Nonlinear Difference-in-Differences and Difference-in-Difference-in-Differences with Placebo and Surrogate Outcomes

Adam N. Glynn* Nahomi Ichino†

July 17, 2018
(Link to up-to-date draft)

Abstract

Difference-in-differences (DiD) and difference-in-difference-in-differences (DiDiD) allow for the correction of unmeasured confounding. However, these techniques require that the same variable be measured prior to treatment or on units for which the treatment could have had no effect. Athey and Imbens (2006) provides a quantile-based DiD approach which Sofer et al. (2016) extends to allow the use of placebo (negative) outcomes in lieu of pre-treatment measurements. We extend these approaches in three ways. First, we demonstrate the use of placebo outcomes in the DiD context for the estimation of the average treatment effect on the controls. Second, we show that surrogate outcomes can be used in an analogous but reverse manner to placebo outcomes. Third, we show that these quantile-based techniques can be implemented in the DiDiD context. We apply these methods to a study of whether exposure to candidate debates affected Nepalese citizens' sense of political efficacy.

*Associate Professor, Department of Political Science, Emory University, 327 Tarbutton Hall, 1555 Dickey Drive, Atlanta, GA 30322. aglynn@emory.edu

†Assistant Professor, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 S. State St., Ann Arbor, MI 48109. nichino@umich.edu

Prepared for presentation at PolMeth XXXV at Brigham Young University, July 18–21, 2018.

1 Introduction

Difference-in-differences (DiD) and difference-in-difference-in-differences (DiDiD) allow for the correction of unmeasured confounding. However, these techniques require the outcome variable to be measured on "treated" observations for which the treatment could have had no effect. Typically, these observations are pre-treatment measures of the outcome for the treated units, and some of these observations may be measures on "treated" units that are not eligible for the treatment. In a typical example, a DiD analysis that makes use of pre- and post-treatment data in a comparison of areas with a program (treated) and areas without that program (control) may be extended to a DiDiD analysis by comparing subjects who would have been either eligible or ineligible for that program due to a minimum age requirement or maximum age limit within those areas.

In this paper, we discuss the extension of DiD and DiDiD techniques with two different types of outcome variables: placebo (negative) outcomes and surrogate outcomes. This builds on Sofer, Richardson, Colicino, Schwartz and Tchetgen (2016), which showed how the changes-in-changes model (CiC) of non-linear DiD (Athey and Imbens 2006) could be extended to allow the use of placebo outcomes in lieu of pre-treatment measurements of the outcome for the estimation of the average treatment effect on the treated (ATT). The Sofer et al. (2016) approach, which they call negative outcome control (NOC), relies on placebo outcomes – outcomes that the treatment should not affect and for which the confounding should equal the confounding of primary outcomes on the quantile scale. For example, Sofer et al. (2016) uses body mass index (BMI) as a placebo outcome to help estimate the effect of black carbon exposure due to air pollution on fibrinogen (blood inflammation). There is likely unmeasured confounding of the relationship between black carbon and fibrinogen because those living in areas exposed to high levels of black carbon are also likely to have other risk factors for fibrinogen. BMI may serve as a useful placebo outcome because it is unlikely to be directly affected by black carbon, but it is likely to share many of the same confounding factors as fibrinogen.

We extend the Athey and Imbens (2006) and Sofer et al. (2016) approaches in three ways. First, we demonstrate the use of placebo outcomes in the DiD context for the estimation of the average treatment effect on the controls (ATC). While ATT addresses the effects of existing exposures or programs on those who have been exposed, ATC addresses the effects of potential program expansions to a currently untreated set of subjects.

Second, we show that the techniques used with placebo outcomes can also be used in reverse for surrogate (proxy) outcomes. In contrast to placebo outcomes, for which the treatment has no effect but the confounding is equal to the primary outcome, a surrogate outcome has no confounding but the treatment is assumed to have the same effect on the surrogate outcome as the primary outcome as measured on the quantile scale. Note that this definition of a surrogate outcome is slightly different from other definitions that have been proposed (see VanderWeele (2013) and associated discussion for a survey). Using this definition, we show that the method for imputing the missing

potential outcomes for ATC using placebo outcomes is identical to the method for imputing the missing potential outcomes for the ATT using surrogate outcomes. Similarly, we show that the method for imputing the missing potential outcomes for ATT using placebo outcomes is identical to the method for imputing the missing potential outcomes for the ATC using surrogate outcomes.

One surprising result is that for the estimation of ATT using surrogate outcomes, we use primary outcome measures only for the treated units. This is similar to how surrogate outcomes are typically used when the primary outcome is not available for the sample. In our context, this property implies that researchers may use these methods for the most extreme forms of differential attrition in randomized experiments. For example, suppose we can randomize election observers to polling stations and that we are interested in the effect of election observers on voter intimidation by political parties, which can only be measured by those election observers. We lack measures of intimidation at the unobserved polling stations, precluding a direct estimation of the effect of observers on voter intimidation. However, if we can measure a proxy outcome like turnout on both observed and unobserved polling stations, then our approach allows us to generate an estimate for the effect of election observers on voter intimidation.

Third, we show how these techniques can be extended to the DiDiD context and demonstrate this in detail using placebo (negative) outcomes. We develop two methods to relax the Sofer et al. (2016) assumption of quantile-quantile primary-placebo equi-confounding by using pre- and post-treatment data. The first method simply takes the difference between the post-treatment NOC estimate and the pre-treatment NOC estimate produces consistent estimates as long as differential confounding between the primary and placebo outcome is time-invariant on the outcome scale. The second method, which we call the NOCNOC estimator, is fully non-linear and scale invariant. We apply these methods to analyze whether exposure to candidate debates affected the political efficacy of Nepalese citizens. We use political knowledge as a placebo outcome, despite the fact that first, exposure to the debates may have had some small effect on knowledge, and second, it is difficult to determine *ex ante* whether cross-sectional confounding would be equal for political efficacy and knowledge. Using pre-treatment measures of both political efficacy and knowledge and our two DiDiD estimators, we demonstrate that exposure to candidate debates likely had no effect on political efficacy, even though traditional DiD analysis with placebo tests in conventional practice suggested a positive effect.

2 Review of Changes-in-Changes (CiC) and Negative Outcome Control (NOC)

Define the outcome variable Y_{at} for action/treatment group $a = \{0, 1\}$ (control, treatment) at time $t = \{0, 1\}$ (pre-treatment, post-treatment). Then the linear DiD estimator can be written as

the following:

$$\begin{aligned}&(\widehat{E}[Y_{11}] - \widehat{E}[Y_{01}]) - (\widehat{E}[Y_{10}] - \widehat{E}[Y_{00}]) \\ &= \widehat{E}[Y_{11}] - (\widehat{E}[Y_{01}] + (\widehat{E}[Y_{10}] - \widehat{E}[Y_{00}]))\end{aligned} \tag{1}$$

This is an estimator for the average treatment effect for the treated (ATT) in the post-treatment period, which can be defined in terms of the missing potential outcomes $Y_{11}(0)$, the outcome that would have occured for the treated units in the post-treatment period if they had been assigned control. With this notation, the ATT is:

$$E[Y_{11}] - E[Y_{11}(0)] \tag{2}$$

Note that $\widehat{E}[Y_{11}]$ in (1) is simply the mean outcome among the treated units in the post-treatment period and is a plug-in estimator for $E[Y_{11}]$ in (2). The difficult task is the estimation of the second term in (2), the mean of the missing potential outcomes $E[Y_{11}(0)]$. The linear DiD approach estimates this quantity with $\widehat{E}[Y_{01}] + (\widehat{E}[Y_{10}] - \widehat{E}[Y_{00}])$ from (1), where $\widehat{E}[Y_{10}] - \widehat{E}[Y_{00}]$ is the correction for the confounding.

2.1 Changes-in-Changes (CiC)

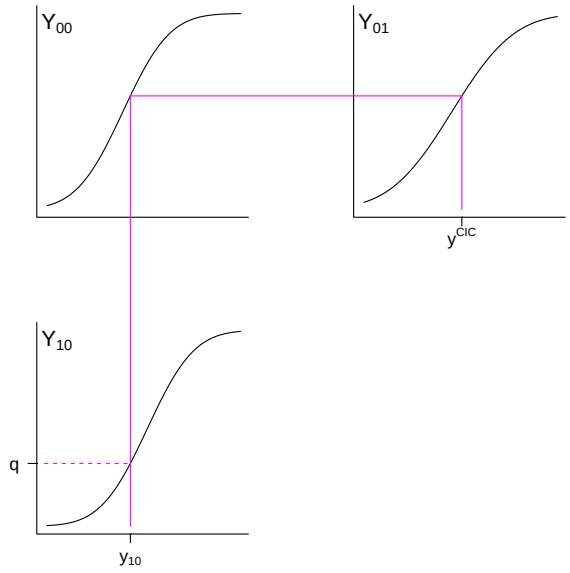
Athey and Imbens (2006)'s CiC procedure is a nonlinear DiD approach to the correction for unmeasured confounding. It generalizes the linear DiD in the following way:

$$\widehat{E}[Y_{11}] - \frac{1}{n_{10}} \sum_{i=1}^{n_{10}} \hat{F}^{-1}_{Y_{01}}(\hat{F}_{Y_{00}}(Y_{10,i})) \tag{3}$$

where n_{10} is the number of treated units in the pre-treatment period, $\hat{F}_{Y_{00}}$ represents a consistent estimator of the CDF from the control units in the pre-treatment period, and $\hat{F}^{-1}_{Y_{01}}$ represents a consistent esimator of the inverse CDF from the control units in the post-treatment period. Note that the second term from (3) is analogous to the second term in the linear DiD estimator in (1) and attempts to estimate the second term of the ATT in (2).

