1 Adding Controls

Consider the linear regression model:

$$Y = X_1'\beta_1 + X_2'\beta_2 + U$$

where (X_1, X_2, U) i.i.d; $(X_1, X_2) \perp U$; $\mathbb{E}[X_1], \mathbb{E}[X_2], \mathbb{E}[U] < \infty$

Problem 1.1. What is the effect of deleting X_2 on the estimated coefficient of β_1?

Solution. We get an omitted variable bias.

- ▷ Specifically, the estimate of β_1 without X_2 (β_1^*) is

$$\beta_1^* = \mathbb{E}\left[X_1X_1'\right]^{-1}\mathbb{E}\left[X_1Y\right]$$

- ▷ Plugging in the expression for Y:

$$\begin{aligned} &= \mathbb{E}\left[X_1X_1'\right]^{-1}\mathbb{E}\left[X_1\left(X_1'\beta_1 + X_2'\beta_2 + U\right)\right] \\ &= \beta_1 + \mathbb{E}\left[X_1X_1'\right]^{-1}\mathbb{E}\left[X_1X_2'\right]\beta_2 \end{aligned}$$

  since $\mathbb{E}[X_1U] = 0$ from our assumtion.

Thus β_1^* will be different from β_1, and the deviation depends on the sign and the magnitude of β and $\mathbb{E}[X_1X_1']^{-1}\mathbb{E}[X_1X_2']$.

Problem 1.2. Suppose that you include X_2 in different ways:

- ▷ (1): Regress X_1 on X_2, work with the residual of X_1 purged of X_2 and regress Y on this residual
- ▷ (2): Regress X_1 on X_2 and Y on X_2, and regress the residual from the second regression on the residual from the first regression.

How do these estimates compare with your result from a)? With the OLS estimates of the initial specification?

Solution. See below:

- ▷ Approach (1):
  - * Denote $\tilde{X}_1$ as the residuals of X_1 purged of X_2: $\tilde{X}_1 = X_1 - BLP(X_1|X_2)$
  - * Then the expression for β is

$$\begin{aligned} \beta &= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1Y\right] \\ &= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\left(X_1'\beta_1 + X_2'\beta_2 + U\right)\right] \\ &= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\left\{\mathbb{E}\left[\tilde{X}_1X_1'\beta_1\right] + \mathbb{E}\left[\tilde{X}_1X_2'\beta_2\right] + \mathbb{E}\left[\tilde{X}_1U\right]\right\} \end{aligned}$$

Since $\mathbb{E}\left[\tilde{X}_1 X_2\right] = 0$ from the property of Best Linear Predictors (BLP) and $\mathbb{E}\left[\tilde{X_1 U}\right] = 0$ from $(X_1, X_2) \perp U$:

$$
\begin{aligned}
=&\mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1 X_1'\beta_1\right] \\
=&\mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\left(\tilde{X}_1 + BLP\left(X_1|X_2\right)\right)'\beta_1\right] \\
=&\mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]\beta_1 + \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1 BLP\left(X_1|X_2\right)\right]\beta_1
\end{aligned}
$$

Since $\mathbb{E}\left[\tilde{X}_1 BLP\left(X_1|X_2\right)\right] = 0$ from the BLP property:

$$
= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]\beta_1 = \beta_1
$$

* This implies that we get an unbiased estimator of β_1, which is consistent with the original specification.

▷ Approach (2)

* Denote $\tilde{X}_1$ as the residuals of X_1 purged of X_2: $\tilde{X}_1 = X_1 - BLP\left(X_1|X_2\right)$
* Denote $\tilde{Y}$ as the residuals of Y_1 purged of X_2: $\tilde{Y} = Y - BLP\left(Y|X_2\right)$
* Then the expression for β is

$$
\begin{aligned}
\beta &= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\tilde{Y}\right] \\
&= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1\left(Y - BLP\left(Y|X_2\right)\right)\right] \\
&= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\left\{\mathbb{E}\left[\tilde{X}_1 Y\right] + \mathbb{E}\left[\tilde{X}_1 BLP\left(Y|X_2\right)\right]\right\}
\end{aligned}
$$

