MBA 753 : Causal Inference Methods for Business Analytics

Dr. Nivedita Bhaktha

22.08.2024

Agenda

- Matching
- Event Studies

Matching

Motivation

- Randomization aids causal inference because in expectation it balances observed & unobserved confounders
- When we cannot randomize, we can design studies to capture the central strength of randomized experiments:
  - have treatment & control groups that are as comparable as possible
  - i.e. we can try to control for observed covariates

- Assume treatment is not randomized, but is independent of potential outcomes so long as other factors are held fixed
  - We are assuming that among units with the same values for some covariate X (i.e. conditional on X), the treatment is "as good as randomly" assigned

Example

- Does teacher training improve university applications?
  - Imagine that some school teachers take specialist training in how to prepare their students for university applications. Teachers select into the training program (i.e. they are not randomly assigned). You believe, however, that conditional on the type of school in which a teacher teaches, training is as good as random.

Y_i: Number of students applying for top universities
D_i: 1 if the teacher did the training, 0 otherwise
X_i: Whether the teacher is at a state, private, or public school

Example – is DIGM an unbiased estimator of ATE?

X_i, D_i joint distribution

	$D_i = 0$	$D_i = 1$
X_i = State	0.30	0.05
X_i = Private	0.15	0.15
X_i = Public	0.05	0.30

Mean outcomes

	$D_i = 0$	$D_i = 1$
X_i = State	0	2
X_i = Private	3	4
X_i = Public	5	5

Y_{0i} Y_{1i}

counterfactual → $\times\ Y_{1i}$ $\times\ Y_{0i}$

τ_i

$2 \times 0.05 - 0 \times 0.30$

$E(\tau_i) = ATE$

$$\text{DIGM} \equiv E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

$$= \frac{(0.05 \times 2 + 0.15 \times 4 + 0.3 \times 5)}{\frac{1}{2}} - \frac{(0.3 \times 0 + 0.15 \times 3 + 0.05 \times 5)}{\frac{1}{2}}$$

$$= 3 \rightarrow ATT + \underset{\overset{\text{"}}{0}}{\text{selection bias}} \rightarrow E(Y_{0i} | D_i = 1) - E(Y_{0i} | D_i = 0)$$

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

T_i	D_i	X_i	Y_i
T_1	1	S	
	1	Pr	
		Pu	
T_n	1		
T_{n+1}	0	S	
		Pr	
T_m	0	Pu	

Matching

- One way to deal with counterfactuals – treat it as missing data problem
- Matching: imputing missing outcomes
- For each unit i, find the "closest" unit j with opposite treatment status and impute j's outcome as the unobserved potential outcome for i

$$\tau_{ATT} = \frac{1}{n}\sum_{i} Y_i - Y_{j(i)} \quad or \quad \frac{1}{n}\{\sum Y_i - \frac{1}{m}\sum_{m=1}^{M} Y_{j_m\ (i)}\}$$

Types of Matching Methods

- **Nearest Neighbour Matching**
  - C T
  - M:1 nearest neighbor matching 4:1
  - it matches control individuals to the treated group and discards controls who are not selected as matches
- Full Matching
- Subclassification

Example

- Do UN interventions Cause Peace?
  - Gilligan and Sergenti (2008) investigate whether UN peacekeeping operations have a causal effect on building sustainable peace after civil wars. They study 87 post-Cold-War conflicts, and evaluate whether peace lasts longer after conflict in 19 situations in which the UN had a peacekeeping mission compared to 68 situations where it did not.

19^{T} $\quad$ $\overset{C}{68}$

Y_i: Peace duration (measured in months)
D_i: 1 if the UN intervened post-conflict, 0 otherwise
$X_{1,i}$: Region of conflict (categorical)
$X_{2,i}$: Democracy in the past (binary, based on polity)
$X_{3,i}$: Ethnic Fractionalization (continuous)

Example source: Lecy & Fusi, ds4ps.org

Example – Matching M = 1 with replacement

without replacement

1:1

NN 1:1 Matching

Country	D	EthFrac	Region	Y_{0i}	Y_{1i}
Liberia	1	83	SS Africa	?3	51
Sierra Leone	1	77	SS Africa	?11	35
Zaire	1	90	SS Africa	?3	23
Chad	0	83	SS Africa	3	
Senegal	0	72	SS Africa	11	
Niger	0	73	SS Africa	11	

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

51 - 3

35 - 11

23 - 3

What is the $\hat{\tau}_{ATT}$?

$$\begin{aligned}\hat{\tau}_{\text{ATT}} &= \frac{1}{N_1}\sum_{D_i=1}(Y_i - Y_{j(i)}) \\ &= (51-3)\times 1/3 + (35-11)\times 1/3 + (23-3)\times 1/3 \\ &= 30.7\end{aligned}$$

2:1

Example – Matching M = 2 with replacement

1 → 4, 6
2 → 5, 6
3 → 4, 6

NN 2:1 Matching

	Country	D	EthFrac	Region	Y_{0i}	Y_{1i}
1	Liberia	1	83	SS Africa	?7	51
2	Sierra Leone	1	77	SS Africa	?11	35
3	Zaire	1	90	SS Africa	?7	23
4	Chad	0	83	SS Africa	3	
5	Senegal	0	72	SS Africa	11	
6	Niger	0	73	SS Africa	11	

ATT : τ_i

51 - 7

35 - 11

23 - 7

What is the $\hat{\tau}_{ATT}$?

$$\begin{aligned}\hat{\tau}_{\text{ATT}} &= \frac{1}{N_1}\sum_{D_i=1}\left(Y_i - \frac{1}{M}\sum_{m=1}^{M} Y_{j_m(i)}\right) \\ &= (51-7)\times 1/3 + (35-11)\times 1/3 + (23-7)\times 1/3 \\ &= 28\end{aligned}$$

