Causality 3

http://users.ox.ac.uk/~sfos0015

1. DAGs.
2. Instrumental variables.
3. Unobservables and selection models.
4. Other “quasi-experimental” designs.
5. Concluding thoughts.

DAGs

Building blocks

Conditional Independence

1. Chain of mediation

$$A \longrightarrow B \longrightarrow C$$

A & C are marginally dependent;
A & C are conditionally independent.

$$
\begin{tikzcd}
 & A \arrow[dl] \arrow[dr] & \\
B & & C
\end{tikzcd}
$$

B & C are marginally dependent;
B & C are conditionally independent.

2. Mutual dependence

3. Collider (Mutual Causation)

$$
\begin{tikzcd}
A \arrow[dr] & & B \arrow[dl] \\
 & C &
\end{tikzcd}
$$

A & B are marginally independent;
A & B are conditionally dependent.

Collider Bias

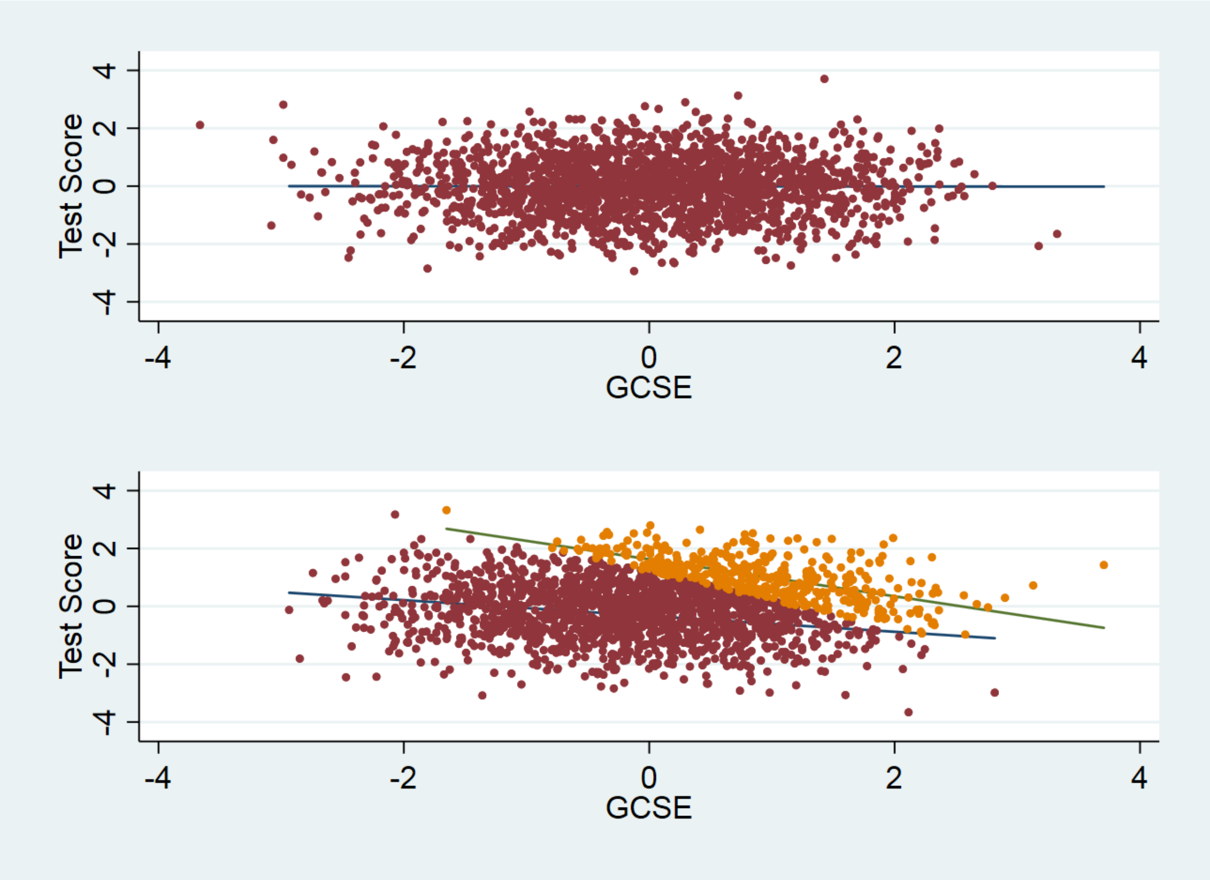
1. Draw sample of 2000 applicants from bivariate normal distribution for GCSE and entrance test scores with means = 0, standard deviations = 1, correlation = 0.
2. Add GCSE score and entrance test score.
3. Admit top 20%.
4. Draw scatterplot of GCSE versus entrance score for:

    Whole population – ie the marginal relationship.

    Whole population conditioning on admission ie the conditional relationship.