Since $\mathbb{E}\left[\tilde{X}_1 BLP\left(Y|X_2\right)\right] = 0$ from the BLP first-order condition (since $\tilde{X}_1$ is purged of X_2 and $BLP\left(Y|X_2\right)$ is a linear function in X_2) and $Y = X_1'\beta_1 + X_2'\beta_2 + U$:

$$
= \mathbb{E}\left[\tilde{X}_1\tilde{X}_1'\right]^{-1}\mathbb{E}\left[\tilde{X}_1 Y\right]
$$

This reduced to the case in Approach (1), which we showed to yield unbiased estimators consistent with the original specification.

* Therefore, we conclude that in this approach, we still get an unbiased estimator of β_1, which is consistent with the original specification.

2 Propensity Score

Let $Y_i(1)$ and $Y_i(0)$ be potential outcomes of individual i if treated or not treated, respectively. Let Y_i be the actual outcome and let D_i be the treatment indicator. We assume:

▷ $D_i \perp (Y_i(1), Y_i(0)) | X_i$

▷ $0 < P[D_i = 1 | X_i = x] < 1, \forall x \in supp(X_i)$

Problem 2.1. Propose an estimator of $\mathbb{E}[Y_i(1) - Y_i(0)]$

Solution. We propose the following estimator:

$$\sum_{k=1}^{K} \{\mathbb{E}[Y_i | D_i = 1, X_i = x_k] - \mathbb{E}[Y_i | D_i = 0, X_i = x_k]\} P(X = x_k)$$

where $\mathbb{E}[\cdot]$ is replaced with the sample averages from data.

▷ Note that this estimator equals the ATT since:

$$\begin{aligned} ATE &\equiv \mathbb{E}[Y_i(1) - Y_i(0)] \\ &= \mathbb{E}[\mathbb{E}[Y_i(1) - Y_i(0)] | X] \\ &= \mathbb{E}[\mathbb{E}[Y_i(1) | X] - \mathbb{E}[Y_i(0) | X]] \\ &= \mathbb{E}[\mathbb{E}[Y_i(1) | X, D = 1] - -\mathbb{E}[Y_i(0) | X, D = 0]] \qquad (\because) \text{ CIA} \\ &= \sum_{k=1}^{K} \{\mathbb{E}[Y_i | D_i = 1, X_i = x_k] - \mathbb{E}[Y_i | D_i = 0, X_i = x_k]\} P(X = x_k) \end{aligned}$$

▷ If X is continuous, we can block the data and get the same estimator.

▷ This estimator will be unbiased (as sample average) and consistent (by LLN).

Note that we require the common support assumption here to enable us to compute the corresponding sample averages from data.

Problem 2.2. Show that the assumptions stated imply that D_i is conditionally independent of $(Y_i(1), Y_i(0))$ given $P[D_i = 1 | X_i]$; $D_i \perp (Y_i(1), Y_i(0)) | P(D_i = 1 | X_i)$. Why is this result important?

Solution. It suffices to show that

$$P[D_i = 1 | Y_i, Y_0, P[D_i = 1 | X_i]] = P[D_i = 1 | P[D_i = 1 | X_i]]$$

▷ Rewriting the LHS as an expectation of the indicator variable:

$$P\left[D_i = 1|Y_i, Y_0, P\left[D_i = 1|X_i\right]\right]$$
$$= \mathbb{E}\left[1\left\{D_i = 1\right\}|Y_i, Y_0, P\left[D_i = 1|X_i\right]\right]$$

Using Law of Iterated Expectations:

$$= \mathbb{E}\left[\mathbb{E}\left[1\left\{D_i = 1\right\}|\underbrace{Y_i, Y_0, P\left[D_i = 1|X_i\right], X_i}_{\mathfrak{F}_2}\right]|\underbrace{Y_i, Y_0, P\left[D_i = 1|X_i\right]}_{\mathfrak{F}_1}\right]$$

which works out since $\mathfrak{F}_2 \subseteq \mathfrak{F}_1$. Furthermore, **assuming selection on observables** (the first assumption provided):

$$= \mathbb{E}\left[\mathbb{E}\left[1\left\{D_i = 1\right\}|P\left[D_i = 1|X_i\right], X_i\right]|Y_i, Y_0, P\left[D_i = 1|X_i\right]\right]$$

