Econ 312: Problem Set 1

Professor Magne Mogstad

Answer Key by Sergei Bazylik

Problem 1

a) In class we assumed that U is continuously distributed and showed that it can always be transformed into $\tilde{U} \sim \mathbb{U}[0,1]$. The assumption and transformation are not restrictive.

b)& c) MTE$= \mathbb{E}[Y_1 - Y_0|U = u]$

$$ATT = \mathbb{E}[Y_1 - Y_0|D = 1] = \mathbb{E}[Y_1 - Y_0|U \leq p(Z)] = \mathbb{E}[\mathbb{E}[Y_1 - Y_0|U = u, U \leq p(Z)]|U \leq p(Z)]$$

$$\underbrace{=}_{Z \perp\!\!\!\perp (Y_1,Y_0,U)} \mathbb{E}[\underbrace{\mathbb{E}[Y_1 - Y_0|U = u]}_{MTE(u)}|U \leq p(Z)] = \frac{\mathbb{E}[MTE(u)\mathbb{1}[U \leq p(Z)]]}{\underbrace{\mathbb{P}[U \leq p(Z)]}_{=\mathbb{P}[D=1]}}$$

$$\underbrace{=}_{LIE} \mathbb{E}[\mathbb{E}\left(MTE(u)\frac{\mathbb{1}[U \leq p(Z)]}{\mathbb{P}[D = 1]}|U = u\right)] = \mathbb{E}[MTE(u)\frac{\mathbb{E}[\mathbb{1}[U \leq p(Z)]|U = u]}{\mathbb{P}[D = 1]}]$$

$$= \int_0^1 MTE(u)\frac{\mathbb{P}[u \leq p(Z)]}{\mathbb{P}[D = 1]}du = \int_0^1 MTE(u)\omega_{ATT}du$$

Consider the weights: $\frac{\mathbb{P}[u \leq p(Z)]}{\mathbb{P}[D=1]} = \frac{1-\mathbb{P}[p(Z)<u]}{\mathbb{P}[D=1]} = \frac{1-\mathbb{P}[p(Z)<u]}{\mathbb{P}[D=1]} = \frac{1-F_p^-(u)}{\mathbb{P}[D=1]}$. Note that $F_p^-(u) = \lim_{x \to u-0} F_p(x)$ is weakly increasing function meaning that weights for smaller values of u are weakly larger. In other words, we put more weight on the individuals with lower values of u. Note, that low value of u corresponds to higher probability of treatment.

$$ATUT = \mathbb{E}[Y_1 - Y_0|D = 0] = \mathbb{E}[Y_1 - Y_0|U > p(Z)] = \mathbb{E}[\mathbb{E}[Y_1 - Y_0|U = u, U > p(Z)]|U > p(Z)]$$

$$
\begin{aligned}
&\underbrace{=}_{Z \perp\!\!\!\perp (Y_1,Y_0,U)} \mathbb{E}[\underbrace{\mathbb{E}[Y_1 - Y_0|U=u]}_{MTE(u)}|U > p(Z)] = \frac{\mathbb{E}[MTE(u)\mathbb{1}[U > p(Z)]]}{\underbrace{\mathbb{P}[U > p(Z)]}_{=\mathbb{P}[D=0]}} \\
&\underbrace{=}_{LIE} \mathbb{E}[\mathbb{E}\left(MTE(u)\frac{\mathbb{1}[U > p(Z)]}{\mathbb{P}[D=0]}|U=u\right)] = \mathbb{E}[MTE(u)\frac{\mathbb{E}[\mathbb{1}[U > p(Z)]|U=u]}{\mathbb{P}[D=0]}] \\
&= \int_0^1 MTE(u)\frac{\mathbb{P}[u > p(Z)]}{\mathbb{P}[D=0]}du = \int_0^1 MTE(u)\omega_{ATUT}du
\end{aligned}
$$

