FMBF: Computer lab 1

Introduction

There are four computer labs for the FMBF module. They will be organized as follows:

Practical 1: ARMA modeling and unit root test

Practical 2: Cointegration test

Practical 3: ARCH/GARCH modeling

Practical 4: Review

ARMA Modeling

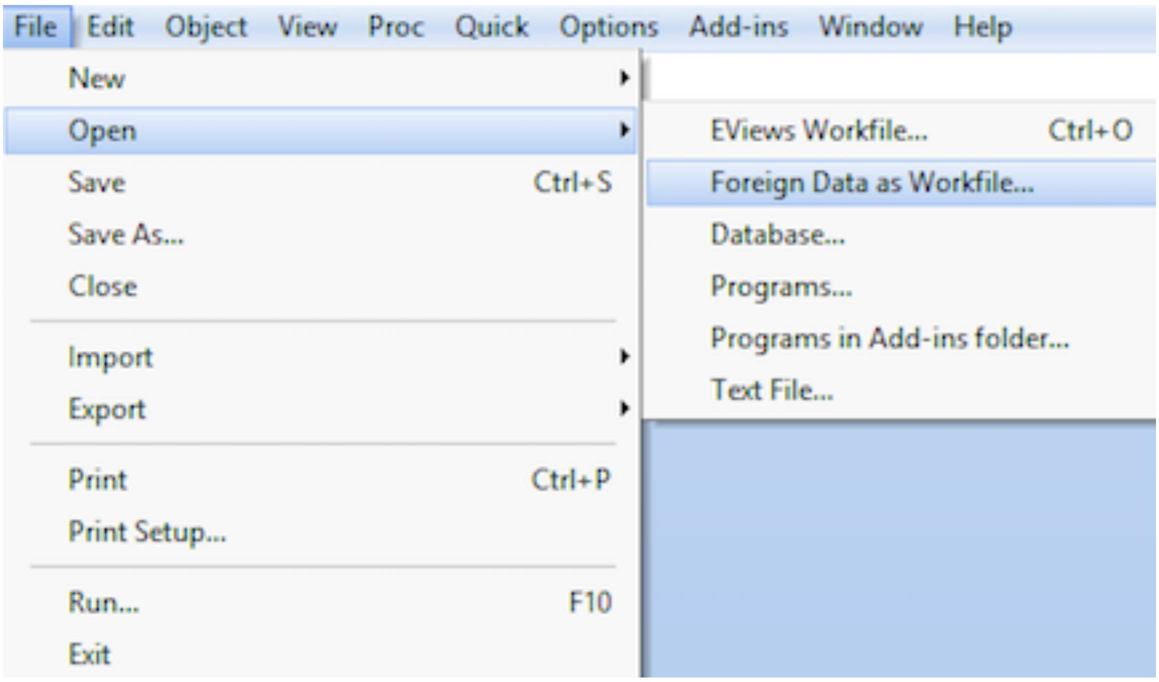
We will firstly load a dataset and build ARMA models by following the Box–Jenkins approach. The Box–Jenkins approach involves three steps:

1. Identification
2. Estimation
3. Diagnostic checking

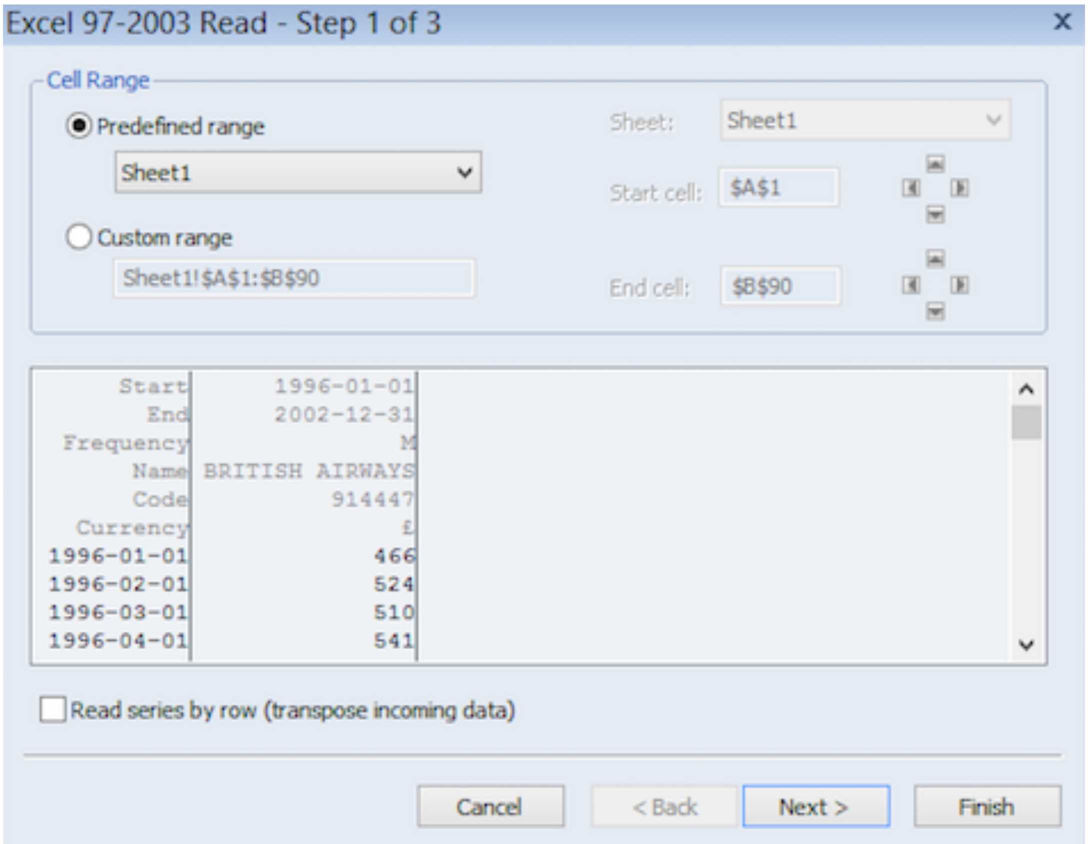
Data Preparation

1. We will use monthly data on the price of British Airways from 1996 to 2002 (Please download the file 'BA.xls' from DUO).
2. Open the file in EViews by following steps:

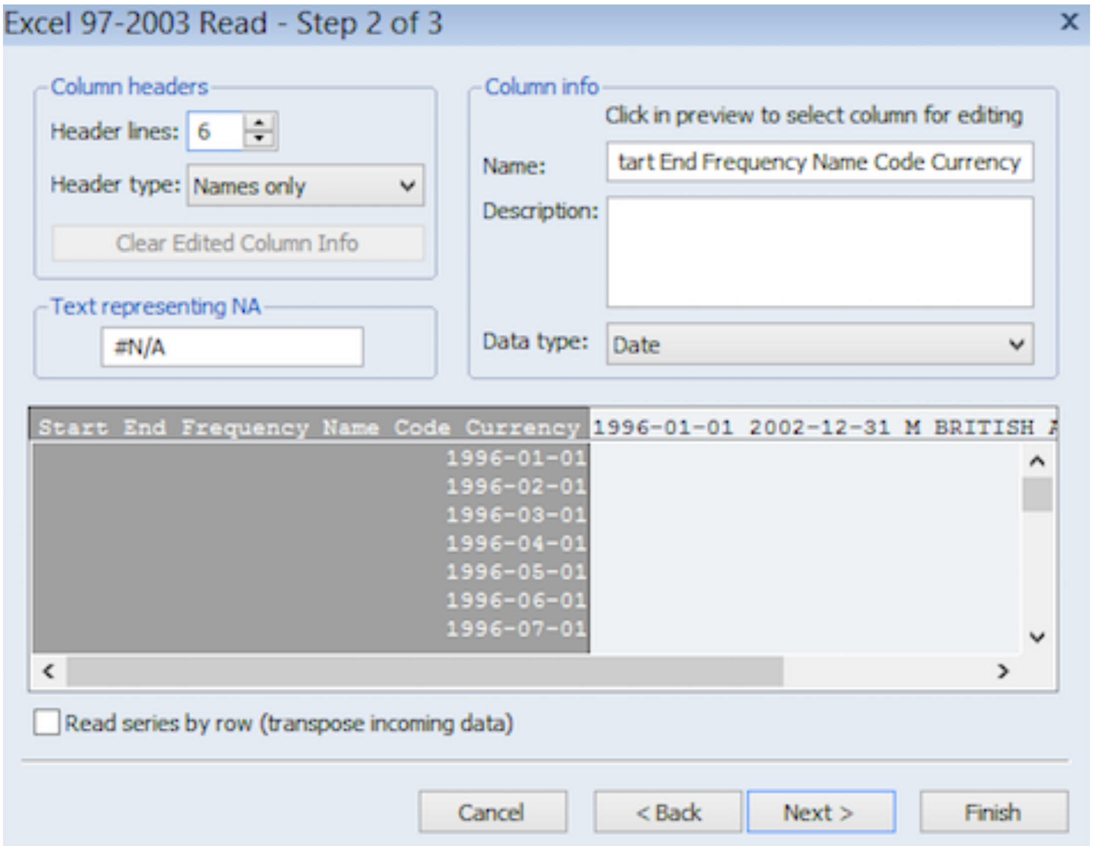
   (1) File > Open > Foreign Data as Workfile



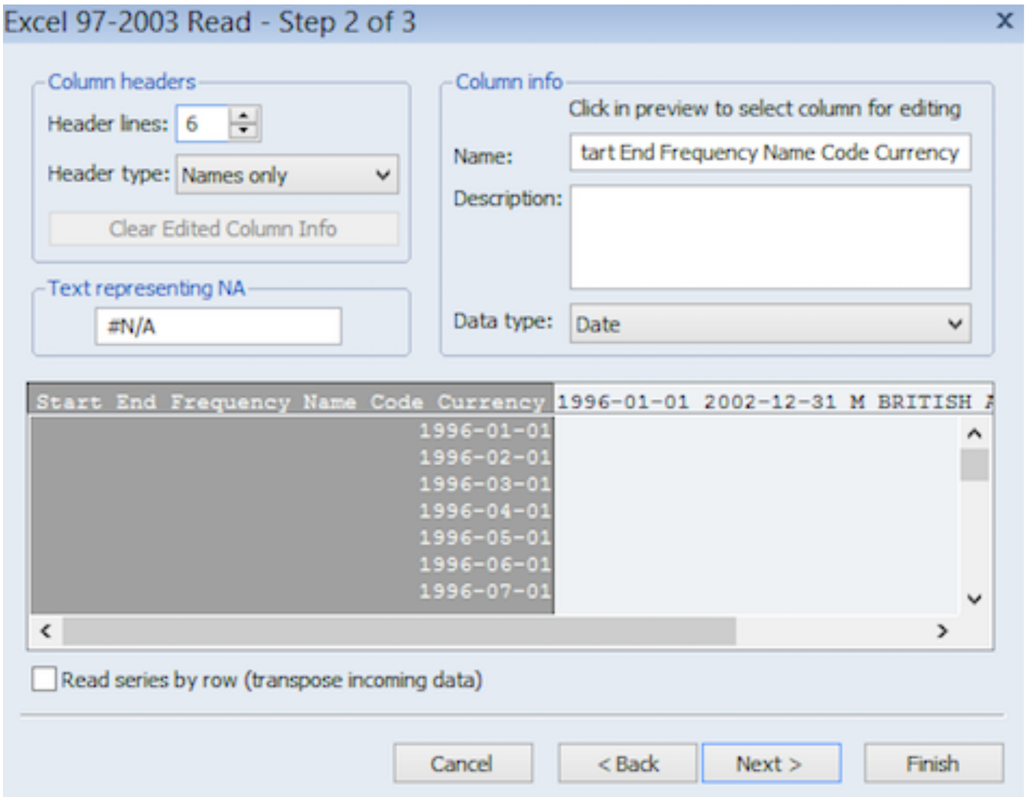
(2) Select the file ‘BA.xls’, and following window will appear. Click ‘Next’.

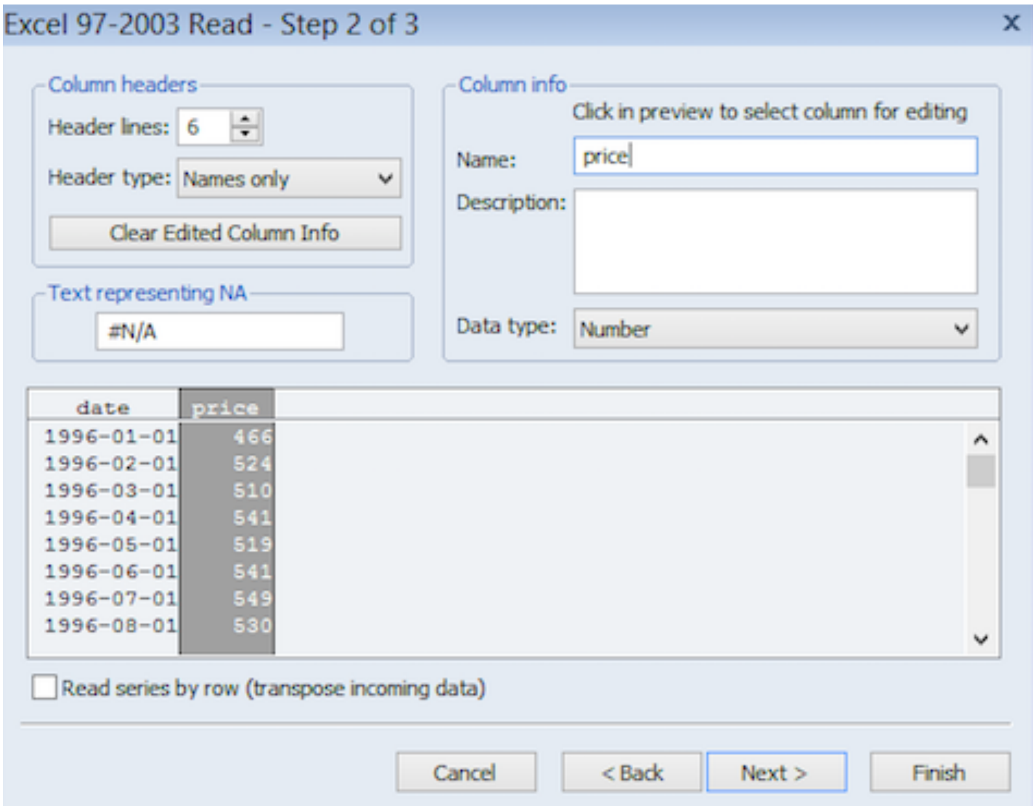


(3) Since the first 6 rows of the table are header lines, input 6 in the ‘Header lines’ box.

