Heterogeneous Treatment Effects

Christopher Taber

University of Wisconsin

October 30, 2016

So far in this course we have focused on the homogeneous treatment case:

$$Y_i = \alpha T_i + \varepsilon_i$$

In allowing for heterogeneous treatment effects, we focus on the case in which T_i is binary

Let

- Y_{1i} denote the value of Y_i for individual i when $T_i = 1$
- Y_{0i} denote the value of Y_i for individual i when $T_i = 0$

It is useful to define the treatment effect as

$$\alpha_i = Y_{1i} - Y_{0i}$$

Note that in the case we have been thinking about so far

$$\begin{aligned}\alpha_i &= \alpha + \varepsilon_i - \varepsilon_i \\ &= \alpha\end{aligned}$$

and thus we have imposed that it can not vary over the population

This seems pretty unreasonable for almost everything we have thought about in this class

A relatively recent literature has tried to study heterogeneous treatment effects in which these things vary across individuals

A clear problem is that even if we have estimated the full distribution what do we present in the paper?

We must focus on a feature of the distribution

The most common:

- Average Treatment Effect (ATE)

$$E(\alpha_i)$$

- Treatment on the Treated (TT)

$$E(\alpha_i \mid T_i = 1)$$

- Treatment on the Untreated (TUT)

$$E(\alpha_i \mid T_i = 0)$$

(Heckman and Vytlacil discuss Policy Relevant Treatment effects, but I need more notation than I currently have to define those)

These each answer very different questions

I will ignore TUT for the rest of these lecture notes because it is symmetric with TT

All we can directly identify from the data is :

$$E(Y_{1i} \mid T_i = 1), E(Y_{0i} \mid T_i = 0), Pr(T_i = 1)$$

There are two key missing pieces:

$$E(Y_{1i} \mid T_i = 0), E(Y_{0i} \mid T_i = 1)$$

Knowledge of these would be sufficient to identify the two parameters:

$$\begin{aligned}
TT =& E(\alpha_i \mid T_i = 1) \\
=& E(Y_{1i} \mid T_i = 1) - E(Y_{0i} \mid T_i = 1) \\
ATE =& E(\alpha_i) \\
=& [E(Y_{1i} \mid T_i = 1) - E(Y_{0i} \mid T_i = 1)]\, Pr(T_i = 1) \\
& + [E(Y_{1i} \mid T_i = 0) - E(Y_{0i} \mid T_i = 0)]\, [1 - Pr(T_i = 1)]
\end{aligned}$$

How do we estimate these?

Selection only on Observables

I next want to consider the case in which we only have selection only on observables by which I mean:

Assumption 1

For all x in the support of X_i and $t \in \{0, 1\}$,

$$E(Y_{1i} \mid X_i = x, T_i = t) = E(Y_{1i} \mid X_i = x)$$
$$E(Y_{0i} \mid X_i = x, T_i = t) = E(Y_{0i} \mid X_i = x)$$

A “slightly” stronger version of this is random assignment of T_i conditional on X_i

This is often also called **unconfoundedness**

A very strong assumption

Interestingly this is still not enough

If there are values of the observables for which $Pr(T_i = 1 \mid X_i \in \chi) = 1$ or $Pr(T_i = 0 \mid X_i \in \chi) = 0$ then the full distribution of treatment effects is not identified.

For example suppose T_i is being pregnant, we could never hope to identify

$$E(\text{Income} \mid \text{Pregnant, Male})$$

This is perhaps not a relevant counterfactual , but if you want to measure the average treatment effect you can't.

Consider a more interesting case:

- the treatment is free preschool
- the outcome is the kids cognitive test score
- the conditioning variable is family income

In that case the elements of the treatment effect make sense for all income levels:

$$E(Y_i \mid T_i = 1, X_i = x), E\left(Y_i \mid T_i = 0, X_i = x\right)$$

(as opposed to $E(\text{Income} \mid \text{Pregnant, Male})$ which doesn't make sense)

However suppose that the program is means tested so that you are only eligible if your family income is below x^*, then for any value $X_i > x^*$ the effect of the program is not identified

Thus the ATE is not identified without further assumptions

We need additional assumptions

Assumption 2

For almost all x in the support of X_i,

$$Pr(T_i = 0 \mid X_i = x) > 0$$

Assumption 3

For almost all x in the support of X_i,

$$Pr(T_i = 1 \mid X_i = x) > 0$$

Theorem 1

Under assumptions 1 and 2 the TT is identified. Under assumptions 1, 2, and 3 the ATE is identified.

It is pretty clear to see why this holds

Consider the treatment on the treated.

Note that $E(Y_{1i} \mid T_i = 1)$ is identified directly from the data so all we need to get is $E(Y_{0i} \mid T_i = 0)$.

