The Triple Difference Estimator

BY Andreas Olden and Jarle Møen

DISCUSSION PAPER

NHH

Institutt for foretaksøkonomi
Department of Business and Management Science

FOR 1/2020

ISSN: 1500-4066
April 2020

The Triple Difference Estimator*

Andreas Olden[†] Jarle Møen[‡]

April 18, 2020

Abstract

Triple difference has become a widely used estimator in empirical work. A close reading of articles in top economics journals reveals that the use of the estimator to a large extent rests on intuition. The identifying assumptions are neither formally derived nor generally agreed on. We give a complete presentation of the triple difference estimator, and show that even though the estimator can be computed as the difference between two difference-in-differences estimators, it does not require two parallel trend assumptions to have a causal interpretation. The reason is that the difference between two biased difference-in-differences estimators will be unbiased as long as the bias is the same in both estimators. This requires only one parallel trend assumption to hold.

Keywords: triple difference, difference-in-difference-in-differences, difference-in-differences, DID, DiDiD, parallel trend assumption

JEL Codes: C10, C18, C21

*This paper is a methodological companion paper to Olden (2018). We are grateful to Erik Øiolf Sørensen and Håkon Otneim for useful discussions and comments.

[†]Research Scholar. NHH Norwegian School of Economics, Department of Business and Management Science. N-5045 Bergen. E-mail: andreasolden@gmail.com

[‡]Corresponding author. NHH Norwegian School of Economics, Department of Business and Management Science. N-5045 Bergen. E-mail: jarle.moen@nhh.no

1 Introduction

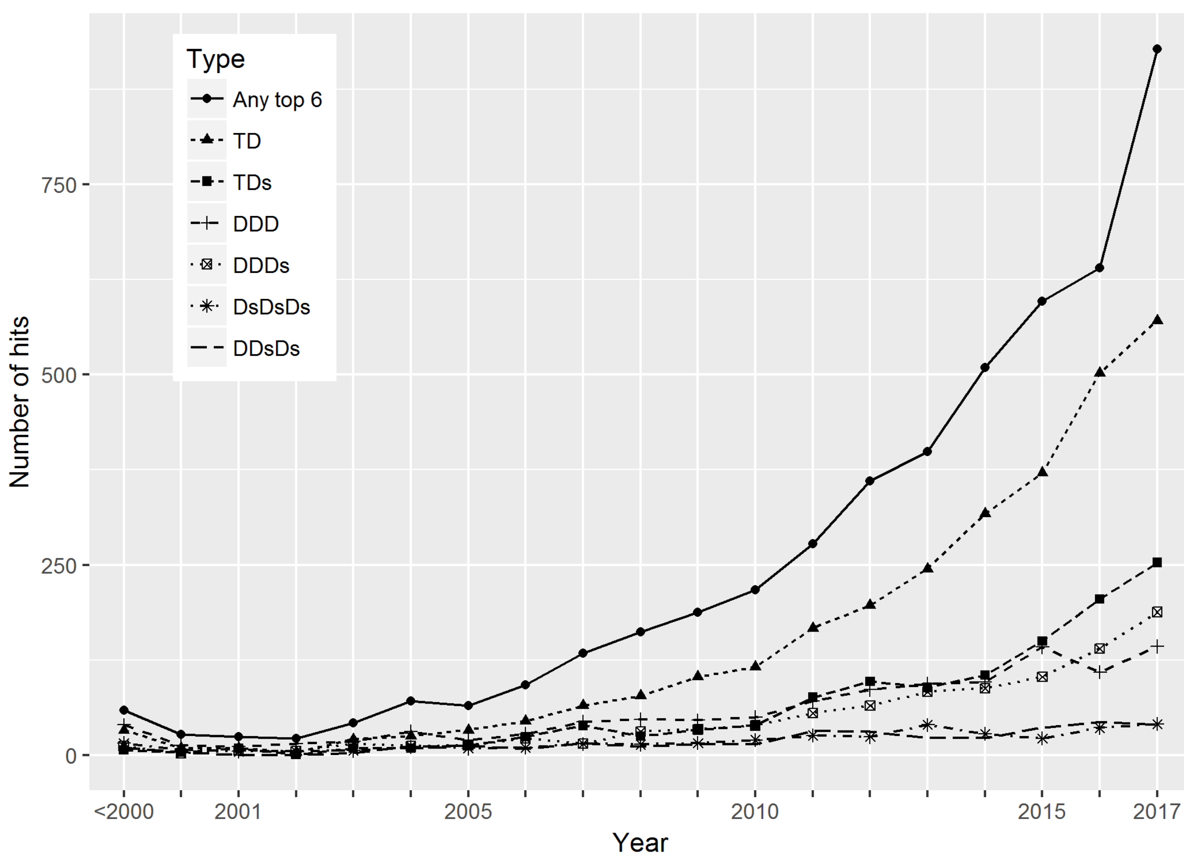
The triple difference estimator is widely used, either under the name "Triple difference" (TD) or the name "difference-in-difference-in-differences" (DDD), or with minor variations of these spellings. Triple difference is an extension of double differences and was introduced by Gruber (1994). Even though Gruber's paper is well cited, very few modern users of triple difference credit him for his methodological contribution. One reason may be that the properties of the triple difference estimator are considered obvious. Another reason may be that triple difference was little more than a curiosity in the first ten years after Gruber's paper. On Google Scholar, the annual number of references to triple difference did not pass one hundred until year 2007. Since then, the use of the estimator has grown rapidly and reached 928 unique works referencing it in the year 2017, see Figure 1.

