Quiz 2 – Chapter 2: statistical uncertainty and inference in linear regressions

(Lucas Girard) – This version: 15 October 2023

Questions

The quizzes are provided as training to help you check your knowledge and understanding of the course; the course and the TD remain the only reference. The quizzes are not necessary, all the less so sufficient, to study Econometrics 1 but might nonetheless be helpful in your learning[1].

Some words about the quiz. As always henceforth and absent contrary indication, the notation used follows that of the course's slides.[2] *Beyond notations, try to be constantly aware of the nature of the objects they denote*: is it a non-stochastic parameter like β_0? Or an estimator, thus a random variable (since it is a function of the stochastic observations), like $\widehat{\beta}$? Likewise, be careful about the dimension of the objects (vectors, matrices, numbers) in computations.
Questions 1, 2, and 3 are about the limit distribution of the OLS estimator (under proper moment conditions) and its resulting precision.
Questions 4 to 10 relate to inference (tests and confidence intervals). In particular, Question 6 deals with the links between bilateral (also known as two-sided) and unilateral (a.k.a. one-sided) tests; it is a bit more advanced, but within the course's material (no asterisk mark here!). Question 9 is also a bit more advanced than the other questions from that set of questions focused on inference, which should be well understood.
Question 11 explores the interpretation of the Stata output of a `regress` command. It is a very classical question for mid-term and final exams in Econometrics 1. You will do that several times during small classes, but it is particularly interesting for you to prepare this question by yourself as you are expected to know that.
Question 12 is marked with an asterisk (as it relates to a slide also marked with an asterisk). However, in your formation as a statistician/econometrician, it is a crucial question conceptually about the difference between two types of objects both called "variance".
Question 13 is about the Gauss-Markov theorem (sub-question (a)). In the continuation of Question 12, it is also the opportunity to clarify the different types of "variance" seen in Chapter 2 (sub-question (b)).
Finally, Question 14, marked with an asterisk, relates to one part of the last section, "Particular cases*", of Chapter 2.
Bonne lecture ! Do not hesitate if you have any questions.

1 Limit distribution of the OLS estimator

With the notation of Chapter 2, if $(Y_i, X_i)_{i=1,\dots,n}$ is an i.i.d. sample with the same distribution as (Y, X), and $\mathbb{E}[\|X\|^2] < +\infty$, $\mathbb{E}[\varepsilon^2\|X\|^2] < +\infty$, and $\mathbb{E}[XX']$ invertible, then

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \xrightarrow[n\to+\infty]{d} \mathcal{N}(?_1,\ ?_2)$$

Give the name of that result, the values of $?_1$ and $?_2$, and the name of the quantity $?_2$.

2 Accuracy of the OLS estimator (simple linear regression)

We assume the homoskedasticity condition $\mathbb{E}[\varepsilon^2 XX'] = \mathbb{E}[\varepsilon^2]\,\mathbb{E}[XX']$ holds, and we consider a simple linear regression (SLR) of Y on D, with $D \in \mathbb{R}^{\Omega}$. $\widehat{\beta}_D$ is the OLS estimator of the slope.

The asymptotic variance of $\widehat{\beta}_D$, denoted $V_a(\widehat{\beta}_D)$, is equal to

[1] See "auto-test", one of the pillars of efficient learning – reference: David Louapre (Science Étonnante)'s video on learning how to learn (*link*). If you have not seen this video yet, I advise you to stop this quiz immediately and first watch it: the returns you can get from this 29-minute video likely eclipse any specific quiz, lecture note, or review.

[2] $\widehat{\beta}$ is the OLS estimator in the linear regression of Y on $X = (X^1, \dots, X^j, \dots X^k)'$ (with $X^1 = 1$ the constant/intercept if not stated otherwise), β_0 is the coefficient in the theoretical linear regression of Y on X, and ε denotes the error term: $\varepsilon := Y - X'\beta_0$. For any $j \in \{1, \dots, k\}$, $\widehat{\beta}_j$ is the j-th component of $\widehat{\beta}$, and β_{0j} is the j-th component of β_0. ***Note:*** *without contrary indications, the moment conditions of Chapter 1 Proposition 5 that are required to uniquely define β_0 as $\beta_0 := \mathbb{E}[XX']^{-1}\,\mathbb{E}[XY]$ are implicitly assumed; note that it is also the case in TD and exams.* Consequently, under i.i.d sampling (also implicitly assumed in general) β_0 is the limit in probability of the OLS estimator $\widehat{\beta}$.

