One-stage robust difference-in-differences regression

John Gardner*

This version: Jan., 2026†

Abstract

I develop a simple method of combining regression with difference-in-difference and event-study designs to obtain treatment effect estimates that are robust to the presence of average treatment-effect heterogeneity under staggered rollout. The resulting estimator can be obtained via a single regression, which automatically produces approximately valid asymptotic standard errors. This one-stage estimator is numerically equivalent to the two-stage difference-in-differences estimator developed in Gardner (2021) and Gardner, Thakral, Tô, and Yap (2023) (which is also the same as the estimators developed in Borusyak, Jaravel and Spiess, 2021 and Liu, Wang and Xu, 2023 and can also be obtained from the estimators developed in Wooldridge, 2025 and Deb et al., 2024), and therefore inherits the robustness (and other) properties of those estimators. The estimator can also be extended to identify other treatment-effect measures and implement placebo tests of parallel trends. I illustrate the properties and application of this approach using simulations and applications from the literature.

Keywords: Differences-in-differences, event-studies, treatment effects, treatment-effect heterogeneity, matching, regression, causal inference.

JEL codes: C01, C10, C21, C22, C23.

*Department of Economics, University of Mississippi, jrgardne@olemiss.edu. I thank Jeff Wooldridge, Len Goff, Matt Webb, and seminar participants at the Midwestern Econometrics Association and Southern Economic Association annual meetings for helpful comments and suggestions.

†Previous versions: July, 2025, 20 March, 2024, 11 Jan., 2023, 1 Dec., 2023.

1 Introduction

It is now widely known that difference-in-differences and event-study estimates based on traditional two-way-fixed-effects regression specifications do not always identify sensible measures of average treatment effects when the adoption of a treatment is staggered over time and duration-specific average treatment effects are heterogeneous across treatment cohorts (see de Chaisemartin and D'Haultfœuille, 2020; Borusyak, Jaravel and Spiess, 2021; Sun and Abraham, 2021; Goodman-Bacon, 2022). These observations have spawned a proliferation of alternative estimators that are robust to the problems facing traditional regression specifications in the staggered and heterogeneous setting (see, e.g., Borusyak, Jaravel and Spiess, 2021; Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020; Dube, Jordà and Taylor, 2023; Gardner, 2021; Gardner, Thakral, Tô, and Yap, 2023; Liu, Wang and Xu, 2023; Sun and Abraham, 2021; Wooldridge, 2025; Deb, Norton, Wooldridge and Norton, 2024).

In this paper, I develop a new approach to robust identification of average treatment effects using difference-in-difference designs in this setting, motivated by analogy to matching and regression methods for identifying causal effects under selection on observables. A key advantage of this new approach is that it only requires the estimation of a single regression, which automatically produces estimates of the overall average effect of the treatment on the treated, along with approximately valid asymptotic standard errors. Moreover, this approach does not really add to the growing list of robust estimators. Instead, point estimates obtained using this approach are identical to those from the two-stage difference-in-differences estimator developed in Gardner (2021) and discussed in greater detail in Gardner, Thakral, Tô, and Yap (2023), which is also the same as the imputation estimator developed in Borusyak, Jaravel and Spiess (2021) and the fixed-effects counterfactual estimator developed in Liu, Wang and Xu (2023). As a consequence, the one-stage estimator developed in this paper automatically inherits the advantages of these estimators, including robustness to treatment effect heterogeneity under staggered adoption, efficiency (under the Gauss-Markdov conditions, see

Borusyak, Jaravel and Spiess, 2021), the ability to control for time-varying covariates (that evolve exogeneously, see Caetano, Callaway, Payne and Rodrigues, 2022), and arbitrary dependence of treatment effects on those covariates. The one-stage approach that I develop here is also related, and complementary, to the estimators developed in Wooldridge (2025) and Deb et al. (2024), which can also recover the two-stage estimate from a regression specification.[1] Relative to those papers, this paper provides an alternative way of obtaining two-stage estimates from a single regression. While, in the specifications in those papers, summary measures of the effect of the treatment are identified as aggregates of group×time-specific ATTs, in the specifications developed below they are identified directly as the coefficients on treatment-status variables. Thus, this paper brings some of the simplicity of differences in differences regression in the 2×2 case to settings with staggered adoption.

This one-stage approach to estimation developed in this paper is flexible. In addition to the overall average effect of the treatment on the treated, variations on the basic one-stage regression specification can be used to identify other objects of interest, including dynamic treatment effects and cohort- or time-specific average treatment effects, as well as to implement robust triple-differences designs and placebo tests of the parallel trends assumption. Furthermore, since it ultimately amounts to estimating a single regression, it can be implemented by anyone using any standard statistical software package, without the assistance of a specialized estimation routine.

The one-stage estimator developed here is a simple extension of the traditional two-way fixed effects specification, which makes that specification robust to treatment-effect heterogeneity under staggered adoption. The one-stage specification consists of the traditional specification, plus a set of interaction terms between treatment status and the other variables included in that specification. The inclusion of these additional terms ensures that the coefficient on treatment status identifies a particular (and particularly sensible) measure

[1]The specifications developed in this paper are different from those in Wooldridge (2025) and Deb et al. (2024), which can also be used to recover point estimates from (variations on) the two-stage estimator. I was not aware of the equivalence between those estimators and the two-stage approach when writing the initial versions of this paper.

of the overall average effect of the treatment on the treated. This overall average effect of the treatment on the treated can also be viewed as what the traditional two-way fixed effects difference-in-differences regression "tries" to identify when average treatment effects are heterogeneous and adoption is staggered.[2]

In Section 2, below, I outline the environment in which the one-stage approach to differences in differences developed in this paper applies. Because I formally establish the consistency of this estimator by showing its equivalence to the two-stage difference-in-differences estimator, I also briefly review the properties of the latter estimator in that section. I motivate and develop the one-stage robust difference-in-difference regression estimator in Section 3. There, I also I show that this one-stage estimator is equivalent to two-stage differences in differences, and introduce several useful variations on the methodology. In Section 5, I present evidence on the performance of the one-stage estimator using Monte Carlo simulations. In Section 6, I illustrate the use of the estimator, and compare it to the two-stage estimator, in the context of two applied examples from the literature. I offer some concluding remarks in Section 7.

2 Setup, and review of the two-stage approach

Suppose that the data consist of observations on outcomes Y_{it}, treatment status $D_{it} \in \{0, 1\}$, and a set of time-varying control variables X_{it}, for units $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$. I assume throughout that the covariates X_{it} are not affected by the treatment. Further suppose that the treatment is irreversible and unanticipated.[3]

Let $C_i \in \{2, \ldots, T, \infty\}$ be the date at which unit i adopts the treatment, and set $C_i = \infty$ if i is never treated during the sample period (drop any observations that are always treated during the sample period, since no treatment effects are identified for always-treated units).

[2] I provide more detail on what the traditional specification can be viewed as attempting to identify in Appendix B.

[3] It is possible to allow for potential anticipation by redefining D_{it} to be an indicator for whether unit i adopts the treatment a certain number of periods after t.

Also define treatment-cohort indicators $C_i^j = 1(C_i = j) \in \{0, 1\}$ for $j \in \mathcal{C} = \{2, \ldots, T, \infty\}$.

Let (Y_{it}^0, Y_{it}^1) be the counterfactual untreated and treated outcomes that i would experience at time t, and $\beta_{it} = Y_{it}^1 - Y_{it}^0$ be the time-t causal effect of the treatment for unit i. I assume for simplicity that the data $\{Y_{it}, X_{it}, C_i\}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, consist of a random panel, although all of the results in this paper also apply to repeated-cross-sectional data.[4]

Suppose that untreated outcomes follow parallel trends in the sense that

$$E(Y_{i,t+1}^0 - Y_{it}^0 | X, C^j) = E(Y_{i,t+1}^0 - Y_{it}^0 | X, C^k) = \Delta\gamma_t + \Delta X_{it}'\delta$$

for all $j, k \in \mathcal{C}$. Under this parallel trends assumption, and maintaining the no-anticipation assumption, expected observed outcomes can be expressed as

$$E(Y_{it} | X_{it}, C_i, D_{it}) = \lambda_{c(i)} + \gamma_t + X_{it}'\delta + \beta_{ct}(X_{it})D_{it},$$

where $\beta_{ct}(X_{it}) = E(Y_{it}^1 - Y_{it}^0 | C_i, X_{it}) = E(\beta_{it} | C_i, X_{it})$, and observed outcomes can be expressed as

$$Y_{it} = \lambda_{c(i)} + \gamma_t + X_{it}'\delta + \beta_{ct}(X_{it})D_{it} + u_{it}. \tag{1}$$

The two-stage difference-in-differences estimator is based on the implication of parallel trends that

$$Y_{it} - \lambda_{c(i)} - \gamma_t - X_{it}'\gamma = \beta_{ct}(X_{it})D_{it} + u_{it} = \beta D_{it} + [\beta_{ct}(X_{it}) - \beta]D_{it} + u_{it} \equiv \beta D_{it} + \varepsilon_{it},$$

where the overall ATT β is defined as the average effect of the treatment on the treated, with the average taken across all cohorts, covariates, and time periods, which I denote by

[4]Since all of the estimands discussed in this paper are conditional on $D_{it} = 1$, causal effects for members of the never-treated cohort can be normalized to zero without loss of generality.

$\beta = E^{\circ}(\beta_{it}|D_{it} = 1) = E^{\circ}[\beta_{ct}(X_{it})|D_{it} = 1]$, in which case $E^{\circ}\{[\beta_{ct}(X_{it}) - \beta]D_{it}|D_{it}\} = D_{it}E^{\circ}[\beta_{ct}(X_{it})|D_{it}] - \beta D_{it} = 0$.[5] While there are multiple ways of summarizing treatment effects that vary across cohorts and time periods (see Callaway and Sant'Anna, 2021, for a good discussion), this notion of the overall ATT is intuitive and, as I discuss in Appendix B, has the additional advantage of being the analog of what the traditional two-way fixed effects difference-in-differences specification arguably "tries" (but fails) to identify when adoption is staggered and treatment effects are heterogeneous.

