Econometrics II (Spring 2024)
Suggested Solutions to Problem Set 1

Shuhei Kainuma*

April 16, 2024

Question 1

1.

We can define the individual treatment effect of i, denoted by τ_i, as

$$\tau_i = Y_i(1) - Y_i(0).$$

However, we will not be able to know τ_i for each i. This is because, as the fundamental problem of causal inference tells, for every i, we only observe either one of $D_i = 0$ or $D_i = 1$; hence, $Y_i(1)$ and $Y_i(0)$ are not observed at the same time.

The individual treatment effect τ_i to be well-defined under the Stable Unit Treatment Value Assumption (SUTVA). This assumption consists of two parts:

1. No hidden treatment variation, i.e., $\forall i,\ \forall D_i = D_i',\ Y_i(D_i, ...) = Y_i(D_i', ...)$;
2. No interference: i.e., $\forall i,\ Y_i(d_1, \ldots, d_N) = Y_i(d_i)$.

If the first part is not satisfied, then there exists some i for which $D_i = d \neq 0, 1$; then, either $Y_i(1)$ or $Y_i(0)$ is not well-defined, nor is τ_i. Similarly, if the no interference condition is violated, then, for some i, $Y_i(d_i), d_i \in \{0, 1\}$ is not well-defined.

2.

We can define the average treatment effect on the treated (ATT), denoted by τ, as

$$\tau = \mathbb{E}\left[\tau_i \mid D_i = 1\right] = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1\right].$$

One assumption under which τ is identified is the following:

$$\mathbb{E}\left[Y_i(0) \mid D_i = 0\right] = \mathbb{E}\left[Y_i(0) \mid D_i = 1\right].$$

*shuhei.kainuma@iies.su.se. Let me know if you spot any typos/errors!

To see this, suppose this holds; then,

$$
\begin{aligned}
\tau &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1\right] \\
&= \mathbb{E}\left[Y_i(1) \mid D_i = 1\right] - \underbrace{\mathbb{E}\left[Y_i(0) \mid D_i = 1\right]}_{=[Y_i(0) \ \mid \ D_i=0]} \\
&= \mathbb{E}\left[Y_i(1) \mid D_i = 1\right] - \mathbb{E}\left[Y_i(0) \mid D_i = 0\right] \\
&= \underbrace{\mathbb{E}\left[Y_i \mid D_i = 1\right]}_{\text{known population quantity}} - \underbrace{\mathbb{E}\left[Y_i \mid D_i = 0\right]}_{\text{known population quantity}} .
\end{aligned} \tag{1}
$$

Intuitively, the assumption means that on average the individuals with $D_i = 1$ and those with $D_i = 0$ are not different; hence, we can use the observed outcome of the $D_i = 0$ group to impute the unobserved quantity $\mathbb{E}[Y_i(0)|D_i = 1]$.

3.

For each type $x \in \{1, 2, 3\}$, we can define conditional ATT, denoted by $\tau(x)$, as

$$
\tau(x) = \mathbb{E}\left[\tau_i \mid D_i = 1, X_i = x\right].
$$

This is identified under the following assumptions: $\forall x \in \{1, 2, 3\}$,

$$
\begin{aligned}
&\mathbb{E}\left[Y_i(0) \mid D_i = 0, X_i = x\right] = \mathbb{E}\left[Y_i(0) \mid D_i = 1, X_i = x\right], \\
&0 < e(X_i) < 1.
\end{aligned}
$$

To see this, for each $x \in \{1, 2, 3\}$,

$$
\begin{aligned}
\tau(x) &= \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1, X_i = x\right] \\
&= \mathbb{E}\left[Y_i(1) \mid D_i = 1, X_i = x\right] - \underbrace{\mathbb{E}\left[Y_i(0) \mid D_i = 1, X_i = x\right]}_{=\mathbb{E}[Y_i(0) \ \mid \ D_i=0, X_i=x]} \\
&= \mathbb{E}\left[Y_i(1) \mid D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i(0) \mid D_i = 1, X_i = x\right] \\
&= \underbrace{\mathbb{E}\left[Y_i \mid D_i = 1, X_i = x\right]}_{\text{known population quantity}} - \underbrace{\mathbb{E}\left[Y_i \mid D_i = 0, X_i = x\right]}_{\text{known population quantity}}.
\end{aligned}
$$

Why do we need the second assumption? Consider the type $X_i = 3$ with $e(3) = 1$, indicating the violation of this assumption. Then, $\mathbb{E}[Y_i \mid D_i = 0, X_i = 3]$ is undefined, because $e(X_i = 3) = \Pr(D_i = 1 \mid X_i = 3) = 3/3 = 1$ and

$$
\begin{aligned}
\mathbb{E}\left[Y_i \mid D_i = 0, X_i = 3\right] &= \int y \Pr(Y_i = y \mid D_i = 0, X_i = 3) dy \\
&= \int y \frac{\Pr(Y_i = y, D_i = 0, X_i = 3)}{\Pr(D_i = 0, X_i = 3)} dy \\
&= \int y \frac{\Pr(Y_i = y, D_i = 0, X_i = 3)}{\underbrace{\Pr(D_i = 0 \mid X_i = 3)}_{=0} \Pr(X_i = 3)} dy
\end{aligned}
$$

The similar intuition as above applies: because the assumption ensures that for a given $x \in \{1, 2, 3\}$, the individuals with $D_1 = 1$ and with $D_i = 0$ are not different on average, we can use $\mathbb{E}[Y_i|D_i = 0, X_i = 0]$ for the unobserved $\mathbb{E}[Y_i(0)|D_i = 1, X_i = 0]$. Besides, the second condition, typically called the common support, ensures that these quantities are well-defined.

4.

An estimator of $\tau(x)$ can be

$$\hat{\tau}(x) = \frac{1}{\sum_{i:X_i=x} D_i} \sum_{i:X_i=x} D_i Y_i - \frac{1}{\sum_{i:X_i=x}(1-D_i)} \sum_{i:X_i=x} Y_i(1-D_i).$$

Given the data, the CATT for $X_i = 1$ and $X_i = 2$ are estimated as

$$\begin{aligned}
\hat{\tau}(1) &= \frac{1}{1} \cdot 2 - \frac{1}{2} \cdot (5+3) = -2, \\
\hat{\tau}(2) &= \frac{1}{1} \cdot 4 - \frac{1}{1} \cdot 5 = -1.
\end{aligned}$$

$\tau(3)$ is not identified in this case, as the common support condition $0 < e(X_i = 3) < 1$ is violated.

