CAMPUS OF
INTERNATIONAL
EXCELLENCE

Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Task T7

Residual analysis

Pilar González and Susan Orbe

Dpt. Applied Economics III (Econometrics and Statistics)

T7. Residual analysis.

Regression models.

1. `umbrellas.gdt`.

$$U_t = \beta_1 + \beta_2\,P_t + \beta_3\,T_t + \beta_4\,WW_t + \beta_5\,time_t + u_t \qquad (1)$$

2. `vehicles.gdt`.

$$RV_t = \beta_1 + \beta_2\,BOP_t + \beta_3\,IPIR_t + \beta_4\,time_t + u_t \qquad (2)$$

3. `wages.gdt`.

$$W_i = \beta_1 + \beta_2\,ED_i + \beta_3\,EX_i + \beta_4\,T_i + \beta_5\,F_i + u_i \qquad (3)$$

4. `cottages.gdt`.

$$RP_i = \beta_1 + \beta_2\,NR_i + \beta_3\,BP_i + u_i \qquad (4)$$

T7. Residual analysis.

Descriptive analysis.

Estimate by OLS the four regression models. For each one:

a. Save the OLS residuals.

b. Compute the main descriptive statistics of the residuals.

c. Draw a boxplot of the residuals.

d. Estimate the density function of the residuals. If the number of observations is not enough to estimate the density function, plot the frequency distribution.

e. Comment on the results obtained.

T7. Residual analysis.

Autocorrelation.

f. Consider the results of estimating models (1) and (2):

    f.a. Plot the OLS residuals against time.

    f.b. Comment on the graph. Does it suggest the presence of autocorrelation in the error term? Why?

T7. Residual analysis.

Heteroskedasticity.

g. Consider the results of estimating model (3):

  g.a. Plot the OLS residuals by observation.

  g.b. Plot the OLS residuals against education.

  g.c. Plot the OLS residuals against experience.

  g.d. Plot the OLS residuals against tenure.

  g.e. Comment on the results obtained. Do these graphs suggest the presence of heteroskedasticity in the error term? Why?

T7. Residual analysis.

Heteroskedasticity.

h. Consider the estimation results of model (4):

   h.a. Plot the OLS residuals by observation.

   h.b. Plot the OLS residuals against BP.

   h.c. Plot the OLS residuals against NR.

   h.d. Do these graphs suggest the presence of heteroskedasticity in the error term? Why?