INTRODUCTION to ECONOMETRICS

Home-work III 2020

Deadline. 18 Dec 11h59pm [Paris time] via email in PDF format and first class after the deadline for the printed version of your work again. Delays will be penalized.

Instructions. I am expecting one PDF file for each student named LAST_NAME.pdf containing the description of the code, the results obtained and their interpretations. This document should be formatted according to the template available on the course web-page. Not respecting the formatting will affect your grade. Unless explicitly stated you cannot use any built-in command in R, I am expecting you to code everything yourself.

Remark. The second part of the exam might be challenging, do not worry if you end up in not solving every single point. A useful reference for this second question is

http://cameron.econ.ucdavis.edu/research/Cameron_Miller_JHR_2015_February.pdf

Together with your textbook this paper contains anything you need to complete this home-work.

Problem 1 [classical measurement error]

Set-up. Consider the following population regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

and assume that the usual 3 OLS assumptions are satisfied. However instead of properly measuring X_i you incorrectly measure it with a measurement error: you observe $\tilde{X}_i$ instead of X_1.
As a consequence you end up estimating

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + \underbrace{\beta_1 (X_i - \tilde{X}_i) + u_i}_{v_i}$$

$$Y_i = \beta_0 + \beta_1 \tilde{X}_i + v_i$$

where $\tilde{X}_i$ and the error term v_i might be **correlated**.

Assume that the measurement error $w_i = \tilde{X}_i - X_i$ has zero mean, that it is uncorrelated with the variable X_i and, also, with the error term of the population regression model u_i:

$$\tilde{X}_i = X_i + w_i \qquad \rho_{w_i, ui} = 0 \qquad \rho_{w_i, X_i} = 0$$

and, as a consequence, that

$$\hat{\beta}_1 \xrightarrow{p} \frac{Var[X]}{Var[X] + Var[w]} \beta_1 \quad .$$

Assume that (X, Y) are jointly Normally distributed according with

$$(X, Y) \sim N\left[\begin{pmatrix} 50 \\ 100 \end{pmatrix}, \begin{pmatrix} 10 & 5 \\ 5 & 10 \end{pmatrix}\right]$$

and that you do not observe X_i but only $\tilde{X}_i = X_i + w_i$ where w_i are iid Normally distributed random variable drawn from N(0,10).

Questions.

1. Set seed to 123, draw an iid sample of (X_i, Y_i), generate $\tilde{X}_i$ and write the R code to estimate β_1 in

   $$Y_i = \beta_0 + \beta_1 X_i + u_i$$
   $$Y_i = \beta_0 + \beta_1 \tilde{X}_i + v_i \quad .$$

   Briefly discuss your results plotting in a graph the population and the two sample regression functions.

2. Use the information provided in the text above to
   - suggest a correction for α_1;
   - discuss what happens when $Var[w_i] \to 0$ and when $Var[w_i]$ grows very large with respect to $Var[X_i]$.

3. Secchi is very suspicious that the results you have obtained above depend on the specific sample you have drawn. Show, using a Monte Carlo simulation, that the OLS estimator for β_1 in the model with the measurement error is biased and inconsistent.

Problem 2 [panel data and SE]

Set-up. In the US there is an on-going debate on the extent to which the right to carry a gun influences crime. Proponents of so-called "Carrying a Concealed Weapon" (CCW) law argue that the deterrent effect of guns prevents crime, whereas opponents argue that the public availability of guns increases their usage and thus makes it easier to commit crimes. With the aim of contributing to this debate you are considering to estimate the regression model

$$\begin{aligned}\log(\mathrm{violent}_i) = {} & \beta_1 \times \mathrm{law}_i + \beta_2 \times \mathrm{density}_i + \beta_3 \times \mathrm{income}_i + \beta_4 \times \mathrm{population}_i \\ & + \beta_5 \times \mathrm{afam}_i + \beta_6 \times \mathrm{cauc}_i + \beta_7 \times \mathrm{male}_i + u_i \quad .\end{aligned}$$

using the "Guns" dataset available in the "AER" package available in R. Use "?Guns" to obtain information about the dataset.

Questions.

1. Consider the following models with one or two-way unobserved heterogeneity

$$\log(Y_{it}) = \alpha_i + \sum_i \beta_i X_{it} + u_{it}$$
$$\log(Y_{it}) = \lambda_t + \sum_i \beta_i X_{it} + u_{it}$$
$$\log(Y_{it}) = \alpha_i + \lambda_t + \sum_i \beta_i X_{it} + u_{it} \quad .$$

   where Y_{it} is the proxied by the variable "violent" in the data-set and the X_{it} are: "law", "density", "income", "population", "afam", "cauc", and "male". Produce a table with descriptive statistics of these variables.

2. Estimate the models above without any fixed-effects, with entity fixed-effects α_i, with time fixed-effects λ_t and with both. Compare the estimated coefficient associated with "law" across specifications. [**Hint**: in the two-way FE with a balanced panel the correct within transformation is $Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} + \bar{Y}$ where $\bar{Y}_{i.}$, $\bar{Y}_{.t}$ and $\bar{Y}$ are the time avg within state, the state avg within year and the total avg respectively.]

3. Replicate the same results you have just obtained using the package "lfe".

From now on consider only the specification with the entity fixed-effects.

4. Write the code to estimate the standard errors under homoskedasticty, those that correct for heteroskedasticity and those clustered at the entity (state) level. Focus again on your variable of interest "law" and comment similarities and differences of the se across estimators. [**Hint**: here you should first write the within estimator in matrix form, then estimate the var-cov matrix of the $\hat{\beta}^{ols}$ in the three cases. Part of the exercise consists in finding the proper expressions of the SE]

5. Replicate the same results using the package "lfe".