

Разность разностей и синтетический контроль

Дмитрий Колодезев
Reliable ML 2025.10.19 v1

Мы здесь

- Ортогонализация
- Difference in Difference и синтетический контроль ←
- Матчинг
- Мета-лернеры
- Хороший и плохой контроль
- CATE
- Инструментальные переменные
- Пропущенные переменные
- Causal Discovery
- Обзор библиотек
- Regression discontinuity design и regression kink design *

Терминология

- АТЕ — ожидаемый средний эффект
 $\mathbb{E}[Y \mid do(A = 1)] - \mathbb{E}[Y \mid do(A = 0)]$.
- ATT — эффект на тестовой группе
 $\mathbb{E}[Y(1) \mid A = 1] - \mathbb{E}[Y(0) \mid A = 1]$
- ATU — эффект на контрольной группе
- LATE — эффект с поправкой на вовлеченность*
- CATE — эффект при условиях
 $\mathbb{E}[Y(1) \mid X = x] - \mathbb{E}[Y(0) \mid X = x]$

* Не факт, что больной ест все таблетки

Предположения CI

- Нет пропущенных конфаундеров
 - Позитивность Positivity
 - Консистентность Consistency
 - Нет сетевого эффекта No interference
-
- SUTVA = Consistency + No interference
 - SPICE = SUTVA, Positivity, Identifiability, Consistency, Exchangeability

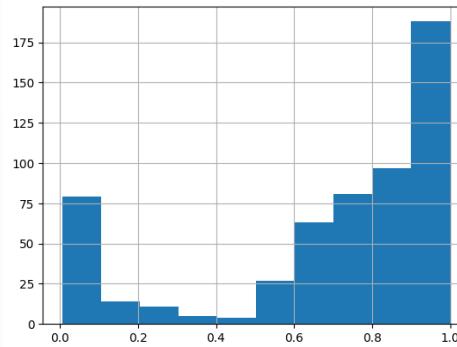
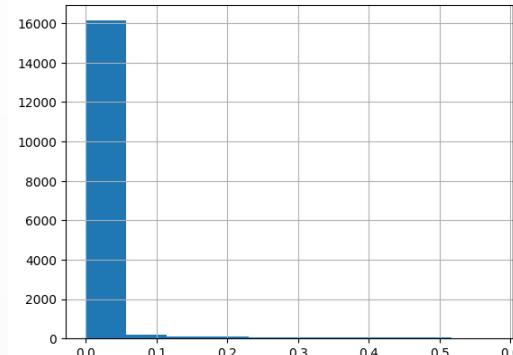
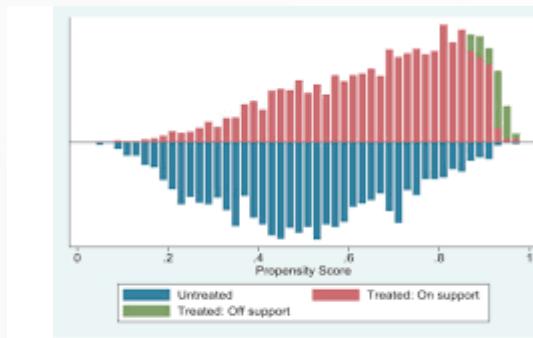
<https://www.stats.ox.ac.uk/~evans/APTS/causassmp.html>

Что там с ИВЛ

- Пропущенных нет (*fate* не влияет на *vent*)
 - Positivity нарушена — легким ИВЛ не давали
 - Консистентность — эффект зависел от тяжести
 - Сетевого эффекта нет (мы не заложили)
-
- В реальных данных вечно что-то нарушено
 - Позитивность нужна всем, некоторым сильнее

Позитивность

- PCI здорового человека и PCI курильщика



Позитивность - у всех был ненулевой шанс попасть под ИВЛ, см positivety.ipynb

<https://www.sciencedirect.com/science/article/abs/pii/S0166046217303186>

Что мы на самом деле оценили

- Как правильно заметил @Andrigeto мы оценили не АТЕ а АТТ
- Что мы могли оценить?
- Ключевой вопрос — источник дисперсии воздействия