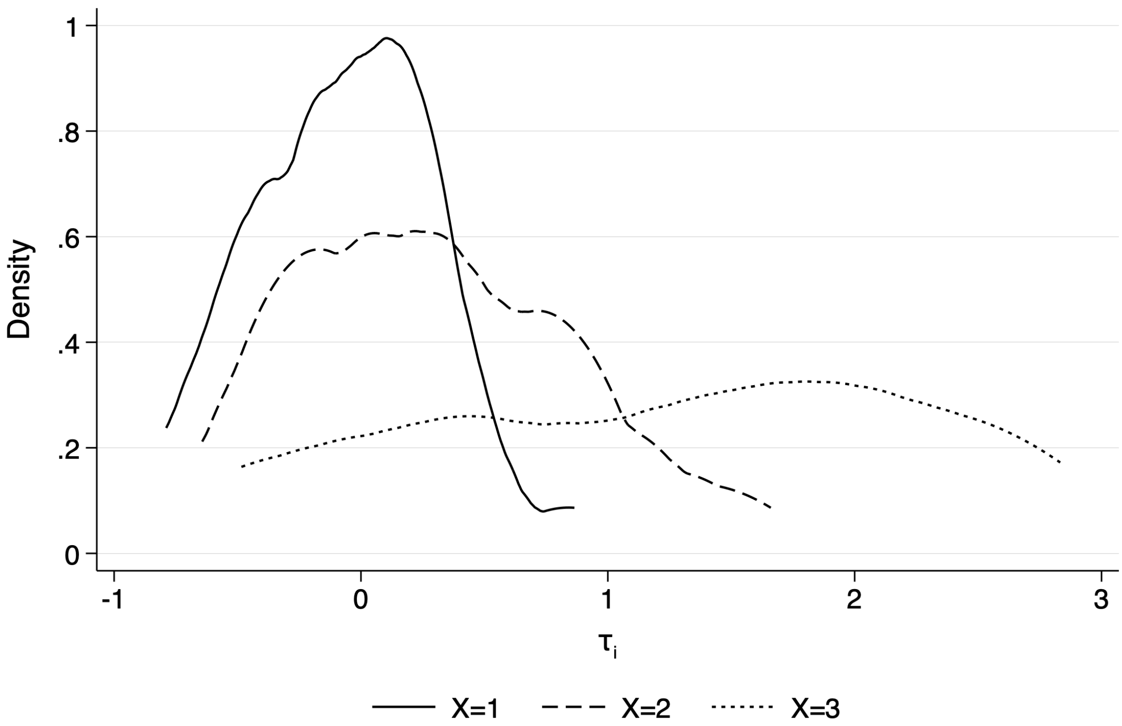
5.

See the do-file for Stata implementation.

6.

See the do-file for Stata implementation. Figure 1 presents the kernel density estimate of the individual treatment effect τ_i distribution for each value of $X_i \in \{1, 2, 3\}$. As we simulate, the distribution centres around 0 for $X_i = 1$, and it shifts rightwards and flattens as the value of X_i becomes larger.

Figure 1: Distribution of τ_i for each $X_i \in \{1, 2, 3\}$



7.

See the do-file for Stata implementation.

Table 1 shows the estimation results along with the true values. Column (1) corresponds to the ATT estimates with the simple regression/difference-in-means estimator. We obtain the estimate close to the true value. This is because $\mathbb{E}[Y_i(0)|D_i = 0] = \mathbb{E}[Y_i(0)|D_i = 1]$ holds, as we have seen in Question 1.2; hence, the sample analogue of Equation 1 works.

Column (2)-(4) present the CATT estimate for each value of X_i. The CATT estimates for $X_i = 1$ and $X_i = 2$ are close to their true values, respectively, while we cannot estimate the CATT for $X_i = 3$ that is not identified.

Note that we can estimate the CATT in a single, fully-saturated regression

$$Y_i = \sum_x \left(\alpha_x \, \mathbb{1}\{X_i = x\} + \beta_x \, D_i \times \mathbb{1}\{X_i = x\}\right) + \varepsilon_i,$$

where $\hat{\beta}_x$ is the estimator of the CATT at each value of $X_i = x$. We can see this in Column (5), where we obtain the identical point estimates for $\tau(1)$ and $\tau(2)$.

While One can say that they estimate the ATE using the difference-in-means estimator, we can also use a different estimator in this particular setting. Using the weight average formula below, we get the ATE estimate = 0.339, close to the true ATE value.

Table 1: 1.7. Estimating ATE, ATT, and CATT

	(1)	(2)	(3)	(4)	(5)
	ATE/ATT	X=1	X=2	X=3	CATT: Stacked Reg.
D	0.577	-0.0593	0.309	0	
	(0.142)	(0.0872)	(0.188)	(.)	
D $\times 1\{X = 1\}$					-0.0593
					(0.172)
D $\times 1\{X = 2\}$					0.309
					(0.243)
D $\times 1\{X = 3\}$					0
					(.)
N	100	50	25	25	100
$\widehat{\text{ATE}}$	0.339				
True ATE	0.345				
True ATT	0.558				
True CATT		-0.118	0.231	1.212	

Standard errors in parentheses

Identification of ATE We can show that the ATE $\equiv \mathbb{E}[\tau_i]$ is point-identified in this case (but with a specific set of assumptions). There are multiple ways (assumptions) to approach identification of the ATE. A typical assumption is the mean independence of $Y_i(d)$, i.e., $\forall d \in \{0,1\}$,

$$\mathbb{E}\left[Y_i(d) \mid D_i = 1\right] = \mathbb{E}\left[Y_i(d) \mid D_i = 0\right].$$

Under this assumption, the difference-in-means in Equation 1 identifies $\mathbb{E}[\tau_i]$. Unfortunately, this is not satisfied in our setting, because the $X_i = 3$ group, which has on average the largest values of $Y_i(1)$, exhibits $\Pr(X_i = 3|D_i = 0) = 1 - e(3) = 0$. This imbalance in X_i causes the unbalanced $Y_i(1)$ distribution across two groups.

