Two-period difference-in-differences

Wojciech Wideł
wojciech.widel[at]gmail.com

January 6, 2025

Abstract

The aim of this document is to provide a concise yet detailed description of the difference-in-differences analysis with measurements taken at two points in time. The methodology and some of the potential pitfalls are described, a step-by-step guide is provided.

1 Introduction

The goal of the difference-in-differences (DID) analysis is to estimate the effect of an intervention (such as a medical treatment or advertising campaign) on an outcome of interest (response variable, e.g., risk of death) in a population, by investigating the change in differences between average outcomes over time in two groups: the treatment and the control group. See Figure 1 for an illustration.

The analysis is performed by modelling the outcome as a function (typically, a linear one) of time, intervention and relevant explanatory variables. The impact of the intervention on the outcome variable is then derived from a fitted model. If a linear regression model is used, then the impact (the *average treatment effect on the treated*, ATT) will be the model coefficient corresponding to the interaction of time and treatment.

1.1 Notation

For this document, the following assumptions are made:

1. The analysis is conducted on a *sample* of n individuals from a *population* of size $N \gg n$.

2. Available data describe two points in time: before and after the treatment has been administered. Thus, for every individual in the sample there are at most two observations in the data. Let m be the total number of observations.

3. The outcome for the ith observation is denoted by y_i. Covariates describing the observation are stored in a vector $x^{(i)}$. They should contain all variables that might influence the outcome. The outcome modelling involves also indicators of time (pre- or post-intervention) and treatment (describing whether or not the treatment has been administered to the corresponding individual). For notational convenience, the two indicators and the interaction between them (i.e., their product) are considered separately and denoted by $x^{(i)}_{time}$, $x^{(i)}_{treatment}$ and $x^{(i)}_{interaction}$, respectively.Again, the goal of the analysis is to estimate the impact of $x_{interaction}$ on the outcome.

1.2 Assumptions

For the causal inferences drawn from a DID analysis to be reliable, several assumptions need to be satisfied.

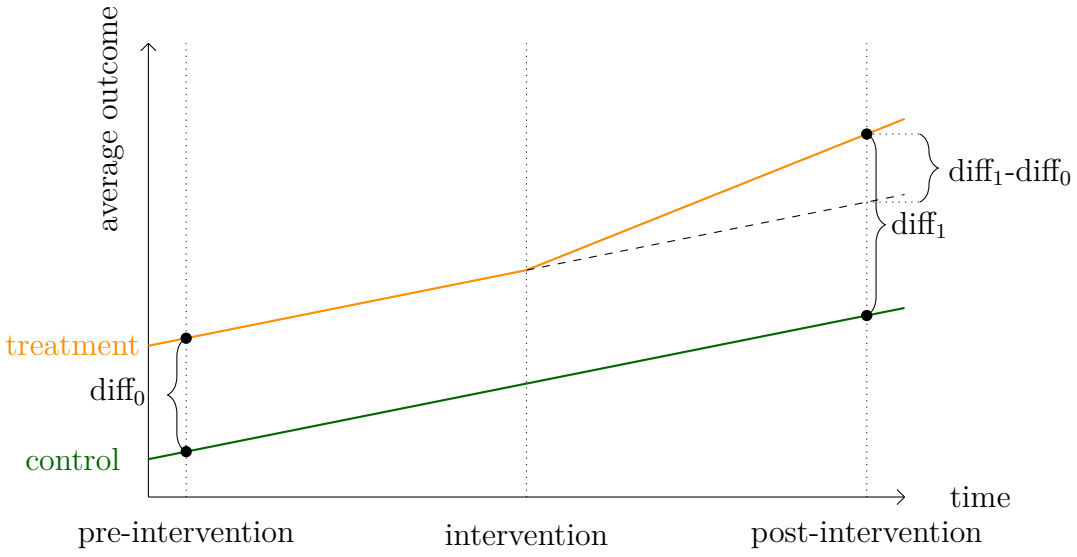

Figure 1: Initially, a difference of diff_0 between average outcomes in the two groups is observed. After the treatment is administered, this difference is equal to diff_1. Can the difference in differences $\text{diff}_1 - \text{diff}_0$ be ascribed to the treatment?

1.2.1 SUTVA

Stable Unit Treatment Value Assumption (SUTVA) states that

1. The (potential) outcome value of an individual should not be affected by the particular assignment of treatments to other individuals.

2. For each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.

Example 1 *Suppose that an individual is exposed to a marketing campaign in social media (i.e., they belong to the treatment group) but a number of their friends or followers are not, and some of them belong to the control group. In this situation the first assumption is violated since the treatment might indirectly affect the untreated friends via the treated individual sharing their experiences or opinions with them.*

Example 2 *Suppose that the treatment is defined as exposure to news articles of a certain type, i.e., an individual is considered treated if an appropriate news article headline appears in their Facebook feed. This violates the second assumption since the impact of being exposed to two headlines is likely to be different than the impact of being exposed to, say, twenty.*

1.2.2 Exogeneity

In the context of DID, the *exogeneity assumption* means that neither the pre-treatment values of the outcome variable nor the covariates are affected by the treatment assignment.