Get the Stata do file from: http://users.ox.ac.uk/~sfos0015/collider_bias.do

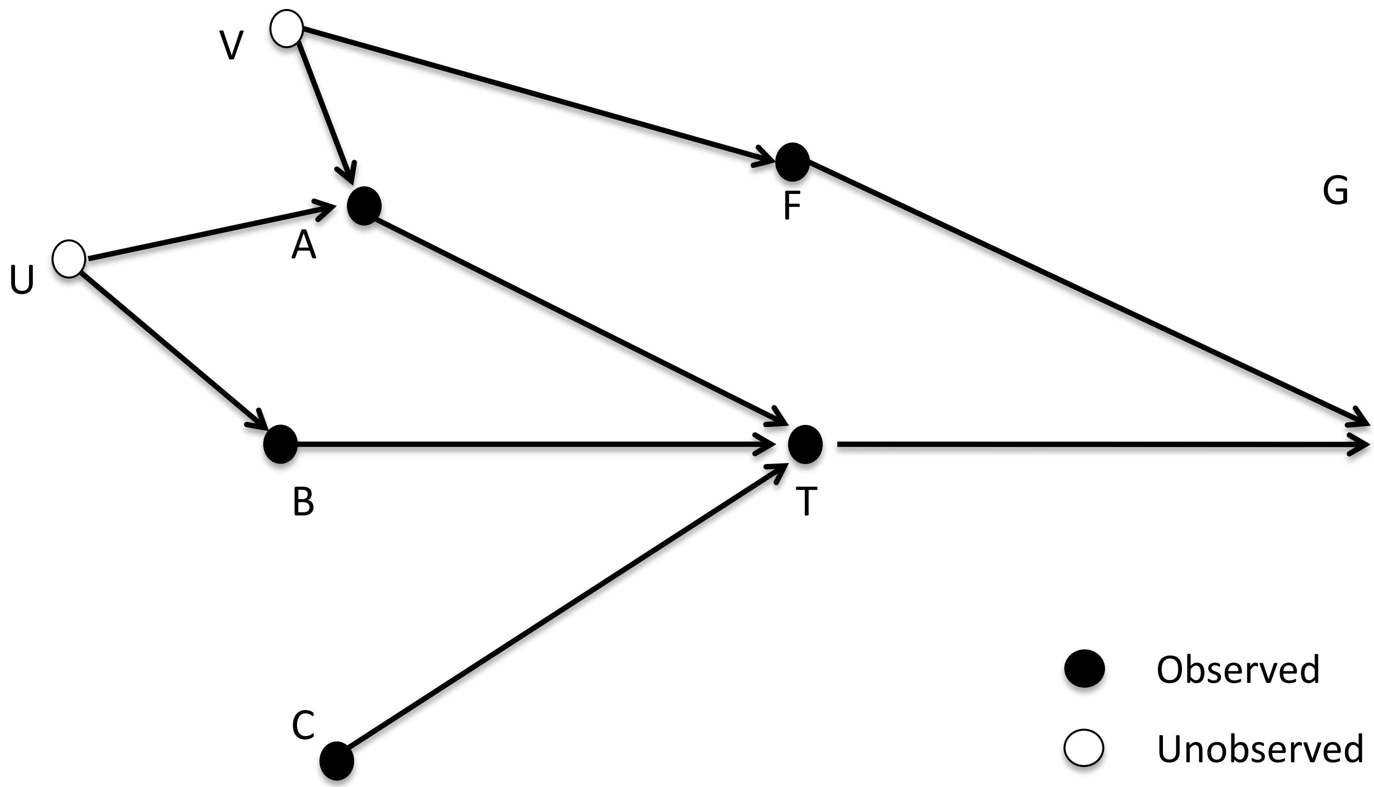
Collider Bias



DAGs

See Morgan & Winship (2007) pp72-73.

Conditioning set to identify ATE of T on Y



DAGs

Rules

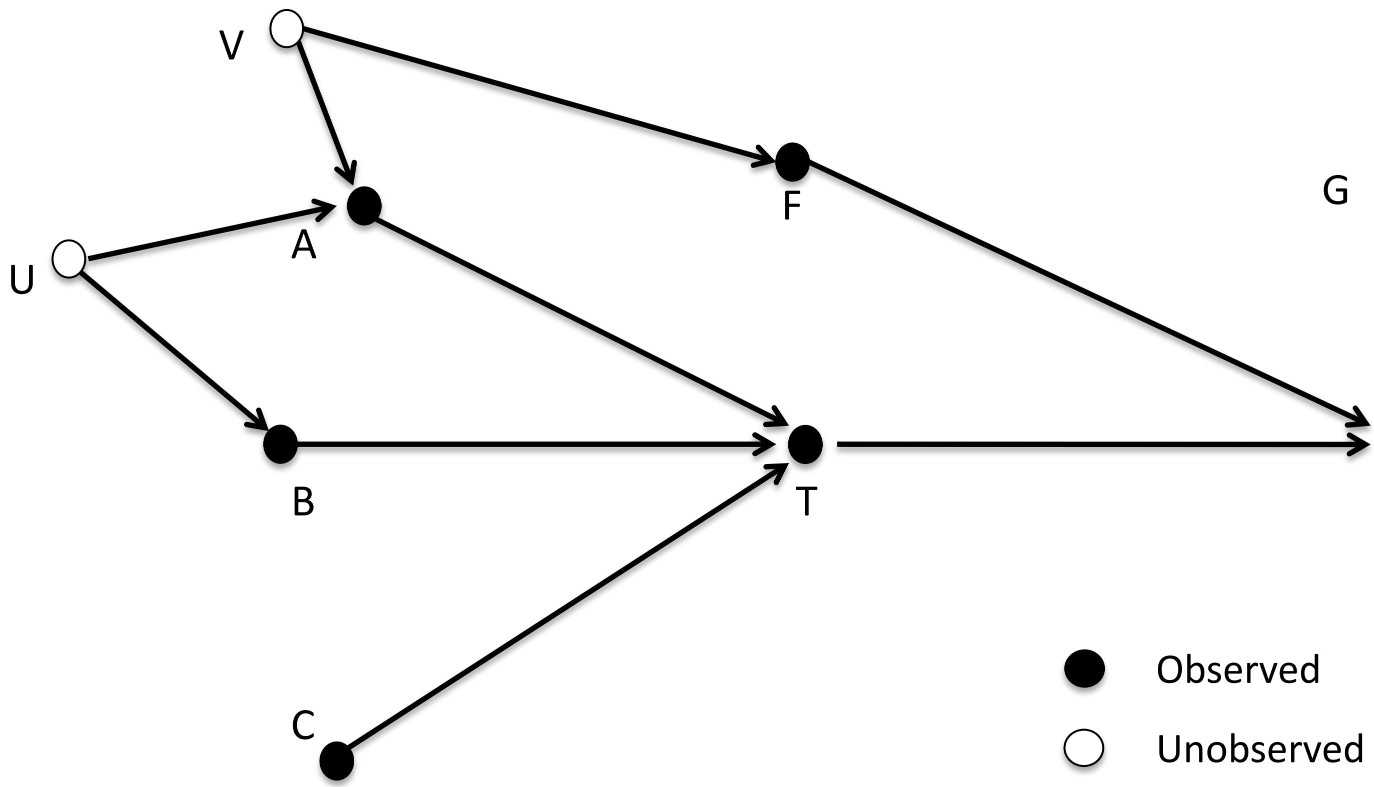
1. Aim: to identify ATE of T on Y.
    1. **Don't condition on an endogenous mediator.**
    2. Do condition on nodes on backdoor pathways.
    3. Backdoor pathways with collider nodes are blocked.
    4. Conditioning on a collider unblocks a pathway.
2. To identify ATE make sure the backdoor is closed.
3. Define a conditioning set.

UNIVERSITY OF OXFORD

DAGs

See Morgan & Winship (2007) pp72-73.

