Topic: Introduction to Multivariate Regression

- This class is an introduction to the **menu-driven features** of EViews 6. The data is available on the website as `cigarette.txt`.
- The problem set covers (i) file import and workfile save, (ii) statistical analysis, (iii) creation of groups, (iv) ordinary least squares, (v) (robust) inference, and (vi) diagnostic testing.
- We use data on (i) the log of cigarette consumption (in packs) per person of smoking age ($>$ 16 years) for 46 U.S. states in 1992: LNC ($\ln(C)$), (ii) the log real price of cigarettes in each state, normalized at 1983$ per pack: LNP ($\ln(P)$), and (iii) the log of real disposable income per capita in each state, in 1983$1000: LNY ($\ln(Y)$).
- Perform all of the steps described in Figures **1–85** (answering any corresponding questions, and carefully considering the methods and output, noting any new or confusing tools), and end by responding in full to the following questions:

1. Regress log consumption on log prices **(eq01)**:

$$\widehat{\ln(C_i)} \approx \boxed{[\qquad\quad]} + \boxed{[\qquad\quad]}\ln(P_i),$$

with estimated residuals $\widehat{u}_i^{(1)} \approx \ln(C_i) + \boxed{[\qquad\quad]} + \boxed{[\qquad\quad]}\ln(P_i)$.

What is the estimated price elasticity of consumption? Answer:

$$\frac{\partial \widehat{\ln(C_i)}}{\partial \ln(P_i)} \approx \boxed{[\qquad\quad]}$$

and so a $\boxed{[\qquad]}$ increase/decrease in price results in a $\boxed{[\qquad]}$ increase/decrease in consumption.

2. Regress log consumption on log prices and log income **(eq02)**:

$$\widehat{\ln(C_i)} \approx \boxed{[\qquad]} + \boxed{[\qquad]}\ln(P_i) + \boxed{[\qquad]}\ln(Y_i).$$

3. Regress log income on log prices **(eq03)**:

$$\widehat{\ln(Y_i)} \approx \boxed{[\qquad]} + \boxed{[\qquad]}\ln(P_i),$$

with estimated residuals $\widehat{u}_i^{(3)} \approx \ln(Y_i) + \boxed{[\qquad]} + \boxed{[\qquad]}\ln(P_i)$.

4. Regress log consumption on $\widehat{u}_i^{(3)}$ **(eq04)**:

$$\widehat{\ln(C_i)} \approx \boxed{[\qquad]} + \boxed{[\qquad]}\widehat{u}_i^{(3)}$$

5. Regress $\widehat{u}_i^{(1)}$ on $\widehat{u}_i^{(3)}$ **(eq05)**:

$$\widehat{\widehat{u}_i^{(1)}} \approx \boxed{[\qquad]} + \boxed{[\qquad]}\widehat{u}_i^{(3)}.$$

6. Regress log consumption on log income, log income squared, and log prices **(eq06)**:

$$\widehat{\ln(C_i)} \approx \boxed{[\qquad]} + \boxed{[\qquad]}\ln(Y_i) + \boxed{[\qquad]}(\ln(Y_i))^2 + \boxed{[\qquad]}\ln(P_i),$$

so that a $\boxed{[\qquad]}$ increase/decrease in price results in a $\boxed{[\qquad]}$ increase/decrease in consumption. The income elasticity of consumption is now:

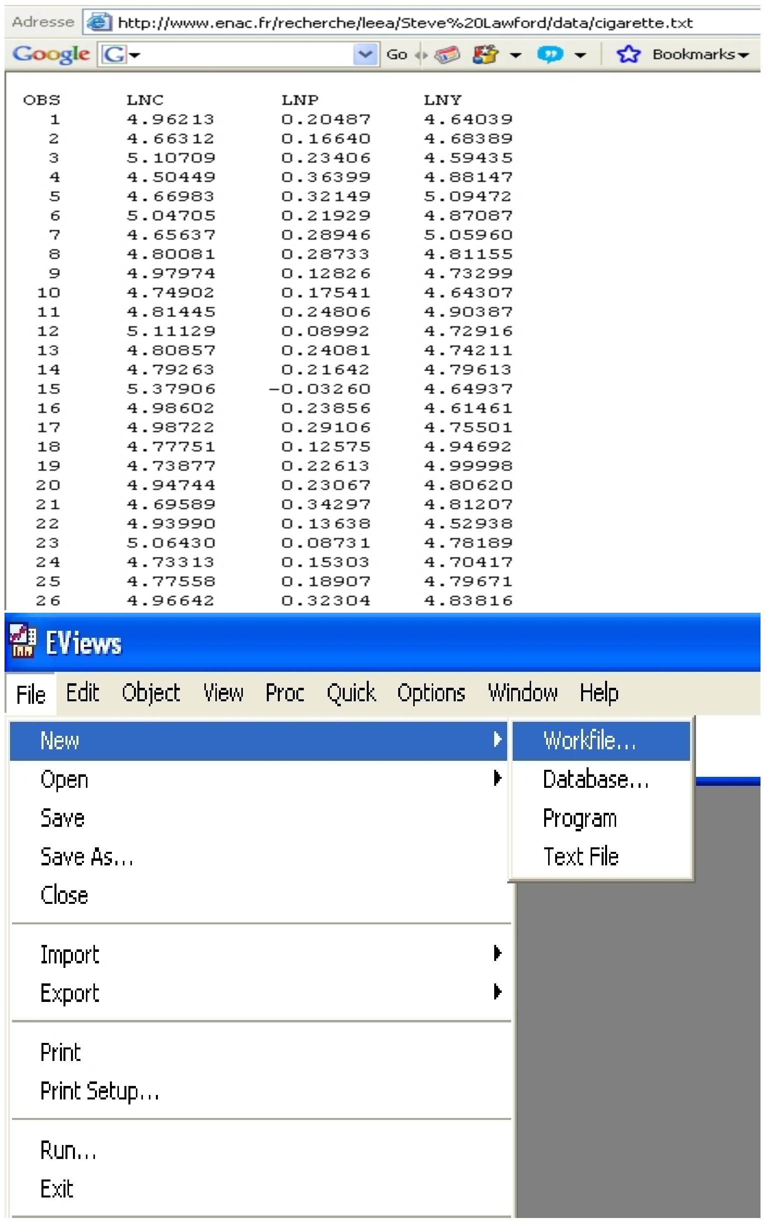
$$\frac{\partial \ln(C_i)}{\partial \ln(Y_i)} \approx \boxed{[\qquad]} + \boxed{[\qquad]} \ln(Y_i),$$

which is greater than (equal to) less than 1 as $\ln(Y_i)$ is *less than (equal to) greater than* [ ] (as Y_i is *less than (equal to) greater than* $[ ]): use the full available accuracy on the estimated coefficients when performing this computation. Interpret your findings carefully.

7. Compare equations eq02, eq04 and eq05. In eq02, [ ] quantifies the impact of log income on log consumption. In eq04, $\widehat{u}_i^{(3)}$ is the part of **log income not explained by log price**, and so [ ] quantifies the impact on log consumption of that part of log income not explained by log price. In eq05, $\widehat{u}_i^{(1)}$ is the part of **log consumption not explained by log price**, and so [ ] quantifies the impact on that part of log consumption not explained by log price, of that part of log income not explained by log price. Carefully explain the intuition behind these results.

- Special attention should be paid to observations 3 (Arkansas), 15 (Kentucky) and 40 (Utah): Arkansas and Kentucky have particularly high sales, Kentucky is a producer with rather low prices, and Utah has especially low sales because of its high Mormon population (which bans smoking). It is important to build a deep understanding of the structure and peculiarities of your data before you start modelling.

Importing Data into a Workfile



OBS	LNC	LNP	LNY
1	4.96213	0.20487	4.64039
2	4.66312	0.16640	4.68389
3	5.10709	0.23406	4.59435
4	4.50449	0.36399	4.88147
5	4.66983	0.32149	5.09472
6	5.04705	0.21929	4.87087
7	4.65637	0.28946	5.05960
8	4.80081	0.28733	4.81155
9	4.97974	0.12826	4.73299
10	4.74902	0.17541	4.64307
11	4.81445	0.24806	4.90387
12	5.11129	0.08992	4.72916
13	4.80857	0.24081	4.74211
14	4.79263	0.21642	4.79613
15	5.37906	-0.03260	4.64937
16	4.98602	0.23856	4.61461
17	4.98722	0.29106	4.75501
18	4.77751	0.12575	4.94692
19	4.73877	0.22613	4.99998
20	4.94744	0.23067	4.80620
21	4.69589	0.34297	4.81207
22	4.93990	0.13638	4.52938
23	5.06430	0.08731	4.78189
24	4.73313	0.15303	4.70417
25	4.77558	0.18907	4.79671
26	4.96642	0.32304	4.83816

Figure 1: Check that the dataset `cigarette.txt` can be opened by double-clicking on the website link (various data formats can be used, including `.txt` and `.xls`). Open a new workfile in EViews 6: it is possible to work on multiple workfiles simultaneously.

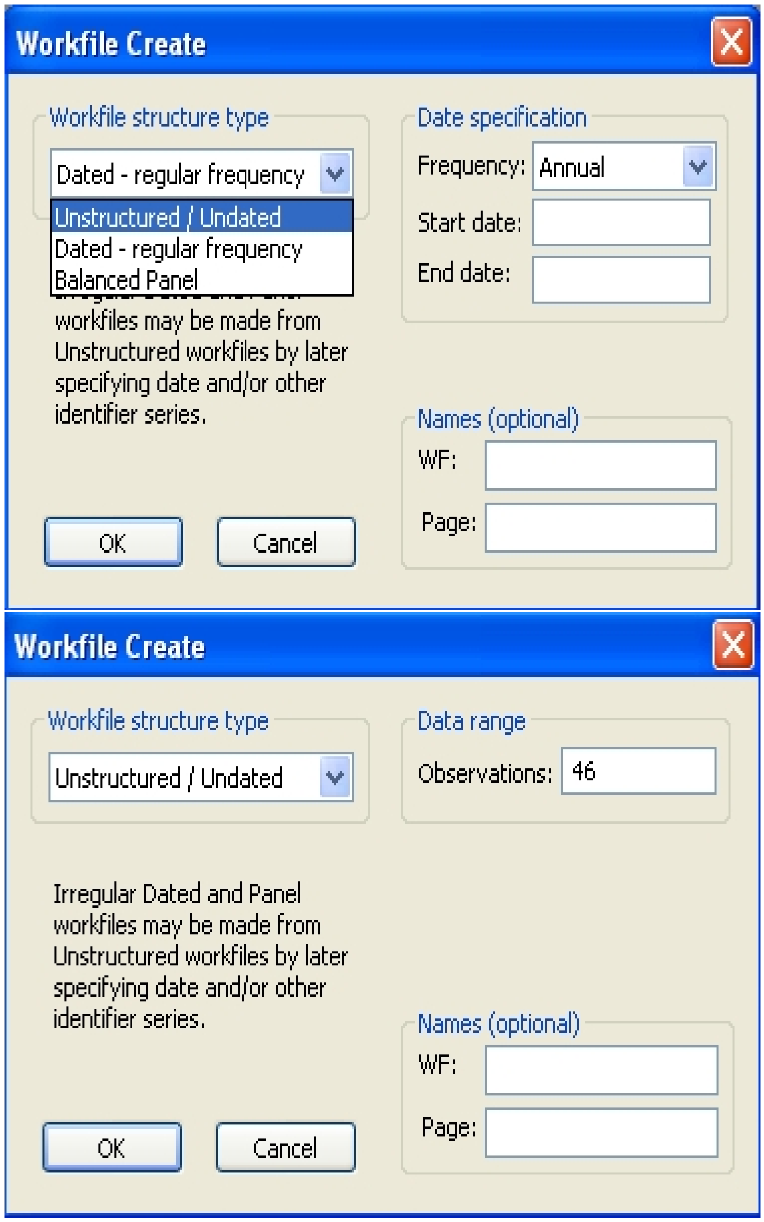


Figure 2: Choose the workfile structure: unstructured/undated (this can often be convenient even for time-series). Choose the number of observations: using this method, it is necessary to know the sample size before importing the data. Alternatively, data can be imported by File - Open - Foreign Data as Workfile . . ., and navigating to the datafile.

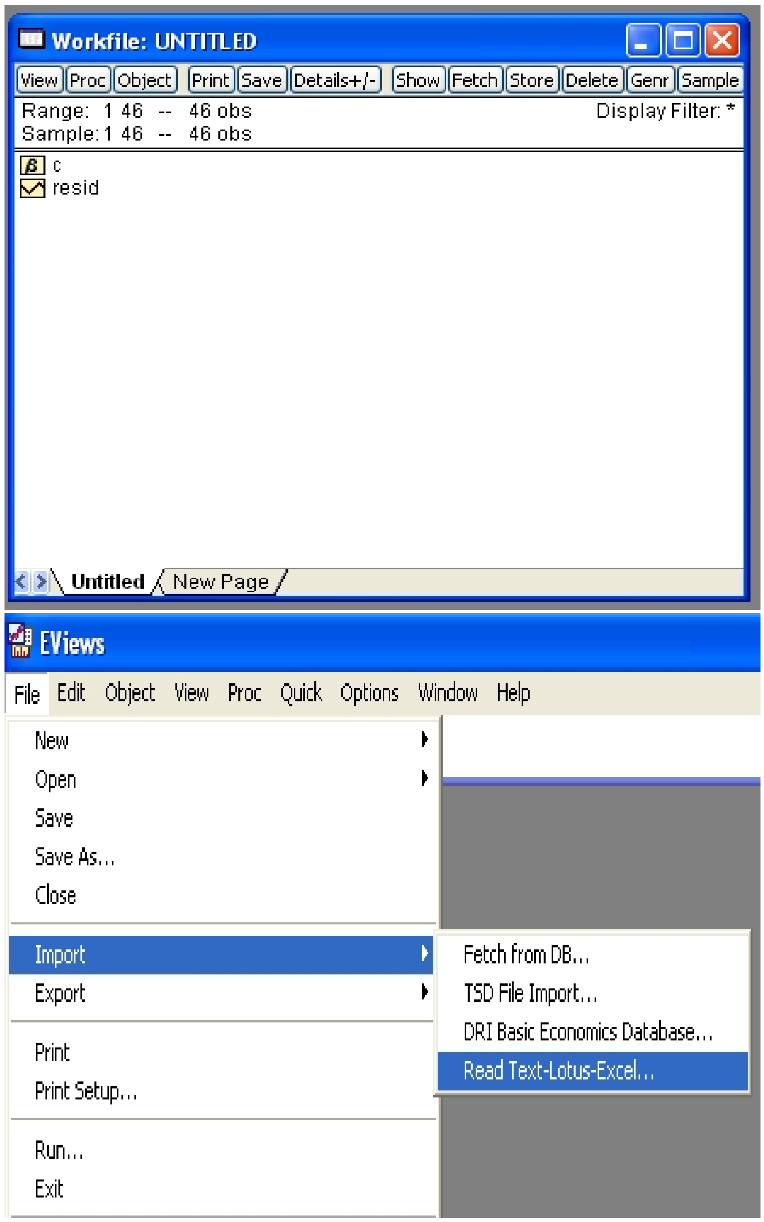


Figure 3: An unnamed workfile is opened, indicating the sample size, and EViews 'objects' that will later contain estimated parameter values (c) and estimated residuals (resid). Import the data into EViews, choosing the appropriate format.

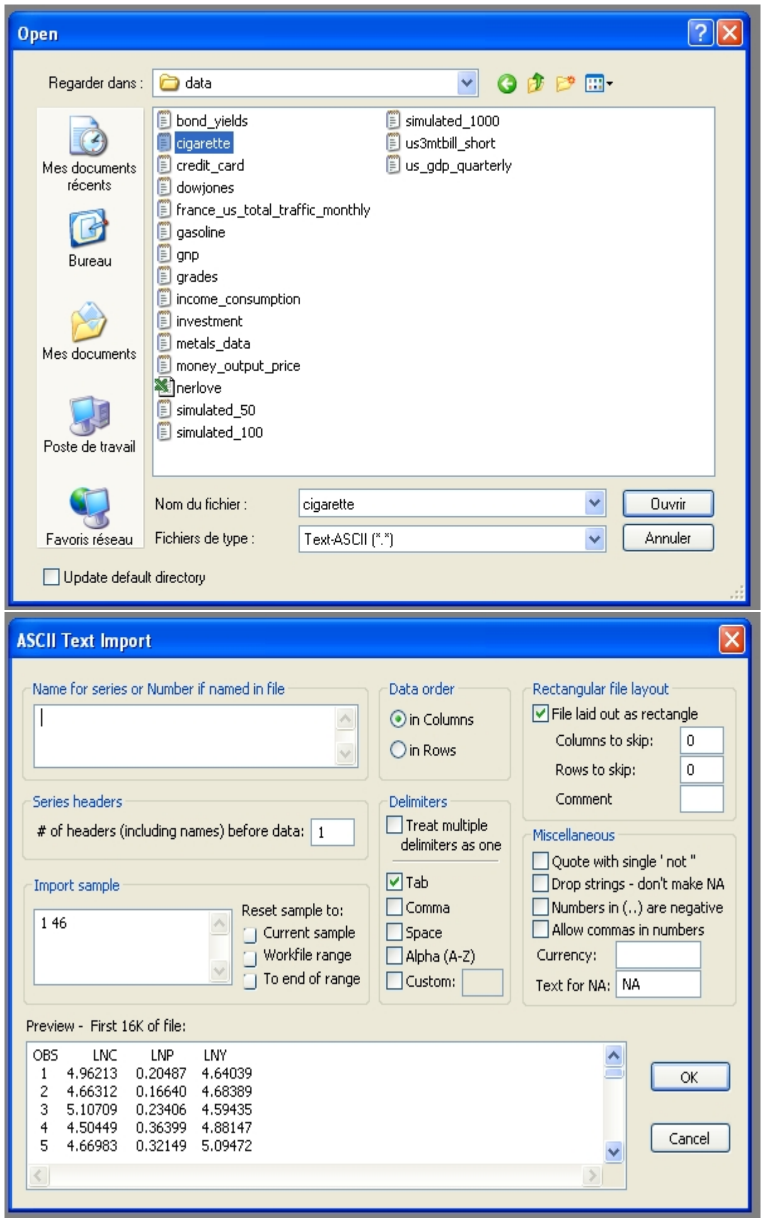


Figure 4: Navigate to the datafile, and open. EViews will preview the datafile, and enables various choices to be made concerning the import, including delimiter type, header, etc.

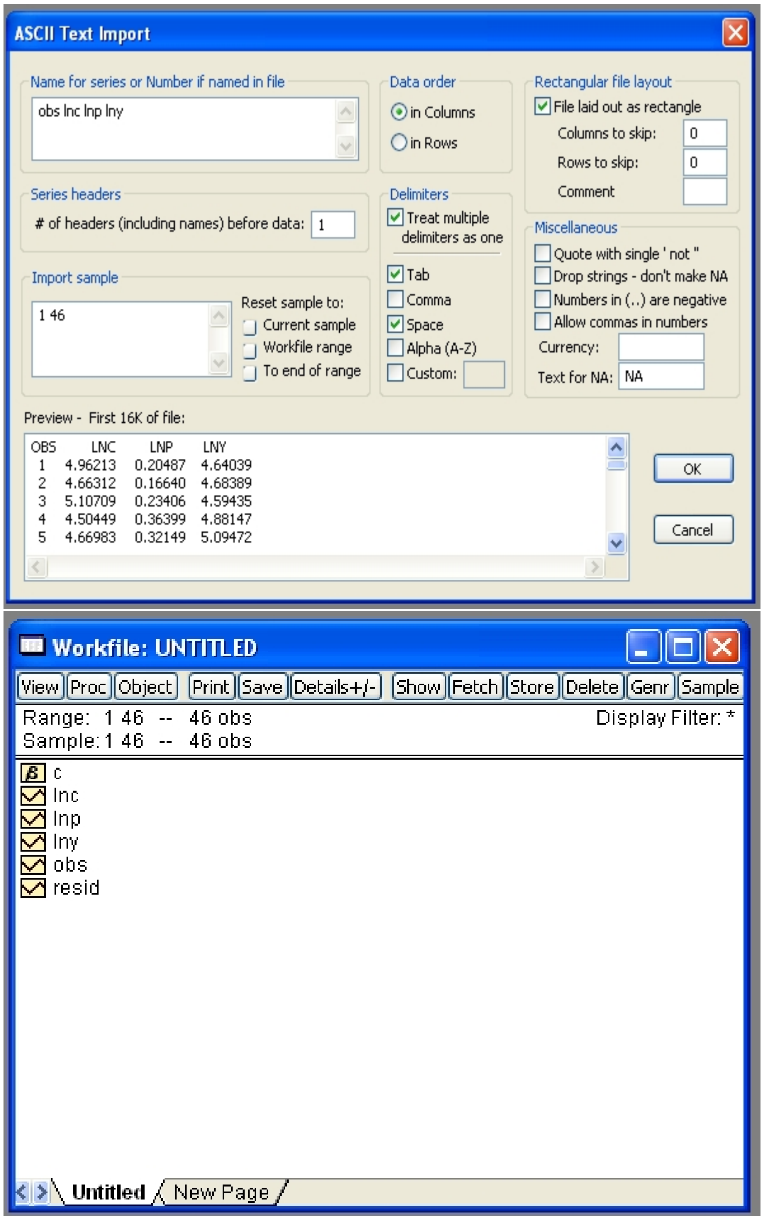


Figure 5: Choose names for the series (stored in columns), and select 'Treat multiple delimiters as one' and both 'Tab' and 'Space'. Verify the other settings. EViews will import the data, creating objects for each of the variables (lnc, lnp, lny, obs).

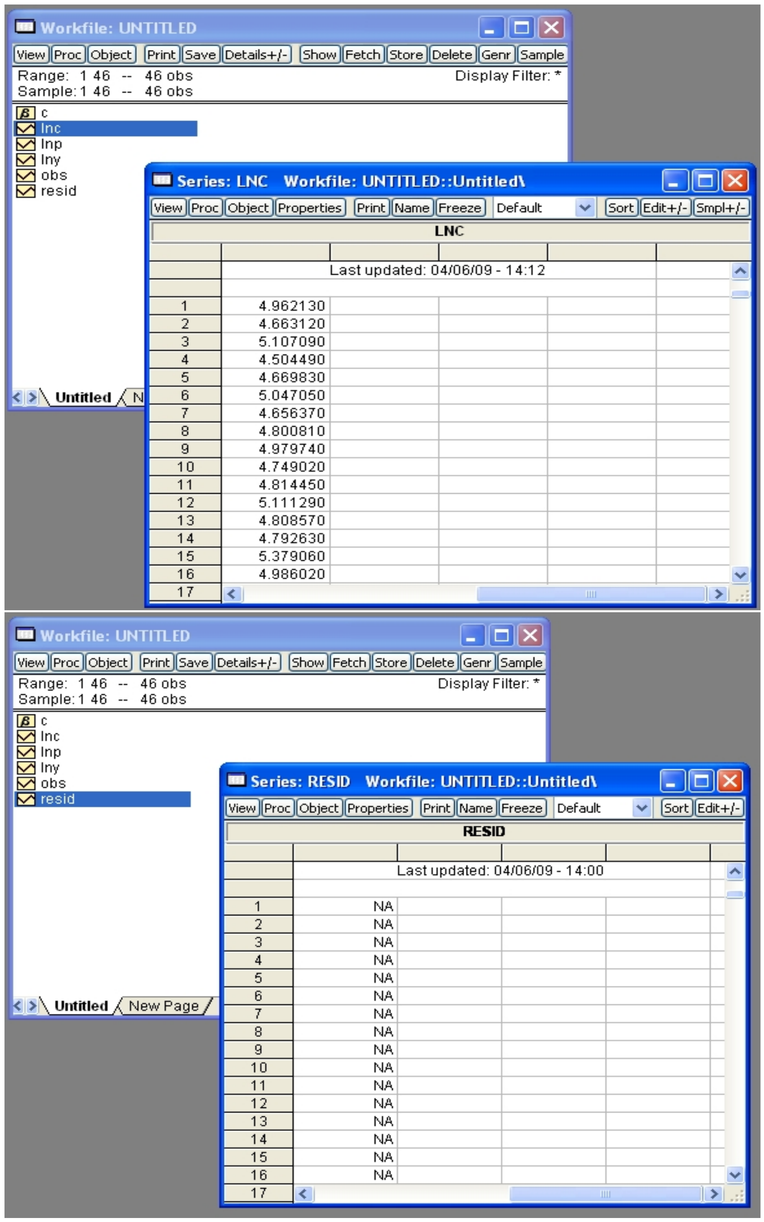


Figure 6: Click on lnc, to view the data on log consumption. Click on resid, to view the estimated residuals $\widehat{u}$: these have not yet been defined, since no model has been estimated.

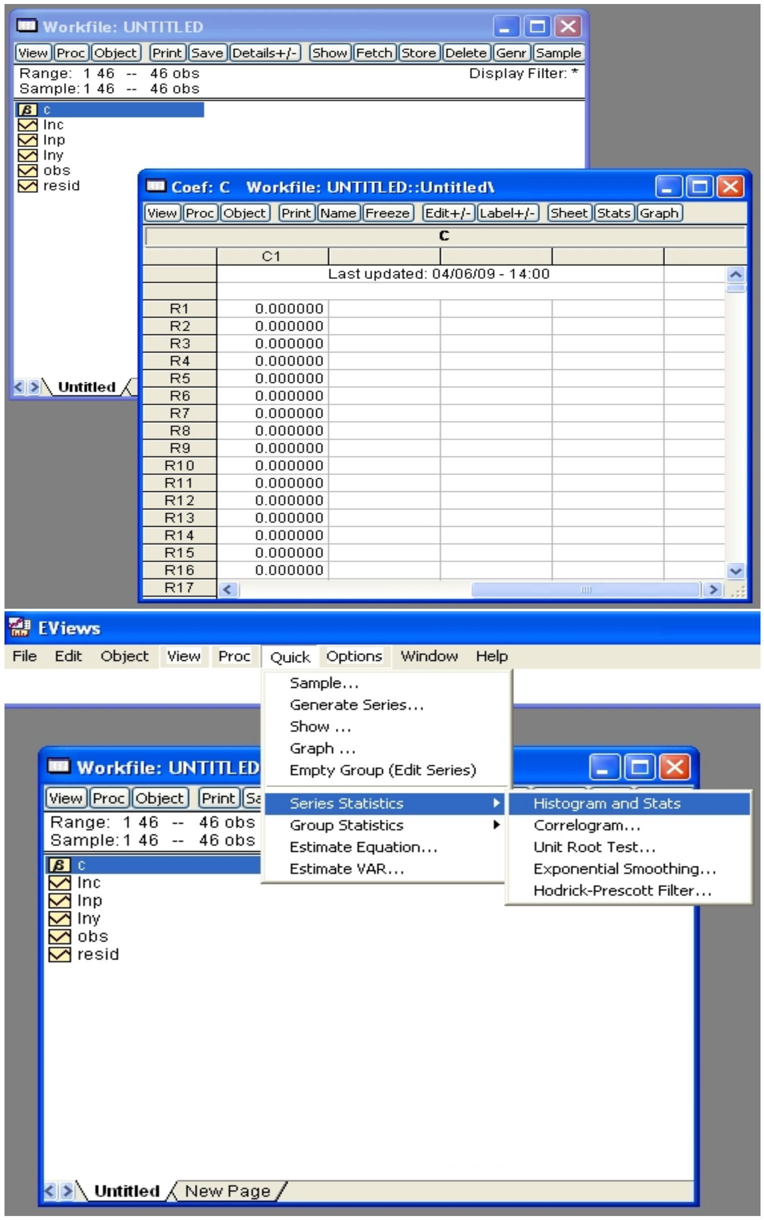


Figure 7: Select the estimated coefficient object c, which has not yet been defined since the model, and $\widehat{\beta}$, have not been estimated. Choose the 'Histogram and Stats' option from the Quick - Series Statistics menu.

Univariate Descriptive Statistics

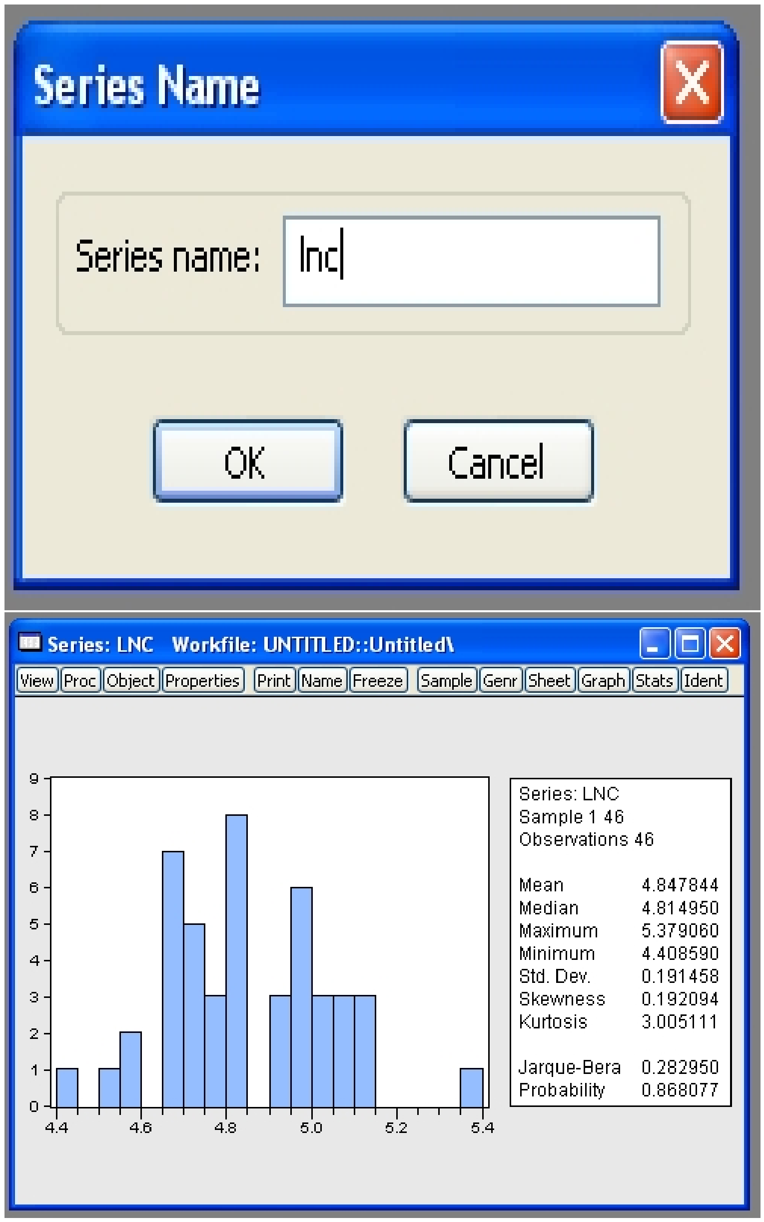
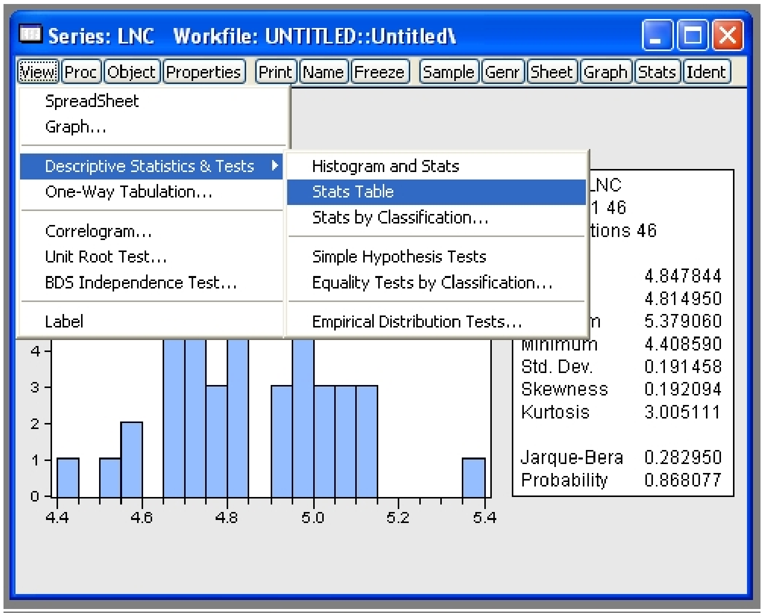
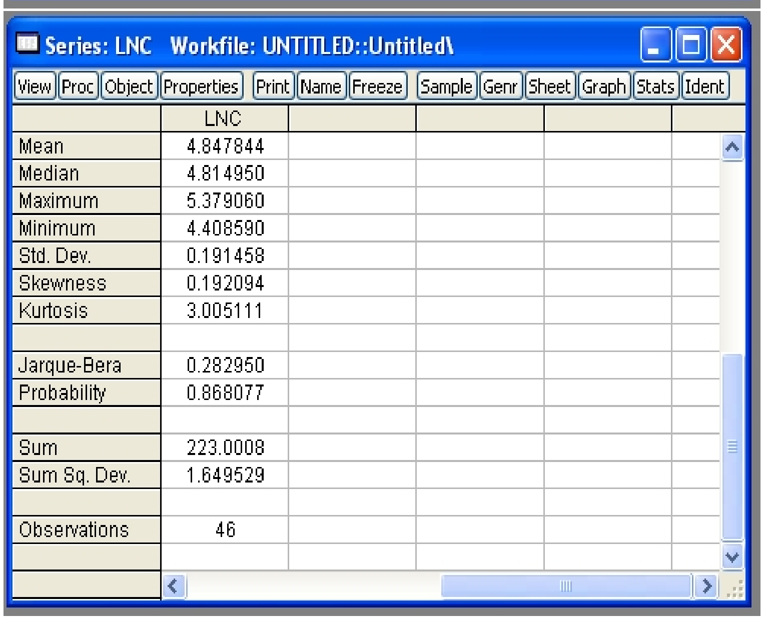


Figure 8: Enter the series name: lnc. The histogram of the data also gives some basic descriptive statistics, up to the standardized fourth central moment (kurtosis), and the Jarque-Bera test for normality of the data: this performs the joint test that the skewness = 0 and the kurtosis = 3.





	LNC
Mean	4.847844
Median	4.814950
Maximum	5.379060
Minimum	4.408590
Std. Dev.	0.191458
Skewness	0.192094
Kurtosis	3.005111
Jarque-Bera	0.282950
Probability	0.868077
Sum	223.0008
Sum Sq. Dev.	1.649529
Observations	46

Figure 9: Choose 'Stats Table', and view the same descriptive statistics in spreadsheet format, with the addition of the sum $\sum_i \text{lnc}_i$ and the sum of squared deviations $\sum_i(\text{lnc}_i - (n^{-1}\sum_i \text{lnc}_i))^2$, where n is the sample size.

