Worksheet **3**
The Multiple Regression Model

Note: In those problems that include estimations and have a reference to a data set the students should check the outputs obtained with Gretl.

1. Let the model be
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$
where $E(\varepsilon|X_1, X_2) = 0$, and assume that we have a sample of size n.

   a) Derive the first order conditions of the OLS estimators $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$ for the β.coefficients

   b) Show that
$$\begin{aligned} \widehat{\beta}_0 &= \overline{Y} - \widehat{\beta}_1 \overline{X}_1 - \widehat{\beta}_2 \overline{X}_2 \\ \widehat{\beta}_1 &= \frac{1}{D}(s_{22}s_{1y} - s_{12}s_{2y}) \\ \widehat{\beta}_2 &= \frac{1}{D}(s_{11}s_{2y} - s_{12}s_{1y}) \end{aligned}$$
   where
$$\begin{array}{lll} s_{11} = \frac{1}{n}\sum_i x_{1i}^2, & s_{12} = \frac{1}{n}\sum_i x_{1i}x_{2i} = s_{21}, & s_{1y} = \frac{1}{n}\sum_i x_{1i}y_i \\ s_{22} = \frac{1}{n}\sum_i x_{2i}^2, & & s_{2y} = \frac{1}{n}\sum_i x_{2i}y_i \\ D = s_{11}s_{22} - s_{12}^2. & & \end{array}$$
   and $y_i = Y_i - \overline{Y}$, $x_{1i} = X_{1i} - \overline{X}_1$, $x_{2i} = X_{2i} - \overline{X}_2$.

   c) What happens when $D = 0$? Interpret the result.

   d) Show that, when $s_{12} = 0$, the estimator $\widehat{\beta}_1$coincides with the estimator $\widehat{\gamma}_1$ in the simple regression
$$\widehat{Y} = \widehat{\gamma}_0 + \widehat{\gamma}_1 X_1,$$
   and interpret the results.

2. Which of these situations, if any, does not comply with the assumptions of the classical regression model?

   a) The variable X_2 is the reciprocal of the variable X_1

   b) The variable X_2 is the variable X_1 squared

   c) The variable X_1 is an artificial variable that takes the value 1 for females and 0 for males and the variable X_2 is an artificial variable that takes the value 1 for males and 0 for females.

3. Even though wine is a consumption good, vintage wines can be considered as an investment good given their characteristics. In particular, we have data on the auction prices of thousands of red Bordeaux vintage wines from 1952 to 1980. These wines are stored for a considerable period of time before being consumed, which leads to an increase in the price given the cost of storage. This entails an opportunity cost given the possibility of investing in other alternatives. Our data file `BORDEAUX.GDT` contains information on LPR (logarithm of the

price of wine), *lluvinv* (Amount of rainfall in the winter preceding the harvest), *tempmed* (Average degree Centigrade while the grapes ripe), *lluvcos* (Amount of rainfall while the grapes ripe), *edad* (Number of years since the harvest)

a) Estimate the linear projection of *lpr* on *edad*. Given the results, what would be the annual profitability from keeping the wine?

b) Carry out the multiple regression of *lpr* on *edad*, *lluvinv*, *lluvcos*, *tempmed*. How does the estimation of the profitability rate change? How do you explain this difference? (**Clue**: Which factors can affect the quality of wine?).

4. Consider Y = logarithm of real money demand, X_1 = logarithm of real GDP, X_2 = logarithm of the interest rate of Treasury bills. Consider the following regression results:

$$\begin{array}{rccc} \widehat{Y} = & 2{,}3296 & +0{,}5573X_1 & -0{,}2032X_2 \\ & (0{,}2054) & (0{,}0264) & (0{,}0210) \end{array}$$

$$R^2 = 0{,}927 \quad s = 0{,}048 \quad \overline{Y} = 6{,}63$$

$$\begin{array}{rcc} \widetilde{Y} = & 2{,}9967 & +0{,}4356X_1 \\ & (0{,}3657) & (0{,}0438) \end{array}$$

$$\widetilde{R}^2 = 0{,}733 \quad \widetilde{s} = 0{,}091 \quad \overline{Y} = 6{,}63$$

$$\widehat{X}_2 = -3{,}2839 \quad +0{,}5988X_1$$

Given these results, derive the slope of the other "short regression", i.e., the regression of Y on X_2. Hint: write the ommitted variable rule swchitng the roles of X_1 and X_2. The slope and the R^2 of the auxiliary regression can be used to calculate the slope of the *other auxiliary regression*.

