Generalized Difference-in-Differences

David B. Richardson,[a] Ting Ye,[b] and Eric J. Tchetgen Tchetgen[c]

Abstract: Difference-in-differences (DID) analyses are used in a variety of research areas as a strategy for estimating the causal effect of a policy, program, intervention, or environmental hazard (hereafter, treatment). The approach offers a strategy for estimating the causal effect of a treatment using observational (i.e., nonrandomized) data in which outcomes on each study unit have been measured both before and after treatment. To identify a causal effect, a DID analysis relies on an assumption that confounding of the treatment effect in the pretreatment period is equivalent to confounding of the treatment effect in the post treatment period. We propose an alternative approach that can yield identification of causal effects under different identifying conditions than those usually required for DID. The proposed approach, which we refer to as generalized DID, has the potential to be used in routine policy evaluation across many disciplines, as it essentially combines two popular quasiexperimental designs, leveraging their strengths while relaxing their usual assumptions. We provide a formal description of the conditions for identification of causal effects, illustrate the method using simulations, and provide an empirical example based on Card and Krueger's landmark study of the impact of an increase in minimum wage in New Jersey on employment.

Keywords: regression analysis; cohort studies; instrumental variables; unmeasured confounding

Difference-in-differences (DID) analyses are widely used in a variety of research areas including economics, public policy, and public health.[1] The approach offers a strategy for estimating the causal effect of a policy, program, intervention, or environmental hazard (hereafter, treatment) on an outcome of interest outside of a randomized trial design. In observational settings, a DID analysis can sometimes be used to obtain unbiased comparisons of outcomes between treatment groups even when those groups are not balanced with respect to unmeasured determinants of the outcome. Specifically, to identify a causal effect in such settings, a DID analysis relies on an assumption that confounding of the treatment effect in the pretreatment period is equivalent to confounding of the treatment effect in the post treatment period. This condition is sometimes referred to as the "parallel trends assumption" and remains a challenging, but necessary, condition for valid inference in DID analyses.[2,3]

Here, we propose an alternative approach that can yield the identification of causal effects under different identifying conditions than those usually required for DID. Observations in the pretreatment period provide information on covariate-outcome associations in a setting where the treatment is set to 0, that is, to its control value; we use that information to repurpose a measured confounder of the association of interest as a "bespoke" instrumental variable,[4] yielding a consistent estimator of the treatment effect in the posttreatment period. We focus on a setting in which a DID analysis might typically be undertaken, where outcomes on each study unit have been measured both before and after treatment. While the assumptions necessary for identification of causal effects in observational studies often may not hold perfectly, access to alternative approaches that can yield identification of causal effects under different identifying conditions can help investigators to triangulate evidence and undertake potentially informative comparative sensitivity analyses.

METHODS

Let $A(i)$ denote treatment status, where $A(i)=1$ if individual i is treated, $A(i)=0$ otherwise. Let $Y(i,t)$ be the outcome of interest for individual i at time t, where a population is observed in two periods: a pre-treatment period, $t=t_0$; and, a post-treatment period, $t=t_1$. Let $Z(i,t)$ and $U(i,t)$ denote measured and unmeasured variables, respectively, that may confound associations between $A(i)$ and $Y(i,t)$. Z may denote a vector of measured variables, $Z_1,\ldots,Z_p$, and similarly U may denote a vector of unmeasured variables $U_1,\ldots,U_Q$. For convenience, we use the term "individual" to refer to observed study units; however, the methods discussed apply equally to

observations on aggregated units (such as employers, counties, or census tracts).

We define the causal effect of interest in terms of potential outcomes. Let $Y^a(i,t)$ denote individual i's potential outcome at time t if $A(i)$ were set, possibly contrary to fact, to a. The effect of the treatment on the outcome for individual i at time t is then $Y^1(i,t) - Y^0(i,t)$. One cannot observe both potential outcomes $Y^1(i,t)$ and $Y^0(i,t)$ for a given individual i at time t and therefore one cannot compute individual treatment causal effects. Here, focusing on the post-treatment period, t_1, we are interested in estimating an average effect of treatment on the treated (ATT), defined as $E[Y^1(i,t_1) - Y^0(i,t_1)|A(i)=1]$. Additional assumptions are needed to identify the average effect of treatment on the total population.

Given our focus on average causal effects, we drop the individual argument i to simplify notation. In subsequent discussion we assume the following conditions hold:

(1) Consistency for the treated, $Y(t_1) = Y^a(t_1)$, if $A=a$;
(2) Positivity (i.e., a small constant $c>0$, such that for any z such that $\Pr(Z=z|A=1)>c$ it must be that $\Pr(Z=z|A=0)>c$; and,
(3) No anticipation of future treatment (i.e., at t_0 individuals do not anticipate the treatment received at t_1), such that $E[Y(t_0)|Z] = E[Y^a(t_0)|Z]$, for all a.