Matching – Many questions

- If we are selecting matches, how many?
  - One best match
  - K best matches
  - All acceptable matches
- Matching with or without replacement?
  - Without replacement - running out of controls
  - With replacement - use the same control multiple times, giving it a weight equal to the number of times it has matched
- How do we use weights?
  - Constructing weights for observations

Matching – Bias-Variance Trade-off

- How many – few or more
  - Few matches – better matches – less bias
  - More matches – less sampling variation – lower standard errors
- With/out replacement –
  - With – reduces bias – each control has more influence – sampling variation is larger
- Weighting –
  - Weight each observation separately
    - Kernel matching
    - Inverse probability weighting

Matching on more variables

- Commonly we will want to match on many X variables, not just one or two
- Adding more covariates creates a problem - We have to define how we measure whether two units are “close” to one another

Treated case:

- Haiti, with polity = -6, region = Latin America, and **ethfrac = 1**

Control cases:

- Panama, with polity = 8, region = Latin America, **ethfrac = 3**
- Egypt, with polity = -7, region = N Africa, **ethfrac = 4**
- El Salvador, with polity = -6, region = Latin America, **ethfrac = 26**

Multiple Matching Variables

- Distance Matching
  - Mahalanobis Distance: distance from the centroid
  - Entropy Balancing: *enforce restrictions* on the distance between treatment and control
  - Curse of dimensionality
    - Limit the number of matching variables
    - Extremely large sample sizes
    - Using caliper or bandwidth for match quality

- Propensity Score Matching

Event Studies

Event Studies

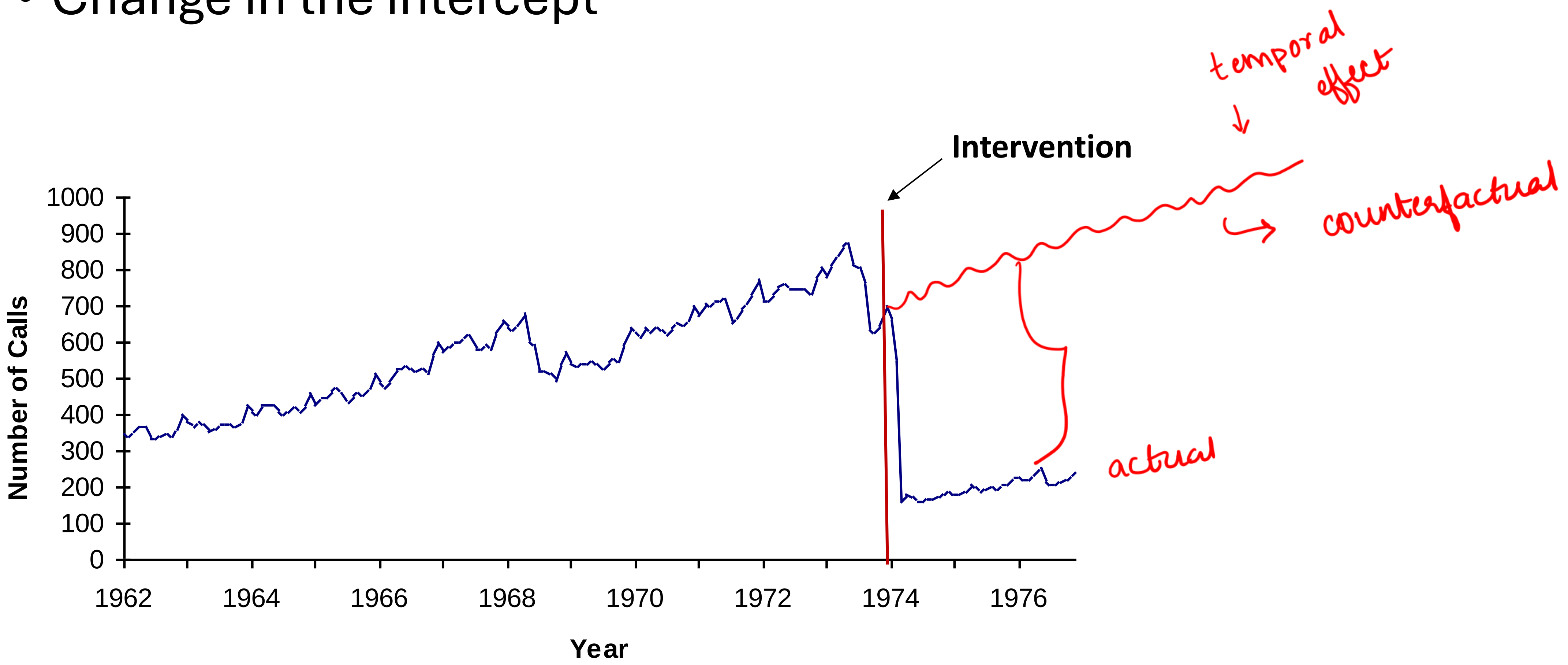
- Also known as interrupted time series design
- Simplest quasi experimental design
  - Treatments that switch from off to on
- Data: Time series data – track an individual across multiple time points
  - Time series collects the variables that changes over time
- We move form before event to after event – a treatment goes into effect

Event Studies Model

- Treatment effect: compare before event and after event
- Caveats: treatment should be the only thing that is changing
  - Remove natural temporal effects – predict counterfactuals
    - Assume whatever was going on before would have continued if not for the treatment
    - Use before event data to predict future without event
  - Deviation of the future without event from the from the actual outcome gives the treatment effect

