Forecasting Methods / Métodos de Previsão

Week 4 - Regression model - Eviews

ISCTE - IUL, Gestão, Econ, Fin, Contab.

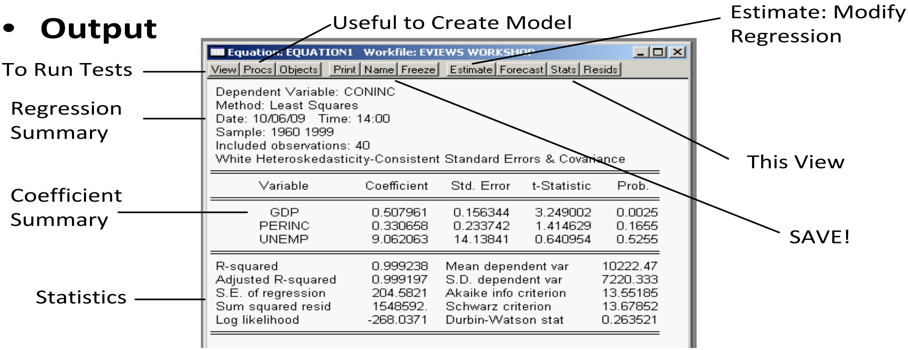
Diana Aldea Mendes

diana.mendes@iscte.pt

February 24, 2011

Regression with Eviews

- Regression output in Eviews



Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP	0.507961	0.156344	3.249002	0.0025
PERINC	0.330658	0.233742	1.414629	0.1655
UNEMP	9.062063	14.13841	0.640954	0.5255

R-squared	0.999238	Mean dependent var	10222.47
Adjusted R-squared	0.999197	S.D. dependent var	7220.333
S.E. of regression	204.5821	Akaike info criterion	13.55185
Sum squared resid	1548592.	Schwarz criterion	13.67852
Log likelihood	-268.0371	Durbin-Watson stat	0.263521

- **Name = Save to Workfile**

- **Equation Output**: When you click OK in the **Equation Specification** dialog, EViews displays the equation window displaying the estimation **output** view:

Dependent Variable: PRICE
Method: Least Squares
Date: 11/05/09 Time: 17:40
Sample: 1964 2003
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13196.12	1699.561	7.764426	0.0000
SIZE	16.67652	0.311663	53.50816	0.0000

R-squared	0.986902	Mean dependent var	87030.00
Adjusted R-squared	0.986557	S.D. dependent var	54123.57
S.E. of regression	6275.317	Akaike info criterion	20.37534
Sum squared resid	1.50E+09	Schwarz criterion	20.45979
Log likelihood	-405.5068	F-statistic	2863.123
Durbin-Watson stat	0.177000	Prob(F-statistic)	0.000000

Regression with Eviews - Coefficients Summary

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13196.12	1699.561	7.764426	0.0000
SIZE	16.67652	0.311663	53.50816	0.0000

- **Variable**: the coefficients will be labeled in the **Variable** column with the name of the corresponding regressor (indep. variable);
- If present, the coefficient of the **C** is the **constant or intercept** in the regression (it is the base level of the prediction when all of the other independent variables are zero)
- The column labeled **Coefficient** depicts the estimated coefficients (computed by the standard OLS formula) - measures the marginal contribution of the independent variable to the dependent variable

Regression with Eviews - Coefficients Summary

- **Standard Errors**: The **Std. Error** column reports the estimated standard errors of the coefficient estimates (measure the statistical reliability of the coefficient estimates—the larger the standard errors, the more statistical noise in the estimates).
- The standard errors of the estimated coefficients are the square roots of the diagonal elements of the coefficient **covariance matrix**. You can view the whole covariance matrix by choosing

View -> Covariance Matrix.