We first describe a standard difference-in-differences approach to identify the ATT. Then, we describe our proposed generalized difference-in-differences approach to identify the ATT.

Standard Difference-In-Differences

In a standard DID analysis, among all individuals the pre-treatment outcome is subtracted from the posttreatment outcome. Specifically, the pretreatment outcome among the treated is subtracted from the posttreatment outcome among the treated; in addition, the pretreatment outcome among the untreated is subtracted from the post-treatment outcome among the untreated. The difference of these differences between the treated and the untreated identifies the ATT,

$$ATT_{DID} = E[Y(t_1) - Y(t_0)|A = 1] - E[Y(t_1) - Y(t_0)|A = 0],$$

that is typically justified by conditions 1–3 above as well as the parallel trends assumption,

$$E[Y^0(t_1) - Y^0(t_0) | A = 1] = E[Y^0(t_1) - Y^0(t_0) | A = 0].$$

The parallel trends assumption, which we ordinarily require for the DID estimand to identify ATT, implies no unmeasured time-varying confounders (i.e., any factor that causes a trend in the outcome over time is independent of treatment, A). The standard DID approach allows that there may be changes in the outcome between the pre- and post-treatment periods for reasons other than treatment. Among the untreated, the outcome may vary from the pre-treatment period to the post-treatment period despite there being no treatment applied. By subtracting the temporal change in the outcome

among the untreated, $E[Y(t_1) - Y(t_0)|A=0]$, from the temporal change in the outcome among the treated, $E[Y(t_1) - Y(t_0)|A=1]$, the DID estimand accounts for change in the outcome over time that is independent of treatment, A.

The standard DID approach also allows that, within each period, there may be confounding of the association between treatment and outcome by measured or unmeasured subject-specific characteristics (Figure A). In the pretreatment period, an association between treatment and outcome may be observed despite there being no causal effect of treatment on outcomes in the pretreatment period; such an association would arise due to confounding. For DID to yield a consistent estimator of the causal effect of treatment on posttreatment outcomes, we ordinarily assume that any confounding bias that occurs in the posttreatment period is exactly equal to the association between the treatment and the pre-treatment outcome mean. Sofer et al. noted that the parallel trends assumption is equivalent to the "additive equi-confounding" assumption

A Parallel trends
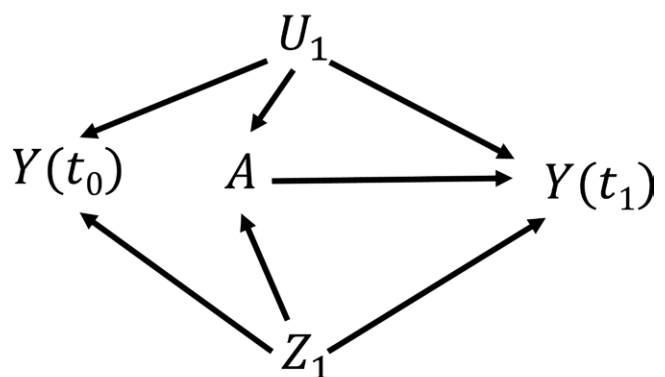
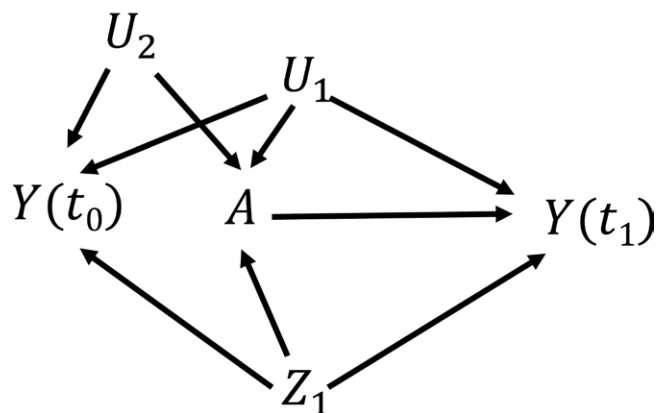**B** Violation of parallel trends

FIGURE. (A) Parallel trends. (B) Violation of parallel trends.

described in the literature on negative controls, noting that the pre-exposure outcome is a negative control outcome that cannot be influenced by subsequent exposure.[5] Differencing post- and pretreatment outcomes induces mean independence between any confounder whose effect on the outcome is constant over time and the outcome, $Y(t_1) - Y(t_0)$. Therefore, the standard DID estimand is unconfounded by any factor whose effect on the outcome is constant (given everything else being used for adjustment) but is susceptible to confounding by a factor whose effect on the outcome varies over time.