Conditioning set to identify ATE of T on Y



DAGs

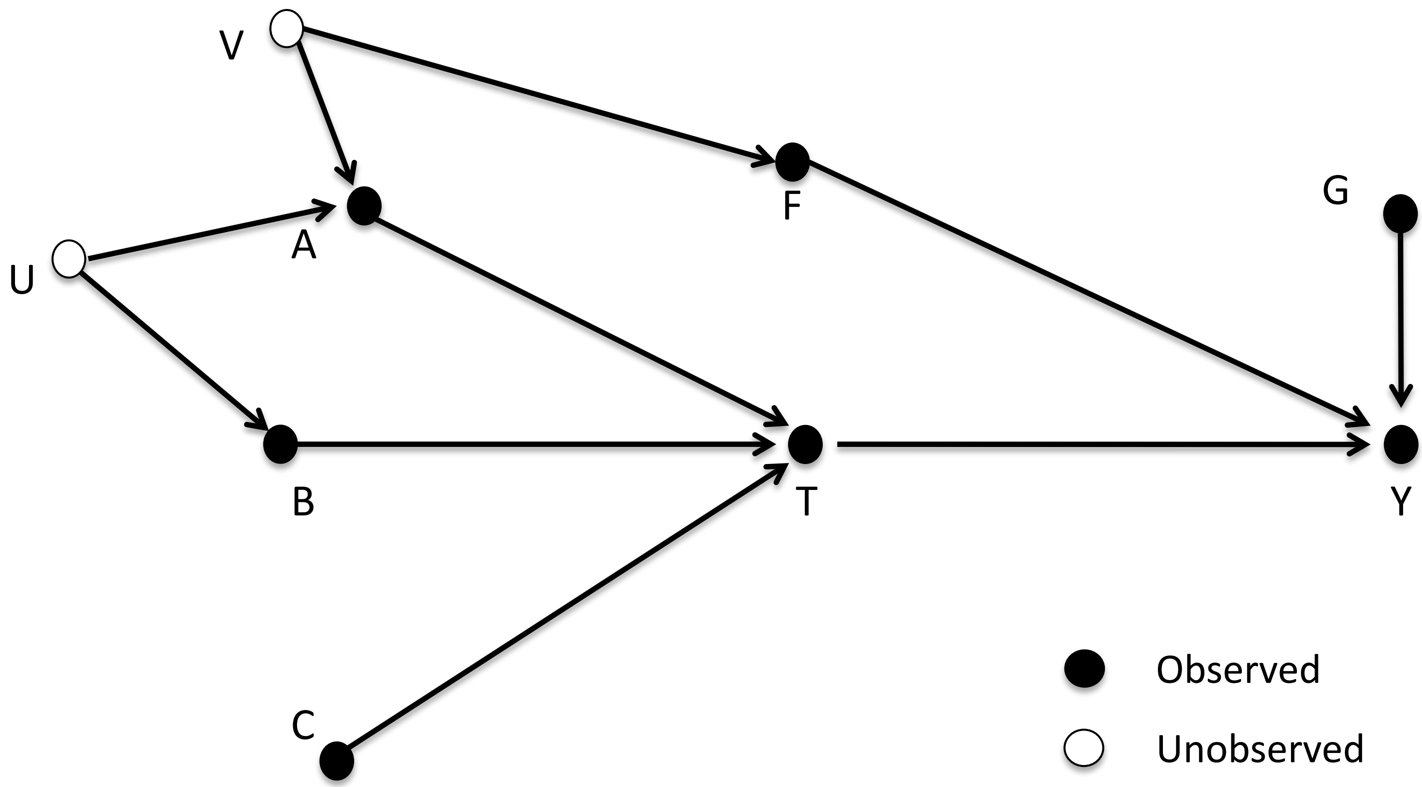
Rules

1. Aim: to identify ATE of T on Y.
    1. Don’t condition on an endogenous mediator.
    2. **Do condition on nodes on backdoor pathways.**
    3. Backdoor pathways with collider nodes are blocked.
    4. Conditioning on a collider unblocks a pathway.
2. To identify ATE make sure the backdoor is closed.
3. Define a conditioning set.

DAGs

See Morgan & Winship (2007) pp72-73.

Conditioning set to identify ATE of T on Y



DAGs

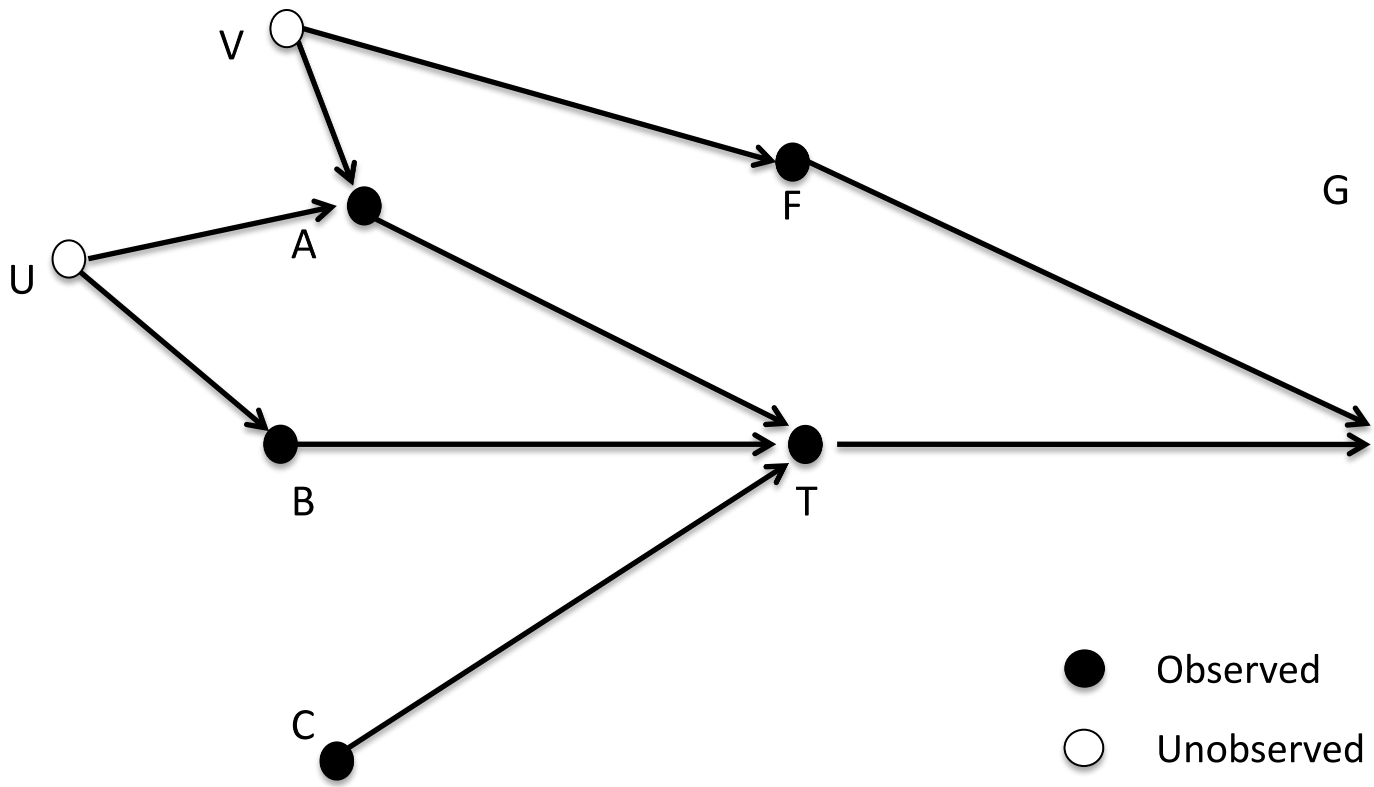
Rules

1. Aim: to identify ATE of T on Y.
    1. Don’t condition on an endogenous mediator.
    2. Do condition on nodes on backdoor pathways.
    3. **Backdoor pathways with collider nodes are blocked.**
    4. **Conditioning on a collider unblocks a pathway.**
2. To identify ATE make sure the backdoor is closed.
3. Define a conditioning set.

DAGs

See Morgan & Winship (2007) pp72-73.

