# EC402 - Problem Set 4

Konrad Burchardi

18th of February 2009

# Introduction

Today we will talk about what is a suitable **control group** in non-experimental settings.

Depending on the data and our level of comfort regarding unobservables we can then apply

- **difference-in-difference** estimation or
- (propensity-score) **matching methods**

and we will see how these can go wrong.

# Introduction

## Why a control group?

Suppose we are interested in the treatment effect on the treated (TOT). If $\tau_i = Y_{i1} - Y_{i0}$ is the treatment effect for individual $i$ then TOT is

$$E[\tau_i|T_i = 1] = E[Y_{i1}|T_i = 1] - E[Y_{i0}|T_i = 1]$$

The trouble is that we never observe $Y_{i0}|T_i = 1$!

That is why we search for a group that *is not treated and behaves the same way as the treated group would have behaved in the absence of treatment*, a **control group**.

# Introduction

## Why a control group?

So, if we need a control group, why do we not just take experimental data?

- Often experiments might not be possible.
- Or collecting data for a control group is too expensive.

# Question 1

## The Difference-in-Difference idea

Suppose we have data on the outcome we are interested in for two groups both before and after some treatment occurred.

- *Group* $1$ was actually treated.
- For some untreated group, *group* $0$, we believe that the **trend of its outcome data** is the same as the trend for *group* $1$ would have been if *group* $1$ had not been treated.

