Econometrics 1

Final Exam, January 2024.

2 hours ; books, notes, slides, and calculators are forbidden.

The number of assigned to each exercise is given for indicative purposes and might be modified. We recall that plim$\widehat{\theta}$ denote the probability limit (thus, implicitly supposed to exist) of $\widehat{\theta}$.

1 Exercise 1 (5 points)

1 point for a correct answer, -0,25 points for the wrong one, 0 points if no answer is provided.

Please, mark the correct answer only. You do not need to justify your answer.

1. Let $(Y_i, X_i)_{i=1,\dots,n}$ be the observed sample, with $\sum_{i=1}^n X_i X_i'$ invertible, and assume that $\widehat{V}(Y) > 0$, where $\widehat{V}(Y)$ is the sample variance of $(Y_i)_{i=1,\dots,n}$. If $\dim(X_i) = n$, then
   (a) $R^2 = 1$.
   (b) $0 < R^2 < 1$.
   (c) $R^2 = 0$.
   (d) We cannot compute the OLS estimator in this case.
2. Let $d \in \{0, 1\}$ be a binary treatment, $Y_i(d)$ be the potential outcome of individual i under treatment status d, D_i be the observed treatment status of individual i, with $V(D) > 0$, and $Y_i(D_i)$ be the observed outcome of individual i. We denote with β_D the OLS coefficient of D in the theoretical regression of Y on $(1, D)$, and let $\delta^T = E[Y(1) - Y(0)|D = 1]$. If $\text{Cov}(Y(0), D) > 0$, then
   (a) $\beta_D > \delta^T$.
   (b) $\beta_D = \delta^T$.
   (c) $\beta_D < \delta^T$.
   (d) We cannot tell the sign of $\beta_D - \delta^T$.
3. In a regression, we obtain an OLS estimator $\widehat{\beta}_k = 3.2$ for the parameter β_k, with a standard error se $= 1.1$. Then, using the quantiles of the standard normal distribution displayed below,
   (a) We reject the null hypothesis $\beta_k = 1$ at the 1% level.
   (b) We reject the null hypothesis $\beta_k = 1$ at the 5% level.
   (c) We reject the null hypothesis $\beta_k = 2$ at the 5% level.
   (d) We cannot conclude with only this information.

TABLE 1 – Quantiles of order $1 - \alpha$ of a $\mathcal{N}(0, 1)$.

α	0.10	0.05	0.025	0.01	0.005
$q_{1-\alpha}$	1.282	1.645	1.96	2.326	2.576

4. Let $Y^* = X'\beta_0 + \varepsilon$, where $E(\varepsilon) = 0$, $E(\varepsilon X) = 0$, Y^* and $\|X\|$ are bounded, and $E(XX')$ is invertible. Assume that Y^* is observed up to an error η, so we observe $Y = Y^* + \eta$, but not Y^*. η is an unobserved variable that is bounded, with $E(\eta) = 0$, $E(\eta X) = 0$ and $V(\eta) > 0$. We observe a sample $(Y_i, X_i)_{i=1,\dots,n}$. Then,
   (a) $\widehat{\beta}$ is always consistent for β_0.
   (b) $\widehat{\beta}$ is never consistent for β_0.
   (c) $\widehat{\beta}$ is consistent for β_0 if $V(\eta|X)$ does not depend on X.
   (d) Since Y^* is mismeasured, the OLS coefficients of the regression of Y on X cannot be computed.