Regression with Eviews - Coefficients Summary

- **t-Statistics**: The $\boxed{\textbf{t-Statistic}}$ (that is, the ratio of an estimated coefficient to its standard error), is used to test the null hypothesis that **a coefficient is equal to zero**.
- **Probability (p-value):** The last column of the output, $\boxed{\textbf{Prob.}}$, shows the probability of drawing a t-statistic as extreme as the one actually observed, under the assumption that the errors are normally distributed, or that the estimated coefficients are asymptotically normally distributed.
- Given a p-value, you can tell if you **reject** or **not reject** the **null hypothesis** that **the true coefficient is zero** against a two-sided alternative that it differs from zero.
- For example, if you are performing the test at the 5% significance level, a p-value lower than 0.05 is taken as evidence to reject the null hypothesis of a zero coefficient.

Regression with Eviews - Summary Statistics

R-squared	0.986902	Mean dependent var	87030.00
Adjusted R-squared	0.986557	S.D. dependent var	54123.57
S.E. of regression	6275.317	Akaike info criterion	20.37534
Sum squared resid	1.50E+09	Schwarz criterion	20.45979
Log likelihood	-405.5068	F-statistic	2863.123
Durbin-Watson stat	0.177000	Prob(F-statistic)	0.000000

- **R-squared**: The R-squared (R^2) statistic measures the success of the regression in predicting the values of the dependent variable within the sample (the statistic will equal **one** if the regression fits perfectly, and **zero** if it fits no better than the simple mean of the dependent variable).
  - It can be negative for a number of reasons. For example, if the regression does not have an intercept or constant, if the regression contains coefficient restrictions, or if the estimation method is two-stage least squares.

- $\boxed{\textbf{Adjusted R-squared}}$: ($\bar{R}^2$) One problem with using R^2 as a measure of goodness of fit is that the R^2 will never decrease as you add more regressors. The adjusted R^2, commonly denoted as $\bar{R}^2$, penalizes the R^2 for the addition of regressors which do not contribute to the explanatory power of the model. The adjusted R^2 is computed as:

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

- The $\bar{R}^2$ is never larger than the R^2, can decrease as you add regressors, and for poorly fitting models, may be negative.

Regression with Eviews - Summary Statistics

- **Standard Error of the Regression** ($\boxed{\textbf{S.E. of regression}}$): is a summary measure based on the estimated variance of the residuals. The standard error of the regression is computed as:

$$s = \sqrt{\frac{u^2}{(T-k)}}$$

- $\boxed{\textbf{Sum-of-Squared Residuals}}$: The sum-of-squared residuals can be used in a variety of statistical calculations (loss function to optimize in OLS estimation)

- $\boxed{\textbf{Log Likelihood}}$: EViews reports the value of the log likelihood function (assuming normally distributed errors) evaluated at the estimated values of the coefficients. The log likelihood is computed as:

$$l = -\frac{T}{2}\left(1 + \log(2\pi) + \log\left(\frac{\sum u_t^2}{T}\right)\right)$$

- $\boxed{\textbf{Durbin-Watson Statistic}}$: The Durbin-Watson (DW) statistic measures the (first order) **serial correlation (autocorrelation)** in the residuals. The statistic is computed as

$$d = \frac{\sum_{t=1}^{T}(u_t - u_{t-1})^2}{\sum_{t=1}^{T} u_t^2}$$

Regression with Eviews - Summary Statistics

- The value of d always lies between 0 and 4
- Since d is approximately equal to $2(1-r)$, where r is the sample autocorrelation of the residuals, $d=2$ indicates **no (auto)correlation**.
- If the Durbin–Watson statistic is substantially **less than 2**, there is **evidence of positive serial correlation**.
- If Durbin–Watson is **less than 1.0**, there may be cause for alarm (indicate successive error terms are, on average, close in value to one another, or positively correlated).
- If $d>2$ successive error terms are negatively correlated. In regressions, this can imply an underestimation of the level of statistical significance.
- Hypothesis setting

$$\begin{aligned} H_0 &: \quad \text{no serial (auto) correlation (independence)} \\ H_1 &: \quad \text{serial (auto) correlation} \end{aligned}$$

Regression with Eviews - Summary Statistics

- $\boxed{\textbf{Mean and Standard Deviation (S.D.) of the Dependent Variable}}$: The mean and standard deviation of are computed using the standard formulae:

$$y = \frac{\sum_{t=1}^{T} y_t}{T};\ s_y = \sqrt{\frac{\sum_{t=1}^{T} (y_t - \bar{y})^2}{T-1}}$$

- $\boxed{\textbf{Akaike Information Criterion}}$: The Akaike Information Criterion (AIC) is computed as:

$$AIC = \frac{2(K-l)}{T}$$

where l is the log likelihood. The AIC is often used in model selection for non-nested alternatives - **smaller values of the AIC are preferred**.