# Question 1

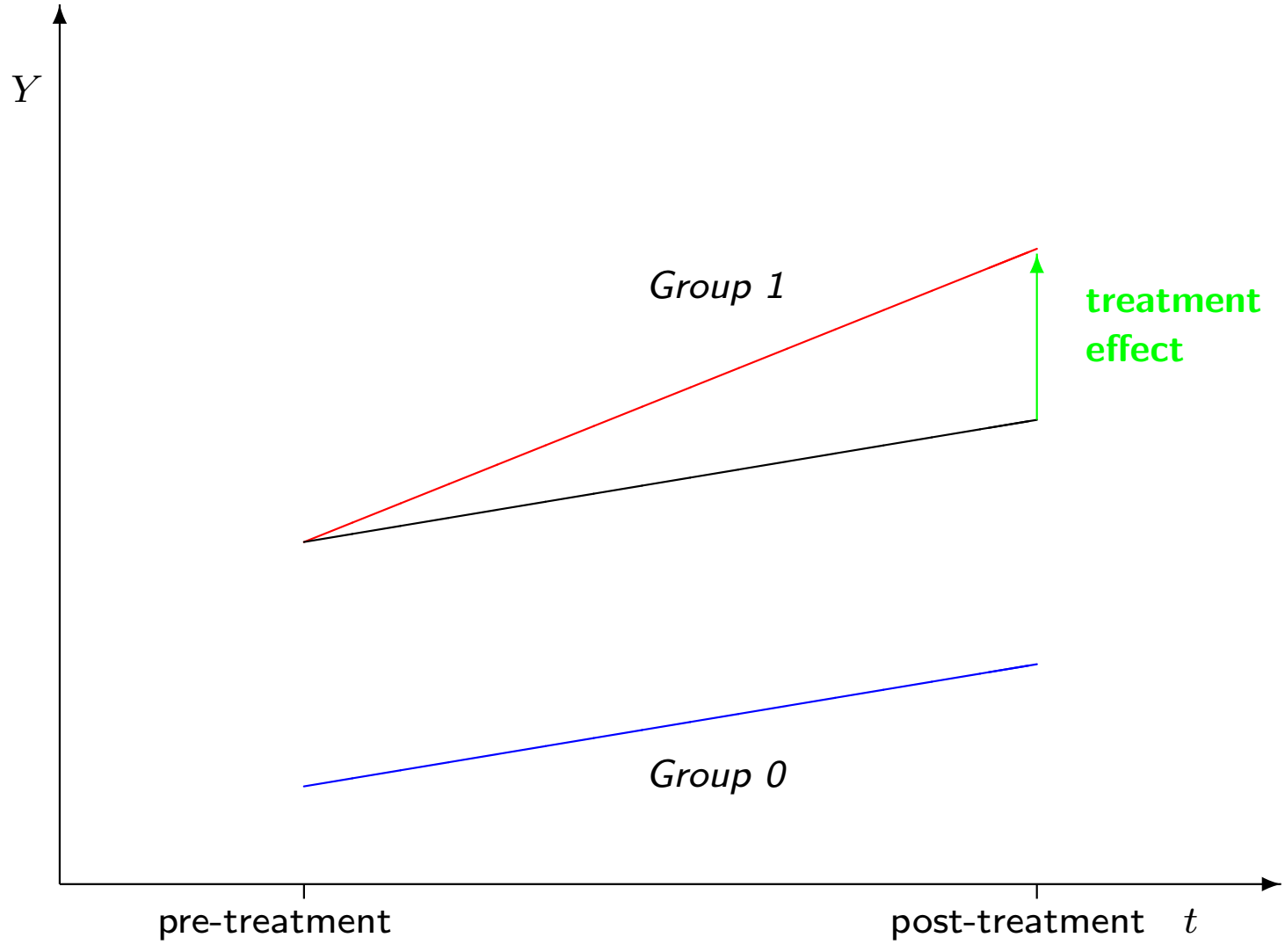
## The Difference-in-Difference idea

Then we can take the actual development of the outcome for *group* $0$ and together with the initial value (pre-treatment) for *group* $1$ to calculate where *group* $1$ would have been, if it had not been treated, i.e. $E[Y_{i0}|T_i = 1]$.

The difference between what the outcome actually was and what it would have been in the absence of treatment is our **difference-in-difference estimate** of the treatment effect.

# Question 1

## The Difference-in-Difference idea



# Question 1

## C. Using the diff-in-diff method

In our data we have earning for treated individuals and match data for untreated individuals from the PSID dataset, the later being our *group* $0$.

```
. ttest ba_diff, by(treatment)

Two-sample t test with equal variances
```

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 | 429 | -4517.685 | 333.9821 | 6917.543 | -5174.135 | -3861.236 |
| 1 | 185 | -4817.088 | 608.4202 | 8275.409 | -6017.465 | -3616.711 |
| combined | 614 | -4607.896 | 296.5206 | 7347.49 | -5190.216 | -4025.577 |
| diff | | 299.4029 | 646.6766 | | -970.5715 | 1569.377 |

```
    diff = mean(0) - mean(1)                                      t =   0.4630
Ho: diff = 0                     degrees of freedom =      612

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6782         Pr(|T| > |t|) = 0.6435          Pr(T > t) = 0.3218
```

Using the diff-in-diff method we estimate a treatment effect of 299 dollars of the training programme under study.

# Question 1

## B. Using experimental data

In our case we actually do have an experimental control group.[1] Using this we estimate the treatment effect to be

```
. reg re78 treatment
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 137332501 | 1 | 137332501 | Number of obs | = 722 |
