# Problem Set #2

Danny Edgel
Econ 717: Applied Econometrics
Spring 2022

May 4, 2022

The attached files, edgel_ps2b.jl and ps2b_functions.jl, jointly conduct all of the required computational analyses for this problem set.

## Part 1: Analytic Exercises

1) a) $ATE = \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[1 + 0.5A - A] = 1 - 0.5\mathbb{E}[A] = 0.75$

   b) $Pr(D = 1) = \mathbb{E}[\mathbb{1}\{A > 0.5\}] = 0.5$

   c)
   $$\min\{1 - 0.5A\} = 0.5 \text{ at } A = 1$$
   $$\max\{1 - 0.5A\} = 1 \text{ at } A = 0$$

   d) If $A \sim \mathcal{N}(0, 1)$, $A$ would have support on the real line and would thus be unbounded, so the treatment effect would have a minimum at negative infinity and a maximum at positive infinity.

   e)
   $$ATET = \mathbb{E}[1 - 0.5A|A > 0.5] = 1 - 0.5(0.75) = 5/8$$
   $$ATEU = \mathbb{E}[1 - 0.5A|A \leq 0.5] = 1 - 0.5(0.25) = 7/8$$

   f) ATEU > ATET because the treatment effect is a decreasing in $A$, while the treatment is given only to those on the upper half of $A$'s support.

   g)
   $$\begin{aligned}\beta^{OLS} &= \mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0] = \mathbb{E}[1 - 0.5A|A > 0.5] - \mathbb{E}[A|A \leq 0.5] \\ &= 1 + 0.5(0.75) - 0.25 = 9/8\end{aligned}$$

   h) The OLS estimand is biased upward relative to the ATE because selection into being untreated requires having a lower value of $A$, and the treatment itself depends on $A$. Thus, by subtracting the outcome of those who are untreated from that of those who are treated necessarily overstates the treatment effect.

## Part 2: Monte Carlo Exercises

### Question 1

a) See the SimulateData() function in `ps2b_q1functions.jl`.

b) See the line 27 in `edgel_ps2b.jl` and the OLS() function in `ps2b_q1functions.jl`. The results are reported in the first column of table 2 below.

c)

d)

e)

f) See Table 2, below. It is puzzling that the OLS parameter estimates, which are for a misspecified model, are so close to the true parameter, as are the large-sample estimates that rely solely on $z_3$, which is a completely weak variable. This is likely because $a_i$ is mean-zero and symmetrically-distributed in the data generating process, so omitting it will create a lot of noise, but in a sufficiently large sample, negligibly biases the parameter estimates.

| | $N = 2,000$ | | | $N = 500,000$ | | |
|---|---|---|---|---|---|---|
| | $\beta_1$ | $\beta_2$ | F-Stat | $\beta_1$ | $\beta_2$ | F-Stat |
| OLS | 1.00 | 0.06 | Inf | 1.00 | 0.06 | Inf |
| | (0.01) | (0.00) | | (0.00) | (0.00) | |
| 1 | 1.97 | 0.69 | 0.0 | 1.00 | 0.05 | 1.7 |
| | (8.60) | (5.63) | | (0.00) | (0.01) | |
| 2 | 0.99 | 0.05 | 4269.9 | 1.00 | 0.05 | 1074303.6 |
| | (0.01) | (0.00) | | (0.00) | (0.00) | |
| 3 | 3.39 | 1.62 | 0.0 | 1.00 | 0.05 | 5.3 |
| | (112.34) | (73.63) | | (0.00) | (0.01) | |
| 1, 2 | 0.99 | 0.05 | 2850.4 | 1.00 | 0.05 | 716254.0 |
| | (0.01) | (0.00) | | (0.00) | (0.00) | |
| 1, 3 | 2.02 | 0.72 | 0.0 | 1.00 | 0.05 | 4.7 |
| | (8.86) | (5.80) | | (0.00) | (0.01) | |
| 2, 3 | 0.99 | 0.05 | 2848.8 | 1.00 | 0.05 | 716204.7 |
| | (0.01) | (0.00) | | (0.00) | (0.00) | |
| 1, 2, 3 | 0.99 | 0.05 | 2139.4 | 1.00 | 0.05 | 537192.3 |
| | (0.01) | (0.00) | | (0.00) | (0.00) | |

**Question 2**

a)