Consider what the second term of (3) is doing for a particular quantile q. For example, suppose we want to impute the missing potential outcome for $q = .4$. The associated CiC procedure is presented in Figure 1. We take the y value associated with the .4 quantile among the treated units in the pre-treatment period (y_{10} in the figure). We calculate what quantile that value would take among the control units in the pre-treatment period (roughly .5 in the figure). We then assume the imputed value should be at the same quantile among the controls in the post-treatment period as in the pre-treatment period (indicated by the horizontal line between the Y_{00} and Y_{01} in the figure). Finally, we calculate the y value associated with the .5 quantile among the controls in the

Figure 1: CiC procedure for imputing $Y(0)$ for quantile q ($F^{-1}_{Y_{11}(0)}(q)$)



post-treatment period (the y^{CiC} in the figure).

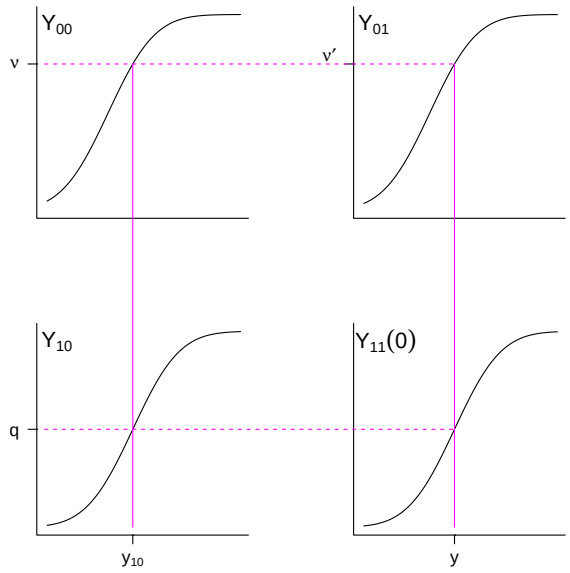
Intuitively, the CiC procedure uses the pre-treatment period to measure confounding on the quantile scale by showing how far the quantiles move from the treatment group to the control group (.4 to .5 in the figure). This is then used to impute the missing potential outcome in the post-treatment period. Athey and Imbens (2006) provides a set of assumptions under which (3) is consistent for ATT, and Sofer et al. (2016) weakens these assumptions. Here we provide a reduced form of the Sofer et al. (2016) assumptions extracted from their proof.[1] The underlying assumption is that the confounding in the pre-treatment period equals the confounding in the post-treatment period on the quantile scale. This is stated explicitly in Assumption 1a for all quantiles. Additionally, the support for the treated units in the pre-treatment period must be contained in the support of the controls as stated in Assumption 1b. That the number of control units is often much larger than the number of treated units can sometimes make this support condition reasonable.

Assumption 1a.

$$F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q)) = F_{Y_{00}}(F^{-1}_{Y_{10}}(q)), q \in [0,1]$$

[1]The more expansive assumptions of Athey and Imbens (2006) and Sofer et al. (2016) provide more detail in terms of data generating processes that might give rise to these reduced form assumptions being satisfied.

Figure 2: CiC quantile-quantile pre-post equi-confounding assumption for quantile q



Assumption 1b.

$$if\ 0 < f_{Y_{10}}(y_{10}),\ then\ 0 < F_{Y_{00}}(y_{10}) < 1$$

For a particular quantile q, Assumption 1a is depicted in Figure 2, where the move from quantile q among the treated units in the pre-treatment period to quantile ν among the control units in the pre-treatment period is assumed to mirror the move from quantile q in the missing potential outcomes to quantile ν' among the control units in the post-treatment period. Note however that we only need this assumption to hold on average across the quantiles. Therefore, the sufficient assumption, along with the aforementioned support conditions for the estimation of ATT can be stated as the following:

Assumption 2.

$$E_q[F^{-1}_{Y_{11}(0)}(q)] = E_q[F^{-1}_{Y_{01}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q)))], q \sim Unif(0,1)$$

2.2 Negative Outcome Control (NOC)

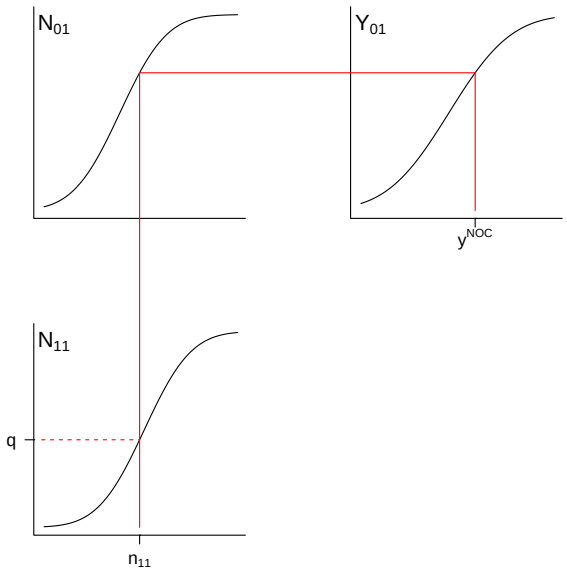
Without covariates, Sofer et al. (2016)'s NOC procedure is very similar to the CiC procedure. The pre-treatment outcomes, Y_{10} and Y_{00}, are replaced with post-treatment placebo (negative)

outcomes, N_{11} and N_{01}. Sofer et al. (2016)'s application, for example, is cross-sectional in the post-treatment period: fibrinogen (the primary outcome Y) and BMI (the placebo/negative outcome N) are both measured after exposure to black carbon. Hence, the NOC procedure is the following:

$$\widehat{E}[Y_{11}] - \frac{1}{k_{11}} \sum_{i=1}^{k_{11}} \hat{F}_{Y_{01}}^{-1}(\hat{F}_{N_{01}}(N_{11,i})) \tag{4}$$

where k_{11} is the number of treated units in the post-treatment period, $\hat{F}_{N_{01}}$ represents a consistent estimator of the CDF for the placebo outcome from the control units in the post-treatment period, and $\hat{F}_{Y_{01}}^{-1}$ represents a consistent estimator of the inverse CDF for the primary outcome for the control units in the post-treatment period. Note that the second term in (4) is analogous the second term in (3).

Figure 3: NOC procedure for imputing $Y(0)$ for quantile q ($F_{Y_{11}(0)}^{-1}(q)$)



Again, we can build intuition about the NOC by considering what the second term of (4) is doing for a particular quantile q. For example, suppose we want to impute the missing potential outcome for $q = .4$. The associated NOC procedure is presented in Figure 3. We take the y value associated with the .4 quantile for the placebo outcome among the treated units in the post-treatment period (n_{11} in the figure). We calculate what quantile that placebo outcome value would take among the control units in the post-treatment period (roughly .6 in the figure). We then assume the imputed value should be at the same quantile among the controls for the primary

outcome in the post-treatment period as among the placebo outcome in the same post-treatment period (horizontal line between the N_{01} and Y_{01} in the figure). Finally, we calculate the y value associated with the .6 quantile among the controls in the post-treatment period (the y_{NOC} in the figure). The formal statement of this assumption is stated below (and as before, the support of the control units must contain the support of the treated units):

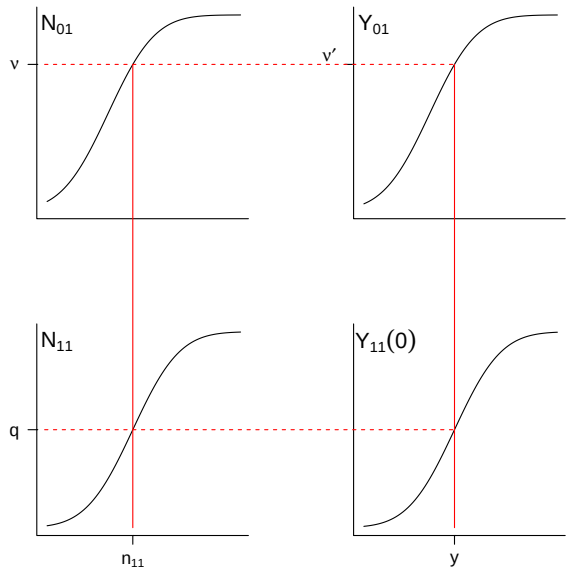
Assumption 3a.

$$F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q)) = F_{N_{01}}(F^{-1}_{N_{11}}(q)), q \in [0,1]$$

Assumption 3b.

$$\textit{if } 0 < f_{N_{11}}(n_{11}), \ \textit{then } 0 < F_{N_{01}}(n_{11}) < 1$$

Figure 4: NOC quantile-quantile placebo-primary equi-confounding assumption for quantile q