| Residual | 2.8053e+10 | 720 | 38962866.3 | F( 1, 720) | = 3.52 |
| | | | | Prob > F | = 0.0609 |
| Total | 2.8191e+10 | 721 | 39099301.3 | R-squared | = 0.0049 |
| | | | | Adj R-squared | = 0.0035 |
| | | | | Root MSE | = 6242 |

| re78 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| treatment | 886.3037 | 472.0863 | 1.88 | 0.061 | -40.52635 | 1813.134 |
| _cons | 5090.048 | 302.7826 | 16.81 | 0.000 | 4495.606 | 5684.491 |

886 dollars!

[1] Question 1A, which I skip, shows how most observable characteristics which might influence the outcome as well are very similar in the control and treatment group. This is what we would expect under random treatment assignment. It is precisely the point of random assignment that all observable and - more importantly - unobservable characteristics are similar in both the treatment and control group! (Then, not surprisingly, including these controls does not change much in 1B.)

# Question 1

## D. Understanding diff-in-diff

**So what went wrong when using the diff-in-diff method**?

- Is the fact that many observable characteristics are different between the treatment group and the control group picked by LaLonde (1986) necessarily a problem for our diff-in-diff estimation?

- What is crucial for the diff-in-diff method to work?

# Question 1

## E. Understanding diff-in-diff

Obviously we never know how the treated would have developed in the absence of treatment, so we cannot test the crucial part of the **common-trend assumption**.

However, if we can show that for some periods prior to the treatment the control and treatment groups were developing similarly, this would be a convincing argument for our control group to be a good control group.

# Question 1

## E. Understanding diff-in-diff

In our case we have data for 1974 and 1975 (the treatment happened between 1975 and 1978).