Regression with Eviews - Summary Statistics

- **Schwarz Criterion**: The Schwarz Criterion (SC) is an alternative to the AIC that imposes a larger penalty for additional coefficients:

$$SIC = \frac{(K \log(T) - 2l)}{T}$$

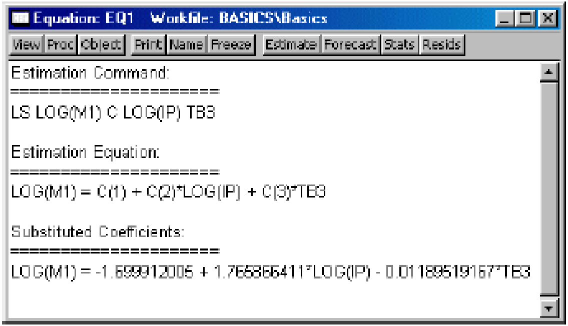
- **F-Statistic**: The F-statistic reported in the regression output is from a test of the hypothesis that all of the slope coefficients (excluding the constant, or intercept) in a regression are zero.
- Under the null hypothesis with normally distributed errors, this statistic has an F-distribution with numerator degrees of freedom and denominator degrees of freedom. The p-value given just below the F-statistic, denoted Prob(F-statistic), is the marginal significance level of the F-test.
- Note that the F-test is a joint test so that even if all the t-statistics are insignificant, the F-statistic can be highly significant.

Regression with Eviews

R^2 and Adjusted R^2	$\rightarrow 1$	$> 0,8$
J-statistic	$\rightarrow 0$	$< 0,1$
Mean dependant variable	$\rightarrow +\infty$	> 100
S.E. of Regression	$\rightarrow 0$	Choose the lower value (comparison)
Residual sum of squares	$\rightarrow 0$	Choose the lower value (comparison)
Prob(F-statistic)	$\rightarrow 0$	$< 0,05$
Durbin-Watson statistic	$\rightarrow 2$	$1 < \text{DW} < 3$ (Under conditions)
Determinant residual covariance	$\rightarrow 0$	Choose the lower value (comparison)
Log-Likelihood	$\rightarrow +\infty$	$> 10^3$
Average Log-Likelihood	$\rightarrow +\infty$	> 10
AIC	$\rightarrow -\infty$	Choose the lower value (comparison)
SIC	$\rightarrow -\infty$	Choose the lower value (comparison)
HQIC	$\rightarrow -\infty$	Choose the lower value (comparison)

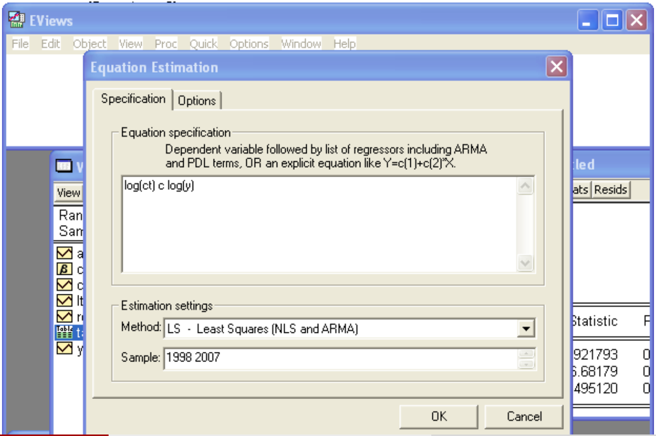
Regression with Eviews - Working with equations

- $\boxed{\textbf{View}}$ of an Equation: **Representations**. Displays the equation in three forms: EViews command form, as an algebraic equation with symbolic coefficients, and as an equation with the estimated values of the coefficients.
- **Estimation Output**. Displays the equation output results described above.