Intuitively, the placebo outcome plots measure the confounding on the quantile scale by showing how far the quantiles move from the treatment group to the control group (.4 to .6 in the figure). This makes sense as a measure of confounding because the treatment should not affect a placebo outcome. This .6 is then used to impute the missing potential outcome in the post-treatment period. Again, the underlying assumption for the NOC procedure is quantile-quantile equi-confounding assumption, except that now the equality of confounding is assumed between the placebo outcome

and the primary outcome. In the CiC procedure, the equal confounding is assumed across time. This NOC assumption is depicted for a particular quantile q in Figure 4. Finally, as before we only need this assumption to hold on average. Thus the sufficient assumption, along with the aforementioned control conditions, is:

Assumption 4.

$$E_q[F^{-1}_{Y_{11}(0)}(q)] = E_q[F^{-1}_{Y_{01}}(F_{N_{01}}(F^{-1}_{N_{11}}(q)))], q \sim Unif(0,1)$$

3 Average Treatment Effect on the Controls (ATC) with NOC

In addition to the CiC procedure for estimating ATT, Athey and Imbens (2006) describe a procedure for estimating the average treatment effect in the post-treatment period for those that received control (ATC):

$$E[Y_{01}(1)] - E[Y_{01}] \tag{5}$$

Note that for (5), the missing potential outcome is the outcome in the post-treatment period that would have been observed for the control units if they had been assigned treatment ($Y_{01}(1)$). In contrast to ATT, which is a measure of the effect of an existing program, ATC is a measure of the effect of a potential program expansion to those currently in the control condition.

The CiC estimator for ATC can be written as the following:

$$\frac{1}{n_{00}} \sum_{i=1}^{n_{00}} \hat{F}^{-1}_{Y_{11}}(\hat{F}_{Y_{10}}(Y_{00,i})) - \widehat{E}[Y_{01}] \tag{6}$$

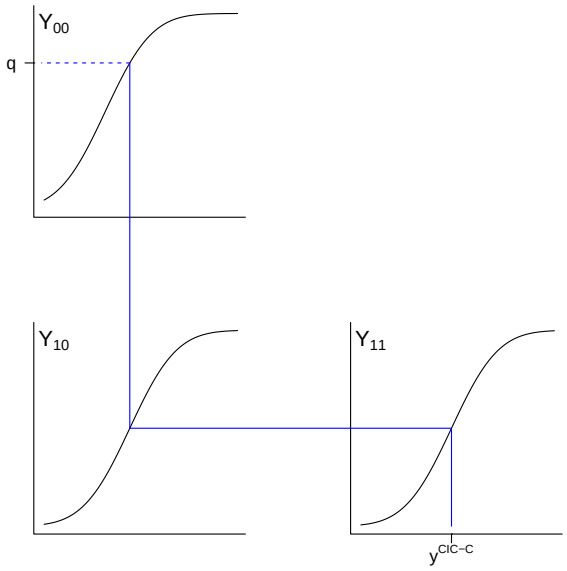
where n_{00} is the number of control units in the pre-treatment period, $\hat{F}_{Y_{10}}$ represents a consistent estimator of the CDF from the treated units in the pre-treatment period, and $\hat{F}^{-1}_{Y_{11}}$ represents a consistent esimator of the inverse CDF from the treated units in the post-treatment period. $\widehat{E}[Y_{01}]$ is simply the mean outcome among the control units in the post-treatment period. This procedure is visualized in Figure 5.

Note that the key assumption for this ATC estimator is the following:

Assumption 5a.

$$F_{Y_{11}}(F^{-1}_{Y_{01}(1)}(q)) = F_{Y_{10}}(F^{-1}_{Y_{00}}(q)), q \in [0,1]$$

Figure 5: CiC procedure for imputing $Y(1)$ for quantile q ($F^{-1}_{Y_{01}(1)}(q)$)



Assumption 5b.

$$if\ 0 < f_{Y_{00}}(y_{00}),\ then\ 0 < F_{Y_{10}}(y_{00}) < 1$$

where treated and control have been switched in the the support condition. Assumption 5b implies that control units are contained within the support of the treated units. This assumption may be plausible when the expansion of a program being considered is incremental and hence limited to a small subset of the control units. In this case, the support of the treated units may reasonably contain the support of control units being considered for the expansion.

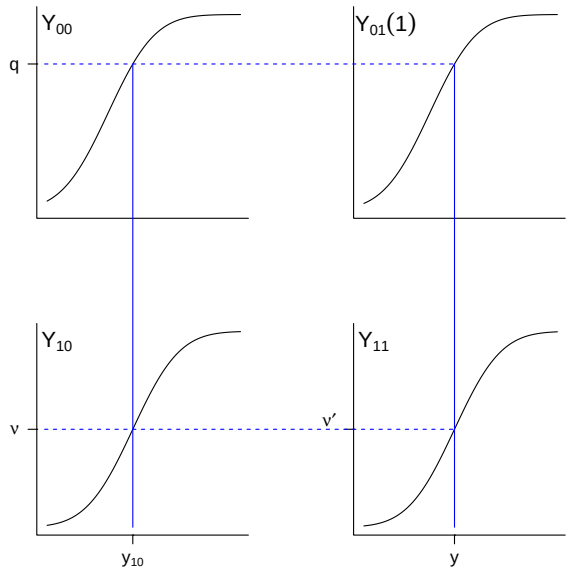
Note again, that for consistent estimation of ATC, Assumption 5a only needs to hold on average. This can be stated as the following:

Assumption 6.

$$E_q[F^{-1}_{Y_{01}(1)}(q)] = E_q[F^{-1}_{Y_{11}}(F_{Y_{10}}(F^{-1}_{Y_{00}}(q)))], q \sim Unif(0,1)$$

Although Sofer et al. (2016) does not comment on the ATC, here we show that it is straightforward to adapt the CiC ATC approach to placebo outcomes.

Figure 6: CiC quantile-quantile pre-post equi-confounding assumption for quantile q for ATC



$$\frac{1}{k_{01}} \sum_{i=1}^{k_{01}} \hat{F}_{Y_{11}}^{-1}(\hat{F}_{N_{11}}(N_{01,i})) - \widehat{E}[Y_{01}] \tag{7}$$

where k_{01} is the number of control units in the post-treatment period for the placebo outcome, and $\hat{F}_{N_{11}}$ represents a consistent estimator of the CDF for the post-treatment placebo outcome for the treated units, and $\hat{F}_{Y_{11}}^{-1}$ represents a consistent estimator the inverse CDF for the post-treatment primary outcome for the treated units. This is visualized in Figure 7.

The key assumption for this ATC estimator is the following:

Assumption 7a.

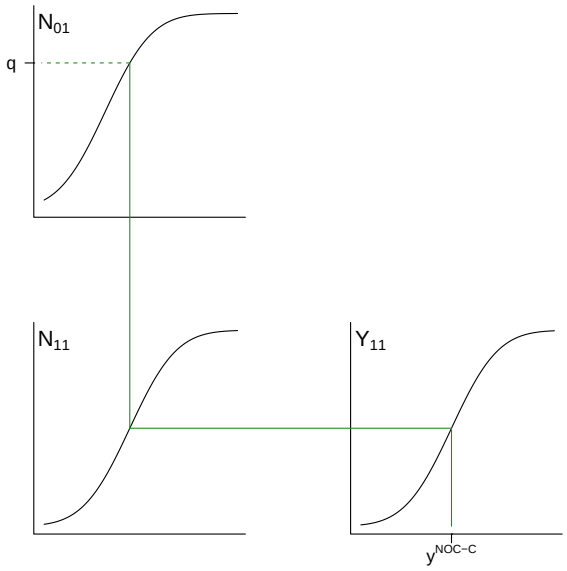
$$F_{Y_{11}}(F_{Y_{01}(1)}^{-1}(q)) = F_{N_{11}}(F_{N_{01}}^{-1}(q)), q \in [0,1]$$

Assumption 7b.

$$\textit{if } 0 < f_{N_{01}}(n_{01}), \ \textit{then } 0 < F_{N_{11}}(n_{01}) < 1$$

where again, treated and control have been switched in the the support condition (represented in Figure 8). Similarly, for consistent estimation of ATC, Assumption 7a only needs to hold on

Figure 7: NOC procedure for imputing $Y(1)$ for quantile q ($F^{-1}_{Y_{01}(1)}(q)$)



average. This can be stated as the following:

Assumption 8.

