Study designs and statistical methods for current observational studies

2. Instrumental variable analysis

27. Mar. 2024

Kazuhiro Abe

Project Assistant Professor,
Graduate School of Medicine,
The University of Tokyo

Study designs and statistical analysis in current observational studies

1. Difference-in-Differences
2. Instrumental Variable Analysis
3. Regression Discontinuity

Natural experimental methods (Quasi-experimental methods)

4. Panel Data Analysis / Interrupted Time-Series

5. Propensity Score Analysis (matching, weighting, stratification, and adjustment using propensity scores)
6. Adjustment(regression), weighting, stratification, and matching

Recap: Difference-in-Differences design

Key assumptions

1. **Parallel trends assumption**

If there is no intervention, the outcomes of the treatment and control group would be a parallel trend.

Demonstrate that the outcomes of the treatment and control groups are parallel at several time points before the intervention.

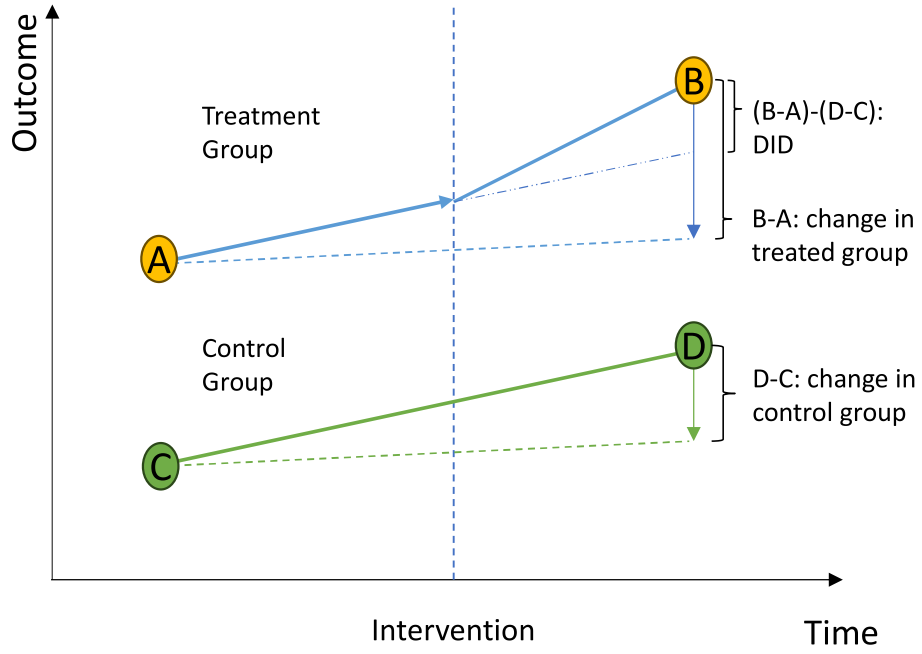
2. **Common shocks assumption**

There are no other exogenous events occurring that affect pre- and post-intervention outcomes, or if they occur, they affect the two groups in the same way.

DID

= (change in treatment) – (change in control)

= ($Treat^{After}$ - $Treat^{Before}$) – ($Control^{After}$ - $Control^{Before}$)



	Treatment	**Control**
Before	A	C
After	B	D
Change	B-A	D-C
DID	(B-A) – (D-C)	

