Identification based on Difference-in-Differences Approaches with Multiple Treatments

AUTHORS

Hans Fricke

Stanford University

ABSTRACT

This paper discusses identification based on difference-in-differences (DiD) approaches with multiple treatments. It shows that an appropriate adaptation of the common trend assumption underlying the DiD strategy for the comparison of two treatments restricts the possibility of effect heterogeneity for at least one of the treatments. The required assumption of effect homogeneity is likely to be violated because of non-random assignment to treatment based on both observables and unobservables. However, this paper shows that, under certain conditions, the DiD estimate comparing two treatments identifies a lower bound in absolute values on the average treatment effect on the treated compared to the unobserved non-treatment state, even if effect homogeneity is violated. This is possible if, in expectation, the effects of both treatments compared to no treatment have the same sign, and one treatment has a stronger effect than the other treatment on the respective recipients. Such assumptions are plausible if treatments are ordered or vary in intensity.

Identification based on Difference-in-Differences Approaches with Multiple Treatments

Hans Fricke*

May 4, 2015

Abstract This paper discusses identification based on difference-in-differences (DiD) approaches with multiple treatments. It shows that an appropriate adaptation of the common trend assumption underlying the DiD strategy for the comparison of two treatments restricts the possibility of effect heterogeneity for at least one of the treatments. The required assumption of effect homogeneity is likely to be violated because of non-random assignment to treatment based on both observables and unobservables. However, this paper shows that, under certain conditions, the DiD estimate comparing two treatments identifies a lower bound in absolute values on the average treatment effect on the treated compared to the unobserved non-treatment state, even if effect homogeneity is violated. This is possible if, in expectation, the effects of both treatments compared to no treatment have the same sign, and one treatment has a stronger effect than the other treatment on the respective recipients. Such assumptions are plausible if treatments are ordered or vary in intensity.

---

*Stanford University & University of St.Gallen, hfricke@stanford.edu

1 Introduction

In policy evaluation, the main challenge is to identify what would have happened to the treated group that experienced the policy in the hypothetical situation absent the policy. Difference-in-Differences (DiD) strategies use the development of a control group which was unaffected by the policy to identify this situation. Identification requires that both groups would have developed equally over time in absence of the policy. In other words, the effect of time (other time varying factors) on the outcome of interest is equal for both groups. This assumption is known as the common trend (CT) assumption.

In case of multiple treatments, the effect of the different treatments in comparison to no treatment can be identified precisely under this standard CT assumption (Frölich, 2004). The comparison of different treatments with DiD has not received much theoretical attention. Frölich (2004) simply argues that DiD estimation is not helpful to compare two treatments, as the pre-treament period is only informative about the non-treatment state, but he does not explore the issue any further.

Despite this recommendation, DiD is used in many applied studies to identify the causal difference between two treatments. For example, Duflo (2001) analyzes the effect of school construction on schooling and labor market outcomes comparing regions with high and low levels of newly constructed schools in Indonesia. Similarly, Havnes and Mogstad (2011b) and Havnes and Mogstad (2011a) use regional variation in child care expansion rates to estimate the effect of formal child care on child development and maternal employment. Moreover, Alatas and Cameron (2008) explore labor market effects of differences in minimum wage increases between two Indonesian regions. In all these studies, the development of a group that is less affected by a policy is used to estimate the development of a more affected group.

This study argues that DiD estimation can be used to compare two treatments as in the studies above. To that end, it first works out the rather restrictive assumptions needed for the conventional interpretation of the DiD estimand - the average treatment effect on the treated (ATET) of one treatment compared to another treatment. Second, it proposes an alternative interpretation of the DiD estimand - the lower bound in magnitude of an ATET in comparison to the non-treatment state - which relies on more realistic assumptions.