The treatment effect we get is almost entirely determined by the source of treatment variation we use. That's pretty much it. Ask where the variation in your treatment is coming from, after removing any variation you choose to remove by controlling for things, etc. and you'll have a pretty good idea whose treatment effects you are averaging, and who is being weighted more heavily. (c) Nick Huntington-Klein

Типа правила из книжки

- ROT1: если у нас случайно назначено воздействие и не нужны поправки, у нас есть ATE
- ROT2: если у нас назначение разное по группам и мы можем их изолировать, у нас есть CATE*
- ROT3: если часть дисперсии воздействия идет через back doors (позже), мы можем ее убрать — и на остатках получить взвешенный средний эффект
- ROT4: если мы решим, что у нас есть взаимозаменяемые подгруппы в контрольной и тестовой группах, у нас есть ATT
- ROT5: если часть дисперсии воздействия идет от экзогенной переменной, и вы изолируете только эту часть, у вас есть LATE*

* не в этот раз

<https://theeffectbook.net/ch-TreatmentEffects.html#i-just-want-an-ate-it-would-make-me-feel-great-what-do-i-get>

К следующему воркшопу

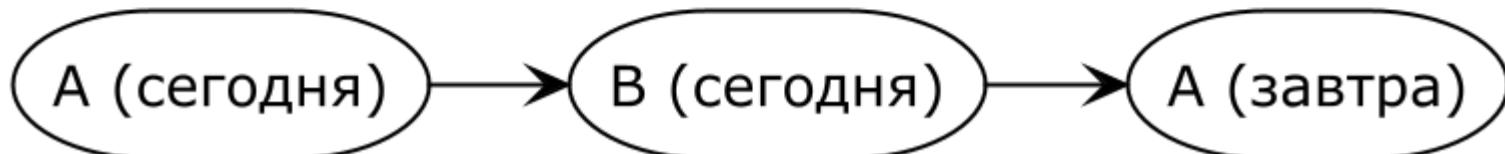
- <https://arxiv.org/abs/2506.04194>
What Makes Treatment Effects Identifiable:
Characterizations and Estimators Beyond
Unconfoundedness
- <https://pubmed.ncbi.nlm.nih.gov/38847059/>
Causal inference in the absence of positivity:
The role of overlap weights

Основная проблема СІ

- Нет параллельной вселенной под рукой
- Не можем наблюдать контрфактический результат
- Можем предположить, не можем проверить
- DiD - у нашей группы есть брат-близнец
- SC – близнеца нет, но есть Франкенштейн

