Faculty of Economics and Social Sciences
of the University of Tuebingen

Seminar thesis

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

Research Seminar on Marketing
Prof. Dr. Dominik Papies with M.Sc. Alexandra Becker

Summer term 2024

Benjamin Herzberger
Goethestr. 20
72076 Tuebingen
Phone: 0176 / 82216024
E-Mail: benjamin.herzberger@
student.uni-tuebingen.de

Study Program: Data Science in Business and Economics (M.Sc.)
4^{th} semester
Matriculation Number: 3919677
Date of Submission: June 17, 2024

Table of contents

List of abbreviations

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

List of Figures

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 interventions, also called treatment, is small. Difference in Differences (DiD) is a popular method, used to estimate causal effects. By comparing the development of units receiving treatment to different, untreated units, meaningful insights can be gained. However, DiD relies on rather restrictive assumptions which are not entirely compatible with the previously mentioned problems. Consequently, different methods to overcome these issues are necessary.

In 2003 Abadie and Gardezabal introduced a new approach, the so-called Synthetic Control (SC). This method is specifically helpful in providing reliable estimates, even in the context of the aforementioned problems.

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, 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 is structured as follows: Section 2 provides a more detailed explanation of all estimation methods and previous literature findings. Section 3 will discuss the simulation process. In Section 4, results will be analyzed. Section 5 then concludes this research, aims to provide a brief overview 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 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 time span, but the hypothetical development of the treated units in the absence of the intervention is unknown. 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. 493). 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 the same in both groups. 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. 4088). 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. 494).

Disregarding these aspects, DiD uses a straightforward way, collecting all non-treated units to use 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. 500). More details about the three different methods will be discussed in the following.

2.1 Difference in Differences

DiD 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:

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

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

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 the groups can 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 et al. 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 states (Roth et al. 2023, p. 5).
Lastly, it requires the absence of simultaneously happening interventions, that might also influence the outcome variable of interest.
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:

$$Y_{i,t} = \alpha_i + \phi_t + \beta \cdot (\mathbf{1}\{t \geq \mathrm{T}_0\} \cdot W_i) + \varepsilon_{i,t} \tag{2}$$

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 is treated. Consequently, β results in the treatment coefficient. The main advantage of running this Ordinary Least Squares (OLS) regression is, that it returns standard errors, which can be used for statistical inference. If random draws from the population are considered and the assumptions of parallel trends and no anticipation hold, the regression returns consistent estimates with asymptotically normal confidence intervals (Roth et al. 2023, p. 8).
Mathematically, solving the TWFE regression is equivalent to solving the following minimization problem.

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

However, this equation, in which the weighted sum of squared residuals is minimized to estimate the 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 et al. 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 and treatment assignment is potentially not random. 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:

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

The development of this weighted sum of untreated units is carried on into the post-treatments 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:

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

Since equation 4 rarely holds in practice, it suffices that this relation can be approximated (Abadie et al. 2010, p. 495). To work with multiple treated units, the artificial control group will 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. 493). 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:

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

Equation 6 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 these weights, the work of Abadie et al. provides meaningful insights. The estimate derived from this method, applied to the real-world intervention of Proposition 99, an anti-smoking campaign, is widely regarded as more credible, thereby validating the SC approach (Arkhangelsky et al. 2021, p. 4093). However, compared to DiD, statistical inference requires further work like placebo studies. Here, the SC approach is applied to every untreated unit in the population. If these studies returned treatment estimates similar to the estimate of the treated group, the effect is considered statistically insignificant (Abadie et al. 2010, p. 501). Another drawback of SC is its inability to accurately construct a control group for observations, which exhibited the highest/lowest outcome prior to the intervention, as this can not be approximated by combining lower/higher values (Abadie et al. 2010, p. 502).

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. SDiD again chooses unit weights to match pre-treatment trends of exposed and unexposed units. They result from the following minimization problem:

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

The authors argue that, although the settings in which DiD and SC are applied are usually different, underlying assumptions are similar (Arkhangelsky et al. 2021, p. 4089), 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 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. 2021, p. 4092).
However, SDiD further incorporates time weights that aim to balance pre and post treatment periods for the unexposed population:

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

At their core, these weights are designed to disregard atypical pre-treatment periods, as these do not hold much predictive power for future periods. 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. Using SC, 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 grants 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 reweighed and matched (Arkhangelsky et al. 2021, p. 4089). For a more detailed derivation of time- and unit-weights, the work by Arkhangelsky et al. can be considered.
The calculation of the treatment effect via the SDiD method can ultimately be expressed as:

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

Similar to DiD, unit level fixed effects α_i are employed, which were disregarded in the SC framework. Furthermore, not only unit-weights ω_i but also time-weights λ_t are considered. 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. 2021, p. 4094). 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. 2021, p. 4089). 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. 2021, p. 4098). However, similarly to SC, additional steps to enable statistical inference are required, presenting a disadvantage compared to DID.