Since the distribution of D is entirely determined by $P\left[D_i = 1|X_i\right]$, and X_i gives no additional information on top of knowing $P\left[D_i = 1|X_i\right]$:

$$= \mathbb{E}\left[\mathbb{E}\left[1\left\{D_i = 1\right\}|P\left[D_i = 1|X_i\right]\right]|Y_i, Y_0, P\left[D_i = 1|X_i\right]\right]$$

Using Law of Iterated Expectations:

$$= \mathbb{E}\left[1\left\{D_i = 1\right\}|P\left[D_i = 1|X_i\right]\right] = P\left[D_i = 1|P\left[D_i = 1|X_i\right]\right]$$

and thus we see that the D_i is indeed conditionally independent.

This result is important because it justifies the reduction in dimensionality through computing the propensity score. Matching on high-dimensional X can lead to sparse data for some values of the covariates, which can lead to high variance. Matching on univariate propensity score allows us to overcome this problem, and the above result provides evidence for this argument.

▷ Note that we still need to estimate these propensity scores, which still puts us subject to the dreaded curse of dimensionality.

Problem 2.3. Define the propensity score $P\left[D_i = 1|X_i\right]$ as the probability of receiving the treatment given the observables variables X_i. Propose an estimator of $\mathbb{E}\left[Y_i(1) - Y_i(0)\right]$ based on the propensity score.

Solution. We propose the following estimator:

$$\int \left\{\mathbb{E}\left[Y_i|D_i = 1, X_i = x_k\right] - \mathbb{E}\left[Y_i|D_i = 0, X_i = x_k\right]\right\} dF_P(p)$$

where the $F_p(p) = P(p(X) \leq p)$ is the CDF of the propensity score and $\mathbb{E}[\cdot]$ is replaced with the corresponding sample averages.

▷ Note that this estimator equals the ATT since:

$$\begin{aligned} ATE &\equiv \mathbb{E}\left[Y_i(1) - Y_i(0)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_i(1) - Y_i(0)\right]|p(X)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_i(1)|p(X)\right] - \mathbb{E}\left[Y_i(0)|p(X)\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_i(1)|p(X), D = 1\right] - -\mathbb{E}\left[Y_i(0)|p(X), D = 0\right]\right] \qquad (\because) \text{ CIA} \\ &= \int \left\{\mathbb{E}\left[Y_i|D_i = 1, X_i = x_k\right] - \mathbb{E}\left[Y_i|D_i = 0, X_i = x_k\right]\right\} dF_P(p) \end{aligned}$$

▷ This estimator will be unbiased (as sample average) and consistent (by LLN).

3 Propensity Score

LaLonde (AER, 1986) investigated whether non-experimental methods could reproduce the experimental estimates based on the National Supported Work (NSW) Demonstration. The following dataset from Smith and Todd (J Ectrics, 2005) includes the NSW sample, as well as two non-experimental samples: one based on the Current Population Survey (CPS) and one on the Michigan Panel of Income Dynamics (PSID).

- ▷ The variable -sample- identifies the relevant observations.
- ▷ The variable -treated- identifies the observations that were treated (participate in a subsidized work experience program) in the NSW (from April 1975 to August 1977)

You are interested in the average effect on Real Earnings in 1978 of the treatment for the treated. Start with the NSW sample:

Problem 3.1. Investigate whether the data is consistent with randomization of the treatment.

Solution. To see if the data is consistent, we run two-sample t-tests across the treated group and the control group in our NSW sample. The following shows the t-statistic associated with the null hypothesis that the mean between the two groups is equal to zero:

Table 1: Two-sample t-test Results

Variable	age	educ	black	married	nodegree	re74	kids18	hisp	kidmiss	early_ra
p-value	0.7209	0.1361	0.9645	0.7014	0.0077	0.8294	0.6355	0.4220	0.2834	0.4127

- ▷ For most variables, the means are statistically indistinguishable, which is consistent with the randomization of treatment.
- ▷ For the "nodegree" variable, it is significantly larger for the untreated group.

