

# Recent Developments in Methods for Causal Inference with Cross-Sectional Time-Series Data\*

## *Workshop Notes*

Roberto Valli<sup>†</sup>

January 31, 2023

## 1 Introduction

Political science is often characterized by research questions that cannot be studied experimentally because ethical or practical reasons prevent administration of theoretically relevant treatments. Given the frequent impossibility of using true randomization in polisci research it is all the more important to find ways to estimate (close-to) causal effects of policies and other interventions with observational data. These notes discuss recent developments in a family of statistical methods that exploit within and across-unit variation over time to estimate counterfactual outcomes, and thereby approximate estimates of average causal effects. In fact the methods reviewed in these notes were developed within slightly different literatures, so one can probably argue the extent to which they all belong together. I try to make the case that they all do.

The goal of these notes is to tie together related strands of literature that highlight the limitations of standard difference-in-differences methods (DiD) and propose corrections and new estimators. The interesting thing about this literature is that not only it develops extremely fast, but it accompanies a harsh critique of common modelling choices to novel solutions, practical advice, and R packages to implement more suitable models. Moreover, contributions to the literature are quickly bridging previously-distinct fields of applied statistical theory. However, the fast-moving and technical nature of the literature can be confusing. This is why I wrote these notes: To guide applied researchers to the right references and modelling choices without having to first read everything under the sun.

Note that this document is meant to be a practical guide for advanced researchers with prior knowledge of statistics, basic causal inference designs and programming in R. It is by no means a technical paper, nor an introduction to any of these subjects. Also, the focus of this paper is on the different methods used to statistically estimate average treatment effects and not on the related ways to compute uncertainty measures. Applied scholars need to learn more about these through the original methods papers and the related R package documentation. Moreover, this note is limited to models dealing with binary treatments and does not cover related topics, such as developments of DiD estimators that drop standard sampling assumptions for design-based inference and the literature on event study models.

---

\*Find the workshop's GitHub repo at [https://github.com/RobertoValli/panel\\_models\\_workshop](https://github.com/RobertoValli/panel_models_workshop).

<sup>†</sup>PhD candidate, International Conflict Research, ETH Zurich: roberto.valli@icr.gess.ethz.ch

The notes proceed as follows. First, I give a brief overview of causal inference methods with cross-sectional time-series data (CSTS) and its limitations. I start from the standard DiD model and its estimators (Section 2), and walk through some of its critique in the literature (Section 3). Next, I present a heuristic decision model to select the right CSTS method based on one’s data structures. Then I provide a description of various methods, their assumptions and implementation in R, and conclude with some remarks and a summary table of each method’s features.

## 2 Causality and Cross-Sectional Time-Series Data

The problem of estimating causal effects can be thought of as the challenge of predicting counterfactual outcomes. The impossibility of observing the outcome of a process had something (the treatment condition) been different is often referred to as the *fundamental problem of causal inference*. The Neymar-Rubin counterfactual outcome framework formalizes one way of thinking about causality in terms of counterfactuals.

Consider an outcome of interest  $Y_{it}$  that is observed for individual  $i$  at time  $t$ . Also, consider a binary treatment  $D_{it} \in \{0, 1\}$ . The individual causal effect of  $D$  on  $Y$  can be defined as

$$\text{ICE} = Y_{it}(1) - Y_{it}(0)$$

meaning that the effect of  $D$  is defined as the individualistic difference between the outcome that would result from treatment exposure  $Y_{it}(1)$  and the one resulting from no treatment exposure  $Y_{it}(0)$  (i.e., control status).

However, since only one of the two terms is ever observed in reality at a given time and for a specific individual, it is extremely hard and noisy to estimate individual effects of policies. Statistical methods for causal inference find ways to estimate *average causal effects* by predicting the counterfactual outcome under the unobserved treatment condition for a population in ways that ensure unbiased predictions *on average*. It is interesting to note that while applied statistics are often divided between causal inference and prediction tasks, prediction is very much at the heart of causal inference methods, although with a strong focus on bias reduction a.k.a. unconfoundedness.