The interpretation of the ATET of one treatment compared to another treatment relies on a common trend assumption which requires both treatment groups to develop equally over time under at least one of the treatments. If the effect of time is the same for both groups, which is generally assumed in DiD approaches, this assumption implies homogenous treatment responses to at least one treatment across the two treatment groups. For instance, if interest is in the ATET for treatment 1 compared to treatment 2, those who received treatment 1 have to respond to treatment 2 in the same way as those who received treatment 2. If the common trend assumption is imposed unconditionally, this requires effect homogeneity of treatment 2. If instead, the common trend assumption is imposed conditionally on some observed variables, this implies effect homogeneity with respect to unobservables.

So far, the implication of effect homogeneity has not been recognized. For instance, Duflo (2001), Havnes and Mogstad (2011b), and Havnes and Mogstad (2011a) explicitly state, as main identifying assumption, merely that the respective groups would have to react equally in absence of the treatment, which is not sufficient for effect homogeneity as argued above. In fact, the homogeneity assumption seems too restrictive because of non-random treatment assignment in most studies.

Consider for example Duflo (2001). New schools were built depending on the number of children not enrolled in primary school. In regions with a high number of enrolled children, other constraints might be more important than the availability of schools for those children still not enrolled. Such constraints might be distance to school or family obligations. Hence, one would expect regions with a low level of new schools to react less to an additional school than regions with a high level of new schools. If that is the case, Duflo (2001) over estimates the ATET of the high level of new schools compared to the low level of new schools.

In Alatas and Cameron (2008) the difference in the minimum wage arises due to a convergence to equal post-treatment levels. That is, the region that experiences the more pronounced increase starts with a substantially lower minimum wage. Possibly, a low initial minimum wage is binding in fewer sectors than a high initial minimum wage. In that case, a marginal increase in the minimum wage would affect the intensive as well as the extensive margin of businesses in the region with the low initial wage. In the region with the high initial wage, the increase would rather affect the intensive than the extensive margin. Consequently, the region that experienced the high increase in the minimum wage would have reacted more strongly to the low increase than the region that experienced the low increase.

This paper proposes an alternative interpretation of the DiD estimand if effect homogeneity is implausible. Instead of the difference between two treatments, the DiD estimand can be interpreted as the lower bound in magnitude of an ATET in comparison to the non-treatment state. This is possible if in expectation the direction of both treatment effects in comparison to no treatment is the same, i.e. either both are negative or both are positive, but one treatment has a stronger effect than the other. For instance, if treatment 1 affects those who receive treatment 1 more strongly than treatment 2 affects those who receive treatment 2, a lower bound of the ATET of treatment 1 compared to no treatment can be identified. Furthermore, this interpretation can be valid in some instances if the CT assumption about the non-treatment state is violated.

These assumptions are credible if the treatments can be naturally ordered or if treatments vary in intensity. Thus, in all of the above examples, such assumptions seem more credible than effect homogeneity. For instance, in Duflo (2001) an additional school should increase the share of enrolled students. Moreover, it is likely that many schools have a larger impact than a few schools. Similarly, in standard economic theory, a minimum wage increase should reduce employment[1] and more likely, a larger increase will have a stronger bite than a smaller increase

---

[1]Of course, this might not be true in the presence of monopsonies.

in Alatas and Cameron (2008).

2 Comparison of one treatment state to the non-treatment state

Let Y_t^d be the potential outcome in period $t \in \{1, 0\}$, where $d \in \{H, L, 0\}$ is a treatment indicator. $D = H$ and $D = L$ refer to two different treatments and $D = 0$ to the non-treatment state. In this section, treatments $D = H$ and $D = L$ can correspond to any set of distinct treatments. In Section 3.2, the assumptions are more credible and the results are easier to interpret if treatments H and L correspond to a high and a low intensity of one treatment, respectively, or to ordered treatments. $t = 1$ indicates the post-treatment and $t = 0$ the pre-treatment period.

