Online Appendix

Using Multiple Pre-treatment Periods to Improve Difference-in-Differences and Staggered Adoption Designs

Naoki Egami Soichiro Yamauchi

Political Analysis

Table of Contents

A Literature Review

A.1 Papers in *APSR* and *AJPS*

We conduct a review of the literature to assess current practices of the difference-in-differences (DID) design. Specifically, we search articles published in *American Political Science Review* and *American Journal of Political Science* from 2015 to 2019. Some of the papers we reviewed were accepted in 2019 and were officially published in 2020. Using Google Scholar, we find articles that contains any of the following keywords: "two-way fixed effect", "two-way fixed effects", "difference in difference" or "difference in differences." We then manually select articles from the list that uses the basic DID design and the staggered adoption design (see the main text for details about the first two design). This procedure left us with a total of 25 articles, 11 from APSR and 14 from AJPS. Table A1 and A2 show the articles in the list published in APSR and AJPS, respectively.

To determine the number of pre-treatment periods, we manually assess the listed articles. Among the 25 articles, 20 articles use the basic DID design, and 5 articles use the staggered adoption design. When a paper uses the basic DID design, we can determine the length of the pre-treatment periods from the data description and the time of the treatment assignment. On the other hand, the pre-treatment periods for the staggered adoption and the general design are set to the total number of time-periods available in the data, as the length of pre-treatment periods varies across units.

We found that most DID applications have less than 10 pre-treatment periods. The median number of pre-treatment periods is 3.5 and, the mean number of pre-treatment periods is about 6 after removing one unique study that has more than 100 pre-treatment periods.

A.2 Examples of Two Common Approaches

As we wrote in Section 1, there are several different popular ways to analyze the DID design with multiple pre-treatment periods. One common approach is to apply the two-way fixed effects regression to the entire time periods, and supplement it with alternative model specifications by including time-trends or leads of the treatment variable to assess possible violations of the parallel trends assumption. Examples include Dube et al. (2013); Truex (2014); Earle and Gehlbach (2015); Hall (2016); Larreguy and Marshall (2017). Another is to stick with the two-time-period DID and limit the use of additional pre-treatment periods only to the assessment of pre-treatment trends. Examples include Ladd and Lenz (2009); Bechtel and Hainmueller (2011); Bullock and Clinton (2011); Keele and Minozzi (2013); Garfias (2018). Note that we list exemplary papers here and thus, we also include papers from journals other than APSR and AJPS.

Authors	Year	Title
O'brien, D. Z., & Rickne J.	2016	Gender Quotas And Women's Political Leadership
Garfias, F.	2018	Elite Competition and State Capacity Development: Theory and Evidence From Post-Revolutionary Mexico.
Martin, G. J., & Mccrain, J.	2019	Local News And National Politics
Blom-Hansen, J., Houlberg, K., Serritzlew, S., & Treisman, D.	2016	Jurisdiction Size and Local Government Policy Expenditure: Assessing The Effect of Municipal Amalgamation
Clinton, J. D., & Sances, M. W.	2018	The Politics of Policy: The Initial Mass Political Effects of Medicaid Expansion in The States
Malesky, E. J. , Nguyen, C. V., & Tran, A.	2014	The Impact of Recentralization on Public Services: A Difference-in-Differences Analysis of the Abolition of Elected Councils in Vietnam.
Larsen, M. V., Hjorth, F., Dinesen, P. T., & Sønderskov, K. M.	2019	When Do Citizens Respond Politically to The Local Economy? Evidence From Registry Data on Local Housing Markets
Becher, M., & González, I. M.	2019	Electoral Reform and Trade-Offs in Representation
Selb, P., & Munzert, S.	2018	Examining A Most Likely Case for Strong Campaign Effects
Enos, R. D., Kaufman, A. R., & Sands, M. L.	2019	Can Violent Protest Change Local Policy Support?
Vasiliki Fouka	2019	How Do Immigrants Respond to Discrimination?

Table A1: DID papers on APSR.

Authors	**Year**	**Title**
Bechtel, M. M., Hangartner, D., & Schmid, L.	2016	Does compulsory voting increase support for leftist policy?
Bisgaard, M., & Slothuus, R.	2018	Partisan elites as culprits? How party cues shape partisan perceptual gaps.
Bischof, D., & Wagner, M.	2019	Do voters polarize when radical parties enter parliament?
Dewan, T., Meriläinen, J., & Tukiainen, J.	2020	Victorian voting: The origins of party orientation and class alignment.
Earle, J. S., & Gehlbach, S.	2015	The Productivity Consequences of Political Turnover: Firm-Level Evidence from Ukraine's Orange Revolution.
Enos, R. D.	2016	What the demolition of public housing teaches us about the impact of racial threat on political behavior.
Gingerich, D. W.	2019	Ballot Reform as Suffrage Restriction: Evidence from Brazil's Second Republic.
Hainmueller, J, & Hangartner, D.	2019	Does direct democracy hurt immigrant minorities? Evidence from naturalization decisions in Switzerland.
Holbein, J. B., & Hillygus, D. S.	2016	Making young voters: the impact of preregistration on youth turnout.
Jäger, K.	2020	When Do Campaign Effects Persist for Years? Evidence from a Natural Experiment.
Lindgren, K. O., Oskarsson, S., & Dawes, C. T.	2017	Can Political Inequalities Be Educated Away? Evidence from a Large-Scale Reform.
Lopes da Fonseca, M.	2017	Identifying the source of incumbency advantage through a constitutional reform.
Paglayan, AS.	2019	Public-Sector Unions and the Size of Government
Pardos-Prado, S., & Xena, C.	2019	Skill specificity and attitudes toward immigration.

Table A2: DID papers on AJPS.

B Comparison with Three Existing Methods

This section clarifies relationships between our proposed double DID and three existing methods: the two-way fixed effects estimator, the sequential DID estimator, and synthetic control methods.

B.1 Relationship with Two-Way Fixed Effects Estimator

While we contrast the double DID with the two-way fixed effects estimator throughout the paper, we summarize our discussion here. First, in the basic DID design, the two-way fixed effects estimator is a special case of the double DID with a specific choice of the weight matrix $\mathbf{W}$ (see Table 1). Therefore, whenever the two-way fixed effects estimator is consistent for the ATT, the double DID is a more efficient, consistent estimator of the ATT. This is because the double DID can choose the optimal weight matrix via the GMM, while the two-way fixed effects uses the pre-determined equal weights over time. Second, in the SA design, a large number of recent papers show that the widely-used two-way fixed effects estimator are in general inconsistent for the ATT due to treatment effect heterogeneity and implicit parametric assumptions (Strezhnev, 2018; Athey and Imbens, 2021; Imai and Kim, 2021; Sun and Abraham, 2020). In contrast, the proposed double DID in the SA design generalizes nonparametric DID estimators to allow for treatment effect heterogeneity, and thus, it does not suffer from the same problem.

B.2 Relationship with Sequential DID Estimator

Our double DID estimator contains the sequential DID estimator (e.g., Lee, 2016; Mora and Reggio, 2019) as a special case. Our proposed double DID improves over the sequential DID estimator in two ways. First, when the parallel trends assumption holds, the double DID optimally combine the standard DID and the sequential DID to improve efficiency, and it is not equal to the sequential DID. Therefore, it avoids a dilemma of the sequential DID — it is consistent under the parallel trends-in-trends assumption (weaker than the parallel trends assumption), but is less efficient when the parallel trends assumption holds. Second, while the sequential DID estimator has only been available for the basic DID design where treatment assignment happens only once, we generalize it to the staggered adoption design and further incorporate it into our staggered-adoption double DID estimator (Section 4).

B.3 Relationship with Synthetic Control Methods

Another relevant popular class of methods is the synthetic control methods. While the method was originally designed to estimate the causal effect on a *single* treated unit, recent extensions allow for multiple treated units and the staggered adoption design (e.g., Xu, 2017; Ben-Michael et al., 2018; Hazlett and Xu, 2018; Athey et al., 2021). Despite a wide variety of innovative extensions, they all share the same core feature: they require long pre-treatment periods to accurately estimate a pre-treatment trajectory of the treated units. For example, Xu (2017) recommends collecting more than ten pre-treatment periods. In contrast, the proposed double DID can be applied as long as there are more than one pre-treatment periods, and is better suited when there are a small to moderate number of pre-treatment periods.

When there are a large number of pre-treatment periods (i.e., long enough to apply the synthetic control methods), we recommend to apply both the synthetic control methods and proposed double DID, and evaluate robustness across those approaches. This is important because they rely on different identification assumptions. In fact, we show in Section H.2, the double DID can recover credible estimates similar to more flexible variants of synthetic control methods even when there are a large number of pre-treatment periods. This robustness provides researchers with additional credibility for their causal estimates and underlying assumptions.

C Nonparametric Equivalence to Regression Estimators

In this section, we provide results on the nonparametric connection between regression estimators and the three DID estimators we discussed in the paper. This section provides methodological foundations for our main methodological contributions, which we prove in Sections E.2 and E.3.

C.1 Standard DID

In practice, we can compute the DID estimator via a linear regression. We regress the outcome Y_{it} on an intercept, treatment group indicator G_i, time indicator I_t (equal to 1 if post-treatment and 0 otherwise) and the interaction between the treatment group indicator and the time indicator $G_i \times I_t$.

$$Y_{it} \sim \alpha + \theta G_i + \gamma I_t + \beta (G_i \times I_t), \tag{A.2}$$

where $(\alpha, \theta, \gamma, \beta)$ are corresponding coefficients. In this case, a coefficient of the interaction term β is numerically equal to the DID estimator, $\widehat{\tau}_{\texttt{DID}}$. Importantly, the linear regression is used here only to compute the nonparametric DID estimator (equation (3)), and thus it does not require any parametric modeling assumption such as constant treatment effects. Furthermore, when we analyze panel data in which the same units are observed repeatedly over time, we obtain exactly the same estimate via a linear regression with unit and time fixed effects. This numerical equivalence in the two-time-period case is often the justification of the two-way fixed effects regression as the DID design (Angrist and Pischke, 2008). The above equivalence is formally shown below for completeness.

C.1.1 Repeated Cross-Sectional Data

For the later use in this Appendix, we report the well-known result that the standard DID estimator $\widehat{\tau}_{\texttt{DID}}$ (equation (3)) is equivalent to coefficient $\widehat{\beta}$ in the regression estimator (equation (A.2)) (Abadie, 2005).

We define O_{it} to be an indicator variable taking the value 1 when individual i is observed in time period t. Using this notation, we prove the following result.

Result 1 (Nonparametric Equivalence of the Standard DID and Regression Estimator) *We write the linear regression estimator (equation* (A.2)*) as a solution to the following least squares problem.*

$$(\widehat{\alpha}, \widehat{\theta}, \widehat{\gamma}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=1}^{2} O_{it} \Big\{ Y_{it} - \alpha - \theta G_i - \gamma I_t - \beta (G_i \times I_t) \Big\}^2.$$

Then, $\widehat{\tau}_{\texttt{DID}} = \widehat{\beta}$.

Proof. By solving the least squares problem, we obtain the following solutions:

$$\widehat{\alpha} = \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}$$

$$\widehat{\theta} = \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}$$

$$\widehat{\gamma} = \frac{\sum_{i:\, G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}$$

$$\widehat{\beta} = \left(\frac{\sum_{i:\, G_i=1} Y_{i2}}{n_{12}} - \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} \right),$$

which completes the proof. □

C.1.2 Panel Data

Again, for the later use in the Appendix, we report the well-known result that the standard DID estimator $\widehat{\tau}_{\texttt{DID}}$ (equation (3)) is equivalent to coefficient $\widehat{\beta}$ in the two-way fixed effects regression estimator in the panel data setting (Abadie, 2005).

Result 2 (Nonparametric Equivalence of the Standard DID and Two-way Fixed Effects Regression Estimator)

We can write the two-way fixed effects regression estimator as a solution to the following least squares problem.

$$(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta}) = \text{argmin} \sum_{i=1}^{n} \sum_{t=1}^{2} (Y_{it} - \alpha_i - \delta_t - \beta D_{it})^2.$$

Then, $\widehat{\tau}_{\texttt{DID}} = \widehat{\beta}$.

Proof. First we define the demeaned treatment and outcome variables, $\overline{Y}_i = \sum_{t=1}^{2} Y_{it}/2$, $\overline{Y}_t = \sum_{i=1}^{n} Y_{it}/n$, $\overline{Y} = \sum_{i=1}^{n} \sum_{t=1}^{2} Y_{it}/2n$, $\overline{D}_i = \sum_{t=1}^{2} D_{it}/2$, $\overline{D}_t = \sum_{i=1}^{n} D_{it}/n$, and $\overline{D} = \sum_{i=1}^{n} \sum_{t=1}^{2} D_{it}/2n$.

Given these transformed variables, we can transform the least squares problem into a well-known demeaned form.

$$\widehat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} \sum_{t=1}^{2} (\widetilde{Y}_{it} - \beta \widetilde{D}_{it})^2$$

where $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}$ and $\widetilde{D}_{it} = D_{it} - \overline{D}_i - \overline{D}_t + \overline{D}$. Using this notation, we can express $\widehat{\beta}$ as

$$\widehat{\beta} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{2} \widetilde{D}_{it} \widetilde{Y}_{it}}{\sum_{i=1}^{n} \sum_{t=1}^{2} \widetilde{D}_{it}^2}$$

where $\widetilde{D}_{it}$ takes the following form,

$$\widetilde{D}_{it} = \begin{cases} 1/2 \cdot n_0/n & \text{if } G_i = 1, t = 2 \\ -(1/2) \cdot n_0/n & \text{if } G_i = 1, t = 1 \\ -(1/2) \cdot n_1/n & \text{if } G_i = 0, t = 2 \\ 1/2 \cdot n_1/n & \text{if } G_i = 0, t = 1, \end{cases}$$

where $n_1 = \sum_{i=1}^{n} G_i$ and $n_0 = \sum_{i=1}^{n} (1 - G_i)$. Then, the numerator can be written as

$$\sum_{i=1}^{n} \sum_{t=1}^{2} \widetilde{D}_{it} \widetilde{Y}_{it} = \frac{n_0}{2n} \left\{ \sum_{i=1}^{n} G_i \widetilde{Y}_{i2} - \sum_{i=1}^{n} G_i \widetilde{Y}_{i1} \right\} - \frac{n_1}{2n} \left\{ \sum_{i=1}^{n} (1 - G_i) \widetilde{Y}_{i2} - \sum_{i=1}^{n} (1 - G_i) \widetilde{Y}_{i1} \right\}$$

and the denominator is given as

$$\sum_{i=1}^{n}\sum_{t=1}^{2}\widetilde{D}_{it}^2 = 2n_1\left(\frac{n_0}{2n}\right)^2 + 2n_0\left(\frac{n_1}{2n}\right)^2 = \frac{n_1 n_0}{2n}.$$

Combining both terms, we get

$$\begin{aligned}
\widehat{\beta} &= \frac{\sum_{i=1}^{n}\sum_{t=1}^{2}\widetilde{D}_{it}\widetilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{2}\widetilde{D}_{it}^2} \\
&= \frac{1}{n_1}\left\{\sum_{i=1}^{n} G_i\widetilde{Y}_{i2} - \sum_{i=1}^{n} G_i\widetilde{Y}_{i1}\right\} - \frac{1}{n_0}\left\{\sum_{i=1}^{n}(1-G_i)\widetilde{Y}_{i2} - \sum_{i=1}^{n}(1-G_i)\widetilde{Y}_{i1}\right\} \\
&= \frac{1}{n_1}\sum_{i=1}^{n} G_i(Y_{i2}-Y_{i1}) - \frac{1}{n_0}\sum_{i=1}^{n}(1-G_i)(Y_{i2}-Y_{i1}) \\
&= \widehat{\tau}_{\texttt{DID}},
\end{aligned}$$

which concludes the proof. □

C.2 Extended DID

C.2.1 Repeated Cross-Sectional Data

We consider a case in which there are two pre-treatment periods $t = \{0, 1\}$ and one post-treatment period $t = 2$. Using this notation, we report the following result.

