Computational Econometrics with R

Heteroskedasticity in Multiple Regression Analysis: Labour Demand

We estimate a simple econometric model to explain labour demand of German firms. To do so, we use a cross-sectional data set of 6326 firms that includes information for 2005 on the total number of employees, their average wage, the amount of capital and a measure of output. The following variables are available in `labour_germany.csv`:

Variable	Explanation
LABOUR	total employment (number of workers)
OUTPUT	added value (in million euro)
WAGE	total wage costs divided by number of workers (in 1000 euro)
CAPITAL	total fixed assets (in million euro)

1. Preliminaries

1.1 Import the file `labour_germany.csv` into R.

1.2 Generate four series containing the natural logarithm of the variables LABOUR, WAGE, OUTPUT and CAPITAL.

1.3 Assuming the production function to be of the Cobb-Douglas type, the labour demand function can be represented by the following log-linear model:

$$\begin{aligned}\log(\text{LABOUR})_i &= \beta_0 + \beta_1 \log(\text{WAGE})_i + \beta_2 \log(\text{OUTPUT})_i + \\ &\quad \beta_3 \log(\text{CAPITAL})_i + \varepsilon_i. \end{aligned} \tag{1}$$

Estimate equation (1) by OLS and interpret your results statistically and economically.

2. Visual Detection of Heteroskedasticity

2.1 Generate a series object containing the predicted values of the variable log(LABOUR)

$$\begin{aligned}\log(\widehat{\text{LABOUR}})_i &= \hat{\beta}_0 + \hat{\beta}_1 \log(\text{WAGE})_i + \hat{\beta}_2 \log(\text{OUTPUT})_i + \\ &\quad \hat{\beta}_3 \log(\text{CAPITAL})_i \end{aligned} \tag{2}$$

2.2 Save the residuals of the recent regression

$$\hat{\varepsilon}_i = \log(\text{LABOUR})_i - \log(\widehat{\text{LABOUR}})_i \tag{3}$$

2.3 Provide a scatter plot of (2) versus (3).

2.4 What do you conclude for the assumption of homoskedastic error terms on the basis of the residual plot from the previous task? What consequences follow for the OLS estimates and inference in the presence of heteroskedasticity?

3. Breusch-Pagan Test for Heteroskedasticity

3.1 Create a series that contains the squared residuals from regression (1).

3.2 To perform a Breusch-Pagan test for heteroskedasticity estimate the following auxiliary regression

$$\begin{aligned} \hat{\varepsilon}_i^2 &= \delta_0 + \delta_1 \log(\text{WAGE})_i + \delta_2 \log(\text{OUTPUT})_i + \\ &\quad \delta_3 \log(\text{CAPITAL})_i + \nu_i \end{aligned} \tag{4}$$

by OLS.

3.3 Under the null hypothesis of homoskedasticity, i.e. $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$, the Breusch-Pagan test statistic

$$\text{BP} = n \cdot R^2_{\hat{\varepsilon}^2}$$

is asymptotically χ^2 distributed with k degrees of freedom; $R^2_{\hat{\varepsilon}^2}$ denotes the coefficient of determination of the auxiliary regression (4) while n is the number of observations and k the number of estimated coefficients except the constant. Compute the Breusch-Pagan test statistic.

3.4 Compute the p-value corresponding to the Breusch-Pagan test statistic from task 3.3 and interpret your test result.

4. Heteroskedasticity-Robust Inference

4.1 One remedy in the case of a violation of the homoskedasticity assumption is to use heteroskedasticity- robust standard errors for inference. Estimate the regression model (1) again with robust standard errors. Compare your results to those of task 1.3.