First, consider the ATET of each treatment separately compared to no treatment as discussed in Frölich (2004). I exemplify the results for treatment H but they readily generalize to all treatments (including treatment L) even in applications with more than two treatments:

$$ATET_1^{H\,0} = E[Y_1^H | D = H] - E[Y_1^0 | D = H]. \tag{1}$$

The main challenge to identification is that the counterfactual situation $E[Y_1^0 | D = H]$ is never observed. The DiD approach solves this problem by assuming parallel trends of the counterfactual situation. To identify (1), the CT is the same as in the standard single treatment case:

$$E[Y_1^0 | D = H] - E[Y_0^0 | D = H] = E[Y_1^0 | D = 0] - E[Y_0^0 | D = 0]$$
$$\Leftrightarrow \underbrace{E[Y_1^0 | D = H]}_{counterfactual} = E[Y_1^0 | D = 0] - E[Y_0^0 | D = 0] + E[Y_0^0 | D = H]. \tag{2}$$

It requires that the expected potential outcome under non-treatment of both groups develops equally over time. In other words, the effect of time needs to be the same for the group $D = H$ and $D = 0$. Under this CT assumption, the counterfactual can be expressed in three moments that can be identified from the data.[2] By plugging in (2) into (1), we can see that the $ATET_1^{H\,0}$ is point identified:

$$ATET_1^{H\,0} = E[Y_1^H | D = H] - E[Y_0^0 | D = H] - \left[E[Y_1^0 | D = 0] - E[Y_0^0 | D = 0] \right]$$
$$= \underbrace{E[Y_1 | D = H]}_{identified} - \underbrace{E[Y_0 | D = H]}_{identified} - \left[\underbrace{E[Y_1 | D = 0]}_{identified} - \underbrace{E[Y_0 | D = 0]}_{identified} \right]$$

---

[2]I implicitly impose two standard assumptions: first, the stable unit treatment value assumption and second, that the treatment has no effect in the pre-treatment period. See Lechner (2011).

In some applications, the CT assumption holds only conditional on trend confounding factors. See Frölich (2004) for identification with DiD approaches conditional on covariates.

3 Comparison of two treatments states

3.1 Point identification

In many applications, all individuals receive a treatment, i.e. treatment 0 is not observed. In these cases, DiD strategies are used to compare two treatments. Consider the average treatment effect of H compared to L for those who receive the treatment H. The results readily generalize to all pairwise treatment comparisons in applications with more than two treatments. $ATET_1^{HL}$ can be expressed as

$$ATET_1^{HL} = E[Y_1^H|D=H] - E[Y_1^L|D=H]. \tag{3}$$

The corresponding CT is

$$E[Y_1^L|D=H] - E[Y_0^0|D=H] = E[Y_1^L|D=L] - E[Y_0^0|D=L], \tag{4}$$
$$\Leftrightarrow \underbrace{E[Y_1^L|D=H]}_{counterfactual} = E[Y_1^L|D=L] - E[Y_0^0|D=L] + E[Y_0^0|D=H].$$

Assumption (4) states that both groups would have developed equally over time under treatment L.[3] To understand the implications of this assumption in comparison to the standard CT (2), it is helpful to separate (4) into two components: CT in the non-treatment state and effect homogeneity. By expanding both sides of the equation, (4) can be expressed as

$$\left[E\left[Y_1^0|D=H\right] - E[Y_0^0|D=H]\right] + \left[E\left[Y_1^L|D=H\right] - E[Y_1^0|D=H]\right]$$
$$= \left[E\left[Y_1^0|D=L\right] - E[Y_0^0|D=L]\right] + \left[E\left[Y_1^L|D=L\right] - E[Y_1^0|D=L]\right].$$

The first differences on each side of the equation correspond to the CT in the non-treatment state. They represent the development of the expected potential outcome under non-treatment, i.e. the effect of time, for groups $D=H$ and $D=L$. The second differences correspond to effect homogeneity. They represent the effect of treatment L in comparison to no treatment for each group post-treatment. Note, that neither component alone is binding. However, if the CT of the counterfactual non-treatment state is true, which is generally assumed, effect homogeneity is necessary for identification and vice versa. Heterogeneous effects would thus be allowed only to the extent that they are exactly compensated by a deviation from the CT under non-treatment. If one were to impose (4) conditional on covariates, effect homogeneity has to hold only with respect to unobservables. As exemplified in Section 1, homogeneity seems