Result 3 (Nonparametric Equivalence of the Extended DID and Regression Estimator) *We focus on a linear regression estimator that is a solution to the following least squares problem.*

$$(\widehat{\theta}, \widehat{\gamma}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n}\sum_{t=0}^{2} O_{it}\left(Y_{it} - \theta G_i - \gamma_t - \beta D_{it}\right)^2.$$

Then, $\widehat{\beta} = \lambda\widehat{\tau}_{\texttt{DID}} + (1-\lambda)\widehat{\tau}_{\texttt{DID(2,0)}}$ *where*

$$\lambda = \frac{n_{11}n_{01}(n_{10}+n_{00})}{n_{11}n_{01}(n_{10}+n_{00}) + n_{10}n_{00}(n_{11}+n_{01})},$$

$$1-\lambda = \frac{n_{10}n_{00}(n_{11}+n_{01})}{n_{11}n_{01}(n_{10}+n_{00}) + n_{10}n_{00}(n_{11}+n_{01})}.$$

When the sample size of each group is fixed over time, i.e., $n_{11} = n_{10}$ *and* $n_{01} = n_{00}$*,* $\lambda = 1/2$ *and therefore,* $\widehat{\beta}$ *is equivalent to the extended DID estimator of equal weights in equation* (8).

Proof. By solving the least squares problem, we obtain

$$\widehat{\theta} = \lambda\left(\frac{\sum_{i:\,G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\,G_i=0} Y_{i1}}{n_{01}}\right) + (1-\lambda)\left(\frac{\sum_{i:\,G_i=1} Y_{i0}}{n_{10}} - \frac{\sum_{i:\,G_i=0} Y_{i0}}{n_{00}}\right)$$

$$\widehat{\gamma}_2 = \frac{\sum_{i:\,G_i=0} Y_{i2}}{n_{02}}$$

$$
\begin{aligned}
\widehat{\gamma}_1 &= \frac{\sum_{i:\,G_i=1} Y_{i1} + \sum_{i:\,G_i=0} Y_{i1}}{n_{11}+n_{01}} - \frac{n_{11}}{n_{11}+n_{01}}\widehat{\theta} \\
\widehat{\gamma}_0 &= \frac{\sum_{i:\,G_i=1} Y_{i0} + \sum_{i:\,G_i=0} Y_{i0}}{n_{10}+n_{00}} - \frac{n_{10}}{n_{10}+n_{00}}\widehat{\theta} \\
\widehat{\beta} &= \lambda \left\{ \left(\frac{\sum_{i:\,G_i=1} Y_{i2}}{n_{12}} - \frac{\sum_{i:\,G_i=1} Y_{i1}}{n_{11}} \right) - \left(\frac{\sum_{i:\,G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\,G_i=0} Y_{i1}}{n_{01}} \right) \right\} \\
&\quad + (1-\lambda) \left\{ \left(\frac{\sum_{i:\,G_i=1} Y_{i2}}{n_{12}} - \frac{\sum_{i:\,G_i=1} Y_{i0}}{n_{10}} \right) - \left(\frac{\sum_{i:\,G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\,G_i=0} Y_{i0}}{n_{00}} \right) \right\},
\end{aligned}
$$

which completes the proof. □

C.2.2 Panel Data

We report that the extended DID estimator $\widehat{\tau}_{\texttt{e-DID}}$ (equation (8)) (equal weights: $\lambda = 1/2$) is equivalent to the estimated coefficient $\widehat{\beta}$ in the two-way fixed effects regression estimator in the panel data setting with $t = \{0, 1, 2\}$.

Result 4 (Nonparametric Equivalence of the Extended DID and Two-way Fixed Effects Regression Estimator) *We can write the two-way fixed effects regression estimator as a solution to the following least squares problem.*

$$
(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{2} (Y_{it} - \alpha_i - \delta_t - \beta D_{it})^2.
$$

Then, $\widehat{\tau}_{\texttt{e-DID}} = \widehat{\beta}$.

Proof. First we define $\overline{Y}_i = \sum_{t=0}^{2} Y_{it}/3$, $\overline{Y}_t = \sum_{i=1}^{n} Y_{it}/n$, $\overline{Y} = \sum_{i=1}^{n}\sum_{t=0}^{2} Y_{it}/3n$, $\overline{D}_i = \sum_{t=0}^{2} D_{it}/3$, $\overline{D}_t = \sum_{i=1}^{n} D_{it}/n$, and $\overline{D} = \sum_{i=1}^{n}\sum_{t=0}^{2} D_{it}/3n$. Then, we can write the two-way fixed effects estimator as a two-way demeaned estimator,

$$
\widehat{\beta} = \operatorname*{argmin}_{\beta} \sum_{i=1}^{n} \sum_{t=0}^{2} (\widetilde{Y}_{it} - \beta \widetilde{D}_{it})^2 = \frac{\sum_{i=1}^{n}\sum_{t=0}^{2} \widetilde{D}_{it}\widetilde{Y}_{it}}{\sum_{i=1}^{n}\sum_{t=0}^{2} \widetilde{D}_{it}^2},
$$

as in Result 2, where $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i - \overline{Y}_t + \overline{Y}$ and $\widetilde{D}_{it} = D_{it} - \overline{D}_i - \overline{D}_t + \overline{D}$. Importantly, $\widetilde{D}_{it}$ takes the following form:

$$
\widetilde{D}_{it} = \begin{cases} 2/3 \cdot n_0/n & \text{if } G_i = 1, t = 2 \\ -1/3 \cdot n_0/n & \text{if } G_i = 1, t = 0, 1 \\ -2/3 \cdot n_1/n & \text{if } G_i = 0, t = 2 \\ 1/3 \cdot n_1/n & \text{if } G_i = 0, t = 0, 1, \end{cases}
$$

where $n_1 = \sum_{i=1}^{n} G_i$ and $n_0 = \sum_{i=1}^{n}(1 - G_i)$. Then, the numerator can be written as

$$
\sum_{i=1}^{n} \sum_{t=0}^{2} \widetilde{D}_{it}\widetilde{Y}_{it}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} G_i \left(\frac{2n_0}{3n}\right) \widetilde{Y}_{i2} - \sum_{i=1}^{n}\sum_{t=0}^{1} G_i \left(\frac{n_0}{3n}\right) \widetilde{Y}_{it} + \sum_{i=1}^{n} (1-G_i) \left(\frac{-2n_1}{3n}\right) \widetilde{Y}_{i2} + \sum_{i=1}^{n}\sum_{t=0}^{1} (1-G_i) \left(\frac{n_1}{3n}\right) \widetilde{Y}_{it} \\
&= \sum_{i=1}^{n} G_i \left(\frac{n_0}{3n}\right) \{\widetilde{Y}_{i2} - \widetilde{Y}_{i1}\} + \sum_{i=1}^{n} G_i \left(\frac{n_0}{3n}\right) \{\widetilde{Y}_{i2} - \widetilde{Y}_{i0}\} \\
&\quad - \left\{ \sum_{i=1}^{n} (1-G_i) \left(\frac{n_1}{3n}\right) \{\widetilde{Y}_{i2} - \widetilde{Y}_{i1}\} + \sum_{i=1}^{n} (1-G_i) \left(\frac{n_1}{3n}\right) \{\widetilde{Y}_{i2} - \widetilde{Y}_{i0}\} \right\} \\
&= \frac{n_0}{3n} \left\{ \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i1}\} + \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i0}\} \right\} - \frac{n_1}{3n} \left\{ \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i1}\} + \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i0}\} \right\}.
\end{aligned}
$$

The denominator can be written as

$$
\sum_{i=1}^{n}\sum_{t=0}^{2} \widetilde{D}_{it}^2 = \frac{n_0 n_1}{n} \cdot \frac{2}{3}.
$$

Combining the two terms, we have

$$
\begin{aligned}
\widehat{\beta} &= \frac{1}{2n_1} \left\{ \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i1}\} + \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i0}\} \right\} \\
&\quad - \frac{1}{2n_0} \left\{ \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i1}\} + \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i0}\} \right\} \\
&= \frac{1}{2} \left\{ \frac{1}{n_1} \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i1}\} - \frac{1}{n_0} \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i1}\} \right\} \\
&\quad + \frac{1}{2} \left\{ \frac{1}{n_1} \sum_{i=1}^{n} G_i \{Y_{i2} - Y_{i0}\} - \frac{1}{n_0} \sum_{i=1}^{n} (1-G_i) \{Y_{i2} - Y_{i0}\} \right\} \\
&= \frac{1}{2} \widehat{\tau}_{\texttt{DID}} + \frac{1}{2} \widehat{\tau}_{\texttt{DID(2,0)}}.
\end{aligned}
$$

By solving the least squares problem, we also obtain

$$
\begin{aligned}
\widehat{\alpha}_i &= \overline{Y}_i - \overline{Y} - \overline{Y}_{t=0} + \widehat{\beta}(\overline{D} - \overline{D}_{t=0}) \\
\widehat{\delta}_t &= \overline{Y}_t - \overline{Y}_{t=0} + \widehat{\beta}(\overline{D}_{t=0} - \overline{D}_t)
\end{aligned}
$$

□

C.3 Sequential DID

The sequential DID estimator is connected to a widely used regression estimator. In particular, the sequential DID estimator (equation (10)) can be computed as a linear regression in which we replace the outcome Y_{it} with a transformed outcome. In panel data, we replace the original outcome with its first difference $Y_{it} - Y_{i,t-1}$ so that we use changes instead of levels. In repeated cross-sectional data, we use the following linear regression.

$$
\Delta Y_{it} \sim \alpha_s + \theta_s G_i + \gamma_s I_t + \beta_s (G_i \times I_t), \tag{A.3}
$$

where $\Delta Y_{it} = Y_{it} - (\sum_{i:\, G_i=1} Y_{i,t-1})/n_{1,t-1}$ if $G_i = 1$ and $\Delta Y_{it} = Y_{it} - (\sum_{i:\, G_i=0} Y_{i,t-1})/n_{0,t-1}$ if $G_i = 0$. Coefficients are denoted by $(\alpha_s, \theta_s, \gamma_s, \beta_s)$. In this case, a coefficient in front of the interaction term β_s is numerically identical to the sequential DID estimator. We provide the proof of this equivalence for both panel and repeated cross-sectional data settings below.

C.3.1 Repeated Cross-Sectional Data

We clarify that the sequential DID estimator $\widehat{\tau}_{\texttt{s-DID}}$ (equation (10)) is equivalent to a coefficient in a regression estimator with transformed outcomes.

Result 5 (Nonparametric Equivalence of the Sequential DID and Regression Estimator) *We focus on a linear regression estimator with a transformed outcome.*

$$(\widehat{\alpha}, \widehat{\theta}, \widehat{\gamma}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=1}^{2} O_{it} \Big\{ \Delta Y_{it} - \alpha - \theta G_i - \gamma I_t - \beta (G_i \times I_t) \Big\}^2,$$

where

$$\Delta Y_{it} = \begin{cases} Y_{i2} - \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} & \text{if } G_i = 1, t = 2 \\ Y_{i1} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}} & \text{if } G_i = 1, t = 1 \\ Y_{i2} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} & \text{if } G_i = 0, t = 2 \\ Y_{i1} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}} & \text{if } G_i = 0, t = 1. \end{cases}$$

Then, $\widehat{\tau}_{\texttt{s-DID}} = \widehat{\beta}$.

Proof. Using Result 1, we obtain

$$\begin{aligned} \widehat{\beta} &= \left(\frac{\sum_{i:\, G_i=1} \Delta Y_{i2}}{n_{12}} - \frac{\sum_{i:\, G_i=1} \Delta Y_{i1}}{n_{11}} \right) - \left(\frac{\sum_{i:\, G_i=0} \Delta Y_{i2}}{n_{02}} - \frac{\sum_{i:\, G_i=0} \Delta Y_{i1}}{n_{01}} \right) \\ &= \left\{ \left(\frac{\sum_{i:\, G_i=1} Y_{i2}}{n_{12}} - \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} \right) \right\} \\ &\qquad - \left\{ \left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}} \right) \right\}, \end{aligned}$$

which completes the proof. □

Next, we clarify that the sequential DID estimator $\widehat{\tau}_{\texttt{s-DID}}$ (equation (10)) is also equivalent to a coefficient in a regression estimator with group-specific time trends. Mora and Reggio (2019) derive similar results by making the parametric assumption of the conditional expectations. We prove nonparametric equivalence without making any assumptions about conditional expectations.

Result 6 (Nonparametric Equivalence of the Sequential DID and Regression Estimator with Group-Specific Time Trends) *We focus on a linear regression estimator with group-specific time trends.*

$$(\widehat{\theta}, \widehat{\gamma}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{2} O_{it} \Big\{ Y_{it} - \theta_0 G_i - \theta_1 (G_i \times t) - \gamma_t - \beta D_{it} \Big\}^2.$$

Then, $\widehat{\tau}_{\texttt{s-DID}} = \widehat{\beta}$.

