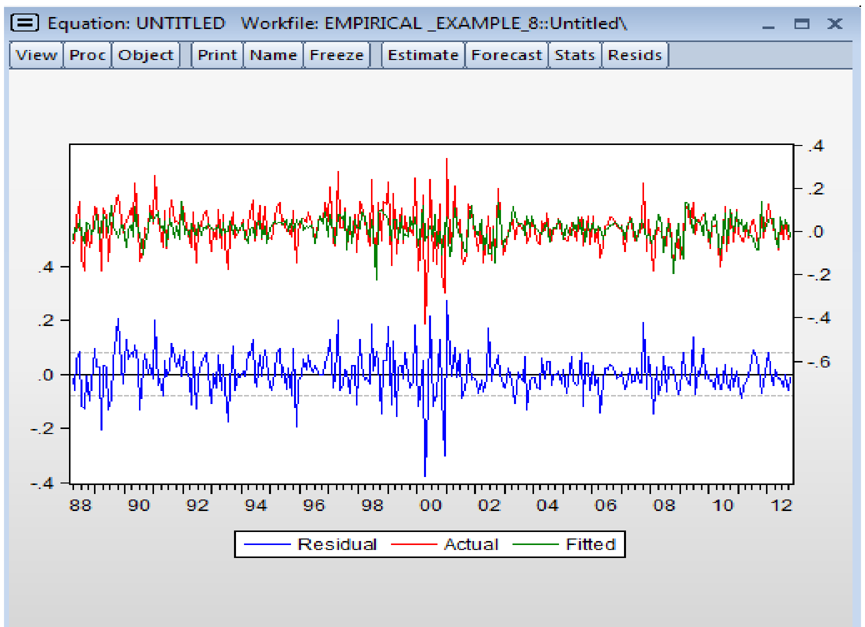
Testing for Non- Normality

What to do if evidence of non-normality is found?

- Central Limit Theory: The test statistics will asymptotically follow the appropriate distribution even in the absence of error normality ; the sample mean converges to a normal distribution.
- Financial/ Economic theory : One or two very extreme residuals cause a rejection of normality assumption (outliers)

A plausible solution : Use of dummy variables

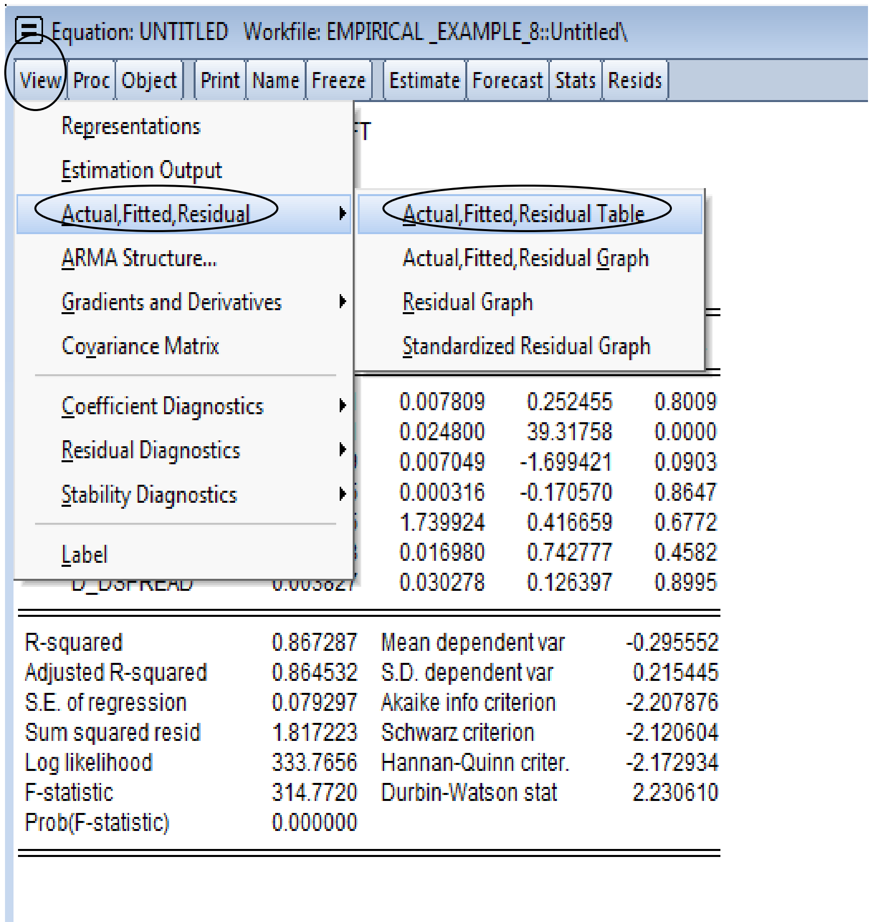




Testing for Non- Normality

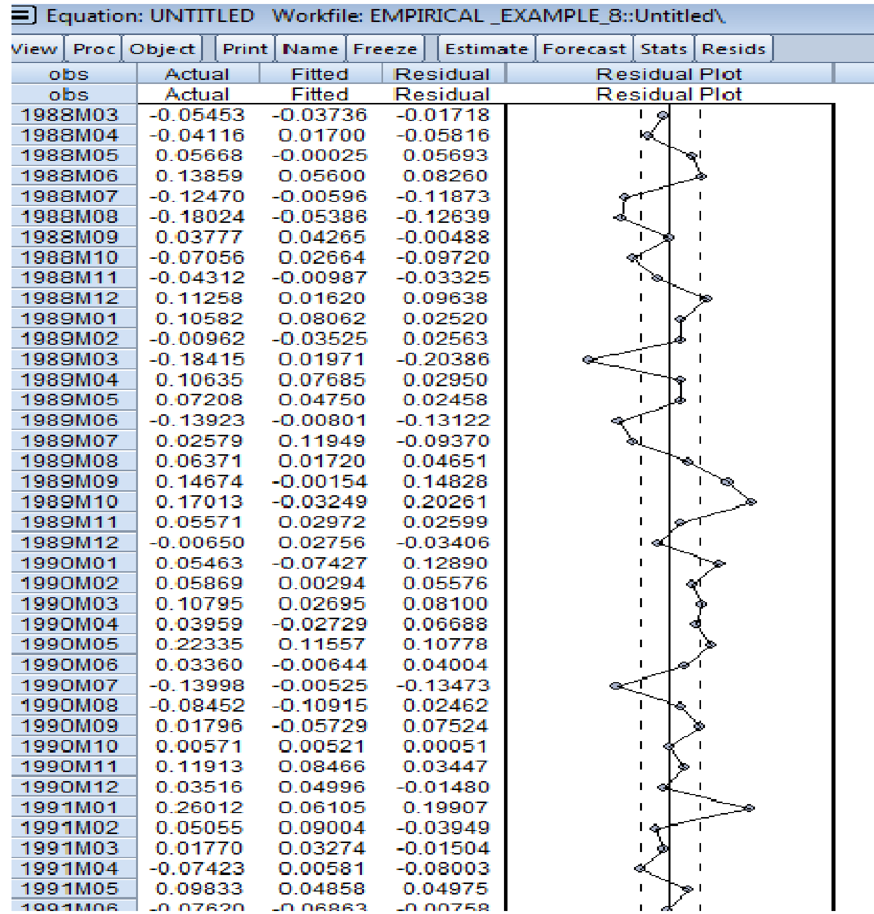
From the graph , we observe a small number of outliers in early 2000.

We can also see the table with the corresponding values



0.007809	0.252455	0.8009
0.024800	39.31758	0.0000
0.007049	-1.699421	0.0903
0.000316	-0.170570	0.8647
1.739924	0.416659	0.6772
0.016980	0.742777	0.4582
0.030278	0.126397	0.8995

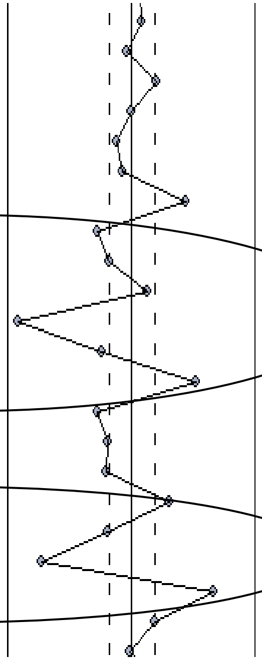
R-squared	0.867287	Mean dependent var	-0.295552
Adjusted R-squared	0.864532	S.D. dependent var	0.215445
S.E. of regression	0.079297	Akaike info criterion	-2.207876
Sum squared resid	1.817223	Schwarz criterion	-2.120604
Log likelihood	333.7656	Hannan-Quinn criter.	-2.172934
F-statistic	314.7720	Durbin-Watson stat	2.230610
Prob(F-statistic)	0.000000		



obs	Actual	Fitted	Residual
1988M03	-0.05453	-0.03736	-0.01718
1988M04	-0.04116	0.01700	-0.05816
1988M05	0.05668	-0.00025	0.05693
1988M06	0.13859	0.05600	0.08260
1988M07	-0.12470	-0.00596	-0.11873
1988M08	-0.18024	-0.05386	-0.12639
1988M09	0.03777	0.04265	-0.00488
1988M10	-0.07056	0.02664	-0.09720
1988M11	-0.04312	-0.00987	-0.03325
1988M12	0.11258	0.01620	0.09638
1989M01	0.10582	0.08062	0.02520
1989M02	-0.00962	-0.03525	0.02563
1989M03	-0.18415	0.01971	-0.20386
1989M04	0.10635	0.07685	0.02950
1989M05	0.07208	0.04750	0.02458
1989M06	-0.13923	-0.00801	-0.13122
1989M07	0.02579	0.11949	-0.09370
1989M08	0.06371	0.01720	0.04651
1989M09	0.14674	-0.00154	0.14828
1989M10	0.17013	-0.03249	0.20261
1989M11	0.05571	0.02972	0.02599
1989M12	-0.00650	0.02756	-0.03406
1990M01	0.05463	-0.07427	0.12890
1990M02	0.05869	0.00294	0.05576
1990M03	0.10795	0.02695	0.08100
1990M04	0.03959	-0.02729	0.06688
1990M05	0.22335	0.11557	0.10778
1990M06	0.03360	-0.00644	0.04004
1990M07	-0.13998	-0.00525	-0.13473
1990M08	-0.08452	-0.10915	0.02462
1990M09	0.01796	-0.05729	0.07524
1990M10	0.00571	0.00521	0.00051
1990M11	0.11913	0.08466	0.03447
1990M12	0.03516	0.04996	-0.01480
1991M01	0.26012	0.06105	0.19907
1991M02	0.05055	0.09004	-0.03949
1991M03	0.01770	0.03274	-0.01504
1991M04	-0.07423	0.00581	-0.08003
1991M05	0.09833	0.04858	0.04975
1991M06	-0.07620	-0.06863	-0.00758

Testing for Non- Normality

1999M06	0.10732	0.07399	0.03334
1999M07	-0.05368	-0.03782	-0.01587
1999M08	0.07157	-0.00983	0.08140
1999M09	-0.02591	-0.02580	-0.00011
1999M10	0.01758	0.06852	-0.05094
1999M11	-0.02093	0.01132	-0.03224
1999M12	0.24422	0.06500	0.17922
2000M01	-0.18114	-0.06500	-0.11614
2000M02	-0.09567	-0.02116	-0.07451
2000M03	0.16805	0.11737	0.05068
2000M04	-0.42573	-0.04511	-0.38062
2000M05	-0.11344	-0.01337	-0.10007
2000M06	0.24096	0.02415	0.21681
2000M07	-0.14144	-0.02575	-0.11569
2000M08	-0.00526	0.07766	-0.08292
2000M09	-0.15147	-0.06686	-0.08460
2000M10	0.12744	0.00055	0.12689
2000M11	-0.18786	-0.10605	-0.08181
2000M12	-0.28463	0.01522	-0.29985
2001M01	0.33786	0.06566	0.27219
2001M02	-0.03840	-0.11280	0.07440
2001M03	-0.07949	-0.07159	-0.00789