We can, however, work with another set of assumptions:

$$\begin{aligned}\forall x \in \text{Supp}(X_i|D_i = 0),\ \mathbb{E}\left[Y_i(1) \mid D_i = 1, X_i = x\right] &= \mathbb{E}\left[Y_i(1) \mid D_i = 0, X_i = x\right], \\ \mathbb{E}\left[Y_i(0) \mid D_i = 1\right] &= \mathbb{E}\left[Y_i(0) \mid D_i = 0\right].\end{aligned}$$

To show the identification of the ATE, first observe that

$$\begin{aligned}\mathbb{E}\left[\tau_i\right] &= \mathbb{E}\left[\mathbb{E}\left[\tau_i \mid D_i\right]\right] \\ &= \underbrace{\mathbb{E}\left[\tau_i \mid D_i = 1\right]}_{=\text{ATT}} \Pr\left(D_i = 1\right) + \mathbb{E}\left[\tau_i \mid D_i = 0\right] \Pr\left(D_i = 0\right).\end{aligned}$$

The ATT is identified under the second assumption, as shown in Question 1.2. Since $\Pr(D_i)$ is known quantity, we want to show the identification of $\mathbb{E}\left[\tau_i \mid D_i = 0\right]$ under the first assumption.[1] Observe that

$$\begin{aligned}\mathbb{E}\left[\tau_i \mid D_i = 0\right] &= \mathbb{E}\left[\mathbb{E}\left[\tau_i \mid D_i = 0, X_i = x\right] \mid D_i = 0\right] \\ &= \sum_{x \in \{1,2\}} \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 0, X_i = x\right] \Pr(X_i = x \mid D_i = 0) \\ &\qquad\qquad (\because \text{ You can show } \Pr(X_i = 3|D_i = 0) = 0) \\ &= \sum_{x \in \{1,2\}} \{\underbrace{\mathbb{E}\left[Y_i(1) \mid D_i = 0, X_i = x\right]}_{=\mathbb{E}[Y_i(1) \mid D_i = 1, X_i = x]} - \mathbb{E}\left[Y_i(0) \mid D_i = 0, X_i = x\right]\} \Pr(X_i = x \mid D_i = 0) \\ &= \sum_{x \in \{1,2\}} \{\mathbb{E}\left[Y_i(1) \mid D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i(0) \mid D_i = 0, X_i = x\right]\} \Pr(X_i = x \mid D_i = 0) \\ &= \sum_{x \in \{1,2\}} \tau(x) \Pr(X_i = x \mid D_i = 0),\end{aligned}$$

Because $\tau(x)$ are shown to be identified in Question 1.3. and $\Pr(X_i|D_i = 0)$ is known quantity, we establish the identification of $\mathbb{E}\left[\tau_i \mid D_i = 0\right]$. It also indicates that under these assumptions, we can estimate this as a weighted average of the CATT.

In our simulation, these two assumptions hold; hence, the ATE is identified. We can then estimate the ATE using the weighted-average formula above, plugging in corresponding quantities. The ATE estimate is shown in Table 1.

[1] It is called average treatment effect on the untreated (ATU), or on the non-treated (ATNT).

Question 2

a.

The average treatment effect on the treated (ATT) is NOT point-identified without further assumptions, because

$$\begin{aligned}\tau \equiv \mathbb{E}\left[\tau_i | D_i = 1\right] &= \mathbb{E}\left[Y_i(1) - Y_i(0) | D_i = 1\right] \\ &= \underbrace{\mathbb{E}\left[Y_i | D_i = 1\right]}_{=\mathbb{E}[Y_i(1)|D_i=1]} - \mathbb{E}\left[Y_i(0) | D_i = 1\right] + \underbrace{\mathbb{E}[Y_i(0) - Y_i(0) | D_i = 0]}_{=0} \\ &= \underbrace{\left\{\mathbb{E}\left[Y_i | D_i = 1\right] - \mathbb{E}\left[Y_i | D_i = 0\right]\right\}}_{\text{known population quantity}} - \underbrace{\left\{\mathbb{E}[Y_i(0) | D_i = 1] - \mathbb{E}[Y_i(0) | D_i = 0]\right\}}_{\neq 0 \quad \text{without further assumptions}}.\end{aligned}$$

b.

We will show that τ is (point-)identified under the following two assumptions:

$$\begin{aligned}&\mathbb{E}[Y_i(0) | D_i = 0, X_i = x] = \mathbb{E}[Y_i(0) | D_i = 1, X_i = x] \\ &\Pr(D_i = 1 | X_i = x) < 1 \ \forall x.\end{aligned}$$

We consider the difference in means of Y_i conditional on X_i and D_i, integrated over the distribution of X_i for the group with $D_i = 1$, i.e.,

$$\mathbb{E}\left[Y_i | D_i = 1\right] - \mathbb{E}\left[\mathbb{E}\left[Y_i | D_i = 0, X_i = x\right] | D_i = 1\right],$$

where $\mathbb{E}\left[Y_i | D_i = 1\right]$ and $\mathbb{E}\left[Y_i | D_i = 0, X_i\right]$ are known quantities. The assumption $\Pr(D_i = 1 | X_i = x) < 1 \ \Leftrightarrow \ \Pr(D_i = 0 | X_i = x) > 1$ for every value of X_i ensures that $\forall x, \mathbb{E}\left[Y_i | D_i = 0, X_i = x\right]$ is well-defined.[2] Then, observe that

$$\begin{aligned}&\mathbb{E}\left[Y_i | D_i = 1\right] - \mathbb{E}\left[\mathbb{E}\left[Y_i | D_i = 0, X_i = x\right] | D_i = 1\right] \\ =&\mathbb{E}[\underbrace{Y_i(1)}_{Y_i(0)+\tau_i} | D_i = 1] - \mathbb{E}[\mathbb{E}[Y_i(0) | D_i = 0, X_i = x] | D_1 = 1] \\ =&\underbrace{\mathbb{E}[\tau_i | D_i = 1]}_{=\tau} + \mathbb{E}\left[\mathbb{E}[Y_i(0) | D_i = 1, X_i = x] - \mathbb{E}[Y_i(0) | D_i = 0, X_i = x] | D_1 = 1\right] \\ =&\tau + \underbrace{\mathbb{E}[\underbrace{\mathbb{E}[Y_i(0) | X_i = 0] - \mathbb{E}[Y_i(0) | X_i = x]}_{=0} | D_1 = 1]}_{=0} \\ =&\tau.\end{aligned}$$

