UNIVERSITY CERTIFICATE IN ECONOMETRICS

2017 · Edition 2

Course : Microeconometrics for policy evaluation

Part 2 – Policy evaluation/treatment analysis

Vincent Vandenberghe

Outline

1. Introduction
2. The endogeneity problem
3. Experimental methods
4. Quasi-experimental methods

- Short panel analysis (day 1)
- Policy evaluation/treatment analysis (day 2)
  - Difference-in-Difference estimation:
    - Treatment effects with non-experimental data using (repeated) cross-sectional data
    - The canonical DiD estimator
    - Assessing the common trend assumption
    - Relaxing the common trend assumption: from DiD
  - Propensity score matching
    - Matching over a single (propensity) score
    - Matching algorithms (nearest neighborhood, caliper, Kernel)
  - Combining DiD and propensity score matching

Outline

Main Stata commands

1. reg
2. psmatch2
3. teffects
4. diff

Useful references

Cameron, A. C. & Trivedi, P. K. (2010). *Microeconometrics using Stata*. College Station: Stata Press.

Policy_eval\Folder _

MATERIAL @ YOUR DISPOSAL

MOODLE@UCL: LECME2FC

TOPIC 2\ Policy_eval\

ECcourse2.ppt

Code…\Stata_code\

#2EC_PSMATCH.do

#2EC_Extra.do

#2EC_Ex_corr.do

#2EC_Ex.do

#2EC_DiD.do

#2EC_data.do _____

(+ corrected version at the end)

Data.zip

the various data sets @ your disposal

via the web: https://perso.uclouvain.be/vincent.vandenberghe/Stata_EC/Stata_EC1.html

4. Quasi-experimental Methods: Policy Evaluation

4.1. Difference-in-Differences (or two-way fixed effects)

Let's think about a simple evaluation of a policy (treatment)

If we have data on a bunch of people right *before* the policy is enacted and on the same group of people *after* it is enacted. How can we try to identify the effect

Suppose we have two years (the shortest possible panel) of data *t=0* and *t=1* and that the policy/treatment is enacted in between

We could try to identify the effect by simply looking at *y* before and after the policy. That is we can identify the effect by resorting to first differences

[Eq 6] $y_t = \beta\, TR_t + Z + u_t$

where TR_t= 1 if *t=1* and 0 if *t=0*, *Z* is the unobserved fixed effect caracterizing the treated, potentially correlated with treatment

One way of dealing with *Z* is to resort to first differences

[Eq 7] $E(y1)\text{-}E(y0)= \beta$

The problem with this "difference model" is that it attributes any changes in time to the policy

Suppose something else happened between *t=0* and *t=1* other than just the program (eg. an economic recession/boom)

We will attribute whatever that is to the program/treatment

How to solve the problem?

Simply adding a time dummy (*t*) would not help us separate the time effect from the treatment effect (ie. perfect collinearity=> $TR_t=t$)

Rather suppose we have two groups:

- People who are affected by the policy changes (*D*=1)
- People who are not affected by the policy change (*D=0*) => the controls

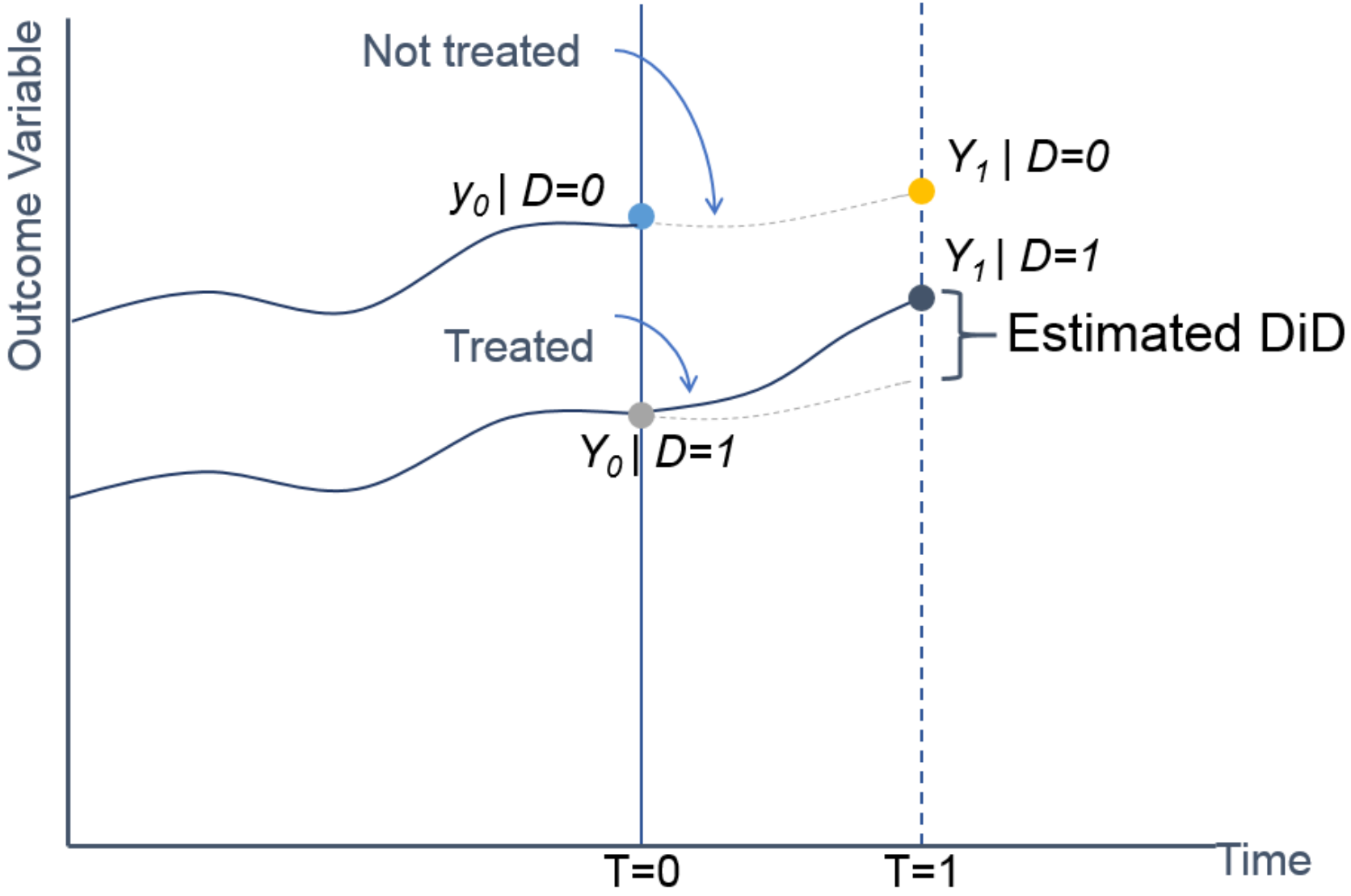
We can think of using the control to pick up the time changes:

