Question 4. In the paper "The Costs of Remoteness: Evidence from German Division and Reunification", the researchers investigate the importance of market access for economic development. They use as a natural experiment the construction of the Iron Curtain which divided Germany into Eastern and Western parts, between which all trade was stopped.
The sample consists of all West German cities with a population of more than 20,000 in a base year. The researchers have observations on a measure of economic development (denoted by y) before and after the construction of the Iron Curtain. Their treatment group is cities within 75km of the border (which are presumed to lose access to markets with division of Germany) and the control group is all other cities.

1. Explain why a simple comparison of y in treatment and control groups after division is unlikely to provide a good estimate of the effect of losing market access.

	Before Iron Curtain	After Iron Curtain
Treatment Group	①	③
Control Group	②	④

① : Treatment Group
② : Control Group
③ : Treatment Group + time effect$_t$ + treatment effect
④ : Control Group + time effect$_c$

A simple comparison of y in treatment and control groups after division = comparison of ③ and ④
A simple comparison is unlikely to provide a good estimate of the treatment effect since we can't ensure that the time effect of the treatment group and the time effect of the control group are the same.

- The procedures for providing a good estimate of the treatment effect
  (a) Assume Random sampling
  (b) Obtain the time effect with ④-②
  (c) Subtract the time effect from ③
  (d) Compare the ① with ③ obtained in (b)

2. What equation would you estimate with this data and what parameters of this equation tell you about the causal effect of interest?

$$T_c = \begin{cases} 1 & \text{if city is treated} \\ 0 & \text{if city is not treated} \end{cases} \quad D_t = \begin{cases} 1 & \text{for periods at or after treatment} \\ 0 & \text{for periods before treatment} \end{cases} \quad x_{ct} = \text{ covariates.}$$

$$y_{ct} = \beta x_{ct} + D_t + T_c + \delta\,(T_c \cdot D_t) + U_{ct}$$

$\therefore \delta$ = treatment effect (causal effect)
The reason for adding x_{ct}(covariates) here is to satisfy the Cetris Paribus.

3. You now also have more observations on y both before and after the construction of the Iron Curtain. Why might this information be useful and how would you use it?

If we have more observations on y, the effect of the outlier can be canceled out, so that the accuracy of the ATE increases and a value close to the true value can be obtained.

Question 7. Suppose we are interested in the effect of treatment T on Y. We estimate a regression of the form

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 T_i + \hat{\epsilon}_i$$

Y is a continuous outcome variable and T is a discrete treatment indicator that is 1 if an individual received the treatment and 0 otherwise. Suppose that you estimate $\hat{\beta_1}$ using OLS.

1. Show that can be decomposed two components, the treatment on the treated (TOT), and a selection bias term (SB).
   (Hint: TOT = $E[Y_{1i} - Y_{0i} \mid T_i = 1]$ and SB = $E[Y_{0i} \mid T_i = 1] - Y_{0i}[T_i = 0]$ where Y_{1i} is the outcome in the state of the world where an individual receives the treatment and Y_{0i} is the outcome in the state of the world where an individual does not receive the treatment.)

$$\begin{aligned}
&\mathbb{E}\left(Y_i \mid T_i=1\right)-\mathbb{E}\left(Y_i \mid T_i=0\right) \text{ observed difference} \\
=&\mathbb{E}\left(Y_{1i} \mid T_i=1\right)-\mathbb{E}\left(Y_{0i} \mid T_i=0\right) \\
=&\mathbb{E}\left(Y_{1i} \mid T_i=1\right)-\mathbb{E}\left(Y_{0i} \mid T_i=1\right)+\mathbb{E}\left(Y_{0i} \mid T_i=1\right)-\mathbb{E}\left(Y_{0i} \mid T_i=0\right) \\
=&\underbrace{\mathbb{E}\left(Y_{1i}-Y_{0i} \mid T_i=1\right)}_{TOT}+\underbrace{\mathbb{E}\left(Y_{0i} \mid T_i=1\right)-\mathbb{E}\left(Y_{0i} \mid T_i=0\right)}_{SB}
\end{aligned}$$

2. Suppose treatment was conditionally randomly assigned based on some characteristic X. Must you include X in the regression to ensure that $\hat{\beta_1}$ converges to true β_1? (MSc Only)

   True. We must include X in the regression.