### 2.1 The Classic DiD Setup as a Motivating Example

Let’s consider the classic two-period DiD setup and how it can be thought of as a prediction exercise. The outcome is assumed to be a continuous variable  $Y_{it}$  observed over two periods  $t = 1, 2$ . The treatment is assumed to be dichotomous, to take value 0 for all units at period  $t = 1$ , and value 1 for some units at time  $t = 2$ . The observed value of the outcome at any given time can be expressed in terms of counterfactual terms as:

$$Y_{it} = Y_{it}(1)D_{it} + Y_{it}(0)(1 - D_{it})$$

In this setting one might describe the average treatment effect on the treated (ATT) as

$$\tau = E(Y_{it}(1) - Y_{it}(0) | D_{it} = 1).$$

Once again,  $\tau$  contains unobserved values in the counterfactual outcomes  $Y_{it}(0) | D_{it} = 1$ . The major contribution of the classic DiD literature (e.g., Angrist & Pischke, 2009) is to demonstrate that under parallel trend and SUTVA<sup>1</sup> assumptions it is possible estimate the unbiased ATT. Parallel trends suggest that the

---

<sup>1</sup>The *stable unit treatment value assumptions* is a set of assumptions such that: i) units composition does not change over time, ii) no treatment heterogeneity exists, iii) no treatment spillovers take place.

individual *change* in outcome under control between  $t = 1, t = 2$  is independent of treatment status and can be expressed as

$$E(Y_{i2}(1) - Y_{i1}(0)|D_{it} = 1) = E(Y_{i2}(1) - Y_{i1}(0)|D_{it} = 0).$$

Note that a major strength of DiD designs is that *they do not rule out selection bias per se*, but only rule out differential selection into treatment between treated and control group. This makes DiD a very flexible and relatively reasonable way of thinking at applied settings where units are observed repeatedly and only part of the population is treated.

The parallel trend assumption is key to understand how DiD estimators predict counterfactuals. In fact, by assuming that the trends in the outcome under control are independent from the realized treatment we justify taking the average change in outcome in the control group  $E(Y_{i2}(1) - Y_{i1}(0)|D_{it} = 0)$  as the trend that the treated units would have had, had they not been treated. Then it becomes easy to make a linear prediction of the counterfactual outcomes under no-treatment condition for the treated units using observed data. It is sufficient to add the control units' trend to the outcome of the treated units at  $t = 1$ :

$$E(Y_{i2}(0)|D_{it} = 1) = E(Y_{i1}|D_{it} = 1) + [E(Y_{i2}(0) - Y_{i1}(0)|D_{it} = 0)].$$

## 2.2 Causal Inference as a Prediction Problem

Once the identification of ATTs with DiD methods is thought of as a prediction problem it becomes easier to see the link between DiD and other CSTS estimators such as the synthetic control method (SCM) (Abadie, Diamond, & Hainmueller, 2010). In fact, as suggested by Roth et al. (2022) the two literatures are getting closer to one another. The differences between SCM and DiD are mostly down to data requirement and the way counterfactual outcomes are estimated. In its original formulation, SCM only allows to have one treated unit (usually a state-like political unit), provides a transparent selection of the donor pool, i.e. the units that make up the control group. On the other hand, DiD designs ideally require large population of individuals, where less emphasis is placed on control group selection. I discuss the SCM more below.

Another family of methods that causal-inference-as-prediction helps linking to DiD is matching. Imai, Kim, and Wang (2021) ingeniously re-define estimation in DiD settings as a problem of choosing the right control group for counterfactuals estimation. They propose an R package with a set of tools to improve estimation of ATTs by explicitly defining the criteria for being an appropriate control unit in observational settings. As we'll see further down, getting the control group right can be complicated in DiD settings that go beyond the classic setup.

