ESTIMATING AVERAGE TREATMENT EFFECTS: REGRESSION DISCONTINUITY DESIGNS

Jeff Wooldridge
Michigan State University
BGSE/IZA Course in Microeconometrics
July 2009

1. Introduction
2. The Sharp RD Design
3. The Fuzzy RD Design
4. Unconfoundedness versus FRD
5. Graphical Analyses

1. Introduction

- Long history dating back to 1960s in Psychology (Thistlewait and Cook, 1960). Key work in econometrics by Van der Klaauw (2001), Hahn, Todd, and Van der Klaauw (2001). Application by Lee (2007).
- Regression discontinuity (RD) designs exploit discontinuities in policy assignment. For example, there might be an age threshold at which one becomes eligible for a buyout plan, or an income threshold at which one becomes eligible for financial aid.

- General idea: Assume that units on different sides of the discontinuity are similar. Their treatment status differs because of the institutional setup, and differences in outcomes can be attributed to the different treatment status.
- We consider the sharp design – where assignment follows a deterministic rule – and the fuzzy design, where the probability of being treated is discontinuous at a known point.

2. The Sharp RD Design

- Usual setup, y_{i0}, y_{i1} as counterfactuals, treatment status w_i. For now, assume a single covariate, x_i, determining treatment (sometimes called the *forcing variable*). In the sharp regression discontinuity (SRD) design case, treatment is determined as

$$w_i = 1[x_i \geq c].$$

- Along with x_i and w_i, observe $y_i = (1 - w_i)y_{i0} + w_i y_{i1}$.
- Define $\mu_g(x) = E(y_g|x)$, $g = 0, 1$.
- Maintain the assumption that $\mu_g(\cdot)$, $g = 0, 1$, are both continuous at c.

- Because w is a deterministic function of x, unconfoundedness necessarily holds. Stated for the means,

$$E(y_g|x,w) = E(y_g|x),\ g = 0,1.$$

- The key is that the overlap assumption fails. By construction, $p(x) = 0$ for all $x < c$ and $p(x) = 1$ for $x \geq c$. Clearly we cannot use propensity score weighting.
- Technically, we can use regression adjustment but we would be relying on extreme forms of extrapolation.

• First, suppose first that the treatment effect is constant, $y_{i1} - y_{i0} = \tau$, so that $y_i = y_{i0} + (y_{i1} - y_{i0})w_i = y_{i0} + \tau w_i$. It also follows that $\mu_1(x) = \mu_0(x) + \tau$ for all x.

• Easy to see that τ is indentified provided $\mu_0(x) = E(y_0|x)$ is continuous at c. Why?

$$E(y|x) = E(y_0|x) + \tau E(w|x) = \mu_0(x) + \tau 1[x \geq c].$$

Write the mean function for the observed variable y as $m(x) \equiv E(y|x)$. Then, if $\mu_0(\cdot)$ is continuous at c

$$m^-(c) \equiv \lim_{x \uparrow c} m(x) = \mu_0(c)$$

$$m^+(c) \equiv \lim_{x \downarrow c} m(x) = \mu_0(c) + \tau$$

- It follows that

$$\tau = m^{+}(c) - m^{-}(c).$$

- We can estimate $E(y|x)$ quite generally at $x = c$, and so τ is identified.
- As an important technical matter, if we want to use nonparametric estimation (in order to avoid extrapolation of parametric functions), then we are estimating two regression functions at a boundary. That is, we estimate $E(y|x)$ for $x \leq c$ and $E(y|x)$ for $x \geq c$ at the boundary value, c.

- Under parametric assumptions, estimation is easy. For example, if

$$E(y_0|x) = \alpha_0 + \beta_0 x$$

then

$$E(y|x) = E(y_0|x) + \tau w = \alpha_0 + \beta_0 x + \tau w$$

So, a standard regression of y on a constant, x, and w consistently estimates τ.

- Even if we use polynomials, or other smooth functions of x, the discontinuity of $w = 1[x \geq c]$ in x identifies τ.
- If we maintain parametric models over the entire range of x, allowing a nonconstant treatment effect is easy.
- Let $m_0(x, \boldsymbol{\delta}_0)$ and $m_1(x, \boldsymbol{\delta}_1)$ be the counterfactual, correctly specified mean functions. Because $E(y|x, w = 0) = m_0(x, \boldsymbol{\delta}_0)$ we can consistently estimate $\boldsymbol{\delta}_0$ using nonlinear least squares or a QMLE for the control sample, $w_i = 0$. Similarly, we estimate $\boldsymbol{\delta}_1$ using the treated subsample.

- Because we have a random sample, τ_{ate} is estimated as before:

$$\hat{\tau}_{ate,reg} = N^{-1}\sum_{i=1}^{N}[m_1(x_i,\hat{\boldsymbol{\delta}}_1) - m_0(x_i,\hat{\boldsymbol{\delta}}_0)].$$

But the extrapolation here is extreme: we obtain $\hat{\boldsymbol{\delta}}_0$ using only data with $x_i < c$ and $\hat{\boldsymbol{\delta}}_1$ using only data with $x_i \geq c$. To obtain, say, $m_1(x_i,\hat{\boldsymbol{\delta}}_1)$ when $w_i = 0$, we are plugging in a value for x that was excluded when obtaining $\hat{\boldsymbol{\delta}}_1$.