1. $\mathbb{E}[\varepsilon^2] \,/\, (\sqrt{n}\, \mathbb{V}[D])$
2. $\mathbb{E}[\varepsilon^2] \,/\, \mathbb{V}[D]$
3. $\mathbb{V}[D] \,/\, \big(n\ \mathbb{E}[\varepsilon^2]\big)$
4. $\mathbb{V}[D] \,/\, \mathbb{E}[\varepsilon^2]$

3 Accuracy of the OLS estimator (multiple linear regression)

We assume the homoskedasticity condition $\mathbb{E}[\varepsilon^2 XX'] = \mathbb{E}[\varepsilon^2]\,\mathbb{E}[XX']$ holds, and we consider a multiple linear regression (MLR) of Y on X, X is a column vector of random variables. $\widehat{\beta}_j$ is the OLS estimator associated with the j-component X^j of X and the shortcut notation X^{-j} denotes the other components of X, namely the vector X except X^j.

Then, the asymptotic variance of $\widehat{\beta}_j$ is minimal when

1. X^j and X^{-j} are linearly dependent
2. X is a Gaussian vector
3. $\mathbb{C}\mathrm{ov}(X^{-j}, Y) = 0$
4. $\mathbb{C}\mathrm{ov}(X^{-j}, X^j) = 0$

4 "Standard errors"

As in Chapter 2, we denote by $V_{a,jj}$ the asymptotic variance of $\widehat{\beta}_j$ and by $\widehat{V}_{a,jj}$ its estimator. Then, se_j, the "standard error" associated with $\widehat{\beta}_j$ reported by statistical software like Stata, is equal to

1. $V_{a,jj}^{1/2} / \sqrt{n}$
2. $\widehat{V}_{a,jj}^{1/2}$
3. $\widehat{V}_{a,jj}^{1/2} / \sqrt{n}$
4. $\mathbb{V}[\widehat{\beta}_j]^{1/2}$

5 A simple bilateral test

We consider the test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$, and we obtain $\widehat{\beta}_j = 5$ and $\mathrm{se}_j = 2.5$. Thus,

1. We accept H_0 at 5%, but we reject H_0 at 1%
2. We accept H_0 at 1%, but we reject H_0 at 5%
3. We accept H_0 at 1% and 5%
4. We reject H_0 at 1% and 5%

Table 1: Quantiles of order $\tau \in (0, 1)$ of a standard Gaussian distribution $\mathcal{N}(0, 1)$.

τ	0.900	0.950	0.975	0.990	0.995
q_τ	1.282	1.645	1.960	2.326	2.576

6 Bilateral (two-sided) and unilateral (one-sided) tests

In the simple bilateral test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$, we obtain $t_j < 0$ and a p-value equal to 0.06. Then,

1. We reject the bilateral test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$ at the 1% level
2. We reject the bilateral test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} \neq 0$ at the 5% level
3. We reject the unilateral test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} > 0$ at the 5% level
4. We likely reject the unilateral test $H_0 : \beta_{0j} = 0$ against $H_1 : \beta_{0j} < 0$ at the 5% level

7 Links between simple and joint/multiple tests

For a given level $\alpha \in (0, 1)$, assume that we accept the simple test $H_0 : \beta_{01} = 0$ against $H_1 : \beta_{01} \neq 0$ and that we reject the same test for β_{02}, $H_0 : \beta_{02} = 0$ against $H_1 : \beta_{02} \neq 0$. Then,