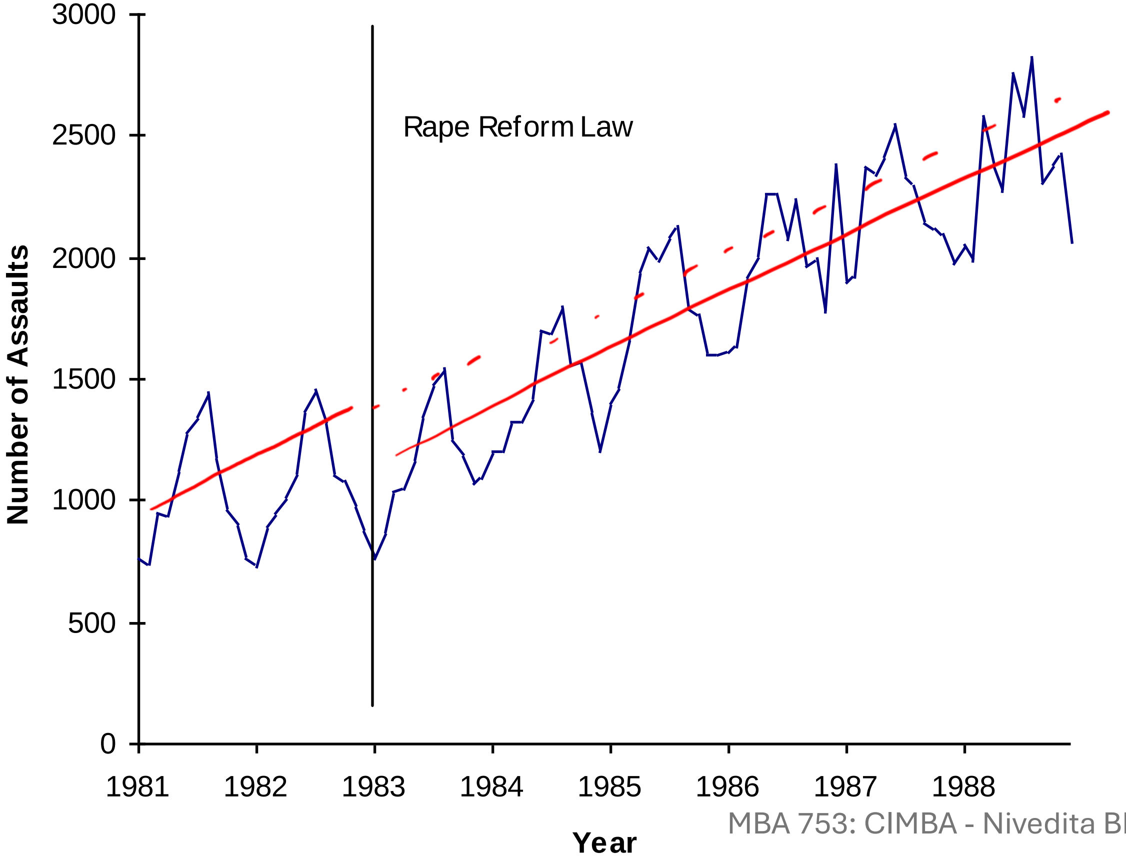
Event Studies Effect

- Change in the intercept



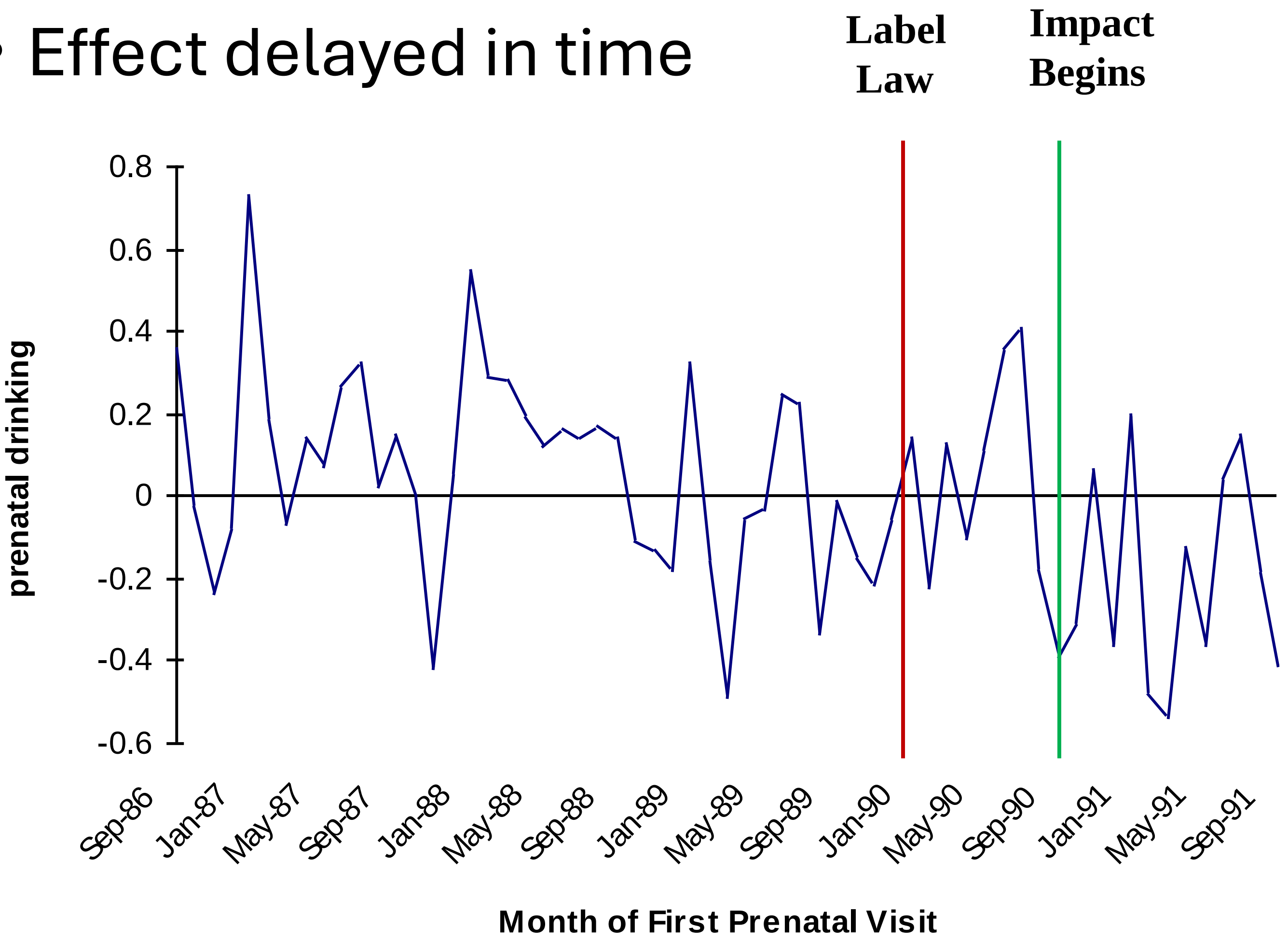
Event Studies Effect

- Change in the slope



Event Studies Effect

- Effect delayed in time



When does Event Studies work well?

- Clear intervention/treatment/test time point
- Huge and immediate effect
- Clear pretest functional form
  - many observations before and after treatment
- No alternative explanation for change

Event Studies - Issues Satisfying Assumptions