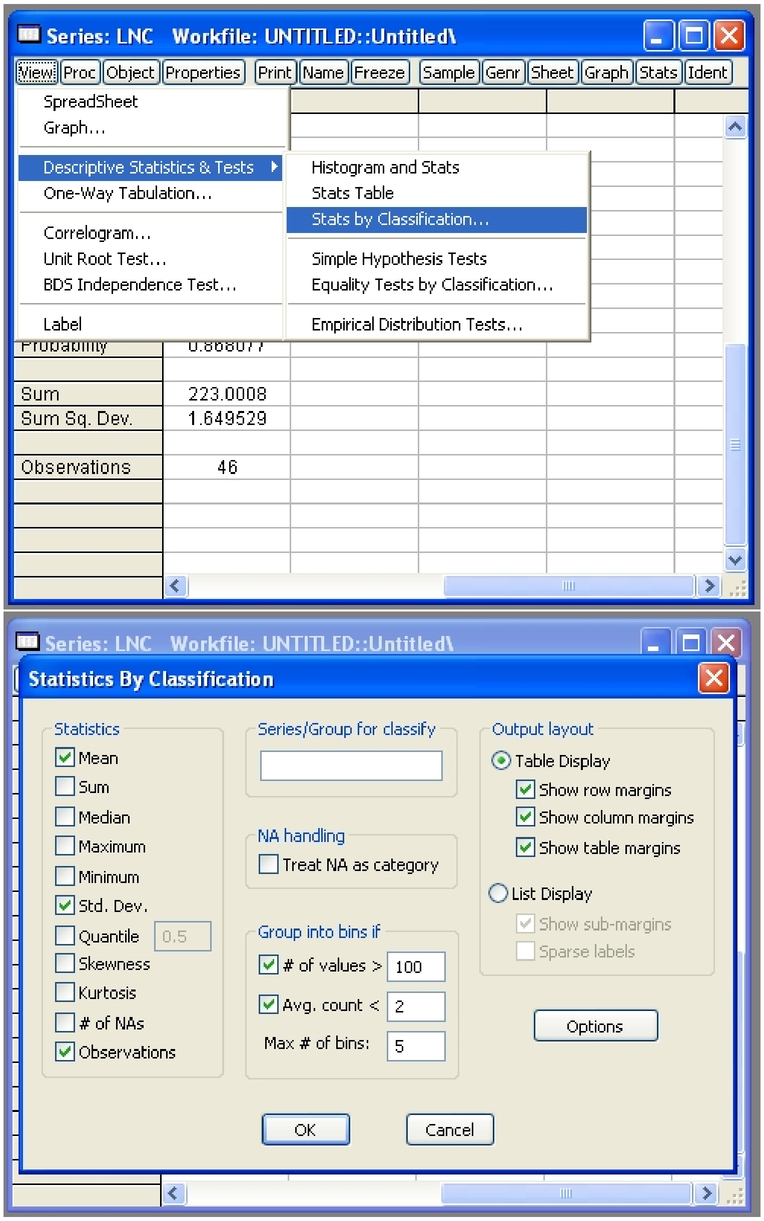
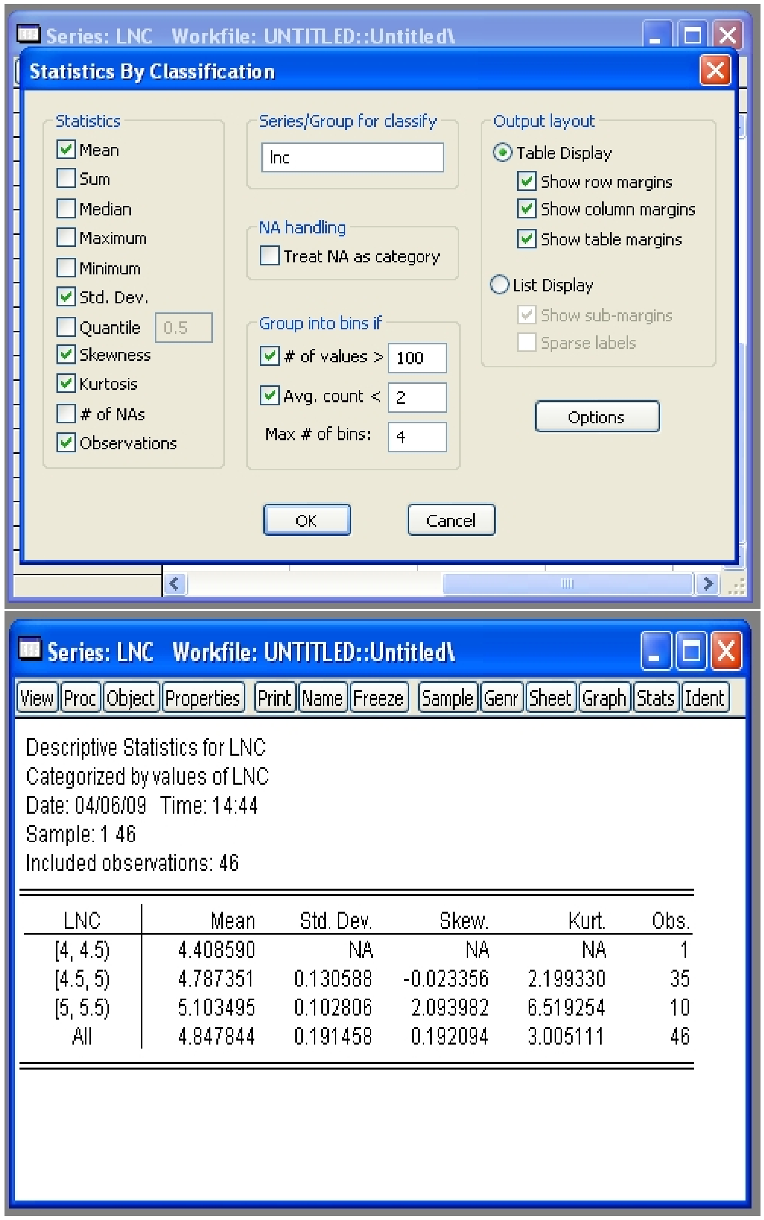


Figure 10: Choose 'Stats by Classification', which enables choice of descriptive statistics, binning of data, and other output options.



LNC	Mean	Std. Dev.	Skew.	Kurt.	Obs.
[4, 4.5)	4.408590	NA	NA	NA	1
[4.5, 5)	4.787351	0.130588	-0.023356	2.199330	35
[5, 5.5)	5.103495	0.102806	2.093982	6.519254	10
All	4.847844	0.191458	0.192094	3.005111	46

Figure 11: Select the mean, standard deviation, skewness, kurtosis, and the number of observations, choose the series lnc, and set the maximum number of bins to 4. The data is grouped (binned) if (a) lnc has more than 100 distinct values (not relevant here, since $n = 46$), or (b) if each distinct value of lnc occurs less than twice. The maximum number of bins only provides approximate control over the actual number that will be selected by EViews. The displayed output results. **Why are some statistics not computed (NA) for the first bin?**

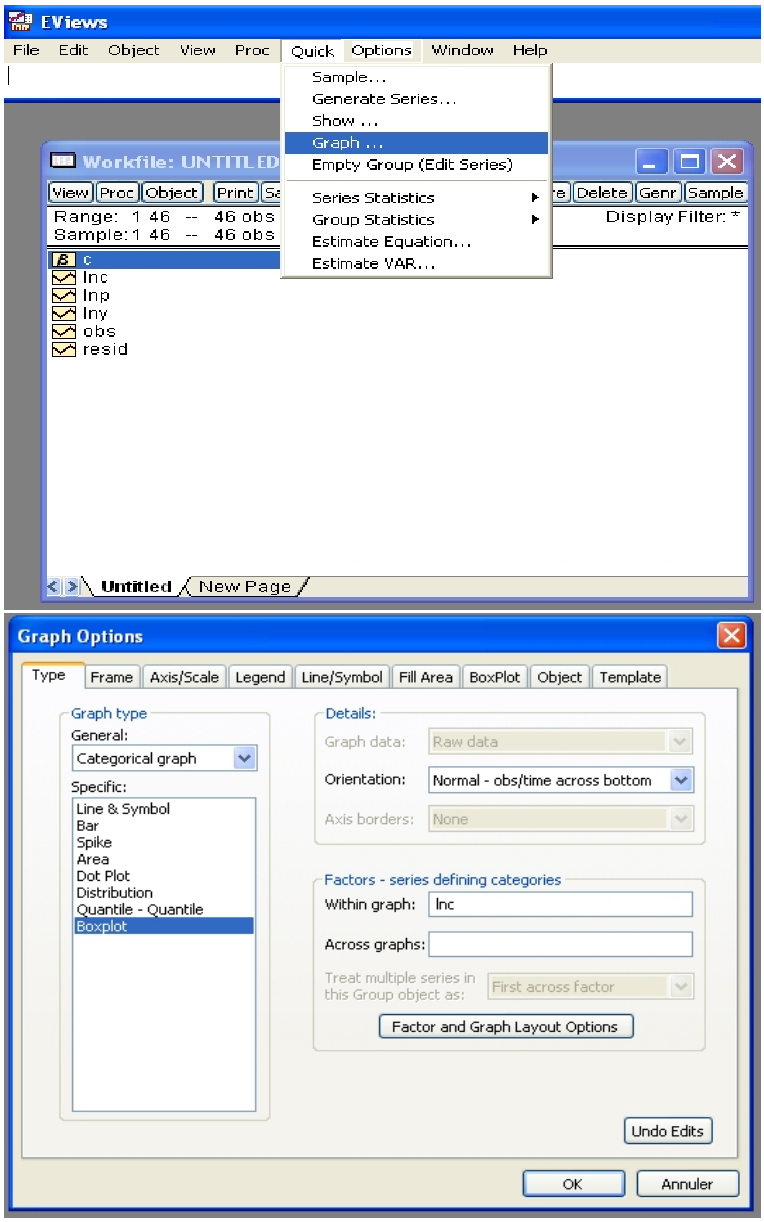


Figure 12: Select Quick - Graph. Choose the series lnc, and then select Categorical graph - Boxplot with lnc for 'Within graph'. Before plotting the graph, choose 'Factor and Graph Layout Options'.

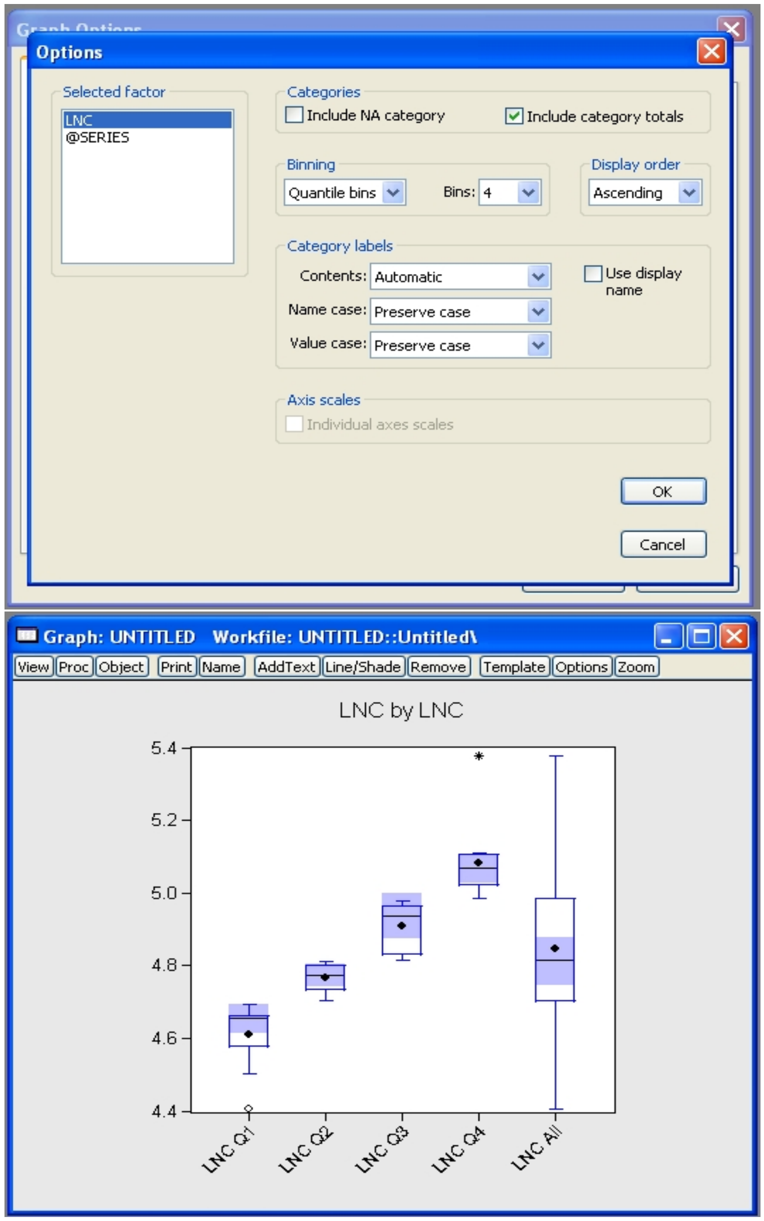


Figure 13: Select 4 quantile bins, and check the 'Include category totals' box. Plot the boxplots for each of the four quartiles of lnc, and for the full sample. **Can you identify the mean, median, first quartile Q1, third quartile Q3, staples (the last datapoints that do not fall below Q1-1.5IQR, or exceed Q3$+1.5$IQR, where IQR is the interquartile range Q3$-$Q1) and far outliers (observations which fall below Q1-3IQR, or exceed Q3$+3$IQR)?** The shaded areas correspond to median± 1.57IQR$/\sqrt{n}$, and give an approximate confidence interval for equality of median across bins (if the shaded areas do not overlap, then the medians are (roughly) not equal).

A Hypothesis Test on a Variable Mean

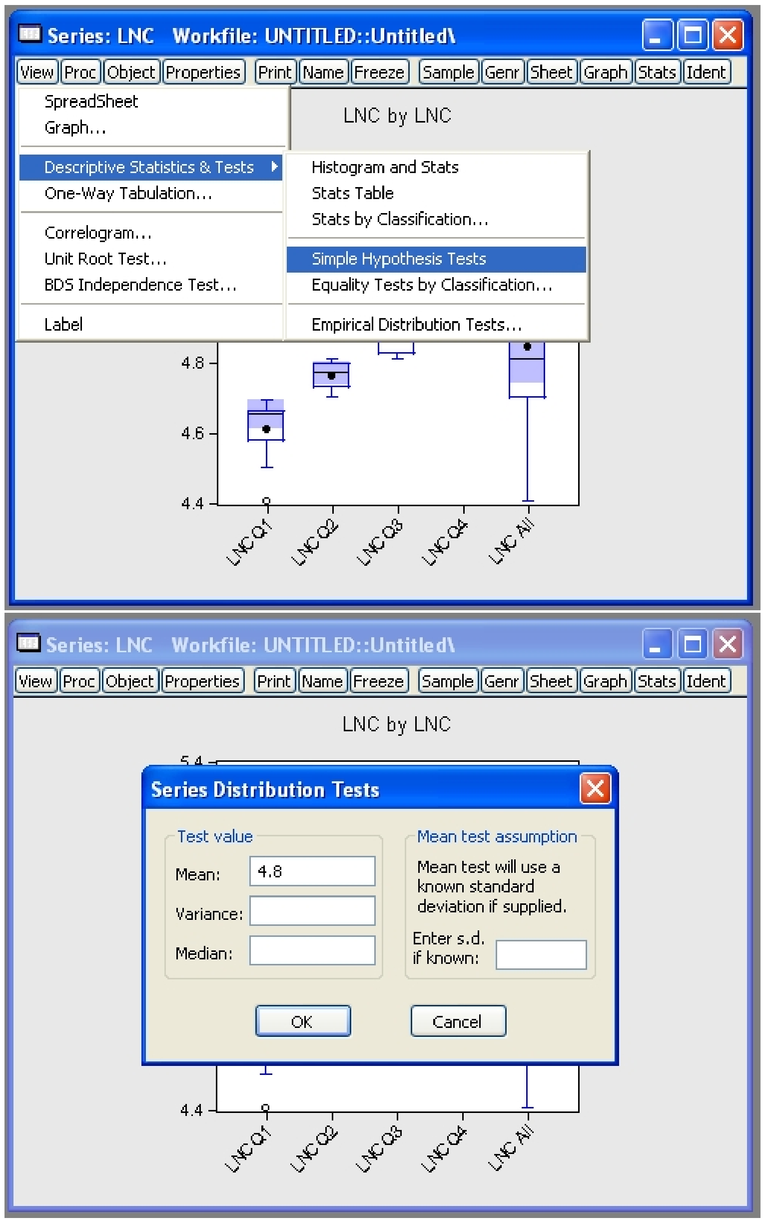


Figure 14: Select 'Simple Hypothesis Tests' and choose $\bar{x}_0 = 4.8$, to test $H_0 : \bar{x} = \bar{x}_0 = 4.8$ using a t statistic $t = \sqrt{n}(\bar{x} - \bar{x}_0)/\widehat{\sigma} \sim t(n-1)$ under the null hypothesis, if x (lnc) is normally distributed.

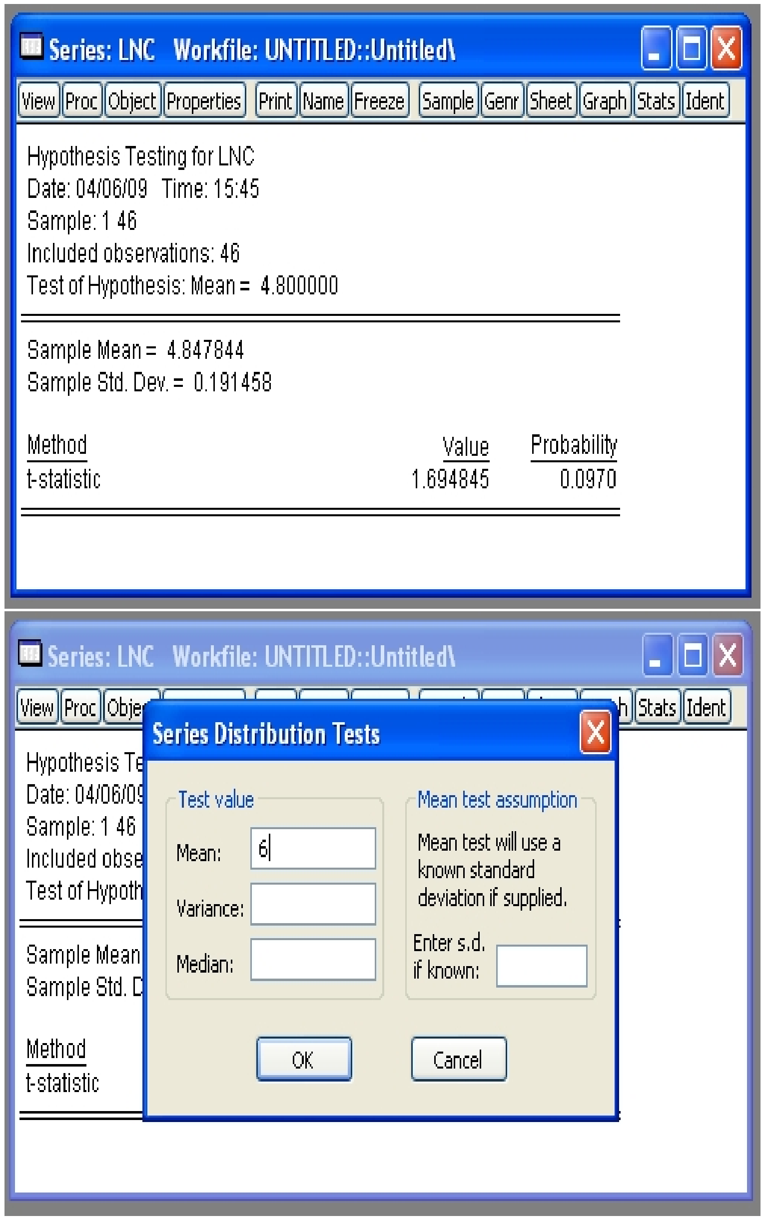


Figure 15: Result of hypothesis test $H_0 : \bar{x} = \bar{x}_0 = 4.8$ against the two-sided alternative. **Use statistical tables to perform this test manually at the 95% level of significance: what is the decision rule and what is the outcome of the test?** Note that the null is rejected at the 90% level of significance (this can be seen easily, since the 'Probability' is less than 0.10). Now set $\bar{x}_0 = 6$, to test $H_0 : \bar{x} = \bar{x}_0 = 6$.

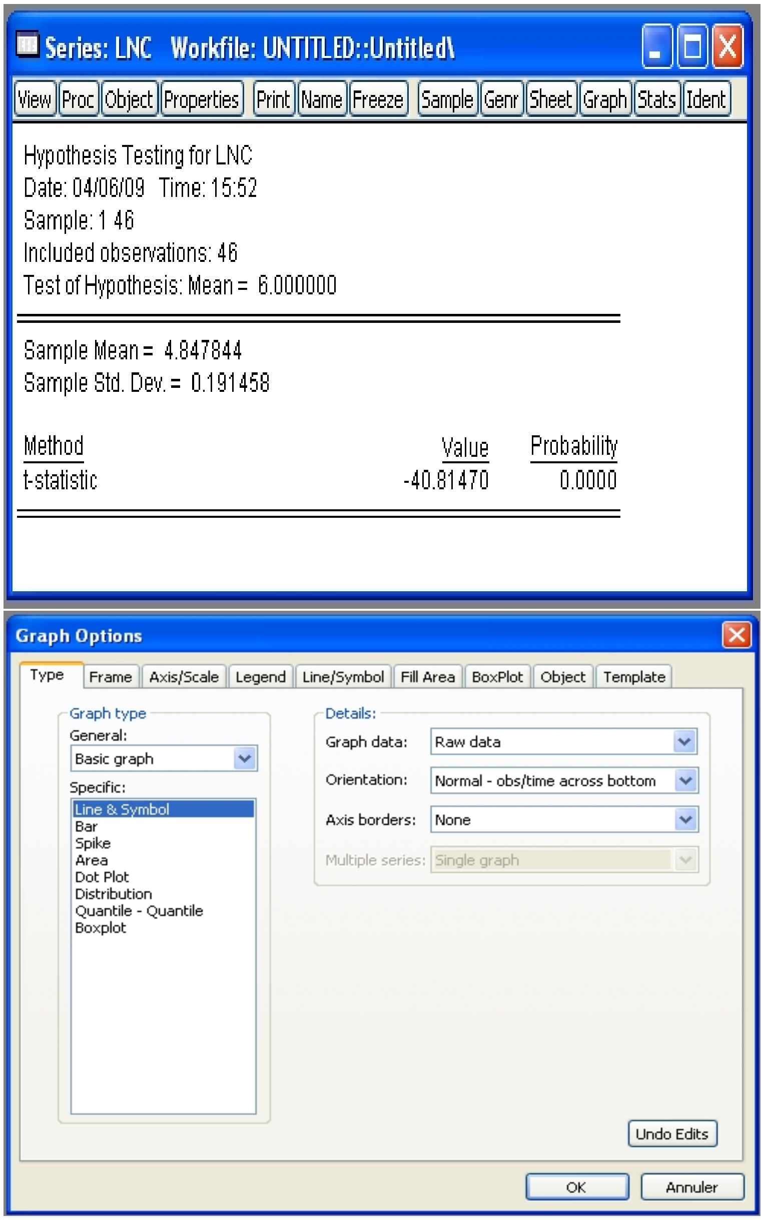


Figure 16: Result of hypothesis test $H_0 : \bar{x} = \bar{x}_0 = 6$ against the two-sided alternative. **Perform this test manually at the 95% level, using statistical tables.** Note that the null is rejected at the 99% level of significance (since the 'Probability' is less than 0.01). For lnc, select a Basic graph - Line & Symbol.

Plot Formatting

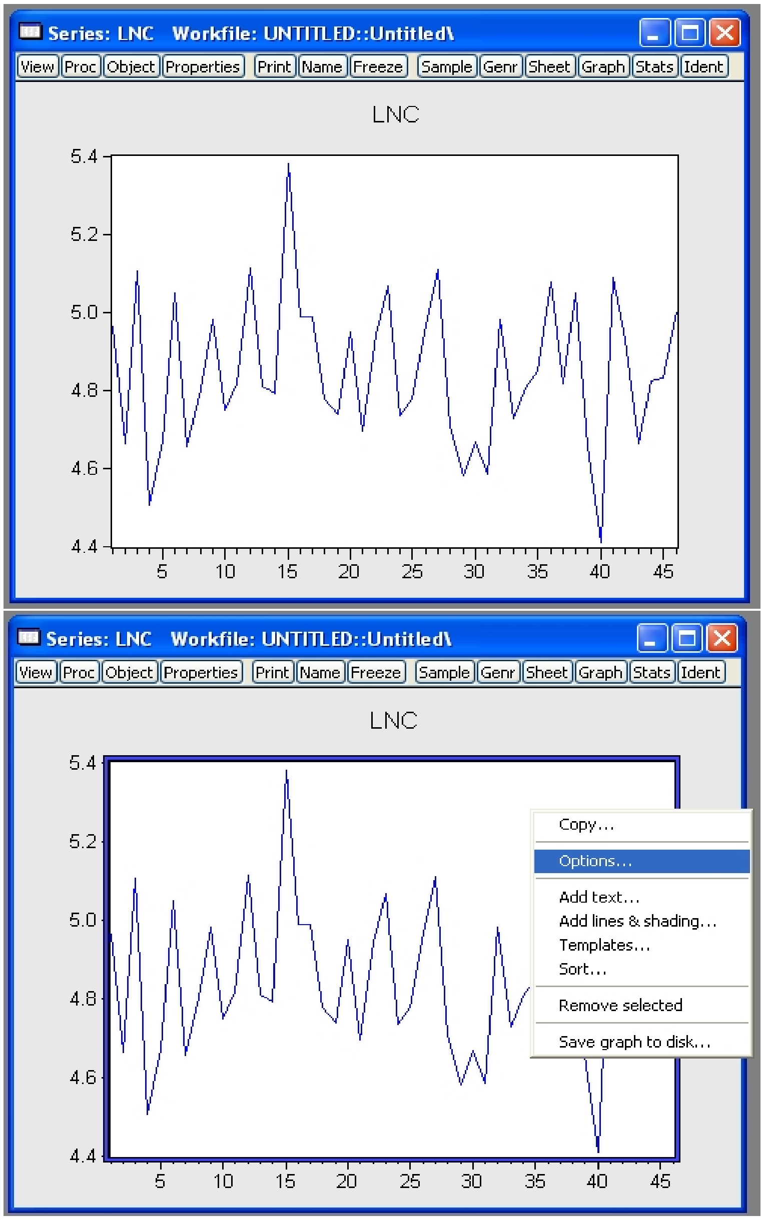


Figure 17: Plot of lnc against observation number (data not ordered). Select the graph (right click), and choose 'Options'.

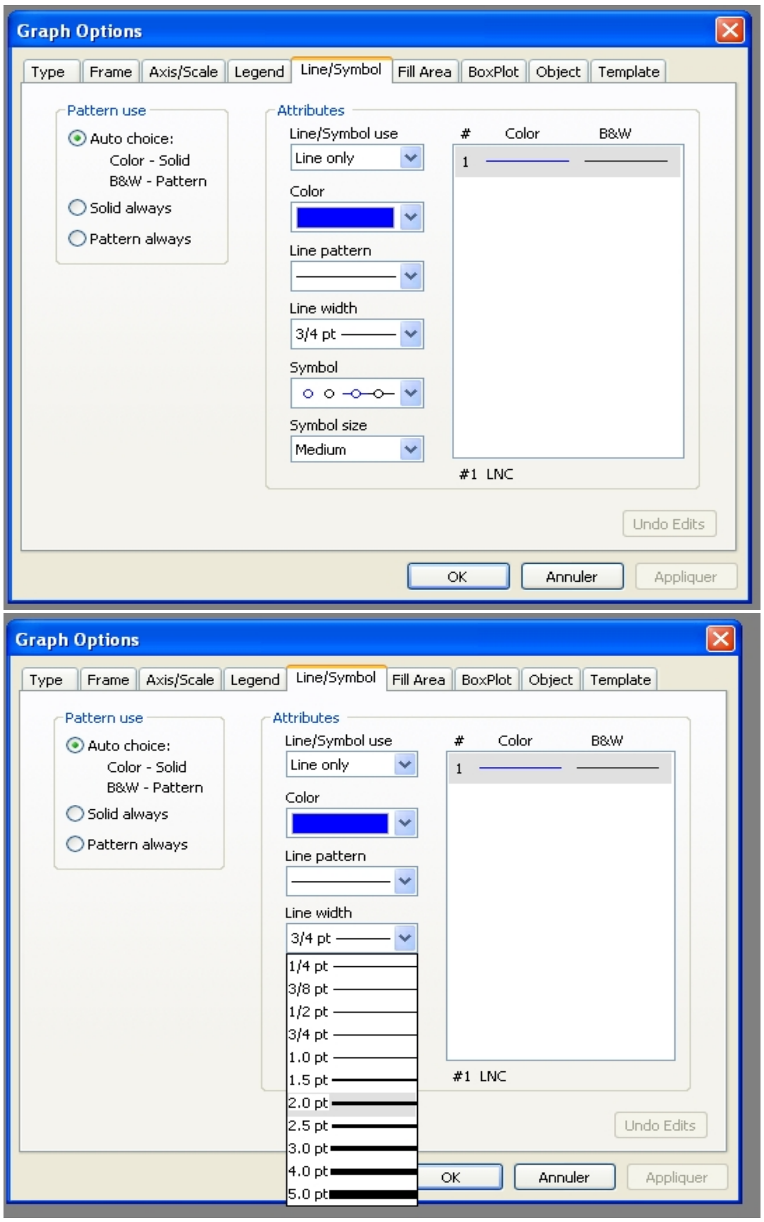


Figure 18: From the option screen, choose line width, and set to 2.0pt.

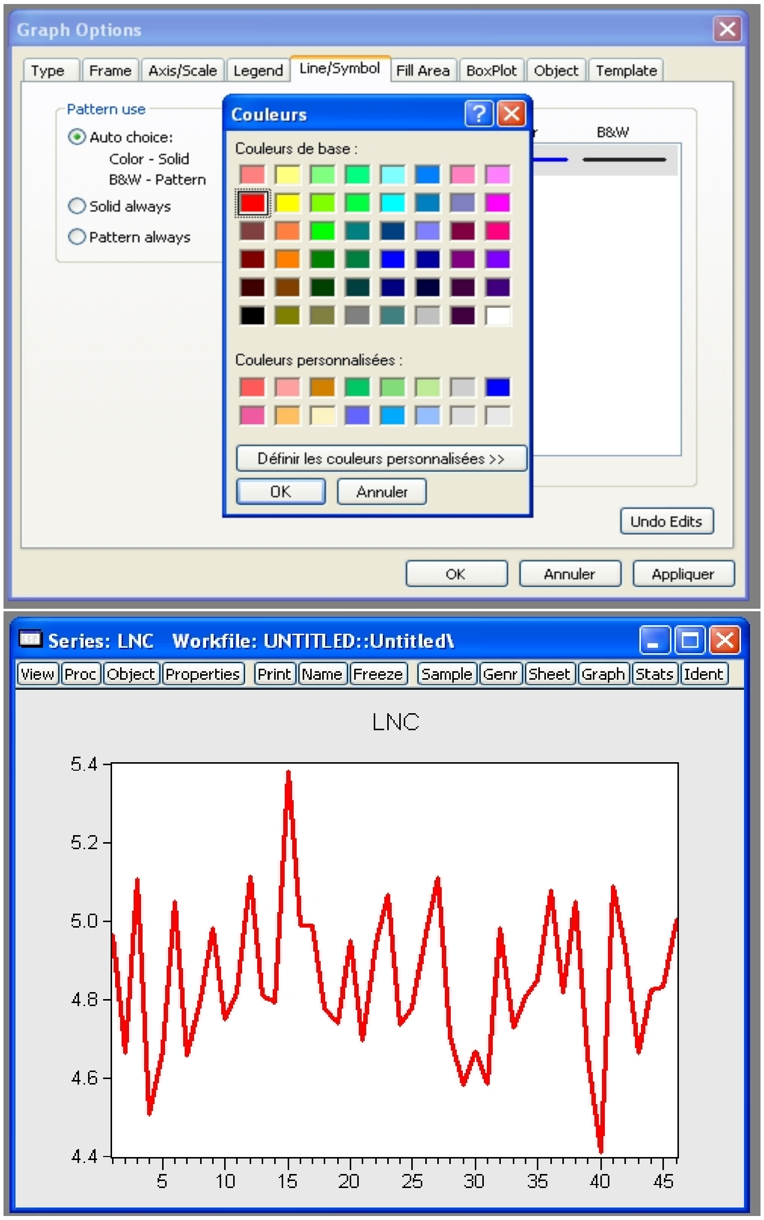


Figure 19: From the option screen, change the line colour. Display the final plot.

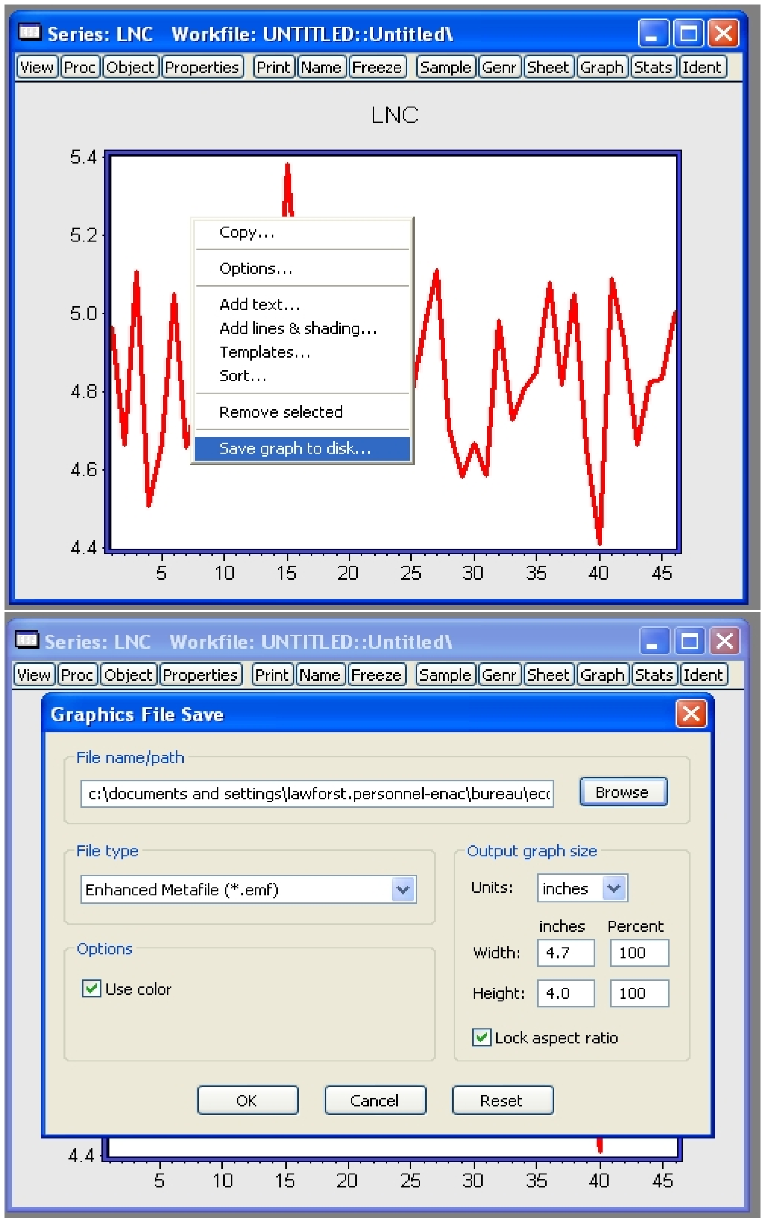


Figure 20: Select the graph, and choose 'Save graph to disk'. Choose a path and filename, and set the filetype to `.emf`. Save the graph.

