Multiple choice examination

M. Bobba, I. Durrmeyer, T. Magnac

Toulouse School of Economics - Master 1
Mid term Policy Evaluation: Semester 2

Wednesday 6 February 2018: 18H35 – 19H55

Question sheet

Please DO NOT WRITE on this sheet, USE the exam sheet.

In the following six pages, you will find a list of 20 questions. To each question, 4 possible answers numbered from A to D are associated. **Only one of the four is the true answer.**

On the exam sheet, you will report first, your surnames and first names, student ID number and second, the answers to questions in columns 2 to 5 and numbered from A to D. The identification number of the question appears in the first column. A wrong answer, multiple answers as well as no answer at all, yields no point.

Beware, the order in which the questions are listed in the exam sheet is not from 1 to 20. This order is therefore different from your neighbors'.

Exercise 1: (Questions 1-4)

Consider the linear model $Y_i(d) = \alpha + X_i\beta(d) + U_i(d)$, with $E(U_i(d) \mid X_i, D_i = d') = 0$.

Question 1 The average treatment on the treated (ATT) is equal to

A. $\alpha + E(X_i \mid D_i = 1)(\beta(1) - \beta(0))$

B. $E(X_i \mid D_i = 1)(\beta(1) - \beta(0))$

C. $\alpha + E(X_i)(\beta(1) - \beta(0))$

D. $E(X_i)(\beta(1) - \beta(0))$

Question 2 It can be consistently estimated by:

A. Regressing Y on a constant and X

B. Regressing Y on a constant, X, D and D×X

C. Regressing Y on a constant, D and D×(X-$\bar{X}^T$) (where $\bar{X}^T$ denotes the average on the treated)

D. Regressing Y on a constant and D

Consider the linear model $Y_i(d) = \alpha(d) + X_i\beta(d) + W_i\gamma + U_i(d)$, with $E(U_i(d) \mid X_i, W_i, D_i = d') = 0$.

Question 3 The average treatment effect (ATE) is equal to

A. $\alpha(1) - \alpha(0) + (\beta(1) - \beta(0))E(X_i) + \gamma\left[E(W_i|D_i = 1) - E(W_i|D_i = 0)\right]$

B. $\alpha(1) - \alpha(0) + (\beta(1) - \beta(0))E(X_i) + \gamma E(W_i)$

C. $\alpha(1) - \alpha(0) + (\beta(1) - \beta(0))E(X_i)$

D. None of the above

Question 4 It can be consistently estimated by:

A. Regressing Y on a constant, X and W

B. Regressing Y on a constant, X-$\bar{X}^T$, W-$\bar{W}^T$, D (where $\bar{X}^T$ and $\bar{W}^T$ denote the averages on the treated)

C. Regressing Y on a constant, X, D and D×(X-$\bar{X}^T$) (where $\bar{X}^T$ denotes the average on the treated)

D. Regressing Y on a constant, X, W, D and D×X

Exercise 2: (Questions 5-7)

Question 5 We consider a randomized experiment where we provide additional teaching sessions to half of the girls in a given school. We want to estimate the effect of these extra sessions (D_i) on the grade (Y_i). We assume that $Y_i(0), Y_i(1) \perp D_i|$ girl.

A. The ATT can be estimated by the difference of grades between girls and boys within a class.

B. The ATT for boys is equal to the ATT for girls.

C. The ATE for girls is equal to the ATT for girls.

D. The ATT can never be estimated.

Question 6 When implementing the experiment described in Question 5, we prefer to randomize the treatment at the class level instead of randomizing at the pupil level because

A. We may expect $(Y_j(1))_{j \neq i} \not\perp D_i$.

B. We want to decrease the variance of the estimate.

C. We want to increase the external validity.

D. We want to stratify.

Question 7 When implementing the experiment described in Question 5, we prefer to stratify by age group and randomize the treatment within each age group instead of randomizing in the whole population because

A. We may expect $(Y_j(1))_{j \neq i} \not\perp D_i$.

B. We want to decrease the variance of the estimate.

C. We want to increase the external validity.

D. We want to make sure SUTVA is satisfied.

Exercise 3 (Questions 8-11): Consider a sample of $i.i.d.$ observations $i = 1, ..., n$ with outcomes Y_i, observable characteristics $X_i = \{X_i^1, X_i^2, ..., X_i^k\}$, and where $D_i \in \{0, 1\}$ denotes whether unit i receives or not a given program or policy. Consider the following linear model:

$$Y_i = \alpha + \beta D_i + X_i'\gamma + \epsilon_i \tag{1}$$

Question 8 The β coefficient denotes:

A. The Average Treatment Effect (ATE) of D_i on Y_i:

B. $E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i)$

C. $E(Y_i|D_i = 1) - E(Y_i|D_i = 0)$

D. None of the above

Question 9 Write potential outcomes associated to equation (1) as: $Y_i(d) = \alpha(d) + X_i'\beta(d) + \epsilon_i(d)$ and assume that $E(\epsilon_i(d)|X_i, D = d') = 0$ (selection on observables). Which statement is true?

A. $(Y_i(1), Y_i(0)) \perp D_i$

B. $ATT(x) = E(Y_i(1) - Y_i(0)|D_i = 1, X_i = x) = ATT$

C. $E(Y_i(0)|X_i, D_i = 1) = E(Y_i(0)|X_i, D_i = 0)$

D. $Pr(D = 1) = Pr(D = 0)$

Question 10 Assume now that $D_i \in \{0, 1\}$ is randomly assigned across units i. Which statement is false?