- Long span of data is not available
  - pretest functional form is often unclear
  - Counterfactuals associated with short pretest time series is often weak
- Implementing the intervention can span several years
- Instantaneous effects are rare
- Effect sizes are usually small

Event Studies – Regression model

- Also known as segmented regression

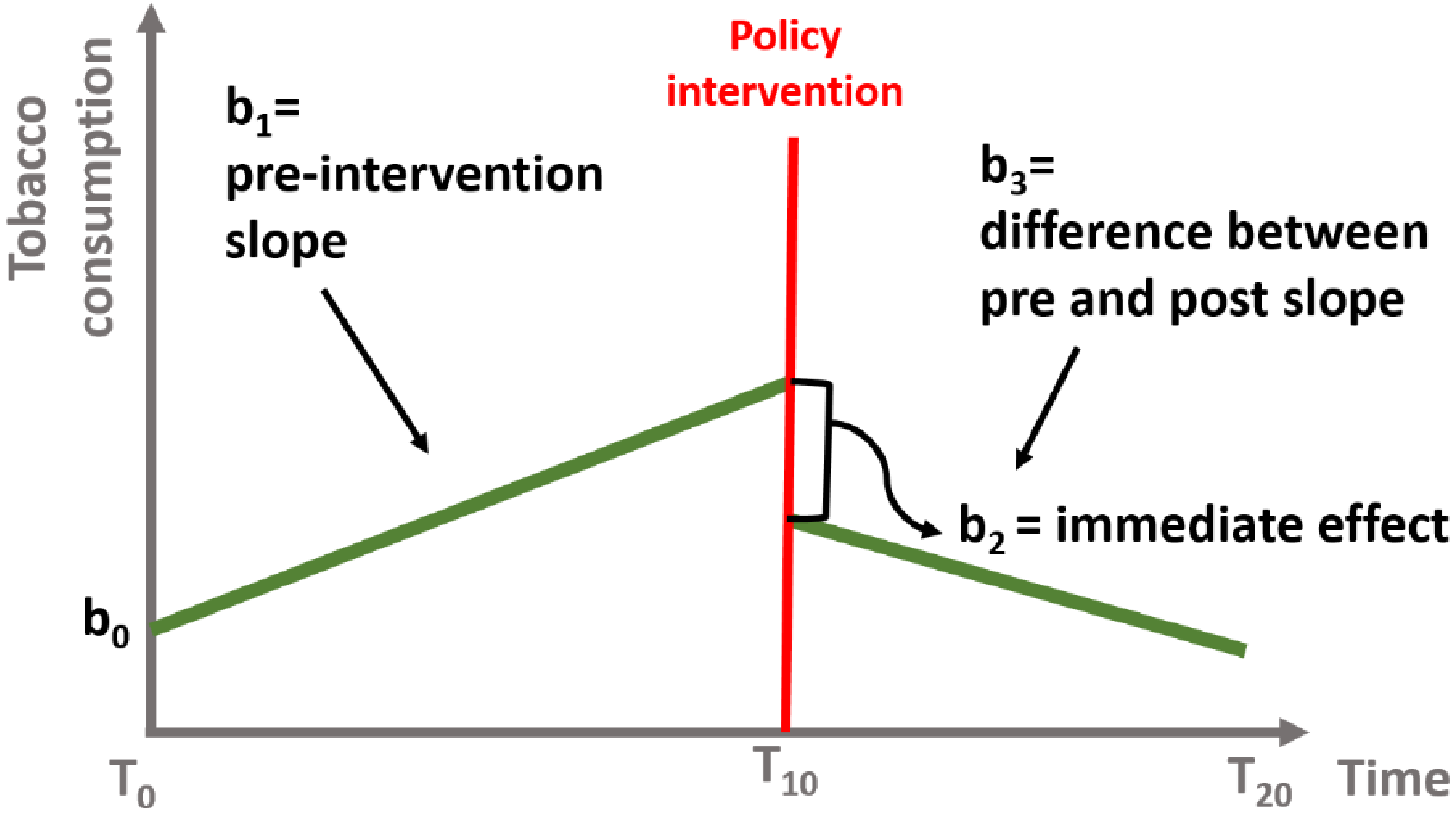
$$Y = \beta_o + \beta_1 time + \beta_2 treatment + \beta_3 time * treatment + \epsilon$$