```
. ttest bb_diff, by(treatment)

Two-sample t test with equal variances
```

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| 0 | 429 | 3152.752 | 280.3635 | 5806.977 | 2601.691 | 3703.813 |
| 1 | 185 | 563.5184 | 278.2071 | 3784.026 | 14.63228 | 1112.404 |
| combined | 614 | 2372.608 | 218.284 | 5408.865 | 1943.933 | 2801.284 |
| diff | | 2589.234 | 464.4893 | | 1677.047 | 3501.42 |

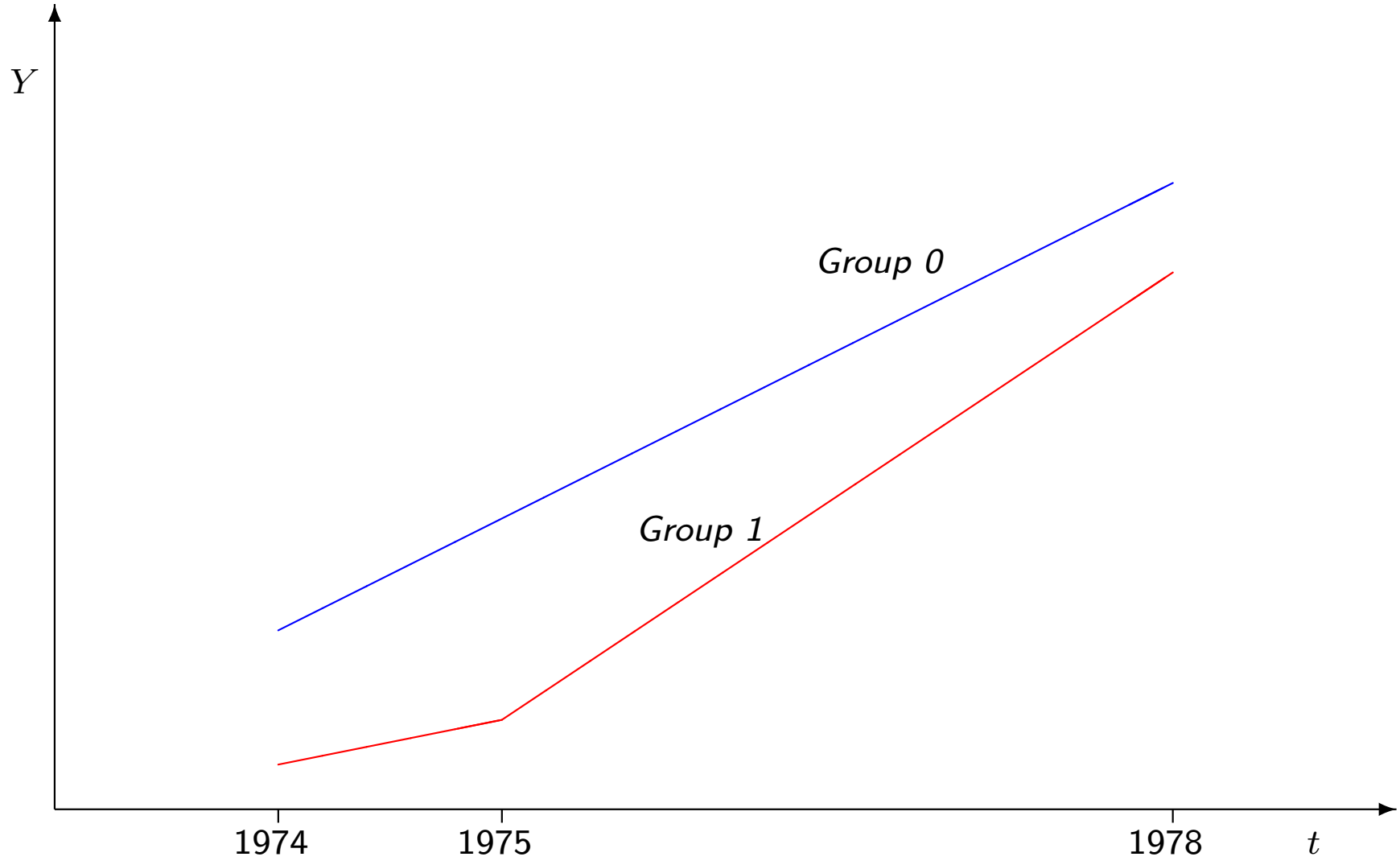
```
    diff = mean(0) - mean(1)                                      t =   5.5744
Ho: diff = 0                     degrees of freedom =      612

