Quiz 4 – Chapter 4: Linear Regressions and Causality

(Lucas Girard) – This version: 6 December 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. Notation. *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. Absent contrary indication, the notation used follows that of the course's slides.
$D \in \mathbb{R}^{\Omega}$, a real random variable, is the treatment variable, either binary (Support$(D) = \{0, 1\}$) or a non-binary *ordered quantitative* variable (Support$(D) = \mathbb{R}$). (Sometimes, we may also consider multivariate treatment: $D \in (\mathbb{R})^{\dim(D)}$, with $\dim(D) \geq 2$.) We are interested in the causal effect of D (the "treatment") on an outcome real variable Y. For any $d \in$ Support(D) (note that d is thus not a random variable, it is just a "free variable" – *variable muette* in French), we denote by $Y(d)$ the *potential* outcome associated with the value d of the treatment D: $Y(d) \in \mathbb{R}^{\Omega}$ is a real random variable. Remember that, in addition to D (and G if present), we only observe $Y := Y(D) \in \mathbb{R}^{\Omega}$, the *observed* outcome. In particular, when the treatment D is binary, we observe $Y := DY(1) + (1 - D)Y(0)$, but not the couple of potential outcomes $(Y(0), Y(1))$. That distinction between *observed* and *potential* outcomes is crucial.
We thus assume to observe an i.i.d. sample $(Y_i, D_i)_{i=1,...,n}$, with $n \in \mathbb{N}^*$, that have the same distribution, denoted $\mathrm{P}_{(Y,D)}$, as a generic couple (Y, D), written without the index i. Sometimes, we also have access to a column vector $G \in (\mathbb{R}^{\dim(G)})^{\Omega}$ of control variables. In this case, we observe $(Y_i, D_i, G_i')_{i=1,...,n}$ i.i.d with distribution $\mathrm{P}_{(Y,D,G)}$.
We consider simple linear regressions of Y on D, or multiple linear regressions of Y on D and G. Absent contrary indications, all regressions include an intercept/constant as usual. As in problem sets and exams, if not stated otherwise, we implicitly assume i.i.d. sampling of the observations and that the three standard moment conditions of Chapter 1, Proposition 5 hold: Y and X admit finite second-order moments, with $X := (1, D)'$ (without controls) or $X := (1, D, G')'$ (with controls), and $\mathbb{E}[XX']$ is invertible.
Questions. As in the course slides, questions marked with an asterisk are more advanced.
Questions 1 to 4 deal with the first section of Chapter 4: binary treatments. Question 1 is to check your knowledge of the fundamental definitions used to formalize causal effects. Question 2 is about Proposition 1. Question 3 presents two concrete examples to apply the notions of Chapter 4; they are interesting exercises. Part (a) of Question 4 is a must-know about Proposition 1, again. Part (b) is more advanced and deals with the issue of testing the absence of selection.
Questions 5 and 6 are concerned with the interpretation of linear regressions and how, despite being linear (*in parameters!*), linear regression can account for non-linear effects, at the cost of using known transformations of the initial outcome and/or the treatment variables.
Question 7 is about Section 2: the case of a single non-binary (but ordered and with a quantitative – as opposed to qualitative – meaning) treatment, notably Proposition 2.
Questions 8 to 9 are concerned with the second part of Chapter 4, sections 3 and 4 with control variables. More precisely, Question 8 is also linked to sections 1 and 2 and compares the assumptions of absence of selection and of absence of *conditional* selection, which are critical conditions for the purpose of Chapter 4, namely, identify causal parameters through linear regressions. For that goal, Question 9 discusses, on an example, whether one should add control variables.
Questions 10 and 11 are proposed to explore several aspects of the causal linear models (Lin. mod. 1) and (Lin. mod. 2) studied in Chapter 4. Question 10 is quite important to understand the distinction between heterogeneous or homogeneous causal effects. Question 11 (marked with an asterisk) is a more open question, but it might be interesting to check your understanding of the course deeper.
Bonne lecture ! Do not hesitate if you have any questions.