$\beta_2 = 0$, $\beta_3 = 0$

Where **time * treatment** implies time after interruption

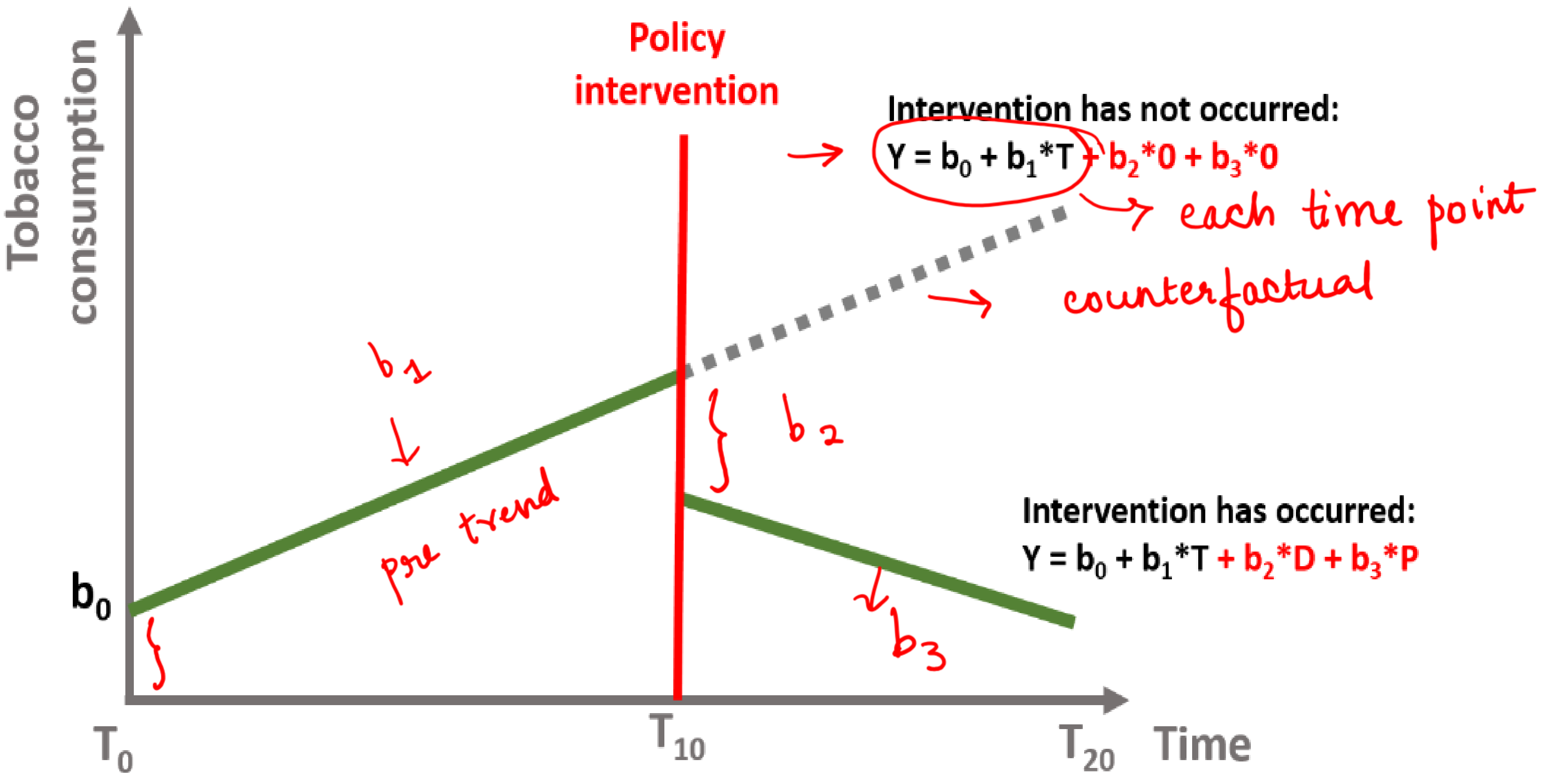
Parameter	Interpretation
β_1	Pre- Trend
β_2	Post- Level Change
β_3	Post- Trend Change
$\beta_1+\beta_3$	Post- Trend

Regression Model



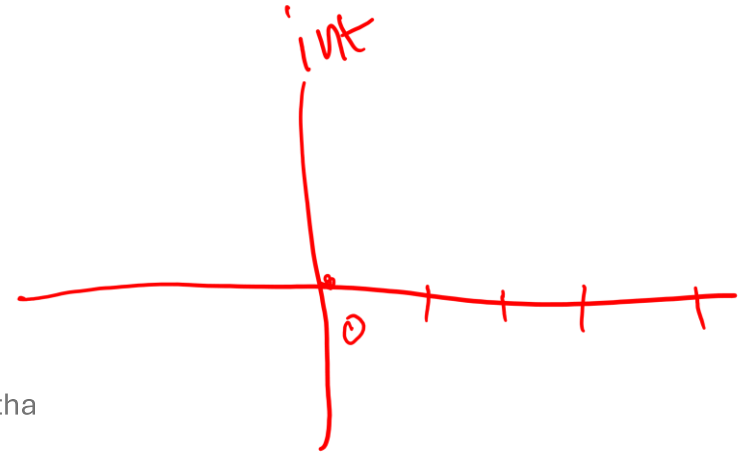
The Counterfactual

- To calculate the counterfactual, we need to assume that the intervention has never occurred
  - there has been no immediate nor sustained effect
- We can calculate the counterfactual for each point in time



