

New developments at the intersection of Machine Learning, Causal Inference, and Marketing

A comparison of Difference in Differences, Synthetic Controls and Synthetic Difference in Differences through simulated data

Benjamin Herzberger

Submission date: 17.06.2024

Table of contents

List of abbreviations	III
1 Introduction	1
2 Methodology	1
2.1 Difference in Differences	2
2.2 Synthetic Control	3
2.3 Synthetic Difference in Differences	4
3 Data and Simulations	5
4 Results and Interpretation	5
4.1 Random treatment assignment	5
4.2 Non-random treatment assignment	5
4.2.1 Inverse probabilty weighting	5
4.3 Further findings	5
5 Conclusion	5
Literature	5

List of abbreviations

DiD	Difference in Differences
SC	Synthetic Control
SDiD	Synthetic Difference in Differencesl
OLS	Ordinary Least Squares
TWFE	Two-way-fixed-effects
IPTW	Inverse Probabilty of Treatment Weighting

1 Introduction

Causal inference in social sciences, such as Marketing, is crucial for understanding the impact of interventions and policy changes. However, researchers often encounter two key problems: First, interventions are not performed on random subjects, but are assigned systematically. Second, the number of observations exposed to those intervention, also called treatment, is small.

In Difference in Differences (DiD), a popular technique used to estimate causal effects, multiple units are observed over multiple time periods. At one point in time, the treatment period, some units will receive a treatment, while others remain untreated and serve as the control group. Comparing the development of these groups over time can provide meaningful insights. However, DiD relies on rather restrictive assumptions e.g. the parallel trends assumption, which are not entirely compatible with the previously mentioned problems. Consequently, different methods to overcome these issues are necessary.

In 2003, Abadie and Gardeazabal introduced a new approach, the so-called Synthetic Control (SC). This method is specifically helpful in providing reliable estimates, when treatment assignment is not random and the number of treated units is small.

More recently, in 2021, Arkhangelsky et al. introduced another approach, the so-called Synthetic Difference in Differences (SDiD), which can be conceived of as a combination and refinement of the formerly mentioned methods.

This work aims to assess the overall performance of the SDiD approach, when compared to the standard DiD and the SC method in the context of simulated data. Therefore, data that varies in treatment assignment, population size, number of treatment periods, violation of the parallel trends assumption and the degree of treatment heterogeneity, will be simulated. Through repeated simulation and subsequent estimation using the three discussed methods, the statistical properties of the resulting estimates can be investigated.

The paper will be structured as follows: Section 2 provides a more detailed explanation on the different estimation methods together with previous literature findings. Section 3 will discuss the simulation process. In Section 4 results will be presented and discussed. Section 5 then concludes this research and aims to provide a brief overview of the findings, and provide recommendations on when to use which estimation method.

2 Methodology

In this context of research, panel data is employed. Therefore, multiple units are observed over multiple time periods. At a given point in time T_0 , some of the units will receive a certain treatment W and might consequently develop differently from non-treated units. The basic logic that all methods, which will be discussed in this research, share, is that the development of treated units is compared to the development of some form of control units to estimate the effect of the treatment intervention. This is precisely where one of the key differences in the three approaches arises: All units are observed over the whole timespan, however we do not know the hypothetical development of the treated units in the absence of the treatment intervention. The imputation of this so-called counterfactual can take many forms and substantially influences the final estimate of the treatment effect. Whether the regarded units are on aggregate or individual level does not matter, as the calculations themselves do not change. However, in social sciences data is often collected at some level of aggregation (Abadie et al. 2010, p. 1). This does not necessarily pose a hindrance, since interventions are also often performed on an aggregate level, for example per state. The absence of large sample sizes does however impact the ability to construct a suitable comparison unit, as researchers can not rely on the assumption that the distribution of observed, as well as unobserved confounders is going to be the same in the treated and the control group. This problem not only stems from the issue of small sample sizes, but also exists because treatment assignment is usually not random in social sciences (Arkhangelsky et al. 2021, p. 1). It might therefore be challenging to find units, unexposed to the treatment, whose characteristics resemble those of the treated units sufficiently well (Abadie et al. 2010, p. 2). The more different

these characteristics are, the less suitable this unit is to serve as a control unit. Consequently, we cannot separate the change in the outcome of interest caused by the treatment from the change caused by unobserved confounding factors.

