HSE and University of London Double Degree Programme in Data Science and Business Analytics

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Seminar 3. Pair linear regression

Problem 1.

Recall Gauss Markov theorem. What problems with can lead to violation of GMT?

Problem 2.

Let Y_t be the aggregated consumption and X_t be aggregated disposable income. The regression model is

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \quad t = 1, \ldots, n$$

	t	Y_t	X_t		t	Y_t	X_t
1986	1	152	170	1992	7	177	200
1987	2	159	179	1993	8	179	207
1988	3	162	187	1994	9	184	215
1989	4	165	189	1995	10	186	216
1990	5	170	193	1996	11	190	220
1991	6	172	199	1997	12	191	225

$$\begin{array}{rr} n = 12, & \bar{X} = 200 \\ \sum X = 2400, & \bar{Y} = 173.92 \\ \sum Y = 2087, & \sum Y^2 = 364741 \\ \sum X^2 = 483236, & \sum XY = 419782 \end{array}$$

(a) (5 points) Calculate coefficient for pair linear regression, give an interpretation.

(b) (10 points) Calculate $s^2, s^2_{\hat{\alpha}}, s^2_{\widehat{\beta}}$.

(c) (5 points) Perform t-tests for both coefficients ($\alpha = 0.05$). Explain the results.

(d) (5 points) Calculate 95%-confidence intervals.

(e) (5 points) Calculate R^2.

Problem 3.

We are interested in investigating the factors governing the precision of regression coefficients. Consider the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

with OLS parameter estimates $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$.

Under the Gauss Markov assumptions, we have

$$\operatorname{Var}\left(\hat{\beta}_2\right) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n \left(X_{2i} - \bar{X}_2\right)^2} \times \frac{1}{1 - r_{X_2 X_3}^2},$$

where σ_ε^2 is the variance of ε and $r_{X_2 X_3}$ is the sample correlation between X_2 and X_3.

(a) Provide four factors that help with obtaining more precise parameter estimates for, say, $\hat{\beta}_2$.

Problem 4.

Consider a model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

(a) Derive confidence interval for prediction $\hat{y}_{n+1}$ given new data x_{n+1}.

Hint: Check if $\hat{y}_{n+1}$ unbiased then consider the variance of $\hat{y}_{n+1} - y_{n+1}$.

(b) Say, we have data for 10 observations:

$$\sum x_i = 20, \sum x_i^2 = 50, \sum y_i = 8, \sum y_i^2 = 26, \sum x_i y_i = 10$$

Find best linear unbiased prediction for y_{11} given that $x_{11} = 5$ and estimate it's precision by constructing confidence interval.

Problem 5.

Suppose that the units of measurement of X are changed so that the new measure, X^*, is related to the original one by $X_i^* = \mu_1 + \mu_2 X_i$. How do the estimates of the slope coefficient and of the intercept change?

Problem 6.

A researcher has international cross-sectional data on aggregate wages, W, aggregate profits, P, and aggregate income, Y, for a sample of n countries. By definition,

$$Y_i = W_i + P_i$$

The regressions

$$\hat{W}_i = \hat{\alpha}_1 + \hat{\alpha}_2 Y_i$$
$$\hat{P}_i = \hat{\beta}_1 + \hat{\beta}_2 Y_i$$

are fitted using OLS regression analysis. Show that the regression coefficients will automatically satisfy the following equations:

$$\hat{\alpha}_2 + \hat{\beta}_2 = 1$$
$$\hat{\alpha}_1 + \hat{\beta}_1 = 0$$

Explain intuitively why this should be so.