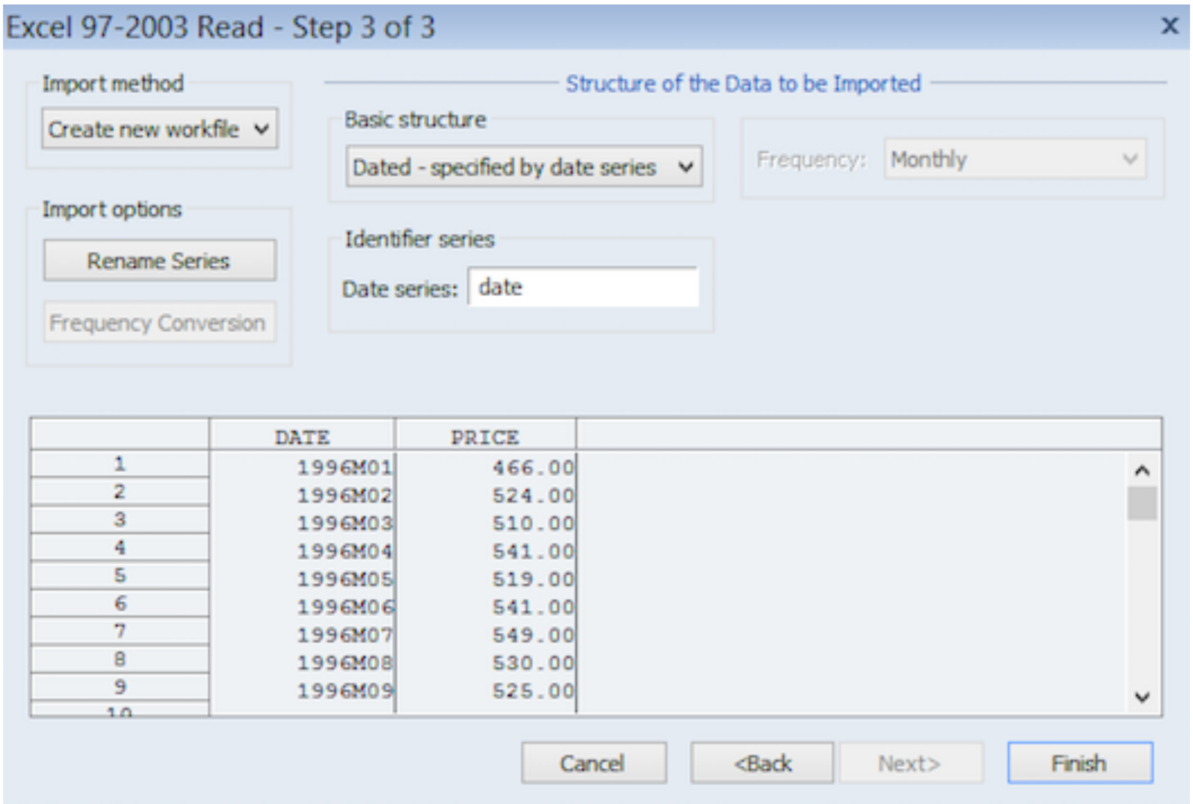


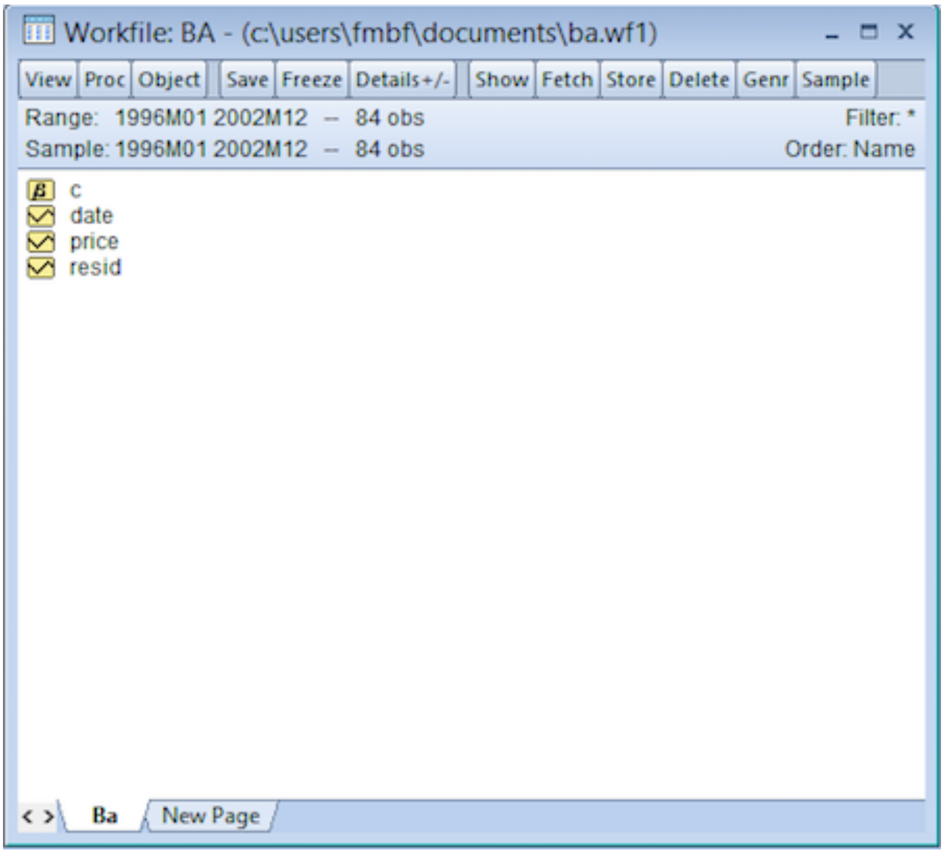
Then, change the variable names to ‘date’ and ‘price’. Please pay attention that the ‘Data type’ should be ‘Date’ and ‘Number’, respectively.





(4) Since the EViews has correctly recognized the 'Structure of the Data to be Imported', keep the 'Basic structure' as the 'Dated – specified by date series'. Click 'Finish', and finally we will get the 'Workfile: BA'.

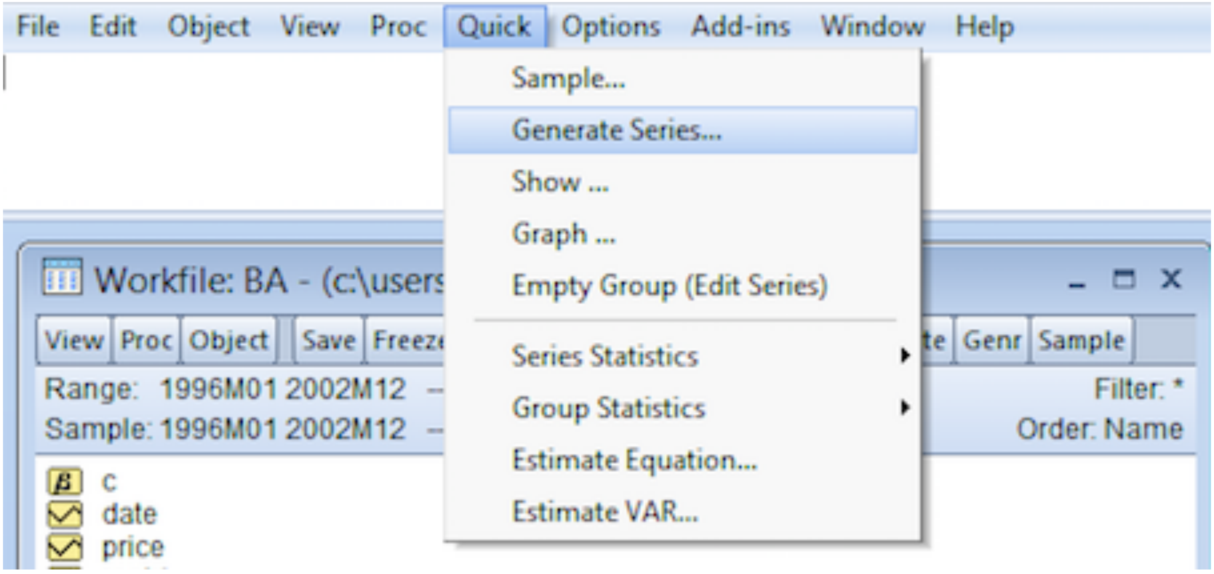




3. Generate two series – 'lnprice' (natural logarithm of price) and 'return' (first difference of lnprice). We calculate return by following equation:

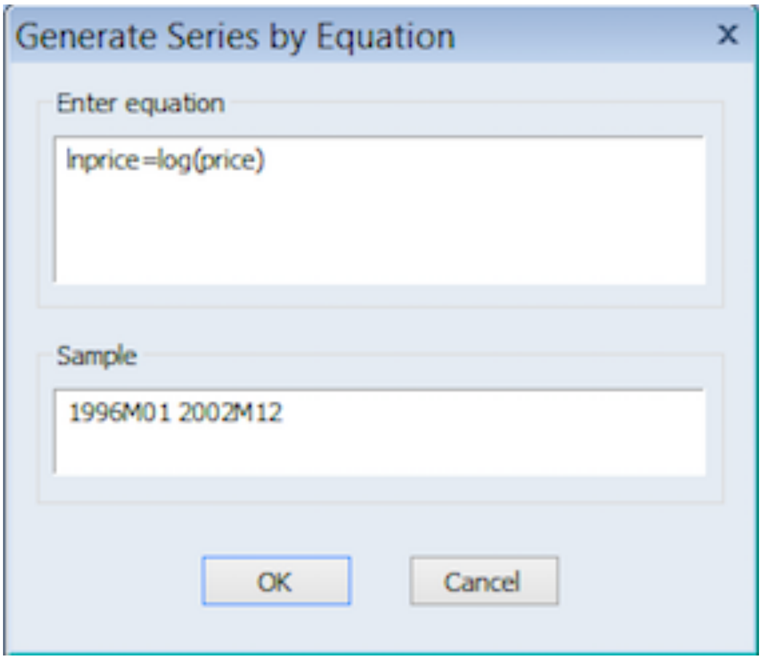
$$return = \ln(price_t) - \ln\,(price_{t-1})$$

(1) Quick > Generate Series



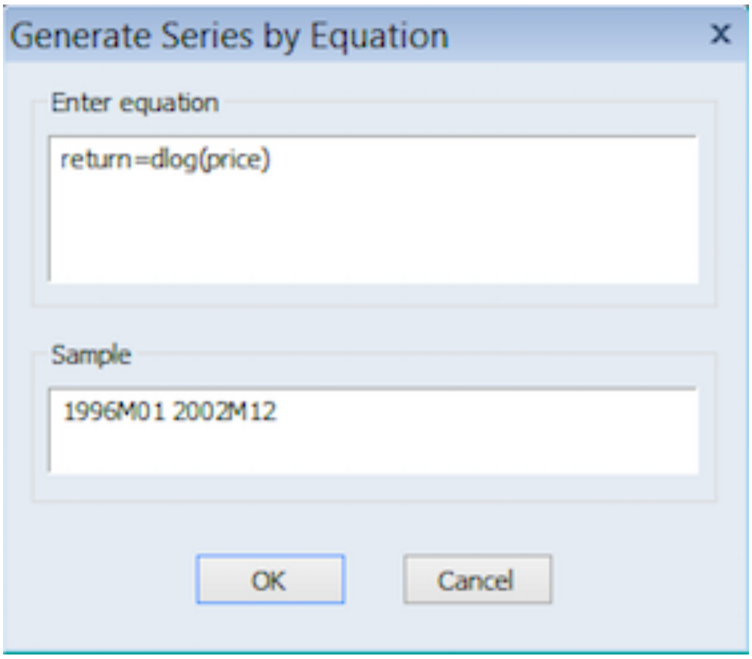
(2) Enter equation to generate 'lnprice':

$$lnprice = \log\,(price)$$



(3) Enter equation to generate 'return':

$$return = dlog(price)$$



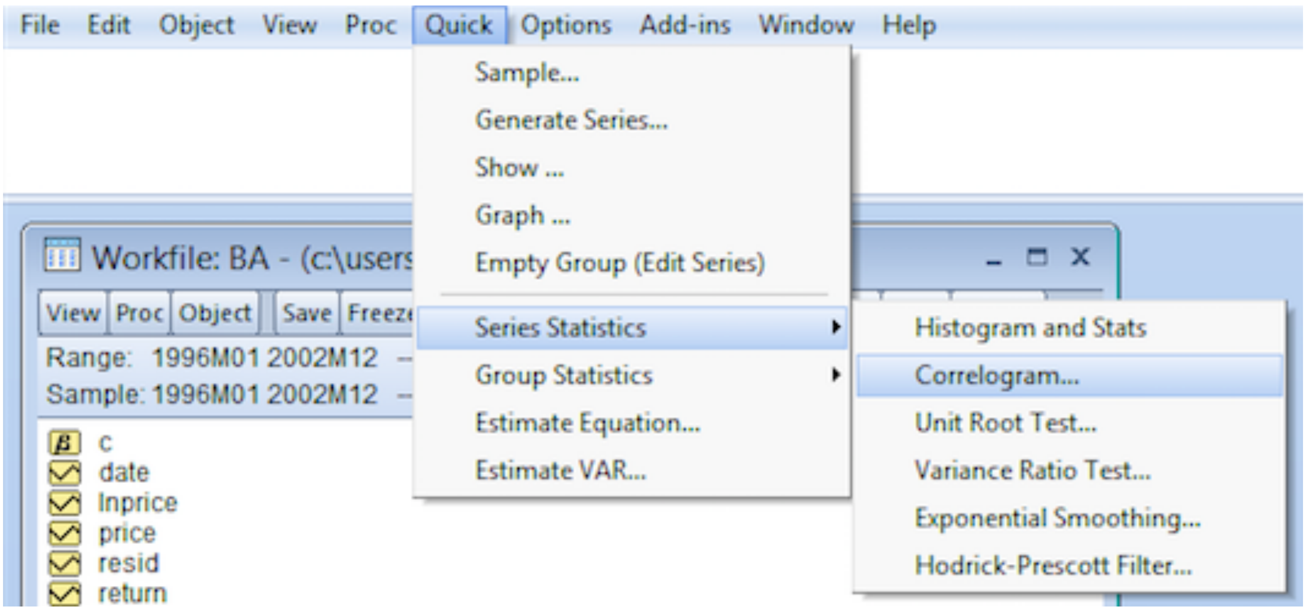
Box–Jenkins approach

Identification

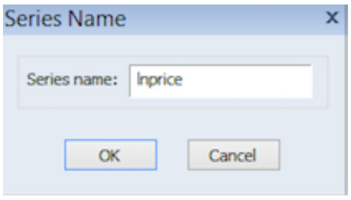
ACF/PACF Plots can be used to identify the order of ARMA(p,q) model.