In fact, as soon as one's case of interest goes beyond this simple setup the assumptions and common estimation methods for the ATT start to be violated, with various identification problems and biases arising in the process. The next section discusses some of the major problems identified by the literature and touch on the solutions out there.

## 3 When the Basic DiD Design Breaks Down

The econometric literature on TWFEs has produced a substantial number of papers discussing the limitations of the workhorse regression specification in much applied work. Most of these papers focus on the shortcomings of TWFE regressions when used to estimate ATTs in cross-sectional time-series data.

### 3.1 Number of Time Periods

The simplest violation of the basic DiD setup emerges when units are observed for more than two periods  $t = 1, 2, \dots, N$ . In this case, Angrist and Pischke (2009) already note how the first-difference model cannot be considered an unbiased estimator of  $\tau$ , since taking first-differences over multiple periods introduces autocorrelation of the outcome and violates the independence of the error term. Nevertheless, the same is not true for the two-way fixed effects model, which can be used in multiple-period settings under the assumption of no effect heterogeneity.

### 3.2 Treatment Timing

The most significant result in the recent literature on CSTS estimators, and probably the one that contributed the most to its quick development, is the demonstration that the TWFE estimator is a biased estimator of  $\tau$  in the common setting of *staggered treatment adoption*. Actually, most contributions to the literature confront the problems related to staggered adoption (see de Chaisemartin & D’Haultfœuille, 2022).

Researchers demonstrated that TWFE estimates are a weighted average of DiD estimates comparing all pairs of treatment statuses and cohorts. Whenever treated units are not exposed to treatment at the same point in time, but rather adopt treatment in several cohorts, several problems negatively affect the estimate of both aggregate (Callaway & Sant’Anna, 2021; de Chaisemartin & D’Haultfœuille, 2022; Goodman-Bacon, 2021) and "dynamic" or event-study ATTs (Sun & Abraham, 2021). Fortunately, the TWFE estimate of  $\hat{\tau}$  is generally biased towards zero.

The first issue with the TWFE estimates of ATTs in staggered settings is that some DiD comparisons are undesirable, meaning that already-treated cohorts serve as counterfactual for other treated units. Since TWFE aggregates all possible DiD comparisons, some of them have inappropriate control units as they were exposed to treatment in the past. This leads TWFE to under-appreciate the effect of treatment as some control units have their outcomes affected by treatment.

A second problem is that some of the cohort-effects’ weights can be problematic. Goodman-Bacon (2021) shows that the weight attributed to cohort effects are proportional to cohort size and the variance of  $D_{it}$  in a cohort. This means that treatment effects on early and late treated cohorts are weighted less. Moreover, TWFE weights of cohort average effects can be negative, which creates problem with the interpretation of the aggregate estimate  $\hat{\tau}$ .

There are various solutions to the weighting and comparison problems, which are solved either by setting constraints on group weights, or by explicitly selecting the control group before estimation. The available options are explained below.

### 3.3 Violations of Parallel Trends

The assumption of parallel trends is hard to motivate, and impossible to prove. There are common tests for the existence of pre-treatment trends, like the common event study plots where the outcome is regressed onto lags and leads of the treatment. However, recent studies demonstrated that these tests are underpowered when it comes to rejecting the presence of pre-trends. In fact, there is a growing literature just on this issue, but I won’t go into it in this note (see literature on “further topics.”

That said, scholars proposed ways to estimate consistent ATTs when parallel trends are valid conditional on

a set of covariates (Sant’Anna & Zhao, 2020). The authors proposed a family of “doubly robust” estimators that have desirable properties when either the outcome model or the covariate set are correctly specified.

A treatment is referred to as being in an absorbing state whenever units that are treated cannot revert to being untreated. This is very common in many policy areas, but it does also not apply to many phenomena political scientists want to know the effects of. In fact, there are multiple phenomena in which units can be treated repeatedly over time. There are various problems with repeated and non-absorbing treatments, the major of which is the selection into (repeated) treatment that might bias estimates. However, scholars proposed various methods that allow to deal with these problems, although at the expense of more demanding assumptions.