3 Data and Simulations

This research focuses 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:

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

G indicates a group membership taking the value 1 for treated units and 0 otherwise. The δs represent linear time-trends for each group prior to and after treatment implementation. Treatment τ is applied if the observed unit belongs to group 1 and t is equal to the treatment period T_0. 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 and set to 0.1. To simulate non-random assignment, pre-treatment trends are adjusted to evolve differently. Data will further vary in sample size, number of pre-treatment periods, treatment heterogeneity and the degree of violation of the parallel trends assumption. Unless stated otherwise, the number of regarded periods is set to 10 and treatment is applied in t = 8 to enhance comparability across simulation designs. Since the simulated datasets are simply customized random walks, no visualizations will be included in the main analysis. All calculations is done are 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

4.1 Random treatment assignment

4.1.1 Effect of sample size on estimates

To examine the effect of the overall population size on the estimation methods, the sample size is systematically increased from 5 to 1000, while only treating one single unit. As figure 1 demonstrates, all methods produce unbiased estimates across all population sizes. This is indicated by the red horizontal line, representing the treatment effect. However, SDiD appears

to be more precise in every simulation design. SC on the other hand, while generally also being more precise than DiD, tends to produce considerable outliers in some iterations.

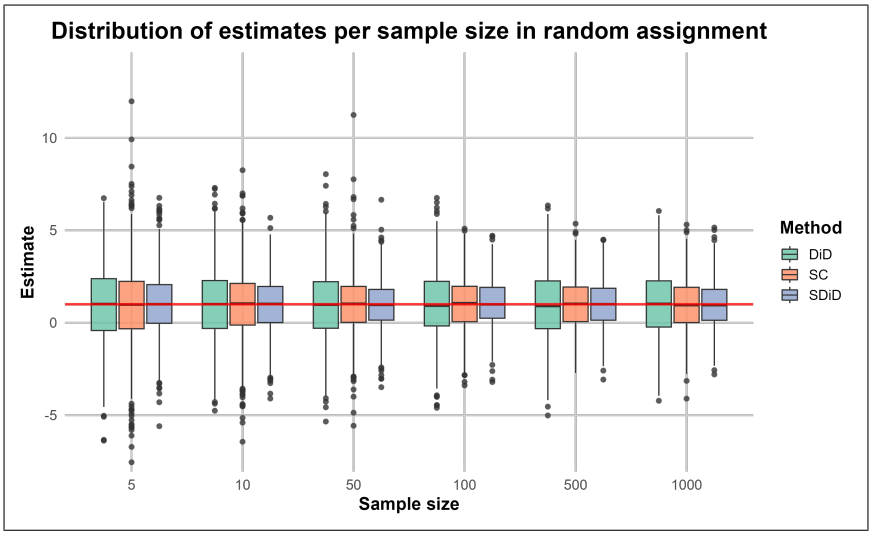


Figure 1: Effect of increasing sample sizes in random treatment assignment

4.1.2 Number of pre-treatment periods

SDiD and SC use pre-treatment periods to construct an artificial control group. With increasing pre-treatment periods more data is available to design a suitable control group. Therefore, it is expected that the number of pre-treatment periods has an effect on the resulting estimates. To isolate this effect, the number of pre-treatment periods is systematically varied from two to nine, while always considering an overall sample of 10 units with 1 treated observation. As Figure 2 shows, all methods are again able to produce unbiased estimates in all simulation settings. However, in this context, DiD is able to perform at least on par with both algorithms until treatment period 5. After that, SDiD again dominates the other methods in terms of precision. The SC approach does again produce heavy outliers and is outperformed by the classical TWFE until treatment period 8.

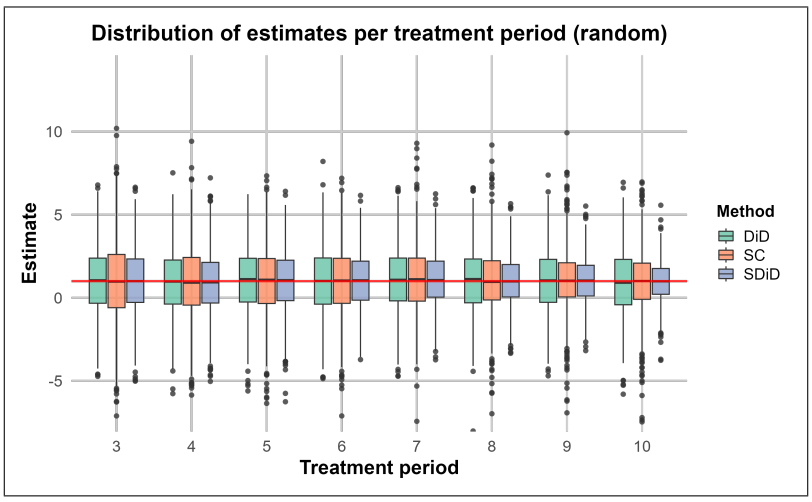


Figure 2: Effect of number of pre-treatment periods in random treatment assignment

4.1.3 Treatment heterogeneity

In order to research the effect of treatment heterogeneity on all methods, treatment is no longer simulated by simply adding 1, but is now drawn from a Normal distribution with mean 1 and increasing standard deviation. Furthermore, for heterogeneity to take effect, more units need to be treated. Therefore, an overall sample size of 100 is considered and 5%/10% will be treated.

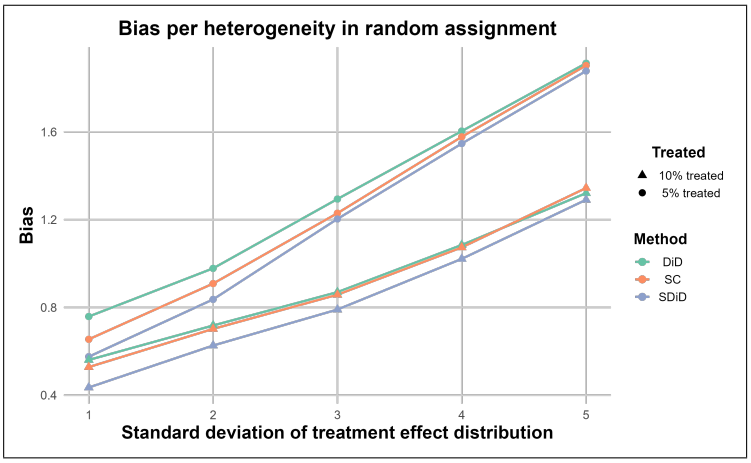


Figure 3: Effect of treatment heterogeneity in random treatment assignment

Figure 3 demonstrates that, in the context of treatment heterogeneity, all methods produce biased estimates. This bias does naturally increase, with rising heterogeneity and is reduced, when more units are treated, as the effect tends towards the mean. Although differences are small, SDiD is best able to mitigate this bias, with DiD being outperformed in almost all configurations.