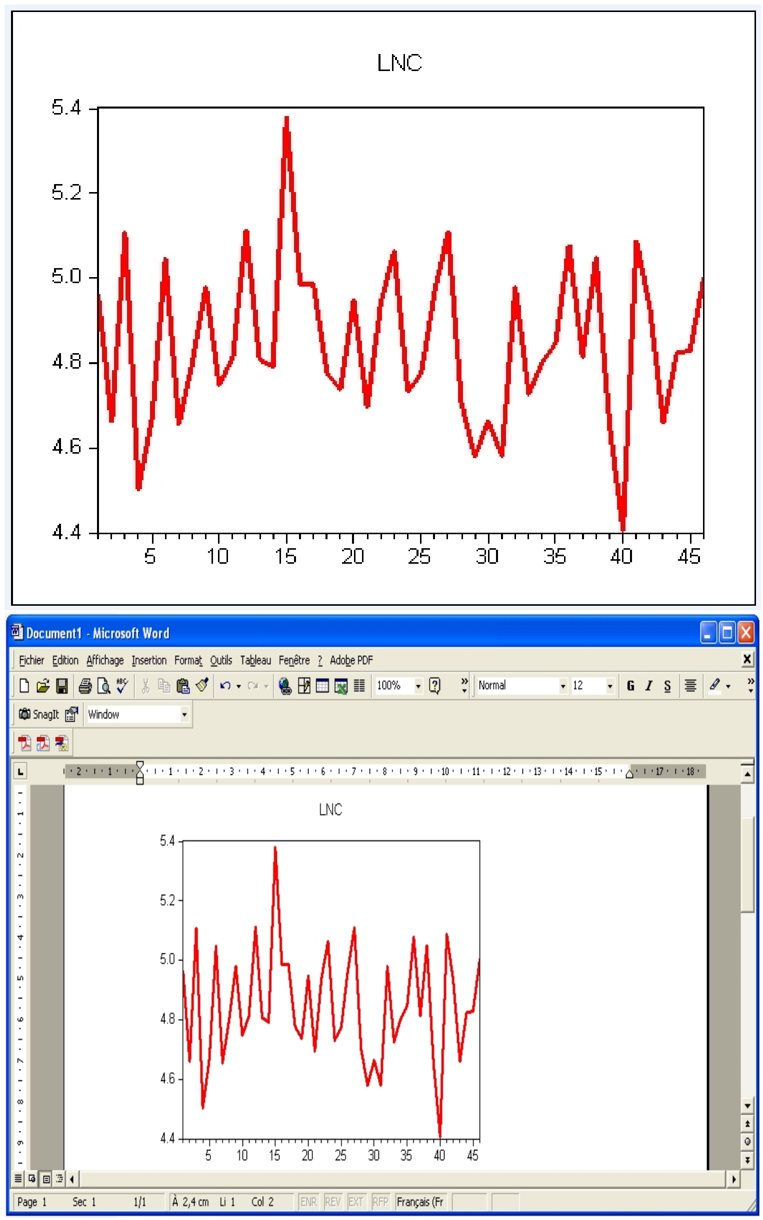


Figure 21: Find the graph file and double-click on it. The graph can also be copied and pasted in the usual way: copy it into a Word document, or similar.

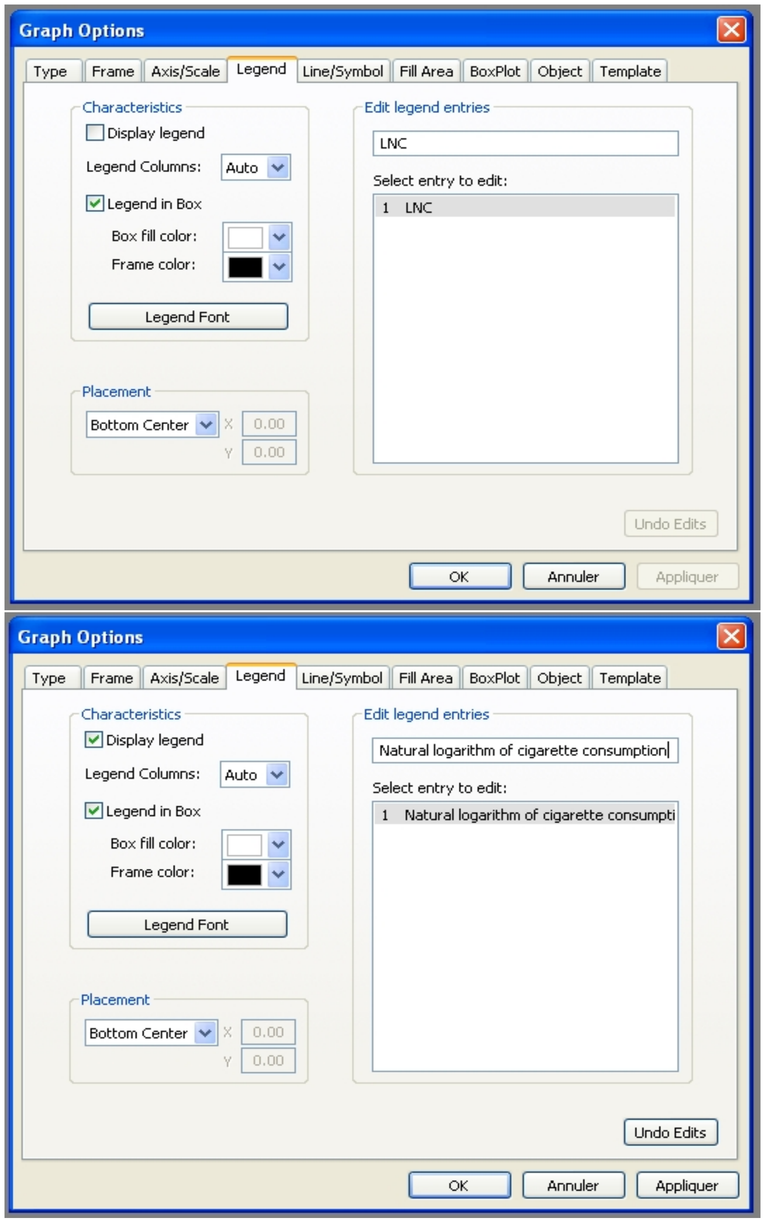


Figure 22: Select Options and Legend. Check 'Display legend' and change the legend to 'Natural logarithm of cigarette consumption'.

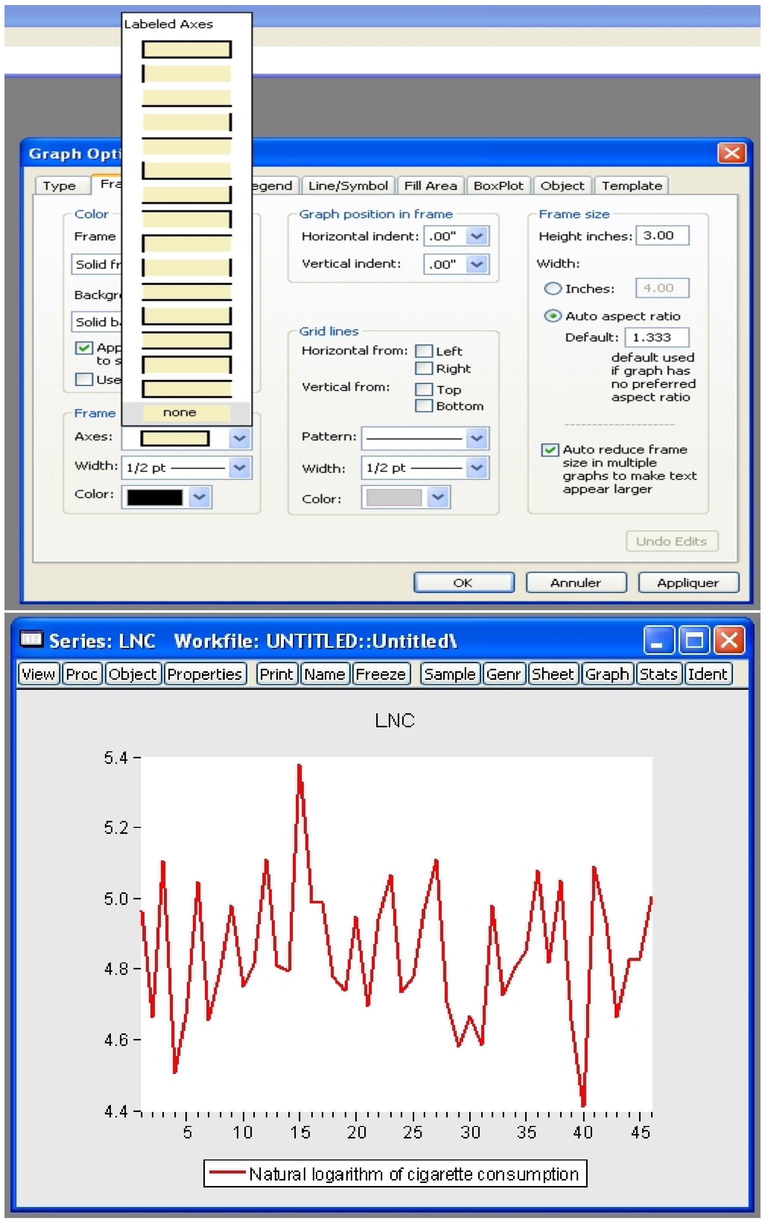


Figure 23: Remove the frame around the graph. Display the plot.

Univariate Descriptive Statistics (Continued)

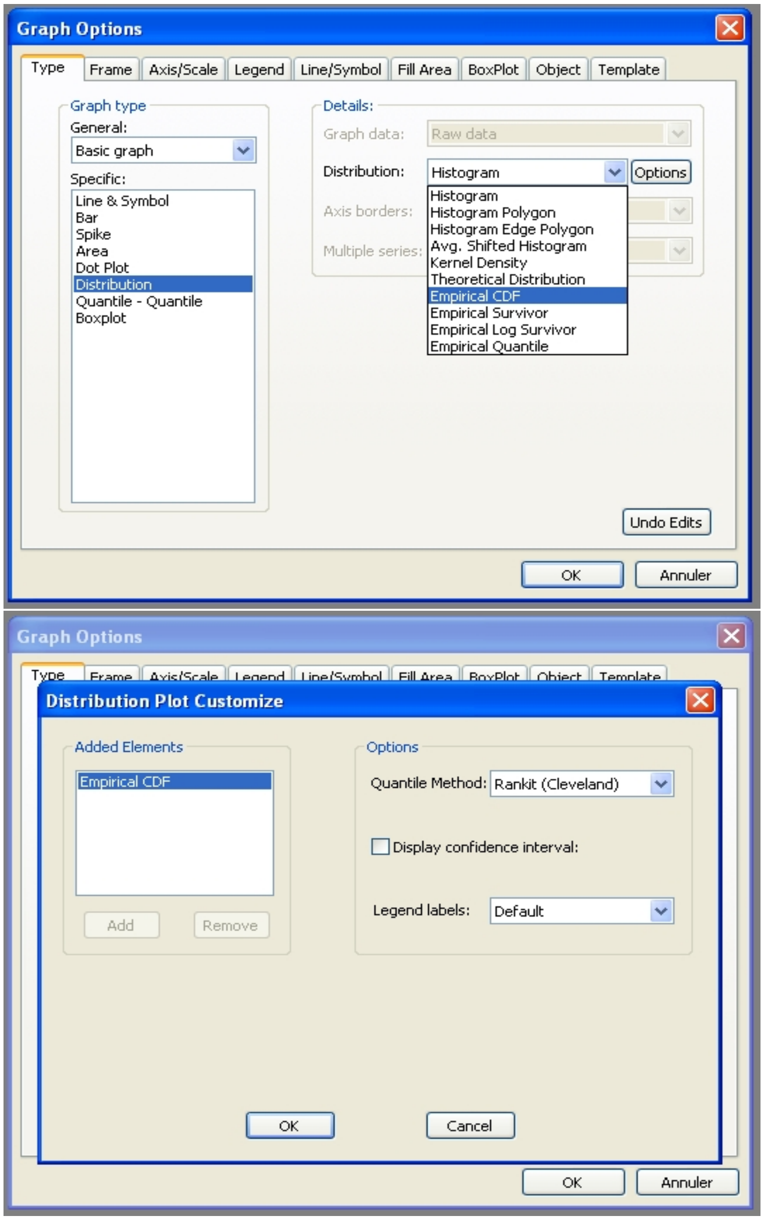


Figure 24: Select an Empirical CDF distribution plot, and uncheck the 'Display confidence interval' box.

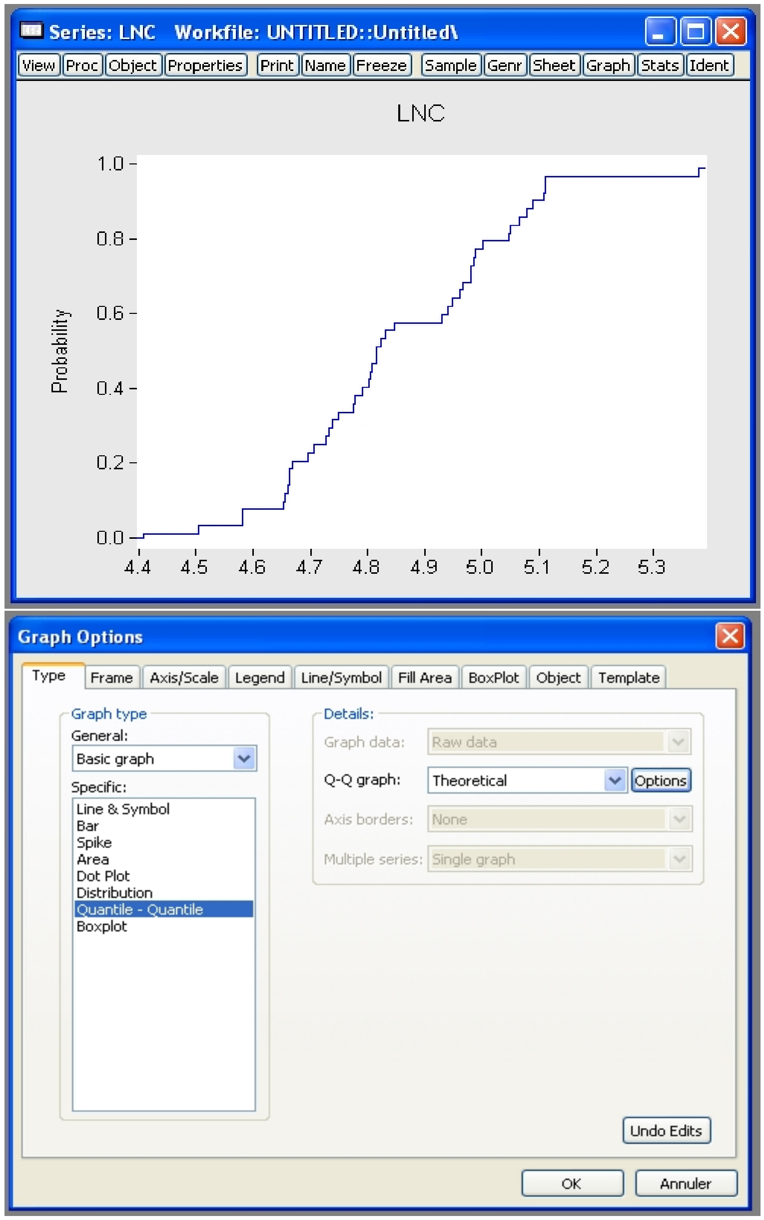


Figure 25: Display the plot. The empirical cumulative distribution plot displays a data-calculated $\text{Prob}(X \leq x)$ for all x. Select a Quantile - Quantile plot.

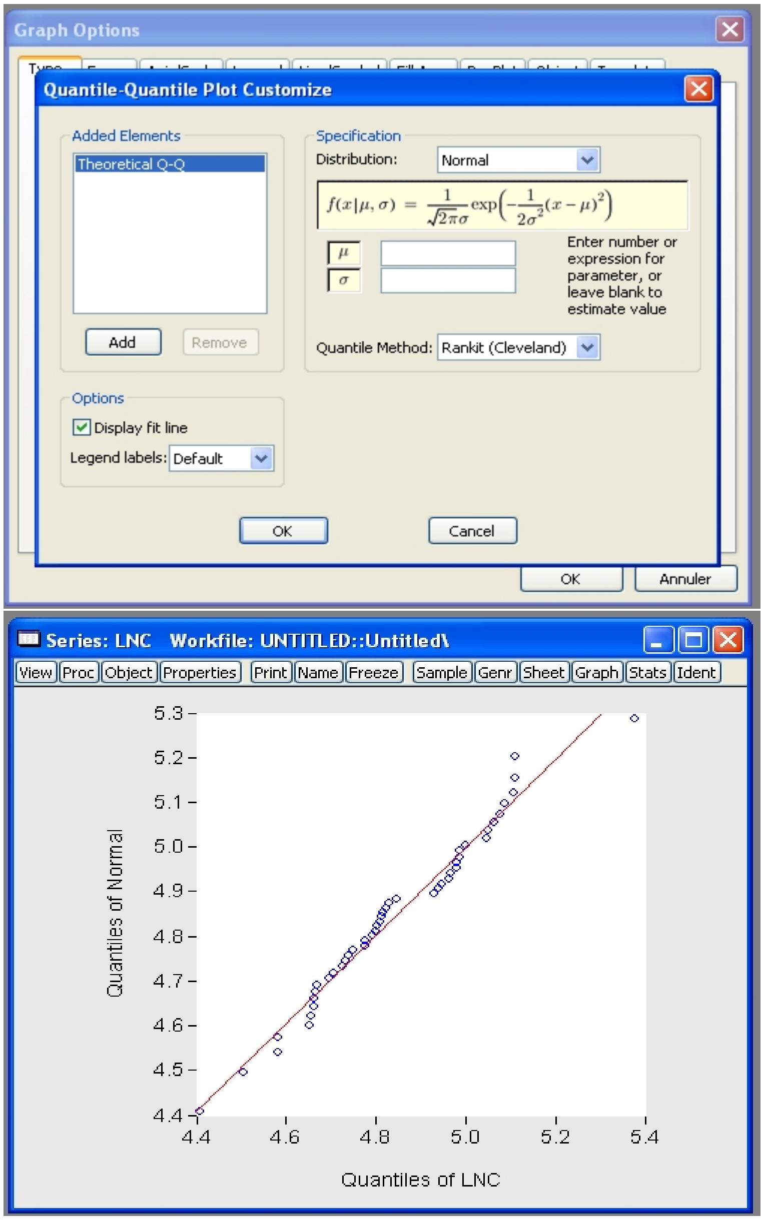


Figure 26: Select 'Normal' distribution and 'Display fit line', to give the quantile-quantile plot of the data against the normal distribution: this plots each quantile of the observed data q_α against the corresponding quantile from the normal distribution (standardized to have the same mean and variance as the observed data).

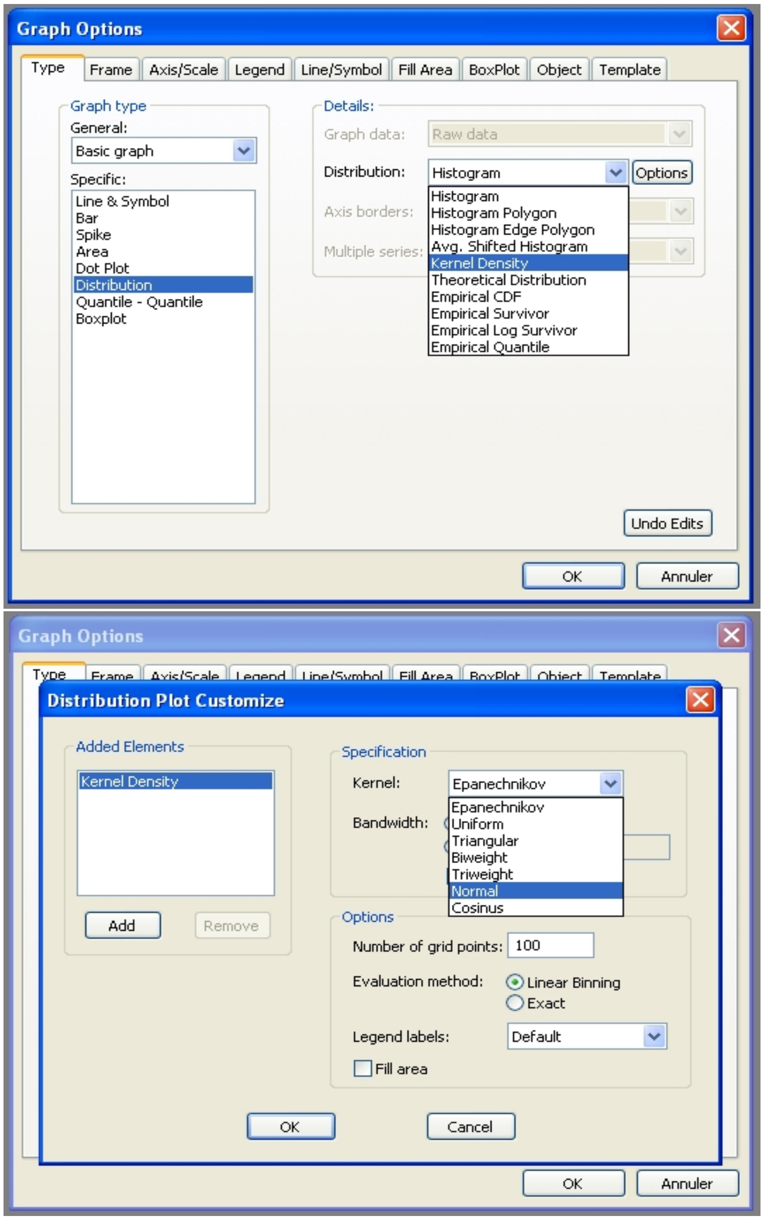


Figure 27: Choose a 'Kernel Density' distribution plot, with 'Normal' kernel.

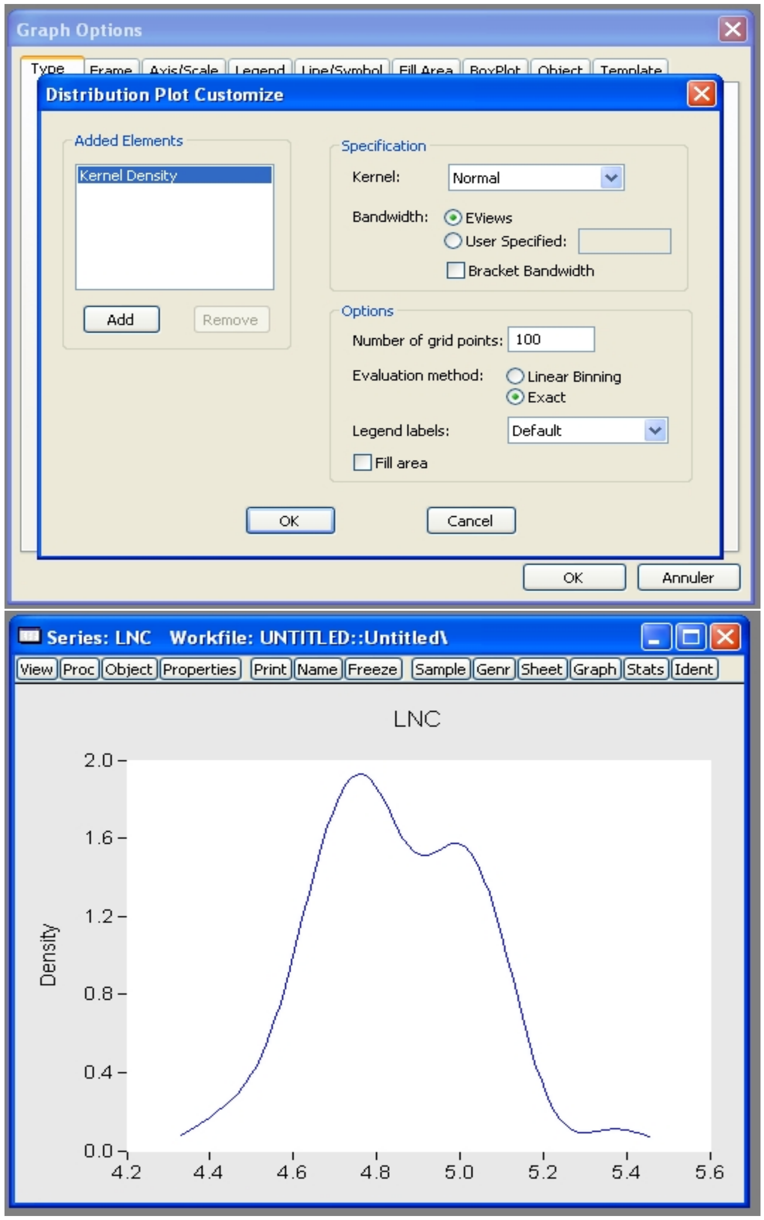


Figure 28: Select 'Exact' instead of 'Linear Binning', to give the Gaussian kernel density plot of the data: $\widehat{f}(x) = (nh)^{-1} \sum_i K((x - X_i)h^{-1})$, where n is the sample size, h is the kernel bandwidth (chosen automatically by EViews), and $K()$ is the kernel (here, the N(0,1) pdf). **Compare this to the histogram of the data that was plotted earlier.**

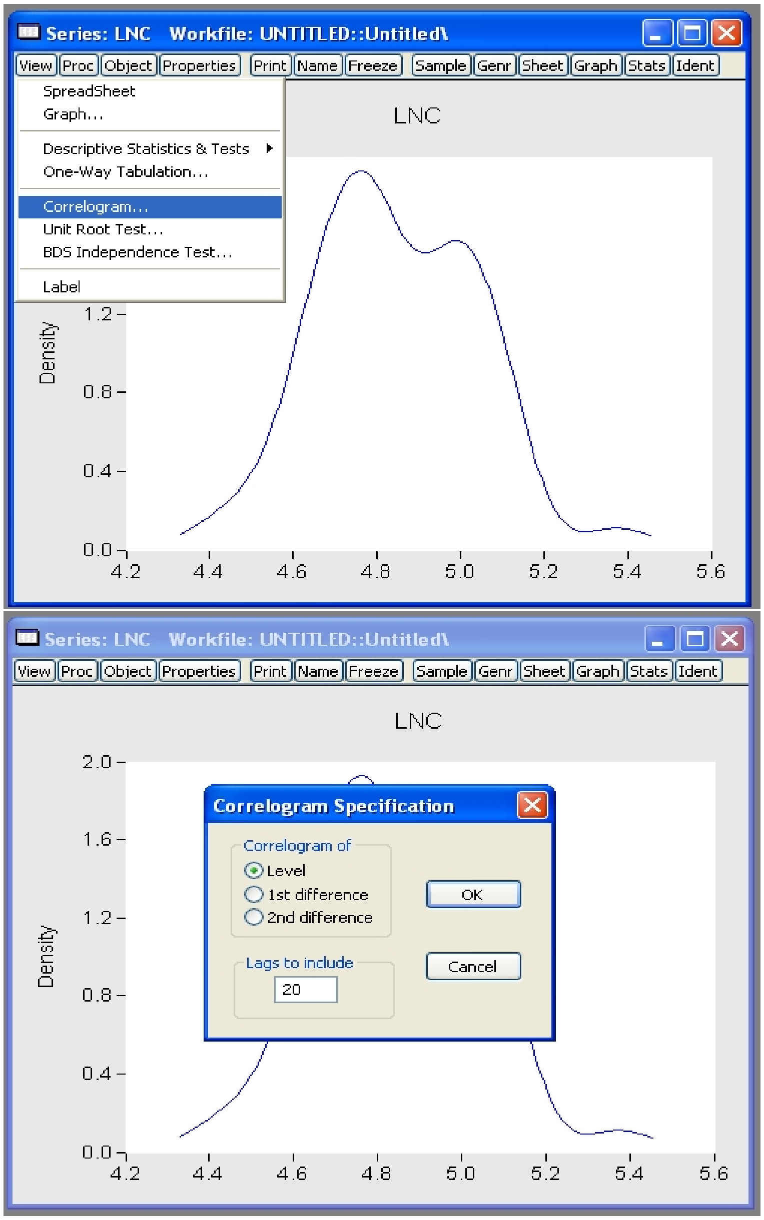


Figure 29: Select ‘Correlogram’ and ‘Level’, with 20 lags.

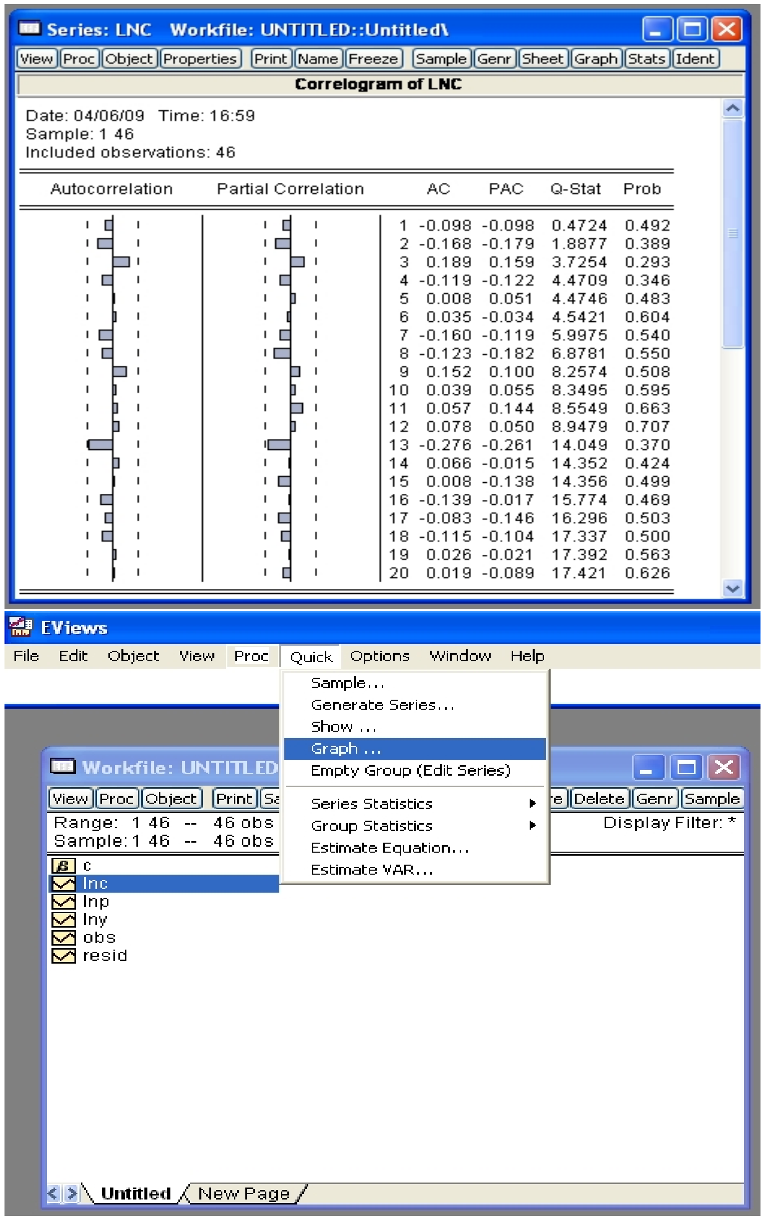


Figure 30: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of lnc, for the first 20 lags (s), with associated Ljung-Box Q statistic, and probability that the autocorrelations up to and including lag s are jointly equal to zero: this suggests that the series is approximately white noise (*note that this test is not really useful, for the cross-sectional data that we are considering here; we will return to this when working with time series data*). Select a Quick - Graph.

Multivariate Descriptive Statistics and Groups

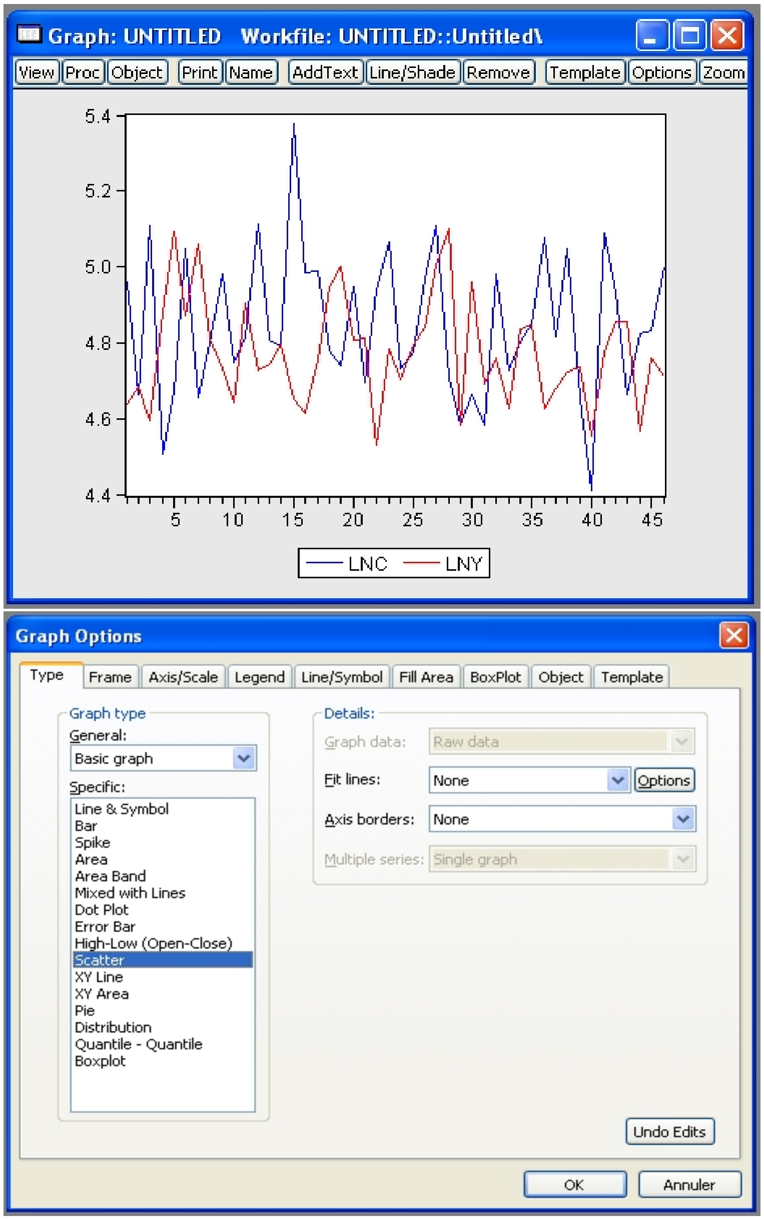


Figure 31: Plot *both* lnc and lny against the observation number. Select a 'Scatter' graph.