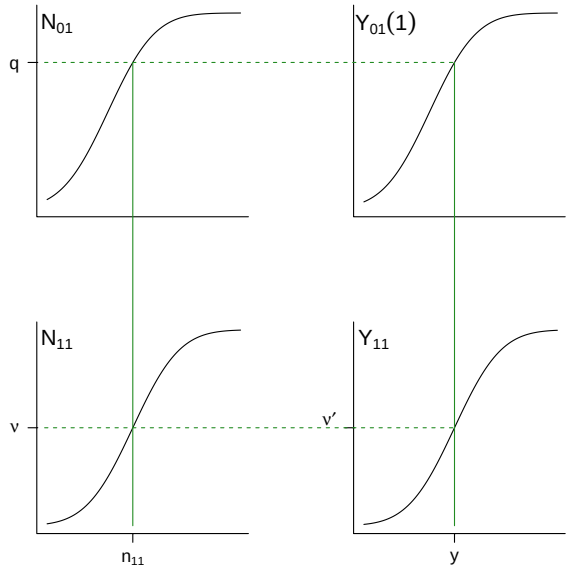
$$E_q[F^{-1}_{Y_{01}(1)}(q)] = E_q[F^{-1}_{Y_{11}}(F_{N_{11}}(F^{-1}_{N_{01}}(q)))], q \sim Unif(0,1)$$

4 Using Surrogate/Proxy Outcomes

The previous sections have established that placebo outcomes can be used to estimate both ATT and ATC when there is unmeasured confounding. Recall that placebo outcomes are outcomes that are unaffected by the treatment but that share the same amount of confounding with the primary outcome, where confounding is measured in quantiles. In this section, we show that the reverse approach can also work – that a proxy outcome, for which there is no confounding, but for which the treatment effect is the same as the primary outcome, can be used to correct for unmeasured confounding. In this sense, proxy outcomes function as a sort of surrogate outcome.

Define the proxy outcome for action/treatment group $a = \{0, 1\}$ (control, treatment) at time $t = \{0, 1\}$ (pre-treatment, post-treatment) as S_{at}. Then the proxy outcome estimator for ATT can

Figure 8: NOC quantile-quantile placebo-primary equi-confounding assumption for quantile q for ATC



be written as:

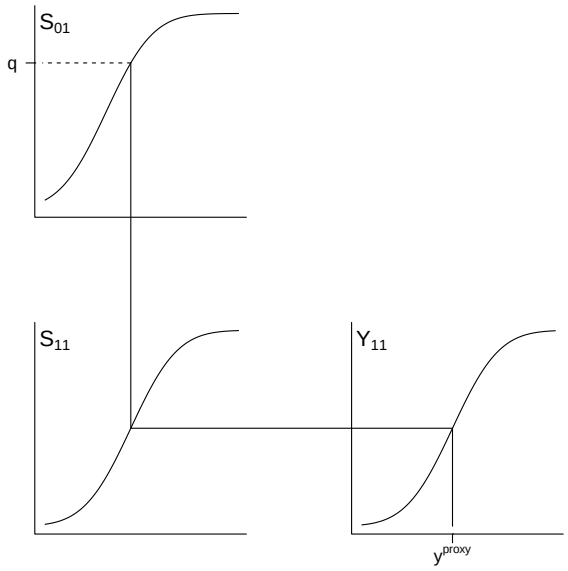
$$\widehat{E}[Y_{11}] - \frac{1}{n_{01}} \sum_{i=1}^{n_{01}} \hat{F}_{Y_{11}}^{-1}(\hat{F}_{S_{11}}(S_{01,i})) \tag{8}$$

where n_{01} is the number of control units in the post-treatment period, $\hat{F}_{S_{11}}$ represents a consistent estimator of the CDF of S from the treated units in the pre-treatment period, and $\hat{F}_{Y_{11}}^{-1}$ represents a consistent esimator of the inverse CDF of the primary outcome from the treated units in the post-treatment period.

We highlight two features of this estimator. First, the second term of (8) is structurally identical to the first term of (7), just that the placebo outcomes N are replaced with the proxy outcomes S. This reflects the symmetry of this estimator for ATT with proxy outcomes (visualized in Figure 9) with the earlier estimator for ATC with placebo outcomes (Figure 7).

Second, the proxy estimator for ATT does not use primary outcomes measured on the control units; neither Y_{01} nor Y_{00} are included in the formula. This implies some useful applications, such as when we have a randomized treatment, but there is severe differential attrition between treatment and control. For example, in the extreme case where Y is only measured on the treated units, this proxy outcome estimator can be used if we have a proxy/surrogate outcome measured on all units. One such scenario is if election observers are randomized to polling stations, and those election observers also record instances of voter intimidation at those polling stations (Y).

Figure 9: Proxy outcome ATT estimator procedure for imputing $Y(0)$ for quantile q ($F^{-1}_{Y_{11}(0)}(q)$)



Treatment status and availability of data on the outcome are inextricably tied, but administrative data available for both observed and unobserved polling stations might be used as a proxy variable S.

Assumption 9a.

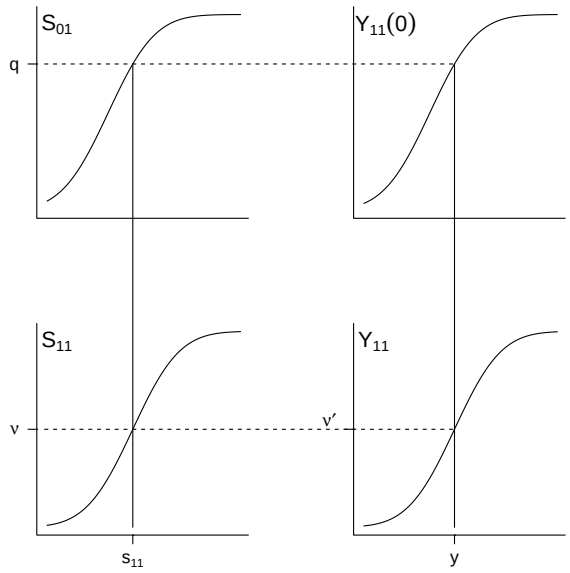
$$\begin{aligned} F_{Y_{11}}(F^{-1}_{Y_{11}(0)}(q)) &= F_{S_{11}}(F^{-1}_{S_{11}(0)}(q)), q \in [0,1] \\ &= F_{S_{11}}(F^{-1}_{S_{01}}(q)), q \in [0,1] \end{aligned}$$

Assumption 9b.

$$\textit{if } 0 < f_{S_{01}}(s_{01}), \ \textit{then } 0 < F_{S_{11}}(s_{01}) < 1$$

The assumptions necessary for consistency of the proxy outcome estimator (visualized in Figure 10) are similar to those for previous estimators, but there are key differences. Assumption 9a states that the effect on Y for the treated units on the quantile scale equals the effect on S (first equation) and that treatment is as-if randomized with respect to S (second equation). Moreover, there may be real concerns with the support condition in Assumption 9b. If the treatment has an effect on S, then the support of S_{01} may not be contained in S_{11}. In this case the overall ATT may not be

Figure 10: Quantile-quantile proxy-primary equi-confounding assumption for quantile q



estimable without additional modeling assumptions. However, if the support condition holds, then Assumption 9a only needs to hold on average for the ATT:

Assumption 10.

$$E_q[F^{-1}_{Y_{11}(0)}(q)] = E_q[F^{-1}_{Y_{11}}(F_{S_{11}}(F^{-1}_{S_{01}}(q)))], q \sim Unif(0,1)$$

Finally, note that an analogous proxy approach can be taken for ATC, with the estimator:

$$\frac{1}{n_{11}} \sum_{i=1}^{n_{11}} \hat{F}^{-1}_{Y_{01}}(\hat{F}_{S_{01}}(S_{11,i})) - \widehat{E}[Y_{01}] \tag{9}$$

Note that the first term in (9) is structurally identical to the second term of the placebo outcome estimator of ATT (4), except that the placebo outcomes have been replaced with proxy outcomes. Assumptions 11a and 12 are also similar to Assumptions 9a and 10, but reversed. Finally, the proxy estimator for the ATC does not use primary outcomes measured on the treated units – neither Y_{11} nor Y_{10} appear in the formula.

Assumption 11a.

$$F_{Y_{01}}(F^{-1}_{Y_{01}(1)}(q)) = F_{S_{01}}(F^{-1}_{S_{01}(1)}(q)), q \in [0,1]$$
$$= F_{S_{01}}(F^{-1}_{S_{11}}(q)), q \in [0,1]$$

Assumption 11b.

$$\textit{if } 0 < f_{S_{11}}(s_{11}), \textit{ then } 0 < F_{S_{01}}(s_{11}) < 1$$

Assumption 12.

$$E_q[F^{-1}_{Y_{01}(1)}(q)] = E_q[F^{-1}_{Y_{01}}(F_{S_{01}}(F^{-1}_{S_{11}}(q)))], q \sim Unif(0,1)$$

5 Difference-Difference-in-Differences (DiDiD)

When both pre-treatment observations and a placebo outcome are available, we have information about the relationship between the confounding on both N and $Y(0)$. This is because prior to the application of the treatment, we can observe $Y(0)$ among the $A = 1$ units. The easiest way to explore this information is to conduct the NOC procedure on the pre-treatment observations. If there is evidence of a treatment effect in the pre-treatment period, then we will know that there is a differential in the confounding for N and $Y(0)$ during this period. In other words, we learn that Assumption 4 does not hold in the pre-treatment period. If we further believe that this differential is time-invariant, then we would have evidence that Assumption 4 is violated in the post-treatment period. This analysis constitutes a placebo test of the NOC procedure.

One simple approach to correct for the confounding differential can be conceptualized as a generalization of a difference-in-difference-in-differences procedure. If Assumption 13 holds, then the bias of the NOC procedure in the pre-treatment period equals the bias of the NOC procedure in the post-treatment period, and we can simply subtract a pre-treatment NOC estimate from the post-treatment NOC estimate to correct for this bias (see Appendix for proof).