1 Fundamental objects for causality (binary treatment)

As in the first section of Chapter 4, we consider a single binary variable $D \in \{0, 1\}^{\Omega}$.

(a) Chapter 4 introduces the following objects and notations: Δ, δ, δ^{T}, B, and β_D. Give their respective name and definition.

[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.

(b) In general, what can we say about δ and δ^{T}?

1. $\delta > \delta^{\mathrm{T}}$
2. $\delta < \delta^{\mathrm{T}}$
3. $\delta \neq \delta^{\mathrm{T}}$
4. $\delta = \delta^{\mathrm{T}}$

(c) In this sub-question, we assume *homogeneous causal effects*, namely $\exists\, \delta_0 \in \mathbb{R} : \Delta = \delta_0$, that is, Δ is a degenerate/constant random variable ($\mathbb{V}[\Delta] = 0$). In this case, what can we say about δ and δ^{T}?

2 Causality and regressions with a single binary covariate

We consider a single binary treatment $D \in \{0,1\}^{\Omega}$ and the simple linear regression of Y on D.

What is the minimal (that is, weakest) assumption for that regression to identify δ^{T}, the average causal effect on the treated? In other words,[2] what is the minimal assumption for the OLS estimator of the slope in the simple linear regression of Y on D to converge in probability to δ^{T}?

1. None, it is always the case
2. $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$
3. $\mathbb{C}\mathrm{ov}(D, Y(d)) = 0$ for all $d \in \mathrm{Support}(D) = \{0,1\}$, that is, $\mathbb{C}\mathrm{ov}(D, Y(0)) = \mathbb{C}\mathrm{ov}(D, Y(1)) = 0$
4. $D \perp\!\!\!\perp (Y(0), Y(1))$

3 Examples to apply the notions of Chapter 4

This question differs from standard Quiz questions: they are rather open questions for you to search, think about, and apply the notions of Chapter 4 in some concrete examples.

(a) During some web navigation or at the library, you encounter the (title of the) following article published in *The Harvard Business Review* in May-June 2021:

Enlightened by your Econometrics 1 course, what questions should you ask yourself?
Some examples/hints:[3]

[2] This is another equivalent formulation of the question to explain the meaning of “identify” here.

[3] Feel free to imagine other questions to apply the concepts of Chapter 4 to this concrete example (or another example you can find).

1. To begin with, you can simplify the analysis by considering a binary treatment D that you will precisely define; in a second step, you can extend your reasoning to a non-binary, ordered, and quantitative treatment that you will specify. Give the concrete meanings in words of the potential outcomes $Y(d)$ for this example.

2. Do you think showing such a causal effect is, a priori, easy? Think about possible omitted variable bias issues. Would a controlled randomized experiment be easy to implement in this case?

3. In words, for this concrete application, discuss what would be a statistically significant but practically insignificant effect.

(b) Below is a quotation[4] from a former version of ENSAE Statistics 1 course:

> *"Certain old men prefer to rise at dawn, taking a cold bath and a long walk with an empty stomach and otherwise mortifying the flesh. They then point with pride to these practices as the cause of their sturdy health and ripe years; the truth being that they are hearty and old, not because of their habits, but in spite of them. The reason we find only robust persons doing this thing is that it has killed all the others who have tried it."*
> *Ambrose Pierce*

1. In this example, what would be the binary treatment D? Specify in words the concrete meaning of the potential outcomes $Y(0)$ and $Y(1)$ here.

2. Discuss the selection bias problem evoked in this example, notably by the phrase: "*not because of their habits, but in spite of them*".[5] The quotation also evokes another problem of selection, namely the selection of units *into the sample*: which units do we observe?[6] as opposed to the selection of observations *into the treatment*: which observations receive the treatment? You shall forget this additional dimension here (see next semester in Econometrics 2) to focus on the selection bias into the treatment studied in Chapter 4, in particular in slides 8 and 9.

4 Correlation between D and $Y(0)$

We are in the setting of Chapter 4, Section 1: in particular, D is the binary treatment, and $Y(0)$ is the potential outcome absent the treatment.