Now we comptue the true value of τ given $\Pr(X_i = 1) = 0.3, \alpha_0 = 1$, and $\alpha_1 = -2$. Suppose that X_i is independent of D_i, then

[2] See Question 2.c. for the necessity of the common support assumption.

$$
\begin{aligned}
\tau &= \mathbb{E}\left[\tau_i | D_i = 1\right] \\
&= \mathbb{E}\left[\mathbb{E}[\alpha_0 + \alpha_1 x | D_i = 1, X_i = x] | D_i = 1\right] \\
&= \alpha_0 + \alpha_1 \cdot \Pr\left(X_i = 1\right) \\
&= 1 - 2 \cdot 0.3 \\
&= 0.4.
\end{aligned}
$$

c.

In addition to the assumption in the previous question, assume that $0 < \Pr(D_i = 1 | X_i = x) < 1,\ \forall x \in \{0, 1\}$. We will then show that the conditional ATT (CATT) $\tau(X_i) = \mathbb{E}[\tau_i | D_i = 1, X_i]$ is identified. Observe that for any $x \in \{0, 1\}$,

$$
\begin{aligned}
\tau(x) &= \mathbb{E}\left[\tau_i | D_i = 1, X_i = x\right] \\
&= \mathbb{E}\left[\tau_i + \left(Y_i(0) - Y_i(0)\right) | D_i = 1, X_i = x\right] \\
&= \mathbb{E}\left[Y_i(0) + \tau_i | D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i(0) | D_i = 1, X_i = x\right] \\
&= \mathbb{E}\left[Y_i(1) | D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i(0) | D_i = 0, X_i = x\right] \\
&= \mathbb{E}\left[Y_i | D_i = 1, X_i = x\right] - \mathbb{E}\left[Y_i | D_i = 0, X_i = x\right],
\end{aligned}
$$

indicating that $\tau(x)$ is point-identified as the difference-in-means between the groups $D_i = 1$ and $D_i = 0$ at $X_i = x$.

Why do we need this condition for the identification of $\tau(X_i)$? Suppose instead that $\Pr(D_i = 1 | X_i = 1) = 0$. Then, $\mathbb{E}[Y_i | D_i = 1, X_i = 1]$ becomes

$$
\begin{aligned}
\mathbb{E}\left[Y_i | D_i = 1, X_i = 1\right] &= \int y \cdot \Pr\left(Y_i = y | D_i = 1, X_i = 1\right) dy \\
&= \int y \cdot \frac{\Pr\left(Y_i = y, D_i = 1, X_i = 1\right)}{\underbrace{\Pr\left(D_i = 1, X_i = 1\right)}_{= \Pr(D_i = 1 | X_i = 1) \Pr(X_i = 1) = 0}} dy,
\end{aligned}
$$

meaning that $\mathbb{E}[Y_i | D_i = 1, X_i = 1]$ is undefined. Therefore, this condition is necessary for identification. The condition $0 < \Pr(D_i = 1 | X_i) < 1$ is called *common support*, *overlap*, or *positivity*. Intuitively, it ensures that we have both $D_i = 1$ and $D_i = 0$ for every value of X_i.

Then, given the numerical values $\alpha_0 = 1$ and $\alpha_1 = -2$, the true values for $\tau(x)$ for $x = 0, 1$ are given by

$$
\begin{aligned}
\tau(0) &= \mathbb{E}\left[\tau_i | D_i = 1, X_i = 0\right] = \mathbb{E}\left[\alpha_0 + \alpha_1 \cdot 0\right] = \alpha_0 = 1, \\
\tau(1) &= \mathbb{E}\left[\tau_i | D_i = 1, X_i = 1\right] = \mathbb{E}\left[\alpha_0 + \alpha_1 \cdot 1\right] = \alpha_0 + \alpha_1 = -1.
\end{aligned}
$$

d.

Since we have $N = 5$ individuals in total and the number of treated individuals is fixed as $n_1 = 3$, the total number of assignments for this setup can be obtained as

$$
\binom{5}{3} = \frac{5 \cdot 4 \cdot 3}{3 \cdot 2} = 10.
$$

e.

Now we consider the case where the treatment is assigned to exactly two individuals with $X_i = 0$ and one individual with $X_i = 1$. Then, the total number of assignments is given by

$$\binom{3}{2} \cdot \binom{2}{1} = 3 \cdot 2 = 6.$$

When it comes to the unit assignment probability, notice that the treatment assignment in this case is a completely randomised experiment *within* each group of X_i. We can hence simply consider assignment separately for each group of X_i. For the group with $X_i = 0$, within which 2 out of 3 observations are assigned to treatment, the unit assignment probability is given by $n_1(x = 0)/N(x = 0) = 2/3$. Likewise, for the group with $X_i = 1$, within which 1 out of 2 observations is treated, the unit assignment probability is given by $n_1(x = 1)/N(x = 1) = 1/2$.

Since the unit assignment probability is constant within each group of $X_i = x$, $\{0, 1\}$, the corresponding propensity score is equal to the unit assignment probability, that is, the propensity score is also given by

$$e(0) = \frac{2}{3},\ e(1) = \frac{1}{2}.$$

f.

See the do-file for the Stata implementation. In either of the simple averaging and the regression approaches, we obtain identical estimates, i.e., $\hat{\tau} = \hat{\tau}_{\text{OLS}} = 0.468$, which is close to the true value of the ATT $\tau = 0.4$ computed in part (b) above. This result shows that a simple difference-in-means estimator and the OLS estimator are equivalent.