Disregarding these aspects, DiD uses a simple and straightforward way, simply collecting all non-treated units to use them as the control group. SC and SDiD on the other hand offer a rather data-driven approach, which aims to provide a more systematic, unambiguous way in constructing the control units, forcing researchers to highlight affinities between treatment and control group (Abadie et al. 2010, p. 8). More details on how the three different methods operate will be discussed in the following.

2.1 Difference in Differences

DiD was supposedly first used in 1855 by John Snow to study the cause of Cholera. Today, it is among the most widely used methods to estimate causal effects, especially in non-experimental settings, which often occur in social sciences (Roth et al. 2023, p. 1). At its core, the average outcome Y for treated units ($W_i = 1$) prior to and after the treatment intervention is computed and the their difference is calculated. The same is done for the control units ($W_i = 0$). Finally, the difference of these two population expectations is taken and this *Difference in Differences* can be conceived of as the treatment effect at time t :

$$\tau_t = \underbrace{\mathbb{E}[Y_{i,\text{post}} - Y_{i,\text{pre}} \mid W_i = 1]}_{\text{Difference for treated units}} - \underbrace{\mathbb{E}[Y_{i,\text{post}} - Y_{i,\text{pre}} \mid W_i = 0]}_{\text{Difference for control units}} \quad (1)$$

Equation 1 yields the so-called Average Treatment effect on the Treated. However, computing the treatment effect using the DiD approach imposes several rather restrictive assumptions on the observed units.

Most importantly, it relies on the so-called *parallel trends* assumption which states that, in absence of the treatment, both the treated as well as the control groups would have developed in parallel (Roth et al. 2023, p. 6). Consequently, any observed difference in post-treatment outcomes between treatment and control group can then be attributed to the treatment intervention.

Second, it requires the absence of any *anticipatory effects*. This means, that the treatment must have no causal effect before its implementation. If anticipatory effects have the same sign as the treatment itself, not accounting for these effects will for example lead to an underestimation of the effect (Malani and Reif, 2015, p. 3).

Furthermore, it imposes the *Stable Unit Treatment Value Assumption (SUTVA)*, a fundamental assumption in the field of causal inference in general. According to SUTVA, the treatment effect of one unit is independent of other units' treatment status (Roth et al. 2023, p. 5).

Lastly, it requires the absence of any kind of intervention, that might also influence the outcome variable of interest. If this assumption is violated, we can no longer disentangle the actual effect of the treatment intervention from other, simultaneously happening, interventions.

Concerning the interpretation, researchers often also require the treatment effect to be *homogeneous*, meaning the treatment effect is assumed to be the same for all treated units.

If these assumptions are met, the treatment effect can be estimated using equation 1. In practice however, it is more common to run a *two-way fixed effects (TWFE) regression* according to the following formula:

$$Y_{i,t} = \alpha_i + \phi_t + \beta \cdot (\mathbf{1}\{t \geq T_0\} \cdot W_i) + \varepsilon_{i,t} \quad (2)$$

The outcome Y_i in period t is regressed on an individual fixed effect α_i , a time fixed effect ϕ_t and an interaction between an indicator function and the treatment status W_i . This term is always 0 in pre-treatment periods and only takes the value 1 in post-treatment periods, if the regarded unit belongs to the treatment population. Consequently β then results in the treatment coefficient. The main advantage of running this Ordinary Least Squares (OLS) regression is, that it also returns standard errors for the estimate, which can be used for statistical inference. If

random draws from the population are considered and the assumption of parallel trends and no anticipatory effects hold, the regression returns consistent estimates with asymptotically normal confidence intervals (Roth et al., 2023, p. 8)

Mathematically, solving the TWFE is equivalent to solving the following minimization problem.

$$\hat{\tau}_{\text{did}}, \hat{\mu}, \hat{\alpha}, \hat{\beta} = \arg \min_{\alpha, \beta, \mu, \tau} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \alpha_i - \beta_t - W_{it} \tau)^2 \right\} \quad (3)$$

However, this equation, in which the weighted sum of squared residuals is minimized to estimate our coefficients, helps highlight the similarities between SC and SDiD, which will become more obvious in the following.