- **Actual, Fitted, Residual**. These views display the actual and fitted values of the dependent variable and the residuals from the regression in tabular and graphical form. **Residual Graph** plots only the residuals, while the **Standardized Residual Graph** plots the residuals divided by the estimated residual standard deviation.
- **Covariance Matrix**. Displays the covariance matrix of the coefficient estimates as a spreadsheet view
- **Coefficient Tests, Residual Tests, and Stability Tests**. These are views for specification and diagnostic tests

- **Procedures of an Equation**: $\boxed{\textbf{Proc}}$
  - **Specify/Estimate:** Brings up the Equation Specification dialog box so that you can modify your specification (edit the equation specification, or change the estimation method or estimation sample).



Regression with Eviews- Working with equations

- **Forecast:** Forecasts or fits values using the estimated equation.
- **Make Residual Series:** Saves the residuals from the regression as a series in the workfile.
- **Make Regressor Group**: Creates an untitled group comprised of all the variables used in the equation (with the exception of the constant).
- **Make Model**: Creates an untitled model containing a link to the estimated equation.
- **Update Coefs from Equation**: Places the estimated coefficients of the equation in the coefficient vector. You can use this procedure to initialize starting values for various estimation procedures.

Regression with Eviews- Working with equations

- Residuals from an Equation: The residuals from the default equation are stored in a series object called **RESID**. RESID may be used directly as if it were a regular series, except in estimation.
- RESID will be overwritten whenever you estimate an equation and will contain the residuals from the latest estimated equation.
- To save the residuals from a particular equation for later analysis, you should save them in a different series so they are not overwritten by the next estimation command.
- For example, you can copy the residuals into a regular EViews series called **RES1** by the command:

series res1 = resid

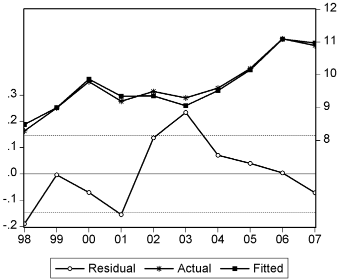
or use **Quick** from the menu of the **command window** (main Eviews window) **Quick -> Generate Series** and insert **res1=resid**

Regression with Eviews - Residual assumptions

- $\boxed{\textbf{Looking at Residuals}}$: In Equation View:

  $\boxed{\textbf{View} \rightarrow \textbf{Actual, Fitted, Residual} \rightarrow \textbf{Actual, Fitted, Residual Tabl}}$

- Click $\boxed{\textbf{Resid}}$ at the menu of the Equation View, to observe the residuals graph
- Plotting Resid Vs. Fitted Values (Scatter plot for Group)

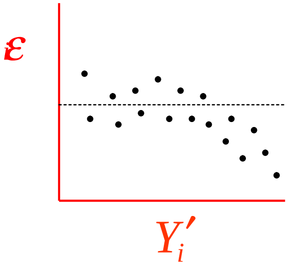


Regression with Eviews - Residual assumptions

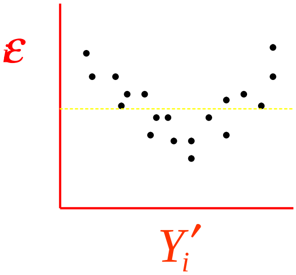
- **Linearity**: if you fit a linear model to data which are nonlinearly related, your predictions are likely to be seriously in error
- **How to detect**: plot of the observed versus predicted values or plot of residuals versus predicted values (the points should be symmetrically distributed around a diagonal line in the former plot or a horizontal line in the latter plot).
- **How to fix**: consider applying a nonlinear transformation to the dependent and/or independent variables. For example, if the data are strictly positive, a log transformation may be feasible. Another possibility to consider is adding another regressor which is a nonlinear function of one of the other variables. For example, if you have regressed Y on X, and the graph of residuals versus predicted suggests a parabolic curve, then it may make sense to regress Y on both X and X^2.

