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

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Class 16: Multicollinearity

Problem 1

(i) Derive variance of vector of OLS coefficient estimators $\hat{\beta}$.

(ii) In case of pair linear regression calculate the variance of estimators

Problem 2

(i) For multiple linear regression state which factors inflate variance of the estimator $\hat{\beta}_j$

$$D\left(\hat{\beta}_j\right) = \frac{\sigma^2}{\text{TSS}_j\left(1 - R_j^2\right)},$$

where TSS $_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2, R_j^2$ - coefficient of determination when regressing x_j on all other features.

(ii) What are the consequences of multicollinearity?

(iii) Explain how VIF works to determine multicollinearity presence?

(iv) How to fight multicollinearity?

Problem 3

Let a regression equation be:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t : t = 1, 2 \ldots, T$$

(i) What are normal equations of OLS for the multiple linear regression in covariance form? Use sample variance and sample covariance functions to represent normal equations of OLS without taking derivatives.

(ii) Derive formulas for regression coefficients in the case of two regressors?

$$b_2 = \frac{\text{Cov}(X_2,Y)\,\text{Var}(X_3) - \text{Cov}(X_3,Y)\,\text{Cov}(X_2,X_3)}{\text{Var}(X_2)\,\text{Var}(X_3) - [\text{Cov}(X_2,X_3)]^2}$$
$$b_3 = \frac{\text{Cov}(X_3,Y)\,\text{Var}(X_2) - \text{Cov}(X_2,Y)\,\text{Cov}(X_2,X_3)}{\text{Var}(X_2)\,\text{Var}(X_3) - [\text{Cov}(X_2,X_3)]^2}$$

Problem 4

A regression of consumption (C) on income (Y) and unemployment (U) (all variables are index numbers) using annual data 1961-82 for the UK produced the following results:

$$\hat{C}_i = 17880 + 0.7527Y_i + 0.930U, R^2 = 0.992$$
$$(2817.0) \quad (0.026) \quad (0.798)$$

(figures in brackets are standard errors) with a table of correlation coefficients between variables of:

	C	Y	U
C	1.00	0.996	0.783
Y	0.996	1.00	0.771
U	0.783	0.771	1.00

(a) Interpret the estimated equation and test their significance. From the analysis of the table of correlation coefficients do you think that multicollinearity is a problem? Explain how high correlation between explanatory variables influence the significance of regression coefficients.

(b) If consumption was to be regressed on income alone what do you think would happen to the coefficient on income? Explain.

(c) The researcher supposed that influence of income on consumption cannot be negative, while the influence of the unemployment on consumption cannot be positive. Does this assumptions change your conclusion on the significance of coefficients? Are both slope coefficients taken together significant? What assumptions have you used for the significance tests?