9
Treatment effects: ignorability

First, we describe a prototypical selection setting. Then, we identify some typical average treatment effects followed by a review of various identification conditions assuming ignorable treatment (sometimes called selection on observables). Ignorable treatment approaches are the simplest to implement but pose the strongest conditions for the data. That is, when the data don't satisfy the conditions it makes it more likely that inferences regarding properties of the *DGP* are erroneous.

9.1 A prototypical selection setting

Suppose the *DGP* is
outcome equations:[1]

$$Y_j = \mu_j(X) + V_j, j = 0, 1$$

selection equation:[2]

$$D^* = \mu_D(Z) - V_D$$

observable response:

$$Y = DY_1 + (1 - D)Y_0$$

[1] Sometimes we'll find it convenient to write the outcome equations as a linear response

$$Y_j = \mu_j + X\beta_j + V_j$$

[2] We'll stick with binary choice for simplicity, though this can be readily generalized to the multinomial case (as discussed in the marginal treatment effects chapter).

where

$$D = \begin{array}{ll} 1 & D^* > 0 \\ 0 & otherwise \end{array}$$

and Y_1 is (potential) outcome with treatment and Y_0 is outcome without treatment.

In the binary case, the treatment effect is the effect on outcome of treatment compared with no treatment, $\Delta = Y_1 - Y_0$. Some typical treatment effects include: *ATE*, *ATT*, and *ATUT*. *ATE* refers to the average treatment effect, by iterated expectations

$$\begin{aligned} ATE &= E_X\left[ATE\left(X\right)\right] \\ &= E_X\left[E\left[\Delta \mid X = x\right]\right] = E\left[Y_1 - Y_0\right] \end{aligned}$$

In other words, the average effect on outcome of treatment for a random draw from the population. *ATT* refers to the average treatment effect on the treated,

$$\begin{aligned} ATT &= E_X\left[ATT\left(X\right)\right] \\ &= E_X\left[E\left[\Delta \mid X = x, D = 1\right]\right] = E\left[Y_1 - Y_0 \mid D = 1\right] \end{aligned}$$

In other words, the average effect on outcome of treatment for a random draw from the subpopulation selecting (or assigned) treatment. *ATUT* refers to the average treatment effect on the untreated,

$$\begin{aligned} ATUT &= E_X\left[ATUT\left(X\right)\right] \\ &= E_X\left[E\left[\Delta \mid X = x, D = 0\right]\right] = E\left[Y_1 - Y_0 \mid D = 0\right] \end{aligned}$$

In other words, the average effect on outcome of treatment for a random draw from the subpopulation selecting (or assigned) no treatment.

The remainder of this chapter is devoted to simple identification and estimation strategies. These simple strategies pose strong conditions for the data that may lead to logically inconsistent inferences.

9.2 Exogenous dummy variable regression

The simplest strategy (strongest data conditions) is exogenous dummy variable regression. Suppose D is independent of (Y_1, Y_0) conditional on X, response is linear, and errors are normally distributed, then *ATE* is identified via exogenous dummy variable (*OLS*) regression.[3] For instance, suppose the *DGP* is

$$Y = \delta + \varsigma D + X\beta_0 + DX\left(\beta_1 - \beta_0\right) + \varepsilon$$

Since Y_1 and Y_0 are conditionally mean independent of D given X

$$\begin{aligned} E\left[Y_1 \mid X, D = 1\right] &= E\left[Y_1 \mid X\right] \\ &= \delta + \varsigma + X\beta_0 + X\left(\beta_1 - \beta_0\right) \end{aligned}$$

[3]These conditions are stronger than necessary as we can get by with conditional mean independence in place of conditional stochastic independence.

and

$$\begin{aligned} E[Y_0 \mid X, D=0] &= E[Y_0 \mid X] \\ &= \delta + X\beta_0 \end{aligned}$$

then

$$\begin{aligned} ATE(X) &= E[Y_1 \mid X] - E[Y_0 \mid X] \\ &= \varsigma + X(\beta_1 - \beta_0) \end{aligned}$$

Then, by iterated expectations, $ATE = \varsigma + E[X](\beta_1 - \beta_0)$. *ATE* can be directly estimated via α if we rewrite the response equation as

$$Y = \delta + \alpha D + X\beta_0 + D(X - E[X])(\beta_1 - \beta_0) + \varepsilon$$

which follows from rewriting the *DGP* as

$$\begin{aligned} Y &= \delta + (\varsigma + E[X](\beta_1 - \beta_0))D + X\beta_0 \\ &\quad + D[X(\beta_1 - \beta_0) - E[X](\beta_1 - \beta_0)] + \varepsilon \end{aligned}$$

9.3 Tuebingen-style examples

To illustrate ignorable treatment, we return to the Tuebingen-style examples of chapter 8 and add regressors to the mix. For each case, we compare treatment effect analyses when the analyst observes the states with when the analyst observes only the regressor, X. The setup involves simple discrete probability and outcome structure. Identification of counterfactuals is feasible if outcome distributions are not affected by treatment selection. Hence, outcomes Y_0 and Y_1 vary only between states (and not by D within a state).

Case 1

The first case depicted in table 9.1 involves extreme homogeneity (no variation in Y_0 and Y_1). Suppose the states are observable to the analyst. Then, we have

Table 9.1: Tuebingen example case 1: extreme homogeneity

State (s)	*one*		*two*		*three*	
$\Pr(Y, D, s)$	0.0272	0.0128	0.224	0.096	0.5888	0.0512
D	0	1	0	1	0	1
Y	0	1	0	1	0	1
Y_0	0	0	0	0	0	0
Y_1	1	1	1	1	1	1
X	1	1	1	1	0	0

a case of perfect regressors and no residual uncertainty. Consequently, we can

identify treatments effects by states. The treatment effect for all three states is homogeneously one.

Now, suppose the states are unobservable but the analyst observes X. Then, conditional average treatment effects are

$$E[Y_1 - Y_0 \mid X = 1] = E[Y_1 - Y_0 \mid X = 0] = 1$$

Key components, unconditional average (integrating out X) treatment effects, and any bias for case 1 are reported in table 9.2. Case 1 exhibits no endogeneity bias.

Table 9.2: Tuebingen example case 1 results: extreme homogeneity

Results	Key components
$ATE = E[Y_1 - Y_0] = 1.0$	$p = \Pr(D = 1) = 0.16$
$ATT = E[Y_1 - Y_0 \mid D = 1] = 1.0$	$E[Y_1 \mid D = 1] = 1.0$
$ATUT = E[Y_1 - Y_0 \mid D = 0] = 1.0$	$E[Y_1 \mid D = 0] = 1.0$
$OLS = E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = 1.0$	$E[Y_1] = 1.0$
$bias_{ATT} = E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0] = 0.0$	$E[Y_0 \mid D = 1] = 0.0$
$bias_{ATUT} = E[Y_1 \mid D = 1] - E[Y_1 \mid D = 0] = 0.0$	$E[Y_0 \mid D = 0] = 0.0$
$bias_{ATE} = pbias_{ATT} + (1-p)\, bias_{ATUT} = 0.0$	$E[Y_0] = 0.0$

Extreme homogeneity implies stochastic independence of (Y_0, Y_1) and D conditional on X.

Case 2

Case 2 adds variation in outcomes but maintains treatment effect homogeneity as displayed in table 9.3. Suppose the states are observable to the analyst. Then, we

Table 9.3: Tuebingen example case 2: homogeneity

State (s)	*one*		*two*		*three*	
$\Pr(Y, D, s)$	0.0272	0.0128	0.224	0.096	0.5888	0.0512
D	0	1	0	1	0	1
Y	0	1	1	2	2	3
Y_0	0	0	1	1	2	2
Y_1	1	1	2	2	3	3
X	1	1	1	1	0	0

can identify treatments effects by states. The treatment effect for all three states is homogeneously one.

Now, suppose the states are unobservable but the analyst observes X. Then, conditional average treatment effects are

$$E[Y_1 - Y_0 \mid X = 1] = E[Y_1 - Y_0 \mid X = 0] = 1$$

which follows from

$$E_X[E[Y_1 \mid X]] = 0.36(1.889) + 0.64(3) = 2.6$$

$$E_X[E[Y_0 \mid X]] = 0.36(0.889) + 0.64(2) = 1.6$$

but *OLS* (or, for that matter, nonparametric regression) estimates

$$E_X[E[Y_1 \mid X, D = 1]] = 0.68(1.882) + 0.32(3) = 2.24$$

and

$$E_X[E[Y_0 \mid X, D = 0]] = 0.299(0.892) + 0.701(2) = 1.669$$

Clearly, outcomes are not conditionally mean independent of treatment given X ($2.6 \neq 2.24$ for Y_1 and $1.6 \neq 1.669$ for Y_0). Key components, unconditional average (integrating out X) treatment effects, and any bias for case 2 are summarized in table 9.4. Hence, homogeneity does not ensure exogenous dummy variable (or

Table 9.4: Tuebingen example case 2 results: homogeneity

Results	Key components
$ATE = E[Y_1 - Y_0] = 1.0$	$p = \Pr(D = 1) = 0.16$
$ATT = E[Y_1 - Y_0 \mid D = 1] = 1.0$	$E[Y_1 \mid D = 1] = 2.24$
$ATUT = E[Y_1 - Y_0 \mid D = 0] = 1.0$	$E[Y_1 \mid D = 0] = 2.669$
$OLS = E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = 0.571$	$E[Y_1] = 2.6$
$bias_{ATT} = E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0] = -0.429$	$E[Y_0 \mid D = 1] = 1.24$
$bias_{ATUT} = E[Y_1 \mid D = 1] - E[Y_1 \mid D = 0] = -0.429$	$E[Y_0 \mid D = 0] = 1.669$
$bias_{ATE} = pbias_{ATT} + (1-p)\, bias_{ATUT} = -0.429$	$E[Y_0] = 1.6$

nonparametric) identification of average treatment effects.

Case 3

Case 3 slightly perturbs outcomes with treatment, Y_1, to create heterogeneous response as depicted in table 9.5. Suppose the states are observable to the analyst. Then, we can identify treatments effects by states. The treatment effect for all three states is homogeneously one.

Table 9.5: Tuebingen example case 3: heterogeneity

State (s)	*one*		*two*		*three*	
$\Pr(Y, D, s)$	0.0272	0.0128	0.224	0.096	0.5888	0.0512
D	0	1	0	1	0	1
Y	0	1	1	1	2	0
Y_0	0	0	1	1	2	2
Y_1	1	1	2	2	2	2
X	1	1	1	1	0	0

But, suppose the states are unobservable and the analyst observes X. Then, conditional average treatment effects are heterogeneous

$$\begin{aligned} E[Y_1 - Y_0 \mid X = 1] &= 1 \\ E[Y_1 - Y_0 \mid X = 0] &= 0 \end{aligned}$$

This follows from

$$E_X[E[Y_1 \mid X]] = 0.36(1.889) + 0.64(2) = 1.96$$

$$E_X[E[Y_0 \mid X]] = 0.36(0.889) + 0.64(2) = 1.6$$

but *OLS* (or nonparametric regression) estimates

$$E_X[E[Y_1 \mid X, D = 1]] = 0.68(1.882) + 0.32(2) = 1.92$$

and

$$E_X[E[Y_0 \mid X, D = 0]] = 0.299(0.892) + 0.701(2) = 1.669$$

Clearly, outcomes are not conditionally mean independent of treatment given X ($1.96 \neq 1.92$ for Y_1 and $1.6 \neq 1.669$ for Y_0). Key components, unconditional average (integrating out X) treatment effects, and any bias for case 3 are summarized in table 9.6. A modest change in outcomes with treatment produces endogeneity bias in all three average treatment effects (ATT, ATE, and $ATUT$). Average treatment effects are not identified by dummy variable regression (or nonparametric regression) in case 3.