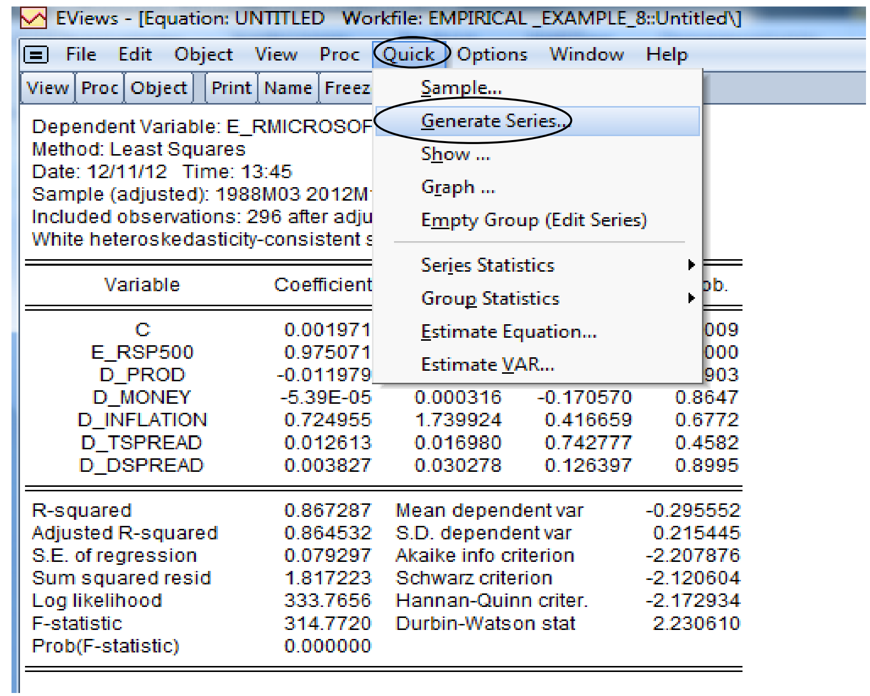


Extreme residuals:

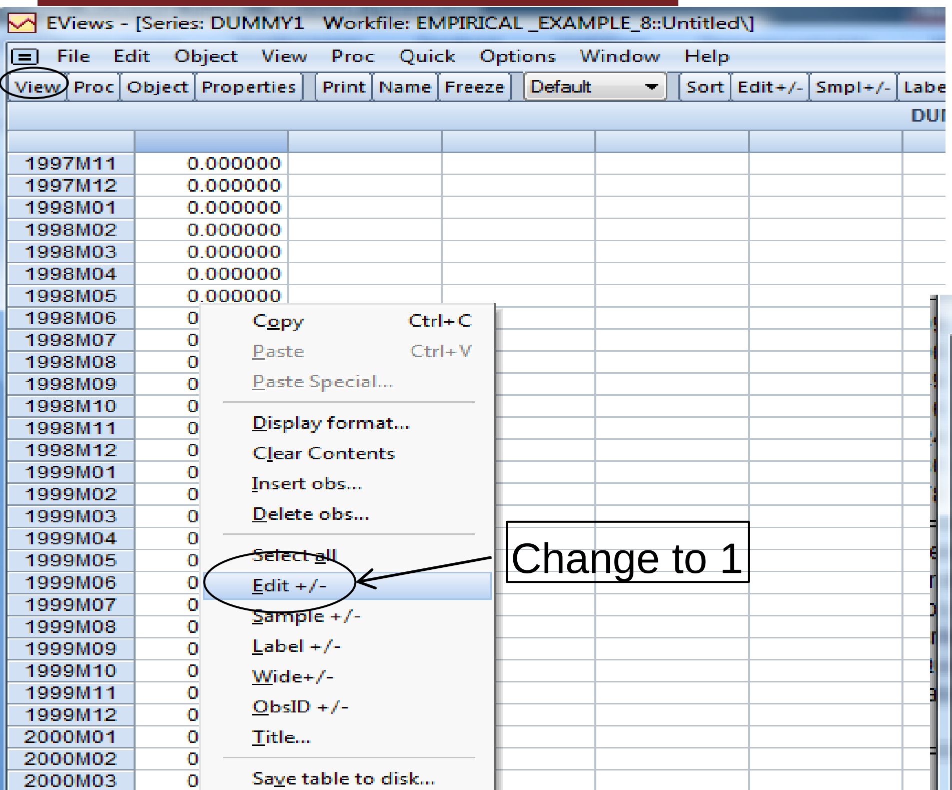
2000M04, 2000M12

Construct the dummy variables in order to remove big outliers

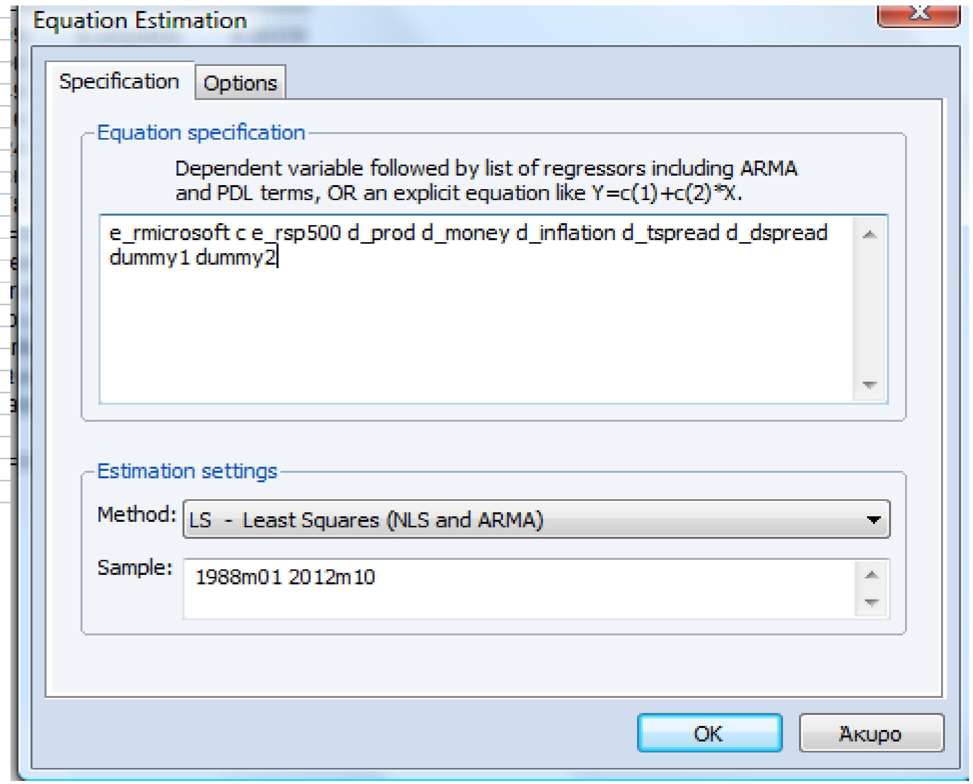
•Create 2000M04 dummy 1:



Testing for Non- Normality



Create also the 2000M12 dummy2 in the same way and run the following regression



Testing for Non- Normality

With Dummies

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:48
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.011695	0.004575	2.556130	0.0111
E_RSP500	1.265716	0.101805	12.43277	0.0000
D_PROD	-0.011751	0.006335	-1.854849	0.0646
D_MONEY	-0.000227	0.000308	-0.735616	0.4626
D_INFLATION	0.367889	1.412260	0.260497	0.7947
D_TSPREAD	0.017380	0.016146	1.076453	0.2826
D_DSPREAD	0.046554	0.026051	1.787039	0.0750
DUMMY1	-0.391750	0.010734	-36.49662	0.0000
DUMMY2	-0.302171	0.006651	-45.43132	0.0000

R-squared	0.421973	Mean dependent var	0.012106
Adjusted R-squared	0.405861	S.D. dependent var	0.095085
S.E. of regression	0.073292	Akaike info criterion	-2.358804
Sum squared resid	1.541672	Schwarz criterion	-2.246597
Log likelihood	358.1030	Hannan-Quinn criter.	-2.313879
F-statistic	26.18956	Durbin-Watson stat	2.189767
Prob(F-statistic)	0.000000		

Without Dummies

EViews - [Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\]

File | Edit | Object | View | Proc | Quick | Options | Window | Help

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/11/12 Time: 11:52
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
White heteroskedasticity-consistent standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001971	0.007809	0.252455	0.8009
E_RSP500	0.975071	0.024800	39.31758	0.0000
D_PROD	-0.011979	0.007049	-1.699421	0.0903
D_MONEY	-5.39E-05	0.000316	-0.170570	0.8647
D_INFLATION	0.724955	1.739924	0.416659	0.6772
D_TSPREAD	0.012613	0.016980	0.742777	0.4582
D_DSPREAD	0.003827	0.030278	0.126397	0.8995

R-squared	0.867287	Mean dependent var	-0.295552
Adjusted R-squared	0.864532	S.D. dependent var	0.215445
S.E. of regression	0.079297	Akaike info criterion	-2.207876
Sum squared resid	1.817223	Schwarz criterion	-2.120604
Log likelihood	333.7656	Hannan-Quinn criter.	-2.172934
F-statistic	314.7720	Durbin-Watson stat	2.230610
Prob(F-statistic)	0.000000		

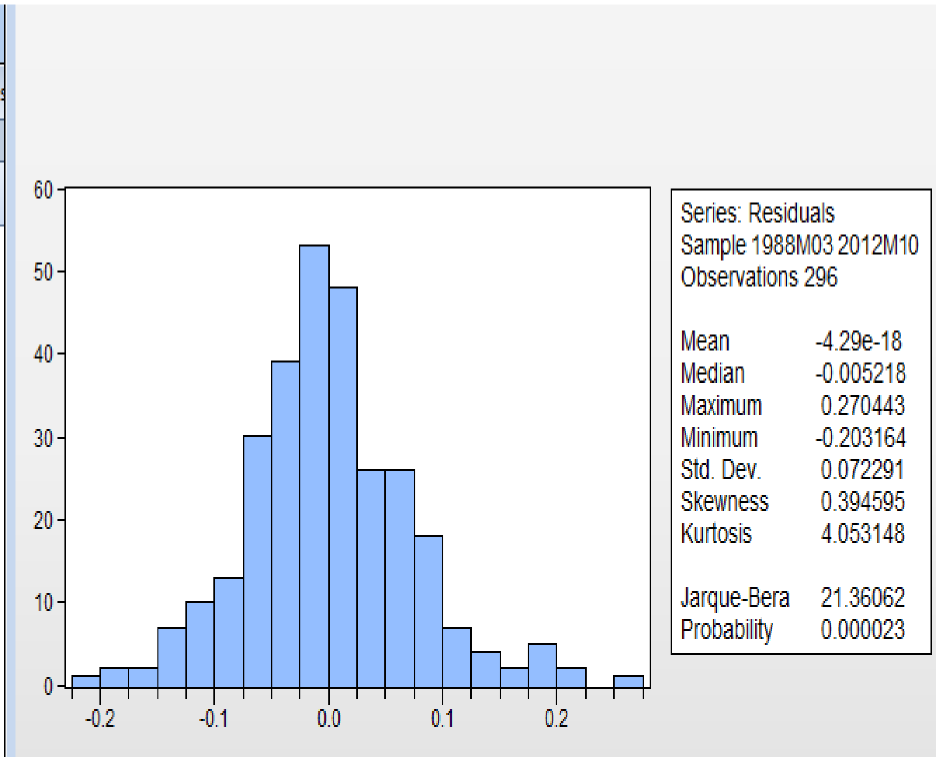
We observe a change in coefficient values as well as in t-stats and p-values