Under the first assumption above

$$E(Y_{0i} \mid T_i = 1) = \sum_j E(Y_{0i} \mid X_i = x_j) Pr(X_i = x_j \mid T_i = 1)$$

As long as assumption 2 holds, $E(Y_{0i} \mid X_i = x)$ is identified so $E(Y_{0i} \mid T_i = 1)$ is identified and thus the TT is identified

Under Assumption 3, you can also get

$$E(Y_{1i} \mid T_i = 0) = \sum_k E(Y_{1i} \mid X_i = x_j) Pr(X_i = x_j \mid T_i = 0)$$

and use this to identify the ATE

Estimation

There are a number of different ways to estimate this model

The most common is to just use OLS defining

$$Y_i = \alpha T_i + X_i' \beta + u_i$$

and run a regression

However this is assuming that the treatment effect is homogeneous

Allowing for heterogeneous treatment effects is straight forward

$$Y_{0i} = X_i'\beta_0 + u_{0i}$$
$$Y_{1i} = X_i'\beta_1 + u_{1i}$$

Then one could estimate

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} X_i'\left(\widehat{\beta}_1 - \widehat{\beta}_0\right)$$

or alternatively:

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} T_i\left[Y_{1i} - X_i'\widehat{\beta}_0\right] + (1 - T_i)\left[X_i'\widehat{\beta}_1 - Y_{0i}\right]$$

TT is analogous (although second method might be more natural)

Matching

Even though regression can be very flexible, many authors argue that matching is better than regression in practice

If you are interested in TT, but the support of X_i conditional on $T_i = 1$ is very different than the unconditional support of X_i than the regression approach can work poorly

Heckman and coauthors made this argument in the context of JTPA where only low income people are eligible for treatment

The idea behind matching can be seen most clearly when X_i has discrete support

Lets focus on the TT case

Let N_0 be the the number of respondents with $T_i = 0$ and let N_1 be the number of respondents with $T_i = 1$

Step 1

Notation is really messy-I don't know of a super clean way to do this

For each observation i with $T_i = 1$ find another observation with exactly the same value of X but for which $T = 0$

You can think of drawing at random from the potential people. Let $I_0(i)$ denote this choice so that for every value of i with $T_i = 1$,

$$X_{I_0(i)} = X_i$$
$$T_{I_0(i)} = 0$$

Example

i	T_i	X_i
1	0	3
2	0	0
3	1	0
4	0	2
5	1	3
6	0	4
7	1	2

Then

$$I_0(3) = 2$$
$$I_0(5) = 1$$
$$I_0(7) = 4$$

Step 2

Estimate the Treatment on the treated using

$$\widehat{TT} = \frac{1}{N_1} \sum_{\{i:T_i=1\}} Y_i - Y_{I_0(i)}$$

To see why this works note that

$$E\left(\widehat{TT}\right) = E\left(Y_{1i} \mid T_i = 1\right) - E\left(Y_{I_0(i)} \mid T_i = 1\right)$$

and

$$\begin{aligned}
E\left(Y_{I_0(i)} \mid T_i = 1\right) =& \sum_{j=1}^{J} E\left(Y_{I_0(i)} \mid T_i = 1, X_i = x_j\right) Pr\left(X_i = x_j \mid T_i = 1\right) \\
=& \sum_{j=1}^{J} E\left(Y_{0\ell} \mid T_\ell = 0, X_\ell = x_j\right) Pr\left(X_i = x_j \mid T_i = 1\right) \\
=& \sum_{j=1}^{J} E\left(Y_{0\ell} \mid X_\ell = x_j\right) Pr\left(X_i = x_j \mid T_i = 1\right) \\
=& \sum_{j=1}^{J} E\left(Y_{0\ell} \mid T_\ell = 1, X_\ell = x_j\right) Pr\left(X_\ell = x_j \mid T_\ell = 1\right) \\
=& E\left(Y_{0i} \mid T_i = 1\right)
\end{aligned}$$

This is difficult to do in practice for two reasons:

1. If X_i is continuous we can't match exactly
2. If X_i is very high dimensional, even with discrete data we probably couldn't match directly because there might be no controls with the same value for every single covariate

Propensity Score Matching

Propensity score matching is a way of getting around the second problem.

Rather than matching on the high dimensional X_i it turns out that we can match on the lower dimensional

$$P(x) \equiv Pr(T_i = 1 \mid X_i = x)$$

The reason why comes from Bayes Theorem

For any x,

$$
\begin{aligned}
& F(x \mid P(X_i) = \rho, T_i = 1) \\
& = Pr(X_i \leq x \mid P(X_i) = \rho, T_i = 1) \\
& = \frac{Pr(T_i = 1 \mid X_i \leq x, P(X_i) = \rho) Pr(X_i \leq x \mid P(X_i) = \rho)}{Pr(T_i = 1 \mid P(X_i) = \rho)} \\
& = \frac{\rho Pr(X_i \leq x \mid P(X_i) = \rho)}{\rho} \\
& = Pr(X_i \leq x \mid P(X_i) = \rho) \\
& = F(x \mid P(X_i) = \rho)
\end{aligned}
$$

and analogously,

$$
\begin{aligned}
& F(x \mid P(X_i) = \rho, T_i = 0) \\
& = Pr(X_i \le x \mid P(X_i) = \rho, T_i = 0) \\
& = \frac{Pr(T_i = 0 \mid X_i \le x, P(X_i) = \rho) Pr(X_i \le x \mid P(X_i) = \rho)}{Pr(T_i = 0 \mid P(X_i) = \rho)} \\
& = \frac{(1 - \rho) Pr(X_i \le x \mid P(X_i) = \rho)}{1 - \rho} \\
& = Pr(X_i \le x \mid P(X_i) = \rho) \\
& = F(x \mid P(X_i) = \rho)
\end{aligned}
$$

thus

$$
F(x \mid P(X_i) = \rho, T_i = 0) = F(x \mid P(X_i) = \rho, T_i = 1)
$$

Thus if we condition on the propensity score, the distribution of X_i is identical for the controls and the treatments.