Case 4

Case 4, described in table 9.7, maintains the probability structure of case 3 but alters outcomes with treatment, Y_1, to produce a Simpson's paradox result. Suppose the states are observable to the analyst. Then, we can identify treatments effects by states. The treatment effect for all three states is homogeneously one. But, suppose the states are unobservable and the analyst observes X. Then, conditional average treatment effects are heterogeneous

$$\begin{aligned} E[Y_1 - Y_0 \mid X = 1] &= 0.111 \\ E[Y_1 - Y_0 \mid X = 0] &= 0.3 \end{aligned}$$

Table 9.6: Tuebingen example case 3 results: heterogeneity

Results	Key components
$ATE = E[Y_1 - Y_0] = 0.36$	$p = \Pr(D = 1) = 0.16$
$ATT = E[Y_1 - Y_0 \mid D = 1] = 0.68$	$E[Y_1 \mid D = 1] = 1.92$
$ATUT = E[Y_1 - Y_0 \mid D = 0] = 0.299$	$E[Y_1 \mid D = 0] = 1.968$
$OLS = E[Y_1 \mid D = 1] - E[Y_0 \mid D = 0] = 0.251$	$E[Y_1] = 1.96$
$bias_{ATT} = E[Y_0 \mid D = 1] - E[Y_0 \mid D = 0] = -0.429$	$E[Y_0 \mid D = 1] = 1.24$
$bias_{ATUT} = E[Y_1 \mid D = 1] - E[Y_1 \mid D = 0] = -0.048$	$E[Y_0 \mid D = 0] = 1.669$
$bias_{ATE} = pbias_{ATT} + (1-p)\, bias_{ATUT} = -0.109$	$E[Y_0] = 1.6$

Table 9.7: Tuebingen example case 4: Simpson's paradox

State (s)	*one*		*two*		*three*	
$\Pr(Y, D, s)$	0.0272	0.0128	0.224	0.096	0.5888	0.0512
D	0	1	0	1	0	1
Y	0	1	1	1	2	2.3
Y_0	0	0	1	1	2	2
Y_1	1	1	1	1	2.3	2.3
X	1	1	1	1	0	0

This follows from

$$E_X[E[Y_1 \mid X]] = 0.36\,(1.0) + 0.64\,(2.3) = 1.832$$

$$E_X[E[Y_0 \mid X]] = 0.36\,(0.889) + 0.64\,(2) = 1.6$$

but *OLS* (or nonparametric regression) estimates

$$E_X[E[Y_1 \mid X, D = 1]] = 0.68\,(1.0) + 0.32\,(2.3) = 1.416$$

and

$$E_X[E[Y_0 \mid X, D = 0]] = 0.299\,(0.892) + 0.701\,(2) = 1.669$$

Clearly, outcomes are not conditionally mean independent of treatment given X ($1.932 \neq 1.416$ for Y_1 and $1.6 \neq 1.669$ for Y_0). Key components, unconditional average (integrating out X) treatment effects, and any bias for case 4 are summarized in table 9.8. Case 4 is particularly noteworthy as dummy variable regression (or nonparametric regression) indicates a negative treatment effect, while all three standard average treatment effects, ATE, ATT,and $ATUT$, are positive. Hence,

Table 9.8: Tuebingen example case 4 results: Simpson's paradox

Results	Key components
$ATE = E[Y_1 - Y_0]$ $= 0.232$	$p = \Pr(D = 1) = 0.16$
$ATT = E[Y_1 - Y_0 \mid D = 1]$ $= 0.176$	$E[Y_1 \mid D = 1] = 1.416$
$ATUT = E[Y_1 - Y_0 \mid D = 0]$ $= 0.243$	$E[Y_1 \mid D = 0] = 1.911$
$OLS = E[Y_1 \mid D = 1]$ $-E[Y_0 \mid D = 0] = -0.253$	$E[Y_1] = 1.832$
$bias_{ATT} = E[Y_0 \mid D = 1]$ $-E[Y_0 \mid D = 0] = -0.429$	$E[Y_0 \mid D = 1] = 1.24$
$bias_{ATUT} = E[Y_1 \mid D = 1]$ $-E[Y_1 \mid D = 0] = -0.495$	$E[Y_0 \mid D = 0] = 1.669$
$bias_{ATE} = pbias_{ATT}$ $+ (1-p)\, bias_{ATUT} = -0.485$	$E[Y_0] = 1.6$

average treatment effects are not identified by exogenous dummy variable regression (or nonparametric regression) for case 4.

How do we proceed when ignorable treatment (conditional mean independence) fails? A common response is to look for instruments and apply *IV* approaches to identify average treatment effects. Chapter 10 explores instrumental variable approaches. The remainder of this chapter surveys some other ignorable treatment approaches and applies them to the asset revaluation regulation problem introduced in chapter 2.

9.4 Nonparametric identification

Suppose treatment is ignorable or, in other words, treatment is conditionally mean independent of outcome,

$$E[Y_1 \mid X, D] = E[Y_1 \mid X]$$

and

$$E[Y_0 \mid X, D] = E[Y_0 \mid X]$$

This is also called "selection on observables" as the regressors are so powerful that we can ignore choice D. For binary treatment, this implies

$$E[Y_1 \mid X, D = 1] = E[Y_1 \mid X, D = 0]$$

and

$$E[Y_0 \mid X, D = 1] = E[Y_0 \mid X, D = 0]$$

The condition is difficult to test directly as it involves $E[Y_1 \mid X, D=0]$ and $E[Y_0 \mid X, D=1]$, the counterfactuals. Let $p(X) = \Pr(D=1 \mid X)$. Ignorable treatment implies the average treatment effect is nonparametrically identified.

$$\begin{aligned} ATE(X) &= E[\Delta \mid X] = E[Y_1 - Y_0 \mid X] \\ &= E[Y_1 \mid X] - E[Y_0 \mid X] \end{aligned}$$

By Bayes' theorem we can rewrite the expression as

$$\begin{aligned} &p(X) E[Y_1 \mid X, D=1] + (1 - p(X)) E[Y_1 \mid X, D=0] \\ &-p(X) E[Y_0 \mid X, D=1] - (1 - p(X)) E[Y_0 \mid X, D=0] \end{aligned}$$

conditional mean independence allows simplification to

$$E[Y_1 \mid X] - E[Y_0 \mid X] = ATE(X)$$

Consider a couple of ignorable treatment examples which distinguish between exogenous dummy variable and nonparametric identification.

Example 9.1 *The first example posits a simple case of stochastic independence between treatment D and response (Y_1, Y_0) conditional on X. The DGP is depicted in table 9.9 (values of D, Y_1, and Y_0 vary randomly at each level of X).*[4] *Clearly, if the response variables are stochastically independent of D conditional*

Table 9.9: Exogenous dummy variable regression example

probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$E[\cdot]$
$(Y_1 \mid X, D=1)$	0	1	0	2	0	3	1
$(Y_1 \mid X, D=0)$	0	1	0	2	0	3	1
$(Y_0 \mid X, D=1)$	−1	0	−2	0	−3	0	−1
$(Y_0 \mid X, D=0)$	−1	0	−2	0	−3	0	−1
X	1	1	2	2	3	3	2
$(D \mid X)$	0	1	0	1	0	1	0.5

on X

$$\Pr(Y_1 = y_1 \mid X = x, D=1) = \Pr(Y_1 = y_1 \mid X = x, D=0)$$

and

$$\Pr(Y_0 = y_0 \mid X = x, D=1) = \Pr(Y_0 = y_0 \mid X = x, D=0)$$

[4]The columns in the table are not states of nature but merely indicate the values the response Y_j and treatment D variables are allowed to take and their likelihoods. Conditional on X, the likelihoods for Y_j and D and independent.

then they are also conditionally mean independent

$$\begin{aligned} E[Y_1 \mid X=1, D=1] &= E[Y_1 \mid X=1, D=0] = 0.5 \\ E[Y_1 \mid X=2, D=1] &= E[Y_1 \mid X=2, D=0] = 1 \\ E[Y_1 \mid X=3, D=1] &= E[Y_1 \mid X=3, D=0] = 1.5 \end{aligned}$$

and

$$\begin{aligned} E[Y_0 \mid X=1, D=1] &= E[Y_0 \mid X=1, D=0] = -0.5 \\ E[Y_0 \mid X=2, D=1] &= E[Y_0 \mid X=2, D=0] = -1 \\ E[Y_0 \mid X=3, D=1] &= E[Y_0 \mid X=3, D=0] = -1.5 \end{aligned}$$

Conditional average treatment effects are

$$ATE(X=1) = 0.5 - (-0.5) = 1$$

$$ATE(X=2) = 1 - (-1) = 2$$

$$ATE(X=3) = 1.5 - (-1.5) = 3$$

and unconditional average treatment effects are

$$ATE = E[Y_1 - Y_0] = 1 - (-1) = 2$$

$$ATT = E[Y_1 - Y_0 \mid D=1] = 1 - (-1) = 2$$

$$ATUT = E[Y_1 - Y_0 \mid D=0] = 1 - (-1) = 2$$

Exogenous dummy variable regression

$$Y = \delta + \alpha D + X\beta_0 + D(X - E[X])(\beta_1 - \beta_0) + \varepsilon$$

consistently estimates ATE via α*. Based on a saturated "sample" of size* 384 *reflecting the DGP, dummy variable regression results are reported in table 9.10.*

Table 9.10: Exogenous dummy variable regression results

parameter	coefficient	se	t-statistic
δ	0.000	0.207	0.000
α	2.000	0.110	18.119
β_0	−0.500	0.096	−5.230
$\beta_1 - \beta_0$	1.000	0.135	7.397

The conditional regression estimates of average treatment effects

$$ATE(X=1) = 2 + 1(1-2) = 1$$

$$ATE(X=2) = 2 + 1(2-2) = 2$$

$$ATE(X=3) = 2 + 1(3-2) = 3$$

correspond well with the DGP. In this case, exogenous dummy variable regression identifies the average treatment effects.

Example 9.2 *The second example relaxes the DGP such that responses are conditionally mean independent but not stochastically independent and, importantly, the relations between outcomes and X are nonlinear. The DGP is depicted in table 9.11 (values of D, Y_1, and Y_0 vary randomly at each level of X).*[5] *Again,*

Table 9.11: Nonparametric treatment effect regression

probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$E[\cdot]$
$(Y_1 \mid X, D=1)$	0	1	0	2	0	3	1
$(Y_1 \mid X, D=0)$	0.5	0.5	1	1	1.5	1.5	1
$(Y_0 \mid X, D=1)$	-1	0	-2	0	-3	0	-1
$(Y_0 \mid X, D=0)$	-0.5	-0.5	-1	-1	-1.5	-1.5	-1
X	-1	-1	-2	-2	3	3	0
$(D \mid X)$	0	1	0	1	0	1	0.5

population average treatment effects are

$$ATE = E[Y_1 - Y_0] = 1 - (-1) = 2$$

$$ATT = E[Y_1 - Y_0 \mid D = 1] = 1 - (-1) = 2$$

$$ATUT = E[Y_1 - Y_0 \mid D = 0] = 1 - (-1) = 2$$

Further, the average treatment effects conditional on X are

$$ATE(X = -1) = 0.5 - (-0.5) = 1$$

$$ATE(X = -2) = 1 - (-1) = 2$$

$$ATE(X = 3) = 1.5 - (-1.5) = 3$$

Average treatment effects are estimated in two ways. First, exogenous dummy variable regression

$$Y = \delta + \alpha D + X\beta_0 + D(X - E[X])(\beta_1 - \beta_0) + \varepsilon$$

consistently estimates ATE via α. A saturated "sample" of 48 observations reflecting the DGP produces the results reported in table 9.12. However, the regression-estimated average treatment effects conditional on X are

$$ATE(X = -1) = 1.714$$

$$ATE(X = -2) = 1.429$$

$$ATE(X = 3) = 2.857$$

[5] Again, the columns of the table are not states of nature but merely indicate the values the variables can take conditional on X.

Table 9.12: Nonparametrically identified treatment effect: exogenous dummy variable regression results

parameter	coefficient	se	t-statistic
δ	-1.000	0.167	-5.991
α	2.000	0.236	8.472
β_0	-0.143	0.077	-1.849
$\beta_1 - \beta_0$	0.286	0.109	2.615

Hence, the conditional average treatment effects are not identified by exogenous dummy variable regression for this case. Second, let $\Im_x$ be an indicator variable for $X = x$. ANOVA is equivalent to nonparametric regression since X is sparse.

$$Y = \alpha D + \gamma_1 \Im_{-1} + \gamma_2 \Im_{-2} + \gamma_3 \Im_3 + \gamma_4 D\Im_{-1} + \gamma_5 D\Im_{-2} + \varepsilon$$

ANOVA results are reported in table 9.13. The ANOVA-estimated conditional av-

Table 9.13: Nonparametric treatment effect regression results

parameter	coefficient	se	t-statistic
α	3.000	0.386	7.774
γ_1	-0.500	0.273	-1.832
γ_2	-1.000	0.273	-3.665
γ_3	-1.500	0.273	-5.497
γ_4	-2.000	0.546	-3.665
γ_5	-1.000	0.546	-1.832

erage treatment effects are

$$ATE\left(X = -1\right) = 3 - 2 = 1$$

$$ATE\left(X = -2\right) = 3 - 1 = 2$$

$$ATE\left(X = 3\right) = 3$$

and the unconditional average treatment effect is

$$ATE = \frac{1}{3}\left(1 + 2 + 3\right) = 2$$

Therefore, even though the estimated average treatment effects for exogenous dummy variable regression are consistent with the DGP, the average treatment effects conditional on X do not correspond well with the DGP. Further, the treatment effects are not even monotonic in X. However, the ANOVA results properly account for the nonlinearity in the data and correspond nicely with the DGP for both unconditional and conditional average treatment effects. Hence, average treatment effects are nonparametrically identified for this case but not identified by exogenous dummy variable regression.

9.5 Propensity score approaches

Suppose the data are conditionally mean independent

$$E[Y_1 \mid X, D] = E[Y_1 \mid X]$$

$$E[Y_0 \mid X, D] = E[Y_0 \mid X]$$

so treatment is ignorable,and common X support leads to nondegenerate propensity scores

$$0 < p(X) = \Pr(D = 1 \mid X) < 1 \text{ for all } X$$

then average treatment effect estimands are

$$\begin{aligned} ATE &= E\left[\frac{(D - p(X))Y}{p(X)(1 - p(X))}\right] \\ ATT &= E\left[\frac{(D - p(X))Y}{(1 - p(X))}\right] / \Pr(D = 1) \\ ATUT &= E\left[\frac{(D - p(X))Y}{p(X)}\right] / \Pr(D = 0) \end{aligned}$$

The econometric procedure is to first estimate the propensity for treatment or propensity score, $p(X)$, via some flexible model (e.g., nonparametric regression; see chapter 6), then *ATE*, *ATT*, and *ATUT* are consistently estimated via sample analogs to the above.