5. Consider the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon,$$

where $E(\varepsilon | X_1) = 0$, we have obtained the following OLS estimates:

$$\begin{aligned} \widehat{Y} &= 2{,}613 + 0{,}30X_1 - 0{,}090X_1^2 \\ & \quad (0{,}429) \quad (0{,}14) \quad (0{,}037) \\ n &= 32, \quad R^2 = 0{,}1484 \end{aligned}$$

a) Given the results, from what value of X_1 is the causal effect of X_1 on Y negative? Justify.

b) We want to test whether we should keep the quadratic term in the model. Write the null and the alternative hypotheses in terms of the model parameters, and write the corresponding test statistic and its approximate distribution. If the available information suffices, implement the test and establish your conclusion.

6. We are interested in a equation to explain a CEO wage as a function of the firm's annual sales, the firm's bond yield (*roe*, in percentage) and the firm's equity value (*ros*, in percentage),

$$log(salary) = \beta_0 + \beta_1 log(sales) + \beta_2 roe + \beta_3 ros + u$$

a) Specify, in terms of the model parameters, the null hypothesis that, once that *sales* and *roe* are accounted for, *ros* does not influence the CEO wage. As alternative hypothesis, consider that, other things equal, a higher equity value tends to increase the CEO wage.

b) Consider the following OLS results:

$$\begin{aligned} \widehat{log(salary)} &= \underset{(0{,}32)}{4{,}32} + \underset{(0{,}035)}{0{,}280} log(sales) + \underset{(0{,}0041)}{0{,}0174} roe + \underset{(0{,}00054)}{0{,}00024} ros \\ n &= 209 \qquad R^2 = 0{,}283 \end{aligned}$$

In what predicted percentage would the wage increase if *ros* increased by 50 points?

c) Test, at the 10 % significance level, the null that *ros* has no effect on *salary*, against the alternative that it has a positive effect.

d) Would you include *ros* in the final model to explain the CEO wage as a function of the firm performance? Justify.

7. Consider the following OLS estimation:

$$\begin{aligned} \widehat{sleep} &= \underset{(235{,}22)}{3840{,}83} - \underset{(0{,}018)}{0{,}163} totwrk - \underset{(5{,}86)}{11{,}71} educ - \underset{(11{,}21)}{8{,}70} age + \underset{(0{,}134)}{0{,}128} age2 + \underset{(34{,}33)}{87{,}75} male \\ n &= 706 \qquad R^2 = 0{,}117 \end{aligned}$$

The variables *sleep* and *totwrk* measure, respectively, the minutes devoted to sleep and to work during the week, *educ* and *age* denote individual's years of education and individual's age in years, and *male* is a binary variable which takes on value 1 if the individual is a man and 0 otherwise.

a) Keeping everything else constant, is there evidence that men sleep more than women? Is it a strong evidence?

b) Is the dichotomy between working and sleeping significant? What is the estimate of such dichotomy?

c) What additional regresion is needed to test the null that the age does not influence the time devoted to sleep (keeping everything else constant)?

