

Contents

Preface	xv
I INTRODUCTION AND SETUP	1
1 Introduction	3
1.1 The classical DID design: Chapters 3 and 4	5
1.1.1 Definition of a classical DID design	5
1.1.2 Potential outcomes and target parameter	6
1.1.3 Three possible estimators: treated versus control, before–after, and DID	8
1.1.4 The parallel trends assumption	9
1.1.5 Parallel trends is weaker than the assumptions underlying treated-versus-control and before–after comparisons	10
1.1.6 Parallel trends remains a strong assumption, whose plausibility should be assessed	11
1.1.7 Relaxations of parallel trends	12
1.1.8 Application to the effect of the 2009 VAT reduction in French restaurants on profits, prices, and wages	13
1.2 Beyond the classical DID design: Chapters 5 to 8	14
1.2.1 Two-way fixed effects regressions	14
1.2.2 What does a TWFER estimate outside of the classical DID design?	15
1.2.3 Heterogeneity-robust DID estimators	18
1.2.3.1 Staggered adoption designs	18
1.2.3.2 Heterogeneous adoption designs	18
1.2.3.3 Staggered first switch designs	19
2 Data, Notation, and Assumptions	20
2.1 Data requirements: group-level panel data	20
2.2 Treatment and potential outcomes	20
2.3 Identifying assumptions	21
2.3.1 Exclusion restrictions	21
2.3.2 Parallel trends	22
2.4 The book’s perspective on statistical inference	24
2.4.1 A “model-based” perspective on statistical inference	24
2.4.2 Assuming independent outcomes between groups does not rule out common shocks* .	26
2.4.3 At what level should we cluster standard errors?*	27

II	THE CLASSICAL DID DESIGN	29
3	The Classical DID Design	31
3.1	Target parameters	33
3.2	Two-way fixed effects regressions	36
3.2.1	Static two-way fixed effects regressions	36
3.2.1.1	Assuming randomized treatment rather than parallel trends?	40
3.2.2	Event-study two-way fixed effects regressions	42
3.2.2.1	Estimating event-study effects	42
3.2.2.2	Pretrend tests	44
3.2.2.3	Application to the effect of compulsory licensing on innovation	50
3.3	Inference	52
3.3.1	Asymptotically valid CIs with many treated and control groups	53
3.3.2	When are G_0 and G_1 large enough to rely on CIs with HC2 standard errors and Bell and McCaffrey's critical values? Some simulations	54
3.3.3	What can researchers do when HC2-BM CIs seem unreliable?*	58
3.3.3.1	Few treated groups but many control groups	58
3.3.3.2	Few treated and control groups	61
3.3.3.3	Recommendations for practitioners	63
3.4	Limitations of pretrend tests	64
3.4.1	We can test for parallel trends before but not after treatment	64
3.4.1.1	Concomitant shocks	64
3.4.1.2	Other policies	65
3.4.2	Pretrend tests often lack power	65
3.4.3	Do pretrend tests lead to a pretesting problem?*	69
3.5	An alternative, imputation estimator	70
3.5.1	Imputation estimator	70
3.5.2	Three numerical equivalences	71
3.5.3	Comparing the properties of $\hat{\beta}_\ell^{\text{ic}}$ and $\hat{\beta}_\ell^{\text{imp}}$	72
3.5.4	Application to the effect of compulsory licensing on innovation	74
3.6	Estimating heterogeneous effects	74
3.6.1	Estimating the correlation between treatment effects and some covariates	74
3.6.2	Estimating the variance of group-specific effects*	77
3.6.3	Estimating the distribution of group-specific effects*	79
3.7	Nonlinear DID	80
3.7.1	Limited dependent variables	80
3.7.2	Sensitivity to functional form*	82
3.7.2.1	The parallel trends assumption may not be invariant to functional form	82
3.7.2.2	A necessary and sufficient condition to have parallel trends for any functional form	83
3.7.2.3	When should researchers worry about sensitivity to functional form?	84
3.7.3	Estimating quantile treatment effects*	86
3.8	Instrumental-variable DID estimators	88
3.9	Further topics	90
3.9.1	Imbalanced panels	90
3.9.2	Weighting	92

3.9.3	Accounting for and estimating spillover effects	95
3.10	Practitioners' dos and don'ts	96
3.11	Appendix*	98
3.11.1	The Frisch–Waugh–Lovell theorem	98
3.11.2	Proof of (3.15).....	99
4	Alternatives to the Parallel Trends Assumption	101
4.1	TWFE and DID estimators with control variables	101
4.1.1	Conditional parallel trends.....	101
4.1.2	TWFERs with control variables	103
4.1.3	DID estimators with control variables	104
4.1.3.1	Parametric estimators.....	104
4.1.3.2	Nonparametric estimators.....	107
4.1.3.3	Matched DID.....	109
4.1.3.4	Controlling for group-specific linear trends	109
4.1.3.5	Triple-difference estimators*	111
4.1.3.6	When to include control variables in the estimation?	112
4.1.4	Controlling for the lagged outcome?	114
4.1.4.1	Autoregressive model for the outcome without treatment*	114
4.1.4.2	Self-selection, and parallel trends conditional on the baseline outcome.....	116
4.1.5	Computing DID estimators with controls in Stata and R	119
4.1.6	Application to the effect of compulsory licensing on innovation	121
4.2	Interactive fixed effects, synthetic controls, and synthetic DID.....	122
4.2.1	Interactive fixed effects	123
4.2.2	Synthetic control and synthetic DID.....	128
4.2.2.1	Synthetic control.....	128
4.2.2.2	Synthetic DID	133
4.2.3	Application to the effect of compulsory licensing on innovation	137
4.3	Bounded differential trends	141
4.4	Practitioners' dos and don'ts	148
4.5	Appendix*	149
4.5.1	Proof of Theorem 4.1	149
4.5.2	Comparing the variances of DID estimators with and without covariates	150
4.5.3	Details on the computation of the TWFE-IFE estimator.....	152

III BEYOND THE CLASSICAL DESIGN 153

5 The Pitfalls of TWFE Estimators Outside of the Classical DID Design 155

5.1	A decomposition of $\hat{\beta}^{\text{fe}}$	157
5.2	$\hat{\beta}^{\text{fe}}$ may be biased for ATT	159
5.3	$\hat{\beta}^{\text{fe}}$ may not estimate a convex combination of effects.....	162
5.4	Decompositions of related estimators.....	164