9.5.1 ATE and propensity score

$ATE = E\left[\frac{(D-p(X))Y}{p(X)(1-p(X))}\right]$ is identified as follows. Observed outcome is

$$Y = DY_1 + (1 - D)Y_0$$

Substitution for Y and evaluation of the conditional expectation produces

$$\begin{aligned} & E[(D - p(X))Y \mid X] \\ = \; & E[DDY_1 + D(1 - D)Y_0 - p(X)DY_1 - p(X)(1 - D)Y_0 \mid X] \\ = \; & E[DY_1 + 0 - p(X)DY_1 - p(X)(1 - D)Y_0 \mid X] \end{aligned}$$

Letting $m_j(X) \equiv E[Y_j \mid X]$ and recognizing

$$\begin{aligned} p(X) &\equiv Pr(D = 1 \mid X) \\ &= E[D = 1 \mid X] \end{aligned}$$

gives

$$\begin{aligned} & E[DY_1 - p(X)DY_1 - p(X)(1 - D)Y_0 \mid X] \\ = \; & p(X)m_1(X) - p^2(X)m_1(X) - p(X)(1 - p(X))m_0(X) \\ = \; & p(X)(1 - p(X))(m_1(X) - m_0(X)) \end{aligned}$$

This leads to the conditional average treatment effect

$$\begin{aligned} E\left[\frac{p(X)(1-p(X))(m_1(X)-m_0(X))}{p(X)(1-p(X))} \mid X\right] &= m_1(X)-m_0(X) \\ &= E[Y_1-Y_0 \mid X] \end{aligned}$$

The final connection to the estimand is made by iterated expectations,

$$\begin{aligned} ATE &= E[Y_1-Y_0] \\ &= E_X[Y_1-Y_0 \mid X] \end{aligned}$$

9.5.2 ATT, ATUT, and propensity score

Similar logic identifies the estimand for the average treatment effect on the treated

$$ATT = E\left[\frac{(D-p(X))Y}{(1-p(X))}\right] / \Pr(D=1)$$

Utilize

$$E[(D-p(X))Y \mid X] = p(X)(1-p(X))(m_1(X)-m_0(X))$$

from the propensity score identification of *ATE*. Eliminating $(1-p(X))$ and rewriting gives

$$\begin{aligned} & \frac{p(X)(1-p(X))(m_1(X)-m_0(X))}{(1-p(X))} \\ = \; & p(X)(m_1(X)-m_0(X)) \\ = \; & \Pr(D=1 \mid X)(E[Y_1 \mid X]-E[Y_0 \mid X]) \end{aligned}$$

Conditional mean independence implies

$$\begin{aligned} & \Pr(D=1 \mid X)(E[Y_1 \mid X]-E[Y_0 \mid X]) \\ = \; & \Pr(D=1 \mid X)(E[Y_1 \mid D=1,X]-E[Y_0 \mid D=1,X]) \\ = \; & \Pr(D=1 \mid X)E[Y_1-Y_0 \mid D=1,X] \end{aligned}$$

Then, by iterated expectations, we have

$$\begin{aligned} & E_X[\Pr(D=1 \mid X)E[Y_1-Y_0 \mid D=1,X]] \\ = \; & \Pr(D=1)E[Y_1-Y_0 \mid D=1] \end{aligned}$$

Putting it all together produces the estimand

$$\begin{aligned} ATT &= E_X\left[\frac{(D-p(X))Y}{(1-p(X))}\right] / \Pr(D=1) \\ &= E[Y_1-Y_0 \mid D=1] \end{aligned}$$

For the average treatment effect on the untreated estimand

$$ATUT = E\left[\frac{(D - p(X))\,Y}{p(X)}\right] / \Pr(D = 0)$$

identification is analogous to that for *ATT*. Eliminating $p(X)$ from

$$E[(D - p(X))\,Y \mid X] = p(X)\,(1 - p(X))\,(m_1(X) - m_0(X))$$

and rewriting gives

$$\begin{aligned} & \frac{p(X)\,(1 - p(X))\,(m_1(X) - m_0(X))}{p(X)} \\ = \;& (1 - p(X))\,(m_1(X) - m_0(X)) \\ = \;& \Pr(D = 0 \mid X)\,(E[Y_1 \mid X] - E[Y_0 \mid X]) \end{aligned}$$

Conditional mean independence implies

$$\begin{aligned} & \Pr(D = 0 \mid X)\,(E[Y_1 \mid X] - E[Y_0 \mid X]) \\ = \;& \Pr(D = 0 \mid X)\,(E[Y_1 \mid D = 0, X] - E[Y_0 \mid D = 0, X]) \\ = \;& \Pr(D = 0 \mid X)\,E[Y_1 - Y_0 \mid D = 0, X] \end{aligned}$$

Iterated expectations yields

$$\begin{aligned} & E_X[\Pr(D = 0 \mid X)\,E[Y_1 - Y_0 \mid D = 0, X]] \\ = \;& \Pr(D = 0)\,E[Y_1 - Y_0 \mid D = 0] \end{aligned}$$

Putting everything together produces the estimand

$$\begin{aligned} ATUT &= E\left[\frac{(D - p(X))\,Y}{p(X)}\right] / \Pr(D = 0) \\ &= E[Y_1 - Y_0 \mid D = 0] \end{aligned}$$

Finally, the average treatment effects are connected as follows.

$$\begin{aligned} ATE &= \Pr(D = 1)\,ATT + \Pr(D = 0)\,ATUT \\ &= \Pr(D = 1)\,E\left[\frac{(D - p(X))\,Y}{(1 - p(X))}\right] / \Pr(D = 1) \\ &\quad + \Pr(D = 0)\,E\left[\frac{(D - p(X))\,Y}{p(X)}\right] / \Pr(D = 0) \\ &= E\left[\frac{(D - p(X))\,Y}{(1 - p(X))}\right] + E\left[\frac{(D - p(X))\,Y}{p(X)}\right] \\ &= E_X[\Pr(D = 1 \mid X)\,(E[Y_1 \mid X] - E[Y_0 \mid X])] \\ &\quad + E_X[\Pr(D = 0 \mid X)\,(E[Y_1 \mid X] - E[Y_0 \mid X])] \\ &= \Pr(D = 1)\,E[Y_1 - Y_0] + \Pr(D = 0)\,E[Y_1 - Y_0] \\ &= E[Y_1 - Y_0] \end{aligned}$$

9.5.3 *Linearity and propensity score*

If we add the condition $E[Y_0 \mid p(X)]$ and $E[Y_1 \mid p(X)]$ are linear in $p(X)$ then α in the expression below consistently estimates *ATE*

$$E[Y \mid X, D] = \varsigma_0 + \alpha D + \varsigma_1 \hat{p} + \varsigma_2 D\left(\hat{p} - \hat{\mu}_p\right)$$

where $\hat{\mu}_p$ is the sample average of the estimated propensity score $\hat{p}$.

9.6 Propensity score matching

Rosenbaum and Rubin's [1983] propensity score matching is a popular propensity score approach. Rosenbaum and Rubin suggest selecting a propensity score at random from the sample, then matching two individuals with this propensity score — one treated and one untreated. The expected outcome difference $E[Y_1 - Y_0 \mid p(X)]$ is ATE conditional on $p(X)$. Hence, by iterated expectations

$$ATE = E_{p(X)}[E[Y_1 - Y_0 \mid p(X)]]$$

ATE identification by propensity score matching poses strong ignorability. That is, outcome (Y_1, Y_0) independence of treatment D given X (a stronger condition than conditional mean independence) and, as before, common X support leads to nondegenerate propensity scores $p(X) \equiv \Pr(D = 1 \mid X)$

$$0 < \Pr(D = 1 \mid X) < 1 \text{ for all } X$$

As demonstrated by Rosenbaum and Rubin, strong ignorability implies index sufficiency. In other words, outcome (Y_1, Y_0) independence of treatment D given $p(X)$ and

$$0 < \Pr(D = 1 \mid p(X)) < 1 \text{ for all } p(X)$$

The latter (inequality) condition is straightforward. Since X is finer than $p(X)$, the first inequality (for X) implies the second (for $p(X)$). The key is conditional stochastic independence given the propensity score

$$\Pr(D = 1 \mid Y_1, Y_0, p(X)) = \Pr(D = 1 \mid p(X))$$

This follows from

$$\begin{aligned}
\Pr(D = 1 \mid Y_1, Y_0, p(X)) &= E[\Pr(D = 1 \mid Y_1, Y_0, X) \mid Y_1, Y_0, p(X)] \\
&= E[p(X) \mid Y_1, Y_0, p(X)] = p(X) \\
&= E[D \mid p(X)] \\
&= \Pr(D = 1 \mid p(X))
\end{aligned}$$

For a general matching strategy on X, Heckman, Ichimura, and Todd [1998] point out that for *ATT*, strong ignorability can be relaxed to conditional mean

independence for outcomes without treatment and full support S for the treated subsample. This allows counterfactuals to be related to observables

$$E[Y_0 \mid D = 1, X] = E[Y_0 \mid D = 0, X] \text{ for } X \in S$$

so that *ATT(X)* can be expressed in terms of observables only

$$\begin{aligned} ATT(X) &= E[Y_1 \mid D = 1, X] - E[Y_0 \mid D = 1, X] \\ &= E[Y_1 \mid D = 1, X] - E[Y_0 \mid D = 0, X] \end{aligned}$$

Iterated expectations gives the unconditional estimand

$$\begin{aligned} ATT &= E_{X \in S}[E[Y_1 \mid D = 1, X] - E[Y_0 \mid D = 1, X]] \\ &= E[Y_1 - Y_0 \mid D = 1] \end{aligned}$$

For *ATUT* the analogous condition applies to outcomes with treatment

$$E[Y_1 \mid D = 0, X] = E[Y_1 \mid D = 1, X] \text{ for } X \in S'$$

so that the counterfactual mean can be identified from observables.

$$\begin{aligned} ATUT(X) &= E[Y_1 \mid D = 0, X] - E[Y_0 \mid D = 0, X] \\ &= E[Y_1 \mid D = 1, X] - E[Y_0 \mid D = 0, X] \end{aligned}$$

Again, iterated expectations gives

$$\begin{aligned} ATUT &= E_{X \in S'}[E[Y_1 \mid D = 0, X] - E[Y_0 \mid D = 0, X]] \\ &= E[Y_1 - Y_0 \mid D = 0] \end{aligned}$$

Heckman, et al relate this general matching strategy to propensity score matching by the following arguments.[6] Partition X into two (not necessarily mutually exclusive) sets of variables, (T, Z), where the T variables determine outcomes and outcomes are additively separable

$$Y_0 = g_0(T) + U_0$$

$$Y_1 = g_1(T) + U_1$$

and the Z variables determine selection.

$$P(X) \equiv \Pr(D = 1 \mid X) = \Pr(D = 1 \mid Z) \equiv P(Z)$$

ATT is identified via propensity score matching if the following conditional mean independence condition for outcomes without treatment is satisfied

$$E[U_0 \mid D = 1, P(Z)] = E[U_0 \mid D = 0, P(Z)]$$

[6] Heckman, Ichimura, and Todd [1998] also discuss trade-offs between general matching on X and propensity score matching.

Then, the counterfactual $E[Y_0 \mid D = 1, P(Z)]$ can be replaced with the mean of the observable

$$\begin{aligned} ATT(P(Z)) &= E[Y_1 - Y_0 \mid D = 1, P(Z)] \\ &= g_1(T) + E[U_1 \mid D = 1, P(Z)] \\ &\quad - \{g_0(T) + E[U_0 \mid D = 1, P(Z)]\} \\ &= g_1(T) + E[U_1 \mid D = 1, P(Z)] \\ &\quad - \{g_0(T) + E[U_0 \mid D = 0, P(Z)]\} \end{aligned}$$

Iterated expectations over $P(Z)$ produces the unconditional estimand

$$ATT = E_{P(Z)}[ATT(P(Z))]$$

Also, *ATUT* is identified if

$$E[U_1 \mid D = 1, P(Z)] = E[U_1 \mid D = 0, P(Z)]$$

is satisfied for outcomes with treatment. Analogous to *ATT*, the counterfactual $E[Y_1 \mid D = 0, P(Z)]$ can be replaced with the mean of the observable

$$\begin{aligned} ATUT(P(Z)) &= E[Y_1 - Y_0 \mid D = 0, P(Z)] \\ &= g_1(T) + E[U_1 \mid D = 0, P(Z)] \\ &\quad - \{g_0(T) + E[U_0 \mid D = 0, P(Z)]\} \\ &= g_1(T) + E[U_1 \mid D = 1, P(Z)] \\ &\quad - \{g_0(T) + E[U_0 \mid D = 0, P(Z)]\} \end{aligned}$$

Iterated expectations over $P(Z)$ produces the unconditional estimand

$$ATUT = E_{P(Z)}[ATUT(P(Z))]$$

Interestingly, the original strategy of Rosenbaum and Rubin implies homogeneous response while the relaxed approach of Heckman, et al allows for heterogeneous response. To see this, notice the above conditions say nothing about

$$E[U_0 \mid D, P(Z)] = E[U_0 \mid P(Z)] = 0$$

or

$$E[U_1 \mid D, P(Z)] = E[U_1 \mid P(Z)] = 0$$

so individual effects (heterogeneity) are identified by conditional mean independence along with additive separability.

A strength of propensity score matching is that it makes the importance of overlaps clear. However, finding matches can be difficult. Heckman, Ichimura, and Todd [1997] discuss trimming strategies in a nonparametric context and derive asymptotically-valid standard errors. Next, we revisit our second example from chapter 2 to explore ignorable treatment implications in a richer accounting setting.