Furthermore, we regress the treatment dummy on all the covariates.

```
. reg treated age educ black married nodegree re74 kids18 hisp kidmiss early_ra, r

Linear regression                               Number of obs     =        722
                                                F(10, 711)        =       1.09
                                                Prob > F          =     0.3643
                                                R-squared         =     0.0154
                                                Root MSE          =     .49204

------------------------------------------------------------------------------
             |               Robust
     treated |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0007713   .0028817    -0.27   0.789    -.0064289    .0048863
        educ |   -.006391   .0146421    -0.44   0.663    -.035138     .022356
       black |   -.029767   .0649968    -0.46   0.647    -.1573756    .0978416
     married |   .0226487     .05524     0.41   0.682    -.0858044    .1311018
    nodegree |  -.1372686   .0600206    -2.29   0.022    -.2551074   -.0194298
        re74 |  -2.16e-06   2.98e-06    -0.73   0.468    -8.00e-06    3.68e-06
      kids18 |   .0067459   .0236859     0.28   0.776    -.0397568    .0532486
        hisp |  -.0666199   .0844727    -0.79   0.431    -.2324658    .0992259
     kidmiss |  -.0840695   .0582959    -1.44   0.150    -.1985222    .0303831
    early_ra |   .0418421   .0399444     1.05   0.295     -.036581    .1202652
       _cons |   .6310442   .2079525     3.03   0.002     .2227698    1.039319
------------------------------------------------------------------------------
```

- ▷ We find that most of the coefficients are statistically insignificant except for the "nodegree" variable.

Problem 3.2. Estimate the effect using the experimental sample.

Solution. We estimate the effect in stata:

```
. reg re78 treated, r

Linear regression                               Number of obs     =        722
                                                F(1, 720)         =       3.30
                                                Prob > F          =     0.0698
                                                R-squared         =     0.0049
                                                Root MSE          =       6242

------------------------------------------------------------------------------
             |               Robust
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   886.3038   488.1385     1.82   0.070    -72.04111    1844.649
       _cons |   5090.048    277.426    18.35   0.000     4545.388    5634.709
------------------------------------------------------------------------------
```

▷ The estimate of the ATT – the difference in means between the treated and control groups – is $886.30.

Now use the sample consisting in the treated from the NSW sample and the comparison individuals from the CPS sample.

Problem 3.3. Estimate the effect using OLS.

Solution. We estimate the effect using OLS by regressing "re78" on treatment and other control variables.

```
. reg re78 treated age educ black married nodegree re74 hisp kids18 kidmiss metro, r

Linear regression                               Number of obs     =     16,289
                                                F(11, 16277)      =    1238.36
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4286
                                                Root MSE          =     7317.6

------------------------------------------------------------------------------
             |               Robust
        re78 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |  -1641.219   499.6991    -3.28   0.001    -2620.684   -661.7538
         age |  -104.7382   6.043211   -17.33   0.000    -116.5836   -92.89289
        educ |   207.5299    30.6089     6.78   0.000     147.5331    267.5267
       black |  -1011.522   207.6853    -4.87   0.000    -1418.608   -604.4366
     married |   501.1817   152.2492     3.29   0.001     202.7566    799.6068
    nodegree |    306.308   182.5228     1.68   0.093    -51.45676    664.0727
        re74 |   .6728306    .007122    94.47   0.000     .6588707    .6867905
        hisp |  -222.2795   229.7224    -0.97   0.333    -672.5606    228.0016
      kids18 |   4.890967   33.85897     0.14   0.885    -61.47633    71.25827
     kidmiss |   334.0832   799.2312     0.42   0.676    -1232.498    1900.664
       metro |   277.1105   122.7767     2.26   0.024     36.45469    517.7662
       _cons |   5850.772   480.6098    12.17   0.000     4908.724     6792.82
------------------------------------------------------------------------------
```

▷ We find that the OLS estimate of ATT – the coefficient on the treatment dummy – is $-\$1641.22$.