4.1.4 Parallel trend violation

To induce a violation of the parallel trends assumption, the time trend of the treated population is changed after treatment is applied. The previously parallel trend of 0.1 will systematically be increased up until the value 1, while keeping every other parameter fixed.

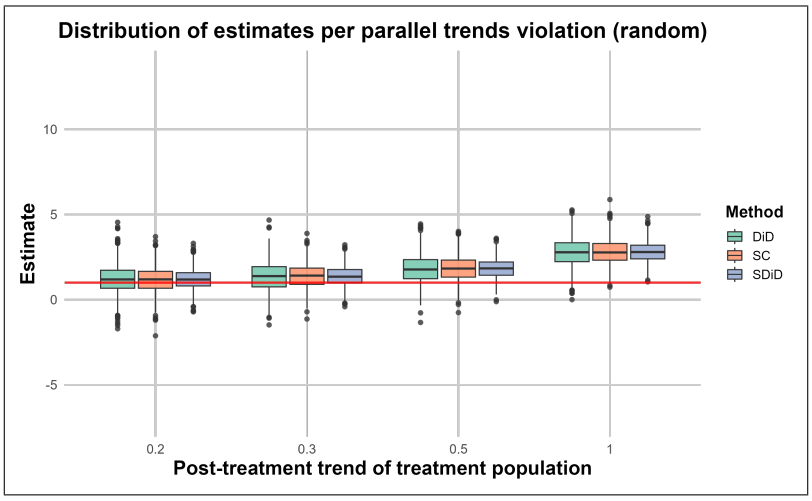


Figure 4: Effect of the degree of parallel trends violation in random treatment assignment

As figure 4 shows, once the parallel trends assumption is violated, all methods produce biased estimates that increase with the degree of trend violation. The expected value for all

three methods seems to be identical and the enhanced precision of SDiD does not necessarily provide an advantage to researchers, if estimates are biased. Consequently, none of the methods are able to disentangle the static treatment from the dynamical time-trends.

4.2 Non-random treatment assignment

In the context of random treatment assignment, the control trend will be set to 0.1, whereas the treated population evolves with a linear time-trend of 0.3.

4.2.1 Effect of sample size on estimates

Figure 5 captures the effect that an increasing overall sample size has on the distribution of our estimates in the context of non-random treatment assignment. Again, while the overall sample increased, only 1 unit was treated. DiD is basically unable to mitigate this bias, no matter the sample size and seems to produce almost identical results across all simulation designs. SC and SDiD however are able to decrease this bias with an increasing sample size, although not eliminating it entirely in the regarded research design. SDiD is however able to mitigate this bias faster, compared to SC and again shows the lowest standard deviation.

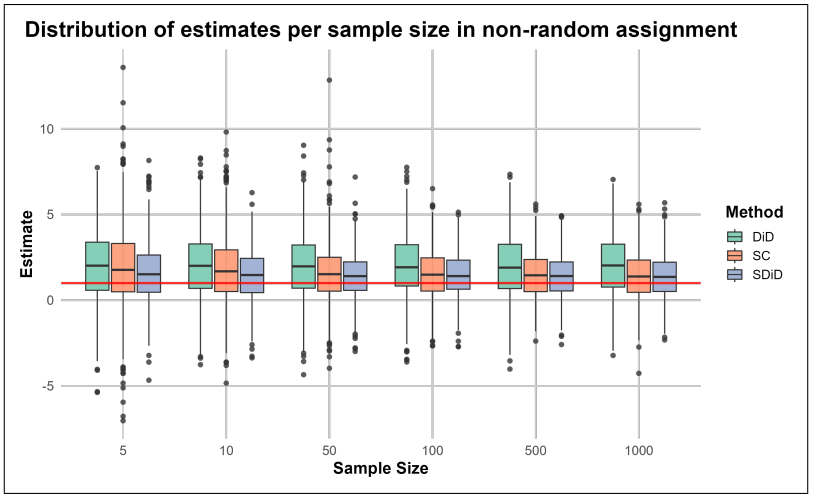


Figure 5: Effect of increasing sample sizes in non-random treatment assignment

4.2.2 Number of pre-treatment periods

To examine the influence of the number of pre-treatment periods in the context of non random treatment assignment, a similar procedure as in random treatment assignment was applied. As Figure 6 shows, DiD is not able to mitigate the bias resulting from different trends, no matter the treatment period. SC again tends to produce wider box plots and considerable outliers, symbolizing less precise estimates. While it is able to decrease the bias gradually, it lacks behind SDiD, which through all simulation designs produces a more narrow estimate distribution and is able to mitigate the bias almost entirely, once the number of pre-treatment periods is large enough to construct a suitable control group.