Generalized Difference-In-Differences

Suppose that some unmeasured confounder of the association between treatment and pre-treatment outcomes differs in the post-treatment period (Figure B) such that the parallel trends assumption does not hold because there is a time-varying pretreatment cause of the outcome that is associated with treatment. We may also allow that some measured confounders violate the parallel trends assumption. We now introduce an alternative identifying condition of the average effect of treatment on the treated without invoking the usual DID assumption of parallel trends.

To ground ideas, suppose that in addition to treatment and pre- and post-treatment outcomes, one has observed a confounder Z. Our proposed generalized DID repurposes a measured confounder, Z, of the association of interest as a bespoke instrumental variable,[4] replacing the standard DID ATT_{DID}, with an alternative that we refer to as a generalized difference-in-differences (GDID) estimand,

$$ATT_{GDID} = E\left[Y(t_1) - Y(t_0) \,|\, t(Z) = 1\right]$$
$$-E\left[Y(t_1) - Y(t_0) \,|\, t(Z) = 0\right]$$

where $t(Z) = E(A|Z)$ and $E\left[Y(t_1) - Y(t_0) \,|\, t(Z)\right]$ are assumed to be linear in Z and $t(Z)$ respectively. These linear models should be interpreted as working models only used to estimate the ATT_{GDID} under formal counterfactual assumptions; the GDID approach is justified by conditions formalized below.

A two-stage least squares approach can be used to estimate ATT_{GDID}:

Stage 1: We first obtain the predicted value of A given Z,

$$\hat{A}(Z) = \hat{E}(A|Z)$$

by fitting a linear regression of A on Z via ordinary least squares (OLS).

Stage 2: Then, via OLS, we fit a linear regression of $Y(t_1) - Y(t_0)$, on $\hat{A}(Z)$,

$$\widehat{ATT}_{GDID} = E(Y(t_1) - Y(t_0) \,|\, \hat{A}(Z)) = \hat{\beta}_0 + \hat{\beta}_1^{IV}\hat{A}(Z)$$

In the eAppendix, http://links.lww.com/EDE/B983, we provide illustrative SAS and R code to implement our proposed approach for the estimation of ATT and associated 95% confidence intervals derived using a recommended robust variance estimator. The estimated parameter, $\hat{\beta}_1^{IV}$, is a consistent

estimator of the average casual effect of interest, under conditions formalized below, even if unobserved variables are time-varying confounders of the association of interest.

Identification

Suppose that $Z = (Z_1, Z_2)$ and instead of taking all of Z as a candidate bespoke instrumental variable, we take Z_1 only as a bespoke instrumental variable, while Z_2 are additional measured covariates that we adjust for.

In this section, we establish formal identification of ATT under conditions 1–3 above, as well as the following conditions:

(4) Z_1 is relevant for predicting treatment: $E[A|Z_1, Z_2]$ depends on Z_1;

(5) No interaction between A and Z_1 in causing $Y^a(t_1)$ conditional on Z_2 and $A = 1$, such that

$$E\left[Y^{a=1}(t_1) - Y^0(t_1)|A = 1, Z_1, Z_2\right]$$
$$= E[Y^{a=1}(t_1) - Y^0(t_1)|A = 1, Z_1 = 0, Z_2]; \text{ and,}$$

(6) The additive association between Z_1 and pre-treatment outcomes is equal to the additive association between Z_1 and posttreatment outcomes (in the absence of treatment):

$$E\left[Y^0(t_0)|Z_1, Z_2\right] - E\left[Y^0(t_0)|Z_1 = 0, Z_2\right]$$
$$= E\left[Y^0(t_1)|Z_1, Z_2\right] - E\left[Y^0(t_1)|Z_1 = 0, Z_2\right]$$

Result: Under conditions 1–6 we have that for all $z_1 \neq 0$

$$E\left[Y^1(t_1) - Y^0(t_1)|A = 1, Z_1 = z_1, Z_2\right] =$$
$$\frac{E[Y(t_1)-Y(t_0)|Z_1=z_1,Z_2]-E[Y(t_1)-Y(t_0)|Z_1=0,Z_2]}{E[A|Z_1=z_1,Z_2]-E[A|Z_1=0,Z_2]}; \text{ and therefore,}$$

$$E\left[Y^1(t_1) - Y^0(t_1)|A = 1\right]$$
$$= E\left[\frac{E\left[Y(t_1) - Y(t_0) \,|\, Z_1 = z_1, Z_2\right] - E[Y(t_1)-Y(t_0)|\,Z_1 = 0, Z_2]}{E[A|\,Z_1 = z_1, Z_2] - E[A|\,Z_1 = 0, Z_2]}\Big|A = 1\right]$$