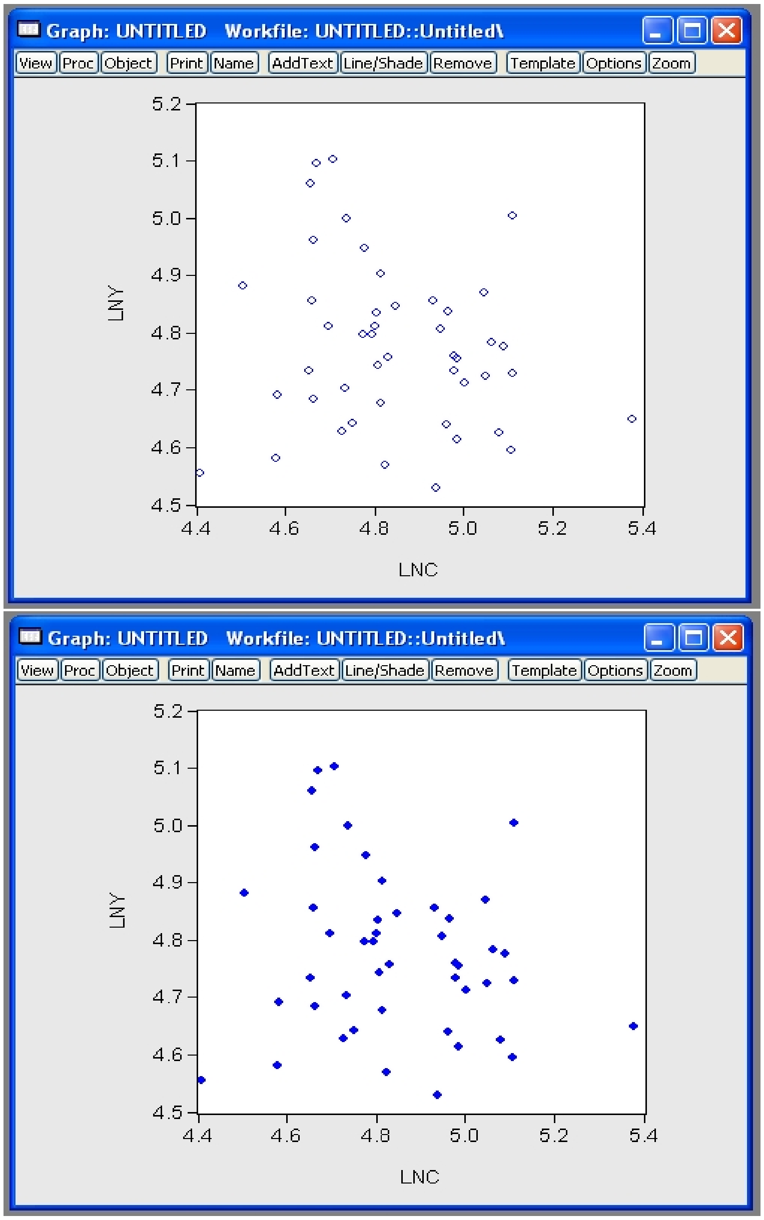


Figure 32: Choose series lnc and lny, to give a scatter plot of log income against log consumption (note how the order of series selection corresponds to the axes). Using 'Options', change the symbol plot type.

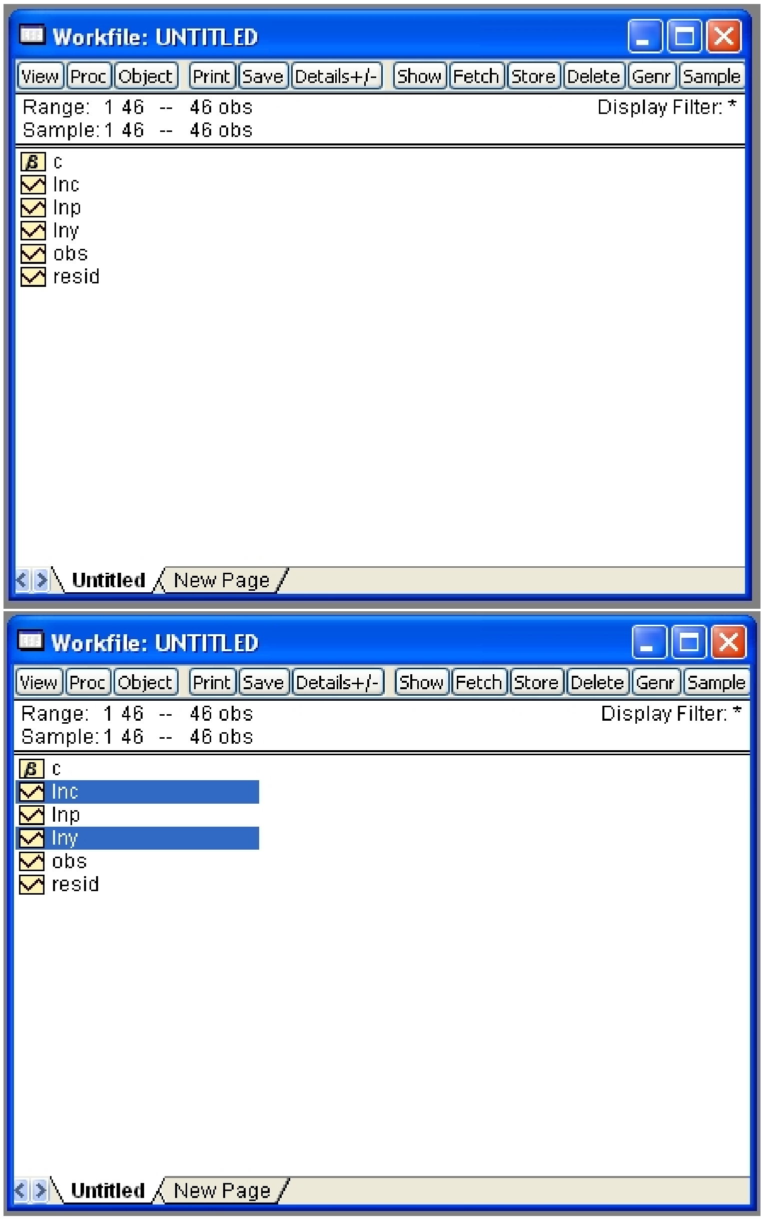


Figure 33: In the workfile, select the series lnc and lny.

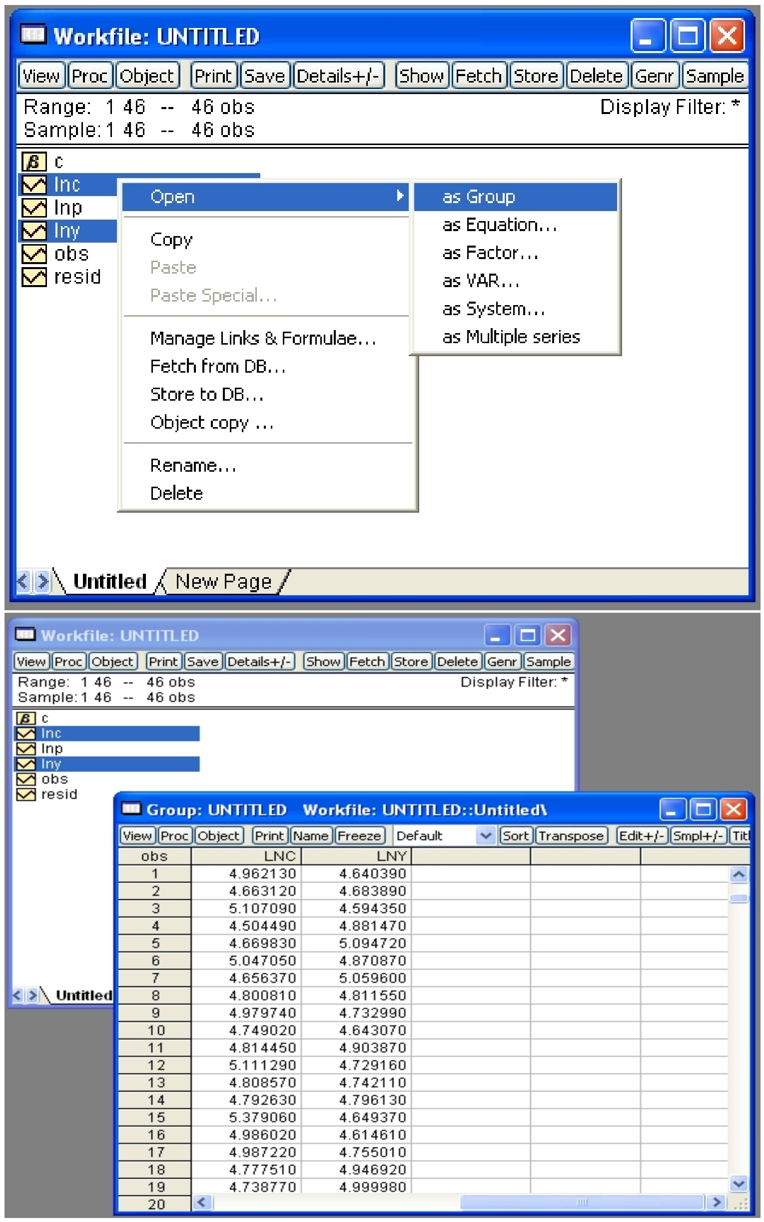


Figure 34: Select 'Open as Group' to create a group containing series lnc and lny.

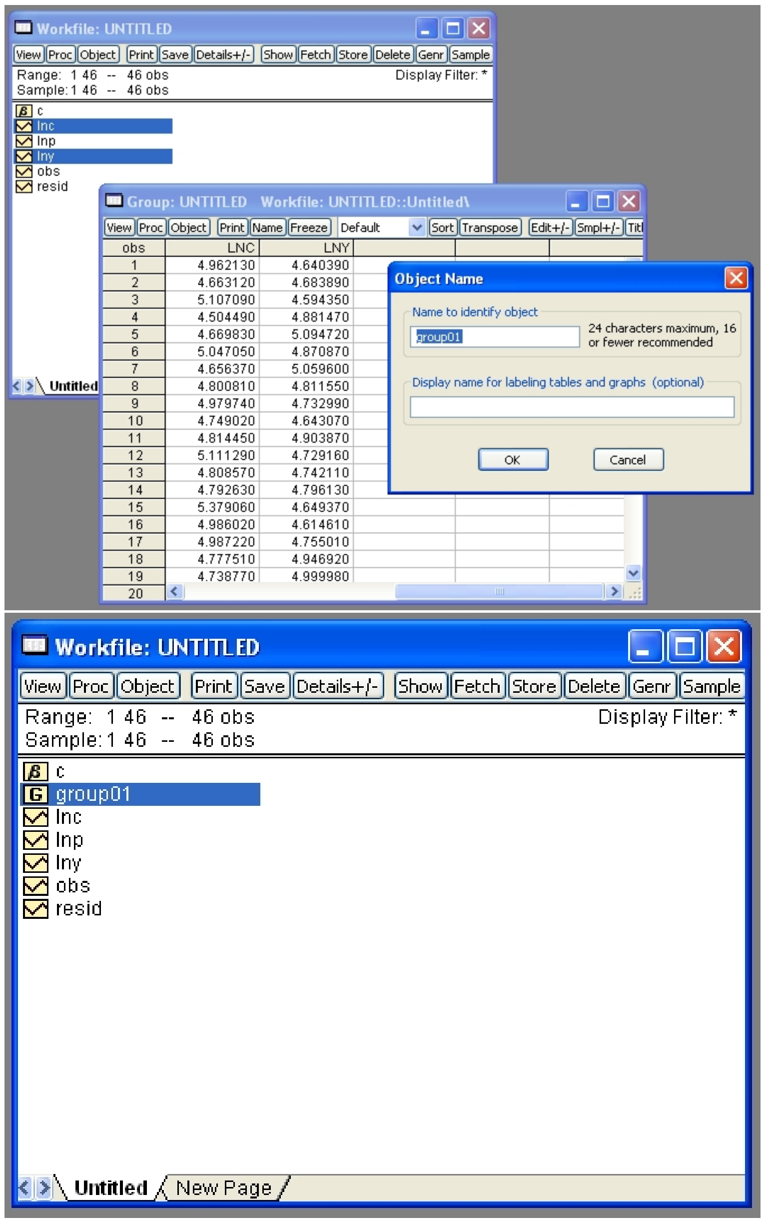


Figure 35: Name this group ‘group01’, and check that this appears in the workfile.

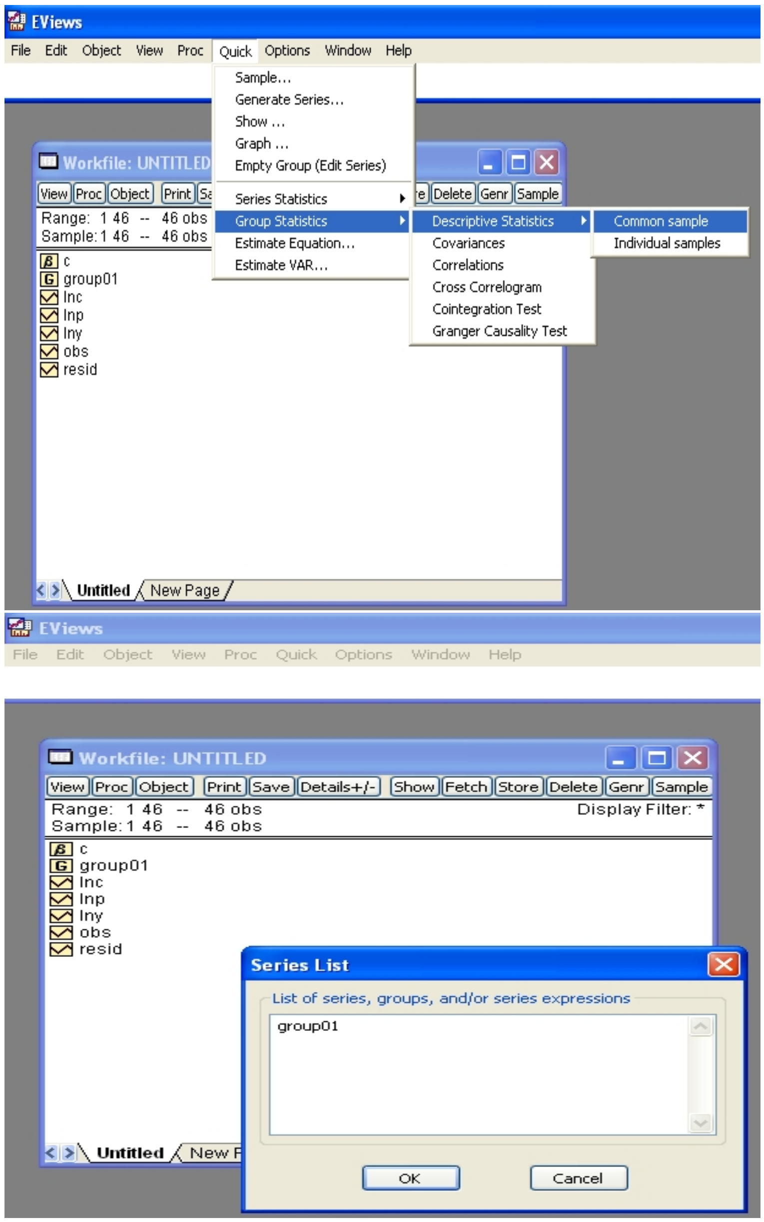
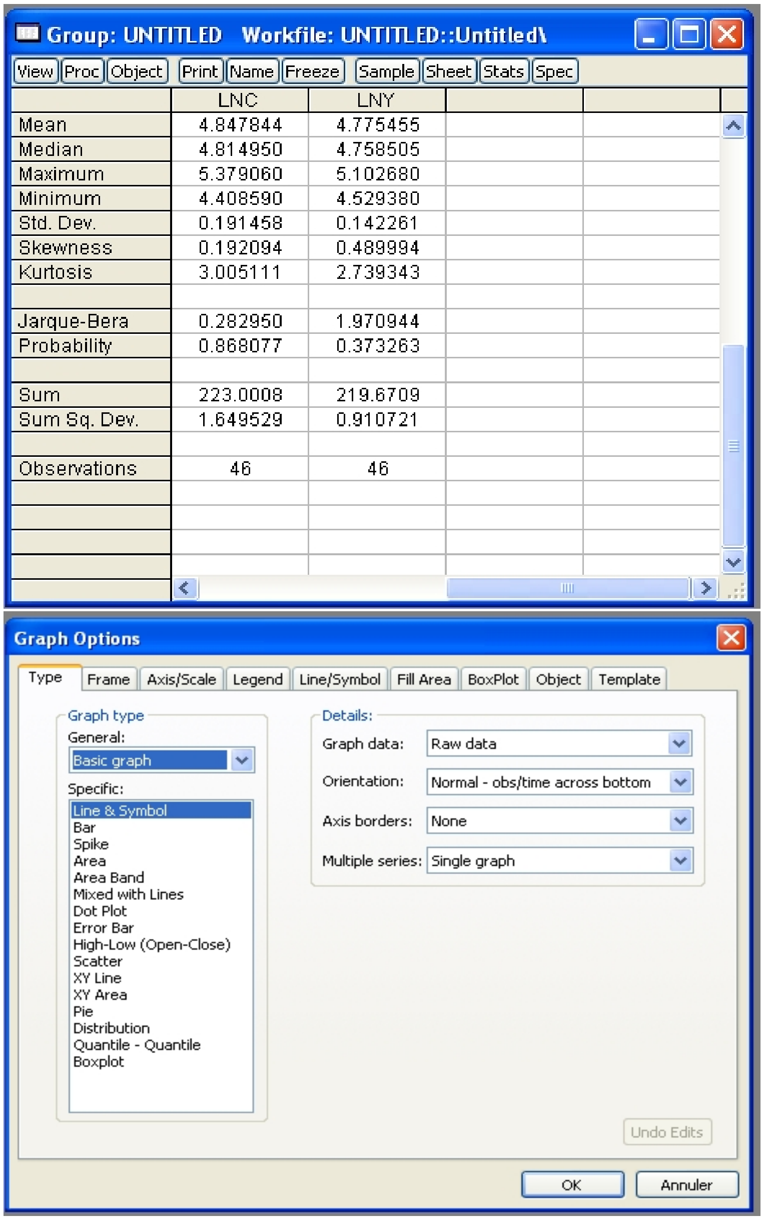


Figure 36: Select 'Common sample' group descriptive statistics and choose group01.



	LNC	LNY
Mean	4.847844	4.775455
Median	4.814950	4.758505
Maximum	5.379060	5.102680
Minimum	4.408590	4.529380
Std. Dev.	0.191458	0.142261
Skewness	0.192094	0.489994
Kurtosis	3.005111	2.739343
Jarque-Bera	0.282950	1.970944
Probability	0.868077	0.373263
Sum	223.0008	219.6709
Sum Sq. Dev.	1.649529	0.910721
Observations	46	46

Figure 37: Descriptive statistics for all series in group01. Select 'Line & Symbol' graph.

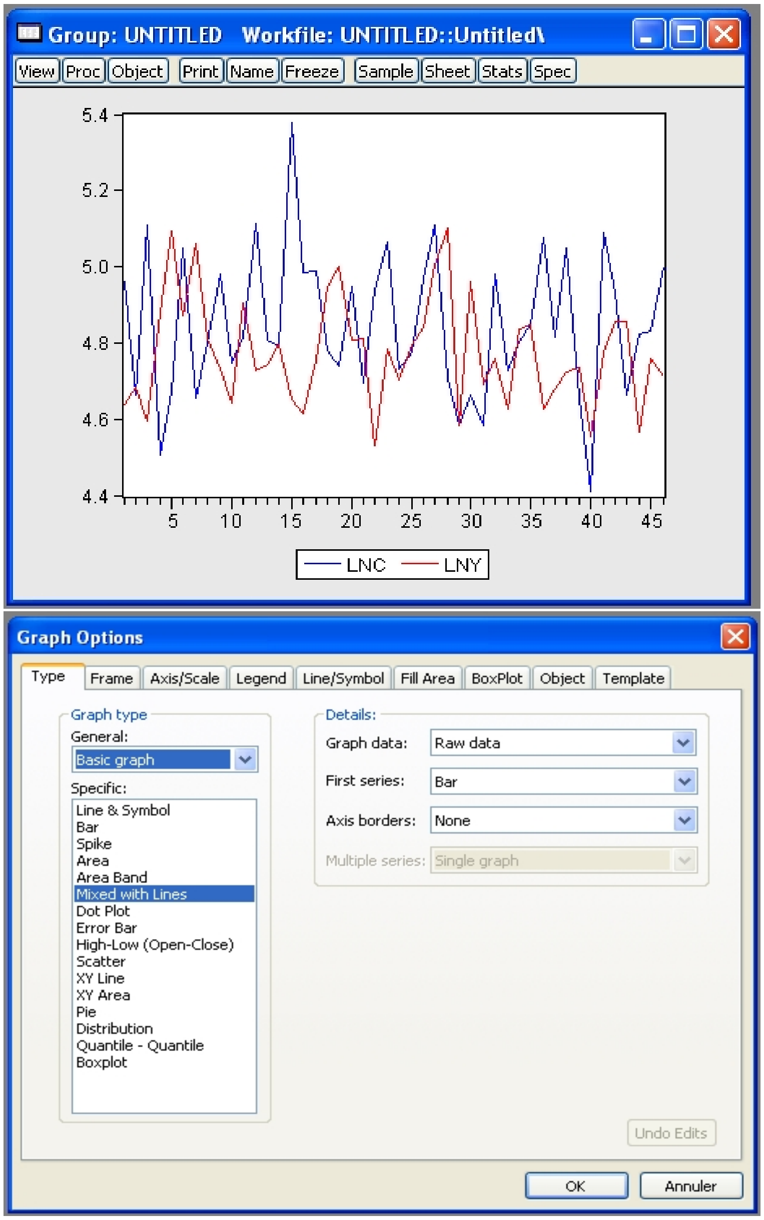


Figure 38: Plot of group01 series against observation. Select a 'Mixed with Lines' graph.

Figure 39: Mixed bar/line plot of lny and lnc. Select 'Scatter with Regression Line'.

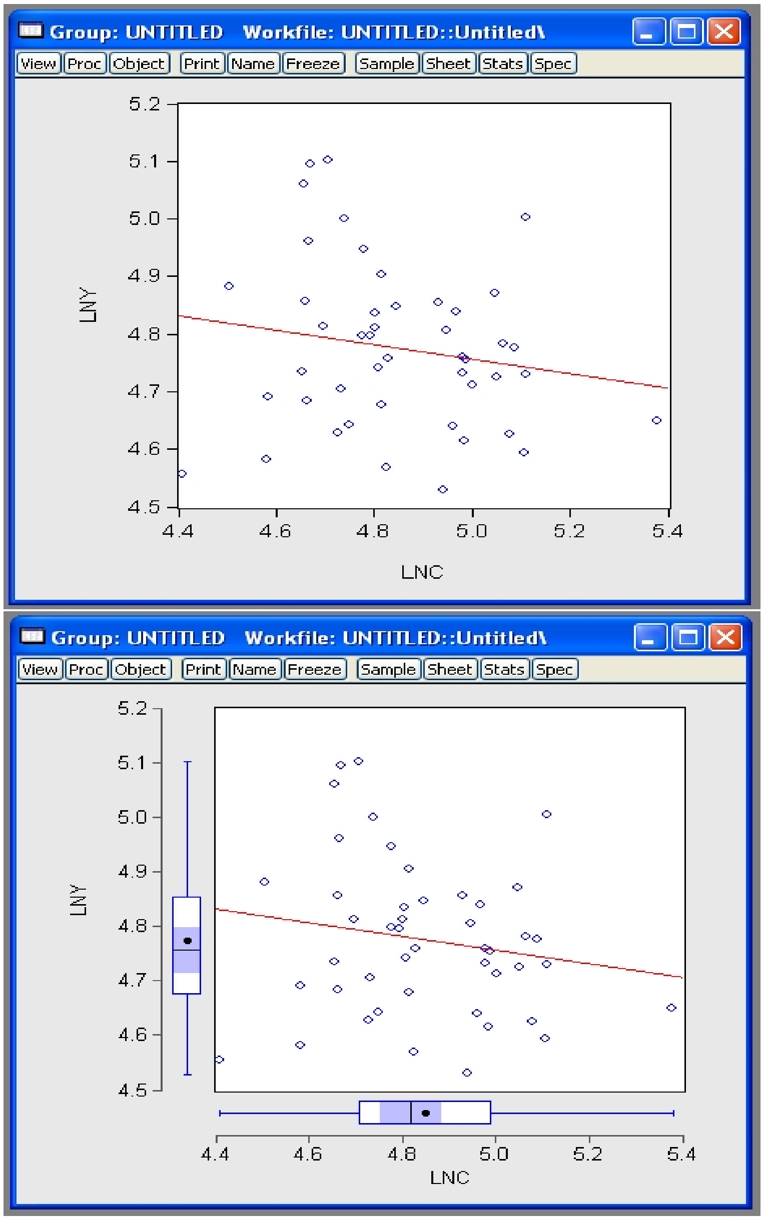


Figure 40: Scatter plot of lny against lnc, with ordinary least squares (OLS) fit from a regression of lny on a constant and lnc. Plot the same figure with boxplots of lny and lnc on the axes!

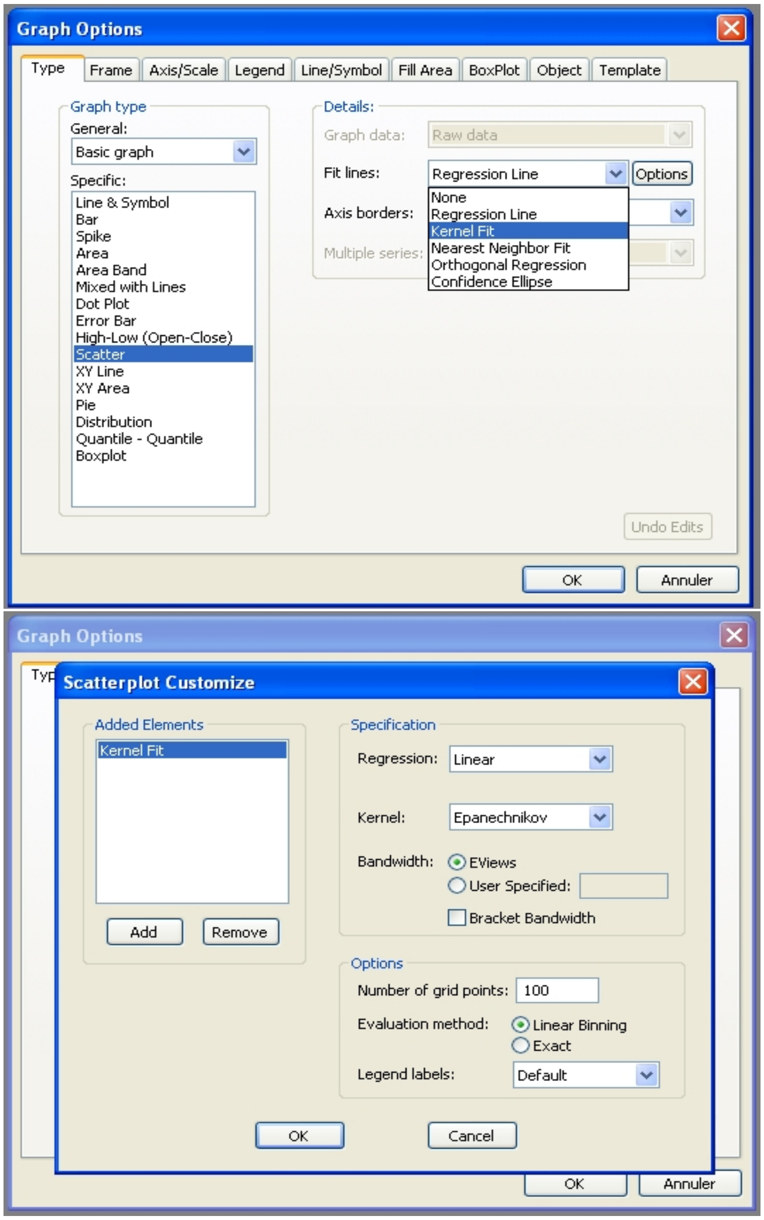


Figure 41: Select 'Scatter with Kernel Fit'.

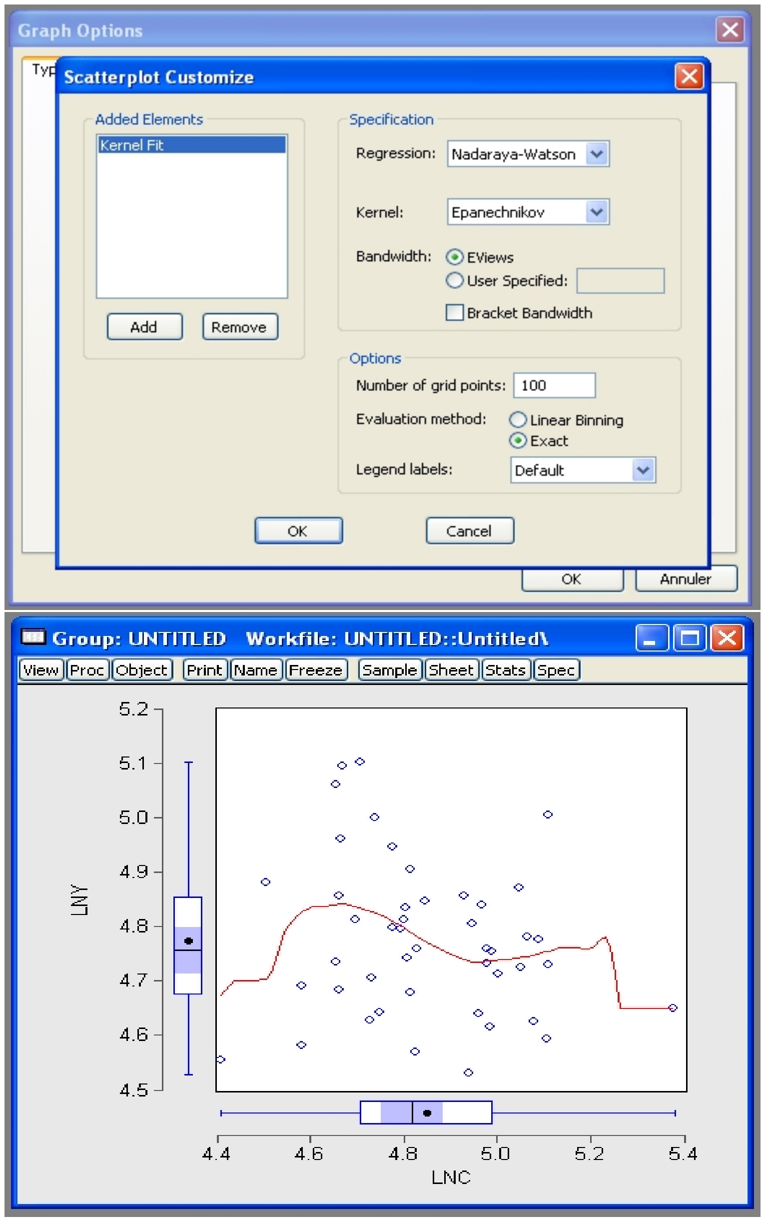


Figure 42: Choose 'Nadaraya-Watson' regression, with 'Epanechnikov' kernel, 'Exact' rather than 'Linear Binning', and bandwidth chosen by 'Eviews', to give the Nadaraya-Watson kernel regression fit, superimposed on the scatter plot of lny against lnc. The Nadaraya-Watson kernel estimator of Y_i on X_i is given by $\widehat{R}(x) = \arg\min_{\psi} \sum_i (Y_i - \psi)^2 K((x - X_i)/h)$, where ψ is a locally fit constant, and $K(u) = (3/4)(1 - u^2)$ on $[-1, 1]$ is the Epanechnikov kernel.

