Inferring Treatment Effects Through Quasi-Experiments

The Difference-in-Differences Estimator and its Extensions

Nils Wlömert[1], Daniel Winkler[1]

[1]Vienna University of Economics & Business

21 February

@UNSW Data Science Hub 2023

Plan for today

1. Preliminaries
1.1 Research design for causal inference
1.2 Causal effects in panel data settings

2. Difference-in-Differences Estimator
2.1 Motivation
2.2 Estimation
2.3 Assumptions

3. Extensions
3.1 Heterogeneous Treatment Effects
3.2 Matched Difference-in-Differences
3.3 Synthetic Control
3.4 Synthetic Difference-in-Differences
3.5 Time-varying Treatment Effects
3.6 Tripple Differences
3.7 Further extensions

Workshop outline

- None of the materials covered in the workshop are new
- We summarize and synthesize current topics around Difference-in-Differences (DiD) estimation from an applied perspective
- The session is designed interactively; please ask questions anytime
- If you have prior knowledge in R, you may follow the practical coding examples in R on our own laptop using the provided code file
- If you can‘t follow the example, you may review the code & outputs via the following link: https://wu-rds.github.io/did/

Recommended materials - Readings

Textbooks

- Huntington-Klein (2022): *The effect: An introduction to research design and causality*
- Cunningham (2021): *Causal inference: The mixtape*
- Wooldridge (2010): *Econometric analysis of cross section and panel data*
- Wickham, Çetinkaya-Rundel, and Grolemund (2023): R-Intro: *R for Data Science*

Papers

- Roth et al. (2023): "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature"*
- Todri (2022): "Frontiers: The Impact of Ad-Blockers on Online Consumer Behavior"*
- Rambachan and Roth (2023): "A More Credible Approach to Parallel Trends"
- Callaway and Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods"
- Arkhangelsky et al. (2021): "Synthetic Difference-in-Differences"
- Varian (2016): "Causal Inference in Economics and Marketing"

Recommended materials - R packages

Estimation

- *fixest: Fast and user-friendly fixed-effects estimation* (Bergé, 2018)
- Callaway and Sant'Anna (2021): *did: Difference in Differences*
- Xu and Liu (2021): *gsynth: Generalized Synthetic Control Method*
- Greifer (2022): *WeightIt: Weighting for Covariate Balance in Observational Studies*
- Arkhangelsky (2023): *synthdid: Synthetic Difference-in-Difference Estimation*
- Brodersen et al. (2014): "Inferring causal impact using Bayesian structural time-series models"

Diagnostics & Visualization

- Rambachan (2023): *HonestDiD: Robust inference in difference-in-differences and event study designs*
- Mou, Liu, and Xu (2023): *panelView: Visualizing Panel Data*
- Flack and Jee (2020): *bacondecomp: Goodman-Bacon Decomposition*
- Huntington-Klein (2021): "DAGitty — draw and analyze causal diagrams"

Recommended materials - Twitter accounts

- Dimitry Arkhangelsky: @ArkhangelskyD
- Susan Athey: @Susan_Athey
- Laurent Bergé: @lrberge
- Scott Cunningham: @causalinf
- Andrew Goodman-Bacon: @agoodmanbacon
- Nick Huntington-Klein: @nickchk
- Ashesh Rambachan: @asheshrambachan
- Jonathan Roth: @jondr44
- Pedro Sant'Anna: @pedrohcgs
- Khoa Vu: @KhoaVuUmn
- Jeffrey Wooldridge: @jmwooldridge
- Paul Goldsmith-Pinkham: @paulgp

Preliminaries

Increasing number of publications on DiD

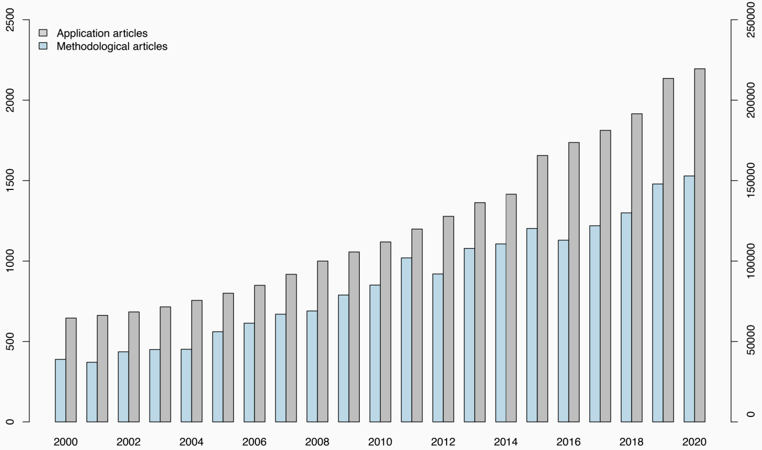
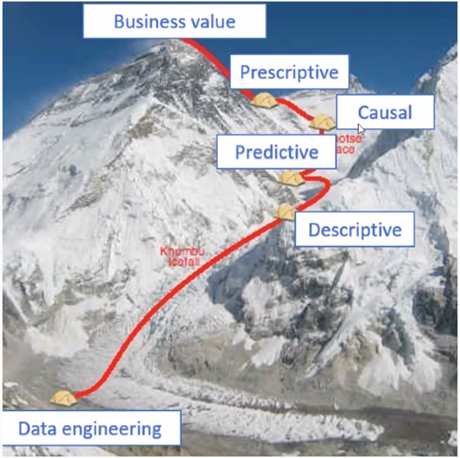


Figure: Number of published DiD-related articles (Hirschberg, 2020)

Current developments in DiD

- DiD estimation is a very active field under intense development from different fields
- Two topics in particular are being tackled
  - Violations of the parallel trend assumptions
  - Differences in treatment timing (staggered adoption)

Climbing the AI Summit



- Increasing relevance of descriptive analytics through big data
- Predictive analytics is where a lot of breakthroughs happened in the field of machine learning over the past years
- Causal inference is rough terrain, but crucial for policy evaluation through counterfactual predictions
- Prescriptive is where we go from (counterfactual) predictions to decisions in the context of organizations and individual-level constraints

Research designs

	Description	Prediction	Causal Inference
Example of scientific questions	How can the customers of our online store be partitioned in classes defined by their characteristics?	What is the probability that users who visited our online store last year will purchase from our store within the next month?	Will behavioral targeting in online advertising increase, on average, the probability of purchasing from our store within the next month?
Data	**Features**: user characteristics (age, gender, location, ...), product characteristics of visited pages, ...;	**Output**: making a purchases within the next month **Inputs**: age, gender, frequency of past purchases, recency of last purchases, monetary value of past purchases, past ad exposures, ...	**Outcome**: making a purchases within the next month **Treatment**: initiation of targeting campaign **Confounders**: for non-experimental settings (interest in product category, eligibility criteria used for targeting ...)
Example of analytics	Cluster Anaylsis ...	Regression Decision trees Random forests Support vector machines Neural networks ...	Experiments with random assignment Regression Instrumental variables Regression discontinuity Difference-in-differences ...