Testing for Non- Normality

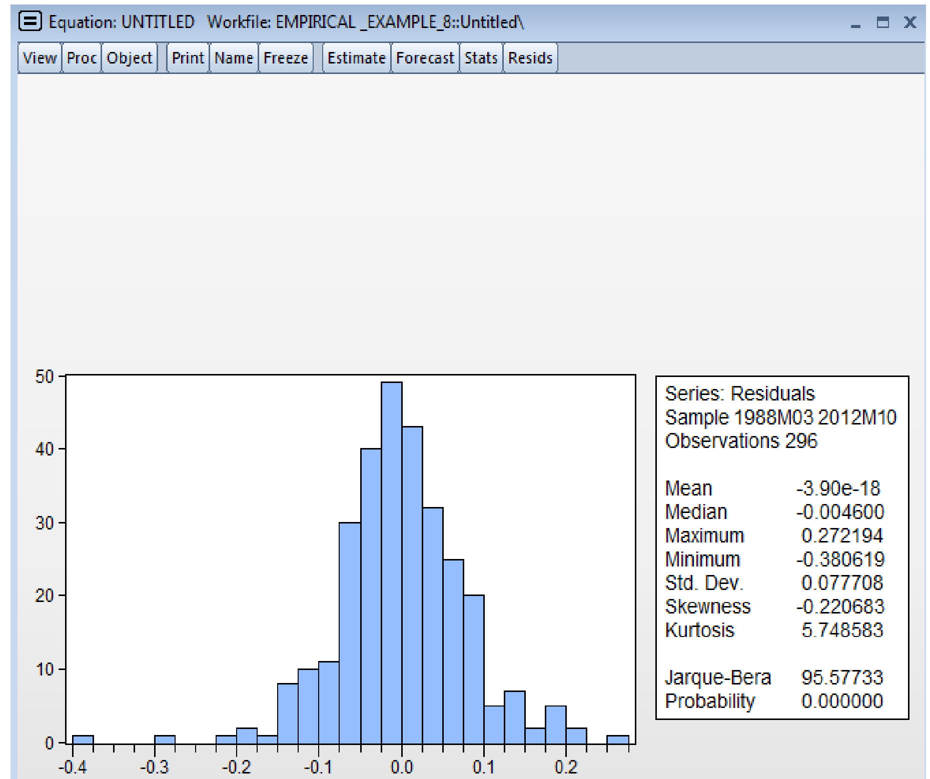
OIKONOMIKO ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

- What about the normality of residuals??

With Dummies



Without Dummies



Skewness and kurtosis are slightly closer to the values that would take under normality

JB value is significantly lower

A long way for residuals to follow a normal distribution…

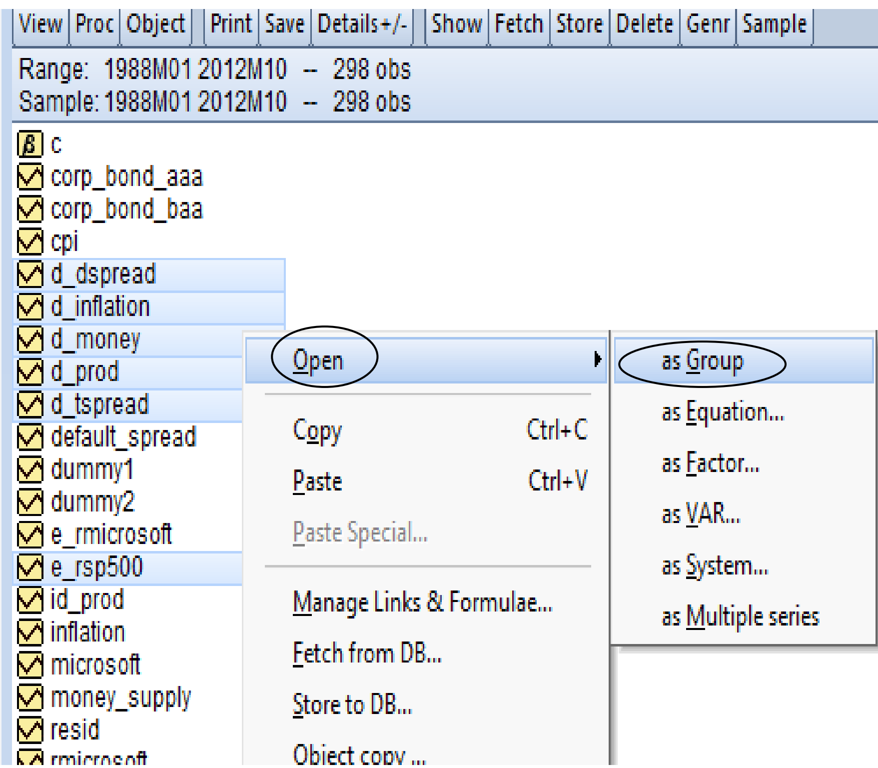
Testing for Multicollinearity

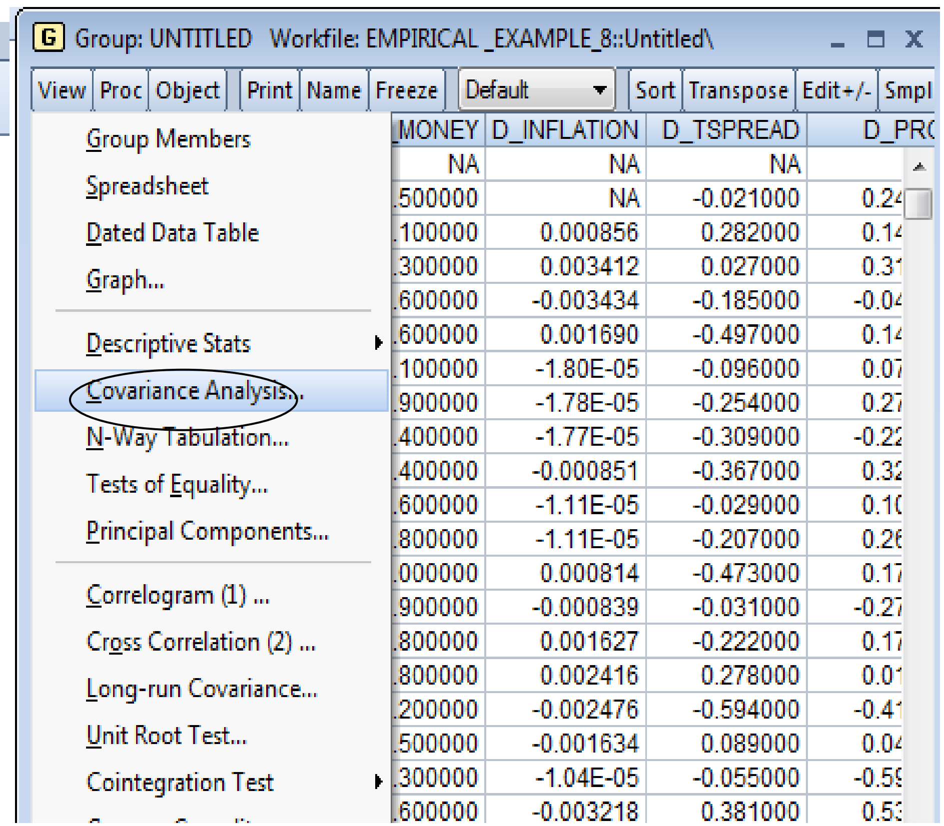
Testing for multicollinearity

Implicit assumption: explanatory variables not correlated/orthogonal with one another.

How detect multicollinearity?? Two easy ways:

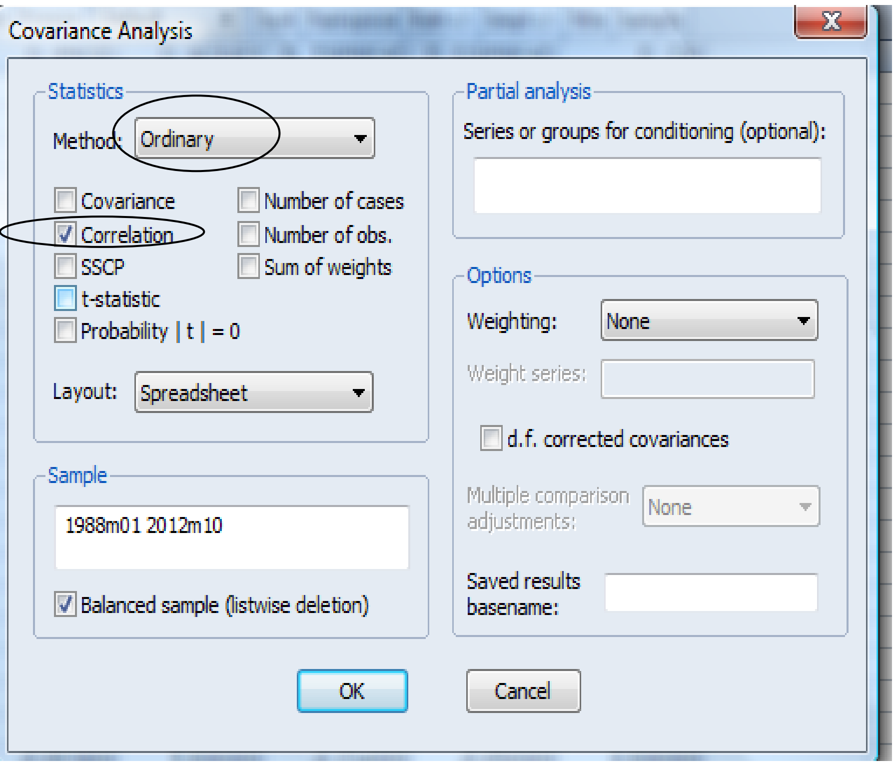
1. Use the correlation matrix of the explanatory variables