Proof. By solving the least squares problem, we obtain

$$\begin{aligned}
\widehat{\theta}_0 &= \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}} \\
\widehat{\theta}_1 &= \left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}\right) - \left(\frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}}\right) \\
\widehat{\gamma}_2 &= \frac{\sum_{i:\, G_i=0} Y_{i2}}{n_{02}}, \quad \widehat{\gamma}_1 = \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}, \quad \widehat{\gamma}_0 = \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}} \\
\widehat{\beta} &= \left\{\left(\frac{\sum_{i:\, G_i=1} Y_{i2}}{n_{12}} - \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}}\right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i2}}{n_{02}} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}}\right)\right\} \\
&\qquad - \left\{\left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}}\right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}}\right)\right\},
\end{aligned}$$

which completes the proof. □

C.3.2 Panel Data

We clarify that the sequential DID estimator $\widehat{\tau}_{\texttt{s-DID}}$ (equation (10)) is equivalent to a coefficient in the two-way fixed effects regression estimator with transformed outcomes.

Result 7 (Nonparametric Equivalence of the Sequential DID and Two-way Fixed Effects Regression Estimator) *We focus on the two-way fixed effects regression estimator with transformed outcomes.*

$$(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=1}^{2} (\Delta Y_{it} - \alpha_i - \delta_t - \beta D_{it})^2,$$

where $\Delta Y_{it} = Y_{it} - Y_{i,t-1}$*. Then,* $\widehat{\tau}_{\texttt{s-DID}} = \widehat{\beta}$*.*

Proof. As in Result 2, we can focus on the demeaned form.

$$\widehat{\beta} = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=1}^{2} (\widetilde{\Delta Y}_{it} - \beta \widetilde{D}_{it})^2,$$

where $\widetilde{\Delta Y}_{it} = \Delta Y_{it} - \overline{\Delta Y}_i - \overline{\Delta Y}_t + \overline{\Delta Y}$, $\overline{\Delta Y}_i = \sum_{t=1}^{2} \Delta Y_{it}/2$, $\overline{\Delta Y}_t = \sum_{i=1}^{n} \Delta Y_{it}/n$, and $\overline{\Delta Y} = \sum_{i=1}^{n} \sum_{t=1}^{2} \Delta Y_{it}/2n$. Similarly, $\widetilde{D}_{it} = D_{it} - \overline{D}_i - \overline{D}_t + \overline{D}$, $\overline{D}_i = \sum_{t=1}^{2} D_{it}/2$, $\overline{D}_t = \sum_{i=1}^{n} D_{it}/n$, and $\overline{D} = \sum_{i=1}^{n} \sum_{t=1}^{2} D_{it}/2n$. Using Result 2,

$$\begin{aligned}
\widehat{\beta} &= \frac{1}{n_1} \sum_{i=1}^{n} G_i(\Delta Y_{i2} - \Delta Y_{i1}) - \frac{1}{n_0} \sum_{i=1}^{n} (1 - G_i)(\Delta Y_{i2} - \Delta Y_{i1}) \\
&= \left\{\frac{1}{n_1} \sum_{i=1}^{n} G_i(Y_{i2} - Y_{i1}) - \frac{1}{n_0} \sum_{i=1}^{n} (1 - G_i)(Y_{i2} - Y_{i1})\right\} \\
&\qquad - \left\{\frac{1}{n_1} \sum_{i=1}^{n} G_i(Y_{i1} - Y_{i0}) - \frac{1}{n_0} \sum_{i=1}^{n} (1 - G_i)(Y_{i1} - Y_{i0})\right\} \\
&\equiv \widehat{\tau}_{\texttt{s-DID}},
\end{aligned}$$

which concludes the proof. □

Next, we clarify that the sequential DID estimator $\widehat{\tau}_{\texttt{s-DID}}$ (equation (10)) is also equivalent to a coefficient in the two-way fixed effects regression estimator with individual-specific time trends.

Result 8 (Nonparametric Equivalence of the Sequential DID and Two-way Fixed Effects Regression Estimator with Individual-Specific Time Trends) *We focus on the two-way fixed effects regression estimator with individual-specific time trends*

$$(\widehat{\alpha}, \widehat{\xi}, \widehat{\delta}, \widehat{\beta}) = \text{argmin} \sum_{i=1}^{n} \sum_{t=0}^{2} (Y_{it} - \alpha_i - (\xi_i \times t) - \delta_t - \beta D_{it})^2.$$

Then, $\widehat{\tau}_{\texttt{s-DID}} = \widehat{\beta}$.

Proof. By solving the least squares problem, we obtain that

$$\begin{aligned}
&\sum_{i:\, G_i=1} Y_{i2} = (\widehat{\beta} + \widehat{\gamma}_2) n_1 + \sum_{i:\, G_i=1} \widehat{\alpha}_i + 2 \sum_{i:\, G_i=1} \widehat{\xi}_i, && \sum_{i:\, G_i=0} Y_{i2} = \widehat{\gamma}_2 n_0 + \sum_{i:\, G_i=0} \widehat{\alpha}_i + 2 \sum_{i:\, G_i=0} \widehat{\xi}_i \\
&\sum_{i:\, G_i=1} Y_{i1} = \widehat{\gamma}_1 n_1 + \sum_{i:\, G_i=1} \widehat{\alpha}_i + \sum_{i:\, G_i=1} \widehat{\xi}_i, && \sum_{i:\, G_i=0} Y_{i1} = \widehat{\gamma}_1 n_0 + \sum_{i:\, G_i=0} \widehat{\alpha}_i + \sum_{i:\, G_i=0} \widehat{\xi}_i \\
&\sum_{i:\, G_i=1} Y_{i0} = \widehat{\gamma}_0 n_1 + \sum_{i:\, G_i=1} \widehat{\alpha}_i, && \sum_{i:\, G_i=0} Y_{i0} = \widehat{\gamma}_0 n_0 + \sum_{i:\, G_i=0} \widehat{\alpha}_i.
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
\widehat{\beta} = &\left\{ \left(\frac{\sum_{i:\, G_i=1} Y_{i2}}{n_1} - \frac{\sum_{i:\, G_i=1} Y_{i1}}{n_1} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i2}}{n_0} - \frac{\sum_{i:\, G_i=0} Y_{i1}}{n_0} \right) \right\} \\
&- \left\{ \left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_1} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_1} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i1}}{n_0} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_0} \right) \right\},
\end{aligned}$$

which completes the proof. □

C.3.3 Alternative Interpretation of Parallel Trends-in-Trends Assumption

We emphasize an alternative way to interpret the parallel trends-in-trends assumption. Unlike the parallel trends assumption that assumes the time-invariant unmeasured confounding, the parallel trends-in-trends assumption can account for *linear time-varying* unmeasured confounding — unobserved confounding increases or decreases over time but with some constant rate. For example, researchers might be worried that some treated communes have higher motivation for reforms, which is not measured, and the infrastructure qualities differ between treated and control communes due to this unobserved motivation. The parallel trends assumption means that the difference in the infrastructure qualities due to this unobserved confounder does not grow or decline over time. In contrast, the parallel trends-in-trends assumption accommodates a simple yet important case in which the unobserved difference in the infrastructure qualities does grow or decline with some fixed rate, which analysts do not need to specify. This interpretation comes from the following equivalent representation of the parallel trends-in-trends assumption.

$$\underbrace{\{\mathbb{E}[Y_{i2}(0) \mid G_i = 1] - \mathbb{E}[Y_{i2}(0) \mid G_i = 0]\}}_{\text{Bias at } t=2} - \underbrace{\{\mathbb{E}[Y_{i1}(0) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 0]\}}_{\text{Bias at } t=1}$$
$$= \underbrace{\{\mathbb{E}[Y_{i1}(0) \mid G_i = 1] - \mathbb{E}[Y_{i1}(0) \mid G_i = 0]\}}_{\text{Bias at } t=1} - \underbrace{\{\mathbb{E}[Y_{i0}(0) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 0]\}}_{\text{Bias at } t=0}. \qquad \text{(A.4)}$$

The difference between the mean potential outcome $Y_{it}(0)$ for the treated and control group at time t, $\mathbb{E}[Y_{it}(0) \mid G_i = 1] - \mathbb{E}[Y_{it}(0) \mid G_i = 0]$, is often called *bias* (or selection bias) in the literature (e.g., Heckman et al., 1998; Cunningham, 2021). Equation (A.4) shows that the parallel trends-in-trends assumption allows for a linear change in bias over time, whereas the bias is assumed to be constant over time in the extended parallel trends assumption. This representation is useful when we generalize our results to K pre-treatment periods where $K > 2$. Importantly, equation (11) and equation (A.4) are equivalent, and therefore, researchers can choose whichever interpretation easy for them to evaluate in each application.

C.4 Connection to the Leads Test

Here we formally prove the connection between the test of pre-treatment periods discussed in Section 2.2 and the well known leads test (Angrist and Pischke, 2008). The leads test includes $D_{i,t+1}$ into a linear regression and check whether a coefficient of $D_{i,t+1}$ is zero.

C.4.1 Repeated Cross-Sectional Data

In the repeated cross-sectional data setting, the leads test considers the following linear regression.

$$(\widehat{\theta}, \widehat{\gamma}, \widehat{\beta}, \widehat{\zeta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{1} O_{it} \left(Y_{it} - \theta G_i - \gamma_t - \beta D_{it} - \zeta D_{i,t+1}\right)^2.$$

Then, because $D_{it} = 0$ for all units in $t = \{0, 1\}$, this least squares problem is the same as

$$(\widehat{\theta}, \widehat{\gamma}, \widehat{\zeta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{1} O_{it} \left(Y_{it} - \theta G_i - \gamma_t - \zeta D_{i,t+1}\right)^2.$$

Finally, using Result 1, we have

$$\widehat{\zeta} = \left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_{11}} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_{10}}\right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i1}}{n_{01}} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_{00}}\right),$$

which is the standard DID estimator to the pre-treatment periods $t = 0, 1$. □

C.4.2 Panel Data

In the panel data setting, the leads test considers the following two-way fixed effects regression.

$$(\widehat{\alpha}, \widehat{\delta}, \widehat{\beta}, \widehat{\zeta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{1} (Y_{it} - \alpha_i - \delta_t - \beta D_{it} - \zeta D_{i,t+1})^2.$$

Again, this least squares problem is the same as

$$(\widehat{\alpha}, \widehat{\delta}, \widehat{\zeta}) = \operatorname{argmin} \sum_{i=1}^{n} \sum_{t=0}^{1} (Y_{it} - \alpha_i - \delta_t - \zeta D_{i,t+1})^2.$$

Then, using Result 2, we have

$$\widehat{\zeta} = \left(\frac{\sum_{i:\, G_i=1} Y_{i1}}{n_1} - \frac{\sum_{i:\, G_i=1} Y_{i0}}{n_1} \right) - \left(\frac{\sum_{i:\, G_i=0} Y_{i1}}{n_0} - \frac{\sum_{i:\, G_i=0} Y_{i0}}{n_0} \right),$$

which is the standard DID estimator to the pre-treatment periods $t = 0, 1$. □

D Details of Double DID Estimator

D.1 Properties of Double DID Estimator

Here, we prove several important properties of the double DID estimator based on the GMM theory (Hansen, 1982).

Theorem 1 *When the extended parallel trends assumption (Assumption 2) holds, the double DID estimator with the optimal weight matrix (equation* (13) *in the main paper) is consistent, and its asymptotic variance is smaller than or equal to that of the standard, extended, and sequential DID estimators, i.e.,* $\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}) \leq min(\mathrm{Var}(\widehat{\tau}_{\texttt{DID}}), \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}), \mathrm{Var}(\widehat{\tau}_{\texttt{e-DID}}))$.

Proof.

Suppose we define a moment function $m_i(\tau)$ as

$$m_i(\tau) = \begin{pmatrix} \tau - \widehat{\tau}_{\texttt{DID}}(i) \\ \tau - \widehat{\tau}_{\texttt{s-DID}}(i) \end{pmatrix}$$

where

$$\begin{aligned}
\widehat{\tau}_{\texttt{DID}}(i) &= \left(\frac{n}{n_{12}} G_i Y_{i2} - \frac{n}{n_{11}} G_i Y_{i1}\right) - \left(\frac{n}{n_{02}}(1-G_i)Y_{i2} - \frac{n}{n_{01}}(1-G_i)Y_{i1}\right) \\
\widehat{\tau}_{\texttt{s-DID}}(i) &= \left\{\left(\frac{n}{n_{12}} G_i Y_{i2} - \frac{n}{n_{11}} G_i Y_{i1}\right) - \left(\frac{n}{n_{02}}(1-G_i)Y_{i2} - \frac{n}{n_{01}}(1-G_i)Y_{i1}\right)\right\} \\
&\quad - \left\{\left(\frac{n}{n_{11}} G_i Y_{i1} - \frac{n}{n_{10}} G_i Y_{i0}\right) - \left(\frac{n}{n_{01}}(1-G_i)Y_{i1} - \frac{n}{n_{00}}(1-G_i)Y_{i0}\right)\right\}
\end{aligned}$$

for the repeated cross-sectional setting. They can be similarly defined in the panel data setting. Then, we can write the double DID estimator as the GMM estimator:

$$\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}) = \underset{\tau}{\mathrm{argmin}} \left(\frac{1}{n}\sum_{i=1}^{n} m_i(\tau)\right)^{\top} \mathbf{W} \left(\frac{1}{n}\sum_{i=1}^{n} m_i(\tau)\right) \tag{A.2}$$

where we index the double DID estimator by $\mathbf{W}$, which is a weight matrix of dimension 2×2.