We can draw ACF/PACF plots by following steps:

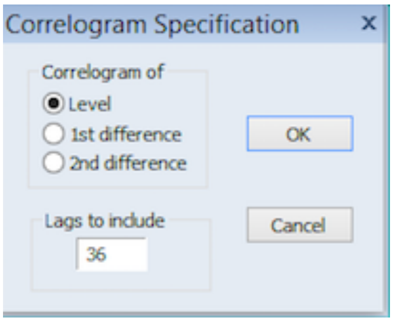
1. Quick > Series > Correlogram



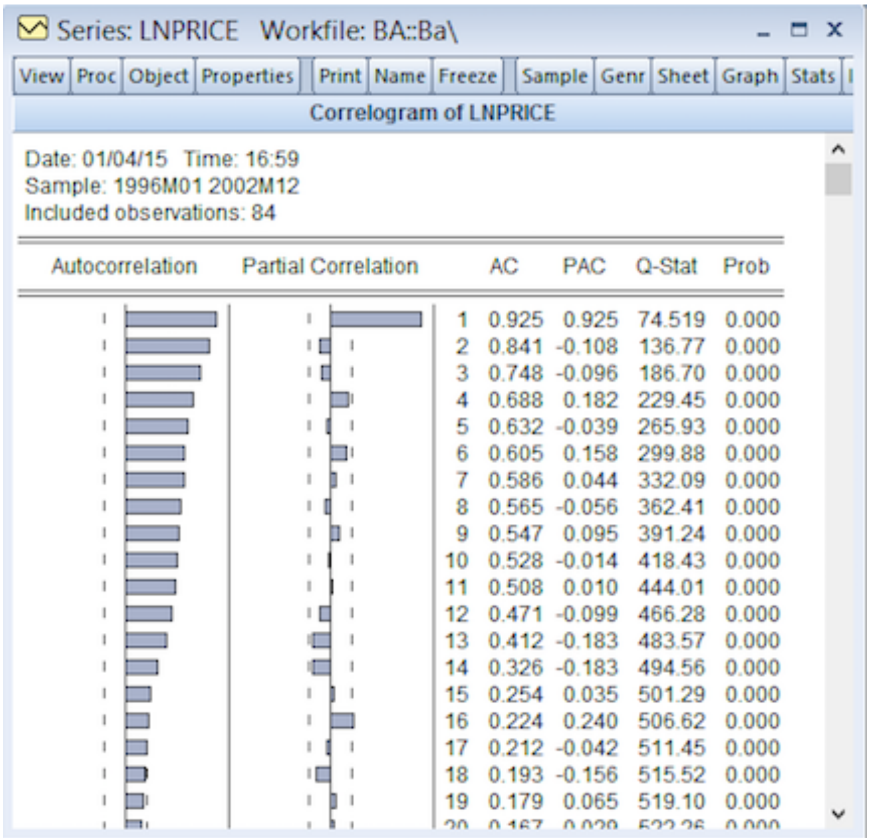
2. Input the name of series that you want to test. If you want to draw the ACF/PACF plots for 'lnprice', please input 'lnprice'. You can also input 'log(price)', and you will get the same results.



3. Choose the 'level' of the variable. If you choose the '1st difference', it will show the results of 'd(lnprice)'. In this case, return=d(lnprice)=dlog(price)

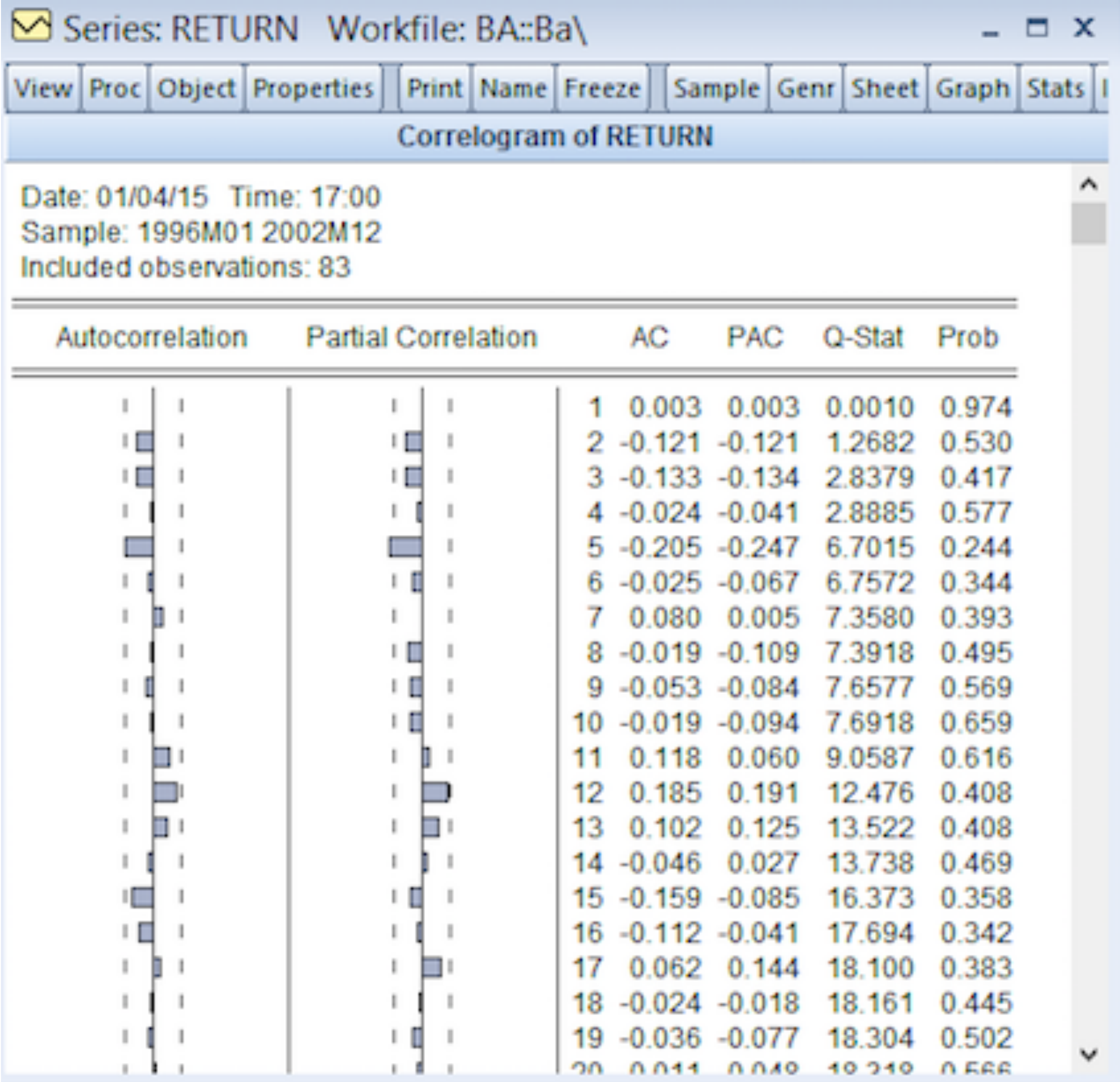


4. Finally, we will get the ACF/PACF plots for 'lnprice'.



The results suggest that the series 'lnprice' is a non-stationary process (random walk: $y_t = y_{t-1} + \varepsilon_t$). Can you explain the reason for this?

5. We can also draw the ACF/PACF plots for ‘return’ by repeating above process.



	AC	PAC	Q-Stat	Prob
1	0.003	0.003	0.0010	0.974
2	-0.121	-0.121	1.2682	0.530
3	-0.133	-0.134	2.8379	0.417
4	-0.024	-0.041	2.8885	0.577
5	-0.205	-0.247	6.7015	0.244
6	-0.025	-0.067	6.7572	0.344
7	0.080	0.005	7.3580	0.393
8	-0.019	-0.109	7.3918	0.495
9	-0.053	-0.084	7.6577	0.569
10	-0.019	-0.094	7.6918	0.659
11	0.118	0.060	9.0587	0.616
12	0.185	0.191	12.476	0.408
13	0.102	0.125	13.522	0.408
14	-0.046	0.027	13.738	0.469
15	-0.159	-0.085	16.373	0.358
16	-0.112	-0.041	17.694	0.342
17	0.062	0.144	18.100	0.383
18	-0.024	-0.018	18.161	0.445
19	-0.036	-0.077	18.304	0.502

Nearly all the ACF/PACF are insignificant. Only the ACF and PACF at the lag five are significant (they are outside the dotted lines in the graph). Is there any particular ARMA(p,q) models suggested by the ACF/PACF plots?

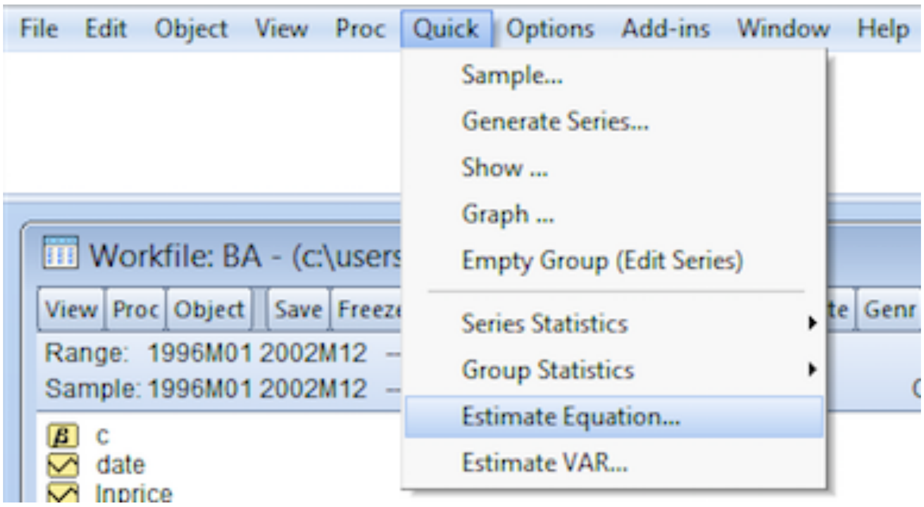
Note:

Sometimes, it is difficult to use ACF/PACF plots to choose the model order. Information criteria, such as AIC, SBIC and HQIC, can be used to specify the model. Specifically, the optimal specification should minimize the value of an information criterion. Information criteria can be got after you estimate the ARMA(p,q) model.

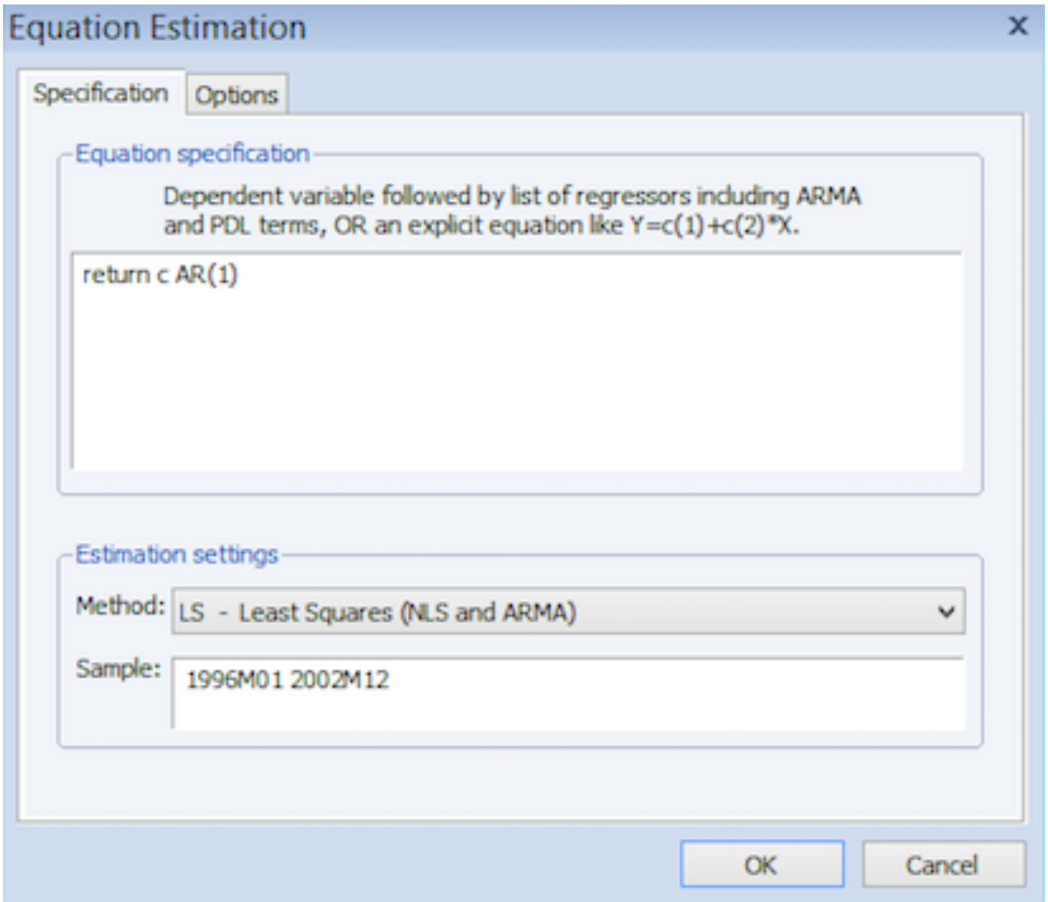
Estimation

Suppose that ARMA models from order (0,0) to (5,5) are plausible for the stock returns of BA. In EViews, this can be done by separately estimating each of the models and noting down the value of the information criteria in each case. We can construct ARMA(p,q) model by following steps:

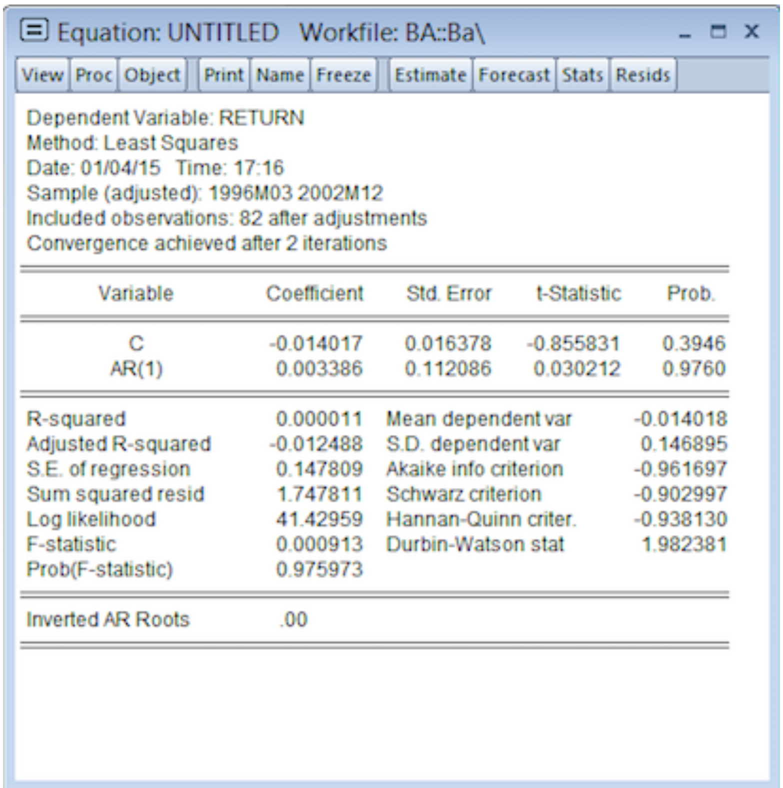
1. Quick > Estimate Equation



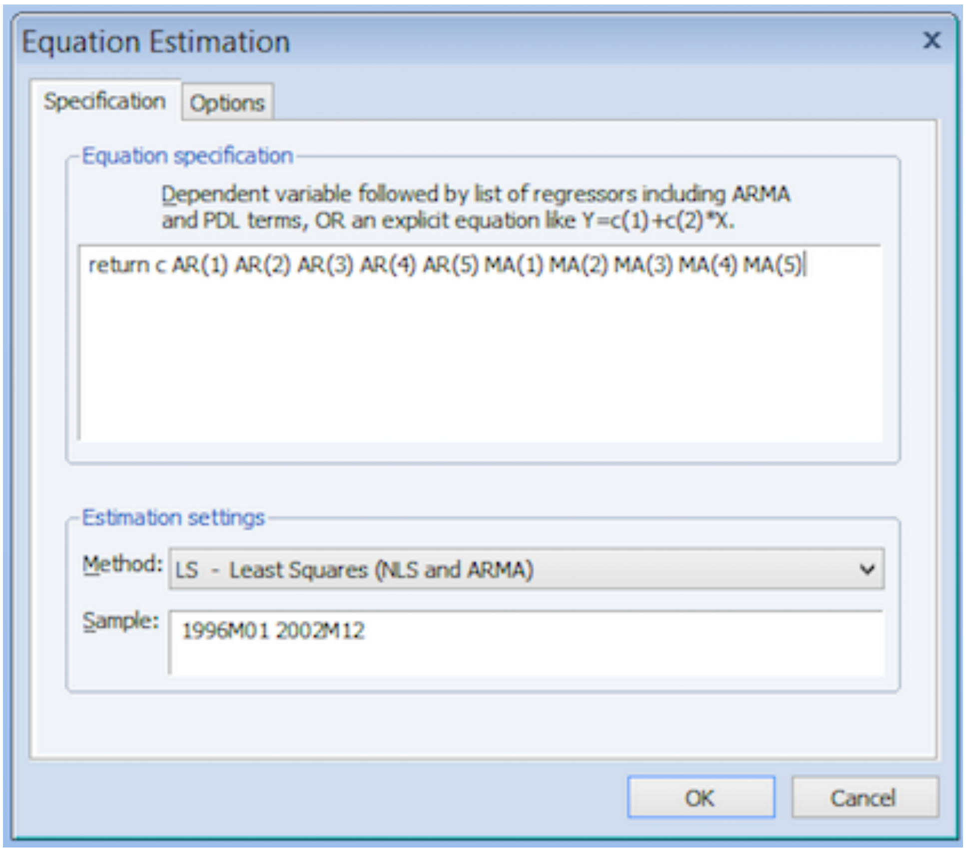
2. For example, if we want to construct a AR(1) model, we should type 'return c AR(1)' in the 'Specification' tab of 'Equation Estimation' window. 'Method' should be 'LS – Least Squares (NLS and ARMA)'. We can also type 'dlog(price) c AR(1)', and we will get the same results.

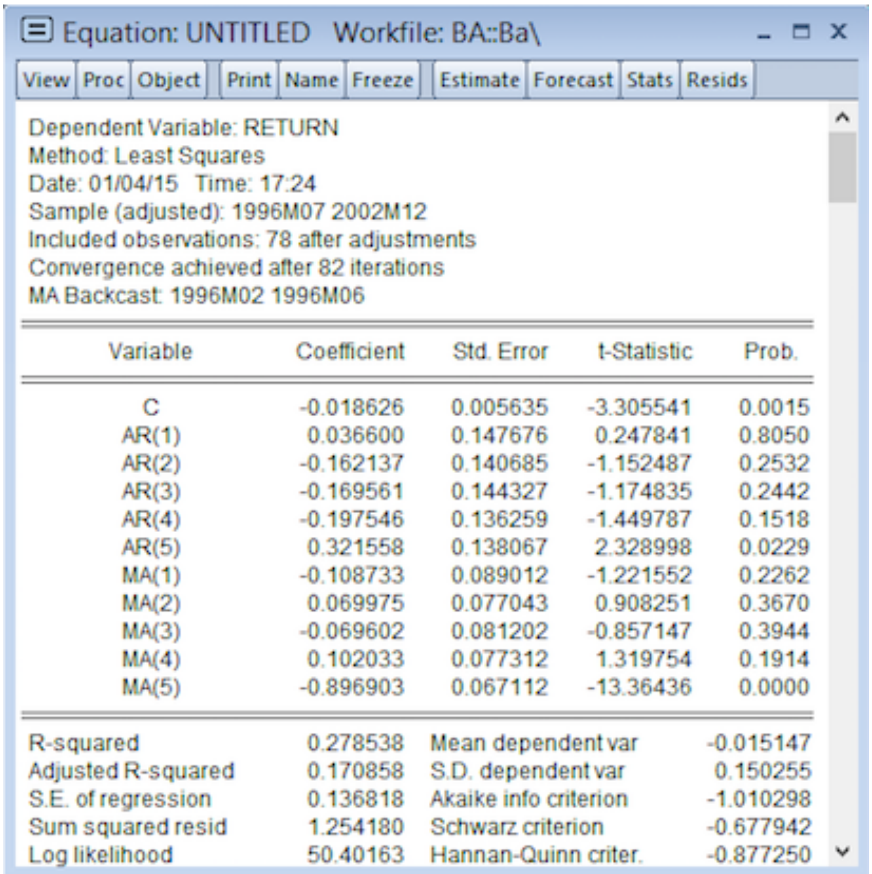


3. Consequently, we get following model, and the result window show the information criteria.



4. If we want to construct a ARMA(5,5) model, we should type ‘return c AR(1) AR(2) AR(3) AR(4) AR(5) MA(1) MA(2) MA(3) MA(4) MA(5)’ in the ‘Specification’ tab of ‘Equation Estimation’ window. Subsequently, we can get following results.




Equation: UNTITLED Workfile: BA::Ba\

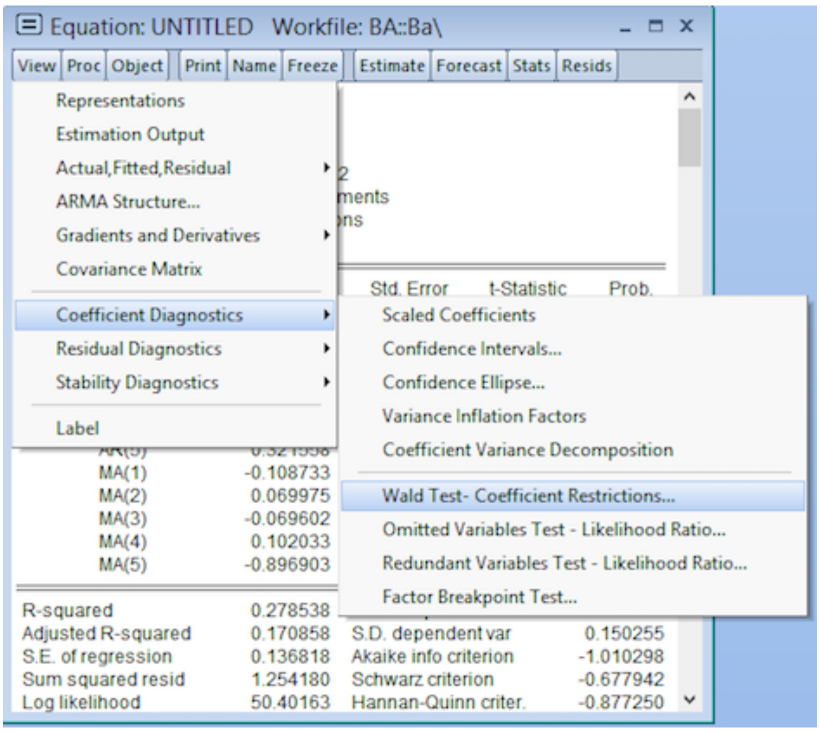
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RETURN
Method: Least Squares
Date: 01/04/15 Time: 17:24
Sample (adjusted): 1996M07 2002M12
Included observations: 78 after adjustments
Convergence achieved after 82 iterations
MA Backcast: 1996M02 1996M06

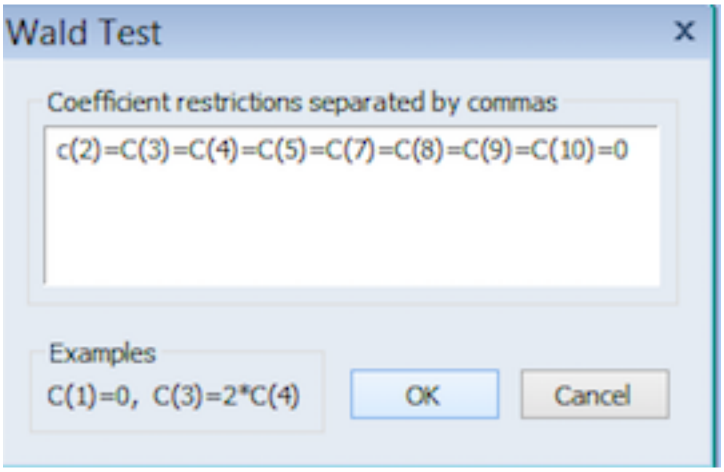
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.018626	0.005635	-3.305541	0.0015
AR(1)	0.036600	0.147676	0.247841	0.8050
AR(2)	-0.162137	0.140685	-1.152487	0.2532
AR(3)	-0.169561	0.144327	-1.174835	0.2442
AR(4)	-0.197546	0.136259	-1.449787	0.1518
AR(5)	0.321558	0.138067	2.328998	0.0229
MA(1)	-0.108733	0.089012	-1.221552	0.2262
MA(2)	0.069975	0.077043	0.908251	0.3670
MA(3)	-0.069602	0.081202	-0.857147	0.3944
MA(4)	0.102033	0.077312	1.319754	0.1914
MA(5)	-0.896903	0.067112	-13.36436	0.0000

R-squared	0.278538	Mean dependent var	-0.015147
Adjusted R-squared	0.170858	S.D. dependent var	0.150255
S.E. of regression	0.136818	Akaike info criterion	-1.010298
Sum squared resid	1.254180	Schwarz criterion	-0.677942
Log likelihood	50.40163	Hannan-Quinn criter.	-0.877250

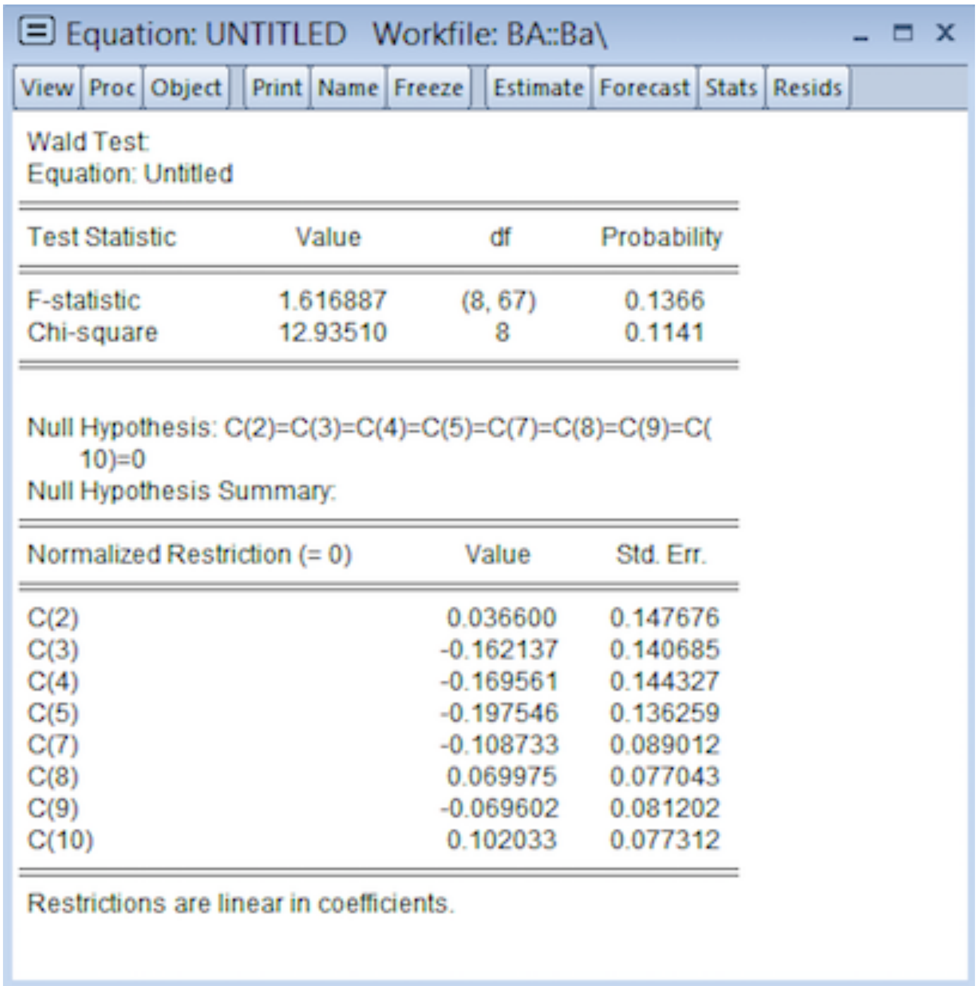
5. In above ARMA(5,5) model, only constant (C), AR(5) and MA(5) components are significant, other variables are insignificant. Wald test can be used to examine whether we can exclude theses variables in the model. In the 'Equation' window, View > Coefficient Diagnostics > Wald Test-Coefficient Restrictions.



In the ‘Wald Test’ window, input the variables that we want to test. For example, if we want to test whether AR(1), AR(2), AR(3), AR(4), MA(1), MA(2), MA(3), and MA(4) components are jointly equal to zero, we should input ‘c(2)=c(3)=c(4)=c(5)=c(7)=c(8)=c(9)=c(10)=0’.



Following result of Wald Test are insignificant. In other words, it cannot reject the null hypothesis that above coefficients are jointly equal to zero. Therefore, we can exclude these variables in the regressions.