Regression with Eviews - Residual assumptions

Residuals for a non-linear fit



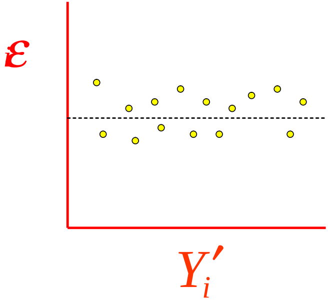
Residuals for a quadratic function or polynomial



Regression with Eviews - Residual assumptions

- $\boxed{\textbf{Mean (expected value) of residuals is zero}}$: $E(u_t) = 0$
- If $\beta_0 \neq 0$, then we have always $E(u_t) = 0$
- If $\beta_0 = 0$, then R^2 can be negative (so, the sample mean explain more about variations in y that the independent variable)
- If $\beta_0 = 0$, biased estimation of β

Expected distribution of residuals for a linear model with normal distribution or residuals (errors).



Regression with Eviews - Residual Assumptions

- $\boxed{\textbf{Homoscedasticity}}$**:** The variance of the residual (u) is constant (**Homoscedasticity**) $Var(u_t) = \sigma^2$: Heteroscedasticity is a term used to the describe the situation when the variance of the residuals from a model is not constant.
- Detection of Heteroscedasticity: graphical representation of residuals versus independent variable
- Detection of Heteroscedasticity: **Breusch-Pegan-Godfrey Test**

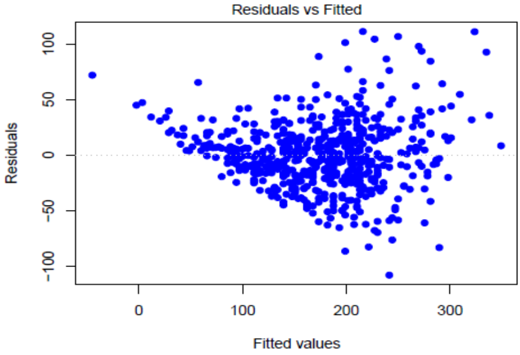
View ->Residual Test -> White Heteroscedasticity

- Hypothesis setting for heteroscedasticity

$$H_0: \text{ Homoscedasticity (the variance of residual } (u) \text{ is constant))}$$
$$H_1: \text{ Heteroscedasticity (the variance of residual } (u) \text{ is not constant)}$$

Regression with Eviews - Residual assumptions

- Residuals are not homogeneous (increasing in variance)



Example

The p-value of Obs*R-squared shows that we **can not reject null**. So residuals do have constant variance which is desirable meaning that residuals are homoscedastic.

F-statistic	1.84	Probability	0.3316
Obs*R-squared	3.600	Probability	0.3080

Regression with Eviews - Residual Assumptions

- **Problems** when $Var(u_t)$ is not constant (Heteroscedasticity)
  - OLS is no longer **efficient** among linear estimators, and this means that hypothesis test and confidence intervals are not truthfully
  - OLS errors
    - to large for the intercept β_0
    - to small (or to large) for β_1 if the residual variance is positively (negatively) related to the independent variable

Regression with Eviews - Residual Assumptions

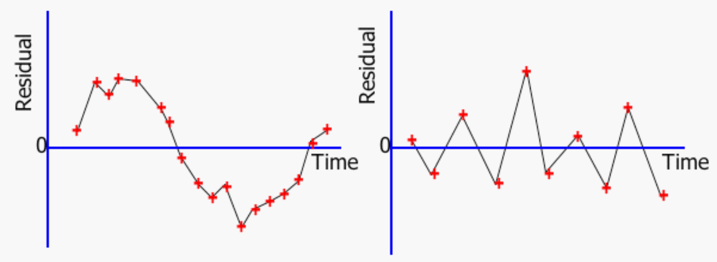
- How to **correct these problems**:
  - If the variance of the residuals appears to be increasing in Y-predicted (and if Y is a positive random variable), then try a Variance-Stabilizing Transformation, such taking the **log** or **square root** of Y to reduce this heteroscedasticity
  - If Y is non-positive, or if you do not wish to transform Y for some reason (such as ease of interpreting the results) then you should try a Weighted Least-Squares procedure.
  - use Maximum likelihood estimation method