In general, the variance of the GMM estimator is given by

$$\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W})) = (M^{\top}\mathbf{W}M)^{-1} M^{\top}\mathbf{W}\Omega\mathbf{W}^{\top} M (M^{\top}\mathbf{W}M)^{-1}.$$

where $M = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{\frac{\partial}{\partial \tau} m_i(\tau)\right\}$, and

$$\Omega = \begin{pmatrix} \mathrm{Var}(\widehat{\tau}_{\texttt{DID}}) & \mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) \\ \mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) & \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}) \end{pmatrix}.$$

Hansen (1982) showed in general that $\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}))$ is minimized when $\mathbf{W}$ is set to Ω^{-1}. We define this optimal weight as $\mathbf{W}^*$

$$\mathbf{W}^* = \Omega^{-1} = \begin{pmatrix} \mathrm{Var}(\widehat{\tau}_{\texttt{DID}}) & \mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) \\ \mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) & \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}) \end{pmatrix}^{-1}.$$

In general, the asymptotic variance of this optimal GMM estimator is given by

$$\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*)) = (M^\top \mathbf{W}^* M)^{-1}.$$

Because $M = \mathbf{1}$, the asymptotic variance of $\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*))$ can be explicitly written as

$$\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*)) = (\mathbf{1}^\top \mathbf{W}^* \mathbf{1})^{-1} = \frac{\mathrm{Var}(\widehat{\tau}_{\texttt{DID}}) \cdot \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}) - \mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})^2}{\mathrm{Var}(\widehat{\tau}_{\texttt{DID}}) + \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}) - 2\mathrm{Cov}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})}.$$

Finally, the standard, sequential, and extended DID estimators are all special cases of the double DID with a specific choice of the weight matrix as described in Table 1 of the main paper. Because for any $\mathbf{W}$, $\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*)) \leq \mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}))$, it implies that

$$\mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*)) \leq \min(\mathrm{Var}(\widehat{\tau}_{\texttt{DID}}), \mathrm{Var}(\widehat{\tau}_{\texttt{s-DID}}), \mathrm{Var}(\widehat{\tau}_{\texttt{e-DID}})).$$

Now, we can show the consistency of the estimator and its variance estimator. The optimal weight matrix $\mathbf{W}^*$ can be estimated by its sample analog:

$$\widehat{\mathbf{W}} = \begin{pmatrix} \widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{DID}}) & \widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) \\ \widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) & \widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{s-DID}}) \end{pmatrix}^{-1}.$$

which is a consistent estimator of $\mathbf{W}^*$ under the standard regularity conditions. Therefore, by solving the GMM optimization problem (equation (A.2)), we can explicitly write the double DID as

$$\widehat{\tau}_{\texttt{d-DID}}(\widehat{\mathbf{W}}) = \widehat{w}_1 \widehat{\tau}_{\texttt{DID}} + \widehat{w}_2 \widehat{\tau}_{\texttt{s-DID}}$$

where $\widehat{w}_1 + \widehat{w}_2 = 1$, and

$$\widehat{w}_1 = \frac{\widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{s-DID}}) - \widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})}{\widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{DID}}) + \widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{s-DID}}) - 2\widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})},$$
$$\widehat{w}_2 = \frac{\widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{DID}}) - \widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})}{\widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{DID}}) + \widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{s-DID}}) - 2\widehat{\mathrm{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}})}.$$

Under the extended parallel trends assumption (Assumption 2), both the standard DID and the sequential DID estimator are consistent to the ATT. Therefore, by the continuous mapping theorem and law of large numbers, we have

$$\widehat{\tau}_{\texttt{d-DID}}(\widehat{\mathbf{W}}) \xrightarrow{p} \tau$$

and

$$\widehat{\mathrm{Var}}(\widehat{\tau}_{\texttt{d-DID}}(\widehat{\mathbf{W}})) \xrightarrow{p} \mathrm{Var}(\widehat{\tau}_{\texttt{d-DID}}(\mathbf{W}^*)),$$

which completes the proof. □

D.2 Standard Error Estimation

As described in Section 3.1.2, we use the block bootstrap.

1. Estimate $\{\widehat{\tau}_{\texttt{DID}}^{(b)}, \widehat{\tau}_{\texttt{s-DID}}^{(b)}\}_{b=1}^{B}$ where B indicates the total number of bootstrap iterations. We recommend the block-bootstrap where the block is taken at the level of treatment assignment.

2. Estimate the optimal weight matrix via computing the variance-covariance matrix:

$$\widehat{\text{Var}}(\widehat{\tau}_{\texttt{DID}}) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\tau}_{\texttt{DID}}^{(b)} - \overline{\widehat{\tau}}_{\texttt{DID}})^2$$

$$\widehat{\text{Var}}(\widehat{\tau}_{\texttt{s-DID}}) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\tau}_{\texttt{s-DID}}^{(b)} - \overline{\widehat{\tau}}_{\texttt{s-DID}})^2$$

$$\widehat{\text{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) = \frac{1}{B}\sum_{b=1}^{B}(\widehat{\tau}_{\texttt{DID}}^{(b)} - \overline{\widehat{\tau}}_{\texttt{DID}})(\widehat{\tau}_{\texttt{s-DID}}^{(b)} - \overline{\widehat{\tau}}_{\texttt{s-DID}})$$

   where $\overline{\widehat{\tau}}_{\texttt{DID}} = \sum_{b=1}^{B}\widehat{\tau}_{\texttt{DID}}^{(b)}/B$, and $\overline{\widehat{\tau}}_{\texttt{s-DID}} = \sum_{b=1}^{B}\widehat{\tau}_{\texttt{s-DID}}^{(b)}/B$ are empirical average of two estimators. Finally, we obtain the estimate of the weight matrix by inverting the variance-covariance matrix (equation (13) in the main text),

$$\widehat{\mathbf{W}} = \begin{pmatrix} \widehat{\text{Var}}(\widehat{\tau}_{\texttt{DID}}) & \widehat{\text{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) \\ \widehat{\text{Cov}}(\widehat{\tau}_{\texttt{DID}}, \widehat{\tau}_{\texttt{s-DID}}) & \widehat{\text{Var}}(\widehat{\tau}_{\texttt{s-DID}}) \end{pmatrix}^{-1}$$

3. The double DID estimator is given by equation (14) in the main paper.

4. The variance of double DID estimator is then obtained via the standard efficient GMM variance formula

$$\widehat{\text{Var}}(\widehat{\tau}_{\texttt{d-DID}}) = (\mathbf{1}^{\top}\widehat{\mathbf{W}}\mathbf{1})^{-1}.$$

E Extensions of Double DID

E.1 Double DID Regression

Like other DID estimators, the double DID estimator has a nice connection to a widely-used regression approach. Using this double DID regression, researchers can include other pre-treatment covariates $\mathbf{X}_{it}$ to make the DID design more robust and efficient. We provide technical details in Appendix

To introduce the regression-based double DID estimator, we begin with the basic DID. As discussed in Appendix C.1, the basic DID estimator is equivalent to a coefficient in the linear regression of equation (A.2). Inspired by this connection, researchers often adjust for additional pre-treatment covariates as:

$$Y_{it} \sim \alpha + \theta G_i + \gamma I_t + \beta(G_i \times I_t) + \mathbf{X}_{it}^\top \boldsymbol{\rho}, \tag{A.3}$$

where we adjust for the additional pre-treatment covariates $\mathbf{X}_{it}$. A coefficient of the interaction term $\widehat{\beta}$ is a consistent estimator for the ATT when the conditional parallel trends assumption holds and the parametric model is correctly specified. Here, we make the parallel trends assumption *conditional* on covariates $\mathbf{X}_{it}$. The idea is that even when the parallel trends assumption might not hold without controlling for any covariates, trends of the two groups might be parallel conditionally after adjusting for observed covariates. For example, the conditional parallel trends assumption means that treatment and control groups have the same trends of the infrastructure quality after controlling for population and GDP per capita.

The sequential DID estimator is extended similarly. Based on the connection to the linear regression of equation (A.3), we can adjust for additional pre-treatment covariates as:

$$\Delta Y_{it} \sim \alpha_s + \theta_s G_i + \gamma_s I_t + \beta_s(G_i \times I_t) + \mathbf{X}_{it}^\top \boldsymbol{\rho}_s, \tag{A.4}$$

where $\Delta Y_{it} = Y_{it} - (\sum_{i:\, G_i=1} Y_{i,t-1})/n_{1,t-1}$ if $G_i = 1$ and $\Delta Y_{it} = Y_{it} - (\sum_{i:\, G_i=0} Y_{i,t-1})/n_{0,t-1}$ if $G_i = 0$. The estimated coefficient $\widehat{\beta_s}$ is consistent for the ATT under the *conditional* parallel trends-in-trends assumption and the conventional assumption of correct specification.

The double DID regression combines the two regression estimators via the GMM:

$$\widehat{\beta}_{\texttt{d-DID}} = \underset{\beta_d}{\text{argmin}} \begin{pmatrix} \beta_d - \widehat{\beta} \\ \beta_d - \widehat{\beta}_s \end{pmatrix}^\top \mathbf{W} \begin{pmatrix} \beta_d - \widehat{\beta} \\ \beta_d - \widehat{\beta}_s \end{pmatrix} \tag{A.5}$$

where $\mathbf{W}$ is a weighting matrix of dimension 2×2.

Thus, as the double DID estimator with no covariates, the double DID regression has two steps. The first step is to assess the underlying assumptions. Here, instead of using the standard DID estimator, we use the standard DID regression on pre-treatment periods to assess the conditional extended parallel trends assumption. The second step is to estimate the ATT, while adjusting for pre-treatment covariates. Instead of using the double DID estimator without covariates, we implement the regression-based double DID estimator (equation (A.5)).

E.2 Generalized K-DID

In this section, we propose the generalized K-DID, which extends the double DID in Section 3 to arbitrary number of *pre-* and *post*-treatment periods in the basic DID setting. We consider the staggered adoption design in Section 4.

E.2.1 The Setup and Causal Quantities of Interest

We first extend the setup to account for arbitrary number of pre- and post-treatment periods. Suppose we observe outcome Y_{it} for $i \in \{1, \ldots, n\}$ and $t \in \{0, 1, \ldots, T\}$. We define the binary treatment variable to be $D_{it} \in \{0, 1\}$. The treatment is assigned right before time period T^*, and thus, time periods $t \in \{T^*, \ldots, T\}$ are the post-treatment periods and time periods $t \in \{0, \ldots, T^* - 1\}$ are the pre-treatment periods. As in Section 2, we denote the treatment group as $G_i = 1$ and $G_i = 0$ otherwise. Note that $D_{it} = 0$ for $t \in \{1, \ldots, T^*\}$ for all units.

We are interested in the causal effect at post-treatment time $T^* + s$ where $s \geq 0$. When $s = 0$, this corresponds to the contemporaneous treatment effect. By specifying different values of $s > 0$, researchers can study a variety of long-term causal effects of the treatment. Formally, our quantity of interest is the average treatment effect on the treated (ATT) at post-treatment time $T^* + s$.

$$\tau(s) \equiv \mathbb{E}[Y_{i,T^*+s}(1) - Y_{i,T^*+s}(0) \mid G_i = 1].$$

For example, when $s = 3$, this could mean the causal effect of the policy after three years from its initial introduction. This definition is a generalization of the standard ATT: when $s = 0$, this quantity is equal to the ATT defined in equation (1).

E.2.2 Generalize Parallel Trends Assumptions

What assumptions do we need to identify the ATT at post-treatment time $T^* + s$? Here, we provide a generalization of the parallel trends assumption, which incorporates both the standard parallel trends assumption and the parallel trends-in-trends assumption.

Assumption A1 (k-th Order Parallel Trends) *For some integer k such that $1 \leq k \leq T^*$,*

$$\Delta_s^k \left(\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = 1]\right) = \Delta_s^k \left(\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = 0]\right),$$

where Δ_s^k is the k-th order difference operator defined recursively as follows. For $g \in \{0, 1\}$,

$$\Delta_s^1 \left(\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = g]\right) \equiv \mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = g] - \mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g],$$

when $k = 1$ and, in general,

$$\begin{aligned}
&\Delta_s^k \left(\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = g]\right) \\
\equiv\ &\Delta_s^{k-1} \left(\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = g]\right) - M_s^k \Delta^{k-1} \left(\mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g]\right), \\
=\ &\mathbb{E}[Y_{i,T^*+s}(0) \mid G_i = g] - \mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g] - \sum_{j=1}^{k-1} M_s^{j+1} \Delta^j \left(\mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g]\right),
\end{aligned}$$

where $M_s^\ell = \prod_{j=1}^{\ell-1}(s+j)/\prod_{j=1}^{\ell-1} j$ for $\ell \geq 2$. $\Delta^k\left(\mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g]\right)$ is also recursively defined as $\Delta^k\left(\mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g]\right) \equiv \Delta^{k-1}\left(\mathbb{E}[Y_{i,T^*-1}(0) \mid G_i = g]\right) - \Delta^{k-1}\left(\mathbb{E}[Y_{i,T^*-2}(0) \mid G_i = g]\right)$, and $\Delta^1\left(\mathbb{E}[Y_{i,T^*-m}(0) \mid G_i = g]\right) = \mathbb{E}[Y_{i,T^*-m}(0) \mid G_i = g] - \mathbb{E}[Y_{i,T^*-m-1}(0) \mid G_i = g]$ for $m = \{1, 2\}$. The standard parallel trends assumption and the parallel-trends-in-trends assumption are both special cases of this assumption. The k-th order parallel trends assumption reduces to the standard parallel trends assumption (Assumption 1) when $s = 1$ and $k = 1$, and to the parallel-trends-in-trends assumption (Assumption 3) when $s = 1$ and $k = 2$.

To further clarify the meaning of Assumption A1, we can consider a simpler but stronger condition. In particular, the k-th order parallel trends assumption (Assumption A1) is implied by the following p-th degree polynomial model of confounding.

$$\mathbb{E}[Y_{it}(0) \mid G_i = 1] - \mathbb{E}[Y_{it}(0) \mid G_i = 0] = \alpha + \sum_{p=1}^{k-1} \Gamma_p t^p,$$

with unknown parameters α and $\mathbf{\Gamma}$. Here, the left hand side of the equality captures the difference between the two groups (treatment and control) in terms of the mean of potential outcomes under the control condition. This representation shows that the standard parallel trends assumption (Assumption 1) is implied by the time-invariant confounding; the parallel trends-in-trends assumption (Assumption 3) is implied by the linear time-varying confounding; and in general, the k-th order parallel trends assumption is implied by the k-th order polynomial confounding.

E.2.3 Estimate ATT with Multiple Pre- and Post-Treatment Periods

We consider the identification and estimation of the ATT at post-treatment time $T^* + s$. Under the k-th order parallel trends assumption (Assumption A1), the ATT is identified as follows.

$$\tau(s) = \Delta_s^k\left(\mathbb{E}[Y_{i,T^*+s} \mid G_i = 1]\right) - \Delta_s^k\left(\mathbb{E}[Y_{i,T^*+s} \mid G_i = 0]\right).$$

Because each conditional expectation can be consistently estimated via its sample analogue,

$$\widehat{\tau}_k(s) = \Delta_s^k\left(\frac{\sum_{i:\, G_i=1} Y_{i,T^*+s}}{n_{1,T^*+s}}\right) - \Delta_s^k\left(\frac{\sum_{i:\, G_i=0} Y_{i,T^*+s}}{n_{0,T^*+s}}\right)$$

is a consistent estimator for the ATT at time $T^* + s$ under the k-th order parallel trends assumption. When $s = 0$ and $k = 1$, this estimator corresponds to the standard DID estimator (equation (3)). When $s = 0$ and $k = 2$, this is equal to the sequential DID estimator (equation (10)). While existing approaches (e.g., Angrist and Pischke, 2008; Mora and Reggio, 2012; Lee, 2016; Mora and Reggio, 2019) consider each estimator separately, we propose combining multiple DID estimators within the GMM framework.