Example 3 *Suppose that a study is examining the effect of a new medication on blood pressure. If the medication causes changes in other health outcomes that are used as covariates, then the exogeneity assumption is violated – the treatment affects the outcome indirectly via covariates, and this effect is not visible in the effect estimate.*

1.2.3 Parallel trends assumption

The *parallel trend assumption* requires that, in the absence of treatment, the difference in average outcomes between the treatment and the control group is constant over time. Violation

of this assumption might mean that another intervention affecting the outcome variable happened between the two points in time. In such a case, estimating the effect of the intervention of interest might be impossible.

Do note that it might happen that the outcome y satisfies the parallel trend assumption, but its transformation, e.g., $\log(y)$, does not, and vice versa [1]. Be careful when applying transformations to the outcome.

It is possible that, while the observations are marked as pre- or post-intervention, the actual dates of the observation are available. In such a case, it is advised to perform *placebo test* to assess whether the parallel trend assumption is violated.

Placebo test: Pick a date from the pre-intervention observations, so that there is sufficiently many treatment and control pre-intervention observations preceding and following the date. Perform a DID analysis using the selected date as a fake intervention date, using only the pre-intervention observations from the original sample. If a statistically significant DID effect is found, then it is highly likely that the parallel trend assumption is not satisfied.

2 Outcome modelling using linear regression

It is assumed that there is a relationship between the outcome and all the covariates, i.e., that there is a function f such that

$$y_i = f(x^{(i)}, x^{(i)}_{time}, x^{(i)}_{treatment}, x^{(i)}_{interaction}) + \epsilon_i, \epsilon_i \overset{\text{iid}}{\sim} (0, \sigma^2), \text{ for } i \in \{1, \dots, m\}.$$

Since f is unknown, either the outcome itself, or – in the case of binary outcome – its logit, is modelled as a linear function of the covariates:

$$\hat{y}_i = \beta x^{(i)} + \beta_{time} x^{(i)}_{time} + \beta_{treatment} x^{(i)}_{treatment} + \beta_{interaction} x^{(i)}_{interaction},$$

or,

$$\log \frac{\hat{p}(y_i = 1)}{1 - \hat{p}(y_i = 1)} = \beta x^{(i)} + \beta_{time} x^{(i)}_{time} + \beta_{treatment} x^{(i)}_{treatment} + \beta_{interaction} x^{(i)}_{interaction}.$$

We are interested in the average treatment effect on the treated. In the case of continuous outcomes this effect is $\beta_{interaction}$ – if an individual received the treatment, then, post-intervention, their outcome will change by $\beta_{interaction}$, on average. In the case of binary outcomes, the treatment changes the odds ratio $\hat{p}(y_i = 1)/(1 - \hat{p}(y_i = 1))$ by the factor of $e^{\beta_{interaction}}$. For example, if $\beta_{interaction} = 0.13$, then the odds ratio will change by $e^{0.13} = 1.14$, or, in other words, it will increase by 14%.

Remark 1 *An estimate of the treatment effect is obtained by fitting a model to data. If the model does not explain data well, the estimate cannot be trusted. Standard measures should be taken to ensure good fit: features engineering (in particular, addition of non-linear transformations of covariates), features selection, candidate models comparison using appropriate metric (e.g., Akaike information criterion, AIC).*

From now on we will focus on continuous outcomes, to shorten the presentation. The same issues need to be tackled in the case of binary outcomes, it's just the loss function that is different.

[1] Suppose that in the absence of treatment the differences between average outcomes are $20 - 10 = 10$ at time 0, and $25 - 15 = 10$ at time 1; the parallel trend condition is satisfied. But on the log scale, it is not: $\log(20) - \log(10) = 0.693$ but $\log(25) - \log(15) = 0.511$. For conditions ensuring that the parallel trend assumption remains satisfied after a transformation, see `https://arxiv.org/pdf/2010.04814`

The estimates of model coefficients are obtained using the ordinary least squares (OLS) method, i.e., by minimizing the sum of the squared differences between the observed outcomes and their linear estimates:

$$\widehat{\beta}, \widehat{\beta}_{time}, \widehat{\beta}_{treatment}, \widehat{\beta}_{interaction} = \operatorname{argmin} \sum_{i=1}^{m} \Big(y_i - \\ - (\beta x^{(i)} + \beta_{time} x^{(i)}_{time} + \beta_{treatment} x^{(i)}_{treatment} + \beta_{interaction} x^{(i)}_{interaction}) \Big)^2.$$

To minimize the bias of the OLS estimates, several potential issues need to be tackled.