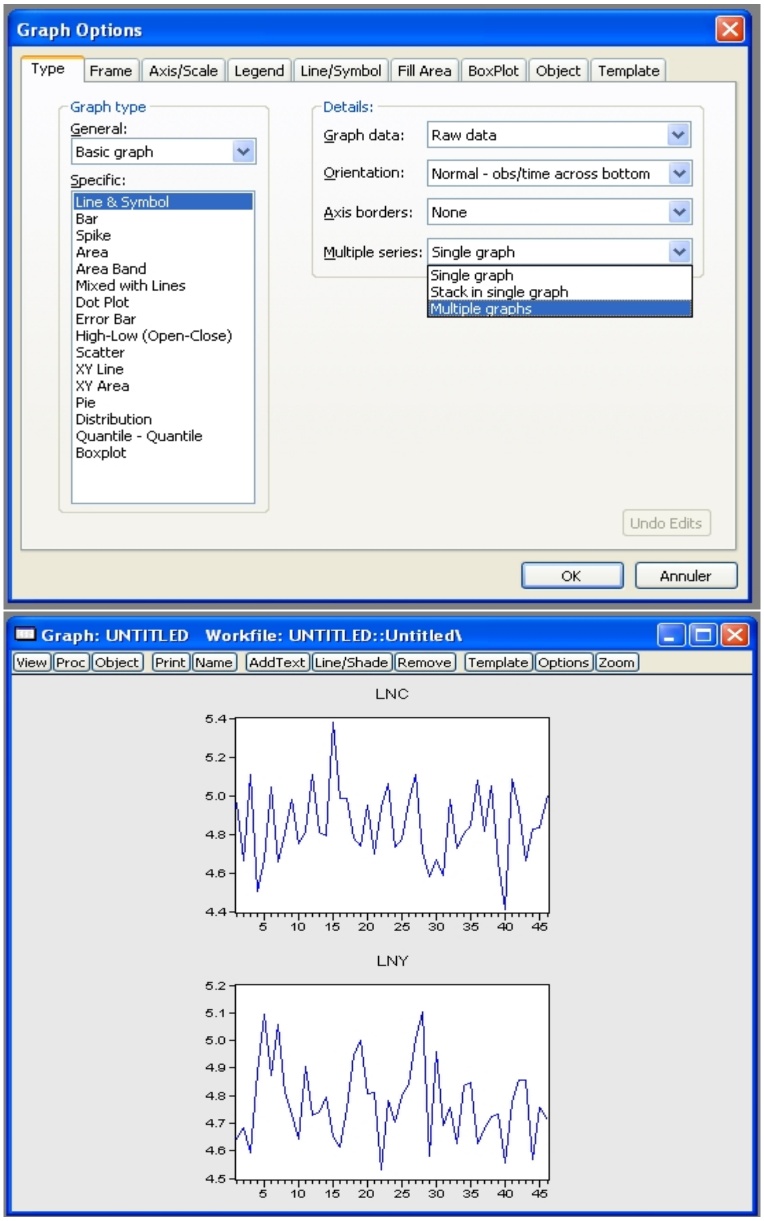


Figure 43: Choose 'Line & Symbol' and 'Multiple graphs' to give separate plots of lnc and lny against observation number.

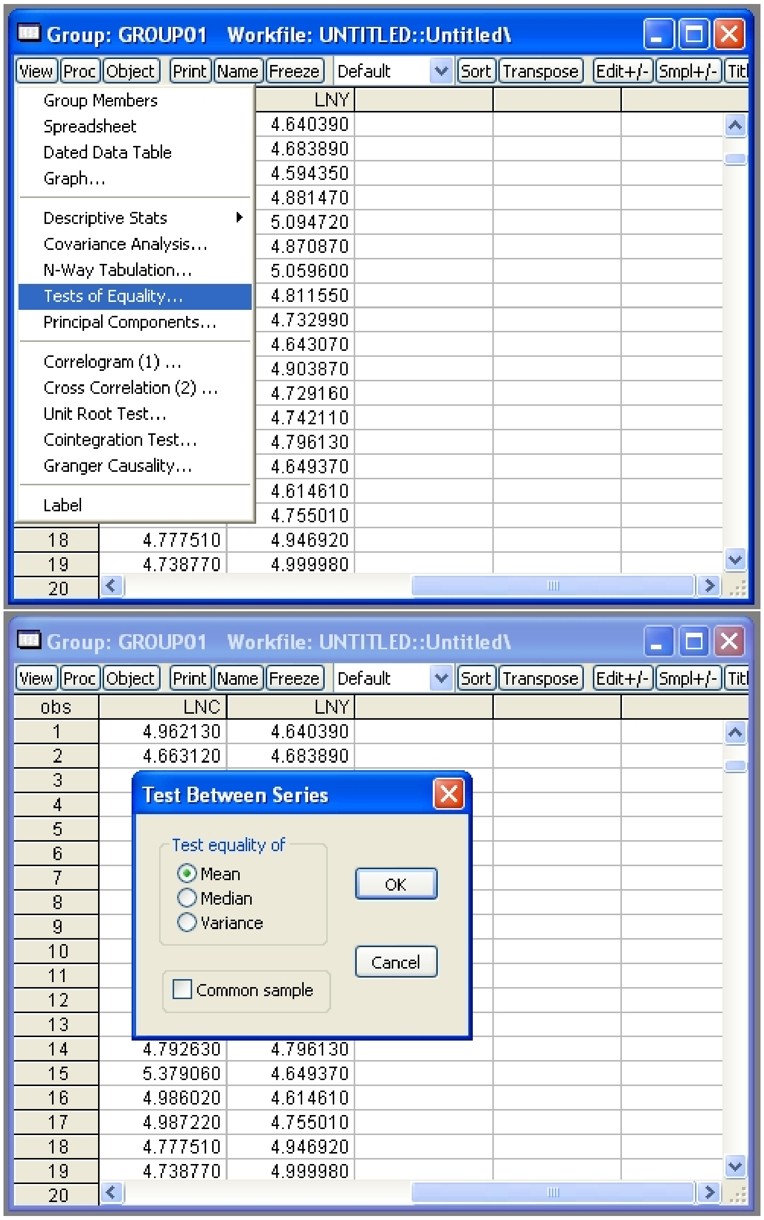


Figure 44: Choose 'Tests of Equality' on group01, and 'Mean', to test $H_0 : \bar{x}_{\text{lnc}} = \bar{x}_{\text{lny}}$.

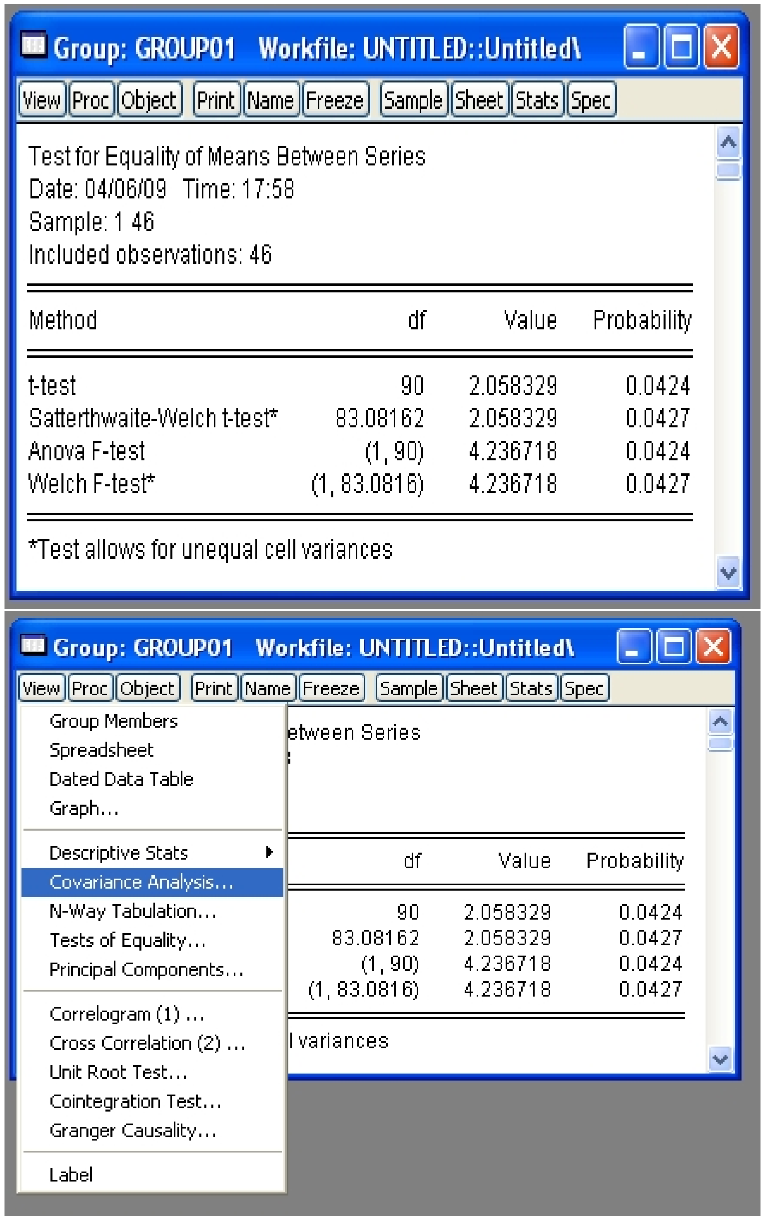


Figure 45: Results of the test $H_0 : \bar{x}_{\text{lnc}} = \bar{x}_{\text{lny}}$: without entering into the details of the tests, note that we reject the null at the 95% level of significance, against the two-sided alternative, using the t test, since 'Probability' is below 0.05. Select 'Covariance Analysis'.

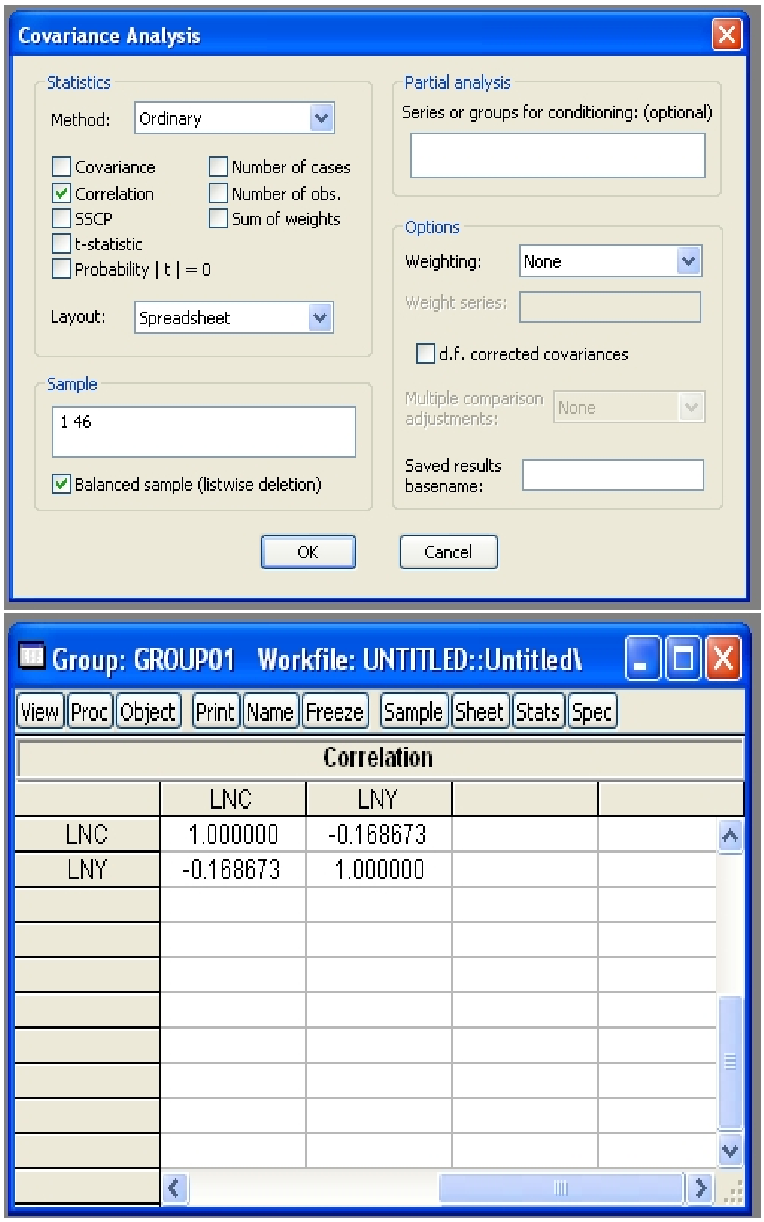


Figure 46: Select 'Correlation' and 'Spreadsheet', to give the sample correlation matrix.

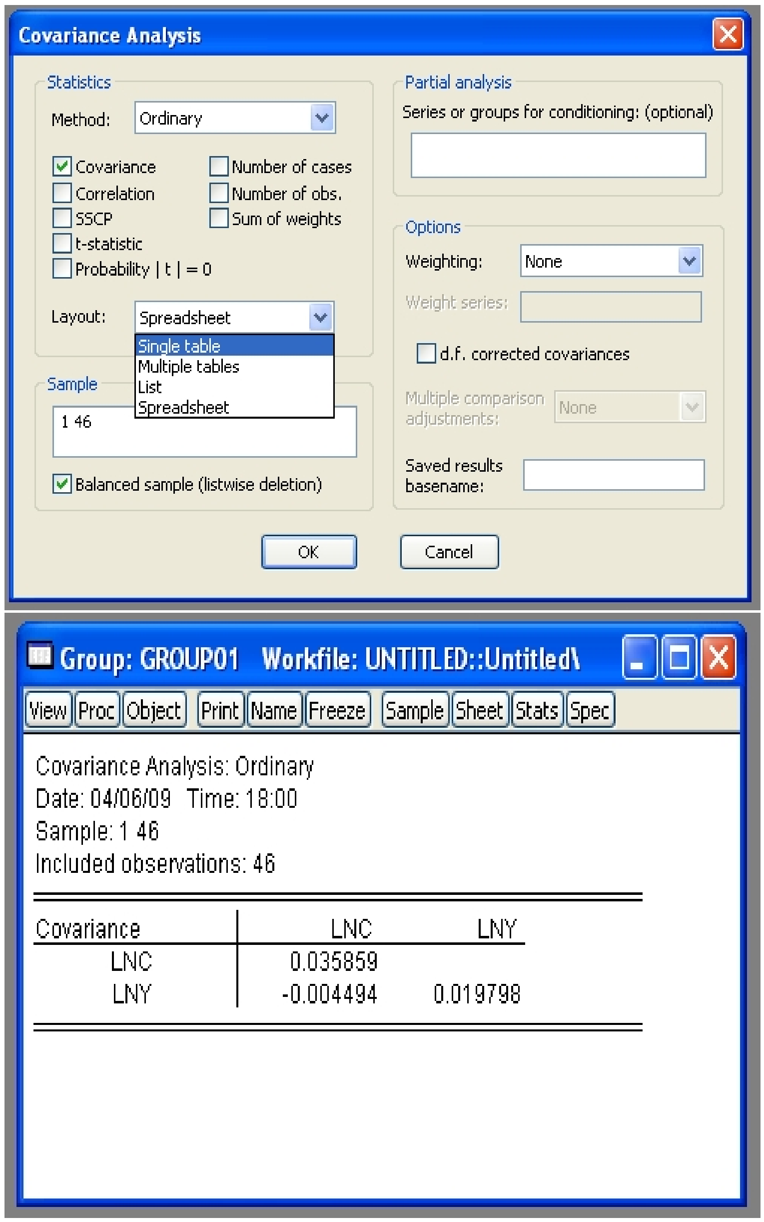
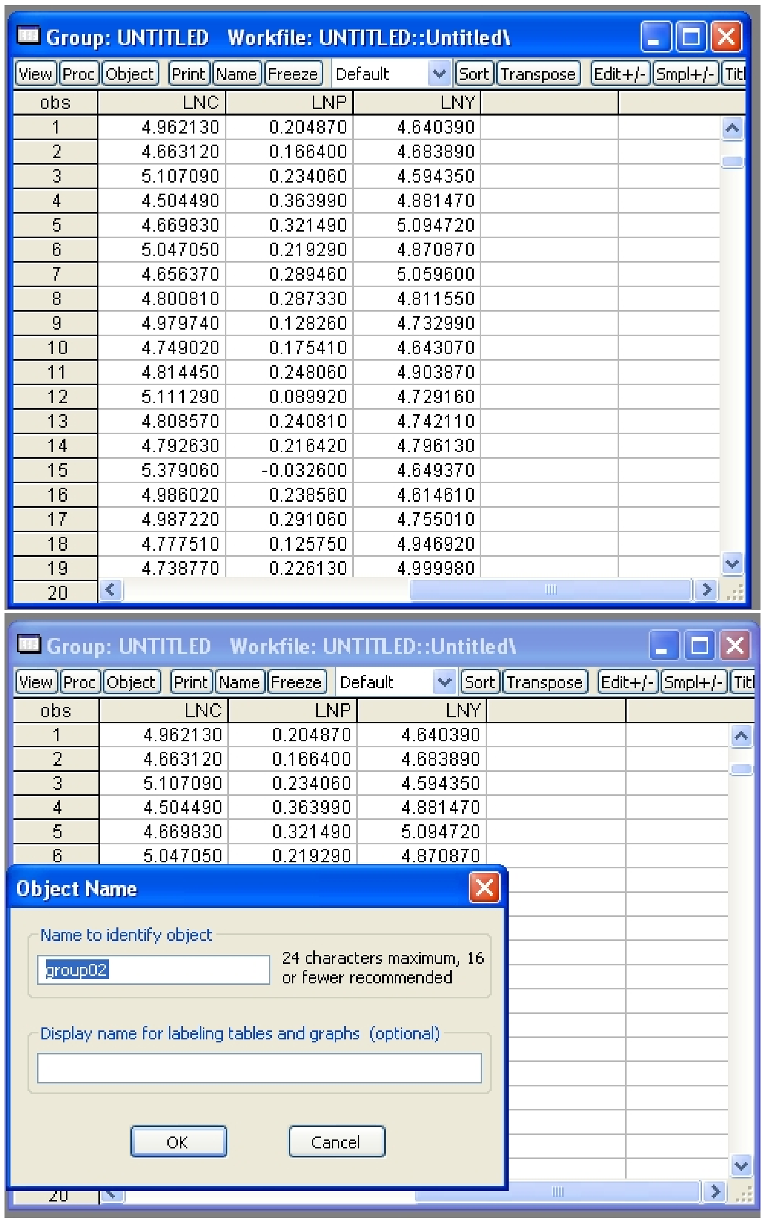


Figure 47: Select 'Covariance' and 'Single table', to give the sample covariances.



obs	LNC	LNP	LNY
1	4.962130	0.204870	4.640390
2	4.663120	0.166400	4.683890
3	5.107090	0.234060	4.594350
4	4.504490	0.363990	4.881470
5	4.669830	0.321490	5.094720
6	5.047050	0.219290	4.870870
7	4.656370	0.289460	5.059600
8	4.800810	0.287330	4.811550
9	4.979740	0.128260	4.732990
10	4.749020	0.175410	4.643070
11	4.814450	0.248060	4.903870
12	5.111290	0.089920	4.729160
13	4.808570	0.240810	4.742110
14	4.792630	0.216420	4.796130
15	5.379060	-0.032600	4.649370
16	4.986020	0.238560	4.614610
17	4.987220	0.291060	4.755010
18	4.777510	0.125750	4.946920
19	4.738770	0.226130	4.999980
20			

Figure 48: Open lnc, lnp and lny as 'group02'.

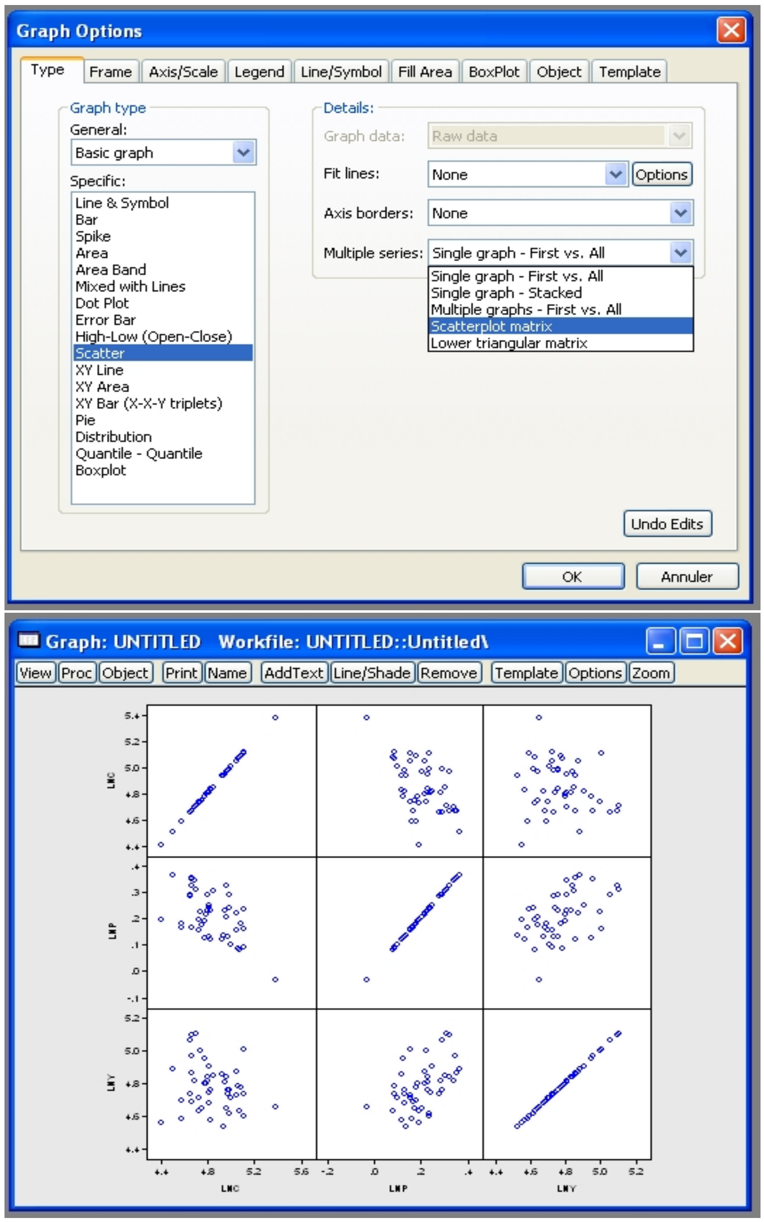


Figure 49: Plot a matrix scatterplot of lnc, lnp and lny, and name the graph graph01.

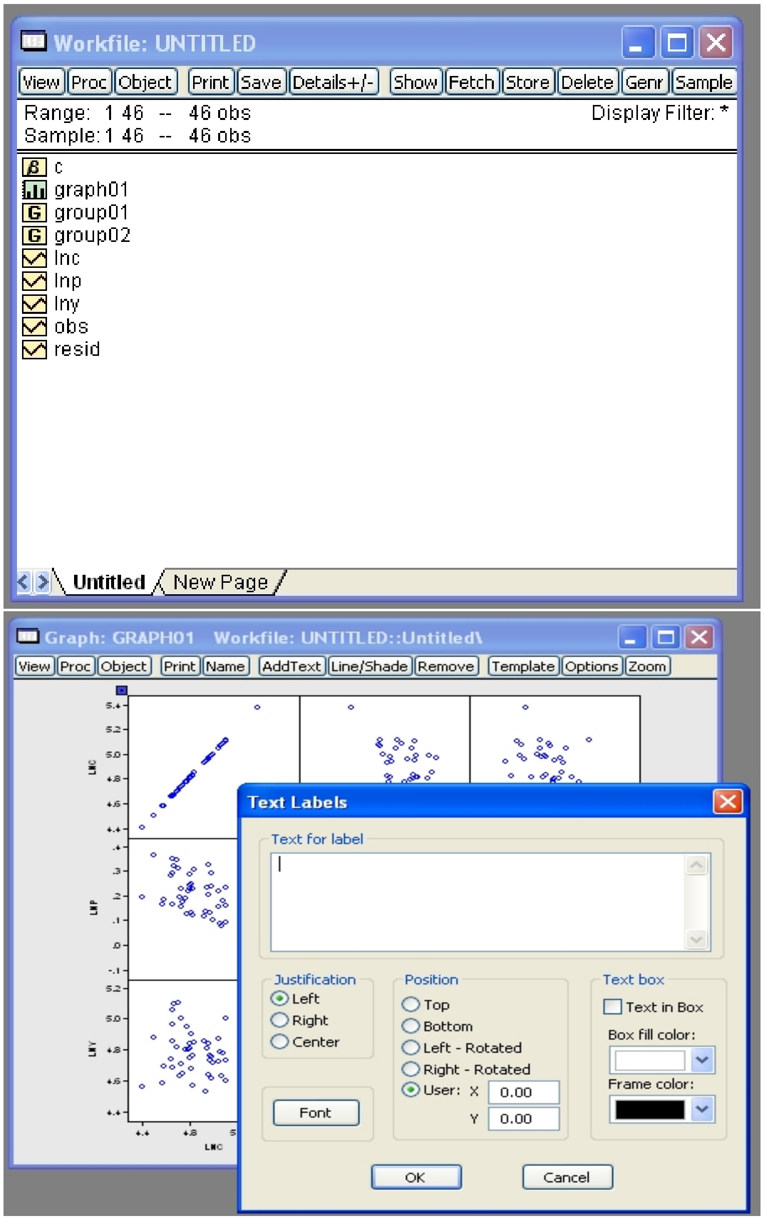


Figure 50: Check that graph01 appears in the workfile. Click the AddText button.

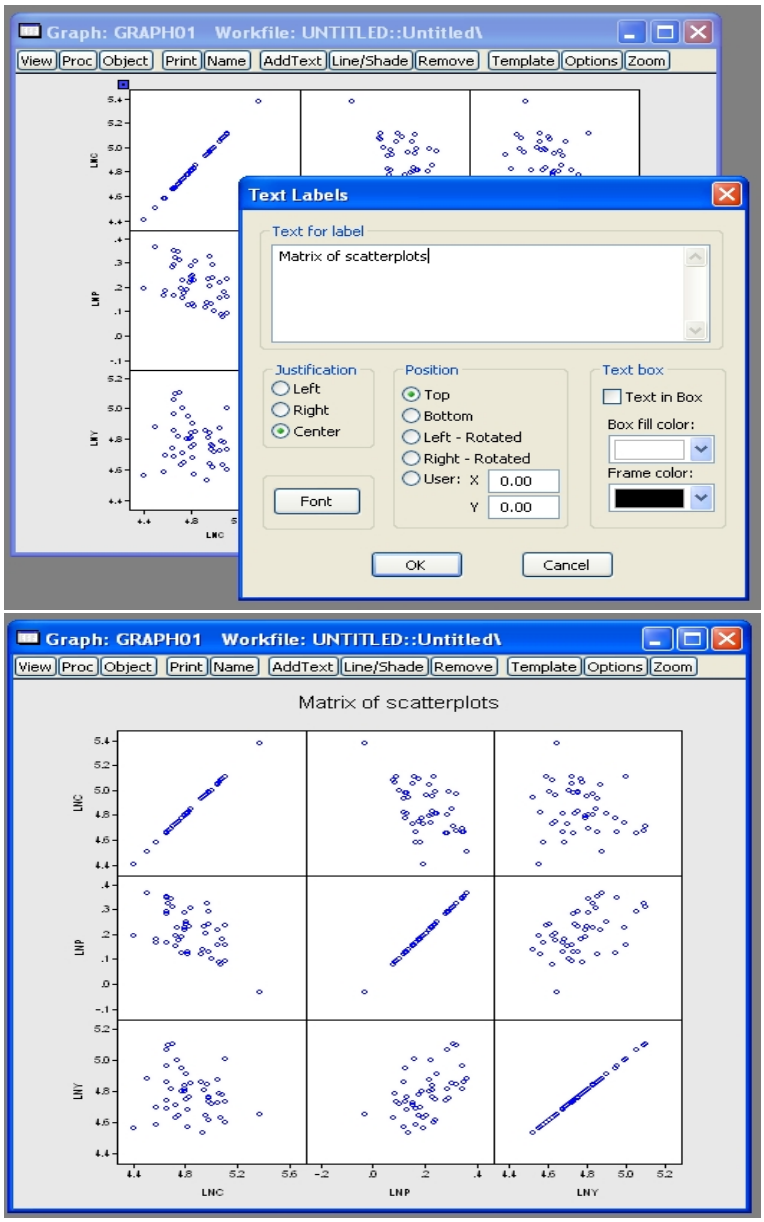


Figure 51: Label the graph as shown, with text font size 20.

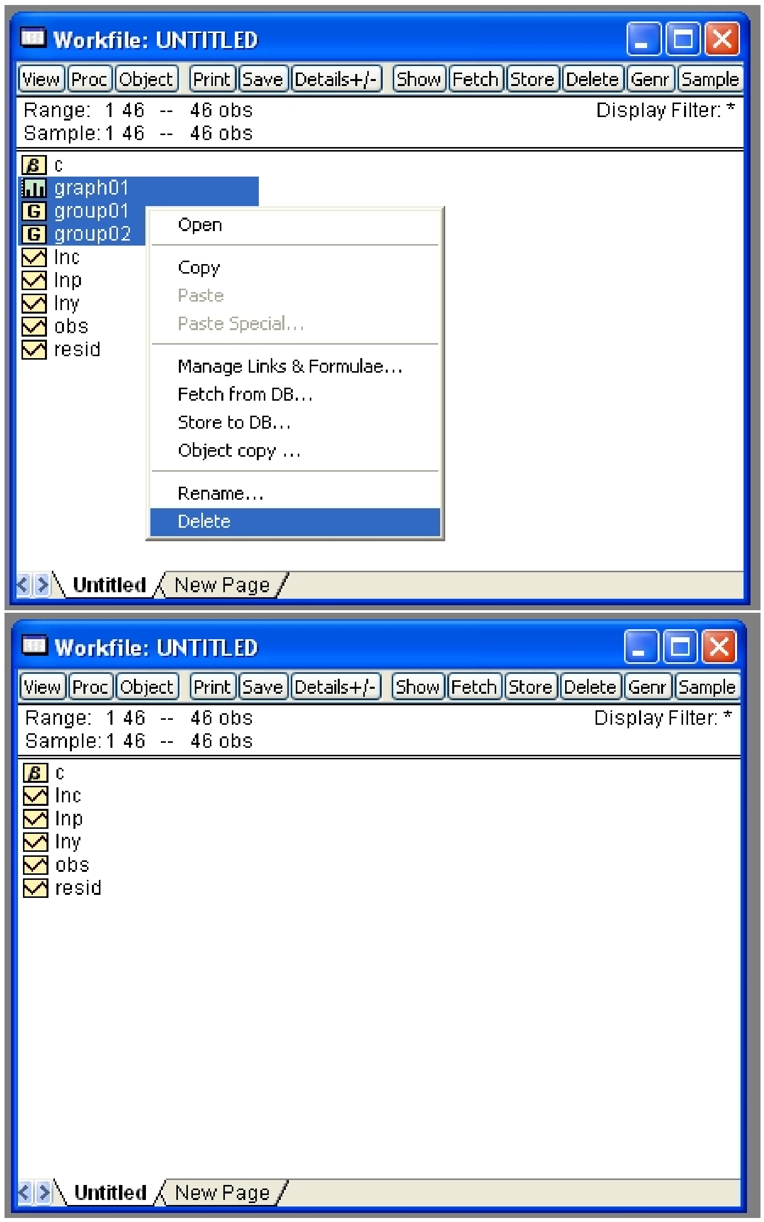


Figure 52: Select graph01 and both group01 and group02, and delete from the workfile.

A First Regression

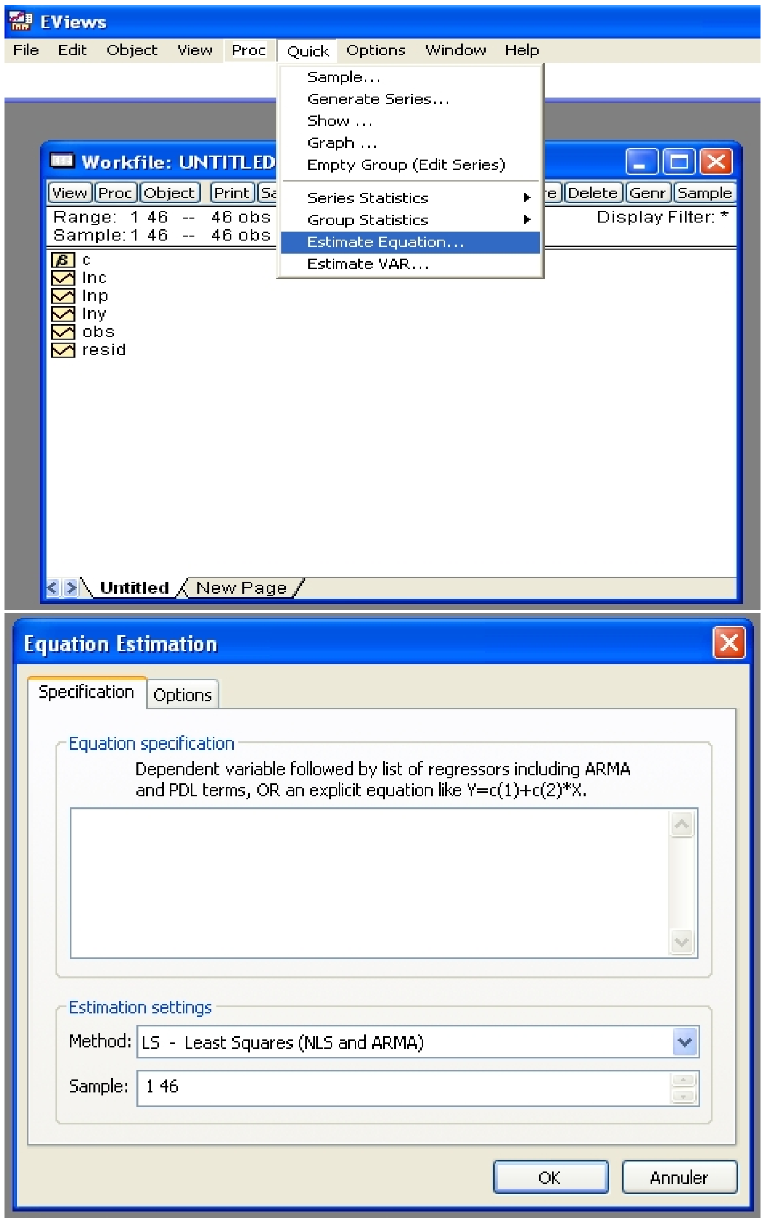


Figure 53: Select 'Estimate Equation...'.