Note: $\omega_{ATUT} = \frac{\mathbb{P}[u>p(Z)]}{\mathbb{P}[D=0]} = \frac{F_p^-(u)}{\mathbb{P}[D=0]}$, where $F_p^-(u)$ is weakly increasing. This means that we put more weight on higher values of u and this corresponds to lower probability of being treated ($D = \mathbb{1}[U \leq p(Z)]$)

Problem 2

a) By definition we can write:

$$
\begin{aligned}
\beta_{OLS} &= \frac{\text{Cov}(Y,D)}{\text{Var}(D)} = \frac{\mathbb{E}[YD] - \mathbb{E}[Y]\mathbb{E}[D]}{\mathbb{E}[D^2] - (\mathbb{E}[D])^2} = \frac{\mathbb{E}[(Y_1 D + Y_0(1-D))D] - \mathbb{E}[Y_1 D + Y_0(1-D)]\mathbb{E}[D]}{\mathbb{E}[D](1-\mathbb{E}[D])} \\
&= \frac{\mathbb{E}[Y_1|D=1]\mathbb{E}[D] - (\mathbb{E}[Y_1|D=1]\mathbb{E}[D] + \mathbb{E}[Y_0|D=0](1-\mathbb{E}[D]))\mathbb{E}[D]}{\mathbb{E}[D](1-\mathbb{E}[D])} \\
&= \frac{(1-\mathbb{E}[D])(\mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0])}{1-\mathbb{E}[D]} = \mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0]
\end{aligned}
$$

Where 4th equality follows from the fact that $\mathbb{E}[Y_1 D + Y_0(1-D)|D=1] = \mathbb{E}[Y_1|D=1]$ and $\mathbb{E}[Y_1 D + Y_0(1-D)|D=0] = \mathbb{E}[Y_0|D=0]$.

Now let us use the structure of bivariate normal:

$$
\begin{aligned}
\mathbb{E}[Y_1|D=1] = \mathbb{E}[U_1|U_0 - U_1 < 0] &= \frac{\rho\sigma - \sigma^2}{\sigma^2 + 1 - 2\rho\sigma}\mathbb{E}[U_0 - U_1|U_0 - U_1 < 0] + \underbrace{\mathbb{E}[\epsilon|U_0 - U_1 < 0]}_{=0} \\
&= \frac{\sigma^2 - \rho\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}}\frac{\phi(0)}{\Phi(0)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[Y_0|D=0] = \mathbb{E}[U_0|U_0-U_1>0] &= \frac{1-\rho\sigma}{\sigma^2+1-2\rho\sigma}\mathbb{E}[U_0-U_1|U_0-U_1>0] + \underbrace{\mathbb{E}[\eta|U_0-U_1<0]}_{=0} \\
&= \frac{1-\rho\sigma}{\sqrt{\sigma^2+1-2\rho\sigma}}\frac{\phi(0)}{1-\Phi(0)} = [1-\Phi(0)=\Phi(0)=\frac{1}{2}] = \frac{1-\rho\sigma}{\sqrt{\sigma^2+1-2\rho\sigma}}\frac{\phi(0)}{\Phi(0)} \\
\beta_{OLS} = \mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0] &= \frac{\sigma^2-1}{\sqrt{\sigma^2+1-2\rho\sigma}}\frac{\phi(0)}{\Phi(0)}
\end{aligned}
$$

In particular we used the decomposition:

$$
U_i = \alpha(U_0-U_1) + \epsilon, \epsilon \perp\!\!\!\perp (U_0-U_1) \implies \alpha = \frac{\mathrm{Cov}(U_i, U_0-U_1)}{\mathrm{Var}(U_0-U_1)}, i \in \{0,1\}
$$

And the expression for truncated normal mean.