[Eq. 8] $E(y_1|D=0) - E(y_0|D=0)$

Then we can estimate our policy effect as a "difference in (time-driven) differences" (DiD) between the treated and the control group

[Eq. 9] $[E(y_1|D=1) - E(y_0|D=1)] - [E(y_1|D=0) - E(y_0|D=0)]$

Graphically



Algebraically

Remember : *D=0* if not treated and *D=1* if treated (with $Z^{D=1} = \alpha + \delta$; $Z^{D=0} = \alpha$)

[Eq. 10] $y_t = \alpha + \beta D*t + \lambda t + \delta D + u_t$

$[E(y_1|D=1) - E(y_0|D=1)] - [E(y_1|D=0) - E(y_0|D=0)]$

$= [\alpha + \beta + \lambda + \delta - \alpha - \delta] -$

$[\alpha + \lambda - \alpha]$

$= \beta$

So estimating β by *OLS* in the above equation will deliver the treatment effect as DiD

Key assumption : the control group (*D=0*) identifies the time path (ie. growth) of outcomes – the trend - that would have happened to the treated (*D=1*) in the absence of the treatment

⇔ control & treated outcome are assumed to be characterized by the *same* trend (common trend assumption) *ex ante*

⇔ Identification of the impact of treatment relies on the change of trend specific to the treated group

⇔ In practice, the common-trend assumption prior to treatment requires a certain element of geographical and political proximity

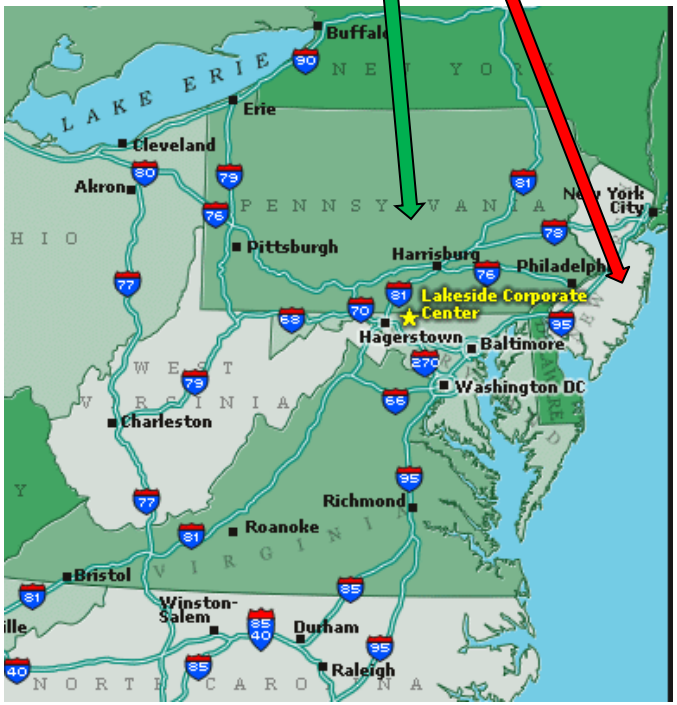
⇔ It can be tested if panels contains several periods of observation prior to treatment

=> #2EC_DiD.do/Case 1

Case 1: Card & Krueger on minimum wages & employment

On April 1, 1992 New Jersey's minimum wage increased from $4.25 to $5.05 per hour (vs Pennsylvania where it stayed at $4.25)

Good or bad for employment in the fast-food sector?



```
Contains data from CD94.dta
  obs:           820                          Dataset from Card&Krueger (1994)
 vars:             8                          27 May 2011 20:36
 size:        12,300
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
id              int     %8.0g                 Store ID
t               byte    %8.0g                 Feb. 1992 = 0; Nov. 1992 = 1
treated         long    %8.0g      treated    New Jersey = 1; Pennsylvania = 0
fte             float   %9.0g                 Output: Full Time Employment
bk              byte    %8.0g                 Burger King == 1
kfc             byte    %8.0g                 Kentuky Fried Chiken == 1
roys            byte    %8.0g                 Roy Rogers == 1
wendys          byte    %8.0g                 Wendy's == 1
-------------------------------------------------------------------------------
```

```
. list id fte t treated z bk kfc roys  wendys if _n<10
```

	id	fte	t	treated	z	bk	kfc	roys	wendys
1.	1	31	0	NJ	0	1	0	0	0
2.	1	40	1	NJ	1	1	0	0	0
3.	2	13	0	NJ	0	1	0	0	0
4.	2	12.5	1	NJ	1	1	0	0	0
5.	3	12.5	0	NJ	0	0	1	0	0
6.	3	7.5	1	NJ	1	0	1	0	0
7.	4	16	0	NJ	0	0	0	1	0
8.	4	20	1	NJ	1	0	0	1	0
9.	5	20	0	NJ	0	0	0	1	0

```
. reg fte t treated itr
```