Finally, one might wonder how DiD designs might be extended to continuous treatment settings. While Angrist and Pischke (2009) suggest that the type of treatment does not influence the estimand or estimation quantity, Callaway and Sant’Anna (2021) demonstrate that continuous treatments make settings with staggered adoption and non-heterogeneous effects extremely hard to handle, as well as deserving a complicated notation. In short, there does not seem to be a set of simple tools and guidelines to be employed in DiDs with continuous treatment.

### 3.4 Other Assumption Violations

As part of SUTVA DiD estimators generally assume homogeneous effects of treatment, meaning that all individual ITE derives from the same distribution. This can be problematic in cases of staggered adoption or underlying and unobserved characteristics of the treatment or treated unit. Goodman-Bacon (2021) and many others describe the consequences of treatment heterogeneity across cohorts for TWFE estimates, usually suggesting that more flexible specifications that compute event-study-like average treatment effects for cohorts or post-treatment periods are less affected by treatment heterogeneity.

Multi-period DiD settings are also more prone to violation of the assumptions of constant treatment effects and of no carryover effect (i.e. past treatments affect current outcomes Liu, Wang, & Xu, n.d.). The assumption of no-carryover effect is generally of less concern if one is interested in the effect of an intervention over the whole observed period, but much more in event-study settings or when the treatment status is allowed to reverse.

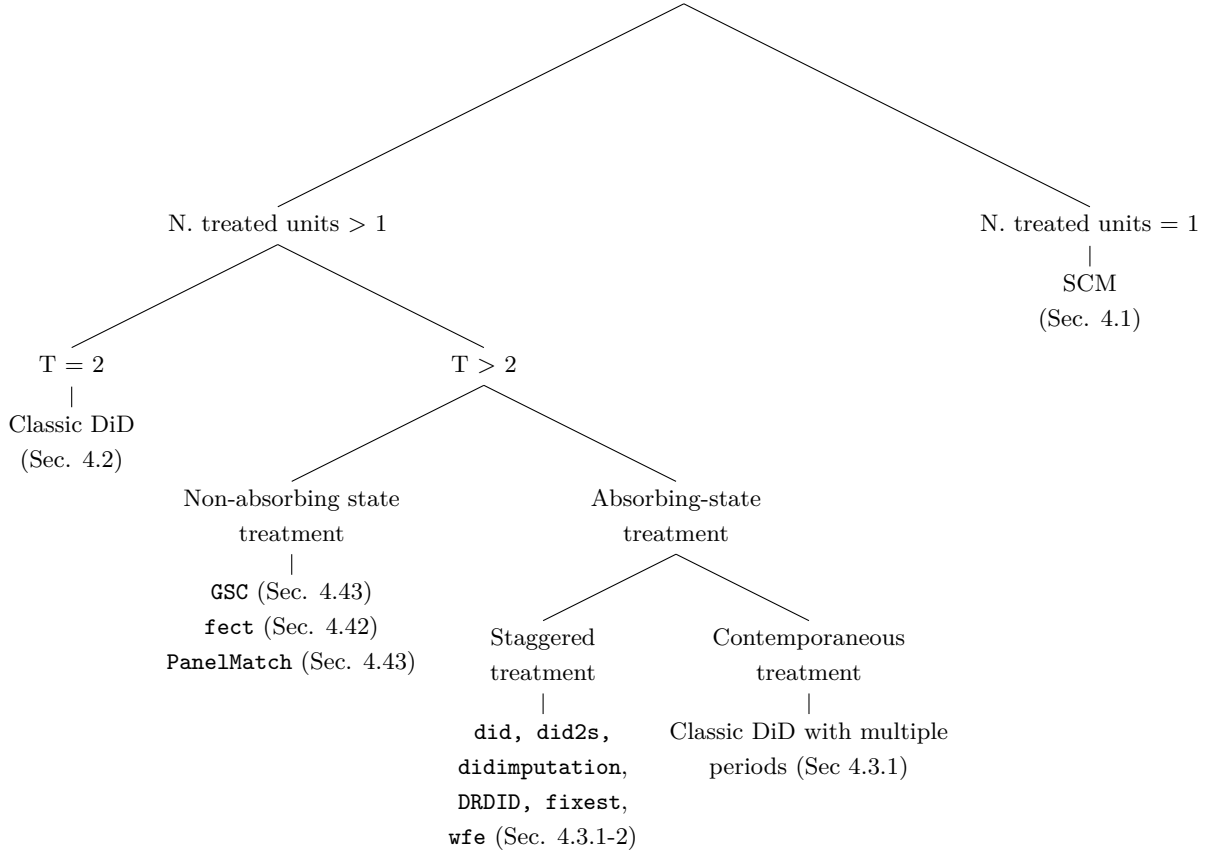
## 4 Which CSTS Model is Right for You?