Figure: Classification of data science tasks based on Hernán, Hsu, and Healy (2019)

Data generating process

- In social sciences, causal questions matter more than associational
  - What is the efficacy of a given drug in a given population?
  - Does personalized advertising lead to higher sales?
  - What fraction of past crimes could have been avoided by a given policy?
  - Do lower prices lead to more demand?
  - Which part of the sales growth is due to the increase in advertising budget?
  - Is advertisement A or B more effective?
  - Do loyalty cards lead to higher sales?
- Interest in the effect of a cause (X) on an outcome (Y)

- Answers to these questions require knowledge about the data-generating process
- They cannot be computed from the data alone or from the distributions
- They reflect a quest for knowledge on how data is generated

Data generating process

The DGP describes how the data came about...

- Who acts to create X?
- In which fashion is X created?
- What do the actors who create X consider in their decisions?
- Do they have any knowledge on Y?
- Who acts to create Y?
- In which way do the actors who create Y use X?
- ...

Research designs - Experimental research

"To find out what happens when you change something, it is necessary to change it."

— Box, W. Hunter, and J. Hunter (1978)

- Experiments are a proper way of establishing a causal relationship
- Typically superior to other methods of controlling extraneous variables
- Experimentation process:
  - Divide test units into homogeneous subsamples
  - Manipulate independent variables and measure dependent variable
  - Random assignment of test units to test and control groups to control for extraneous (potentially confounding) variables

Research designs - Observational research

"Behind any causal conclusion there must be some causal assumption, untested in observational studies."

— Pearl (2009)

- Observe what happens naturally without interfering
- Concerns about unobserved confounders (X = idependent variable, Y = dependent variable, U = unobserved confounder)

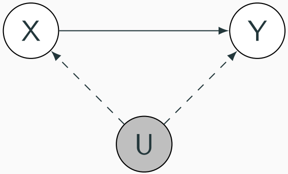


Figure: DAG - Selection bias due to unobserved confounder

Formal representation

- Motivating example: smart ice cream seller
  - Good weather – high prices; bad weather – low price
  - Challenge: weather (plus other factors) is unobserved in the model and influences price and demand This leads to inconsistent estimates of the effect of price
- Consider the following equation, where Y_i is demand, X_1 is price and X_2 is weather:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \phi \tag{1}$$

- If weather is unobserved, then:

$$Y = \beta_0 + \beta_1 * X_1 + \epsilon \tag{2}$$

- ... where $\epsilon = \beta_2 * X_2 + \phi$, and thus $COV(X;\epsilon) \neq 0$ (exogeneity assumption is violated)
- The model is not identified $\rightarrow$ the estimate of $\beta 1$ is biased and inconsistent

Estimation challenges in observational research

Causal question	**Potential problem**
Does personalized advertising lead to higher sales?	Customers who are targeted by personalized ads may have a higher interest in the product categories
Do loyalty cards lead to higher sales?	Customers who sign up for a loyalty program likely spend more than other customers even without the loyalty card
Do lower prices lead to more demand?	Managers set prices based on unobserved characteristics (e.g., quality)

Research designs - Observational research

"Given the manifest virtues of experiments, why do I almost always analyze observational data? The short answer is almost all data out there are observational."

— Gelman (2020)

- Randomized trials are widely considered the 'gold standard' to investigate causal effects
- In many situations, experiments are often not feasible, not appropriate, or simply too costly to conduct (e.g., channel & product development research, price experiments)

Identification strategy

"Underlying this is the recognition, description, and presentation of the identification strategy, or the manner in which a researcher uses observational data (i.e., data not generated by a randomized trial) to approximate a real experiment"

— Angrist and Pischke (2009)

- In the absence of a randomized trial, you need to rely on statistical methods to estimate a 'counterfactual'
- In many situations, experiments are often not feasible, not appropriate, or simply too costly to conduct (e.g., channel & product development research, price experiments)
- The approach how this is achieved, is detailed in the identification strategy

Identification strategy

- "Furious Five methods of causal inference" (Angrist and Pischke, 2009):
- 1) random assignment, 2) regression, 3) differences-in-differences, 4) instrumental variables, 5) regression discontinuity