In general, the generalized double DID combines K moment conditions where K is the number of pre-treatment periods researchers use. When there are more than two pre-treatment periods, we can naturally combine more than two DID estimators, improving upon the double DID in Section 3. Formally, the generalized double DID is defined as,

$$\widehat{\tau}(s) = \underset{\tau}{\operatorname{argmin}}\, \boldsymbol{g}(\tau)^\top \widehat{\mathbf{W}} \boldsymbol{g}(\tau)$$

where $\boldsymbol{g}(\tau) = (\tau - \widehat{\tau}_1(s), \ldots, \tau - \widehat{\tau}_K(s))^\top$. Based on the theory of the efficient GMM (Hansen, 1982), the optimal weight matrix is $\widehat{\mathbf{W}} = \text{Var}(\widehat{\tau}_{(1:K)}(s))^{-1}$ where $\text{Var}(\cdot)$ is the variance-covariance matrix and $\widehat{\tau}_{(1:K)}(s) = (\widehat{\tau}_1(s), \ldots, \widehat{\tau}_K(s))^\top$. When $T^* = 2$, this converges to the standard DID estimator (equation (3)). When $T^* = 3$, this corresponds to the basic form of the double DID estimator (equation (12)). Within the GMM framework, we can select moment conditions using the J-statistics (Hansen, 1982). We can similarly generalize the double DID regression.

To assess the extended parallel trends assumption, we can apply the generalized double DID to pre-treatment periods $t \in \{1, \ldots, T^* - 1\}$ as if the last pre-treatment period $T^* - 1$ is the target time period. Moments are $\boldsymbol{g}(\tau) = (\tau - \widehat{\tau}_1(0), \ldots, \tau - \widehat{\tau}_K(0))^\top$ where $\widehat{\tau}_k(0) = \Delta_s^k \left(\frac{\sum_{i:\, G_i=1} Y_{i,T^*-1}}{n_{1,T^*-1}} \right) - \Delta_s^k \left(\frac{\sum_{i:\, G_i=0} Y_{i,T^*-1}}{n_{0,T^*-1}} \right)$. Similarly, to assess the extended parallel trends-in-trends assumption, we can apply the generalized double DID to pre-treatment periods with moments $\boldsymbol{g}(\tau) = (\tau - \widehat{\tau}_2(0), \ldots, \tau - \widehat{\tau}_K(0))^\top$.

E.3 Generalized K-DID for Staggered Adoption Design

Combining the setup introduced in Section E.2.1 and the one in Section 4.1, we propose the generalized K-DID for the SA design, which allows researchers to estimate long-term causal effects in the SA design. We focus on the SA-ATT at post-treatment time $t + s$ where t is the timing of the treatment assignment and $s \geq 0$ represents how far in the future we want estimate the ATT for. We first redefine the group indicator G to estimate the long-term SA-ATT at post-treatment time $t + s$. In particular, we define

$$G_{its} = \begin{cases} 1 & \text{if } A_i = t \\ 0 & \text{if } A_i > t + s \\ -1 & \text{otherwise} \end{cases}$$

where $G_{its} = 1$ represents units who receive the treatment at time t, and $G_{its} = 0$ indicates units who do not receive the treatment by time $t + s$. $G_{its} = -1$ includes other units who receive the treatment before time t or receive the treatment between $t+1$ and $t+s$. When $s = 0$, this definition corresponds to the group indicator in equation (15).

Formally, our first quantity of interest is the *staggered-adoption average treatment effect on the treated* (SA-ATT) at post-treatment time $t + s$.

$$\tau^{\mathsf{SA}}(s, t) \equiv \mathbb{E}[Y_{i,t+s}(1) - Y_{i,t+s}(0) \mid G_{its} = 1].$$

By averaging over time, we can also define the *time-average staggered-adoption average treatment effect on the treated* (time-average SA-ATT) at s periods after treatment onset.

$$\overline{\tau}^{\mathsf{SA}}(s) \equiv \sum_{t \in \mathcal{T}} \pi_t \tau^{\mathsf{SA}}(s, t),$$

where $\mathcal{T}$ represents a set of the time periods for which researchers want to estimate the ATT. The SA-ATT in period t, $\tau^{\mathsf{SA}}(t)$, is weighted by the proportion of units who receive the treatment at time t: $\pi_t = \sum_{i=1}^n \mathbf{1}\{A_i = t\} / \sum_{i=1}^n \mathbf{1}\{A_i \in \mathcal{T}\}$.

Here, we provide a generalization of the parallel trends assumption, which incorporates both the standard parallel trends assumption and the parallel trends-in-trends assumption.

Assumption A2 (k-th Order Parallel Trends for Staggered Adoption Design) *For some integer k such that $1 \leq k \leq T$, and for $k \leq t \leq T-s$,*

$$\Delta_s^k \left(\mathbb{E}[Y_{i,t+s}(0) \mid G_{its}=1]\right) = \Delta_s^k \left(\mathbb{E}[Y_{i,t+s}(0) \mid G_{its}=0]\right),$$

where Δ_s^k is the k-th order difference operator defined in Assumption A1.

Under Assumption A2, the SA-ATT at post-treatment time $t+s$ is identified as follows.

$$\tau^{\mathsf{SA}}(s,t) = \Delta_s^k \left(\mathbb{E}[Y_{i,t+s} \mid G_{its}=1]\right) - \Delta_s^k \left(\mathbb{E}[Y_{i,t+s} \mid G_{its}=0]\right).$$

Since conditional expectations can be consistently estimated via the sample analogue,

$$\widehat{\tau}_k^{\mathsf{SA}}(s,t) = \Delta_s^k \left(\frac{\sum_{i:\, G_{its}=1} Y_{i,t+s}}{n_{1,t+s}}\right) - \Delta_s^k \left(\frac{\sum_{i:\, G_{its}=0} Y_{i,t+s}}{n_{0,t+s}}\right)$$

is a consistent estimator for the SA-ATT at post-treatment time $t+s$ under Assumption A2.

In general, we combine K DID estimators to obtain the generalized K-DID for the SA-ATT at post-treatment time $t+s$ as follows.

$$\widehat{\tau}^{\mathsf{SA}}(s,t) = \operatorname*{argmin}_{\tau^{\mathsf{SA}}} \boldsymbol{g}(\tau^{\mathsf{SA}})^\top \widehat{\mathbf{W}} \boldsymbol{g}(\tau^{\mathsf{SA}})$$

where $\boldsymbol{g}(\tau^{\mathsf{SA}}) = (\tau^{\mathsf{SA}} - \widehat{\tau}_1^{\mathsf{SA}}(s), \dots, \tau^{\mathsf{SA}} - \widehat{\tau}_K^{\mathsf{SA}}(s))^\top$. The optimal weight matrix is $\widehat{\mathbf{W}} = \mathrm{Var}(\widehat{\tau}_{(1:K)}^{\mathsf{SA}}(s))^{-1}$ where $\widehat{\tau}_{(1:K)}^{\mathsf{SA}}(s) = (\widehat{\tau}_1^{\mathsf{SA}}(s), \dots, \widehat{\tau}_K^{\mathsf{SA}}(s))^\top$.

To estimate the time-average SA-ATT, we first define the time-average k-th order time-average DID estimator as,

$$\widehat{\overline{\tau}}_k^{\mathsf{SA}}(s) = \sum_{t \in \mathcal{T}} \pi_t \widehat{\tau}_k^{\mathsf{SA}}(s,t).$$

Finally, the generalized K-DID combines K moment conditions as follows.

$$\widehat{\overline{\tau}}^{\mathsf{SA}}(s) = \operatorname*{argmin}_{\overline{\tau}^{\mathsf{SA}}} \boldsymbol{g}(\overline{\tau}^{\mathsf{SA}})^\top \widehat{\overline{\mathbf{W}}} \boldsymbol{g}(\overline{\tau}^{\mathsf{SA}})$$

where $\boldsymbol{g}(\overline{\tau}^{\mathsf{SA}}) = (\overline{\tau}^{\mathsf{SA}} - \widehat{\overline{\tau}}_1^{\mathsf{SA}}(s), \dots, \overline{\tau}^{\mathsf{SA}} - \widehat{\overline{\tau}}_K^{\mathsf{SA}}(s))^\top$. The optimal weight matrix is $\widehat{\overline{\mathbf{W}}} = \mathrm{Var}(\widehat{\overline{\tau}}_{(1:K)}^{\mathsf{SA}}(s))^{-1}$ where $\widehat{\overline{\tau}}_{(1:K)}^{\mathsf{SA}}(s) = (\widehat{\overline{\tau}}_1^{\mathsf{SA}}(s), \dots, \widehat{\overline{\tau}}_K^{\mathsf{SA}}(s))^\top$.

E.4 Double DID Regression for Staggered Adoption Design

We now extend the double DID regression to the SA design setting. We first extend the standard DID regression (Appendix E.1) to the SA design. In particular, to estimate the SA-ATT at time t, we can fit the following regression for units who are not yet treated at time $t-1$, that is, $\{i : G_{it} \geq 0\}$.

$$Y_{iv} \sim \alpha + \theta G_{it} + \gamma I_v + \beta^{\mathsf{SA}}(t)(G_{it} \times I_v) + \mathbf{X}_{iv}^\top \boldsymbol{\rho},$$

where $v \in \{t-1, t\}$ and the time indicator I_v (equal to 1 if $v = t$ and 0 if $v = t-1$). Note that G_{it} defines the treatment and control group at time t, and thus, it does not depend on time index v. The estimated coefficient $\widehat{\beta}^{\mathsf{SA}}(t)$ is consistent for the SA-ATT under the conditional parallel trends assumption.

Similarly, we can extend the sequential DID regression to the SA design. Using the connection to the linear regression of equation (A.3), we can adjust for additional pre-treatment covariates as:

$$\Delta Y_{iv} \sim \alpha_s + \theta_s G_{it} + \gamma_s I_v + \beta_s^{\mathsf{SA}}(t)(G_{it} \times I_v) + \mathbf{X}_{iv}^\top \boldsymbol{\rho}_s,$$

where $v \in \{t-1, t\}$ and $\Delta Y_{iv} = Y_{iv} - (\sum_{i:\, G_{it}=1} Y_{i,v-1})/n_{1,v-1}$ if $G_{it} = 1$ and $\Delta Y_{iv} = Y_{iv} - (\sum_{i:\, G_{it}=0} Y_{i,v-1})/n_{0,v-1}$ if $G_{it} = 0$. The estimated coefficient $\widehat{\beta}_s^{\mathsf{SA}}(t)$ is consistent for the SA-ATT under the conditional parallel trends-in-trends assumption.

Therefore, the double DID regression for the SA design combines the two regression estimators via the GMM:

$$\widehat{\beta}^{\mathsf{SA}}_{\texttt{d-DID}}(t) = \underset{\beta_d^{\mathsf{SA}}(t)}{\operatorname{argmin}} \begin{pmatrix} \beta_d^{\mathsf{SA}}(t) - \widehat{\beta}^{\mathsf{SA}}(t) \\ \beta_d^{\mathsf{SA}}(t) - \widehat{\beta}_s^{\mathsf{SA}}(t) \end{pmatrix}^\top \mathbf{W}(t) \begin{pmatrix} \beta_d^{\mathsf{SA}}(t) - \widehat{\beta}^{\mathsf{SA}}(t) \\ \beta_d^{\mathsf{SA}}(t) - \widehat{\beta}_s^{\mathsf{SA}}(t). \end{pmatrix}$$

where the choice of the weight matrix follows the same two-step procedure as Section 4.2. We also provide further details in Appendix E.3. The optimal weight matrix $\mathbf{W}(t)$ is equal to $\mathrm{Var}(\widehat{\beta}^{\mathsf{SA}}_{(1:2)}(t))^{-1}$ where $\widehat{\beta}^{\mathsf{SA}}_{(1:2)}(t) = (\widehat{\beta}^{\mathsf{SA}}(t), \widehat{\beta}_s^{\mathsf{SA}}(t))^\top$.

To estimate the time-average SA-ATT, we extend the double DID regression as follows.

$$\widehat{\overline{\beta}}^{\mathsf{SA}}_{\texttt{d-DID}} = \underset{\overline{\beta}_d^{\mathsf{SA}}}{\operatorname{argmin}} \begin{pmatrix} \overline{\beta}_d^{\mathsf{SA}} - \widehat{\overline{\beta}}^{\mathsf{SA}} \\ \overline{\beta}_d^{\mathsf{SA}} - \widehat{\overline{\beta}}_s^{\mathsf{SA}} \end{pmatrix}^\top \overline{\mathbf{W}} \begin{pmatrix} \overline{\beta}_d^{\mathsf{SA}} - \widehat{\overline{\beta}}^{\mathsf{SA}} \\ \overline{\beta}_d^{\mathsf{SA}} - \widehat{\overline{\beta}}_s^{\mathsf{SA}} \end{pmatrix}$$

where

$$\widehat{\overline{\beta}}^{\mathsf{SA}} = \sum_{t \in \mathcal{T}} \pi_t \widehat{\beta}^{\mathsf{SA}}(t), \quad \text{and} \quad \widehat{\overline{\beta}}_s^{\mathsf{SA}} = \sum_{t \in \mathcal{T}} \pi_t \widehat{\beta}_s^{\mathsf{SA}}(t).$$

The optimal weight matrix $\overline{\mathbf{W}}$ is equal to $\mathrm{Var}(\widehat{\overline{\beta}}_{(1:2)})^{-1}$ where $\widehat{\overline{\beta}}_{(1:2)} = (\widehat{\overline{\beta}}^{\mathsf{SA}}, \widehat{\overline{\beta}}_s^{\mathsf{SA}})^\top$.