Conditioning set to identify ATE of T on Y



DAGs

Rules

1. Aim: to identify ATE of T on Y.
    1. Don't condition on an endogenous mediator.
    2. Do condition on nodes on backdoor pathways.
    3. Backdoor pathways with collider nodes are blocked.
    4. Conditioning on a collider unblocks a pathway.
2. **To identify ATE make sure the backdoor is closed.**
3. **Define a conditioning set.**

DAGs

Conditioning set to identify ATE of T on Y

Summary

1. Backdoor paths are:
   1. T<-B<-U->A<-V->F->Y;
   2. T<-A<-V->F->Y.
2. A is a collider on first backdoor path.

3. Conditioning sets {F] or {A, B} close the backdoor.

DAGs

Experimental investigation of recidivism

Reading:

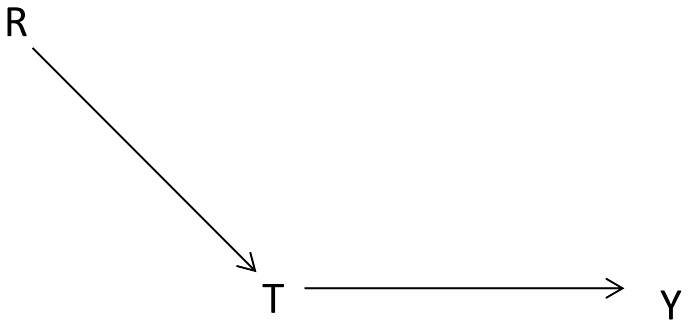
Berk, R. A. et al. (1980) "Crime and Poverty: Some Experimental Evidence from Ex-Offenders." *ASR.* 45:766-86;

Replies and discussion in *AJS*, 88, 2, 1982, 378-396.

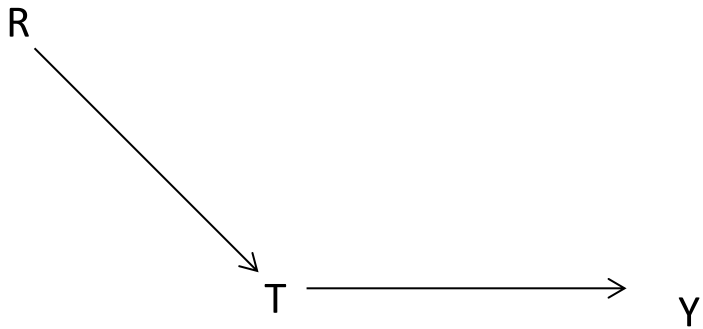
UNIVERSITY OF OXFORD

Instrumental variable

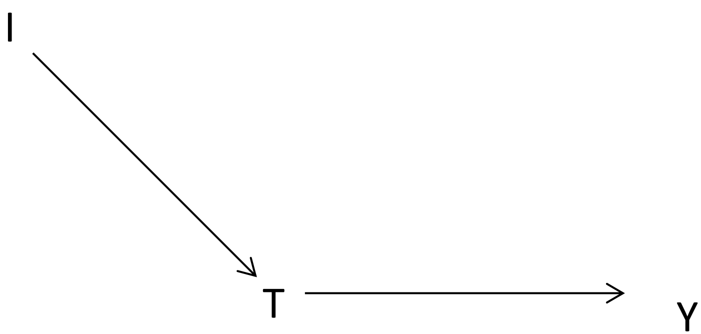
Source of exogenous variation in T



Classic randomized experiment



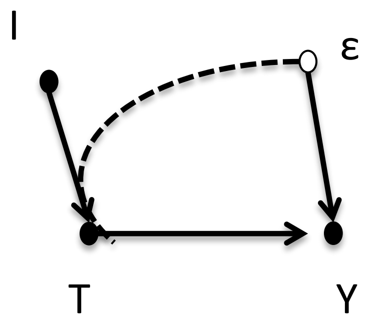
“Natural” experiment – draft lottery



Month of birth as instrument for years of education in estimating returns to education?

Instrumental variable

What's the problem?



1. T is suspected correlated with ε because of:

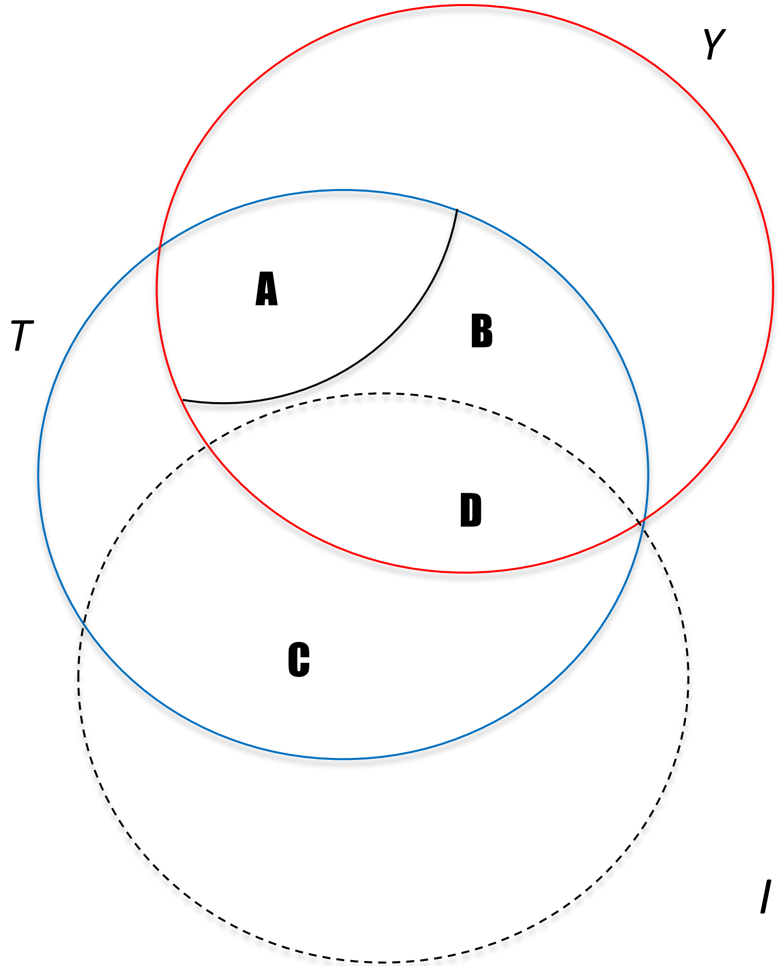
    1. Measurement error in T;
    2. Self-selection into T based on unobservables;
    3. Missing observables that should be conditioned on.

2. If we had complete and error free measurement there would be no problem…

Instrumental variable

How does it work?

1. Red circle = variation in *Y*.
2. Blue circle = variation in *T*.
3. Dashed circle = variation in *I*.
4. **A** = variation in *T* shared with ε.
5. **B** = variation in *T* shared with *Y* but not shared with ε or I.
6. **C** = variation in *T* shared with *I* but not shared with *Y*
7. **D = variation in *T* shared with *Y* and *I*.**



Instrumental variable

How does it work?

1. Needed:

    1. A variable I which is:
        1. Correlated with T;
        2. Not correlated with ε;
        3. Only affects Y through T (which is another way of putting 2.).

2. If I exists then estimate:

$$T_i = \alpha + \beta I_i + \varepsilon_i$$

$$Y_i = \alpha + \beta \widehat{T}_i + \varepsilon_i$$

$$\beta_{IV} = \beta_{YI} / \beta_{TI}$$

Stata do file walk-through at: http://users.ox.ac.uk/~sfos0015/iv_estimation.do

Instrumental variable

Examples

Becker, S. O. and L. Woesmann (2009) ‘Was Weber Wrong? A Human Capital Theory of Protestant Economic History’, *Quarterly Journal of Economics*, May: 531-596.