Example – wellbeing classes

Column	Variable name	Description
Y	Wellbeing	Wellbeing index (from 0 to 300)
T	Time	Time (from 1 to 365)
D	Treatment	Observation post (=1) and pre (=0) intervention
P	Time Since Treatment	Time passed since the intervention



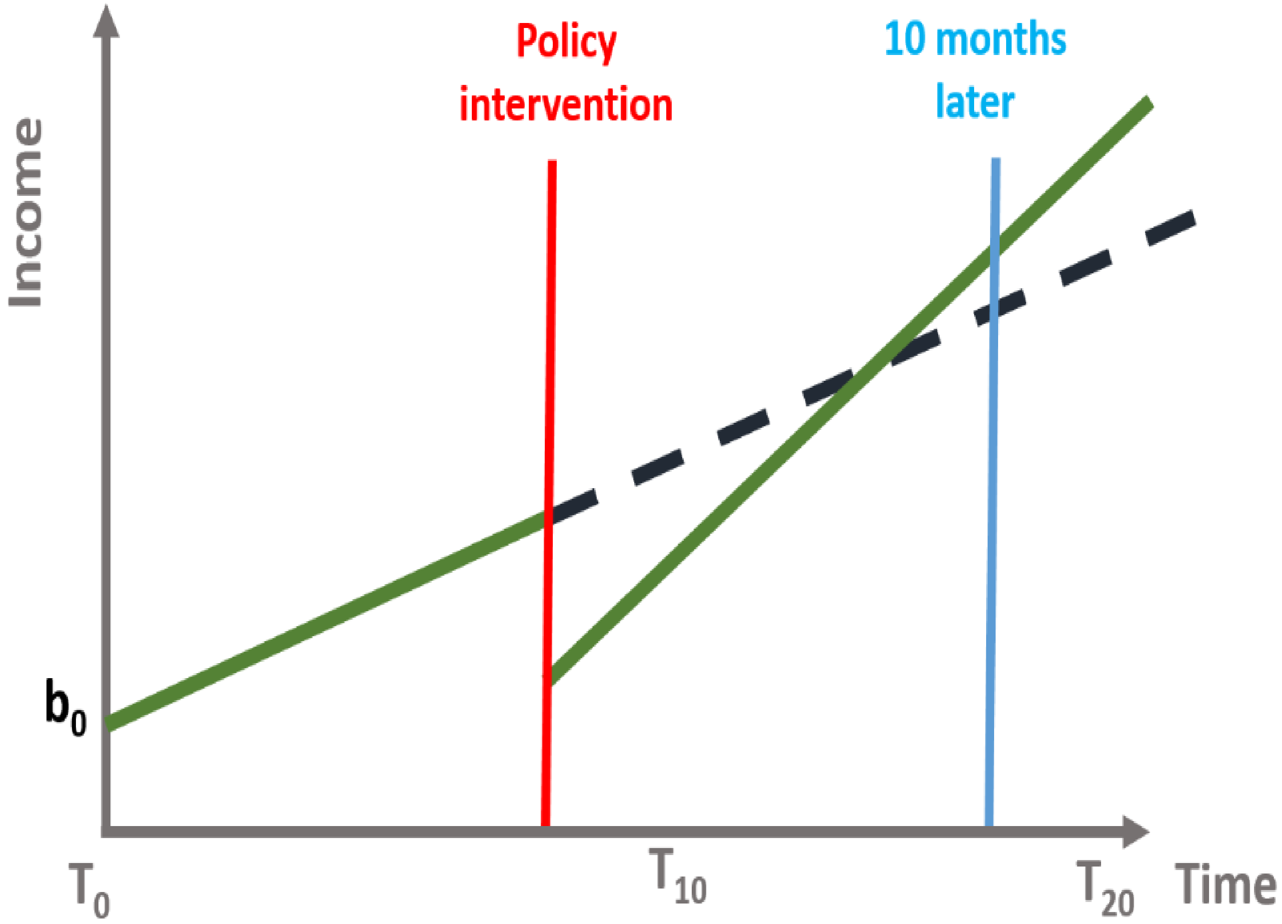
Autocorrelation

- Autocorrelation is a major issue when working with time series
  - Autocorrelation occurs when observation at one point in time depends from observations at another point in time
- OLS assumes that error terms associated with each observation are uncorrelated
  - Violated in presence of autocorrelation
- Impact - underestimated the standard errors
  - Overestimating the statistical significance