F Equivalence Approach

Here, we provide technical details on the equivalence approach we introduced in Section 3.1. In the standard hypothesis testing, researchers usually evaluate the two-sided null hypothesis $H_0 : \delta = 0$ where $\delta = \{\mathbb{E}[Y_{i1}(0) \mid G_i = 1] - \mathbb{E}[Y_{i0}(0) \mid G_i = 1]\} - \{\mathbb{E}[Y_{i1}(0) \mid G_i = 0] - \mathbb{E}[Y_{i0}(0) \mid G_i = 0]\}$ when we are conducting the pre-treament-trends test. However, this approach has a risk of conflating evidence for parallel trends and statistical inefficiency. For example, when sample size is small, even if pre-treatment trends of the treatment and control groups differ (i.e., the null hypothesis is false), a test of the difference might not be statistically significant due to large standard error. And, analysts might "pass" the pre-treatment-trends test by not finding enough evidence for the difference.

The equivalence approach can mitigate this concern by flipping the null hypothesis, so that the rejection of the null can be the evidence for parallel trends. In particular, we consider two one-sided tests:

$$H_0 : \theta \geq \gamma_U, \quad \text{or} \quad \theta \leq \gamma_L$$

where (γ_U, γ_L) is a user-specified equivalence range. By rejecting this null hypothesis, researchers can provide statistical evidence for the alternative hypothesis:

$$H_0 : \gamma_L < \theta < \gamma_U,$$

which means that θ (i.e., the difference in pre-treament-trends across treatment and control groups) are within an interval $[\gamma_L, \gamma_U]$.

One difficulty of the equivalence approach is that researchers have to choose this equivalence range (γ_U, γ_L), which might not be straightforward in practice. To overcome this challenge, we follow Hartman and Hidalgo (2018) to estimate the 95% equivalence confidence interval, which is the smallest equivalence range supported by the observed data. Suppose we obtain $[-c, c]$ as the symmetric 95% equivalence confidence interval where $c > 0$ is some positive constant. Then, this means that if researchers think the absolute value of θ smaller than c is substantively negligible, the 5% equivalence test would reject the null hypothesis and provide the evidence for the parallel pre-treatment-trends. In contrast, if researchers think the absolute value of θ being c is substantively too large as bias in practice, the 5% equivalence test would fail to reject the null hypothesis and cannot provide the evidence for the parallel pre-treatment-trends. In sum, by estimating the equivalence confidence interval, readers of the analysis can decide how much evidence for the parallel pre-treatment-trends exists in the observed data. Researchers can estimate the 95% equivalence confidence interval by the following general two steps. First, estimate 90% confidence interval, which we denote by $[b_L, b_U]$. Second, we can obtain the symmetric 95% equivalence confidence interval as $[-b, b]$ where we define $b = \max\{|b_L|, |b_U|\}$. See Wellek (2010); Hartman and Hidalgo (2018) for more details.

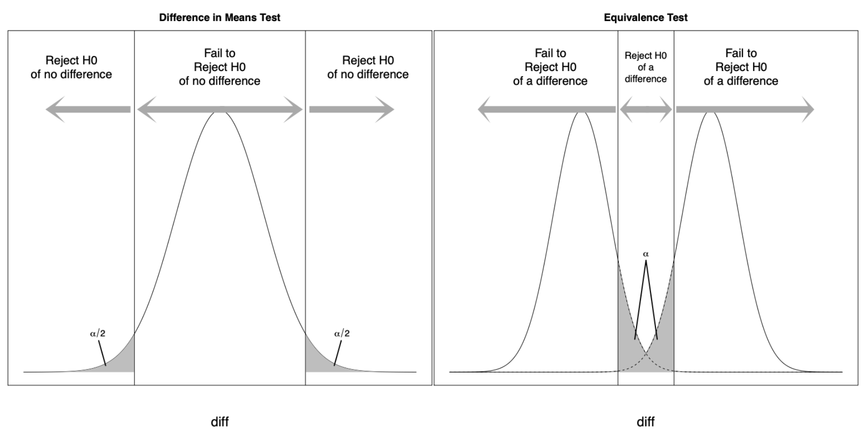

Figure A1: Figure 1 from Hartman and Hidalgo (2018) on the difference between the standard hypothesis testing and the equivalence testing.

G Simulation Study

We conduct a simulation study to compare the performance of the various DID estimators discussed in this paper. We demonstrate two key results. First, the double DID is unbiased under the extended parallel trends assumption or under the parallel trends-in-trends assumption. Second, the double DID has the smallest standard errors among unbiased DID estimators. In particular, standard errors of the double DID are smaller than those of the extended DID (i.e., the two-way fixed effects estimator) even under the extended parallel trends assumption.

We compare three DID estimators — the double DID, the extended DID, and the sequential DID — using two scenarios. In the first scenario, the extended parallel trends assumption (Assumption 2) holds where the difference between potential outcomes under control $\mathbb{E}[Y_{it}(0) \mid G_i = 1] - \mathbb{E}[Y_{it}(0) \mid G_i = 0]$ is constant over time. This corresponds to time-invariant unmeasured confounding, and we expect that all the DID estimators are unbiased in this scenario. The second scenario represents the parallel-trends-in-trends assumption (Assumption 3) where unmeasured confounding varies over time linearly. Here, we expect that the double DID and the sequential DID are unbiased, whereas the extended DID is biased.

For each of the two scenarios, we consider the balanced panel data with n units and five-time periods where treatments are assigned at the last time period. We vary the number of units (n) from 100 to 1000 and evaluate the quality of estimators by absolute bias and standard errors over 2000 Monte Carlo simulations. We describe the details of the simulation setup next.

G.1 Simulation Design

We consider the balanced panel data with $T = 5$ ($t = \{0, 1, 2, 3, 4\}$) where the last period ($t = 4$) is treated as the post-treatment period. We vary the number of units at each time period as $n \in \{100, 250, 500, 1000\}$. Thus, the total number of observations are $nT \in \{500, 1250, 2500, 5000\}$. We compare three estimators: the double DID, the extended DID, and the sequential DID.

Note that we consider four pre-treatment periods here, and thus the generalized double DID is not equal to the sequential DID even under the parallel trends-in-trends assumption because it combines two other moments and optimally weight them (see Appendix E.2). The equivalence between the sequential DID and the double DID holds only when there are two pre-treatment periods. We see below that the generalized double DID improves upon the sequential DID even under the parallel trends-in-trends assumption as they optimally weight observations from different time periods.

We study two scenarios: one under the extended parallel trends assumption (Assumption 2) and the other under the parallel-trends-in-trends assumption (Assumption 3). In the first scenario, the difference between potential outcomes under control $\mathbb{E}[Y_{it}(0) \mid G_i = 1] - \mathbb{E}[Y_{it}(0) \mid G_i = 0]$ is constant over time. In particular, we set

$$\mathbb{E}[Y_{it}(0) \mid G_i = g] = \alpha_t + 0.05 \times g \tag{A.6}$$

where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1, 2, 3, 4, 5)$. In the second scenario, we allow for linear time-varying confounding. In particular, we set

$$\mathbb{E}[Y_{it}(0) \mid G_i = g] = \alpha_t + 0.1 \times g \times (t+1) \tag{A.7}$$

where $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1, 2, 3, 4, 5)$.

Then, potential outcomes under control are drawn as follows. $Y_{it}(0) = \mathbb{E}[Y_{it}(0) \mid G_i] + \epsilon_{it}$ where ϵ_{it} follows the AR(1) process with autocorrelation parameter ρ. That is,

$$\begin{aligned}
\epsilon_{it} &= \rho\epsilon_{i,t-1} + \xi_{it}, \\
\epsilon_{i0} &= \mathcal{N}(0, 3/(1-\rho^2)), \\
\xi_{it} &= \mathcal{N}(0, 3).
\end{aligned}$$

The causal effect is denoted by τ and thus, $Y_{it}(1) = \tau + Y_{it}(0)$ where we set $\tau = 0.2$. Finally, $Y_{it} = Y_{it}(0)$ for $t \leq 3$ (pre-treatment periods) and $Y_{it} = G_i Y_{it}(1) + (1 - G_i) Y_{it}(0)$ for $t = 4$ (post-treatment period). The half of the samples are in the treatment group ($G_i = 1$) and the other half is in the control group ($G_i = 0$).

In Figure A2, we set the autocorrelation parameter $\rho = 0.6$. This value is similar to the autocorrelation parameter used in famous simulation studies in Bertrand et al. (2004) ($\rho = 0.8$). We pick a smaller value to make our simulations harder as we see below. In Figure A3, we also provide additional results where we consider a full range of the autocorrelation parameters $\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$ (the same positive autocorrelation values considered in Bertrand et al. (2004)). Both figures show the absolute bias and the standard errors which are defined as

$$\text{absolute bias} = \left| \frac{1}{M} \sum_{m=1}^{M} (\widehat{\tau}_m - \tau) \right| \quad \text{and} \quad \text{standard error} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (\widehat{\tau}_m - \tau)^2},$$

where M is the total number of Monte Carlo iterations. Note that this standard error is a true standard error over the sampling distribution.

G.2 Results

Figure A2 shows the results when the autocorrelation parameter $\rho = 0.6$. To begin with the absolute bias, visualized in the first row, all estimators have little bias under the extended parallel trends assumption (Scenario 1), as expected from theoretical results. In contrast, under the parallel-trends-in-trends assumption (Scenario 2), the extended DID (white circle with dotted line) is biased, while the double DID (black circle with solid line) and the sequential DID (white triangle with dotted line) are unbiased.

The second row represents the standard errors of each estimator. Under the extended parallel trends assumption (the first column), the double DID estimator has the smallest standard error, smaller than the extended DID estimator (i.e., the two-way fixed effects estimator). This efficiency gain comes from the fact that the double DID uses the GMM framework to optimally weight observations from different time periods, although the two-way fixed effects estimator uses equal weights to all pre-treatment periods.

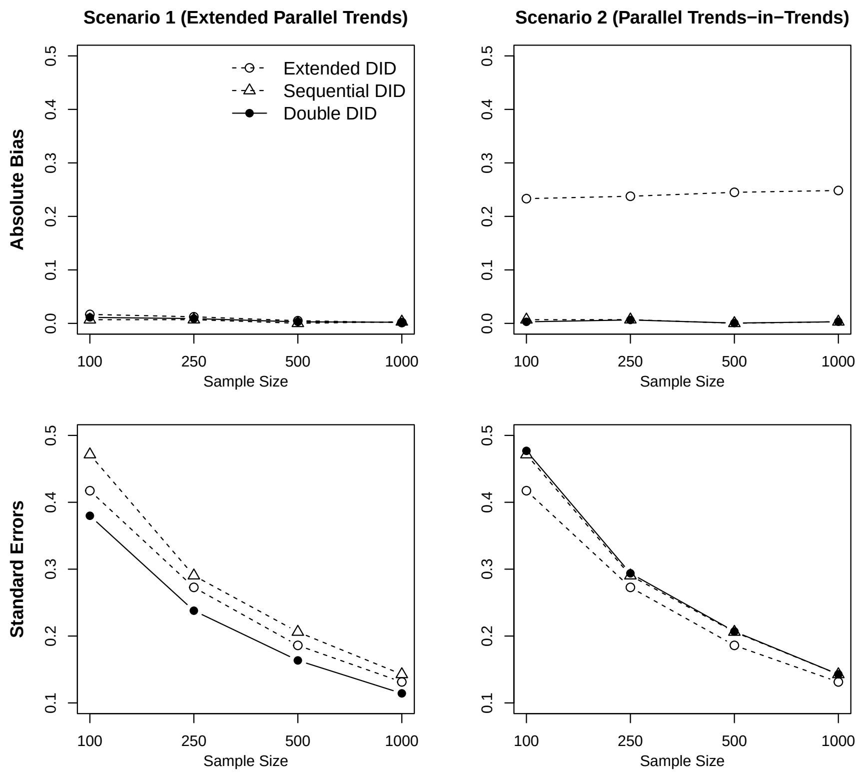

Figure A2: Comparing DID estimators in terms of the absolute bias and the standard errors. The first row shows that the double DID estimator (black circle with solid line) is unbiased under both scenarios. The second row demonstrates that the double DID has the smallest standard errors among unbiased DID estimators.

Under the parallel trends-in-trends assumption (the second row; the second column), the double DID has almost the same standard error as the sequential DID. This shows that the double DID changes weights according to scenarios and solves a practical dilemma of the sequential DID — it is unbiased under the weaker assumption of the parallel trends-in-trends, but not efficient under the extended parallel trends.

In Figure A3, we provide additional results where we consider a full range of the autocorrelation parameters $\rho \in \{0, 0.2, 0.4, 0.6, 0.8\}$ (the same positive autocorrelation values considered in Bertrand et al. (2004)). We find that when the autocorrelation of errors is small, standard errors of the double DID are smaller than those of the sequential DID even under the parallel trends-in-trends assumption.

The first row of Figure A3 shows that our results on the (absolute) bias do not change regardless of the autocorrelation of errors. In particular, the double DID is unbiased under the extended parallel trends assumption (the first column) or under the parallel trends-in-trends assumption (the second column). In terms of the standard errors (the second row), two results are important. First, under the extended parallel trends assumption (the first column), the standard errors of the double DID is the smallest for all the values of ρ and the efficiency gain relative to the extended DID (i.e., two-way fixed effects estimator) is large when the there is high auto-correlations (i.e., ρ is large).