Thus, if λ_c, γ_t, and δ were known, the overall ATT could be estimated from a regression of adjusted outcomes $Y_{it} - \lambda_{c(i)} - \gamma_t - X'\delta$ on treatment status D_{it}. Although they are not known, as long as (i) there are untreated observations in every period, and (ii) there are pre-treatment observations for every eventually-treated unit, λ_c, γ_t, and δ can be estimated from a regression of outcomes on cohort fixed effects, time fixed effects, and time-varying controls using the sample of untreated observations.[6] The two-stage difference-in-differences estimate $\hat{\beta}^{2SDD}$ of the overall ATT β is the estimated coefficient on treatment status from a regression of $Y_{it} - \hat{\lambda}_{c(i)} - \hat{\gamma}_t - X'_{it}\hat{\delta}$ on D_{it}, where $\hat{\lambda}_c$, $\hat{\gamma}_t$, $\hat{\delta}$ are estimates of those parameters obtained from a first-stage regression using the sample of untreated observations. When this procedure is implemented with cohort fixed effects, the consistency of $\hat{\beta}^{2SDD}$ follows directly from the consistency of the first-stage estimates of λ_c, γ_t, and δ. In the interest of generality,

[5] I use the symbol "$E^{\circ}(\cdot)$" to denote expectations over time periods (in addition to other random variables). This notation implicitly treats $\beta_{ct}(X_{it})$ as a single random variable whose distribution varies across covariates, cohorts, and time (as opposed to a sequence of random variables indexed by time), which can motivated by thinking of a regression of Y on cohort and time indicators as specifying the conditional expectation of outcomes conditional on time and cohort membership. Note that when some Z_{it} is considered as a single random variable whose distribution varies across time, $E^{\circ}(Z_{it}|D_{it} = 1) = \sum_t E^{\circ}(Z_{it}D_{it}|t)/\sum_t E^{\circ}(D_{it}|t)$ [this follows because, suppressing the subscripts, $E^{\circ}(Z|D = 1) = E^{\circ}(ZD)/E^{\circ}(D)$, where $E^{\circ}(ZD) = \sum_t E^{\circ}(ZD|t)/T$, and similarly for $E^{\circ}(D)$], which is also the probability limit of $\bar{Z}^1 = (\sum_i \sum_t Z_{it}D_{it})/(\sum_i \sum_t D_{it})$. Alternatively, we could maintain the convention that the Z_{it} are separate random variables for each time period and simply define the estimand of interest to be $E[\sum_t (Y_{it}^1 - Y_{it}^0)D_{it}]/E(\sum_t D_{it})$. I prefer the former approach since it makes clear the connection between the overall ATT and the ATT in the cross-sectional case and provides a formal justification for the latter approach, although both lead to the same estimators, and none of this paper's formal results rely on the use of the former approach.

[6] Observed outcomes can also be expressed in terms of unit (rather than cohort) fixed effects as

$$Y_{it} = \lambda_i + \gamma_t + X'_{it}\delta + \beta_{it}D_{it} + u_{it},$$

where $\lambda_c = E(\lambda_i|C_i = c)$ and $\beta_{it} = \beta_{it}(X_{it})$.

the following result establishes the consistency and asymptotic distribution of the two-stage estimator when it is implemented using unit, rather than cohort, fixed effects (although the argument can be easily be adapted to the case of cohort fixed effects).

Proposition 1. *Suppose that (i)* $E\left[\sum_t (1 - D_{it})\ddot{X}_{it}^0 \ddot{X}_{it}^{0\prime}\right]$ *is invertible, where* $\ddot{X}_{it}^0$ *denotes the vector of deviations in unit i's covariates (including time indicators) from their means across all periods where i is untreated, (ii)* $E(u_{is}|D_{it}, X_{it}) = 0$ *for all s and t, and (iii)* $E\left(\sum_t D_{it}\right) \neq 0$*. Then, under parallel trends and no anticipation,* $\hat{\beta}^{2SDD} \xrightarrow{p} \beta$ *and* $\sqrt{N}(\hat{\beta}^{2SDD} - \beta) \overset{a}{\sim} N\left(0, A_0^{-1} B_0 A_0^{-1}\right)$*, where* A_0 *and* B_0 *are defined in Appendix A.*[7]

The proof is given in Appendix A.[8]

3 A robust one-stage regression approach

3.1 Motivation

The motivation for the robust one-stage estimator comes from the literature on matching and selection on observables. In a cross-sectional setting, if counterfactual outcomes (Y_0, Y_1) are independent of treatment status $D \in \{0, 1\}$ conditional on a set X of covariates, then the conditional counterfactual mean outcome functions $E(Y_d|X = x)$, $d \in \{0, 1\}$, can be estimated from separate regressions of outcomes on covariates for the treated and untreated samples, or from a pooled regression of outcomes on the covariates and their interaction with

[7] Parallel trends implies that $E(u_{it}|D_{it}, X_{it}) = 0$. The assumption that u_{it} is strictly exogeneous is consistent with the idea that the covariates are not affected by the treatment – there is no feedback from the errors or the treatment to the covariates. The proof of Proposition 1 can be extended to show the consistency and asymptotic distribution of a version of the estimator that uses cohort, rather than unit, fixed effects, by absorbing the cohort dummies into the vector X_{it} and replacing $\ddot{X}_{it}$ with the (un-demeaned) vector X_{it}. Furthermore, some tedious algebra shows that the estimated variances for the dummy-variable and within-transformation versions of the estimator are identical, so the asymptotic distribution with cohort fixed effects can also be obtained from Proposition 1, redefining $\ddot{X}_{it}^0$ to be the deviation from the cohort-specific untreated mean.

[8] Butts and Gardner (2022) and Gardner, Thakral, Tô, and Yap (2023) derive the asymptotic distribution of the estimator (for the dummy-variable and within-transformation cases, respectively) by treating the first- and second-stages as a joint GMM estimator.

treatment status:

$$Y = X'\theta_0 + D \cdot X'\theta_1 + q, \tag{2}$$

where $E(q|X, D) = 0$. If the counterfactual mean outcome functions are indeed linear in the covariates, they are identified by these regressions, and the ATT can be estimated as the sample analog of

$$ATT = E[E(Y_1|X = x) - E(Y_0|X = x)|D = 1] = E(X|D = 1)'\theta_1.$$

The second, aggregation, step of this procedure can be avoided by replacing the regression specification (2) with

$$Y = X'\rho_0 + D[X - E(X|D = 1)]'\rho_1 + \beta D + r. \tag{3}$$

In this case, the ATT can be estimated as the sample analog of

$$ATT = \beta + E[X - E(X|D = 1)|D = 1]'\rho_1 = \beta,$$

after replacing $E(X|D = 1)$ with its sample analog $\bar{X}^1 = \sum_i D_i X_i / (\sum_i D_i)$.[9] This modified approach has at least two practical advantages. First, it may be easier to obtain the treated means $\bar{X}^1$ than to aggregate the covariate-specific ATTs. Second, regression estimates of specification (3) will automatically produce asymptotic standard errors for the ATT, which can be used for hypothesis testing and other statistical inference (as Wooldridge, 2010, notes, the standard errors should technically account for the estimation of $\bar{X}^1$, although this unlikely to make much of a difference).

The traditional difference-in-differences estimator regresses outcomes on treatment-cohort (or unit) and time indicators (abstracting away from any other potential control variables), in addition to time-varying treatment status. The nexus between the regression-adjustment

[9] To the best of my knowledge, this observation is due to Wooldridge (2010, Ch. 21).

matching approach described above and difference-in-differences estimation comes from pretending that these indicators are true covariates (i.e., that they are quasiexperimentally manipulable). If this were the case, the overall average effect of the treatment could be identified by matching observations for treated and untreated units belonging to the same cohort and recorded in the same period.

Extending the regression-adjustment approach to differences in differences presents two challenges. The first is that treated units can never be matched to untreated versions of themselves in the same period. All difference-in-differences methodologies circumvent the impossibility of this thought experiment under a parallel trends assumption, which allows the evolution of outcomes for untreated units to be used in place of the counterfactual evolution of untreated outcomes for treated units. While this kind of extrapolation between units may be suspect in the general context of selection on observables, in difference-in-differences designs it is actually desirable.

The second challenge is that, while parallel trends implies that untreated mean outcomes are linear in cohort/unit and time indicators, in the presence of arbitrary heterogeneity, treated outcomes will be nonlinear in those variables if treatment effects vary at the cohort/unit×time level. If the covariates W_{it} consist of a full set of cohort or unit indicators, $T-1$ relative time indicators, and time-varying controls X_{it}, parallel trends implies (along with no anticipation) that untreated outcomes satisfy[10]

$$E(Y_{it}^0|W_{it}) = \lambda_{c(i)} + \gamma_t + X_{it}'\delta \equiv W_{it}'\theta_0,$$

while treated outcomes satisfy

$$E(Y_{it}^1|W_{it}) = \lambda_{c(i)} + \gamma_t + X_{it}'\delta + \beta_{ct}(X_{it}) \equiv W_{it}'\theta_0 + \beta_{ct}(X_{it}).$$

Although the latter expression is nonlinear in the covariates, a closer look at the regression-

[10]If W_{it} includes unit rather than cohort indicators, then the cohort becomes the unit, so that $\lambda_c = \lambda_i$.

adjustment approach reveals that what it really requires is that counterfactual outcomes are linear on average across the treated population. To see that the logic of this approach carries over to the case of differences in differences, express $\beta_{ct}(X_{it})$ in terms of its projection onto cohort and time indicators (and time-varying controls) as

$$\beta_{ct}(X_{it}) = \beta_c + \beta_t + X_{it}'\beta_x + (\beta_{ct} - \beta_c - \beta_t - X_{it}'\beta_x) \equiv \beta_c + \beta_t + X_{it}'\beta_x + \tilde{\beta}_{ct} \equiv W_{it}'\theta_1 + \tilde{\beta}_{ct},$$

where, by definition, $E^\circ(\tilde{\beta}_{ct}|D_{it} = 1) = 0$ (with the expectation taken across covariates, cohorts, and time periods, as in the definition of the overall ATT). Using this decomposition, we have that

$$E^\circ[E(Y_{it}^1 - Y_{it}^0|W_{it})|D_{it} = 1] = E^\circ(\beta_c + \beta_t + X_{it}'\beta_x|D_{it} = 1) = E^\circ(W_{it}|D_{it} = 1)'\theta_1.$$

Thus, the overall ATT β can be estimated as $\bar{W}^{1'}\hat{\theta}_1$, where $\bar{W}^1 = (\sum_i \sum_t D_{it}W_{it})/\sum_i \sum_t D_{it}$ is the average of the covariates among treated observations and $\hat{\theta}_1$ is the estimated pooled least-squares regression coefficient vector on $D_{it}W_{it}$ from the specification

$$Y_{it} = W_{it}'\theta_0 + D_{it}W_{it}'\theta_1 + s_{it}.$$

Moreover, the aggregation step (calculating $\bar{W}^{1'}\hat{\theta}_1$) of this procedure is obviated by using the alternative specification

$$Y_{it} = W_{it}'\rho_0 + D_{it}(W_{it} - \bar{W}^1)\rho_1 + \beta D_{it} + r_{it}, \tag{4}$$

from which β represents the overall ATT. Intuitively, including D_{it} forces the cohort and time effects to measure deviations from a reference group$\times$period, while demeaning W_{it} forces this to be the overall average among the treated.

In other words, the overall ATT can be estimated as the coefficient on treatment status

from a regression of outcomes on

(i) cohort/unit and and time-period indicators, as well as any time-varying control variables,

(ii) interactions between treatment status and deviations in cohort/unit indicators, time indicators, and time-varying controls from their means among treated observations, and

(iii) treatment status.

Note that (ii) above is the only difference between this robust one-stage specification and the traditional two-way fixed effects specification.