Source	SS	df	MS
Model	524.003099	3	174.6677
Residual	64600.6458	797	81.0547626
Total	65124.6489	800	81.4058111

Number of obs	801
F(3, 797)	2.15
Prob > F	0.0919
R-squared	0.0080
Adj R-squared	0.0043
Root MSE	9.003

fte	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t	-2.40651	1.446314	-1.66	0.097	-5.245544	.4325237
treated	-2.883534	1.134812	-2.54	0.011	-5.111107	-.6559608
itr	2.913982	1.610513	1.81	0.071	-.2473667	6.075331
_cons	19.94872	1.019394	19.57	0.000	17.9477	21.94973

➔ *#2EC_DID.do/Case 2*

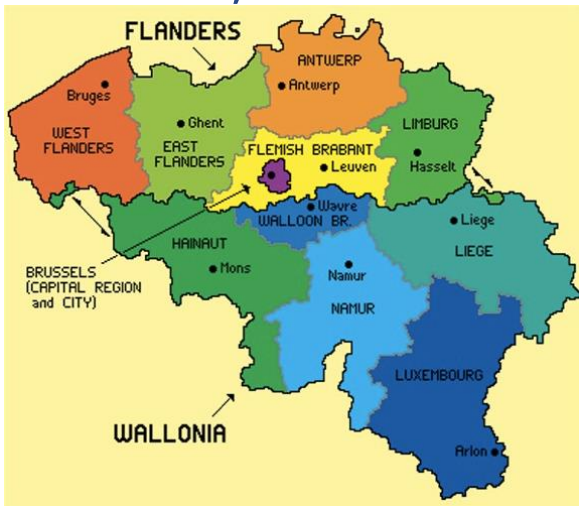
Case 2: VVDB – Hainaut – Objective 1

As part of EU's regional policy, **Objective 1** aims at helping regions lagging behind (<75% of EU average GDP/head)

Treatment/Policy : 1994-1999, Hainaut receives 2.4 billions EURO [5% of the province's GDP for each of the year ranging from 1994 to 1999]

Evidence of impact on income ?

A DID analysis using taxable income in each municipality before and after 1994-1999, and the rest of Wallonia as control



```
        *Without population weights
. reg meanc t treat itr   if inlist(year, 1993, 2000) //Control=rest of Belgium
```

Source	SS	df	MS	Number of obs	=	1,178
				F(3, 1174)	=	342.64
Model	2.2546e+09	3	751529473	Prob > F	=	0.0000
Residual	2.5750e+09	1,174	2193368.96	R-squared	=	0.4668
				Adj R-squared	=	0.4655
Total	4.8296e+09	1,177	4103316.55	Root MSE	=	1481

meanc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t	2723.035	91.84789	29.65	0.000	2542.83	2903.239
treat	-866.2954	189.7524	-4.57	0.000	-1238.587	-494.0037
itr	-341.2793	268.3504	-1.27	0.204	-867.7793	185.2206
_cons	8314.756	64.94626	128.03	0.000	8187.332	8442.179

```
.
.                *With population weights
. reg meanc t treat itr   [fweight=pop] if inlist(year, 1993, 2000) //with municipality population as weight
```