1. We necessarily reject at level α the joint/multiple test $H_0 : \beta_{01} = \beta_{02} = 0$ against $H_1 : \beta_{01} \neq 0$ or $\beta_{02} \neq 0$
2. We necessarily accept at level α the joint/multiple test $H_0 : \beta_{01} = \beta_{02} = 0$ against $H_1 : \beta_{01} \neq 0$ or $\beta_{02} \neq 0$
3. 0 belongs to $\mathrm{CI}_{1-\alpha}(\beta_{01})$, the $(100 - \alpha)$%-level confidence interval for β_{01}
4. With that information, we cannot decide on the previous three assertions

8 Statistical significance

Let β_D and β_G be the coefficients associated with D and G in the theoretical linear regression of Y on $X = (1, D, G)'$, with $D \in \mathbb{R}^\Omega$ and $G \in \mathbb{R}^\Omega$ two real random variables.

D is statistically significant at the 5%-level if

1. The R^2 of the regression is greater than a certain threshold
2. D is statistically significant at the 1%-level
3. D is statistically significant at the 10%-level
4. The test of joint nullity of β_D and β_G is rejected at the 5%-level

9 Test of the equality of two coefficients

β_{01} and β_{02} are two components of β_0. We want to test $H_0 : \beta_{01} = \beta_{02}$ against $H_1 : \beta_{01} \neq \beta_{02}$. We denote by $q_{1-\alpha}(r)$ the quantile of order $1 - \alpha$ of a $\chi^2(r)$. To do that test

1. We can use a F-test (Fisher's test, Proposition 5) with the critical region $\{F > q_{1-\alpha}(2)/2\}$
2. We can use a F-test with the critical region $\{F > q_{1-\alpha}(1)\}$
3. We cannot use a F-test because the tested value of β_{01} and β_{02} is not specified
4. We can use a simple bilateral t-test (Proposition 3) with the test statistic $t = (\widehat{\beta}_1 - \widehat{\beta}_2) \,/ \sqrt{\mathrm{se}(\widehat{\beta}_1)^2 - \mathrm{se}(\widehat{\beta}_2)^2}$

10 Power of tests

When the sample size n increases to infinity, the simple and joint/multiple tests studied in Chapter 2 have more power, namely

1. We tend to reject the null hypothesis H_0 when it is false more often
2. We tend to accept the null hypothesis H_0 when it is true more often
3. The probability that the null hypothesis H_0 is true tends to 0
4. The probability that the null hypothesis H_0 is true tends to 1

11 Interpretation of a Stata `regress` output table

We obtained the following Stata output by doing the multiple linear regression of `fr_ce2` (french test score in 3rd grade (US system) = CE2 (French system)) on `mois_naiss` (the birth month[3]), `nbelev1` (the number of pupils, i.e., the class size on 1st grade = CP), `nbelev2` (the class size on 2nd grade = CE1), and a constant.

Answer the following questions (write the answers for open questions; otherwise, True or False?).

```
Linear regression                               Number of obs     =      3,644
                                                F(3, 3640)        =       5.43
                                                Prob > F          =     0.0010
                                                R-squared         =     0.0050
                                                Root MSE          =     14.334
```

fr_ce2	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
mois_naiss	-.0757238	.0340534	-2.22	0.026	-.1424894	-.0089583
nbelev1	.2005965	.0777009	2.58	0.010	.0482548	.3529382
nbelev2	.0935055	.0768507	1.22	0.224	-.0571693	.2441803
_cons	72.58695	1.976025	36.73	0.000	68.71272	76.46117

1. The standard errors are computed under the homoscedasticity assumption (Hom) of Chapter 2 (slide 10).
2. We accept at 1% the null hypothesis that the coefficients of the three regressors are jointly equal to 0.
3. Which regressors are statistically significant at 1%?
4. Which regressors are statistically significant at 5%?
5. The low R^2 of the regression entails that the regression does not identify causal effects.
6. Discuss the sign of the estimated coefficients: are they expected? Do you think the regression identifies the causal effect of class size?[4]
7. Write a sentence to interpret the estimated coefficient of the variable `mois_naiss`.
   Indication: here, `fr_ce2` is a grade measured over 100 points.
8. Same question for the variable `nbelev1`.