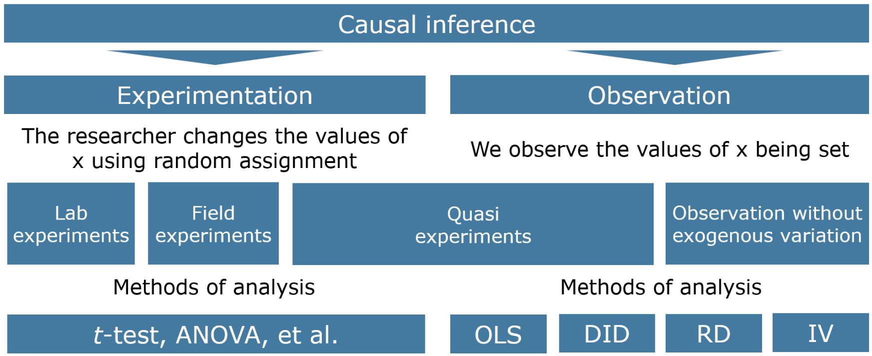


Figure: Data excerpt - Panel data

Paths to your identification strategy

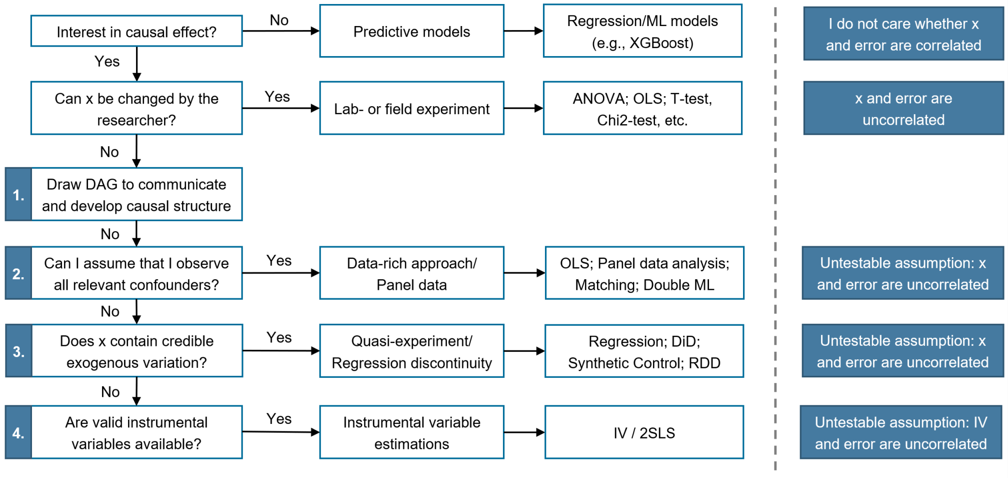


Figure: Causal inference flow chart

Directed Acyclic Graphs (DAG)

- Nodes (variables)
  - White circles mean variable is observed
  - Gray circles mean variable is unobserved
- Directed edges connect nodes (effects/coefficients)
  - Solid lines connect observed entities
  - Dashed lines involve unobserved entities
- Arrow indicates (causal) effect – absence of arrow indicates absence of (causal) effect Nodes can be described as ancestors/descendants and parents/children

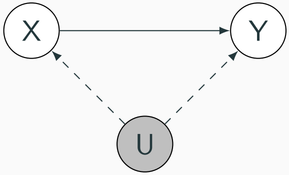


Figure: DAG - Selection bias due to unobserved confounder

Causal effects in panel data settings

Motivating example:

The global recorded music industry revenue is over $23 billion annually. With the rise of streaming services, new marketing tools, such as music playlists, became relevant and traditional marketing tools (e.g., advertising, radio) possible lose relevance.

As a marketing manager at a major music label, you are interested to estimate the elasticity of demand for 1) advertising, 2) radio, and compare their effectiveness to the elasticity of 3) playlist placements on streaming services. For this task, you collected weekly data, containing information regarding the number of streams, and the key independent variables (i.e., advertising, radio, playlist followers) for a sample of 97 songs and a duration of 3 years (i.e., 156 weeks).

The data for this exercise appears in the file "music_data.csv".

Causal effects in panel data settings

- Panel data set contains multiple observations per unit over time
- Included variables:
  - song_id: Unique song ID
  - week: Observation week
  - streams: Number of weekly streams for song i in week t
  - release_date: Release date of the song
  - weeks_since_release: Number of weeks since the song had been released
  - playlist_follower: Sum of followers across all playlists for song i in week t
  - radio: Audience-weighted number of radio plays for song i in week t
  - adspend: Advertising expenditures related so i is listed on in week t
  - genre: Genre associated with song i

	song_id	week	streams	release_date	weeks_since_release	playlist_follower	radio	adspend	genre
149	1	2019-11-04	1136	2013-01-14	355	59026.0	0	0	rock
150	1	2019-11-11	1027	2013-01-14	356	59356.0	0	0	rock
151	1	2019-11-18	870	2013-01-14	357	59662.0	0	0	rock
152	1	2019-11-25	901	2013-01-14	358	59967.0	0	0	rock
153	1	2019-12-02	970	2013-01-14	359	60420.0	0	0	rock
154	1	2019-12-09	967	2013-01-14	360	60824.0	0	0	rock
155	1	2019-12-16	889	2013-01-14	361	61239.0	0	0	rock
156	1	2019-12-23	652	2013-01-14	362	61582.0	0	0	rock
157	2	2017-01-02	3345	2015-03-16	94	186189.0	0	0	classic
158	2	2017-01-09	3590	2015-03-16	95	186843.0	0	0	classic
159	2	2017-01-16	3110	2015-03-16	96	187400.0	0	0	classic
160	2	2017-01-23	3217	2015-03-16	97	187935.0	0	0	classic
161	2	2017-01-30	2939	2015-03-16	98	188459.0	0	0	classic

Figure: Data excerpt - Panel data

Causal effects in panel data settings

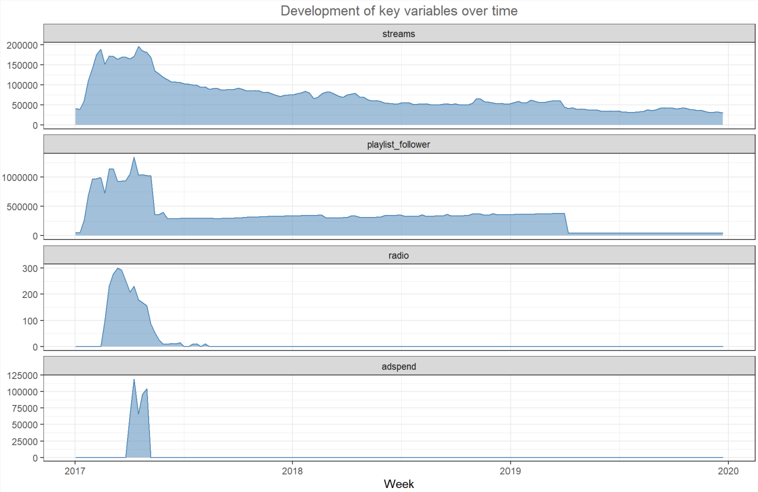


Figure: Data for one example song - Market response model

Causal effects in panel data settings

Consider the following market response model for a set of songs (i) and weeks (t):

$$Y_{it} = \beta_0 + \beta_1 * X_{it} + c_i + \epsilon_{it} \tag{3}$$

- Following our example, let Y_{it} represent the number of streams and X_{it} the number of playlist followers associated with song i in week t
- Let's assume, c_i is unobserved, i.e., it contains unobserved song characteristics (e.g., music quality, musical features, artist status, etc.)
- Because c_i is unobserved it becomes part of the error term
- If the unobserved song-specific effects affect both the number of streams and the number of playlists, then $COV(X;\epsilon) \neq 0$
- Panel data are a powerful way of removing unobserved time-invariant unit-specific effects

Fixed effects panel data model

- Demeaning removes the unobserved effect (see also here):

$$Y_{it} - \bar{Y_i} = \beta_1 * (X_{it} - \bar{X_i}) + (c_i - c_i) + (\epsilon_{it} - \bar{\epsilon_i}) \tag{4}$$

