Elements of Econometrics. 2022-2023.
Class 5. Multiple Regression Continued

RESTRICTIONS

Problem 1. (UoL and ICEF Exams)
Let a regression equation be:

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

Outline briefly, how would you test

(a) $\beta_2 = 0$

(b) $\beta_2 = 1$

(c) jointly β_2 and β_3 are significantly different from zero.

(d) $\beta_2 = \beta_3$

(e) $\beta_2 + \beta_3 = 1$

(f) $\beta_2 + \beta_3 = 1$ against $\beta_2 + \beta_3 < 1$

PREDICTION

Problem 2. Suppose that $Y_1, Y_2, ..., Y_n$ are independent normal variables with $E(Y_i) = \beta_1 + \beta_2 X_i$ and $V(Y_i) = \sigma_u^2$ for $n = 1, 2, ..., n$.

a) Prove that $\text{cov}(\overline{Y}, \hat{\beta}_2) = 0$.

b) Prove that $\text{cov}(\overline{Y}, \hat{\beta}_1) = 0$.

Problem 3. Suppose that $Y_1, Y_2, ..., Y_n$ are independent normal variables with $E(Y_i) = \beta_1 + \beta_2 X_i$ and $V(Y_i) = \sigma_u^2$ for $n = 1, 2, ..., n$. Prove that $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = \dfrac{-\overline{X}\sigma_u^2}{\sum (X_i - \overline{X})^2}$.

Problem 4. Let X^* be arbitrary value of explanatory variable X and $Y^* = \beta_1 + \beta_2 X^* + u^*$.
Expected value of $Y^* = \beta_1 + \beta_2 X^* + u^*$ is $EY^* = \beta_1 + \beta_2 X^* + Eu^* = \beta_1 + \beta_2 X^*$.
We do not know it but we can get **predicted value** $\hat{Y}^* = \hat{\beta}_1 + \hat{\beta}_2 X^*$
Show that variance of the predicted value of the regression $\text{var}(\hat{Y}^*) = \text{var}(\hat{\beta}_1 + \hat{\beta}_2 X^*)$ is equal to $\left(\dfrac{1}{n} + \dfrac{(X^* - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right)\sigma_u^2$.

Problem 5.
Prove that $\text{cov}(Y^*, \hat{Y}^*) = 0$

Problem 6.

a) What is prediction error (***PE***)?

b) What is expected value of prediction error?

c) What is variance of prediction error?

d) Show that the variance of the prediction error $\text{var}(PE)$ is equal to $\left(1 + \dfrac{1}{n} + \dfrac{(X^* - \overline{X})^2}{\sum (X_i - \overline{X})^2}\right)\sigma_u^2$.

FINAL CHALLENGE (Problems from real examinations)

Problem 7. (UoL Exam) 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:

$$\hat{C}_t = 17880 + 0.7527 Y_t + 0.930 U_t \quad R^2 = 0.992$$
$$(2817.0)\ (0.026) \qquad (0.798)$$

(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?