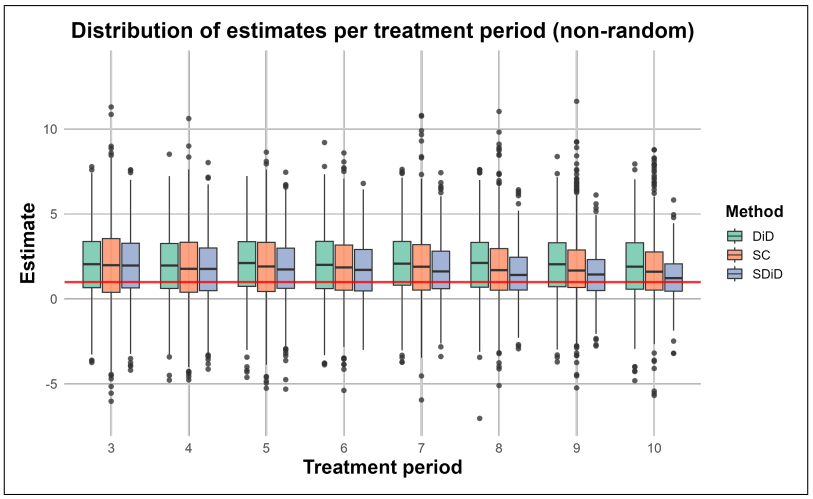


Figure 6: Effect of number of pre-treatment periods in non-random assignment

4.2.3 Treatment heterogeneity

Concerning the effect of treatment heterogeneity in the context of non random treatment assignment, the previous observations of section 4.1.3 almost entirely transfer. Bias increases, once the heterogeneity rises and it can be decreased when considering larger sample sizes. However, the extent to which the three methods suffer from bias has now changed.

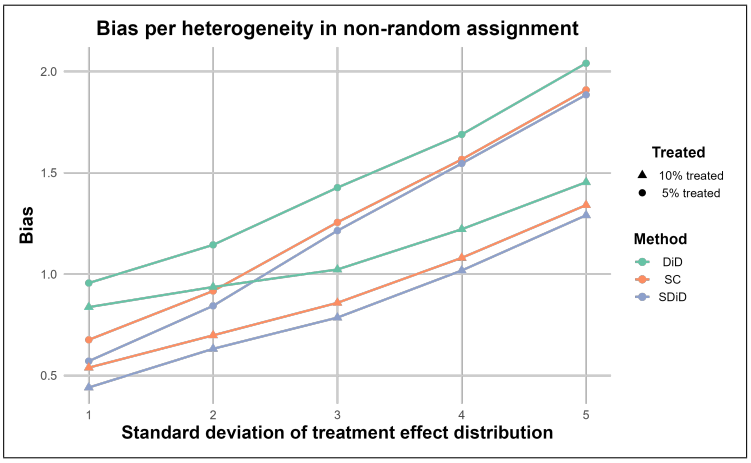


Figure 7: Effect of treatment heterogeneity in non-random treatment assignment

Sorting by bias, the same order of performance as before results. However, with only 5% treated units, SDiD is even able to outperform DiD with 10% treated units up until a standard deviation of 2, in terms of bias.

4.2.4 Parallel trend violation

In contrast to section 4.1.4, trends will be non-parallel to begin with. The control group will therefore develop with a linear time trend of 0.1, where as this value is systematically increased for the treated population. Both trends remain unchanged after the treatment intervention, as Section 4.1.4 already demonstrates the inability of all three methods to handle this violation of the parallel trends assumption. 100 units will be observed overall, 5 of which are treated, to ensure reliable estimates. All other parameters are kept constant. As figure 8 illustrates, DiD

is unable to produce unbiased results in this context. SC and SDiD are able to mitigate the bias resulting from the groups not evolving in parallel to some extent. While they both perform almost identically in most scenarios, SDiD outperforms SC in the final scenario, although it remains biased.

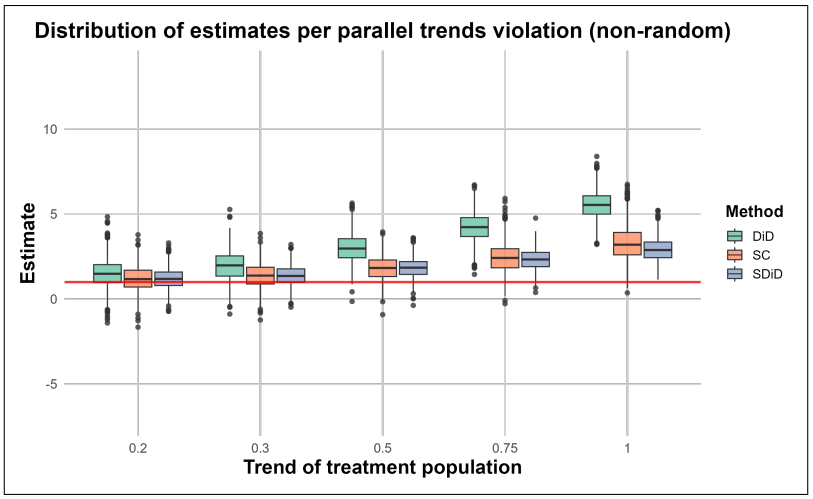


Figure 8: Effect of the degree of parallel trends violation in non-random assignment

4.3 Further findings

4.3.1 Computational constraints

In applied research, computational constraints, such as processing power, can heavily limit researchers' possibilities. It is therefore interesting to examine whether DiD, SC and SDiD differ significantly in terms of computation time. Of course, this is also influenced by efficient code design and hardware, but since the algorithms of SC and SDiD are more complex than DiD, natural differences remain. In order to examine the computation time of each method, the simple setting of random treatment assignment was considered. The overall population size was gradually increased up to 10.000 while always treating 10%. Each simulation was only carried out 100 times and the mean computation time per method was calculated. As Table 1 demonstrates, DiD outperforms both alternatives by far. Regarding the largest sample size, SC does manage to stay under 1 second, but still takes 35 times as long as DiD. Due to its computational complexity however, SDiD is clearly outperformed by both alternatives, taking more than 300 times as long as DiD. These findings are especially interesting, once the aforementioned placebo studies are considered. Redoing the estimation process multiple times to enable statistical inference in SC and SDiD takes an enormous amount of time.