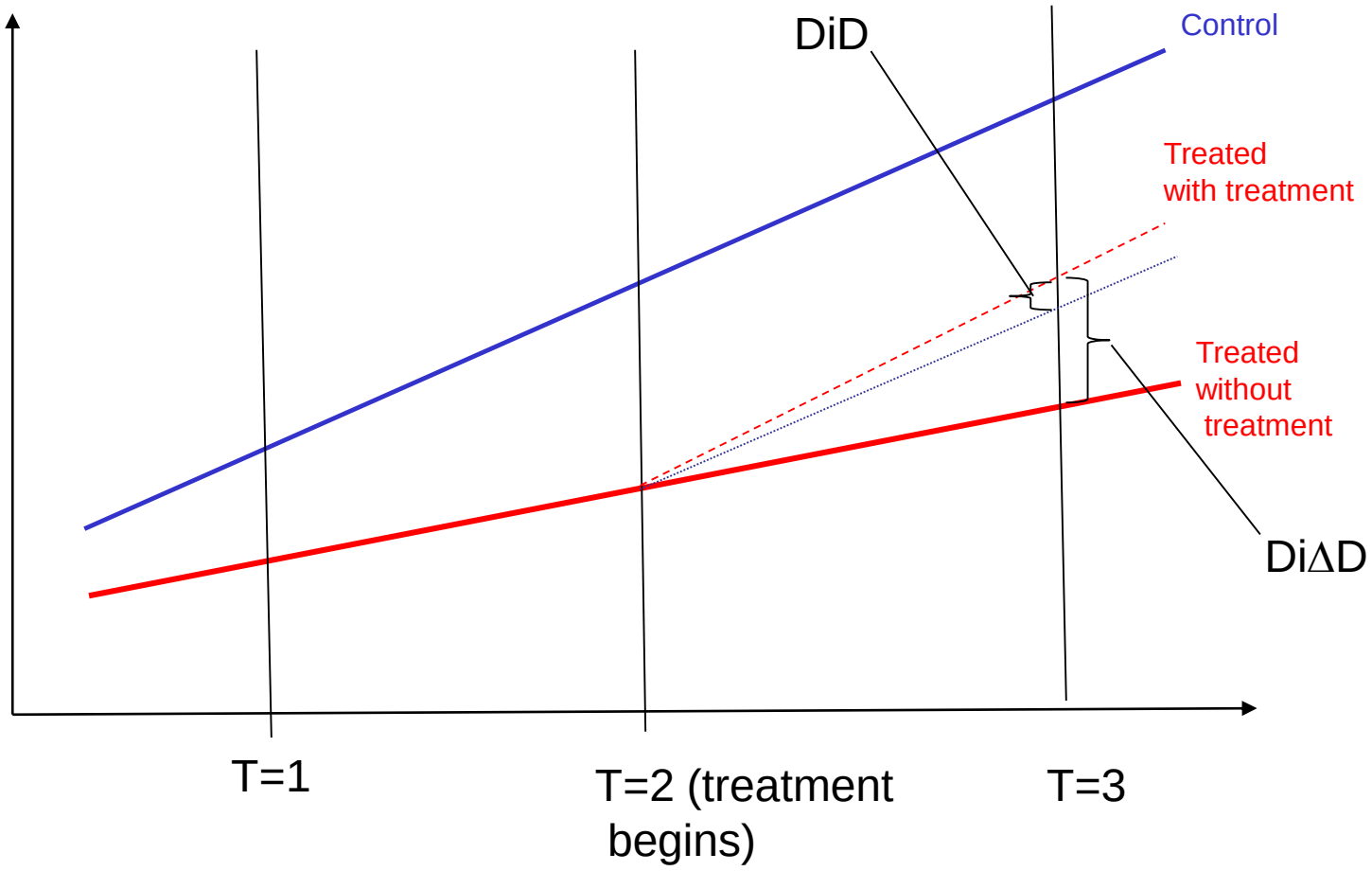
Source	SS	df	MS	Number of obs	=	20296090
				F(3, 20296086)	>	99999.00
Model	3.9602e+13	3	1.3201e+13	Prob > F	=	0.0000
Residual	3.7786e+13	20296086	1861747.31	R-squared	=	0.5117
				Adj R-squared	=	0.5117
Total	7.7388e+13	20296089	3812964.31	Root MSE	=	1364.5

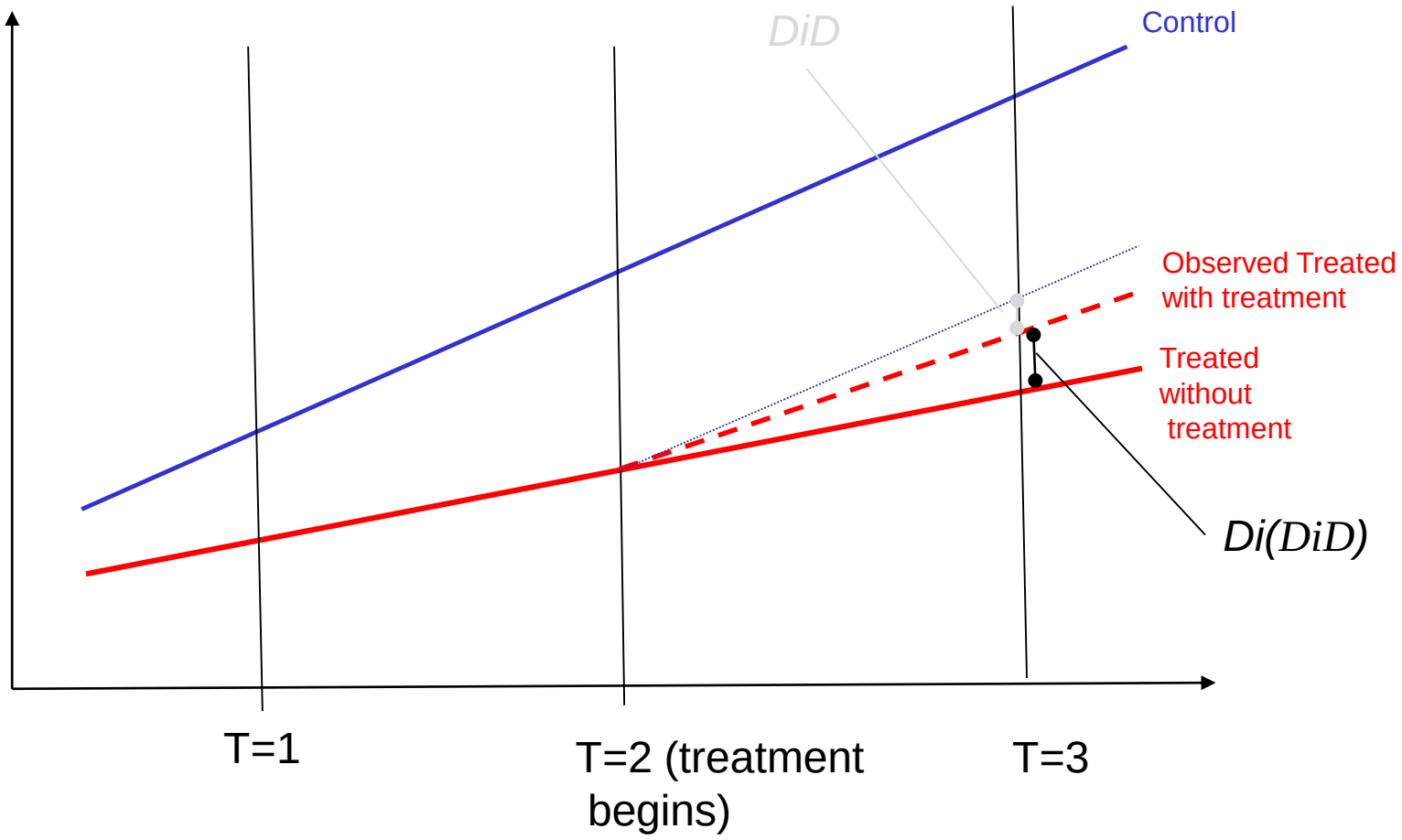
meanc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t	2679.322	.6481021	4134.11	0.000	2678.052	2680.593
treat	-1188.12	1.289127	-921.65	0.000	-1190.647	-1185.593
itr	-473.2427	1.82326	-259.56	0.000	-476.8162	-469.6692
_cons	8590.68	.4609965	1.9e+04	0.000	8589.776	8591.583