5.5	Stata and R commands to compute the implicit weights of TWFERs and first-difference regressions	164
5.6	Application to the effect of newspapers on turnout in US elections	165
5.6.1	Basic TWFER	165
5.6.2	TWFER with state-year FEs	166
5.6.3	First-difference regression with state-year FEs	167
5.7	Next steps	169
6	Variation in Treatment Timing: Staggered Adoption Designs	170
6.1	Target parameters	172
6.2	Two-way fixed effects estimators	174
6.2.1	Static two-way fixed effects estimator	174
6.2.1.1	Decomposition of $\hat{\beta}^{\text{fe}}$	174
6.2.1.2	Application to the effect of unilateral divorce laws on divorces	176
6.2.1.3	The origin of the negative weights	177
6.2.1.4	Assuming randomized treatment timing instead of parallel trends?	182
6.2.2	Event-study two-way fixed effects regressions	185
6.2.2.1	Decomposition of ES-TWFERs	188
6.2.2.2	Application to the effect of unilateral divorce laws on divorces	192
6.2.3	Local-projection regressions*	193
6.3	Heterogeneity-robust estimators	194
6.3.1	Target parameters	194
6.3.2	DID estimators	195
6.3.2.1	Estimators	195
6.3.2.2	Extensions	196
6.3.2.3	Inference	198
6.3.2.4	Numerical equivalences with regression coefficients	199
6.3.2.5	Computation in Stata, R, and Python	199
6.3.3	Imputation estimators	201
6.3.3.1	Some numerical equivalences	201
6.3.3.2	Extensions	203
6.3.3.3	Inference	203
6.3.3.4	Computation in Stata and R	203
6.3.4	A comparison of heterogeneity-robust estimators	204
6.3.4.1	Variance	204
6.3.4.2	Confidence intervals coverage	205
6.3.4.3	Bias	206
6.3.4.4	Implementation in Stata, R, and Python	207
6.3.5	Application to the effect of unilateral divorce laws on divorces	209
6.4	TWFE and HR estimators in 13 SADs in political science	213
6.5	Estimating heterogeneous effects	218
6.6	Nonlinear DID	220
6.6.1	Limited dependent variables	220
6.6.1.1	Application to the effect of free trade agreements on trade	221
6.6.2	Estimating quantile treatment effects*	225

6.7	Further topics*	226
6.7.1	Using always-treated groups as controls?.....	226
6.7.2	Imbalanced panels.....	226
6.7.3	Weighting.....	227
6.7.4	Accounting for and estimating spillover effects	227
6.8	Practitioners' dos and don'ts	228
6.9	Appendix*	230
7	Variation in Treatment Dose: Heterogeneous Adoption Designs	232
7.1	Target parameters	236
7.1.1	The conditional-average-slope function	237
7.1.2	Two unconditional averages of slopes.....	238
7.2	TWFERs in heterogeneous adoption designs	240
7.2.1	Parallel trends assumption.....	240
7.2.2	$\hat{\beta}^{\text{te}}$ may not identify a convex combination of slopes	241
7.2.3	The origin of the negative weights	244
7.2.4	Assuming randomized treatment dose rather than parallel trends*	246
7.3	Heterogeneity-robust estimators	248
7.3.1	Parallel trends assumptions and a fundamental decomposition.....	248
7.3.2	Designs with stayers or quasi-stayers	250
7.3.2.1	Identification	250
7.3.2.2	Estimation with stayers	251
7.3.2.3	Estimation without stayers but with quasi-stayers	254
7.3.3	Designs without stayers or quasi-stayers.....	257
7.3.4	Testing the null that there are quasi-stayers.....	259
7.3.5	Application to the effect, on US employment, of eliminating a potential tariffs' spike on Chinese imports	260
7.4	Practitioners' dos and don'ts	261
7.5	Appendix*	262
7.5.1	Proof of Theorem 7.1	262
7.5.2	Proof of Theorem 7.2.....	263
7.5.3	Proof of Theorem 7.3.....	264
8	General Designs	265
8.1	Static TWFER	267
8.1.1	Decomposition of TWFERs under the usual parallel trends assumption	267
8.1.2	Decomposition of TWFERs under a parallel trends assumption in a counterfactual where groups' treatment does not change.....	269
8.1.3	Decomposition of TWFERs with randomly assigned treatments*	271
8.1.4	Extensions	274
8.1.4.1	TWFERs with several treatments.....	274
8.1.4.2	Two-stage least squares TWFERs, and Bartik regressions*	275
8.2	Distributed-lag TWFER	276
8.3	Heterogeneity-robust estimators for staggered first switch designs	279
8.3.1	Staggered first switch designs.....	279
8.3.2	Parallel trends assumption.....	280

8.3.3	Target parameters and estimators.....	282
8.3.3.1	Building blocks: the actual-versus-status-quo effects	282
8.3.3.2	A generalization of the ATT_t event-study effects to staggered first switch designs	284
8.3.3.3	Path-specific event-study effects	286
8.3.3.4	Normalized event-study effects.....	287
8.3.3.5	Distributed-lag regressions allowing for heterogeneous effects across groups	291
8.3.3.6	A generalization of ATT to staggered first switch designs.....	291
8.3.3.7	Pretrend estimators.....	294
8.3.3.8	Testing if lagged treatments affect the outcome	295
8.3.4	Inference.....	296
8.3.5	Extensions.....	296
8.3.5.1	Continuous treatment at period one	296
8.3.5.2	Estimators with control variables.....	297
8.3.5.3	Estimating heterogeneous effects	297
8.3.5.4	Estimators with several treatments*	298
8.3.5.5	Instrumental-variable DID estimators*	301
8.3.5.6	The initial-conditions problem, and allowing for dynamic effects only up to a prespecified number of lags*	301
8.3.6	Application to the effect of newspapers on turnout in US elections	302
8.4	Heterogeneity-robust estimators, without dynamic effects	308
8.4.1	Switchers and stayers designs.....	308
8.4.2	Parallel trends assumption.....	309
8.4.3	Target parameters and estimators.....	310
8.4.3.1	Target parameters	310
8.4.3.2	Estimators.....	311
8.4.3.3	Pretrend estimators.....	312
8.4.4	Imputation estimators.....	313
8.4.5	Extensions.....	314
8.4.5.1	Continuous treatment at period one	314
8.4.5.2	Estimators with several treatments*	315
8.4.5.3	Instrumental-variable DID estimators*	316
8.4.6	Computation in Stata and R	317
8.4.7	Application to the effect of newspapers on turnout in US elections	317
8.5	Conclusion	318
8.5.1	TWFE and HR estimators in general designs.....	318
8.5.2	HR estimators in designs without stayers?	320
8.5.2.1	Using quasi-stayers.....	320
8.5.2.2	Changing the treatment's definition.....	321
8.5.2.3	Functional-form assumptions	321
8.6	Practitioners' dos and don'ts	321
8.7	Appendix*	323
8.7.1	Proof of Theorem 8.1	323
8.7.2	Proof of Theorem 8.2.....	323
8.7.3	Proof of Theorem 8.3.....	323
8.7.4	Proof of Theorem 8.4.....	324
8.7.5	Proof of Theorem 8.5.....	324
8.7.6	Proof of Theorem 8.6.....	324

Bibliography	327
Index	339

CHAPTER 1 Introduction

Many research questions fall under the umbrella of causal inference, that is, estimating the effect of a *treatment* on an *outcome*. To this end, the gold standard method consists of conducting a randomized controlled trial (RCT), in which units' exposure to treatment is determined randomly. We can then compare the average outcomes between treated and untreated units to estimate, without bias, the average treatment effect (ATE) in the population under study.

Nevertheless, there are questions for which it is difficult, or even impossible, to conduct an RCT. For example, it is not possible to randomly increase imports from China for certain countries and not for others, in order to measure the causal effect of these imports on employment in importing countries. Similarly, randomly assigning firms to groups subject to a high or low minimum wage, in order to study the effect of the minimum wage on employment, has never been done. Moreover, even when RCTs are feasible, they sometimes suffer from a lack of *external validity*: their results cannot always be generalized from the experimental sample to the entire population for which the researcher wishes to estimate the ATE. For example, research ethics require informed consent from participants in an RCT. Thus, in our hypothetical minimum wage example, the RCT would provide an unbiased estimate of the minimum wage effect among firms that agreed to participate, but this effect would likely differ from the effect among all firms.