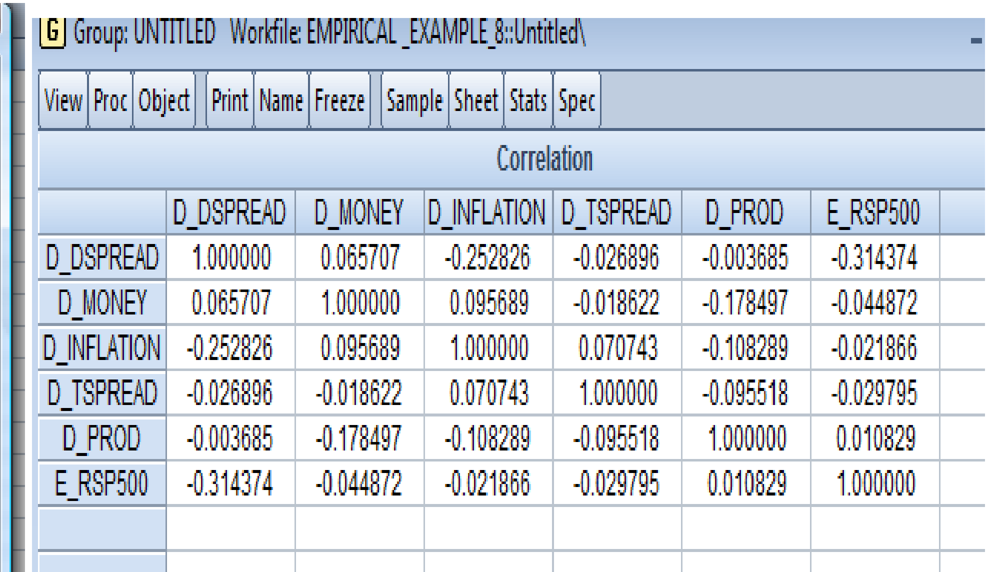




Testing for multicollinearity

Correlation Matrix





Correlation

	D_DSPREAD	D_MONEY	D_INFLATION	D_TSPREAD	D_PROD	E_RSP500
D_DSPREAD	1.000000	0.065707	-0.252826	-0.026896	-0.003685	-0.314374
D_MONEY	0.065707	1.000000	0.095689	-0.018622	-0.178497	-0.044872
D_INFLATION	-0.252826	0.095689	1.000000	0.070743	-0.108289	-0.021866
D_TSPREAD	-0.026896	-0.018622	0.070743	1.000000	-0.095518	-0.029795
D_PROD	-0.003685	-0.178497	-0.108289	-0.095518	1.000000	0.010829
E_RSP500	-0.314374	-0.044872	-0.021866	-0.029795	0.010829	1.000000

No problem of multicollinearity, since the correlation among the explanatory variables is relatively small

Testing for multicollinearity

2. Add/Remove an explanatory variables, identifying a big change in the coefficient values

Drop *d_dspread* from our regression equation

Without d_dspread

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:57
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009018	0.004850	1.859321	0.0640
E_RSP500	1.254566	0.111124	11.28984	0.0000
D_PROD	-0.012005	0.006233	-1.926016	0.0551
D_MONEY	-0.000161	0.000317	-0.507025	0.6125
D_INFLATION	0.812608	1.985283	0.409316	0.6826
D_TSPREAD	0.012783	0.017931	0.712904	0.4765

R-squared	0.329369	Mean dependent var	0.012106
Adjusted R-squared	0.317807	S.D. dependent var	0.095085
S.E. of regression	0.078535	Akaike info criterion	-2.230476
Sum squared resid	1.788658	Schwarz criterion	-2.155672
Log likelihood	336.1105	Hannan-Quinn criter.	-2.200526
F-statistic	28.48573	Durbin-Watson stat	2.188938
Prob(F-statistic)	0.000000		

With d_dspread

Equation: UNTITLED Workfile: EMPIRICAL _EXAMPLE_8::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:56
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009011	0.004833	1.864355	0.0633
E_RSP500	1.294789	0.111250	11.63852	0.0000
D_PROD	-0.011834	0.006198	-1.909107	0.0572
D_MONEY	-0.000189	0.000320	-0.591941	0.5544
D_INFLATION	1.320730	1.915675	0.689433	0.4911
D_TSPREAD	0.013109	0.017875	0.733379	0.4639
D_DSPREAD	0.043376	0.026522	1.635447	0.1030

R-squared	0.332096	Mean dependent var	0.012106
Adjusted R-squared	0.318229	S.D. dependent var	0.095085
S.E. of regression	0.078511	Akaike info criterion	-2.227793
Sum squared resid	1.781386	Schwarz criterion	-2.140521
Log likelihood	336.7134	Hannan-Quinn criter.	-2.192851
F-statistic	23.94944	Durbin-Watson stat	2.196550
Prob(F-statistic)	0.000000		

No big changes in coeff. values, thus no present of multicollinearity

Testing for multicollinearity

- ***Problems if near Multicollinearity is present but ignored***

1. R-squared will be high, but the individual coeff. will have high standard errors, so that regression "looks good" as a whole, but the individual variables are not significant.

*Remark: Multicollinearity does **not** affect the value of R-squared in the regression.*

2. Regression becomes very sensitive to small changes in the specification; add/remove an independent variable leads to large changes in the coeff. values or significances of other variables.

3. Wide confidence intervals for the parameters; inappropriate results for significance tests.

Testing for multicollinearity

- ***Solutions to the problem of multicollinearity***

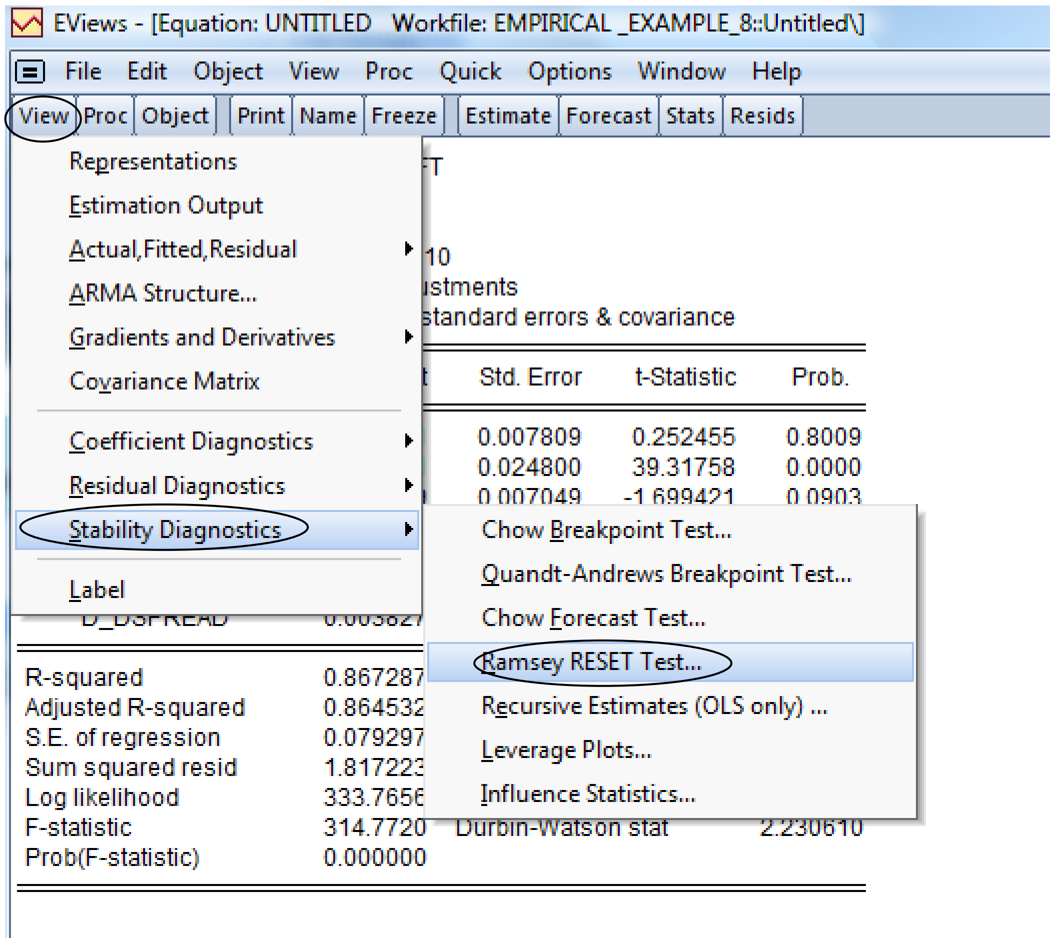
1. Use of ridge Regressions
2. Use of Principal Component Analysis.
3. **Ignorance** of multicollinearity **if the model is statistically appropriate.**
4. **Drop** one of the collinear variables
5. **Transform** the highly correlated variables into a **ratio** and **include** the **ratio** and **not** the individual explanatory variables.
6. A sufficient history of data : longer run of data/ higher frequent data/pooled data.

Testing for linear relationship between Y and X

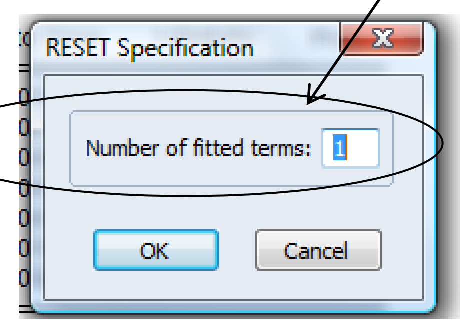
Testing for linear relationship between Y and X

Linearity or not???

Ramsey RESET test : View → Stability Diagnostics → Ramsey RESET Test



The default of EViews



Testing for linear relationship between Y and X

Omitted Variables: Squares of fitted values

	Value	df	Probability
t-statistic	1.125077	288	0.2615
F-statistic	1.265799	(1, 288)	0.2615
Likelihood ratio	1.298109	1	0.2546

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	0.007795	1	0.007795
Restricted SSR	1.781386	289	0.006164
Unrestricted SSR	1.773591	288	0.006158
Unrestricted SSR	1.773591	288	0.006158

LR test summary:

	Value	df
Restricted LogL	336.7134	289
Unrestricted LogL	337.3625	288

Unrestricted Test Equation:
Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:59
Sample: 1988M03 2012M10
Included observations: 296
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005799	0.005455	1.062927	0.2887
E_RSP500	1.298691	0.110054	11.80051	0.0000
D_PROD	-0.011925	0.006576	-1.813385	0.0708
D_MONEY	-0.000199	0.000317	-0.627211	0.5310
D_INFLATION	1.389351	1.928037	0.720604	0.4717
D_TSPREAD	0.011931	0.017602	0.677800	0.4984
D_DSPREAD	0.032900	0.026370	1.247624	0.2132
FITTED^2	1.035949	0.683401	1.515874	0.1306

R-squared	0.335018	Mean dependent var	0.012106
Adjusted R-squared	0.318856	S.D. dependent var	0.095085
S.E. of regression	0.078475	Akaike info criterion	-2.225422
Sum squared resid	1.773591	Schwarz criterion	-2.125683
Log likelihood	337.3625	Hannan-Quinn criter.	-2.185488
F-statistic	20.72780	Durbin-Watson stat	2.197602
Prob(F-statistic)	0.000000		

$$H_0 : Linearity$$

$$H_A : Non-Linearity$$

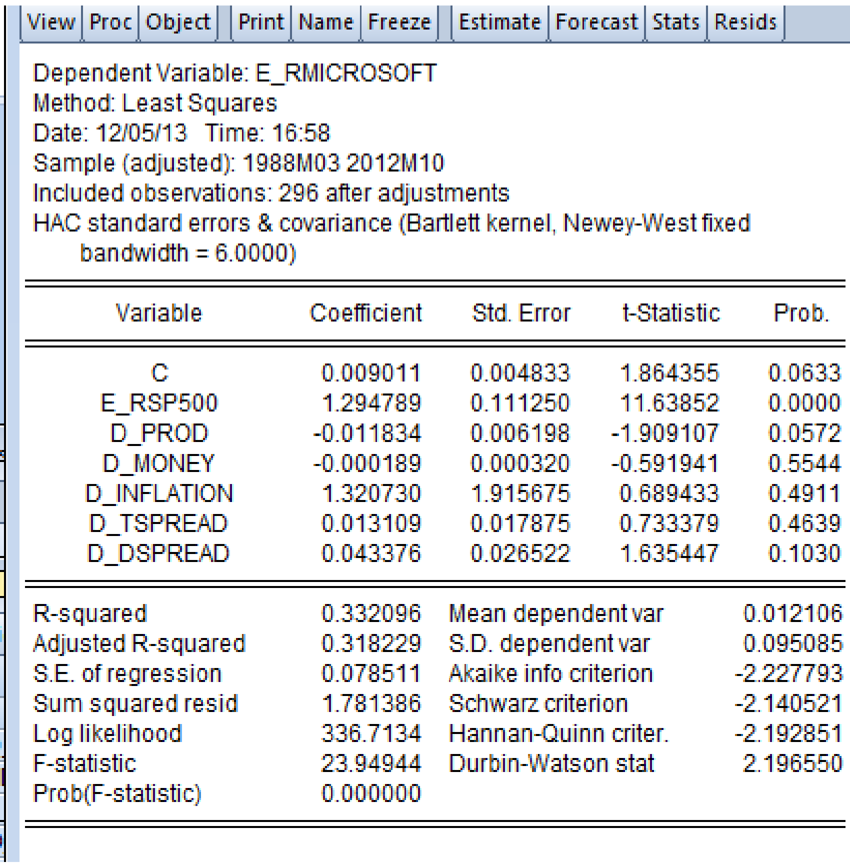
Both **F and chi-squared versions** of the test show that there is **no** apparent **non-linearity equation**; the relationship between the Microsoft excess returns and the explanatory variables is linear. The linear Model for the Microsoft returns is appropriate.