9.7 Asset revaluation regulation example

Our second example from chapter 2 explores the ex ante impact of accounting asset revaluation policies on owners' welfare through their investment decisions (a treatment effect) in an economy of, on average, price protected buyers.[7] Prior to investment, an owner evaluates both investment prospects from asset retention and the market for resale in the event the owner becomes liquidity stressed. The payoff from investment I is distributed uniformly and centered at $\hat{x} = \frac{\beta}{\alpha} I^{\alpha}$ where $\alpha, \beta > 0$ and $\alpha < 1$. That is, support for investment payoff is $x : \hat{x} \pm f = [\underline{x}, \overline{x}]$. A potential problem with the resale market is the owner will have private information — knowledge of the asset value. However, since there is some positive probability the owner becomes distressed π (as in Dye [1985]) the market will not collapse. The equilibrium price is based on distressed sellers marketing potentially healthy assets combined with non-distressed sellers opportunistically marketing impaired assets. Regulators may choose to prop-up the price to support distressed sellers by requiring certification of assets at cost k [8] with values below some cutoff x_c.[9] The owner's ex ante expected payoff from investment I and certification cutoff x_c is

$$\begin{aligned} E[V \mid I, x_c] &= \pi \frac{1}{2f}\left[\frac{1}{2}\left(x_c^2 - \underline{x}^2\right) - k\left(x_c - \underline{x}\right) + P\left(\overline{x} - x_c\right)\right] \\ &\quad + (1-\pi)\frac{1}{2f}\left[\frac{1}{2}\left(x_c^2 - \underline{x}^2\right) + P\left(P - x_c\right) + \frac{1}{2}\left(\overline{x}^2 - P^2\right)\right] \\ &\quad - I \end{aligned}$$

The equilibrium uncertified asset price is

$$P = \frac{x_c + \sqrt{\pi}\overline{x}}{1 + \sqrt{\pi}}$$

This follows from the equilibrium condition

$$P = \frac{1}{4fq}\left[\pi\left(\overline{x}^2 - x_c^2\right) + (1-\pi)\left(P^2 - x_c^2\right)\right]$$

where

$$q = \frac{1}{2f}\left[\pi\left(\overline{x} - x_c\right) + (1-\pi)\left(P - x_c\right)\right]$$

is the probability that an uncertified asset is marketed. When evaluating the welfare effects of their policies, regulators may differentially weight the welfare,

---

[7] This example draws heavily from Demski, Lin, and Sappington [2008].

[8] This cost is incremental to normal audit cost. As such, even if audit fee data is available, k may be difficult for the analyst to observe.

[9] Owners never find it ex ante beneficial to commit to any certified revaluation because of the certification cost. We restrict attention to targeted certification but certification could be proportional rather than targeted (see Demski, et al [2008] for details). For simplicity, we explore only targeted certification.

$W(I, x_c)$, of distressed sellers and non-distressed sellers. Specifically, regulators may value distressed seller's net gains dollar-for-dollar but value non-distressed seller's gains at a fraction w on the dollar.

$$\begin{aligned} W(I, x_c) &= \pi \frac{1}{2f}\left[\frac{1}{2}\left(x_c^2 - \underline{x}^2\right) - k\left(x_c - \underline{x}\right) + P\left(\overline{x} - x_c\right)\right] \\ &\quad + w(1-\pi)\frac{1}{2f}\left[\frac{1}{2}\left(x_c^2 - \underline{x}^2\right) + P\left(P - x_c\right) + \frac{1}{2}\left(\overline{x}^2 - P^2\right)\right] \\ &\quad - I\left[\pi + (1-\pi)w\right] \end{aligned}$$

9.7.1 Numerical example

Consider the following parameters

$$\left\{\alpha = \frac{1}{2}, \beta = 10, \pi = 0.7, k = 2, f = 100\right\}$$

Owners will choose to never certify asset values. No certification ($x_c = \underline{x}$) results in investment $I = 100$, owner's expected payoff $E[V \mid I, x_c] = 100$, and equilibrium uncertified asset price $P \approx 191.1$. However, regulators may favor distressed sellers and require selective certification. Continuing with the same parameters, if regulators give zero consideration ($w = 0$) to the expected payoffs of non-distressed sellers, then the welfare maximizing certification cutoff $x_c = \overline{x} - \frac{\left(1+\sqrt{\pi}\right)k}{\left(1-\sqrt{\pi}\right)(1-w)} \approx 278.9$. This induces investment $I = \left[\frac{\beta(2f+\pi k)}{2f}\right]^{\frac{1}{1-\alpha}} \approx 101.4$, owner's expected payoff approximately equal to 98.8, and equilibrium uncertified asset price $P \approx 289.2$ (an uncertified price more favorable to distressed sellers). To get a sense of the impact of certification, we tabulate investment choices and expected payoffs for no and selective certification regulations and varied certification costs in table 9.14 and for full certification regulation and varied certification costs and stress likelihood in table 9.15.

Table 9.14: Investment choice and payoffs for no certification and selective certification

	$x_c = \underline{x}$, $k = 2$	$x_c = \underline{x}$, $k = 20$	$x_c = 200$, $k = 2$	$x_c = 200$, $k = 20$
π	0.7	0.7	0.7	0.7
I	100	100	101.4	114.5
P	191.1	191.1	246.2	251.9
$E[x-k]$	200	200	200.7	208
$E[V]$	100	100	99.3	93.5

Table 9.15: Investment choice and payoffs for full certification

	$x_c = \overline{x}$, $k = 2$	$x_c = \overline{x}$, $k = 20$	$x_c = \overline{x}$, $k = 2$	$x_c = \overline{x}$, $k = 20$
π	0.7	0.7	1.0	1.0
I	100	100	100	100
P	NA	NA	NA	NA
$E[x-k]$	198.6	186	198	180
$E[V]$	98.6	86	98	80

9.7.2 Full certification

The base case involves full certification $x_c = \overline{x}$ and all owners market their assets, $\pi = 1$. This setting ensures outcome data availability (excluding investment cost) which may be an issue when we relax these conditions. There are two firm types: one with low mean certification costs $\widehat{k}^L = 2$ and the other with high mean certification costs $\widehat{k}^H = 20$.

Full certification doesn't present an interesting experiment if owners anticipate full certification[10] but suppose owners choose their investment levels anticipating selective certification with $x_c = 200$ and forced sale is less than certain $\pi = 0.7$. Then, ex ante optimal investment levels for a selective certification environment are $I^L = 101.4$ (for low certification cost type) and $I^H = 114.5$ (for high certification cost type), and expected asset values including certification costs are $E\left[x^L - k^L\right] = 199.4$ and $E\left[x^H - k^H\right] = 194$. Treatment (investment level) is chosen based on ex ante beliefs of selective certification. As a result of two certification cost types, treatment is binary and the analyst observes low or high investment but not the investment level.[11] Treatment is denoted $D = 1$ when $I^L = 101.4$ while non-treatment is denoted $D = 0$ when $I^H = 114.5$. For this base case, outcome is ex post value in an always certify, always trade environment $Y_j = x^j - k^j$.

To summarize, the treatment effect of interest is the difference in outcome with treatment and outcome without treatment. For the base case, outcome with treatment is defined as realized value associated with the (ex ante) equilibrium investment choice when certification cost type is low (I^L). And, outcome with no treatment is defined as realized value associated with (ex ante) equilibrium investment choice when certification cost type is high (I^H). Variations from the base case retain the definition for treatment (low versus high investment) but alter outcomes based on data availability given the setting (e.g., assets are not always traded so values may not be directly observed).

[10] As seen in the table, for full certification there is no variation in equilibrium investment level.

[11] If the analyst observes the investment level, then outcome includes investment cost and we work with a more complete measure of the owner's welfare.

Since the equilibrium investment choice for low certification cost type is treatment (I^L), the average treatment effect on the treated is

$$\begin{aligned} ATT &= E\left[Y_1 - Y_0 \mid D = 1\right] \\ &= E\left[x^L - k^L \mid D = 1\right] - E\left[x^H - k^L \mid D = 1\right] \\ &= E\left[x^L - x^H \mid D = 1\right] \\ &= 201.4 - 214 = -12.6 \end{aligned}$$

Similarly, the equilibrium investment choice for high certification cost type is no treatment (I^H). Therefore, the average treatment effect on the untreated is

$$\begin{aligned} ATUT &= E\left[Y_1 - Y_0 \mid D = 0\right] \\ &= E\left[x^L - k^H \mid D = 0\right] - E\left[x^H - k^H \mid D = 0\right] \\ &= E\left[x^L - x^H \mid D = 0\right] \\ &= 201.4 - 214 = -12.6 \end{aligned}$$

The above implies outcome is homogeneous,[12] $ATE = ATT = ATUT = -12.6$. With no covariates and outcome not mean independent of treatment, the *OLS* estimand is[13]

$$\begin{aligned} OLS &= E\left[Y_1 \mid D = 1\right] - E\left[Y_0 \mid D = 0\right] \\ &= E\left[x^L - k^L \mid D = 1\right] - E\left[x^H - k^H \mid D = 0\right] \\ &= 5.4 \end{aligned}$$

The regression is

$$E\left[Y \mid D\right] = \beta_0 + \beta_1 D$$

where $Y = D\left(x^L - k^L\right) + (1 - D)\left(x^H - k^H\right)$ (ex post payoff), β_1 is the estimand of interest, and

$$D = \begin{array}{ll} 1 & I^L = 101.4 \\ 0 & I^H = 114.5 \end{array}$$

A simple experiment supports the analysis above. We simulate 200 samples of $2,000$ draws where traded market values are

$$x^j \sim uniform\left(\widehat{x}^j - 100, \widehat{x}^j + 100\right)$$

certification costs are

$$k^j \sim uniform\left(\widehat{k}^j - 1, \widehat{k}^j + 1\right)$$

---

[12] If k is unobservable, then outcome Y may be measured by x only (discussed later) and treatment effects represent gross rather than gains net of certification cost. In any case, we must exercise care in interpreting the treatment effects because of limitations in our outcome measure — more to come on the importance of outcome observability.

[13] Notice the difference in the treatment effects and what is estimated via *OLS* is $k^L - k^H = 2 - 20 = -18 = -12.6 - 5.4$.

and assignment of certification cost type is

$$L - type \sim Bernoulli\,(0.5)$$

Simulation results for the above *OLS* model including the estimated average treatment effect are reported in table 9.16. As simulation allows us to observe both the factual data and counterfactual data in the experiment, the sample statistics described in table 9.17 are "observed" average treatment effects.

Table 9.16: OLS results for full certification setting

$statistics$	β_0	$\beta_1\,(estATE)$
$mean$	193.8	5.797
$median$	193.7	5.805
$stand.dev.$	1.831	2.684
$minimum$	188.2	-1.778
$maximum$	198.9	13.32
$E\,[Y \mid D] = \beta_0 + \beta_1 D$		

Table 9.17: Average treatment effect sample statistics for full certification setting

$statistics$	ATE	ATT	$ATUT$
$mean$	-12.54	-12.49	-12.59
$median$	-12.55	-12.44	-12.68
$stand.dev.$	1.947	2.579	2.794
$minimum$	-17.62	-19.53	-21.53
$maximum$	-7.718	-6.014	-6.083

OLS clearly produces biased estimates of the treatment effect in this simple base case. This can be explained as low or high certification cost type is a perfect predictor of treatment. That is, $\Pr\left(D = 1 \mid k^L\right) = 1$ and $\Pr\left(D = 1 \mid k^H\right) = 0$. Therefore, the common support condition for identifying counterfactuals fails and standard approaches (ignorable treatment or even instrumental variables) don't identify treatment effects.[14]

[14] An alternative analysis tests the common support condition. Suppose everything remains as above except $k^H \sim uniform\,(1, 19)$ and sometimes the owners perceive certification cost to be low when it is high, hence $\Pr\,(D = 1 \mid \text{type} = H) = 0.1$. This setup implies observed outcome is

$$Y = D\,[(Y_1 \mid \text{type} = L) + (Y_1 \mid \text{type} = H)] + (1 - D)\,(Y_0 \mid \text{type} = H)$$

such that

$$E\,[Y] = 0.5E\left[x^L - k^L\right] + 0.5\left\{0.1E\left[x^L - k^H\right] + 0.9E\left[x^H - k^H\right]\right\}$$

Suppose the analyst ex post observes the actual certification cost type and let $T = 1$ if type $= L$. The common support condition is satisfied and the outcome mean is conditionally independent of treatment given T implies treatment is ignorable. *OLS* simulation results are tabulated below.

Adjusted outcomes

However, from the above we can manipulate the outcome variable to identify the treatment effects via *OLS*. Observed outcome is

$$\begin{aligned} Y &= D\left(x^L - k^L\right) + (1-D)\left(x^H - k^H\right) \\ &= \left(x^H - k^H\right) + D\left(x^L - x^H\right) - D\left(k^L - k^H\right) \end{aligned}$$

Applying expectations, the first term is captured via the regression intercept and the second term is the average treatment effect. Therefore, if we add the last term $DE\left[k^L - k^H\right]$ to Y we can identify the treatment effect from the coefficient on D. If the analyst observes $k = Dk^L + (1-D)\,k^H$, then we can utilize a two-stage regression approach. The first stage is

$$E\left[k \mid D\right] = \alpha_0 + \alpha_1 D$$

where $\alpha_0 = E\left[k^H\right]$ and $\alpha_1 = E\left[k^L - k^H\right]$. Now, the second stage regression employs the sample statistic for α_1, $\widehat{\alpha}_1 = \overline{k}^L - \overline{k}^H$.

$$\begin{aligned} Y' &= Y + D\widehat{\alpha}_1 \\ &= Y + D\left(\overline{k}^L - \overline{k}^H\right) \end{aligned}$$

and estimate the treatment effect via the analogous regression to the above[15]

$$E\left[Y' \mid D\right] = \beta_0 + \beta_1 D$$

OLS parameter estimates with common support for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	196.9	7.667	-5.141
$median$	196.9	7.896	-5.223
$stand.dev.$	1.812	6.516	6.630
$minimum$	191.5	-10.62	-23.54
$maximum$	201.6	25.56	14.25
$E\left[Y \mid T, D\right] = \beta_0 + \beta_1 T + \beta_2 D$			

Average treatment effects sample statistics with common support for full certification setting

$statistics$	ATE	ATT	$ATUT$
$mean$	-5.637	-5.522	-5.782
$median$	-5.792	-5.469	-5.832
$stand.dev.$	1.947	2.361	2.770
$minimum$	-9.930	-12.05	-12.12
$maximum$	0.118	0.182	0.983

The estimated average treatment effect is slightly attenuated and has high variability that may compromise its finite sample utility. Nevertheless, the results are a dramatic departure and improvement from the results above where the common support condition fails.