To answer causal inference questions for which RCTs are impossible to implement, or would not be generalizable to the entire population, researchers typically rely on what are called *natural experiments*, in which a treated group and a control group arise spontaneously, without intervention by the researcher. Typically, natural experiments take the form of public policy changes. For example, one US state raises its minimum wage while a neighboring state does not, thus providing researchers with a treated group subject to a higher minimum wage, and a control group facing a lower minimum wage. Natural experiments often concern an entire region or province, so that results obtained from these studies generally apply to large representative populations, unlike the results of RCTs.

Yet, in a natural experiment it is not the researcher but the policymaker who decides who will be treated and who will not. This has two crucial consequences. First, since policymakers do not randomly choose where to apply a reform, the treated and control groups may not be comparable. Therefore, a simple comparison of their average outcomes may provide a biased estimator of the ATE. In the minimum wage example, the treated and control states may have different employment levels even before the treated state raises its minimum wage. Comparing their employment

levels after this increase then amounts to measuring the sum of the minimum wage effect and pre-existing differences between the two states. This comparison does not isolate the effect of the minimum wage itself.

Second, the evolution of public policies is often full of twists and turns. A researcher conducting an RCT would simply assign some units to treatment and others to control, and then observe their trajectories without changing this initial assignment, for as long as their research question requires. By contrast, in our example, policymakers in states that raised their minimum wage may subsequently decide to lower it, or to raise it again. In control states, which initially made no changes, authorities may choose to modify the minimum wage later. Furthermore, some states may implement large minimum wage increases, while others make only modest adjustments. These multiple variations complicate the analysis and may prevent researchers from estimating the effects they originally intended to measure. The complexity of natural experiments leveraged by researchers is a central theme of this book.

Faced with these difficulties inherent to natural experiments—imperfect comparability between groups and sometimes complex evolution of public policies—researchers have progressively developed tools to still be able to isolate the treatment’s causal effect. Among these tools, differences-in-differences (DID) occupy a central place.

The use of DID is not recent. In the 1840s, Ignaz Semmelweis, a young assistant physician at the Vienna maternity hospital, was confronted with high mortality among his patients due to childbed fever. The city then had two clinics, and mortality was much higher in the first, run by physicians, than in the second, run by midwives. Semmelweis suspected a hygiene problem: physicians performed autopsies before attending deliveries, without always washing their hands afterward. Perhaps the “cadaverous particles” on their hands were contaminating their patients? To test his hypothesis, Semmelweis required all caregivers to wash their hands with chlorinated water before performing a delivery. After implementing this protocol, mortality dropped sharply in the clinic run by physicians, while it remained unchanged in the clinic run by midwives. This comparison of the evolution of an outcome (mortality) before and after the implementation of a treatment (handwashing) in two groups constitutes the first known example of a DID estimator (Semmelweis, 1983; Baker, Callaway, Cunningham, Goodman-Bacon and Sant’Anna, 2025). While Semmelweis’s study demonstrated that handwashing is a simple but extremely effective means of preventing childbed fever, his theory was ridiculed by his peers, who refused to admit that their own practices were responsible for their patients’ deaths. In 1865, Semmelweis, facing increasingly virulent criticism, reportedly suffered a nervous breakdown and was committed to an asylum by his colleagues, where he died fourteen days later, under obscure circumstances.

Another early example of a DID is John Snow’s study, aimed at determining whether cholera is transmitted through air or water, the two dominant theories in the 1850s. Snow exploited a change in the water supply of a London neighborhood,

from polluted water drawn from central London, to cleaner water taken upstream. He showed that after this change, cholera epidemics became less frequent in this neighborhood, while their frequency remained unchanged in neighboring areas that continued to draw water from central London. Since air quality had remained stable over time in both the treated and control neighborhoods, his study demonstrated that cholera is transmitted through water (Snow, 1856; Lechner, 2011).

While DID is an old method, recent developments have considerably expanded its applicability. These advances have notably adapted the method to complex natural experiments, where units can be exposed to treatment at staggered dates over time, can switch out of treatment, and can receive heterogeneous treatment doses. These extensions are important, because complex natural experiments are very common in practice. This complexity is already present in the pioneering work of Semmelweis, where both groups are treated but with different intensities: handwashing improves hygiene more in the physicians' clinic, because they perform autopsies.

The purpose of this book is to introduce applied researchers to modern Differences-in-Differences (DID) estimators, tailored to potentially complicated natural experiments. This introduction gives an overview of the book, in a simplified setup that ignores statistical uncertainty. Section 1.1 describes the DID method in the context of the canonical “two-by-two” design, elaborating on the so-called *parallel trends* assumption that underlies the method. Section 1.2 is devoted to recent extensions of the DID method to more complex designs.

1.1 The classical DID design: Chapters 3 and 4

1.1.1 Definition of a classical DID design

In a so-called classical DID design, the treatment is a binary variable: one can receive or not receive the treatment, but its intensity does not vary. The treatment is also *absorbing*: once treated, a unit cannot become untreated again. Finally, there is no variation in the timing of treatment: all treated units become treated at the same time. In this framework, one can define two groups of units and two periods: group s switches from untreated to treated between periods 1 and 2, while group n remains untreated at both dates. A “cell” refers to a pair (g, t) , that is, a given group at a given period. We thus have four cells: $(s, 1)$, $(s, 2)$, $(n, 1)$, and $(n, 2)$.

Let us illustrate this notation in the context of our running example in this chapter. In July 2009, the value-added tax¹ (VAT) rate in sit-down restaurants dropped from 19.6% to 5.5% in France. Before the reform, a large number of restaurant owners had signed a contract with the government, committing to share the benefits of the VAT reduction equally among their customers, their employees, and themselves. To determine whether they kept their promise, Benzarti and Carloni (2019) evaluate the effect of this reform on three outcome variables: the prices that sit-down

¹For readers unfamiliar with the VAT, it is similar to a sales tax.

restaurants charge their customers, the wages they pay their employees, and their profits. To isolate the causal effect of the reform, they use takeaway restaurants, a very similar economic sector that did not benefit from the VAT reduction, as the control group. Thus, in this example, s represents sit-down restaurants, n represents takeaway restaurants, period 1 could for example correspond to year 2008, the year before the reform, while period 2 would correspond to year 2009, the year the reform was implemented.

1.1.2 Potential outcomes and target parameter

Measuring the impact of a treatment requires comparing what happens with the treatment to what would have happened without it. To formalize what we are trying to measure, suppose that each cell has two *potential* outcomes, without treatment and with treatment. For $g \in \{s, n\}$ and $t \in \{1, 2\}$, let $Y_{g,t}(0)$ be the potential outcome of group g at period t if cell (g, t) is not treated, and let $Y_{g,t}(1)$ be its potential outcome if it is treated (Neyman, Dabrowska and Speed, 1990; Rubin, 1974). $Y_{g,t}(0)$ and $Y_{g,t}(1)$ are referred to as the *untreated and treated outcomes* for cell (g, t) , respectively.