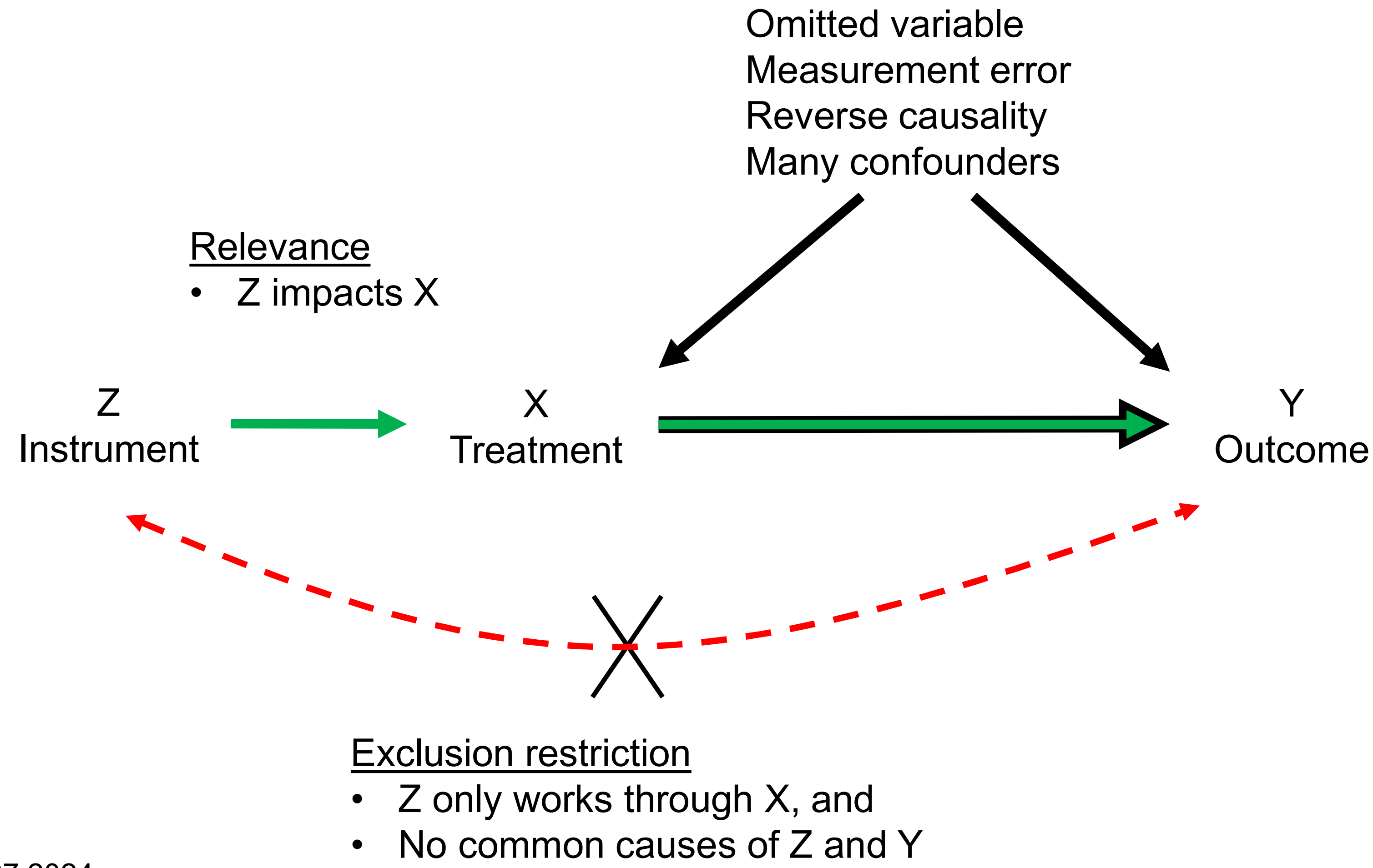
Instrumental Variable Analysis (IV)

Textbook

Wooldridge MJ. Introductory Econometrics: A Modern Approach. 2019

K. Abe, 3.27.2024

The IV idea



K. Abe, 3.27.2024

Key IV assumptions

Relevance

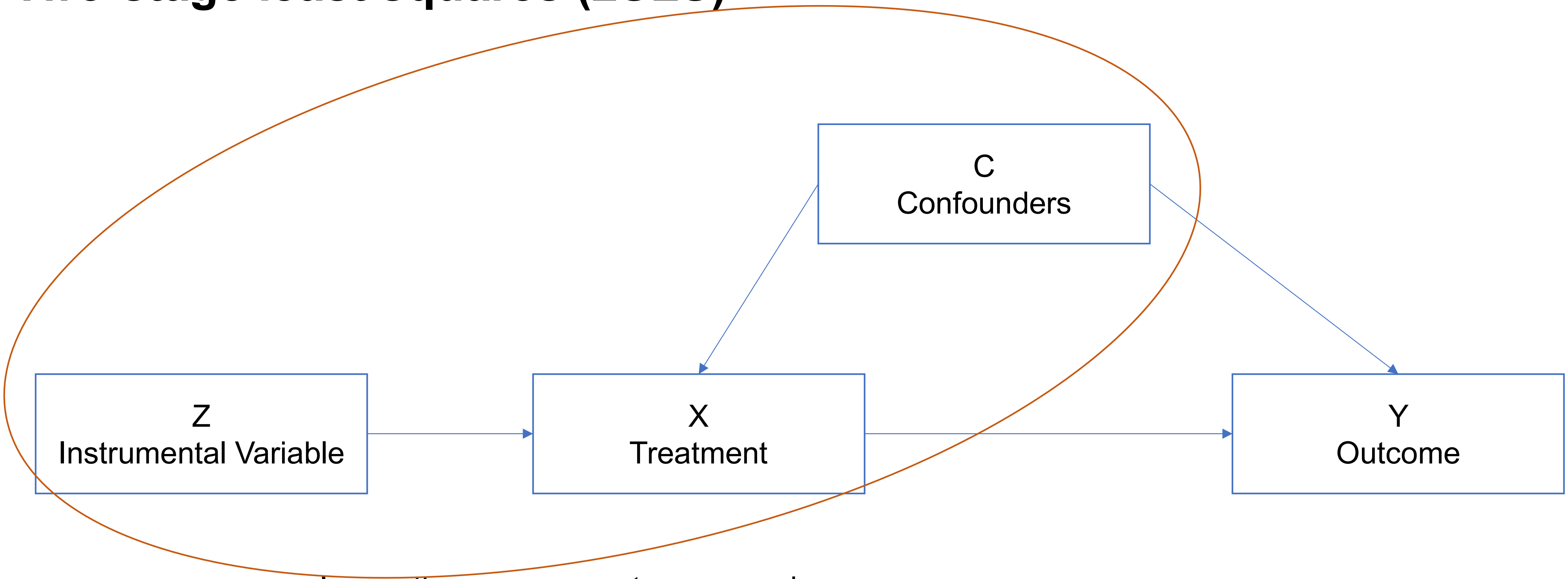
- The instrument Z is “strongly” correlated with the treatment status X: $\mathrm{cov}(X_i, Z_i) \neq 0$
- That is, **Z explains a significant part of the variation in treatment variable X** *once we control for all other factors in our empirical model*