- In the linear case with two different mean functions we can write

$$E(y|x,w) = \alpha_0 + \tau w + \beta_0 x + \delta w \cdot (x - \mu_x)$$

- Again, if we believe the linear conditional means hold over all x, the regression

$$y_i \text{ on } 1,\ w_i,\ x_i, w_i(x_i - \bar{x}), i = 1, \ldots, N$$

consistently estimates $\tau = \tau_{ate}$. (Naturally, we replace μ_x with the sample average, $\bar{x}$.)

- Without parametric assumptions we cannot estimate $\mu_0(x)$ for $x \geq c$ and cannot estimate $\mu_1(x)$ for $x < c$. Therefore, in general, τ_{ate} is nonparametrically unidentified (unless we assume a constant treatment effect or parametric functional forms over the range of x.)
- Consequently, the focus in the SRD setting us usually on a very specific parameter, the ATE when $x = c$. Define

$$\tau_c \equiv E(y_1 - y_0 | x = c) = \mu_1(c) - \mu_0(c).$$

- For the constant treatment effect case, of course $\tau_c = \tau$. Generally, we can only estimate the ATE for those at the margin of receiving the treatment.
- Thus, even in the SRD case there are issues of external validity.

- It turns out that τ_c is *nonparametrically identified* by the SRD design.

How? Write $y = (1 - w)y_0 + wy_1 = 1[x < c]y_0 + 1[x \geq c]y_1$, and so

$$E(y|x) = 1[x < c]\mu_0(x) + 1[x \geq c]\mu_1(x)$$

- Then, using continuity of $\mu_0(\cdot)$ and $\mu_1(\cdot)$ at c,

$$m^-(c) \equiv \lim_{x \uparrow c} m(x) = \mu_0(c)$$

$$m^+(c) \equiv \lim_{x \downarrow c} m(x) = \mu_1(c)$$

and so

$$\tau_c = m^+(c) - m^-(c)$$

- Several approaches to estimating $m^-(c)$ and $m^+(c)$ have been proposed. A simple is **local linear regression**. $\mu_{0c} = \mu_0(c)$ and $\mu_{1c} = \mu_1(c)$, and write

$$\begin{aligned} y_0 &= \mu_{0c} + \beta_0(x - c) + u_0 \\ y_1 &= \mu_{1c} + \beta_1(x - c) + u_1 \end{aligned}$$

so that

$$y = \mu_{0c} + \tau_c w + \beta_0(x - c) + \delta w \cdot (x - c) + r,$$

where $r = u_0 + w(u_1 - u_0)$.

- The estimate of τ_c is just the jump in the linear function at $x = c$. We could use the entire data set to run the regression

$$y_i \text{ on } 1,\ w_i,\ (x_i - c), w_i \cdot (x_i - c)$$

and obtain $\hat{\tau}_c$ as the coefficient on w_i. But then it would be global estimation.

- Instead, choose a "small" value $h > 0$ and only use the data satisfying $c - h < x_i < c + h$. This is why it is "local."

- Equivalently, estimate two separate regressions, y_i on $1, (x_i - c)$ for $c - h < x_i < c$ and then y_i on $1, (x_i - c)$ for $c \leq x_i < c + h$, and then $\hat{\tau}_c = \hat{\mu}_{1c} - \hat{\mu}_{0c}$, the difference in the two intercepts.

- Can see how sensitive estimates are to h. Tradeoff between bias and variance: as h decreases, the bias shrinks but the variance increases.
- Can use quadratic or cubic in $(x_i - c)$, too, also interacted with w_i.
- Inference is standard when h is viewed as fixed: use a heteroskedasticity-robust t statistic. (No "sample selection" problem in choosing the sample on the basis of x because we are estimating $E(y|x)$.)

- Imbens and Lemieux (2008, *Journal of Econometrics*) show that if h shrinks to zero quickly enough, the usual inference is still valid.
- Adding regressors is no problem: if the regressors are $\mathbf{r}_i$, just run

$$y_i \text{ on } 1,\ w_i,\ (x_i - c), w_i \cdot (x_i - c),\ \mathbf{r}_i$$

again only using data $c - h < x_i < c + h$.

- Using extra regressors is likely to have more of an impact when h is large; it might help reduce the bias from arising from the detereoration of the linear approximation.
- If $\mathbf{r}_i$ helps explain a lot of the variation in y_i, adding $\mathbf{r}_i$ can shrink the error variance and improve the precision of $\hat{\tau}_c$.

• For response variables with special characteristics, can use local versions of other estimation methods. For example, suppose y_g are count variables. Then use the observations with $c - h < x_i < c$ to estimate a Poisson regression $E(y|x, w = 0) = \exp(\alpha_0 + \beta_0 x)$ and use $c \leq x_i < c + h$ to estimate a Poisson regression $E(y|x, w = 1) = \exp(\alpha_1 + \beta_1 x)$.