Assumption 13. *For $t = \{0,1\}$ corresponding to pre- and post-treatment periods, a violation of the equi-confouding assumption (Assumption 3a) can be written as the following:*

$$F^{-1}_{Y_{1t}(0)}(q) + \delta(q)_t = F^{-1}_{Y_{0t}(0)}(F_{N_{0t}}(F^{-1}_{N_{1t}}(q))), q \in [0,1]$$

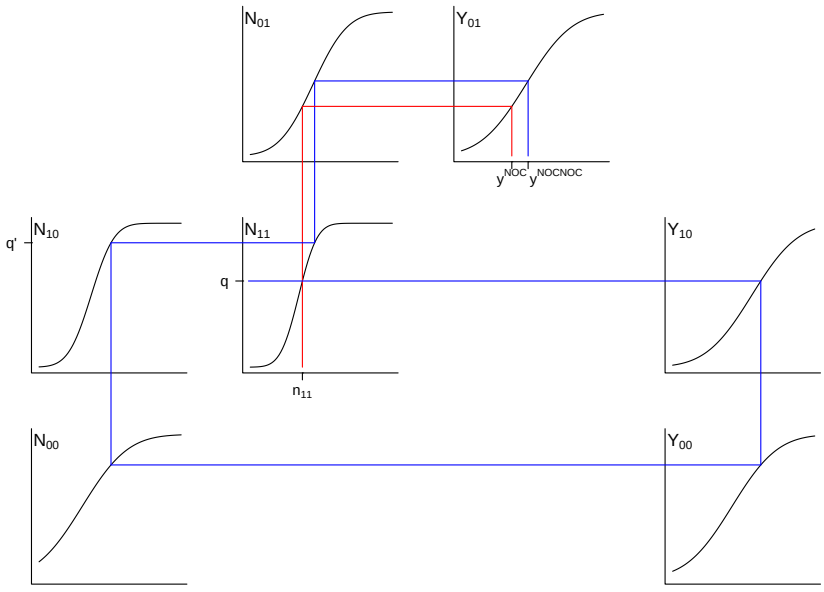
where $\delta(q)_t$ represents the confounding differential in time t. Assume that the average confounding differential is time invariant such that $E_q[\delta(q)_0] = E_q[\delta(q)_1]$.

However, this post-minus-pre NOC approach relies on the average bias in the pre and post-treatment periods to be equal on the Y scale. Similar to the logic behind extending linear DiD to CiC or NOC, we can produce a nonlinear DiDiD estimator that works directly on the quantiles. Hence we introduce a nonlinear DiDiD estimator for ATT that we call the NOCNOC estimator:

$$\frac{1}{n_{11}}\sum_{i=1}^{n_{11}} Y_{11,i} - \frac{1}{n_{10}}\sum_{i=1}^{n_{10}} \hat{F}^{-1}_{Y_{01}}(\hat{F}_{N_{01}}(\hat{F}^{-1}_{N_{11}}(\hat{F}_{N_{10}}(\hat{F}^{-1}_{N_{00}}(\hat{F}_{Y_{00}}(Y_{10,i})))))) \tag{10}$$

The fundamental idea behind the NOCNOC estimator is depicted in Figure 11. Suppose we want to estimate $Y_{11}(0)$ for a particular quantile q. This is accomplished by starting with the quantile q pre-treatment primary outcome in the treated group (Y_{10}) and asking what quantile q' among the pre-treatment values of the placebo in the treated group (N_{10}) would have produced that Y_{10}. This process can be seen by following the blue path in the lower part of Figure 11. Then the estimated q' is used instead of q to start the NOC process in the post-treatment period. This process can be seen by the blue path in the upper part of Figure 11. For comparative purposes, NOC process is represented by the red path in the upper part of of Figure 11. The fundamental difference between the procedures is that NOC starts with quantile q in the post-treatment placebo distribution (N_{11}) and NOCNOC uses quantile q' in the post-treatment placebo distribution (N_{11}), having estimated this quantile in the pre-treatment period.

Figure 11: Placebo outcome NOCNOC estimator procedure for imputing $Y(0)$ for quantile q ($F^{-1}_{Y_{11}(0)}(q)$)



The fundamental assumptions of the NOCNOC estimator are presented in Assumptions 14a and 14b below:

Assumption 14a.

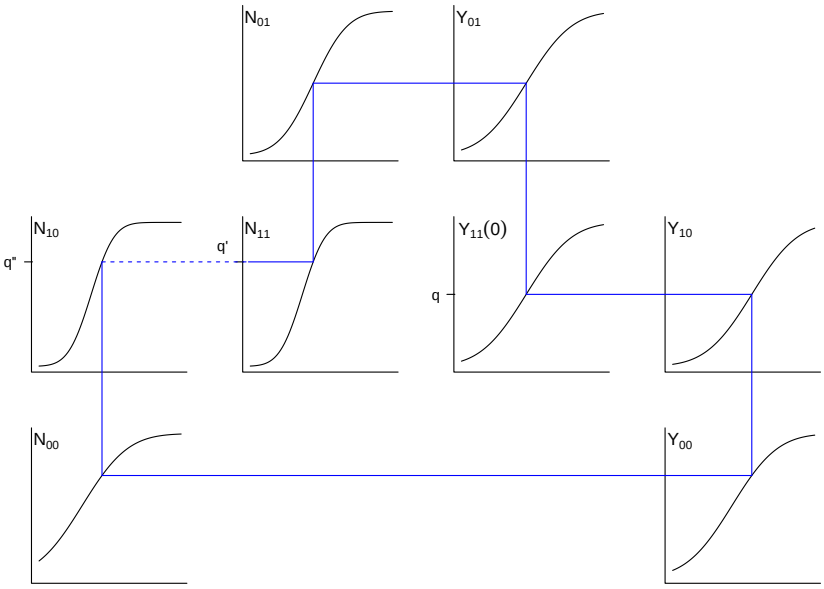
$$F_{N_{11}}(F^{-1}_{N_{01}}(F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q)))) = F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q)))), q \in [0,1]$$

Assumption 14b.

$$\begin{aligned} \textit{if } 0 < f_{Y_{10}}(y_{10}), \ \textit{then } & 0 < F_{Y_{00}}(y_{10}) < 1, \\ & 0 < F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(y_{10}))) < 1, \\ & 0 < F_{N_{01}}(F^{-1}_{N_{11}}(F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(y_{10})))))) < 1 \end{aligned}$$

Assumption 14a is visualized in Figure 12 for a particular value of q. The assumption holds in this case, with the output from the left hand side of Assumption 14a depicted on the figure as q', the output from the right hand side of Assumption 14a depicted on the figure as q'', and $q' = q''$. The assumption means that the confounding generated on the quantile scale is equal for the NOC process in the post-treatment period and the pre-treatment period.

Figure 12: Placebo outcome NOCNOC assumption for imputing $Y(0)$ for quantile q ($F^{-1}_{Y_{11}(0)}(q)$)



Assumption 14b states the support conditions for the NOCNOC estimator, and these ensure

that the cdfs never output negative or positive infinity. For the NOCNOC, the support for the treated units in the pre-treatment period must again be contained in the support of the controls in the pre-treatment period. In addition, quantiles implied by the treated units in the pre-treatment period must also be in the support of the placebo outcomes in the manner presented. See the Appendix for an identification proof.

6 Simulation Study

In order to assess the performance of the Post-minus-Pre NOC estimator and the NOCNOC estimator, we ran a number of simulations where pre- and post-treatment data were available for both a primary outcome and a placebo outcome. In the first set of simulations, we generated data consistent with the model presented in Athey and Imbens (2006), with some deviations. We then attempted to estimate the ATT from that model with five different estimators: linear DiD, CiC, NOC, Post-minus-Pre NOC, and NOCNOC. The results of those simulations are presented in Table 1. In this table, d is the mean difference in the distributions of the unmeasured confounders for both the primary and placebo outcomes between the treated and control groups. The second set of simulations are similar, except we generated data consistent with the model presented in Sofer et al. (2016) as well as some deviations from that model. The results of those simulations are presented in Table 2. In this table, d is the mean difference in distributions of the unmeasured confounders between the primary and placebo outcomes for a given treatment group and time period. The final set of simulations are also similar, but we generated data consistent with the CiC-r model presented in Athey and Imbens (2006), which reverses the treatment and time dimensions. The results of those simulations are presented in Table 3. Details of all the simulations are presented in the appendix.

The results of these simulation studies provide some guidance as to the potential choice of estimator. The main finding is that for bias, the NOCNOC estimator nearly weakly dominates all other estimators over all simulations at large sample sizes. In particular, all the other estimators have catastrophic failures for at least some of the simulations, while NOCNOC never does. Furthermore, the NOCNOC has reasonable root mean squared error, so not a great amount of efficiency is lost in using the most robust model. The second finding is that while Post-minus-Pre NOC performs well for some models and improves on NOC in many of the simulations, it has some catastrophic failures and is almost always dominated by CiC and NOCNOC. Therefore the Post-minus-pre NOC estimator should probably only be used as a robustness check.