Equation: UNTITLED Workfile: BA::Ba\

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

Wald Test:
Equation: Untitled

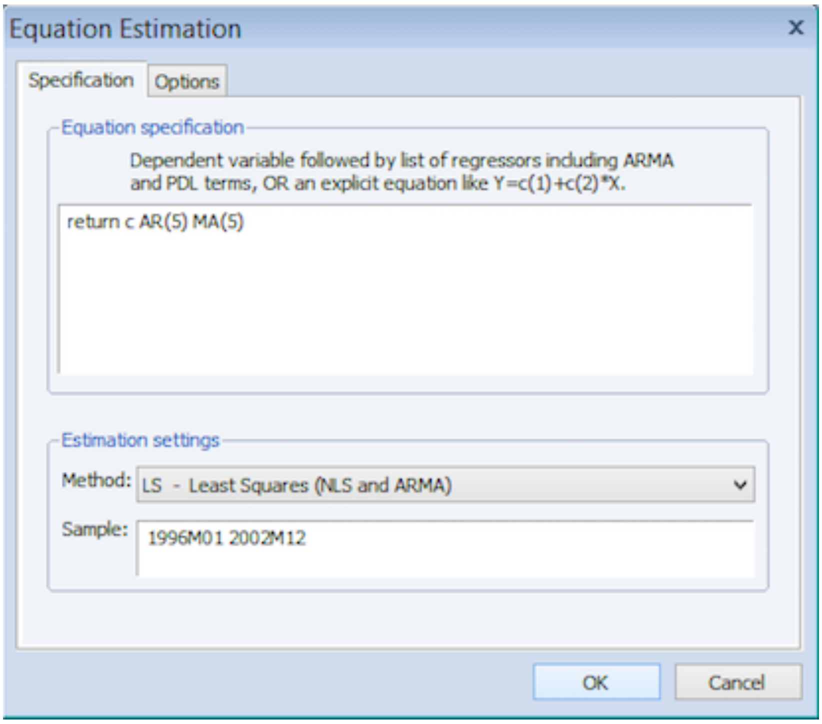
Test Statistic	Value	df	Probability
F-statistic	1.616887	(8, 67)	0.1366
Chi-square	12.93510	8	0.1141

Null Hypothesis: C(2)=C(3)=C(4)=C(5)=C(7)=C(8)=C(9)=C(10)=0
Null Hypothesis Summary:

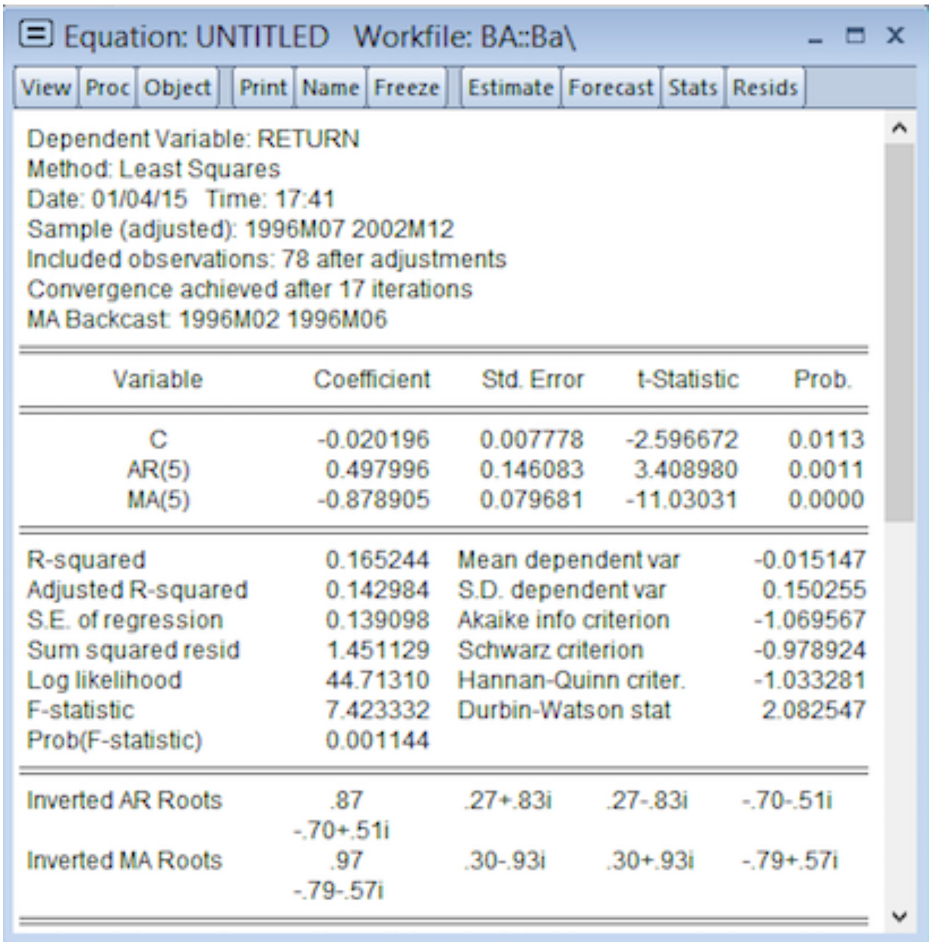
Normalized Restriction (= 0)	Value	Std. Err.
C(2)	0.036600	0.147676
C(3)	-0.162137	0.140685
C(4)	-0.169561	0.144327
C(5)	-0.197546	0.136259
C(7)	-0.108733	0.089012
C(8)	0.069975	0.077043
C(9)	-0.069602	0.081202
C(10)	0.102033	0.077312

Restrictions are linear in coefficients.

6. According to above results, we construct the ARMA model only include constant, AR(5) and MA(5) components. Type 'return c AR(5) MA(5)' in the 'Equation Specification' window.



Then, we get following model.



Dependent Variable: RETURN
Method: Least Squares
Date: 01/04/15 Time: 17:41
Sample (adjusted): 1996M07 2002M12
Included observations: 78 after adjustments
Convergence achieved after 17 iterations
MA Backcast: 1996M02 1996M06

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.020196	0.007778	-2.596672	0.0113
AR(5)	0.497996	0.146083	3.408980	0.0011
MA(5)	-0.878905	0.079681	-11.03031	0.0000

R-squared	0.165244	Mean dependent var	-0.015147
Adjusted R-squared	0.142984	S.D. dependent var	0.150255
S.E. of regression	0.139098	Akaike info criterion	-1.069567
Sum squared resid	1.451129	Schwarz criterion	-0.978924
Log likelihood	44.71310	Hannan-Quinn criter.	-1.033281
F-statistic	7.423332	Durbin-Watson stat	2.082547
Prob(F-statistic)	0.001144		

Inverted AR Roots	.87	.27+.83i	.27-.83i	-.70-.51i
	-.70+.51i			
Inverted MA Roots	.97	.30-.93i	.30+.93i	-.79+.57i
	-.79-.57i			

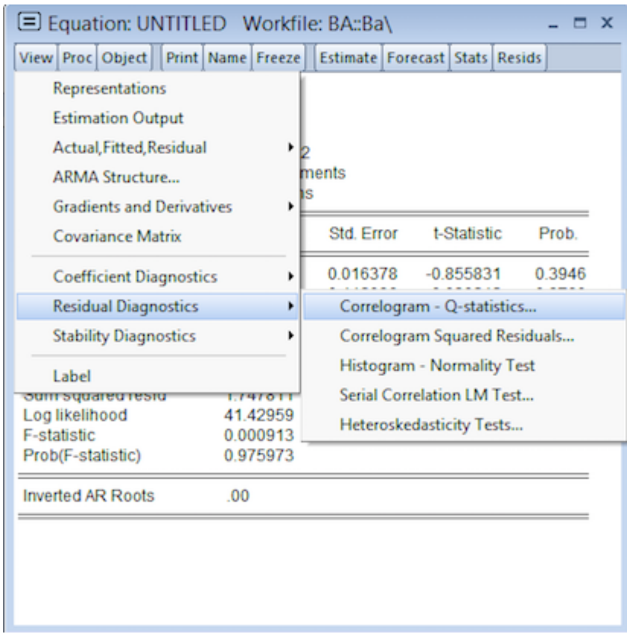
Diagnostic checking

We can examine whether the model constructed is adequate by two methods - overfitting and residual diagnostics.

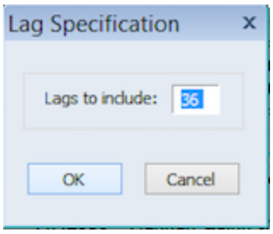
Overfitting refers to fitting a larger model than the model identified in first stage. The model is adequate, if any extra components added to the ARMA model are insignificant.

Residual diagnostics refers to checking whether the residuals are free from autocorrelation. The model is adequate, if autocorrelations of residuals are zero. The Ljung–Box tests (Q-statistics) can be used. In Eviews, Q-statistics can be conducted by following steps:

1. In the Equation window, View > Residual Diagnostics > Correlogram – Q-statistics



2. Then, input the lags to include. For ARMA(p,q) model, the number of lags to include should be greater than p+q+1



Note:

The residuals of an adequate model should be approximately white noise for which the autocorrelations are zero.

If the residuals are close to a white noise all ACF and PACF should be approximately within two standard error bounds $\pm 2/\sqrt{T}$

To check the overall acceptability of the residual autocorrelations, the Ljung-Box (1978) test statistic maybe used:

$$Q_k = T(T+2)\sum_{k=1}^{k}\frac{1}{T-k}\hat{\rho}_k^2$$

Here, the $\hat{\rho}_k$ is the estimated autocorrelation coefficients of the residuals and k is the number of lags examined.

For an ARMA(p,q) process the statistic Q_k is approximately Chi-squared distributed with k-p-q-1 degrees of freedom.

Note that the model only makes sense if k>p+q+1

The null hypothesis and alternative hypothesis of Ljung–Box test:

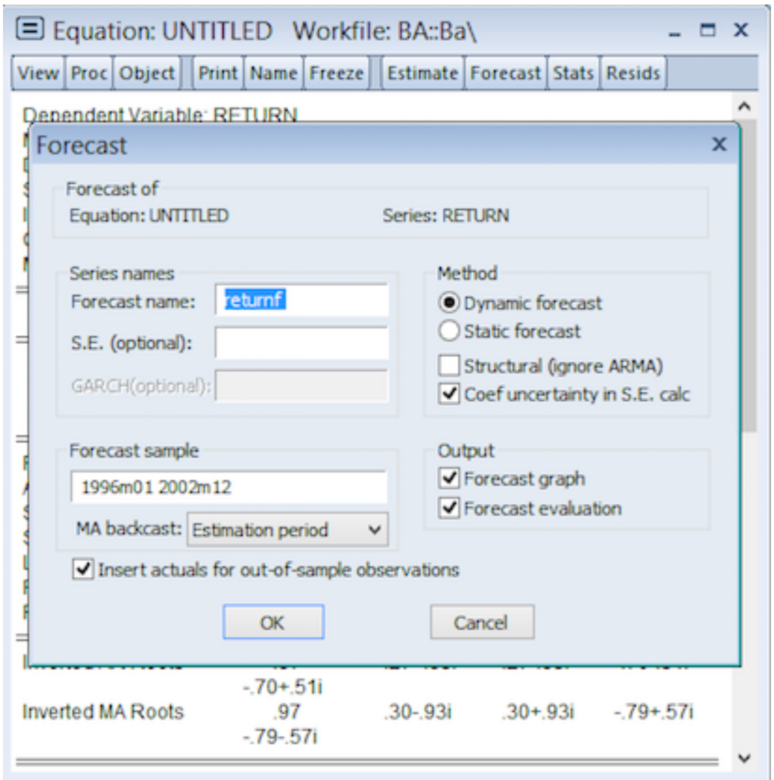
H_0: The data are independently distributed.

H_a: The data are not independently distributed.

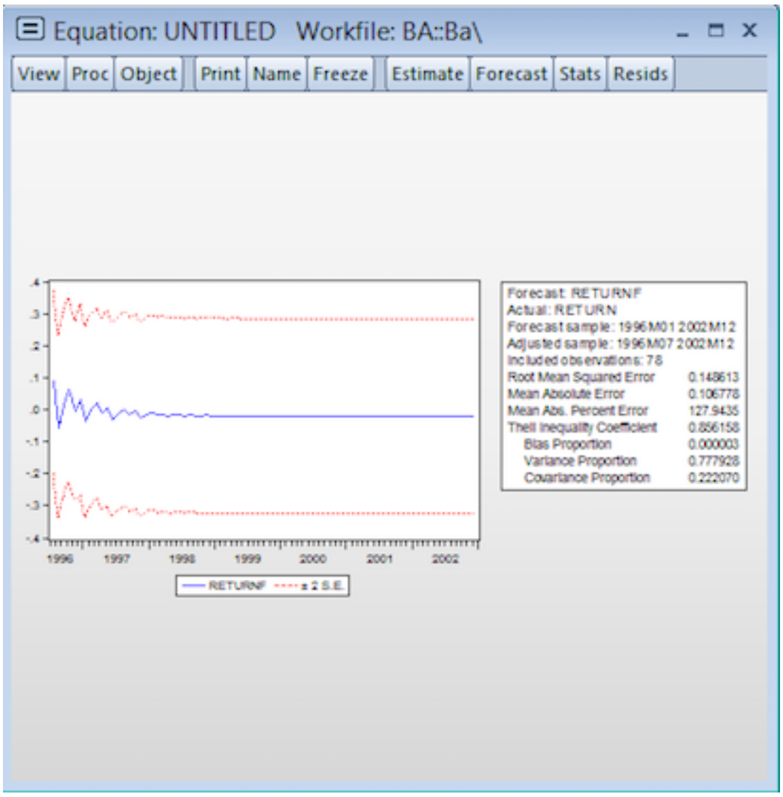
Forecasting

After we determine the proper model, we can make forecasting based on it.

In Equation window, click ‘Forecast’



Finally, we get the results on forecasting.



We will discuss forecasting further in computer lab 3.

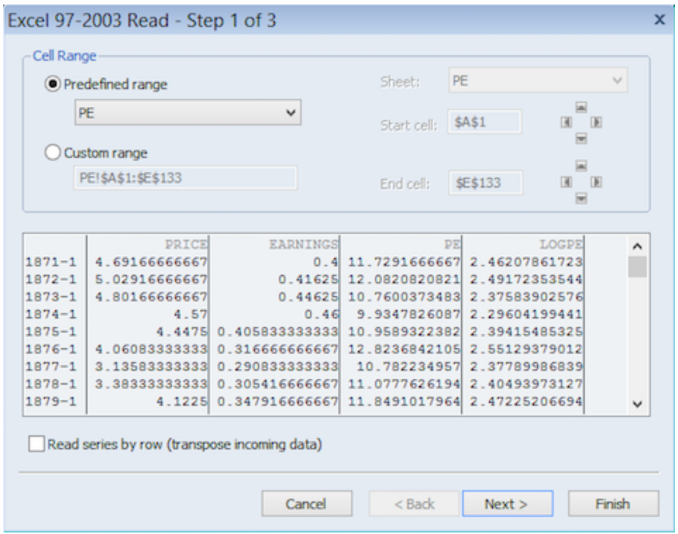
Unit Root Test

In the second part of this session, we will conduct unit root tests by using annual data on the price/earnings ratio of the S&P Composite Index over the period 1871–2002 (see Verbeek 2004, p.274).