Martin, G. J. and A. Yurukoglu (2017) ‘Bias in Cable News: Persuasion and Polarization’, *American Economic Review*, 107(9): 2565-2599.

Unobservables

Selection models

1. Oxford University admission tutors wish to estimate the relationship between the entrance examination mark of u/g applicants and average grade in Finals.

2. OU gets about 19000 applications a year to read for BA/BSc degrees.

3. About 1 in 6 are admitted.

4. The good (or bad) fairy has arranged the world so that FHS grades are generated by:

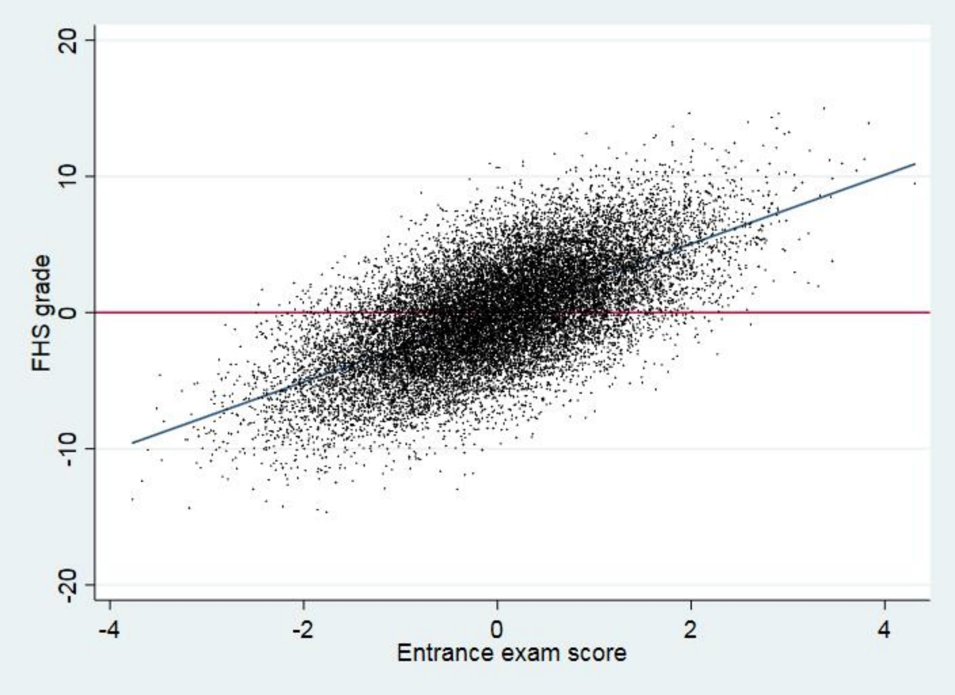
   $FHS\_grade^* = 2.5 \cdot ent\_exam + 3 \cdot u\_1$

   where *ent_exam* and u_1 are random draws from N(0,1).

Unobservables

Selection models

1. This year the relationship between *FHS_grade** and *ent_exam*, if OU admissions tutors were able to observe it, would look like this:



$\hat{\beta}$=2.536

2. But this is **not** what the admission tutors **observe**.

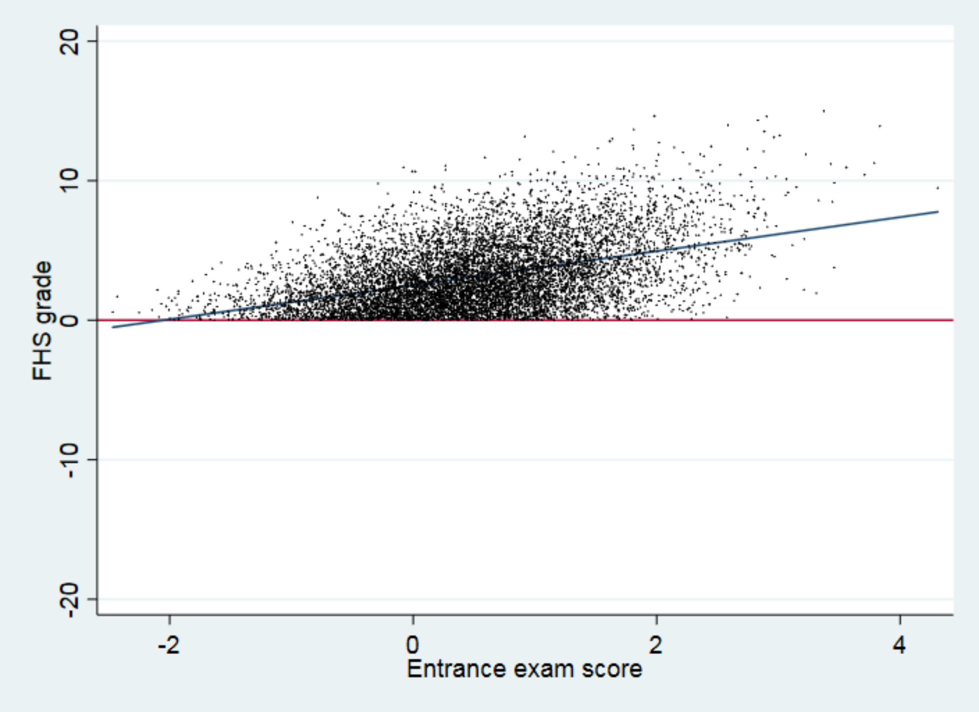
3. Imagine they are blessed with foresight. They never admit anyone who would get less than the average FHS grade*.

4. Then they only admit applicants who will score above the red line.

Unobservables

Selection models

1. Admission tutors don't observe *FHS_grade**. Instead they observe *FHS_grade* which is the grade recorded for those admitted.



$\hat{\beta}$=1.221

2. They begin to feel nervous. Have their gifts deserted them?

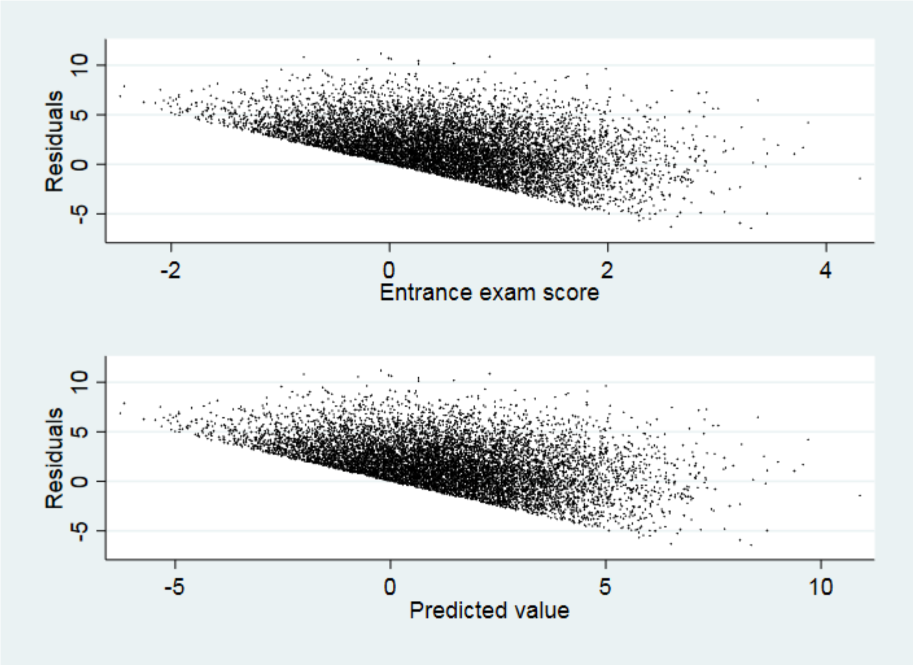
3. Student radicals urge them to abandon using the entrance exam because it is a poor predictor of FHS grade.