---

[3]The proof of identification is analogous to the one in Section 2.

too restrictive even conditionally in many applications because the treatment groups are not randomly selected but rather based on observable and unobservable characteristics.[4]

DiD applications commonly argue the plausibility of the common trend by testing for differential trends in pre-treatment periods. These placebo tests estimate treatment effects in periods without a treatment. Hence, they are only informative about the common trend in the non-treatment state and they cannot address concerns about effect heterogeneity. Therefore, in case of multiple treatments, arguing the validity of the required common trend assumption (4) is not as straight forward as in the standard case.

3.2 Partial identification

This section shows that the DiD estimand can be interpreted as the lower bound in magnitude of an ATET compared to no treatment. For ease of interpretation, now treatments H and L correspond to a high and a low intensity of one treatment, respectively. Consider again the average treatment effect of treatment H compared to treatment 0:

$$ATET_1^{H\,0} = E[Y_1^H|D = H] - E[Y_1^0|D = H]. \tag{5}$$

First, assume that the expected potential outcomes under non-treatment of groups $D = H$ and $D = L$ develop equally over time:

$$E[Y_1^0|D = H] - E[Y_0^0|D = H] = E[Y_1^0|D = L] - E[Y_0^0|D = L]$$
$$\Leftrightarrow \underbrace{E[Y_1^0|D = H]}_{counterfactual} = \underbrace{E[Y_1^0|D = L]}_{counterfactual} - E[Y_0^0|D = L] + E[Y_0^0|D = H]. \tag{6}$$

This assumption is stronger than necessary for the partial identification result and will be relaxed at the end of this section. Even under effect homogeneity, this CT assumption cannot point identify (5) because it includes two counterfactual situations. The potential outcome without treatment in period $t = 1$ is not observed for either group.[5]

However, the lower bound in absolute values can be identified under an assumption about the relation between the counterfactual $E[Y_1^0|D = L]$ and the observed moment $E[Y_1^L|D = L]$. The following assumption depends on the sign of the DiD estimand, i.e. the observed difference between the development of groups $D = H$ and $D = L$.

---

[4]Duflo (2001) controls for pre-treatment levels of schools in one specification, which are arguably an important dimension of effect heterogeneity. However, the level of schools also perfectly determines the regions' treatment status. Hence, common support is violated by construction.

[5]Only if the effect of L is zero for both groups compared to no treatment, the $ATET_1^{H\,0}$ can be identified. Then, $ATET_1^{H\,0} = ATET_1^{H\,L}$.

Second, assume that either

$$E[Y_1^0|D=L] \leq E[Y_1^L|D=L] \tag{7}$$
$$\text{if} \quad E[Y_1^H|D=H] - E[Y_0^0|D=H] > E[Y_1^L|D=L] - E[Y_0^0|D=L]$$

or

$$E[Y_1^0|D=L] \geq E[Y_1^L|D=L] \tag{8}$$
$$\text{if} \quad E[Y_1^H|D=H] - E[Y_0^0|D=H] < E[Y_1^L|D=L] - E[Y_0^0|D=L].$$

Together with the CT in (6), these assumptions imply that, in expectation, both treatments affect the respective groups in the same direction compared to no treatment, and the effect of treatment H is stronger on group $D=H$ than the effect of treatment L on group $D=L$. These assumptions are credible if the treatments can be naturally ordered or if treatments vary in intensity. For example in Duflo (2001), a low level and a high level of new schools have to impact schooling either both positively or both negatively, but the effect size of the low level must be smaller than the effect size of the high level for the respective groups. It is reasonable that an additional school will increase the share of enrolled students. Moreover, it is likely that many schools have a larger impact than a few schools. The same argumentation goes for the other aforementioned studies as they compare differences in the share of child care facilities, or in minimum wage changes.