#1 In the VAT example, what does $Y_{s,2}(0)$ represent?

Solution. In the VAT example, $Y_{s,2}(0)$ represents the average profit of sit-down restaurants in 2009, had those businesses not benefited from a reduced VAT rate.

$Y_{s,2}(1)$ represents the average profit in the same businesses at the same date, had they benefited from a reduced VAT rate.

Let $Y_{g,t}$ denote the observed outcome of group g at period t , and let $D_{g,t}$ be an indicator variable equal to 1 if the cell is treated and to 0 otherwise. If cell (g, t) is untreated, its observed outcome corresponds to its untreated outcome: $D_{g,t} = 0$ implies that $Y_{g,t} = Y_{g,t}(0)$. If (g, t) is treated, its observed outcome corresponds to its treated outcome: $D_{g,t} = 1$ implies that $Y_{g,t} = Y_{g,t}(1)$. Thus, we have the following relationship between the observed and potential outcomes:

$$Y_{g,t} = (1 - D_{g,t})Y_{g,t}(0) + D_{g,t}Y_{g,t}(1).$$

For each cell (g, t) , only one of its two potential outcomes is observed. The unobserved potential outcome of (g, t) is called its *counterfactual outcome*.

 **#2** In the VAT example, which of the two potential outcomes is unobserved for sit-down restaurants in 2009?

Solution. $Y_{s,2}(0)$, the average profit that sit-down restaurants would have earned in 2009 if they had not benefited from a reduced VAT rate, is not observed: this is their counterfactual outcome.

Our goal will be to estimate

$$Y_{s,2}(1) - Y_{s,2}(0). \quad (1.1)$$

 **#3** Interpret this quantity, in general and in the context of the VAT example.

Solution. $Y_{s,2}(1) - Y_{s,2}(0)$ is the difference between the two potential outcomes of group s at period 2, with and without treatment. This difference thus represents the *effect* of the treatment on the outcome in this cell. In the VAT example, $Y_{s,2}(1) - Y_{s,2}(0)$ represents the difference between average profits among sit-down restaurants in 2009, with and without the VAT reduction.

As $(s, 2)$ is the only treated cell, $Y_{s,2}(1) - Y_{s,2}(0)$ is the average treatment effect on the treated (ATT).

 **#4** Can we easily compute ATT?

Solution. Computing ATT requires computing the difference between two quantities, $Y_{s,2}(1)$ and $Y_{s,2}(0)$. Since group s is treated at period 2, $Y_{s,2}(1)$ is equal to $Y_{s,2}$, the observed outcome of group s at period 2. Therefore, $Y_{s,2}(1)$ is observed and does not need to be estimated. On the other hand, $Y_{s,2}(0)$ —the

outcome that s would have had at period 2 if it had not been treated—is unobserved and must be estimated: this is what is called the *fundamental problem of causal inference*.

1.1.3 Three possible estimators: treated versus control, before–after, and DID

To estimate $Y_{s,2}(0)$, a first possibility is to use $Y_{n,2}$. Replacing $Y_{s,2}(0)$ with its estimator in (1.1), we can use

$$Y_{s,2} - Y_{n,2} \quad (1.2)$$

to estimate ATT. This estimator is a treated-versus-control comparison of the outcomes of s and n at period 2. In the VAT example, (1.2) compares average profits in 2009, among sit-down restaurants and among takeaway restaurants. However, since the treated group was not chosen randomly, $Y_{n,2}$ might not be a good estimator of $Y_{s,2}(0)$ —the untreated outcome of s at period 2. In the VAT example, the profits of takeaway restaurants in 2009 may not be a good counterfactual of the profits that sit-down restaurants would have had in 2009 without the VAT reduction. For instance, perhaps sit-down and takeaway restaurants already had different profit levels in 2008, the year before the reform.

To estimate $Y_{s,2}(0)$, a second possibility is to use $Y_{s,1}$. Replacing $Y_{s,2}(0)$ with its estimator in (1.1), we can use

$$Y_{s,2} - Y_{s,1} \quad (1.3)$$

to estimate ATT. This estimator is a before–after comparison, equal to the change in the outcome of group s between periods 1 and 2. In the VAT example, (1.3) compares the average profits of sit-down restaurants in 2009 and in 2008. Nevertheless, since the economic context changed considerably between the two periods, it is possible that the profits of sit-down restaurants would have changed even without the VAT reduction. Thus, their profits in 2008 might not be a good estimator of what their profits would have been in 2009 without the VAT reduction.

Finally, a third possibility to estimate $Y_{s,2}(0)$ is to use $Y_{s,1} + Y_{n,2} - Y_{n,1}$, that is, the sum of the outcome of s at period 1 and the change in the outcome of n between the two periods. Replacing $Y_{s,2}(0)$ with its estimator in (1.1), we obtain the following estimator:

$$\text{DID} := Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1}). \quad (1.4)$$

This estimator is called a *difference-in-differences*. It combines the before–after difference in (1.3) and the treated-versus-control difference in (1.2), to ultimately compare the *outcome changes* of s and n between periods 1 and 2. In the VAT example, (1.4) compares the change in average profits, from 2008 to 2009, between sit-down and takeaway restaurants.



#5 For DID to be a reliable estimator of ATT, what assumption must be satisfied?

Solution. The parallel trends assumption, discussed in detail in the next section.

1.1.4 The parallel trends assumption

DID compares the outcome changes of s and n . Imagine that even if group s had not been treated, its outcome would have increased more than that of group n . For example, perhaps profits would have increased more (or decreased less) between 2008 and 2009 among sit-down restaurants, even if they had not benefited from a reduced VAT rate in 2009. In this case, DID will be strictly positive even if the treatment has no effect. Thus, for DID to be reliable, a fundamental assumption must be satisfied: the *parallel trends* (PT) assumption.

This assumption states that without treatment, the average outcome would have followed the same evolution in both groups. This is illustrated in Figure 1.1. The black line represents the actual outcome evolution of s . The solid gray line represents the actual evolution of n , without treatment at any date. The dashed gray line represents the counterfactual evolution of s , if it had not been treated at any date. In the scenario presented in the figure, the solid and dashed gray lines are parallel, so

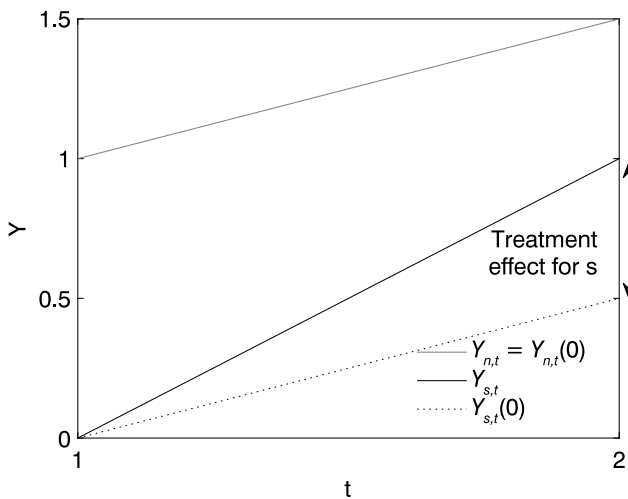


Figure 1.1: Illustration of the parallel trends assumption

the PT assumption is satisfied. In this case, DID correctly estimates the treatment effect. The outcome of s increases from 0 to 1, that of n increases from 1 to 1.5, so that $\text{DID} = 1 - 0 - (1.5 - 1) = 0.5$. Also, 0.5 indeed corresponds to the true treatment effect for s at period 2: $Y_{s,2}(1) = 1$, $Y_{s,2}(0) = 0.5$, so $Y_{s,2}(1) - Y_{s,2}(0) = 1 - 0.5 = 0.5$.