But since we have selection on observables only:

$$
\begin{aligned}
&E(Y_{0i} \mid T_i = 1, P(X_i) = \rho) \\
&\qquad = \int E(Y_{0i} \mid X_i = x) dF(x \mid T_i = 1, P(X_i) = \rho) \\
&\qquad = \int E(Y_{0i} \mid X_i = x) dF(x \mid T_i = 0, P(X_i) = \rho) \\
&\qquad = E(Y_{0i} \mid T_i = 0, P(X_i) = \rho)
\end{aligned}
$$

Consider matching on propensity scores rather than X_i

We do something similar to before. For each observation i with $T_i = 1$ we find another observation with the same propensity score but $T_i = 0$.

Analogous to before we let $I_0(i)$ denote this choice so that for every value of i with $T_i = 1$,

$$\begin{aligned} p\left(X_{I_0(i)}\right) =& p(X_i) \\ T_{I_0(i)} =& 0 \end{aligned}$$

Then

$$
\begin{aligned}
& E\left(Y_i - Y_{l_0(i)} \mid T_i = 1\right) \\
=& \int E\left(Y_i - Y_{l_0(i)} \mid T_i = 1, P(X_i) = \rho\right) f\left(\rho \mid T_i = 1\right) d\rho \\
=& \int E\left(Y_i \mid T_i = 1, P(X_i) = \rho\right) f\left(\rho \mid T_i = 1\right) d\rho \\
& - \int E\left(Y_{l_0(i)} \mid T_i = 1, P(X_i) = \rho\right) f\left(\rho \mid T_i = 1\right) d\rho \\
=& \int E\left(Y_{1i} \mid T_i = 1, P(X_i) = \rho\right) f\left(\rho \mid T_i = 1\right) d\rho \\
& - \int E\left(Y_{0\ell} \mid T_\ell = 0, P(X_\ell) = \rho\right) f\left(\rho \mid T_\ell = 1\right) d\rho \\
=& E\left(Y_{1i} - Y_{0i} \mid T_i = 1\right)
\end{aligned}
$$

This makes the problem much simpler, but

- You still need to estimate the propensity score which is a high dimensional non-parametric problem. People typically just use a logit
- You still have the first problem above that for a continuous propensity score you are not going to be able to get an exact match.

There are essentially 3 ways to deal with this second problem:

- Just take nearest neighbor (or perhaps caliper which throws out observations without a close neighbor)
- Use all of the observations that are sufficiently close
- Estimate $E(Y_{0j} \mid T_j = 0, P(X_j) = P(X_i))$ directly with some semiparametric method

Lets look at two papers that use this approach

How Robust is the Evidence on the Effects of College Quality? Evidence from Matching

by Dan Black and Jeff Smith, *Journal of Econometrics*, 2004

They want to look at the effects of college quality in the U.S. on wages

They use the National Longitudinal Survey of Youth, 1979

A representative panel data that looks at kids 14-21 in 1979 and is still following them

Table 1: NLSY Descriptive Statistics, 1998

Full sample	Men	Women
age	36.7	36.8
black	0.239	0.280
Hispanic	0.166	0.167
years of education	14.91	14.79
Associate degree	0.116	0.156
Bachelor's degree	0.411	0.363
Master's degree	0.148	0.157
N	1504	1695
Representative sample	**Men**	**Women**
Age	36.7	36.8
Black	0.083	0.106
Hispanic	0.057	0.070
years of education	15.15	14.92
Associate degree	0.101	0.149
Bachelor's degree	0.481	0.413
Master's degree	0.175	0.182
N	1012	1136

They rank colleges using SAT scores, faculty salary and the freshman retention rate

You can see there is substantial selection

Table 2: Variables for Propensity Score and Wage Equations

log wage	Log of average real wage (1982 dollars) on all jobs held during the year
Basic Characteristics:	
region of birth	a vector of 10 dummy variables indicating region in which respondent was born
age	respondent's age at the interview, quadratic in age is used
years of education	highest grade or year of school the respondent completed as of the 1998 interview. Only those who attended a college are in the sample
black	dummy variable indicating the respondent is black
Hispanic	dummy variable indicating the respondent is Hispanic (black & Hispanic are mutually exclusive)
ASVAB test scores	Scores on the ten components of the Armed Services Vocational Aptitude Battery, administered in 1980. We use the first two principal components of the age-adjusted scores.

Home Characteristics:

magazine	"When you were about 14 years old, did you or anyone else living with you get magazines regularly?"
newspaper	"When you were about 14 years old, did you or anyone else living with you get a newspaper regularly?"
library card	"When you were about 14 years old, did you or anyone else living with you have a library card?"
mom education	Highest grade or year of school completed by respondent's mother.
mom living	Was the respondent's mother living at the 1979 interview (when respondents were between 14 and 22 years old)?
mom age	At the 1987 interview.
dad education	Highest grade or year of school completed by respondent's father
dad living	Was the respondent's father living at the 1979 interview?

dad age	At the 1987 interview
living together	Indicator for whether the respondent's mother and father lived in the same household at the 1979 interview
mom occupation	Occupation of job held longest by mother or stepmother in 1978, represented by dummy variables for each Census 1-digit occupation
dad occupation	Occupation of job held longest by father or stepfather in 1978, represented by dummy variables for each Census 1-digit occupation.