3.2 Properties

When specification (4) includes cohort fixed effects, it is easy to see that the one-stage robust difference-in-differences regression estimator is consistent for the overall ATT: in this case, since $E(Y_{it}|W_{it}) = W_{it}'\rho_0 + D_{it}[W_{it} - E^{\circ}(W_{it}|D_{it} = 1)]'\rho_1 + \beta D_{it}$, and the vector $\bar{W}^1$ of average cohort indicators, time indicators, and controls among the treated converges to its population analog $E^{\circ}(W_{it}|D_{it} = 1)$ by a law of large numbers, $\hat{\beta}^{1SDD}$ is consistent by an application of the continuous mapping theorem and pooled OLS arguments.[11]

More formally, the consistency and asymptotic distribution of the one-stage estimator can be established by the following result (I also provide an alternative, direct analysis of the one-stage estimator in Appendix C).

Proposition 2. *The one-stage robust difference-in-differences regression estimator is numerically equivalent to the two-stage difference-in-differences estimator:* $\hat{\beta}^{1SDD} = \hat{\beta}^{2SDD}$.

[11] Alternatively, one can appeal to general consistency results for two-stage estimators (see Newey and McFadden, 1994; Wooldridge, 2010, chapter 12). Identification for pooled OLS also requires that $E\left(\sum_t Q_{it}Q_{it}'\right)$ is invertible, where $Q_{it} = [W_{it}, D_{it}(W_{it} - \bar{W}^1), D_{it}]$ is the vector of observations on all covariates for unit i at time t. This requires that the cohort sizes increase with the sample size.

Proof. Let $\hat{\lambda}_c^0$, $c \in \mathcal{C}$, $\hat{\gamma}_t^0$, $t = 1, \ldots, T$, and $\hat{\delta}^0$ denote the estimated cohort fixed effects, time fixed effects, and coefficients on time-varying controls from a first-stage regression of outcomes on those variables, obtained from the sample of untreated observations.[12] The two stage difference-in-differences estimator is the coefficient on D_{it} from a second-stage regression of $Y_{it} - \hat{\lambda}_{c(i)}^0 - \hat{\gamma}_t^0 - X_{it}'\hat{\delta}^0$ on D_{it} (with no constant term). Since D_{it} and $D_{it}(W_{it} - \bar{W}^1)$ are orthogonal, the term $D_{it}(W_{it} - \bar{W}^1)$ can be added to the second-stage regression without changing the estimated coefficient on D_{it}. Now, if $\hat{\lambda}_c^0$, $\hat{\lambda}_t^0$, and $\hat{\delta}^0$ were the same as the estimated cohort effects, time effects, and control coefficients $\hat{\lambda}_c^{1SDD}$, $\hat{\gamma}_t^{1SDD}$, and $\hat{\delta}^{1SDD}$ from the one-stage regression specification of Y_{it} on W_{it}, $D_{it}(W_{it} - \bar{W}_{it})$, and D_{it}, then the estimated coefficient $\hat{\beta}^{1SDD}$ on D_{it} from the one-stage specification would be identical to $\hat{\beta}^{2SDD}$ (this is an exercise in partitioned regression mechanics; see, for example, Greene, 2018, Ch. 3) .

By the Frisch-Waugh-Lovell theorem, $\hat{\lambda}_c^{1SDD}$, $\hat{\lambda}_t^{1SDD}$, and $\hat{\delta}^{1SDD}$ can be obtained from a regression of Y_{it} on the residuals from auxiliary regressions of the elements of W_{it} on D_{it} and $D_{it}(W_{it} - \bar{W})^1$. However, since D_{it} and $D_{it}(W_{it} - \bar{W})$ perfectly predict W_{it} for treated observations (if $D_{it} = 1$, we can always write the kth element of W_{it} as $W_{kit} = D_{it}(W_{kit} - \bar{W}_k) + \bar{W}_k D_{it}$), these residuals will be zero for all treated observations. Therefore, $\hat{\lambda}_c^{1SDD}$, $\hat{\lambda}_t^{1SDD}$, and $\hat{\delta}^{1SDD}$ can also be obtained by regressing Y_{it} on W_{it} in the sample of untreated observations. That is, $\hat{\lambda}_c^{1SDD}$, $\hat{\lambda}_t^{1SDD}$, and $\hat{\delta}^{1SDD}$ equal $\hat{\lambda}_c^0$, $\hat{\lambda}_t^0$, and $\hat{\delta}^0$. □

Thus, the one-stage robust regression estimator is identical to the two-stage difference-in-differences estimator, and the consistency of the former follows formally from that of the latter. The implication of Proposition 2 is that the one-stage approach is another way of obtaining two-stage difference-in-differences estimates. The primary advantage of the one-stage approach is that regression estimates of specification (4) automatically produce standard error estimates that do not need to be adjusted to account for the first-stage estimation of the fixed effects and control coefficients (as Wooldridge, 2010, Ch. 21, notes, the

[12] The following argument also applies if the cohort fixed effects are replaced with unit fixed effects $\hat{\lambda}_i^0$, $i = 1, \ldots, N$.

standard errors should technically be adjusted for the use of $\bar{W}^1$ in place of $E^{\circ}(W_{it}|D_{it}=1)$, although this likely has a small effect on the resulting standard errors).[13] Thus, the estimator can easily be implemented in any statistical package, without any specialized estimation routine.

The proof of Proposition 2 shows that the one- and two-stage estimators are equivalent regardless of whether they are specified using individual or cohort fixed effects, and the proof of Proposition 1 shows that both estimators are consistent in either case. However, there are practical reasons to prefer specifications with cohort fixed effects when using the one-stage approach. The first is computational: while a within transformation can be used to remove unit fixed effects, their demeaned interactions with treatment status must be included in the estimating equation, which may be impractical when there are many units. The second concerns inference: because OLS with unit fixed effects forces the residuals to sum to zero within units, standard errors that are clustered at or above the individual level will be mechanically biased.[14] The two-stage approach may therefore be preferable in estimating models that include unit fixed effects, although the parallel trends conditions that motivate the use of difference-in-difference analyses usually support the use of cohort fixed effects. Furthermore, regardless of whether point estimates are obtained by the one- or two-stage procedure, standard errors can also be bootstrapped (which may provide better inference

[13] In Stata, at least, it is also relatively simple to account for the sampling variation in $\bar{X}^1$ by estimating the ATT as $\bar{X}^{1'}\hat{\theta}_1$, then calculating delta-method standard errors using the margins command with the `unconditional` option.

[14] More specifically, the issue arises because the unit-specific sums of the residuals from a model that includes unit fixed effects must be zero, but these sums are a component of the cluster-robust estimate of the variance of those fixed effects. To see this, recall that the one-stage estimator can also be obtained by estimating treatment-status-specific regressions of outcomes on unit and time fixed effects, taking treated-untreated differences in those effects, then averaging the sum of differential unit and time effects over the treated sample. From the OLS first-order conditions, the estimated unit fixed effects are (ignoring treatment status for the sake of simplicity) $\hat{\lambda}_i = \bar{Y}_i - \bar{X}_i'\hat{\delta}$, where $\hat{\delta}$ includes the coefficients on the time effects (as well as any other covariates). Hence, the conditional variance of $\hat{\lambda}_i$ is $Var(\bar{\varepsilon}_i|X) + \bar{X}_i' Var(\hat{\delta}|X)\bar{X}_i$ (for example, a tedious calculation shows that with homoskedastic errors, $Var(\hat{\lambda}_i) = \sigma^2/T + \bar{X}_i' Var(\hat{\delta}|X)\bar{X}_i$). Under clustering at any level more coarse than the unit, the natural estimate of the first component is $\left(\sum_t e_{it}\right)^2/T^2$ which is zero by the first-order conditions for OLS (and another calculation shows that the actual estimated variance is $\bar{X}_i' Var(\hat{\delta}|X)\bar{X}_i$), underestimating the variance of the unit effect and, in the one-stage differences in differences context, the estimated ATT.

with few or few treated clusters, due to a similar mechanical bias, see MacKinnon and Webb, 2017; MacKinnon, Nielson and Webb, 2023).

4 Extensions

4.1 Dynamic effects

Difference-in-differences analyses are usually accompanied by "event-study" estimates that indicate the dynamic path of the treatment over time and provide placebo tests for the plausibility of parallel trends. Let T_i denote the period when treated unit i adopts the treatment, and $D_{it}^r = 1(t - T_i + 1 = r \text{ and } C_i \neq \infty)$ for $r \in \{-(T-2), \ldots, 1, \ldots, T-1\}$ be $(|r|+1)$-period leads (for $r < 1$) or $(r-1)$-period lags (for $r \geq 1$) of treatment status (i.e., D_{it}^1 represents the first period of treatment and D_{it}^0 represents the first period prior to treatment). Also let Y_{it}^{1r} be the counterfactual effect that i would experience at time t after being treated for r periods, and define $\beta^r = E^\circ(Y_{it}^{1r} - Y_{it}^0 | D_{it}^r = 1)$.

First, consider the case where $r \geq 1$. Under the two-stage difference-in-differences methodology, regressing adjusted outcomes $Y_{it} - \hat{\lambda}_{c(i)} - \hat{\gamma}_t$ on the D_{it}^r, $r \geq 1$, produces estimates of the average effect of being treated for r periods (on units treated for at least r periods). The one-stage robust estimator can be adapted to incorporate these dynamic estimands by replacing treatment status D_{it} with a set of r-period treatment status indicators D_{it}^r, $r \geq 1$. Following the logic of Section 3.1, under parallel trends

$$Y_{it} = W_{it}'\rho_0 + \sum_{r \geq 1} \{D_{it}^r [W_{it} - E^\circ(W_{it} | D_{it}^r = 1)]'\rho_1 + \beta^r D_{it}^r\} + v_{it}, \tag{5}$$

so that

$$E^\circ[E(Y_{it}^{1r} - Y_{it}^0 | W_{it}) | D_{it}^r = 1] = E^\circ[W_{it} - E^\circ(W_{it} | D_{it}^r = 1) | D_{it}^r = 1]'\rho_1 + \beta^r = \beta^r,$$

and hence β^r can estimated consistently as the coefficient on D_{it}^r from a feasible version of (5) that regresses outcomes on W_{it}, $D_{it}^r(W_{it} - \bar{W}^{1r})$, and D_{it}^r, for all $r \geq 1$, where $\bar{W}^{1r} = (\sum_i \sum_t W_{it} D_{it}^r) / \sum_i \sum_t D_{it}^r$ is the vector of average covariates among units treated for r periods.

To establish the consistency of the dynamic effects, first note that since the D_{it}^r are mutually orthogonal, the two-stage estimates of β^r can be obtained from a version of the two-stage procedure for the overall ATT that replaces D_{it} in the second stage equation with D_{it}^r (but still estimates the first stage using the subsample of untreated observations). Consequently, the coefficient on D_{it}^r from the dynamic one-stage regression specification (i.e., the sample analog of (5)) is identical to the coefficient that would obtain from first subsetting the data to contain only observations that are untreated or have been treated for exactly r periods, then estimating a version of the one-stage specification (4) for the overall ATT that replaces overall treatment status D_{it} with r-period treatment status D_{it}^r. Hence, the consistency of the one-stage dynamic-effect estimates follows from the consistency of the overall ATT estimates.