Mathematically, the PT assumption requires that

$$Y_{s,2}(0) - Y_{s,1}(0) = Y_{n,2}(0) - Y_{n,1}(0) : \quad (1.5)$$

without treatment, s and n would have experienced the same change in their outcome. In the VAT example, this means that the average profits of restaurants in both groups would have followed the same evolution if sit-down restaurants had not benefited from a reduced VAT rate. Under this assumption, DID identifies the treatment effect in group s at period 2. Specifically,

$$\begin{aligned} \text{DID} &= Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1}) \\ &= Y_{s,2}(1) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0)) \\ &= Y_{s,2}(1) - Y_{s,2}(0) + Y_{s,2}(0) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0)) \\ &= Y_{s,2}(1) - Y_{s,2}(0). \end{aligned} \quad (1.6)$$



#6 Justify each equality in this derivation.

Solution. The first equality follows from the definition of DID. The second holds because $(s, 2)$ is treated and the three other cells are not. The third is obtained by adding and subtracting $Y_{s,2}(0)$. The fourth relies on the PT assumption.²

1.1.5 Parallel trends is weaker than the assumptions underlying treated-versus-control and before-after comparisons

DID identifies ATT under the PT assumption. For the treated-versus-control comparison in (1.2) to identify ATT, a sufficient condition is that

$$Y_{s,t}(0) = Y_{n,t}(0) \quad \text{for all } t. \quad (1.7)$$

This condition requires that s and n have the same untreated outcome *levels*, whereas the PT assumption in (1.5) only requires that both groups have the same

²As mentioned earlier, we abstract here from statistical uncertainty. When accounting for it, as we do in the next chapters, the DID estimator differs from ATT though it is equal to it in expectation.

untreated outcome *trends*. Thus, (1.7) is a stronger assumption than (1.5): if (1.7) holds, (1.5) holds as well, but the converse is not true. For example, in Figure 1.1, (1.5) holds but (1.7) does not: s and n have different outcome levels without treatment.³

Similarly, a sufficient condition for the before–after comparison in (1.3) to identify ATT is

$$Y_{g,2}(0) = Y_{g,1}(0) \quad \text{for all } g. \quad (1.8)$$

This condition requires that the untreated outcomes of s and n do not change over time, whereas the PT assumption only requires that s and n experience the same change in their untreated outcomes. For example, in Figure 1.1, (1.5) is satisfied while (1.8) is not: in both groups, the outcome without treatment changes over time.

1.1.6 Parallel trends remains a strong assumption, whose plausibility should be assessed

To assess the plausibility of the PT assumption, researchers conduct *pretrend tests*, also sometimes referred to as *placebo tests*, which consist of checking whether the outcomes of s and n followed parallel trends before s was treated. Figure 1.2 illustrates two cases with three periods ($t = 0, 1, 2$). In the left panel, the trends of s and n are perfectly parallel from period 0 to 1, before s was treated. This suggests that the PT assumption is plausible: if the two groups followed parallel trends before treatment, it may be reasonable to assume that they would have continued to evolve in parallel between periods 1 and 2 had s not been treated. By contrast, in the right panel, the trends of the two groups clearly diverge from period 0 to 1, which calls into question the validity of the PT assumption.

Although useful, pretrend tests have two main weaknesses. First, one can only test the PT assumption before treatment, not after, since the outcome without treatment of s is no longer observable once s is treated. Parallel trends before treatment suggests, but does not guarantee, that trends would have been parallel afterward in the absence of treatment. Second, recent work has shown that these tests often have limited statistical power: they may fail to detect trend differences large enough to substantially bias DID estimators (Roth, 2022). Thus, the results of a DID estimation will be all the more credible when the pretreatment DID is close to zero and estimated precisely, with a small statistical margin of error.

³Equation (1.7) is sufficient but not necessary for the treated-versus-control comparison in (1.2) to identify ATT. The necessary and sufficient condition for this result is that (1.7) holds at $t = 2$. This condition is neither stronger nor weaker than (1.5). However, assuming (1.7) only at $t = 2$ does not seem substantively weaker than assuming this condition for all $t \in \{1, 2\}$. Indeed, (1.7) is testable at $t = 1$, since s and n are not treated at that date. It seems to us that when s and n have different outcome levels at $t = 1$, few researchers would be willing to assume that their outcome levels would have been identical at $t = 2$ in the absence of treatment.

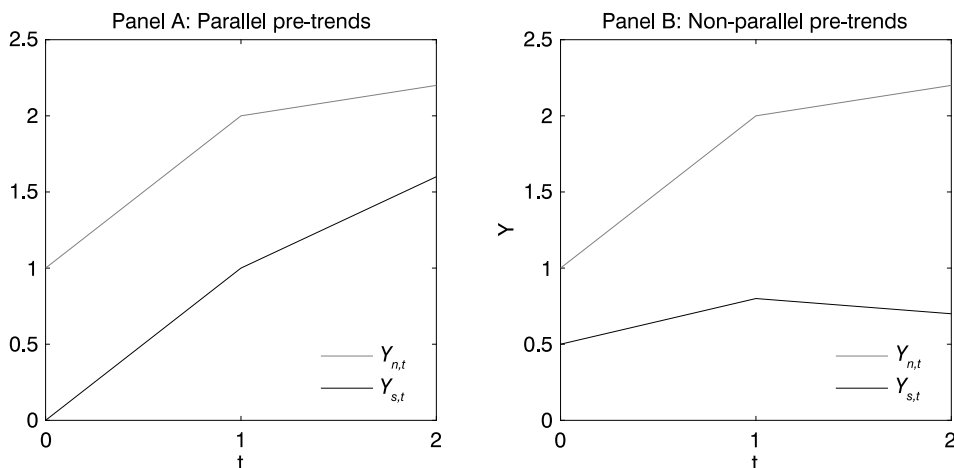


Figure 1.2: Illustrations of pretrend tests

1.1.7 Relaxations of parallel trends

In light of the above, there are at least three instances where one may want to relax the PT assumption: (1) pretrend tests show that the outcomes of the two groups were not following parallel evolutions before treatment; (2) pretrend tests fail to reject the hypothesis that evolutions were parallel before treatment, but these tests lack power; (3) pretrend tests are not rejected and have good statistical power, but one suspects that even without treatment, trends would have diverged in the posttreatment period, for example due to a shock (an economic or political crisis, an epidemic) occurring at the same time as treatment, and affecting the treatment and control groups differently.