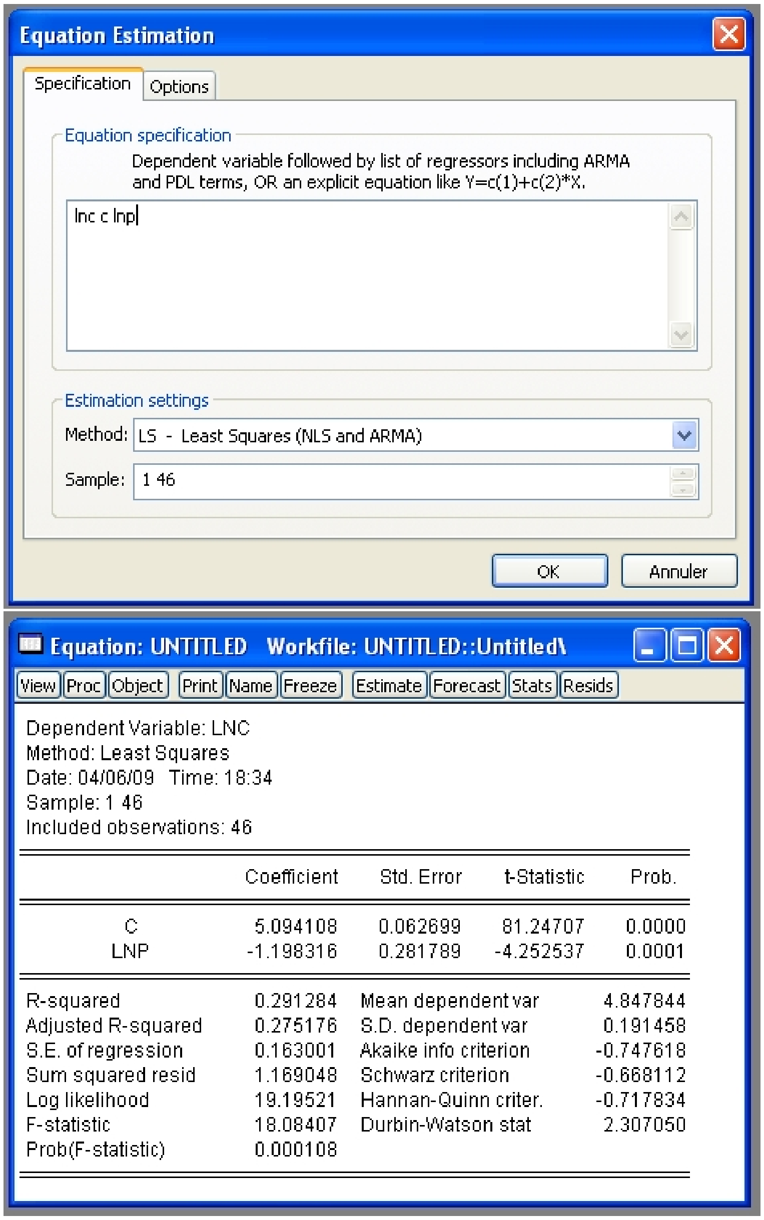


Figure 54: **(eq01)** Enter 'lnc c lnp' for a regression of log consumption on a constant and log price. Observe that the regression output gives estimated coefficients (by ordinary least squares), standard errors, probabilities for individual tests of significance, and various statistics, including coefficients of determination, the sum of squared residuals, the Durbin-Watson statistic for autocorrelation, the Akaike and Schwarz Information Criteria, and the F statistic for significance of the entire regression (*we will return to these statistics later in the course*).

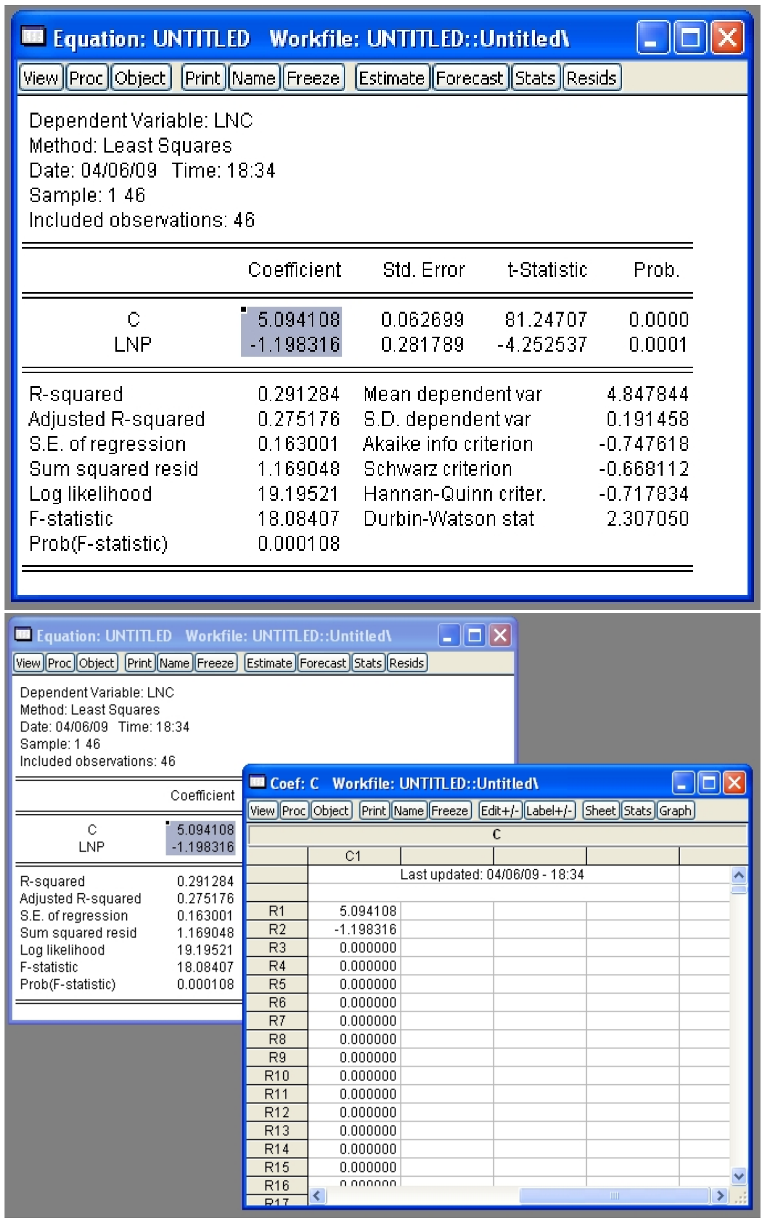


Figure 55: Check that the estimated coefficients have been saved in the object c.

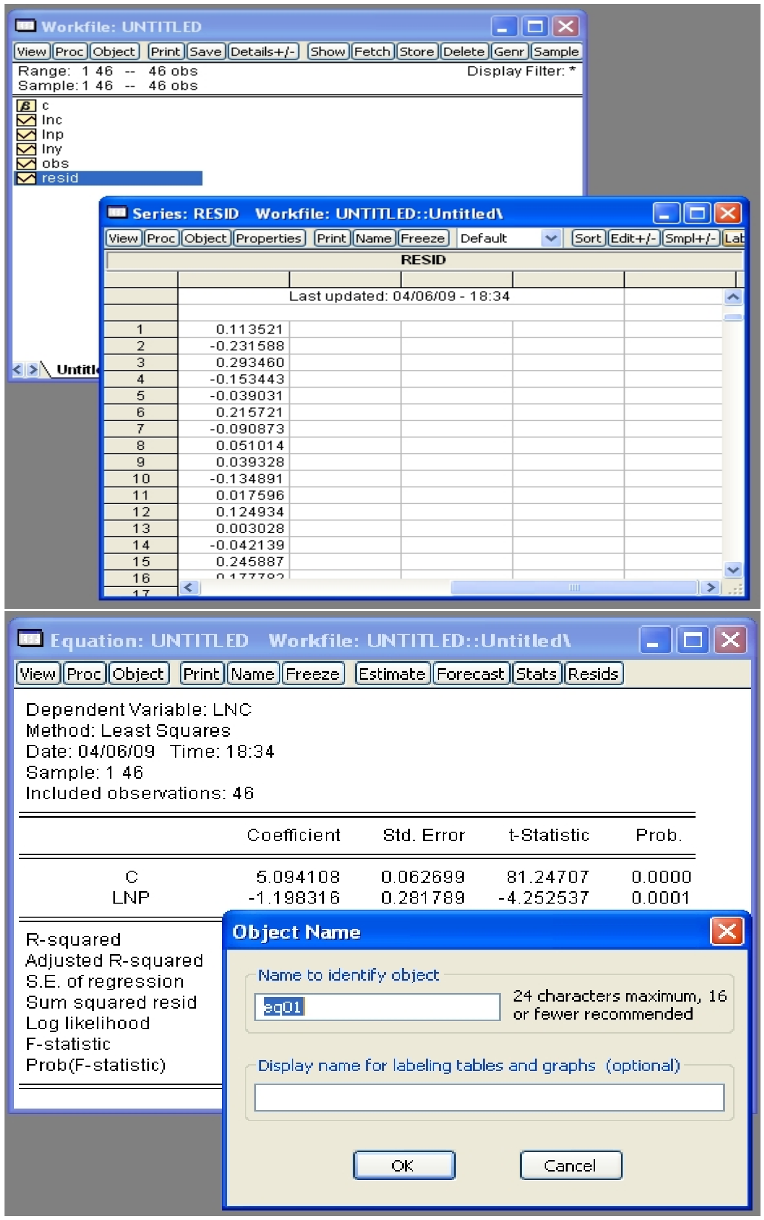


Figure 56: Check that the estimated residuals have been recorded in the workfile object resid: **note that the estimated residuals are overwritten each time that a new regression is performed**. Name the equation ‘eq01’.

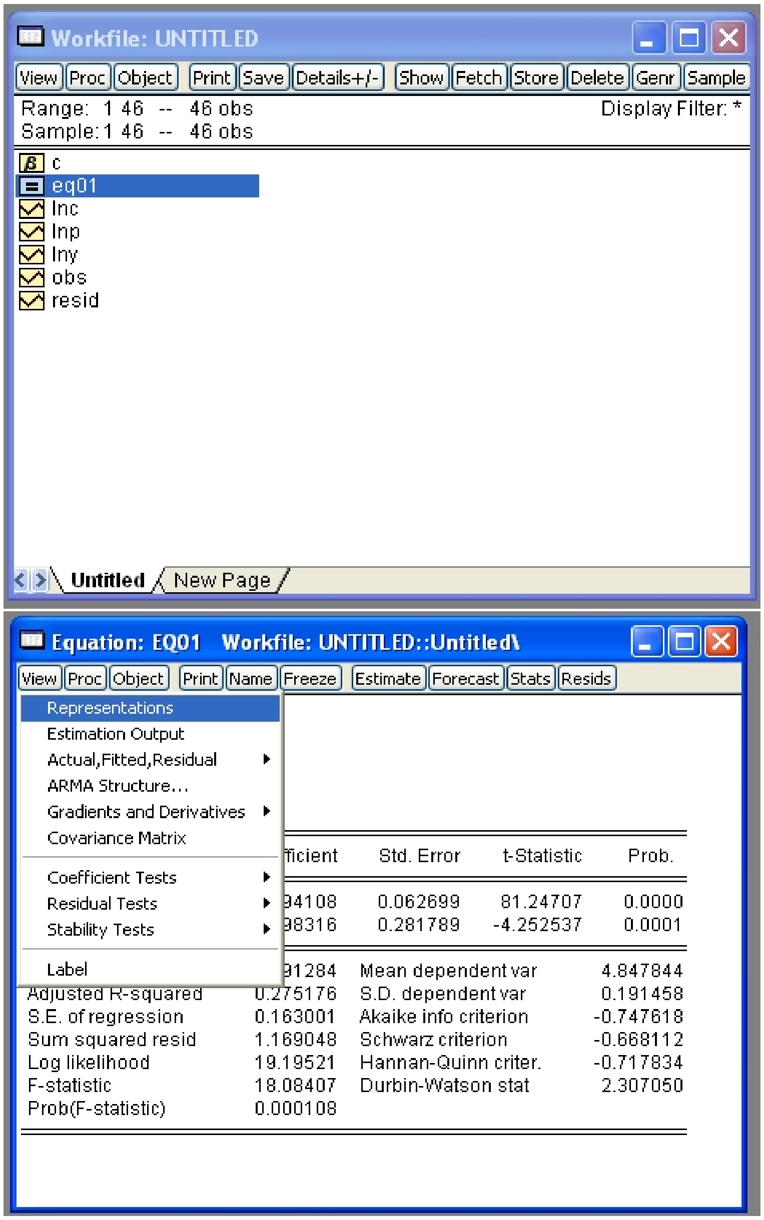


Figure 57: Select ‘eq01’ and ‘Representations’.

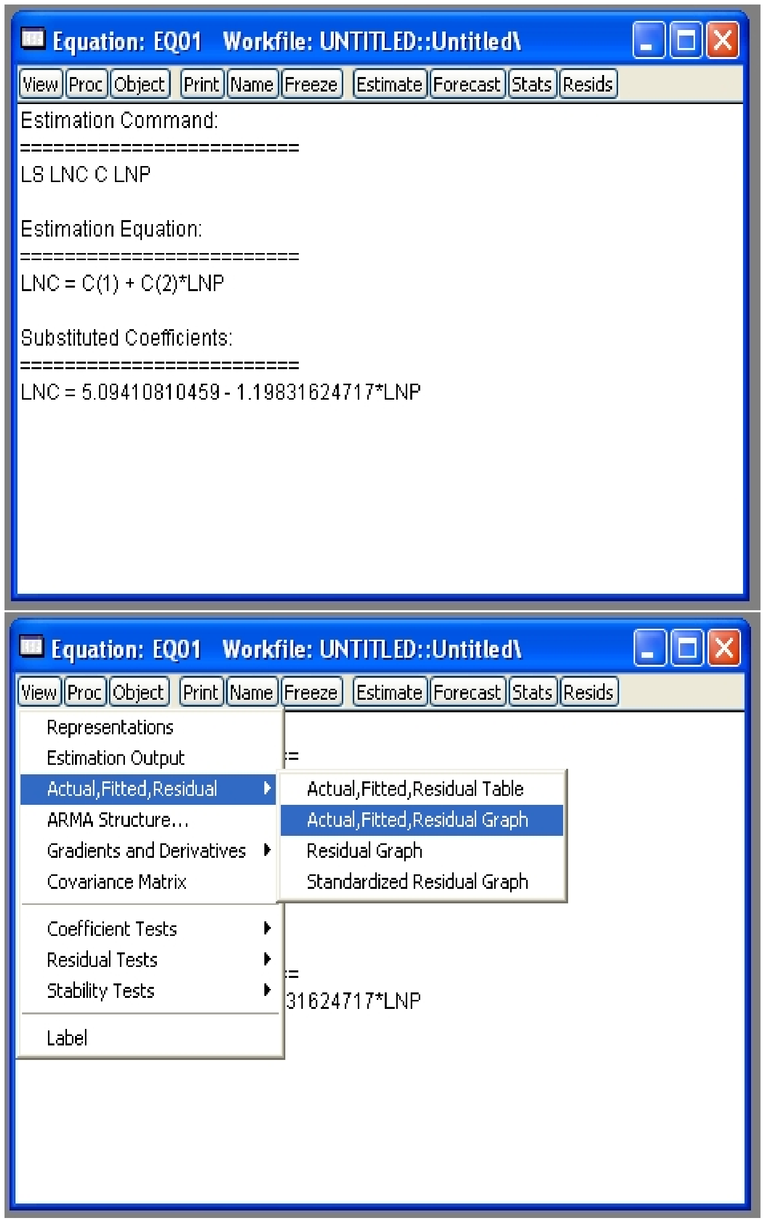


Figure 58: Alternative representations of estimated equation eq01. Select 'Actual, Fitted, Residual Graph', and 'Residual Graph', and check the output.

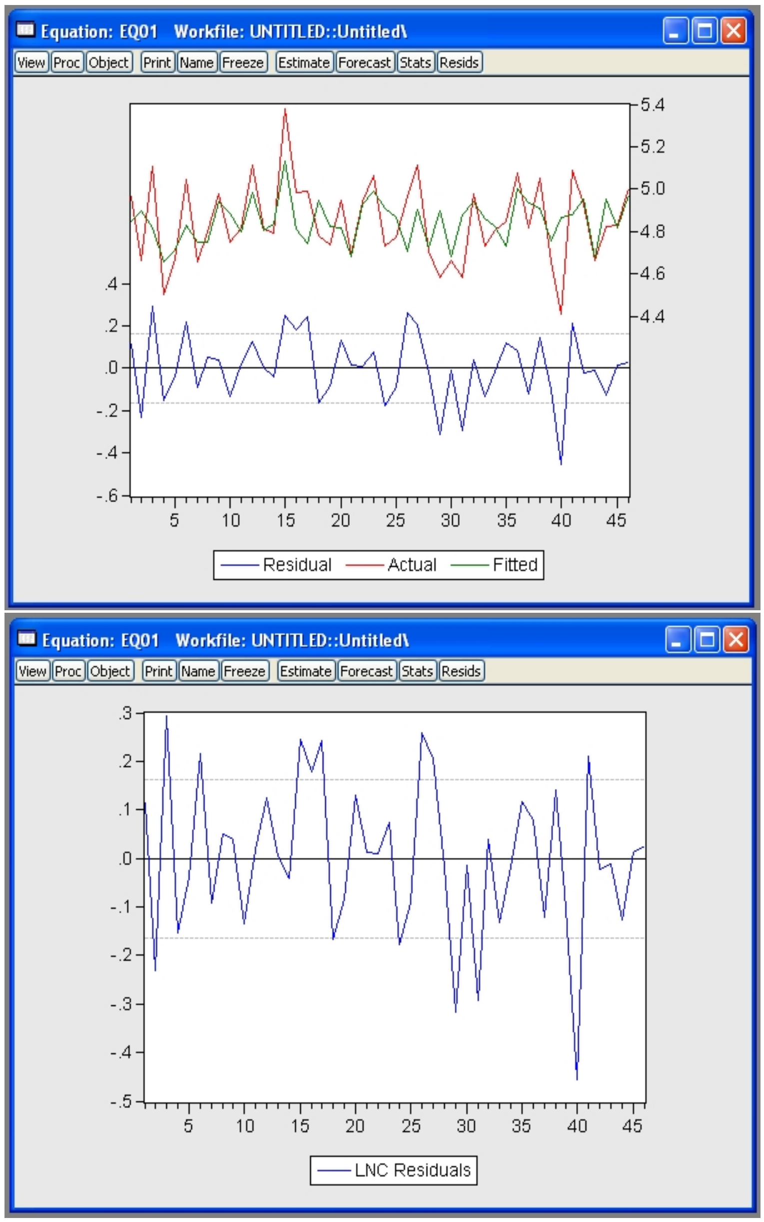


Figure 59: Several representations of the fitted residuals $\widehat{u}$.

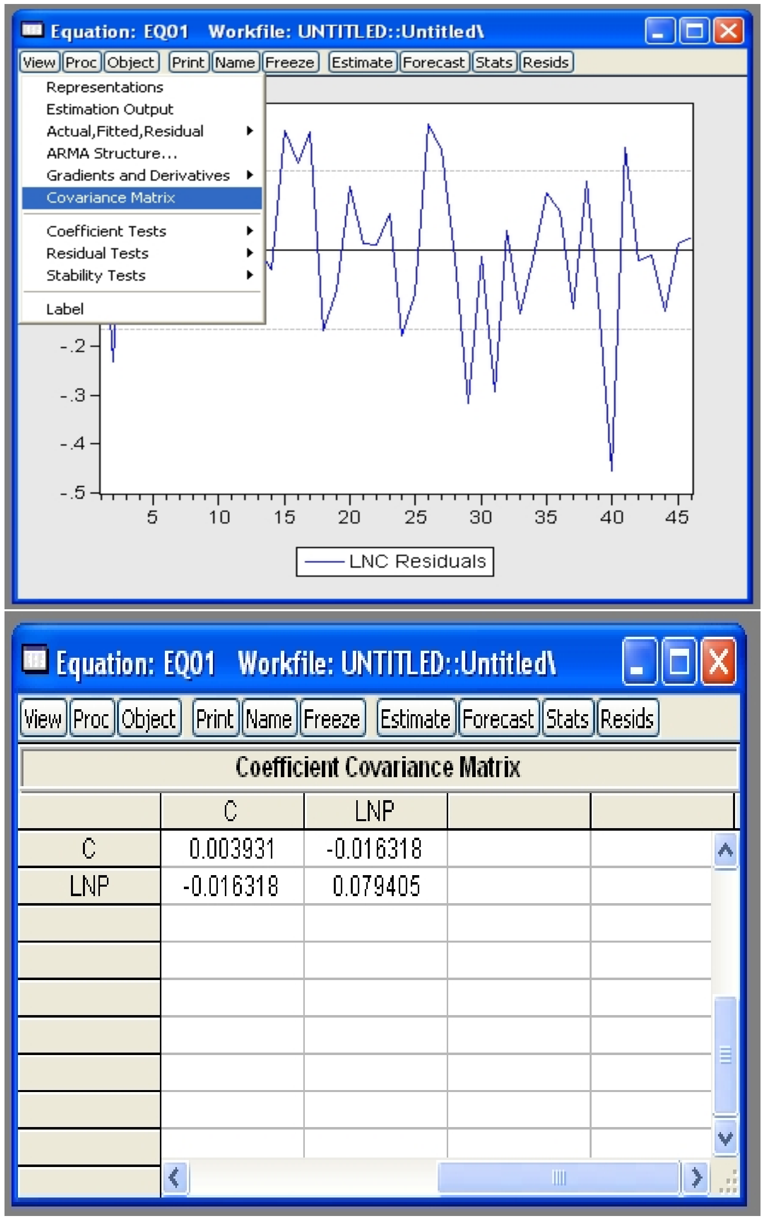


Figure 60: Select 'Covariance Matrix' and check the output. This gives $\widehat{\sigma}^2(X'X)^{-1}$.

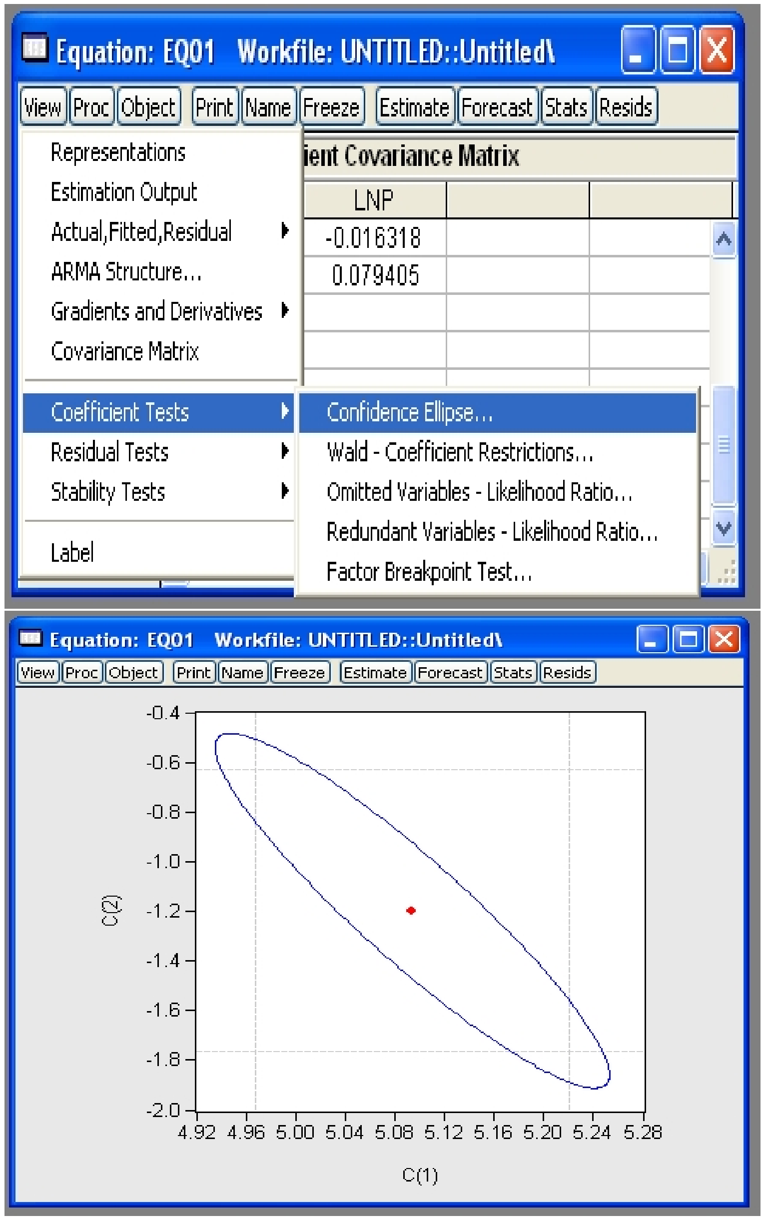


Figure 61: Select 'Confidence Ellipse…', and check the output (*this will be discussed in class when we cover joint hypothesis tests*).

A Hypothesis Test on the Estimated Coefficients

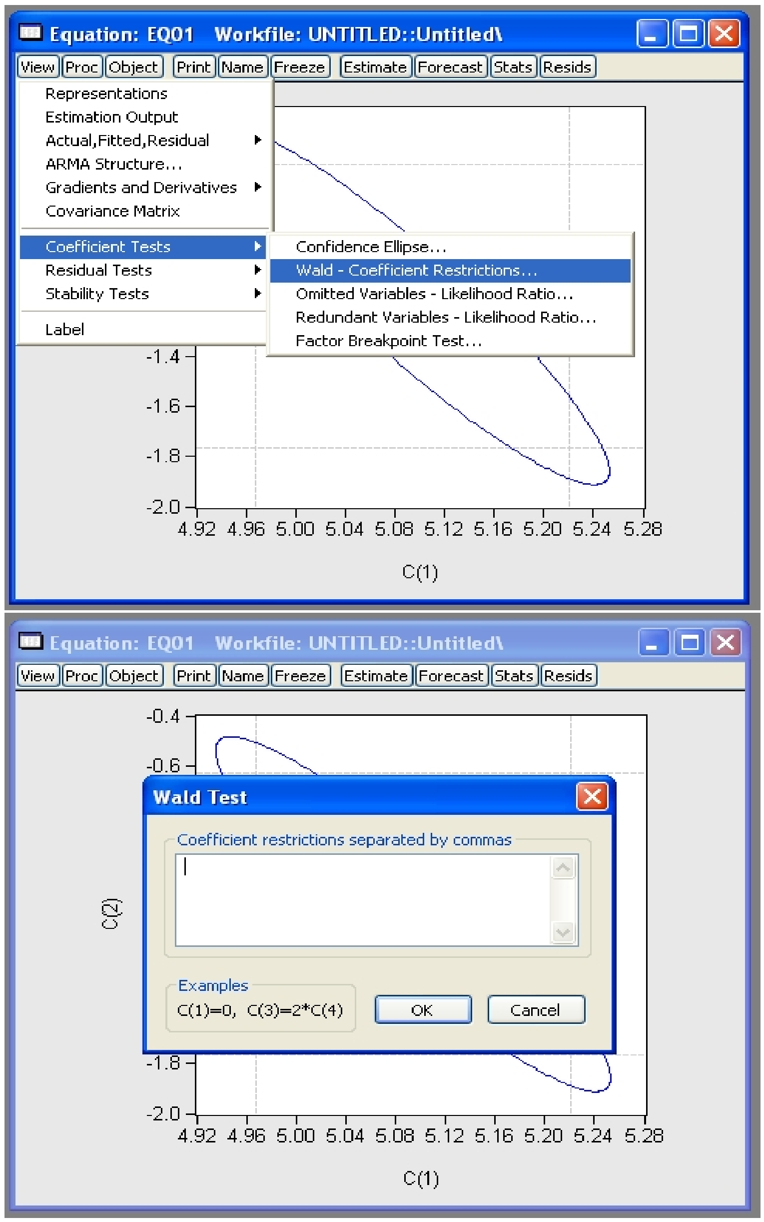


Figure 62: Select ‘Wald - Coefficient Restrictions’.

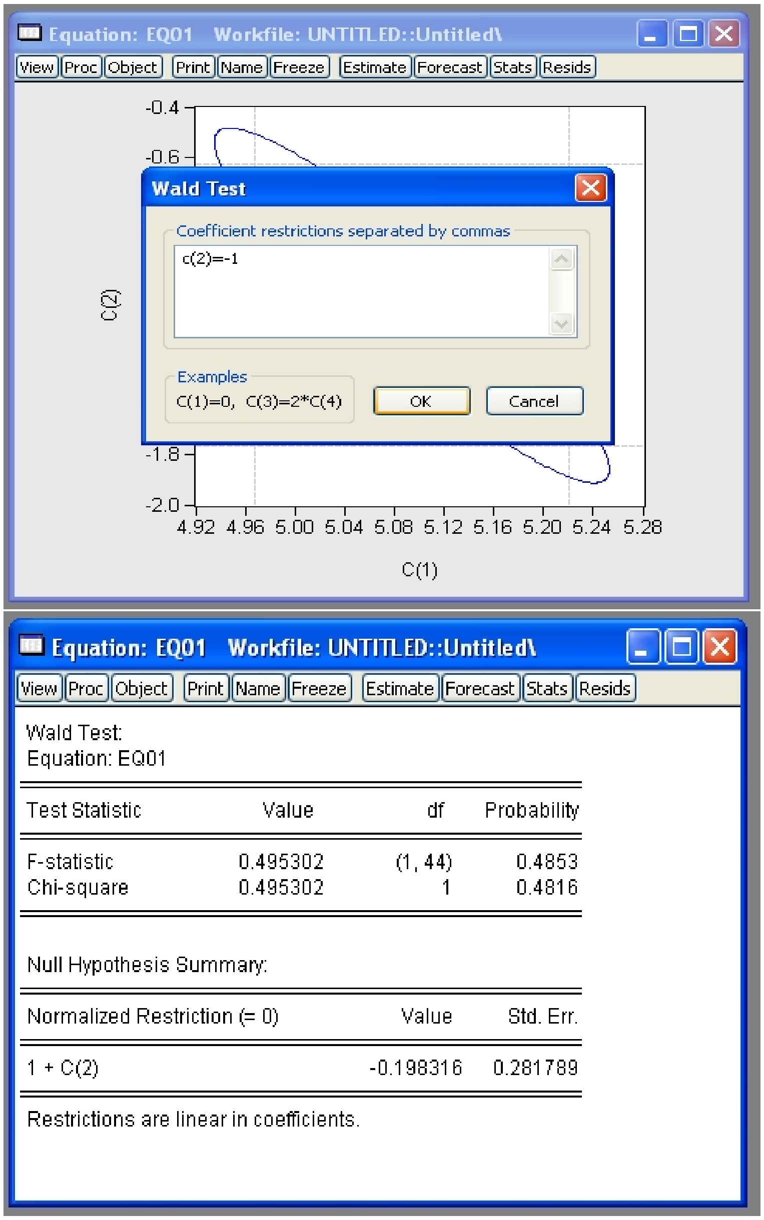


Figure 63: Enter ‘c(2)=-1’ to perform a Wald test of $H_0 : \beta_1 = -1$. Observe that both exact (F) and asymptotic (χ^2) statistics are reported: the result of the test is that we do not reject the null at the 90% level (say). **Why is an F test reported, not a t test?**

Asking for Help!

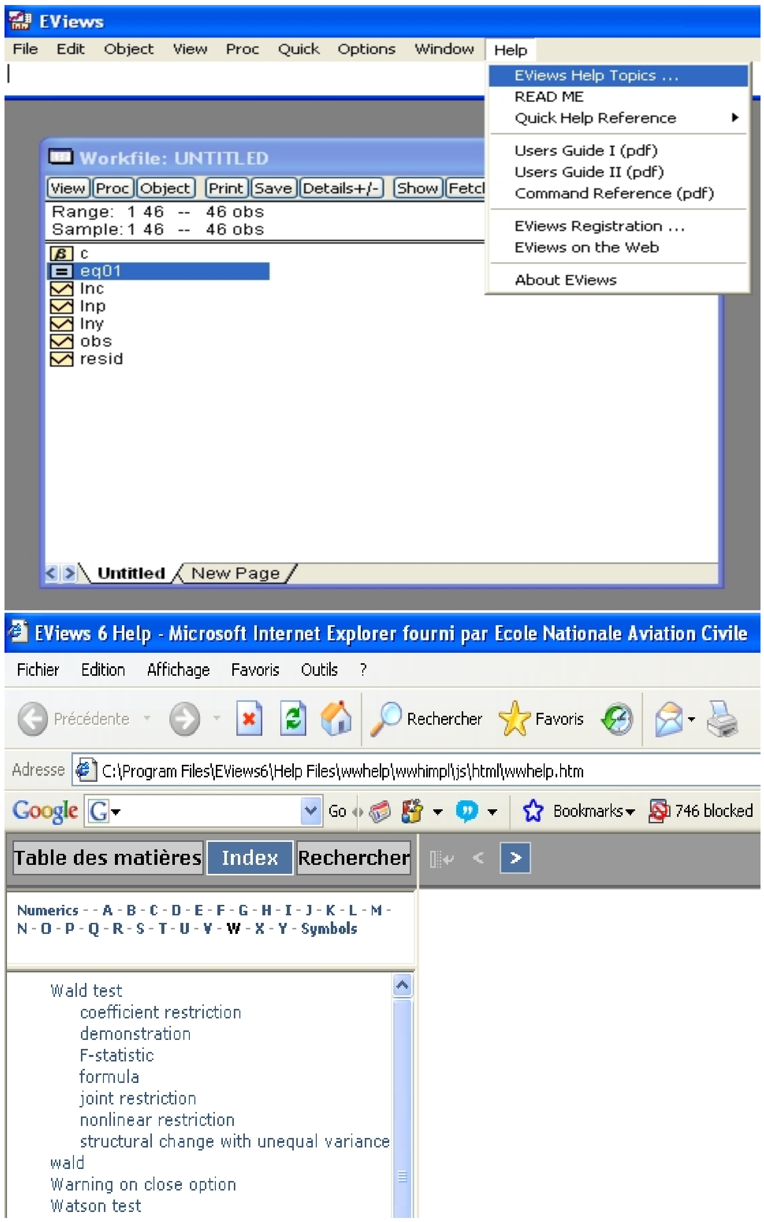
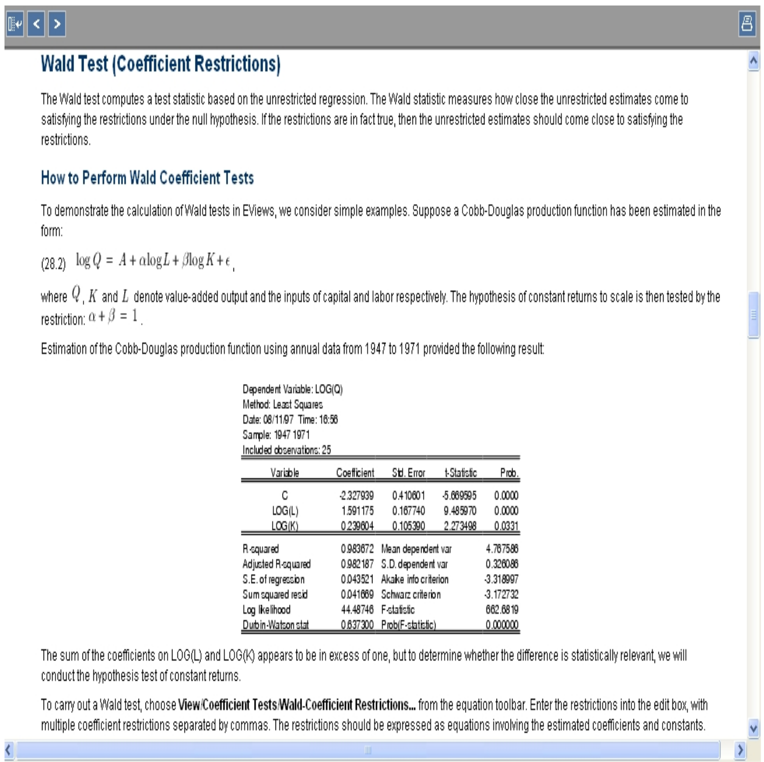


Figure 64: Select 'Help'.



