

**Elements of Econometrics. 2022-2023.**  
**Class 9. Model Specification**

**Problem 1 (UoL Exam).** (a) Explain what you understand by omitted variable bias using regression model without intercept.

Consider two equations

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \quad (1)$$

$$y_t = \beta_1 x_{1t} + v_t \quad (2)$$

(b) Let (1) be a **true model** while (2) be a **false model**. A researcher, using ordinary least squares (*OLS*), estimates  $\beta_1$  from the **false model**. Examine the properties of *OLS* estimator of  $\beta_1$  in equation (2).

(c) Let now (1) be a **false model** while (2) be a **true model**. A researcher, using ordinary least squares (*OLS*), estimates  $\beta_1$  from the **false model** (1). Examine the properties for this *OLS* estimate of  $\beta_1$ .

**Problem 2 (ICEF Exam).** □ Explain why omitting explanatory variable from the regression equation can lead to the violation of Gauss-Markov conditions so standard errors and all tests become invalid.

□ What are consequences of omitting explanatory variable from the regression equation for the estimation regression equations.

**Problem 3 (Dourgerty textbook 1<sup>st</sup> edition).** A social scientist thinks that the level of activity in the shadow economy,  $Y_t$ , depends either positively on the level of the tax burden,  $X_t$ , or negatively on the level of government expenditure to discourage shadow economy activity,  $Z_t$ . The value  $Y_t$  may also depend on the  $X_t$  and  $Z_t$  simultaneously. There are annual time series data for 20 years, where the value of  $Y_t$ ,  $X_t$  and  $Z_t$  are measured in the same units. Sociologist builds regression dependence (1):  $Y_t$  only on the value of  $X_t$ , (2):  $Y_t$  only on the value of  $Z_t$  and (3):  $Y_t$  from both variables  $X_t$  and  $Z_t$ , in relation to each city, with the following results (in parentheses are standard errors). Having carried out the appropriate statistical tests, write a short report advising the social scientist how to interpret these results.

	Constant	Estimated coefficients		$R^2$
		$X_t$	$Z_t$	
City A				
1	315.7 (18.5)	1.54 (0.97)	-	0.12
2	128.6 (50.9)	-	-0.96 (0.06)	0.94
3	218.0 (76.6)	2.85 (0.25)	-1.21 (0.03)	0.99
City B				
1	197.6 (16.8)	2.86 (0.25)	-	0.88
2	512.2 (202.6)	-	-0.05 (0.08)	0.02
3	230.8 (82.5)	2.94 (0.27)	-0.01 (0.03)	0.88

**Question 4 (from the lecture).** Let  $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$  – true model,  $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2$  – estimated model. Is the estimator of  $\beta_1$  biased? If yes, what is the value of bias?

**Question 5 (from the lecture).** Let  $Y = \beta_1 + \beta_2 X_2 + u$  – true model,  $\hat{Y} = \hat{\beta}_2 X_2$  – estimated model. Is the estimator of  $\beta_2$  biased? If yes, what is the value of bias?

## FINAL CHALLENGE

**Question 6 (UoL Exam).** 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:

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

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

### Question 7 (ICEF Exam).

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):

$$\hat{y} = 1.02 + 0.32k \quad R^2 = 0.749$$

(0.45) (0.04)

$$\hat{y} = 0.34 + 0.29k + 0.05exp \quad R^2 = 0.750$$

(0.61) (0.22) (0.15)

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  $k$ .