4. What has happened?

Unobservables

Selection models

Explicit selection



1. E(u_1| ent_exam) ≠ 0.
2. E(u_1| y_hat) ≠ 0
3. u_1 in the population is not correlated with *ent_exam* but in the selected sample it is.
4. Students with relatively low *ent_exam* scores have (on average) relatively large positive residuals.
5. Perhaps they are charming and know how to get the best teaching out of their tutors.
6. But charm is unobserved and uncorrelated with *ent_exam* in the parent population.
7. Selection has given us an observed population that is more charming than is typical of the parent population.
8. Is there anything we can do?

Unobservables

Selection models

Consider a more realistic censoring mechanism.

$Z = -5 + 2 \cdot \mathit{ent\_exam} + 2 \cdot IQ + 3 \cdot u\_2$

$FHS\_grade^* = 2.5 \cdot ent\_exam + 3 \cdot u\_1$

Where: *FHS_grade** is observed only if $Z > 0$

Assume that u_1 and u_2 are correlated 0.5. This amounts to assuming that in the parent population the unobserved things that are predictive of being selected are correlated with the unobserved things that predict higher marks in FHS. This doesn't seem completely unreasonable. Perhaps charm, as well as getting you better teaching also increases your chances of doing well in the entrance interview.

Those who are selected get better FHS grades than those who were not if they (counterfactually) had been selected.

Unobservables

Heckman's selection model

1. What we have is really a missing variables problem. If we could control for missing variables then we would have a solution.

2. In this case it turns out that we do actually have some information about the missing variables if we are prepared to assume that u_1 and u_2 are correlated.

3. We can use that information to correct the estimate that we get from the selected sample.

4. This was James Heckman's insight.

Unobservables

Heckman’s recipe

1. Estimate a prediction equation with a probit model for the probability of being selected into the observed sample: prob (y = 1) F(ent_exam, IQ).

2. Use the predicted values from 1. to construct something (the inverse Mills-ratio) that represents the omitted variables.

3. Estimate the equation for FHS_grade* **using just the selected sample** as: FHS_grade* = F(ent_exam, inverse Mills-ratio).

4. To see how this works with some simulated data see the Stata do file at:

   http://users.ox.ac.uk/~sfos0015/Heckman_simulation.do

```
*** estimate FHS_grade* ent_exam score regression for parent population
. reg FHS_grade ent_exam

      Source |       SS           df       MS      Number of obs   =    19,000
-------------+----------------------------------   F(1, 18998)     =    971.07
       Model |  8905.91281         1  8905.91281   Prob > F        =    0.0000
    Residual |  174234.467    18,998  9.17120048   R-squared       =    0.0486
-------------+----------------------------------   Adj R-squared   =    0.0486
       Total |   183140.38    18,999  9.63947469   Root MSE        =    3.0284

------------------------------------------------------------------------------
   FHS_score |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ent_exam |   2.377136    .076283    31.16   0.000     2.227614    2.526657
       _cons |   .0844771   .0439611     1.92   0.055    -.0016905    .1706447
------------------------------------------------------------------------------
```

```
*** estimate FHS ent_exam score regression for selected cases
.
. reg y1 ent_exam

      Source |       SS           df       MS      Number of obs   =     3,179
-------------+----------------------------------   F(1, 3177)      =     48.35
       Model |  361.618669         1  361.618669   Prob > F        =    0.0000
    Residual |   23760.315     3,177  7.47885269   R-squared       =    0.0150
-------------+----------------------------------   Adj R-squared   =    0.0147
       Total |  24121.9337     3,178   7.5902875   Root MSE        =    2.7347

------------------------------------------------------------------------------
          y1|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ent_exam |   1.197908   .1722723     6.95   0.000     .8601318    1.535684
       _cons |   2.916018   .1104772    26.39   0.000     2.699404    3.132632
------------------------------------------------------------------------------
```

```
reg y1 ent_exam invmills

      Source |       SS           df       MS      Number of obs   =     3,179
-------------+----------------------------------   F(2, 3176)      =     36.98
       Model |  548.917151         2  274.458575   Prob > F        =    0.0000
    Residual |  23573.0165     3,176  7.42223442   R-squared       =    0.0228
-------------+----------------------------------   Adj R-squared   =    0.0221
       Total |  24121.9337     3,178   7.5902875   Root MSE        =    2.7244

------------------------------------------------------------------------------
          y1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ent_exam |   1.910755   .2226881     8.58   0.000     1.474128    2.347382
    invmills |   1.490454   .2967008     5.02   0.000     .9087091    2.072198
       _cons |   .3583128    .520915     0.69   0.492    -.663051     1.379677
------------------------------------------------------------------------------
```