5. Let $Z_i = 1$ if individual i is allocated to the treatment arm and $Z_i = 0$ otherwise ; let $D_i(z) = 1$ if individual i takes the treatment under allocation $z \in \{0, 1\}$ and $D_i(z) = 0$ otherwise ; let $Y_i(d)$ be the potential outcome of individual i under treatment status d ; let $Y_i = Y_i(D_i)$ and $D_i = D_i(Z_i)$ be the observed outcome and treatment for individual i. We assume that $D(1) \geq D(0)$ and $E(D|Z = 1) - E(D|Z = 0) > 0$. Also, $Z \perp (Y(1), Y(0), D(1), D(0))$.
   We also observe an iid sample $(Y_i, D_i, Z_i)_{i=1,\dots,n}$, where Y is a bounded variable. Let us denote with $\widehat{D}$ the fitted values from the OLS regression of D on $(1, Z)$. We assume $\text{plim}\widehat{V}(\widehat{D}) > 0$. We define
   $$\widehat{\beta}_{2SLS} = \frac{\widehat{\text{Cov}}(Y, \widehat{D})}{\widehat{V}(\widehat{D})},$$
   and $\delta^T = E[Y(1) - Y(0)|D = 1]$. If $E(D|Z = 0) = 0$, then
   (a) plim $\widehat{\beta}_{2SLS} = \delta^T$.
   (b) plim $\widehat{\beta}_{2SLS} > \delta^T$.
   (c) since $E(D|Z = 1) - E(D|Z = 0) \neq 0$ we have a selection bias, so $\widehat{\beta}_{2SLS}$ is not consistent for $\delta^C = E[Y(1) - Y(0)|D(1) > D(0)]$.
   (d) For $\widehat{\beta}_{2SLS}$ to have a causal interpretation we need D_i to be independent of Z_i for all $i = 1, \dots, n$.

2 Exercice 2 (10 points)

We aim at measuring the causal effect of microcredit on the profit of individual businesses in Morocco [1]. For this purpose, we have a survey where several variables are measured : whether or not an individual received microcredit (`client`, a binary variable equal to 1 if the person has microcredit), a profit variable (`profit_total`, in Moroccan dirhams), a production variable (`output_total`, in Moroccan dirhams), and household characteristics such as size (`members_resid_bl`), the number of adults living there (`nadults_resid_bl`), and the age of the household head (`head_age_bl`).

We present the following descriptive statistics :

1. The data used here are from the article « Estimating the impact of microcredit on those who take it up : Evidence from a randomized experiment in Morocco » by Bruno Crépon, Florencia Devoto, Esther Duflo, and William Parienté, published in the *American Economic Journal : Applied* in 2015. Some elements of this article have been omitted for simplicity.

Variable	Sub-sample client=1	Sub-sample client=0
profit_total	20 246	9 608
output_total	63 619	31 182
members_resid_bl	2,98	3,95
nadults_resid_bl	2,00	2,64
head_age_bl	27,2	36,3

TABLE 2 – Averages of some selected variables by sub-sample

1. *(2 points)* First, consider the following output. Specify the regression performed, interpret the coefficient, and comment on its statistical significance. Could we obtain the coefficient directly from the table 2 above ? If yes, indicate how ; if not, explain the missing quantities to do so.

```
Linear regression                               Number of obs     =      2,448
                                                F(1, 2446)        =       8.07
                                                Prob > F          =     0.0045
                                                R-squared         =     0.0054
                                                Root MSE          =      53869

------------------------------------------------------------------------------
             |               Robust
profit_total |  Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      client |   10638.13   3743.959     2.84   0.005     3296.468    17979.78
       _cons |   9608.301   1091.424     8.80   0.000     7468.091    11748.51
------------------------------------------------------------------------------
```

2. *(2 points)* Consider two other regressions whose results are shown below. Comment on the changes in the coefficient of client in the two tables. Which of the two tables seems more relevant to measure the causal effect of microcredit ?

```
Linear regression                               Number of obs     =      2,448
                                                F(4, 2443)        =       2.79
                                                Prob > F          =     0.0252
                                                R-squared         =     0.0065
                                                Root MSE          =      53872
```

profit_total	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
client	10166.33	3782.965	2.69	0.007	2748.177	17584.48
members_resid_bl	-139.7577	631.0221	-0.22	0.825	-1377.151	1097.636
nadults_resid_bl	844.9663	936.6103	0.90	0.367	-991.6662	2681.599
head_age_bl	-96.86993	66.01396	-1.47	0.142	-226.319	32.57918
_cons	11438.94	2299.164	4.98	0.000	6930.423	15947.45

```
Linear regression                               Number of obs     =      2,448
                                                F(5, 2442)        =       8.77
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3613
                                                Root MSE          =      43202
```

profit_total	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
client	-394.5067	2877.581	-0.14	0.891	-6037.258	5248.244
members_resid_bl	-104.334	480.983	-0.22	0.828	-1047.511	838.8428
nadults_resid_bl	-1128.86	815.8121	-1.38	0.167	-2728.615	470.8955
head_age_bl	40.73926	61.17533	0.67	0.506	-79.22164	160.7002
output_total	.3260994	.0536785	6.08	0.000	.2208393	.4313595
_cons	1358.109	2073.096	0.66	0.512	-2707.098	5423.316