Let’s turn to a very applied guide to the right modelling choice depending on one’s variation of interest. Note that this typology is strictly limited to binary treatment settings.

### 4.1 Number of Treated Units and the Synthetic Control Method

The first modelling choice is dictated by whether a scholar is interested in the effect of an intervention that treated one (or few) aggregate unit. In this case, the standard synthetic control method (SCM) allows to model the counterfactual outcome of the treated unit flexibly and with extensive diagnostic tools. The idea behind the SCM is to build a model of  $Y_{it}(0) \mid D = 1$  from a pool of untreated units (the “donor pool”). The model is tuned on the pre-treatment period, and then used to predict the treated units’ outcomes under no-treatment counterfactual.

Figure 1: Heuristic tree of CSTS methods.



The data requirements for ATT estimates with the SCM are i) a sufficiently long pre-treatment period to fit the synthetic control observation, ii) a sufficiently large donor pool of comparable units, iii) the treated unit’s outcomes must be within the control units’ common support, i.e. the treated unit cannot have outcomes more extreme than the control group. Note that a recent development of SCM by Ben-Michael, Feller, and Rothstein (2021) suggests that by fitting a regularized outcome model it is possible to make valid predictions of counterfactual outcomes even when the treated unit is outside of the donors’ common support. Additionally, SUTVA assumptions must hold, which can be hard to assess especially when it comes to the absence of spillovers.

There are currently two R packages for SCM analyses and one for Ben-Michael, Feller, and Rothstein’s “augmented” SCM: `tidysynth`, `Synth`, `augsynth`.

## 4.2 Number of Treatment Periods and the Classic DiD Setup

Whenever there are only two periods and treated units are only treated in the second period, scholars are sure to find themselves in the “classic DiD” setting. This means that there are three equivalent ways to estimate the ATT: The non-parametric difference in means, the first-difference OLS and the TWFE model. These can be all estimated with base R tools and packages for linear regressions.

The three models are statistically equivalent for the estimation of ATT because the predicted outcome of the treated under control can only be predicted linearly as the pre-treatment outcome of the treated plus the

trend of the control. The R code below demonstrates it with help of example data.

```
# Load packages
library(fixest)
library(tidyverse)

# Filter example data (from fixest package) to only 2 periods
dat_did <- fixest::base_did %>%
  filter(period %in% c(5, 6))

# Create first-difference of treatment and outcome
dat_did <- dat_did %>%
  mutate(post_treat = post*treat) %>%
  group_by(id) %>%
  mutate(y_diff = y - lag(y),
         x_diff = post_treat - lag(post_treat))

# N. unit
length(unique(dat_did$id))

## [1] 108

# N. treated units
length(unique(dat_did$id[dat_did$treat == 1]))

## [1] 55

# Compute non-parametric DiD
tau_nonp <- mean(dat_did$y[dat_did$period == 6 & dat_did$treat == 1]) -
  mean(dat_did$y[dat_did$period == 5 & dat_did$treat == 1]) -
  (mean(dat_did$y[dat_did$period == 6 & dat_did$treat == 0]) -
   mean(dat_did$y[dat_did$period == 5 & dat_did$treat == 0]))

# Fit first-difference model
mod_first <- feols(y_diff ~ x_diff, data = dat_did)

# Fit TWFE model
mod_twfe <- feols(y ~ post_treat | id + period, data = dat_did)

# Compare estimated ATT
tau_nonp

## [1] 1.159133

mod_first$coefficients
```

```
## (Intercept)      x_diff
##      1.173937      1.159133

mod_twfe$coefficients

## post_treat
##      1.159133
```