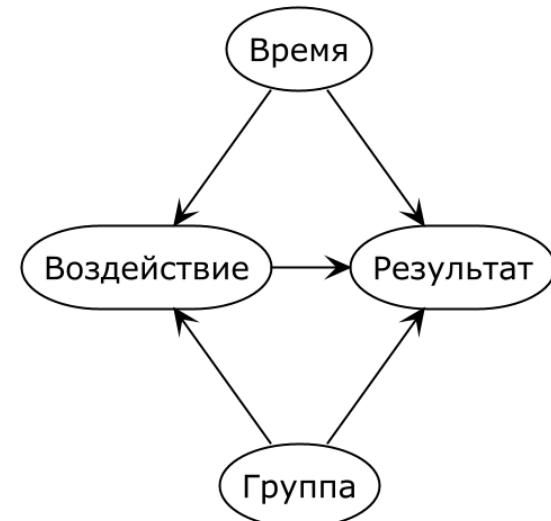
Ряды и панели

- Циклы нельзя, мы их разделим



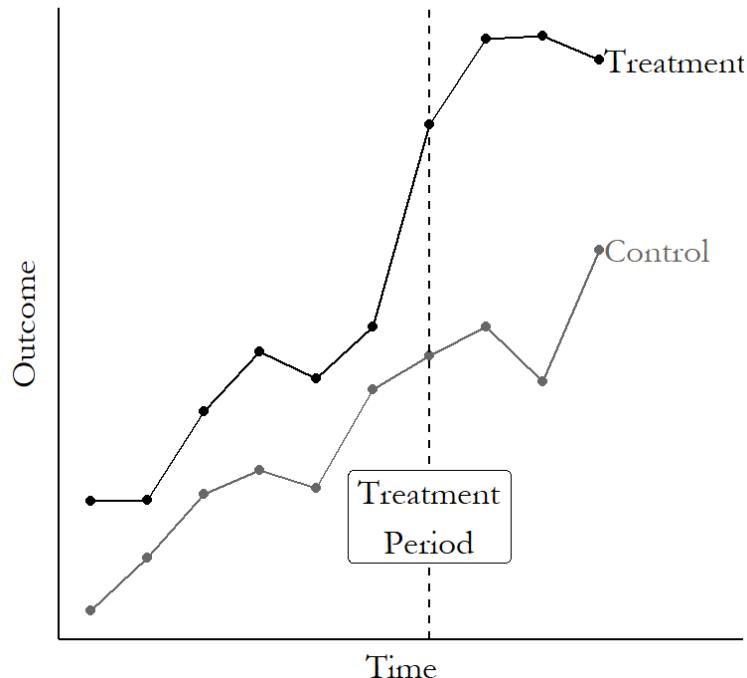
Брат-близнец

- Брат-близнец меняется так же как и мы
- Если бы не было воздействия, наши разницы до и после совпадали бы (параллельность)
- Разница разниц = эффект

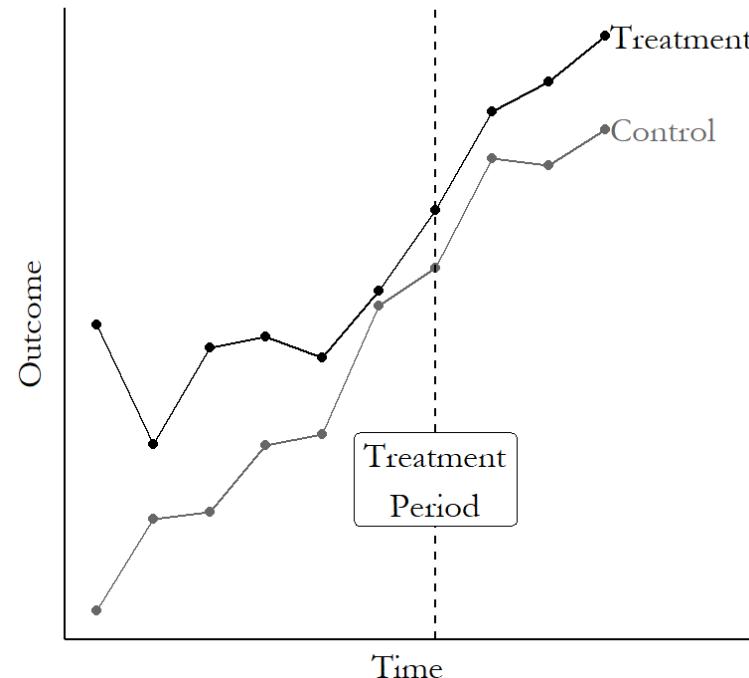


Видел бы нас Евклид

(a) Parallel Prior Trends



(b) Converging Prior Trends



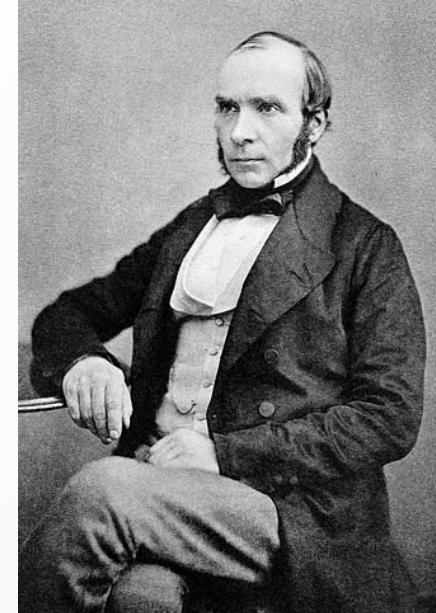
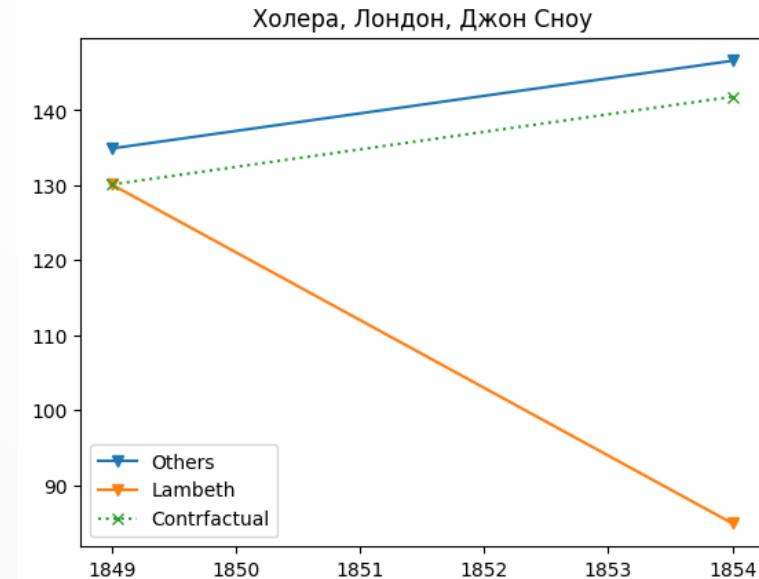
Глупый вопрос