Testing for Stability

Testing for Stability

Stability or not???

Chow Tests : View → Stability Diagnostics → Chow Breakpoint Test

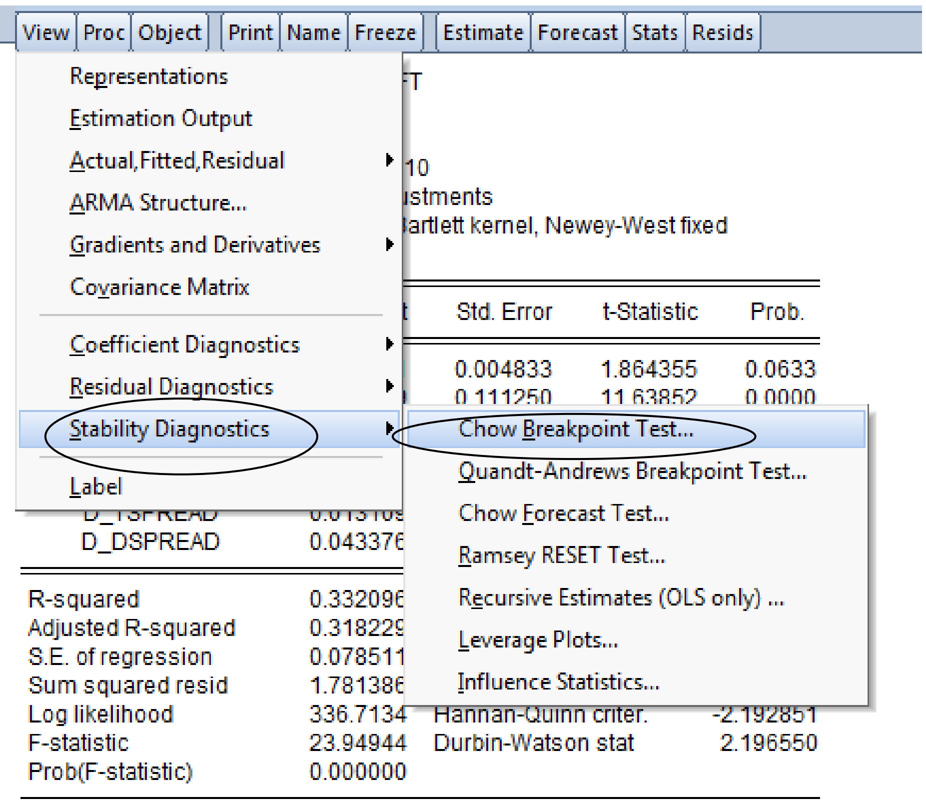


View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

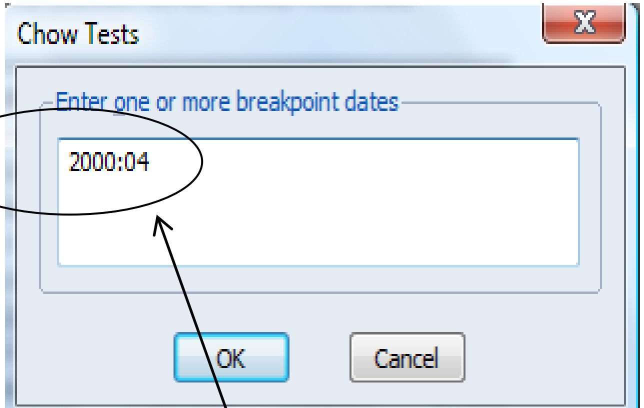
Dependent Variable: E_RMICROSOFT
Method: Least Squares
Date: 12/05/13 Time: 16:58
Sample (adjusted): 1988M03 2012M10
Included observations: 296 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

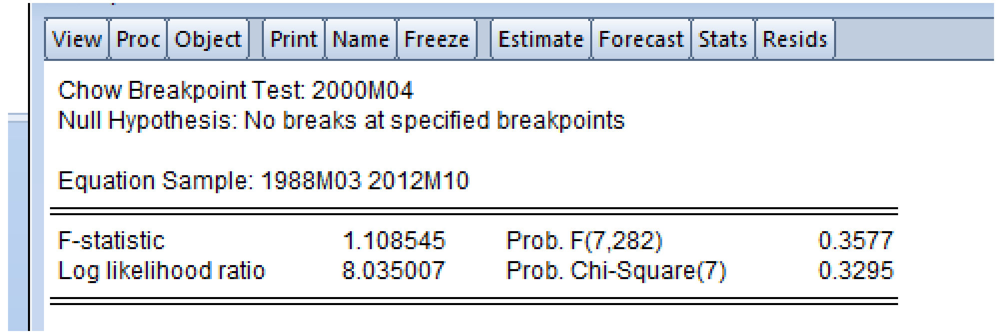
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.009011	0.004833	1.864355	0.0633
E_RSP500	1.294789	0.111250	11.63852	0.0000
D_PROD	-0.011834	0.006198	-1.909107	0.0572
D_MONEY	-0.000189	0.000320	-0.591941	0.5544
D_INFLATION	1.320730	1.915675	0.689433	0.4911
D_TSPREAD	0.013109	0.017875	0.733379	0.4639
D_DSPREAD	0.043376	0.026522	1.635447	0.1030

R-squared	0.332096	Mean dependent var	0.012106
Adjusted R-squared	0.318229	S.D. dependent var	0.095085
S.E. of regression	0.078511	Akaike info criterion	-2.227793
Sum squared resid	1.781386	Schwarz criterion	-2.140521
Log likelihood	336.7134	Hannan-Quinn criter.	-2.192851
F-statistic	23.94944	Durbin-Watson stat	2.196550
Prob(F-statistic)	0.000000		



Testing for Stability




View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Chow Breakpoint Test: 2000M04
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 1988M03 2012M10

F-statistic	1.108545	Prob. F(7,282)	0.3577
Log likelihood ratio	8.035007	Prob. Chi-Square(7)	0.3295

We split the sample roughly in half. You can also select a date of i.e. stock market crash etc…

P-values are in excess of 0.05
We cannot reject the Null- Hypothesis;
thus parameters are stable
across the two sub-samples

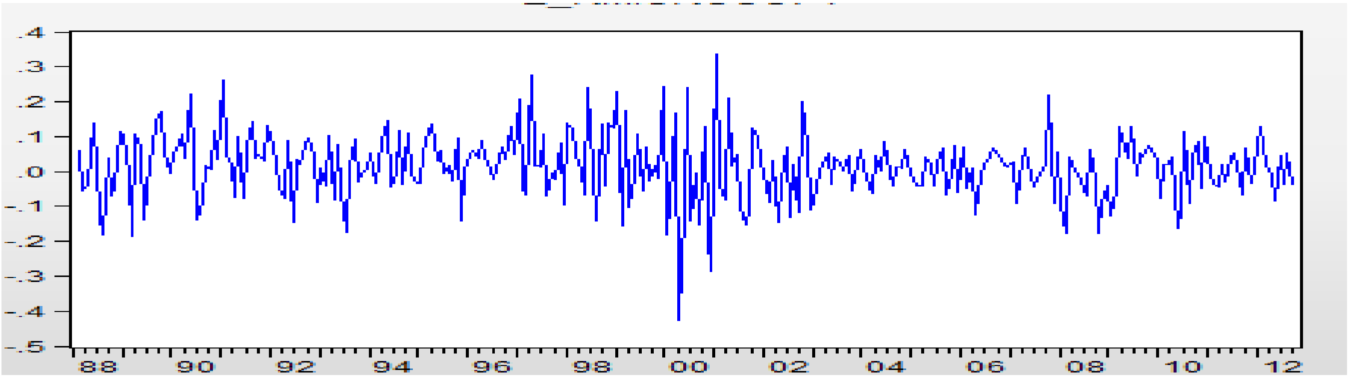
$$H_0 : parameters \text{ are stable across the two sub-samples}$$

$$H_A : parameters \text{ are not stable across the two sub-samples}$$

Testing for Stability

How can the appropriate sub-samples to use be decided?

1. Plot the dependent variable over time and split the data accordingly **to any obvious structural changes** in the series.



2. Split the data accordingly **to any known important historical event** (stock market crash, market microstructure change, new government elected)

You can also use **the last few observations** and perform **a forwards predictive failure test** or the *first few observations* for a *backwards predictive failure test.*

Univariate Time Series Modelling

Univariate Time Series Modeling

- Autoregressive process AR(1)

Open empirical_example_9.wf1 → Quick → Estimate equation

$$Y_t = a + b_1 Y_{t-1} + u_t$$

When you are estimating an AR equation, you do not say that Y follows an AR process, but you say that the error terms follow an AR process.