Looking only at the core economics journals *American Economic Review*, *Journal of Political Economy* and *Quarterly Journal of Economics*, we have found 32 articles using triple difference between 2010 and 2017, see Table A2. A close reading of these articles reveals that the use of the triple difference estimator to a large extent rests on intuition. The identifying assumptions are neither formally derived nor generally agreed on. We fill this void in the literature and give a complete presentation of the triple difference estimator.

The triple difference estimator can be computed as the difference between two difference-in-differences estimators. Despite this, we show that the triple difference estimator does not require two parallel trend assumptions to have a causal interpretation. The intuition is that the difference between two biased difference-in-differences estimators will be unbiased as long as the bias is the same in both estimators. In that case, the bias will be differenced out when the triple difference is computed. This requires only one parallel trend assumption, in ratios, to hold. In fact, the sole purpose of subtracting the second difference-in-differences is to remove bias in the first. Gruber (1994) states the identification requirement verbally, but the result has not been formalized in the econometric literature, and it is overlooked in most of the recent applications.

The rest of the paper is organized as follows: Section 2 gives a short overview of the use of the triple difference estimator. Section 3 derives the triple difference estimator. Section 4 shows that the triple difference estimator can be viewed as the difference between two difference-in-differences estimators. Section 5 derives the identifying assumptions. Section 6 shows that the triple difference estimator can also be viewed as a difference-in-differences using a ratio between two outcome variables. Section 7 discusses the naming of the estimator and provides a short overview of common naming practices. Section 8 provides concluding remarks.

Figure 1: Historical development of the use of the triple difference estimator



Note: *T* denotes triple, *D* denotes difference, and *s* denotes a plural s. *Any top 6* is created by an OR-statement with the six most common ways to reference the model, making it the most accurate estimate of number of works using the estimator.

2 The triple difference literature

The most authoritative and formal treatment of the triple difference estimator seems to be an NBER summer institute lecture note on difference-in-differences estimation by Imbens and Wooldridge (2007). In the introductory "Review of Basic Methodology" chapter they include a simple triple difference estimator. We expand and complement their note in two important and related ways. First, we discuss the assumptions needed to identify a causal effect, while they only present an estimator. Second, we present a fully general estimator allowing for eight different conditional outcomes, while their estimator represents a special case with only six conditional outcomes.[1]

Other authoriative sources treat the topic only in passing. In their famous text book, *Mostly Harmless Econometrics*, Angrist and Pischke (2008, p. 242) write that "A modification of the two-by-two DD setup with possibly improved control groups uses higher-order contrast to draw causal inference". The authors then go on to explain the basic setup using Yelowitz (1995) as an example. They do not discuss or present the estimator, nor the identifying assumption. They simply conclude that "This triple-difference model may generate a more convincing set of results than a traditional DD analysis".

Lechner (2011, p. 3) follows a similar avenue in his monograph *The estimation of causal effects by difference-in-difference methods.* He uses Yelowitz (1995) as an example of triple difference, and states that "the basic ideas of the approach of taking multiple differences are already apparent with two dimensions. Thus, we refrain from addressing these higher dimensions to keep the discussion as focused as possible."

A look at Yelowitz (1995) reveals that he does not go into depth on the estimator and

[1] A general triple difference setup has two groups (A and B), two states (treatment and control), and two time periods (pre and post). This gives eight conditional outcomes. Even though Imbens and Wooldridge (2007) start out with a setup that is identical to ours in all respects except notation (compare their Equation 1.3 to our Equation 1), the estimator presented in their Equation 1.4, lacks the term $(\bar{Y}_{C,A,Post} - \bar{Y}_{C,A,Pre})$, which is the last term in our Equation 4. Hence, they implicitly assume that this term is zero, i.e they assume that there are no time trends or shocks that are specific either to group A or to the control state, C. Note also that their parameter of interest, δ_3, cannot be calculated using the triple difference regression framework specified in their equation 1.3 as they have eight parameters, but only six identifying groups (lacking C, A, Pre and $C, A, Post$).

the identifying assumptions. Instead, he cites Gruber (1994) and Gruber and Poterba (1994). Gruber and Poterba (1994), however, refer back to Gruber (1994).

In his single-authored 1994 article, Gruber analyzes the labour market effects of mandated maternity benefits. Gruber explains the setup as follows:

> I compare the treatment individuals in the experimental states to a set of control individuals in those same states and measure the change in the treatments' relative outcomes, relative to states that did not pass maternity mandates. The identifying assumption of this "differences-in-differences-in-differences" (DDD) estimator are fairly weak: it simply requires that there be no contemporaneous shock that affects the relative outcomes of the treatment group in the same state-years as the law".

We have also looked at all articles applying triple difference (using one of the six most common ways of referencing the estimator) in *American Economic Review*, *Journal of Political Economy*, and *Quarterly Journal of Economics* between 2010 and 2017. As seen in Table A2, we found a total of 32 articles, 16 articles in AER, five in JPE and 11 in QJE. Of these articles Muehlenbachs et al. (2015), Hornbeck (2010), and Shayo and Zussman (2011) show some version of the estimator itself, indicating that it is not entirely obvious. In a similar spirit, Walker (2013) shows the error term of the triple difference estimator and uses it for discussion of robustness. Only Nilsson (2017) cites Gruber (1994).