Identification is best explained graphically. Figure 1 depicts the case of the positve DiD estimand and Figure 2 the case of the negative DiD estimand. Both figures contain a graph for a positive and a negative underlying common trend. The black and gray solid lines represent the observed expected outcomes of groups $D=H$ and $D=L$. The black and gray dashed lines correspond to the expected potential outcomes without a treatment. The black dotted line depicts the development of the expected observed outcome of group $D=L$ from the initial expected level of the observed outcome of group $D=H$. DiD always estimates the difference of the observed development between groups $D=H$ and $D=L$, i.e. the difference between the solid and dotted black lines in the post-treatment period.

In the first case, this difference is positive (see Figure 1). That is, the expected outcome of interest of group $D=H$ either increases more (left graph) or decreases less (right graph) over time than the outcome of group $D=L$. Hence, assumption (7) states that in $t=1$ the expected potential outcome without treatment is below the expected potential outcome under treatment L. Assumption (6) states that the expected potential outcomes in the non-treatment state of both groups develop equally, i.e. that both dashed lines run parallel. This assumption ensures that the expected potential outcomes under non-treatment of group $D=H$ will be smaller than the observed development of group $D=L$ from the initial level of group $D=H$, the dotted black line. Therefore, the $ATET_1^{H0}$, which is the difference between the solid and dashed black lines, will always be greater than or equal to the DiD estimand.
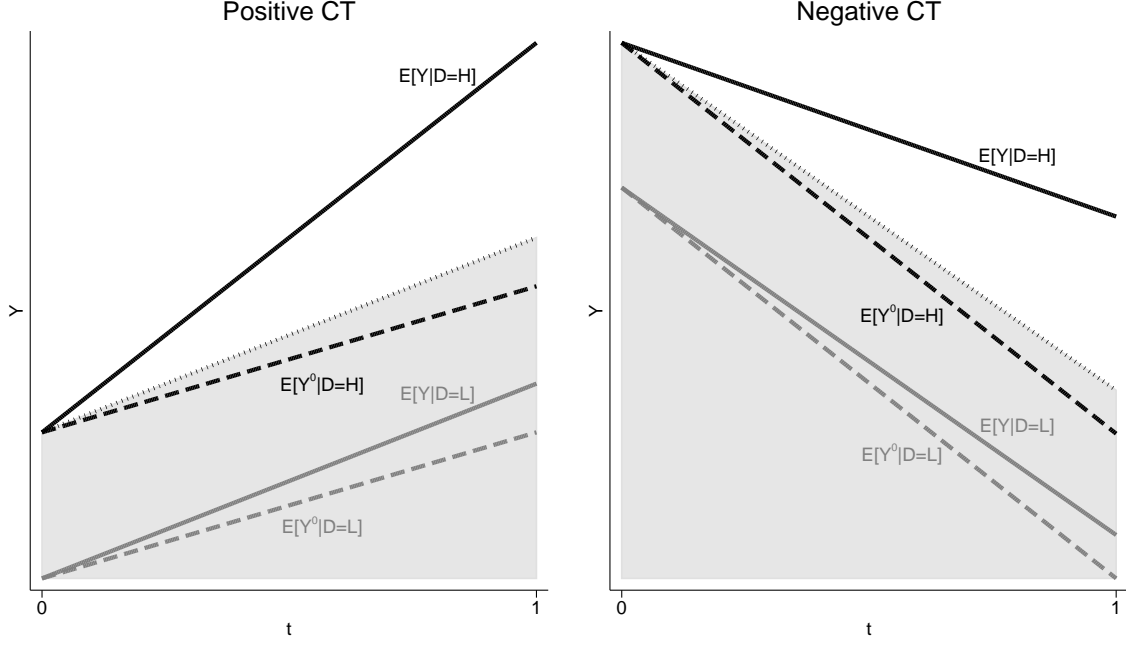
Figure 1: Positive DiD Estimand, Assumption (7)

Figure 2 depicts the second case. The DiD estimand is negative. That is, the outcome of group $D = H$ either decreases more (left graph) or increases less (right graph) than that of the group $D = L$. (8) states the expected potential outcome in absence of the treatment is above the expected outcome under treatment L. Again, this holds for both groups because of the CT assumption without treatment. Hence, the $ATET_1^{H0}$ will always be smaller than or equal to the DiD estimand.