- This is equivalent to adding one dummy variable for each unit (i)
- In our example:

$$\begin{aligned} Y_{it} =& \beta_1 * PlaylistFollowers_{it} + \beta_2 * Advertising_{it} + \beta_3 * Radio_{it} \\ &+ \mu_i + \epsilon_{it} \end{aligned} \tag{5}$$

- where μ_i and are unit fixed effects (i.e., one dummy variable per unit, which is equivalent to de-meaning the data on this dimension)

Fixed effects panel data model

- Add fixed effect on aggregation levels that are likely sources of unobserved heterogeneity
- In our example, we may also add week fixed-effects:

$$\begin{aligned} Y_{it} = &\beta_1 * PlaylistFollowers_{it} + \beta_2 * Advertising_{it} + \beta_3 * Radio_{it} \\ &+ \mu_i + \gamma_t + \epsilon_{it} \end{aligned} \tag{6}$$

- where γ_t are time fixed effects (i.e., one dummy variable per week, which is equivalent to de-meaning the data on this dimension)

Mixed effects panel data model

- Heterogeneity in intercept and/or regressors is explicitly modeled
- For example, using random intercept and slope for the $PlaylistFollower$ variable in our model:

$$
\begin{aligned}
&Level \quad 1: \\
&\qquad Y_{it} = \beta_{0i} + \beta_{1i} * PlaylistFollowers_{it} + \beta_2 * Advertising_{it} \\
&\qquad\qquad + \beta_3 * Radio_{it} + \epsilon_{it}
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
&Level \quad 2: \\
&\qquad \beta_{0i} = \alpha_{00} + \phi_{0i} \\
&\qquad \beta_{1i} = \alpha_{01} + \phi_{1i}
\end{aligned} \tag{8}
$$

- ...where β_{0i} and β_{1i} are random effects that are explicitely modeled

Difference-in-Differences Estimator

When and why

"Quasi-experimental tools mimic the random assignment that is inherent in lab experiments and that is often referred to as the 'gold standard' for identifying causal relationships."

— Goldfarb and Tucker (2014)

- Group assignment due to some exogenous shock (not controlled by the researcher)
- Since observational units self-select into test and control groups the challenge is to assess if the quasi-experiment credibly proxies for random assignment
- Often fewer external validity concerns compared to lab experiments
- The Difference-in-Differences estimator is a popular approach to estimating treatment effects in quasi-experimental settings

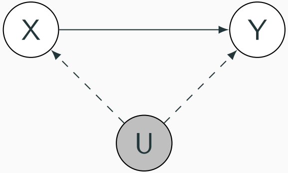


Figure: Selection bias due to unobserved confounder

When and why

"In each case, an external shock creates a source of exogenous variation that the researcher uses to establish a causal relationship between the variation and the outcome of interest."

— Goldfarb, Tucker, and Wang (2022)

- Sources of variation
  - Contractual
  - Ecological
  - Geographical
  - Macroeconomic
  - Individual-level
  - Organizational
  - Regulatory

When and why

- We require a treatment group and a control group of untreated units
- Group assignment as good as random $\rightarrow$ often used to analyze quasi-experiments
- We use repeated observations (at least two) per unit to remove unobserved heterogeneity (across time and across units)
- Estimator exploits longitudinal and cross-sectional dimensions of the data

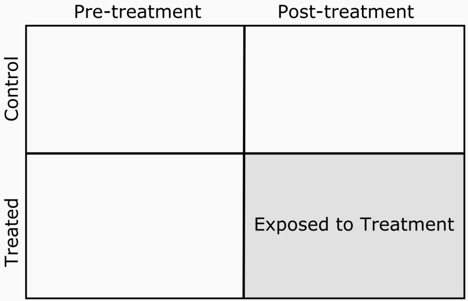


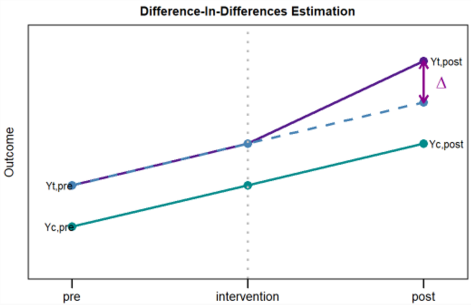
Figure: DiD-Matrix (Hirschberg, 2020)

Potential outcomes framework

- We can never observe the same unit with and without treatment
- We need to come up with a way to make counterfactual predictions
- Average treatment effect on the treated (ATT) (see also here) :

$$ATT(\Delta) = (Y_{t,post} - Y_{t,pre}) - (Y_{c,post} - Y_{c,pre}) \quad (9)$$

- See also work by Guido Imbens, Donald Rubin, and Josh Angrist, e.g., Imbens and Rubin (2015), Angrist and Pischke (2009)



	Pre-treatment	Post-treatment
Control	$Y_{c,pre}$	$Y_{c,post}$
Treated	$Y_{t,pre}$	Exposed to Treatment $Y_{t,post}$

Assumptions

- Untestable key assumption: Treatment group would have developed as control group if not for the treatment
- Surrogate test: carefully check that prior to treatment groups behave similarly (parallel pre-treatment trends assumption; see e.g., Todri (2022))
  - Falsification: conduct placebo tests (e.g., assign treatment/untreated units to different groups; shift treatment period)
  - Leads- & lags: plot "treatment effects" prior and post to the treatment
- In addition:
  - Treatment does not coincide with unobservable confounders
  - The controls units are not directly or indirectly affected by the treatment
  - Treatment has no causal effect prior to its implementation (no-anticipation)
- Potential solutions to violations of assumptions:
  - Matched DiD (e.g., matching on pre-treatment outcomes)
  - Utilize more information to match pre-treatment trends (synthetic control/DiD)

Motivating example

- The relevance of playlists to stimulate demand is increasing
- Research question: What is the effect of a playlist listing on the number of streams that a song receives?
- The data for this exercise appears in the file "did_data.csv"