[3] 1 = January, 2 = February, ..., 12 = December; children born earlier on the year are more mature a priori and it might impact their scores in primary school.

[4] More to come in Chapter 4 for causal effects. For the moment, you can simply wonder whether the sign of the estimated coefficients for `nbelev1` and `nbelev2` are logical or surprising.

12 *Finite-sample variance and asymptotic variance

We have an asymptotically normal estimator $\widehat{\theta}$ of a scalar (1×1) parameter θ_0, that is,

$$\sqrt{n}(\widehat{\theta} - \theta_0) \underset{n\to+\infty}{\xrightarrow{d}} \mathcal{N}(0, V) \text{ for some } V \in \mathbb{R}_+^*.$$

We also have at our disposal two estimators $\widehat{V}_1$ and $\widehat{V}_2$, that are such that

$$\widehat{V}_1 / \mathbb{V}[\widehat{\theta}] \underset{n\to+\infty}{\xrightarrow{P}} 1, \quad \text{and} \quad \widehat{V}_2 \underset{n\to+\infty}{\xrightarrow{P}} V.$$

We can construct a test with asymptotic level α of $H_0 : \theta_0 = 0$ against $H_1 : \theta_0 \neq 0$ with the critical region $W_\alpha = \{|t| > q_{1-\alpha/2}\}$ where

1. $t = \widehat{\theta} \,/\, \widehat{V}_1^{1/2}$
2. $t = \widehat{\theta} \,/\, \widehat{V}_2^{1/2}$
3. $t = \widehat{\theta} \,/\, (\widehat{V}_1/n)^{1/2}$
4. $t = \widehat{\theta} \,/\, (\widehat{V}_2/n)^{1/2}$

13 The Gauss-Markov theorem and different types of variance

(a) State the Gauss-Markov theorem; be precise about its conditions.
Hint: the result of this theorem is also known as the OLS estimator being the BLUE, with BLUE, an acronym, meaning ...

(b) With the notation of Chapter 2, $V(\widehat{\beta}|X_1, \dots, X_n)$ or, with our font,$\mathbb{V}[\widehat{\beta} \,|\, X_1, \dots, X_n]$, intervenes in the statement of the Gauss-Markov theorem. Explain the meaning of this object and, in contrast, of these different types of "variance".
Important: be precise about the nature of the objects, in particular, whether they are stochastic or not.

1. $\mathbb{V}[\widehat{\beta} \,|\, X_1, \dots, X_n]$, or $\mathbb{V}[\widehat{\beta}_j \,|\, X_1, \dots, X_n]$ when considering one particular component, the j-th one, of the vector $\widehat{\beta}$ of OLS estimators
2. V_a also denoted $V_a(\widehat{\beta}) := \mathbb{E}[XX']^{-1}\, \mathbb{E}[\varepsilon^2 XX']\, \mathbb{E}[XX']^{-1}$, or $V_{a,jj}$
3. $\widehat{V}_a := \widehat{\mathbb{E}}[XX']^{-1}\, \widehat{\mathbb{E}}[\widehat{\varepsilon}^2 XX']\, \widehat{\mathbb{E}}[XX']^{-1}$, or $\widehat{V}_{a,jj}$
4. $\mathbb{V}[\widehat{\beta}]$, or $\mathbb{V}[\widehat{\beta}_j]$ (with the font of Chapter 2's slides: $V(\widehat{\beta}_j)$)

14 *Independent and Gaussian error terms

The last section of Chapter 2 introduces the assumption $\varepsilon \,|\, X \sim \mathcal{N}(0, \sigma^2)$. What is the interest of such a hypothesis?

1. It allows us to construct consistent tests
2. It is weaker than the hypothesis $\mathbb{E}[\varepsilon^2 XX'] = \mathbb{E}[\varepsilon^2]\, \mathbb{E}[XX']$
3. It allows us to construct tests with non-asymptotic level (a.k.a. exact or finite-sample level)
4. It allows us to relax the hypothesis that the random variables associated with the different individuals are jointly independent