We will later show formally that a parallel trend assumption very similar to the difference-in-differences approach is needed for the estimated effect to have a causal interpretation. The parallel trend in DDD is, however, on a differential between two categories. In some applications this is stated verbally. Walker (2013, p. 1805) writes e.g. that "[t]he identifying assumption in this class of models is that there are no other factors generating a difference in differential trends between production decisions in regulated and unregulated manufacturing firms." [2]

[2] Some other articles in our sample have similar formulations. Hoynes et al. (2016, p. 925-926) write that

Most of the other 32 top journal articles present some intuition of what the estimator is robust against, but otherwise the information presented varies considerably. Only a few of the authors discuss a common trend or parallel trend assumption, and as the triple difference is based on a strong parallel trend assumption, it is also disturbing to see that a large part of the articles do not include unconditional plots of the outcome series they are studying. This makes it impossible to visually assess potential trends.

In tables A3a and A3b in the appendix, we present the 50 most cited articles referencing the estimator, numbered and ordered by number of citations. There has been almost 5000 papers referencing the estimator since 1994, and it is natural to think that some of the most cited triple difference articles are methodological or represent early use of the methodology. Seven of the 50 most cited articles list Gruber as a co-author.[3] Six articles are covered in the review of articles in AER/QJE/JPE.[4] Among the rest, seven have methodological-sounding names.[5] A close reading of the articles with methodological-sounding names reveals that they do not give a formal exposition of the triple difference estimator, nor its identifying assumption. However, Ravallion (2007) cites Ravallion et al. (2005) which shows a very special case of the triple difference estimator and the identifying assumptions for that special case. [6]

"[i]n this triple-difference model, the maintained assumption is that there are no differential trends for high participation versus low participation groups within early versus late implementing counties". Deschênes et al. (2017, p. 2970) state that "[o]ur identifying assumption is that such policies did not change differentially in NBP versus non-NBP states, in winter versus summer, over this period". Finally, Kleven et al. (2013, p. 1908) write that "[i]n that case, the identifying assumption would be that there is no contemporaneous change in the differential trend between Spain and the synthetic control country".

[3] These are the articles 4, 9, 17, 25, 31, 34, and 39, in which 4 is Gruber (1994) and 31 is Gruber and Poterba (1994). Note also that number 30 is Yelowitz (1995).

[4]These are the articles 7, 11, 21, 35, 42 and 46.

[5]These are the articles 1, 5, 6, 10, 12, 24, and 40. Note that number 24 is Lechner (2011) which is covered previously.

[6]This scenario does not have pre-periods, only post-periods, and two treatment groups that are treated with differential intensity. This requires a set of identifying assumptions that in general are not needed in the triple difference estimator.

3 The triple difference estimator

For the sake of exposition let us assume that we are talking about two American states, and that the Treatment state (T) introduces a health-care measure, while the Control state (C) does not. Further, the population of the states can be subdivided into two groups, group A and group B. The health-care measure we intend to study is only introduced to group B, i.e. group B is the group that can Benefit from the measure. Finally, there are two time periods, namely Pre- and Post-implementation of the health-care measure.

To establish a counterfactual it might seem convenient to compare group A and group B within the treatment state. This will not be valid if the health-care reform has within-state spillovers from group B to group A. Another option is to compare group B in the treatment state with group B in the control state. This will not be valid if different states have different economic conditions, so that group B in the treatment state would have trended differently from group B in the control state, regardless of the health-care measure. However, we may reasonably assume that the general economic differences will not affect the relative outcomes of group A and group B. In that case, we can use the relative difference to estimate what would have happened to the relative outcomes of group A and group B in the treatment state in the absence of treatment.

Equation 1 is a basic triple difference specification in accordance with the above exposition. All variables in this basic setup are dummy variables.

$$Y_{sit} = \beta_0 + \beta_1 T + \beta_2 B + \beta_3 Post + \beta_4 T * B + \beta_5 T * Post + \beta_6 B * Post + \beta_7 T * B * Post + \epsilon_{sit} \tag{1}$$

The conditional mean function of Equation 1 is $E[Y_{sit}|T, C, Post]$, which can take on eight values. Since the model has eight values and eight coefficients, the model is saturated (Angrist and Pischke, 2008). Under standard OLS assumptions and an additive effect, we can use $E[\epsilon_{sit}|T, C, Post] = 0$ to show the eight expected values as in Equations 2.

$$
\begin{aligned}
&E[Y|T=0, B=0, Post=0] = \beta_0 \\
&E[Y|T=1, B=0, Post=0] = \beta_0 + \beta_1 \\
&E[Y|T=0, B=1, Post=0] = \beta_0 + \beta_2 \\
&E[Y|T=0, B=0, Post=1] = \beta_0 + \beta_3 \\
&E[Y|T=1, B=1, Post=0] = \beta_0 + \beta_1 + \beta_2 + \beta_4 \\
&E[Y|T=1, B=0, Post=1] = \beta_0 + \beta_1 + \beta_3 + \beta_5 \\
&E[Y|T=0, B=1, Post=1] = \beta_0 + \beta_2 + \beta_3 + \beta_6 \\
&E[Y|T=1, B=1, Post=1] = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7
\end{aligned} \tag{2}
$$

Starting at the top of equation set 2, we can solve for the $\beta's$.