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

**The trends prior to treatment are in fact not at all similar!**

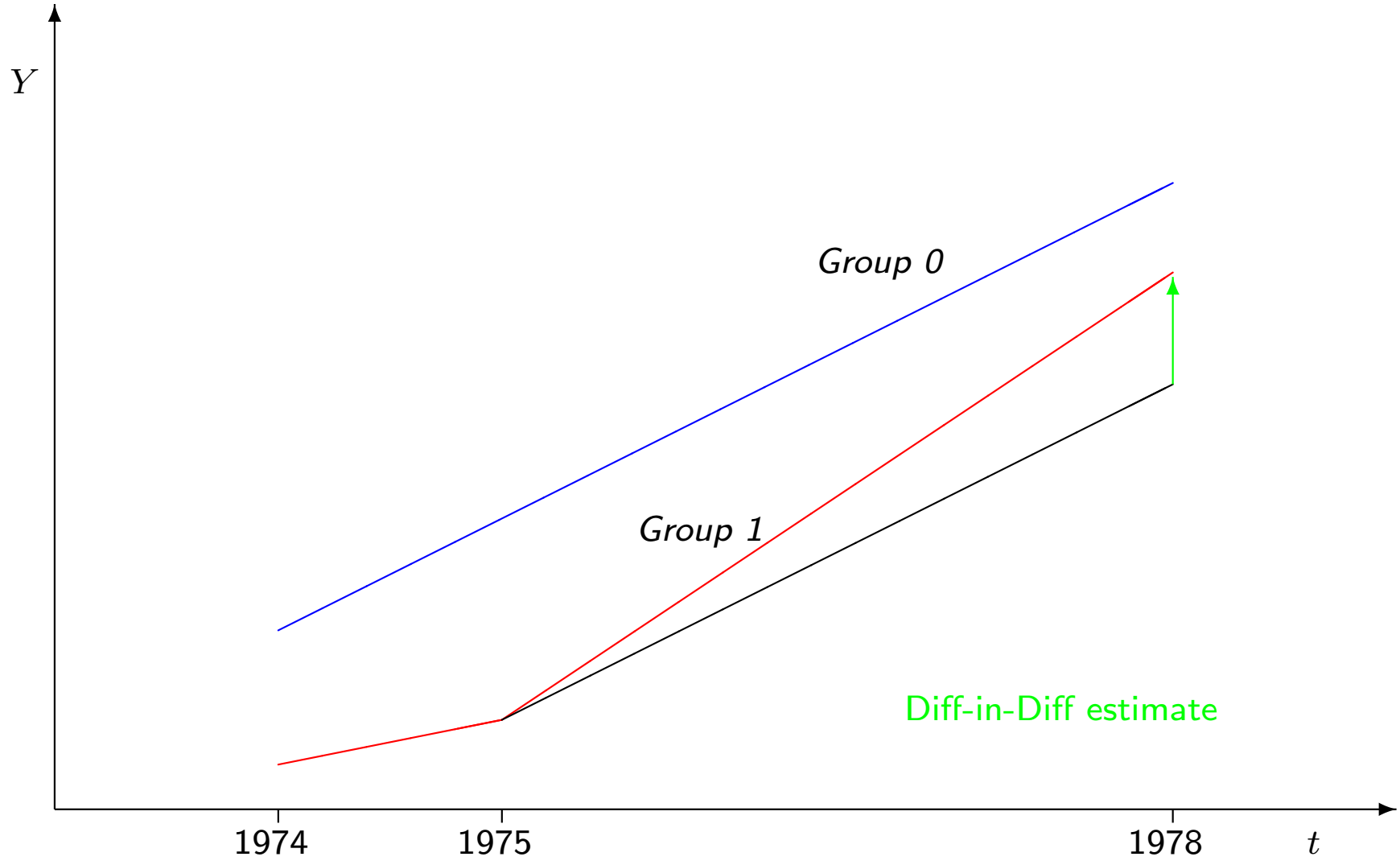
# Question 1

## E. Understanding diff-in-diff



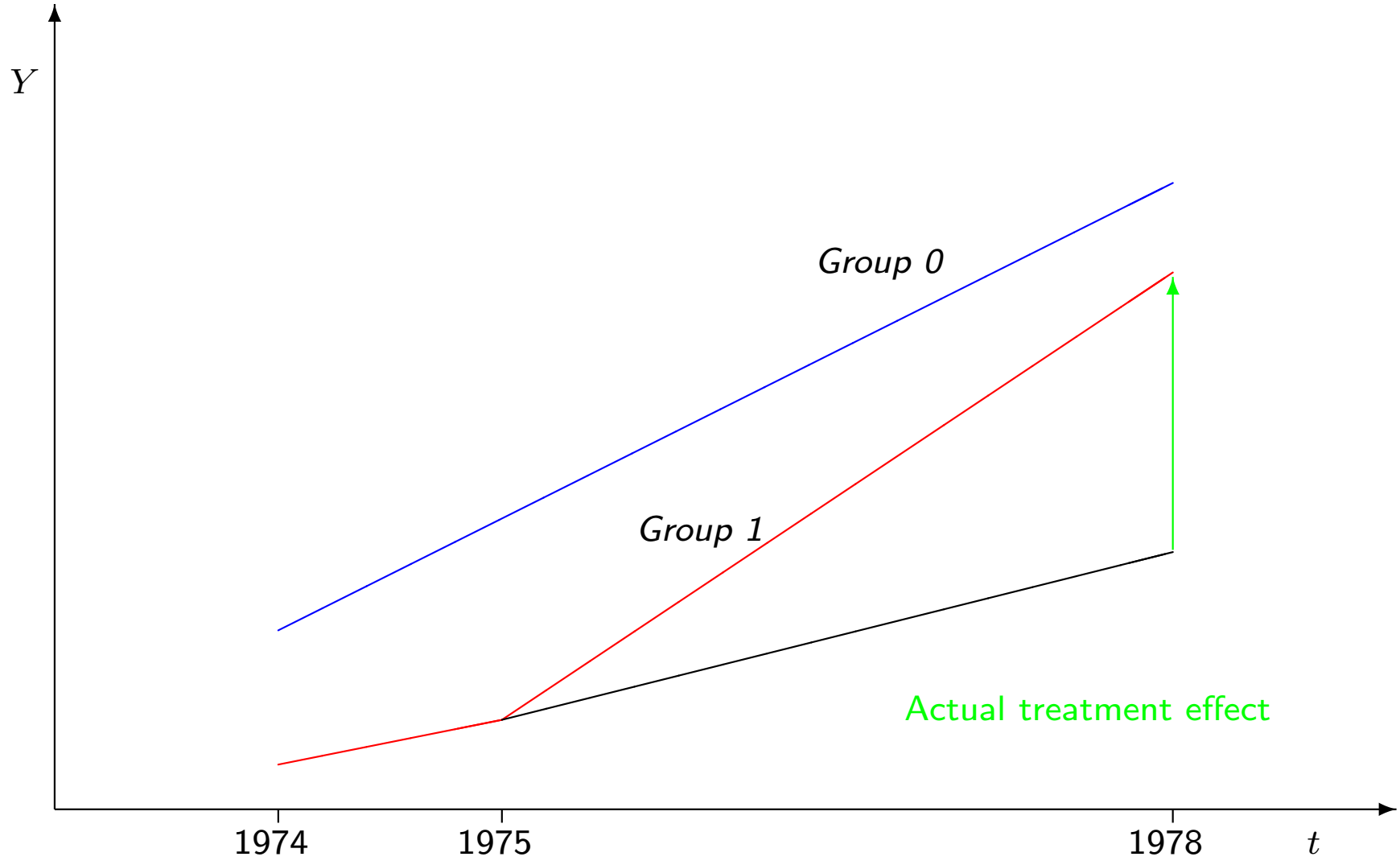
# Question 1

E. Understanding diff-in-diff



# Question 1

E. Understanding diff-in-diff



# Question 1

**Lessons**

- If we have pre- and post-treatment data for treatment and control group, diff-in-diff might be a suitable method.
- Common trends assumption needs to be satisfied in the example at hand!
- We can gain some confidence about this by looking at pre- and post-treatment trends.

# Question 2

## Introduction

A different method to create a control group is a matching estimator. There we *match* treated observations to non-treated observations with similar observable characteristics $X$. **If** the treatment conditional on $X_i$ is as good as random, comparing the outcomes for both groups gives an unbiased estimator. This assumption is called 'conditional independence assumption' (CIA).

However, if $X$ is multidimensional this requires a lot of data - and is hence often not possible.

# Question 2

## Introduction

**Propensity-score matching** allows us to reduce the 'dimensionality of the problem': It has been proven that, rather than matching on all kind of combination of the multidimensional $X$, we can just match on $P(X_i) = E[T_i|X_i]$, which is one-dimensional.

*If the CIA indeed holds* (this is important), then this will give an unbiased estimate of the treatment effect.

# Question 2

## A. Implementation

Suppose we know the true propensity scores (which we never do). Then we can define

- a set of dummy variables $p_m$ where $p_m = 1$ if an individuals propensity score $p(X)$ satisfies $m < p(X) \leq m + 0.1$ and $m \in M = 0, 0.1, 0.2, ..., 0.9$.
- a set of interaction terms $T_i \cdot p_{m,i}$ which is $1$ if the individual is in the treatment group **and** the propensity score is between $m$ and $m + 0.1$.

We can then estimate

$$Y_i = \delta T_i + \sum_{m \in M} \gamma_m p_{m,i} + \sum_{m \in M} \beta_m (T_i \cdot p_{m,i}) + \epsilon_i$$

**Intuition?**

# Question 2

## B. Assumption

**Note again:**

This will only work if the Conditional Independence Assumptions (CIA) does hold!

# Question 2

## C. Implementation

Is the propensity score method flexible?

# Question 3

## A./B./C. How to estimate $P(X)$?

In practice the **propensity score** is unknown. We do know whether somebody was treated or not, but not what was the a priori probability that he will be treated.

We can try to estimate how this probability depends on $X$. We can use e.g.

- a **probit model** or
- a simple **linear probability model**.