2.2 Synthetic Control

The Synthetic Control (SC) method was first introduced by Abadie and Gardeazabal in 2003 and developed further by Abadie et al. in their 2010 paper to provide a more robust method for causal inference in empirical research when the number of treated units is small or even only 1. The main idea of SC is to use the observed pre-treatment periods and untreated population to construct an artificial control unit, that most closely matches the pre-treatment characteristics of the to be treated population. Assuming there exists an optimal set of weights, in the context of one treated unit, this approach can be expressed as:

$$Y_{1t} = \sum_{j=2}^J w_j Y_{jt} \quad \text{for } t < T_0 \quad (4)$$

The development of this weighted sum of untreated units is then carried on into the post-treatment period ($t > T_0$) to construct the counterfactual Y_{1t}^N . Any differences in the development of this SC and the treated population is then interpreted as the treatment effect in period t :

$$\tau_{it} = Y_{1t} - \sum_{j=2}^{J+1} w_j \cdot Y_{jt}^N \quad (5)$$

Since equation 4 however rarely holds in practice, it is sufficient that this relation can at least be approximated (Abadie et al. 2010, p. 3). This method can also be extended to work with multiple treated units. The artificial control group will then be constructed to match the average of all treated units at time t . In practice, researchers tend to choose comparison groups based on some perceived affinity between them and the treated population (Abadie et al. 2010, p. 1). SC on the other hand offers a data-driven approach which aims to generate a fitting control group, effectively minimizing the degree of ambiguity of the process.

The calculation of the treatment effect via the SC method can also be expressed as:

$$\hat{\tau}_{\text{sc}}, \hat{\mu}, \hat{\beta} = \arg \min_{\mu, \beta, \tau} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \beta_t - W_{it} \tau)^2 \hat{\omega}_{iSC} \right\} \quad (6)$$

Equation 6 already demonstrates that the SC framework no longer includes unit fixed effects, but unit-weights $\hat{\omega}_{iSC}$. These weights are restricted to be positive and sum up to 1. For a detailed derivation of the unit-weights, the work of Abadie et al. provides meaningful insights. The estimate obtained via this method, applied to the real-life intervention of Proposition 99, an anti-smoking campaign including a tax-increase, is generally considered to be more credible, thus validating the SC approach (Arkhangelsky et al. 2023, p. 6). However, compared to DiD, statistical inference requires further work like *placebo studies*. Here, the SC approach is applied to every untreated unit in the overall population. If these studies returned treatment effects similar to the estimate of the treated group, the treatment effect is considered to not be statistically significant (Abadie et al. 2010, p. 9).

2.3 Synthetic Difference in Differences

The SDiD framework can be seen, not only as a combination, but also a refinement of DiD and SC. Similarly to SC, SDiD chooses unit weights to match pretreatment trends of exposed and unexposed units. They result from the following minimization problem:

$$(\hat{\omega}_0, \hat{\omega}_{\text{sdid}}) = \arg \min_{\omega_0, \omega} \sum_{t=1}^{T_{\text{pre}}} \left(\omega_0 + \sum_{i=1}^{N_{\text{co}}} \omega_i Y_{it} - \frac{1}{N_{\text{tr}}} \sum_{i=N_{\text{co}}+1}^N Y_{it} \right)^2 + \zeta^2 T_{\text{pre}} \|\omega\|_2^2 \quad (7)$$

The authors argue that, although the settings in which DiD and SC are applied are usually different, underlying assumptions are similar (Arkhangelsky et al. 2023, p. 2), suggesting the combination of the two approaches. The additive *Zeta*, not seen in SC, is used as a regularization term. Apart from this, the main difference to the unit weights used in the SC framework is, that SDiD contains an intercept term ω_0 . Consequently the pre-treatment trends no longer have to be perfectly matched, but making them parallel is sufficient. This is feasible, because SDiD will, in contrast to SC, again employ unit level fixed effects that can capture constant differences between observed units (Arkhangelsky et al. 2023).

However, SDiD further incorporates time weights that aim to balance pre and post treatment periods for the unexposed population. At their core, these weights are designed to disregard atypical pre-treatment periods, as these do not hold much predictive power for future periods. Incorporating these time weights ensures, that time-specific variations are adequately accounted for, ideally leading to more accurate estimates of the treatment effect. They are computed as:

$$(\hat{\lambda}_0, \hat{\lambda}_{\text{sdid}}) = \arg \min_{\lambda_0, \lambda} \sum_{i=1}^{N_{\text{co}}} \left(\lambda_0 + \sum_{t=1}^{T_{\text{pre}}} \lambda_t Y_{it} - \frac{1}{T_{\text{post}}} \sum_{t=T_{\text{pre}}+1}^T Y_{it} \right)^2 \quad (8)$$

Similar to the unit weights in the SC framework, pre-treatment λ are restricted to be positive and sum up to 1, whereas post-treatment λ are required to have equal weights. In the SC framework it is unclear, how many pre-treatment periods to regard in order to construct an appropriate control group. In DiD all periods are simply given equal weight. SDiD now offers a data-driven approach on how to incorporate pre-treatment periods in the estimation process and offers researchers the possibility to back up their reasoning. Similar to SC, the reliance on the *parallel trend assumption* is decreased, because pre-treatment trends are reweighted and matched (Arkhangelsky et al. 2023, p. 2). For a more detailed derivation of time- and unit-weights, the work by Arkhangelsky et al. from 2023 can be considered.