7 Debate Exposure and Political Efficacy in Nepal

Nepal is a country of approximately 29 million that experienced civil war from 1996 to 2006 and was a monarchy until 2008. In December 2017, Nepal held local and national elections for the

Table 1: Simulation study with data consistent with Athey and Imbens (2006)

	DiD		CiC		NOC		Post-Pre NOC		NOCNOC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d=0										
$n_1 = 500$	0.044	2.043	0.148	2.905	22.993	23.105	−0.051	3.160	0.520	4.272
$n_1 = 1000$	−0.030	1.467	0.000	2.092	23.002	23.057	−0.021	2.236	0.276	3.066
$n_1 = 4000$	−0.004	0.733	0.003	1.030	22.985	22.998	−0.013	1.131	0.085	1.534
d=1										
$n_1 = 500$	4.003	4.718	0.139	3.556	29.052	29.187	2.019	4.305	0.762	5.211
$n_1 = 1000$	4.008	4.374	0.107	2.498	29.041	29.106	2.008	3.357	0.437	3.763
$n_1 = 4000$	3.984	4.081	0.002	1.293	28.998	29.015	1.979	2.414	0.108	1.920
d=2										
$n_1 = 500$	8.050	8.397	0.186	3.558	29.069	29.204	4.005	5.445	0.820	5.242
$n_1 = 1000$	8.004	8.183	0.073	2.565	29.037	29.104	4.019	4.793	0.426	3.765
$n_1 = 4000$	7.984	8.028	−0.007	1.247	28.959	28.975	3.958	4.166	0.071	1.863
d=3										
$n_1 = 500$	12.038	12.261	0.191	3.594	29.081	29.215	6.052	7.047	0.799	5.330
$n_1 = 1000$	12.005	12.119	0.086	2.584	29.039	29.109	6.020	6.539	0.428	3.780
$n_1 = 4000$	11.983	12.011	−0.012	1.260	29.013	29.031	5.972	6.104	0.060	1.886

first time under a new constitution that established it as a federal, multiethnic republic. As part of an effort to strengthen the connections between citizens and their representatives and to encourage issue-oriented rather than personality-centered politics, the Samriddhi Foundation, a Nepalese civil society organization, hosted televised debates among candidates for the House of Representatives (the directly-elected lower house of parliament) for three single-member constituencies within Kanchanpur, Jhapa, and Sunsari districts in November 2017. A community radio station in each of these rural districts invited 1000 randomly-selected citizens to a public venue to view a screening of the recorded candidate debate for their area and/or participate in small-group discussions about the candidates and issues.[2] Recent randomized studies in Ghana, Sierra Leone, and Uganda, where candidate debates are also relatively novel, have found that exposure to candidate debates (sometimes followed by community discussion) increased voters' knowledge about the candidates and their policies and affected how they voted in some contexts (Bidwell, Casey and Glennerster 2016, Brierley, Kramon and Ofosu 2018, Platas and Raffler 2017).

To illustrate our method for DiDiD with placebo outcomes, we assess whether exposure to the candidate debates ahead of these historic elections affected a related but different outcome – citizens' sense of political efficacy. Political efficacy – the sense that one can influence politics

[2]The debates were edited for length and to even out the screen time across candidates.

Table 2: Simulation study with data consistent with Sofer et al. (2016)

	DiD		CiC		NOC		Post-Pre NOC		NOCNOC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d=0										
$n_1 = 500$	23.043	23.129	23.082	23.190	0.041	2.918	19.722	20.117	0.447	2.979
$n_2 = 1000$	22.972	23.017	22.989	23.045	0.026	2.059	19.752	19.951	0.229	2.128
$n_3 = 4000$	23.000	23.011	23.004	23.018	−0.009	1.012	19.737	19.786	0.076	1.052
d=1										
$n_1 = 500$	29.000	29.111	29.057	29.192	0.145	3.568	25.831	26.287	0.560	3.660
$n_1 = 1000$	29.007	29.062	29.046	29.112	0.080	2.494	25.807	26.034	0.323	2.597
$n_1 = 4000$	28.984	28.998	28.991	29.008	0.012	1.278	25.745	25.806	0.092	1.325
d=2										
$n_1 = 500$	29.050	29.159	29.092	29.226	0.134	3.598	25.806	26.267	0.592	3.628
$n_1 = 1000$	29.006	29.062	29.027	29.096	0.078	2.519	25.824	26.057	0.357	2.605
$n_1 = 4000$	28.984	28.998	28.989	29.005	−0.038	1.260	25.686	25.744	0.059	1.303
d=3										
$n_1 = 500$	29.045	29.156	29.085	29.222	0.193	3.557	25.905	26.372	0.584	3.675
$n_1 = 1000$	29.010	29.066	29.042	29.112	0.055	2.587	25.770	26.009	0.356	2.636
$n_1 = 4000$	28.981	28.995	28.987	29.004	0.019	1.280	25.716	25.775	0.060	1.321

and government (external) and that one can understand politics and government (internal) – is associated with political participation (Campbell, Gurin and Miller 1954, Almond and Verba 1963, Rosenstone and Hansen 1993, Verba, Schlozman and Brady 1995). Low efficacy could create the danger that citizens will fail participate in politics and hold their politicians accountable, leading politicians to learn that they can fail to serve voters and instead serve private interests with little electoral consequence. But involvement in political activities itself could enhance citizens' sense of efficacy (Finkel 1985, Valentino, Gregorowicz and Groenendyk 2009), and we consider the impact of exposure to these debates.

This analysis uses data for 223 respondents who attended one of the events and were randomized into the debate screening condition, along with 510 respondents who signed up for but did not attend the events, for whom we have measures of all items used to construct primary and placebo outcomes both pre- and post-treatment. These respondents were initially interviewed at their homes in November 2017; signed up for an event to be held between November 21 and 27, 2017; and were re-interviewed mostly at their homes in January/February 2018. Excluded are those attendees who were assigned to other treatment arms that included small-group discussions as part of the larger study and those participants who had signed up for dates for which we had to cancel the events. We use pre- and post-treatment data collected on the same units, but this is not necessary for our

Table 3: Simulation study with data consistent with CiC-r model in Athey and Imbens (2006)

	DiD		CiC		NOC		Post-Pre NOC		NOCNOC	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
d=0										
$n_1 = 500$	−0.051	2.393	−0.000	2.642	0.005	2.641	−0.081	3.438	0.160	3.729
$n_1 = 1000$	0.005	1.700	0.031	1.882	0.060	1.864	0.060	2.451	0.190	2.651
$n_1 = 4000$	−0.005	0.838	0.006	0.930	−0.005	0.942	−0.008	1.215	0.025	1.308
d=1										
$n_1 = 500$	4.080	4.960	0.126	3.357	0.112	3.447	0.092	4.612	0.391	4.927
$n_1 = 1000$	4.010	4.472	0.020	2.372	0.033	2.391	−0.005	3.141	0.147	3.343
$n_1 = 4000$	3.984	4.112	−0.009	1.208	0.022	1.204	0.002	1.610	0.040	1.714
d=2										
$n_1 = 500$	8.040	8.487	0.147	3.913	0.080	3.915	0.033	5.206	0.393	5.567
$n_1 = 1000$	8.022	8.253	0.084	2.787	0.030	2.740	0.034	3.633	0.217	3.867
$n_1 = 4000$	7.998	8.055	−0.002	1.382	−0.001	1.376	−0.000	1.857	0.045	1.975
d=3										
$n_1 = 500$	12.032	12.316	0.341	5.312	0.348	5.237	0.009	6.894	0.890	7.317
$n_1 = 1000$	12.022	12.168	0.150	3.738	0.032	3.735	−0.077	4.884	0.340	5.167
$n_1 = 4000$	11.980	12.017	0.005	1.917	0.024	1.916	−0.025	2.558	0.059	2.721

methods. It is only necessary that the primary and placebo outcomes are measured on the same units within each cross-section.

A first tack at this question would compare the efficacy of attendees and non-attendees of these events after the treatment, and we find that the difference-in-means is 0.0733 (0.0247).[3] Although all respondents signed up for an event date that was convenient for them, attendance was not randomized and we expect those who attended these events to differ from those who do not in important respects. Those who have less interest in politics or have lower efficacy and think that the event will have smaller benefits for them are less likely to take the time to travel and participate in these events. Indeed, differences between attendees and non-attendees in efficacy at baseline provide some evidence for this confounding problem. At baseline, the mean efficacy index were 0.5669 for attendees and 0.5345 for non-attendees, respectively (Table 4). This gives us a difference-in-difference estimate of 0.0408 (0.0231), an apparent effect significant at the 90% level.

[3]Our political efficacy index is a measure of both external and internal efficacy. External efficacy, the sense that one can influence politics and government, is measured on a 5-point scale from strongly agree to strongly disagree with the statement "I feel I can influence political decisions that affect my life." Internal political efficacy, which is the sense that one can understand political affairs, is measured in two ways. The first is on a 5-point scale from strongly agree to strongly disagree with the statement "I feel I am as well-informed about politics and government as most people." The second is whether the respondent agrees more with the statement "Politics is complicated and I usually do not understand what politicians are doing," or "Most of the time I understand what politicians are doing." These items are rescaled so that each has a minimum of 0 and maximum of 1 and averaged to generate an index of political efficacy.