$$\beta_0 = E[Y|T = 0, B = 0, Post = 0]$$

$$\beta_1 = E[Y|T = 1, B = 0, Post = 0] - E[Y|T = 0, B = 0, Post = 0]$$

$$\beta_2 = E[Y|T = 0, B = 1, Post = 0] - E[Y|T = 0, B = 0, Post = 0]$$

$$\beta_3 = E[Y|T = 0, B = 0, Post = 1] - E[Y|T = 0, B = 0, Post = 0]$$

$$\beta_4 = E[Y|T = 1, B = 1, Post = 0] + E[Y|T = 0, B = 0, Post = 0]-$$

$$E[Y|T = 1, B = 0, Post = 0] - E[Y|T = 0, B = 1, Post = 0]$$

$$\beta_5 = E[Y|T = 1, B = 0, Post = 1] + E[Y|T = 0, B = 0, Post = 0]-$$

$$E[Y|T = 1, B = 0, Post = 0] - E[Y|T = 0, B = 0, Post = 1]$$

$$\beta_6 = E[Y|T = 0, B = 1, Post = 1] + E[Y|T = 0, B = 0, Post = 0]-$$

$$E[Y|T = 0, B = 1, Post = 0] - E[Y|T = 0, B = 0, Post = 1]$$

$$\beta_7 = \big(E[Y|T = 1, B = 1, Post = 1] - E[Y|T = 1, B = 1, Post = 0]\big)-$$

$$\big(E[Y|T = 1, B = 0, Post = 1] - E[Y|T = 1, B = 0, Post = 0]\big)-$$

$$\big(E[Y|T = 0, B = 1, Post = 1] - E[Y|T = 0, B = 1, Post = 0]\big)+$$

$$\big(E[Y|T = 0, B = 0, Post = 1] - E[Y|T = 0, B = 0, Post = 0]\big) \tag{3}$$

By rearranging the expression for β_7 and substituting the expected values with their sample equivalents (the mean values), we get Equation 4. This is the triple difference estimator for the effect of the treatment for group B.

$$\hat{\beta}_7 = [(\bar{Y}_{T,B,Post} - \bar{Y}_{T,B,Pre}) - (\bar{Y}_{C,B,Post} - \bar{Y}_{C,B,Pre})] - [(\bar{Y}_{T,A,Post} - \bar{Y}_{T,A,Pre}) - (\bar{Y}_{C,A,Post} - \bar{Y}_{C,A,Pre})] \tag{4}$$

4 The difference between two difference-in-differences

The classical difference-in-differences estimator is presented in Equation 5.

$$\hat{\delta} = [(\bar{Y}_{T,Post} - \bar{Y}_{T,Pre}) - (\bar{Y}_{C,Post} - \bar{Y}_{C,Pre})] \tag{5}$$

Clearly, the triple difference estimator of Equation 4 is equivalent to the difference between two difference-in-differences. The first difference-in-differences is for group B, and is given by the first square brackets, while the second difference-in-differences is for group A, given by the second square brackets. It is also worth mentioning that due to the additive nature of the triple difference estimator of Equation 4, we could alternatively have presented it as a difference-in-differences for the treatment state, comparing the eligible group B and group A, minus a difference-in-differences in the control state, comparing group B and group A there. Mathematically this is equivalent, though when thinking about a specific application one is often preferred over the other.

5 Identifying assumptions

The triple difference estimator requires a parallel trend assumption for the estimated effect to have a causal interpretation. Even though the triple difference is the difference between two difference-in-differences, it does not need two parallel trend assumptions. Rather, it requires the relative outcome of group B and group A in the treatment state to trend in the same way as the relative outcome of group B and group A in the control state, in the absence of treatment. To see this, first take the β_7 in Equations 3 and rearrange it to create Equation 6.

$$\beta_7 = \Bigg[\bigg(E[Y|T=1,B=1,Post=1] - E[Y|T=1,B=1,Post=0]\bigg) - \bigg(E[Y|T=1,B=0,Post=1] - E[Y|T=1,B=0,Post=0]\bigg)\Bigg] - \Bigg[\bigg(E[Y|T=0,B=1,Post=1] - E[Y|T=0,B=1,Post=0]\bigg) - \bigg(E[Y|T=0,B=0,Post=1] - E[Y|T=0,B=0,Post=0]\bigg)\Bigg] \tag{6}$$

Now, introduce the potential outcomes framework (see for instance Angrist and Pischke (2008)). In this framework $E[Y_{1,sit}]$ is the expected outcome of a state, group, and time if treated, while $E[Y_{0,sit}]$ is the expected outcome of a state, group, and time if not treated. Potential outcomes mean that we either observe $\overline{Y}_{1,sit}$ or $\overline{Y}_{0,sit}$, but never both. Expressions like $E[Y_{0,T=1,B=1,Post=1}]$ are the expectation of non-observed potential outcomes; in our case the outcome of group B in the treatment state (T), in the treatment period (Post), had it not been treated.

We can use the potential outcome framework to define δ, the true causal effect of treatment in the treatment state (T), on the treatment group B, in the treatment period (Post) as:

$$\delta = E[Y_1 - Y_0|T=1,B=1,Post=1] \tag{7}$$

Equation 7 states that the true treatment effect is the difference between the outcome of state T, group B in period 2 as treated, and the outcome of state T, group B in period 2, had it not been treated.

To show which parallel trend assumption that identifies δ, we may rewrite Equation 6 using the notation from the potential outcome framework.

$$
\begin{aligned}
\beta_7 = \Bigg[\bigg(&E[Y_1|T=1,B=1,Post=1] - E[Y_0|T=1,B=1,Post=0]\bigg) - \\
&\bigg(E[Y_0|T=1,B=0,Post=1] - E[Y_0|T=1,B=0,Post=0]\bigg)\Bigg] - \\
&\Bigg[\bigg(E[Y_0|T=0,B=1,Post=1] - E[Y_0|T=0,B=1,Post=0]\bigg) - \\
&\bigg(E[Y_0|T=0,B=0,Post=1] - E[Y_0|T=0,B=0,Post=0]\bigg)\Bigg]
\end{aligned} \tag{8}
$$