• If these regression functions are correctly specified for x near c, $\tau_c = \exp(\alpha_1 + \beta_1 c) - \exp(\alpha_0 + \beta_0 c)$ and so

$$\hat{\tau}_c = \exp(\hat{\alpha}_1 + \hat{\beta}_1 c) - \exp(\hat{\alpha}_0 + \hat{\beta}_0 c).$$

- In the linear regression case, Imbens and Lemieux summarize cross-validation methods for choosing the bandwidth, h. (In principle, one could allow two bandwidths, h_L and h_U and then the data used in local estimation satisfies $c - h_L < x_i < c + h_U$, but of course this complicates the problem.)
- The key is that typical methods of cross validation focus on estimating $E(y|x)$ over the entire range of x, whereas here one is interested in $E(y|x, x < c)$ and $E(y|x, x \geq c)$ for $x = c$.
- Of course, can try different rules to check sensitivity of results for τ_c.

- Imbens and Kalraynaram (2008) explicitly look at minimizing

$$E\{[\hat{\mu}_0(c) - \mu_0(c)]^2 + [\hat{\mu}_1(c) - \mu_1(c)]^2\}$$

a mean squared error for the two regression functions at the jump point. Not the same as the MSE for the actual estimand, which is $E\{([\hat{\mu}_1(c) - \hat{\mu}_0(c)] - [\mu_1(c) - \mu_0(c)])^2\}$

- Optimal bandwidth choice depends on second derivatives of the regression functions at $x = c$, the density of x_i at $x = c$, the conditional variances, and the kernel used in local linear regression. But IK have shown how to make the choice of h data-dependent.

3. The Fuzzy RD Design

- In the FRD case, the probability of treatment changes discontinuously at $x = c$. Define the propensity score as

$$P(w = 1|x) \equiv F(x).$$

- In addition to assuming $\mu_0(\cdot)$ and $\mu_1(\cdot)$ are continuous at c, the key assumption for the FRD design is that $F(\cdot)$ is *discontinuous* at c, so that there is a discrete jump in the probability of treatment at the cutoff.

- To identify τ_c, we assume that $(y_1 - y_0)$ is independent of w, conditional on x. This allows treatment, w, to be correlated with y_0 (after conditioning on x) but not with the unobserved gain from treatment. (Note: The unconfoundedness assumption for estimating ATT is that w is unconfounded with respect to y_0.)
- It is possible to relax $(y_1 - y_0) \perp w \mid x$ and estimate a different parameter, but here consider τ_c, as before.
- Again write $y = y_0 + w(y_1 - y_0)$ and use conditional independence:

$$\begin{aligned} E(y|x) &= E(y_0|x) + E(w|x)E(y_1 - y_0|x) \\ &= \mu_0(x) + E(w|x) \cdot \tau(x). \end{aligned}$$

- As before, take limits from the right and left and use continuity of

$\mu_0(\bullet)$ and $\tau(\bullet)$ at c:

$$m^+(c) = \mu_0(c) + F^+(c)\tau_c$$
$$m^-(c) = \mu_0(c) + F^-(c)\tau_c$$

It follows that, if $F^+(c) \neq F^-(c)$, then

$$\tau_c = \frac{[m^+(c) - m^-(c)]}{[F^+(c) - F^-(c)]}.$$

• So, to identify τ_c in the FRD case, continuity of $\mu_0(\bullet)$ and $\mu_1(\bullet)$ are needed, and conditional independence between $(y_1 - y_0)$ and w are used.

• We can estimate $m^{-}(c)$ and $m^{+}(c)$ by, say, local linear regression. Imbens and Lemieux suggest estimating $F^{-}(c)$ and $F^{+}(c)$ in the same way. In other words, use

$$\hat{\tau}_c = \frac{[\hat{m}^{+}(c) - \hat{m}^{-}(c)]}{[\hat{F}^{+}(c) - \hat{F}^{-}(c)]},$$

where $\hat{m}^{+}(c) = \hat{\alpha}_{1c}$, $\hat{m}^{-}(c) = \hat{\alpha}_{0c}$, $\hat{F}^{+}(c) = \hat{\theta}_{1c}$, and $\hat{F}^{-}(c) = \hat{\theta}_{0c}$ are the intercepts from four local linear regressions. For example, $\hat{\alpha}_{1c}$ is from y_i on $1,\ (x_i - c)$, $c \leq x_i < c + h$ and $\hat{\theta}_{1c}$ is from w_i on $1,\ (x_i - c)$, $c \leq x_i < c + h$.

• Conveniently, HTV (2001) show that the IV estimator of

$$y = \alpha_{0c} + \tau_c w + \beta_0(x - c) + \delta 1[x \geq c] \cdot (x - c) + e$$

using $z \equiv 1[x \geq c]$ as the IV for w produces $\hat{\tau}_c$ as the coefficient on w. (In other words, this is an algebraic equivalance.) One uses the data such that $h - c < x_i < c + h$.

• Obtaining $\hat{\tau}_c$ as a standard IV estimator is helpful in performing inference: if h is fixed or is decreasing "fast enough," can use the usual heteroskedasticity-robust IV standard error.