If you have several time periods **before** treatment begins….

- You can test the crucial common trend/parallelism assumption; simply by estimating DiD using two pre-treatment periods

- You can relax the parallel path assumption, assume parallel growth and estimate “treatment” as differences-in-(difference-in-difference) (i.e. change in growth rate differences instead of level differences over time)


DiD
Control
Treated
with treatment
Treated
without
treatment
DiΔD
T=1
T=2 (treatment
begins)
T=3

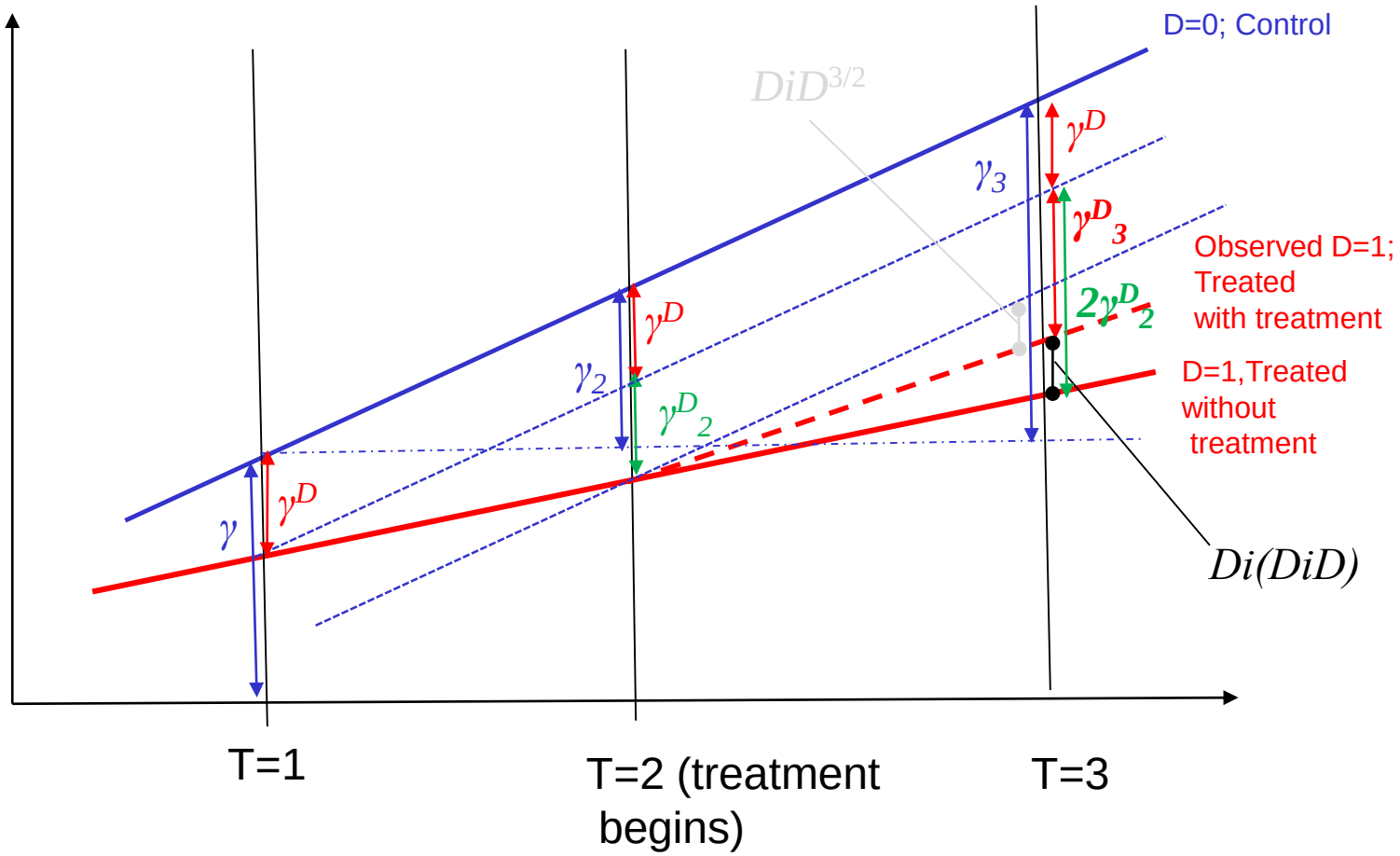

DiD
Control
Observed Treated with treatment
Treated without treatment
Di(DiD)
T=1
T=2 (treatment begins)
T=3

[Eq. 11] $Y_t = \gamma + \gamma_2 T_2 + \gamma_3 T_3 + \gamma^D D + \gamma^D{}_2 D.T_2 + \boldsymbol{\gamma^D{}_3} D.T_3$