For β_7 to equal δ, we need the differential in the outcomes of group A and group B in the treatment state to trend similarly to the differential in the outcomes of group A and group B in the control state, in the absence of treatment. This is the parallel trend assumption. A formal exposition of this statement is given in Equation 9. The first line is the change between the two periods in the outcomes of group B in the treatment state had it not been treated. The second line is the same change for group A. The difference between these two expressions is equated with an expression that is equivalent, except that it gives realized outcomes in the control state.

$$
\begin{aligned}
&\bigg(E[Y_0|T=1,B=1,Post=1] - E[Y_0|T=1,B=1,Post=0]\bigg) - \\
&\bigg(E[Y_0|T=1,B=0,Post=1] - E[Y_0|T=1,B=0,Post=0]\bigg) \\
&= \\
&\bigg(E[Y_0|T=0,B=1,Post=1] - E[Y_0|T=0,B=1,Post=0]\bigg) - \\
&\bigg(E[Y_0|T=0,B=0,Post=1] - E[Y_0|T=0,B=0,Post=0]\bigg)
\end{aligned} \tag{9}
$$

To show that this parallel trend assumption identifies δ, the causal effect, we can substitute Equation 9 into Equation 8.

$$\beta_7 = \Bigg[\Bigg(E[Y_1|T=1,B=1,Post=1] - E[Y_0|T=1,B=1,Post=0]\Bigg) - \\ \Bigg(E[Y_0|T=1,B=0,Post=1] - E[Y_0|T=1,B=0,Post=0]\Bigg)\Bigg] - \\ \Bigg[\Bigg(E[Y_0|T=1,B=1,Post=1] - E[Y_0|T=1,B=1,Post=0]\Bigg) - \\ \Bigg(E[Y_0|T=1,B=0,Post=1] - E[Y_0|T=1,B=0,Post=0]\Bigg)\Bigg] \tag{10}$$

Rearranging and rewriting Equation 10 we get

$$\begin{aligned}\beta_7 = \quad & E[Y_1 - Y_0|T=1,B=1,Post=1] \\ & + E[Y_0|T=1,B=0,Post=1] - E[Y_0|T=1,B=0,Post=1] \\ & + E[Y_0|T=1,B=1,Post=0] - E[Y_0|T=1,B=1,Post=0] \\ & + E[Y_0|T=1,B=0,Post=0] - E[Y_0|T=1,B=0,Post=0]\end{aligned} \tag{11}$$

By canceling out the redundant terms of Equation 11 we find that

$$\beta_7 = (E[Y_1 - Y_0|T=1,B=1,Post=1] = \delta \quad qed. \tag{12}$$

6 Triple difference as difference-in-differences

Take the difference-in-differences estimator of Equation 5 and define the outcome variable, $\bar{Y}$, as:

$$\bar{Y}_{ij} = \bar{Y}_{aij} - \bar{Y}_{bij} \tag{13}$$

Substituting this definition into Equation 5 gives us

$$
\begin{aligned}
&\hat{\hat{\delta}} = \\
&[(\bar{Y}_{a,pre,treat} - \bar{Y}_{b,pre,treat}) - (\bar{Y}_{a,post,treat} - \bar{Y}_{b,post,treat})] - \\
&[(\bar{Y}_{a,pre,cont} - \bar{Y}_{b,pre,cont}) - (\bar{Y}_{a,post,cont} - \bar{Y}_{b,post,cont})] \\
&= \hat{\delta}_{triple} \qquad (14)
\end{aligned}
$$

This shows clearly that a basic difference-in-differences with a differential as the outcome and a symmetric structure, is a triple difference, and the other way around. This implies that all procedures for difference-in-differences can be applied to a transformed triple difference. For instance, standard robustness checks for difference-in-differences can be applied, see for instance Angrist and Pischke (2008). Also, semi-parametric versions of the difference-in-differences estimator are available (Abadie, 2005), as well as non-linear models (Athey and Imbens, 2006) can be directly applied to the transformed problem. Finally, knowing that difference-in-differences models struggle with standard errors when there are few clusters, see Bertrand et al. (2004), this will apply to the transformed triple difference, as well as to the triple difference estimator, though to a smaller extent due to more degrees of freedom.

7 How to name the estimator

Using T as shorthand for triple, D for difference, and s for plural form, the six most common ways of referencing the triple difference estimator are: TD (2911), TDs (1187), DDD (1104), DDDs (928), DsDsDs (351), and DDsDs (332). The numbers in parenthesis are the number of articles that use that particular way of referencing the estimator, equivalent to the cumulative sums from Figure 1. Note that the search string includes the word economics, and excludes all articles after 2017. The total number of articles is found through the same process as above, except that it uses an inclusive *OR* statement for the six ways of referencing the estimator, and yields 4813 unique papers. There are an additional four possible ways to combine *D* and *s*. These combinations, which we believe to be erroneous, have a total of 36 hits, and are excluded throughout our paper.

Going back to Figure 1, we see that most of the growth in the references to the estimator take place after 2010. There are two main ways of referencing the estimator, TD or DDD. Both come with variations in plural *s*. Of the different ways of referencing the model, TD is the most common and also seems to be the fastest growing. [7]

In Table 1 we show a frequency table of different ways to reference the estimator that occur together. Of the 2910 works that reference the triple difference estimator as TD, 256 also reference it as DDD in the same paper and 1713 rely solely on TD[8]. The results in Table 1 strongly suggest that there is a need to unify the terminology. Without taking a strong stand on what is the most logical name, we recommend triple difference (TP) or difference-in-difference-in-differences (DDDs). [9]

8 Concluding remarks

In this paper we document the rise of the triple difference estimator. The use of the estimator has grown exponentially, yet it lacks formal derivation and is often carelessly applied in the literature, for instance by largely ignoring its parallel trend assumption, and by omitting unconditional plots, making model validation difficult. We also document a need to unify the terminology and suggest 'triple difference' or 'difference-in-difference-in-differences'.