### 4.3 Absorbing-state Treatments and Their Different Flavors

In settings where the scholar observes the units over multiple periods it is important to separate the cases of absorbing state-treatments, i.e. those in which treated units remain treated after exposure, and those that do not. Next, I discuss the simplest case and the one covered by much of the econometric literature on DiD, that is absorbing-state treatment settings.

#### 4.3.1 Contemporaneous or Staggered Adoption

In this family of settings, the first distinction to be made concerns whether the scholar believes that the basic DiD assumptions (contemporaneous treatment, no effect heterogeneity, synchronous treatment, parallel trends, no anticipation) hold in a specific case. If these all hold, and in particular if all treated units are exposed to treatment at the same time, then in principle the TWFE approach returns the unbiased ATT. However, some of the assumptions are impossible to test and therefore it is worth running tests and alternative estimations that are robust to violations of the basic DiD setup.

In particular, DiD in settings with staggered adoption almost always present the weighting and comparison problems introduced above. Starting from the comparison problem, scholars proposed ways to limit estimates to “clean” comparisons. This usually entails transitioning from a single-stage estimation as in TWFE regressions to more complicated procedures. For instance, Gardner (2021) and Callaway and Sant’Anna (2021) propose to estimate treatment-cohort specific ATTs. In the first stage is also possible to choose whether to use never-treated or not-yet-treated observations as control group. These cohort-specific estimates are free from comparison issues and can then be aggregated in different ways to estimate treatment period, cohort, or overall ATTs depending on the researcher’s goals. Similarly, multiple-stage estimation helps solve the implicit weighting problem of TWFE by explicitly setting unitary or other weights on each cohort-ATT.

Estimation for DiDs with staggered adoption is implemented in multiple packages, each with its own perks and quirks: `did`, `did2s`, `didimputation`, `DRDID`, `fixest`, `wfe`. Probably the most important differences among packages regard available specification tests, covariate adjustment and ability to aggregate effects.

#### 4.3.2 Failure of Parallel Trends Assumption

The parallel trend assumption is impossible to demonstrate, and even common pre-trend tests fail to really rule out the existence of differential selection bias (Roth et al., 2022). A typical correction used by applied scholars who suspect a violation of the parallel trends assumption is to run an additional set of TWFE models with linear or quadratic time trends (i.e. interacting unit fixed effects with years and years squared). Another solution is to add a set of time-varying covariates to TWFE regressions hoping to control away the selection



bias. The problem with both of these approaches is that they may fail if scholars misspecify the functional form of the time trends or covariate set.

However, currently there are various proposed solutions to this problem, which might be more or less appropriated depending on the specific application. One reason for the violation of the parallel trend assumption might be that the control group is too heterogeneous and hence make inappropriate comparisons. In this case, Abadie (2005) and Imai, Kim, and Wang (2021) suggest to match units on their pre-treatment covariates to reduce heterogeneity and make the parallel trends more credible. Another case that might undermine parallel trends is the staggered treatment adoption, whereby TWFE mechanically produce unwanted comparisons. The solutions to this problem are described below.