3. *(1.5 points)* The output below presents a joint equality test. Provide details about such a test and draw conclusions regarding the test result. Explain the idea behind the statement « We could know the result of this test given the results of question 2, » and indicate whether this statement is true.

```
( 1)  members_resid_bl = 0
( 2)  nadults_resid_bl = 0
( 3)  head_age_bl = 0

      F(  3,  2443) =    0.78
           Prob > F =    0.5078
```

4. *(1 point)* Explain why there might be a selection bias problem in the regression of question 1 and the regression considered the most relevant in question 2.

5. *(2 points)* The data used here actually come from a randomized experiment. In this experiment, a company offered microcredit to members of « treated » villages (`vil_trai` variable

equals 1 in the database), while « control » villages (`vil_trai`=0) did not have access to microcredit.

Explain how, from the table below, we can estimate the average causal effect of microcredit on the profit of those benefiting from microcredit, under assumptions that you will have to specify. Provide the theoretical formula to use, replace theoretical expressions with the appropriate numbers contained in the table, but do not perform the final calculation. Also, if necessary, explain the usefulness of the other lines in the table that are not used to estimate the causal effect.

Variable	Sub-sample `vil_trai`=1	Sub-sample `vil_trai`=0
`client`	0,16	0
`profit_total`	11 035	9 191
`output_total`	35 148	20 911
`members_resid_bl`	3,87	3,89
`nadults_resid_bl`	2,60	2,61
`head_age_bl`	35,9	36,0

TABLE 3 – Selected averages for « treated » and « control » villages.

6. *(1.5 points)* Finally, consider the output below. Explain how the estimators in the table were obtained and which instrumental variable was used (assuming the analyst made a judicious choice !). Comment on the changes in the coefficient of `client` compared to the first output from question 2. Between these two coefficients, which one seems more credible for measuring the causal effect of microcredit ?

```
Instrumental variables 2SLS regression          Number of obs   =      4,934
                                                Wald chi2(4)    =       6.72
                                                Prob > chi2     =     0.1513
                                                R-squared       =     0.0044
                                                Root MSE        =      47702

------------------------------------------------------------------------------
             |               Robust
profit_total |  Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      client |   14061.32   8128.605     1.73   0.084    -1870.456    29993.09
members_resid_bl |  -238.8202   445.9306    -0.54   0.592    -1112.828    635.1878
nadults_resid_bl |   699.4414   717.6571     0.97   0.330    -707.1406    2106.023
  head_age_bl |  -64.21065   41.32376    -1.55   0.120    -145.2037    16.78244
        _cons |   10396.08   1382.013     7.52   0.000     7687.386    13104.78
------------------------------------------------------------------------------
```

3 Exercise 3 (5 points)

Let d the potential treatment status, and let $Y_i(d)$ be the potential outcome of individual i under treatment d. Let D_i be the observed treatment status for individual i. We denote $Y = Y(D)$. We also observe a variable Z. The model for the potential outcome is

$$Y_i(d) = c + \Delta_i d^2 + \eta_i \text{ with } E(\Delta_i | D_i, Z_i) = \delta_0$$

with c a constant, $E(\eta) = 0$, $\text{Cov}(\eta, D^2) \neq 0$, $\text{Cov}(D, Z) \neq 0$, and $E(\eta Z) = 0$. The variables Y_i, D_i, Z_i are all bounded and the treatment D is continuous with $E[D] > 0$. We assume to observe an iid sample $(Y_i, D_i, Z_i)_{i=1,\ldots,n}$ having the same law as (Y, D, Z). The variables Δ_i, η_i are unobserved.

1. *(1 point)* Show that the average marginal causal effect of D on Y, called γ_0, satisfies $\gamma_0 = 2\delta_0 E[D]$.
2. *(1 point)* How can we estimate consistently δ_0 ?
3. *(1 point)* Propose an estimator of γ_0 and show that it is consistent.
4. *(2 points)* Propose a simple test of the hypothesis

$$H_0 : \gamma_0 = 0 \text{ versus } H_1 : \gamma_0 \neq 0$$