A. $(Y_i(1), Y_i(0)) \perp D_i$

B. $ATT(x) = E(Y_i(1) - Y_i(0)|D_i = 1, X_i = x) = ATT$

C. $E(Y_i(0)|X_i, D_i = 1) = E(Y_i(0)|X_i, D_i = 0)$

D. $Pr(D = 1) = Pr(D = 0)$

Question 11 SUTVA states that potential outcomes of each individual are independent of the treatment status of any other individual in the population. Which of the following situation does not violate SUTVA:

A. Physical contagion between observations

B. Equilibrium effects

C. Peer effects

D. Partial compliance with the treatment $D_i = 1$

Exercise 5: (Questions 12-15) In a Randomized Control Trial, with individuals $i = 1, ., n$, we consider a treatment, D_i, whose intensity can vary in $\{1, ., K\}$ and $D_i = 0$ denotes the absence of treatment. For each $d \in \{0, 1, ., K\}$, we denote $Y_i(d)$ the corresponding potential outcome.

Question 12 The randomized nature of treatment D_i implies that

A. $Y_i(d)$ is independent of D_i

B. $Y_i(d)$ and $Y_i(d')$ are independent for $d \neq d'$

C. $Y_i(D_i)$ is independent of D_i

D. None of the above

Question 13 $E(Y_i(0) \mid D_i = d)$ is equal to

A. $E(Y_i(d) \mid D_i = d)$

B. $E(Y_i(0))$

C. $E(Y_i(d) - Y_i(0) \mid D_i = 0)$

D. None of the above

Question 14 We define $ATT_d = E(Y_i(d) - Y_i(0) \mid D_i = d)$ which, under no additional conditions, is equal to :

A. $ATT_{d'}$ for $d' \neq d$

B. $E(Y_i(d) \mid D_i = d)$

C. $E(Y_i(d) \mid D_i = 0) - E(Y_i(0) \mid D_i = 0)$

D. None of the above

Question 15 Suppose that $Var(Y_i(d)) = \sigma_d^2$ and that $\sigma_d^2 < \sigma_{d'}^2$ for any $d < d'$. Denote n_d the number of observations in sample $S_d = \{i; D_i = d\}$. We would like to equalize, if possible, the precision of the estimates of ATT_d for all $d > 0$ for fixed $n = \sum_0^K n_d$. We should choose:

A.n_d increasing with d

B.n_d decreasing with d

C. $n_0 = 0$

D. None of the above

Exercise 6: (Questions 16-18) Assume that we run two randomized control experiments in the same sample and denote the first treatment as $D_i \in \{0, 1\}$ and the second treatment as $E_i \in \{0, 1\}$. Some observations can be treated by both treatments, others by one only, and the rest receiving no treatment.

We define potential outcomes as $Y_i(d, e)$ for $(d, e) \in \{0, 1\}^2$. Define the treatment effect on the treated for treatment D when the other treatment has value e as

$$ATT_D(e) = E(Y_i(1, e) - Y_i(0, e) \mid D_i = 1)$$

Question 16 Which statement is true?

A. $ATT_D(0) = E(Y_i(1, E_i) - Y_i(0, E_i) \mid D_i = 1)$

B. $ATT_D(0) = ATT_D(1)$

C. $ATT_D(0) = E(Y_i(1, E_i) - Y_i(0, E_i) \mid D_i = 1, E_i = 1)$

D. None of the above

Question 17 Denote sample mean $\bar{y}_{de}$ as the mean of observed outcome y in the sample defined by $D_i = d$ and $E_i = e$ and sample mean $\bar{y}_{d.}$ (respectively $\bar{y}_{.e}$) as the mean of the observed outcome in sample $D_i = d$ (resp. $E_i = e$). A consistent estimate of $ATT_D(0)$ is:

A. $\bar{y}_{1.} - \bar{y}_{1.}$

B. $\bar{y}_{11} - \bar{y}_{01}$

C. $\bar{y}_{10} - \bar{y}_{00}$

D. None of the above

Question 18 If we randomize in such a way that $D_i = E_i$ for all observations:

A. $ATT_D(0) = ATT_E(0)$

B. $ATT_D(0) = ATT_D(1)$

C. $ATT_D(0)$ is not identified

D. None of the above

Exercise 7 (Questions 19-20): Suppose you are interested in the linear model:

$$Y_i = \alpha + \beta S_i + \gamma A_i + \epsilon_i,$$

where Y_i is some measure of labor earnings, S_i is schooling and A_i is an innate ability measure such that $E(\epsilon_i|S_i, A_i) = 0$. Unfortunately, data on A_i are unavailable. Instead, you have a second ability measure, a test score T_i, which is collected after schooling is completed. In general, schooling increases test scores conditional on innate ability. We assume that $T_i = \pi_0 + \pi_1 S_i + \pi_2 A_i$.

Question 19 The OLS coefficient of S_i in a regression of Y_i on a constant and S_i is equal to:

A. $\beta + \gamma \dfrac{Cov(A_i, S_i)}{Var(S_i)}$

B. $\beta + \gamma$

C. $\dfrac{Cov(A_i, S_i)}{Var(S_i)}$

D. β

Question 20 The OLS coefficient of S_i in a regression of Y_i on a constant, S_i and T_i is equal to:

A. β

B. $\beta + \gamma$

C. $\beta - \gamma \dfrac{\pi_1}{\pi_2}$

D. $\beta + \gamma \dfrac{\pi_1}{\pi_2}$