4.2 Assessing parallel trends

Placebo adoption There are several ways to implement placebo tests for parallel trends from within the one-stage approach developed above. To extend the approach to dynamic effects described above to test for parallel trends, choose some pre-treatment period $k \leq 0$, and consider a feasible version of (5) (i.e., replacing $E^{\circ}(W_{it}|D_{it}^{1r} = 1)$ with $\bar{W}^{1r}$) that sums over all $r \geq k$ (rather than over all $r \geq 1$). By the preceding argument, this specification is equivalent to estimating the dynamic specification after redefining the onset of the treatment as being $|k| + 1$ periods *before* its actual onset.[15] Consequently, the estimated coefficients on D^r, $r \in \{k, \ldots, 0\}$, represent consistent estimates of $|k| + 1$ pre-treatment placebo ATTs,

[15] An implementation note: While the researcher is free to choose k, since both the one- and two-stage procedures require that there are pre-treatment observations for all units, this placebo test requires that all units are observed at least $|k| + 1$ periods before adopting the treatment.

which can be used to test the plausibility of parallel trends (under which these pre-treatment ATTs should be zero). The consistency of these estimates follows from the preceding discussion, which implies that this procedure is equivalent to estimating dynamic treatment effects by two-stage differences in differences, after redefining treatment status as being $|k|+1$ periods prior the actual adoption of the treatment.[16]

The estimated coefficients on D^r, $r \geq 1$, from a version of specification (5) that sums over all $r \geq k$ represent consistent estimates of the r-period post-treatment ATTs β^r. However, because this specification only uses observations that are more than $|k|+1$ periods away from the actual onset of the treatment as the untreated sample, estimates of β^r, $r \geq 1$, from this specification will differ from those based on a version of (5) that only sums over $r \geq 1$. The dynamic ATT estimates β^r obtained from the original specification may therefore be preferable when the data are consistent with parallel trends, since those estimates compare the same treated observations to a larger sample of untreated observations.

To summarize, the dynamic and placebo effects of the treatment are identified from regressions of outcomes on

(i) cohort/unit indicators, time indicators, and control variables, collected into the vector of covariates W_{it},

(ii) interactions $D_{it}^r(W_{it} - \bar{W}_{it}^r)$ between deviations $(W_{it} - \bar{W}^r)$ in these covariates from their duration-specific means and duration-specific treatment-status indicators D_{it}^r, for all $r \geq k$ and some $k \leq 1$, and

(iii) duration-specific treatment status indicators D_{it}^r, for all $r \geq k$ and some $k \leq 1$.

The coefficients on D^r represent r-period ATTs for $r \geq 1$ and placebo tests of parallel trends for $r < 1$ (to estimate the dynamic ATTs using as many observations as possible, set $k = 1$).

[16] This is one of several approaches to testing parallel trends using two-stage differences in differences. For other approaches, see Gardner, Thakral, Tô, and Yap (2023), Borusyak, Jaravel and Spiess, 2021 and Liu, Wang and Xu (2023).

Adapting the traditional specification The simple approach to placebo testing developed above has some some disadvantages. First, it requires the researcher to specify the number of placebo periods ahead of time. Second, while under parallel trends, the (population) coefficients on the D_{it}^r should be zero, when parallel trends is violated, those coefficients cannot be interpreted as estimates of the average "effects" of being $|r|+1$ periods away from the treatment (except under the rather strict assumption that parallel trends holds whenever the adoption of the treatment is more than $|k|+1$ periods in the future). In the traditional (two-way fixed effects) event-study specification, the coefficients on the leads of treatment status can be interpreted as the average deviation from trends among units that are $|r|+1$ periods away from adopting the treatment, subject of course to the normalization that there is no deviation from trend in the omitted period (usually, $r=0$).[17]

Gardner, Thakral, Tô, and Yap (2023) note that the two-stage procedure can be adapted to share this property of the traditional event-study specification, even when adoption is staggered and treatment effects are heterogeneous. Specifically, they note that when the first-stage is estimated using observations for units that are either never treated or exactly one period ahead of adopting the treatment, the coefficients on the leads of treatment status can be interpreted as the average effects of being $|r|+1$ periods away from adopting the treatment (subject to the normalization that this effect is zero in period $r=0$).[18]

It follows directly from the logic of Section 4.1 that this approach can also be implemented from within the one-stage framework developed in this paper. In particular, modifying the dynamic specification (5) to include duration-specific treatment status indicators D_{it}^r and interactions between treatment status and deviations in the covariates (cohort indicators, time indicators, and control variables) from their duration-specific means $D_{it}^r(W_{it}-\bar{W}^r)$ for all $r \neq 0$ produces estimates that are numerically identical to the modified two-stage

[17]Of course, this interpretation is only valid if the adoption is not staggered or treatment effects are homogeneous.

[18]It is also easy to verify that, in the absence of covariates, average treatment effects obtained using this modified version of the two-stage procedure are numerically identical to those from the "never treated" variant of the Callaway and Sant'Anna (2021) estimator.

procedure described above. As before, although this modified procedure produces estimates of the dynamic effects of the treatment (as well as a more readily interpretable test of parallel trends), estimates of these effects from a version of specification (5) that only uses periods $r \geq 1$ may be more precise, since they compare post-treatment outcomes to a larger pool of pre-treatment outcomes.

4.3 Other extensions

Triple differences in differences Triple difference specifications are often used in difference-in-difference analyses when there is reason to believe that parallel trends may be violated. In the interest of concreteness, suppose that in states that adopt a policy, only those belonging to a particular group $G_i \in \{0, 1\}$ are affected by the treatment. If there is concern that, regardless of group membership, outcomes evolve differently between treated and untreated states, the differential outcomes of group non-members can be used to account for the parallel trends violation among group members.

In the two-stage framework, this triple-differenced strategy can be implemented by modifying the first stage to include cohort×group and time×group fixed effects, as well as cohort×time fixed effects. The same point estimates can be obtained from within the one-stage framework by regressing outcomes on

(i) cohort×group, time×group, cohort×time indicators, as well as time-varying control variables X_{it},

(ii) the interaction $D_{it}(W_{it} - \bar{W}^1_{it})$ between treatment status and deviations $(W_{it} - \bar{W}^1)$ in cohort, time, and cohort×time indicators, as well as time-varying controls, from their means among all treated units, and

(iii) treatment status, D_{it}.

Group- and time-averaged ATTs Difference-in-differences analyses sometimes also include estimates of cohort-specific ATTs $\beta^j = E(Y_{it}^1 - Y_{it}^0 | D_{it} = 1, C_i^j = 1)$, $j \in \mathcal{C}$. These ATTs are easy to estimate using the robust one-stage approach: simply replace D_{it} and $D_{it}(W_{it} - \bar{W}^1)$ with $D_{it}C_i^j$ and $D_{it}C_i^j(W_{it} - \bar{W}^{1j})$, $j \in \mathcal{C}$, where $\bar{W}^{1j}$ is the average of the covariates among treated observations corresponding to cohort j. These cohort-specific ATTs can also be averaged (perhaps weighting by relative cohort sizes) using commands available in standard software.[19] An analogous variation on the one-stage approach can be used to estimate calendar-time specific average treatment effects. Both of these approaches can also be implemented from within the two-stage framework (e.g., by replacing D_{it} with $D_{it}C_i^j$ in the second stage of the estimator to obtain cohort-specific average treatment effects).

5 Simulations

In order to illustrate the properties of the robust one-stage estimator, I present results from a number of Monte Carlo simulations corresponding to different data-generating processes. In each simulation, there are a total of five time periods, the treatment adoption times $T_i \in \{2, 3, 4, 5, 6\}$ are drawn from a discrete uniform distribution, and there are $N \in \{100, 500, 1000\}$ panel units (units with $T_i = 6$ are never treated). For each simulation, observed outcomes are determined by $Y_{it} = Y_{it}(0) + D_{it}\beta_{it}$, where

$$Y_{it}(0) = \lambda_i + \gamma_t + X_{it}'\delta + \varepsilon_{it},$$

$\lambda_i = T_i + \nu_i$, $\nu_i \sim N(0, 1)$, $\gamma_t \sim N(0, 1)$, and $\varepsilon_{it} \sim N(0, 3)$.

To illustrate the effectiveness of the estimator(s) in different settings, I run simulations across four different configurations of the data-generating process. In simulation (1), there are no control covariates, and the treatment effects $\beta_{it} \sim N(2, 1)$ are drawn independently of treatment timing. In simulation (2), there are no control covariates, and $\beta_{it} = t - T_i + 1$.

[19] For example, using Stata's `lincom` command.

In simulation (3), there is a time-varying control covariate $X_{it} \sim N(1,1)$, and the treatment effect depends on the value of the covariate according to $\beta_{it} = t - T_i + 1 + X_{it}/4$. Finally, in simulation (4), the covariates $X_{it} \sim N(\lambda_i/25, 1)$ are correlated with the unit fixed effects, and the treatment effects $\beta_{it} = (t - T_i + 1)X_{it}/4$ interact with the covariates multiplicatively. For each configuration of the simulations, I report results across 1,000 simulated datasets.

For each simulation, I estimate the overall and duration-specific ATTs using the two- and one-stage estimators.[20] To illustrate its numerical equivalence with the other estimators, I also estimate the overall ATT using a "manual" version of the one-stage estimator that computes the ATT as $\bar{W}_{it}^{1}\hat{\theta}_1$, as described in Section 3.1 (along with delta-method standard errors). For all estimators, I cluster all standard errors at the unit level.

The results from the primary simulations are summarized in Table 1. The entries under "2SDD," "1SDD," and "Manual" are rejection rates for size-.05 tests of the nulls that the overall and duration-specific ATTs equal their true values. In the interest of completeness, the table also reports average bias and RMSE across different simulation sizes (these statistics are only reported once per simulation because the one-stage, two-stage, and manual estimators are all numerically equivalent).

Across sample sizes and data-generating processes, the conclusion from the simulations is that the performance of the one- and two-stage estimators is highly comparable, with both producing rejection rates close to the theoretical 5% rate, particularly as the sample size increases. Beyond the results for rejection rates, the simulation results also show that both point estimates and inference for the one-stage approach are numerically equivalent to its "manual aggregation" variant, and that both the one- and two-stage approaches produce highly accurate estimates of the average effects of the treatment.