Proof of the result is given in eAppendix 1, http://links.lww.com/EDE/B983. We refer to the identifying formula obtained in the result as GDID. Under linear model specifications

$$E\left[Y^a(t_1) - Y^0(t_1)|A = a, Z\right] = b_1 a, \quad E\left[Y^0(t_1)|Z_2\right] = d_0 + d_1 Z_2;$$

in eAppendix 1, http://links.lww.com/EDE/B983, we show that the standard two-stage least squares approach described in the previous section, further adjusted for Z_2 in both stages, obtains a consistent estimator of $b_1 = E\left[Y^1(t_1) - Y^0(t_1)|A = 1\right]$.

Condition 4, which states that Z_1 is associated with treatment, A, holds by definition when Z_1 is a confounder of the association of interest. Condition 5 is analogous to a no-interaction assumption routinely made in the instrumental variable setting.[6,7] This holds by definition under the null hypothesis of no conditional effect of treatment in the treated; it follows that the proposed approach produces a valid test of the sharp causal null hypothesis provided that the remaining assumptions hold. Condition 6 can essentially be interpreted

as a Z_1 parallel trend assumption conditional on Z_2; it states that the additive association between Z_1 and $Y^0(t_0)$ is equal to the additive association between Z_1 and $Y^0(t_1)$. Standard DID fails if there is an unmeasured time-varying confounder (i.e., a violation of the parallel trends assumption). Condition 6 is a modest alternative identifying condition: it holds if we can identify a measured covariate, Z_1, whose association with the treatment-free outcome does not vary over times t_0, t_1. When condition 6 holds, $Y^0(t_1)-Y^0(t_0)$, by definition, is mean independent of Z_1; this permits our bespoke instrumental variable[4] approach to the estimation of ATT.

An intuition for our identification result (suppressing Z_2 here) may follow by noting that the change in observed outcomes between t_0 and t_1 conditional on $Z_1 = z_1$ is the result of (1) the change in untreated potential outcomes between t_0 and t_1; (2) the causal effect of the treatment at t_1 on the treated; and, (3) the proportion of treated units, $P(A = 1|Z_1 = z_1)$. Under our identifying conditions, the change in untreated potential outcomes conditional on $Z_1 = z_1$ is equal to the change in untreated potential outcomes conditional on $Z_1 = 0$. Therefore, to recover the causal effect of the treatment at t_1 for treated units at $Z_1 = z_1$ we just need to subtract off the change in observed outcomes among untreated units conditional $Z_1 = 0$ and then account for the proportion of treated units, $P(A = 1|Z_1 = z_1)$.

Simulation

We simulated data for 1,000 studies, with 5,000 people in each study sample, with people observed in pre- ($t=0$) and posttreatment ($t=1$) periods. We generated simulations for two scenarios. In the first scenario, which conformed to the "parallel trends" assumption (Figure A), we generated a measured covariate, denoted Z_1, and an unmeasured covariate denoted U_1. Z_1 and U_1 were random binary variables, both taking values of 1 with a probability of 0.5. We assigned A as a random binary variable that took a value of 1 with probability $1/(1+\exp(-(-0.1 -0.5 \times U_1 + \gamma_z \times Z_1)))$. We considered the case when Z_1 is strongly ($\gamma_z =2$), moderately ($\gamma_z =1$), and weakly ($\gamma_z =0.4$) associated with A. The pre-treatment outcome variable, $Y(t=0)$, took a value of $(1+1\times U_1 +1\times Z_1 +\varepsilon)$, where $\varepsilon \sim N(0,1)$; and, the post-treatment outcome variable, $Y(t=1)$, took a value of $(1+1\times U_1 +1\times Z_1 +1\times A+\varepsilon)$. In the second scenario, which violated the parallel trends assumption (Figure B), we generated an additional covariate, denoted U_2, that was a continuous variable assigned by sampling from a normal $(0,1)$ distribution. We assigned A as a random binary variable that took a value of 1 with probability $1/(1+\exp(-(-0.1 -0.5 \times U_1 -0.5 \times U_2 + \gamma_z \times Z_1)))$. The pretreatment outcome, $Y(t=0)$, took a value of $(1+1\times U_1 +1\times U_2 +1\times Z_1 +\varepsilon)$; and, the posttreatment outcome, $Y(t=1)$, took a value of $(1+1\times U_1 +1\times Z_1 +1\times A+\varepsilon)$.