In this context, Proposition 1 of Chapter 4 states a sufficient and necessary condition for the simple linear regression of Y on D to identify the quantity denoted δ^{T} (that is, a sufficient and necessary condition for the OLS estimator of the slope in that regression to converge in probability to δ^{T}). This condition can be written $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$.

(a) Write the relationship between β_D, δ^{T}, and B given by Proposition 1 and deduce another equivalent formulation of the condition $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$.

(b) The condition $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$ is thus fundamental to determine whether a linear regression identifies a causal effect. Hence, we would like to test this condition against the alternative $\mathbb{C}\mathrm{ov}(D, Y(0)) \neq 0$.

Is it possible to do so? Choose the unique correct assertion below and justify your answer.

[4] The author is probably Ambrose *B*ierce rather than Pierce – Wikipedia page *link*.

[5] For a French literary analogue: "Ma mère s'émerveillait qu'il [M. de Norpois] fût si exact quoique si occupé, si aimable quoique si répandu, sans songer que *les « quoique » sont toujours des « parce que » méconnus*", Marcel Proust, *À l'ombre des jeunes filles en fleur* (je mets en italique, et non le texte original).

[6] See, for instance, Quiz 1, Question 13 on that topic.

1. We can *always* test $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$; it amounts to looking at the empirical covariance between D and Y in the subsample of non-treated units (namely, with D equal to 0).

2. In general, we cannot test $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$; however, we can test it in the specific set-up of randomized experiments.

3. We can *never* test $\mathbb{C}\mathrm{ov}(D, Y(0)) = 0$ (even in the set-up of randomized experiments).

4. None of the previous assertions; if so, indicate below the correct one.

5 Accounting for nonlinearities

We consider the model

$$\forall p \in \mathbb{R}_+^*, \; \log Y(p) = \gamma_0 - \delta_0 \log p + \eta, \tag{1}$$

where,

- for any positive price $p \in \mathbb{R}_+^*$, the potential outcome variable $Y(p) \in (\mathbb{R}_+^*)^\Omega$ is the demand (assumed positive to take its logarithm) for some good when the price of the good is equal to p, that is, the quantity asked (measured in some physical unit of measure; for instance, 3 tons of apples) when the price is equal to p;
- $\gamma_0 \in \mathbb{R}$ and $\delta_0 \in \mathbb{R}$ are two scalar non-stochastic parameters;
- and $\eta \in \mathbb{R}^\Omega$ is a real random variable.

In other words, more formally, we assume the following proposition holds:

$$\exists\, (\delta_0, \gamma_0) \in \mathbb{R}^2, \exists\, \eta \in \mathbb{R}^\Omega : \; \forall p \in \mathbb{R}_+^*, \; \log Y(p) = \gamma_0 - \delta_0 \log p + \eta.$$

In model (1), how can we interpret the parameter δ_0?

1. An increase of 1% of the price decreases the demand by δ_0 units.

2. An increase of 1 euro in the price decreases the demand by δ_0 units.

3. An increase of 1 euro in the price decreases the demand by $100\delta_0\%$ approximately.

4. δ_0 is the elasticity of the demand with respect to the price; that is, an increase of 1% of the price decreases the demand by $\delta_0\%$.

6 Interpretation of a linear regression

We consider the simple linear regression, where an observation is an American town, of Y on D, where Y is the share of votes expressed in % for the Republican party in the town, and D is a binary variable equal to 1 if the media Fox News is available in the town, 0 otherwise.

As in the course (see, for instance, the non-causal linear representation in slide 14), we denote by α_0 and β_D the intercept and the slope coefficient in the theoretical regression. As a reminder: they are the limit in probability of the OLS estimators $(\widehat{\alpha}, \widehat{\beta}_D)$ under the appropriate moment conditions and i.i.d sampling of Chapter 1, Proposition 5.

What can we say from this regression?

1. We predict an increase of β_D percentage points (1 percentage point = 0.01) in favor of the Republican party in a town with Fox News compared to a town without Fox News.

2. We predict an increase of $\beta_D\%$ in the share of the vote in favor of the Republican party in a town with Fox News compared to a town without Fox News.