In the appendix, I present two additional sets of simulations. In Appendix Table 4, I present simulation results from a "fixed design" setting in which treatment status and the covariates are fixed, so that only the treatment effect (when it is random) and error term are

[20] I obtained the 2SDD estimates using Kyle Butts' Stata package `did2s` (Butts, 2021).

drawn anew for each simulated dataset. These fixed design simulations design show that both the one- and two-stage procedures can produce valid inference in this setting, and illustrate that the comparative performance of the one- and two-stage estimators has little to do with the fact that former estimator uses the sample value of $\bar{W}^1$ in place of the true treated mean $E^{\circ}(W_{it}|D_{it}=1)$ (which is equal to its sample value in a fixed design). As the results in the table show, both the absolute and relative performance of the one- and two-stage estimators are similar to in the random design. In Appendix Table 5, I present rejection rates for the extended two-way fixed effects (ETWFE) estimator (Wooldridge, 2025), which can also recover the two-stage point estimates, from the simulations summarized above in Table 1. Although my focus here is on illustrating the one-stage estimator as an alternative way of obtaining the two-stage estimates (as opposed to drawing comparisons among alternative ways of obtaining this estimate), the results in Appendix Table 5 show that the rejection rates from the one-stage and ETWFE estimators are very similar.[21]

6 Empirical applications

6.1 Cheng and Hoekstra (2013)

To illustrate the application of the one-stage estimator, I revisit Cheng and Hoekstra's (2013) analysis of the effects of strengthening the Castle Doctrine, also known as "stand your ground" laws, on violent crime. In this analysis, the key treatment status variable is an indicator for whether a state has adopted a stand your ground law in a given year, and the dependent variable is the state-level average of the log of the number of homicides committed per 100,000 people in that year. The data span 2000 to 2010, during which 21 states, divided

[21] I obtain the ETWFE estimates using the `jwdid` Stata package (Rios-Avila, Nagengast and Yotov, 2024). In models without covariates, the "simple" aggregate of the cohort×time-specific ETWFE treatment effect estimates are numerically identical to the one- and two-stage estimates (i.e., the coefficients on treatment status). With covariates, the estimates are different because the `jwdid` package includes covariate-specific time trends by default, and these are the specifications I use in the simulation results. However, I obtain similar results when I manually estimate a variant of the ETWFE specification that recreates the one- and two-stage estimates in settings with covariates.

into five treatment cohorts, adopt the treatment between 2006 and 2010.

Table 2 compares the results from one- and two-stage difference-in-difference estimates of the impact of these laws on homicides.[22] All of the point estimates discussed in this section are based on models that use cohort fixed effects and also control for the number of police employed in a given state-year, and all of the standard errors are clustered at the state level. Columns (1) and (4) report one- and two-stage estimates of the overall ATT. The point estimates are identical (for the reasons detailed above). The one-stage estimate has an estimated standard error of about .033, while the estimated standard error of the two-stage estimate is somewhat larger, at about .041, with both suggesting a statistically significant effect at conventional levels. From a practical perspective, the results from tests based on these methods agree with each other.

Columns (2) and (5) present estimates of the dynamic effects of the treatment. The one-stage estimates are derived from a regression of outcomes on cohort and time indicators, duration-specific treatment status indicators D^r, $r \in \{1, \ldots, 5\}$, and interactions $D^r(W_{it} - \bar{W}^r)$ between duration-specific treatment status and deviations in the cohort and time indicators from their duration-specific treated means. The two-stage estimates are based on second-stage regressions of adjusted outcomes on the duration-specific treatment status variables. I also include leads of treatment status in these second-stage regressions, the coefficients on which represent tests of parallel trends, since the two-stage approach makes this easy, and their inclusion does not affect the estimated coefficients on the duration-specific treatment-status indicators (note that these placebo tests are notionally different from those described in Section 4, which are implemented below).[23] As before, the point estimates are identical (also note that the one-stage estimates reported in column (2) do not include comparable tests of parallel trends). The one- and two-stage estimators agree on the statistical

[22] All of the results presented in Table 2 are weighted by state×year populations size. As a practical matter, when applying weights using the one-stage approach, it is necessary to use weighted regressions and deviations in the covariates from their weighted means.

[23] I obtained these estimates using the Stata package `did2s` (Butts, 2021) to obtain standard errors that reflect the first-stage estimation of the adjusted outcomes.

significance of four of the five duration-specific treatment effects, although they do not always agree on the precise level of significance (also note that the one-stage standard errors are not always smaller, though they tend to be). Both approaches lead to the same practical conclusions about the effect of the treatment.

Columns (3) and (6) present placebo tests of parallel trends that redefine treatment status to mean being four periods *before* the actual onset of the treatment. For the one-stage approach, I implement these tests using the regression specification detailed in Section 4, setting $k = -3$ (i.e., including four leads of treatment status, as well as interactions between these leads and de-meaned cohort and time indicators in the regression). For the two-stage approach, I implement these tests by estimating the first stage on a sample of observations that are never treated or more than four periods away from adopting the treatment, then including four leads of treatment status (in addition to the five lags) in the second-stage regression. In this case, tests based on the one- and two-stage approaches agree on the level of statistical significance for all of the placebo coefficients D^r, $r < 1$, although most of the one-stage standard errors remain smaller than their two-stage counterparts. These placebo tests also produce "collateral" estimates of the duration-specific treatment effects D^r, $r \geq 1$ (which compare treated observations to a smaller control sample of untreated observations). Here, the one- and two-stage approaches agree on the level of statistical significance of all but one of the coefficients. Setting comparisons between the one- and two-stage estimators aside, it is comforting that signs and significance levels of these collateral estimates agree with the full-sample estimates presented in columns (2) and (5), and that the placebo estimates presented in columns (3) and (6) generally agree with the alternative test of parallel trends presented for the two-stage estimator in column (5).

6.2 Autor (2003)

To further illustrate the application of the one-stage approach, I apply it, along with the two-stage estimator, to Autor's (2003) analysis of exceptions to the employment-at-will doctrine

on temporary employment. The key variables for this analysis are state×year-level log employment in the temporary help services industry and time-varying indicators for whether a state currently has an exception to the doctrine. Following Autor (2003), I limit the sample to the period spanning 1979-1995. I also drop one state which reversed the treatment after adopting it. After making these sample selections, a total of 34 states, organized into nine treatment cohorts, adopt the treatment between 1980 and 1988.

I then estimate the effects of the exception using both the one- and two-stage approaches, using cohort fixed effects and clustering the standard errors at the state level. The results are presented in Table 3. The one- and two-stage estimates of the overall ATT are shown in columns (1) and (4). The point estimates are, naturally, identical and, like Autor's original baseline estimate, they are also both statistically insignificant.[24] In this case, the standard error for the one-stage estimator is slightly larger than that for the two-stage estimator, though they are very similar.

Columns (2) and (5) of Table 3 present estimates of the duration-specific effects of the treatment (I only present estimates for the first ten post-treatment periods, although the comparative results for longer durations are similar). As in the Cheng and Hoekstra (2013) example, the one- and two-stage estimators both lead to the same practical conclusions about the dynamic effects of the policy (in this case, that they are not statistically different from zero). In this case, however, the one-stage standard errors remain slightly larger, and the differences between the one- and two-stage standard errors are much smaller for earlier durations.[25] As in the previous application, column (5) also includes second-stage estimates

[24]Autor's baseline estimate (in a model with no covariates) is .112 with a standard error of .099. He later presents statistically significant estimates from models that include covariates and state-specific linear trends. I do not estimate more complicated models here because my purpose is to compare the one- and two-stage estimates.

[25]One potential reason why there is greater divergence between some of the one- and two-stage standard errors in the Cheng and Hoekstra application than in the Autor application (and in the simulation results from Section 5) is that the Cheng and Hoekstra data contain a relatively large number of small cohorts (3/5 of the cohorts contain 2 or fewer states, compared to 4/9 for for the Autor data). Since the cluster-robust variance estimator is based on the sums of residuals across all units in a state, while the inclusion of cohort fixed effects means that all residuals in a cohort must sum to zero, these small cohort sizes increase the likelihood of mechanical bias in the clustered variance estimates. Despite this, the one- and two-stage estimators point to the same practical conclusions about the overall and dynamic effects of the treatment in

of the coefficients on the leads of treatment (which should be zero under parallel trends).

Finally, columns (3) and (6) of the table present one- and two-stage estimates of four periods of placebo effects (as well as ten periods of dynamic effects), i.e., after redefining the treatment variable to equal one if the treatment begins fewer than four periods in the future. To ensure that all states are observed prior to the beginning of their "placebo" treatment, I drop cohorts that adopted the treatment prior to 1983. Once again, the one- and two-stage estimates lead to the same conclusions (about parallel trends as well as the dynamic effects of the treatment), though in this application the one-stage standard errors tend to be a bit larger, and are more different from the corresponding two-stage standard errors than they are for the overall ATT or for earlier duration-specific effects in columns (2) and (4).

7 Conclusion

The problems associated with traditional regression-based difference-in-differences and event-study estimators have sparked the development of several alternative, robust estimators that reliably identify average treatment effect measures, even when adoption is staggered and average treatment effects are heterogenous. While all of the recently devised alternative estimators offer some robustness to treatment-effect heterogeneity under staggered adoption, some also have characteristics that make them particularly well-suited for specific environments.

The one-stage regression approach to difference-in-difference and event-study analysis developed in this paper is motivated intuitively by analogy to matching and regression methods for selection on observables. It is simple and easy to implement in any statistical package, using only a single regression that extends the traditional two-way fixed effects specification. It is also flexible, and can be extended to identify duration-, group- and time- specific average treatment effects, to test the plausibility of parallel trends, and to accommodate triple difference-in-differences designs. Because it is also identical to the two-stage difference-in-differences estimator (and its numerical equivalents), it enjoys many of the advantages of

both applications.

that estimator. It is, *inter alia*, robust to the presence of staggered adoption and heterogeneous average treatment effects, efficient (in some circumstances), and readily able to handle settings where parallel trends only holds conditional on time-varying covariates (and when treatment effects depend arbitrarily on those covariates). On the other hand, the two-stage approach is better-suited to models that include many unit fixed effects, and offers a broader menu of options for testing whether parallel trends holds, while estimators that produce intermediate estimates of cohort×period-specific treatment effects (such as those developed in Callaway and Sant'Anna, 2021 and Wooldridge, 2025) allow researchers to use those estimates to report any aggregate of them that they wish.

The one- and two-stage estimators perform similarly in simulation exercises and applied examples, with both leading to the same practical conclusions. In a simulation setting, both approaches to estimation result in rejection rates close to the theoretical target rates. In the empirical examples, not only do the results from the one- and two-stage estimators point to the same broad conclusions, they also exhibit a high degree of consistency: each methodology offers multiple ways of estimating average treatment effects and testing the validity of parallel trends, and the resulting estimates agree both within and across estimation approaches.