To show formally that these assumptions allow us to interpret the DiD estimand as the lower bound of the $ATET_1^{H0}$ in absolute values, we start by plugging in (7) in (6):

$$E[Y_1^0|D = H] \leq E[Y_1^L|D = L] - E[Y_0^0|D = L] + E[Y_0^0|D = H]. \tag{9}$$

Then, if we plug (9) into (5), we get

$$ATET_1^{H0} \geq E[Y_1^H|D = H] - E[Y_0^0|D = H] - \left[E[Y_1^L|D = L] - E[Y_0^0|D = L]\right]$$

$$= \underbrace{E[Y_1|D = H]}_{identified} - \underbrace{E[Y_0|D = H]}_{identified} - \left[\underbrace{E[Y_1|D = L]}_{identified} - \underbrace{E[Y_0|D = L]}_{identified}\right]$$
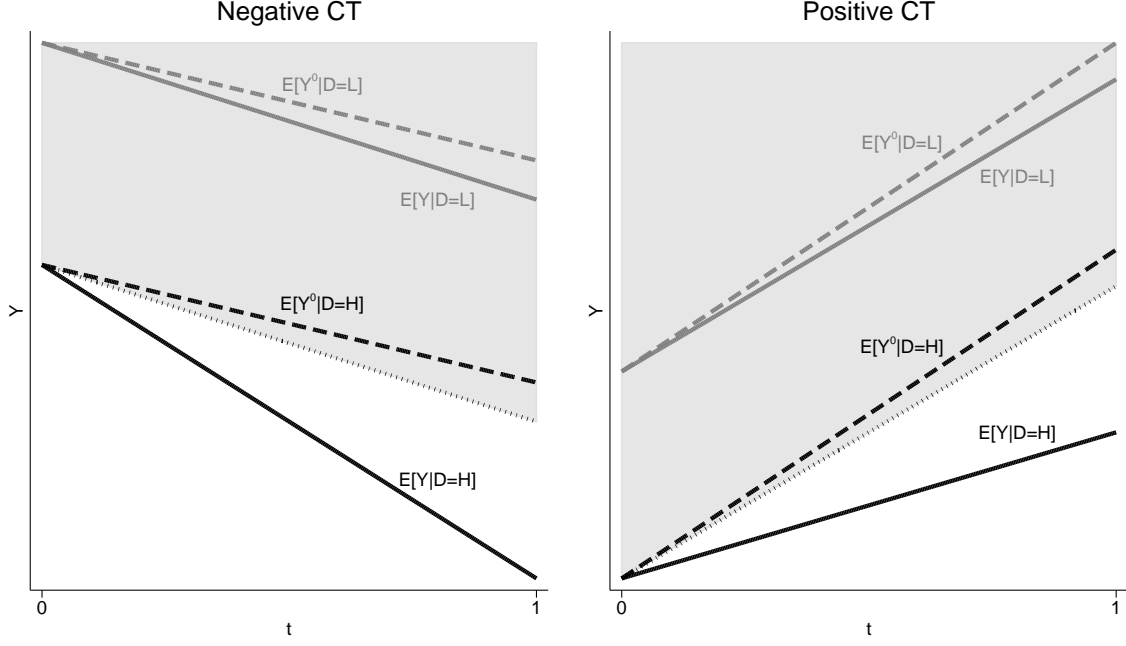
$$> 0.$$

7

Figure 2: Negative DiD Estimand, Assumption (8)

Similarly, for the second case:

$$ATET_1^{H0} \leq E[Y_1|D = H] - E[Y_0|D = H] - [E[Y_1|D = L] - E[Y_0|D = L]] < 0.$$

Given that the CT assumption in (6) holds, assumptions (7) and (8) would be violated in two scenarios. First, group $D = L$ reacts in the opposite direction to treatment L than group $D = H$ to treatment H compared to no treatment. In that case, the true effect would be smaller in magnitude than the DiD estimand. Second, the group $D = L$ reacts more strongly to treatment L than group $D = H$ to treatment H compared to no treatment. The sign of the DiD estimand would be opposite to that of the true effect. That is, if unordered treatments are compared with DiD approaches the above assumptions will be less plausible. If, for example, Duflo (2001) explored different changes in the curriculum instead, it would be difficult to say anything about the sign or the magnitude of the different treatment effects.