[7]While DDD has historically been the most common way of referencing the model when avoiding the word *triple*, this has reversed for the last two years, and in 2017 DDDs was referenced 188 times, while DDD was only referenced 143 times.

[8]If we look at occurrences of TD with any plural variation of DDD, only 604 out of 2910 papers also use a DDD variation, which is only about 20 percent. This is confirmed by looking at any of the DDD variations and the co-occurrence of TD and TDs as well, meaning that the majority of papers that reference the estimator rely on only one of the two main ways of referencing it.

[9]Gruber (1994), the father of the triple difference estimator, used the terminology *differences in differences in differences* or DsDsDs. This way of referencing the estimator has only 360 hits throughout time, and only 41 hits in 2017. This suggests that the 1152 citations to his paper are not primarily methodological. If they were, we would expect his choice of terminology to be more common.

Table 1: Combinations of triple difference referencing

	TD	TDs	DDD	DDDs	DsDsDs	DDsDs
TD	2910	593	256	205	69	74
TDs	593	1130	62	105	40	49
DDD	256	62	1080	97	31	23
DDDs	205	105	97	915	19	41
DsDsDs	69	40	31	19	352	29
DDsDs	74	49	23	41	29	334

Note: T denotes triple, D denotes difference, and s denotes a plural s. All searches are from Google Scholar and require the result to contain the word economics and to be from the period between 1994-2017. Google Scholar treats spaces and hyphens as the same. Note also that all numbers are upper bounds, as a single paper might use more than two ways of referencing the estimator.

Our main contribution is to show that the triple difference estimator does not require two parallel trend assumptions to have a causal interpretation, even though it can be computed as the difference between two difference-in-differences estimators. We also show that the triple difference parallel trend assumption is equivalent to the parallel trend assumption in a difference-in-differences model based on ratios.

When choosing between a triple difference and a difference-in-differences on a ratio-variable, there are several things to consider. The difference-in-differences estimator is much better understood, and there is a large literature that addresses the estimator and its shortcomings. However, it comes at the cost of degrees of freedom, and provides less information than the triple difference. The triple difference will for instance provide an estimate of spillover-effects i.e. β_5 in Equation 1, which is the effect on the non-treated in the treatment state in the treatment period. This information is lost in the difference-in-differences estimator.

The triple difference estimator is often used as a heterogeneity test or as a robustness check. When comparing it with a standard difference-in-differences, Berck and Villas-Boas (2016) show conditions for when the triple difference estimator reduces bias relative to a difference-in-differences approach in the presence of omitted variable bias.

Finally, our reading of the literature points to some other key issues that demand more awareness. Many of the articles examined spend considerable time on control variables, which will not affect unbiasedness, only precision. This is easily shown by deducting any mean from the estimator. Such means will cancel out, a point previously made for difference-in-differences by Angrist and Pischke (2008, p.237). Much less time, if any, is spent on functional form. In the triple difference estimator we make an assumption on how the outcomes of two groups co-move relative to the co-movement in two other groups in the control state. Both a ratio and its log-transformed counterpart can be a natural choice of functional form, depending on the situation. This requires thought, however. Particularly since if the parallel trend assumption holds in logs it will not hold in levels, and vice versa (Angrist and Pischke, 2008).

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Berck, P. and Villas-Boas, S. B. (2016). A note on the triple difference in economic models. *Applied Economics Letters*, 23(4):239–242.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Deschênes, O., Greenstone, M., and Shapiro, J. S. (2017). Defensive investments and the demand for air quality: Evidence from the nox budget program. *American Economic Review*, 107(10):2958–89.

Gruber, J. (1994). The incidence of mandated maternity benefits. *American Economic Review*, 84(3):622–641.

Gruber, J. and Poterba, J. (1994). Tax incentives and the decision to purchase health insurance: Evidence from the self-employed. *Quarterly Journal of Economics*, 109(3):701–733.

Hornbeck, R. (2010). Barbed wire: Property rights and agricultural development. *Quarterly Journal of Economics*, 125(2):767–810.

Hoynes, H., Schanzenbach, D. W., and Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review*, 106(4):903–34.

Imbens, G. W. and Wooldridge, J. M. (2007). What's new in econometrics? lecture 10 difference-in-differences estimation. *NBER Summer Institute, available at: www. nber. org/WNE/Slides7–31–07/slides_10_diffindiffs. pdf, accessed April 2018*, 9:2011.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the european football market. *American Economic Review*, 103(5):1892–1924.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, 4(3):165–224.

Muehlenbachs, L., Spiller, E., and Timmins, C. (2015). The housing market impacts of shale gas development. *American Economic Review*, 105(12):3633–59.

Nilsson, J. P. (2017). Alcohol availability, prenatal conditions, and long-term economic outcomes. *Journal of Political Economy*, 125(4):1149–1207.

Olden, A. (2018). What do you buy when no one's watching? the effect of self-service checkouts on the composition of sales in retail. NHH FOR DP 3/18, Norwegian School of Economics.

Ravallion, M. (2007). Evaluating anti-poverty programs. *Handbook of Development Economics*, 4:3787–3846.

Ravallion, M., Galasso, E., Lazo, T., and Philipp, E. (2005). What can ex-participants reveal about a program's impact? *Journal of Human Resources*, 40(1):208–230.

Shayo, M. and Zussman, A. (2011). Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics*, 126(3):1447–1484.