Population Size	DiD	SC	SDiD
10	0.0111	0.0636	0.0399
50	0.0115	0.1030	0.2215
100	0.0115	0.0185	0.2685
250	0.0115	0.0219	0.3698
500	0.0114	0.0390	0.5674
1000	0.0120	0.0750	0.9329
5000	0.0171	0.3875	4.2691
10000	0.0225	0.7993	7.5231

Table 1: Comparison of computation times for DiD (TWFE), SC, and SDiD in seconds

5 Conclusion

This research demonstrated some of the advantages and disadvantages of the classical DiD approach, when compared to the more refined SC and SDiD approach. Confirming previous findings, the analysis was able to demonstrate that, in the case of random treatment assignment, all methods are able to produce unbiased estimates. SDiD strictly outperformed the other methods in terms of precision. Once the number of pre-treatment periods or sample size is large enough, SC outperforms DiD as well. However, once the assumption of parallel trends is violated, all methods are biased. In the context of non-random treatment assignment, again, confirming previous findings, SDiD performed best in mitigating potential estimation biases. If the number of control units and pre-treatment periods is large enough, SDiD could almost eliminate bias entirely. However, in doing so, it is limited by the degree of similarity between the development of treated and control group. On top of that, in both treatment configurations, effect heterogeneity did pose a problem for all methods, especially if the number of treated units is small. Researchers should however keep in mind, that running a TWFE regression in the context of DiD provides several benefits: It is easily implemented, very fast and returns standard errors for statistical inference. Both other methods rely on alternative methods for inference like placebo studies. If the underlying assumptions of the DiD framework are expected to hold, running a TWFE should be the first choice. However, if there is doubt about the validity of these assumptions, which can especially arise in scenarios with few observations and non-random treatment assignment, SC, and if computational constraints are irrelevant, SDiD should be used. Concerning further research, regarding Inverse Probability of Treatment Weighting (IPTW) that mitigates the influence of confounders, putting DiD more on par with SDiD and SC, would be interesting. Unfortunately, since only the dependent variable was simulated here, conditioning on pre-treatment means alone is not sufficient for IPTW to produce meaningful results. Conditioning and matching on observed confounders is expected to work better. Furthermore, this work only regarded static treatment. Although it can be expected, that all findings transfer to the context of dynamically applied treatment, this should ideally be confirmed through further work, thus also presenting interesting further fields of research in this context.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (June 1, 2010). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program". In: *Journal of the American Statistical Association*. Publisher: Taylor & Francis. DOI: 10.1198/jasa.2009.ap08746. URL: https://www.tandfonline.com/doi/abs/10.1198/jasa.2009.ap08746 (visited on 06/13/2024).

Abadie, Alberto and Javier Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country". In: *The American Economic Review* 93.1. Publisher: American Economic Association, pp. 113–132. ISSN: 0002-8282. URL: https://www.jstor.org/stable/3132164 (visited on 06/13/2024).

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager (Dec. 2021). "Synthetic Difference-in-Differences". In: *American Economic Review* 111.12, pp. 4088–4118. ISSN: 0002-8282. DOI: 10.1257/aer.20190159. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20190159 (visited on 06/13/2024).

Malani, Anup and Julian Reif (Apr. 1, 2015). "Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform". In: *Journal of Public Economics* 124, pp. 1–17. ISSN: 0047-2727. DOI: 10.1016/j.jpubeco.2015.01.001. URL: https://www.sciencedirect.com/science/article/pii/S0047272715000122 (visited on 06/13/2024).