The calculation of the treatment effect via the SDiD method can ultimately be expressed as:

$$(\hat{\tau}_{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}) = \arg \min_{\tau, \mu, \alpha, \beta} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \mu - \alpha_i - \beta_t - W_{it} \tau)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_t^{\text{sdid}} \right\} \quad (9)$$

As already mentioned, similar to DiD, we again employ unit level fixed effects α_i , which were disregarded in the SC framework. Furthermore, we now not only utilize unit-weights ω_i but also time-weights λ_t . Utilizing these time weights encourages researchers to only regard a subset of pre-treatment periods. In a practical application, the authors found, that SDiD emphasizes periods closer to the intervention. It is therefore interesting to contrast this finding to event studies, in which researchers often deliberately emphasize periods closer to the intervention, or even only regard the last period before treatment as a benchmark (Arkhangelsky et al., 2023, p. 5). Arkhangelsky et al. found that the SDiD approach performs at least as good as DiD in DiD-typical settings and similarly performs at least as good as SC in settings, where SC is typically used (Arkhangelsky et al. 2023, p. 2). Furthermore they found that if treatment assignment is random, DiD, SC and SDiD all yield unbiased results, with SDiD being more precise. However, once treatment assignment is no longer random, SDiD is most successful in mitigating bias (Arkhangelsky et al. 2023, p 11). However, similarly to SC, additional steps to enable statistical inference are

required, presenting a disadvantage compared to DID.

3 Data and Simulations

This research focusses on static block treatment and views treatment as an absorbing state. Therefore, treatment is received only once at the same point in time and once it has been applied, will never be revoked. Mainly two settings will be distinguished: Random and non-random treatment assignment. Every observed unit basically follows a random walk with adjustments: A pre-treatment trend, as well as a post-treatment trend will be added for both groups. Furthermore, the treated population will receive a simple treatment of 1 in the treatment period T_0 . Consequently, the development of one observation can be summarized in the following formula:

$$Y_{it} = Y_{i,t-1} + \mathbf{1}\{G = 0 \ \& \ t < T_0\} \cdot \delta_{0,\text{pre}} + \mathbf{1}\{G = 0 \ \& \ t \geq T_0\} \cdot \delta_{0,\text{post}} \\ + \mathbf{1}\{G = 1 \ \& \ t < T_0\} \cdot \delta_{1,\text{pre}} + \mathbf{1}\{G = 1 \ \& \ t \geq T_0\} \cdot \delta_{1,\text{post}} + \mathbf{1}\{G = 1 \ \& \ t = T_0\} \cdot \tau + \varepsilon_{it}$$

G indicates a group membership taking the value 1 for treated units and 0 otherwise. The δ s represent time-trends for each group prior to and after treatment implementation. Treatment is applied if the observed unit belongs to group 1 and t is equal to the treatment period T_0 . In most applications, initial draws, and noise terms ε_{it} are drawn from the Standard Normal distribution.

To simulate random treatment assignment pre-treatment trends will be parallel. To simulate non-random assignment either pre-treatment trends are adjusted, or the initial draws are taken from a different distribution. Data will further vary in overall sample size, number of pre-treatment periods, treatment heterogeneity and the degree of violation of the parallel trends assumption. To recreate heterogeneous treatment effects, treatment will be drawn from a distribution with increasing standard deviation. The number of regarded periods is set to be 10 in almost all simulations to enhance comparability. All calculations is done in R, Version 2024.04.0+ 735 using the *synthdid*-package by Arkhangelsky et al. while setting a seed of 100. Each simulation will be run 1000 times.

4 Results and Interpretation

[abadie'economic'2003]

4.1 Random treatment assignment

4.2 Non-random treatment assignment

4.2.1 Inverse probability weighting

4.3 Further findings

5 Conclusion