Note that this is why we often run the OLS in place of a difference-in-means; the OLS makes it easier to do things beyond computing the difference, such as inference. Yet, the equivalence does not hold in general in the presence of covariates.

g.

See the do-file for the Stata implementation. In our simulation, we obtain $\hat{\tau}(x = 0) = 1.138$, $\hat{\tau}(x = 1) = -0.859$. They are close to the true values compute in part (c), i.e., $\tau(x = 0) = 1$, $\tau(x = 1) = -1$. As in part (f), the estimated CATTs are identical between the diffrence-in-means and OLS estimators of the regression of Y_i on D_i, X_i, $D_i \times X_i$, and the constant term.[3] Note that the OLS coefficent on D_i corresponds to $\hat{\tau}(x = 0)$ and the sum of the coefficents on D_i and D_iX_i represents $\hat{\tau}(x = 1)$.

h.

See the do-file for the Stata implementation. In our simulation, we generate the data so that exactly 30 observations are female ($X_i = 1$), and 15 of them are treated in Experiment 2, instead of strictly following $X_i \sim$ Bernoulli(p). In Experiment 2, we follow the fixed-number-of-women

[3] We can also run the regression of Y_i on $D_i \times X_i$, $D_i \times (1 - X_i)$, X_i, and the constant. In this case, you could simply take the coefficents of the interaction terms for the corresponding CATT estimates.

requirement by treating 15 of oberavtions with $X_i = 1$ and $15 \times 2 = 30$ observations with $X_i = 0$. Hence, the total number of the treated observations is $15 + 30 = 45$.

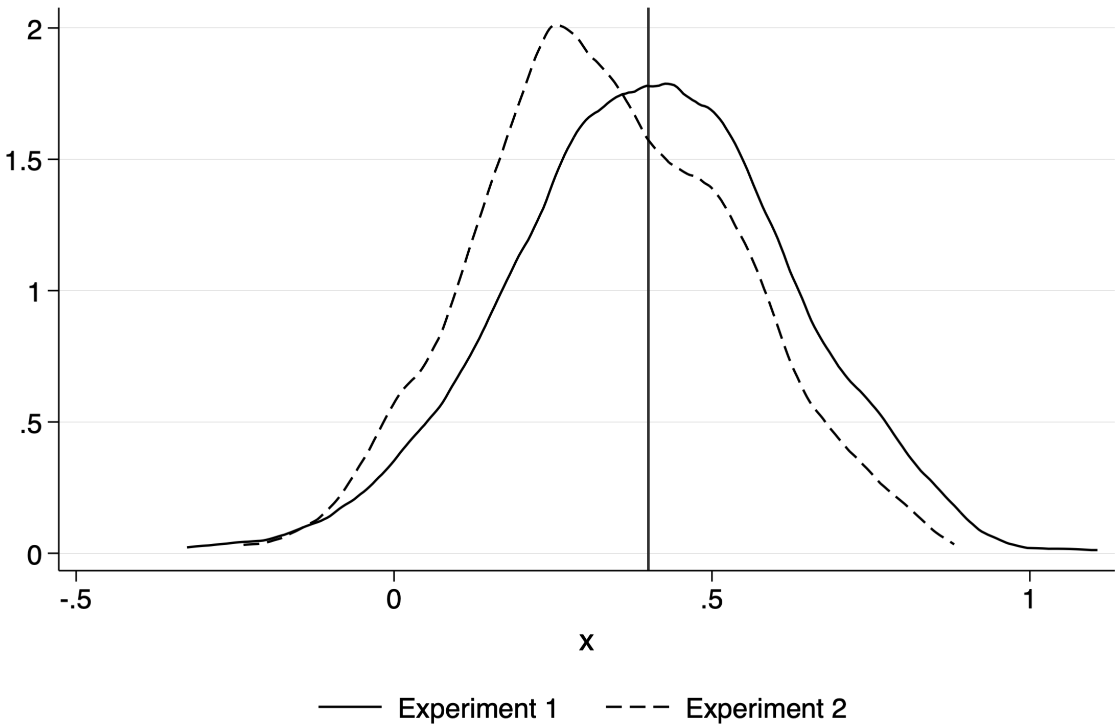
Figure 2 presents the distribution of the estimates of ATT and CATTs from Experiment 1 and 2. Figure 2a corresponds to the ATT estimates. The ATT estimates in Experiment 2 are slightly biased downwards. This is because, in Experiment 2, the share of male observations $X_i = 0$ are imbalaned across the treatment and control groups (40 vs. 30), leading to the incorrect weights on subgroups' CATTs.

On the other hand, looking at Figure 2b, the estimates of each subgroup's CATT do NOT seem biased. Rather, particularly for $x = 1$, the estimates are somewhat more precisely estimated in Experiment 2. This is because, since the share of the treatment and control observations in the subgroup is fixed in Experiment 2, we effectively eliminate the cases where the treatment and control observations are imbalanced within each subgroup. For $x = 1$, specifically the ratio of the numbers of treated to control units is one-to-one, which minimises the variance.[4]

