Class 4. Multiple regression

Problem 1. Train yourself to prove the properties of 'auxiliary coefficients'

1) $\sum_{i=1}^{n} a_i = 0$

2) $\sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{1}{\sum_{i=1}^{n} x_i^2}$ where $x_i = X_i - \overline{X}$

3) $\sum_{i=1}^{n} a_i X_i = 1$

Problem 2. Let the regression be $Y_t = \beta_1 + \beta_2 X_t + u_t;\ t = 1, 2, \ldots, T$ where $E(u_t) = 0;\ E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$. Prove that $b_2 = \hat{\beta}_2 = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sum(X_i - \overline{X})^2}$ is unbiased estimator of β_2

Now the same in concise form $b_2 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$

<u>**Decomposition of coefficient estimators**</u>

Problem 3. Let the regression be $Y_t = \beta_1 + \beta_2 X_t + u_t;\ t = 1, 2, \ldots, T$ where $E(u_t) = 0;\ E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$. Derive formula for decomposition of regression estimator $b_2 = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ into fixed and random components: $b_2 = \beta_2 + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}$

Problem 4. Under the same assumptions prove that OLS estimator of the slope coefficient is unbiased.

Problem 5. Let a regression equation be:

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

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**.

Problem 6. Derive formulas for regression coefficients in the case of two regressors?

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

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

Problem 7. By definition $R^2_{adj} = \overline{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$. Compare it with the definition of conventional R^2, what is the difference? Comment on the meaning of $(n-k)$ and $(n-1)$ in this formula.

Problem 8. Show that $\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$.

Problem 9. Show that $\overline{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2) = \frac{ESS}{TSS} - \frac{k-1}{n-k} \cdot \frac{RSS}{TSS}$.

Problem 10. The $\overline{R}^2$ coefficient increases if and only if the absolute value of t-statistic of the added variable coefficient is greater than 1.

PROBLEMS FROM EXAM

Problem 1 (UoL Exam). The following equation was estimated by Ordinary Least Squares using 37 annual observations of UK aggregate data. The dependent variable ($cloth_t$) is the log of expenditure on clothing at 1995 prices, yd_t is the log of aggregate disposable income at 1995 prices, pc_t is the log of the price of clothing relative to all consumer prices, ps_t is the log of the price of shoes relative to all consumer prices.

$$\underset{}{cloth_t} = \underset{(1.531)}{-3.256} + \underset{(0.118)}{1.021} yd_t - \underset{(0.132)}{0.240} pc_t - \underset{(0.185)}{0.429} ps_t + e_t$$

standard errors in brackets, e_t is an OLS residual. $R^2 = 0.992$

(a) Test the hypothesis that the coefficient of yd_t is one.

(b) Construct a 95% confidence interval for the coefficient pc_t.

(c) Test the hypothesis that all slope coefficients in the equation above are zero. Give any assumptions which your results in b) and c) require.

Problem 2 (ICEF Exam). Two students A and B are trying to answer the following question: what is better for future earnings – to study or to work, and so to get working experience. They collect data on 28 people working in different companies on their schooling S_i (in years), their working experience W_i (also in years) and current hourly earnings $EARN_i$ (in dollars). They also calculate the total number of years of active life spent on work or study $A_i = S_i + W_i$. It is assumed that one can not work and study at the same time.

The student A runs the following equation

$$\widehat{EARN}_i = \underset{(17.48)}{-22.96} + \underset{(0.89)}{2.44} \cdot S_i + \underset{(0.85)}{0.92} \cdot W_i \qquad R^2 = 0.26 \qquad \textbf{(1)}$$

The student B using the same data estimates another equation

$$\widehat{EARN}_i = \underset{(17.48)}{-22.96} + \underset{(0.89)}{2.44} \cdot A_i - \underset{(0.68)}{1.52} \cdot W_i \qquad R^2 = 0.26 \qquad \textbf{(2)}$$

(a) . Give interpretation to the coefficients of equation (1) and to all coefficients except that of W_i in the second equation. Explain why the coefficient of S_i in equation **(1)** is equal to the coefficient of A_i in equation **(2).** Explain why the constant terms in equations **(1)** and **(2)** are equal.

(b) The student A claims that spending one extra year to study is 2.5 times more useful for future earnings than spend it on work. Is he right? Give interpretation to the coefficient of W_i in the equation **(2)** and explain why it has negative sign. Is it possible to evaluate this coefficient using information from equation **(1)**?

(c) Evaluate the significance of the coefficients. Evaluate the significance of the equation **(1)** using F-statistic**.** Is it possible to get F-statistic for equation **(2)** without calculation? Why the equations **(1)** and **(2)** have the same R^2 ? Why the standard error of the coefficient of S_i in the equation **(1)** is equal to the standard error of the coefficient of A_i in the equation **(2)**?