High School Characteristics:

size of high school	Asked of respondents' high schools: "As of 10/1/79 [or nearest date] what was [your] total enrollment?"
books	Asked of respondents' high schools: "What is the approximate number of catalogued volumes in the school library (enter 0 if your school has no library)." [in 1979]
teacher salary	Asked of respondents' high schools: "What is the first step on an annual salary contract schedule for a beginning certified teacher with a bachelor's degree?" [in 1979]
disadvantaged	Asked of respondents' high schools: "What percentage of the students in [the respondent's high school] are classified as disadvantaged according to ESEA [or other] guidelines?" [in 1979]

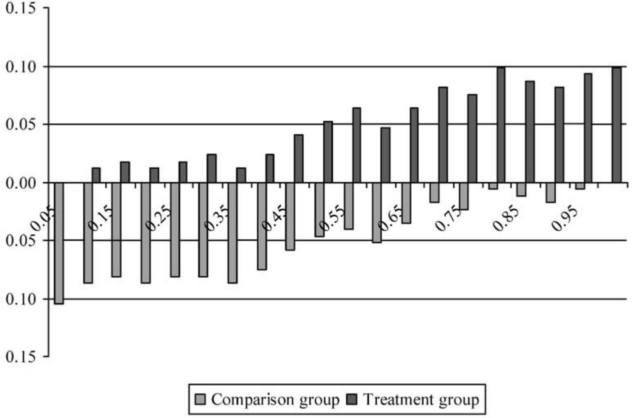
Panel A: Men **Ability quintiles**

Quality index quintiles	First quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile	Total
First quintile	(32.38) [32.38] 6.48	(21.90) [21.90] 4.38	(16.19) [16.19] 3.24	(14.29) [14.29] 2.86	(15.24) [15.24] 3.05	(100.0) (N=105)
Second quintile	(23.81) [23.81] 4.76	(20.95) [20.95] 4.19	(20.95) [20.95] 4.19	(20.95) [20.95] 4.19	(13.33) [13.33] 2.67	(100.0) (N=105)
Third quintile	(24.76) [24.76] 4.95	(15.24) [15.24] 3.05	(21.90) [21.90] 4.38	(17.14) [17.14] 3.43	(20.95) [20.95] 4.19	(100.0) (N=105)
Fourth quintile	(11.54) [11.43] 2.29	(18.27) [18.10] 3.62	(27.88) [27.62] 5.52	(20.19) [20.00] 4.00	(22.12) [21.90] 4.38	(100.0) (N=104)
Fifth quintile	(7.55) [7.62] 1.52	(23.58) [23.81] 4.76	(13.21) [13.33] 2.67	(27.36) [27.62] 5.52	(28.30) [28.57] 5.71	(100.0) (N=106)
Total	[100.0] [N = 105]	[100.0] [N = 105]	[100.0] [N =105]	[100.0] [N = 105]	[100.0] [N = 105]	100.0 N = 525

Panel B: Women

Quality index quintiles	Ability quintiles: First quintile	Second quintile	Third quintile	Fourth quintile	Fifth quintile	Total
First quintile	(31.07) [31.07] 6.21	(19.42) [19.42] 3.88	(20.39) [20.39] 4.08	(15.53) [15.53] 3.11	(13.59) [13.59] 2.72	(100.0) (N=103)
Second quintile	(22.22) [21.36] 4.27	(25.25) [24.27] 4.85	(26.26) [25.24] 5.05	(10.10) [9.71] 1.94	(16.16) [15.53] 3.11	(100.0) (N=99)
Third quintile	(25.71) [26.21] 5.24	(19.05) [19.42] 3.88	(20.95) [21.36] 4.27	(19.05) [19.42] 3.88	(15.24) [15.53] 3.11	(100.0) (N=105)
Fourth quintile	(14.85) [14.56] 2.91	(21.78) [21.36] 4.27	(17.82) [17.48] 3.50	(24.75) [24.27] 4.85	(20.790 [20.39] 4.08	(100.0) (N=101)
Fifth quintile	(6.54) [6.80] 1.36	(14.95) [15.53] 3.11	(14.95) [15.53] 3.11	(29.91) [31.07] 6.21	(33.64) [34.95] 6.99	(100.0) (N=107)
Total	[100.0] [N = 103]	[100.0] [N = 103]	[100.0] [N =103]	[100.0] [N = 103]	[100.0] [N = 103]	100.0 N = 515

A. Men



0.15
0.10
0.05
0.00
0.05
0.10
0.15
0.05
0.15
0.25
0.35
0.45
0.55
0.65
0.75
0.85
0.95
Comparison group
Treatment group