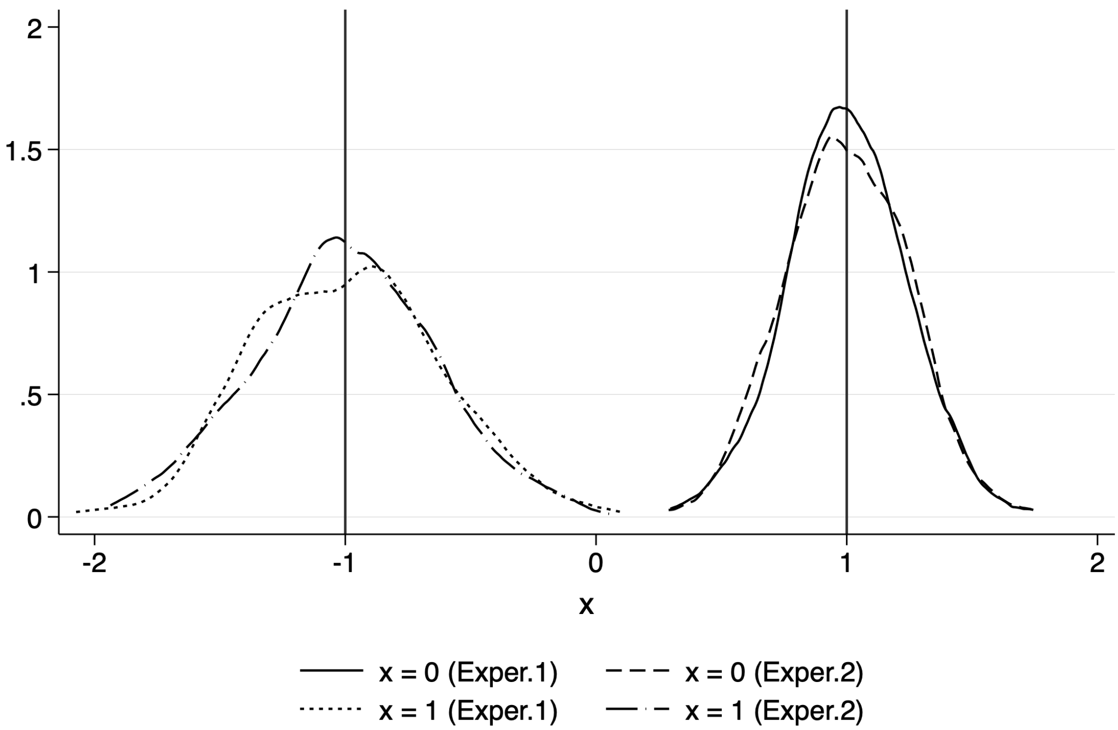
[4]In a simple complete random experiment, the optimal ratio of the treated and control observations, which minimises the variance of the difference-in-means estimator, depends on the variance of the potential outcomes under treatment and control, $\{Y_i(1), Y_i(0)\}_i$. Since in our case, the variance of the potential outcomes is identical, $n_1/N = 1/2$ minimises the variance of the estimator. Note that in a given simulated sample, the variance may not be the same.

A more subtle difference of the CATT estimates for $x = 0$ comes from the fact that while the numbers of treated and control observations are 30 and 40, so not $n_1/N \neq 1/2$, fixing these numbers eliminates the draws with highly unbalanced sizes of treatment and control.

Figure 2: Distribution of the Estimates from Experiment 1 and 2



(a) Distribution of $\hat{\tau}$ in Experiment 1 and 2



(b) Distribution of $\hat{\tau}(x)$ in Experiment 1 and 2

Question 3

a.

First, we consider the probability of the first unit being treated. The probability that the first unit is treated is equal to the share of the assignments in which the first unit is treated within all the potential assignments. The number of the assignments in which the first unit is treated is equal to the total number of the assignments for the rest of four units, since we fix the treatment status of the first unit, for a given n_1, that is, $\binom{N-1}{n_1-1} = \binom{4}{n_1-1}$. Given $n_1 < N$, the total number of the treatment assignments is $\binom{N}{n_1} = \binom{5}{n_1}$. Therefore, given n_1, the probability of the first unit being treated is given by

$$\frac{\binom{N-1}{n_1-1}}{\binom{N}{n_1}} = \frac{\frac{(N-1)!}{(n_1-1)!(N-n_1)!}}{\frac{N!}{n_1!(N-n_1)!}} = \frac{n_1}{N} = \frac{n_1}{5}.$$

Note that since this applies to the other units, the unit assignment probability is equal to $n_1/N = n_1/5$.

Second, similarly, the probability that the first two units are both treated equals the total number of the assignments for the other three units (fixing the *treated* status of the first two units) divided by the total number of the assignments. Hence, we have

$$\frac{\binom{N-2}{n_1-2}}{\binom{N}{n_1}} = \frac{\frac{(N-2)!}{(n_1-2)!(N-n_1)!}}{\frac{N!}{n_1!(N-n_1)!}} = \frac{n_1(n_1-1)}{N(N-1)} = \frac{n_1(n_1-1)}{20}.$$

b.

We will show that the Sample Average Treatment Effect (SATE) is identified. Let $\mathcal{O} \equiv \{\mathbf{Y}(\mathbf{0}), \mathbf{Y}(\mathbf{1})\}_{i=1}^{N}$. In this case, we define a quantity of interest to be *identified* if it is represented by the known quantities with data containing all the possible treatment assignment under the given assignment mechanism. That is, assuming that we know all the information from every assignment $\left(\{Y_i, D_i\}_{i,d\in\{\mathbf{D}:\ \sum_i D_i = n_1\}}\right)$, we ask whether we can express the quantity of interest with the information at hand.

First note that since this is a compelete randomised experiment,

$$\mathbb{E}\left[D_i|\mathcal{O}\right] = \Pr(D_i = 1|\mathcal{O}) = \frac{n_1}{N}.$$

Besides, by assumption, $\sum_i D_i = n_1$ and $\sum_i (1-D_i) = n_0$. Then, observe that

$$
\begin{aligned}
\tau \equiv N^{-1}\sum_{i=1}^{N}\{Y_i(1)-Y_i(0)\} &= \frac{1}{N}\sum_i Y_i(1) - \frac{1}{N}\sum Y_i(0) \\
&= \frac{1}{n_1}\cdot\frac{n_1}{N}\sum_i Y_i(1) - \frac{1}{n_0}\frac{n_0}{N}\sum_i Y_i(0) \\
&= \frac{1}{n_1}\sum_i \Pr(D_i=1|\mathcal{O}) \underbrace{Y_i(1)}_{=\ \mathbb{E}_D[Y_i(1)|D_i=1,\mathcal{O}]\ =\ \mathbb{E}_D[Y_i|D_i=1,\mathcal{O}]} \\
&\quad - \frac{1}{n_0}\sum_i \Pr(D_i=0|\mathcal{O}) \underbrace{Y_i(0)}_{=\ \mathbb{E}[Y_i(0)|D_i=0,\mathcal{O}]\ =\ \mathbb{E}[Y_i|D_i=0,\mathcal{O}]} \\
&= \frac{1}{\underbrace{n_1}_{\sum_j D_j}}\sum_i \mathbb{E}_D\left[Y_iD_i|\mathcal{O}\right] - \frac{1}{\underbrace{n_0}_{\sum_j(1-D_j)}}\sum_i \mathbb{E}_D\left[Y_i(1-D_i)|\mathcal{O}\right] \\
&= \mathbb{E}_D\left[\frac{1}{\sum_j D_j}\sum_{i:\ D_i=1} Y_i \ - \ \frac{1}{\sum_j(1-D_j)}\sum_{i:\ D_i=0} Y_i|\mathcal{O}\right],
\end{aligned}
$$