• An alternative IV estimator is the one we discussed under control function estimation. Under linearity of the conditional means (which we eventually only need to hold locally), we can write

$$y = \mu_{0c} + \tau_c w + \beta_0(x - c) + \delta w \cdot (x - c) + u_0 + w(u_1 - u_0).$$

• As in the previous equation, and unlike in the SRD design, w can be "endogenous" in this equation because it can be correlated with u_0 or $w(u_1 - u_0)$.

• This means w and $w \cdot (x - c)$ are generally endogenous in the equation.

• Nevertheless, if the treatment is unconfounded with respect to the gain $u_1 - u_0$, that is,

$$E(u_1 - u_0|x,w) = E(u_1 - u_0|x) = 0,$$

then $w(u_1 - u_0)$ is uncorrelated with any function of x.

• That means any function of x is exogenous in this equation. Therefore, we can use $z \equiv 1[x \geq c]$ as an IV for w and $1[x \geq c] \cdot (x - c) = z \cdot (x - c)$ as an IV for $w \cdot (x - c)$. The entire IV list is

$$\{1, 1[x \geq c], (x - c), 1[x \geq c] \cdot (x - c)\}$$

or, equivalently, $\{1, 1[x \geq c], x, 1[x \geq c] \cdot x\}$.

- Using either approach, could estimate $F^{+}(c)$ and $F^{-}(c)$ by local logit or probit rather than local linear regression. For example,

$$P(w = 1|x) = \Lambda(\eta_{c0} + \psi_0(x - c)), x < c$$
$$P(w = 1|x) = \Lambda(\eta_{c1} + \psi_1(x - c)), x \geq c$$

and then use

$$\hat{F}^{+}(c) - \hat{F}^{-}(c) = \Lambda(\hat{\eta}_{c1}) - \Lambda(\hat{\eta}_{c0})$$

where $(\hat{\eta}_{c0}, \hat{\psi}_0)$ are from a logit of w_i on 1, $(x_i - c)$ using $h - c < x_i < c$, and similarly for $(\hat{\eta}_{c1}, \hat{\psi}_1)$.

- Alternatively, $\hat{F}(x_i)$ can be used as an IV for w_i, and $\hat{F}(x_i) \cdot (x_i - c)$ as an IV for $w_i(x_i - c)$.
- Estimation of F would necessarily recognize the jump at c. This could be done flexibly using probit or logit, say,

$$F(x) = \Lambda(\pi_1 + \pi_2 1[x \geq c] + \pi_3(x - c) + \pi_4 1[x \geq c](x - c)).$$

Then,

$$F^+(c) = \Lambda(\pi_1 + \pi_2), F^-(c) = \Lambda(\pi_1),$$

so can test $H_0 : \pi_2 = 0$ to see if the jump in treatment probability is really there.

- Now have two choices of bandwidths because need to estimate $E(w|x)$ for $x < c$ and $x \geq c$ (and assume this results in a single bandwidth choice) in addition to $E(y|x)$ for $x < c$ and $x \geq c$. Could just, say, choose one based on $E(y|x)$ and use it for both, or choose them separately using, say, Imbens and Kalraynaram (2008)

4. Unconfoundedness versus FRD

- In the FRD case, overlap can hold (although it might be weak in practice). We can compare regression adjustment to the methods of the previous section.
- Useful to return to the linear formulation:

$$y = \eta_c + \tau_c w + \beta_0(x - c) + \delta w \cdot (x - c) + u_0 + w(u_1 - u_0).$$

Under unconfoundedness, composite error has zero mean conditional on (w, x), and so OLS (or local regression) would consistently estimate τ_c. In fact, if we believe unconfoundedness and the linear functional form, we can use all of the data and average across x_i to estimate τ_{ate}.

- If we assume the less restrictive version of unconfoundedness, that is, $D(y_1 - y_0|w,x) = D(y_1 - y_0|x)$ – but allow u_0 to be correlated with w – then the OLS estimator is inconsistent. For example, if we drop the interaction $w \cdot (x - c)$, the estimate of τ_c using (say) the entire sample is just the difference in means between the treated and control (so information on the threshold is not even used; w_i can be zero of $x_i \geq c$ and w_i can be unity if $w_i < c$).
- But the IV method developed above is consistent:

$$\hat{\tau}_c = \frac{[\hat{m}^+(c) - \hat{m}^-(c)]}{[\hat{F}^+(c) - \hat{F}^-(c)]}.$$

The IV estimator exploits the threshold because the means and treatment probabilities exploit the jumps at $x = c$. Namely, the "+" quantities use data only with $x_i \geq c$ and the "–" quantities use data only with $x_i < c$.:

- Importantly, the IV estimator is consistent for ATE for compliers at $x = c$ *without* unconfoundedness, provided we add a monotonicity assumption. Let $w(a)$ denote treatment status if the cutoff point were a, and think of this as a function of potential cutoff points at least over some interval that includes c. The monotonicity assumption is that $w(\cdot)$ is nonincreasing at $a = c$.

- Suppose that the cutoff is determined by age, so that, initially, those with $age = x \geq c$ are eligible. Now suppose the eligibility age is lowered to $c - 1$. The monotonicity assumption is (a local version of) $w(c - 1) \geq w(c)$, which rules out the possibility that a person would participate if eligible at age c, $w(c) = 1$, but would refuse to participate if the eligibility age were lowered , $w(c - 1) = 0$.
- See Imbens and Lemieux (2008) for derivations and further discussion.