The answer for the second part of the question is the following:

$$
\begin{aligned}
D \perp\!\!\!\perp (Y_1, Y_0) &\implies (\mathbb{E}[Y_1|D=1] = \mathbb{E}[Y_1]) \& (\mathbb{E}[Y_0|D=0] = \mathbb{E}[Y_0]) \\
&\implies \mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0] = \mathbb{E}[Y_1-Y_0] \\
&\implies \beta_{OLS} = \mathrm{ATE}
\end{aligned}
$$

b) By definition:

$$
\begin{aligned}
\mathrm{ATT} &= \mathbb{E}[Y_1-Y_0|D=1] = \mathbb{E}[U_1-U_0|U_1>U_0] = \mathbb{E}[U_1-U_0|U_1-U_0>0] \\
&= \sqrt{\sigma^2+1-2\rho\sigma}\frac{\phi(0)}{\Phi(0)} = 2\phi(0)\sqrt{\sigma^2+1-2\rho\sigma}
\end{aligned}
$$

Where 4th equality from the fact that $U_1-U_0 \sim \mathcal{N}(0, \sigma^2+1-2\rho\sigma)$ and the formula for the mean of the truncated normal distribution.

Similarly:

$$
\mathrm{ATUT} = \mathbb{E}[Y_1-Y_0|D=0] = \mathbb{E}[U_1-U_0|U_1<U_0] = \mathbb{E}[U_1-U_0|U_1-U_0<0]
$$

$$= -\sqrt{\sigma^2 + 1 - 2\rho\sigma}\frac{\phi(0)}{1 - \Phi(0)} = -2\phi(0)\sqrt{\sigma^2 + 1 - 2\rho\sigma}$$

Alternatively, one can simply note:

$$\begin{aligned}
\mathbb{E}[U_1 - U_0|U_1 - U_0 < 0] &= \frac{\mathbb{E}[(U_1 - U_0)\mathbb{1}(U_1 - U_0 < 0)]}{\mathbb{P}[U_1 - U_0 < 0]} = \frac{\mathbb{E}[(U_1 - U_0)(1 - \mathbb{1}(U_1 - U_0 \geq 0)])}{\Phi(0)} \\
&= \frac{\mathbb{E}[(U_1 - U_0)] - \mathbb{E}[(U_1 - U_0)\mathbb{1}(U_1 - U_0 \geq 0)]}{\Phi(0)} = -\mathbb{E}[U_1 - U_0|U_1 > U_0]
\end{aligned}$$

Which automatically answers the second part of the question: ATT$= -$ATUT

c) Simply: ATE $= \mathbb{E}[Y_1 - Y_0] = \mathbb{E}[U_1 - U_0] = 0$

d)

$$\begin{aligned}
\frac{\partial ATT}{\partial \rho} &= -\frac{\partial ATUT}{\partial \rho} = -2\phi(0)\frac{\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} \\
\frac{\partial \beta_{OLS}}{\partial \rho} &= 2\phi(0)\frac{\sigma(\sigma^2 - 1)}{(\sigma^2 + 1 - 2\rho\sigma)^{\frac{3}{2}}}
\end{aligned}$$

Meaning that ATT decreases in ρ and ATUT increases in ρ (but it is negative, so its absolute value decreases!). We would expect to see these results since higher value of ρ means that when U_1 is high U_0 is also likely to be high, meaning that difference between them is smaller (irrespective of which ends up to be larger) so both ATT and ATUT decrease in their absolute value. Roughly speaking, the options become more similar when correlation increases.

For β_{OLS} situation is little bit more complicated. Recall, on the intermediate stage of the derivation we had an expression:

$$\beta_{OLS} = \mathbb{E}[Y_1|D = 1] - \mathbb{E}[Y_0|D = 0] \iff$$

$$\beta_{OLS} = \mathbb{E}[U_1|U_1 > U_0] - \mathbb{E}[U_0|U_0 > U_1]$$

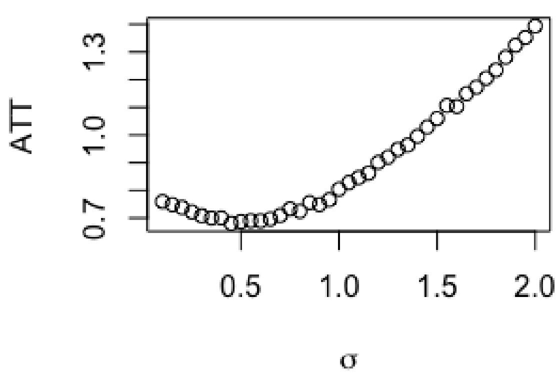
So when we increase correlation both values in each conditioning event increases, so both conditional expectations are larger. But normal distribution with larger variance has larger mass to the right and more likely to increase more than normal distribution with lower variance. So when $\sigma > 1$ increase in correlation leads to increase in β_{OLS} and when $\sigma \in (0, 1)$ increase in correlation leads to decrease in β_{OLS}.

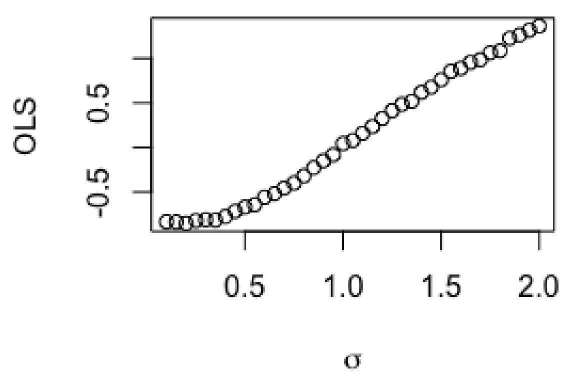
$$\frac{\partial ATT}{\partial \sigma} = -\frac{\partial ATUT}{\partial \sigma} = 2\phi(0)\frac{\sigma - \rho}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}}$$
$$\frac{\partial \beta_{OLS}}{\partial \sigma} = 2\phi(0)\frac{\sigma^3 + 3\sigma - 3\sigma\rho^2 - \rho}{(\sigma^2 + 1 - 2\rho\sigma)^{\frac{3}{2}}}$$

e) For the analytical part of the question:

$$\mathbb{E}[Y|D = 1] - \mathbb{E}[Y|D = 0] = \mathbb{E}[Y_1|D = 1] - \mathbb{E}[Y_0|D = 0] = \beta_{OLS} \neq \text{ATE}$$

	$\rho = 0.5$	$\rho = 0$	$\rho = -0.5$
OLS	1.406	1.076	0.920
ATE	-0.007	-0.003	0.013
ATT	1.383	1.785	2.145
ATUT	-1.392	-1.779	-2.139
$\mathbb{E}[Y\|D = 1] - \mathbb{E}[Y\|D = 0]$	1.406	1.076	0.920





f) The claim is incorrect.

First, consider $D \perp\!\!\!\perp (U_1, U_0)$, this does not imply anything about the dependence between U_1 and

U_0. They could be correlated positively or negatively or not correlated at all.

Second, consider the case, when $\rho = 0$:

$$\begin{pmatrix} U_1 \\ U_0 \end{pmatrix} \sim \mathcal{N}\Bigg(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 1 \end{pmatrix} \Bigg)$$

$$D = \mathbb{1}[U_1 > U_0]$$

Note U_1 is still not independent from D. To show this note that mean independence is implied by independence, meaning if there is no mean independence then there is no independence. Borrowing from **a)**:

$$\mathbb{E}[U_1|D=1] = \mathbb{E}[U_1|U_0 - U_1 < 0] = \frac{\sigma^2 - \rho\sigma}{\sqrt{\sigma^2 + 1 - 2\rho\sigma}} \frac{\phi(0)}{\Phi(0)} = [\rho = 0] = \frac{\sigma^2}{\sqrt{\sigma^2 + 1}} \frac{\phi(0)}{\Phi(0)} \neq 0 = \mathbb{E}[U_1]$$

Note also that $D \perp\!\!\!\perp (U_1, U_0)$ implies $ATE = ATT = ATUT$ which does not hold in the computational exercise for $\sigma = 2$ and $\rho = 0$.