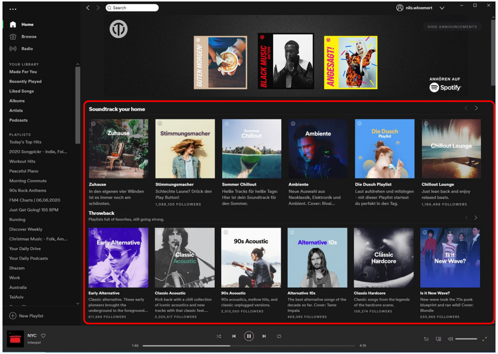


Figure: Spotify playlist

Motivating example

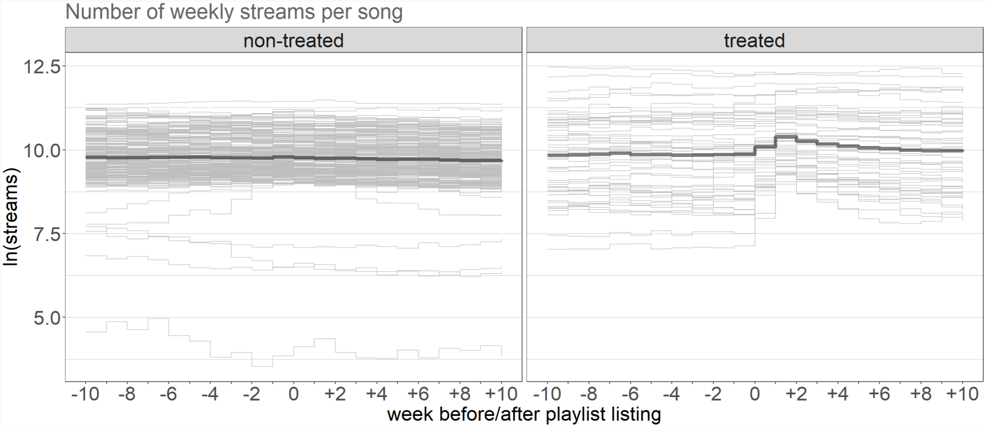
- song_id: Unique song ID
- week: Observation week
- streams: Number of streams for song *i* in week *t*
- post: indicator variable that is 1 for all periods after playlist listing, and 0 else
- treated: indicator variable that is 1 for all songs that got listed on the playlist, and 0 for control songs
- treated_post: interaction between *post* and *treated* variables
- song features: danceability, specechiness, acousticness, instrumentalness, liveness, valence, tempo, duration, genre
- audience age: age of an artist's audience
- artist fame: previous chart placements of an artist
- ig_followers: Instagram followers
- n_playlists: number of playlist placements
- playlist_followers: sum of playlist followers across all playlists a song is listed on
- n_postings: number of social media postings

	week	song_id	streams	treated_post	post	treated	major_label	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	genre	audience_age	artist_fame	ig_followers	n_playlists	playlist_followers	n_postings
80	2018-04-30	4	18975	0	1	0	1	0.538	0.742	-7.924	0.0380	0.0188	0.09000000	0.0749	0.273	150.029	348.053	pop	51.00122	1.21030204	14055	41	765871	5
81	2018-05-07	4	17253	0	1	0	1	0.538	0.742	-7.924	0.0380	0.0188	0.09000000	0.0749	0.273	150.029	348.053	pop	51.00122	1.21030204	14069	41	780094	7
82	2018-05-14	4	16337	0	1	0	1	0.538	0.742	-7.924	0.0380	0.0188	0.09000000	0.0749	0.273	150.029	348.053	pop	51.00122	1.21030204	14092	41	789983	11
83	2018-05-21	4	18434	0	1	0	1	0.538	0.742	-7.924	0.0380	0.0188	0.09000000	0.0749	0.273	150.029	348.053	pop	51.00122	1.21030204	14103	42	801541	9
84	2018-05-28	4	19377	0	1	0	1	0.538	0.742	-7.924	0.0380	0.0188	0.09000000	0.0749	0.273	150.029	348.053	pop	51.00122	1.21030204	14152	42	812700	14
85	2018-01-08	5	14014	0	0	0	1	0.241	0.272	-7.629	0.0297	0.3460	0.00000000	0.2010	0.128	179.336	238.706	rock	49.92508	1.28084324	424429	63	520723	3
86	2018-01-15	5	14509	0	0	0	1	0.241	0.272	-7.629	0.0297	0.3460	0.00000000	0.2010	0.128	179.336	238.706	rock	49.92508	1.28084324	425027	63	526669	0
87	2018-01-22	5	14431	0	0	0	1	0.241	0.272	-7.629	0.0297	0.3460	0.00000000	0.2010	0.128	179.336	238.706	rock	49.92508	1.28084324	425137	63	532022	6
88	2018-01-29	5	14908	0	0	0	1	0.241	0.272	-7.629	0.0297	0.3460	0.00000000	0.2010	0.128	179.336	238.706	rock	49.92508	1.28084324	427757	62	536951	6

Figure: Data excerpt - Difference-in-Differences

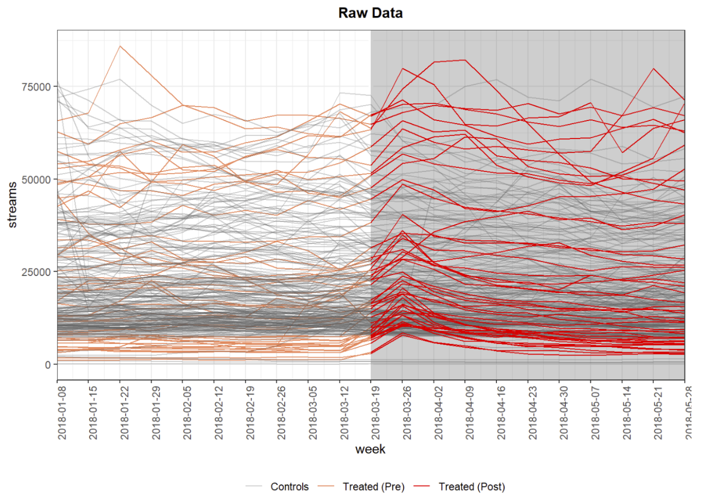
Model free evidence

Always visualize your data first!



Model free evidence

Always visualize your data first!