5. Graphical Analyses

• As a supplement to formal estimation – and probably prior to estimation – several graphs can be useful. First, put the forcing variable x into bins and compute the average outcome in each bin. (The bin choices should not smooth across the threshold. So, if the threshold is $c = 5$, choose bins such as $\ldots[4, 4.5), [4.5, 5), [5, 5.6), \ldots.$) Should be able to detect a shift in the mean y at the threshold, and one that is as substantial as mean differences across other bins.

• Can do the same for other covariates that should not be affected by the threshold as a check.

• A histogram of the forcing variable to verify it is not being manipulated around the threshold.

EXAMPLE: **Generated Data**. The data in REGDISC.DTA were generated to follow an FRD design. The forcing variable is x (uniform on $[0, 10]$), the rule that predicts treatment is $z = 1[x \geq 5]$, and w is the actual treatment indicator. The outcome variable is y.

```
. des x z w y

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------
x               float  %9.0g                  forcing variable
z               byte   %9.0g                  =1 if x >= 5
w               byte   %9.0g                  =1 if treated
y               float  %9.0g                  response variable

. tab w z

     =1 if |     =1 if x >= 5
   treated |         0          1 |     Total
-----------+----------------------+----------
         0 |       727        111 |       838
         1 |       273        889 |     1,162
-----------+----------------------+----------
```

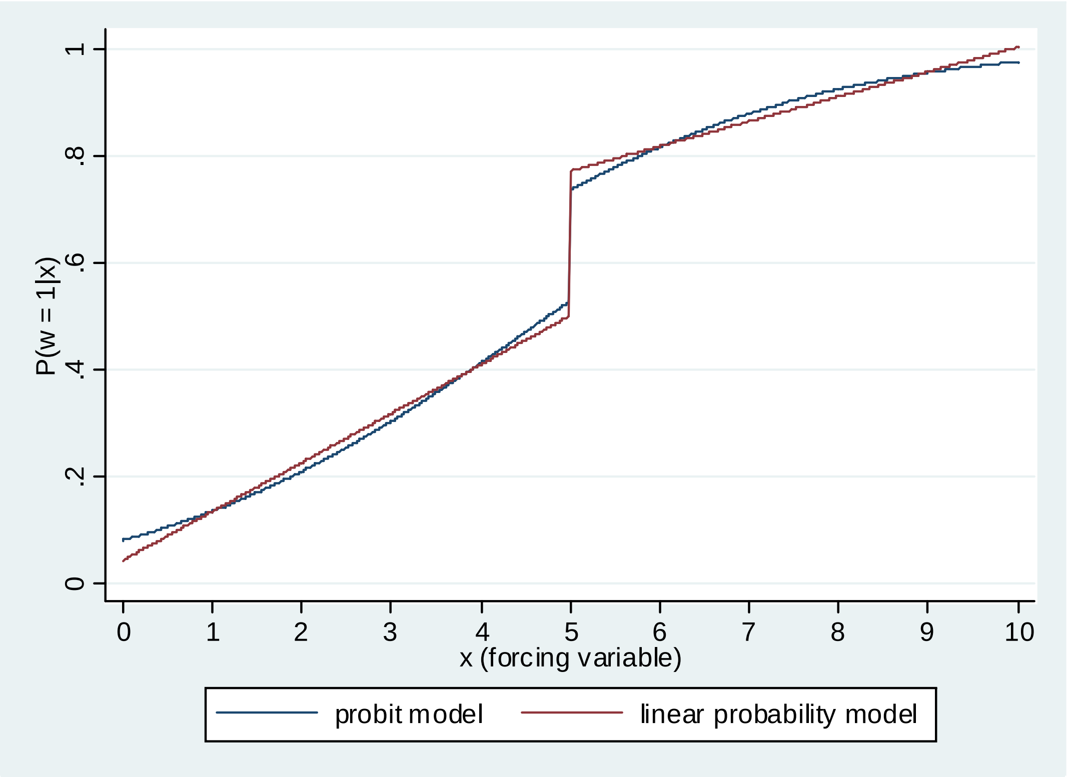
```
     Total |     1,000       1,000 |     2,000

. gen x_5 = x - 5

. gen zx_5 = z*x_5

. gen wx_5 = w*x_5
```

```
. reg w z

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  1,  1998) = 1275.
       Model |     189.728     1     189.728           Prob > F      =  0.0000
    Residual |      297.15  1998  .148723724           R-squared     =  0.3897
-------------+------------------------------           Adj R-squared =  0.3894
       Total |     486.878  1999   .24356078           Root MSE      =  .38565

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           z |       .616   .0172467    35.72   0.000     .5821767    .6498233
       _cons |       .273   .0121952    22.39   0.000     .2490833    .2969167
------------------------------------------------------------------------------

. reg w x if ~z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =   96.
       Model |  17.5177744     1  17.5177744           Prob > F      =  0.0000
    Residual |  180.953226   998  .181315857           R-squared     =  0.0883
-------------+------------------------------           Adj R-squared =  0.0874
       Total |     198.471   999   .19866967           Root MSE      =  .42581

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           x |   .0916522   .0093244     9.83   0.000     .0733545    .1099499
       _cons |    .043984   .0269105     1.63   0.102    -.0088237    .0967917
```

```
-----------------------------------------------------------------------------

. predict what0
(option xb assumed; fitted values)
```

```
. reg w x if z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =   47.
       Model |  4.4914493     1  4.4914493             Prob > F      =  0.0000
    Residual |  94.1875507   998  .094376303           R-squared     =  0.0455
-------------+------------------------------           Adj R-squared =  0.0446
       Total |     98.679   999  .098777778            Root MSE      =  .30721

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           x |   .0464084   .0067272     6.90   0.000     .0332073    .0596095
       _cons |   .5408787   .0513891    10.53   0.000     .4400356    .6417218
------------------------------------------------------------------------------

. predict what1
(option xb assumed; fitted values)

. gen what = what0 if ~z
(1000 missing values generated)

. replace what = what1 if z
(1000 real changes made)

. qui probit w x if ~z

. predict phat0
(option pr assumed; Pr(w))
```

```
. qui probit w x if z

. predict phat1
(option pr assumed; Pr(w))

. gen pshat = phat0 if ~z
(1000 missing values generated)

. replace pshat = phat1 if z
(1000 real changes made)
```