# Question 3

## A. How to estimate $P(X)$?

If we use the **probit model** (or just **pscore** in stata) we have to drop the *nodegree* variable for the balancing property to be satisfied.

- What is the balancing property?
- When is it a problem to drop some variables, for which it is not satisfied?

# Question 3

## B. How to estimate $P(X)$?

How are the estimated propensity scores for the treated and untreated distributed?

| p2 | treatment 0 | 1 | Total |
|---|---|---|---|
| 0 | 14,333 | 20 | 14,353 |
| .01 | 346 | 2 | 348 |
| .02 | 215 | 4 | 219 |
| .03 | 234 | 7 | 241 |
| .04 | 189 | 7 | 196 |
| .05 | 118 | 8 | 126 |
| .06 | 54 | 5 | 59 |
| .07 | 37 | 5 | 42 |
| .08 | 16 | 0 | 16 |
| .09 | 13 | 2 | 15 |
| .1 | 21 | 0 | 21 |
| .11 | 21 | 1 | 22 |
| .12 | 13 | 2 | 15 |
| .13 | 20 | 2 | 22 |
| .14 | 20 | 3 | 23 |
| .15 | 16 | 3 | 19 |
| .16 | 12 | 3 | 15 |
| .17 | 15 | 6 | 21 |
| .18 | 18 | 4 | 22 |
| .19 | 24 | 8 | 32 |
| .2 | 30 | 13 | 43 |
| .21 | 22 | 8 | 30 |
| .22 | 40 | 13 | 53 |
| .23 | 28 | 7 | 35 |
| .24 | 9 | 6 | 15 |
| .25 | 27 | 7 | 34 |
| .26 | 12 | 2 | 14 |
| .27 | 15 | 11 | 26 |
| .28 | 10 | 6 | 16 |
| .29 | 18 | 2 | 20 |
| .3 | 14 | 6 | 20 |
| .31 | 14 | 3 | 17 |
| .32 | 9 | 4 | 13 |
| .33 | 1 | 3 | 4 |
| .34 | 1 | 0 | 1 |
| .35 | 3 | 1 | 4 |
| .36 | 1 | 1 | 2 |
| .38 | 1 | 0 | 1 |
| .39 | 1 | 0 | 1 |
| .4 | 1 | 0 | 1 |
| Total | 15,992 | 185 | 16,177 |

# Question 3

## C. How to estimate $P(X)$?

If we use a **linear probability model** instead, we get

```
. reg treatment age education black hispanic married
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 18.1851766 | 5 | 3.63703532 | Number of obs | = 16177 |
| Residual | 164.699165 | 16171 | .010184847 | F( 5, 16171) | = 357.10 |
| | | | | Prob > F | = 0.0000 |
| | | | | R-squared | = 0.0994 |
| | | | | Adj R-squared | = 0.0992 |
| Total | 182.884342 | 16176 | .011305906 | Root MSE | = .10092 |

| treatment | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| age | -.0002708 | .000081 | -3.34 | 0.001 | -.0004296 | -.0001119 |
| education | -.0010997 | .0002873 | -3.83 | 0.000 | -.0016628 | -.0005366 |
| black | .1110172 | .0029347 | 37.83 | 0.000 | .1052648 | .1167696 |
| hispanic | .0057716 | .0031484 | 1.83 | 0.067 | -.0003996 | .0119428 |
| married | -.0195236 | .0019524 | -10.00 | 0.000 | -.0233506 | -.0156966 |
| _cons | .037838 | .0046064 | 8.21 | 0.000 | .028809 | .046867 |

- Why are the coefficient easier to interpret?
- What is the problem with the linear probability model?

# Question 3

## D. Using the Propensity-Score method

Finally, if we use the estimated propensity scores and match the treated and untreated individuals in 4 bins of propensity scores (0.0-0.1, 0.1-0.2,...), what are the **treatment effects in each bin of the propensity score**?

# Question 3

## D. Using the Propensity-Score method