B. Women

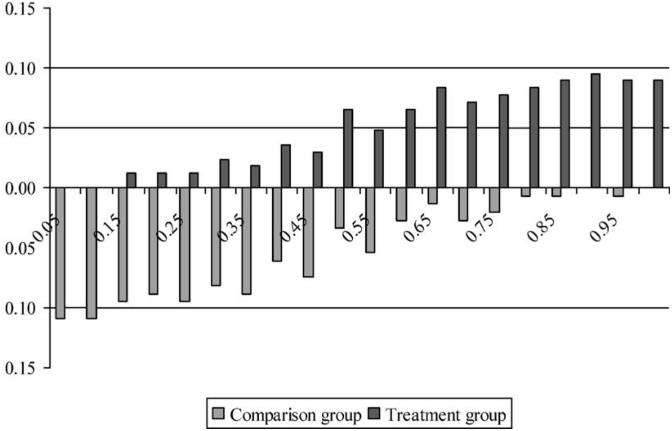


Fig. 1. The distributions of the propensity scores.

They then do propensity score estimation-what they do is somewhat complicated-more than I think is worth getting into here

Table 7
Propensity score estimates of the effects of college quality: fourth and first quartiles, NLSY 1998

$\Delta_{41} = Y_{i4} - Y_{i1}$	Men		Women	
	Using years of education in propensity score estimation	Not using years of education in propensity score estimation	Using years of education in propensity score estimation	Not using years of education in propensity score estimation
Epanechnikov kernel, bandwidth 0.40 for men and 0.30 for women	0.120 (0.0867) [$n = 158$]	0.139 (0.0767) [$n = 152$]	0.067 (0.0862) [$n = 145$]	0.078 (0.0830) [$n = 155$]
OLS estimates	0.122 (0.0584)	0.159 (0.0584)	0.112 (0.0557)	0.155 (0.0552)
Thick support region	0.199 (0.1357) [$n = 44$]	0.250 (0.1181) [$n = 44$]	0.124 (0.1407) [$n = 39$]	0.157 (0.1418) [$n = 39$]
OLS estimates, thick support region	0.121 (0.0639)	0.156 (0.0653)	0.144 (0.0724)	0.184 (0.0720)

Table 8
Propensity score estimates of the effects of college quality, NLSY 1998

	Not using years of education in propensity score estimation	
	Men	Women
$\Delta_{41} = Y_{i4} - Y_{i1}$		
Epanechnikov kernel, bandwidth 0.40 for men and 0.30 for women	0.139	0.078
	(0.0767)	(0.0830)
	[$n = 152$]	[$n = 155$]
OLS estimates	0.159	0.155
	(0.0584)	(0.0552)
$\Delta_{31} = Y_{i3} - Y_{i1}$		
Epanechnikov kernel, bandwidth 0.30 men and 0.50 women	0.056	0.118
	(0.0695)	(0.0561)
	[$n = 166$]	[$n = 133$]
OLS estimates	0.082	0.104
	(0.0541)	(0.0498)
$\Delta_{21} = Y_{i3} - Y_{i1}$		
Epanechnikov kernel, bandwidth 0.20 for men and 0.50 for women	0.006	0.123
	(0.0863)	(0.506)
	[$n = 147$]	[$n = 159$]
OLS estimates	0.072	0.094
	(0.0584)	(0.0458)

Does Piped Water Reduce Diarrhea for Children in Rural India?

by Jalan and Ravallion, *Journal of Econometrics*, 2003

Unsafe drinking water is one of the biggest health risks in the world

This paper studies the effects of piped water on health in rural India using propensity scores

they use the closest five matches as long as they were close enough

Table 1
Access to piped water across the income distribution and by education

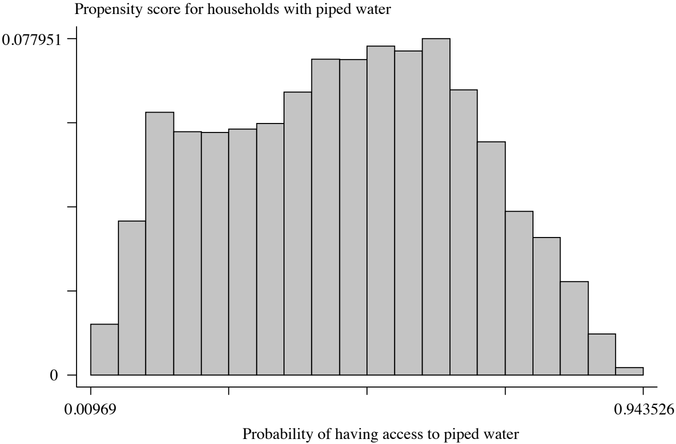
Income quintiles (stratified by household income per person)	Number of observations	Percentage of people with piped water	Households with piped water stratified by highest education of female members				
			Illiterate	At most primary	At most matriculation	Higher secondary or more	Full sample
Bottom 20th percentile	6581	27.18	768	655	251	33	1707
20–40th percentile	6508	25.40	674	590	274	29	1567
40–60th percentile	6543	26.96	667	560	371	60	1658
60–80th percentile	6694	29.62	660	602	462	90	1814
Top 20th percentile	6904	33.63	665	593	638	185	2081
Full sample	33230	28.62	3434	3000	1996	397	8827