Furthermore, even if we use the more robust Athey and Imbens (2006)'s CiC model we get an ATT estimate of 0.0448 (0.0245, standard error from 1000 bootstrap samples).[4] It appears from these estimates that debates have an effect on efficacy.

Table 4: Pre- and Post-Treatment Data from Nepal Debate Study

Number of attendees (treated)	223
Number of non-attendees (control)	510
Post-treatment:	
Political efficacy index (treated)	mean 0.5643, min 0, max 1
Political efficacy index (control)	mean 0.4910, min 0, max 1
Knowledge index (treated)	mean 3.4215, min 0, max 11
Knowledge index (control)	mean 2.9980, min 0, max 12
Pre-treatment:	
Political efficacy index (treated)	mean 0.5669, min 0.0833, max 1
Political efficacy index (control)	mean 0.5345, min 0, max 1
Knowledge index (treated)	mean 3.1300, min 1, max 14
Knowledge index (control)	mean 2.7922, min 1, max 13

However, even the CiC estimate is only valid if the quantile-quantile pre-post equiconfounding assumption holds on average (Assumption 2). Fortunately, we have a placebo outcome that can be used to assess this assumption. This is an index of knowledge of aspects of politics and government that were *not* discussed in the debate, which is likely to suffer from confounding similar to political efficacy. Respondents were asked how many levels of government Nepal has under the new constitution, how many legislative bodies Nepal has at the national level under the new constitution, and asked to list as many as they could of the responsibilities and power of local governments under the new constitution.[5] The debates were held after the local government elections and featured only candidates for the federal-level House of Representatives. They did not mention the elections for the provincial-level State Assemblies being held concurrently or the federal-level National Assembly which were to be held later. Therefore, we expect the debates or discussion to not affect knowledge on these particular items, unlike for knowledge on candidate platforms and other information that were presented during the debates. Political efficacy and knowledge are closely related, since those with less interest in politics or lower efficacy are less likely to seek out information, pay attention to information, or participate in activities that would expose them to information that they don't expect to understand well or find useful.

A DiD analysis on knowledge constitutes a classic placebo test. It appears that we "pass" this placebo test with a statistically insignificant estimate of 0.0856 (0.1855). Although these tests

[4] To apply this model, we add a small amount of random noise (0.04σ) to break ties and create a more continuous measure of our outcomes. Then for each attendee, we determine at what quantile this pre-treatment efficacy level would fall in the distribution of pre-treatment efficacy for non-attendees.

[5] Each correct answer is given one point, with a possible maximum of 16.

are widely used, it is unclear how well they protect us against unmeasured confounding. In a recent advance, Hartman and Hidalgo (2018) have proposed moving away from null hypotheses of no difference in variables that should be unaffected by the treatment, and instead testing a null hypothesis of difference in those variables against an alternative hypothesis of equivalence between treated and control groups.

Instead of using placebo outcomes to test for the validity of the research design, Sofer et al. (2016)'s negative outcome control (NOC) approach uses a placebo outcome that is assumed to have the same confounding as the primary outcome on the quantile scale to correct for the confounding. The Sofer et al. (2016) procedure is represented in Figure 3. We would take the knowledge index level for an attendee n_{11} and find its quantile amongst the distribution of knowledge for the non-attendees. Then the imputed counterfactual outcome y^{NOC} for the attendee with n_{11} would be the value at that same quantile in the distribution of political efficacy for the non-attendees. Attendees with knowledge index values greater (smaller) than that observed amongst non-attendees will be assigned the largest (smallest) efficacy value observed amongst non-attendees as their counterfactual efficacy values. The average of the differences between the observed outcomes and these counterfactual outcomes for the attendees is the NOC approach's estimated ATT, 0.0040 (0.0287).

This null finding contradicts our earlier results, which emphasizes the importance of assessing whether the quantile-quantile primary-placebo equi-confounding assumption for the Sofer et al. (2016) procedure is reasonable. This assumption is visualized in Figure 4. For example, an attendee whose knowledge is at the median for attendees ($q = .5$) might be at the 70th percentile for knowledge for non-attendees ($\nu = .7$). Would this attendee be at the 70th percentile for efficacy for non-attendees if he had not attended the debate event ($\nu' = .7$)? If efficacy is more strongly associated with attendance than is knowledge, then this attendee's counterfactual efficacy might be greater than the 70th percentile for efficacy amongst non-attendees. But if a general curiosity means that an attendee is more willing to make time for events like the debate screening, there may be a stronger association between knowledge and attendance at a debate event than efficacy and attendance, then this attendee might be at lower than the 70th percentile.

Table 5: Summary of Results

	Estimate	S.E.
Mean difference in post-treatment political efficacy (Y)	0.0733	(0.0247)
Linear DiD on political efficacy (Y)	0.0408	(0.0231)
CiC on political efficacy (Y)	0.0448	(0.0245)
Linear DiD on knowledge (N)	0.0856	(0.1855)
Post-treatment NOC	0.0040	(0.0287)
Pre-treatment NOC	-0.0031	(0.0249)
Post–Pre NOC	0.0071	(0.0336)
NOCNOC	0.0258	(0.0387)

We can relax the quantile-quantile primary-placebo equi-confounding assumption (Assumption 4) by moving to a DiDiD context and using both pre- and post-treatment data. That is, we could assume that there is differential confounding for efficacy and knowledge, but that the differences in confounding are time-invariant on the scale of the efficacy index (Assumption 13). There is no particular reason for us to expect that the pattern of confounding would change from before to after the debate events. Almost all respondents from the pre-treatment survey were recontacted post-treatment, and the data collection was separated by only two months. This assumption may not hold if, for example, confounding on knowledge is unaffected but there is less confounding on efficacy post-treatment because individuals with low baseline efficacy, who are disproportionately in the non-attendee group, get a boost in efficacy from experiencing the national election even if they did not attend the debate events. However, if this assumption of time-invariant differential primary-placebo confounding on the outcome scale is reasonable, then we can simply take the difference between the NOC ATT estimate for the post-treatment data and the NOC ATT estimate for the pre-treatment data. The pre-treatment NOC ATT estimate is -0.0031 (0.0249, standard error from 1000 bootstrap samples), and this DiDiD method gives us a post–pre NOC difference of 0.0071 (0.0336, standard error from 1000 bootstrap samples). The apparent effect from the original DiD and CiC analyses is not robust, and we conclude that exposure to debate screenings had no effect on political efficacy in this study (Table 5).

We can also use our NOCNOC estimator by further relaxing the assumption of time-invariant differential primary-placebo confounding to be on the quantile scale rather than on the scale of the outcome. The NOCNOC procedure is visualized with the blue line segments in Figure 11 in earlier Section 5. For each attendee's pre-treatment efficacy value, we determine at what quantile in the distribution of the pre-treatment efficacy for the non-attendees it would fall. Then we find the knowledge level associated with that quantile, and then where this knowledge level would be in the distribution of pre-treatment knowledge for attendees. This second quantile then becomes the starting point for the second portion of the procedure, which is the NOC procedure. There is no treatment effect in the pre-treatment period, so we are effectively backing out the confounding on the quantile scale through the first portion of the procedure to carry through in the second portion. By subtracting the mean NOCNOC estimates for the counterfactual outcomes from the mean post-treatment efficacy for the attendees, we get an ATT estimate of 0.0258 (0.0387, standard error from 1000 bootstrap samples). This again indicates that exposure to debate screenings had no effect on political efficacy.

8 Conclusion

In this paper, we discussed the extension of DiD and DiDiD techniques with two different types of outcome variables, placebo (negative) outcomes and surrogate outcomes, and we extended the Athey and Imbens (2006) and Sofer et al. (2016) approaches in three ways. First, we demonstrated

the use of placebo outcomes in the DiD context for the estimation of the average treatment effect on the controls (ATC). Second, we showed that the techniques used with placebo outcomes can also be used in reverse for surrogate (proxy) outcomes. Third, we showed how these techniques can be extended to the DiDiD context and demonstrated this in detail using placebo (negative) outcomes. In the application, we applied these methods to analyze whether exposure to candidate debates discussions affected the political efficacy of Nepalese citizens, using political knowledge as a placebo outcome. Using pre-treatment measures of both political efficacy and knowledge, we found that exposure to candidate debates likely had no effect on political efficacy, even though traditional DiD analysis with placebo tests in conventional practice suggested a positive effect.

Two additional discussions will be part of future work. First, for the DiD techniques, covariates can be included in the manner described in Sofer et al. (2016), and next steps include extending this approach to the DiDiD. Second, it would be useful to weaken the support assumptions for these techniques in the manner prescribed in Athey and Imbens (2006) given potential problems with these assumptions.

A Mathematical Derivations

Most of the results in this paper follow directly from the proofs in Sofer et al. (2016). However, the results regarding averages in Assumptions 2, 4, 6, 8, 10, 12, and 13 require some elaboration. Because the proofs corresponding to Assumptions 2, 4, 6, 8, 10, 12 are analogous, we present only the proof for Assumption 4.