Estimation

Two possible specifications:

$$Y_{it} = \delta * (treated_i \times post_t) + \alpha_1 * treated_i + \alpha_2 * post_t + \epsilon_{it} \tag{10}$$

- $treated_i$ is an indicator variable that is 1 for the treated units
- $post_t$ is an indicator variable that is 1 for the treated periods
- δ is our estimate of the treatment effect

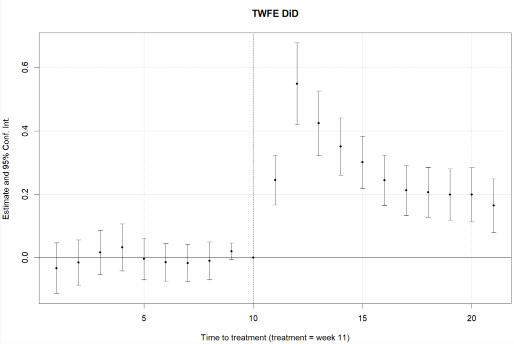
$$Y_{it} = \delta Treatment_{it} + \mu_i + \gamma_t + \epsilon_{it} \tag{11}$$

- $Treatment_{it}$ is an indicator variable that is 1 for the treated units in the post-treatment periods, and 0 else (i.e., $(treated_i \times post_t)$)
- μ_i and γ_t are unit and time fixed effects

Testing assumptions

- Conduct leads & lags analysis to assess plausibility of parallel pre-treatment trends

$$Y_{it} = \sum_{t=1}^{T} \delta_t (period_t \times treated_i) + \mu_i + \gamma_t + \epsilon_{it} \tag{12}$$



- In addition, conduct falsification (placebo) tests (see code)

Testing assumptions

Rambachan and Roth (2023): "A More Credible Approach to Parallel Trends"

- Impose restrictions on how different the post-treatment violations of parallel trends can be from the pre-treatment differences in trends
- Causal parameter of interest is partially identified under these restrictions
- Formally: impose that the post-treatment violation of parallel trends is not more than $\bar{M}$ longer than the worst pre-treatment violation of parallel trends (between consecutive periods)
- For sensitivity analyses report standard errors under different assumptions for $\bar{M}$:

lb	ub	method	Delta	Mbar
0.344514861	0.8091479	C-LF	DeltaRM	0.5
0.249476279	0.9622656	C-LF	DeltaRM	1.0
0.135957972	1.1180233	C-LF	DeltaRM	1.5
0.003959941	1.2737810	C-LF	DeltaRM	2.0
-0.143877854	1.3186603	C-LF	DeltaRM	2.5

Figure: Assessment of significance of causal parameter under different $\bar{M}$

Extensions

Heterogeneous Treatment Effects

Heterogeneity across units:

- Treatment effects are rarely the same across all the observed treated units
- Heterogeneous treatment effects can be used to test behavioral mechanisms (Goldfarb, Tucker, and Wang, 2022)
- Moderation analysis can be used to analyse heterogeneity:

$$Y_{it} = \delta Treatment_{it} + \sum_{j=1}^{J} \phi_j (Treatment_{it} * X_i) + \mu_i + \gamma_t + \epsilon_{it} \tag{13}$$

... where X_i is a vector of covariates that are interacted with the treatment indicator

Heterogeneous Treatment Effects

Heterogeneity across time:

- Treatment effects may vary as a function of the time after the treatment occurred (Callaway and Sant'Anna, 2021)
- The specification of multiple treatment indicator variables allows us to test how the treatment effect evolves; see Datta, Knox, and Bronnenberg (2018) for an example application in the context of music streaming service adoption
- Model specification:

$$\begin{aligned} Y_{it} = {} & \delta^{ST} * I(t_0 \leq periods\_since\_treatment_{it} \leq t_1) \\ & + \delta^{MT} * I(t_2 \leq periods\_since\_treatment_{it} \leq t_3) \\ & + \delta^{LT} * I(t_4 \leq periods\_since\_treatment_{it} \leq t_5) \\ & + \mu_i + \gamma_t + \epsilon_{it} \end{aligned} \tag{14}$$

... where δ is an exemplary parameter vector denoting the short-term, mid-term and long-term treatment effects

Matched Difference-in-Difference

- Problem of causal inference: we can't observe the same unit with and without treatment
- Key idea: from a pool of untreated units, we find a unit that is as similar as possible to the treated unit, except for the treatment status → Differences in outcome can then be attributed to treatment (see also here)
- The key appeal of matching is that adjusting for many covariates is difficult in high-dimensional setting ("curse of dimensionality")
- Pre-treatment covariates: Think about variables that may explain the self-selection of units to treated and control conditions, e.g., pre-treatment outcomes may be predictive of latent characteristics if parallel trend assumption is violated (Miratrix, 2022); future treatment status of other individuals may (partly) adjust for unobserved characteristics, i.e., use future treated units as controls (Bapna, Ramaprasad, and Umyarov, 2018)

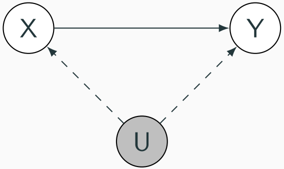


Figure: Self-selection into treated and control conditions

Matched Difference-in-Difference

- Step 1: Propensity score estimation:

$$Pr(Treated_i = 1) = Pr(\beta_0 + \alpha * Z_i + \eta_i > 0) \tag{15}$$

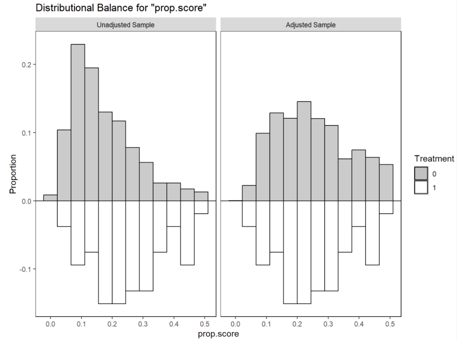
- Step 2: Compute treatment weight

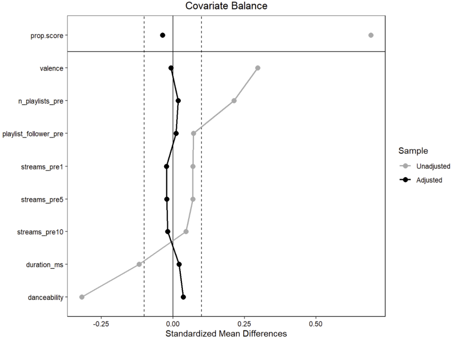
$$w_i = Treated_i + \frac{Pr(Treated_i = 1)(1 - Treated_i)}{(1 - Pr(Treated_i = 1))} \tag{16}$$

- Step 3: DiD estimation (estimated with *weighted* least squares):

$$Y_{it} = \delta Treatment_{it} + \mu_i + \gamma_t + \epsilon_{it} \tag{17}$$

Matched Difference-in-Differences





- Assessment of matching results by inspecting (1) the distributions of propensity scores in the treated and untreated groups, and (2) the standardized mean differences w.r.t. the matching covariates

