POLSCI 787 **Due Date:**
Fall 2017 **Problem Set 3** **Fri Nov 3, 2017, 11:59pm**

- You may work in (small) groups while solving this assignment.
- Submit individual solutions via Canvas in one PDF file collecting everything (e.g., derivations, figures, tables, computer code).
- Use `RMarkdown` to make all calculations and create a report with all the solutions and answers.
- Please provide written explanations whenever those are requested; reporting only code and output automatically leads to $\checkmark^{-}$.

Question 1

This question requires computation. You should start with this question first, as you may need to wait for simulations to run. You are asked to simulate data according to the following data generating process (DGP):

$$y_i = \beta_1 + \beta_2 T_i + x_i \beta_3 + \varepsilon_i$$

with $i = 1, \ldots, n$, $x_i \sim \mathcal{N}(0, 1)$, and $\varepsilon_i | (x_i, T_i) \sim \mathcal{N}(0, 1)$, and x_i and T_i independent of each other (so $\mathbb{C}\text{ov}(x_i, T_i) = 0$). The parameter of interest is β_2, the effect of the binary "treatment" T_i that is randomly assigned according to

$$T_i = \begin{cases} 1 & \text{with probability } 0.3 \\ 0 & \text{with probability } 0.7. \end{cases}$$

Note that this random assignment assumption means that T_i is also independent of the error term. A simple way to ensure all conditions above is to generate all the regressors and the error term in your simulation separately, i.e. as independent random variables.

Set $n = 1000$ and the true values to

$$\beta_1 = 2$$
$$\beta_2 = 0$$
$$\beta_3 = 3$$

(a) Perform $M = 1000$ simulations to assess whether OLS confidence intervals for β_2 have correct coverage; remember to set the seed. This is, simulate data according to the linear model above, estimate the parameters with OLS, and build a 1-α confidence interval for β_2. Repeat this 1000 times and calculate the proportion of times that you cover the true value of β_2. Set $\alpha = 0.05$. Do your confidence intervals have correct coverage? Why should you expect/not expect to observe correct coverage in this case?

(b) Do the same simulations you did in the previous exercise, but this time estimate β_2 using a nearest-neighbor matching algorithm; in each iteration, calculated confidence intervals via non-parametric bootstrapping with two different methods: the percentile method and the basic bootstrap method. Calculate coverage of each of the two bootstrapped confidence intervals. Also, save in every iteration of the simulations the estimated value of β_2 according to three methods: OLS as above, a simple difference in means, and the matching estimate. To implement the matching, use the `R` function `Match` from the `Matching` library), setting the estimand to ATT (average treatment effect on the treated). In other words, in this exercise, we change only the estimation method for β_2 with respect to Exercise a, and keep constant the model that generates the data; indeed, you should set the seed at the same value.

You should anticipate that this will be computation intensive, because in every simulation iteration we are taking B random samples with replacement from the data. Ideally, we would set $M = 1000$ and $B = 999$. But this may not be feasible depending on the speed of your computer. In that case, set $M = 500$ and $B = 500$; if this is not feasible, set $B = 500$ and perform as many simulations as is feasible. You should first try your code with very low M and B to make sure it runs without errors, then run it with the desired values.

Answer the following questions: (i) Plot the distribution of the estimated β_2 across all simulations for the three methods separately, what do you observe?; (ii) do your bootstrapped confidence intervals have correct coverage?; (iii) Why should you expect/not expect to observe correct coverage in this case?; (iv) why is the ATT equal to the average treatment effect in this case?

Question 2

This question is about exploring the selection on observables assumption. We assume there are N units, $i = 1, 2, \ldots, N$, viewed as a random sample of a larger population. There is a binary treatment variable, D_i, and every unit i has two potential outcomes: $Y_i(1)$ if $D_i = 1$ and $Y_i(0)$ if $D_i = 0$—the potential outcomes under treatment and

control, respectively. The number of units in treatment and control are, respectively, $N_t = \sum_{i=1}^{N} D_i$ and $N_c = \sum_{i=1}^{N}(1 - D_i)$, with $N = N_t + N_c$.

The observed outcome is $Y_i^{\texttt{obs}} = D_i Y_i(1) + (1 - D_i)Y_i(0)$. In addition, for each unit i there is a K-component covariate, X_i, with $X_i \in \mathbb{X} \in \mathbb{R}^K$. The triplet $(Y_i^{\texttt{obs}}, X_i, D_i)$ is observed for all units in the sample.