Regression with Eviews - Residual Assumptions

- $\boxed{\textbf{No serial or (auto)correlation in the residual } (u)}$:
  $Cov\left(u_i, u_j\right) = 0, i \neq j$. Serial correlation is a statistical term used to describe the situation when the residual is correlated with lagged values of itself. In other words, If residuals are correlated, we call this situation serial correlation which is not desirable.
- How serial correlation can be formed in the model?
  - Incorrect model specification,
  - omitted variables,
  - incorrect functional form,
  - incorrectly transformed data.
- Detection of serial correlation: **Breusch-Godfrey serial correlation LM test**

View ->Residual Test -> Serial Correlation LM test

Regression with Eviews - Residual Assumptions



Note the runs of positive residuals, replaced by runs of negative residuals

Note the oscillating behavior of the residuals around zero.

Regression with Eviews - Residual Assumptions

- Hypothesis setting

$$H_0: \text{ no serial correlation (no correlation between residuals } u_i \text{ and } u_j)$$
$$H_1: \text{ serial correlation (correlation between residuals } u_i \text{ and } u_j)$$

Example

There is serial correlation in the residuals (u) since the p-value (0.3185) of Obs*R-squared is more than 5 percent ($p > 0.05$), we can not reject null hypothesis meaning that residuals (u) are not serially correlated which is desirable.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	**1.01**	**Prob. F(2,29)**	**0.3751**
Obs*R-squared	**2.288**	**Prob. Chi-Square(2)**	**0.3185**

Regression with Eviews - Residual Assumptions

- **Problems** when the residuals are correlated
  - OLS is no longer **efficient** among linear estimators, and this means that hypothesis test and confidence intervals are not truthfully
- How to solve these problems
  - estimate the model for the first difference of variables ($\Delta y_t = y_t - y_{t-1}$) instead of levels
  - use other estimation method
  - use other econometric model

Regression with Eviews - Residual Assumptions

- $\boxed{\textbf{Normality}}$: Residuals ($u$) should be normally distributed: **Jarque Bera statistics**

$\boxed{\textbf{View ->Residual Test -> Histogram - Normality test}}$

- Setting the hypothesis:

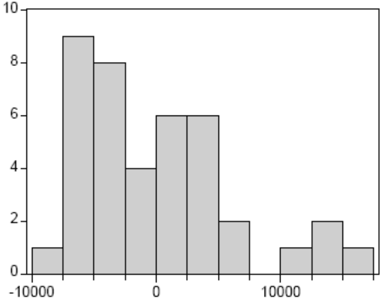
$$H_0: \text{ Normal distribution (the residual } (u) \text{ follows a normal distribut}$$
$$H_1: \text{ Not normal distribution (the residual } (u) \text{follows not normal distrib}$$

- If the p-value of Jarque-Bera statistics is less than 5 percent (0.05) we can reject null and accept the alternative, that is residuals (u) are not normally distributed.
- Note that the DW statistic is not appropriate as a test for serial correlation, if there is a lagged dependent variable on the right-hand side of the equation.

Regression with Eviews - Residual Assumptions

Example

Jarque Berra statistics is 5.4731 and the corresponding $p-$value is 0.0647. Since p value is more than 5 percent we accept null meaning that population residual (u) is normally distributed which fulfills the assumption of a good regression line.



Series: Residuals	
Sample 1964 2003	
Observations 40	
Mean	1.17e-11
Median	-1176.539
Maximum	15343.17
Minimum	-7909.530
Std. Dev.	6194.342
Skewness	0.904941
Kurtosis	3.090487
Jarque-Bera	5.473106
Probability	0.064793

Regression with Eviews - Residual Assumptions

- If the residuals are not normal and this is due to some outliers, use dummy variables to remove the outliers (In some cases, however, it may be that the extreme values in the data provide the most useful information about values of some of the coefficients and/or provide the most realistic guide to the magnitudes of forecast errors)
- Nonlinear transformation of variables might cure this problem
- Use other estimation method or other econometric model