3. We predict an increase of β_D percentage points in favor of the Republican party if the audience of Fox News increases by 1% in the town.

4. None of the previous interpretations is valid in general: to make them, we have to assume that the selection bias is null.

7 Regressions and causal effects with a non-binary treatment

This question relates to Section 2 of Chapter 4: the case of a *single* non-binary but ordered and quantitative treatment D.

(a) Write the definitions of W and δ^W as defined in Chapter 4.

(b) Give a sufficient condition (but possibly strong assumption) that implies $\delta^W = \delta$.

(c) Is the following assertion true or false?
The parameter δ^W does not vary if the distribution P_D of the treatment D changes.

(d) As a comparison, same question for δ:
Does the parameter δ vary if the distribution P_D of the treatment D changes?

(e) Give the conditions stated in Chapter 4 (Proposition 2) for the limit in probability β_D of the OLS estimator $\widehat{\beta}_D$ of the slope in the simple linear regression of Y on D to be equal to the causal parameter δ^W.

8 The absence of (conditional) selection

We are in the setting of Section 2 of Chapter 4 with D the treatment variable (a real random variable, not necessarily binary), G the control variables (a vector of real random variables), and, for any $d \in \text{Support}(D)$, $Y(d)$ the potential outcome corresponding to the value d of D ($Y(d)$ is a real random variable).

We consider the following two assumptions:

(i) the absence of selection: $\mathbb{C}\text{ov}(Y(d), D) = 0$ for all $d \in \text{Support}(D)$;

(ii) the absence of *conditional* selection: $\mathbb{C}\text{ov}(Y(d), D \mid G) = 0$ for all $d \in \text{Support}(D)$.

What can you say about those two conditions?

1. $\mathbb{C}\text{ov}(Y(d), D) = 0$ is the weakest: it is implied by $\mathbb{C}\text{ov}(Y(d), D \mid G) = 0$
2. $\mathbb{C}\text{ov}(Y(d), D \mid G) = 0$ is the weakest: it is implied by $\mathbb{C}\text{ov}(Y(d), D) = 0$
3. The two conditions are equivalent when D is binary
4. $\mathbb{C}\text{ov}(Y(d), D) = 0$ neither implies nor is implied by $\mathbb{C}\text{ov}(Y(d), D \mid G) = 0$, but the latter (conditional on G) is often more credible
5. $\mathbb{C}\text{ov}(Y(d), D) = 0$ neither implies nor is implied by $\mathbb{C}\text{ov}(Y(d), D \mid G) = 0$, but the former (unconditional) is often more credible

9 Short or long regressions in a randomized experiment

We are interested in the causal effect of a training program[7] on jobseekers' income. To do so, we implement a *randomized controlled experiment* (RCT) on a representative sample of jobseekers: we randomly provide the training to some of them, then, one year later, we measure their monthly income as well as other individual covariates (age, gender, education, having or not a college degree, etc.)

We contemplate two linear regressions:

[7] like the NSWD Program we will see in a Problem Set with personalized support to write CVs, make online job searches, prepare for interviews, etc.

(i) a "short" regression of the monthly income Y on the indicator D of receiving the training;

(ii) a "long" regression of the monthly income Y on the indicator D of receiving the training and the indicator C of having a college degree (diplôme post-bac).

In order to *estimate* the average causal effect of the training program, it is preferable to do

1. the "short" regression because the "long" one suffers from the included variable bias
2. the "long" regression because the "short" one suffers from the omitted variable bias
3. it is indifferent: we can do the "short" regression or the "long" regression since the estimates of the coefficients of D will tend to be close in the two regressions (more precisely, the estimators of the two regressions converge to the same probability limit)
4. none of the previous answer: neither the "short" nor the "long" regressions are useful to estimate the average causal effect of the training; they can only estimate the average causal effect *on the treated*

10 Heterogeneous or homogeneous causal effects

We consider the linear causal model (Lin. mod. 1) (slide 28) and its multivariate extension (Lin. mod. 2) (slide 31) of Chapter 4. Remember that m is the dimension of D and of Δ.