A.1 NOC proof under Assumption 4

Recall that the post-treatment NOC estimator can be written as the following:

$$\widehat{E}[Y_{11}] - \frac{1}{k_{11}} \sum_{i=1}^{k_{11}} \hat{F}^{-1}_{Y_{01}}(\hat{F}_{N_{01}}(N_{11,i}))$$

In large samples, this estimator can be written as $E[Y_{11}] - E_q[F^{-1}_{Y_{01}}(F_{N_{01}}(F^{-1}_{N_{11}}(q)))]$ where $q \sim Unif(0,1)$. Under Assumption 4, this equals $E[Y_{11}] - E_q[F^{-1}_{Y_{11}(0)}(q)]$ where $q \sim Unif(0,1)$. Finally note that when $q \sim Unif(0,1)$, $E_q[F^{-1}_{Y_{11}(0)}(q)] = E[Y_{11}(0)]$.

A.2 Post-Pre NOC proof under Assumption 13

Following the NOC proof above, under Assumption 13, the post-treatment NOC estimator identifies the following (where $q \sim Unif(0,1)$ throughout):

$$\begin{aligned}
&E[Y_{11}] - E_q[F^{-1}_{Y_{01}}(F_{N_{01}}(F^{-1}_{N_{11}}(q)))] \\
&\quad = E[Y_{11}] - (E_q[F^{-1}_{Y_{11}(0)}(q)] + E_q[\delta(q)_1]) \\
&\quad = E[Y_{11}] - E[Y_{11}(0)] - E_q[\delta(q)_1]
\end{aligned}$$

while the pre-treatment NOC estimator identifies the following:

$$\begin{aligned}
&E[Y_{10}] - E_q[F^{-1}_{Y_{00}}(F_{N_{00}}(F^{-1}_{N_{10}}(q)))] \\
&\quad = E[Y_{10}] - (E_q[F^{-1}_{Y_{10}(0)}(q)] + E_q[\delta(q)_0]) \\
&\quad = E[Y_{10}] - E[Y_{10}(0)] - E_q[\delta(q)_0] \\
&\quad = E[Y_{10}] - E[Y_{10}] - E_q[\delta(q)_0] \\
&\quad = E_q[\delta(q)_0]
\end{aligned}$$

Note that the second to last equation here is due to the fact that treatment is not applied in the pre-treatment time period. Finally, under Assumption 13, $E_q[\delta(q)_1] = E_q[\delta(q)_0]$ and hence the difference between these two estimators will produce the ATT.

A.3 NOCNOC proof under Assumption 14a and 14b

We can rewrite Assumption 14a in the following way:

$$
\begin{aligned}
F_{N_{11}}(F^{-1}_{N_{01}}(F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q)))) &= F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q)))), q \in [0,1] \\
F^{-1}_{N_{01}}(F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q))) &= F^{-1}_{N_{11}}(F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q))))), q \in [0,1] \\
F_{Y_{01}}(F^{-1}_{Y_{11}(0)}(q)) &= F_{N_{01}}(F^{-1}_{N_{11}}(F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q)))))), q \in [0,1] \\
F^{-1}_{Y_{11}(0)}(q) &= F^{-1}_{Y_{01}}(F_{N_{01}}(F^{-1}_{N_{11}}(F_{N_{10}}(F^{-1}_{N_{00}}(F_{Y_{00}}(F^{-1}_{Y_{10}}(q))))))), q \in [0,1]
\end{aligned}
$$

where the right hand side of the last equation corresponds to the second term in the NOCNOC estimator. If we further have the support conditions in Assumption 14b, then all observed values of Y_{10} will produce non-infinite values of this expression.

Additionally, note that Assumption 14a can be microfounded in terms of unmeasured confounders as in the Sofer et al. (2016) proof:

Assumption 15a.

$$
\begin{aligned}
N_{at}(a) &= N_{at} \text{ for } a = 0,1 \\
Y_{at}(a) &= Y_{at} \text{ if } A = a
\end{aligned}
$$

Assumption 15b.

$$
\begin{aligned}
A_t &\perp\!\!\!\perp Y_t(0) | U_t \text{ for } t = 0,1 \\
A_t &\perp\!\!\!\perp N_t | W_t \text{ for } t = 0,1
\end{aligned}
$$

Assumption 15c.

$$
\begin{aligned}
Y_{at}(0) &= h_{yt}(U_{at}) \text{ where } h_{yt}(U_{at}) \text{ is monotone increasing} \\
N_{at} &= h_{nt}(W_{at}) \text{ where } h_{nt}(W_{at}) \text{ is monotone increasing}
\end{aligned}
$$

Assumption 15d.

$$
F_{W_{11}}(F^{-1}_{W_{01}}(F_{U_{01}}(F^{-1}_{U_{11}(0)}(q)))) = F_{W_{10}}(F^{-1}_{W_{00}}(F_{U_{00}}(F^{-1}_{U_{10}}(q)))), q \in [0,1]
$$

If Assumptions 15a, 15b, 15c, and 15d hold, then the proof in Sofer et al. (2016) implies that Assumption 14a holds.

B Simulation Details

For Table 1, we generated data from the following model:

$Y_{at} = (U_{at} + 1)^2 + a\beta$, where $U|A \sim N(\eta_0 + \delta t + 2a, \frac{3-a}{2})$

$N_{at} = \sqrt{W_{at}}$, where $W|A \sim N(\eta_0 + \delta t + a, \frac{3-a}{2})$

for $a = \{0, 1\}$, $t = \{0, 1\}$. The treatment effect is β and additive. Y and N are each functions of time (pre/post-treatment) and an unobserved confounder, and both strictly monotonically increasing in the confounder. U and W are the unobserved confounders for the treatment with the primary and placebo outcomes, respectively. The confounding bias across treatment groups, over time, and across outcomes is given by the differences in the means of U and W. In this set up, U and W are the same when $a = 0$ and $t = 0$ (control group, pre-treatment) and diverge with a and t. The standard deviations of the unmeasured confounders are set to be larger for the controls than for the treated group to ensure support.

We set the treatment effect at $\beta = 1$ and set $\eta_0 = 9$ and $\delta = \{0, 1, 2, 3\}$ to ensure that the support conditions hold. We generate data for $n_1 = 500$, 1000, and 4000 treated observations and $n_0 = 2n_1$ control observations. We compare bias and RMSE of the linear differences-in-differences estimator, the CiC estimator, the NOC estimator, the Post - Pre NOC estimator, and the NOCNOC estimator, with 5000 simulations.

The data for Table 2 are generated by the model:

$Y_{at} = (U_{at} + 1)^2 + a\beta$, where $U|A \sim N(\eta_0 + a(1 + t), \frac{3-a}{2})$

$N_{at} = \sqrt{W_{at}}$, where $W|A \sim N(\eta_0 + \delta + a(1 + t), \frac{3-a}{2})$

with $a = \{0, 1\}$, $t = \{0, 1\}$. As for the first simulation study, we set the treatment effect at $\beta = 1$ and set $\eta_0 = 9$ and $\delta = \{0, 1, 2, 3\}$ to ensure that the support conditions hold. We generate data for $n_1 = 500$, 1000, and 4000 treated observations and $n_0 = 2n_1$ control observations for 5000 simulations.

The data for Table 3 are generated as in the first model, except that $\eta_0 = 13$ and U and W have standard deviation of 1 when $t = 1$ and 1.7 when $t = 0$. This ensures positivity and support for the CiC-r estimator. We consider $\delta = \{0, 1, 2, 3\}$ and generate data for $n_1 = 500$, 1000, and 4000 treated observations and $n_0 = 2n_1$ control observations for 5000 simulations.

References

Almond, Gabriel A. and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Princeton: Princeton University Press.

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

Bidwell, Kelly, Katherine Casey and Rachel Glennerster. 2016. "Debates: Voting and Expenditure Responses to Political Communication." Working paper.

Brierley, Sarah, Eric Kramon and George Ofosu. 2018. "The Moderating Effect of Debates on Political Attitudes." Working paper.

Campbell, Angus, Gerald Gurin and Warren E. Miller. 1954. *The Voter Decides*. Evanston, IL: Row, Peterson, and Co.

Finkel, Steven E. 1985. "Reciprocal Effects of Participation and Political Efficacy: A Panel Analysis." *American Journal of Political Science* 29(4):891–913.

Hartman, Erin and F Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." Working paper.

Platas, Melina and Pia Raffler. 2017. "Meet the Candidates: Information and Accountability in Ugandan Elections." Working paper.

Rosenstone, S. J. and M. J. Hansen. 1993. *Mobilization, Participation, and Democracy in America*. New York: MacMillan.

Sofer, Tamar, David B. Richardson, Elena Colicino, Joel Schwartz and Eric J. Tchetgen Tchetgen. 2016. "On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences." *Statistical Science* 31(3):348 – 361.

Valentino, Nicholas A., Krysha Gregorowicz and Eric W. Groenendyk. 2009. "Efficacy, Emotions and the Habit of Participation." *Political Behavior* 31(3):307–330.

VanderWeele, Tyler J. 2013. "Surrogate Measures and Consistent Surrogates." *Biometrics* 69:561 – 581.

Verba, Sidney, Kay Lehman Schlozman and Henry Brady. 1995. *Voice and equality: Civic voluntarism in American politics*. Cambridge: Harvard University Press.