Exclusion restriction

- Z is not correlated with the error term in the main regression: $\mathrm{cov}(Z_i, \varepsilon_i) = 0$

- The instrument must <u>not</u> have a direct causal effect on Y, and can also not be correlated with any other unobserved determinant of Y: it must be orthogonal to the residual
- Said differently: **the instrument Z affects outcome Y <u>only</u> through the relation of Z on X**
- Because of this, we can legitimately exclude the instrument from the regression

Two-stage least squares (2SLS)

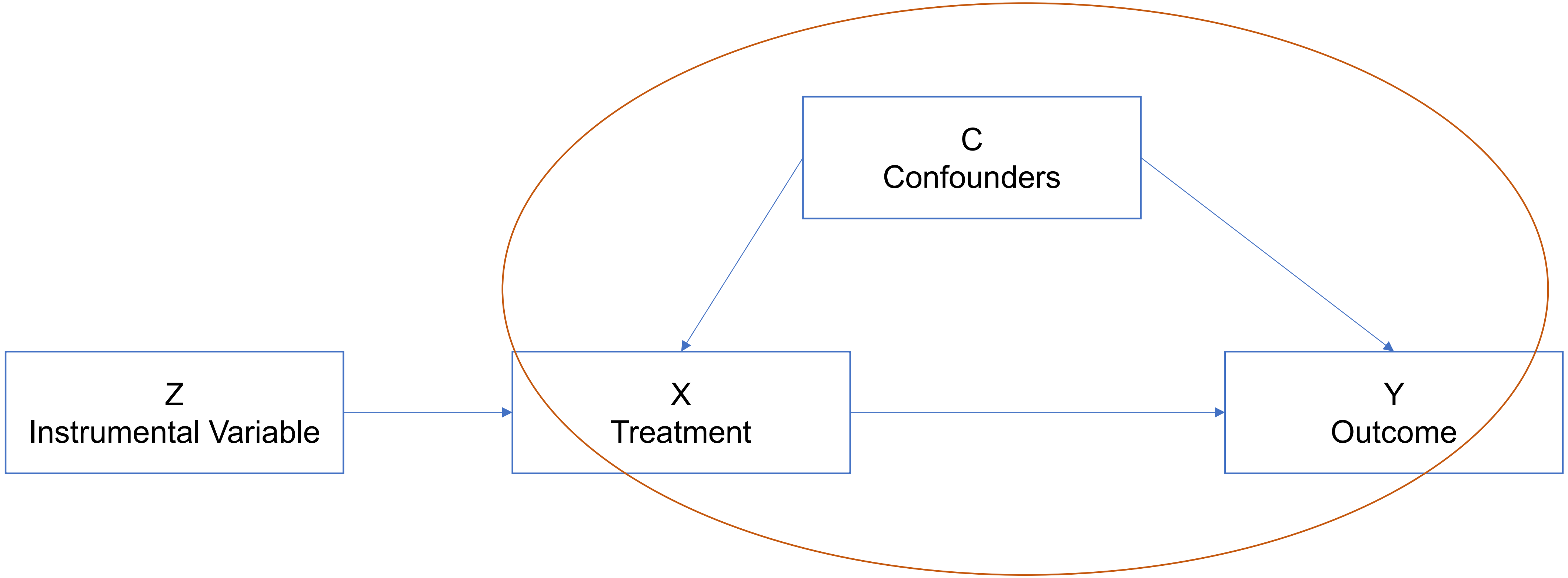


- In practice we can use two regressions :

First-stage: estimate effect of instrument Z on treatment X $X_i = \theta + \gamma Z_i + u_i$

Use this to predict the treatment status $\hat{X}_i$ for different values of Z

Two-stage least squares (2SLS)



- Second-stage: estimate effect of predicted status $\hat{X}_i$ on outcome Y

$$Y_i = \pi + \mu \hat{X}_i + \eta_i$$

K. Abe, 3.27.2024

Continued only for students