We used the GDID method described in this paper to obtain an estimate of the average change in $Y(t=1)$ with A by a two-stage regression. In all simulations, in the first-stage model Z_1 was the measured variable used to predict A in linear regression; then, in the second stage, we fitted a linear

regression model for $Y(t=1)-Y(t=0)$ as a function of the predicted value of A given Z_1. Following the influential work by Bound et al.,[8] we use the rule of thumb that the first-stage F-statistic needs to be larger than 10 for the usual asymptotic inference to be reliable.[9] For comparison, we fitted a difference-in-differences model; a linear regression model was fitted to each simulated cohort for $Y(t=1)-Y(t=0)$ as a function of A. We summarized results from the simulated studies by computing the Monte Carlo mean and Monte Carlo standard deviation (SD) of the estimates, square root of the mean of squared difference between the estimated associations, and the specified true effect of A on Y (the root mean squared error, RMSE), average of standard errors (SEs), and coverage probability (CP) of 95% confidence intervals from normal approximation. We also reported the Monte Carlo mean of the first-stage F-statistic. The eAppendix, http://links.lww.com/ EDE/B983, reports results of additional simulations in which: i) U_1 was associated with Z_1; and, ii) U_2 was associated with Z_1 (i.e., violating the condition of Z_1 additive equi-confounding). R code for the GDID method and for reproducing numerical examples can be found in the supplementary material, http:// links.lww.com/EDE/B983.

Empirical example 1: Card and Krueger

This empirical example is based on Card and Krueger's landmark study[10] of the impact on employment of an increase in minimum wage in New Jersey (NJ). In early 1990 the NJ legislature increased the state minimum wage to \$5.05 per hour effective April 1, 1992. Card and Krueger surveyed fast-food restaurants (Burger King, KFC, Wendy's, and Roy Rogers chains) in NJ and eastern Pennsylvania before the increase in the minimum wage (February 15–March 4, 1992, denoted $t=0$) and after the increase in the minimum wage (November 5– December 31, 1992, denoted $t=1$). The survey was conducted primarily by telephone and included questions on employment, starting wages, the price of a full meal (medium drink, small fries, and an entree), and other store characteristics. The outcome variable of primary interest, denoted $Y(t)$, is employment per store measured in full-time-equivalent and calculated as the number of full-time workers (including managers) plus 0.5 times the number of part-time workers. The treatment variable of primary interest states, coded 1 for NJ, else 0. First, we fitted a standard difference-in-differences model; a linear regression model was fitted for $Y(t=1)-Y(t=0)$ as a function of *state*. Next, we fitted our proposed GDID. The "bespoke IV" was price of a full meal. The variable was chosen both because it was relevant for predicting treatment (i.e., *state*), and because it is reasonable to posit that the association between full meal price and employment levels before the increase in the NJ minimum wage is likely to be approximately equal to the association between full meal price and employment levels in the posttreatment period (had the NJ minimum wage not changed), as the two associations are ascertained within a relatively short period (approximately 9 months). In a first-stage model, indicators for

quintiles of the price of a full meal were used to predict *state* (NJ=1, else 0) in a linear regression; we note that the price of a full meal is associated with *state* (being higher in NJ than PA), and changed little between pre-and post-increase in minimum wage periods (i.e., the average reported price of a full meal differed by 1 cent in PA over the survey periods, and differed by 6 cents in NJ over the survey periods). In the second stage, a linear regression model was fitted for $Y(t=1)-Y(t=0)$ as a function of the predicted value of *state* given the price of a full meal.

Empirical example 2: Health Insurance Subsidy Program

We use the Health Insurance Subsidy Program (HISP), a case example modeled after real-world example data of impact evaluations that were developed by the World Bank.[11] One of the primary objectives of HISP is to reduce the burden of health-related out-of-pocket expenditures for low-income households. The data are at the level of household and period and include a baseline ($t=0$) and follow-up survey ($t=1$). The outcome variable of primary interest denoted $Y(t)$, is out-of-pocket health expenditure (per capita per year); the intervention of primary interest is a binary indicator of whether the household enrolled in HISP (0=no, 1=yes). Here we restrict to data from localities where the program has been offered. First, a standard DID estimate is obtained by linear regression in which $Y(t=1)-Y(t=0)$ is regressed on HISP. However, one might question whether the parallel trends assumption required for standard DID holds. Next, we fitted our proposed GDID. The bespoke IV was age of the head of the household (in years). The variable was chosen both because it was relevant for predicting treatment (i.e., whether the household enrolled in HISP), and because it is plausible that the association between age of head of household and health expenditure before enrolling in HISP is equal to the association between age of head of household and health expenditure in the posttreatment period (in the absence of HISP). In a first-stage model, indicators for quintiles of age of the head of the household were used to predict HISP in linear regression. In the second stage, a linear regression model was fitted for $Y(t=1)-Y(t=0)$ as a function of the predicted value of HISP.