Estimators = $DiD^{2/1} = \gamma^D{}_2 + \gamma^D - \gamma^D = \gamma^D{}_2$ reflecting the « normal » growth rate difference

$DiD^{3/2} = (\boldsymbol{\gamma^D{}_3} + \gamma^D) - (\gamma^D{}_2 + \gamma^D) = \boldsymbol{\gamma^D{}_3} - \gamma^D{}_2$;

$Di(DiD) = DiD^{3/2} - DiD^{2/1} = (\boldsymbol{\gamma^D{}_3} - \gamma^D{}_2) - \gamma^D{}_2 = \boldsymbol{\gamma^D{}_3} - 2\gamma^D{}_2$



#2EC_ex.do/Ex 1

4.2. Propensity score matching methods

In non-experimental economic data, we observe whether "individuals" were treated or not, but in the absence of random assignment, we must be concerned with differences between the treated and non-treated

One very appealing idea is to match "individuals" with maximal similarity

With a **single** measure (e.g. *X*=education), we can readily compute a measure of distance between a treated unit and each candidate match. With **multiple** measures defining similarity, how are we to balance similarity along each of those dimensions (e.g.: *X*=education, wealth, #siblings….)?

Propensity score matching (PSM): match treated and untreated observations on the estimated probability of being treated (propensity score).

$$P(X) = Pr(D=1|X)$$

with *D=1,0* indicates (non)treatment

=> Key idea : rather than matching on all values of the variables, individual units can be matched solely on the basis of their **propensity to be treated *P(X)***

Advantages of PSM

- Solves the “dimensionaliy” problem

- balances treatment and control groups on a large number of covariates(*X*) without losing a large number of observations
- If treatment & control were balanced one at a time, large numbers of observations would be needed = > *#X +1* increases the minimum necessary # of obs. geometrically

- PSM is non-parametric

does not rest on the validiy of a particular functional form like

[Eq. 12] $y = X\beta + \gamma D + \varepsilon$

where y is the outcome, X are covariates and D is the treatment indicator.

OLS estimation of [12] assumes that the effects of treatment γ are additive are constant across individuals.

PSM limitations

- Conditional independence: PSM validity rests on

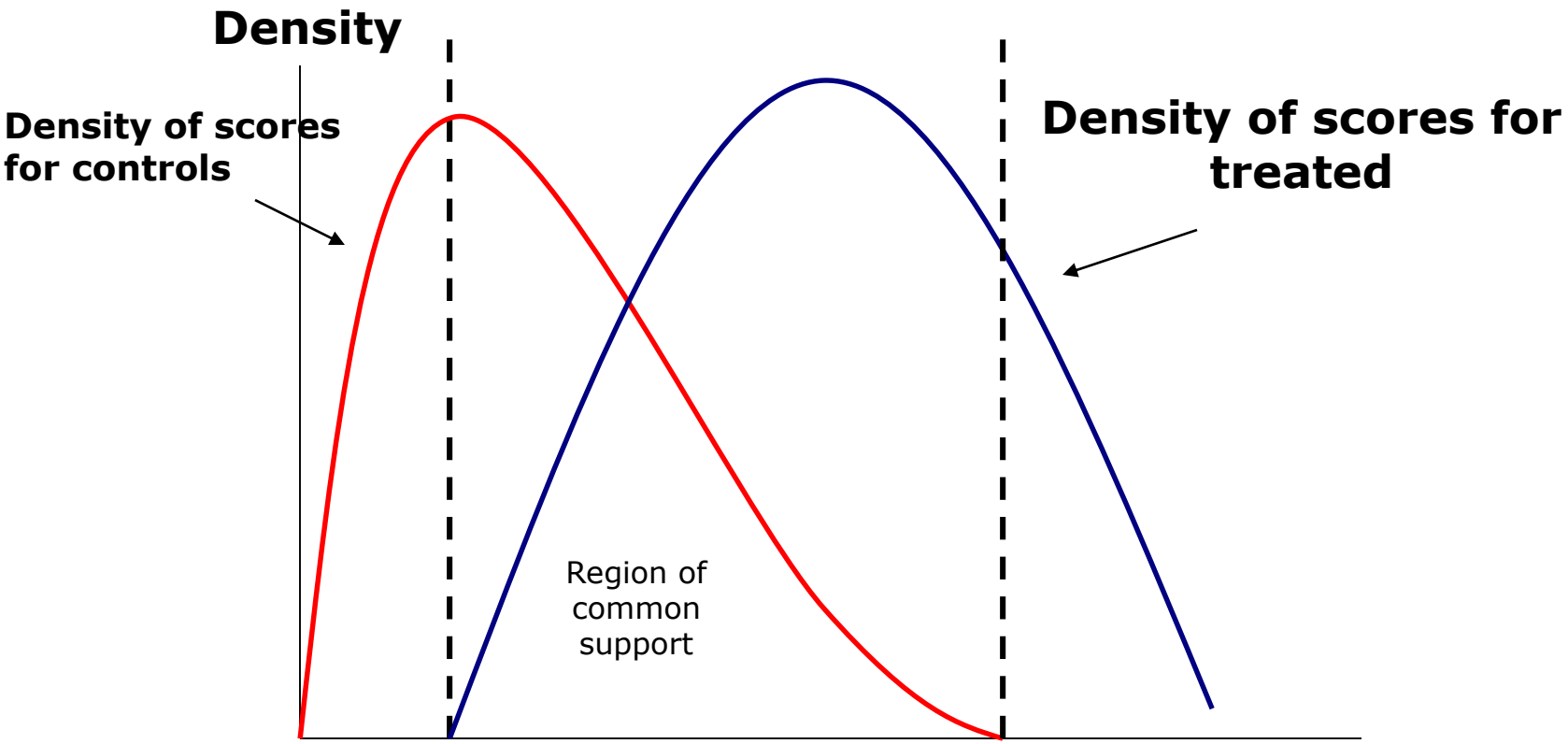
[Eq. 13] $(Y_1, Y_0) \perp D|X$

where Y_1, Y_0 denote treated/control potential outcomes under treatment

[and $\perp$ statistical independence].