Problem 3.4. Investigate covariate balancing and support between the treated and the CPS sample.

Solution. To investigate covariate balancing and common support for the covariates, we plot distributions of all covariates for both the treated and the control variables.

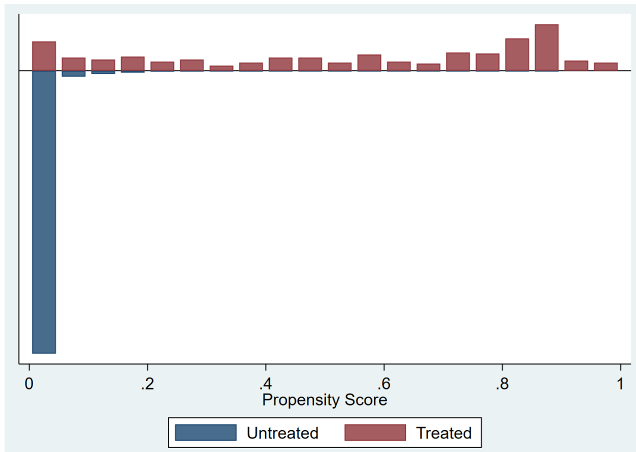
▷ Most continuous variables do not seem to have common support, while most binary variables do have common support.

Problem 3.5. Estimate the effect using 1-nearest neighbor propensity score matching (Use -psmatch2- which can be installed using: ssc install psmatch2, if you use Stata)

Solution. We use psmatch2 to implement propensity score. We first implement matching with replacement:

▷ We drop metro variable since $P\left(D=0|metro=1\right)=0$ violating the common support assumption.

▷ We implement the matching in Stata and check for common support:



▷ Based on this graph, we could potentially consider trimming the data and re-estimating the propensity score. But given the criticisms about dropping the data in further papers discussing Lalonde (1986), we keep all data and proceed with the estimation.

▷ We assess the matching quality using the $pstest$ command.

```
. pstest age educ black hisp married nodegree re74 kids18 kidmiss, both
```

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|	V(T)/ V(C)
age	U	24.626	33.225	-94.2		-13.37	0.000	0.37*
	M	24.626	27.01	-26.1	72.3	-3.54	0.000	0.50*
educ	U	10.38	12.028	-68.5		-9.85	0.000	0.40*
	M	10.38	10.731	-14.6	78.7	-1.61	0.107	0.31*
black	U	.80135	.07354	215.6		47.04	0.000	.
	M	.80135	.80135	0.0	100.0	0.00	1.000	.
hisp	U	.09428	.07204	8.1		1.47	0.143	.
	M	.09428	.09764	-1.2	84.9	-0.14	0.889	.
married	U	.16835	.71173	-130.7		-20.54	0.000	.
	M	.16835	.25926	-21.9	83.3	-2.71	0.007	.
nodegree	U	.73064	.29584	96.5		16.27	0.000	.
	M	.73064	.58923	31.4	67.5	3.67	0.000	.
re74	U	3571	14017	-132.2		-18.75	0.000	0.36*
	M	3571	5372.6	-22.8	82.8	-3.76	0.000	0.96
kids18	U	.34343	1.8341	-111.3		-15.12	0.000	0.25*
	M	.34343	.60269	-19.4	82.6	-3.41	0.001	0.71*
kidmiss	U	.10101	.00413	44.4		21.93	0.000	.
	M	.10101	.04377	26.2	40.9	2.70	0.007	.

* if variance ratio outside [0.80; 1.26] for U and [0.80; 1.26] for M

* We achieve significant bias reduction for most of the variables (this can be seen by comparing U and M row for each of the variables.)

▷ We obtain the following estimate of ATT:

re78	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
treated	-381.1362	1083.499	-0.35	0.725	-2510.546	1748.274
_cons	6357.488	1006.17	6.32	0.000	4380.052	8334.924

* The ATT calculated via propensity score matching *with replacement* is $-\$381.14$ but it has a huge standard error of \$1083.50.