Issue 1: covariates imbalance between the sample and the population If distributions of (some of) the covariates in the sample and in the population are very different, then the causal inferences drawn from the analysis do not apply to the population.
Example: In the collected sample there are 90 twenty-year-olds and 10 fifty-year-olds. The observations corresponding to the young individuals dominate in the sample, and so the estimates of linear regression coefficients will be biased towards them. However, the average age in the population of interest is 45. It cannot be assumed that the average treatment effect in the population will be the same as its estimate obtained from the sample.

Issue 2: covariates imbalance between the treatment and the control group If distributions of (some of) the covariates in the treatment and in the control group are very different, then the estimate of the causal effect of the treatment cannot be trusted.
Example: In the same sample, 85 young individuals did not receive treatment, and the remaining individuals did. The treatment is in fact far less effective in older population. Furthermore, it so happens that all the untreated young individuals lead a very healthy lifestyle, which by itself improves their condition. The combination of this two factors leads to a severe underestimation of the actual average treatment effect in the population.

3 Improving covariates balance

There are many methods that can be applied to improve covariates balance. Here, we focus only on one of them, namely, the *inverse probability weighting*.

Assume that A and B are the two groups of individuals that we want to balance. In practice, we are interested in two cases: A is the sample and B is the target population, or A and B are the treatment and the control group. Let

$$e(x) = Pr(\text{individual with covariates } x \text{ belongs to } A)$$

be the (unknown) *propensity score* of x. It can be used to define weights

$$w(x^{(j)}) = \begin{cases} 1/e(x^{(j)}), \text{ if } j\text{th individual belongs to } A, \\ 1/(1 - e(x^{(j)})), \text{ otherwise} \end{cases}$$

for every individual in the sample. A weight is the inverse of the probability of the individual being in their group. Thus, if values of covariates of an individual are rare/underrepresented in their group, the propensity score value is low, and the weight is big. Conversely, if the values occur often in the group, the corresponding propensity score will be big, and thus the weight will be small. Using these weights in the OLS loss function leads to the following estimates of the model parameters.

$$\widehat{\beta}, \widehat{\beta}_{time}, \widehat{\beta}_{treatment}, \widehat{\beta}_{interaction} = \operatorname{argmin} \sum_{i=1}^{m} w(x^{(i)}) \Big(y_i - \\ - (\beta x^{(i)} + \beta_{time} x^{(i)}_{time} + \beta_{treatment} x^{(i)}_{treatment} + \beta_{interaction} x^{(i)}_{interaction}) \Big)^2.$$

Now, in the minimized expression, the contribution of the ith individual is as if the original sample contained $w(x^{(i)})$ individuals with the same values of covariates.

Example 4 *Recall the sample consisting of 90 twenty-year-olds and 10 fifty-year-olds and assume that age is the only covariate. If the population consists of a 100 fifty-year-olds and a 100 twenty-year-olds, then the propensity scores are* $e(20) = 0.9, e(50) = 0.1$*, and the corresponding weights are* $w(20) = 10/9, w(50) = 10$*. That is, every young individual in the sample represents* $10/9$ *individuals from the population, and every middle-aged individual in the sample represents* 10 *individuals from the population. Employing the weights in the OLS loss function ensures that the fitted coefficients generate similarly accurate predictions for all individuals from the population and that the average treatment effect estimate is more accurate.*

Remark 2 *In practice, the propensity score function is unknown. To compute IPW weights, an estimate* $\hat{e}(x)$ *is created using a logistic regression model fitted to data.*

Remark 3 *If 80% of the surveyed answered one question, and only 20% answered another one, the answers come from essentially different samples. Should one try using weights created for all the surveyed individuals, the two weighted samples would correspond to different populations. So, if the response rates for different questions in survey data vary, separate weights should be created for every question.*

One last step to be performed before modelling is to normalize the weights within the treatment arms (treatment or control group):

$$\widetilde{w}(x^{(j)}) = \frac{w(x^{(j)})}{\sum\limits_{\text{individual } i \text{ in the group of } j} w(x^{(i)})}.$$

Intuitively, they are proportions, within their treatment groups, of observations in the target population that the individual represents. These are the weights to be used in the modelling. Formally, they lead to more stable treatment effect estimate[2].

4 Step-by-step two-period difference-in-differences

1. Get and clean data, including outcomes and covariates.
2. If the data contain answers to multiple questions, check the response rates and decide whether separate weights should be created for every question.
3. Perform placebo test, if possible.
4. Check the covariates balance between the sample and the target population.
5. Check the covariates balance between the treated and the control group.
6. Create appropriate IPW weights. Prioritize treated vs control balance, since in its absence the treatment effect estimate can be severely biased. If the weights achieving this balance destroy the sample vs population balance, use them and the sample to create a description of the population to which the inference does apply (e.g., using weighted means of numerical covariates).
7. Add non-linear transformations of covariates to the set of candidate features. Use AIC or BIC for features and model selection. Fit the best model.
8. Report on the estimated average treatment effect, along with the corresponding p-value, confidence interval and whatever else your client expects.

[2]See, e.g., page 5 in `https://www2.stat.duke.edu/~fl35/teaching/640/Chap3.4_observational_weighting.pdf`