Table 2
Logit regression for piped water

	Coefficient	t-statistic
Village variables		
Village size (log)	0.08212	4.269
Proportion of gross cropped area which is irrigated: > 0.75	−0.04824	−1.185
Proportion of gross cropped area which is irrigated: 0.5–0.75	0.19399	4.178
Whether village has a day care center	−0.07249	−2.225
Whether village has a primary school	−0.08136	−1.434
Whether village has a middle school	−0.09019	−2.578
Whether village has a high school	0.26460	7.405
Female to male students in the village	0.10637	3.010
Female to male students for minority groups	−0.07661	−2.111
Main approachable road to village: pucca road	0.19441	3.637
jeepable/kuchha road	−0.00163	−0.033
Whether bus-stoop is within the village	0.11423	2.951
Whether railway station is within the village	0.00920	0.179
Whether there is a post-office within the village	0.02193	0.550
Whether the village has a telephone facility	0.33059	9.655
Whether there is a community TV center in the village	0.09859	2.661
Whether there is a library in the village	−0.04153	−1.116
Whether there is a bank in the village	0.19084	4.655
Whether there is a market in the village	0.31690	6.092
Student teacher ratio in the village	0.00242	5.295
Household variables		
Whether household belongs to the Scheduled Tribe	−0.21288	−4.203
Whether household belongs to the Scheduled Caste	−0.01045	−0.288
Whether it is a Hindu household	−0.24195	−1.709
Whether it is a Muslim household	−0.21631	−1.427
Whether it is a Christian household	0.40367	2.426
Whether it is a Sikh household	−0.86645	−4.531
Household size	0.00337	0.571
Utilization of landholdings: used for cultivation?	0.17109	1.914
Whether the house belongs to the household	−0.18988	−2.854
Whether the household owns other property	0.00181	0.044
Whether the household has a bicycle	−0.26514	−8.243
Whether the household has a sewing machine	0.01183	0.252
Whether the household owns a thresher	−0.05790	−0.577
Whether the household owns a winnower	0.21842	1.820
Whether the household owns a bullock-cart	−0.25900	−5.430
Whether the household owns a radio	0.01036	0.251
Whether the household owns a TV	0.08095	1.335
Whether the household owns a fan	0.01336	0.321
Whether the household owns any livestock	−0.07780	−2.339
Nature of house: Kuchha	−0.10004	−2.775
Pucca	0.12039	2.709
Condition of house: Good	0.00230	0.036
Livable	0.09268	1.756

Table 2 (*continued*)

	Coefficient	t-statistic
Rooms in house: One	−0.10771	−1.371
Two	0.06822	0.952
Three to five	0.07514	1.112
Whether household has a separate kitchen	−0.01993	−0.533
Whether the kitchen is ventilated	0.08103	2.212
Whether the household has electricity	0.40641	11.217
Occupation of the head: Cultivator	−0.02425	−0.481
Agricultural wage labor	0.02432	0.429
Non-agricultural wage labor	0.14628	2.254
Self-employed	-0.06921	−0.955
Whether male members listen to radio	0.20089	3.484
Whether female members listen to radio	−0.12415	−2.177
Whether male members watch TV	0.09365	1.291
Whether female members watch TV	0.03863	0.493
Whether male members read newspapers	0.08950	1.813
Whether female members read newspapers	−0.04066	−0.631
Proportion of household members who are 60+	−0.11370	−1.067
Proportion of females among adults	0.04646	0.331
Proportion of males among children	0.08436	0.779
Proportion of females among children	0.05498	0.498
Whether household head is male	−0.18041	−2.321
Whether household head is single	−0.16659	−1.268
Whether household head is married	−0.02603	−0.422
Whether household head is illiterate	−0.13048	−1.454
Whether household head is primary school educated	−0.03694	−0.416
Whether household head is matriculation educated	−0.03364	−0.385
Whether household head is higher secondary	−0.05545	−0.475
Gross cropped area	−0.00020	−0.666
Gross irrigated area	−0.00050	−1.342
Landholding size: Landless	−0.32849	−3.996
Marginal	−0.31056	−3.987
Small	−0.22129	−2.916
Constant	−1.49531	−5.396
Log-likelihood function	−16236.565	
Number of observations	33216	


Propensity score for households with piped water
0.077951
0
0.00969
0.943526
Probability of having access to piped water

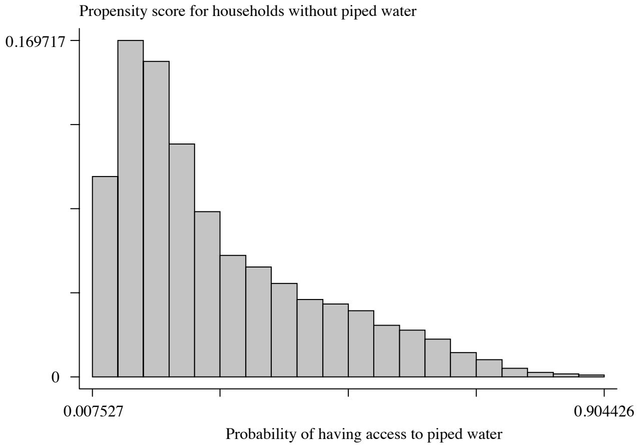

Propensity score for households without piped water
0.169717
0
0.007527
0.904426
Probability of having access to piped water