Note the assumption of parallel trends under non-treatment in (6) is stronger than necessary. The above results hold also if

$$E[Y_1^0|D = H] - E[Y_0^0|D = H] \leq E[Y_1^L|D = L] - E[Y_0^0|D = L] \tag{10}$$
$$\text{if} \ \ E[Y_1^H|D = H] - E[Y_0^0|D = H] > E[Y_1^L|D = L] - E[Y_0^0|D = L]$$

or

$$E[Y_1^0|D = H] - E[Y_0^0|D = H] \geq E[Y_1^L|D = L] - E[Y_0^0|D = L] \tag{11}$$
$$\text{if} \quad E[Y_1^H|D = H] - E[Y_0^0|D = H] < E[Y_1^L|D = L] - E[Y_0^0|D = L].$$

Hence, even if unobserved factors influence both groups differently, the interpretation of DiD estimand as the lower bound is still valid in certain cases. The development of the expected potential outcome of group $D = H$ under non-treatment needs to be below the development of the expected observed outcome of group $D = L$ from the initial level of group $D = H$ in the first case, and above it in the second case. That is, it has to be anywhere in the gray shaded area in Figures 1 and 2. For example, consider negative regional economic shocks in Duflo (2001). Her results could still be interpreted as lower bound as long as regions with low levels of new schools were affected less than regions with high levels of new schools.

In Sections (2) and (3.1), it is easy to show that identification of the respective parameter can be improved by appropriately including covariates if the assumptions do not hold unconditionally. However, the estimand of the lower bound does not necessarily approach the $ATET_1^{H0}$ with the inclusion of covariates. It may even depart further from the true value. By controlling for important dimensions of effect heterogeneity the estimand approaches the true value of the $ATET_1^{HL}$. Yet, there is no guarantee that this effect is not smaller in absolute values than the unconditional lower bound of the $ATET_1^{H0}$.

4   Conclusion

Studies using DiD approaches to compare two or more distinct treatments have not recognized that such a strategy imposes stricter identifying assumptions than a simple comparison of a treated group to a non-treated control group. If the CT of the expected potential outcome in the non-treatment state is true, the effect of at least one treatment needs to be homogeneous, which is not credible in most applications due to non-random treatment assignment. In that case, the DiD estimand cannot identify the difference between two treatment effects. Instead, as has been shown in the present paper, it can be interpreted as the lower bound in magnitude of the ATET in comparison to the non-treatment case under weaker assumptions. This is possible if in expectation the direction of both treatment effects on the respective groups compared to no treatment is the same, i.e. either both are negative or both are positive, but one treatment has a stronger effect than the other. In general, these assumptions are plausible for ordered treatments or treatments that vary in intensity.

References

Alatas, V., and L. A. Cameron (2008): "The Impact of Minimum Wages on Employment in a Low-Income Country: A Quasi-Natural Experiment in Indonesia," *Industrial and Labor Relations Review*, 61(2), 201–223.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4), 795–813.

Frölich, M. (2004): "Programme Evaluation with Multiple Treatments," *Journal of Economic Surveys*, 18(2), 181–224.

Havnes, T., and M. Mogstad (2011a): "Money for nothing? Universal child care and maternal employment," *Journal of Public Economics*, 95(11–12), 1455–1465.

——— (2011b): "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes," *American Economic Journal: Economic Policy*, 3(2), 97–129.

Lechner, M. (2011): "The Estimation of Causal Effects by Difference-in-Difference Methods," *Foundations and Trends in Econometrics*, 4(3), 165–224.