P(w = 1|x)
0
.2
.4
.6
.8
.1
0
1
2
3
4
5
6
7
8
9
10
x (forcing variable)
probit model
linear probability model

```
. * Now estimate the ATE at x = 5:

. ivreg y x zx_5 (w = z), robust

Instrumental variables (2SLS) regression               Number of obs =    2000
                                                       F(  3,  1996) = 3588.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8722
                                                       Root MSE      =   .5959

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           w |   1.963177   .2046892     9.59   0.000      1.56175    2.364604
           x |    .263328   .0295197     8.92   0.000     .2054354    .3212206
        zx_5 |  -.0217891   .0214587    -1.02   0.310    -.0638729    .0202947
       _cons |   .9802505   .0363406    26.97   0.000      .908981     1.05152
------------------------------------------------------------------------------
Instrumented:  w
Instruments:   x zx_5 z
------------------------------------------------------------------------------

. * True value is 2, so 1.96 is very close.
. * Verify this is the same as the ratio of difference
. * in estimated means at the cutoff, 5:
```

```
. reg y x_5 if z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  310.
       Model |  230.759468     1  230.759468           Prob > F      =  0.0000
    Residual |  742.020178   998  .743507193           R-squared     =  0.2372
-------------+------------------------------           Adj R-squared =  0.2365
       Total |  972.779646   999  .973753399           Root MSE      =  .86227

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
         x_5 |   .3326468   .0188819    17.62   0.000      .295594    .3696997
       _cons |   3.814271   .0545347    69.94   0.000     3.707255    3.921287
------------------------------------------------------------------------------

. reg y x_5 if ~z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  368.
       Model |  409.736839     1  409.736839           Prob > F      =  0.0000
    Residual |  1109.52773   998  1.11175123           R-squared     =  0.2697
-------------+------------------------------           Adj R-squared =  0.2690
       Total |  1519.26457   999  1.52078535           Root MSE      =  1.0544

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
         x_5 |   .4432575   .0230891    19.20   0.000     .3979488    .4885663
       _cons |   3.282887   .0666859    49.23   0.000     3.152026    3.413747
```

```
. reg w x_5 if z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =   47.
       Model |  4.4914493     1  4.4914493              Prob > F      =  0.0000
    Residual | 94.1875507   998  .094376303            R-squared     =  0.0455
-------------+------------------------------           Adj R-squared =  0.0446
       Total |     98.679   999  .098777778            Root MSE      =  .30721

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
         x_5 |   .0464084   .0067272     6.90   0.000     .0332073    .0596095
       _cons |   .7729209   .0194295    39.78   0.000     .7347935    .8110482
------------------------------------------------------------------------------

. reg w x_5 if ~z

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =   96.
       Model | 17.5177745     1  17.5177745            Prob > F      =  0.0000
    Residual | 180.953226   998  .181315857            R-squared     =  0.0883
-------------+------------------------------           Adj R-squared =  0.0874
       Total |    198.471   999   .19866967            Root MSE      =  .42581

------------------------------------------------------------------------------
           w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
         x_5 |   .0916522   .0093244     9.83   0.000     .0733545    .1099499
       _cons |   .5022452   .0269307    18.65   0.000     .4493979    .5550926
```

```
------------------------------------------------------------------------------

. di ( 3.814271 - 3.282887)/( .7729209 - .5022452)
1.9631759

. * Same as IV estimate, subject to rounding.
```

```
. * Alternative IV estimate:

. ivreg y x (w wx_5 = z zx_5), robust

Instrumental variables (2SLS) regression               Number of obs =    2000
                                                       F(  3,  1996) = 3591.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8723
                                                       Root MSE      =  .59584

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           w |   1.976194   .1989026     9.94   0.000     1.586116    2.366273
        wx_5 |   -.022558   .0222246    -1.01   0.310    -.0661438    .0210279
           x |   .2635112   .0296645     8.88   0.000     .2053346    .3216877
       _cons |   .9651471   .0472286    20.44   0.000     .8725245     1.05777
------------------------------------------------------------------------------
Instrumented:  w wx_5
Instruments:   x z zx_5
------------------------------------------------------------------------------

. * Very similar, slightly more efficient. (True treatment effect is 2, so even
```