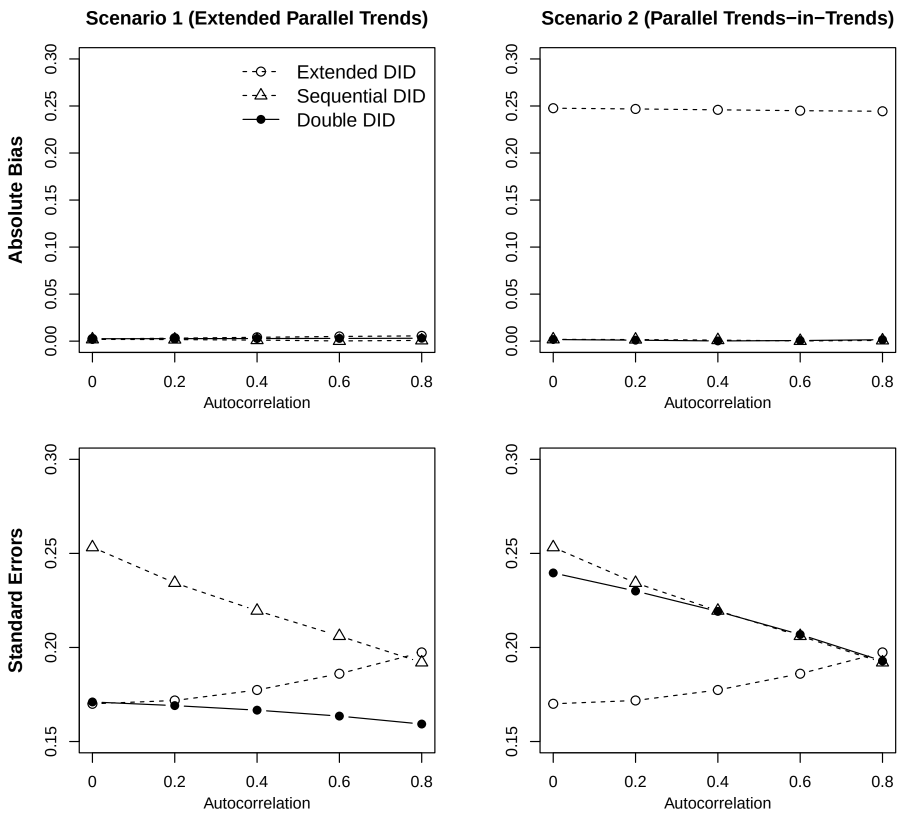

Figure A3: Comparing DID estimators in terms of the absolute bias and the standard errors according to the autocorrelation of errors. *Note*: The first row shows that the double DID estimator (black circle with solid line) is unbiased under both scenarios. The second row demonstrates that the double DID has the smallest standard errors among unbiased DID estimators. Under the extended parallel trends assumption (the first column), the efficiency gain relative to the extended DID (i.e., two-way fixed effects estimator) is large when the autocorrelation parameter ρ is large. Under the parallel trends-in-trends assumption (the second column), the efficiency gain relative to the sequential DID is large when ρ is small.

Second, under the parallel trends-in-trends assumption (the second column), the standard errors of the double DID is the smallest among unbiased DID estimators (the extended DID is biased). The efficiency gain relative to the sequential DID is large when ρ is small.

H Empirical Application

H.1 Malesky, Nguyen, and Tran (2014): DID Design

In Section 3.4, we have focused on three outcomes to illustrate the advantage of the double DID estimator. Each outcome is defined as follows. "Education and Cultural Program" (binary): This variable takes one if there is a program that invests in culture and education in the commune. "Tap Water" (binary): What is the main source of drinking /cooking water for most people in this commune? "Agricultural Center" (binary): Is there any agriculture extension center in a given commune? Please see Malesky et al. (2014) for further details.

In this section, we provide results for all thirty outcomes analyzed in the original paper. To assess the underlying parallel trends assumptions, we combine visualization and formal tests, as recommended in the main text. The assessment suggests that we can make the extended parallel trends assumption for fifteen outcomes. Specifically, for those fifteen outcomes, p-values for the null of pre-treatment parallel trends are above 0.10 (i.e., fail to reject the null at the conventional level), and the 95% standardize equivalence confidence interval is contained in the interval $[-0.2, 0.2]$. This means that the deviation from the parallel trends in the pre-treatment periods are less than 0.2 standard deviation of the control mean in 2006.

Figure A4 shows estimated treatment effects under the extended parallel trends assumption. As in Section 3.4, the double DID estimates are similar to those from the standard DID, and yet, standard errors are smaller because the double DID effectively uses pre-treatment periods within the GMM. Here, we only have two pre-treatment periods, but when there are more pre-treatment periods, the efficiency gain of the double DID can be even larger.

We rely on the parallel trends-in-trends assumption for eight outcomes out of the fifteen remaining outcomes. These outcomes have the 95% standardized equivalence confidence interval wider than $[-0.20, 0.20]$, but show that treatment and control groups' pre-treatment trends have the same sign. The same sign of the pre-treatment trends suggests that parallel trends-in-trends assumption, which can account for the linear time-varying unmeasured confounder, can be plausible for these outcomes, even though the stronger parallel trends assumption is possibly violated.

Figure A5 shows results under the parallel trends-in-trends assumption. As in Section 3.4, the double DID estimates are often different from those of the standard DID because the extended parallel trends assumption is implausible for these outcomes. Importantly, standard errors of the double DID are often larger than the standard DID. This is because the double DID needs to adjust for biases in the standard DID by using pre-treatment trends.

For the remaining seven outcomes of which treatment and control groups' pre-treatment trends have the opposite sign, it is difficult to justify either the extended parallel trends or parallel trends-in-trends assumption without additional information. Thus, there is no credible estimator for the ATT without making stronger assumptions. When there are more than two pre-treatment periods, researchers can apply the sequential DID estimator to pre-treatment periods in order to formally assess the extended parallel trends-in-trends assumption. We emphasize that, although we use the equivalence range of $[-0.20, 0.20]$ as a cutoff for an illustration, it is recommended to base this

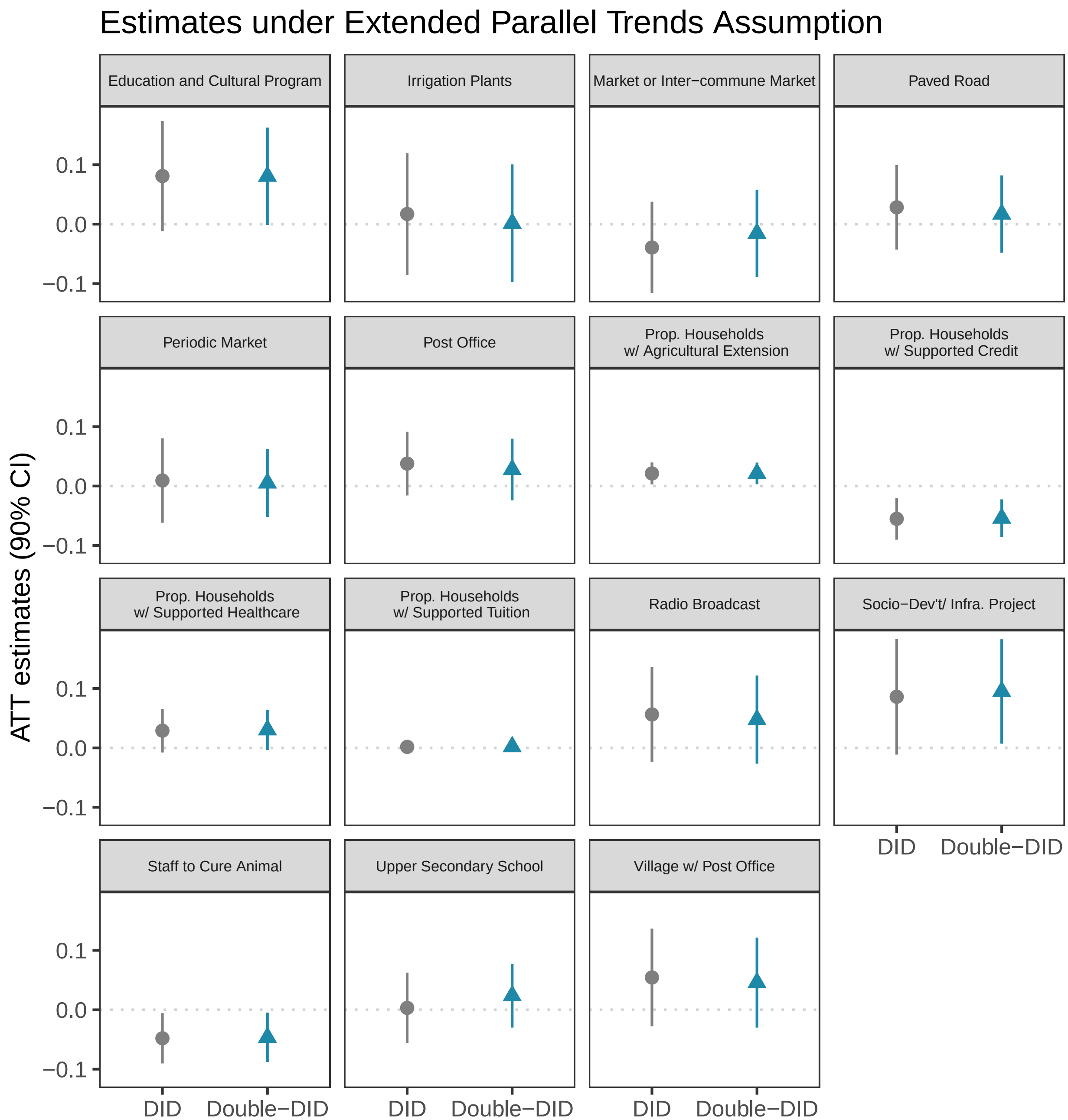

Figure A4: Comparing Standard DID and Double DID under Extended Parallel Trends Assumption. The double DID estimates are similar to those from the standard DID, and yet, standard errors are smaller because the double DID effectively uses pre-treatment periods within the GMM.

decision on substantive domain knowledge whenever possible in practice.

H.2 Paglayan (2019): Staggered Adoption Design

In this section, we apply the proposed double DID estimator to revisit Paglayan (2019), which uses the staggered adoption (SA) design to study the effect of granting collective bargaining rights

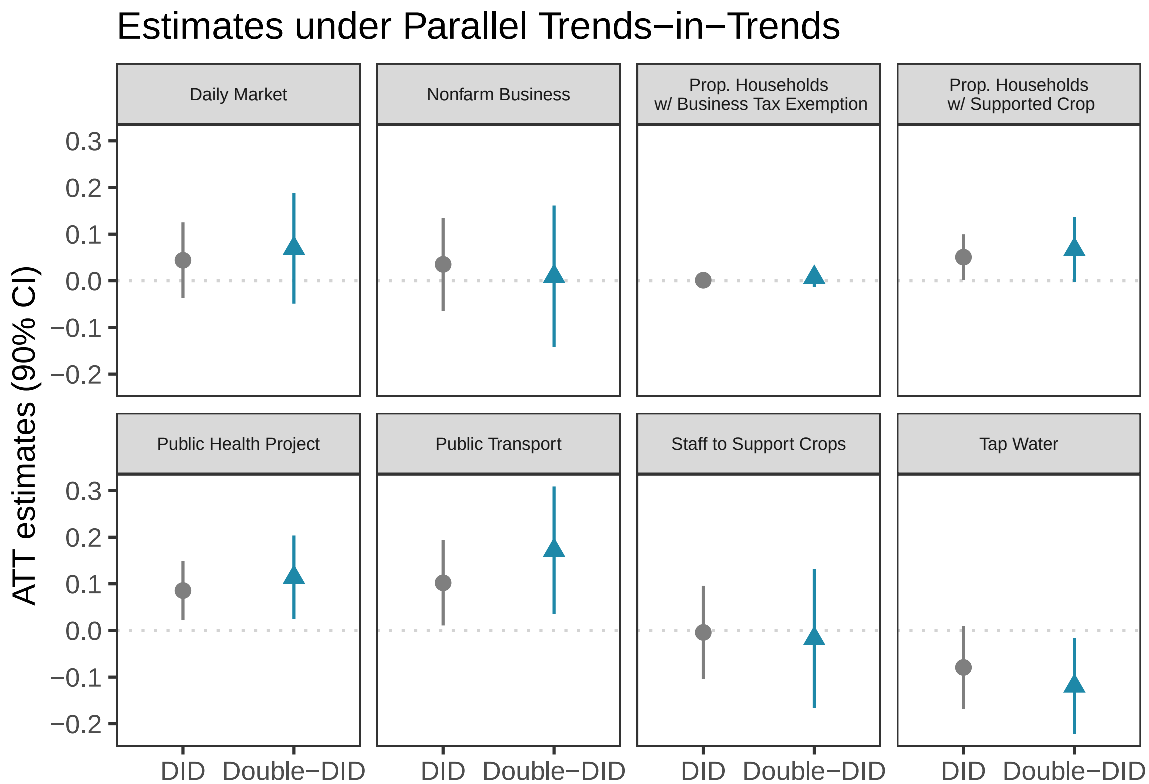

Figure A5: Comparing Standard DID and Double DID under Parallel Trends-in-Trends Assumption. The double DID estimates are often different from those of the standard DID because the extended parallel trends assumption is implausible for these outcomes.

to teacher's union on educational expenditures and teacher's salary. Paglayan (2019) applies the standard two-way fixed effect models to estimate the effect of the introduction of the mandatory bargaining law in the US states on the two outcome. The original author exploits the variation induced by the different introduction timing of the law: A few states introduced the law as early as in the mid 1960's, while some states, such as Arizona or Kentucky, never introduced the mandate. Among the states that granted the bargaining rights, the introduction timing varies from the mid 1960's to the mid 1980's (Nebraska was the last state that adopted the law).

H.2.1 Assessing Underlying Assumptions

We apply the proposed double DID for the SA design to the panel data consists of state-year observations. A state is treated at a particular year, if the state passes the law or has already passed the law of mandatory bargaining. Following the original study, we study two outcome: Per-pupil expenditure and annual teacher salary, both are on a log scale. There are 2,058 observations, containing 49 states (excluding Washington DC and Wisconsin, due to the short availability of the pre-treatment outcomes) and spanning from 1959 through 2000.