[15]This is similar to a regression discontinuity design (for example, see Angrist and Lavy [1999] and Angrist and Pischke [2009]). However, the jump in cost of certification k^j violates the regression continuity in X condition (assuming $k = Dk^L + (1-D)\,k^H$ is observed and included in X). If the

Simulation results for the adjusted outcome *OLS* model are reported in table 9.18. With adjusted outcomes, *OLS* estimates correspond quite well with the treatment effects. Next, we explore propensity score approaches.

Table 9.18: Adjusted outcomes OLS results for full certification setting

$statistics$	β_0	$\beta_1\,(estATE)$
$mean$	193.8	-12.21
$median$	193.7	-12.21
$stand.dev.$	1.831	2.687
$minimum$	188.2	-19.74
$maximum$	198.9	-64.691
$E[Y' \mid D] = \beta_0 + \beta_1 D$		

Propensity score

Based on adjusted outcomes, the data are conditionally mean independent (i.e., satisfy ignorability of treatment). Therefore average treatment effects can be estimated via the propensity score as discussed earlier in chapter 9. Propensity score is the estimated probability of treatment conditional on the regressors $m_j = \Pr(D_j = 1 \mid Z_j)$. For simulation purposes, we employ an imperfect predictor in the probit regression

$$Z_j = z_{1j} D_j + z_{0j} (1 - D_j) + \varepsilon_j$$

---

support of k^L and k^H is adjacent, then the regression discontinuity design

$$E[Y \mid X, D] = \beta_0 + \beta_1 k + \beta_2 D$$

effectively identifies the treatment effects but fails with the current *DGP*. Typical results for the current *DGP* (where *ATE* is the average treatment effect sample statistic for the simulation) are tabulated below.

OLS parameter estimates with jump in support for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$	ATE
$mean$	218.4	-1.210	-16.59	-12.54
$median$	213.2	-0.952	-12.88	-12.55
$stand.dev.$	42.35	2.119	38.26	1.947
$minimum$	122.9	-6.603	-115.4	-17.62
$maximum$	325.9	3.573	71.56	-7.718
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$				

The coefficient on D represents a biased and erratic estimate of the average treatment effect. Given the variability of the estimates, a regression discontinuity design has limited small sample utility for this *DGP*. However, we later return to regression discontinuity designs when modified *DGP*s are considered. For the current *DGP*, we employ the approach discussed above, which is essentially restricted least squares.

where

$$\begin{aligned} z_{1j} &\sim Bernoulli\,(0.99) \\ z_{0_j} &\sim Bernoulli\,(0.01) \\ \varepsilon_j &\sim N\,(0,1) \end{aligned}$$

Some average treatment effects estimated via propensity score are

$$\begin{aligned} estATE &= n^{-1}\sum_{j=1}^{n}\frac{(D_j - m_j)\,Y_j'}{m_j\,(1-m_j)} \\ estATT &= \frac{n^{-1}\sum_{j=1}^{n}\frac{(D_j-m_j)Y_j'}{(1-m_j)}}{n^{-1}\sum_{j=1}^{n}D_j} \\ estATUT &= \frac{n^{-1}\sum_{j=1}^{n}\frac{(D_j-m_j)Y_j'}{m_j}}{n^{-1}\sum_{j=1}^{n}(1-D_j)} \end{aligned}$$

Propensity score estimates of average treatment effects are reported in table 9.19. The estimates are somewhat more variable than we would like but they are

Table 9.19: Propensity score treatment effect estimates for full certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.42	-13.96	-10.87
$median$	-12.50	-13.60	-11.40
$stand.dev.$	5.287	6.399	5.832
$minimum$	-31.83	-45.83	-25.61
$maximum$	-1.721	0.209	10.56

consistent with the sample statistics on average. Further, we cannot reject homogeneity even though the treatment effect means are not as similar as we might expect.

Propensity score matching

Propensity score matching is a simple and intuitively appealing approach where we match treated and untreated on propensity score then compute the average treatment effect based on the matched-pair outcome differences. We follow Sekhon [2008] by employing the "Matching" library for **R.**[16] We find optimal matches of

[16]We don't go into details regarding matching since we employ only one regressor in the propensity score model. Matching is a rich study in itself. For instance, Sekhon [2008] discusses a genetic matching algorithm. Heckman, Ichimura, and Todd [1998] discuss nonparametric kernel matching.

treated with untreated (within 0.01) using replacement sampling. Simulation results for propensity score matching average treatment effects are reported in table 9.20.[17] The matched propensity score results correspond well with the sample sta-

Table 9.20: Propensity score matching average treatment effect estimates for full certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.46	-12.36	-12.56
$median$	-12.54	-12.34	-12.36
$stand.dev.$	3.530	4.256	4.138
$minimum$	-23.49	-24.18	-22.81
$maximum$	-3.409	-2.552	-0.659

tistics. In this setting, the matched propensity score estimates of causal effects are less variable than the previous propensity score results. Further, they are more uniform across treatment effects (consistent with homogeneity). Next, we turn to the more interesting, but potentially more challenging, selective certification setting.

9.7.3 Selective certification

Suppose the owners' ex ante perceptions of the certification threshold, $x_c = 200$, and likelihood of stress, $\pi = 0.7$, are consistent with ex post outcomes. Then, if outcomes x, k^j, and P^j for $j = L$ or H are fully observable to the analyst, expected outcome conditional on asset revaluation experience is[18]

$$\begin{aligned} E[Y \mid X] &= 251.93 \Pr\left(P^H\right) - 5.740 \Pr\left(P^L\right) D - 94.93 \Pr\left(\Im_c^H\right) \Im_c^H \\ &\quad -95.49 \Pr\left(\Im_c^L\right) \Im_c^L - 20 \Pr\left(\Im_{ck}^H\right) \Im_{ck}^H \\ &\quad -2 \Pr\left(\Im_{ck}^L\right) \Im_{ck}^L + 31.03 \Pr\left(\Im_u^H\right) \Im_u^H + 27.60 \Pr\left(\Im_u^L\right) \Im_u^L \\ E[Y \mid X] &= 251.93 (0.477) - 5.740 (0.424) D - 94.93 (0.129) \Im_c^H \\ &\quad -95.49 (0.148) \Im_c^L - 20 (0.301) \Im_{ck}^H - 2 (0.345) \Im_{ck}^L \\ &\quad +31.03 (0.093) \Im_u^H + 27.60 (0.083) \Im_u^L \end{aligned}$$

[17] *ATE*, *ATT*, and *ATUT* may be different because their regions of common support may differ. For example, *ATT* draws on common support only in the $D = 1$ region and *ATUT* draws on common support only in the $D = 0$ region.

[18] The probabilities reflect likelihood of the asset condition rather than incremental likelihood and hence sum to one for each investment level (treatment choice).

where the equilibrium price of traded, uncertified, high investment assets, P^H, is the reference outcome level, and X denotes the matrix of regressors

$$\begin{aligned}
D &= \begin{matrix} 1 & \text{low investment, } I^L \\ 0 & \text{high investment, } I^H \end{matrix} \\
\Im_c^j &= \begin{matrix} 1 & \text{certified range, } x < x_c \\ 0 & \text{otherwise} \end{matrix} \\
\Im_{ck}^j &= \begin{matrix} 1 & \text{certified traded} \\ 0 & \text{otherwise} \end{matrix} \\
\Im_u^j &= \begin{matrix} 1 & \text{untraded asset, } x > P \quad j \in \{L, H\} \\ 0 & \text{otherwise} \end{matrix}
\end{aligned}$$

This implies the average treatment effect estimands are

$$\begin{aligned}
ATT &\equiv E\left[Y^L - Y^H \mid D = 1\right] = -12.7 \\
ATUT &\equiv E\left[Y^L - Y^H \mid D = 0\right] = -13.5
\end{aligned}$$

and

$$\begin{aligned}
ATE &\equiv E\left[Y^L - Y^H\right] \\
&= \Pr(D = 1)\, ATT + \Pr(D = 0)\, ATUT = -13.1
\end{aligned}$$

Hence, in the selective certification setting we encounter modest heterogeneity. Why don't we observe self-selection through the treatment effects? Remember, we have a limited outcome measure. In particular, outcome excludes investment cost. If we include investment cost, then self-selection is supported by the average treatment effect estimands. That is, low investment outcome is greater than high investment outcome for low certification cost firms

$$ATT = -12.7 - (101.4 - 114.5) = 0.4 > 0$$

and high investment outcome is greater than low investment outcome for high certification cost firms

$$ATUT = -13.5 - (101.4 - 114.5) = -0.4 < 0$$

With this background for the selective certification setting, it's time to revisit identification. Average treatment effect identification is somewhat more challenging than the base case. For instance, the average treatment effect on the treated, *ATT*, is the difference between the mean of outcome with low investment and the

mean of outcome with high investment for low certification cost firms.

$$\begin{aligned} ATT &= \pi \frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)-k^L\left(x_c-\underline{x}^L\right)+P^L\left(\overline{x}^L-x_c\right)\right] \\ &\quad +(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)+P^L\left(P^L-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^L\right)^2-\left(P^L\right)^2\right)\end{array}\right] \\ &\quad -\pi\frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)-k^L\left(x_c-\underline{x}^H\right)+P^H\left(\overline{x}^H-x_c\right)\right] \\ &\quad -(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)+P^H\left(P^H-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^H\right)^2-\left(P^H\right)^2\right)\end{array}\right] \end{aligned}$$

The average treatment effect on the untreated, *ATUT*, is the difference between the mean of outcome with low investment and the mean of outcome with high investment for high certification cost firms.

$$\begin{aligned} ATUT &= \pi \frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)-k^H\left(x_c-\underline{x}^L\right)+P^L\left(\overline{x}^L-x_c\right)\right] \\ &\quad +(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)+P^L\left(P^L-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^L\right)^2-\left(P^L\right)^2\right)\end{array}\right] \\ &\quad -\pi\frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)-k^H\left(x_c-\underline{x}^H\right)+P^H\left(\overline{x}^H-x_c\right)\right] \\ &\quad -(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)+P^H\left(P^H-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^H\right)^2-\left(P^H\right)^2\right)\end{array}\right] \end{aligned}$$

But the *OLS* estimand is the difference between the mean of outcome with low investment for firms with low certification cost and the mean of outcome with high investment for firms with high certification cost.

$$\begin{aligned} OLS &= \pi \frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)-k^L\left(x_c-\underline{x}^L\right)+P^L\left(\overline{x}^L-x_c\right)\right] \\ &\quad +(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^L\right)^2\right)+P^L\left(P^L-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^L\right)^2-\left(P^L\right)^2\right)\end{array}\right] \\ &\quad -\pi\frac{1}{2f}\left[\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)-k^H\left(x_c-\underline{x}^H\right)+P^H\left(\overline{x}^H-x_c\right)\right] \\ &\quad -(1-\pi)\frac{1}{2f}\left[\begin{array}{c}\frac{1}{2}\left(x_c^2-\left(\underline{x}^H\right)^2\right)+P^H\left(P^H-x_c\right)\\ +\frac{1}{2}\left(\left(\overline{x}^H\right)^2-\left(P^H\right)^2\right)\end{array}\right] \end{aligned}$$

As in the full certification setting, the key differences revolve around the costly certification terms. The costly certification term for the *ATT* estimand simplifies

as

$$\begin{aligned} & -\pi \frac{1}{2f}\left[k^L\left(x_c-\underline{x}^L\right)-k^L\left(x_c-\underline{x}^H\right)\right] \\ = \quad & -\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}k^L \end{aligned}$$

and the costly certification term for the *ATUT* estimand simplifies as

$$\begin{aligned} & -\pi \frac{1}{2f}\left[k^H\left(x_c-\underline{x}^L\right)-k^H\left(x_c-\underline{x}^H\right)\right] \\ = \quad & -\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}k^H \end{aligned}$$

While the costly certification term in the estimand for *OLS* is

$$-\pi \frac{1}{2f}\left[k^L\left(x_c-\underline{x}^L\right)-k^H\left(x_c-\underline{x}^H\right)\right]$$

Adjusted outcomes

Similar to our approach in the full certification setting, we eliminate the costly certification term for *OLS* by adding this *OLS* bias to observed outcomes

$$Y' = Y + D\pi \frac{1}{2f}\left[k^L\left(x_c-\underline{x}^L\right)-k^H\left(x_c-\underline{x}^H\right)\right]$$

However, now we add back the terms to recover the average treatment effects

$$ATT = E\left[Y_1^{'}-Y_0^{'} \mid D=1\right]-\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}E\left[k^L\right]$$

$$ATUT = E\left[Y_1^{'}-Y_0^{'} \mid D=0\right]-\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}E\left[k^H\right]$$

$$\begin{aligned} ATE \quad = \quad & \Pr\left(D=1\right)ATT+\Pr\left(D=1\right)ATUT \\ = \quad & E\left[Y_1^{'}-Y_0^{'}\right]-\Pr\left(D=1\right)\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}E\left[k^L\right] \\ & -\Pr\left(D=0\right)\pi \frac{\left(\underline{x}^H-\underline{x}^L\right)}{2f}E\left[k^H\right] \end{aligned}$$

These terms account for heterogeneity in this asset revaluation setting but are likely to be much smaller than the *OLS* selection bias.[19]

[19] In our running numerical example, the certification cost term for *ATT* is -0.0882 and for *ATUT* is -0.882, while the *OLS* selection bias is 5.3298.