Walker, W. R. (2013). The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce. *Quarterly Journal of Economics*, 128(4):1787–1835.

Yelowitz, A. S. (1995). The medicaid notch, labor supply, and welfare participation: Evidence from eligibility expansions. *The Quarterly Journal of Economics*, 110(4):909–939.

A Appendix

Table A1: Title abbreviations for Tables A2-A3b

Abbreviaton	Full title
AEJAE	American Economic Journal: Applied Economics
AER	The American Economic Review
ARS	Annual Review of Sociology
EE	Energy Economics
FTE	Foundations and Trends® in Econometrics
HDE	Handbook of Development Economics
HE	Health Economics
HEF	Handbook of the Economics of Finance
HHE	Handbook of HE
HLE	Handbook of Labor Economics
ISR	Information Systems Research
JDE	Journal of Development Economics
JFE	Journal of Financial Economics
JLaE	Journal of Law and Economics
JLE	Journal of Labor Economics
JMR	Journal of Marketing Research
JPE	Journal of Political Economy
JPuE	Journal of Public Economics
JUE	Journal of Urban Economics
MS	Management Science
NBER	NBER Working Paper Series
NEJM	New England Journal of Medicine
NTJ	National Tax Journal
QJE	Quarterly Journal of Economics
RES	Review of Economics and Statistics
RFS	The Review of Financial Studies
SJ	Stata Journal
TEJ	The Economic Journal

Table A2: Use of triple difference estimation in *AER*, *JPE* and *QJE* from 2010-2017

Cit.	Authors	Title		
829	Mian, Sufi	House prices, home equity-based borrowing, and the US household leverage crisis	2011	AER
103	Moser, Voena	Compulsory licensing: Evidence from the trading with the enemy act	2012	AER
293	Hornbeck	The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe	2012	AER
146	Simcoe	Standard setting committees: Consensus governance for shared technology platforms	2012	AER
243	Kleven, Landais, Saez	Taxation and international migration of superstars: Evidence from the European football market	2013	AER
320	Busso, Gregory, Kline	Assessing the incidence and efficiency of a prominent place based policy	2013	AER
57	Aaronson, Lange, Mazumder	Fertility transitions along the extensive and intensive margins	2014	AER
129	Yagan	Capital tax reform and the real economy: The effects of the 2003 dividend tax cut	2015	AER
90	Casey	Crossing party lines: The effects of information on redistributive politics	2015	AER
212	Muehlenbachs, Spiller, Timmins	The housing market impacts of shale gas development	2015	AER
291	Hoynes, Schanzenbach, Almond	Long-run impacts of childhood access to the safety net	2016	AER
440	Pierce, Schott	The surprisingly swift decline of US manufacturing employment	2016	AER
37	Duggan, Garthwaite, Goyal	The market impacts of pharmaceutical product patents in developing countries: Evidence from India	2016	AER
65	Egan, Hortaçsu, Matvos	Deposit competition and financial fragility: Evidence from the us banking sector	2017	AER
30	Deschênes, Greenstone, Shapiro	Defensive investments and the demand for air quality: Evidence from the NOx budget program	2017	AER
122	Besley, Folke, Persson, Rickne	Gender quotas and the crisis of the mediocre man: Theory and evidence from Sweden	2017	AER
79	Aaronson, Mazumder	The impact of Rosenwald schools on black achievement	2011	JPE
50	Autor, Palmer, Pathak	Housing market spillovers: Evidence from the end of rent control in Cambridge, Massachusetts	2014	JPE
163	Carneiro, Løken, Salvanes	A flying start? Maternity leave benefits and long-run outcomes of children	2015	JPE
37	Casas-Arce, Saiz	Women and power: unpopular, unwilling, or held back?	2015	JPE
47	Nilsson	Alcohol availability, prenatal conditions, and long-term economic outcomes	2017	JPE
143	Hornbeck	Barbed wire: Property rights and agricultural development	2010	QJE
179	Shayo, Zussman	Judicial ingroup bias in the shadow of terrorism	2011	QJE
772	Ahern, Dittmar	The changing of the boards: The impact on firm valuation of mandated female board representation	2012	QJE
73	Cascio, Washington	Valuing the vote: The redistribution of voting rights and state funds following the voting rights act of 1965	2013	QJE
150	Walker	The transitional costs of sectoral reallocation: Evidence from the clean air act and the workforce	2013	QJE
155	Garthwaite, Gross, Notowidigdo	Public health insurance, labor supply, and employment lock	2014	QJE
52	Casaburi, Troiano	Ghost-house busters: The electoral response to a large anti–tax evasion program	2015	QJE
16	Agan, Starr	Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment	2017	QJE
25	Alsan, Wanamaker	Tuskegee and the health of black men	2017	QJE
44	Bandiera, Burgess, Das, Gulesci, Rasul, Sulaiman	Labor markets and poverty in village economies	2017	QJE
20	Larcom, Rauch, Willems	The benefits of forced experimentation: striking evidence from the London underground network	2017	QJE