RESULTS

Simulation Example

Under the first simulation scenario, which conformed to the parallel trends assumption, our GDID two-stage regression estimator of the exposure–outcome association suffered no bias due to confounding by U, even though U is an unmeasured variable. Similarly, when using a difference-in-differences approach, the estimator of the exposure–outcome association suffered no bias due to confounding by U. The statistical efficiency of the GDID estimator diminished with diminishing magnitude of association between Z and A (as reflected by the Monte Carlo standard deviations). When the association between Z and A was weak or moderate, the

statistical efficiency of the GDID estimator was lower than that of the standard DID estimator, and the estimated RMSE was larger for the GDID estimator than for the standard DID estimator. When the association between Z and A was strong, the GDID method had efficiency and RMSE close to that of the standard DID estimator under the condition of parallel trends.

Under the second simulation scenario, which violated the "parallel trends" assumption, the GDID estimator of the exposure–outcome association suffered no bias. In contrast, the DID estimator suffered bias; in addition, the estimated RMSE was larger for the standard DID estimator than for the GDID estimator when the association between Z and A is moderate or strong, and was similar for the two estimators when the association between Z and A was weak.

The eTable, http://links.lww.com/EDE/B983, in eAppendix 3, http://links.lww.com/EDE/B983, provides results of additional simulation scenarios. Simulation A1 conforms to the "parallel trends" assumption except that U_1 affects Z_1. Neither the standard DID nor the GDID estimator was biased. Simulation A2 violates the parallel trends assumption, U_1 affects Z_1, and we included an additional covariate U_2. The standard DID estimator was biased while GDID suffered no bias. Simulation A3 violates the parallel trends assumption and U_2 affects Z_1 (violating the condition of Z_1 additive equiconfounding). The DID estimator suffered bias and the GDID estimator suffered bias. In simulations A1 and A3, and in simulation A2 when the association between Z_1 and A was weak, the estimated RMSE was smaller for the DID estimator than for the GDID estimator. In contrast, in simulation A2 when the association between Z_1 and A is strong or moderate, the estimated RMSE was larger for the DID estimator than for GDID.

Empirical Example 1

In the pre-period, average employment was 23.33 full-time equivalent workers per store in Pennsylvania and 20.4 full-time equivalent workers per store in New Jersey. In the post-period, full-time equivalent employment increased in New Jersey relative to Pennsylvania. A standard DID estimator yields an estimate of the relative gain in employment of 2.33 (s.e.=1.19) full-time equivalent employees. Our proposed GDID yielded a slightly larger magnitude of estimate of the relative gain in employment albeit with less precision (estimate=3.12, s.e.=3.69), with the first-stage F-statistic being 10.29.

Empirical Example 2

Estimates of mean household health expenditures (in dollars) for enrolled households before HISP was 14.49 and after introduction of HISP was 7.84. Estimates of mean household health expenditures for non-enrolled households before HISP was 20.79 and after introduction of HISP was 22.30. A standard difference-in-differences estimate of the causal effect of HISP on household health expenditures was −8.16 (s.e.=0.32). The GDID approach yielded an estimate of −8.50 (s.e.=0.77), with the first-stage F-statistic being 168.13.

DISCUSSION

We propose a novel approach to the analysis of data that conform to the DID design. A standard DID estimator allows identification of the average causal effect of treatment on the treated under the parallel trends assumption (Figure A). However, it may yield a biased estimate of the effect of treatment on the treated if an unmeasured confounder varies between pre- and posttreatment periods (Figure B). The GDID approach allows identification of the average causal effect of treatment on the treated under the causal structures illustrated in both parts of the Figure.

