

Econ 312 Part A, Spring 2023

**Problem Set 4**

James J. Heckman

**Due April 17th, 2023 at Midnight**

This draft: April 10, 2023

1. [**2 pts**] In a two sector Roy model for possible outcomes  $(Y_0, Y_1)$ , how can it be that agents who choose sector 1 can be below average compared to the population distribution of  $Y_1$ ?
2. [**3 pts**] Answer questions in the “Normal Generalized Roy Model” hand-out.
3. [**5 pts**] Answer the questions in the “Notes on Identification of the Roy and Generalized Roy Model” handout. Give conditions under which you can identify the joint distribution of  $(Y_1, Y_0)$  in a Generalized Roy model.
4. [**5 pts**] Give conditions under which you can identify the parameters “ $C$ ” in the generalized Roy Model.  $Y = D Y_1 + (1 - D) Y_0$  is observed.

$$Y_1 = \mu_1(X) + U_1$$

$$Y_0 = \mu_U(X) + U_0$$

$$C = \mu_C(Z) + U_C$$

$$(U_1, U_0, U_C) \sim N(0, \Sigma)$$

$$D = 1(Y_1 - Y_0 - C > 0)$$

Specifically, assume you observe  $D, X, Y$  and  $Z$ . Explain how to inden-

tify  $\mu_1(X), \mu_0(X), \mu_C(Z)$  and the distribution of  $(U_0, U_1, U_C)$ .

5. [5 pts] Define and Compare ITT, PRTE, ATE, AMTE, PRTE, TT, the voting criterion for the model of Question 4. When are they the same? Express in terms of MTE.
6. [60 pts] Consider a normal selection model as in notes on generalized Roy Model and as presented in problem (4). Using the dataset titled 'ps4\_dt.csv', fit a parametric selection model. Schooling  $D$  is:

$$D = 1(Y_1 - Y_0 - C \geq 0)$$

You have data on  $Z, D, Y$  for iid samples. Consider two cases:  $Y_{1i} - Y_{0i} = \beta$ , a constant, and  $Y_{1i} - Y_{0i} = \beta_i$ . *In the dataset, variables without subscript were generated assuming  $Y_{1i} - Y_{0i} = \beta_i$ , and variables with the  $\_hom$  subscript were generated assuming  $Y_{1i} - Y_{0i} = \beta$ .*

- (i) What is: the person-specific causal effect, the quantile treatment effect, AMTE, MPRTE, ATE, TT, MTE, PRTE (all policy changes are for a unit change in  $Z$ ), the voting criterion, the distribution of  $C$ ?
- (ii) Compare your estimates with estimates from matching (nearest neighbor). When should these estimates be the same? Why are they the same or different?
- (iii) Are the data generated from a model with essential heterogeneity? (Give a non-parametric test.)

- (iv) What does LATE estimate? (Compare the LATE estimates with the ones previously derived for Question (i).)
  - (v) What is the social surplus of treatment for a given  $P(Z)$ ? (Define and estimate.)
7. **[5 pts]** When does RDD estimate the same parameter(s) as estimated by LATE? Matching? Selection models?
  8. **[5 pts]** Express in terms of MTE. Explain and contrast the role of the probability of selection ( $P(Z)$ ) in LATE, matching and Selection Models.
  9. **[5 pts]** Answer the questions embedded in notes “The Principles Underlying Evaluation Estimators – An Extract.”
  10. Explain how factor models can aid in identification of the joint distributions of potential outcomes in a Roy model.