Appendix A: Large-sample properties of 2SDD

Proof of Proposition 1. *Consistency.* Redefine X_{it} to include time indicators, so that parallel trends implies

$$Y_{it} - \lambda_i - X_{it}'\delta = \beta D_{it} + \varepsilon_{it}.$$

Taking deviations from untreated means eliminates the λ_i, so that the second stage of the estimator can be expressed as

$$(Y_{it} - \bar{Y}_i^0) - (X_{it} - \bar{X}_i^0)'\delta = \beta D_{it} - (\varepsilon_{it} - \bar{\varepsilon}_i^0),$$

where $\bar{Y}_i^0 = [\sum_{t=1}^T (1 - D_{it})Y_{it}] / \sum_{t=1}^T (1 - D_{it})$, and similarly for the elements of $\bar{X}_i^0$ and the error $\bar{\varepsilon}_i^0$.[26] Note that since $\varepsilon_{it} = u_{it} + (\beta_{it} - \beta)D_{it}$ where D_{it} and X_{it} are strictly exogenous with respect to u_{it} in all untreated periods, those variables are also strictly exogenous with respect to ε_{it} in those periods.[27]

Next, write the first-stage estimate of δ as

$$\begin{aligned}
\hat{\delta} &= \delta + \left(\frac{1}{N} \sum_i \sum_t (1 - D_{it}) \ddot{X}_{it}^0 \ddot{X}_{it}^{0\prime} \right)^{-1} \left(\frac{1}{N} \sum_i \sum_t (1 - D_{it}) \ddot{X}_{it}^0 \ddot{\varepsilon}_{it}^0 \right) \\
&\xrightarrow{p} \delta + E\left(\sum_t (1 - D_{it}) \ddot{X}_{it}^0 \ddot{X}_{it}^{0\prime} \right)^{-1} E\left(\sum_t (1 - D_{it}) \ddot{X}_{it}^0 \ddot{\varepsilon}_{it}^0 \right) \\
&= \delta,
\end{aligned}$$

where the second line follows by the weak law of large numbers and the assumption that the inverse exists, and the third from the continuous mapping theorem and the strict exogeneity of X_{it}. Thus, the first stage estimate of δ (which includes the time fixed effects and the coefficients on the covariates) is consistent.

Next, write the second-stage estimate as

$$\begin{aligned}
\hat{\beta}^{2SDD} &= \beta + \left(\frac{1}{N} \sum_i \sum_t D_{it} \right)^{-1} \left(\frac{1}{N} \sum_i \sum_t D_{it} (\ddot{Y}_{it}^0 - \ddot{X}_{it}^{0\prime} \hat{\delta}) \right) \\
&= \beta + \left(\frac{1}{N} \sum_i \sum_t D_{it} \right)^{-1} \left(\frac{1}{N} \sum_i \sum_t D_{it} [\ddot{Y}_{it}^0 - \ddot{X}_{it}^{0\prime} \delta + \ddot{X}_{it}^{0\prime} (\delta - \hat{\delta})] \right) \\
&= \beta + \left(\frac{1}{N} \sum_i \sum_t D_{it} \right)^{-1} \left(\frac{1}{N} \sum_i \sum_t D_{it} \ddot{\varepsilon}_{it}^0 + \frac{1}{N} \sum_i \sum_t D_{it} \ddot{X}_{it}^{0\prime} (\delta - \hat{\delta})] \right) \\
&\xrightarrow{p} \beta + E\left(\sum_t D_{it} \right)^{-1} \left[E\left(\sum_t D_{it} \ddot{\varepsilon}_{it}^0 \right) + E\left(\sum_t D_{it} \ddot{X}_{it}^{0\prime} \right) \cdot 0 \right] = \beta,
\end{aligned}$$

[26] The foregoing argument can also be applied to a version of the estimator that uses cohort fixed effects by absorbing cohort dummies into X_{it}, then writing the second stage of the estimator as $Y_{it} - X_{it}'\delta = \beta D_{it} + \varepsilon_{it}$ (i.e., rather than expressing it in terms of deviations from untreated means).

[27] Also note that, since $\hat{\lambda}_i = \bar{Y}_i^0 - \bar{X}_i^{0\prime} \hat{\delta}$, this form of the estimator is equivalent to regressing $Y_{it} - \hat{\lambda}_i - X_i' \hat{\delta}$ on D_{it}.

where the last line follows from the weak law of large numbers, the continuous mapping theorem/Slutsky's theorem (together with the consistency of the first stage), the assumption that the inverse exists, and the exogeneity of D_{it}.[28]

Asymptotic normality. Consistency implies that

$$\sqrt{N}(\hat{\beta}^{2SDD} - \beta) = \left(\frac{1}{N}\sum_i\sum_t D_{it}\right)^{-1}\left[\frac{\sqrt{N}}{N}\sum_i\sum_t D_{it}\ddot{\varepsilon}^0_{it} + \left(\frac{1}{N}\sum_i\sum_t D_{it}\ddot{X}^{0\prime}_{it}\right)\sqrt{N}(\delta - \hat{\delta})\right]$$
$$\xrightarrow{d} N\left(0, A_0^{-1}B_0A_0^{-1}\right)$$

where $A_0 = E(\sum_t D_{tt})$ and

$$B_0 = E\left[\sum_t D_{it}(\ddot{\varepsilon}^0_{it})^2\right] + E\left(\sum_t D_{it}\ddot{X}^{0\prime}_{it}\right)$$
$$E\left(\sum_t (1-D_{it})\ddot{X}^0_{it}\ddot{X}^{0\prime}_{it}\right)^{-1} E\left[\sum_t \ddot{X}^0_{it}\ddot{\varepsilon}^0_i(1-D_{it})\ddot{\varepsilon}^0_i\ddot{X}^{0\prime}_{it}]\right] E\left(\sum_t (1-D_{it})\ddot{X}^0_{it}\ddot{X}^{0\prime}_{it}\right)^{-1}$$
$$E\left(\sum_t D_{it}\ddot{X}^0_{it}\right).$$

In the above, the convergence in distribution comes from a weak law of large numbers, a central limit theorem, the continuous mapping theorem (for convergence in distribution), and the fact that $\hat{\delta}$ is uncorrelated with $D_{it}\ddot{\varepsilon}^0_{it}$ because they are drawn from different samples. □

[28] Note that the existence of the inverse implies that the number of treated units increases without bound with N. For more detail on the last equality, recall that $\varepsilon_{it} = u_{it} + (\beta_{it} - \beta)D_{it}$, so $\ddot{\varepsilon}^0_{it} = \ddot{u}^0_{it} + (\beta_{it} - \beta)D_{it}$ and $E\left(\sum_t D_{it}\ddot{\varepsilon}_{it}\right) = E\left(\sum_t D_{it}\ddot{u}^0_{it} + \sum_t(\beta_{it} - \beta)D_{it}\right)$. The first term is zero by the strict exogeneity of D_{it} with respect to u_{it}, while the second term equals $\sum_t E(\beta_{it}D_{it}) - \beta\sum_t E(D_{it}) = \sum_t E(\beta_{it}D_{it}) - (\sum_t E(\beta_{it}D_{it})/\sum_t E(D_{it}))\sum_t E(D_{it}) = 0$ (this is another way of expressing the idea that $E^\circ[(\beta_{it}-\beta)D_{it}] = 0$ when the expectation is taken over all time periods).

Appendix B: What TWFE "tries" to estimate

Without covariates, parallel trends implies that (after conditioning on cohort membership) observed outcomes satisfy

$$Y_{cit} = \lambda_c + \gamma_t + \sum_{c=1}^{C} \sum_{t=1}^{T} \beta_{ct} D_{ct} + \varepsilon_{cit},$$

where D_{ct} is an indicator for whether members of cohort c are treated at time t and $E[\varepsilon_{cit}|c, t, (D_{ct})] = 0$ (and $\beta_{ct} = 0$ for periods when cohort c is not yet treated). The traditional TWFE DD specification is

$$Y_{cit} = \lambda_c + \gamma_t + \beta^{TWFE} D_{it} + u_{cit}.$$

Embedding this specification within the true model implied by parallel trends, the term $\sum_c \sum_t \beta_{ct} D_{ct}$ can be considered an omitted variable, while the "true" coefficient on D_{it} is zero.

By the (population) Frisch-Waugh-Lovell theorem and the usual omitted-variable bias formula, the population regression coefficient on D_{it} from the traditional specification can be recovered as the coefficient from a population regression of $\sum_c \sum_t \beta_{ct} D_{ct}$ on the residual $\tilde{D}_{it}$ from a population regression of D_{it} on cohort and time indicators, or

$$\frac{Cov\left(\sum_c \sum_t \beta_{ct} D_{ct}, \tilde{D}_{it}\right)}{Var(\tilde{D}_{it})} = \sum_c \sum_t \beta_{ct} \frac{Cov(D_{ct}, \tilde{D}_{it})}{Var(\tilde{D}_{it})} = \sum_c \sum_t \beta_{ct} \rho_{ct},$$

where ρ_{ct} is the coefficient on D_{it} from a population regression of D_{ct} on D_{it} and sets of cohort and time indicators. As has been noted in the literature (see de Chaisemartin and D'Haultfœuille, 2020; Borusyak, Jaravel and Spiess, 2021), one view of the problems associated with the traditional TWFE specification is that, because of the inclusion of cohort and time fixed effects, the ρ_{ct} are not constrained to be positive (and hence TWFE does not

identify a convex weighted average of cohort×time-average treatment effects).

This problem arises because the ρ_{ct} represent population coefficients from linear models of the probability of being treated as a member of cohort c at time t that are misspecified because they include cohort and time fixed effects in addition to treatment status. If the fixed effects were omitted from the regressions that produce these weights, the ρ_{ct} would instead correctly identify the population shares $\pi_{ct} = P(D_{ct} = 1|D_{it} = 1)$ of treated unit-times that correspond to cohort c and time t, and the coefficient on treatment status from the TWFE regression would identify $\sum_c \sum_t \beta_{ct}\pi_{ct}$, which is precisely the overall ATT $\beta = E^{\circ}(\beta_{it}|D_{it} = 1)$. In this sense, the overall ATT is what the traditional TWFE DD specification is "trying" to estimate, and what the estimators discussed in this paper succeed in estimating after overcoming the problems associated with the traditional approach.

Appendix C: Additional details on the large-sample properties of 1SDD

In this section, I present an alternative proof of the consistency of the estimator. Here, I assume that the estimator is implemented using unit fixed effects, although I also provide parenthetical justifications for the cohort fixed effects case as well.

The one-stage difference-in-difference regression estimate can alternatively be obtained by (i) estimating separate regressions on samples of untreated and treated observations

$$Y_{it} = \lambda_i^d + X_{it}'\delta^d + u_{it}^d, \quad d \in \{0, 1\}, \tag{6}$$

where the time fixed effects γ_t^d have been absorbed into the covariates X_{it}, (ii) forming the differences $\beta_i = \lambda_i^1 - \lambda_i^0$ and $\beta_x = \delta^1 - \delta^0$, and (iii) calculating the average effect of the

treatment on the treated as[29]

$$\frac{1}{\sum_i \sum_t D_{it}} \left(\sum_i \sum_t D_{it} \hat{\beta}_i + \sum_i \sum_t D_{it} X'_{it} \hat{\beta}_x \right).$$

Moreover, each of the treatment-status-specific regressions (6) can be estimated using a within transformation, and the unit fixed effects recovered as

$$\hat{\lambda}_i^d = \bar{Y}_i^d - \bar{X}_i^{d\prime} \hat{\delta}^d,$$

where $\bar{Y}_i^d = [\sum_t Y_{it} 1(D_{it} = d)]/[\sum_t 1(D_{it} = d)]$ (and similarly for the vector $\bar{X}_i^d$).