Controlling for *X is* 'as good as random.' This assumption is also known as *selection on observables*

- Common support: there is sufficient overlap in propensity scores of treated and untreated units to find good matches



=> treated units whose *p(X)* is larger (lower) than the largest (lowest) *p* in the non-treated group should be excluded

PSM = three key steps:

1 Estimate the propensity score (logit, probit) : $p(X)$ for both treated and controls

2 Choose a matching algorithm that will use the estimated $\widehat{p(X)}$ to match untreated units to treated units

3 Estimate the impact of the intervention with the matched sample (usually as mean differences) and calculate standard errors

To estimate the propensity score, a logit or probit model is usually employed. Use **flexible functional form** to allow for possible nonlinearities in the participation model (introduce of higher-order & interaction terms)

In choosing a matching algorithm, choose w**ith or without replacement**. Without replacement, a given untreated unit can only be matched with one treated unit

A criterion for assessing the quality of the match must also be defined

ONE-TO-ONE MATCHING WITH REPLACEMENT (WITHIN CALIPER)

- Nearest-neighbour matching

The *p*'s are the propensity score delivered by logit/probit model

Treated unit *i* is matched to that non-treated unit *j* such that:

$$|p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

- Caliper matching

For a pre-specified $\delta>0$, treated unit *i* is matched to that non-treated unit *j* such that:

$$\delta > |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

KERNEL-BASED MATCHING

Idea

associate to the outcome y_i of treated unit i
a matched outcome given by a kernel-weighted average of the outcome of all non-treated units,
where the weight given to non-treated unit j is in proportion to the closeness between i and j:

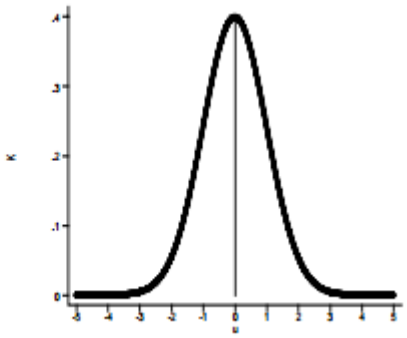
All non-treated *j* are used in the match…

$$\hat{y}_i = \frac{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right) y_j}{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$

… but their weight *K(…)* is inversely proportional to the distance between the propensity score *p*

$K(.)$

- non-negative
- symmetric
- unimodal



Post matching, it is recommended to test

- $D \perp X|p(X)$, meaning that after matching, there should be no statistically significant differences between covariate means of the treated and comparison units

- The common support condition. This can be done by visual inspection of the densities of propensity scores of treated and non-treated groups

➔ *#2EC_PSmatch.do/Case 1*

```
. reg wle priv   //no controls
```

Source	SS	df	MS		
Model	3047624.25	1	3047624.25	Number of obs	= 2700
Residual	18353531.7	2698	6802.64333	F(1, 2698)	= 448.01
				Prob > F	= 0.0000
				R-squared	= 0.1424
				Adj R-squared	= 0.1421
Total	21401156	2699	7929.29083	Root MSE	= 82.478

wle	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
priv	95.77514	4.524921	21.17	0.000	86.90248	104.6478
_cons	338.7235	1.715319	197.47	0.000	335.3601	342.087

```
. reg wle priv $ctl $ctl_i  //controls
note: int46 omitted because of collinearity
```

Source	SS	df	MS		
Model	4607752.71	35	131650.077	Number of obs	= 2522
Residual	15671773.3	2486	6304.01179	F(35, 2486)	= 20.88
				Prob > F	= 0.0000
				R-squared	= 0.2272
				Adj R-squared	= 0.2163
Total	20279526	2521	8044.2388	Root MSE	= 79.398