We focus here on one of the conditions of those models, namely:

$$\exists\, \delta_0 \in \mathbb{R}^m \text{ (that is, non-stochastic)} : \mathbb{E}[\Delta \mid D, G] = \delta_0. \quad (*)$$

Are the following assertions true or false?

1. Condition $(*)$ allows Δ to be random, namely to vary from one individual to another (case of *heterogeneous* causal effects).
2. If Δ is a degenerate constant random variable, that is, Δ is equal across individuals (case of *homogeneous* causal effects), then condition $(*)$ is necessarily satisfied.
3. Condition $(*)$ rules out nonlinear effects of D on Y.
4. Condition $(*)$ implies that $\mathbb{E}[\Delta] = \delta_0$ and $\mathbb{E}[\Delta \mid D] = \delta_0$.

11 *Causal linear models and linear conditional expectations

We consider the linear causal model (Lin. mod. 1) (slide 28) or its multivariate extension (Lin. mod. 2) (slide 31) of Chapter 4.

Let us consider the univariate case (Lin. mod. 1) for simplicity.

(a) Given an outcome real random variable $Y \in \mathbb{R}^\Omega$, a univariate treatment $D \in \mathbb{R}^\Omega$, and control variables $G \in (\mathbb{R}^{\dim(G)})^\Omega$, write Linear Model 1 (Lin. mod. 1) formally with quantifiers ($\exists$ and $\forall$).

Warning: the answer is provided on the next page to ask the following questions; try to answer before turning the page!

(Lin. mod. 1) writes

$$\exists\,\Delta \in \mathbb{R}^{\Omega},\ \exists\,\eta \in \mathbb{R}^{\Omega},\ \exists\,\delta_0 \in \mathbb{R},\ \exists\,\zeta_0 \in \mathbb{R},\ \exists\,\gamma_0 \in \mathbb{R}^{\dim(G)},\ \exists\,d_0 \in \text{Support}(D):$$

$$\forall\, d \in \text{Support}(D),\ Y(d) = \zeta_0 + G'\gamma_0 + \Delta(d - d_0) + \eta \qquad \text{(LASCE)}$$

$$\mathbb{E}[\eta \mid D, G] = 0 \qquad \text{(ACS)}$$

$$\mathbb{E}[\Delta \mid D, G] = \delta_0. \qquad \text{(IHCE)}$$

(b) Under (Lin. mod. 1), what can you say for the observed outcome $Y := Y(D)$? That is, write an equation satisfied by Y.

(c) Comment each of the assumptions (LASCE), (ACS), and (IHCE) of (Lin. mod. 1).

In particular, try to guess why the related equations are called as such; that is, for which words do the acronyms LASCE, ACS, and IHCE stand? Each letter represents one word, and link words, such as "of", "the", etc., are not represented by a letter. An example: ATE = Average Treatment Effect.

***(d)** Why do we *not* consider $Y(d, g)$, for $d \in \text{Support}(D) \subseteq \mathbb{R}$ and $g \in \text{Support}(G) \subseteq \mathbb{R}^{\dim(G)}$?

First, specify what the notation $Y(d, g)$ means. A related question that may help (or not): in the models studied in Chapter 4, is it possible to consider $Y(d)$ with d multivariate? If so, quote the associated model.

(e) What is the implication of (Lin. mod. 1) for the conditional expectation of $\mathbb{E}[Y \mid D, G]$, or, equivalently, of $\mathbb{E}[Y \mid X]$ with $X := (1, D, G')'$?
Hint: look at the title of this question and use your result to question (b).

(f) Is that condition (the one of question (e)) on $\mathbb{E}[Y \mid D, G]$ necessary to identify the causal parameter of interest δ_0? If not, propose an alternative to (Lin. mod. 1) that does not imply the condition of question (e) on $\mathbb{E}[Y \mid D, G]$.

(e and f bis) The same questions as (e) and (f) in a model without control variables G.

You can start by writing the corresponding model (formally, with quantifiers, as in question (a)).

***(g)** Discuss if and why the condition of question (e) on $\mathbb{E}[Y \mid D, G]$ might be interesting nonetheless.[8]

[8] Like question (d), question (g) is quite open.