Since $\beta_i = \lambda_i^1 - \lambda_i^0$ where $E(\hat{\lambda}^d) = E(\bar{Y}_i^d) - E(\bar{X}_i^d)'\delta^d$, a law of large numbers and the continuous mapping theorem give[30]

$$\begin{aligned}
\hat{\beta}^{1SDD} &= \frac{1}{\sum_i \sum_t D_{it}} \left[\sum_i \left(\sum_t D_{it} \hat{\beta}_i \right) + \sum_i \left(\sum_t D_{it} X'_{it} \hat{\beta}_x \right) \right] \\
&= \frac{N}{\sum_i \sum_t D_{it}} \left[\frac{1}{N} \sum_i \left(\sum_t D_{it} [(\bar{Y}_i^1 - \bar{X}_i^{1\prime} \hat{\delta}^d) - (\bar{Y}_i^0 - \bar{X}_i^{0\prime} \hat{\delta}^0)] \right) + \frac{1}{N} \sum_i \left(\sum_t D_{it} X'_{it} \hat{\beta}_x \right) \right] \\
&\xrightarrow{p} \frac{1}{E(\sum_t D_{it})} E \left(\sum_t D_{it} \beta_i + \sum_t D_{it} X'_{it} \beta_x \right) \\
&= E^\circ(\beta_{it} | D_{it} = 1),
\end{aligned}$$

where the third line uses the fact that $N^{-1} \sum_i D_{it} (\bar{Y}_i^d - \bar{X}_i^{d\prime} \delta) \xrightarrow{p} E(\bar{Y}_i^d - \bar{X}_i^{d\prime} \delta | D_{it} = 1) P(D_{it} = 1) = E(\lambda_i | D_{it} = 1) P(D_{it} = 1) = E(D_{it} \lambda_i)$.

[29] In the case of cohort fixed effects, this can also be expressed as $\sum_{j \in \mathcal{C}} \bar{w}_j \hat{\beta}_j + \bar{X}^{1\prime} \hat{\beta}_x$ where $\bar{w}_j$ is the fraction of treated observations that correspond to cohort j and $\bar{X}^1$ is the vector of averages of the controls (including time indicators) among all treated observations.

[30] In the case of cohort fixed effects, $\bar{w}_j = [\sum_i \sum_t D_{it} 1(C_i = j)]/(\sum_i \sum_t D_{it}) \xrightarrow{p} E[\sum_t D_t 1(C_i = j)]/E(\sum_t D_t) \equiv \pi_j$ and $\bar{X}^1 = (\sum \sum_t D_{it} X_{it})/(\sum_i \sum_t D_{it}) \xrightarrow{p} E(\sum_t D_{it} X_{it})/E(\sum_t D_{it})$, so that $\hat{\beta}^{1SDD} = \sum_j \bar{w}_j \hat{\beta}_j + \bar{X}^1 \hat{\beta}_x \xrightarrow{p} \sum_j \pi_j \beta_j + E(\sum_t D_{it} X'_{it} \beta_x)/E(\sum_t D_{it})$.

References

Autor, David. 2003. "Outsourcing at will." Journal of Labor Economics, 21(1): 1-42.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2021. "Revisiting event study designs: Robust and efficient." Working paper.

Kyle Butts, 2021. "DID2S: Stata module to estimate a TWFE model using the two-stage difference-in-differences approach." Statistical Software Components S458951, Boston College Department of Economics, revised 28 Apr 2023.

Butts, Kyle and John Gardner. 2022. "did2s: Two-stage difference-in-differences." R Journal, 14 (3): 162-173.

Callaway, Brantly, and Pedro Sant'Anna. 2021. "Difference-in-differences with multiple time periods and an application on the minimum wage and employment." Journal of Econometrics, 225(2):200-230.

Caetano, Carolina, Brantly Callaway, Stroud Payne, and Hugo Sant'Anna Rodrigues. 2022. "Difference in differences with time-varying covariates." Working paper.

Cheng, Cheng and Mark Hoekstra. 2012. "Does strengthening self-defense law deter crime or escalate violence? Evidence from the expansions to the Castle Doctrine." Journal of Human Resources, 48(3):821-853.

Deb, Partha, Norton, Edward C., Wooldridge, Jeffrey M., and Jeffrey E. Zabel. "A flexible, heterogeneous treatment effects difference-in-differences estimator for repeated cross-sections." NBER Working Paper no. 33026.

Dube, Arandrajit, Girardi, Daniele, Jordà, Òscar and Alan M. Taylor. 2023. "A local projections approach to difference-in-differences event studies." NBER Working Paper no. 31184.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." American Economic Review, 110(9): 2964-2996.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." Journal of Econometrics, 225(2): 254-277.

Gardner, John. 2021. "Two-stage differences in differences." Working paper.

Gardner, John, Thakral, Neil, Tô, Linh and Luther Yap. 2023. "Two-stage differences in differences." Working paper.

Greene, William. 2018. *Econometric Analysis.* 8th Edition, Pearson Education Limited, London.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2023. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data." American Journal of Political Science.

James G. Mackinnon and Matthew D. Webb. 2017. "The wild bootstrap for few (treated) clusters." Econometrics Journal, 21(2): 114-135.

James G. MacKinnon, Morten Ørregaard Nielson, and Matthew D. Webb. 2023. "Fast and reliable jackknife and bootstrap methods for cluster-robust inference." Journal of Applied Econometrics, 38(5): 671-694.

Newey, Whitney K., and Daniel McFadden. 1994. "Large sample estimation and hypothesis testing." In Handbook of Econometrics, edited by R.F. Engle and D.L. McFadden, IV:2112–2245. Elsevier Science.

Fernando Rios-Avila, Arne J. Nagengast, and Yoto V. Yotov. 2024. "JWDID: Stata module to estimate Difference-in-Difference models using Mundlak approach." Statistical Software Components S459114, Boston College Department of Economics.

Sun, Liyang and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." Journal of Econometrics, 225(2):175-199.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. The MIT Press. Cambridge, Mass.: MIT Press.

Wooldridge, Jeffrey M. 2025. "Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators." Empirical Economics, 69: 2545-2587.

Tables

Table 1: Simulations

		Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	N	100	500	1000	100	500	1000	100	500	1000	100	500	1000
2SDD	D	0.064	0.047	0.045	0.051	0.044	0.050	0.057	0.050	0.048	0.058	0.052	0.047
	D^1	0.059	0.038	0.041	0.051	0.041	0.054	0.056	0.048	0.043	0.057	0.047	0.043
	D^2	0.063	0.049	0.042	0.054	0.045	0.050	0.059	0.050	0.045	0.059	0.051	0.045
	D^3	0.068	0.056	0.048	0.057	0.054	0.042	0.069	0.051	0.039	0.067	0.050	0.037
	D^4	0.073	0.051	0.055	0.060	0.053	0.048	0.066	0.052	0.058	0.067	0.051	0.052
1SDD	D	0.061	0.046	0.045	0.051	0.044	0.050	0.057	0.051	0.048	0.055	0.050	0.048
	D^1	0.059	0.038	0.041	0.047	0.041	0.054	0.056	0.048	0.043	0.056	0.046	0.043
	D^2	0.063	0.048	0.042	0.053	0.044	0.050	0.058	0.050	0.045	0.060	0.052	0.045
	D^3	0.063	0.055	0.047	0.051	0.054	0.042	0.070	0.051	0.039	0.070	0.050	0.039
	D^4	0.063	0.050	0.055	0.054	0.050	0.046	0.064	0.053	0.057	0.066	0.054	0.057
Manual	D	0.061	0.046	0.045	0.051	0.044	0.050	0.057	0.051	0.048	0.055	0.050	0.048
Bias	D	-0.018	-0.005	0.010	0.005	-0.004	-0.002	-0.022	-0.005	0.010	-0.022	-0.006	0.010
	D^1	0.012	0.010	-0.008	0.004	0.005	0.001	0.015	0.011	-0.007	0.016	0.012	-0.007
	D^2	0.012	0.002	-0.005	-0.010	0.007	0.001	0.015	0.002	-0.006	0.015	0.002	-0.005
	D^3	0.035	0.005	-0.017	-0.006	0.007	0.007	0.039	0.006	-0.017	0.039	0.007	-0.017
	D^4	0.016	-0.012	-0.019	-0.017	-0.011	-0.004	0.031	-0.012	-0.017	0.031	-0.012	-0.017
MSE	D	0.340	0.065	0.031	0.321	0.060	0.031	0.332	0.064	0.031	0.332	0.064	0.031
	D^1	0.280	0.049	0.025	0.254	0.046	0.026	0.266	0.047	0.024	0.266	0.047	0.024
	D^2	0.436	0.088	0.041	0.424	0.080	0.041	0.413	0.085	0.039	0.413	0.085	0.039
	D^3	0.768	0.144	0.068	0.700	0.136	0.064	0.728	0.137	0.066	0.727	0.137	0.066
	D^4	1.767	0.322	0.164	1.579	0.310	0.154	1.704	0.315	0.160	1.705	0.315	0.161

Notes: Rejection rates, bias, and MSE from 1,000 simulations. D denotes the overall average treatment effect, and D^r denotes the effect of being treated for r periods. In simulation (1), there are no covariates and the treatment effect is distributed $N(2,1)$. In simulation (2), there are no covariates, and the treatment effect $\beta_{it} = t - T_i + 1$, where T_i is the period when i adopts the treatment. In simulation (3), the covariate X_{it} is distributed $N(1,1)$ and the treatment effect is $\beta_{it} = t - T_i + 1 + X_{it}/4$. In simulation (4), the covariate X_{it} is distributed $N(\lambda_i/25, 1)$ where λ_i is unit i's individual fixed effect, and the treatment effect is $\beta_{it} = (t - T_i + 1)x_{it}/4$. "2SDD" is two-stage differences in differences and "1SDD" is robust one-stage difference in differences. "Manual" denotes estimates obtained by regressing outcomes on unit and time fixed effects, controls, and interactions between treatment status and those variables (without converting them to differences from treated means), then averaging the estimated interaction terms, multiplied by their estimated coefficients, over all treated observations (i.e., $D_{it}\bar{W}'_{it}\hat{\theta}_1$ in the notation of Section 3.1). Bias and RMSE are the same for all estimators.