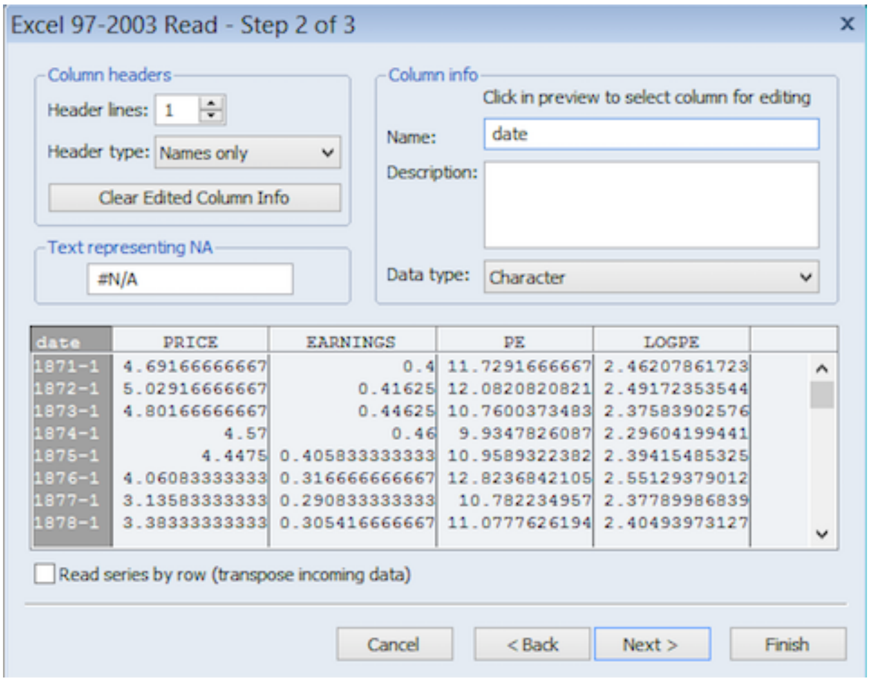
Data Preparation

1. Please download the file 'PE.xls' from DUO, and use EViews to open it.

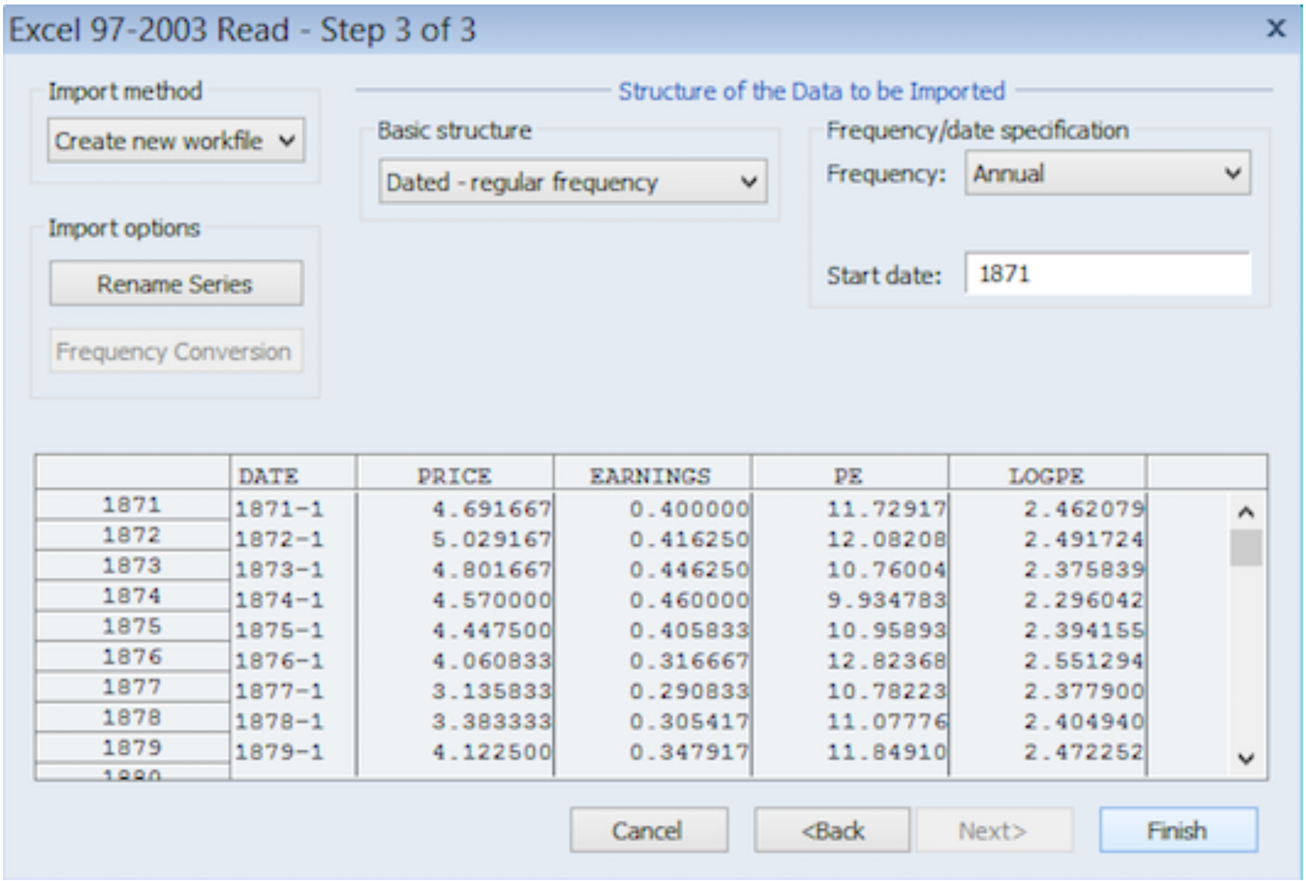
   Click 'Next'



2. Give the first column of data a variable name 'date'. However, the 'Data type' is recognized as the 'Character'. We should adjust the data type later. Other variables (PRICE, EARNINGS, PE, and LOGPE) are correctly identified by the EViews.

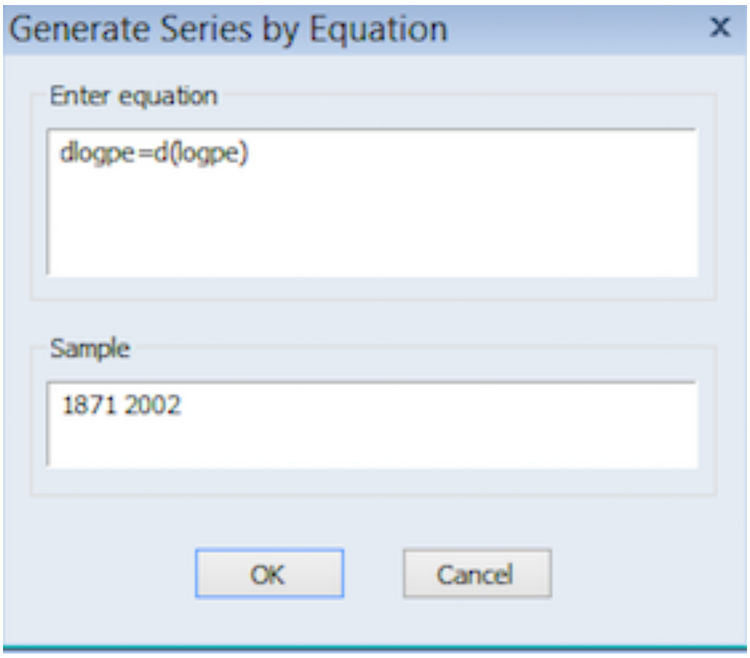


3. Adjust the ‘Structure of the Data to be Imported’. For ‘Basic structure’, we should choose ‘Dated – regular frequency’. For ‘Frequency’, we should choose ‘Annual’. For ‘Start date’, we should input ‘1871’.



4. Generate a new series ‘dlogpe’. Enter following equation:

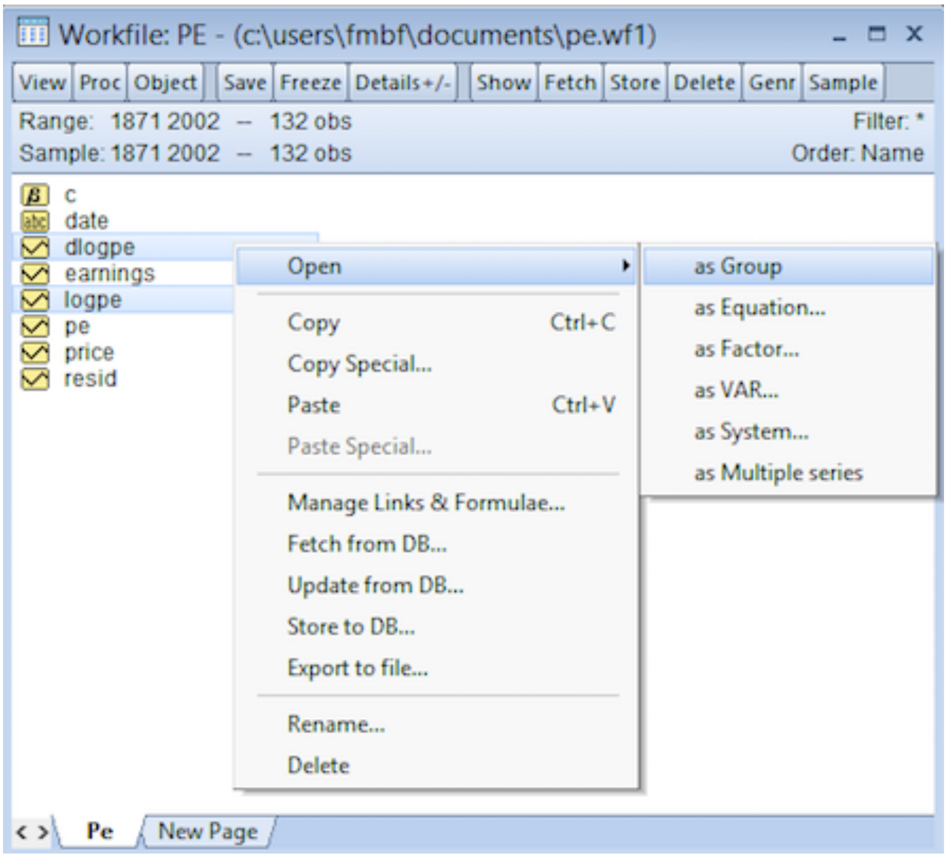
$$dlogpe = d(logpe)$$



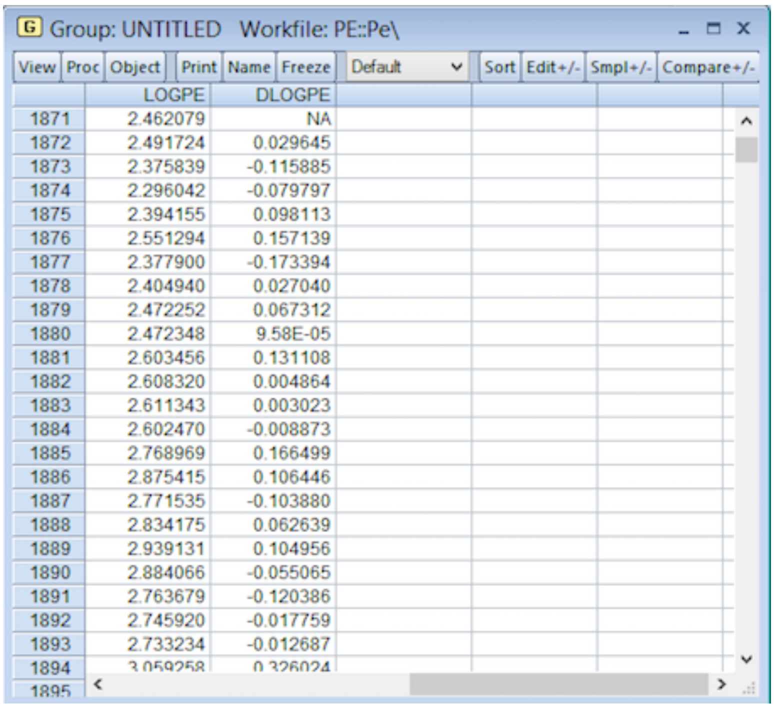
Line Graph

We draw the line graph for the two series – logpe (level) and dlogpe (first difference).

1. Press ‘control’ to select the two series, and right click. Select ‘Open > as Group’.

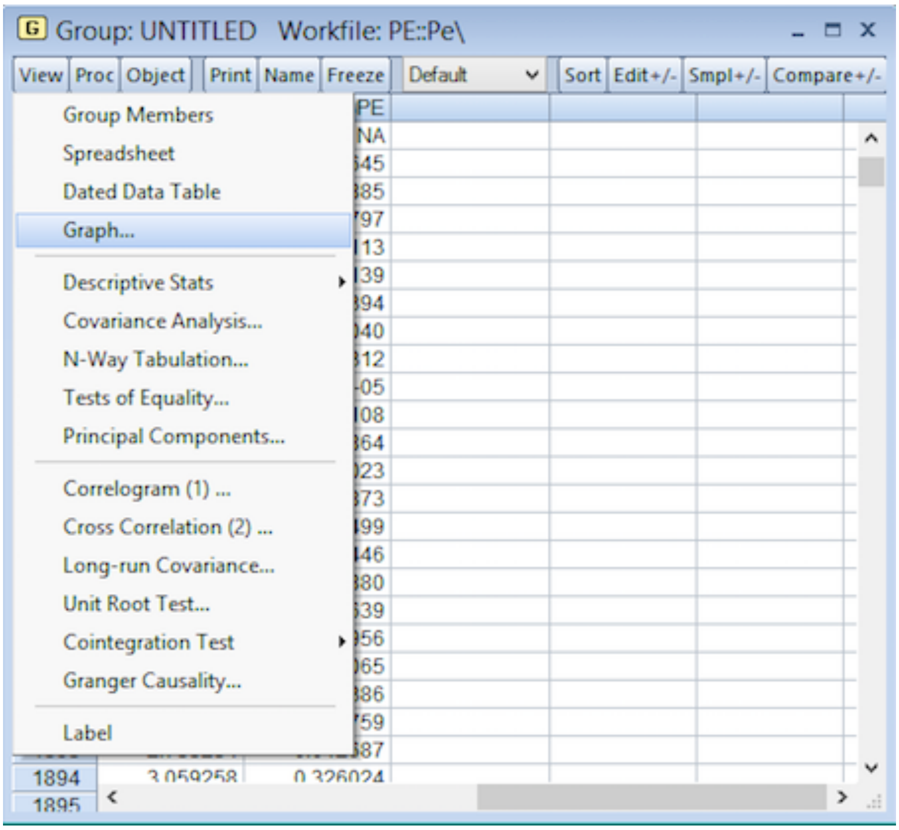


We will get following ‘Group’ window.