Conditional as well as unconditional average treatment effects can be identified from the following regression.

$$\begin{aligned} E\left[Y' \mid X\right] &= \beta_0 + \beta_1 D + \beta_2 \Im_c^H + \beta_3 \Im_c^L \\ &\quad + \beta_4 \Im_{ck}^H + \beta_5 \Im_{ck}^L + \beta_6 \Im_u^H + \beta_7 \Im_u^L \end{aligned}$$

where

$$Y' = Y + D\pi \left[\Im_{ck}^L k^L - \overline{\Im}_{ck}^H \overline{k}^H\right]$$

$\overline{\Im}_{ck}^H$ and $\overline{k}^H$ are sample averages taken from the $D = 0$ regime.[20] The incremental impact on mean value of assets in the certification region is reflected in β_2 for high investment and β_3 for low investment firms, while the mean incremental impact of costly certification of assets, k^j, is conveyed via β_4 and β_5 for high and low investment firms, respectively. Finally, the mean incremental impact of untraded assets with values greater than the equilibrium price are conveyed via β_6 and β_7 for high and low investment firms, respectively.

Simulation results for the *OLS* model are reported in table 9.21 and sample treatment effect statistics are reported in table 9.22. *OLS* effectively estimates the average treatment effects (*ATE*, *ATT*, *ATUT*) in this (modestly heterogeneous) case. However, we're unlikely to be able to detect heterogeneity when the various treatment effect differences are this small. Note in this setting, while outcome is the ex post value net of certification cost, a random sample allows us to assess the owner's ex ante welfare excluding the cost of investment.[21]

Model-estimated treatment effects are derived in a non-standard manner as the regressors are treatment-type specific and we rely on sample evidence from each regime to estimate the probabilities associated with different ranges of support[22]

$$\begin{aligned} estATT &= \beta_1 - \beta_2 \overline{\Im}_c^H + \beta_3 \overline{\Im}_c^L - \beta_4 \overline{\Im}_{ck}^H + \beta_5 \overline{\Im}_{ck}^L - \beta_6 \overline{\Im}_u^H + \beta_7 \overline{\Im}_u^L \\ &\quad - \pi \overline{k}^L \left(\overline{\Im}_{ck}^L - \overline{\Im}_{ck}^H\right) \end{aligned}$$

$$\begin{aligned} estATUT &= \beta_1 - \beta_2 \overline{\Im}_c^H + \beta_3 \overline{\Im}_c^L - \beta_4 \overline{\Im}_{ck}^H + \beta_5 \overline{\Im}_{ck}^L - \beta_6 \overline{\Im}_u^H + \beta_7 \overline{\Im}_u^L \\ &\quad - \pi \overline{k}^H \left(\overline{\Im}_{ck}^L - \overline{\Im}_{ck}^H\right) \end{aligned}$$

and

$$\begin{aligned} estATE &= \beta_1 - \beta_2 \overline{\Im}_c^H + \beta_3 \overline{\Im}_c^L - \beta_4 \overline{\Im}_{ck}^H + \beta_5 \overline{\Im}_{ck}^L - \beta_6 \overline{\Im}_u^H + \beta_7 \overline{\Im}_u^L \\ &\quad - \overline{D} \pi \overline{k}^L \left(\overline{\Im}_{ck}^L - \overline{\Im}_{ck}^H\right) - \left(1 - \overline{D}\right) \pi \overline{k}^H \left(\overline{\Im}_{ck}^L - \overline{\Im}_{ck}^H\right) \end{aligned}$$

---

[20] Sample averages of certification cost, $\overline{k}^H$, and likelihood that an asset is certified and traded, $\overline{\Im}_{ck}^H$, for $D = 0$ (high investment) are employed as these are counterfactuals in the $D = 1$ (low investment) regime.

[21] Investment cost may also be observed or estimable by the analyst.

[22] Expected value of indicator variables equals the event probability and probabilities vary by treatment. Since there is no common support (across regimes) for the regressors, we effectively assume the analyst can extrapolate to identify counterfactuals (that is, from observed treated to unobserved treated and from observed untreated to unobserved untreated).

Table 9.21: OLS parameter estimates for selective certification setting

$statistics$	β_0	β_1	β_2	β_3
$mean$	251.9	-11.78	-94.78	-93.81
$median$	251.9	-11.78	-94.70	-93.85
$stand.dev.$	0.000	0.157	2.251	2.414
$minimum$	251.9	-12.15	-102.7	-100.9
$maximum$	251.9	-11.41	-88.98	-86.90
$statistics$	β_4	β_5	β_6	β_7
$mean$	-20.12	-2.087	31.20	27.66
$median$	-20.15	-2.160	31.23	27.72
$stand.dev.$	2.697	2.849	1.723	1.896
$minimum$	-28.67	-9.747	26.91	22.44
$maximum$	-12.69	8.217	37.14	32.81
$statistics$	$estATE$	$estATT$	$estATUT$	
$mean$	-12.67	-12.29	-13.06	
$median$	-12.73	-12.33	-13.10	
$stand.dev.$	2.825	2.686	2.965	
$minimum$	-21.25	-20.45	-22.03	
$maximum$	-3.972	-3.960	-3.984	

$$E[Y' \mid X] = \beta_0 + \beta_1 D + \beta_2 \Im_c^H + \beta_3 \Im_c^L + \beta_4 \Im_{ck}^H + \beta_5 \Im_{ck}^L + \beta_6 \Im_u^H + \beta_7 \Im_u^L$$

where

$$\overline{\Im}_j^L = \frac{\sum Di \Im_{ji}^L}{\sum Di}$$

and

$$\overline{\Im}_j^H = \frac{\sum (1 - D_i) \Im_{ji}^H}{\sum (1 - D_i)}$$

for indicator j.

We can say a bit more about conditional average treatment effects from the above analysis. On average, owners who select high investment and trade the assets at their equilibrium price sell the assets for 11.78 more than owners who select low investment. Owners who select high investment and retain their assets earn $31.20 - 27.66 = 3.54$ higher proceeds, on average, than owners who select low investment. On the other hand, owners who select high investment and are forced to certify and sell their assets receive lower net proceeds by $20.12 - 2.09 = 18.03$, on average, than owners who select low investment. Recall all outcomes exclude investment cost which, of course, is an important component of owner's welfare.

As we can effectively randomize over the indicator variables, for simplicity, we focus on identification and estimation of unconditional average treatment effects and the remaining analyses are explored without covariates. Next, we demonstrate the above randomization claim via a reduced (no covariates except treatment) *OLS* model, then we explore propensity score approaches applied to selective certification.

Table 9.22: Average treatment effect sample statistics for selective certification setting

$statistics$	ATE	ATT	$ATUT$	OLS
$mean$	-13.01	-12.57	-13.45	-6.861
$median$	-13.08	-12.53	-13.46	-6.933
$stand.dev.$	1.962	2.444	2.947	2.744
$minimum$	-17.90	-19.52	-22.46	-15.15
$maximum$	-8.695	-5.786	-6.247	1.466

Reduced OLS model

We estimate unconditional average treatment effects via a reduced *OLS* model.

$$E\left[Y' \mid D\right] = \beta_0 + \beta_1 D$$

Results from the simulation, reported in table 9.23, indicate that reduced *OLS*, with the adjustments discussed above to recover the treatment effect, effectively recovers unconditional average treatment effects in the selective certification setting.

Table 9.23: Reduced OLS parameter estimates for selective certification setting

$statistics$	β_0	β_1	
$mean$	207.7	-12.21	
$median$	207.50	-12.24	
$stand.dev.$	1.991	2.655	
$minimum$	202.8	-20.28	
$maximum$	212.8	-3.957	
$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.67	-12.29	-13.06
$median$	-12.73	-12.33	-13.10
$stand.dev.$	2.825	2.686	2.965
$minimum$	-21.25	-20.45	-22.03
$maximum$	-3.972	-3.960	-3.984
	$E\left[Y' \mid D\right] = \beta_0 + \beta_1 D$		

Propensity score

As in the full certification setting, propensity score, $\Pr\left(D = 1 \mid Z\right)$, is estimated via probit with predictor Z. Propensity score estimates, based on adjusted outcomes and treatment effect adjustments as discussed for *OLS*, of average treatment effects in the selective certification setting are reported in table 9.24. As in the full certification setting, the estimates are more variable than we prefer but, on average, correspond with the sample statistics. Again, homogeneity cannot be rejected but estimated differences in treatment effects do not correspond well with

Table 9.24: Propensity score average treatment effect estimates for selective certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.84	-14.18	-11.47
$median$	-13.09	-13.71	-11.87
$stand.dev.$	5.680	6.862	6.262
$minimum$	-33.93	-49.88	-25.06
$maximum$	-0.213	1.378	13.80

the sample statistics (e.g., estimated *ATT* is the largest in absolute value but *ATT* is the smallest sample statistic as well as estimand).

Propensity score matching

Simulation results, based on outcome and treatment effect adjustments, for propensity score matching estimates of average treatment effects in the selective certification setting are reported in table 9.25. Again, propensity score matching results

Table 9.25: Propensity score matching average treatment effect estimates for selective certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.90	-12.54	-13.27
$median$	-13.20	-12.89	-13.09
$stand.dev.$	3.702	4.478	4.335
$minimum$	-25.87	-25.54	-26.20
$maximum$	-4.622	-2.431	-2.532

correspond well with the sample statistics and are less variable than the propensity score approach above but cannot reject outcome homogeneity.

9.7.4 Outcomes measured by value x only

Now, we revisit selective certification when the analyst cannot observe the incremental cost of certification, k, but only asset value, x. Consequently, outcomes and therefore treatment effects reflect only $Y = x$. For instance, the *DGP* now

yields

$$\begin{aligned}
ATT &= E\left[Y^L - Y^H \mid D=1\right] \\
&= E\left[x^L - x^H \mid D=1\right] \\
&= 201.4 - 214 = -12.6 \\
ATUT &= E\left[Y^L - Y^H \mid D=0\right] \\
&= E\left[x^L - x^H \mid D=0\right] \\
&= 201.4 - 214 = -12.6 \\
ATE &= E\left[Y^L - Y^H\right] \\
&= E\left[x^L - x^H\right] \\
&= \Pr(D=1)\, ATT + \Pr(D=0)\, ATUT \\
&= 201.4 - 214 = -12.6 \\
OLS &= E\left[Y^L \mid D=1\right] - E\left[Y^H \mid D=0\right] \\
&= E\left[x^L \mid D=1\right] - E\left[x^H \mid D=0\right] \\
&= 201.4 - 214 = -12.6
\end{aligned}$$

The apparent advantage to high investment is even more distorted because not only are investment costs excluded but now also the incremental certification costs are excluded. In other words, we have a more limited outcome measure. We briefly summarize treatment effect analyses similar to those reported above but for the alternative, data limited, outcome measure $Y = x$. Notice, no outcome adjustment is applied.

OLS results

Simulation results for the *OLS* model are reported in table 9.26 and sample average treatment effect statistics are reported in table 9.27.

Table 9.26: OLS parameter estimates for Y=x in selective certification setting

$statistics$	β_0	$\beta_1\,(estATE)$
$mean$	214.0	-12.70
$median$	214.1	-12.70
$stand.dev.$	1.594	2.355
$minimum$	209.3	-18.5
$maximum$	218.11	-5.430
$E[Y \mid D] = \beta_0 + \beta_1 D$		

OLS effectively estimates the treatment effects and outcome homogeneity is supported.

Propensity score

Propensity score estimates for average treatment effects are reported in table 9.28.

Table 9.27: Average treatment effect sample statistics for Y = x in selective certification setting

$statistics$	ATE	ATT	$ATUT$
$mean$	-12.73	-12.72	-12.75
$median$	-12.86	-12.78	-12.62
$stand.dev.$	1.735	2.418	2.384
$minimum$	-17.26	-19.02	-18.96
$maximum$	-7.924	-5.563	-6.636

Table 9.28: Propensity score average treatment effect for Y = x in selective certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-13.02	-14.18	-11.86
$median$	-13.49	-13.96	-11.20
$stand.dev.$	5.058	5.764	5.680
$minimum$	-27.00	-34.39	-24.25
$maximum$	2.451	0.263	7.621

Similar to previous propensity score analyses, the limited outcome propensity score results are more variable than we'd like but generally correspond with average treatment effect sample statistics.

Propensity score matching

Propensity score matching simulation results are reported in table 9.29. Propen-

Table 9.29: Propensity score matching average treatment effect for Y = x in selective certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-12.61	-12.43	-12.76
$median$	-12.83	-12.40	-13.10
$stand.dev.$	3.239	3.727	4.090
$minimum$	-20.57	-21.79	-24.24
$maximum$	-4.025	0.558	-1.800

sity score matching results are generally consistent with other results. For $Y = x$, matching effectively identifies average treatment effects, supports homogeneous outcome, and is less variable than the (immediately) above propensity score results.