Table 2: Empirical example (Cheng and Hoekstra, 2013)

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Overall ATT	D	0.0901***			0.0901**		
		(0.0332)			(0.0412)		
Dynamic effects	D^1		0.102***	0.101**		0.102***	0.101**
			(0.0297)	(0.0417)		(0.0355)	(0.0463)
	D^2		0.0754**	0.0725*		0.0754*	0.0725
			(0.0342)	(0.0410)		(0.0414)	(0.0449)
	D^3		0.0853*	0.0655		0.0853	0.0655
			(0.0446)	(0.0528)		(0.0557)	(0.0583)
	D^4		0.0771	0.0459		0.0771	0.0459
			(0.0509)	(0.0594)		(0.0599)	(0.0626)
	D^5		0.193***	0.133**		0.193***	0.133***
			(0.0556)	(0.0547)		(0.0533)	(0.0512)
Placebo effects	D^0			0.0192		0.0250	0.0192
				(0.0409)		(0.0239)	(0.0468)
	D^{-1}			-0.0373		-0.0219	-0.0373
				(0.0358)		(0.0169)	(0.0371)
	D^{-2}			-0.0229		-0.00113	-0.0229
				(0.0345)		(0.0146)	(0.0315)
	D^{-3}			-0.0218		-0.00123	-0.0218
				(0.0242)		(0.0157)	(0.0248)
	D^{-4}					0.00395	
						(0.0200)	
	D^{-5}					0.00965	
						(0.0175)	
	D^{-6}					0.0360**	
						(0.0173)	
	D^{-7}					-0.0529	
						(0.0430)	
	D^{-8}					-0.207***	
						(0.0436)	
	D^{-9}					-0.189***	
						(0.0237)	
	N	550	550	550	550	550	550

Notes: Columns (1) and (4) are 1SDD and 2SDD overall ATT estimates, respectively. Columns (2) and (5) contain the 1SDD and 2SDD estimates of the dynamic post-treatment effects (column (5) also contains the default 2SDD placebo tests of parallel trends). Columns (3) and (6) contain 1SDD and 2SDD placebo tests of parallel trends that assume the treatment begins four periods before its actual adoption (i.e., pretends that $D = 1$ if $r \geq -3$), as well as the post-treatment dynamic effects implied by this placebo assumption. All estimates control for police per capita, use cohort fixed effects, and cluster at the state level.

Table 3: Empirical example (Autor, 2003)

		One stage			Two stage		
		(1)	(2)	(3)	(4)	(5)	(6)
Overall ATT	D	0.0628			0.0628		
		(0.175)			(0.171)		
Dynamic effects	D^1		0.0843	-0.0557		0.0843	-0.0557
			(0.0680)	(0.184)		(0.0650)	(0.170)
	D^2		0.0908	-0.0190		0.0908	-0.0190
			(0.0879)	(0.208)		(0.0802)	(0.187)
	D^3		0.144	0.0901		0.144	0.0901
			(0.120)	(0.239)		(0.110)	(0.219)
	D^4		0.0159	-0.0746		0.0159	-0.0746
			(0.141)	(0.270)		(0.131)	(0.245)
	D^5		0.0789	0.0278		0.0789	0.0278
			(0.169)	(0.293)		(0.155)	(0.264)
	D^6		0.123	0.128		0.123	0.128
			(0.204)	(0.324)		(0.190)	(0.294)
	D^7		0.0911	0.0633		0.0911	0.0633
			(0.222)	(0.343)		(0.203)	(0.310)
	D^8		0.0877	0.0564		0.0877	0.0564
			(0.252)	(0.362)		(0.231)	(0.326)
	D^9		0.0378	0.0467		0.0378	0.0467
			(0.264)	(0.372)		(0.241)	(0.335)
	D^{10}		-0.0730	-0.143		-0.0730	-0.143
			(0.276)	(0.417)		(0.252)	(0.373)
Placebo effects	D^0			-0.0889		-0.0254	-0.0889
				(0.432)		(0.0311)	(0.394)
	D^{-1}			0.393		-0.0252	0.393
				(0.516)		(0.0291)	(0.455)
	D^{-2}			-0.133		-0.0338	-0.133
				(0.136)		(0.0437)	(0.125)
	D^{-3}			0.0145		0.0616	0.0145
				(0.0801)		(0.0391)	(0.0813)
	D^{-4}					0.00181	
						(0.0553)	
	D^{-5}					-0.0121	
						(0.0482)	
	D^{-6}					0.0718*	
						(0.0435)	
	D^{-7}					0.0983	
						(0.114)	
	D^{-8}					-0.151	
						(0.157)	
	N	714	714	476	714	714	476

Notes: Columns (1) and (4) are 1SDD and 2SDD overall ATT estimates, respectively. Columns (2) and (5) contain the 1SDD and 2SDD estimates of the dynamic post-treatment effects (column (5) also contains the default 2SDD placebo tests of parallel trends). Columns (3) and (6) contain 1SDD and 2SDD placebo tests of parallel trends that assume the treatment begins four periods before its actual adoption (i.e., pretends that $D = 1$ if $r \geq -3$), as well as the post-treatment dynamic effects implied by this placebo assumption. All estimates use cohort fixed effects, and cluster at the state level. Only the first ten post-treatment periods are shown (results for omitted periods are similar). Samples for columns (3) and (6) exclude observations that do not have at least five pre-treatment periods.

Table 4: Simulations (fixed design)

		Simulation 1			Simulation 2			Simulation 3			Simulation 4		
	N	100	500	1000	100	500	1000	100	500	1000	100	500	1000
2SDD	D	0.084	0.054	0.057	0.077	0.067	0.050	0.076	0.065	0.041	0.078	0.064	0.041
	D^1	0.086	0.053	0.050	0.088	0.046	0.046	0.084	0.047	0.039	0.085	0.047	0.039
	D^2	0.068	0.046	0.051	0.068	0.060	0.042	0.068	0.061	0.043	0.065	0.059	0.042
	D^3	0.104	0.047	0.053	0.109	0.055	0.052	0.105	0.054	0.052	0.101	0.053	0.050
	D^4	0.169	0.060	0.061	0.154	0.066	0.048	0.163	0.068	0.048	0.158	0.063	0.043
1SDD	D	0.078	0.051	0.057	0.071	0.065	0.046	0.069	0.064	0.040	0.067	0.064	0.039
	D^1	0.077	0.051	0.051	0.082	0.048	0.046	0.079	0.048	0.039	0.079	0.048	0.039
	D^2	0.063	0.043	0.049	0.065	0.057	0.041	0.067	0.058	0.042	0.067	0.058	0.042
	D^3	0.106	0.040	0.048	0.112	0.049	0.049	0.104	0.049	0.049	0.106	0.050	0.049
	D^4	0.123	0.055	0.058	0.114	0.061	0.047	0.128	0.061	0.047	0.130	0.060	0.047
Manual	D	0.078	0.051	0.057	0.071	0.065	0.046	0.069	0.064	0.040	0.067	0.064	0.039
Bias	D	0.003	0.010	0.015	0.158	-0.068	-0.044	0.029	0.007	0.000	0.029	0.010	-0.003
	D^1	-0.002	-0.019	-0.011	-0.029	-0.005	-0.012	-0.033	-0.004	-0.004	-0.032	-0.005	0.000
	D^2	0.023	0.002	0.000	-0.050	0.001	0.008	-0.051	-0.005	0.008	-0.048	-0.013	0.007
	D^3	-0.055	-0.006	-0.031	-0.031	-0.003	-0.007	0.101	-0.008	-0.009	0.090	-0.016	-0.009
	D^4	-0.013	-0.023	-0.040	-0.075	-0.021	-0.018	-0.149	-0.020	0.009	-0.142	-0.011	0.022
MSE	D	1.560	0.328	0.161	1.500	0.365	0.151	1.491	0.360	0.149	1.491	0.360	0.149
	D^1	1.434	0.271	0.131	1.339	0.258	0.121	1.348	0.257	0.121	1.348	0.257	0.121
	D^2	2.291	0.406	0.206	2.021	0.428	0.191	2.053	0.429	0.192	2.053	0.429	0.192
	D^3	4.292	0.667	0.331	4.297	0.722	0.326	4.429	0.720	0.327	4.427	0.720	0.327
	D^4	8.902	1.388	0.766	8.231	1.495	0.669	8.374	1.497	0.668	8.372	1.496	0.668

Notes: Rejection rates, bias, and MSE from 1,000 simulations. For all simulations, treatment status and covariates are fixed across samples (errors and treatment effects are redrawn with each sample). D denotes the overall average treatment effect, and D^r denotes the effect of being treated for r periods. In simulation (1), there are no covariates and the treatment effect is distributed $N(2,1)$. In simulation (2), there are no covariates, and the treatment effect $\beta_{it} = t - T_i + 1$, where T_i is the period when i adopts the treatment. In simulation (3), the covariate X_{it} is distributed $N(1,1)$ and the treatment effect is $\beta_{it} = t - T_i + 1 + X_{it}/4$. In simulation (4), the covariate X_{it} is distributed $N(\lambda_i/25, 1)$ where λ_i is unit i's individual fixed effect, and the treatment effect is $\beta_{it} = (t - T_i + 1)x_{it}/4$. "2SDD" is two-stage differences in differences and "1SDD" is robust one-stage difference in differences. "Manual" denotes estimates obtained by regressing outcomes on unit and time fixed effects, controls, and interactions between treatment status and those variables (without converting them to differences from treated means), then averaging the estimated interaction terms, multiplied by their estimated coefficients, over all treated observations (i.e., $D_{it}\bar{W}'_{it}\hat{\theta}_1$ in the notation of Section 3.1). Bias and RMSE are the same for all estimators.

Table 5: Simulations (ETWFE)

	Simulation 1			Simulation 2			Simulation 3			Simulation 4		
N	100	500	1000	100	500	1000	100	500	1000	100	500	1000
D	0.061	0.045	0.045	0.05	0.044	0.05	0.05	0.051	0.048	0.057	0.052	0.047
D^1	0.056	0.054	0.054	0.047	0.041	0.054	0.054	0.046	0.041	0.058	0.046	0.041
D^2	0.057	0.053	0.046	0.053	0.044	0.05	0.05	0.052	0.045	0.057	0.05	0.045
D^3	0.055	0.049	0.047	0.051	0.054	0.042	0.042	0.052	0.038	0.066	0.052	0.037
D^4	0.06	0.044	0.05	0.054	0.05	0.046	0.046	0.053	0.054	0.062	0.049	0.052

Notes: Rejection rates from 1,000 simulations from ETWFE estimator. Estimates obtained using the `jwdid` Stata package with default options. D denotes the overall average treatment effect, and D^r denotes the effect of being treated for r periods. In simulation (1), there are no covariates and the treatment effect is distributed $N(2,1)$. In simulation (2), there are no covariates, and the treatment effect $\beta_{it} = t - T_i + 1$, where T_i is the period when i adopts the treatment. In simulation (3), the covariate X_{it} is distributed $N(1,1)$ and the treatment effect is $\beta_{it} = t - T_i + 1 + X_{it}/4$. In simulation (4), the covariate X_{it} is distributed $N(\lambda_i/25, 1)$ where λ_i is unit i's individual fixed effect, and the treatment effect is $\beta_{it} = (t - T_i + 1)x_{it}/4$.