Wald Test (Coefficient Restrictions)

The Wald test computes a test statistic based on the unrestricted regression. The Wald statistic measures how close the unrestricted estimates come to satisfying the restrictions under the null hypothesis. If the restrictions are in fact true, then the unrestricted estimates should come close to satisfying the restrictions.

How to Perform Wald Coefficient Tests

To demonstrate the calculation of Wald tests in EViews, we consider simple examples. Suppose a Cobb-Douglas production function has been estimated in the form:

(28.2) $\log Q = A + \alpha \log L + \beta \log K + \epsilon$,

where Q, K and L denote value-added output and the inputs of capital and labor respectively. The hypothesis of constant returns to scale is then tested by the restriction: $\alpha + \beta = 1$.

Estimation of the Cobb-Douglas production function using annual data from 1947 to 1971 provided the following result:

Dependent Variable: LOG(Q)
Method: Least Squares
Date: 08/11/97 Time: 16:56
Sample: 1947 1971
Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.327939	0.410601	-5.669595	0.0000
LOG(L)	1.591175	0.167740	9.485970	0.0000
LOG(K)	0.239604	0.105390	2.273498	0.0331

R-squared	0.983672	Mean dependent var	4.767586
Adjusted R-squared	0.982187	S.D. dependent var	0.326086
S.E. of regression	0.043521	Akaike info criterion	-3.318997
Sum squared resid	0.041669	Schwarz criterion	-3.172732
Log likelihood	44.48746	F-statistic	662.6819
Durbin-Watson stat	0.637300	Prob(F-statistic)	0.000000

The sum of the coefficients on LOG(L) and LOG(K) appears to be in excess of one, but to determine whether the difference is statistically relevant, we will conduct the hypothesis test of constant returns.

To carry out a Wald test, choose **View/Coefficient Tests/Wald-Coefficient Restrictions...** from the equation toolbar. Enter the restrictions into the edit box, with multiple coefficient restrictions separated by commas. The restrictions should be expressed as equations involving the estimated coefficients and constants.

Figure 65: Choose help on 'Wald test (Coefficient Restrictions)'. (*Read it later!*)

Diagnostic Tests on the Residuals

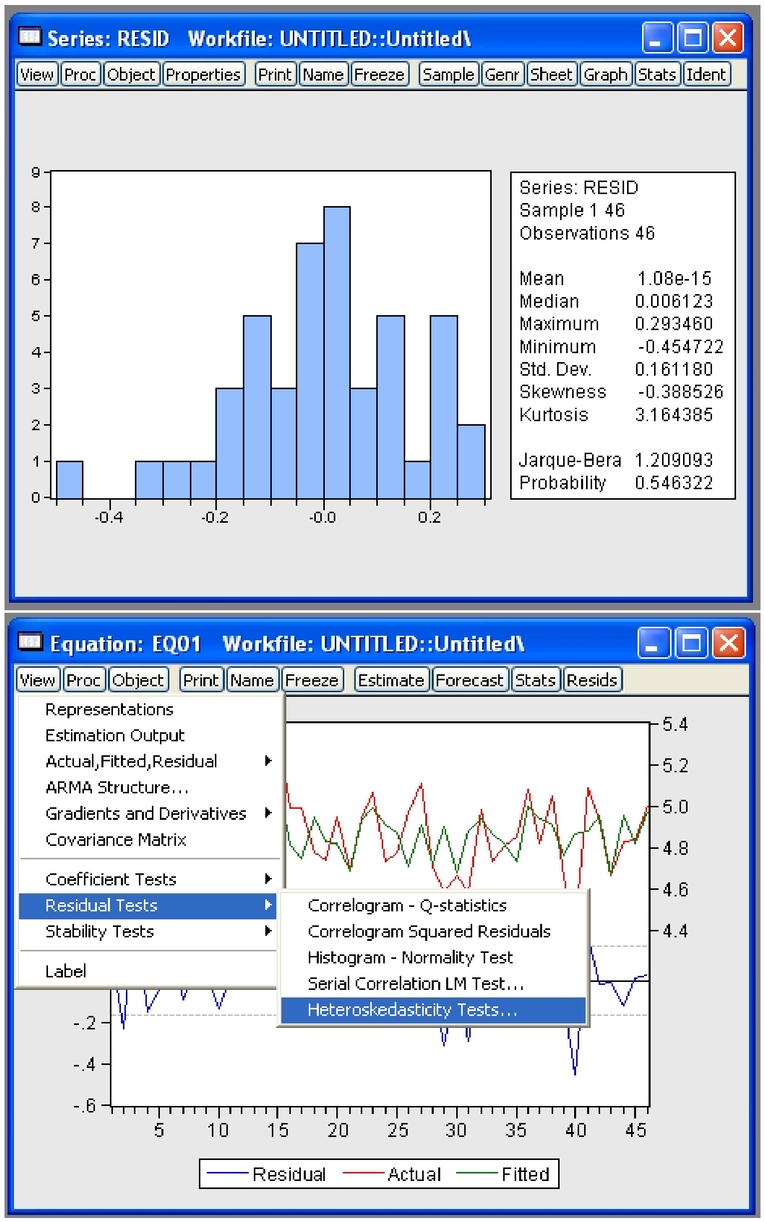
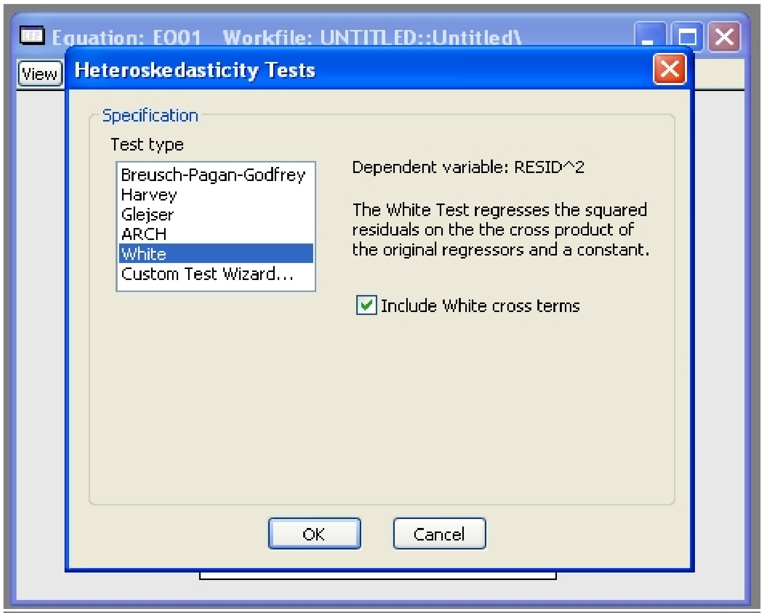
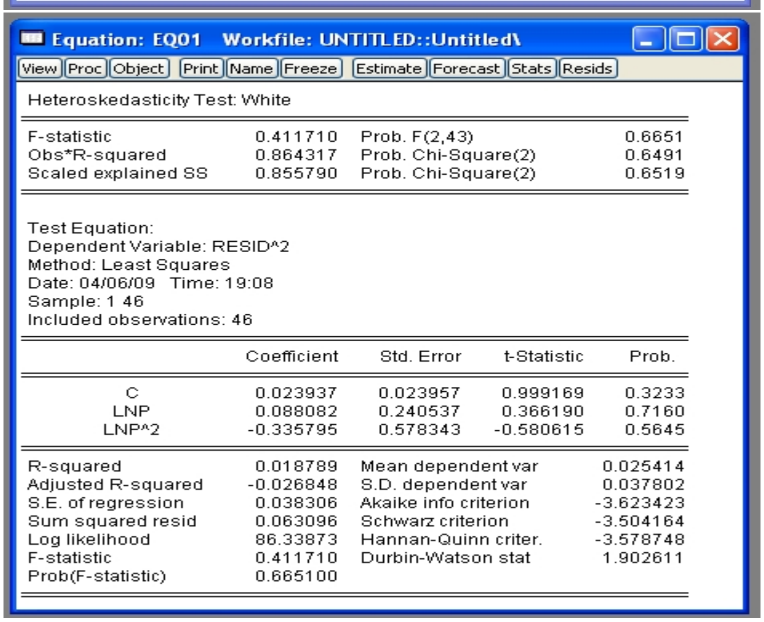


Figure 66: Descriptive statistics, and Jarque-Bera test on estimated residuals from eq01, suggesting that the null of normality of residuals cannot be rejected at usual levels of significance. Select ‘Heteroskedasticity Tests’.





Heteroskedasticity Test: White

F-statistic	0.411710	Prob. F(2,43)	0.6651
Obs*R-squared	0.864317	Prob. Chi-Square(2)	0.6491
Scaled explained SS	0.855790	Prob. Chi-Square(2)	0.6519

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/06/09 Time: 19:08
Sample: 1 46
Included observations: 46

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.023937	0.023957	0.999169	0.3233
LNP	0.088082	0.240537	0.366190	0.7160
LNP^2	-0.335795	0.578343	-0.580615	0.5645

R-squared	0.018789	Mean dependent var	0.025414
Adjusted R-squared	-0.026848	S.D. dependent var	0.037802
S.E. of regression	0.038306	Akaike info criterion	-3.623423
Sum squared resid	0.063096	Schwarz criterion	-3.504164
Log likelihood	86.33873	Hannan-Quinn criter.	-3.578748
F-statistic	0.411710	Durbin-Watson stat	1.902611
Prob(F-statistic)	0.665100		

Figure 67: Choose White's nR^2 test for heteroscedasticity, including cross product terms. (This checks one of the classical assumptions: $Var(u) = \sigma^2 I_n$). The results suggest that the null of homoscedasticity should not be rejected at usual levels of significance.

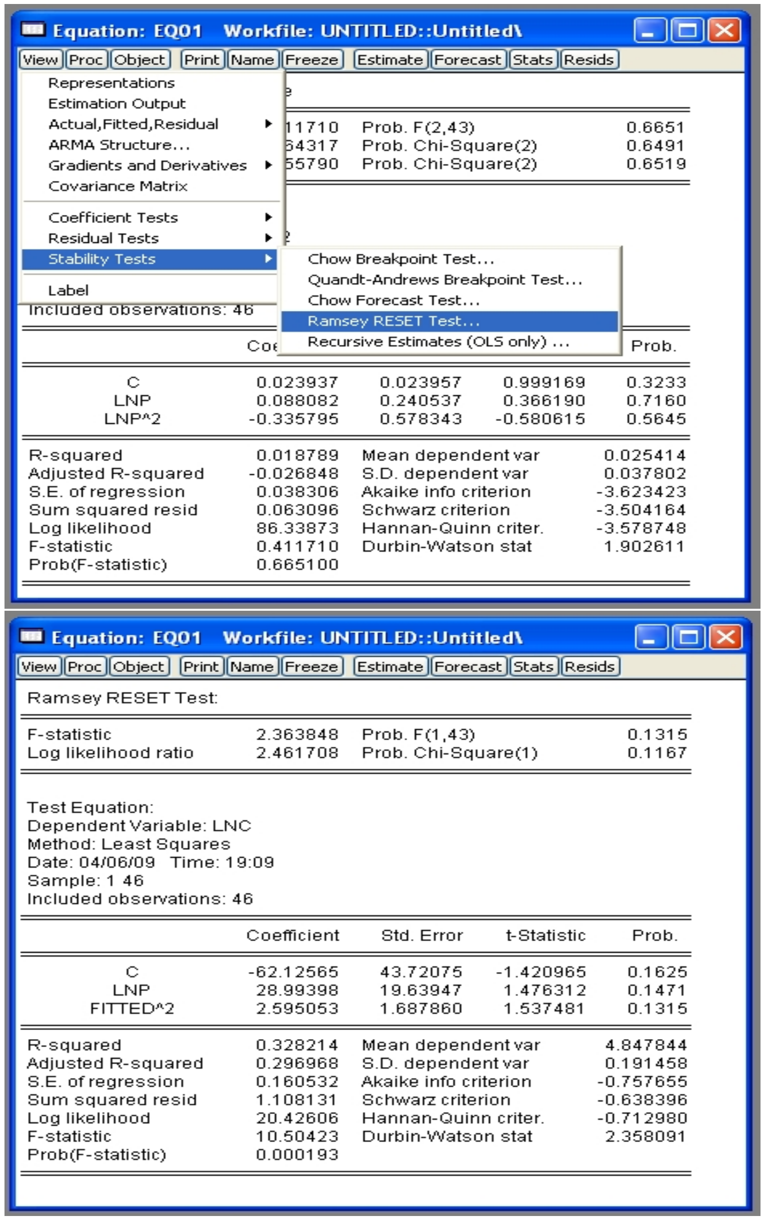


Figure 68: Select the 'Ramsey RESET Test' with one fitted term, to include the squared fitted dependent variable in the RESET test, and observe that the RESET test does not reject the null of no omitted nonlinearity/correct functional form, at usual levels of significance. (This provides some support for the classical assumption $y = X\beta + u$).

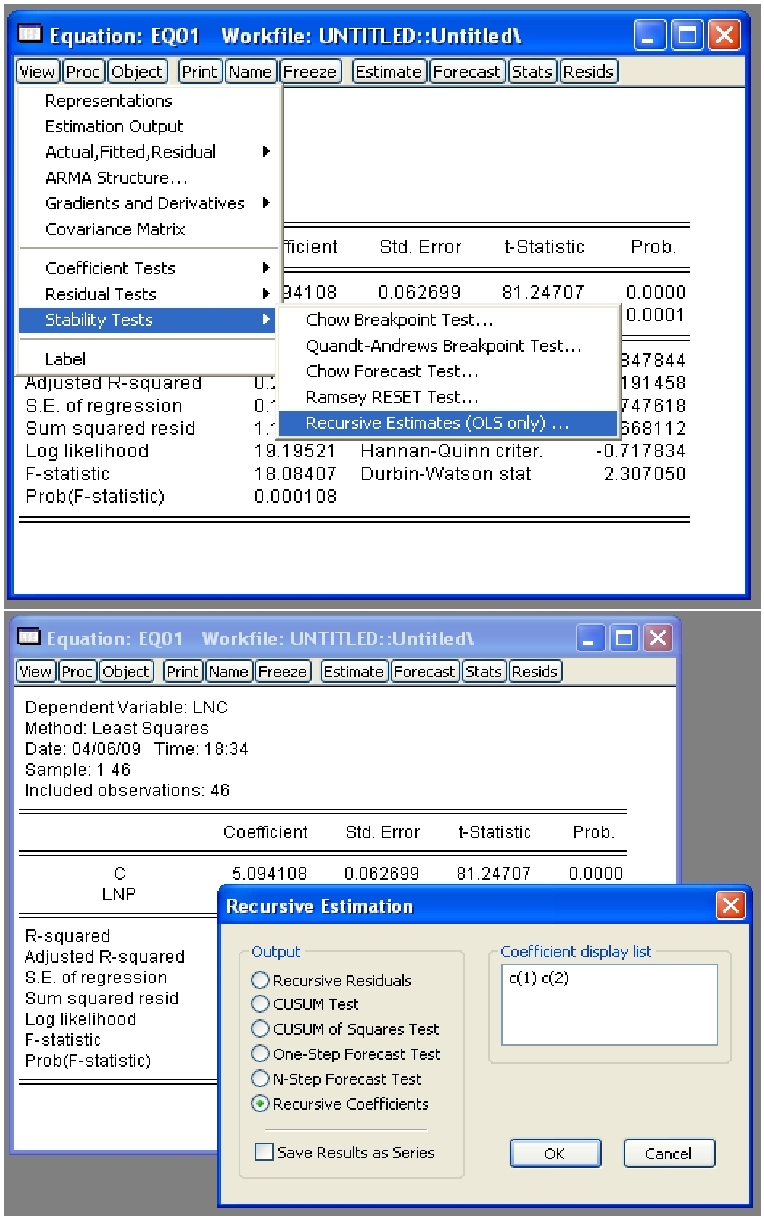


Figure 69: Choose 'Recursive Estimates (OLS only)...', select 'Recursive Coefficients', and set coefficient display list to 'c(1) c(2)'.

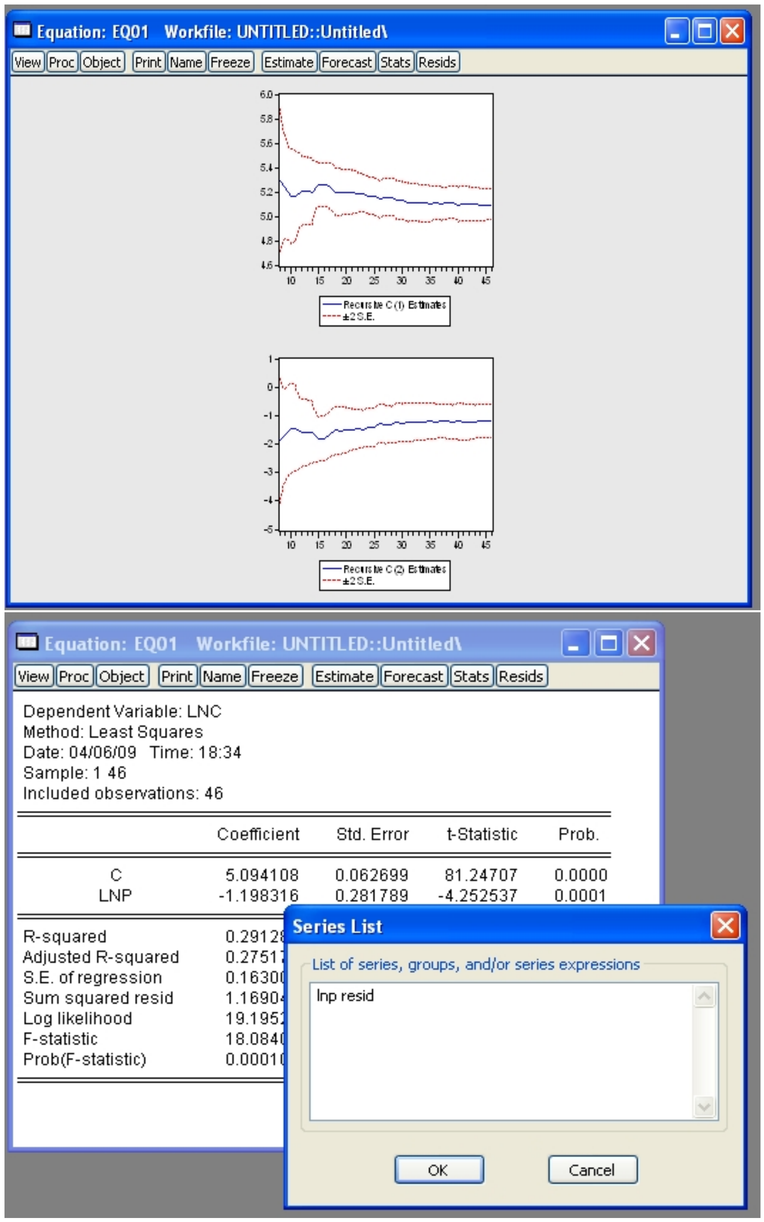


Figure 70: Recursive ordinary least squares estimates for eq01, including an additional datapoint at each step. Select scatter plot of lnp and resid from eq01.

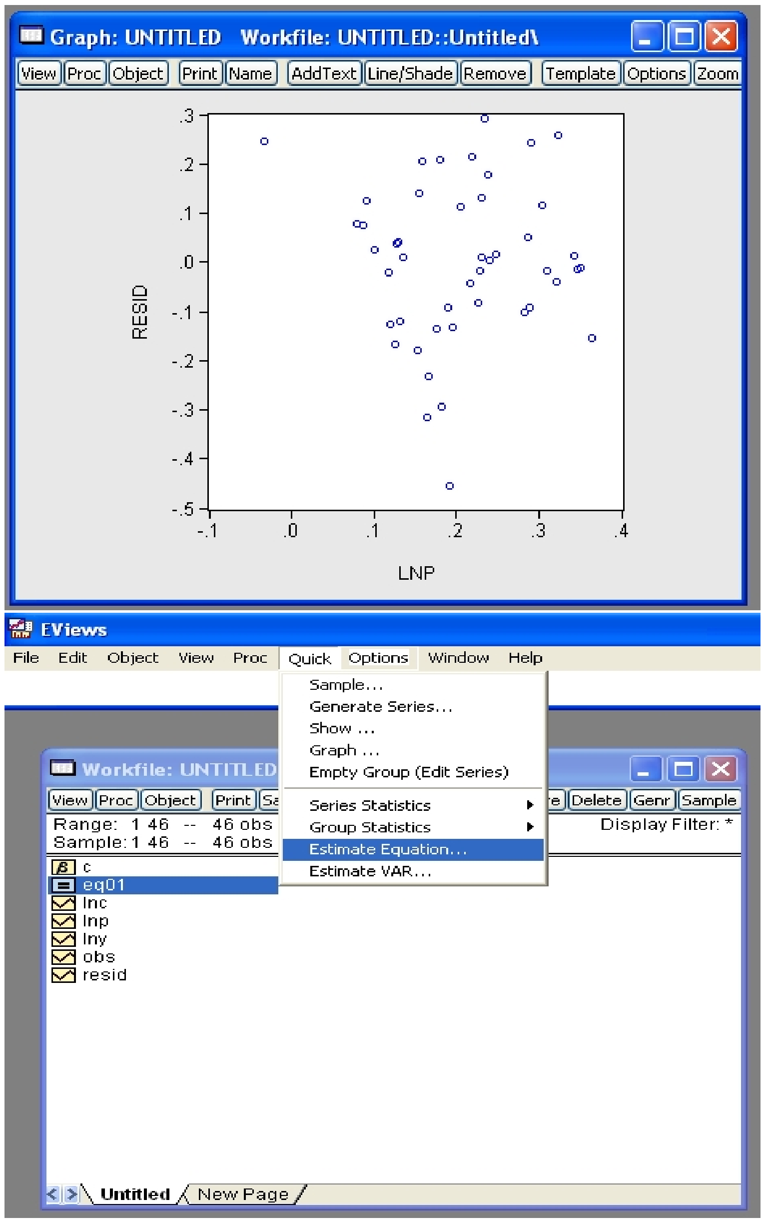


Figure 71: Scatterplot of residuals from eq01 against lnp. Estimate a new equation.

Further Regressions

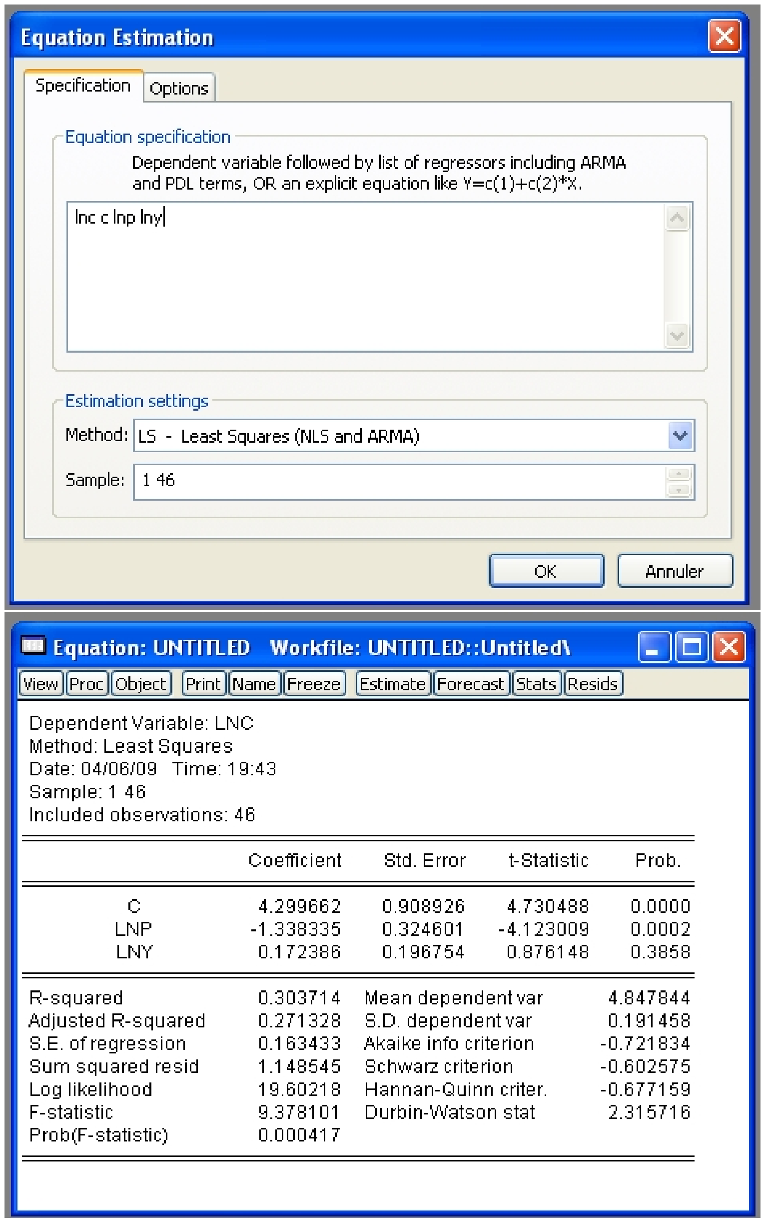


Figure 72: **(eq02)** Run the regression of log consumption on a constant, log price, and log income, using ordinary least squares, and consider the regression output.

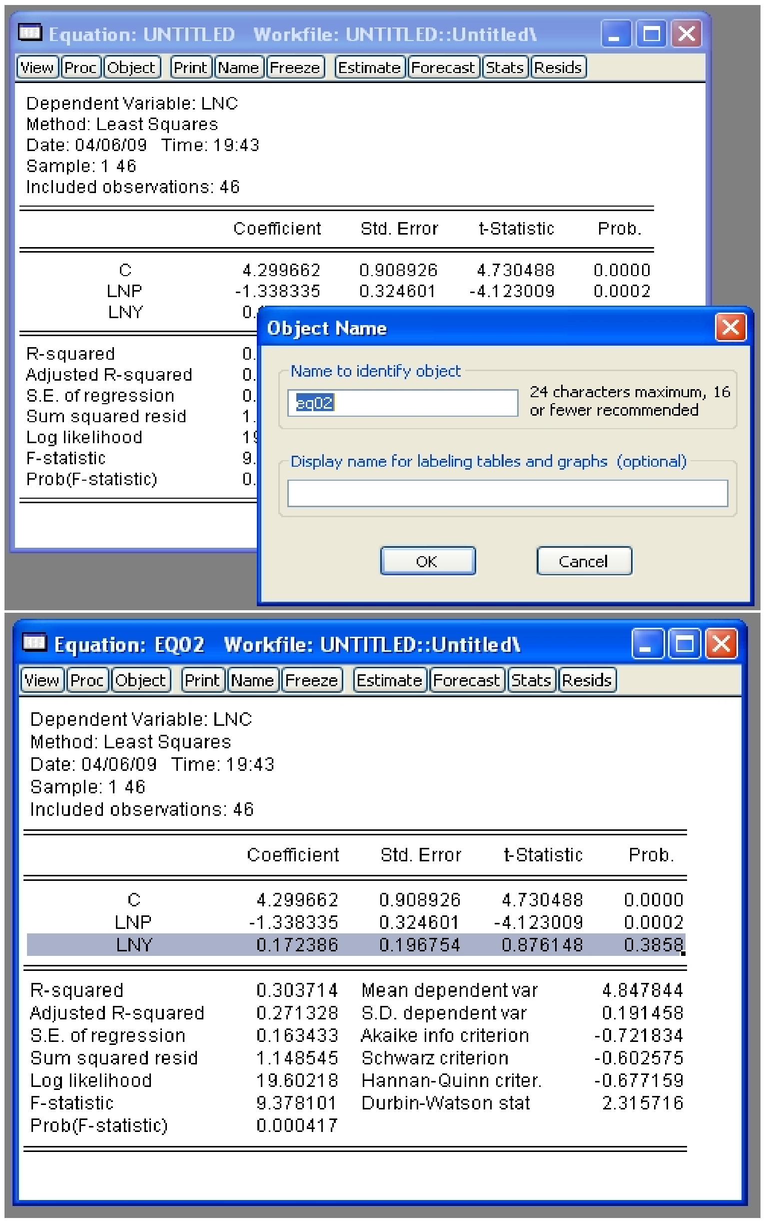


Figure 73: Name the equation 'eq02'.

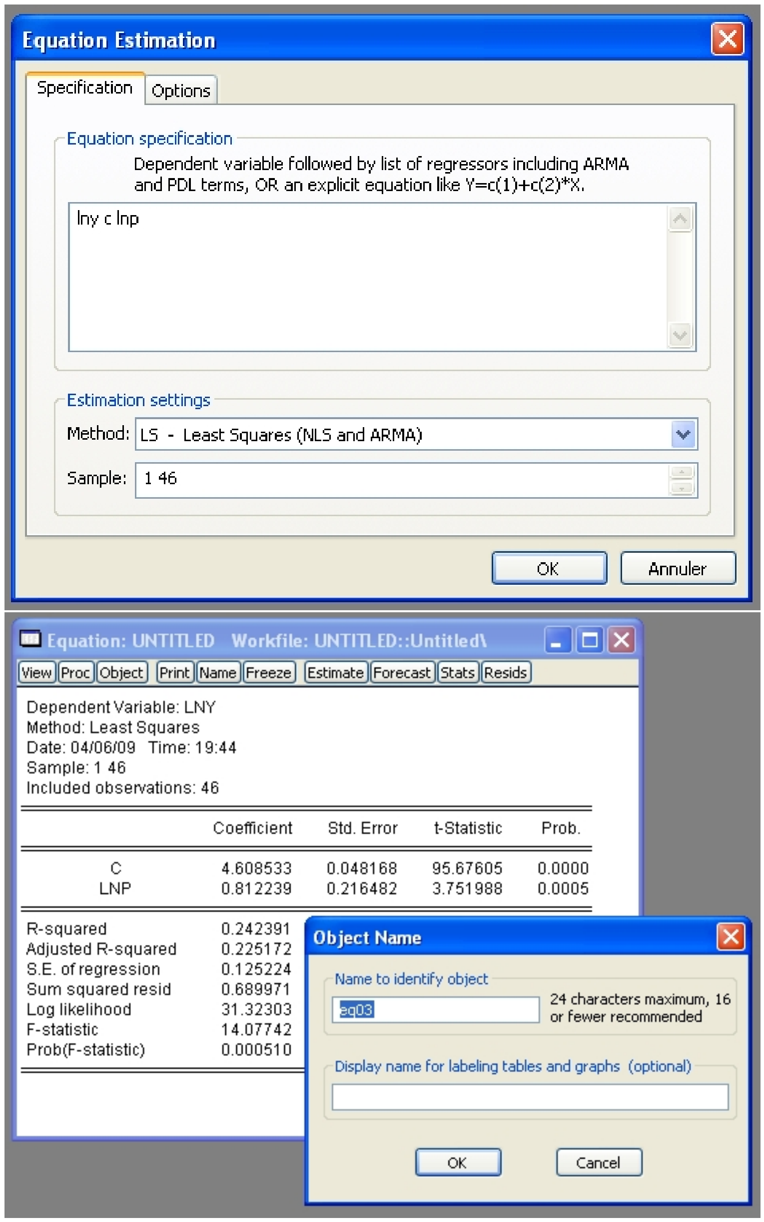


Figure 74: **(eq03)** Regress log income on a constant and log price, using ordinary least squares, and consider the regression output. Name this equation ‘eq03’.

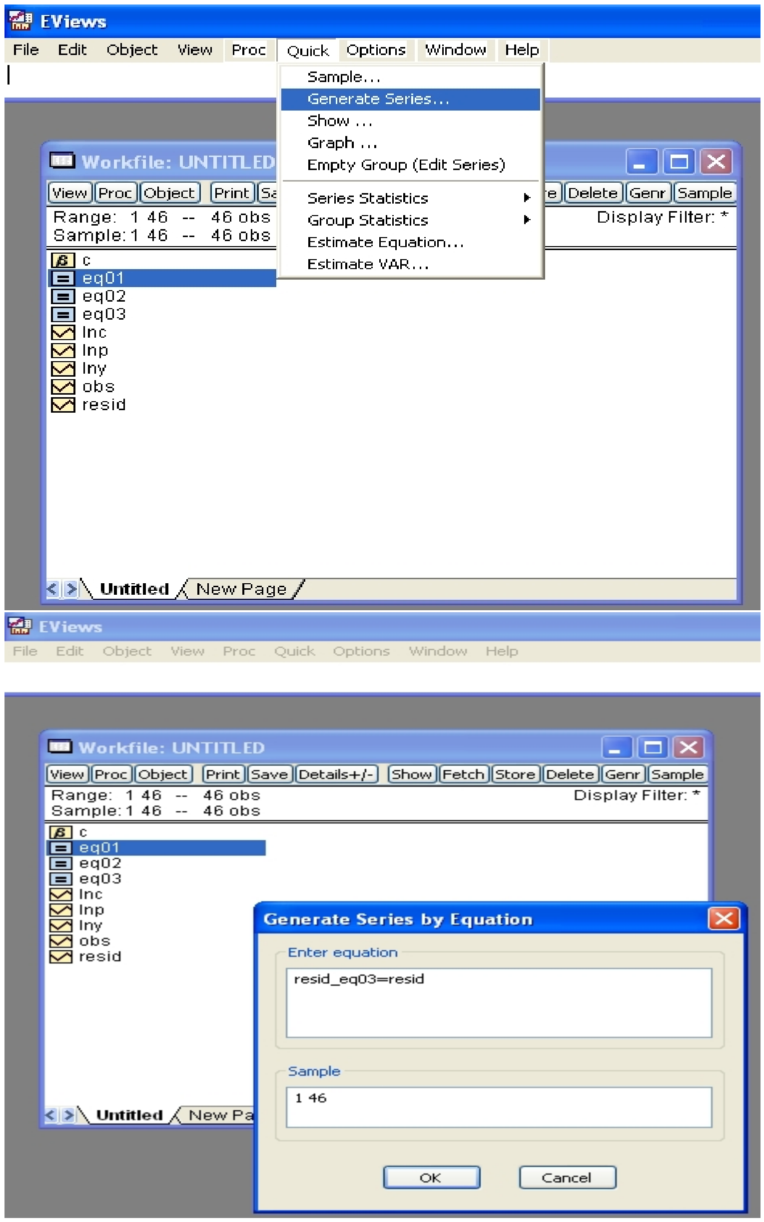


Figure 75: Select 'Generate Series' and enter 'resid_eq03=resid', to store the residuals from eq03, and to prevent them from being overwritten by a new regression.