▷ Note that we can also repeat the matching without replacement and obtain the following results:

re78	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
treated	-1729.848	595.9476	-2.90	0.004	-2900.276	-559.4191
_cons	7706.2	440.1624	17.51	0.000	6841.73	8570.67

* The ATT calculated via propensity score matching without replacement is $-\$1729.848$ with lower standard error.

▷ Combining these two results, we find that the matching with replacement yields a less biased (closer to experimental value) but higher variance (higher standard error) ATT estimate, while matching without replacement yields a more biased but lower variance estimate. This demonstrates the bias-variance tradeoff.

Problem 3.6. Estimate the effect using the propensity score and local linear regression.

▷ We obtain the following estimate of ATT using the local linear regression:

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
re78	Unmatched	5976.35201	14846.6596	-8870.30761	562.477666	-15.77
	ATT	5976.35201	6481.38607	-505.034062	1596.81032	-0.32

Note: S.E. does not take into account that the propensity score is estimated.

▷ The ATT calculated via local linear regression with propensity score is -505.034. Note that we still have a high standard error.

4 Math Homework

You are interested in estimating the effect of doing mathematics homework (yes/no) on mathematics test scores. You have data on all 10th graders in Oslo for the school year 2010/2011. The dataset contains the following variables:

- ▷ students: end year test scores, homework, # of missed classes, gender, age, test scores in 2009/2010
- ▷ parents: education
- ▷ teacher: gender, age, education

About half of the students do their homework.

Problem 4.1. If you want to abolish homework, what effect would you want to estimate?

Solution. We introduce some notation.

- ▷ Denote D as whether the person did mathematics homework.
- ▷ Denote Y as the mathematics test scores and the potential outcomes as Y_1 and Y_0.

If we want to abolish homework, we want to estimate the average treatment effect on the treated (ATT): $\mathbb{E}\left[Y_1 - Y_0 | D = 1\right]$.

- ▷ This will allow us to know how much test scores would decline among people who did their homework if homework were to be abolished.

Problem 4.2. If you want to make homework mandatory, what effect would you want to estimate?

Solution. We want to estimate the average treatment effect on the untreated (ATUT): $\mathbb{E}\left[Y_1 - Y_0 | D = 0\right]$.

- ▷ This will allow us to know how much test scores would increase among people who didn't do their homework if homework were to be made mandatory.

Problem 4.3. You want to compare the effect of doing homework as compared to an extra hour of math teaching. What effect of homework would you like to know?

Solution. We want to estimate the average treatment effect (ATE) for each treatment (independently) and compare the two.

- ▷ This will allow us to compare the effect of homework on the entire population versus the effect of an extra hour of teaching on the whole population.

You want to estimate how well students that are currently not doing their homework would do, if they did their homework. You decide to use matching, and will therefore rely on a conditional independence assumption (CIA).

Problem 4.4. Explain your CIA. Be explicit about the counterfactual outcomes and the variables that you want to control for. Why might your CIA not hold? Can you think of examples where you get upward biased estimates? And downward biased estimates?

Solution. We are interested in the ATUT, i.e. $\mathbb{E}\left[Y_1 - Y_0 | D = 0\right]$.

▷ For CIA, we want to find a set of controls that satisfies $D \perp \left(Y_i\left(1\right), Y_i\left(0\right)\right) | X$

  * This means that conditioned on the controls, whether or not the student does homework should be random.

▷ The following variables could be included as control:

  * Student age (within the same grade, older students may know more and do better on the tests, and they may also be more responsible and likely to do homework more) – upward bias
  * Teacher age (with tenure, typically older teachers teach tougher courses, so students in these courses will be highly educated and may not be comparable across different student groups) – upward bias
  * Number of absence (lower ability students may be less likely to do homework and more likely to skip class) – downward bias
  * Parents' education (Students with parents of high education may have higher ability and do better on test scores and also have a better work ethic) – upward bias
  * Student gender (Under gender discrimination, women may feel more obligated to complete homework in order to not be discriminated against by their teachers or fellow classmates. This may drive them to try harder in school leading to high test scores, where the higher test score wasn't necessarilly due to doing homework.) – upward bias in this example, but could go either way

▷ CIA may not hold if we don't condition all such factors.

