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

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Class 12: Omitted variable bias

Problem 1

(a) Recall the for GLS

(b) Let's consider the linear model $y = X\beta + \varepsilon$, for which all the premises of the classical linear multiple regression model are met with one exception: the variance of the random error is directly proportional to the square of some known variable

$$\text{var}(\varepsilon_i) = \sigma_i^2 = \sigma_0^2 z_i^2 > 0.$$

1.. Find the covariance matrix of the random error vector for the presented model.

2. Write down in explicit form the formula for GLS estimation of the vector of model coefficients.

(c) Let's consider a linear model $y = X\beta + \varepsilon$, for which the variance of random errors is constant, but so-called first-order autocorrelation is observed:

$$\varepsilon_i = \rho \cdot \varepsilon_{i-1} + u_i$$

where u_i are independent and identically distributed random variables with variance $\sigma_u^2; \rho \in (-1, 1)-$ autocorrelation coefficient.

1. Find the covariance matrix of the random error vector for the presented model.

2. Write down an explicit formula for the GLS estimation of the vector of model coefficients, assuming that the coefficient ρ is known.

Problem 2

Consider a true model:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 w_i + \varepsilon_i$$

And the estimated model is:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

Say, regressors are deterministic.

– Derive the formula for the omitted variable bias for $\hat{\beta}_2$.

– What defines the sign of OVB?

– When OVB equals to 0?

Problem 3

A researcher has data on output per worker, y, and capital per worker, k, both measured in thousands of pounds, for 50 firms in the manufacturing sector of the U.K. for 2016. She

hypothesizes that output per worker depends on capital per worker and perhaps also the technological sophistication of the firm, tech:

$$y = \beta_1 + \beta_2 k + \beta_3 tech + u$$

where u is a disturbance term. She is unable to measure tech and decides to use expenditure per worker on research and development in 1998, exp, as a proxy for it.

(a) What do you mean by good or poor proxy?

– Explain the consequences of using exp as a proxy for tech if it is a good proxy.

– Explain the consequences of using exp as a proxy for tech if it is a poor proxy.

The researcher fits the following regressions (standard errors in parentheses):

$$\begin{array}{rl} \hat{y} = & 1.02 + 0.32k \quad R^2 = 0.749 \\ & \quad (0.45)(0.04) \end{array}$$

$$\begin{array}{rl} \hat{y} = & 0.34 + 0.29k + 0.05\,\text{exp} \quad R^2 = 0.750 \\ & (0.61)(0.22)(0.15) \end{array}$$

The correlation coefficient for k and exp was 0.92 .

(b)

– Discuss these regression results assuming that y does depend on both k and tech.

– Discuss these regression results assuming that y depends only on k.

Problem 4

Explain the RESET test as a general test for functional form misspecification and discuss the drawbacks and advantages of this test. In your answer consider the following multiple linear regression model:

$$y_i = \gamma_1 + \gamma_2 x_{2i} + \gamma_3 x_{3i} + u_i, i = 1, \ldots, n,$$

where $x_{2,}$ and x_{3i} are exogenous variables known to affect $E(y_i)$.