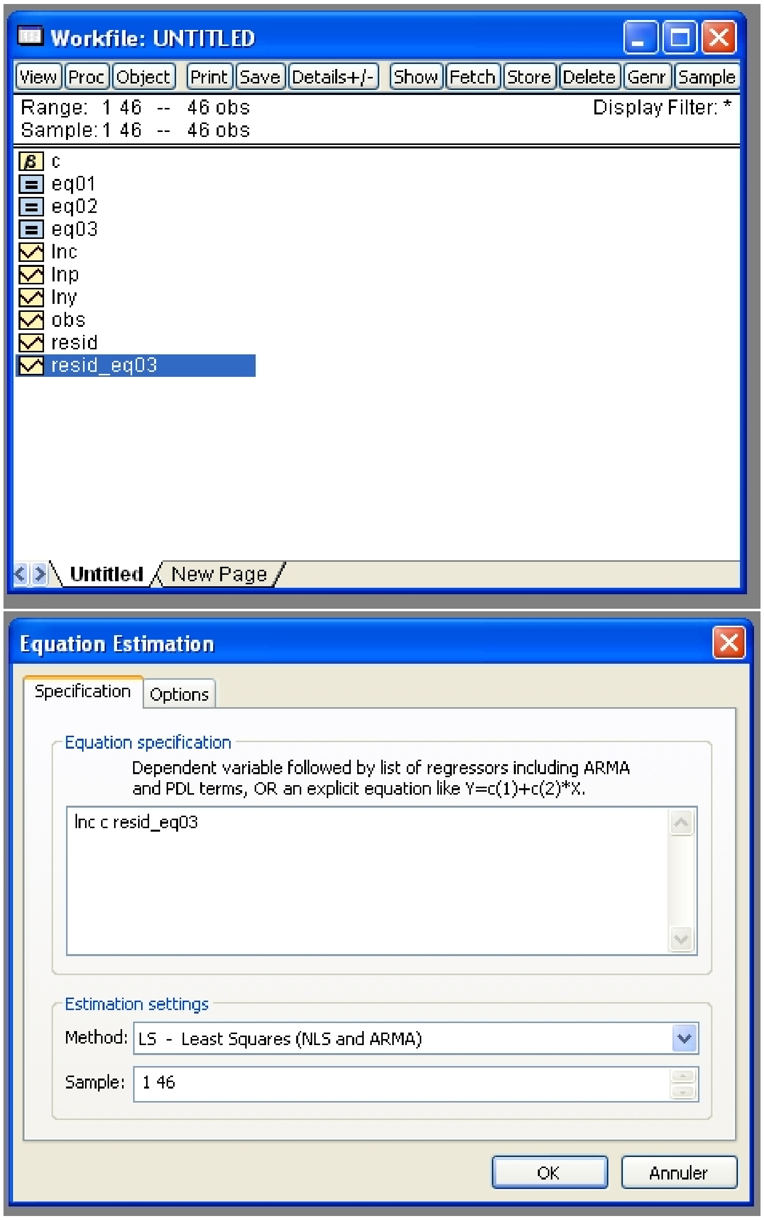


Figure 76: **(eq04)** Check that resid_eq03 appears in the workfile, and run the regression of log consumption on a constant and resid_eq03.

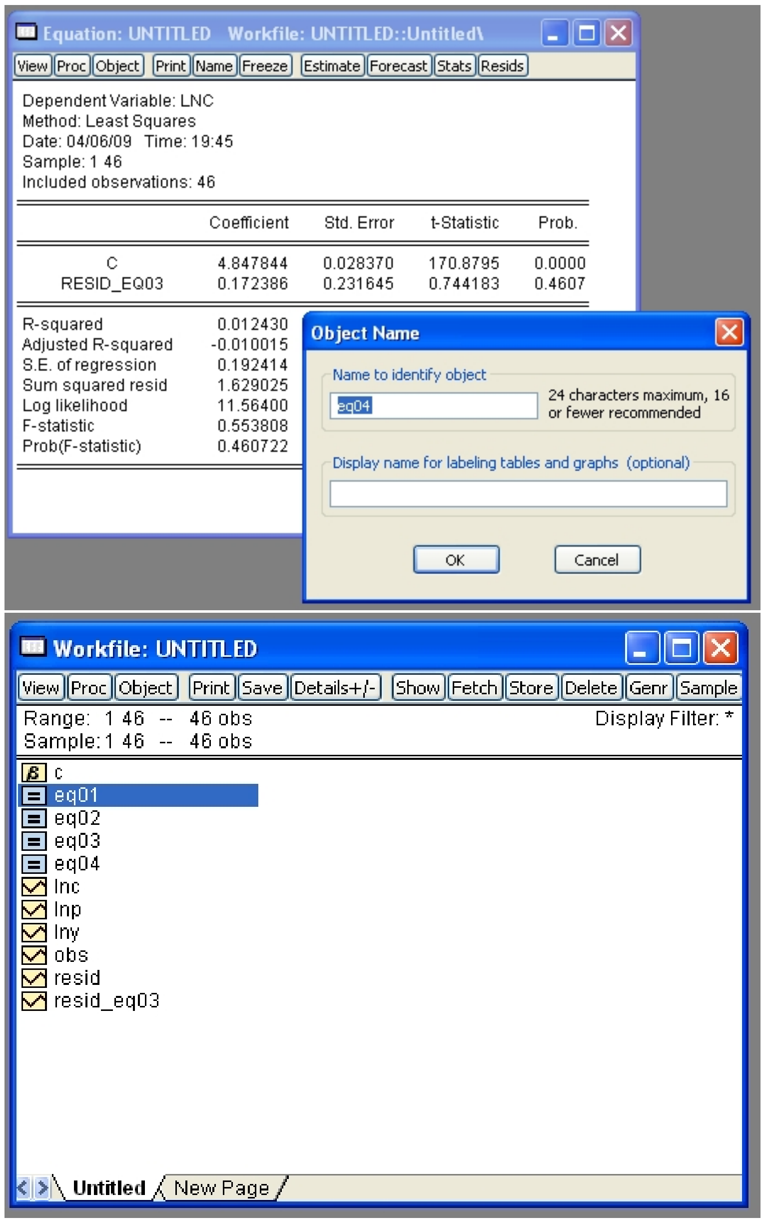


Figure 77: Name this equation 'eq04', and then select the eq01 object.

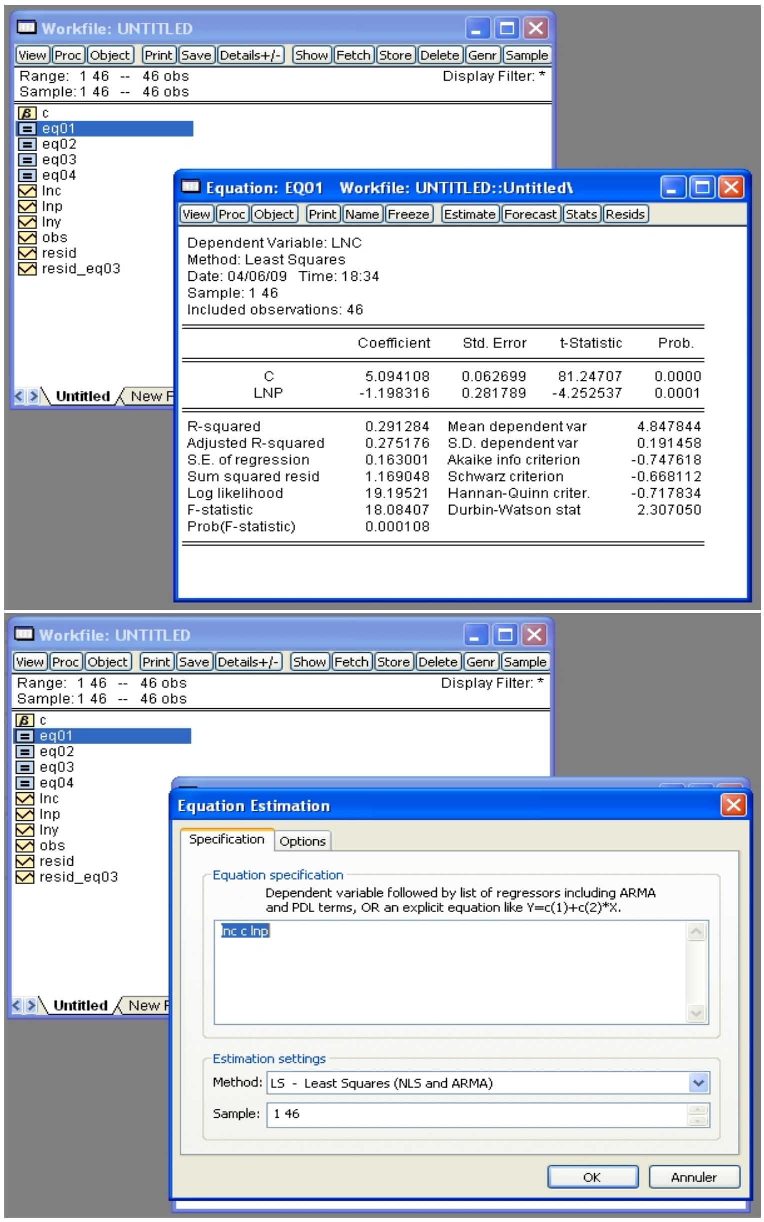


Figure 78: Perform the eq01 regression once again.

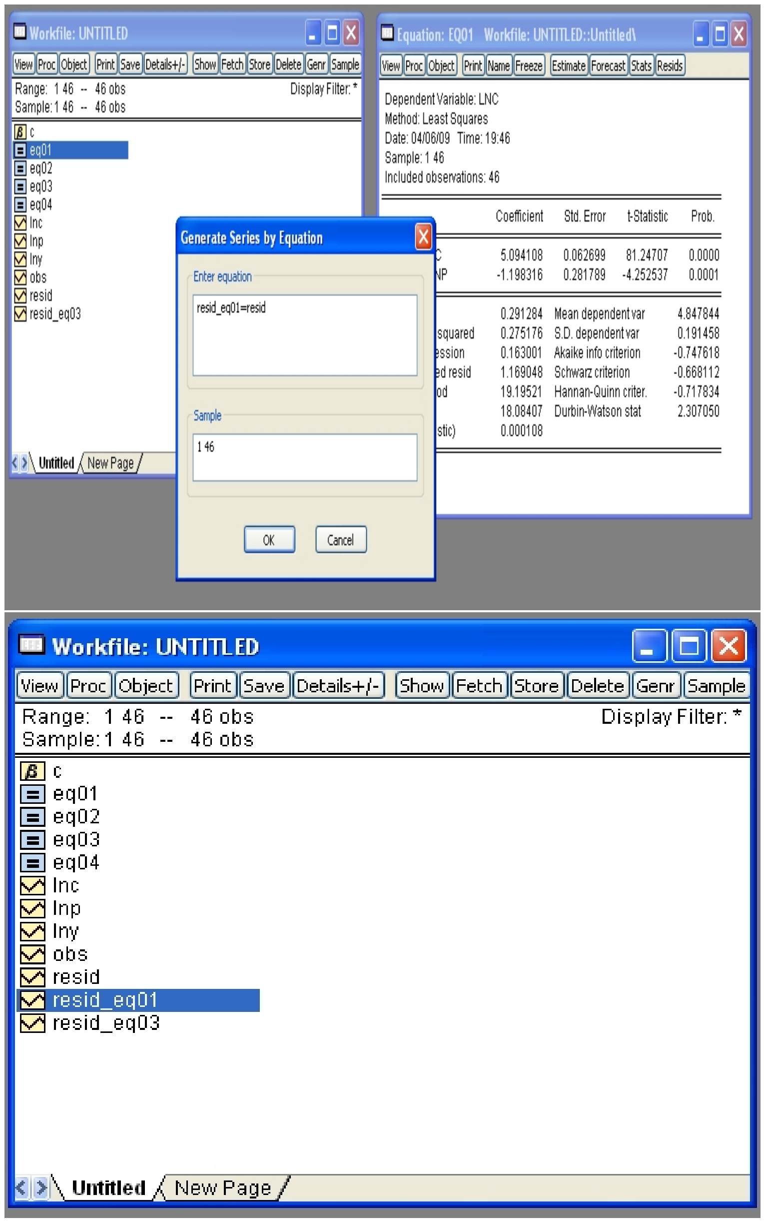


Figure 79: Store the regression residuals in the object 'resid_eq01', and check that they appear in the workfile.

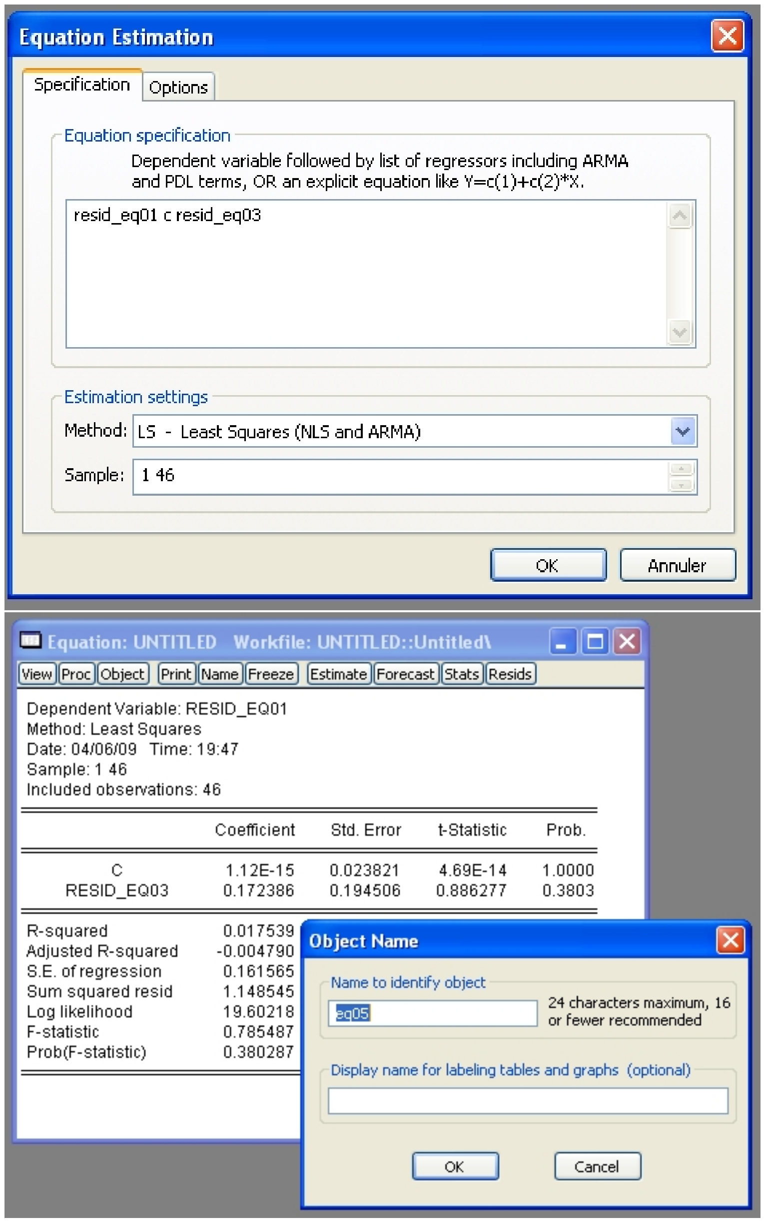


Figure 80: **(eq05)** Run the regression of resid_eq01 on a constant and resid_eq03, and name this equation 'eq05'.

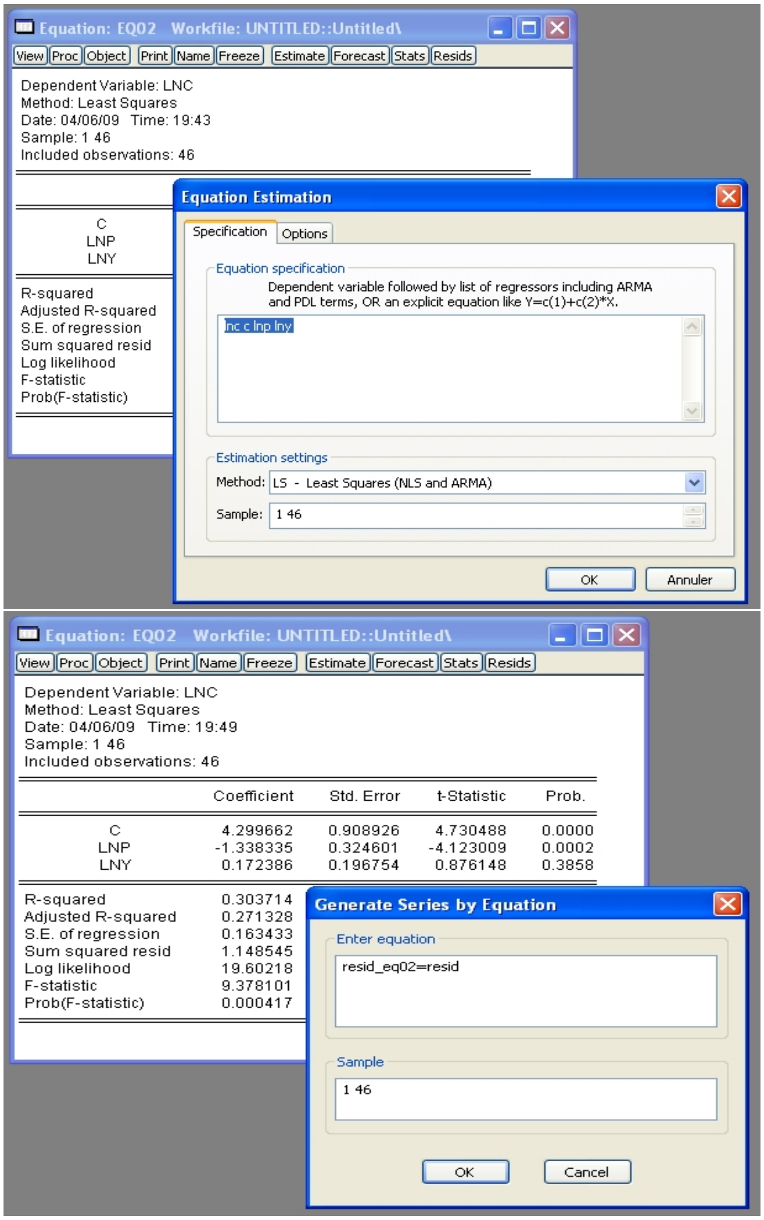
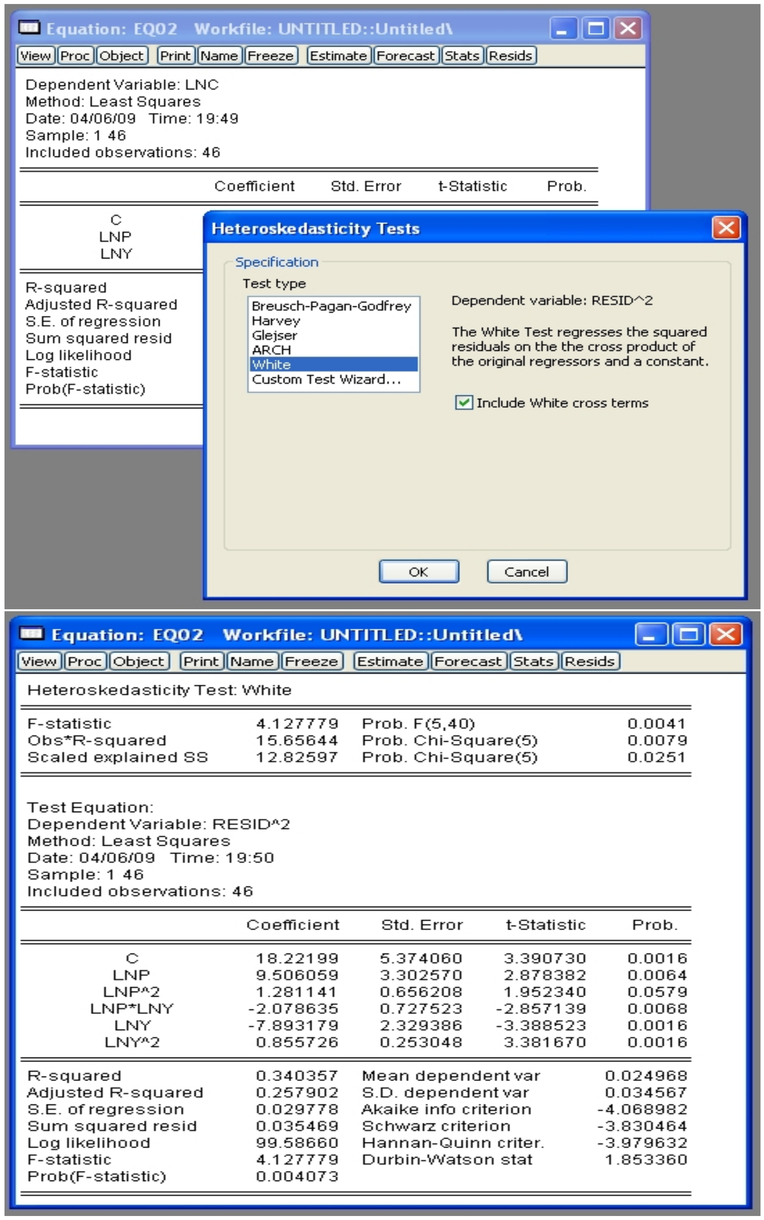


Figure 81: Select the eq02 object, and re-run the regression once again. Store the residuals as 'resid_eq02'.



Heteroskedasticity Test: White

F-statistic	4.127779	Prob. F(5,40)	0.0041
Obs*R-squared	15.65644	Prob. Chi-Square(5)	0.0079
Scaled explained SS	12.82597	Prob. Chi-Square(5)	0.0251

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/06/09 Time: 19:50
Sample: 1 46
Included observations: 46

	Coefficient	Std. Error	t-Statistic	Prob.
C	18.22199	5.374060	3.390730	0.0016
LNP	9.506059	3.302570	2.878382	0.0064
LNP^2	1.281141	0.656208	1.952340	0.0579
LNP*LNY	-2.078635	0.727523	-2.857139	0.0068
LNY	-7.893179	2.329386	-3.388523	0.0016
LNY^2	0.855726	0.253048	3.381670	0.0016

R-squared	0.340357	Mean dependent var	0.024968
Adjusted R-squared	0.257902	S.D. dependent var	0.034567
S.E. of regression	0.029778	Akaike info criterion	-4.068982
Sum squared resid	0.035469	Schwarz criterion	-3.830464
Log likelihood	99.58660	Hannan-Quinn criter.	-3.979632
F-statistic	4.127779	Durbin-Watson stat	1.853360
Prob(F-statistic)	0.004073		

Figure 82: For eq02, perform White's nR^2 test (with cross terms) for heteroscedasticity: homoscedasticity is rejected at all usual levels of significance. (This provides some evidence against the classical assumption $Var(u) = \sigma^2 I_n$).

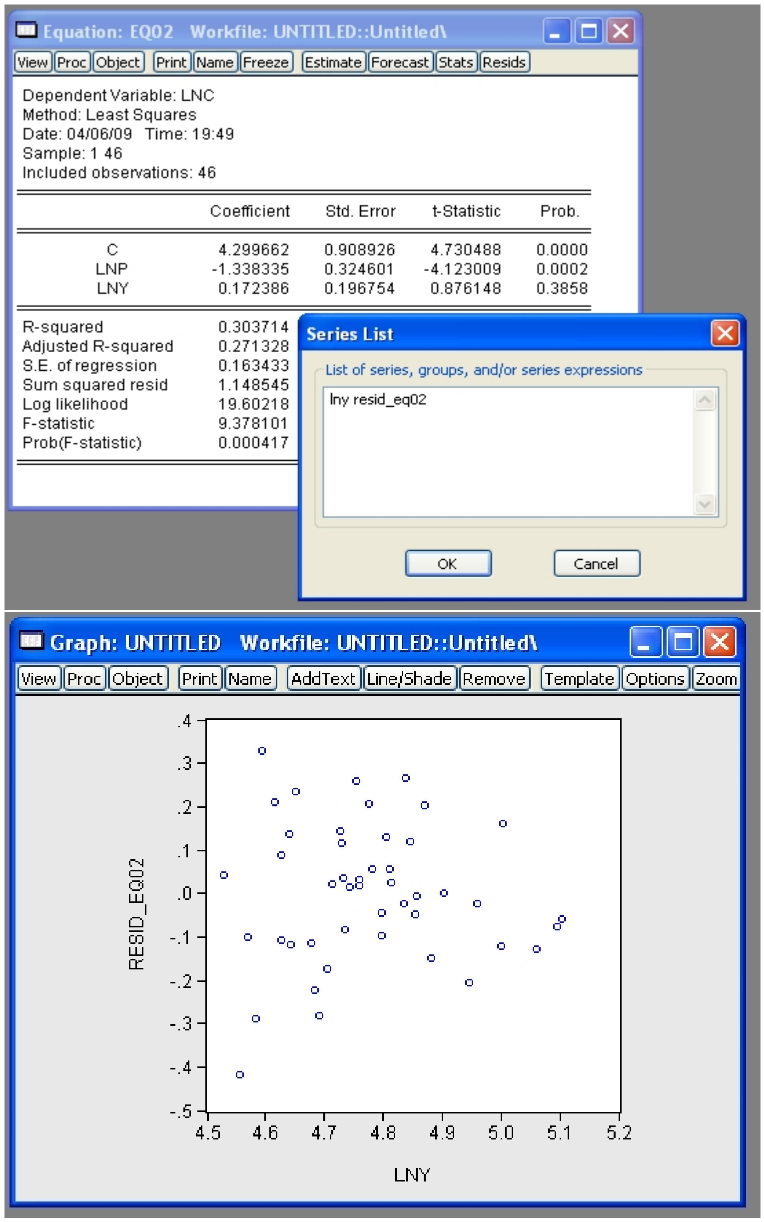
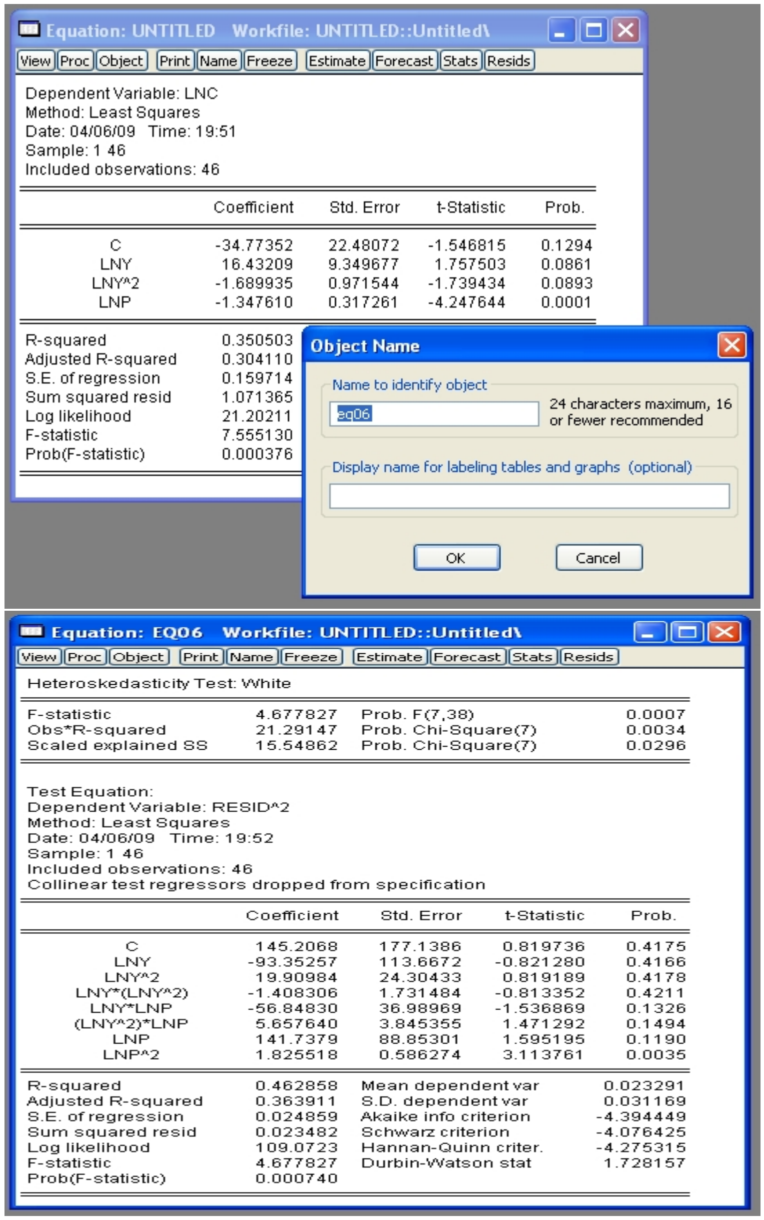


Figure 83: Plot a scatter of resid_eq02 against lny, and note that the variation appears to decrease with lny.



Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LNC
Method: Least Squares
Date: 04/06/09 Time: 19:51
Sample: 1 46
Included observations: 46

	Coefficient	Std. Error	t-Statistic	Prob.
C	-34.77352	22.48072	-1.546815	0.1294
LNY	16.43209	9.349677	1.757503	0.0861
LNY^2	-1.689935	0.971544	-1.739434	0.0893
LNP	-1.347610	0.317261	-4.247644	0.0001

R-squared	0.350503
Adjusted R-squared	0.304110
S.E. of regression	0.159714
Sum squared resid	1.071365
Log likelihood	21.20211
F-statistic	7.555130
Prob(F-statistic)	0.000376

Object Name

Name to identify object: eq06 — 24 characters maximum, 16 or fewer recommended

Display name for labeling tables and graphs (optional)

OK Cancel

Equation: EQ06 Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: White

F-statistic	4.677827	Prob. F(7,38)	0.0007
Obs*R-squared	21.29147	Prob. Chi-Square(7)	0.0034
Scaled explained SS	15.54862	Prob. Chi-Square(7)	0.0296

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 04/06/09 Time: 19:52
Sample: 1 46
Included observations: 46
Collinear test regressors dropped from specification

	Coefficient	Std. Error	t-Statistic	Prob.
C	145.2068	177.1386	0.819736	0.4175
LNY	-93.35257	113.6672	-0.821280	0.4166
LNY^2	19.90984	24.30433	0.819189	0.4178
LNY*(LNY^2)	-1.408306	1.731484	-0.813352	0.4211
LNY*LNP	-56.84830	36.98969	-1.536869	0.1326
(LNY^2)*LNP	5.657640	3.845355	1.471292	0.1494
LNP	141.7379	88.85301	1.595195	0.1190
LNP^2	1.825518	0.586274	3.113761	0.0035

R-squared	0.462858	Mean dependent var	0.023291
Adjusted R-squared	0.363911	S.D. dependent var	0.031169
S.E. of regression	0.024859	Akaike info criterion	-4.394449
Sum squared resid	0.023482	Schwarz criterion	-4.076425
Log likelihood	109.0723	Hannan-Quinn criter.	-4.275315
F-statistic	4.677827	Durbin-Watson stat	1.728157
Prob(F-statistic)	0.000740		

Figure 84: **(eq06)** Perform the regression of log consumption on a constant, log income, log income squared, and log price, and name the equation 'eq06'. Perform White's nR^2 test for heteroscedasticity: note that homoscedasticity is rejected at all usual levels of significance. (This is evidence against the classical assumption $Var(u) = \sigma^2 I_n$).

Saving the Workfile

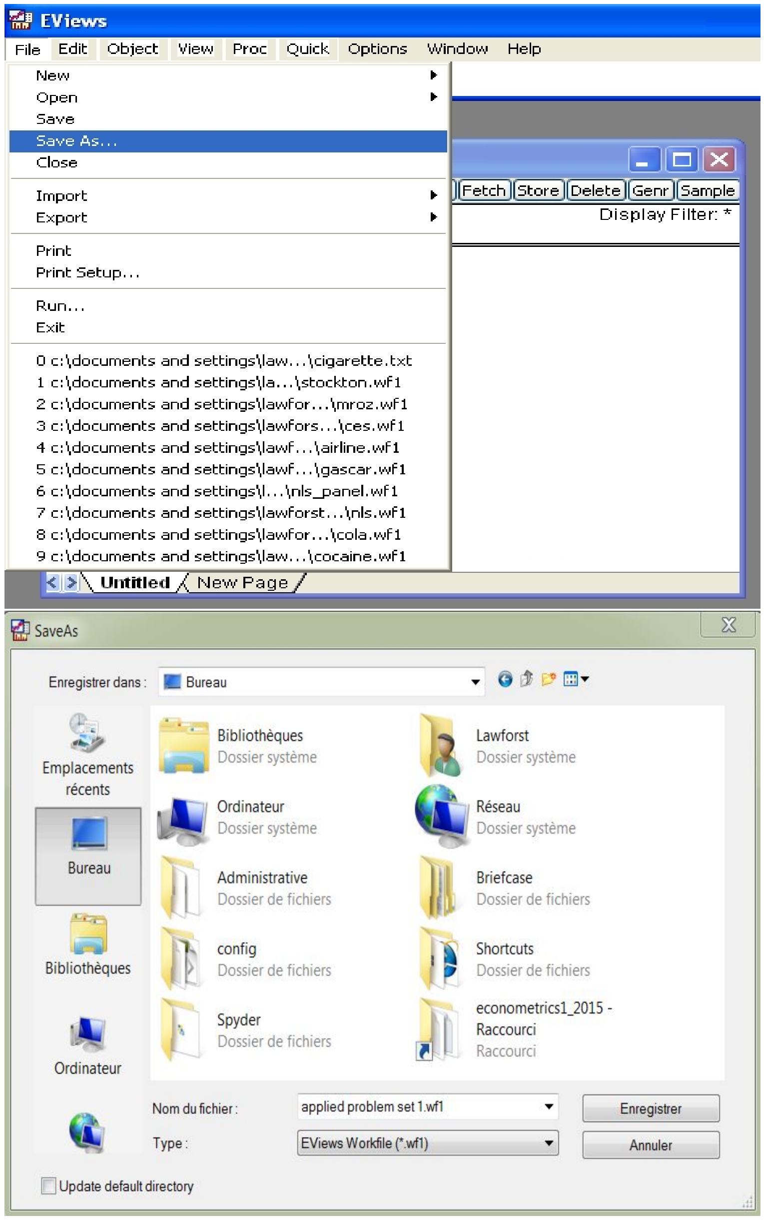


Figure 85: Select 'Save As'. Name the workfile 'applied problem set 1', and select type '.wf1' (Eviews workfile). Store with 'Double precision'. You have saved your work!