



**Barcelona School of Economics**

## Assignment 4

Foundations of Econometrics

Group 7



October 21, 2025

## Question 1

Class slide 3(33) (Unit 3) illustrated, via simulation, the effects of collinearity. The script used to generate the sample is included in file `data33.R`.

### Part (a)

- (i) Estimate the regression model included in the slide, presenting OLS estimates and the 95% confidence intervals for each parameter; Include the output in your answer.

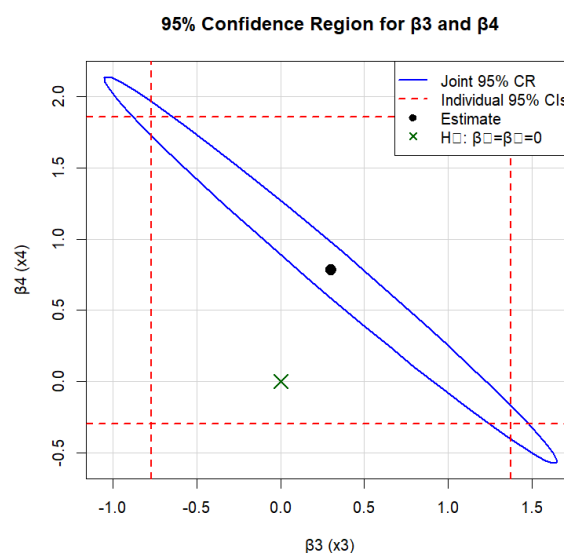
**Answer:**

**Table 1:** Regression Results with 95% CI

	Estimate	Std..Error	Lower	Upper
(Intercept)	7.9365	1.6426	4.5863	11.2867
x2	0.5953	0.0747	0.4429	0.7476
x3	0.2996	0.5264	-0.7740	1.3733
x4	0.7818	0.5277	-0.2945	1.8582

- (ii) Using `confidenceEllipse()` function, or equivalent, draw the 95% confidence region for parameters  $\beta_3$ ,  $\beta_4$ . Include also in the drawing the confidence intervals for each parameter.

**Answer:**



**Figure 1:** 95% Confidence Region for  $\beta_3$  and  $\beta_4$

- (iii) Describe what the confidence region you just drew provides.

**Answer:**

- (iv) Use the figure of confidence intervals and confidence region to show the difference between testing statistical significance of regressors separately or jointly, and explain why this is so relevant under the presence of collinear regressors.

**Answer:**

## Part (b)

Modify the script used to estimate now the same regression with data generated from the same dgp but now using a sample of 3500 observations.

- (i) Surprised with how the estimates have changed? Rigorously justify.

**Answer:**

- (ii) Surprised of the change of the 95% confidence intervals? Rigorously justify.

**Answer:**

- (iii) Surprised of the change of the 95% confidence region for parameters  $\beta_3, \beta_4$ ? Rigorously justify.

**Answer:**

- (iv) Using the variance decomposition expression for  $\text{var}(\hat{\beta}_3 | X)$ , or  $\text{var}(\hat{\beta}_4 | X)$ , discuss why increasing  $n$  can explain the changes observed. Be specific.

**Answer:**

## Part (c)

Now, go back to the original script generating 35 observations, and modify the script so that now  $x_{i3} + 2x_{i4} = 0$ .

- (i) Run the script again. Include the output in your answer.

**Answer:**

- (ii) How many estimates did you get an estimate for  $\beta_3$ ? And for  $\beta_4$ ? You should be able to show, using the proper derivation, that in fact you got an infinite number of estimates for  $\beta_3$  and  $\beta_4$ .

**Answer:**

## Question 2

Data file `microsoft.csv` includes monthly data from May 1986 to April 2013 on  $RP_{msft}$  (excess return of Microsoft stock),  $RP_{s\&p}$  (excess return on the S&P500 portfolio),  $Dprod$  (variation of Industrial production),  $Dinflation$  (change in inflation rate),  $Dterm$  (change in interest rate) and  $m1$  (an indicator variable that takes value 1 if  $t$  is the month of January and 0 otherwise). The following regression is set to measure the reaction of the excess return of Microsoft stocks to changes in macroeconomic variables:

$$RP_{msft,t} = \beta_1 + \beta_2 RP_{s\&p,t} + \beta_3 Dprod_t + \beta_4 Dinflation_t + \beta_5 Dterm_t + \beta_6 m1_t + \epsilon_t$$

- (a) Estimate the model above by OLS. Present the complete output (estimates, standard errors, p-values) as your answer.

**Answer:**

- (b) The January effect states that on average, every else equal, the returns (or excess returns) are larger in the month of January than the rest of the months. Test, at  $\alpha = 1\%$ , the presence of the January effect using the exact  $t$ -test statistic. Would you say the data supports the presence of this effect?

**Answer:**

- (c) Aside from normality, list the assumptions needed to justify the use of the  $t$  test statistic. Justify your answer.

**Answer:**

- (d) For the test you performed in questions (2b), you were asked to use the  $t$  test statistic. Using this test requires, among others, for disturbances to be normally distributed. One of the available tests of normality of the distribution of a given random variable is the Jarque-Bera test. Under the null hypothesis of normality, the Jarque-Bera (JB) test statistic is:

$$JB \equiv \frac{n}{6} \left[ sk^2 + \frac{(kur - 3)^2}{4} \right] \stackrel{a}{\sim} \chi^2(2),$$

where  $sk$  is the sample coefficient of skewness of the variable and  $kur$  is its sample coefficient of kurtosis. Using a significance level of 1%, draw (by hand absolutely fine) the distribution of JB under  $H_0$  and the corresponding acceptance and rejection regions. Provide an intuition for the location of the acceptance region.

**Answer:**

- (e) Now, we want to test for the presence of normality in our disturbances using the JB test. Ideally, to test normality of disturbances, JB test should be applied to a sample of disturbances, but, given that they are unobservable, the JB test is usually applied to our OLS residuals. Explain how, if all the assumptions regarding the dgp

for consistency of OLS estimator are met, it would be justified for  $\hat{\epsilon}_t$ 's to take the place of  $\epsilon_t$ 's to perform the test.

**Answer:**

- (f) Perform the JB test on the OLS residuals. What do you conclude? Comment.

**Answer:**

- (g) Repeat the test performed in (2b) using the asymptotic  $T$ -test statistic. Use a 1% significance level.

**Answer:**

- (h) Is the use of asymptotic tests justified in this case? Rigorously argue.

**Answer:**

- (i) Consider the following statement: "Using the exact  $t$  test statistic leads to slightly more conservative inference, because we get larger acceptance regions and larger  $p$ -values than if we used the asymptotic version." Do you agree? Rigorously argue.

**Answer:**