Define

$$\bar{Y}_c = \frac{1}{N_c} \sum_{i:D_i=0} Y_i^{\texttt{obs}} \quad \bar{Y}_t = \frac{1}{N_t} \sum_{i:D_i=1} Y_i^{\texttt{obs}}$$

$$\bar{X}_c = \frac{1}{N_c} \sum_{i:D_i=0} X_i \quad \bar{X}_t = \frac{1}{N_t} \sum_{i:D_i=1} X_i$$

The average treatment effect on the treated is

$$\tau_{\texttt{treat}} = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$$

We make the assumption of strong ignorability:

A1. [*Unconfoundness*] $D_i \perp\!\!\!\perp (Y_i(1), Y_i(0))|X_i$.

A2. [*Overlap*] $0 < \Pr(D_i = 1|X_i = x) < 1$ for all x in the support of the covariates.

The empirical parts of this question will use two versions of the Lalonde dataset. This data was used to estimate the impact of the National Supported Work Demonstration (NSW) employment training program on earnings. The outcome of interest is earnings in 1978 (`re78`), the treatment is whether the person was assigned to the training program (`treat`), and the covariate we are interested is past earnings in 1975 (`re75`). You can read the details of the experiment in LaLonde (1986) and Dehejia and Wahba (2002). You will have two versions of the data to work with. `lalonde-exp.dta` has the experimental dataset, with 260 experimental controls and 185 experimental treated observations. `lalonde-NOexp.dta` is a non-experimental version of the dataset, where you have the 185 experimental treated observations, but the experimental control group is missing. Instead, there is a non-experimental control group of 2490 observations created by Lalonde from the Population Survey of Income Dynamics and the Current Population Survey. Thus, both datasets have the same treatment group but differ in their control group.

(a) Using the assumptions above, show that

$$\tau_{\texttt{treat}} = \mathbb{E}[Y_i^{\texttt{obs}}|D_i = 1] - \mathbb{E}\left[\mathbb{E}[Y_i^{\texttt{obs}}|D_i = 0, X_i]\,|D_i = 1\right] \tag{1}$$

(b) In Equation 1, explain which of two terms in the right-hand-side is most difficult to estimate, and why.

(c) Assume that

$$\mathbb{E}[Y_i(0)|X_i = x] = \alpha_c + \beta_c \cdot x$$

The OLS estimator for β_c is

$$\widehat{\beta}_c = \frac{\sum_{i:D_i=0}(X_i - \bar{X}_c) \cdot (Y_i^{\texttt{obs}} - \bar{Y}_c)}{\sum_{i:D_i=0}(X_i - \bar{X}_c)^2} \quad \text{and} \quad \widehat{\alpha}_c = \bar{Y}_c - \widehat{\beta}_c \cdot \bar{X}_c$$

Estimate β_c and α_c by OLS in the non-experimental Lalonde data, using only the observed data for the control group. It is easy to see that you can use the control group to estimate $\mathbb{E}[Y_i(0)|X_i = x, D_i = 0]$. But why is it that you can also use the control group to estimate $\mathbb{E}[Y_i(0)|X_i = x]$?

(d) Show that the OLS estimator of the average of the potential outcomes under control for the treated, $\mathbb{E}[Y_i(0)|D_i = 1]$, can be written as

$$\widehat{\mathbb{E}[Y_i(0)|D_i = 1]} = \bar{Y}_c + \widehat{\beta}_c \cdot (\bar{X}_t - \bar{X}_c) \tag{2}$$

(e) Based on the finding above, explain whether the following statement is true of false: "If we assume a linear functional form for $\mathbb{E}[Y_i(0)|X_i]$, we can estimate $\mathbb{E}[Y_i(0)|D_i = 1]$ with the average outcome in the control group as long as the means of the covariates are equal in the treatment and control groups."

(f) Calculate the linear estimator of $\widehat{\mathbb{E}[Y_i(0)|D_i = 1]}$ in Equation 2 using the non-experimental Lalonde data. Also estimate $\mathbb{E}[Y_i(1)|D_i = 1]$ and use these two estimators to provide a point estimator for $\tau_{\texttt{treat}}$.

(g) Compare the linear estimator of $\widehat{\mathbb{E}[Y_i(0)|D_i = 1]}$ in Equation 2 in the non-experimental Lalonde data (which you just calculated above), with two other alternative estimators: the mean of the experimental control group, and the unadjusted mean of the non-experimental control group. Why do think the differences arise?