Since outcome based on x only is more limited than $Y = x - k$, for the remaining discussion of this asset revaluation regulation example we refer to the broader outcome measure $Y = x - k$.

9.7.5 Selective certification with missing "factual" data

It is likely the analyst will not have access to ex post values when the assets are not traded. Then, the only outcome data observed is when assets are certified or when traded at the equilibrium price. In addition to not observing counterfactuals, we now face missing factual data. Missing outcome data produces a challenging treatment effect identification problem. The treatment effects are the same as the above observed data case but require some creative data augmentation to recover. We begin our exploration by examining model-based estimates if we ignore the missing data problem.

If we ignore missing data but adjust outcomes and treatment effects (as discussed earlier) and estimate the model via *OLS* we find the simulation results reported in table 9.30. The average model-estimated treatment effects are biased

Table 9.30: OLS parameter estimates ignoring missing data for selective certification setting

$statistics$	β_0	β_1	
$mean$	207.2	-9.992	
$median$	207.2	-9.811	
$stand.dev.$	2.459	3.255	
$minimum$	200.9	-18.30	
$maximum$	213.2	-2.627	
$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-10.45	-10.07	-10.81
$median$	-9.871	-5.270	-14.92
$stand.dev.$	3.423	3.285	3.561
$minimum$	-19.11	-18.44	-19.75
$maximum$	-2.700	-2.640	-2.762
$E[Y' \mid D] = \beta_0 + \beta_1 D$			

toward zero due to the missing outcome data.

Data augmentation

The above results suggest attending to the missing data. The observed data may not, in general, be representative of the missing factual data. We might attempt to model the missing data process and augment the observed data. Though, data augmentation might introduce more error than do the missing data and consequently generate poorer estimates of the average treatment effects. The observed data are

$$Y_1^o = \Im_{ck}^L \left(x^L - k^L\right) + \Im_p^L P^L$$

and

$$Y_0^o = \Im_{ck}^H \left(x^H - k^H\right) + \Im_p^H P^H$$

where

$$\Im_p^j = \begin{array}{ll} 1 & \text{asset traded at uncertified, equilibrium price for choice } j \\ 0 & \text{otherwise} \end{array}$$

and $\Im_{ck}^j$ refers to assets certified and traded for choice j, as before.

For the region $x < x_c$, we have outcome data for firms forced to sell, $x^j - k^j$, but we are missing untraded asset values, x^j. Based on the *DGP* for our continuing example, the contribution to treatment effects from this missing quantity is $22.289 - 20.253 = 2.036$. If we know k^j or can estimate it, we can model the missing data for this region. Since $I^H > I^L$, $E\left[x^H \mid x^H < x_c\right] > E\left[x^L \mid x^L < x_c\right]$ and $\Pr\left(x^L < x_c\right) > \Pr\left(x^H < x_c\right)$. That is, the adjustment to recover x_j is identified as

$$\frac{(1-\pi)}{\pi}\left(\frac{\sum \Im_{ck,i}^L \left(x_i^L - k_i^L\right)}{\sum D_i} - \frac{\sum \Im_{ck,i}^H \left(x_i^H - k_i^H\right)}{\sum (1 - D_i)}\right)$$
$$+\frac{(1-\pi)}{\pi}\left(\frac{\sum \Im_{ck,i}^L k_i^L}{\sum D_i} - \frac{\sum \Im_{ck,i}^H k_i^H}{\sum (1 - D_i)}\right)$$

The other untraded assets region, $x^j > P^j$, is more delicate as we have no direct evidence, the conditional expectation over this region differs by investment choice, and $P^H > P^L$, it is likely $E\left[x^H \mid x^H > P^H\right] > E\left[x^L \mid x^L > P^L\right]$. Based on the *DGP* for our continuing example, the contribution to treatment effects from this missing quantity is $22.674 - 26.345 = -3.671$.

How do we model missing data in this region? This is not a typical censoring problem as we don't observe the sample size for either missing data region. Missing samples make estimating the probability of each mean level more problematic — recall this is important for estimating average treatment effects in the data observed, selective certification case.[23] Conditional expectations and probabilities of mean levels are almost surely related which implies any augmentation errors will be amplified in the treatment effect estimate.

We cannot infer the probability distribution for x by nonparametric methods since x is unobserved. To see this, recall the equilibrium pricing of uncertified assets satisfies

$$\begin{aligned} P \quad = \quad & \frac{\pi \Pr\left(x_c < x < \overline{x}\right) E\left[x \mid x_c < x < \overline{x}\right]}{\pi \Pr\left(x_c < x < \overline{x}\right) + (1-\pi) \Pr\left(x_c < x < P\right)} \\ & + \frac{(1-\pi) \Pr\left(x_c < x < P\right) E\left[x \mid x_c < x < P\right]}{\pi \Pr\left(x_c < x < \overline{x}\right) + (1-\pi) \Pr\left(x_c < x < P\right)} \end{aligned}$$

For instance, if all the probability mass in these intervals for x is associated with P, then the equilibrium condition is satisfied. But the equilibrium condition is

[23] As is typical, identification and estimation of average treatment effects is more delicate than identification and estimation of model parameters in this selective certification setting.

satisfied for other varieties of distributions for x as well. Hence, the distribution for x cannot be inferred when x is unobserved. If π is known we can estimate $\Pr(x_c < x < \overline{x})$ from certification frequency scaled by π. However, this still leaves much of the missing factual data process unidentified when x is unobserved or the distribution for x is unknown.

On the other hand, consistent probability assignment for x allows π to be inferred from observable data, P and x_c as well as the support for x: $\underline{x} < x < \overline{x}$. Further, consistent probability assignment for x enables us to model the *DGP* for the missing factual data. In particular, based on consistent probability assignment for x we can infer π and identify $\Pr(\underline{x} < x < x_c)$, $E[x \mid \underline{x} < x < x_c]$, $\Pr(P < x < \overline{x})$, and $E[x \mid P < x < \overline{x}]$.

To model missing factual data, suppose π is known and k^j is observed, consistent probability assignment suggests

$$\Pr(P < x < \overline{x}) = \Pr(x_c < x < P)$$

and

$$E[x \mid P < x < \overline{x}] = P + \frac{P - x_c}{2} = \frac{3P - x_c}{2}$$

are reasonable approximations. Then, our model for missing factual data suggests the following adjustments to estimate average treatment effects (*TE*).

$$\begin{aligned} estTE \;=\; & \textit{TE} \text{ estimated based on missing factual data} \\ & + \frac{(1-\pi)}{\pi}\left(\frac{\sum \Im^L_{ck,i}\left(x^L_i - k^L_i\right)}{\sum D_i} - \frac{\sum \Im^H_{ck,i}\left(x^H_i - k^H_i\right)}{\sum (1 - D_i)}\right) \\ & + \frac{(1-\pi)}{\pi}\left(\frac{\sum \Im^L_{ck,i} k^L_i}{\sum D_i} - \frac{\sum \Im^H_{ck,i} k^H_i}{\sum (1 - D_i)}\right) \\ & + \frac{(1-\pi)}{1+\pi}\left[\frac{3P^L - x_c}{2}\frac{\sum \Im^L_{P,i}}{\sum D_i} - \frac{3P^H - x_c}{2}\frac{\sum \Im^H_{P,i}}{\sum (1 - D_i)}\right] \end{aligned}$$

Results adjusted by the augmented factual missing data based on the previous *OLS* parameter estimates are reported in table 9.31. These augmented-*OLS* results

Table 9.31: Treatment effect OLS model estimates based on augmentation of missing data for selective certification setting

$statistics$	$estATE$	$estATT$	$estATUT$
$mean$	-11.80	-11.43	-12.18
$median$	-11.76	-11.36	-12.06
$stand.dev.$	3.165	3.041	3.290
$minimum$	-20.37	-19.58	-21.15
$maximum$	-2.375	-2.467	-2.280
$E[Y' \mid D] = \beta_0 + \beta_1 D$			

are less biased, on average, than results that ignore missing factual data. Thus, it appears data augmentation has modestly aided our analysis of this asset revaluation with selective certification setting.

9.7.6 Sharp regression discontinuity design

Suppose the *DGP* is altered only in that

$$k^L \sim uniform\,(1,3)$$

and

$$k^H \sim uniform\,(3,37)$$

The means for k remain 2 and 20 but we have adjacent support. There is a crisp break at $k = 3$ but the regression function excluding the treatment effect (the regression as a function of k) is continuous. That is, the treatment effect fully accounts for the discontinuity in the regression function. This is a classic "sharp" regression discontinuity design (Trochim [1984] and Angrist and Pischke [2009]) where β_2 estimates the average treatment effect via *OLS*.

$$E\,[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$$

With the previous *DGP*, there was discontinuity as a function of both the regressor k and treatment D. This creates a problem for the regression as least squares is unable to distinguish the treatment effect from the jump in the outcome regression and leads to poor estimation results. In this revised setting, we anticipate substantially improved (finite sample) results.

Full certification setting

Simulation results for the revised *DGP* in the full certification setting are reported in table 9.32 and average treatment effect sample statistics are reported in table

Table 9.32: Sharp RD OLS parameter estimates for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	214.2	-1.007	-12.93
$median$	214.5	-1.019	-13.04
$stand.dev.$	4.198	0.190	4.519
$minimum$	203.4	-1.503	-26.18
$maximum$	226.3	-0.539	-1.959
$E\,[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$			

9.33.

Unlike the previous *DGP*, sharp regression discontinuity (*RD*) design effectively identifies the average treatment effect and *OLS* produces reliable estimates for the (simple) full certification setting. Next, we re-evaluate *RD* with the same adjacent support *DGP* but in the more challenging selective certification setting.

Table 9.33: Average treatment effect sample statistics for full certification setting

$statistics$	ATE	ATT	$ATUT$
$mean$	-12.54	-12.49	-12.59
$median$	-12.55	-12.44	-12.68
$stand.dev.$	1.947	2.579	2.794
$minimum$	-17.62	-19.53	-21.53
$maximum$	-7.718	-6.014	-6.083

Selective certification setting

To satisfy the continuity condition for the regression, suppose cost of certification $k = Dk^L + (1 - D)\, k^H$ is always observed whether assets are certified or not in the regression discontinuity analysis of selective certification. Simulation results for the revised *DGP* in the selective certification setting are reported in table 9.34.[24] In the selective certification setting, *RD* again identifies the average

Table 9.34: Sharp RD OLS parameter estimates for selective certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	214.2	-0.299	-13.00
$median$	214.5	-0.324	-12.89
$stand.dev.$	4.273	0.197	4.546
$minimum$	202.0	-0.788	-25.81
$maximum$	225.5	0.226	-1.886
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$			

treatment effect and *OLS* provides effective estimates. Next, we employ *RD* in the missing factual data setting.

Missing factual data

If some outcome data are unobserved by the analyst, it may be imprudent to ignore the issue. We employ the same missing data model as before and estimate the average treatment effect ignoring missing outcome data (β_2) and the average treatment effect adjusted for missing outcome data ($\beta_2^{'}$). Simulation results for the revised *DGP* (with adjacent support) analyzed via a sharp RD design in the selective certification setting with missing outcome data are reported in table 9.35.

[24] We report results only for the reduced model. If the analyst knows where support changes (i.e., can identify the indicator variables) for the full model, the results are similar and the estimates have greater precision.

Table 9.35: Sharp RD OLS parameter estimates with missing data for selective certification setting

$statistics$	β_0	β_1	β_2	$\beta_2^{'}(estATE)$
$mean$	214.4	-0.342	-11.35	-12.22
$median$	214.5	-0.336	-11.50	-12.47
$stand.dev.$	4.800	0.232	5.408	5.237
$minimum$	201.3	-0.928	-25.92	-26.50
$maximum$	227.9	0.325	2.542	1.383
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$				

9.7.7 Fuzzy regression discontinuity design

Now, suppose the DGP is altered only in that support is overlapping as follows:

$$k^L \sim uniform\,(1,3)$$

and

$$k^H \sim uniform\,(1,39)$$

The means for k remain 2 and 20 but we have overlapping support. There is a crisp break in $E[D \mid k]$ at $k = 3$ but the regression function excluding the treatment effect (the regression as a function of k) is continuous. This leads to a fuzzy discontinuity regression design (van der Klaauw [2002]). Angrist and Lavy [1999] argue that *2SLS-IV* consistently estimates a local average treatment effect in such cases where

$$T = \begin{array}{ll} 1 & k \leq 3 \\ 0 & k > 3 \end{array}$$

serves as an instrument for treatment. In the first stage, we estimate the propensity score[25]

$$\widehat{D} \equiv E[D \mid k, T] = \gamma_0 + \gamma_1 k + \gamma_2 T$$

The second stage is then

$$E[Y \mid k, D] = \gamma_0 + \gamma_1 k + \gamma_2 \widehat{D}$$

Full certification setting

First, we estimate *RD* via *OLS* then we employ *2SLS-IV*. Simulation results for the overlapping support *DGP* in the full certification setting are reported in table 9.36.

Perhaps surprisingly, *OLS* effectively estimates the average treatment effect in this fuzzy *RD* setting. Recall the selection bias is entirely due to the expected difference in certification cost, $E\left[k^H - k^L\right]$. *RD* models outcome as a (regression)

[25] In this asset revaluation setting, the relations are linear. More generally, high order polynomial or nonparametric regressions are employed to accommodate nonlinearities (see Angrist and Pischke [2009]).