Matched Difference-in-Difference

Key identifying assumptions:

- Unconfoundedness: the outcome is independent of the treatment after adjusting for observable covariates
- Overlap. Value range of covariates should be observed for both

$\rightarrow$ Both assumptions together imply “strong ignorability”

Limitation:

- Requires that the matching covariates are stable over time

Synthetic Control

"Arguably the most important innovation in the policy evaluation literature in the last 15 years."
— Athey and Imbens (2017)

- Generalization of DiD in settings with one treated unit and multiple untreated units
- We construct one synthetic control unit as a weighted average of all available control units
- Weights are chosen to maximize similarity between treated and control prior to treatment
- Key identifying assumptions: 1) Treated and synthetic control unit are similar prior to treatment (parallel trends assumption), 2) no unobserved events coincide with treatment that differentially impact treated and control units, and 3) control units unaffected by treated units.
- Assumption 1 is usually verified graphically, while 2 and 3 must be argued theoretically.

Synthetic Difference-in-Differences

- Synthetic Control weights
  - Weights ω to aggregate controls to be similar to treated
  - Learned pre-treatment, applied post-treatment
  - Assumption: pre-treatment relation holds post-treatment
- Synthetic Time weights
  - Weights λ to balance pre-treatment time periods with post-treatment ones
  - Learned using controls, applied to treated
  - Assumption: predictive periods for controls are also predictive for treated
- Double-robust: bias is corrected with either weights
  - Automatically addresses concerns regarding parallel trends (like SC)
  - Allows for unit fixed effects (unlike SC)
- See e.g., Berman and Israeli (2022) for an application

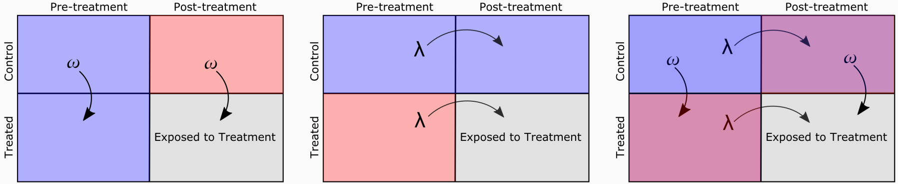


Figure: From Synthetic Control to Synthetic Diff-in-Diff (Hirshberg, 2020)

Comparison of Difference-in-Difference problems

Standard Diff-in-Diff

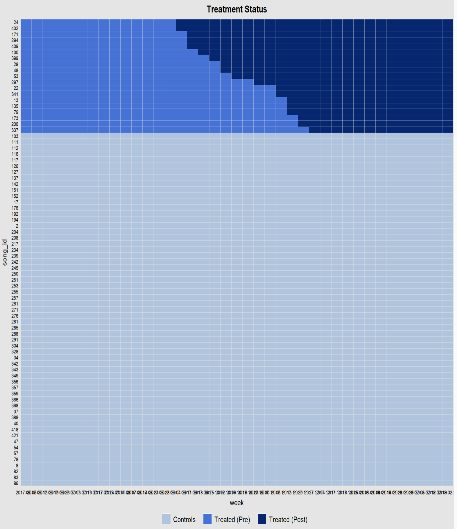
$$\hat{\tau}^{\mathrm{DID}} = \underset{\tau,\gamma,\mu}{\operatorname{argmin}} \sum_{i,t} \left(Y_{it} - \gamma_t - \mu_i - \tau W_{it}\right)^2$$

Synthetic Control

$$\hat{\tau}^{\mathrm{SC}} = \underset{\tau,\gamma}{\operatorname{argmin}} \sum_{i,t} \left(Y_{it} - \gamma_t - \tau W_{it}\right)^2 \times \omega_i^{\mathrm{SC}}$$

Synthetic Diff-in-Diff

$$\hat{\tau}^{\mathrm{SDID}} = \underset{\tau,\gamma,\mu}{\operatorname{argmin}} \sum_{i,t} \left(Y_{it} - \gamma_t - \mu_i - \tau W_{it}\right)^2 \times \omega_i \times \lambda_t$$



- Absorbing treatment may occur in different time period
- Units are part of a treatment group based on their first treatment period

$$G_i = \begin{cases} \min\{t : Treated_{i,t} = 1\} \text{ if } Treated_{i,\cdot} \\ \infty \text{ if } Treated_{i,t} = 0 \; \forall \; t \end{cases}$$

- TWFE requires no variation in treatment effects (Goodman-Bacon, 2021)!
- $\delta^{\cdot,T}$ in the heterogeneous case can be "cross contaminated"
- Solution: Separate $ATT(G_i, t)$ using only never or not yet treated units (Callaway and Sant'Anna, 2021)
- ATTs of interest (e.g., 2 periods after adoption) estimated as weighted average of $ATT(G_i, t)$s

Tripple differences

- Refinement of the definition of treated and control groups to "two dimensions"
- An eligible subset of a treatment-implementing unit receives treatment in $t = 2$

- Assessing the effect of a Diff-in-Diffs Workshop on research output:

implemented

UNSW

DASH faculty

eligible

Other faculty

not eligible

$$Y_{it} = \delta(\text{DASH}_i \times post_t) + \mu_i + \gamma_t + \epsilon_{it}$$

Tripple differences

- Refinement of the definition of treated and control groups to "two dimensions"
- An eligible subset of a treatment-implementing unit receives treatment in $t = 2$
- A similar unit does not implement treatment for any subset
- Assessing the effect of a Diff-in-Diffs Workshop on research output:

implemented

- UNSW
  - DASH faculty — eligible
  - Other faculty — not eligible

not implemented

- WU
  - DASH equiv. faculty — eligible
  - Other faculty — not eligible

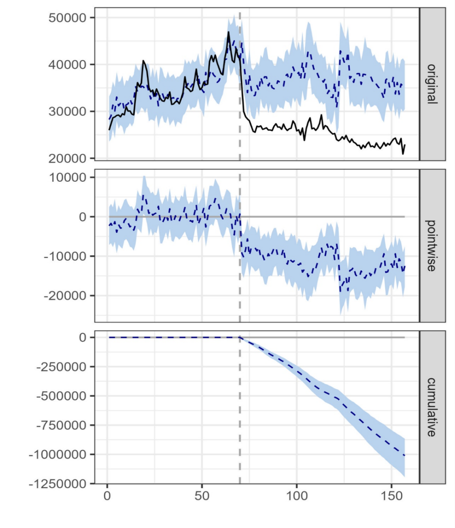
Tripple differences

- Refinement of the definition of treated and control groups to "two dimensions"
- An eligible subset of a treatment-implementing unit receives treatment in $t = 2$
- A similar unit does not implement treatment for any subset
- Assessing the effect of a Diff-in-Diffs Workshop on research output:

implemented		not implemented	
UNSW		WU	
DASH faculty	Other faculty	DASH equiv. faculty	Other faculty
eligible	not eligible	eligible	not eligible
treated	**not treated**		

$$Y_{it} = \delta(\text{UNSW}_i \times \text{DASH (equiv.)}_i \times post_t) + \mu_i + \gamma_t + \phi_{\text{UNSW}\times t} + \eta_{\text{DASH (equiv.)}\times t} + \epsilon_{it}$$