Pretest-Postest Nonequivalent Control Group Design

GROUP 1 Y_1 X Y_2

\- -

GROUP 2 Y_3 Y_4

\- - - - - - - - = non random allocation to groups

Pretest-Postest Nonequivalent Control Group Design

Example 1

1. A large management consultancy organizes work on the basis of “project teams” with a team leader.

2. Senior partners are worried about morale level (as reflected in absenteeism rate).

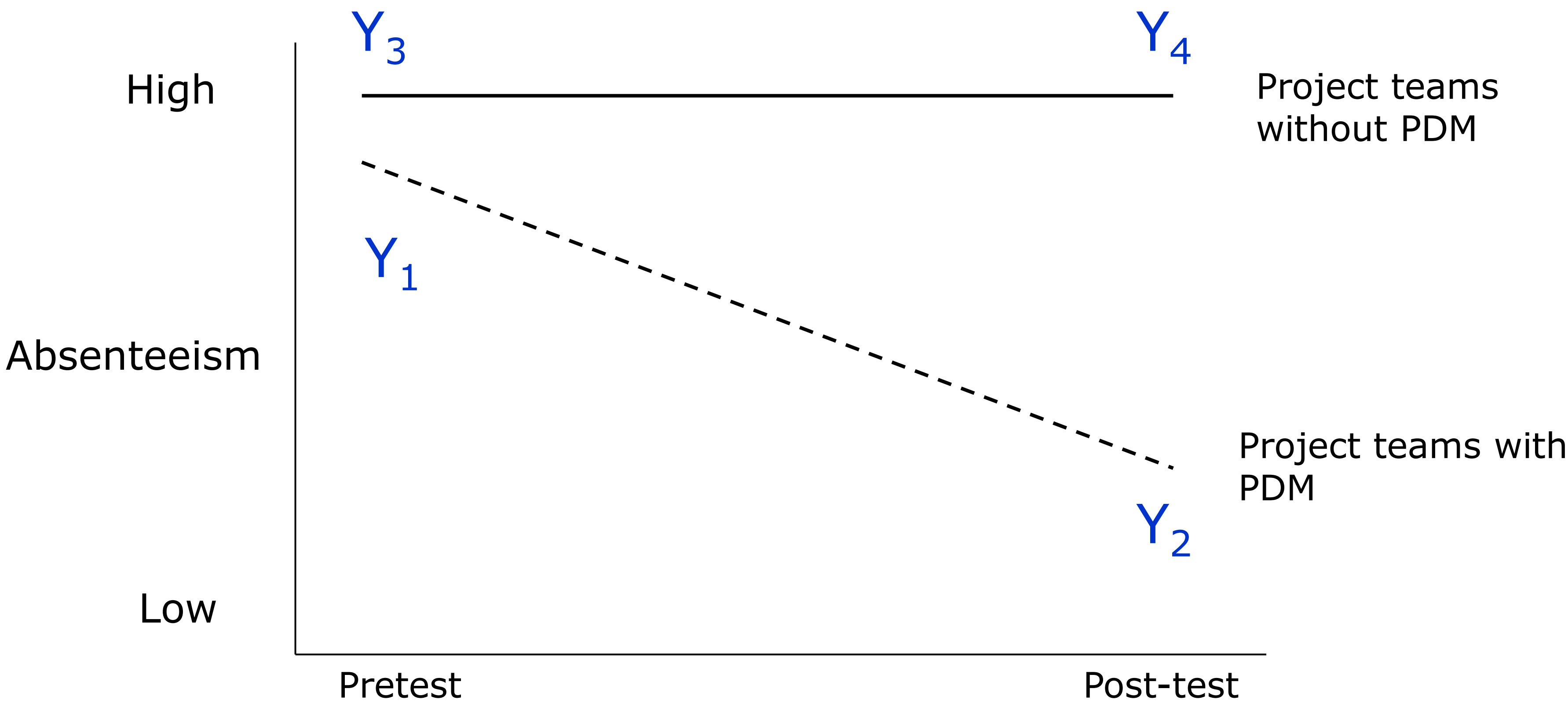
3. Team leaders are allowed to adopt (if they wish) more participatory ways of making decisions.

4. Some do and others don’t.

5. Do participatory teams do better than others?

Pretest-Postest Nonequivalent Control Group Design

Example 1



UNIVERSITY OF OXFORD

Pretest-Postest Nonequivalent Control Group Design

Example 1

1. PDM adopters already have lower absenteeism rates.

2. Difference becomes bigger after adoption of PDM.

    1. But PDM adopters might already be on a downward trend (perhaps team leaders are more easy going) .

    2. aka **selection by maturation** threat.
        1. Can’t be ruled out without more pretest observations.

3. Conclusion depends on the plausibility of the selection by maturation threat **in this particular case.**

Pretest-Postest Nonequivalent Control Group Design

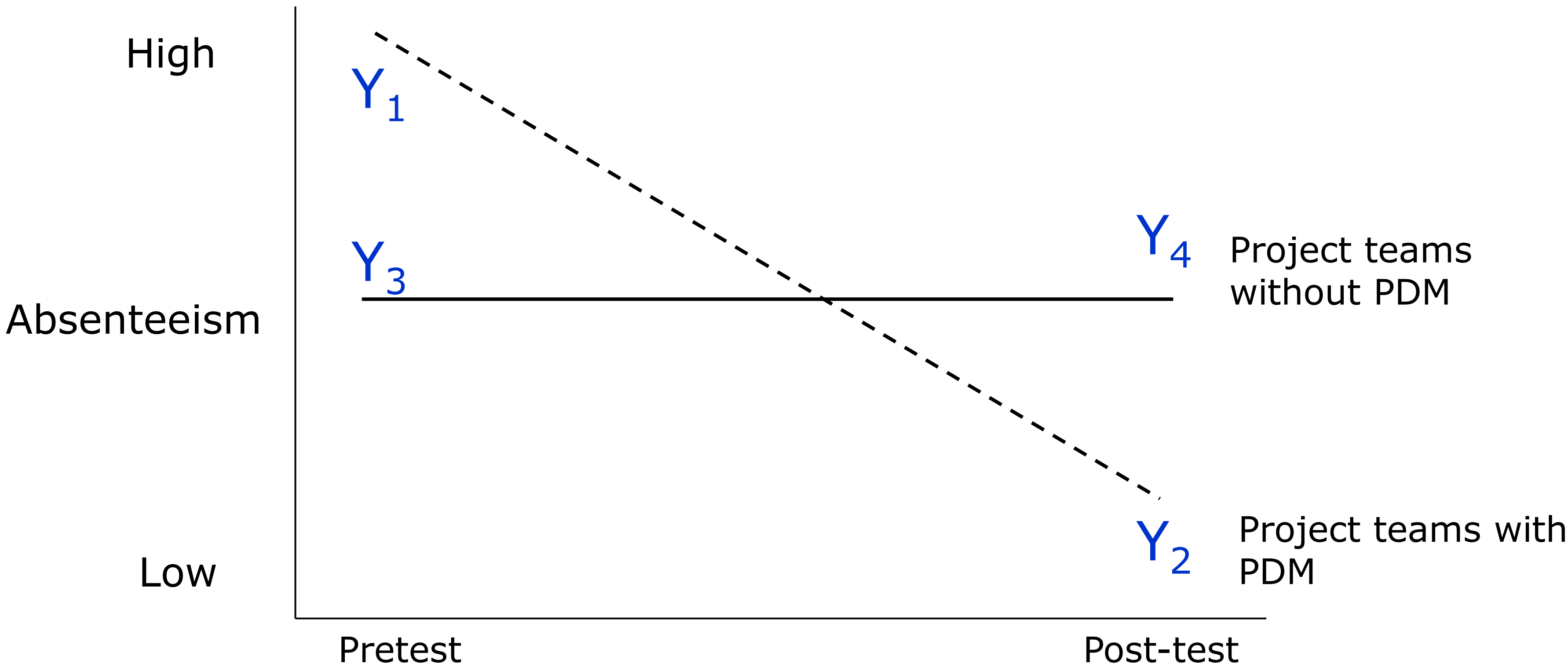
Example 2

1. Imagine a different set up.

2. PDM is imposed by the senior partners on the project teams with the highest absenteeism rates.

Pretest-Postest Nonequivalent Control Group Design

Example 2



UNIVERSITY OF OXFORD

Pretest-Postest Nonequivalent Control Group Design

Example 2

1. Genuine “treatment” effect looks slightly more plausible.
2. Why should differential “maturation” lead to a crossover?
3. Why should “regression to the mean” lead to crossover?
4. Effect of PDM still not “proven” but case looks a little stronger.