Table A3a: Top 50 most cited articles referencing triple difference

	Cites	Authors	Title	Year	Source
1	7550	M Bertrand, E Duflo, S Mullainathan	How much should we trust differences-in-differences estimates?	2004	QJE
2	1418	EA Verhoogen	Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector	2008	QJE
3	1306	J Currie, D Almond	Human capital development before age five	2011	HLE
4	1177	J Gruber	The incidence of mandated maternity benefits	1994	AER
5	989	MR Roberts, TM Whited	Endogeneity in empirical corporate finance1	2013	HEF
6	943	C Winship, SL Morgan	The estimation of causal effects from observational data	1999	ARS
7	824	A Mian, A Sufi	House prices, home equity-based borrowing, and the US household leverage crisis	2011	AER
8	809	CJ Ruhm	The economic consequences of parental leave mandates: Lessons from Europe	1998	QJE
9	807	J Currie, J Gruber	Health insurance eligibility, utilization of medical care, and child health	1996	QJE
10	774	M Ravallion	Evaluating anti-poverty programs	2007	HDE
11	763	KR Ahern, AK Dittmar	The changing of the boards: The impact on firm valuation of mandated female board representation	2012	QJE
12	697	T Besley, A Case	Unnatural experiments? Estimating the incidence of endogenous policies	2000	TEJ
13	690	X Giroud, HM Mueller	Does corporate governance matter in competitive industries?	2010	JFE
14	659	G Zervas, D Proserpio, JW Byers	The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry	2017	JMR
15	648	S Dynarski	Hope for whom? Financial aid for the middle class and its impact on college attendance	2000	NTJ
16	552	DL Costa, ME Kahn	Power couples: changes in the locational choice of the college educated, 1940–1990	2000	QJE
17	526	J Gruber	The incidence of payroll taxation: Evidence from Chile	1997	JLE
18	512	A Purnanandam	Originate-to-distribute model and the subprime mortgage crisis	2010	RFS
19	505	A Low	Managerial risk-taking behavior and equity-based compensation	2009	JFE
20	500	M Puri, J Rocholl, S Steffen	Global retail lending in the aftermath of the US financial crisis: Distinguishing between supply and demand effects	2011	JFE
21	436	JR Pierce, PK Schott	The surprisingly swift decline of US manufacturing employment	2016	AER
22	388	LF Katz	Wage subsidies for the disadvantaged	1996	NBER
23	387	BD Sommers, K Baicker, AM Epstein	Mortality and access to care among adults after state Medicaid expansions	2012	NEJM
24	384	M Lechner	The estimation of causal effects by difference-in-difference methods	2011	FTE
25	377	J Gruber	Disability insurance benefits and labor supply	2000	JPE
26	359	A Goldfarb, CE Tucker	Privacy regulation and online advertising	2011	MS

Table A3b: Top 50 most cited articles referencing triple difference, continued

	Cites	Authors	Title	Year	Source
27	354	J Strauss, D Thomas	Health over the life course	2007	HDE
28	353	DA Matsa, AR Miller	A female style in corporate leadership? Evidence from quotas	2013	AEJAE
29	350	A Seru	Firm boundaries matter: Evidence from conglomerates and R&D activity	2014	JFE
30	343	AS Yelowitz	The Medicaid notch, labor supply, and welfare participation: Evidence from eligibility expansions	1995	QJE
31	333	J Gruber, J Poterba	Tax incentives and the decision to purchase health insurance: Evidence from the self-employed	1994	QJE
32	332	K Milligan	Subsidizing the stork: New evidence on tax incentives and fertility	2005	RES
33	330	J Currie	Inequality at birth: Some causes and consequences	2011	AER
34	328	J Gruber, BC Madrian	Health insurance, labor supply, and job mobility: A critical review of the literature	2002	NBER
35	319	M Busso, J Gregory, P Kline	Assessing the incidence and efficiency of a prominent place based policy	2013	AER
36	318	K Eggleston, L Ling, M Qingyue, M Lindelow, A Wagstaff	Health service delivery in China: A literature review	2008	HE
37	314	D Neumark, J Zhang, S Ciccarella	The effects of Wal-Mart on local labor markets	2008	JUE
38	311	DN Figlio	Testing, crime and punishment	2006	JPuE
39	309	J Gruber	Health insurance and the labor market	2000	HHE
40	296	A Nichols	Causal inference with observational data	2007	SJ
41	291	RT Jensen	Do private transfers 'displace'the benefits of public transfers? Evidence from South Africa	2004	JPuE
42	290	R Hornbeck	The enduring impact of the American Dust Bowl: Short-and long-run adjustments to environmental catastrophe	2012	AER
43	287	D Thomas, K Beegle, E Frankenberg, B Sikoki, J Strauss, G Teruel	Education in a Crisis	2004	JDE
44	286	R Rishika, A Kumar, R Janakiraman, R Bezawada	The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation	2013	ISR
45	282	C Clotfelter, E Glennie, H Ladd, J Vigdor	Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina	2008	JPuE
46	281	H Hoynes, DW Schanzenbach, D Almond	Long-run impacts of childhood access to the safety net	2016	AER
47	277	A Morse	Payday lenders: Heroes or villains?	2011	JFE
48	277	H Cai, Y Chen, H Fang	Observational learning: Evidence from a randomized natural field experiment	2009	AER
49	273	VV Acharya, RP Baghai, KV Subramanian	Labor laws and innovation	2013	JLaE
50	267	JG Weber	The effects of a natural gas boom on employment and income in Colorado, Texas, and Wyoming	2012	EE

This table is produced using the software Harzinger's Publish or Perish 6. A search using each of the six most common ways to reference the triple difference estimator is conducted from 1994 until October 2018, covering almost all results for the triple difference estimator. Each search is combined with the word economics. When removing books and duplicates, this yields 3481 articles. The articles are sorted according to the number of citations, and the top 50 most cited articles are presented here. Full journal titles are found in Table A1.

NHH

NORGES HANDELSHØYSKOLE
Norwegian School of Economics

Helleveien 30
NO-5045 Bergen
Norway

T +47 55 95 90 00
E nhh.postmottak@nhh.no
W www.nhh.no