Bayesian structural time series model



- Bayesian state-space model approach combining 3 sources of information:
  1. Treated unit's systematic evolution (e.g., trend, seasonality)
  2. Evolution of other units that were predictive of the treated one's (like SC)
  3. Prior information for the parameters (e.g., expert knowledge on relationships)
- Parameters are learned in the pre-treatment period
- Treatment effect is the observed treated outcome minus the prediction

Questions?

We hope you enjoyed the workshop! :)

References I

- Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arkhangelsky, Dmitry (2023). *synthdid: Synthetic Difference-in-Difference Estimation*. R package version 0.0.9. URL: https://github.com/synth-inference/synthdid.
- Arkhangelsky, Dmitry et al. (Dec. 2021). "Synthetic Difference-in-Differences". In: *American Economic Review* 111.12, pp. 4088–4118. ISSN: 0002-8282. DOI: 10.1257/aer.20190159.
- Athey, Susan and Guido W. Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation". In: *Journal of Economic Perspectives* 31.2, pp. 3–32.
- Bapna, Ravi, Jui Ramaprasad, and Akhmed Umyarov (2018). "Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement?" In: *MIS Quarterly* 79.47, pp. 719–735.
- Bergé, Laurent (2018). "Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm". In: *CREA Discussion Papers* 13. URL: https://lrberge.github.io/fixest/.

- Berman, Ron and Ayelet Israeli (2022). "The value of descriptive analytics: Evidence from online retailers". In: *Marketing Sci.* forthcoming.
- Box, G.E.P., W.G. Hunter, and J.S. Hunter (1978). *Statistics for Experimenters*. John Wiley Sons, Inc., New York.
- Brodersen, Kay H. et al. (2014). "Inferring causal impact using Bayesian structural time-series models". In: *Annals of Applied Statistics* 9, pp. 247–274. URL: https://research.google/pubs/pub41854/.
- Callaway, Brantly and Pedro H.C. Sant'Anna (2021). *did: Difference in Differences*. R package version 2.1.2. URL: https://bcallaway11.github.io/did/.
- Callaway, Brantly and Pedro H.C. Sant'Anna (2021). "Difference-in-Differences with Multiple Time Periods". In: *Journal of Econometrics* 225.2, pp. 200–230.
- Cunningham, Scott (2021). *Causal inference: The mixtape*. Yale university press. URL: https://mixtape.scunning.com.
- Datta, Hannes, George Knox, and Bart J. Bronnenberg (2018). "Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery". In: *Marketing Science* 37.1, pp. 5–21.

- Flack, Evan and Edward Jee (2020). *bacondecomp: Goodman-Bacon Decomposition*. R package version 0.1.1. URL: https://CRAN.R-project.org/package=bacondecomp.
- Gelman, Andrew (2020). "Experimental Reasoning in Social Science". In: *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. Yale University Press, pp. 185–195.
- Goldfarb, Avi and Catherine Tucker (2014). "Conducting Research with Quasi-Experiments: A Guide for Marketers". In: *working paper*.
- Goldfarb, Avi, Catherine Tucker, and Yanwen Wang (2022). "Conducting Research in Marketing with Quasi-Experiments". In: *Journal of Marketing* 86.3, pp. 1–20.
- Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.
- Greifer, Noah (2022). *WeightIt: Weighting for Covariate Balance in Observational Studies*. R package version 0.13.1. URL: https://CRAN.R-project.org/package=WeightIt.

- Hernán, Miguel A., John Hsu, and Brian Healy (2019). "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks". In: *Chance* 32.1, pp. 42–49.
- Huntington-Klein, Nick (2021). "DAGitty — draw and analyze causal diagrams". In: *CREA Discussion Papers* 13. URL: http://www.dagitty.net/.
- — (2022). *The effect: An introduction to research design and causality*. Chapman and Hall/CRC. URL: https://theeffectbook.net.
- Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Miratrix, Luke (2022). "A devil's bargain? Repairing a Diff-in-Diff parallel trends assumption with matching". In: *https://www.youtube.com/watch?v=tKoxMovw9uY*.
- Mou, Hongyu, Licheng Liu, and Yiqing Xu (2023). *panelView: Visualizing Panel Data*. R package version 1.1.16. URL: https://CRAN.R-project.org/package=panelView.
- Pearl, Judea (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

- Rambachan, Ashesh (2023). *HonestDiD: Robust inference in difference-in-differences and event study designs*. R package version 0.2.1.
- Rambachan, Ashesh and Jonathan Roth (Feb. 15, 2023). "A More Credible Approach to Parallel Trends". In: *The Review of Economic Studies*, rdad018. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdad018. URL: https://academic.oup.com/restud/advance-article/doi/10.1093/restud/rdad018/7039335 (visited on 02/16/2023).
- Roth, Jonathan et al. (2023). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature". In: *arxiv: https://arxiv.org/abs/2201.01194*.
- Todri, Vilma (2022). "Frontiers: The Impact of Ad-Blockers on Online Consumer Behavior". In: *Marketing Science* 41.1, pp. 7–18.
- Varian, Hal R. (July 2016). "Causal Inference in Economics and Marketing". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.27, pp. 7310–7315. ISSN: 1091-6490. DOI: 10.1073/pnas.1510479113.
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund (2023). *R for Data Science*. 2nd ed. "O'Reilly Media, Inc.". URL: https://r4ds.hadley.nz.

References VI

- Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.
- Xu, Yiqing and Licheng Liu (2021). *gsynth: Generalized Synthetic Control Method*. R package version 1.2.1. URL: `https://yiqingxu.org/packages/gsynth/`.