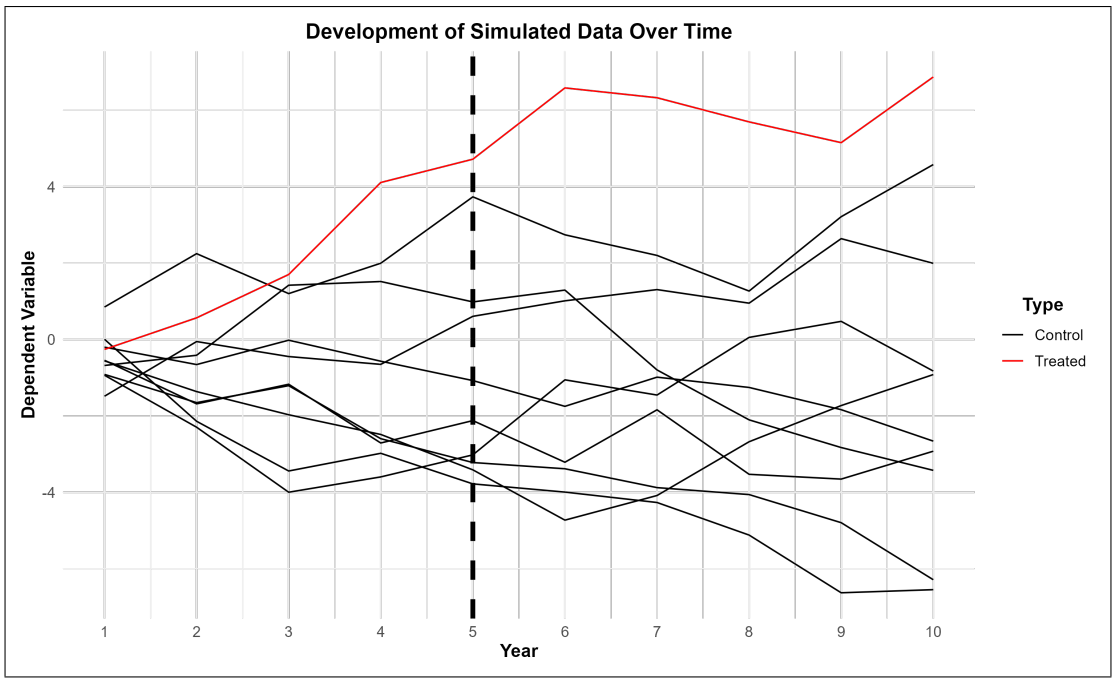
Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe (Aug. 1, 2023). "What's trending in difference-in-differences? A synthesis of the recent econometrics literature". In: *Journal of Econometrics* 235.2, pp. 2218–2244. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2023.03.008. URL: https://www.sciencedirect.com/science/article/pii/S0304407623001318 (visited on 06/13/2024).

Appendix A: Exemplary visualizations

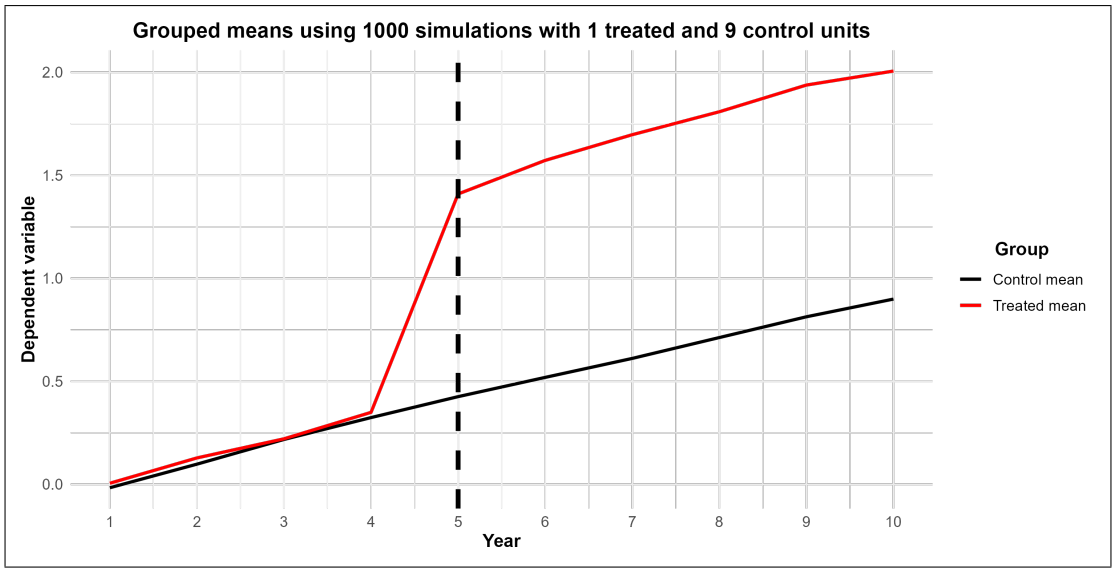
Since all simulated datasets in this research were simply adjustments of random walks, no visualizations were included. The nature of a random walk is expected to be well-known and straightforward, which is why individual visualizations were left out. Furthermore, because of the sheer number of simulation designs in this research, the number of plots to include would have exceeded any rational limit for presentation. However, as an exemplary visualization for both treatment assignment options, the six plots below were included: For each type of assignment, one simulation design was regarded. Furthermore, out of these 1000 respective simulations, 1 random individual data frame will be plotted. The aggregate development of all observed units across the 1000 simulations and an alternative depiction of the estimates' distribution is also included.