  * Data for some of these variables are difficult to obtain or measure.
  * Note that we do not want to control for outcome variables of HW such as the end-of-year test scores for this year.

Problem 4.5. Explain how you use the CIA to estimate the counterfactual outcome, how you take into account that students that do their homework have different characteristics, and what support condition you need.

Solution. Note that under CIA, $ATUT\left(x\right) = ATE\left(x\right)$:

$$ATUT\left(x\right) = \mathbb{E}\left[Y_i\left(1\right) - Y_i\left(0\right)|D = 0, X = x\right] = \mathbb{E}\left[Y_i\left(1\right) - Y_i\left(0\right)|X = x\right] = ATE\left(x\right)$$

which yields:

$$\begin{aligned} ATUT &= \mathbb{E}\left[ATUT\left(x\right)|D = 0\right] \\ &= \mathbb{E}\left[ATE(x)|D = 0\right] \\ &= \sum_x ATE\left(x\right) P\left(X = x|D = 0\right) \end{aligned}$$

▷ To estimate ATUT, we can estimate $ATE(x)$ and obtain a weighted average across all values of x where the probabilities are given by $P\left(X = x|D = 0\right)$

- ▷ Specifically, we can match each treated individual with an untreated individual based on the control variables via:
  - * Computing distnace between the individuals' control vectors X_i or
  - * Computing the propensity score and matching on the fitted values
- ▷ We compute the difference in their outcomes and average across the differences to get our estimate of the ATE.

We need a common support condition to ensure that we have treated and untreated individuals for each value of x. Mathematically, this amounts to:

$$0 < P\left(D = 1|X = x\right) < 1, \forall x \in supp(X)$$

Recall that in problem 3, the "metro" variable did not satisfy the common support assumption and thus had to be dropped in our matching process.

Problem 4.6. How would you estimate your effect using OLS?

Solution. Using OLS, we would run a local linear regression where we estimate β within each band of covariates and weight them using $P\left(X = x|D = 0\right)$.

- ▷ If X is already discrete, this interpretation is natural. If it's continuous, we partition the covariates into bands and estimate $ATE(x)$ in each cell.
- ▷ Note that this step is guided by this expression:

$$\begin{aligned} ATUT &= \mathbb{E}\left[ATUT\left(x\right)|D = 0\right] \\ &= \mathbb{E}\left[ATE(x)|D = 0\right] \\ &= \sum_x ATE\left(x\right)P\left(X = x|D = 0\right) \end{aligned}$$

  where $ATE\left(x\right)$ corresponds to the OLS β in each partition.

Problem 4.7. You see in your data that boys never do their homework. What implications does this have for your research?

Solution. This is a violation of the common support assumption.

- ▷ If CIA still holds after dropping gender from the set of controls:
  - * Then the common support holds in the new reduced setup. We can then implement the same process as above, but the matching will be potentially of lower quality, which may lead to higher variance estimates.
- ▷ If gender is required for CIA, then the estimators will not be able to estimate ATUT.

You discover that not all teachers assign homework, and you get a new variable from Oslo municipality with information $(0/1)$ on whether the teacher assigned homework or not. They tell you that teachers were assigned to give homework (or not) in a randomized experiment.

Problem 4.8. First you add this new information to your matching variables. What will happen to your estimates and standard errors?

Solution. First, note that if CIA held in the old set of conditional variables, then it will still hold by adding this new information since it was randomly assigned. Therefore, the mean of your estimate will be unchanged in expectation. But now the standard errors will be reduced since we can get a better model of propensity score, thus enabling better quality matching.

Problem 4.9. How will you use this new data and what effects can you estimate?

Solution. This new data will allow us to estimate LATE, the average effect among compliers.

- ▷ If you don't have faith in selection on observables, you can try to measure LATE.
- ▷ To obtain LATE, we can run two-stage least squares by using this new variable as an instrument for doing homework.