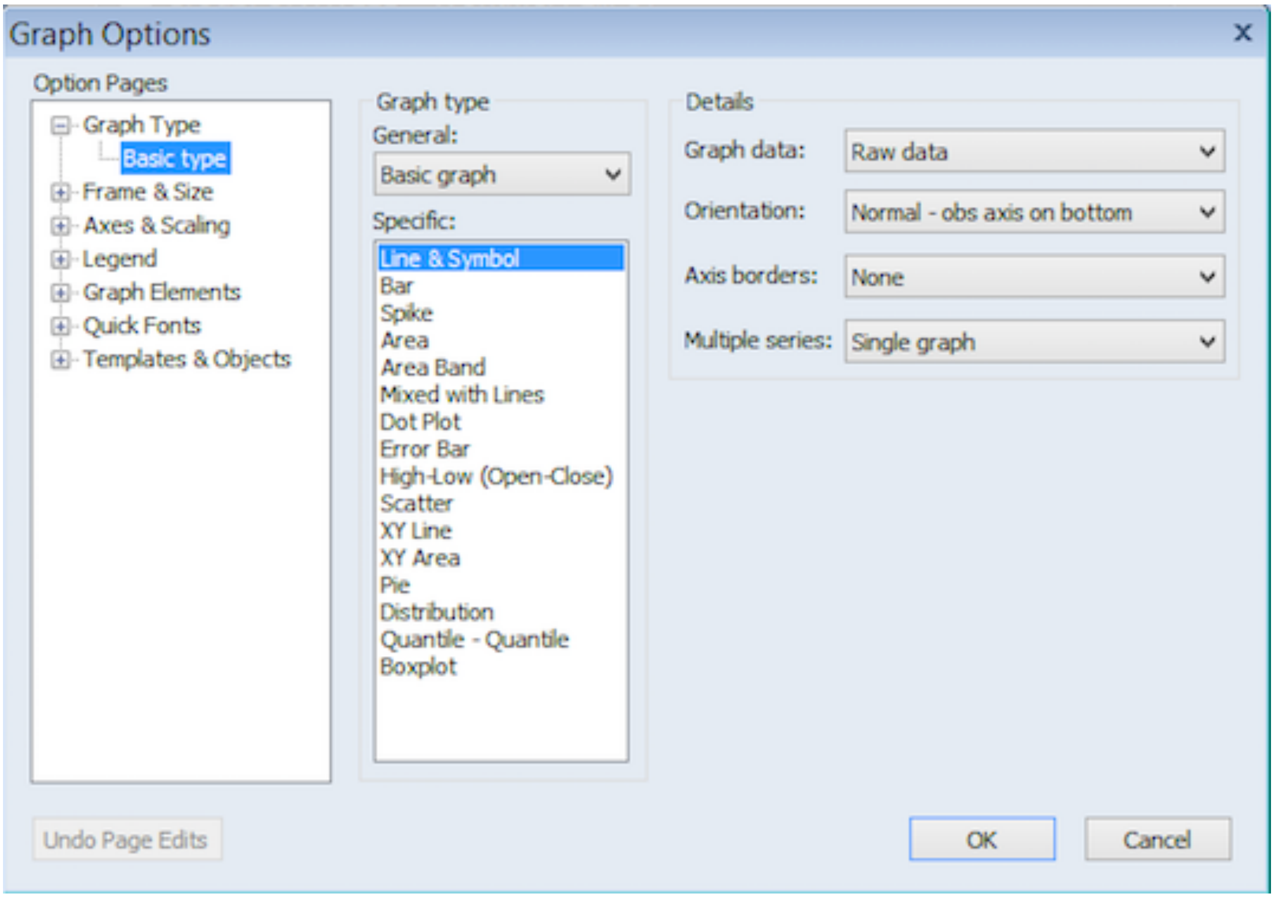


	LOGPE	DLOGPE
1871	2.462079	NA
1872	2.491724	0.029645
1873	2.375839	-0.115885
1874	2.296042	-0.079797
1875	2.394155	0.098113
1876	2.551294	0.157139
1877	2.377900	-0.173394
1878	2.404940	0.027040
1879	2.472252	0.067312
1880	2.472348	9.58E-05
1881	2.603456	0.131108
1882	2.608320	0.004864
1883	2.611343	0.003023
1884	2.602470	-0.008873
1885	2.768969	0.166499
1886	2.875415	0.106446
1887	2.771535	-0.103880
1888	2.834175	0.062639
1889	2.939131	0.104956
1890	2.884066	-0.055065
1891	2.763679	-0.120386
1892	2.745920	-0.017759
1893	2.733234	-0.012687
1894	3.059258	0.326024
1895		

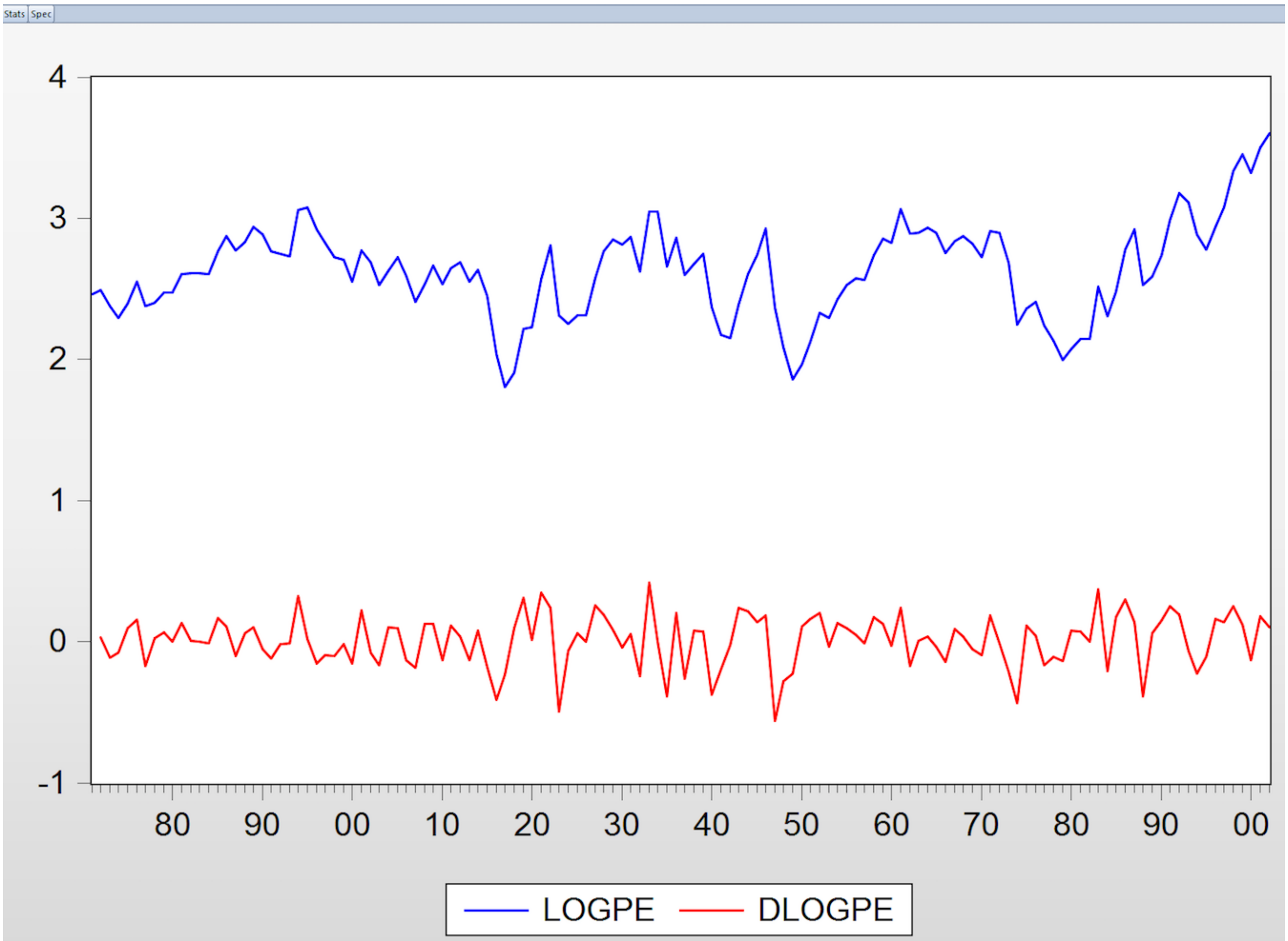
2. In Group window, View > Graph



3. Select 'Basic type' and 'Line&Symbol'.



4. Finally, we get following graph



Based on above graph, which series do you think is a stationary process – logpe or dlogpe?

In fact, to examine whether a process is stationary, we should conduct unit root test, such as Augmented Dickey–Fuller test (ADF), Phillips–Perron test, and Kwiatkowski–Phillips–Schmidt–Shin tests (KPSS test is a stationarity test).

Augmented Dickey–Fuller test

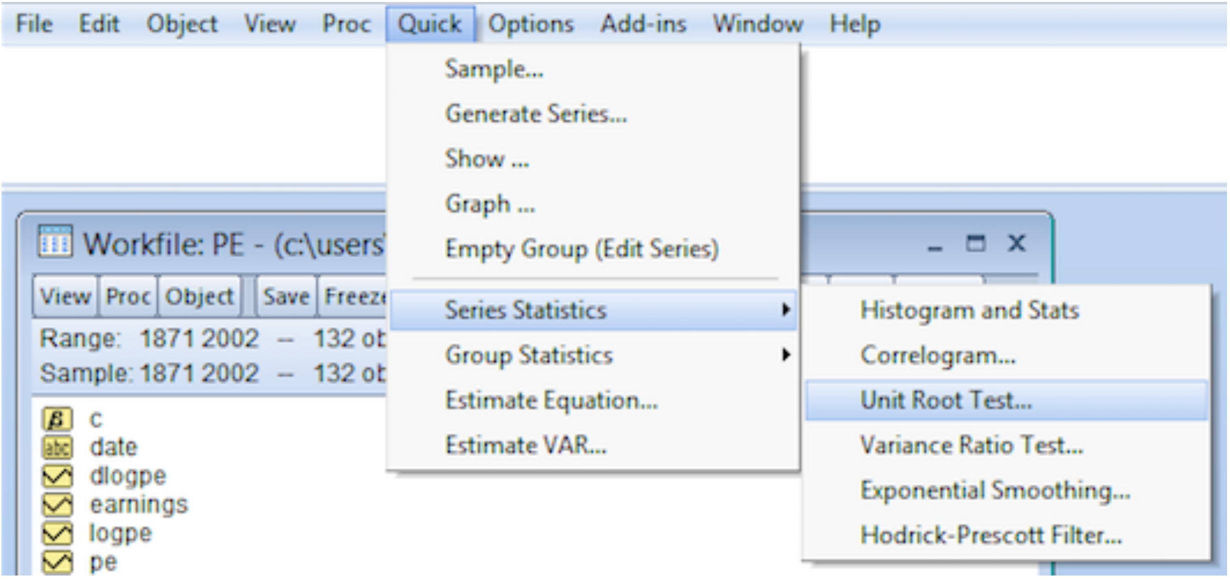
$$\Delta y_t = \delta + \psi y_{t-1} + \gamma t + \sum_{i=1}^{k} \beta_i \Delta y_{t-i} + \varepsilon_t \qquad \text{(Equation 1)}$$

H_0: $\psi = 0$ i.e. unit root

H_a: $\psi < 0$ i.e. stationarity

The choice of a value of k is a specification issue. In general, information criteria like AIC or SBIC are used to determine a value for k.

1. Quick > Series Statistics > Unit Root Test

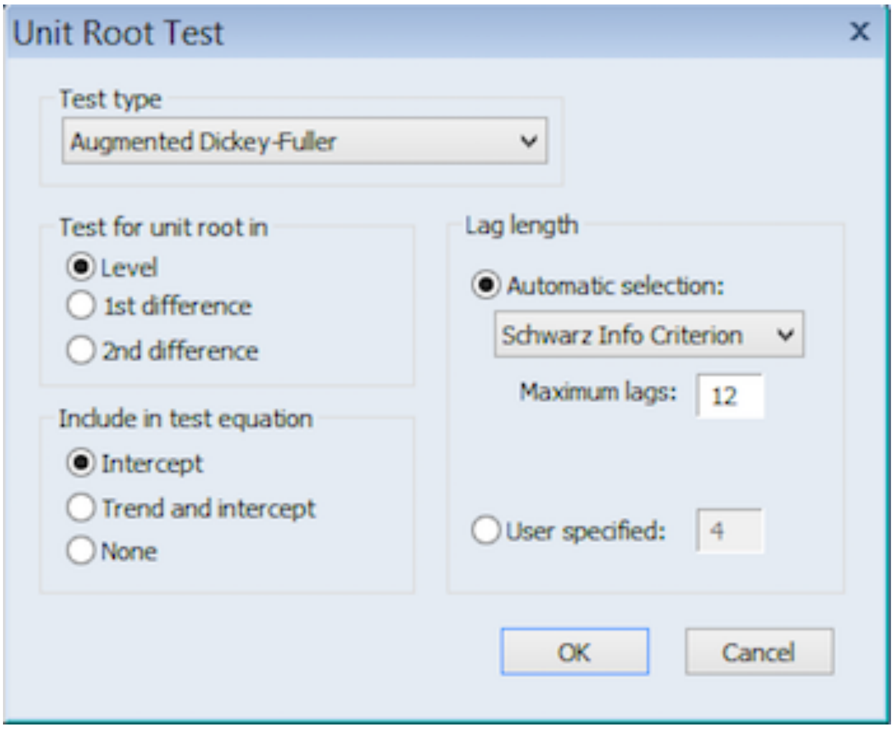


3. Test type: Augmented Dickey-Fuller

   Test for unit root in: level

   Include in test equation: Intercept

   Lag length: Automatic selection: Schwarz Info Criterion Maximum lags: 12



By choosing these options, we will conduct ADF test for 'logpe' with up to 12 lags of the dependent variable (dlogpe). We will include a constant in the test equation. The EViews will use Schwarz Info Criterion (also known as SBIC or BIC) to determine the optimal lag length (the value of k in Equation 1 shown above) in the ADF test.

4. We get following results. The p-value of ADF test is 0.1021 (insignificant). The result cannot reject the null hypothesis that 'logpe' has a unit root. In other words, the 'logpe' is non-stationary. The results window also shows the detailed information on test equation. The optimal specification based on SBIC does not include the lagged term of dependent variable.

   Note: In the ADF test of 'logpe', the dependent variable is d(logpe).

Series: LOGPE Workfile: PE::Pe\

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

Augmented Dickey-Fuller Unit Root Test on LOGPE

Null Hypothesis: LOGPE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.568963	0.1021
Test critical values:	1% level	-3.480818	
	5% level	-2.883579	
	10% level	-2.578601	

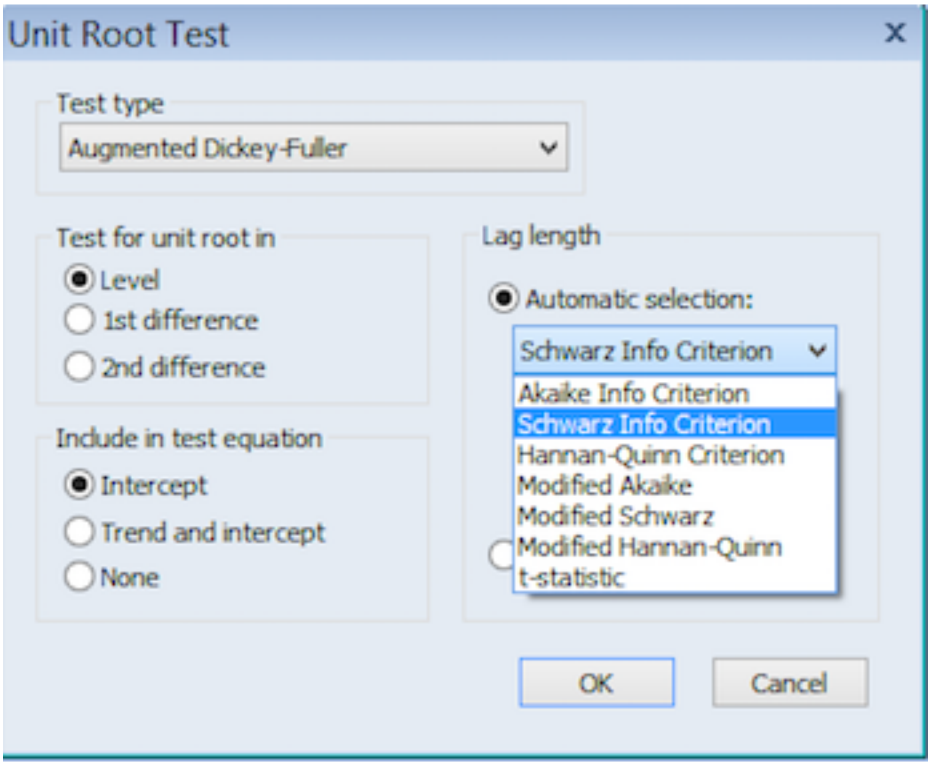
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGPE)
Method: Least Squares
Date: 01/05/15 Time: 15:20
Sample (adjusted): 1872 2002
Included observations: 131 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGPE(-1)	-0.124521	0.048471	-2.568963	0.0113
C	0.334860	0.127902	2.618106	0.0099