where $\frac{1}{\sum_j D_j}\sum_{i:\ D_i=1} Y_i$ and $\frac{1}{\sum_j(1-D_j)}\sum_{i:\ D_i=0} Y_i$ are the average outcomes for the observations with $D_i = 1$ and $D_i = 0$ *given* a treatment assignment. Thus, τ can be represented by the average of average differences in the observed outcomes of treatment and control groups over all the possible assignments. Because these are known quantities, it establishes that τ can be uniquely represented with these know quantities, and does the identification of τ.

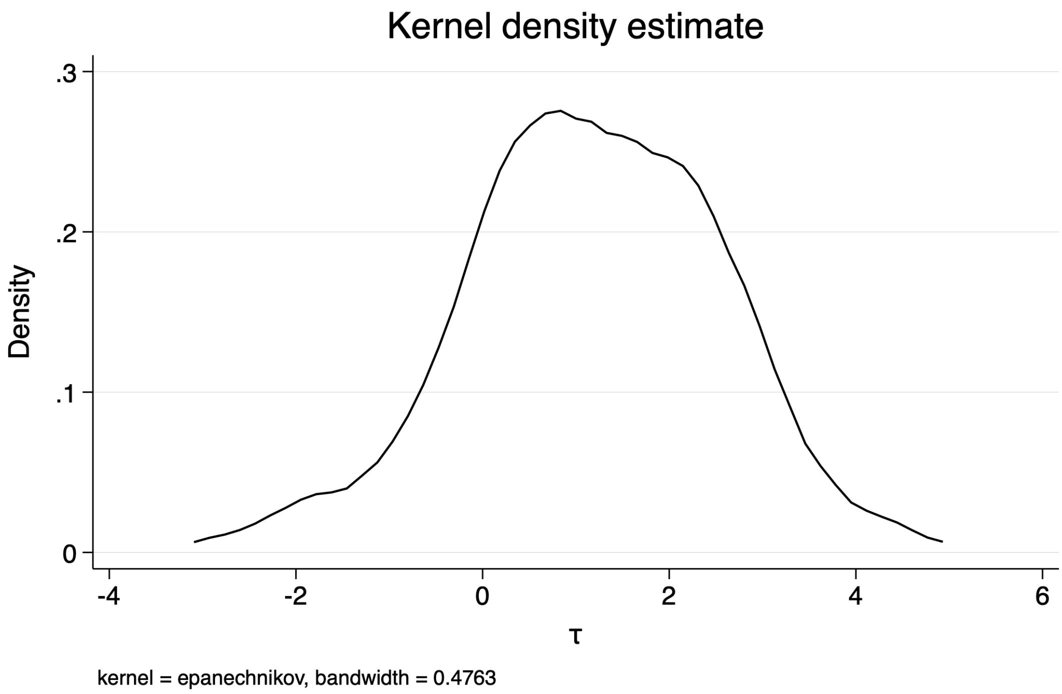
c.

We will show that $\hat{\tau}$ is an unbiased estimator for τ conditional on potential outcome $\mathcal{O}$, that is, $\mathbb{E}_D[\hat{\tau}|\mathcal{O}] \equiv \mathbb{E}[\hat{\tau}|\mathcal{O}] = \tau$, where the expectation is taken over the repeated treatment assignment. Note that the answer part (b) above immediately implies the unbiasedness. Nonetheless, let us prove it again.

Similar to part (b), we have

$$
\begin{aligned}
\mathbb{E}[\hat{\tau}|\mathcal{O}] &= \mathbb{E}\left[\frac{\sum_i Y_iD_i}{\sum_j D_j} - \frac{\sum_i Y_i(1-D_i)}{\sum_j(1-D_i)}|\mathcal{O}\right] \\
&= \mathbb{E}\left[\frac{\sum_i Y_iD_i}{n_1} - \frac{Y_i(1-D_i)}{n_0}|\mathcal{O}\right] \\
&= \frac{1}{n_1}\sum_i \mathbb{E}\left[\mathbb{E}\left[Y_i|D_i,\mathcal{O}\right]D_i|\mathcal{O}\right] - \frac{1}{n_0}\mathbb{E}\left[\mathbb{E}\left[Y_i(1-D_i)|D_i,\mathcal{O}\right]|\mathcal{O}\right] \\
&= \frac{1}{n_1}\sum_i \underbrace{\mathbb{E}\left[Y_i(1)|D_i=1,\mathcal{O}\right]}_{=Y_i(1)}\underbrace{\Pr(D_i=1|\mathcal{O})}_{=n_1/N} - \frac{1}{n_0}\sum_i \underbrace{\mathbb{E}[Y_i(0)|D_i=0,\mathcal{O}]}_{=Y_i(0)}\underbrace{\Pr(D_i=0|\mathcal{O})}_{=n_0/N} \\
&= \frac{1}{N}\sum_i \{Y_i(1)-Y_i(0)\} = \tau.
\end{aligned}
$$

Figure 3: 3.d. Nonparametric density estimate of simulated τ_i



d.

See the do-file for the Stata implementation. Figure 3 presents the nonparametric density estimate of the distribution of individual treatment effects τ_i.