(h) Now estimate $\tau_{\texttt{treat}}$ in the experimental Lalonde data. Compare this experimental estimate to the estimate for $\tau_{\texttt{treat}}$ that you found in question (f) above. Why the discrepancy?

(i) Now estimate $\tau_{\texttt{treat}}$ using a nearest-neighbor matching estimator, where the only covariate is `re75`. Use 1-1 matching, that is, just one neighbor per treated observation (to make your life easier, you can use the `R` function `Match` from the `Matching` library). Compare your estimated $\tau_{\texttt{treat}}$ to the non-experimental estimate obtained

in question (f) above. What do you think is causing the discrepancy? Does matching get you closer to the experimental "benchmark"?

Question 3

This question explores the relationship between the selection on observables and the random assignment assumptions. We are often interested in estimating the average effect of a binary treatment on some outcome of interest. Let i index the number of units, $i = 1, 2, ...N$, obtained via random sampling from a larger super-population. Let $Y_i(0)$ and $Y_i(1)$ denote the two potential outcomes of unit i under the control treatment and the active treatment, respectively. The variable $T_i \in \{0, 1\}$ indicates the type of treatment received. We are interested, among other parameters, in the average treatment effect $\mathbb{E}[Y_i(1) - Y_i(0)]$. In what follows, we drop the i subscripts for convenience.

If the treatment is assigned randomly, treatment assignment will be independent of $(Y(1), Y(0))$ and also of all predetermined covariates $\mathbf{X}$. We denote this assumption by $T \perp\!\!\!\perp (Y(1), Y(0), \mathbf{X})$. In this case, estimating the average treatment effect is straightforward.

Sometimes, however, treatment is not randomly assigned. In those cases, researchers are some times willing to assume that, conditional on predetermined characteristics $\mathbf{X}$, treatment assignment is independent of $(Y(1), Y(0))$. This is called the selection on observables assumption, and we denote it by $\{T \perp\!\!\!\perp (Y(1), Y(0))\} \mid \mathbf{X}$.

We will explore the relationship between the random assignment assumption and the selection on observables assumption (i.e., is one weaker than the other?, does one imply the other?). In the following subquestions, you will show the relationship between these two assumptions mathematically.

We will proceed in three steps:

- Show $(W, Z) \perp\!\!\!\perp V \Longrightarrow W \perp\!\!\!\perp V$, $Z \perp\!\!\!\perp V$.
- Show $f_{z|w,v}(z \mid w, v) = f_{z,w|v}(z, w \mid v) / f_{w|v}(w \mid v)$
- Apply prior two steps to the relationship between random assignment and selection on observables

The following questions take you through the three steps.

(a) Let W, Z and V be three random variables with marginal densities $f_w(w)$, $f_z(z)$ and $f_v(v)$ such that (W, Z) is independent of V (which we denote by $(W, Z) \perp\!\!\!\perp V$). Show that $(W, Z) \perp\!\!\!\perp V$ implies $W \perp\!\!\!\perp V$ and $Z \perp\!\!\!\perp V$.

Definitions: $(W, Z) \perp\!\!\!\perp V$ means that $f_{w,z,v}(w, z, v) = f_{w,z}(w, z) \cdot f_v(v)$; $W \perp\!\!\!\perp V$ means that $f_{w,v}(w, v) = f_w(w) \cdot f_v(v)$ and $Z \perp\!\!\!\perp V$ means that $f_{z,v}(z, v) = f_z(z) \cdot f_v(v)$.

(b) Let $f_{z|w,v}(z \mid w, v)$ be the conditional density of Z given (W, V). Show that $f_{z|w,v}(z \mid w, v) = f_{z,w|v}(z, w \mid v) / f_{w|v}(w \mid v)$, where $f_{w|v}(w \mid v)$ is the conditional density of W given V and $f_{z,w|v}(z, w \mid v)$ is the conditional density of (Z, W) given V.

(c) Let W, Z and V be three random variables. Using what you proved in the first two parts of this question, show that $(Z, W) \perp\!\!\!\perp V$ implies $(Z \perp\!\!\!\perp V) \mid W$. What does this imply about the relationship between random assignment and the selection on observables assumption?

References

Abadie, Alberto, and Guido W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–1557.

Dehejia, Rajeev H., and Sadek Wahba. 2002. "Propensity Score Matching Methods for Non-experimental Causal Studies." *Review of Economics and Statistics* 84 (February): 151–161.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review* 76 (September): 604–20.