Random treatment assignment

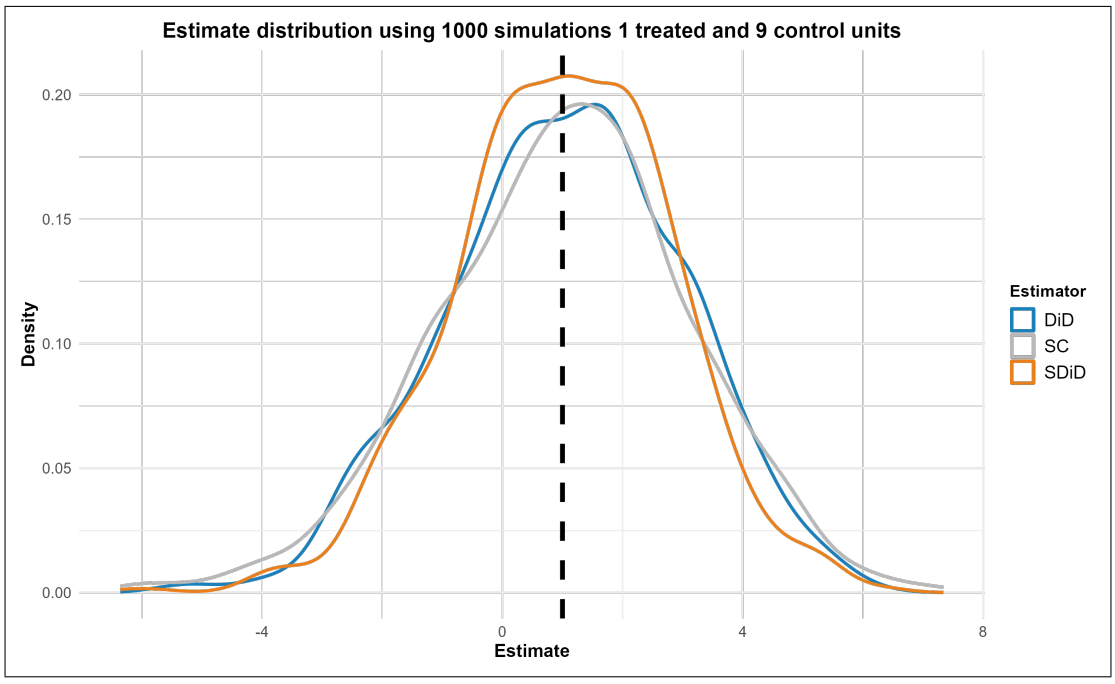
In the context of random treatment assignment, the configuration of varying treatment periods fromn Section 4.1.2 will be considered. More precisely, the simulation design with treatment period 5 will be visualized.



Development of one single simulated data frame over time in random treatment assignment. Overall sample size is 10, number of treated units 1. Treatment was applied in period 5.



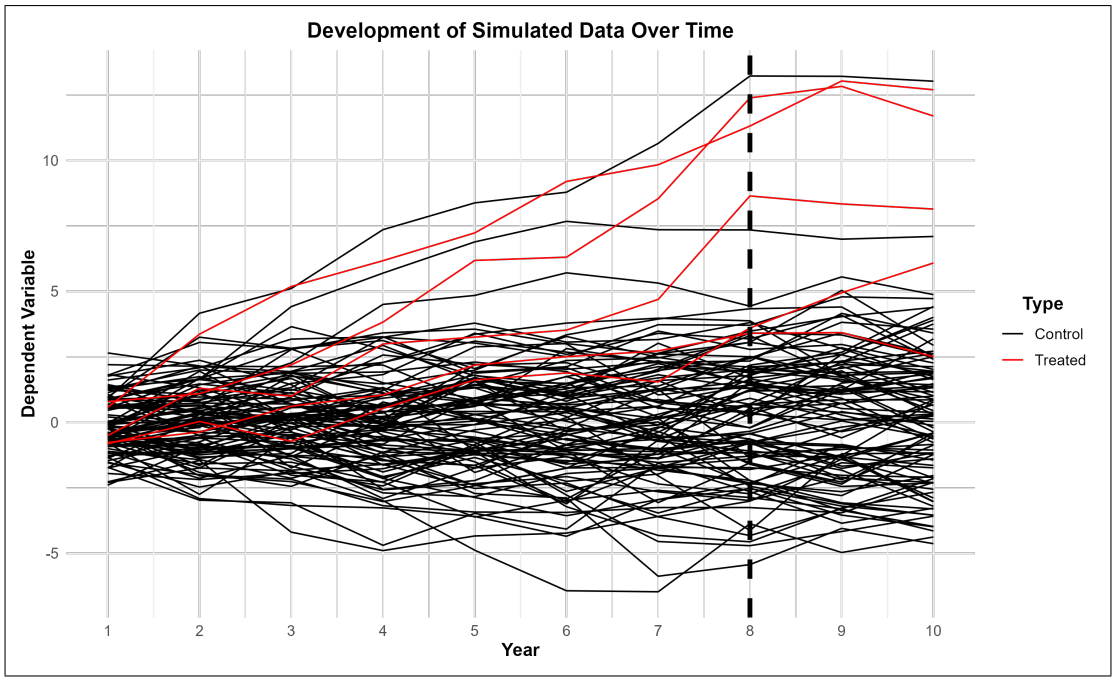
Development of observed units over time in random treatment assignment, averaged over all 1000 simulations. Overall sample size is 10, number of treated units 1. Treatment was applied in period 5.



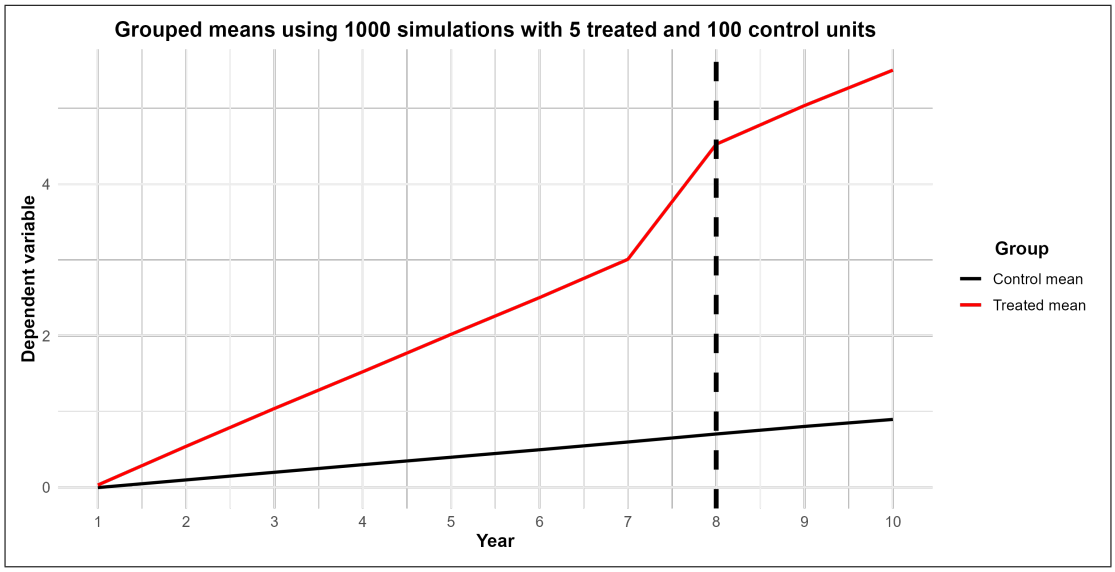
Distribution of estimates in random treatment assignment with treatment period 5, sample size of 10 and 1 treated unit.