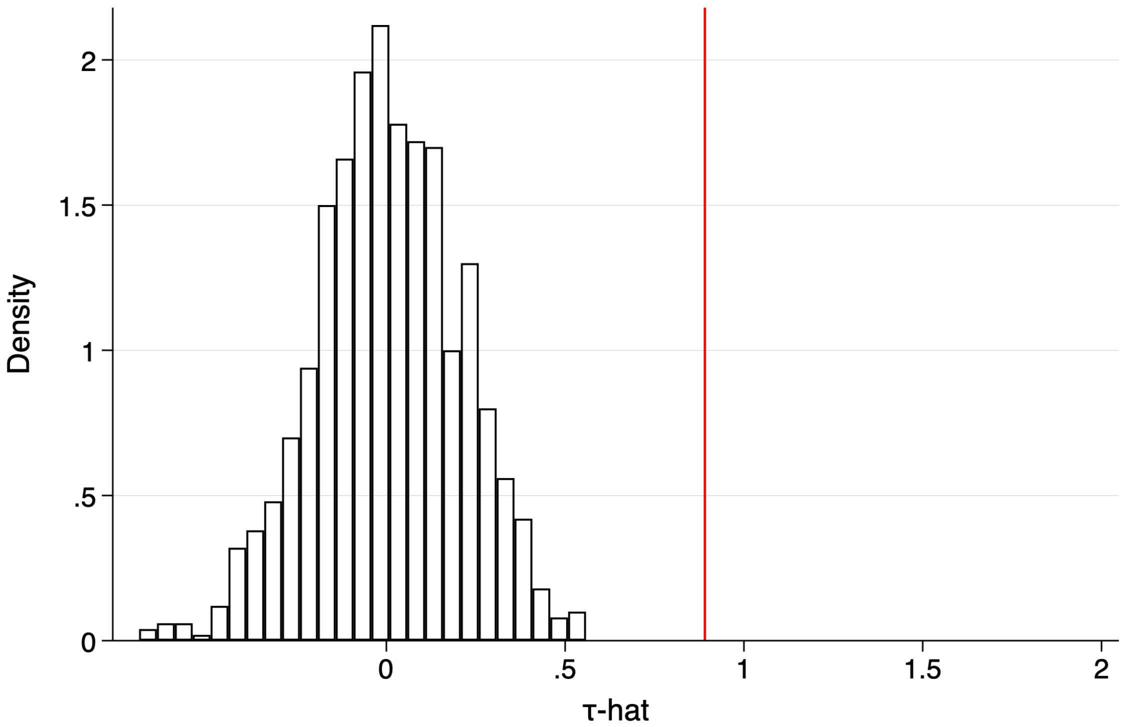
e.

See the do-file for the Stata implementation. Figure 4a presents the distribution of $\hat{\tau}$ from 1000 alternative treatment assignments (denoted by $\hat{\tau}^a$). The red vertical line represents the estimate from the observed data, $\hat{\tau}$. Since none of $\hat{\tau}^a$ is larger than $\hat{\tau}$, the fraction of $\hat{\tau}^a$ from alternative assignments that is larger than $\hat{\tau}$ is 0. Thus, we reject the sharp null $H_0 : \tau_i = 0 \ \forall i$ at any statistical significance level.

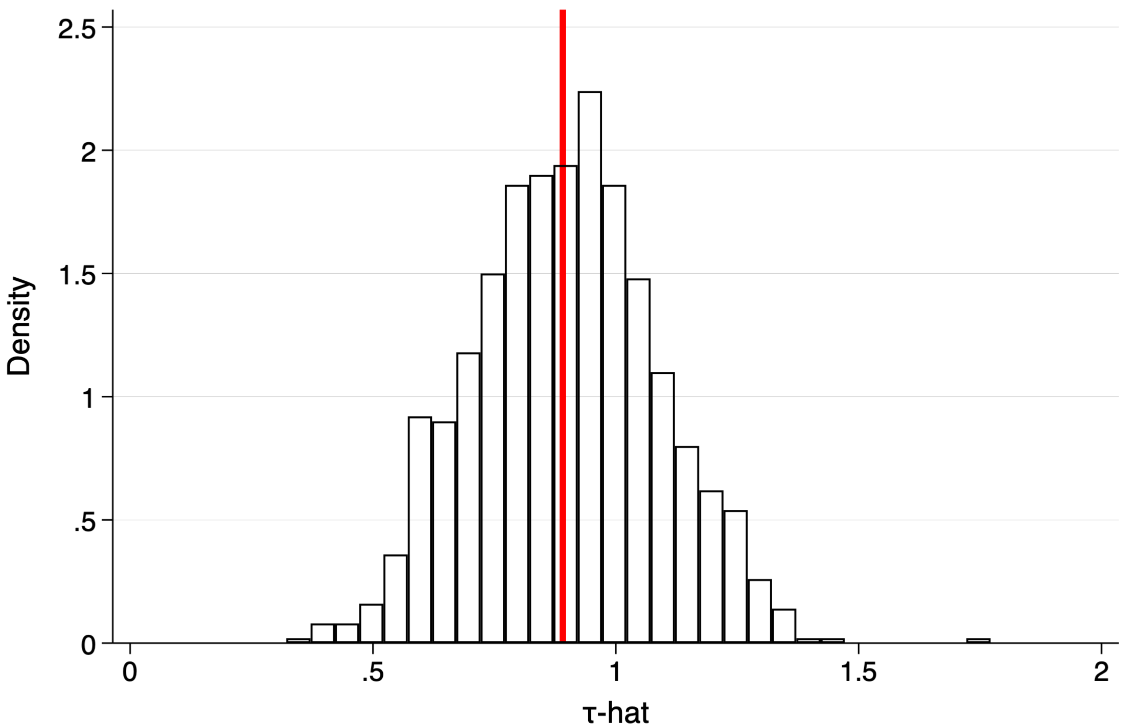
f.

See the do-file for the Stata implementation. Figure 4b presents the bootstrap distribution of $\hat{\tau}$, denoted by $\hat{\tau}^b$. The red vertical line indicates $\hat{\tau}$ from the observed data. This figure shows that no $\hat{\tau}^b$ is smaller than 0, thus the fraction of the coefficient estimates from alternative treatment assignments is smaller than 0 is 0. Thus, we reject the null hypothesis $H_0 : \tau = 0$ (at a reasonable significance level).

Figure 4: Randomisation Inference and Bootstrapping



(a) 3.e. Distribution of $\hat{\tau}$ under alternative assignments



(b) 3.f. Bootstrap distribution of $\hat{\tau}$