Table 3
Impacts of piped water on diarrhea prevalence and duration for children under five

	Prevalence of diarrhea		Duration of illness	
	Mean for those with piped water (st. dev.)	Impact of piped water (st. error)	Mean for those with piped water (st. dev.)	Impact of piped water (st. error)
Full sample	0.0108	−0.0023*	0.3254	−0.0957*
	(0.046)	(0.001)	(1.650)	(0.021)
Stratified by household income per capita (quintiles)				
1 (poorest)	0.0155	0.0032*	0.4805	0.0713
	(0.055)	(0.001)	(2.030)	(0.053)
2	0.0136	0.0007	0.4170	0.0312
	(0.051)	(0.001)	(1.805)	(0.051)
3	0.0083	−0.0039*	0.2636	−0.1258*
	(0.038)	(0.001)	(1.418)	(0.042)
4	0.0100	−0.0036*	0.3195	−0.1392*
	(0.044)	(0.001)	(1.703)	(0.048)
5	0.0076	−0.0068*	0.1848	−0.2682*
	(0.042)	(0.001)	(1.254)	(0.036)
Stratified by highest education level of a female member				
Illiterate	0.0131	−0.0000	0.3588	−0.0904*
	(0.053)	(0.001)	(1.710)	(0.036)
At most primary school educated	0.0112	−0.0015	0.3502	−0.0465
	(0.045)	(0.001)	1.739)	(0.036)
At most matriculation educated	0.0074	−0.0065*	0.2573	−0.1708*
	(0.038)	(0.001)	(1.476)	(0.039)
Higher secondary or more	0.0050	−0.0080*	0.1880	−0.2077*
	(0.027)	(0.002)	(1.158)	(0.076)

Table 4
Child-health impacts of piped water by income and education

	Illiterate		At most primary		At most matriculation		Higher secondary or more	
	Prevalence of diarrhea	Duration of illness	Prevalence of diarrhea	Duration of illness	Prevalence of diarrhea	Duration of illness	Prevalence of diarrhea	Duration of illness
1 (poorest quintile)	0.0100*	0.1028	0.0010	0.0548	−0.0118*	−0.1091	Small Sample	
	(0.002)	(0.089)	(0.002)	(0.094)	(0.003)	(0.132)		
2	0.0057*	0.0777	0.0013	0.1061	−0.0121*	−0.2580*	Small Sample	
	(0.003)	(0.083)	(0.002)	(0.083)	(0.002)	(0.087)		
3	−0.0038*	−0.1503*	−0.0008	0.0056	−0.0069*	−0.1659*	Small Sample	
	(0.002)	(0.069)	(0.002)	(0.081)	(0.002)	(0.059)		
4	−0.0062*	−0.2224*	−0.0041*	−0.1691	0.0008	−0.0186	Small Sample	
	(0.002)	(0.097)	(0.002)	(0.070)	(0.003)	(0.091)		
5	−0.0075*	−0.2932*	−0.0051*	−0.2435*	−0.0063*	−0.2578*	−0.010*	−0.2637*
	(0.000)	(0.045)	(0.002)	(0.075)	(0.002)	(0.008)	(0.003)	(0.085)

Note: Figures in parentheses are the respective standard errors.
*Indicates significance at 5% or lower.

Propensity Score Matching vs Regression

When I think about this too hard I start to get a bit confused about the fundamental difference.

At some level when we do matching we do

$$\widehat{TT} = \frac{1}{N_1} \sum_{\{i:T_i=1\}} Y_i - \widehat{Y}_{0i}$$

where $\widehat{Y}_{0i}$ is an unbiased estimate of $E(Y_{0j} \mid X_j = X_i)$

We can get this estimate by taking one person with the same value of the propensity score or by using the forecast from OLS as above: $X_i'\widehat{\beta}_0$

We can then think about nonparametric regression for our estimate of $\widehat{Y}_{0i}$, but this is kind of a more flexible version of both

Reweighting

Another approach is reweighting

Let $f_t(x)$ be the density of X_i conditional on $T_i = t$.

Using Bayes theorem

$$f_1(x) = \frac{P(x)f(x)}{Pr(T_i = 1)}$$
$$f_0(x) = \frac{(1 - P(x))\,f(x)}{Pr(T_i = 0)}$$

so

$$\begin{aligned} E(Y_{0i} \mid T_i = 1) &= \int E(Y_{0i} \mid X_i = x) f_1(x) dx \\ &= \int E(Y_{0i} \mid X_i = x) \frac{f_1(x)}{f_0(x)} f_0(x) dx \\ &= E\left(Y_{0i} \frac{P(X_i)}{1 - P(X_i)} \mid T_i = 0\right) \frac{Pr(T_i = 0)}{Pr(T_i = 1)} \end{aligned}$$

Putting this together we can use the estimator

$$\begin{aligned} & \frac{\sum_{i=1}^{N_1} Y_{1i}}{N_1} - \frac{\sum_{j=1}^{N_0} Y_{0j} \frac{P(X_j)}{1-P(X_j)}}{N_1} \\ =& \frac{\sum_{i=1}^{N_1} Y_{1i}}{N_1} - \frac{\frac{1}{N_0} \sum_{j=1}^{N_0} Y_{0j} \frac{P(X_j)}{1-P(X_j)}}{\frac{N_1}{N_0}} \\ \approx& E(Y_{1i} \mid T_i = 1) - \frac{E(Y_{0i} \mid T_i = 1) \frac{Pr(T_i=1)}{Pr(T_i=0)}}{\frac{Pr(T_i=1)}{Pr(T_i=0)}} \\ =& TT \end{aligned}$$

Instrumental Variables

What about selection on unobservables?