We exchange the usual DID parallel trends assumption for a different set of assumptions. Specifically, the GDID approach requires assumptions about a measured covariate, Z_1, that we can select (from among those measured variables): Z_1 predicts treatment, A; no statistical interaction between Z_1 and A in causing $Y(t_1)$; in addition, the additive association of Z_1 with $Y^0(t_0)$ and $Y^0(t_1)$ is equivalent. This set of identifying conditions may be an appealing alternative to the usual DID parallel trends assumption. Several prior publications have described methods for researchers to draw causal inferences if parallel trends is violated, including an inverse probability weighting method for DID,[12] an outcome regression modeling approach,[13] and a doubly robust approach.[14] Those methods accommodate violations of parallel trends by modeling measured covariates, so that parallel trends are assumed to hold upon conditioning on measured covariates. Other proposed methods relax the parallel trends assumption by placing bounds on how much unmeasured confounders may affect the untreated potential outcomes.[15,16] In contrast, our approach further accommodates violations of the parallel trends assumption by unmeasured covariates. Our proposed GDID method is a new alternative in this broader suite of methods that seek to relax the parallel trends assumption.

While we have described the approach, and illustrated it, with examples where we identify a single covariate, Z_1, whose effect on $Y^0(t)$ is invariant over $t=0,1$, the GDID approach readily extends to incorporating several covariates; leveraging several covariates may be appealing because it may offer a way to strengthen their relevance for predicting treatment. Our empirical examples leverage publicly available data sets to illustrate the method; we recognize that the assumptions necessary for the identification of causal effects in these empirical examples may not hold perfectly (e.g., a condition may be violated for a provided all BSIV conditions hold selected variable, Z_1, employed in a specific empirical setting), but this serves to underscore the value of sensitivity analyses and access to methods that leverage alternative identifying conditions. Similarly, the approach also may be extended to explicitly model the effects of other measured covariates Z_2 in the second-stage regression. Although our primary focus is on analysis of nonexperimental data, the GDID method also may have utility for analyses of data derived from experimental designs, such A/B tests, where a classical DID analysis is sometimes used to adjust for pre-test differences in covariates between the groups under comparison (i.e., unanticipated imbalance across treatment groups in covariates).

Connections between instrumental variables and DID have been discussed by prior authors. Ye (2021) discusses an "instrumented" DID for example, which leverages exogenous random variation in treatment within a standard DID framework.[17] The proposed GDID has some similarities to the instrumented DID proposed by Ye et al. (2021), but also has notable differences. First, the GDID considers the same setting as the standard DID, where data are observed for treated and control units before and after the treated units adopt the treatment while the control units are never treated. The GDID also considers the same parameter of interest as the standard DID, that is the average treatment effect for the treated in the post-treatment period. In contrast, the instrumented DID considers the setting where units can be treated or untreated at either time point and the average treatment effect as the parameter of interest. Second, the GDID assumes that Z_1 does not modify the average treatment for the treated in the post-treatment period, while the instrumented DID assumes that the instrument variable for DID is independent of the treatment effect and the treatment effect is time-invariant. The "Z_1 parallel trends assumption" made by the GDID is essentially shared by the instrumented DID. Interestingly, the identification formulas of these two methods are the same when Z_1 is binary and no units are treated in the pretreatment period, which provides the identification formula with two different interpretations; this is analogous to the familiar observation that the standard Wald ratio estimator used in a standard IV analysis can be interpreted as the average treatment effect under a no unmeasured common effect modifier assumption[18] and can be interpreted as the average treatment effect for the treated under a no current value interaction assumption.[7]

As our simulations illustrated, when the parallel trends assumption holds, the standard DID approach is more statistically efficient than our proposed two-stage least squares estimator of the GDID identifying estimand, although when the association between Z_1 and A is moderate or strong, our proposed estimator has RMSE close to that of the standard DID. Usefully, nested within the GDID model is a reduced model that implies parallel trends; a constraint can be imposed on β_0 in equation 3 to reflect the stronger (but empirically testable) condition that the mean of the treatment-free potential outcomes does not differ between the pre-treatment period and post-treatment period, $E\left[Y^0(1)|Z=0\right] = E\left[Y^0(0)|Z=0\right]$. Under this constraint one obtains improved precision of the proposed estimator (equivalent to that of standard DID) and identification of the effect of treatment on the treated causal effect, without appeal to the assumption of no additive interaction between A and Z_1 in causing $Y^a(t_1)$. When

TABLE. Monte Carlo Mean, Standard Deviation (SD), root MSE (RMSE), Average Standard Error (SE), and Coverage Probability (CP) of 95% Asymptotic Confidence Interval for 1000 Cohorts with 5,000 Observations Each. Results of Simulations of Association Between Exposure, *A*, Measured Covariate, *Z*, Unmeasured Covariate, *U*, and Outcome, *Y*