```
. reg re78 p_1 p_2 p_3 treatment p_treat1 p_treat2 p_treat3
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 2.9769e+10 | 7 | 4.2527e+09 | Number of obs | = 16177 |
| Residual | 1.4831e+12 | 16169 | 91727540 | F( 7, 16169) | = 46.36 |
| | | | | Prob > F | = 0.0000 |
| | | | | R-squared | = 0.0197 |
| | | | | Adj R-squared | = 0.0193 |
| Total | 1.5129e+12 | 16176 | 93528158.7 | Root MSE | = 9577.4 |

| re78 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| p_1 | -4101.933 | 700.8708 | -5.85 | 0.000 | -5475.717 | -2728.148 |
| p_2 | -6940.625 | 690.1353 | -10.06 | 0.000 | -8293.367 | -5587.884 |
| p_3 | -10853.74 | 1694.808 | -6.40 | 0.000 | -14175.75 | -7531.728 |
| treatment | -7105.638 | 1238.823 | -5.74 | 0.000 | -9533.868 | -4677.409 |
| p_treat1 | 4166.125 | 2014.548 | 2.07 | 0.039 | 217.3883 | 8114.861 |
| p_treat2 | 3273.189 | 1831.397 | 1.79 | 0.074 | -316.5515 | 6862.929 |
| p_treat3 | 7550.119 | 3470.609 | 2.18 | 0.030 | 747.3407 | 14352.9 |
| _cons | 15001.49 | 76.74002 | 195.48 | 0.000 | 14851.07 | 15151.91 |

- For whom are the estimated treatment effects highest highest?
- Does this make sense?

# Question 3

## E. Using the Propensity-Score method

If we now include the covariates $X$, nothing should change if the CIA indeed holds.

```
. reg re78 p_1 p_2 p_3 treatment p_treat1 p_treat2 p_treat3 age education black hispanic married
```

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 1.4761e+11 | 12 | 1.2301e+10 | Number of obs | = 16177 |
| Residual | 1.3653e+12 | 16164 | 84465418.8 | F( 12, 16164) | = 145.63 |
| | | | | Prob > F | = 0.0000 |
| Total | 1.5129e+12 | 16176 | 93528158.7 | R-squared | = 0.0976 |
| | | | | Adj R-squared | = 0.0969 |
| | | | | Root MSE | = 9190.5 |

| re78 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| p_1 | 1632.492 | 763.606 | 2.14 | 0.033 | 135.7398 | 3129.245 |
| p_2 | 191.3484 | 755.1 | 0.25 | 0.800 | -1288.731 | 1671.428 |
| p_3 | -2109.405 | 1670.47 | -1.26 | 0.207 | -5383.711 | 1164.902 |
| treatment | -4191.879 | 1201.372 | -3.49 | 0.000 | -6546.701 | -1837.058 |
| p_treat1 | 1563.486 | 1940.636 | 0.81 | 0.420 | -2240.376 | 5367.349 |
| p_treat2 | 464.183 | 1765.967 | 0.26 | 0.793 | -2997.308 | 3925.674 |
| p_treat3 | 4733.499 | 3334.938 | 1.42 | 0.156 | -1803.348 | 11270.35 |
| age | 64.86673 | 7.426033 | 8.74 | 0.000 | 50.31088 | 79.42258 |
| education | 446.6367 | 26.30531 | 16.98 | 0.000 | 395.0754 | 498.198 |
| black | -2381.292 | 342.9629 | -6.94 | 0.000 | -3053.537 | -1709.047 |
| hispanic | -858.5486 | 286.9083 | -2.99 | 0.003 | -1420.921 | -296.1765 |
| married | 4637.042 | 184.4558 | 25.14 | 0.000 | 4275.488 | 4998.596 |
| _cons | 4238.746 | 425.3136 | 9.97 | 0.000 | 3405.084 | 5072.407 |

- What changes, what not?
- Does this make you have doubts about whether the CIA holds?

# Question 3

## Conclusion

Are the results from propensity score matching **close to the experimental results**?