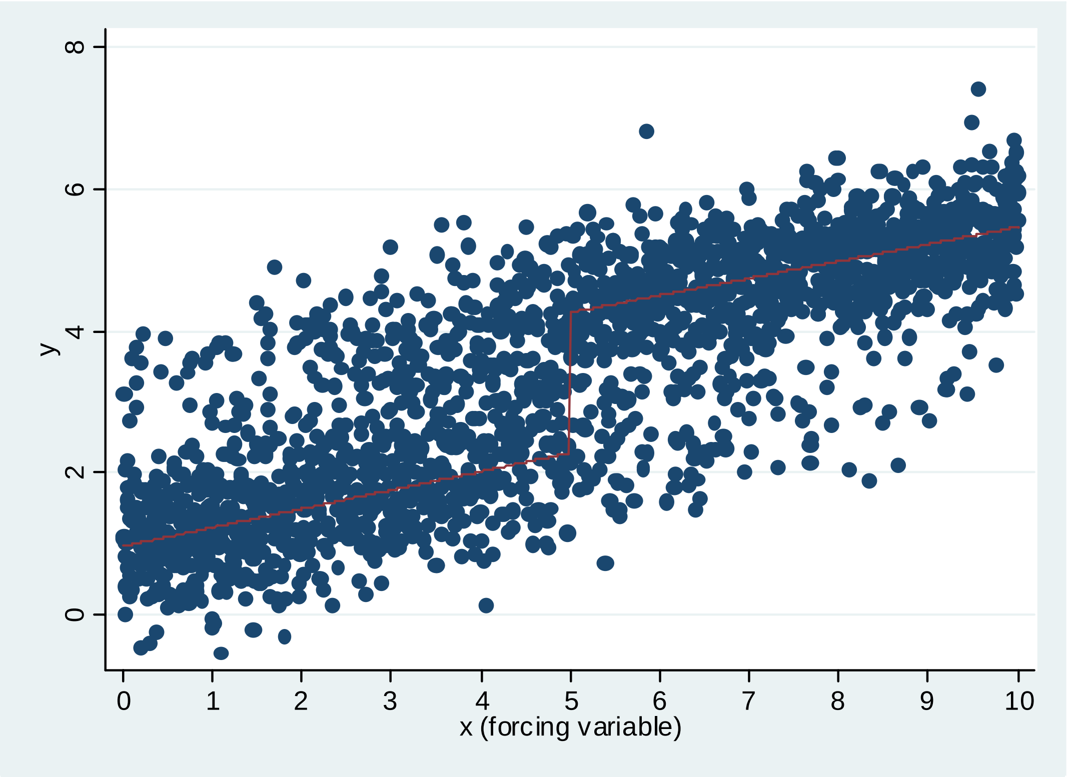
```
. * Now do local versions:

. ivreg y x zx_5 (w = z) if x > 4 & x < 6, robust

Instrumental variables (2SLS) regression               Number of obs =     400
                                                       F(  3,   396) =   62.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6377
                                                       Root MSE      =  .73069

------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           w |   1.241897   .5772794     2.15   0.032     .1069813    2.376812
           x |   .5988192   .1930259     3.10   0.002     .2193356    .9783028
        zx_5 |  -.2123672   .2431103    -0.87   0.383    -.6903155    .2655811
       _cons |  -.1820881   .7057876    -0.26   0.797    -1.569647    1.205471
------------------------------------------------------------------------------
Instrumented:  w
Instruments:   x zx_5 z
------------------------------------------------------------------------------

. ivreg y x (w wx_5 = z zx_5) if x > 4 & x < 6, robust

Instrumental variables (2SLS) regression               Number of obs =     400
                                                       F(  3,   396) =   61.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.6217
                                                       Root MSE      =  .74663
```

```
------------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           w |   1.181874   .5767097     2.05   0.041     .0480781    2.315669
        wx_5 |  -.4146889   .4858873    -0.85   0.394     -1.36993    .5405521
           x |   .7815237     .34224     2.28   0.023     .1086892    1.454358
       _cons |  -1.071853   1.570893    -0.68   0.495    -4.160185    2.016479
------------------------------------------------------------------------------
Instrumented:  w wx_5
Instruments:   x z zx_5
------------------------------------------------------------------------------

. * There are clear costs of dropping 1,600 observations.
```