Finally, it might be the case that potential outcomes of treated and control observations really follow parallel trends, but only conditional on some covariates. In such cases Sant’Anna and Zhao (2020) propose ways to estimate consistent and efficient ATTs conditional on the correct covariate specifications by reweighting units based on the propensity score. The method is implemented in the `DRDID` package. Callaway and Sant’Anna (2021) propose a way to condition parallel trends on time-invariant covariates implemented in the `did` package.

## 4.4 Non-Absorbing Treatments and Their Estimators

For settings that go beyond the typical policy implementation where a set of treated units receives an absorbing state treatment, methodologists have developed different approaches to extend the DiD logic of counterfactual estimation while allowing for more complicated data structures. In this section I discuss three methods: The generalized synthetic control (Xu, 2017, GSCM), the PanelMatch (Imai, Kim, & Wang, 2021; Kim et al., 2021), and the fec (Liu, Wang, & Xu, n.d.) counterfactual methods.

All three methods allow treatments to last for a certain period and be reversed, and offer generally more flexible predictions of counterfactuals than standard linear DiD methods. Moreover, they all have great R implementation and useful diagnostic tools.

### 4.4.1 The Generalized Synthetic Control Method

Xu (2017) introduces a flexible approach to estimating causal effects with CSTS data that can be thought of as a generalization of the synthetic control method, but takes inspiration from the literature on interactive fixed effects (Bai, 2009) and applies them to DiD settings. The GSCM computes counterfactual outcomes for the treated group by first fitting a linear model with fixed effects for time and unit interactions on the control group, then estimating unit-specific intercepts using the pre-treatment outcomes of the treated group, and finally combines the to data to predict the control outcomes for the treated units in the post-treatment period. Then the ATT is estimated as the mean difference between observed and imputed counterfactuals.

In order to make valid estimates the method needs the following assumptions: i) Strict exogeneity of one unit’s error term to other units  $E[\varepsilon_{it} \mid D_{ij}, x_{ij}, \gamma_{ij}, f_{ij}] = E[\varepsilon_{it} \mid \gamma_{ij}, f_{ij}] = 0$ , ii) serial independence of one unit from other units. Moreover, the GSCM requires scholars to observe at least 10 pre-treatment periods and a minimum of 40 control units. The method is implemented through the `gsynth` R package.

#### 4.4.2 The `fect` Method

Liu, Wang, and Xu (n.d.) propose what can be seen as a development of GSCM because it covers a set of estimators of ATTs that fit predictive models of counterfactual outcomes on the control population. Similarly to GSCM, scholars can use one of three methods to fit models of the control group’s outcomes (linear fixed effects, interactive fixed effects, and matrix completion), then use these models to infer the counterfactual outcomes under control for the treated group.

The advantages of `fect` over other methods include the relaxation of the strict exogeneity assumption at least for IFE and MC estimators, which are both robust to the existence of potential time-varying confounders are captured with latent factors, and the large set of diagnostic tools that allow to test for the existence of anticipation, carryover effects, and pre-trends.

The biggest single drawback of the method is probably the novelty of most of the estimators included in the R package `fect`, which might not be received enthusiastically by most political scientists.

#### 4.4.3 The Panel Matching Method

Imai, Kim, and Wang (2021) develop a CSTS method that improves on existing DiD estimators by drawing on matching methods for causal inference. PanelMatch tries to solve the known problems of TWFE estimation of ATTs with a three-step procedure: 1) Scholars match treated and control units for a set of time-varying covariates as well as the outcome over a certain number of periods, 2) each treated unit’s counterfactual outcome  $l$  periods since treatment is estimated nonparametrically as the outcome at  $l - 1$  net of the average change in outcome in its matched set at the same period, 3) the period-specific ATT is computed as the average difference between observed and counterfactual outcomes for a given period since treatment, and similarly the overall ATT is the average of period-specific ATTs.

The advantages of PanelMatch are its intuitive selection of control observations, its robustness to model misspecification, the ability of using different matching and weighting procedures (Mahalanobis, propensity score, covariate-balancing propensity score), the ability to assess the matching quality before seeing effect estimates (potentially less selection bias in the model choice), and finally the fully flexible non-parametric imputation of counterfactual outcomes.