8. The file RELOJES.GDT contains data from an annual auction of antique clocks, organized by the German company Triberg Clock. We have considered the following model:

$$P = \beta_0 + \beta_1 A + \beta_2 C + \beta_3 A^2 + \beta_4 C^2 + \beta_5 (A \times C) + \varepsilon, \qquad \text{(R1)}$$

where P is the price, in hundred euros, of the winner bid, A is the clock's age (in years) and C is the number of bidders. Besides, the error term satisfies, for any age and any number of bidders, $E(\varepsilon | A, C) = 0$ and $V(\varepsilon | A, C) = \sigma^2$.
We assume that the population characteristics coincide with their corresponding sample counterparts, and that we can rely on asymptotic results to make approximate inference. Be aware that certain population characteristics, namely the range of values taken by each variable of interest, can be needed to answer some questions.

a) Compute the descriptive statistics of P, A and C.

b) Estimate model (R1) by OLS. Propose and obtain consistent estimates of $V(P)$ and $V(P|A,C)$.

c) Given the results on (R1), compute the *ceteris paribus* effect of age on the clock's price. Is that effect constant? Justify.

d) Consider the following statement: "The price does not depend on age nor the number of bidders". Write the null (H_0) and the alternative (H_1) hypotheses in terms of the parameters of model (R1). Write the test statistic and indicate its approximate distribution under H_0. If possible, implement the test and establish your conclusion, If additional information is missing, indicate what you needed.

e) Consider the following specification:

$$P = \delta_0 + \delta_1 (4A + C) + \delta_2 A^2 + \delta_3 C^2 + \delta_4 (A \times C) + \varepsilon. \qquad \text{(R2)}$$

Estimate (R2) by OLS.

f) Consider the null $H_0 : \beta_1 = 4\beta_2$ vs the alternative $H_1 : \beta_1 \neq 4\beta_2$. Write the constrained model (under H_0), and relate it with model (R2).

9. Consider the linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i.$$

Explain how you would test the following hypotheses:

a) $\beta_1 = 0$.

b) $\beta_1 = 0$ and $\beta_4 = \beta_5$

c) $\beta_1 = 0$, $\beta_3 = 2$, and $\beta_4 = \beta_5$.

10. Consider the multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon, \qquad (1)$$

that satisfy the assumptions 1 to 4. We want to test the null hypothesis $H_0 : \beta_1 - 3\beta_2 = 1$.

a) Let $\widehat{\beta}_1$ and $\widehat{\beta}_2$ be the OLS estimators of β_1 and β_2 respectively. Express $V\left(\widehat{\beta}_1 - 3\widehat{\beta}_2\right)$ in terms of $V\left(\widehat{\beta}_1\right)$, $V\left(\widehat{\beta}_2\right)$ and $C\left(\widehat{\beta}_1, \widehat{\beta}_2\right)$.

b) Write down the t-statistic to test $H_0 : \beta_1 - 3\beta_2 = 1$.

c) Defining $\theta = \beta_1 - 3\beta_2$ and its estimate (based on the OLS estimators of β_1 and β_2), $\widehat{\theta} = \widehat{\beta}_1 - 3\widehat{\beta}_2$, write down an equivalent specification of (1) where β_0, θ, β_2 and β_3 are present, and it allows us to directly obtain $\widehat{\theta}$ and its standard error from a data sample.

d) Explain an alternative strategy to that of section (b) to test $H_0 : \beta_1 - 3\beta_2 = 1$.

11. Consider the following specification for a production function:

$$y_i = \beta_0 + \beta_1 l_i + \beta_2 k_i + \varepsilon_i \quad (i = 1, \ldots, n), \qquad (2)$$

where y = log of output, l = log of labor input, k = log of capital. Suppose, also, that $E(\varepsilon_i | l_i, k_i) = 0$ for any combination l_i, k_i.
We want to test for constant returns to scale, i.e., $\beta_1 + \beta_2 = 1$. Explain how to implement the test:

a) If we have the OLS estimates of the regression (2) and its corresponding variance-covariance matrix of the estimated parameters.

b) If we have the OLS estimates of the regression of $(y_i - k_i)$ on a constant, $(l_i - k_i)$ and k_i.

c) If we have the RSS (sum of squared residuals) of the OLS estimation of (2) and of the OLS estimation of the regression of $(y_i - k_i)$ on a constant and $(l_i - k_i)$.