3. Now suppose that treatment was not randomly assigned but you find a control group which looks similar to your treatment group. You collect some more data so you can estimate a regression of the form:

$$Y_{it} = \hat{\beta}_0 + \hat{\beta}_1 T_i + \hat{\beta}_2 A_t + \hat{\beta}_3 (A_t \times T_i) + \hat{\epsilon_{it}}$$

   The variable A is 0 at time $t = 1$ and 1 at time $t = 2$. The variable T is 1 for the treatment group (in all periods) and 0 for the control group. The treatment occurs between time $t = 1$ and $t = 2$. Show how this regression estimates the effect of T on Y. What assumptions does this estimation method imply?

$$A_t = \begin{cases} 0 & \text{at time } t=1 \\ 1 & \text{at time } t=2 \end{cases} \quad T_i = \begin{cases} 1 & \text{for the treatment group} \\ 0 & \text{for the control group} \end{cases}$$

   Number of cases (i,t) : (0,1) (0,2) (1,1) (1,2)

   When (i,t) = (0,1) : $T_i = 0$, $A_t = 0$, $A_t \times T_i = 0$
   $Y_{01} = \hat{\beta}_0 + \hat{\epsilon_{01}}$
   control group at time t=1

   When (i,t) = (0,2) : $T_i = 0$, $A_t = 1$, $A_t \times T_i = 0$
   $Y_{02} = \hat{\beta}_0 + \hat{\beta}_2 + \hat{\epsilon_{02}}$
   control group at time t=2, including time effect $\hat{\beta}_2$

   When (i,t) = (1,1) : $T_i = 1$, $A_t = 0$, $A_t \times T_i = 0$
   $Y_{11} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\epsilon_{11}}$
   treatment group at time t=1, $\hat{\beta}_1$ is difference between treatment group and control group

   When (i,t) = (1,2) : $T_i = 1$, $A_t = 1$, $A_t \times T_i = 1$
   $Y_{12} = \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\epsilon_{12}}$
   treatment group at time t=2, including time effect $\hat{\beta}_2$, treatment effect $\hat{\beta}_3$

In summary,
$\hat{\beta_1}$: difference between treatment group and control group
$\hat{\beta_2}$: time effect
$\hat{\beta_3}$: treatment effect

This estimation method imply assumption that $\hat{\beta_1}$ should be 0.
If there is no selection bias, $\hat{\beta_1} = 0$, then $Y_{01} = Y_{11}$

4. Suppose you had 3 periods of data now, so that you could estimate a regression of the form

$$Y_{it} = \hat{\beta}_0 + \hat{\beta}_1 T_i + \hat{\beta}_2 A_{2t} + \hat{\beta}_3(A_{2t} \times T_i) + \hat{\beta}_4 A_{1t} + \hat{\beta}_5(A_{1t} \times T_i) + \hat{\epsilon_{it}}$$

The variable A_{2t} is 1 at time $t = 2$ and 0 otherwise. The variable A_{1t} is 1 at time $t = 1$ and 0 otherwise. How does the coefficient $\hat{\beta_5}$ help you test the identifying assumptions from (3)?

$$A_{2t} = \begin{cases} 1 & t = 2 \\ 0 & \text{otherwise} \end{cases} \quad A_{1t} = \begin{cases} 1 & t = 1 \\ 0 & \text{otherwise} \end{cases} \quad T_i = \begin{cases} 1 & \text{for the treatment group} \\ 0 & \text{for the control group} \end{cases}$$

If we have 3 periods, $t = 0, 1, 2$
Number of cases (i, t) : (0,0) (0,1) (0,2) (1,0) (1,1) (1,2)

When (i, t) = (0,0)
$Y_{00} = \hat{\beta_0} + \hat{\epsilon_{00}}$
control group at t=0

When (i, t) = (0,1)
$Y_{01} = \hat{\beta_0} + \hat{\beta_4} + \hat{\epsilon_{01}}$
control group at t=1,
where $\hat{\beta_4}$: time effect between t=0 and t=1

When (i, t) = (0,2)
$Y_{02} = \hat{\beta_0} + \hat{\beta_2} + \hat{\epsilon_{02}}$
control group at t=2,
where $\hat{\beta_2}$: time effect between t=0 and t=2

When (i, t) = (1,0)
$Y_{10} = \hat{\beta_0} + \hat{\beta_1} + \hat{\epsilon_{10}}$
treatment group at t=0,
where $\hat{\beta_1}$: difference between treatment group and control group (before treatment)

When (i, t) = (1,1)
$Y_{11} = \hat{\beta_0} + \hat{\beta_1} + \hat{\beta_4} + \hat{\beta_5} + \hat{\epsilon_{11}}$
treatment group at t=1,
where $\hat{\beta_5}$: Treatment group's effect at t=1

When (i, t) = (1,2)
$Y_{12} = \hat{\beta_0} + \hat{\beta_1} + \hat{\beta_2} + \hat{\beta_3} + \hat{\epsilon_{12}}$
treatment group at t=2,
where $\hat{\beta_3}$: Treatment effect

In (3), The treatment occurs between time $t = 1$ and $t = 2$.
It means there is no treatment at t=1.
Hence, $\hat{\beta_5}$ should be 0, then we can assume there is no selection bias, which satisfies assumptions(Cetris Paribus) from (3).