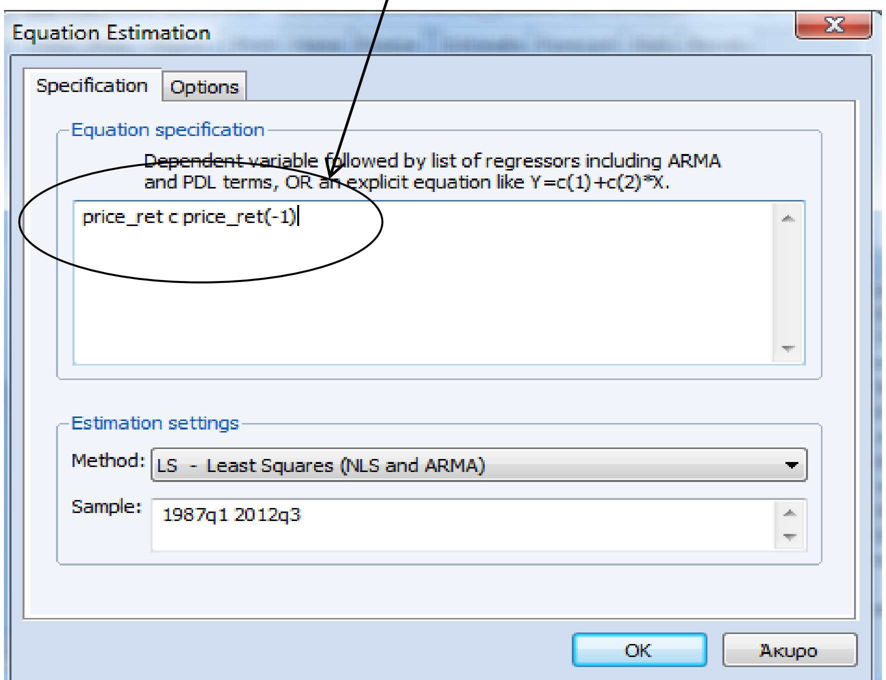
$$Y_t = a + u_t$$

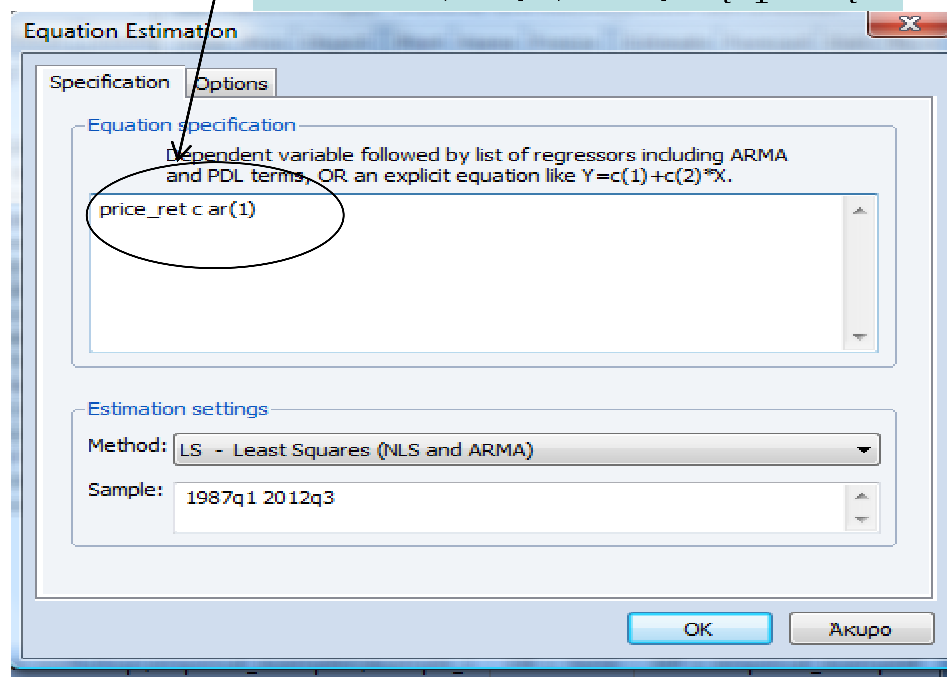
$$u_t = \rho u_{t-1} + v_t$$

$$Y_t = a + \rho u_{t-1} + v_t =$$

$$= a + \rho(Y_{t-1} - a) + v_t =$$

$$= (1-\rho)a + \rho Y_{t-1} + v_t$$





Estimation Outputs

Equation: UNTITLED   Workfile: EMPIRICAL_EXAMPLE9::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PRICE_RET
Method: Least Squares
Date: 12/11/12   Time: 16:15
Sample (adjusted): 1987Q3 2012Q3
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003559	0.002249	1.582286	0.1168
PRICE_RET(-1)	0.512973	0.085864	5.974276	0.0000

R-squared	0.264990	Mean dependent var	0.007425
Adjusted R-squared	0.257565	S.D. dependent var	0.025121
S.E. of regression	0.021646	Akaike info criterion	-4.808412
Sum squared resid	0.046385	Schwarz criterion	-4.756627
Log likelihood	244.8248	Hannan-Quinn criter.	-4.787448
F-statistic	35.69198	Durbin-Watson stat	1.888812
Prob(F-statistic)	0.000000		

Equation: EQ01   Workfile: EMPIRICAL_EXAMPLE9::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PRICE_RET
Method: Least Squares
Date: 12/11/12   Time: 17:10
Sample (adjusted): 1987Q3 2012Q3
Included observations: 101 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007307	0.004423	1.652109	0.1017
AR(1)	0.512973	0.085864	5.974276	0.0000

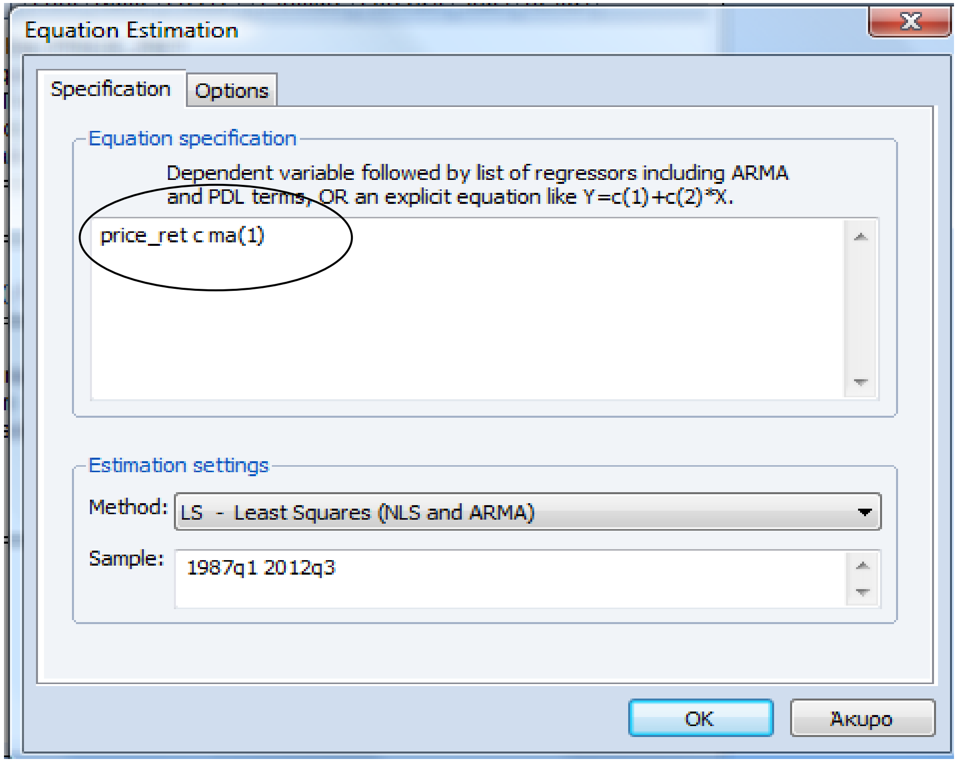
R-squared	0.264990	Mean dependent var	0.007425
Adjusted R-squared	0.257565	S.D. dependent var	0.025121
S.E. of regression	0.021646	Akaike info criterion	-4.808412
Sum squared resid	0.046385	Schwarz criterion	-4.756627
Log likelihood	244.8248	Hannan-Quinn criter.	-4.787448
F-statistic	35.69198	Durbin-Watson stat	1.888812
Prob(F-statistic)	0.000000		

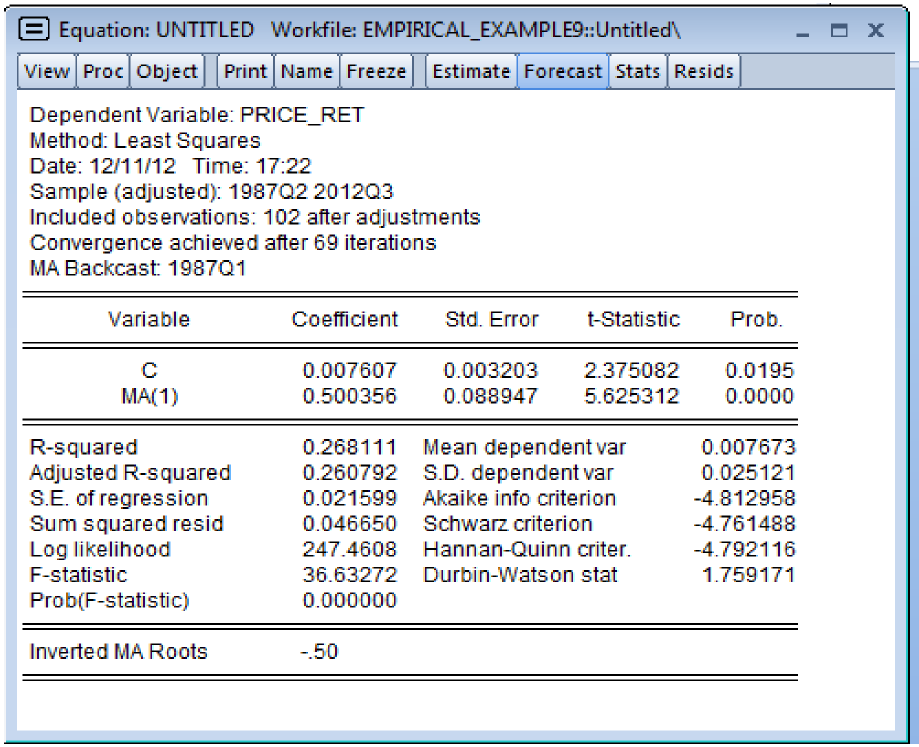
Inverted AR Roots   .51

Identical slope coefficients; different constant terms

- Moving Average Process MA(1)

$$Y_t = a + u_t$$
$$u_t = v_t + \theta v_{t-1}$$





Dependent Variable: PRICE_RET
Method: Least Squares
Date: 12/11/12 Time: 17:22
Sample (adjusted): 1987Q2 2012Q3
Included observations: 102 after adjustments
Convergence achieved after 69 iterations
MA Backcast: 1987Q1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007607	0.003203	2.375082	0.0195
MA(1)	0.500356	0.088947	5.625312	0.0000

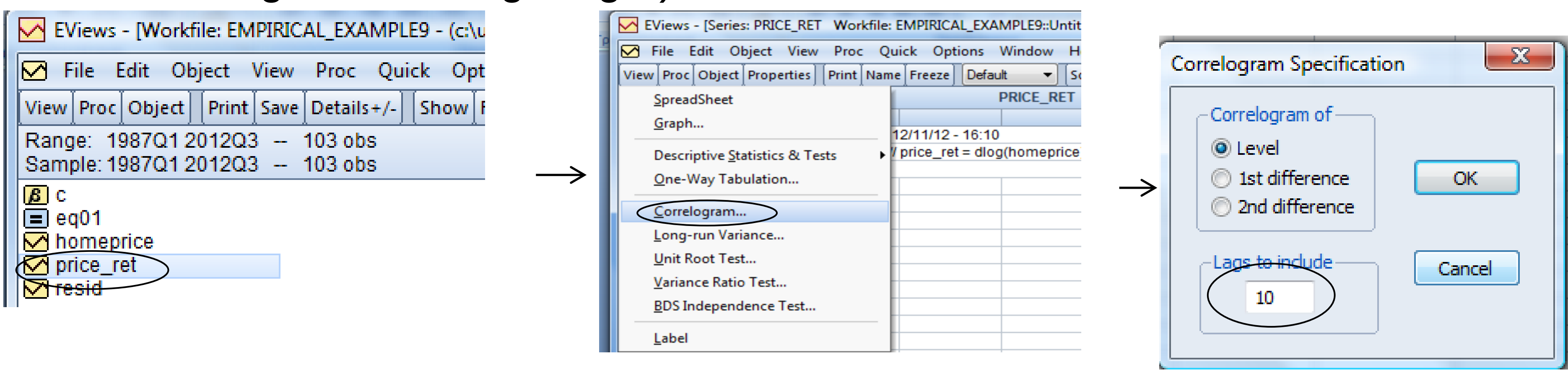
R-squared	0.268111	Mean dependent var	0.007673
Adjusted R-squared	0.260792	S.D. dependent var	0.025121
S.E. of regression	0.021599	Akaike info criterion	-4.812958
Sum squared resid	0.046650	Schwarz criterion	-4.761488
Log likelihood	247.4608	Hannan-Quinn criter.	-4.792116
F-statistic	36.63272	Durbin-Watson stat	1.759171
Prob(F-statistic)	0.000000		