Chapter 4 reviews the three main approaches in the literature to relax the PT assumption. A first approach replaces the PT assumption with a conditional parallel trends (CPT) assumption: without treatment, treated and untreated units *with the same sociodemographic characteristics* (for example) would have experienced the same change in their outcome variable (see, e.g., Heckman, Ichimura and Todd, 1997; Abadie, 2005). This assumption is weaker than the PT assumption: under the CPT assumption, treated and untreated units with different sociodemographic characteristics can have different outcome trends. Nevertheless, controlling for sociodemographic variables does not automatically increase results' credibility. Choosing which control variables to include gives the researcher significant leeway to manipulate the estimated effect or its statistical significance, and there is evidence that researchers do engage in p-hacking when making that choice. Adding controls may only be justified when doing so makes pretrends substantially smaller.

A second approach uses so-called “interactive fixed effects”, “synthetic control”, or “synthetic DID” estimators (see, in particular, Xu, 2017; Abadie, Diamond and Hainmueller, 2010; Arkhangelsky, Athey, Hirshberg, Imbens and Wager, 2021). These approaches substantially relax the PT assumption by assuming that the

potential outcome without treatment follows a factor model. At the same time, these estimators only have proven theoretical guarantees when the number of pretreatment periods tends to infinity. Simulations suggest that their behavior can be satisfactory with approximately 30 to 35 periods before treatment, but researchers do not always have that many pretreatment periods in their data. An additional difficulty is that these methods require choosing tuning parameters, but available theory provides few operational guidelines for making these choices. Ongoing work is addressing these two limitations, and constitutes promising research avenues.

Finally, a third approach assumes that posttreatment differential trends cannot exceed M times the largest differential trend observed before treatment, a so-called bounded differential trends assumption (Manski and Pepper, 2018; Rambachan and Roth, 2023). While intuitively appealing, this approach may lead to statistically insignificant results, including when pretrends are small and precisely estimated. Proposing estimation methods with greater statistical power under the bounded differential trends assumption is another promising research avenue.

1.1.8 Application to the effect of the 2009 VAT reduction in French restaurants on profits, prices, and wages

Figure 1.3, reproduced from Benzarti and Carloni (2019), presents DID estimates comparing the evolution of the average of the logarithm of profits, in sit-down restaurants and in takeaway restaurants, from year 2008, that precedes the reform, to year t .

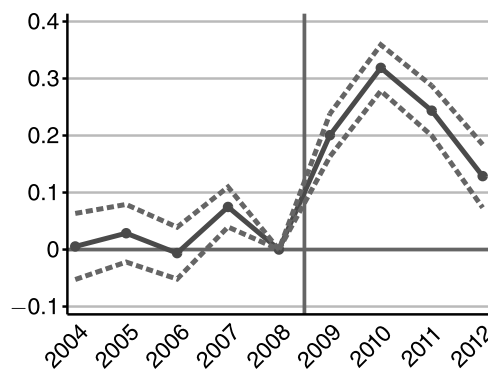


Figure 1.3: The effect of the 2009 VAT reduction on restaurants' profits

Note: This figure, reproduced with permission from Panel E of Figure 2 in Benzarti and Carloni (2019) © American Economic Association, shows DID estimates of, and 95% confidence intervals for, the effects of the 2009 VAT reduction on restaurants' profits, as well as pretrend estimates.

 **#7** How is each point on the figure computed? Does the parallel trends assumption seem plausible in this application? Did the reform increase, decrease, or have no effect on, restaurants' profits?

Solution. At $t = 2008$, the estimate equals zero by definition, because 2008 is the reference year. At $t = 2009$, the estimate equals 0.2. This means that, from 2008 to 2009, the average log profit increased by 0.2 more in sit-down restaurants than in takeaway restaurants. Logarithmic increases can approximately be interpreted as increases in relative terms. Then, profits increased by 20% more in sit-down restaurants than in takeaway restaurants. Under the PT assumption, this differential increase is caused by the reform, and one can conclude that the reform increased the profits of sit-down restaurants by 20%. This DID is statistically significant: the confidence interval represented by the dashed lines around it does not contain zero. At $t = 2010$, the estimate equals approximately 0.3. This means that from 2008 to 2010, profits increased by 30% more in sit-down restaurants than in takeaway restaurants. The DID remains above 0.2 in 2011, before decreasing in 2012. At that date, it remains positive and above 10%.

The interpretation of these DID estimates as effects of the VAT reduction relies on the PT assumption: without the VAT reduction, profits would have followed the same evolution in both groups of businesses. The DID estimates at $t = 2007, 2006, 2005$, and 2004 test this assumption, by comparing the evolution of profits in both groups before the reform. All these estimates are much closer to zero than those after the reform, and except for that at $t = 2007$, they are not statistically significant. This suggests that the PT assumption is approximately satisfied in this application.

Benzarti and Carloni (2019) then estimate the effect of the reform on the wages of employees working in sit-down restaurants. They find a positive effect, but much smaller than the effect on restaurants' profits. Similarly, they find a small negative effect of the reform on restaurants' prices. Thus, the reform benefited restaurant owners more than employees and consumers.

1.2 Beyond the classical DID design: Chapters 5 to 8

1.2.1 Two-way fixed effects regressions

In a classical DID design, it can be shown that the DID estimator equals the coefficient on the treatment variable $D_{g,t}$ in a linear regression of $Y_{g,t}$ on group fixed effects, period fixed effects, and $D_{g,t}$. This regression is commonly called a

“two-way fixed effects regression,” TWFER hereafter. Motivated by the equivalence between TWFER and DID in a classical design, researchers have started running TWFERs in more complex natural experiments, with non-binary treatments, and where groups can experience several treatment changes at different times. Without having explicitly formulated it, researchers probably hoped that in these contexts as well, the TWFER estimator would be equivalent to a DID, and would rely only on a partly testable PT assumption. Unfortunately, and as we will soon see, this hope proved unfounded.

But first, let us illustrate the notion of a “complex” natural experiment, using a concrete illustration that we will use as a running example in this section. During his visit to the United States at the beginning of the 19th century, Alexis de Tocqueville was impressed both by the vitality of American democracy and by the dynamism of the press sector. He then formulates the following conjecture: newspapers have a positive effect on democracy (de Tocqueville, 1850). Almost two centuries later, Gentzkow, Shapiro and Sinkinson (2011a) provide a quantitative test of this conjecture. They create a database containing, for a large number of American counties, the voter turnout in presidential elections from 1868 to 1928, and the number of newspapers in circulation in each county at each election. With these data, one can use a TWFER to estimate the effect of newspapers on voter turnout. The treatment here is the number of newspapers in circulation in county g and year t . This is a non-binary treatment. This treatment is also *non-absorbing*: a county can gain and then lose newspapers multiple times. We are thus very far from the classical DID design described previously.

Beyond this example, TWFERs have become an extremely common method to estimate treatment effects in economics. Of the 100 most cited articles published by the American Economic Review between 2015 and 2019, 26 use a TWFER, whereas only six use an RCT. Among these 26 articles, only two consider a classical DID design (de Chaisemartin and D’Haultfoeulle, 2026). TWFERs are also very common in political science: among all articles published between 2017 and 2022 in the American Political Science Review, the American Journal of Political Science, and the Journal of Politics, 93 estimate a TWFER, but only nine do so in the context of a classical DID design (Chiu, Lan, Liu and Xu, 2026). TWFERs are also commonly used in sociology (see, e.g., Rüttenauer and Aksoy, 2024) and in medicine and epidemiology (see, e.g., McNamara, Wiegand, Burke, Sharma, Sheppard, Adjemian, Ahmad, Anderson, Barbour, Binder et al., 2022).