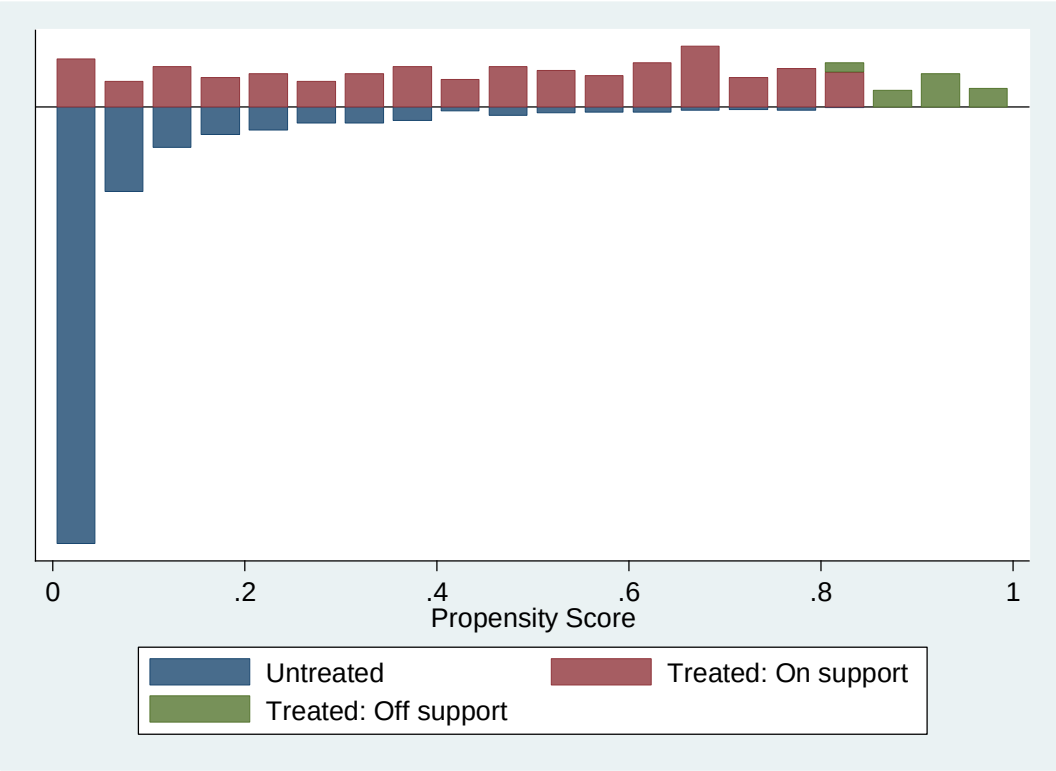
wle	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
priv	48.01915	5.526745	8.69	0.000	37.18165	58.85665
girl	-4.810277	17.12966	-0.28	0.779	-38.40014	28.77959
nsib	16.53822	25.10774	0.66	0.510	-32.69602	65.77246
brthord	-14.25147	12.37078	-1.15	0.249	-38.50956	10.00662
wealth	11.13111	1.883867	5.91	0.000	7.436995	14.82522
misced	-19.20848	21.35036	-0.90	0.368	-61.0748	22.65784
fathim	1.407819	78.50649	0.02	0.986	-152.537	155.3527
hisei	-3.618524	5.07347	-0.71	0.476	-13.56719	6.330138
hedres	.1800147	8.340129	0.02	0.983	-16.1743	16.53433
sq2	-7.257307	4.550105	-1.59	0.111	-16.17969	1.665079

```
. psmatch2 priv $ctl $ctl_i , outcome(wle) common quietly
There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling psmatch2.
```

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
wle	Unmatched	434.716201	340.539584	94.176617	4.63346595	20.33
	ATT	430.124956	381.892135	48.2328214	10.0531743	4.80

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

psmatch2: Treatment assignment	psmatch2: Common support		
	Off suppo	On suppor	Total
Untreated	0	2,143	2,143
Treated	42	337	379
Total	42	2,480	2,522


0
.2
.4
.6
.8
1
Propensity Score
Untreated
Treated: On support
Treated: Off support

```
. pstest $ctl, both                    //assess balancing of controls
```

Variable	Unmatched / Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|	V(T)/ V(C)
girl	U	.4934	.52636	-6.6		-1.18	0.237	.
	M	.48071	.4362	8.9	-35.0	1.16	0.247	.
nsib	U	.92084	.9468	-10.4		-2.01	0.045	.
	M	.92582	.94659	-8.4	20.0	-1.10	0.271	.
brthord	U	.38522	.32524	12.5		2.28	0.023	.
	M	.39169	.39763	-1.2	90.1	-0.16	0.875	.
wealth	U	-.12211	-1.7055	155.4		28.90	0.000	1.22
	M	-.32323	-.32374	0.0	100.0	0.01	0.995	0.84
misced	U	.75462	.23985	120.0		21.60	0.000	.
	M	.72404	.74481	-4.8	96.0	-0.61	0.542	.
fathim	U	.01319	.01307	0.1		0.02	0.984	.
	M	.0089	.0089	0.0	100.0	0.00	1.000	.
hisei	U	60.148	39.743	133.9		23.49	0.000	0.88
	M	58.407	58.769	-2.4	98.2	-0.33	0.740	0.97
hedres	U	-.27826	-1.6209	113.0		19.34	0.000	0.75*
	M	-.3897	-.34555	-3.7	96.7	-0.51	0.608	0.96

```
. tab _weight if priv==1
```

psmatch2: weight of matched controls	Freq.	Percent	Cum.
1	337	100.00	100.00
Total	337	100.00	

```
. tab _weight if priv==0          //freq of use among controls
```

psmatch2: weight of matched controls	Freq.	Percent	Cum.
1	130	66.67	66.67
2	34	17.44	84.10
3	16	8.21	92.31
4	6	3.08	95.38
5	3	1.54	96.92
6	1	0.51	97.44
7	2	1.03	98.46
9	1	0.51	98.97
10	1	0.51	99.49
13	1	0.51	100.00
Total	195	100.00	

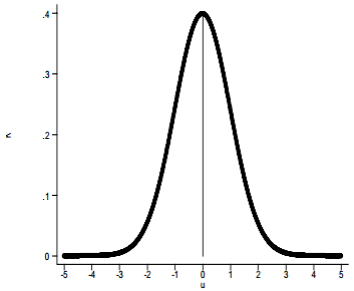
#2EC_Ex.do/Ex 2

4.3. Combining matching & DID

DID is a flexible form of causal inference because it can be combined with some other procedures, such as the Kernel Propensity Score (Heckman et al., 1997, 1998)

Step 1- compute kernel propensity score matching and retain the weights w_{ij}

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$



Step 2- run the traditional DID equ with non treated entities weighted by w_j [and treated one by 1]

#2EC_Ex.do/Ex 3

with

. reg y x1 x2 [aweight=*w*]

being equivalent to estimating

$$Y\sqrt{w} = \eta + \lambda_1 x_1\sqrt{w} + \lambda_2 x_2\sqrt{w}$$