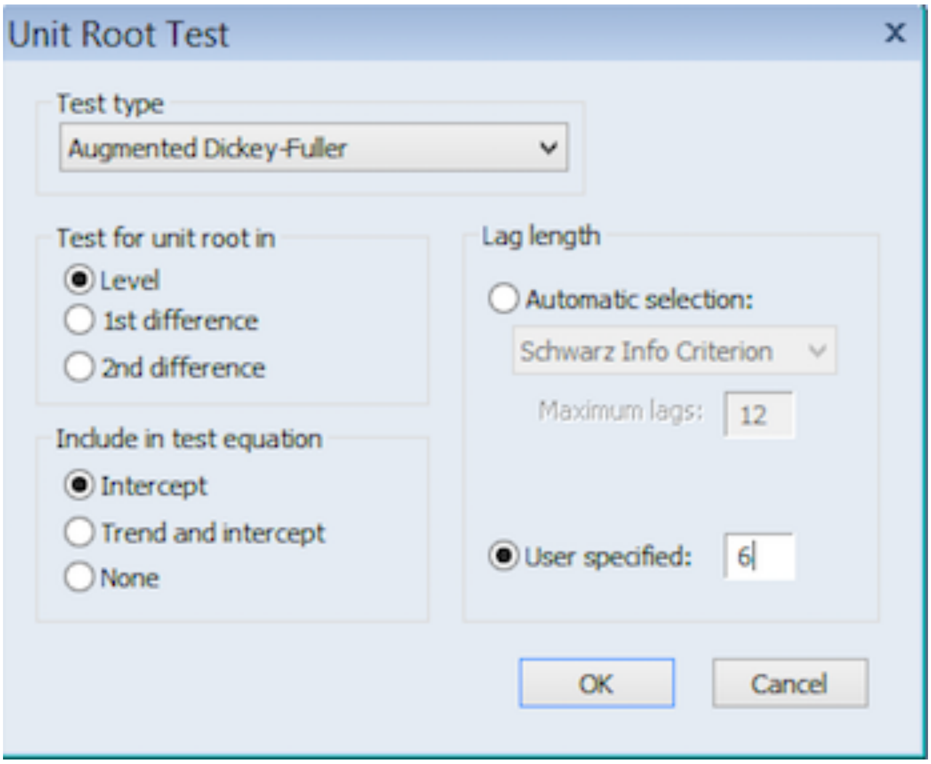
R-squared	0.048670	Mean dependent var	0.008716
Adjusted R-squared	0.041295	S.D. dependent var	0.181506
S.E. of regression	0.177719	Akaike info criterion	-0.602079
Sum squared resid	4.074333	Schwarz criterion	-0.558183
Log likelihood	41.43616	Hannan-Quinn criter.	-0.584242
F-statistic	6.599573	Durbin-Watson stat	1.735417
Prob(F-statistic)	0.011338		

5. We can also use other information criteria to determine the optimal specification. Please check whether it will lead to different results.



6. We can also specify a test model with given number of lags of dependent variables in the regression equation.

For example, we construct an ADF equation with 6 lagged terms.



We get following results. The p-value is 0.3668, which cannot reject the null hypothesis of unit root, confirming that ‘logpe’ is non-stationary. The results also show the detailed information on test equation.

Series: LOGPE Workfile: PE::Pe\

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

Augmented Dickey-Fuller Unit Root Test on LOGPE

Null Hypothesis: LOGPE has a unit root
Exogenous: Constant
Lag Length: 6 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.825316	0.3668
Test critical values:	1% level	-3.483312	
	5% level	-2.884665	
	10% level	-2.579180	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LOGPE)
Method: Least Squares
Date: 01/05/15 Time: 15:21
Sample (adjusted): 1878 2002
Included observations: 125 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGPE(-1)	-0.125162	0.068570	-1.825316	0.0705
D(LOGPE(-1))	0.151893	0.105324	1.442151	0.1519
D(LOGPE(-2))	-0.103754	0.103827	-0.999299	0.3197
D(LOGPE(-3))	0.044115	0.100050	0.440932	0.6601
D(LOGPE(-4))	-0.139051	0.098943	-1.405358	0.1626
D(LOGPE(-5))	0.000147	0.095827	0.001534	0.9988
D(LOGPE(-6))	0.070460	0.095687	0.736366	0.4630
C	0.338777	0.179268	1.889780	0.0613

R-squared	0.106674	Mean dependent var	0.009807
Adjusted R-squared	0.053227	S.D. dependent var	0.183945
S.E. of regression	0.178983	Akaike info criterion	-0.541194
Sum squared resid	3.748075	Schwarz criterion	-0.360182
Log likelihood	41.82464	Hannan-Quinn criter.	-0.467659
F-statistic	1.995894	Durbin-Watson stat	2.009387
Prob(F-statistic)	0.061333		

7. We repeat the test for ‘dlogpe’ (the first difference of ‘logpe’). We can directly conduct ADF test for ‘dlogpe’. Alternatively, we can conduct ADF test for ‘logpe’, and choose the option ‘Test for unit root in: 1st difference’. We will get the same results.

   Following table show the results of ADF test for ‘dlogpe’.

Series: DLOGPE Workfile: PE::Pe\

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

Augmented Dickey-Fuller Unit Root Test on DLOGPE

Null Hypothesis: DLOGPE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-10.58840	0.0000
Test critical values:	1% level	-3.481217	
	5% level	-2.883753	
	10% level	-2.578694	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLOGPE)
Method: Least Squares
Date: 01/05/15 Time: 15:30
Sample (adjusted): 1873 2002
Included observations: 130 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOGPE(-1)	-0.934805	0.088286	-10.58840	0.0000
C	0.008033	0.016024	0.501322	0.6170

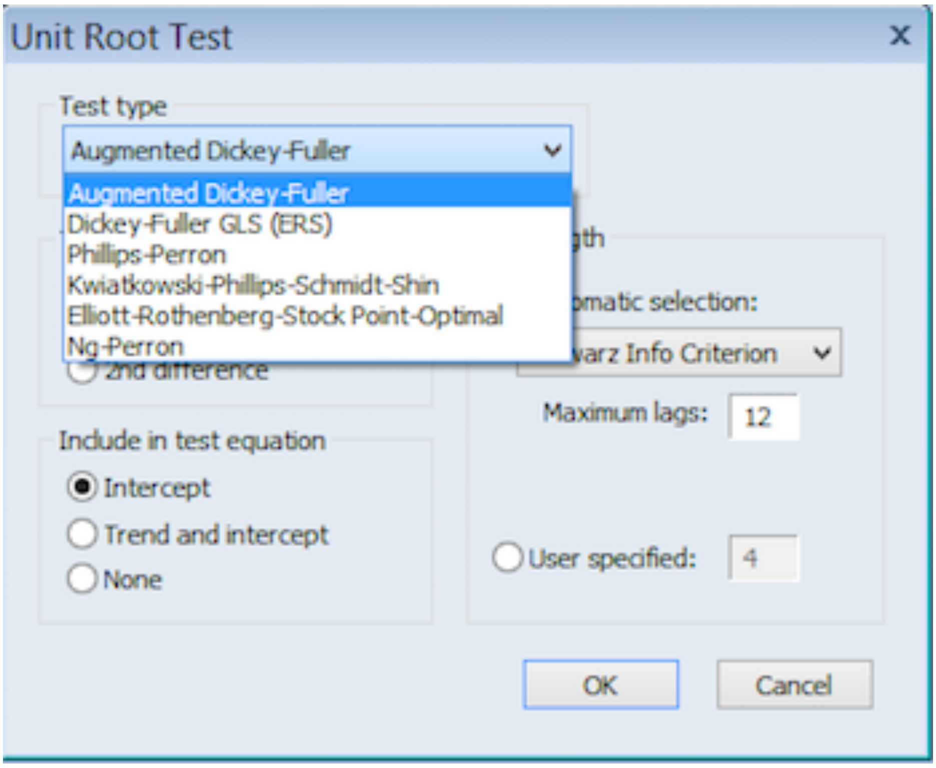
R-squared	0.466920	Mean dependent var	0.000553
Adjusted R-squared	0.462756	S.D. dependent var	0.249015
S.E. of regression	0.182521	Akaike info criterion	-0.548642
Sum squared resid	4.264166	Schwarz criterion	-0.504526
Log likelihood	37.66170	Hannan-Quinn criter.	-0.530716
F-statistic	112.1142	Durbin-Watson stat	1.971193
Prob(F-statistic)	0.000000		

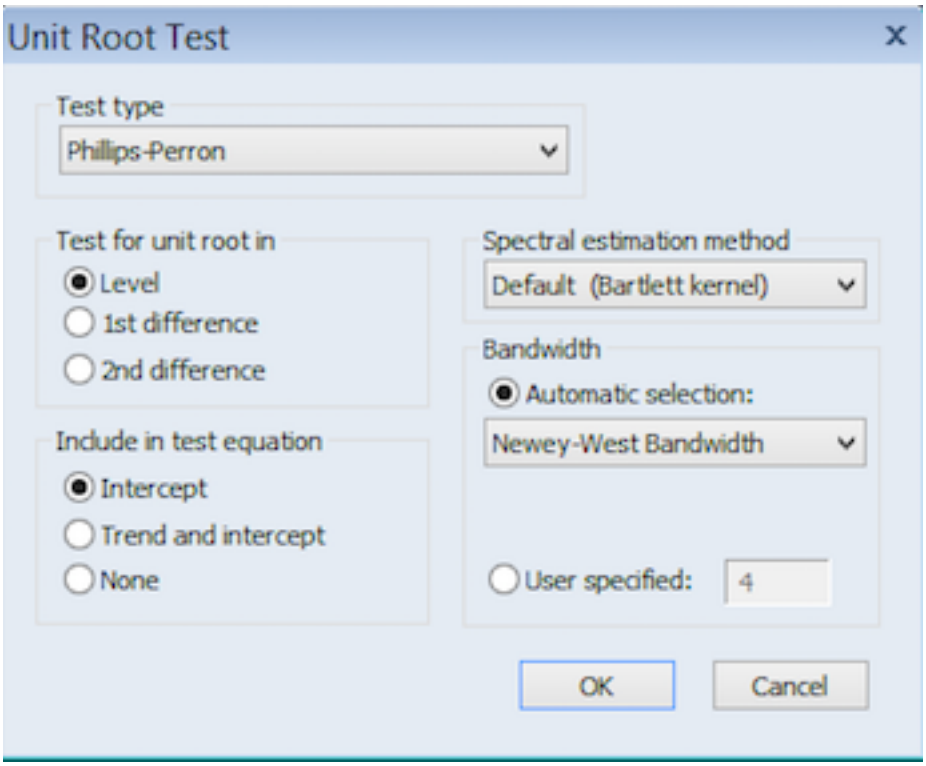
   The p-value is 0.000 (highly significant), which reject the null hypothesis that ‘dlogpe’ has a unit root. In other words, ‘dlogpe’ is stationary.

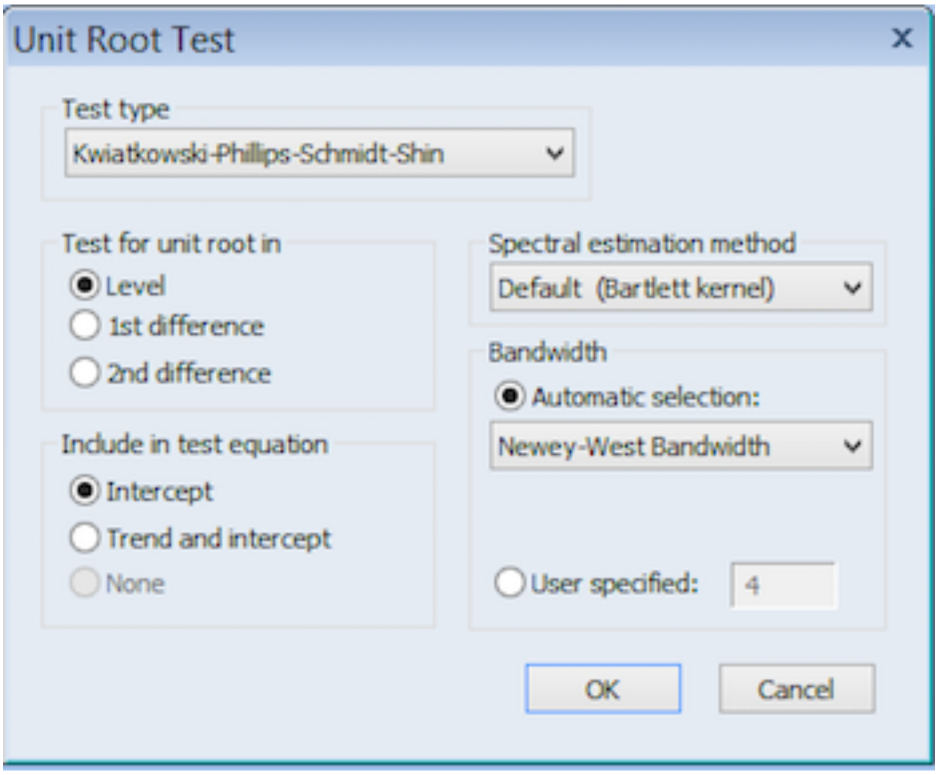
According to above results, level of the series (‘logpe’) is non-stationary, but the first difference of the series (‘dlogpe’) is stationary. The results suggest that the variable ‘logpe’ is a I(1) process.

Other Unit Root/Stationarity Tests

In the 'Unit Root Test' window, we can select different test type, such as Phillips–Perron and KPSS test show as follows.







Please pay attention that the KPSS is a stationarity test that is different from unit root test like ADF test and Phillips–Perron test. Specifically, the null hypothesis for the KPSS test is that the variable is stationary.

Note:

In different tests, we always face a question about how to determine the number of lag terms in the test equation. We can decide the number based on data frequency. For example, we can use 4 and 12 lags for quarterly data and monthly data, respectively. More importantly, we can use information criteria, such as SBIC and AIC, to choose the optimal specification. The optimal specification should minimize the value of an information criterion.

Further Exercise

1. Load the data 'FTSEDATA.xls' that is on duo. This contains monthly data for the FTSE 100 and ALL SHARE from 1985.
2. Create logarithms of the two indices, naming them LFTSE100 and LFTALLSH.
3. Plot the series then test for stationarity adding an appropriate number of lags.
4. Create the first difference of LFTSE100 and LFTALLSH and test for stationarity after plotting the differenced series.

5. Come to conclusions about the presence of a unit root in the two series.
6. Construct ARMA models for the two series following the Box–Jenkins approach.

References

M. Verbeek. *A Guide to Modern Econometrics*. John Wiley & Sons, Inc., 2004.