1.2.2 What does a TWFER estimate outside of the classical DID design?

Under the PT assumption, a TWFER estimates a weighted combination of the effects in each treated cell. Specifically, one can show that

$$\text{TWFER} = \sum_{(g,t): D_{g,t} \neq 0} W_{g,t} \text{TE}_{g,t}, \quad (1.9)$$

where $TE_{g,t}$ is the treatment effect in cell (g, t) and $W_{g,t}$ is the weight that the TWFER gives to this effect (de Chaisemartin and D'Haultfoeuille, 2020). The weights $W_{g,t}$ have three properties: i) they sum to one, ii) they can be negative, iii) they can be computed from the data. We compute the weights in our running newspaper example and find that the effects of newspapers receive negative weights in half of the *county-year* cells. Moreover, these negative weights are substantial, summing to -0.47 .

The result in (1.9) has two consequences. First, outside of the classical DID design, the PT assumption is not sufficient for a TWFER to estimate ATT. Second, when some weights are negative, a TWFER does not estimate a convex combination of causal effects. Then, it is possible to obtain, say, a negative coefficient even if the treatment effect is positive in each cell (g, t) . For example, we have that

$$2 \times 1 + (-1) \times 3$$

is a non-convex combination of two positive numbers, 1 and 3, with weights 2 and -1 that sum to 1. Yet, this non-convex combination yields a negative result, -1 . Thus, a TWFER can give a very biased measure of the treatment effect, potentially even producing a sign opposite to the true effect. In our example, if the effect of newspapers on turnout is positive in all *county-year* cells, but this positive effect is larger in cells whose effect is weighted negatively by the TWFER, the regression can yield a negative coefficient.

Let us give some intuition for the result in (1.9), by returning to our simple example with two groups of units s and n . Now, instead of assuming that we are in a classical DID design, where s goes from untreated to treated from period one to two ($D_{s,1} = 0$, $D_{s,2} = 1$), while n remains untreated ($D_{n,1} = 0$, $D_{n,2} = 0$), we leave the design nearly unrestricted. The treatment may be non-binary, both s and n might already receive a non-zero treatment dose at period one, and both s and n could experience a change in their treatment from period one to two: $D_{s,1} = d_{s,1}$, $D_{s,2} = d_{s,2}$, $D_{n,1} = d_{n,1}$, and $D_{n,2} = d_{n,2}$, where $d_{s,1}$, $d_{s,2}$, $d_{n,1}$, $d_{n,2}$ are arbitrary positive real numbers. In the newspaper example, s could be a county going from, say, one to three newspapers, while n has, say, two newspapers in both periods. Letting $DID_D = d_{s,2} - d_{s,1} - (d_{n,2} - d_{n,1})$ denote a DID comparing the evolution of the treatment of s and n from period one to two, we only assume that $DID_D \neq 0$: s and n do not experience the same treatment evolutions. Then, one can show that the treatment coefficient in a TWFER is equal to

$$\frac{Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1})}{DID_D}, \quad (1.10)$$

our usual DID estimator divided by DID_D . Let $Y_{g,t}(d)$ denote the potential outcome of cell (g, t) under treatment d . Let $TE_{g,t} = (Y_{g,t}(d_{g,t}) - Y_{g,t}(0))/d_{g,t}$ be the slope of the potential outcome of group g at period t , between a treatment of zero and

its actual treatment $d_{g,t}$. As $\text{TE}_{g,t}$ is a slope, it can be interpreted as an effect of increasing the treatment by one unit. Then, one has that

$$\begin{aligned}
 & \frac{Y_{s,2} - Y_{s,1} - (Y_{n,2} - Y_{n,1})}{\text{DID}_D} \\
 = & \frac{Y_{s,2}(d_{s,2}) - Y_{s,1}(d_{s,1}) - (Y_{n,2}(d_{n,2}) - Y_{n,1}(d_{n,1}))}{\text{DID}_D} \\
 = & \frac{Y_{s,2}(d_{s,2}) - Y_{s,2}(0) - (Y_{s,1}(d_{s,1}) - Y_{s,1}(0)) - (Y_{n,2}(d_{n,2}) - Y_{n,2}(0) - (Y_{n,1}(d_{n,1}) - Y_{n,1}(0)))}{\text{DID}_D} \\
 & + \frac{Y_{s,2}(0) - Y_{s,1}(0) - (Y_{n,2}(0) - Y_{n,1}(0))}{\text{DID}_D} \\
 = & \frac{Y_{s,2}(d_{s,2}) - Y_{s,2}(0) - (Y_{s,1}(d_{s,1}) - Y_{s,1}(0)) - (Y_{n,2}(d_{n,2}) - Y_{n,2}(0) - (Y_{n,1}(d_{n,1}) - Y_{n,1}(0)))}{\text{DID}_D} \\
 = & \frac{d_{s,2}}{\text{DID}_D} \frac{Y_{s,2}(d_{s,2}) - Y_{s,2}(0)}{d_{s,2}} + \frac{-d_{s,1}}{\text{DID}_D} \frac{Y_{s,1}(d_{s,1}) - Y_{s,1}(0)}{d_{s,1}} \\
 & + \frac{-d_{n,2}}{\text{DID}_D} \frac{Y_{n,2}(d_{n,2}) - Y_{n,2}(0)}{d_{n,2}} + \frac{d_{n,1}}{\text{DID}_D} \frac{Y_{n,1}(d_{n,1}) - Y_{n,1}(0)}{d_{n,1}} \\
 = & \frac{d_{s,2}}{\text{DID}_D} \text{TE}_{s,2} + \frac{-d_{s,1}}{\text{DID}_D} \text{TE}_{s,1} + \frac{-d_{n,2}}{\text{DID}_D} \text{TE}_{n,2} + \frac{d_{n,1}}{\text{DID}_D} \text{TE}_{n,1}. \tag{1.11}
 \end{aligned}$$



#8 Justify each step in the previous derivation.

Solution. The first equality holds because $D_{s,1} = d_{s,1}$, $D_{s,2} = d_{s,2}$, $D_{n,1} = d_{n,1}$, and $D_{n,2} = d_{n,2}$. The second equality follows from adding and subtracting the untreated potential outcomes $Y_{s,2}(0)$, $Y_{s,1}(0)$, $Y_{n,2}(0)$, and $Y_{n,1}(0)$. The third equality follows from the PT assumption. The fourth equality follows from rewriting, say, $Y_{s,2}(d_{s,2}) - Y_{s,2}(0)$ as $d_{s,2} \times [(Y_{s,2}(d_{s,2}) - Y_{s,2}(0))/d_{s,2}]$, and from rearranging terms. The last equality follows from the definition of $\text{TE}_{g,t}$.

Equation (1.11) implies that in this simple example, under the PT assumption the TWFER estimates a weighted combination of $\text{TE}_{s,2}$, $\text{TE}_{s,1}$, $\text{TE}_{n,2}$, and $\text{TE}_{n,1}$, with weights $d_{s,2}/\text{DID}_D$, $-d_{s,1}/\text{DID}_D$, $-d_{n,2}/\text{DID}_D$, and $d_{n,1}/\text{DID}_D$, that sum to one but where two weights may be negative. See Lemma 1 of de Chaisemartin (2011) for a closely related result.