Regression-Discontinuity Design

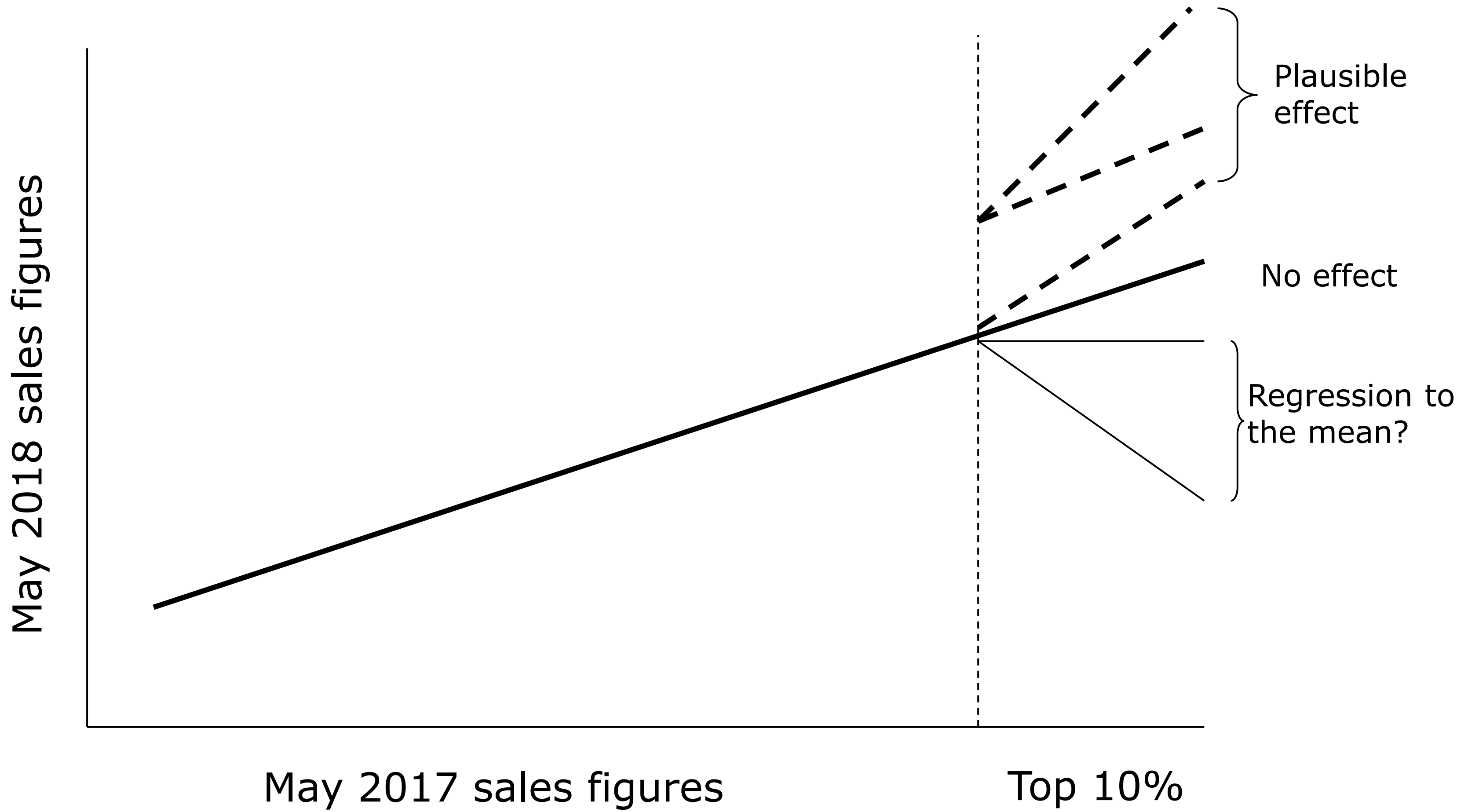
1. If selection into treatment is based in a known way on pre-test score then R-D design possible.

2. Say top 10% of a sales force are given a bonus over and above their commission.

3. Does it affect their performance?

UNIVERSITY OF OXFORD

Regression-Discontinuity Design



Interrupted Time-Series Design

If you can't control who is exposed to the treatment…

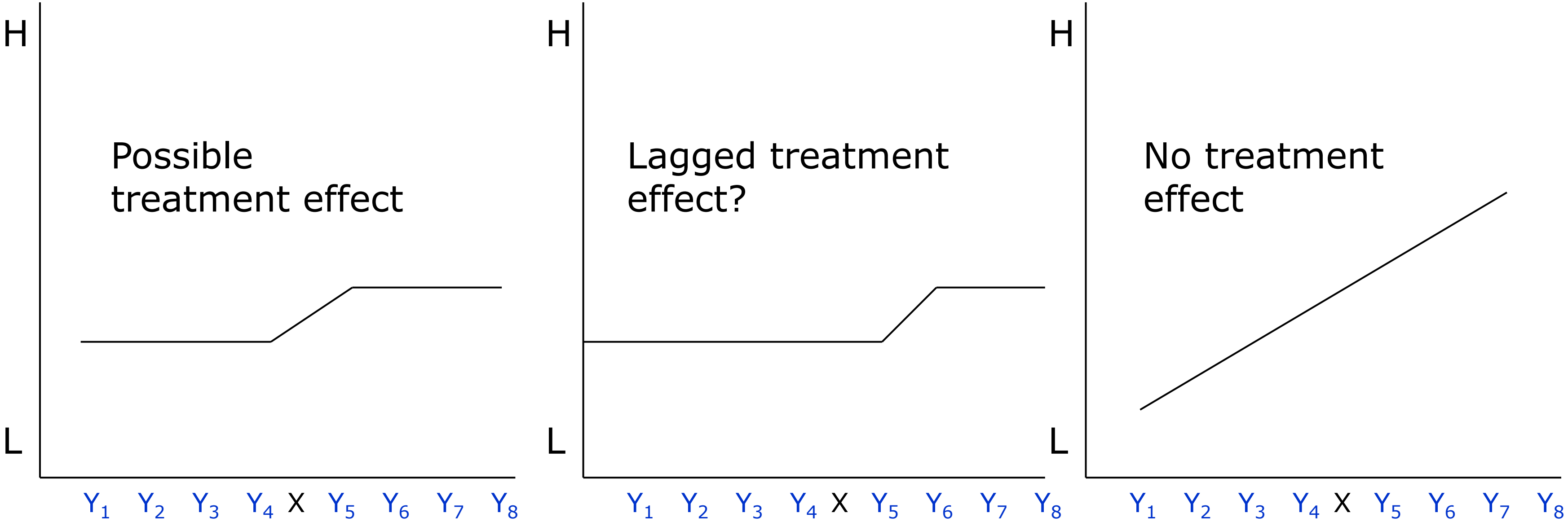
Try to control **when** observations are made

GROUP 1 Y_1 Y_2 Y_3 Y_4 X Y_5 Y_6 Y_7 Y_8

Even spacing is nice

Can rule out maturation by looking at the trend in the pre-tests

Interrupted Time Series Design



Effects of causes AND causes of effects

1. In the context of a scientific research programme there is no conflict;
2. Both ways of looking at and for causes will be useful.
3. Rather than conflict there is complementarity.
4. At each stage though you still need to have a clear view of your scientific goal!