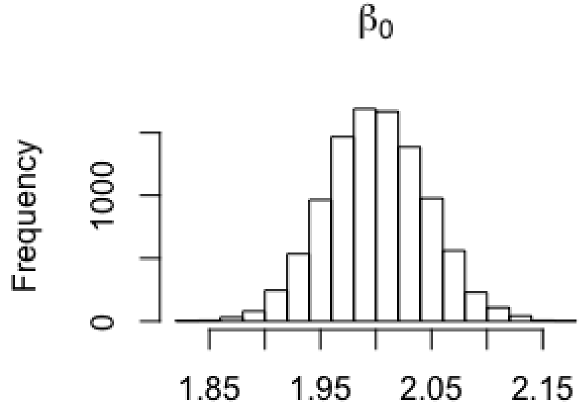
Problem 3

Monte Carlo Simulations

1) OLS estimates: $\beta = \begin{pmatrix} 1.987 \\ 2.9998 \end{pmatrix}$, std. err. $= \begin{pmatrix} 0.046 \\ 0.0045 \end{pmatrix}$

2) Monte Carlo simulations results: std. err. $= \begin{pmatrix} 0.045 \\ 0.0044 \end{pmatrix}$

To justify the transition "?" we need to apply LLN to each of the summations to show its convergence in probability to corresponding expectations and Continuous Mapping Theorem (or Slutsky's Lemma in some places) for convergence in probability to justify operations(squares, square roots).

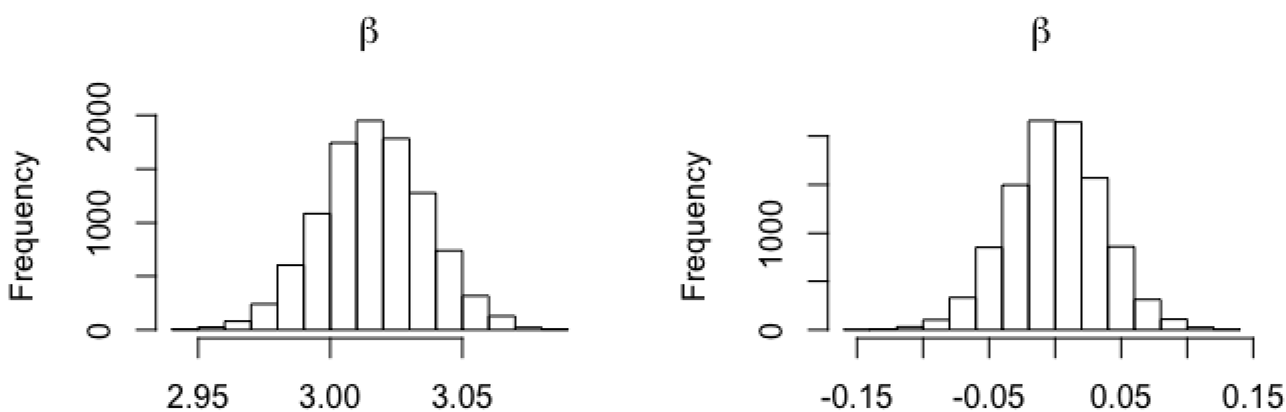


Results for MC simulation of β_0

Nonparametric Bootstrap

1) OLS estimates: $\beta = 3.016$ std.err.$= 0.020$. As we showed in class $\beta_{OLS} = \mathbb{E}[Y_1|D=1] - \mathbb{E}[Y_0|D=0] \underbrace{=}_{D \perp\!\!\!\perp (Y_1,Y_0)} \mathbb{E}[Y_1 - Y_0] = 3$. One way to convince to "feel" the independence of D and (Y_1, Y_0) in this example is to change the value of $\mathbb{P}[D=1]$ and observe that the estimate for β does not change.

2) Bootstrap estimates: std. err.$= 0.020$. If one decides to draw Y and D separately then they end up with a sample of independent explanatory variables and outcomes, meaning $\beta = \frac{\text{Cov}(Y,D)}{\text{Var}(D)} = 0$.



Left: Results of Bootstrap for β
Right: Separate draws of Y and D