The intuition for (1.11) is as follows. Under the PT assumption, if none of the four cells had been treated, the outcome trends of s and n would have been parallel, so DID would have been equal to zero. In a classical design, under the PT assumption the only reason why s and n can exhibit different trends is that cell $(s, 2)$ is treated, so DID identifies the treatment effect in that cell. By contrast, in a more complex design, where all four cells may be treated, differences in trends can be driven by the treatment effect in any of the four cells. Accordingly, (1.11) expresses DID as a weighted sum of the treatment effects in the four cells.

Equation (1.9) calls into question the use of TWFERs to estimate treatment effects in complex natural experiments. Consequently, the methodological literature has endeavored to propose so-called heterogeneity-robust (HR) DID estimators, that rely on a PT assumption, like TWFERs, but that remain valid even if the treatment effect varies across space and / or time, unlike TWFERs.

1.2.3 Heterogeneity-robust DID estimators

1.2.3.1 Staggered adoption designs

Consider first the special case of a binary, absorbing treatment with staggered adoption timing, a so-called staggered adoption design. For example, free trade agreements fall into this framework: the treatment is binary (either two countries have an agreement binding them, or they do not), absorbing (since World War II, there are very few examples of countries signing an agreement and then reversing course, although this situation may change soon), with staggered adoption timing (not all countries sign agreements at the same time).

In this special case, it can be shown that the treatment coefficient in a TWFER can be written as a convex combination of two types of DIDs: DIDs that compare a group becoming treated to a group that is not, and DIDs that compare a group becoming treated to a group that is already treated (Borusyak and Jaravel, 2017; de Chaisemartin and D’Haultfoeulle, 2020; Goodman-Bacon, 2021). The undesirable properties of TWFERs come from this second type of DIDs, which have since been dubbed “forbidden comparisons”. Returning to our simple example with two groups, if s becomes treated ($D_{s,1} = 0$ and $D_{s,2} = 1$) while n is already treated ($D_{n,1} = 1$ and $D_{n,2} = 1$), then $DID_D = 1$ and (1.11) reduces to

$$DID = TE_{s,2} - (TE_{n,2} - TE_{n,1}).$$

The previous equation shows that under the PT assumption, a DID comparing a group becoming treated to a group that is already treated identifies the effect of the treatment in the group becoming treated, plus the change of the treatment effect in the group already treated. Then, such DIDs are only valid if the treatment effect in the already-treated group does not change over time, an assumption that is often implausible.

Therefore, to obtain an HR estimator, it suffices to use only the first type of DIDs, those that compare a group becoming treated to a group that is not or not yet treated (de Chaisemartin and D’Haultfoeulle, 2020; Sun and Abraham, 2021; Callaway and Sant’Anna, 2021b).

1.2.3.2 Heterogeneous adoption designs

Consider next the special case of a natural experiment with two periods, where no unit is treated during the first period, and where treated units receive different *doses* of treatment in the second period, a so-called heterogeneous adoption design. For example, when a country raises its minimum wage, firms are more or less

(continued...)

Index

A

Ashenfelter's dip, 48, 116
average treatment effect on the treated (ATT), 33

B

Bell and McCaffrey's (BM) critical values, 54, 56, 58, 63, 96, 97, 242

C

cell, 20
classical DID design, 31
clustering, 27, 28, 52, 96, 198
condition number, 46, 51, 212
conditional parallel trends assumption (CPT), 101
convex combinations of effects, 104
coverage or coverage rate, 53, 54, 56, 58, 62, 63, 69, 70, 96, 97, 126, 137, 143, 205, 206, 229, 296
csdid, 200, 204–212, 217, 226, 229

D

debiased machine learning (DML), 108, 148, 197
did_imputation, 203–205, 207–210, 213, 215, 217, 226, 229, 317
did_multiplengt_dyn, 120, 121, 201, 204–210, 212, 215, 217, 219, 226, 227, 229, 230, 279, 284, 286, 290, 294–307

E

estimating heterogeneous effects, 74–80, 218–220, 297, 298
event-study effects, 34
event-study two-way fixed effects regression (ES-TWFER), 42
eventstudyinteract, 199, 200, 204, 205, 207–212, 229

F

F-test, 40, 45–47, 51, 54, 97, 112, 184, 185, 187, 207, 212, 225, 249
fect, 128, 137, 138, 204, 296, 317
forbidden comparison, 18, 178, 179, 181, 244–246, 250
Frisch–Waugh–Lovell theorem, 37, 98, 116, 158, 189
fuzzy DID designs, 233, 246

G

group-specific linear trends, 49, 73, 109, 110, 132, 148, 203, 208, 209, 215, 300

H

HC2 standard errors, 53, 54, 56, 58, 63, 96, 97, 242
heterogeneity-robust DID (HR DID), 18

heterogeneous adoption design (HAD), 234
honestdid, 144–147

I

imbalanced panel, 90–92, 98, 205, 226, 227, 230, 262, 322
imputation, 201–204, 206, 218, 221, 227, 229, 313, 314, 317
initial conditions, 22, 301
instrumental variable (IV), 88–90, 98, 116, 252, 259, 275, 301, 316, 317
interactive fixed effects (IFE), 12, 122, 123, 125–128, 137–139, 141, 149, 152, 313, 314

L

limited dependent variable, 80, 220
local projection, 193, 194, 199, 201, 228, 319
logit, 98, 106, 107, 220, 221, 230

M

mean reversion, 117–119, 149
model-based perspective, 24–27, 96

N

natural experiment, 3–5, 15, 18, 19, 40, 73
Nickell bias, 115, 116
no dynamic effects assumption (ND), 22
no-anticipation assumption (NA), 21
no-sign reversal property, 104, 163

P

parallel trends assumption (PT), 22
permutation test, 41, 62–64
Poisson regression, 82, 85, 221–223
potential outcome, 20
practitioners' dos and don'ts, 96, 148, 228, 261, 321
pretrend tests or placebo tests, 11
probit, 98, 106, 107, 220, 221, 230
propensity-score reweighting, 106, 107

Q

quantile treatment effect (QTE), 86, 87, 225

R

randomized controlled trial (RCT), 3, 4, 15, 112, 152, 232, 247
relative time, 42

S

sdid_event, 137, 139, 140
sensitivity to functional form, 82–85

several treatments, 274, 275, 278, 298, 315, 316, 319
simulations, 13, 45, 54, 55, 58, 63–66, 94, 96, 97, 123,
126, 127, 136, 137, 149, 203, 205, 206, 216,
229, 296
spillover, 21, 95, 96, 227
staggered adoption design (SAD), 170
staggered first switch design (SFSD), 279
static two-way fixed effects regression, 36
sup-t tests, 45, 46, 51, 97, 212
synthetic control (SC), 12, 119, 122, 123, 128–135, 138–141,
149
synthetic difference-in-differences (SDID), 122, 123, 128,
133–136, 140, 141, 149

T

triple difference, 111, 203
two-way fixed effects regression (TWFER), 36
twowayfeweights, 164–168, 177, 191, 192, 243, 275, 278, 279

W

weighted averages of effects, 104
weighted combinations of effects, 104

Z

zero-sum-weight combinations of effects, 104