Table 9.36: Fuzzy RD OLS parameter estimates for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	214.3	-1.012	-12.79
$median$	214.2	-1.011	-12.56
$stand.dev.$	3.634	0.163	3.769
$minimum$	204.9	-1.415	-23.51
$maximum$	222.5	-0.625	-3.001
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$			

function of k, $E[Y \mid k]$; hence, the selection bias is eliminated from the treatment effect. Next, we use *2SLS-IV* to estimate *LATE*.[26]

Binary instrument

Now, we utilize T as a binary instrument. Simulation results for the overlapping support *DGP* in the full certification setting are reported in table 9.37. As ex-

Table 9.37: Fuzzy RD 2SLS-IV parameter estimates for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estLATE)$
$mean$	214.5	-1.020	-13.07
$median$	214.6	-1.021	-13.27
$stand.dev.$	4.139	0.181	4.456
$minimum$	202.7	-1.461	-27.60
$maximum$	226.0	-0.630	-1.669
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 \widehat{D}$			

pected, *2SLS-IV* effectively identifies *LATE* in this fuzzy *RD*, full certification setting. Next, we revisit selective certification with this overlapping support *DGP*.

9.7.8 Selective certification setting

First, we estimate *RD* via *OLS* then we employ *2SLS-IV*. Simulation results for the overlapping support *DGP* in the selective certification setting are reported in table 9.38. Since *RD* effectively controls the selection bias (as discussed above), *OLS* effectively estimates the average treatment effect.

Binary instrument

Using T as a binary instrument, *2SLS-IV* simulation results for the overlapping support *DGP* in the selective certification setting are reported in table 9.39. In the selective certification setting, *2SLS-IV* effectively estimates *LATE*, as anticipated.

[26] *LATE* is developed more fully in chapter 10.

Table 9.38: Fuzzy RD OLS parameter estimates for selective certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	214.3	-0.315	-12.93
$median$	214.1	-0.311	-12.73
$stand.dev.$	3.896	0.179	3.950
$minimum$	202.5	-0.758	-24.54
$maximum$	223.3	0.078	-3.201
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$			

Table 9.39: Fuzzy RD 2SLS-IV parameter estimates for selective certification setting

$statistics$	β_0	β_1	$\beta_2\,(estLATE)$
$mean$	214.4	-0.321	-13.09
$median$	214.5	-0.317	-13.03
$stand.dev.$	4.438	0.200	4.631
$minimum$	201.1	-0.805	-27.23
$maximum$	225.6	-0.131	1.742
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 \widehat{D}$			

Missing factual data

Continue with the overlapping support *DGP* and employ the same missing data model as before to address unobserved outcomes (by the analyst) when the assets are untraded. First, we report *OLS* simulation results in table 9.40 then we tabulate *2SLS-IV* simulation results where β_2 is the estimated for the local average treatment effect ignoring missing outcome data and $\beta_2^{'}$ is the local average treatment effect adjusted for missing outcome data. This *OLS RD* model for missing

Table 9.40: Fuzzy RD OLS parameter estimates with missing data for selective certification setting

$statistics$	β_0	β_1	β_2	$\beta_2^{'}\,(estATE)$
$mean$	215.9	-0.426	-12.74	-13.60
$median$	216.2	-0.424	-12.63	-13.52
$stand.dev.$	4.765	0.223	4.792	4.612
$minimum$	201.9	-1.132	-24.20	-23.85
$maximum$	226.3	0.117	0.119	-0.817
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 D$				

outcome data does not offer any clear advantages. Rather, the results seem to be slightly better without the missing data adjustments.

2SLS-IV with T as a binary instrument and missing outcome data adjustments are considered next. Simulation results for the overlapping support *DGP* in the

selective certification, missing outcome data setting are reported in table 9.41.

Table 9.41: Fuzzy RD 2SLS-IV parameter estimates with missing data for selective certification setting

$statistics$	β_0	β_1	β_2	$\beta_2^{'}\,(estLATE)$
$mean$	217.7	-0.428	-12.80	-13.67
$median$	214.8	-0.425	-13.12	-14.30
$stand.dev.$	25.50	0.256	5.919	5.773
$minimum$	139.2	-1.147	-25.24	-25.97
$maximum$	293.9	0.212	6.808	6.010
$E[Y \mid k, D] = \beta_0 + \beta_1 k + \beta_2 \widehat{D}$				

Again, modeling the missing outcome data offers no apparent advantage in this fuzzy *RD*, *2SLS-IV* setting. In summary, when we have adjacent or overlapping support, sharp or fuzzy regression discontinuity designs appear to be very effective for controlling selection bias and identifying average treatment effects in this asset revaluation setting.

9.7.9 Common support

Standard identification conditions associated with ignorable treatment (and *IV* approaches as well) except for regression discontinuity designs include common support $0 < \Pr(D = 1 \mid X) < 1$. As indicated earlier, this condition fails in the asset revaluation setting as certification cost type is a perfect predictor of treatment $\Pr(D = 1 \mid T = 1) = 1$ and $\Pr(D = 1 \mid T = 0) = 0$ where $T = 1$ if type is L and zero otherwise. The foregoing discussion has addressed this issue in two ways. First, we employed an ad hoc adjustment of outcome to eliminate selection bias. This may be difficult or impractical to implement. Second, we employed a regression discontinuity design. The second approach may be unsatisfactory as the analyst needs full support access to adjacent or overlapping regressor k whether assets are certified or not.

However, if there is some noise in the relation between certification cost type and treatment (perhaps, due to nonpecuniary cost or benefit), then a third option may be available. We briefly illustrate this third possibility for the full certification setting.

Suppose everything remains as in the original full certification setting except $k^H \sim uniform(1, 19)$ and some owners select treatment (lower investment) when certification cost is high, hence $\Pr(D = 1 \mid \text{type} = H) = 0.1$. This setup implies observed outcome is

$$Y = D[(Y_1 \mid T = 1) + (Y_1 \mid T = 0)] + (1 - D)(Y_0 \mid T = 0)$$

such that

$$E[Y] = 0.5E\left[x^L - k^L\right] + 0.5\left\{0.1E\left[x^L - k^H\right] + 0.9E\left[x^H - k^H\right]\right\}$$

Suppose the analyst ex post observes the actual certification cost type. The common support condition is satisfied as $0 < \Pr(D = 1 \mid T = 0) < 1$ and if outcomes are conditionally mean independent of treatment given T then treatment is ignorable. The intuition is the type variable, T, controls the selection bias and allows D to capture the treatment effect. This involves a delicate balance as T and D must be closely but imperfectly related.

OLS common support results are reported in table 9.42 and simulation results for average treatment effect sample statistics are reported in table 9.43. The esti-

Table 9.42: Fuzzy RD OLS parameter estimates for full certification setting

$statistics$	β_0	β_1	$\beta_2\,(estATE)$
$mean$	196.9	7.667	-5.141
$median$	196.9	7.896	-5.223
$stand.dev.$	1.812	6.516	6.630
$minimum$	191.5	-10.62	-23.54
$maximum$	201.6	25.56	14.25
$E[Y \mid T, D] = \beta_0 + \beta_1 T + \beta_2 D$			

Table 9.43: Average treatment effect sample statistics for full certification setting

$statistics$	ATE	ATT	$ATUT$
$mean$	-5.637	-5.522	-5.782
$median$	-5.792	-5.469	-5.832
$stand.dev.$	1.947	2.361	2.770
$minimum$	-9.930	-12.05	-12.12
$maximum$	0.118	0.182	0.983

mated average treatment effect is slightly attenuated and has high variability that may compromise its finite sample utility. Nevertheless, the results are a dramatic departure and improvement from the results above where the common support condition fails and is ignored.

9.7.10 Summary

Outcomes at our disposal in this asset revaluation setting limit our ability to assess welfare implications for the owners. Nonetheless, the example effectively points to the importance of recognizing differences in data available to the analyst compared with information in the hands of the economic agents whose actions and welfare is the subject of study. To wit, treatment effects in this setting are uniformly negative. This is a product of comparing net gains associated with equilibrium investment levels, but net gains exclude investment cost. The benefits of higher investment when certification costs are low are not sufficient to overcome the cost of investment but this latter feature is not reflected in our outcome

measure. Hence, if care is not exercised in interpreting the results we might draw erroneous conclusions from the data.

9.8 Control function approaches

Our final stop in the world of ignorable treatment involves the use of control functions. Control functions are functions that capture or control selection so effectively as to overcome the otherwise omitted, correlated variable concern created by endogenous selection. Various approaches can be employed. The simplest (strongest for the data) conditions employ conditional mean independence

$$E[Y_1 \mid X, D] = E[Y_1 \mid X]$$

and

$$E[Y_0 \mid X, D] = E[Y_0 \mid X]$$

and no expected individual-specific gain, $E[V_1 \mid X] = E[V_0 \mid X]$. Then,

$$E[Y \mid X, D] = \mu_0 + \alpha D + g_0(X)$$

where $g_0(X) = E[V_0 \mid X]$ is a control function and $\alpha = ATE = ATT = ATUT$.

9.8.1 Linear control functions

If we add the condition $E[V_0 \mid X] = g_0(X) = \eta_0 + h_0(X)\beta_0$ for some vector control function $h_0(X)$, then

$$E[Y \mid X, D] = \mu_0 + \eta_0 + \alpha D + h_0(X)\beta_0$$

That is, when the predicted individual-specific gain given X, $E[V_1 - V_0 \mid X]$, is zero and the control function is linear in its parameters, we can consistently estimate *ATE* via standard (linear) regression.

9.8.2 Control functions with expected individual-specific gain

Suppose we relax the restriction to allow expected individual specific-gain, that is allow $E[V_1 \mid X] \neq E[V_0 \mid X]$, then

$$E[Y \mid X, D] = \mu_0 + \alpha D + g_0(X) + D[g_1(X) - g_0(X)]$$

where $g_0(X) = E[V_0 \mid X]$ and $g_1(X) = E[V_1 \mid X]$ and $ATE = \alpha$ (but not necessarily equal to ATT).

9.8.3 *Linear control functions with expected individual-specific gain*

Continue with the idea that we allow expected individual specific-gain, $E[V_1 \mid X] \neq E[V_0 \mid X]$ and add the condition that the control functions are linear in parameters $E[V_0 \mid X] = g_0(X) = \eta_0 + h_0(X)\beta_0$ and $E[V_1 \mid X] = g_1(X) = \eta_1 + h_1(X)\beta_1$ for some vector control functions $h_0(X)$ and $h_1(X)$. Hence,

$$E[Y \mid X, D] = \phi + \alpha D + X\beta_0 + D(X - E[X])\delta$$

Now, conditional on X the average treatment effect, $ATE(X)$, is a function of X

$$\alpha + (X - E[X])\delta$$

When we average over all X, the second term is integrated out and $ATE = \alpha$. By similar reasoning, the average treatment effect on the treated can be estimated by integrating over the $D = 1$ subsample

$$ATT = \alpha + \left[\sum_{i=1}^{n} D_i\right]^{-1} \left[\sum_{i=1}^{n} D_i \left(X_i - \overline{X}\right)\delta\right]$$

and the average treatment effect on the untreated can be estimated by integrating over the $D = 0$ subsample

$$ATUT = \alpha - \left[\sum_{i=1}^{n} (1 - D_i)\right]^{-1} \left[\sum_{i=1}^{n} D_i \left(X_i - \overline{X}\right)\delta\right]$$

9.9 Summary

The key element for ignorable treatment identification of treatment effects is outcomes are conditionally mean independent of treatment given the regressors. How do we proceed when ignorable treatment (conditional mean independence) fails? A common response is to look for instruments and apply *IV* strategies to identify average treatment effects. Chapter 10 surveys some instrumental variable approaches and applies a subset of IV identification strategies in an accounting setting — report precision regulation.

9.10 Additional reading

Amemiya [1985] and Wooldridge [2002] provide extensive reviews of the econometrics of selection. Wooldridge [2002] discusses estimating average treatment effects in his chapter 18 (and sample selection earlier). Amemiya [1985] discusses qualitative response models in his chapter 9. Recent volumes of the *Handbook of*

Econometrics are filled with economic policy evaluation and treatment effects. Dawid [2000] offers an alternative view on causal inference.

Heckman, Ichimura, Smith, and Todd [1998] utilize experimental (as well as non-experimental) data to evaluate non-experimental methods (matching, differences - in - differences, and inverse-Mills selection models) for program evaluation. Their results indicate selection bias is mitigated, but not eliminated, by non-experimental methods that invoke common support and common weighting. In fact, they decompose conventional bias into (a) differences in the support of the regressors between treated and untreated, (b) differences in the shape of the distributions of regressors for the two groups in the region of common support, and (c) selection bias at common values of the regressors for both groups. Further, they find that matching cannot eliminate selection bias[27] but their data support the index sufficiency condition underlying standard control function models and a conditional version of differences-in-differences. Heckman and Navarro-Lozano [2004] succinctly review differences amongst matching, control function, and instrumental variable (the latter two are discussed in chapter 10 and the various strategies are compared in chapter 12) approaches to identification and estimation of treatment effects. In addition, they identify the bias produced by matching when the analyst's data fail to meet in the minimally sufficient information for ignorable treatment and when and how other approaches may be more robust to data omissions than matching. They also demonstrate that commonly-employed ad hoc "fixes" such as adding information to increase the goodness of fit of the propensity score model (when minimal information conditions are not satisfied) do not, in general, produce lower bias but rather may increase bias associated with matching.

---

[27]Heckman, Ichimura, and Todd [1997] find that matching sometimes increases selection bias, at least for some conditioning variables.