However, these advantages come at significant costs. First, PanelMatch requires the most demanding version of the parallel trends assumption, known as the “sequential ignorability assumption,” implying that treatment assignment at a given time must be independent from outcome, covariate and treatment histories in all previous periods. Second, because samples are explicitly restricted through the matching procedure, the estimator throws out quite some data and is therefore less efficient than alternative estimators (i.e. estimates are less precise and standard errors wider).

## 5 Recap and Overview Table

## References

Abadie, A. (2005). Semiparametric Difference-in-Differences Estimators. *The Review of Economic Studies*, 72(1), 1–19. <https://doi.org/10.1111/0034-6527.00321>

Method	Treated $N$	Staggered	Exogeneity	Heterogeneity rob.	Treatment	Functional form	Controls	Autocorr./spillovers	Data needs
TWFE	$N >= 1$	No	Strict	No	Absorbing	Linear	Yes		
SCM	$N = 1$	–	Relaxed	Yes	Absorbing	Nonlinear	Yes		
GSCM	$N >= 10$	Yes	Relaxed	Yes	Non-Absorbing	Nonlinear	Yes		
fect	$N >= 1$	Yes	Relaxed	Yes	Non-Absorbing	Nonlinear*	Yes		
PanelMatch	$N >= 1$	Yes	Rep. ignorability	Yes	Non-Absorbing	Nonlinear	Yes		
Two-stage DiD	$N >= 1$	Yes	Strict	Yes	Absorbing	Linear	Yes		
did	$N >= 1$	Yes	Strict	Yes	Absorbing	Linear	Yes		

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493–505. <https://doi.org/10.1198/jasa.2009.ap08746>
- Angrist, J., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion* (1st ed.). Princeton University Press. <https://EconPapers.repec.org/RePEc:pup:pbooks:8769>
- Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica*, 77(4), 1229–1279. <https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6135>
- Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The Augmented Synthetic Control Method. *Journal of the American Statistical Association*, 116(536), 1789–1803. <https://doi.org/10.1080/01621459.2021.1929245>
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods [Themed Issue: Treatment Effect 1]. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/https://doi.org/10.1016/j.jeconom.2020.12.001>
- de Chaisemartin, C., & D’Haultfoeulle, X. (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*. <https://doi.org/10.1093/ectj/utac017>
- Gardner, J. (2021). *Two-stage differences in differences* [Working Paper].
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing [Themed Issue: Treatment Effect 1]. *Journal of Econometrics*, 225(2), 254–277. <https://doi.org/https://doi.org/10.1016/j.jeconom.2021.03.014>
- Imai, K., Kim, I. S., & Wang, E. H. (2021). Matching Methods for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*, n/a(n/a). <https://doi.org/https://doi.org/10.1111/ajps.12685>
- Kim, I. S., Rauh, A., Wang, E., & Imai, K. (2021). *PanelMatch: Matching Methods for Causal Inference with Time-Series Cross-Sectional Data* [R package version 2.0.0]. <https://CRAN.R-project.org/package=PanelMatch>
- Liu, L., Wang, Y., & Xu, Y. (n.d.). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*, n/a(n/a). <https://doi.org/https://doi.org/10.1111/ajps.12723>
- Roth, J., Sant’Anna, P. H. C., Bilinskiz, A., & Poe, J. (2022). What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Working Paper*.
- Sant’Anna, P. H. C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), 101–122. <https://doi.org/https://doi.org/10.1016/j.jeconom.2020.06.003>
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects [Themed Issue: Treatment Effect 1]. *Journal of Econometrics*, 225(2), 175–199. <https://doi.org/https://doi.org/10.1016/j.jeconom.2020.09.006>
- Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25(1), 57–76. <https://doi.org/10.1017/pan.2016.2>