- Сколько непараллельности можно?
- Линейный эффект в популяции
- Линейный эффект при разных базовых уровнях
- Немножко логарифмируем туда-сюда, и Ок
- Всегда нужно думать об аддитивности и мультипликативности эффектов
- DiD выявляют аддитивный эффект
- И, кстати, мы оцениваем ATT

Еще более глупое наблюдение

- Если группы с разным базовым уровнем меняются параллельно, они не взаимозаменяемы
- За аддитивным эффектом может скрываться аддитивный эффект в подгруппе, тогда все отлично
- Или мультипликативный эффект, тогда все плохо

Джон Сноу



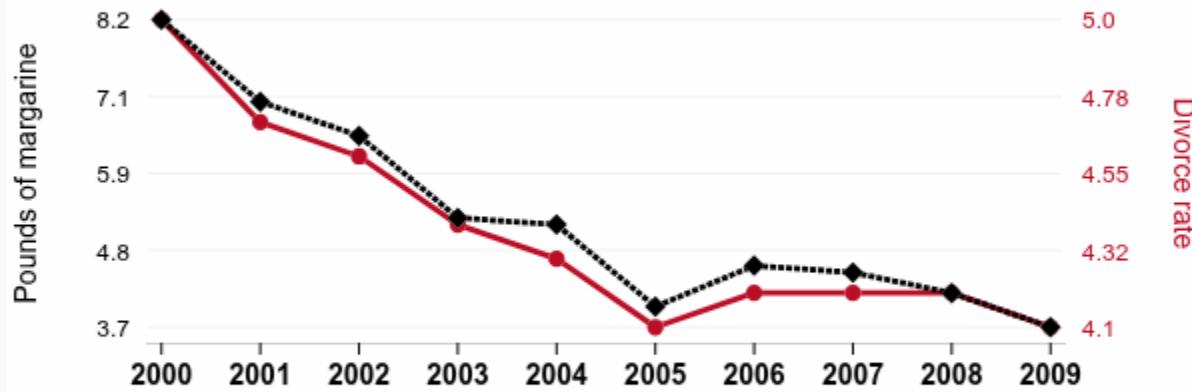
Оба из википедии, график сам рисовал. У сэра Джона это был пятый проект

Параллельность

Per capita consumption of margarine

correlates with

The divorce rate in Maine



◆ Per capita consumption of margarine in the United States · Source: US Department of Agriculture

● The divorce rate in Maine · Source: CDC National Vital Statistics

2000-2009, $r=0.993$, $r^2=0.985$, $p<0.01$ · tylervigen.com/spurious/correlation/5920

Как проверить параллельность

- Никак. Параллельность непроверяема
- Статтест на разницу групп до того (DiD плацебо)
- Верить, что контрольная группа не получила неучтенного воздействия
- Верить, что группы во многом похожи - EDA
- Проверить, что траектории до вмешательства совпадали - EDA