Non-random treatment assignment

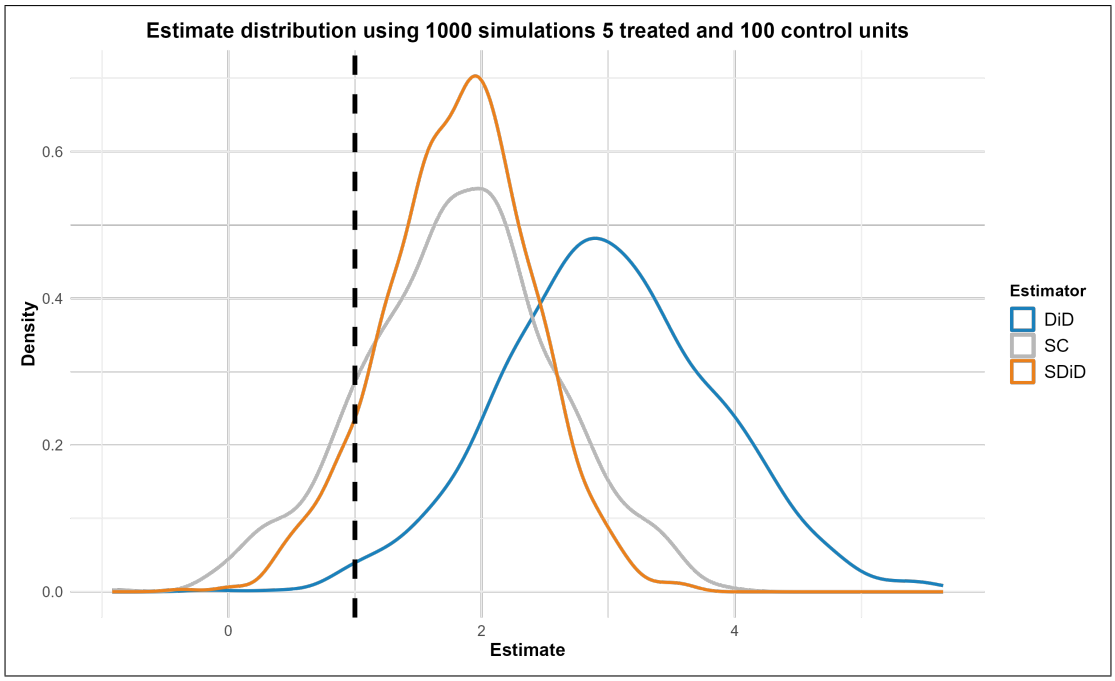
In the context of non-random treatment assignment, the configuration of varying trends for the treated population from Section 4.2.4 will be considered. To be more accurate, the setting with treatment population trend set to 0.5 from the beginning will be regarded.



Development of one single simulated data frame over time in non-random treatment assignment. Overall sample size is 105, number of treated units 5. The treatment-group trend was set to 0.5, the control-group trend to 0.1 and treatment was applied in period 8.



Development of observed units over time in non-random treatment assignment, averaged over all 1000 simulations. Overall sample size is 105, number of treated units 5. The treatment-group trend was set to 0.5, the control-group trend to 0.1 and treatment was applied in period 8.



Distribution of estimates in non-random treatment assignment. Overall sample size is 105, number of treated units 5. The treatment-group trend was set to 0.5, the control-group trend to 0.2 and treatment was applied in period 8.

Examination and Assessment Honor Code

All members of the School of Business and Economics at the University of Tübingen (faculty, students, and alumni) share a commitment to honesty and integrity. In particular, all members follow the standards of scholarship and professionalism in all examinations and assessments.

By submitting these assignments, students agree to comply with this Examination and Assessment Honor Code.

Students who violate this Honor Code are in breach of this agreement and are subject to sanctions imposed by the School of Business and Economics, the University and its responsible bodies (e.g., the board of examiners ("Prüfungsausschuss")).

1. All members of the School of Business and Economics at the University of Tübingen (faculty, students, and alumni) have the obligation to report known violations to the responsible bodies (e.g., the board of examiners ("Prüfungsausschuss") or the Dean of Programs)
2. You must not represent another's work as your own
3. You must not receive inadmissible assistance of any sort before, during, or after an examination or other forms of course work that is subject to assessment by faculty members
4. You must not provide inadmissible assistance of any sort before, during, or after an examination or other forms of course work that is subject to assessment by faculty members
5. Violations of this Honor Code will be handled according to the rules and regulations laid out in the rules for this course

I hereby assure that my submission is in line with the Examination and Assessment Honor Code outlined above.

Tübingen, 16.06.2024, B. Herzberger