- Check residuals plot and Durbin-Watson test

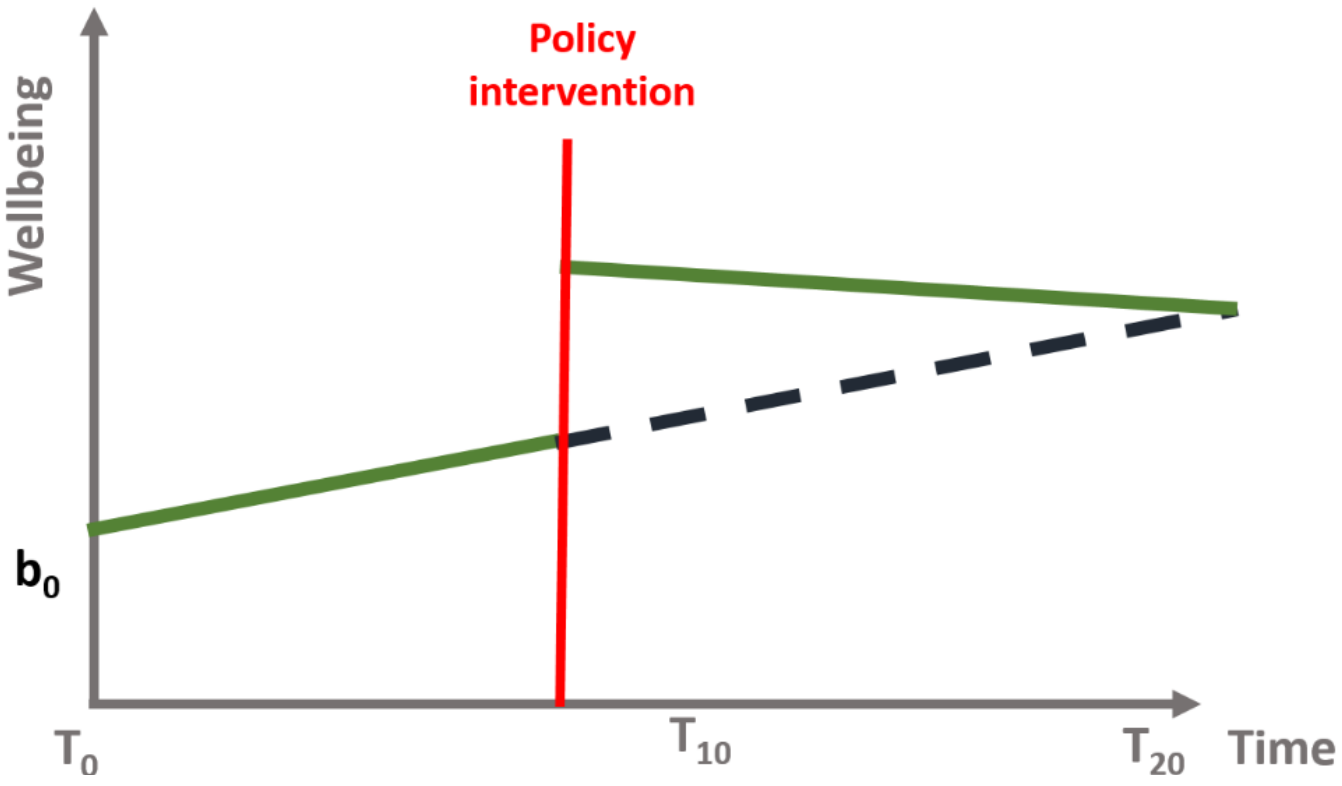
Issues with the design

- Delayed intervention effect
  - Misleading intervention effec
- look at immediate and sustained effects



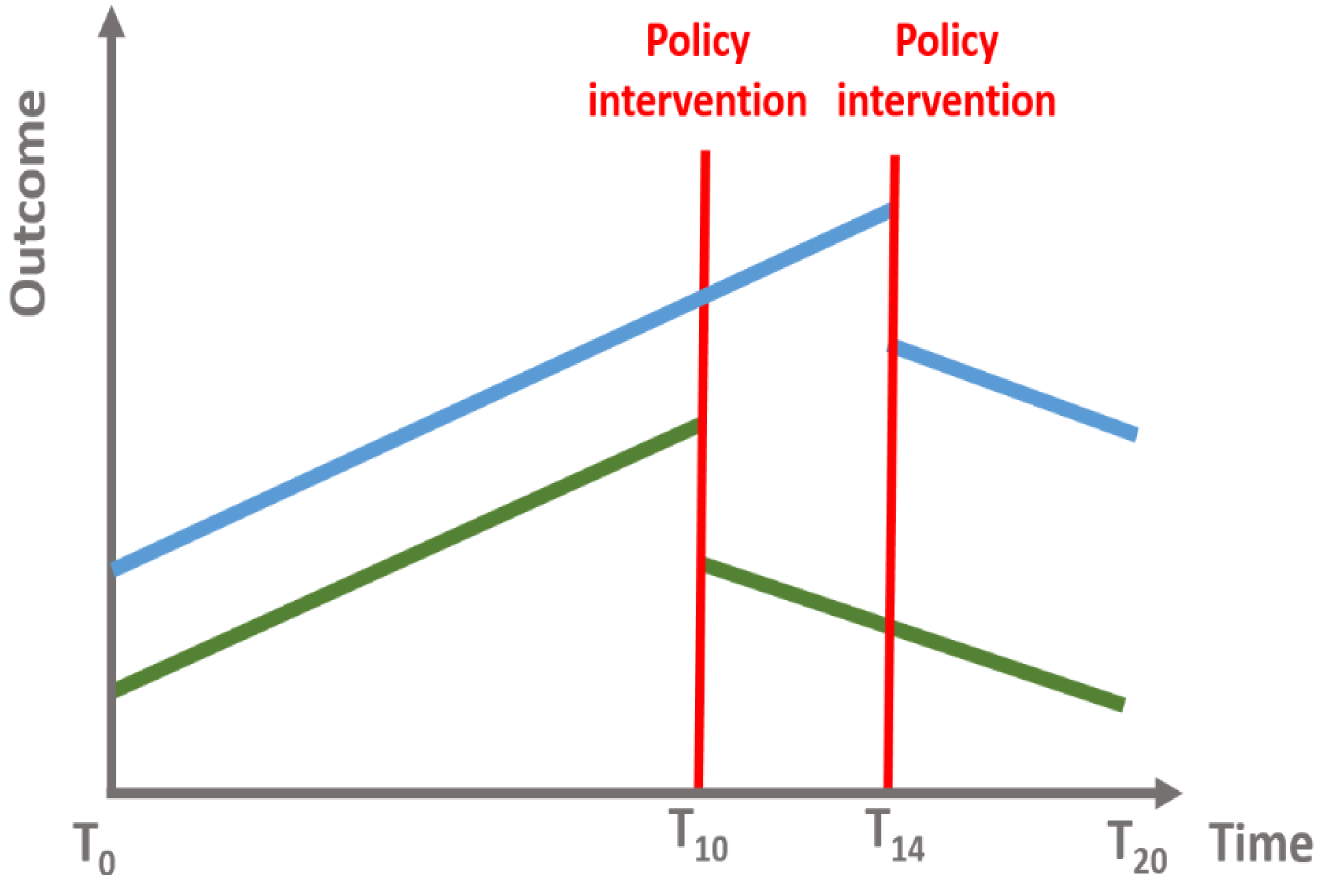
Issues with the design

- Regression to the mean
- How long to observe the effects of intervention?
  - how long the effect of an intervention will be sustained?