Scenario	Mean	SD	RMSE	SE	CP
Scenario 1 (Conforms to "parallel trends")					
Strong bespoke IV (mean F-statistic 1171.8)					
GDID method	1.00	0.09	0.09	0.10	95.2
Standard DID method	1.00	0.04	0.04	0.04	96.1
Moderate bespoke IV (mean F-statistic 308.5)					
GDID method	1.00	0.16	0.16	0.17	95.3
Standard DID method	1.00	0.04	0.04	0.04	96.1
Weak bespoke IV (mean F-statistic 49.7)					
GDID method	1.00	0.42	0.42	0.42	96.0
Standard DID method	1.00	0.04	0.04	0.04	94.5
Scenario 2 (Violates "parallel trends")					
Strong bespoke IV (Mean F-statistic 1067.2)					
GDID method	1.00	0.12	0.12	0.12	95.7
Standard DID method	1.39	0.05	0.39	0.05	0.0
Moderate bespoke IV (mean F-statistic 277.6)					
GDID method	1.01	0.21	0.21	0.22	95.7
Standard DID method	1.44	0.05	0.44	0.05	0.0
Weak bespoke IV (Mean F-statistic 44.0)					
GDID method	1.01	0.55	0.55	0.55	96.5
Standard DID method	1.46	0.05	0.47	0.05	0.0

the parallel trends assumption does not hold, our proposed method will be unbiased and, when the association between Z_1 and A is moderate or strong, our proposed approach will have a smaller RMSE than the standard DID. While the GDID has a cost in terms of statistical precision reflecting the different identifying conditions and bespoke IV estimation approach, it will allow for avoiding bias in certain settings of unmeasured time-varying confounding. Of course, if the cost in terms of precision is high, the root means square error of the GDID approach may exceed that of the standard DID (as illustrated in some simulations); this is particularly true when the parallel trends assumption holds (or nearly holds).

The GDID approach has the potential to be used in routine policy evaluation across many disciplines, as it essentially combines two popular quasiexperimental designs, leveraging their strengths while relaxing their usual assumptions (which also provides overidentification specification tests). The GDID approach can be used with any available measured confounder without requiring unmeasured confounding, and the stronger the association of the measured confounder with the treatment, the stronger the resulting instrument. The GDID approach will tend to sacrifice some statistical efficiency to reduce potential bias due to time-varying unmeasured confounders. In non-experimental studies, this may often be a desirable trade-off.

ACKNOWLEDGMENTS

REFERENCES

1. Wing C, Simon K, Bello-Gomez RA. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health*. 2018;39:453–469.
2. Bilinski A, Hatfield LA. Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. *arXiv*. 2019.
3. Rambachan A, Roth J. An honest approach to parallel trends. *Working Paper*. 2020. http://jonathandroth.github.io/assets/files/HonestParallelTrends_Main.pdf.
4. Richardson DB, Tchetgen Tchetgen EJ. Bespoke instruments: a new tool for addressing unmeasured confounders. *Am J Epidemiol*. 2022;191:939–947.
5. Sofer T, Richardson DB, Colicino E, Schwartz J, Tchetgen Tchetgen EJ. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Stat Sci*. 2016;31:348–361.
6. Robins J. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat*. 1994;23:2379–2412.
7. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17:360–372.
8. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J Am Stat Assoc*. 1995;90:443–450.
9. Andrews I, Stock J, Sun L. Weak instruments in IV regression: theory and practice. *Ann Rev Econ*. 2019;11:727–753.
10. Card D, Krueger AB. *Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania*. National Bureau of Economic Research Cambridge, Mass., USA; 1993.
11. Gertler P; World Bank. *Impact evaluation in practice*. World Bank, Washington, D.C.; 2016:1 online resource.
12. Abadie A. Semiparametric difference-in-difference estimators. *Rev Econ Stud*. 2005;72:1–19.
13. Heckman JJ, Ichimura H, Todd PM. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev Econ Stud*. 1997;64:605–654.
14. Sant'anna P, Zhao J. Doubly robust difference-in-differences estimators. *J Econometr*. 2020;219:101–122.
15. Manski CF, Pepper JV. How do right-to-carry laws affect crime rates? Coping with ambiguity using bounded-variation assumptions. *Rev Econ Stat*. 2018;100:232–244.

16. Bilinski A, Hatfield LA. Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. *arXiv*. 2020:1805.03273 [stat.ME].

17. Ye T, Ertefaie A, Flory J, Hennessy S, Small DS. Instrumented difference-in-differences. *arXiv*. 2020:2011.03593.

18. Wang L, Tchetgen Tchetgen E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J R Stat Soc Series B Stat Methodol*. 2018;80:531–550.

Data for the empirical example are publicly available (https://davidcard.berkeley.edu/data_sets.html).