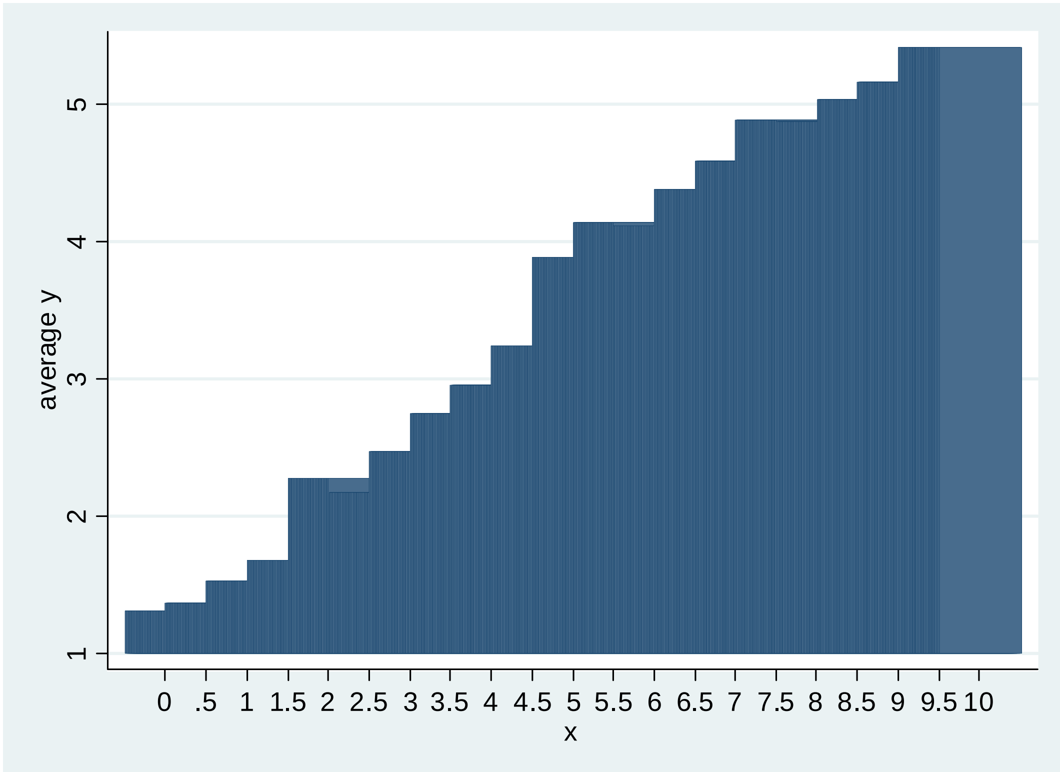
```
. ivreg y x zx_5 (w = z) if x > 3 & x < 7, robust

Instrumental variables (2SLS) regression               Number of obs =     800
                                                       F(  3,   796) =  351.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7662
                                                       Root MSE      =  .61919

----------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+--------------------------------------------------------------
           w |   1.775465   .3267695     5.43   0.000     1.134033    2.416897
           x |   .3471895   .0726118     4.78   0.000     .2046563    .4897226
        zx_5 |  -.0991082   .0772654    -1.28   0.200    -.2507762    .0525599
       _cons |   .7060606   .1912344     3.69   0.000     .3306773    1.081444
----------------------------------------------------------------------------
Instrumented:  w
Instruments:   x zx_5 z
----------------------------------------------------------------------------

. ivreg y x (w wx_5 = z zx_5) if x > 3 & x < 7, robust

Instrumental variables (2SLS) regression               Number of obs =     800
                                                       F(  3,   796) =  338.
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.7601
                                                       Root MSE      =  .62716

----------------------------------------------------------------------------
```

```
             |               Robust
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------+----------------------------------------------------------------
           w |   1.805868   .3227892     5.59   0.000      1.17225    2.439487
        wx_5 |  -.1972305   .1552741    -1.27   0.204    -.5020255    .1075645
           x |   .4119783   .1144717     3.60   0.000     .1872763    .6366803
       _cons |   .3549284   .4503478     0.79   0.431    -.5290812    1.238938
------------------------------------------------------------------------------
Instrumented:  w wx_5
Instruments:   x z zx_5
------------------------------------------------------------------------------

. * Not suprisingly, the estimates for a given data set can be sensitive
. * to the bandwidth.
```

```
. ivreg y x (w wx_5 = z zx_5), robust

Instrumental variables (2SLS) regression                Number of obs =    2000
                                                        F(  3,  1996) = 3591.
                                                        Prob > F      =  0.0000
                                                        R-squared     =  0.8723
                                                        Root MSE      =  .59584

----------------------------------------------------------------------------
             |               Robust
           y |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval
-------------+--------------------------------------------------------------
           w |   1.976194   .1989026     9.94   0.000     1.586116    2.366273
        wx_5 |   -.022558   .0222246    -1.01   0.310    -.0661438    .0210279
           x |   .2635112   .0296645     8.88   0.000     .2053346    .3216877
       _cons |   .9651471   .0472286    20.44   0.000     .8725245     1.05777
----------------------------------------------------------------------------
Instrumented:  w wx_5
Instruments:   x z zx_5
----------------------------------------------------------------------------

. replace muhat = _b[_cons] + _b[w]*z + _b[wx_5]*zx_5 + _b[x]*x
(2000 real changes made)

. twoway (scatter y x, sort) (line muhat x, sort), ytitle(y)
   xtitle(x (forcing variable)) xlabel(#10) legend(off)
```



8
6
4
2
0
y
0
1
2
3
4
5
6
7
8
9
10
x (forcing variable)


average y
1
2
3
4
5
0 .5 1 1.5 2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5 10
x