Issues with the design

- Validity Threats: selection bias and other related events
- Solution:
  - Use a control group
  - Study multiple effect time points



Recap

Summary

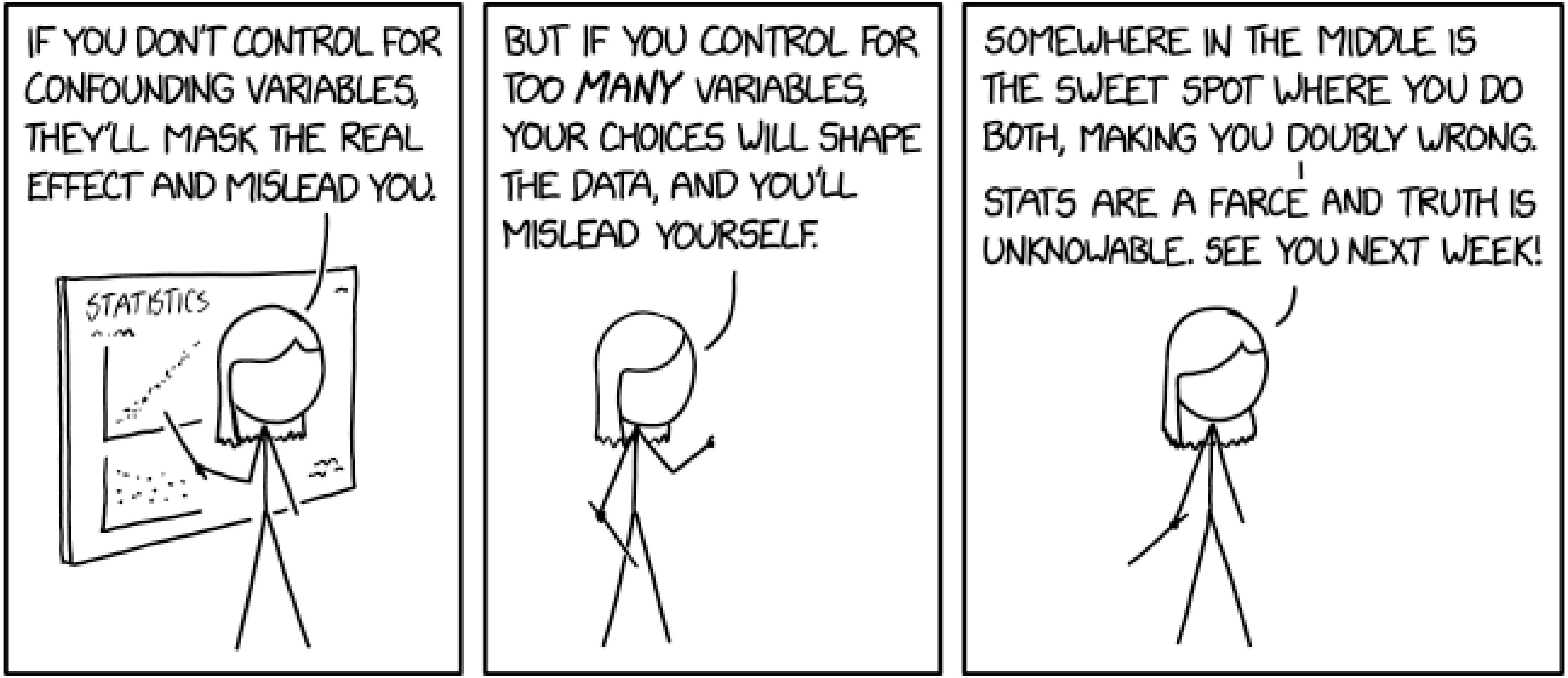
- By assuming treatments are "as good as random" conditional on X, we can make causal claims from non-experimental data
  - We should condition on all potentially confounding variables
- When we want to study the effect on an intervention at a particular time point, we use time series data
  - Event studies - compare before event and after event
  - There are no alternative explanations for change

- Objectives achieved:
  - Can match on observed confounders
  - Can understand time series design, its advantages and disadvantages
  - Can interpret an event studies regression model

References

- Scott Cunningham, Causal Inference: The Mix Tape, Yale University Press.
- Cook, T. D., & Campbell, D. T. (2007). *Experimental and quasi-experimental designs for generalized causal inference*.
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Thank You ☺


IF YOU DON'T CONTROL FOR CONFOUNDING VARIABLES, THEY'LL MASK THE REAL EFFECT AND MISLEAD YOU.
STATISTICS
BUT IF YOU CONTROL FOR TOO MANY VARIABLES, YOUR CHOICES WILL SHAPE THE DATA, AND YOU'LL MISLEAD YOURSELF.
SOMEWHERE IN THE MIDDLE IS THE SWEET SPOT WHERE YOU DO BOTH, MAKING YOU DOUBLY WRONG.
STATS ARE A FARCE AND TRUTH IS UNKNOWABLE. SEE YOU NEXT WEEK!