Inverted MA Roots -.50

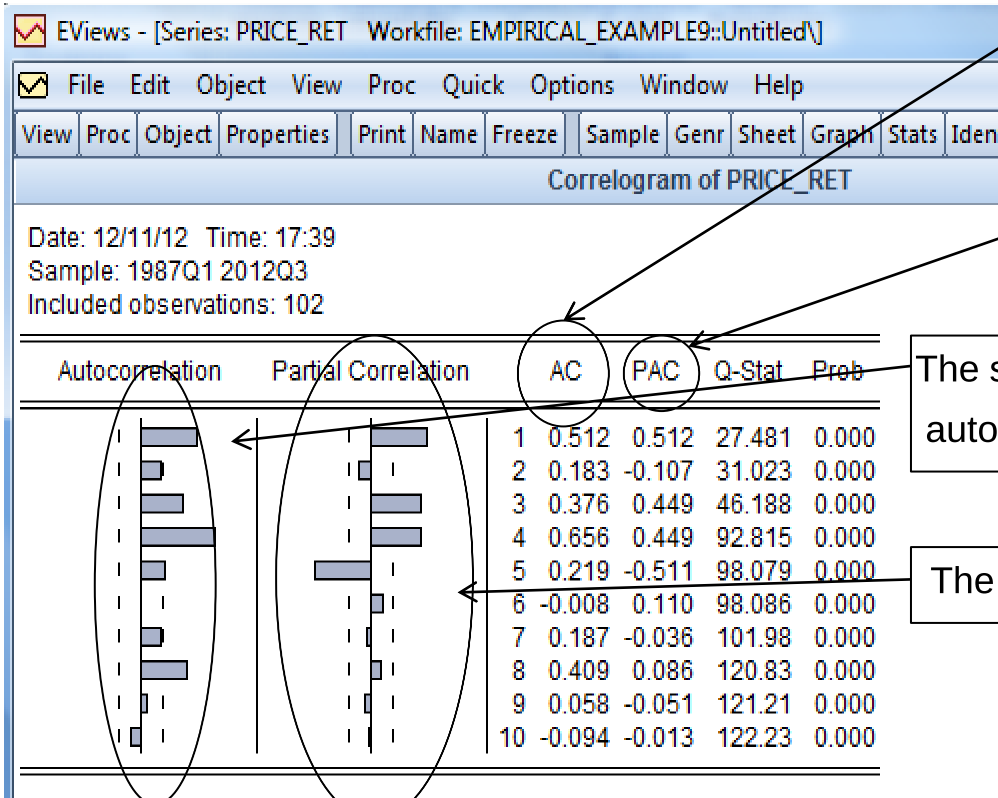
- **ARMA model :** Combination of autoregressive and moving average processes

How to build an ARMA model ??

a. Identification (autocorrelation function **(ACF)**, the partial autocorrelation function **(PACF),** and the resulting **correlograms**(plots of ACFs and PACFs against the lag length).



ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS


EViews - [Series: PRICE_RET Workfile: EMPIRICAL_EXAMPLE9::Untitled\]

File Edit Object View Proc Quick Options Window Help

View Proc Object Properties Print Name Freeze Sample Genr Sheet Graph Stats Ident

Correlogram of PRICE_RET

Date: 12/11/12 Time: 17:39
Sample: 1987Q1 2012Q3
Included observations: 102

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.512	0.512	27.481	0.000
		2	0.183	-0.107	31.023	0.000
		3	0.376	0.449	46.188	0.000
		4	0.656	0.449	92.815	0.000
		5	0.219	-0.511	98.079	0.000
		6	-0.008	0.110	98.086	0.000
		7	0.187	-0.036	101.98	0.000
		8	0.409	0.086	120.83	0.000
		9	0.058	-0.051	121.21	0.000
		10	-0.094	-0.013	122.23	0.000

Autocorrelation function

Partial autocorrelation function

The series are persistent;
autocorrelation dies away quite slowly

The first 5 partial autocorrelation coeffs. are significant

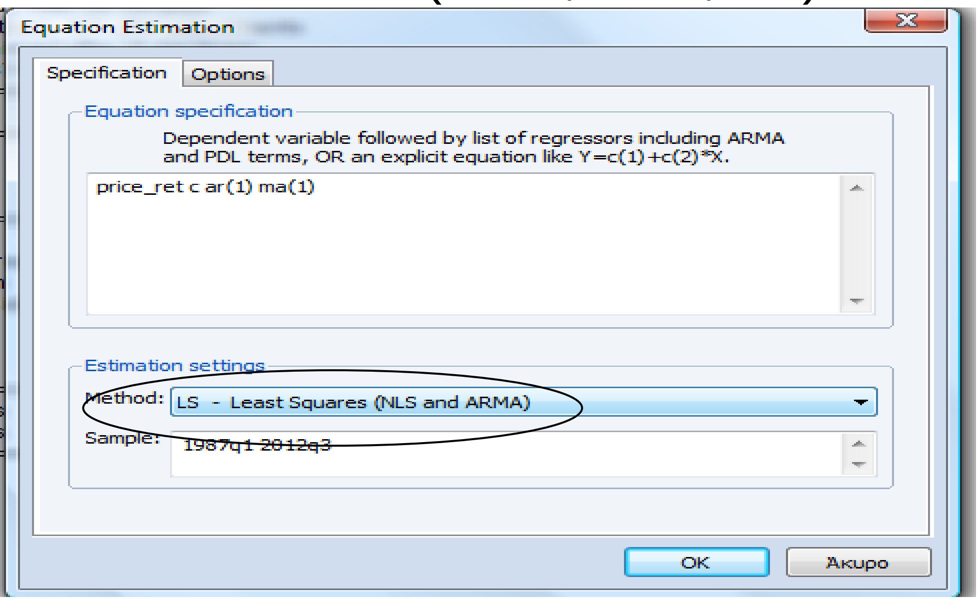
Rule of thumb : An autocorrelation coef .is classed as significant if it **outside** $\pm 1.96 \times \frac{1}{T^{1/2}}$

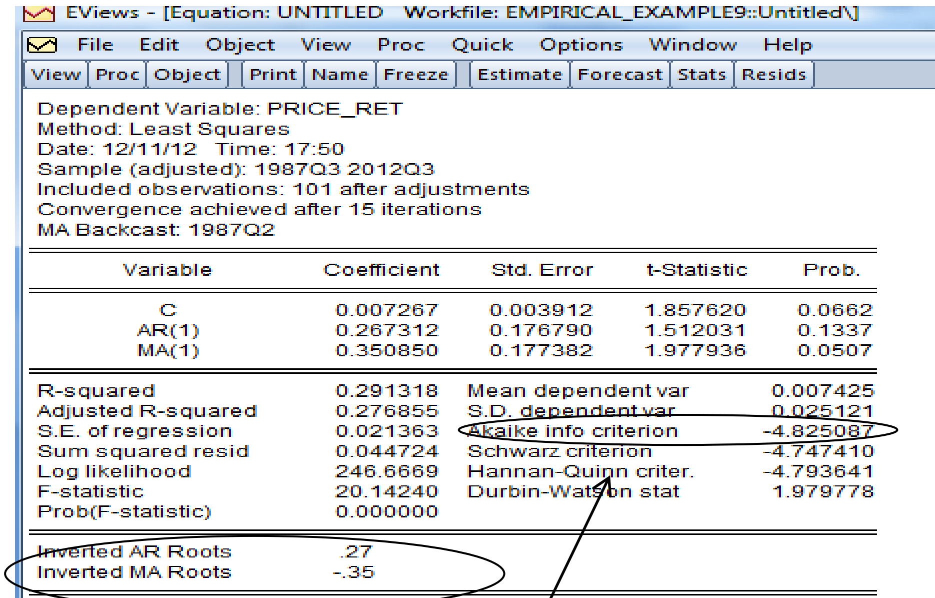
Here ± 0.19

Since the first acf is highly significant , Ljung –Box joint test statistic reject the Null hypothesis of no autocorrelation for all numbers of lags considered

Maybe an ARMA model is appropriate…

b. Estimation (OLS , MLE, etc)





Dependent Variable: PRICE_RET
Method: Least Squares
Date: 12/11/12 Time: 17:50
Sample (adjusted): 1987Q3 2012Q3
Included observations: 101 after adjustments
Convergence achieved after 15 iterations
MA Backcast: 1987Q2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.007267	0.003912	1.857620	0.0662
AR(1)	0.267312	0.176790	1.512031	0.1337
MA(1)	0.350850	0.177382	1.977936	0.0507

R-squared	0.291318	Mean dependent var	0.007425
Adjusted R-squared	0.276855	S.D. dependent var	0.025121
S.E. of regression	0.021363	Akaike info criterion	-4.825087
Sum squared resid	0.044724	Schwarz criterion	-4.747410
Log likelihood	246.6669	Hannan-Quinn criter.	-4.793641
F-statistic	20.14240	Durbin-Watson stat	1.979778
Prob(F-statistic)	0.000000		

Inverted AR Roots	.27
Inverted MA Roots	-.35

How to define the order??

Answer :Choose the model that ***minimizes*** *the value of Information Criteria (IC):* ***We use Akaike Information Criterion***

Bellow 1 = stationarity

- c. Diagnostic Checking (**Overfitting** and **residual diagnostics**)

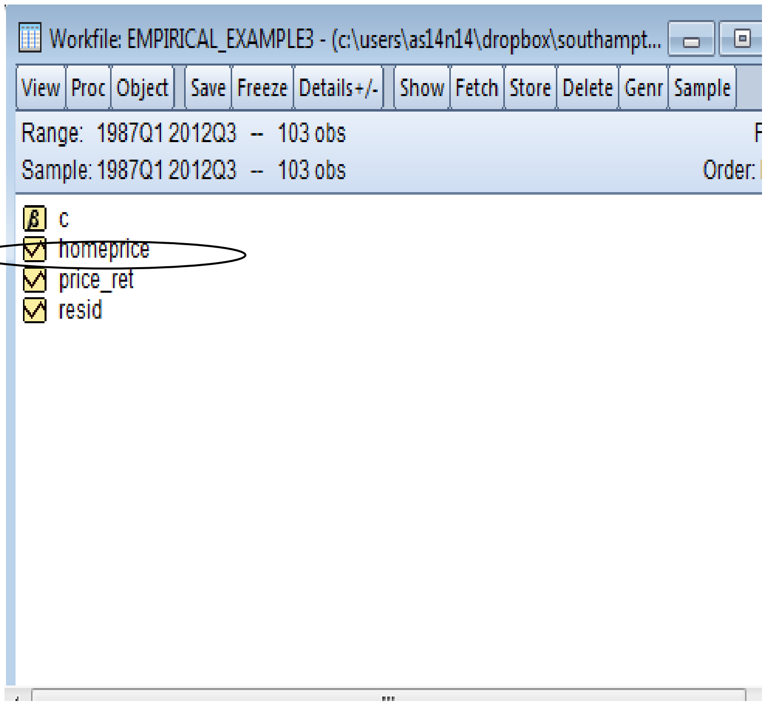
Overfitting : *If the model you specified at step 1 is adequate any extra terms added to the ARMA model will be insignificant*