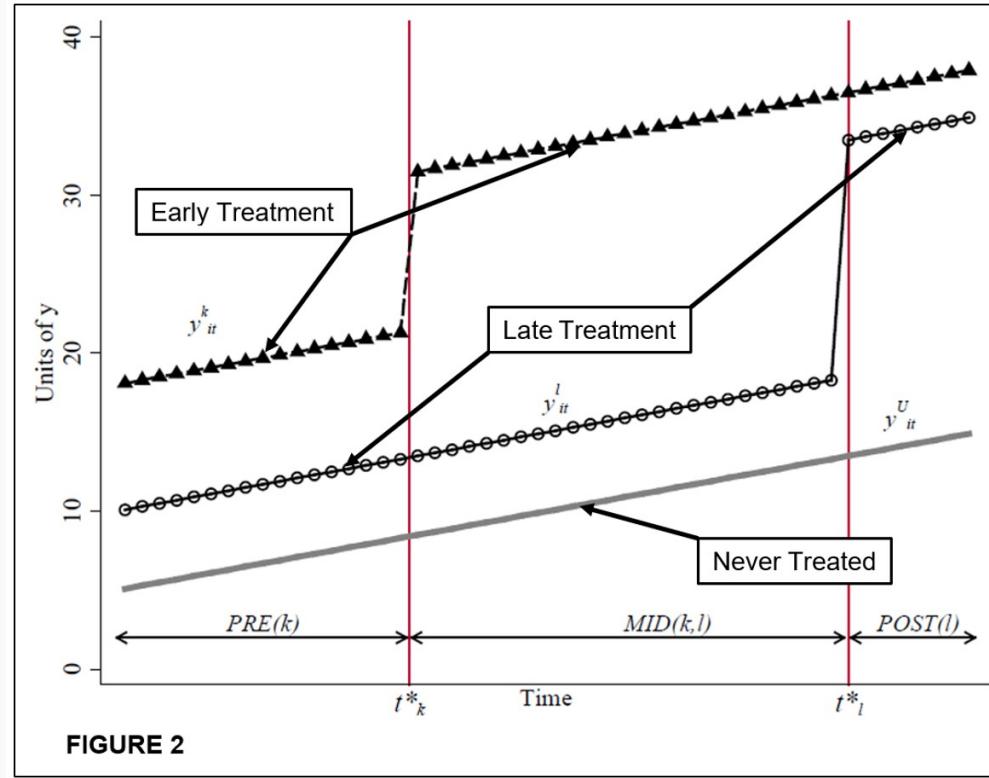
Где доверительный интервал?

- Есть сгруппированные данные — нет интервалов
- Ну или есть, если есть данные о дисперсиях
- Если доступны данные по каждому случаю — ols
- Модель с фиксированным эффектом `pyfixest.feols`
 - Есть смещение у каждой группы
 - Есть смещение у каждого момента времени

Что, если

- Все менялось не одновременно
- У нас есть несколько контрольных групп
- У нас есть несколько тестовых групп
- Эффект наступал не сразу — тут проблема **staggered treatment adoption**
- Примерно с 2021: РАНЬШЕ все использовали TWFE
- А теперь мы рекомендуем НАШ МЕТОД

TWFE - Примерно Так



Декомпозиция в 2x2DiD

Difference-in-Differences Designs: A Practitioner's Guide

Andrew Baker* Brantly Callaway† Scott Cunningham‡

Andrew Goodman-Bacon§ Pedro H. C. Sant'Anna¶

June 18, 2025

Abstract

Difference-in-differences (DiD) is arguably the most popular quasi-experimental research design. Its canonical form, with two groups and two periods, is well-understood. However, empirical practices can be ad hoc when researchers go beyond that simple case. This article provides an organizing framework for discussing different types of DiD designs and their associated DiD estimators. It discusses covariates, weights, handling multiple periods, and staggered treatments. The organizational framework, however, applies to other extensions of DiD methods as well.

Если брата-близнеца нет

- Если есть несколько похожих рядов
- Соберем линейную комбинацию
- Неотрицательные коэффициенты, сумма= 1
- Получившегося франкенштейна в DiD
- Интерпретируемость падает
- Синтетический контроль

Итого, что делать-то?

- Где много юнитов и много событий - `pyfixest`
- Там, где есть хорошие панельные данные — использовать синтетический контроль `pysyncon red dit`
- Там, где есть очень похожая группа и нужна интерпретируемость для не-технарей — DiD
- Там где нужны доверительные интервалы — ~~купить чернил и плакать~~ либо иметь данные по каждой точке либо посчитать их руками
- Можно написать SC руками, **не очень сложно**