Lets first think about what IV does in this case

Define

$$\begin{aligned} Y_i &\equiv T_i Y_{1i} + (1 - T_i)\, Y_{0i} \\ &= T_i\,(Y_{1i} - Y_{0i}) + Y_{0i} \\ &= \beta_0 + \alpha_i T_i + \varepsilon_i \end{aligned}$$

(where $\beta_0 = E(Y_{0i})$ and $\varepsilon_i = Y_i - \beta_0$)

Assume that we have an instrument Z_i that is correlated with T_i but not with α_i or ε_i (or equivalently Y_{0i} or Y_{1i})

Does IV estimate the ATE?

Lets abstract from other regressors

IV yields

$$\begin{aligned}
\text{plim}\widehat{\beta}_1 &= \frac{Cov(Z_i, Y_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \varepsilon_i + \alpha_i T_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \varepsilon_i)}{Cov(Z_i, T_i)} + \frac{Cov(Z_i, \alpha_i T_i)}{Cov(Z_i, T_i)} \\
&= \frac{Cov(Z_i, \alpha_i T_i)}{Cov(Z_i, T_i)}.
\end{aligned}$$

In the case in which treatment effects are constant so that $\alpha_i = \alpha$ for everyone

$$\text{plim}\widehat{\beta}_1 = \frac{Cov(Z_i, \alpha T_i)}{Cov(Z_i, T_i)} = \alpha$$

However, more generally IV does not converge to the Average treatment effect

Local Average Treatment Effects

Imbens and Angrist (1994) consider the case in which there are not constant treatment effects

The consider a simple version of the model in which Z_i takes on 2 values, call them 0 and 1 for simplicity and without loss of generality assume that
$Pr(T_i = 1 \mid Z_i = 1) > Pr(T_i = 1 \mid Z_i = 0)$

There are 4 different types of people those for whom $T_i = 1$ when:

1. $Z_i = 1, Z_i = 0$
2. never
3. $Z_i = 1$ only
4. $Z_i = 0$ only

Imbens and Angrist's monotonicity rules out 4 as a possibility

Let $\mu_1, \mu_2,$ and μ_3 represent the sample proportions of the three groups

and G_i an indicator of the group

Note that

$$\widehat{\beta}_1 \overset{p}{\to} \frac{Cov(Z_i, \alpha_i T_i)}{Cov(Z_i, T_i)}$$
$$= \frac{E(\alpha_i T_i Z_i) - E(\alpha_i T_i) E(Z_i)}{E(T_i Z_i) - E(T_i) E(Z_i)}$$

Let ρ denote the probability that $Z_i = 1$. Lets look at the pieces first the numerator

$$
\begin{aligned}
& E(\alpha_i T_i Z_i) - E\left(\alpha_i T_i\right) E\left(Z_i\right) \\
=& \rho E(\alpha_i T_i \mid Z_i = 1) - E\left(\alpha_i T_i\right) \rho \\
=& \rho E(\alpha_i T_i \mid Z_i = 1) \\
& - \left[\rho E(\alpha_i T_i \mid Z_i = 1) + (1-\rho)\, E(\alpha_i T_i \mid Z_i = 0)\right] \rho \\
=& \rho(1-\rho)\left[E(\alpha_i T_i \mid Z_i = 1) - E(\alpha_i T_i \mid Z_i = 0)\right] \\
=& \rho(1-\rho)\left[E(\alpha_i \mid G_i = 1)\mu_1 + E(\alpha_i \mid G_i = 3)\mu_3 - E(\alpha_i \mid G_i = 1)\mu_1\right] \\
=& \rho(1-\rho) E(\alpha_i \mid G_i = 3)\mu_3
\end{aligned}
$$

Next consider the denominator

$$
\begin{aligned}
& E(T_i Z_i) - E(T_i) E(Z_i) \\
=& \rho E(T_i \mid Z_i = 1) - E(T_i) \rho \\
=& \rho E(T_i \mid Z_i = 1) \\
& - [\rho E(T_i \mid Z_i = 1) + (1 - \rho) E(T_i \mid Z_i = 0)] \rho \\
=& \rho(1 - \rho) [E(T_i \mid Z_i = 1) - E(T_i \mid Z_i = 0)] \\
=& \rho(1 - \rho) [\mu_1 + \mu_3 - \mu_1] \\
=& \rho(1 - \rho)\mu_3
\end{aligned}
$$

Thus

$$\begin{aligned}\widehat{\beta}_1 \xrightarrow{p} & \frac{\rho(1-\rho)E(\alpha_i \mid G_i = 3)\mu_3}{\rho(1-\rho)\mu_3} \\ = & E(\alpha_i \mid G_i = 3)\end{aligned}$$

They call this the local average treatment effect