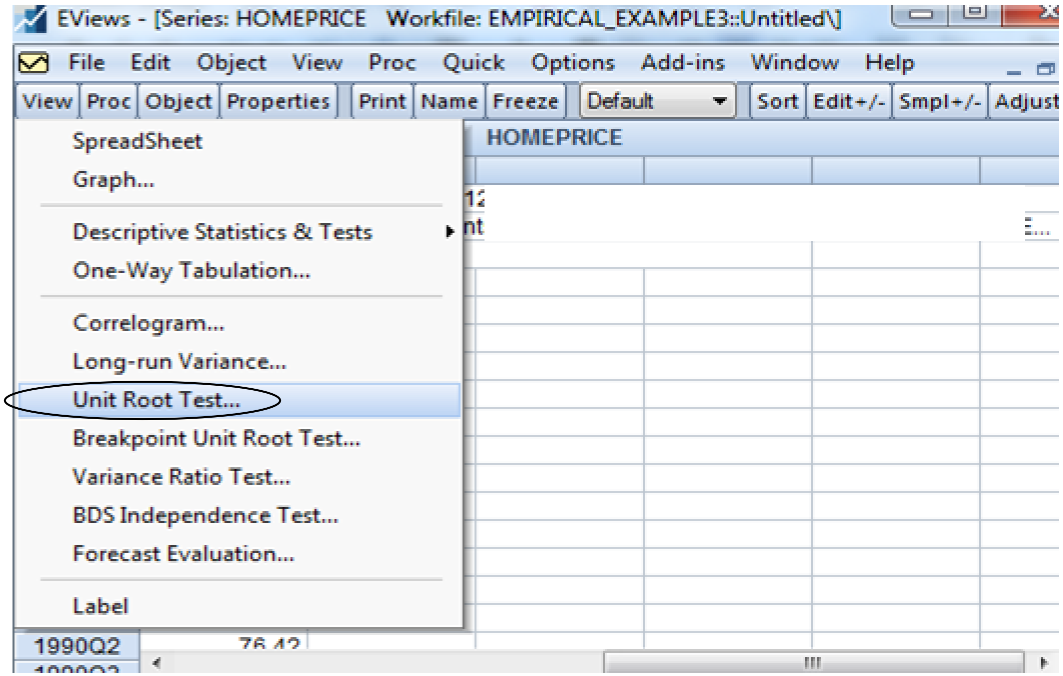
Residual Diagnostics: *If residuals are free from autocorrelation the model you specified at step 1 is adequate*

Stationarity

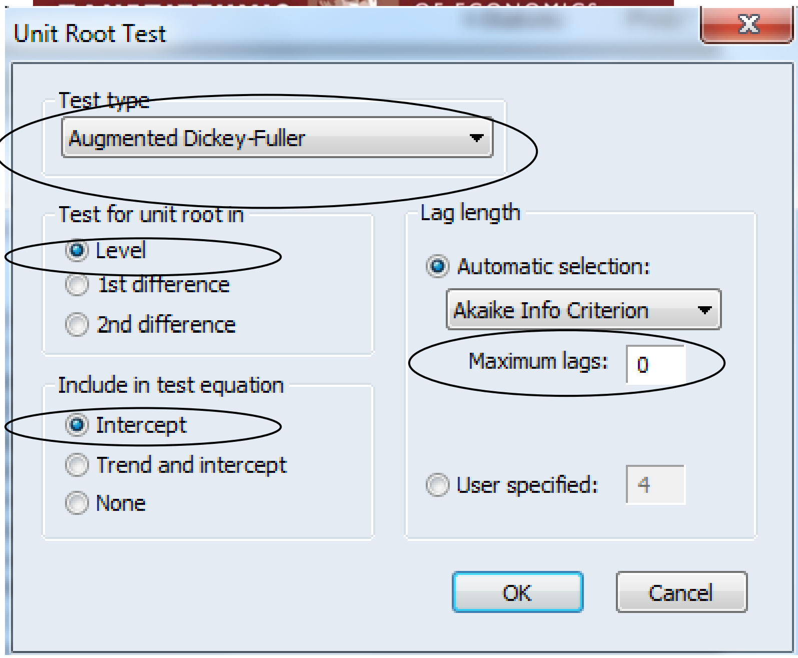
- ***Open Example 9.wf1***

- Click on *homeprice* ⟶ View ⟶ Unit Root Test





Stationarity – case 1



Null Hypothesis: HOMEPRICE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.103222	0.7124
Test critical values:	1% level	-3.495677	
	5% level	-2.890037	
	10% level	-2.582041	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(HOMEPRICE)
Method: Least Squares
Date: 12/02/17 Time: 10:43
Sample (adjusted): 1987Q2 2012Q3
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HOMEPRICE(-1)	-0.010030	0.009092	-1.103222	0.2726
C	1.825857	1.057967	1.725816	0.0875

R-squared	0.012025	Mean dependent var	0.721961
Adjusted R-squared	0.002145	S.D. dependent var	3.474102
S.E. of regression	3.470374	Akaike info criterion	5.345815
Sum squared resid	1204.349	Schwarz criterion	5.397285
Log likelihood	-270.6366	Hannan-Quinn criter.	5.366657
F-statistic	1.217100	Durbin-Watson stat	0.830034
Prob(F-statistic)	0.272578		

Null hypothesis: Unit Root

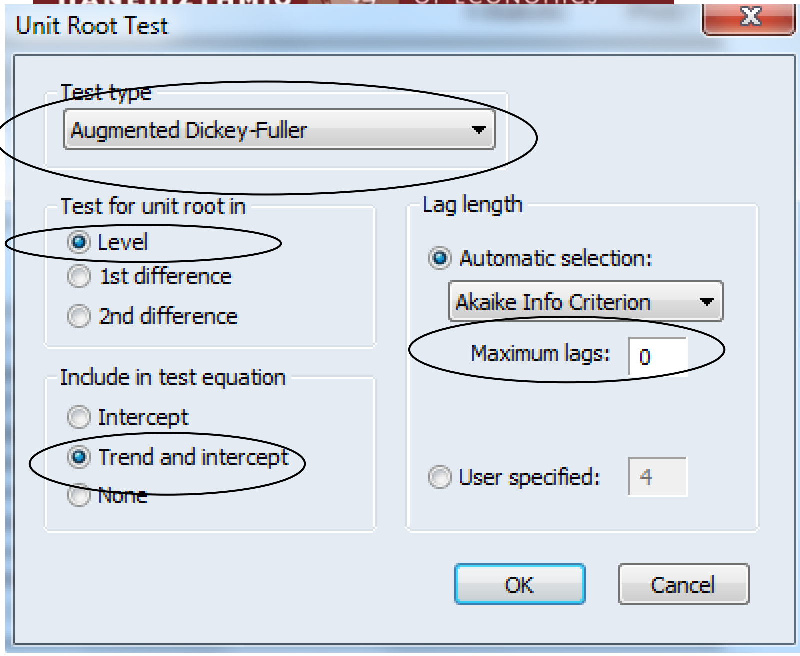
Alt hypothesis: Not unit Root

P-value = 0.7124> 0.05

Do not reject the Null Hypothesis, hence HOMEPRICE is non-stationary

$$\Delta y_t = \mu + y_{t-1} + u_t$$

Stationarity – case 2



Null Hypothesis: HOMEPRICE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.262097	0.9908
Test critical values:	1% level	-4.050509	
	5% level	-3.454471	
	10% level	-3.152909	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(HOMEPRICE)
Method: Least Squares
Date: 12/02/17 Time: 10:42
Sample (adjusted): 1987Q2 2012Q3
Included observations: 102 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
HOMEPRICE(-1)	-0.004409	0.016823	-0.262097	0.7938
C	1.649642	1.151084	1.433120	0.1550
@TREND("1987Q1")	-0.008591	0.021594	-0.397828	0.6916

R-squared	0.013602	Mean dependent var	0.721961
Adjusted R-squared	-0.006326	S.D. dependent var	3.474102
S.E. of regression	3.485072	Akaike info criterion	5.363825
Sum squared resid	1202.427	Schwarz criterion	5.441030
Log likelihood	-270.5551	Hannan-Quinn criter.	5.395088
F-statistic	0.682561	Durbin-Watson stat	0.835999
Prob(F-statistic)	0.507683		

Null hypothesis: Unit Root

Alt hypothesis: Not unit Root

P-value = 0.9908> 0.05

Do not reject the Null Hypothesis, hence HOMEPRICE is non-stationary

$$\Delta y_t = \mu + y_{t-1} + \beta t + u_t$$

Stationarity

Question: **How might you transform the series to remove any unit root?**

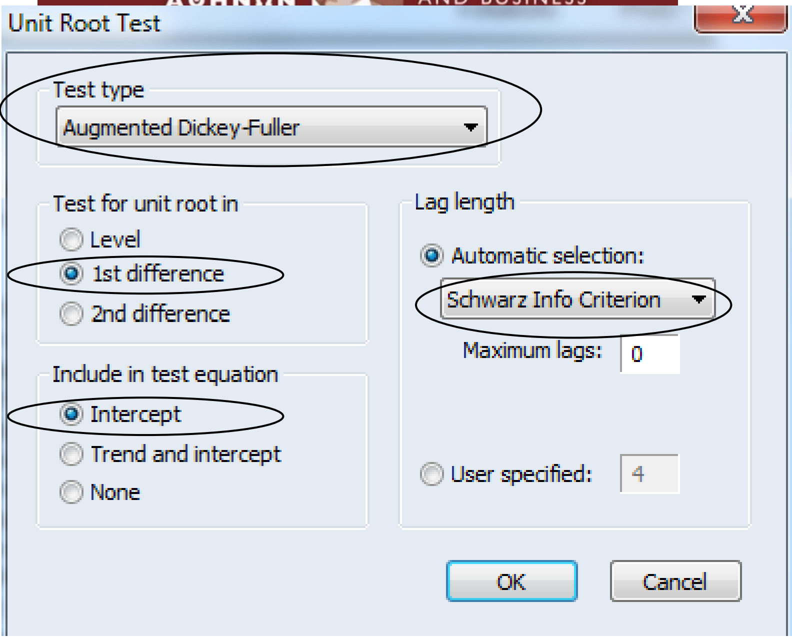
Answer: **Variables can usually be made stationary by transforming them into their differences or by constructing percentage changes of them. It is common price indices to be converted to returns by calculating the logarithmic differences.**

Example:

Click on *homeprice* ⟶ View ⟶ Unit Root

and the select 1st difference

Stationarity



Null Hypothesis: D(HOMEPRICE) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.078693	0.0000
Test critical values:	1% level	-3.496346	
	5% level	-2.890327	
	10% level	-2.582196	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(HOMEPRICE,2)
Method: Least Squares
Date: 12/02/17 Time: 10:53
Sample (adjusted): 1987Q3 2012Q3
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(HOMEPRICE(-1))	-0.414527	0.081621	-5.078693	0.0000
C	0.298418	0.288744	1.033505	0.3039

R-squared	0.206687	Mean dependent var	0.007921
Adjusted R-squared	0.198674	S.D. dependent var	3.177428
S.E. of regression	2.844332	Akaike info criterion	4.948137
Sum squared resid	800.9325	Schwarz criterion	4.999922
Log likelihood	-247.8809	Hannan-Quinn criter.	4.969101
F-statistic	25.79312	Durbin-Watson stat	1.895662
Prob(F-statistic)	0.000002		

Null hypothesis: Unit Root

Alt hypothesis: Not unit Root

P-value = 0.000<0.05

Reject the Null Hypothesis, hence D(HOMEPRICE) is stationary

The end