Figure A6 shows the variation of the treatment across states and over time. Cells in gray indicate

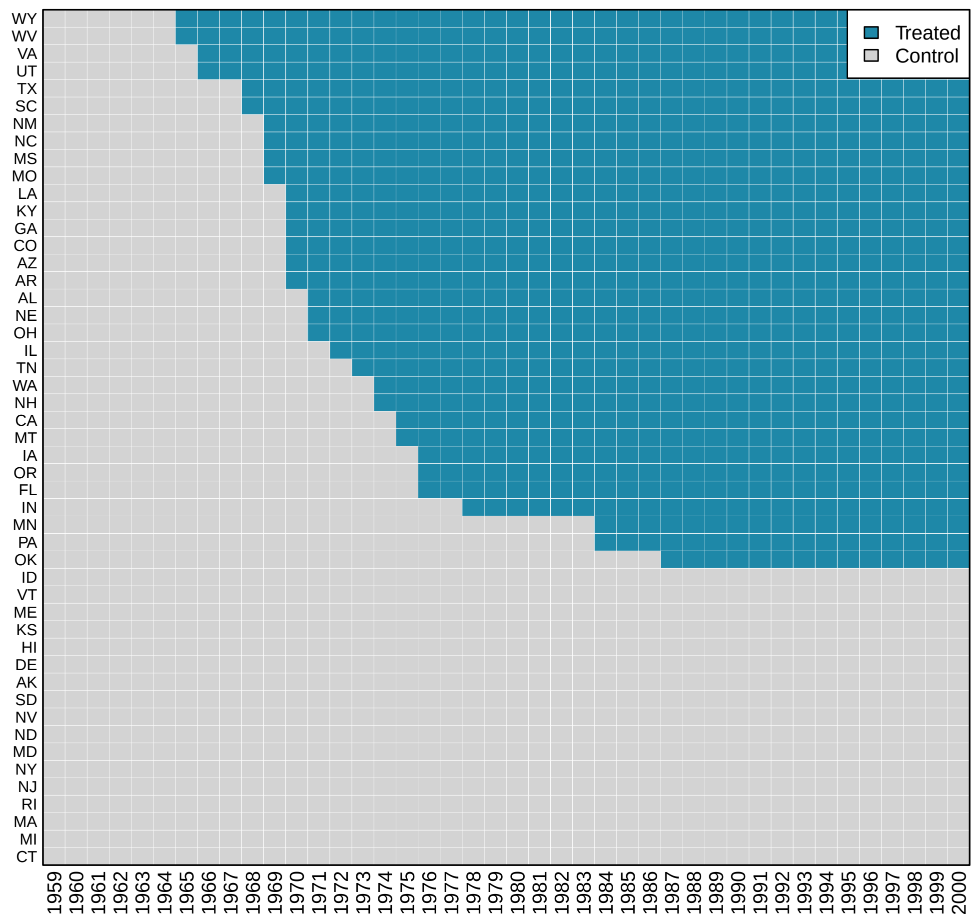

Figure A6: Treatment Variation Plot. *Note:* Cells in gray are state-year observations that are not treated (i.e., the mandatory bargaining law is not implemented), while cells in blue are observations that are under the treatment condition. Rows are sorted such that states that adopt the policy at earlier years are shown near the top, while states that never adopt the policy are shown near the bottom. The figure indicates that there are variations across states in adoption timings, and that some states never adopt the policy.

state-year observations that are not treated and blue cells indicate the treated observations. We can observe that there are 14 unique treatment timings (the earliest is 1965 and the latest is 1987) where the number of states at each treatment timing varies from one to six (the average number of states at a treatment timing is 2.3). We can also see that there is no reversal of a treatment status in that once a state adopts the policy, the state has never abolished it during the sample period.

We assess the underlying parallel trends assumption for the SA design by utilizing the pre-treatment outcome. As in the pre-treatment-trends test in the basic DID design, we apply the standard DID estimator for the SA design to pre-treatment periods. For example, to test the pre-treatment trends from $t-1$ to t for units who receive the treatment at time t, we estimate the SA-ATT using the outcome from $t-2$ and $t-1$ (See Section 4.2 for more details). To further facilitate interpretation, we standardize the outcome by the mean and standard deviation of the

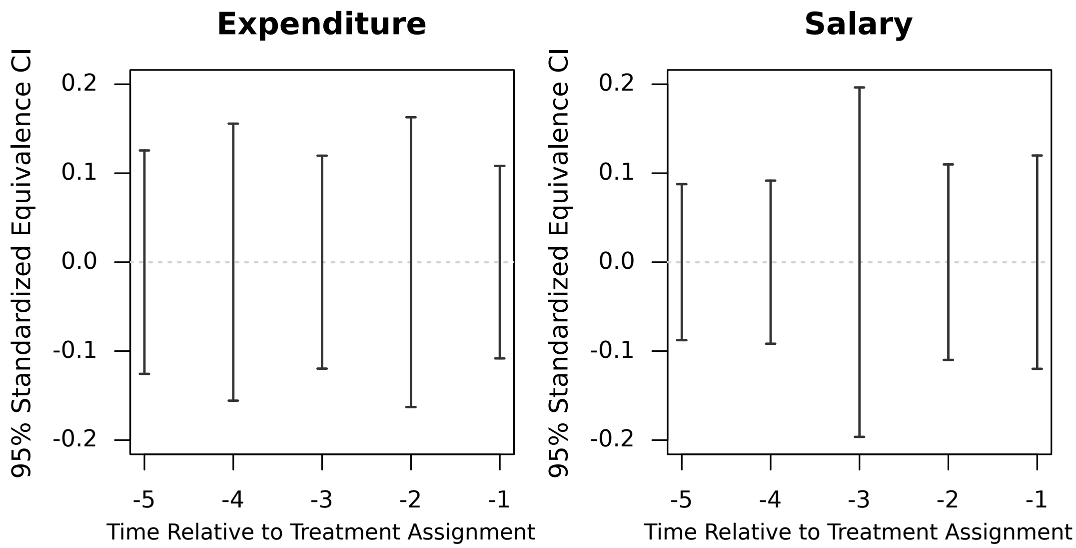

Figure A7: Assessing Underlying Assumptions Using the Pre-treatment Outcomes (Left: logged expenditure; Right: logged teacher salary). *Note:* We report the 95% standardized equivalence confidence intervals.

baseline control group, so that the effect can be interpreted relative to the control group.

Figure A7 shows 95% standardized equivalence confidence intervals for the two outcomes of interest (See Section 3.1 for details on the standardization procedure). It shows that for both outcomes, the equivalence confidence intervals are within 0.2 standard deviation from the means of the baseline control groups through $t-5$ to $t-1$. This suggests that the extended parallel trends assumption is plausible for both outcomes.

H.2.2 Estimating Causal Effects

We apply the double DID for the SA design as described in Section 4. The standard errors are computed by conducting the block bootstrap where the block is taken at the state level and we take 2000 bootstrap iterations. Analyses for the two outcomes are conducted separately. In addition to the proposed method, we apply two existing variants of synthetic control methods that can handle the staggered adoption design: the generalized synthetic control method, `gsynth` (Xu, 2017), and the augmented synthetic control method, `augsynth` (Ben-Michael et al., 2019). While the proposed double DID is better suited for settings where there are a small to moderate number of pre-treatment periods, we evaluate, in the setting of long pre-treatment periods, whether it can achieve comparable performance to these variants of synthetic control methods that are primarily designed to deal with long pre-treatment periods (see more discussions in Section B.3).

Figure A8 shows the estimates of the treatment on the per-pupil expenditure (the first row) and the teacher's salary (the second row), where both effects are on a log scale. We estimated the average treatment effect on the two outcomes ℓ periods after the treatment assignment where $\ell = \{0, 1, \ldots, 9\}$. Note that $\ell = 0$ corresponds to the contemporaneous effect. Each column corresponds to different estimators. The first column shows the proposed double DID estimator for

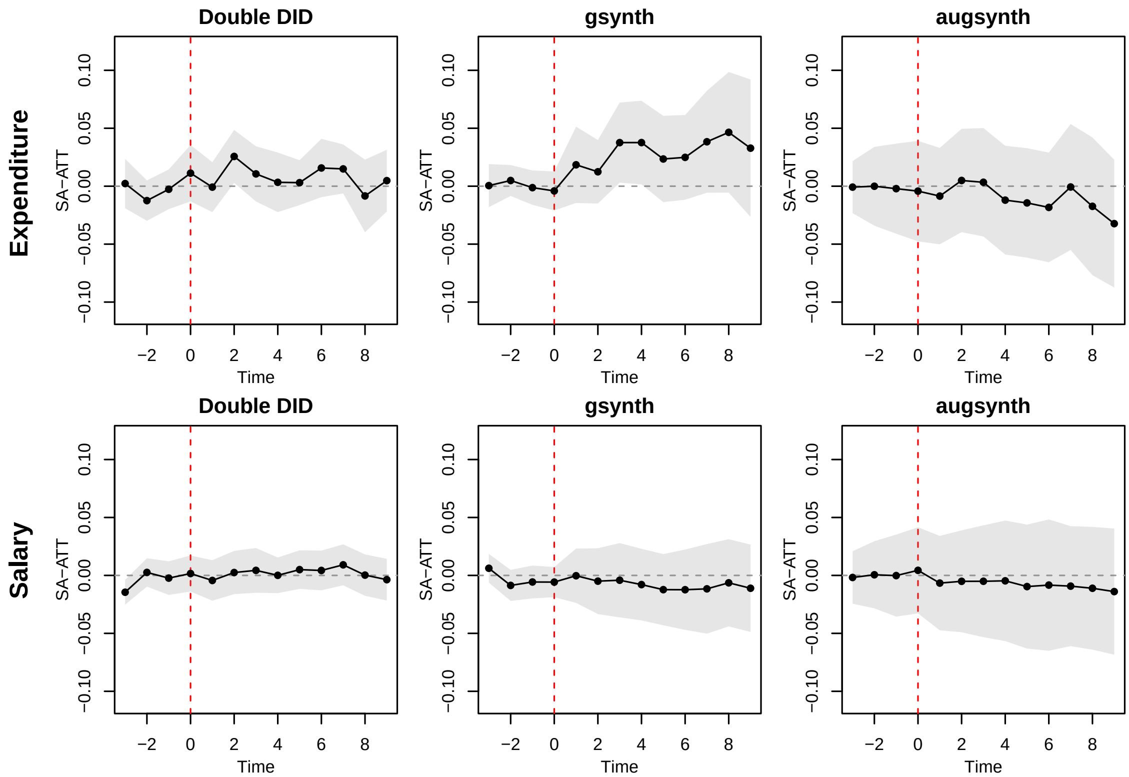

Figure A8: Plot of the Average Treatment Effect on the Treated on Two Outcomes. *Note:* We compare estimates from the double DID, the generalized synthetic control method, and the augmented synthetic control method. The causal estimates are similar across methods for both outcomes and treatment effects are not statistically significant at the conventional 5% level for most of the time periods.

the staggered adoption design, whereas the second (third) column shows estimates based on the generalized synthetic control method (the augmented synthetic control method). We can see that estimates are similar across methods for both outcomes and treatment effects are not statistically significant at the 5% level for most of the time periods. This result is consistent with the original finding of Paglayan (2019) that the granting collective bargaining rights did not increase the level of resources devoted to education.

As in this example, when there are a large number of pre-treatment periods, it is important to apply both synthetic control methods and the proposed double DID, and evaluate robustness across those approaches. This is critical because they rely on different identification assumptions. We found such robustness in this application, which provides us with additional credibility.

References

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *The Review of Economic Studies* 72(1):1–19.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Athey, Susan and Guido W Imbens. 2021. "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption." *Journal of Econometrics* .

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens and Khashayar Khosravi. 2021. "Matrix Completion Methods for Causal Panel Data Models." *Journal of the American Statistical Association* .

Bechtel, Michael M and Jens Hainmueller. 2011. "How Lasting is Voter Gratitude? An Analysis of the Short-and Long-Term Electoral Returns to Beneficial Policy." *American Journal of Political Science* 55(4):852–868.

Ben-Michael, Eli, Avi Feller and Jesse Rothstein. 2018. "The Augmented Synthetic Control Method." Available at `https://arxiv.org/abs/1811.04170`.

Ben-Michael, Eli, Avi Feller and Jesse Rothstein. 2019. "Synthetic Controls and Weighted Event Studies with Staggered Adoption." *arXiv preprint arXiv:1912.03290* .

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119(1):249–275.

Bullock, Will and Joshua D Clinton. 2011. "More a Molehill Than a Mountain: The Effects of the Blanket Primary on Elected Officials' Behavior from California." *The Journal of Politics* 73(3):915–930.

Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press.

Dube, Arindrajit, Oeindrila Dube and Omar García-Ponce. 2013. "Cross-Border Spillover: US Gun Laws and Violence in Mexico." *American Political Science Review* 107(3):397–417.

Earle, John S and Scott Gehlbach. 2015. "The Productivity Consequences of Political Turnover: Firm-Level Evidence from Ukraine's Orange Revolution." *American Journal of Political Science* 59(3):708–723.

Garfias, Francisco. 2018. "Elite Competition and State Capacity Development: Theory and Evidence from Post-Revolutionary Mexico." *American Political Science Review* 112(2):339–357.

Hall, Andrew B. 2016. "Systemic Effects of Campaign Spending: Evidence From Corporate Contribution Bans in US State Legislatures." *Political Science Research and Methods* 4(2):343–359.

Hansen, Lars Peter. 1982. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50(4):1029–1054.

Hartman, Erin and F Daniel Hidalgo. 2018. "An Equivalence Approach to Balance and Placebo Tests." *American Journal of Political Science* 62(4):1000–1013.

Hazlett, Chad and Yiqing Xu. 2018. "Trajectory Balancing: A General Reweighting Approach to Causal Inference with Time-Series Cross-Sectional Data." Available at SSRN: `https://ssrn.com/abstract=3214231`.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5):1017–1098.

Imai, Kosuke and In Song Kim. 2021. "On the Use of Two-way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29(3):405–415.

Keele, Luke and William Minozzi. 2013. "How Much is Minnesota like Wisconsin? Assumptions and Counterfactuals in Causal Inference with Observational Data." *Political Analysis* 21(2):193–216.

Ladd, Jonathan McDonald and Gabriel S Lenz. 2009. "Exploiting A Rare Communication Shift to Document the Persuasive Power of The News Media." *American Journal of Political Science* 53(2):394–410.

Larreguy, Horacio and John Marshall. 2017. "The Effect of Education on Civic and Political Engagement in Nonconsolidated Democracies: Evidence from Nigeria." *Review of Economics and Statistics* 99(3):387–401.

Lee, Myoung-jae. 2016. "Generalized Difference in Differences With Panel Data and Least Squares Estimator." *Sociological Methods & Research* 45(1):134–157.

Malesky, Edmund J, Cuong Viet Nguyen and Anh Tran. 2014. "The Impact of Recentralization on Public Services: A Difference-in-Differences Analysis of the Abolition of Elected Councils in Vietnam." *American Political Science Review* 108(1):144–168.

Mora, Ricardo and Iliana Reggio. 2012. "Treatment Effect Identification Using Alternative Parallel Assumptions." Working Paper 12-33 Economic Series (48), Universidad Carlos III. Available at `https://e-archivo.uc3m.es/bitstream/handle/10016/16065/we1233.pdf?sequence=1`.

Mora, Ricardo and Iliana Reggio. 2019. "Alternative Diff-in-Diffs Estimators with Several Pretreatment Periods." *Econometric Reviews* 38(5):465–486.

Paglayan, Agustina S. 2019. "Public-Sector Unions and the Size of Government." *American Journal of Political Science* 63(1):21–36.

Strezhnev, Anton. 2018. "Semiparametric Weighting Estimators for Multi-Period Difference-in-Differences Designs." Presented at the 2018 American Political Science Association Meeting.

Sun, Liyang and Sarah Abraham. 2020. “Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects.” *Journal of Econometrics* .

Truex, Rory. 2014. “The Returns to Office in a “Rubber Stamp” Parliament.” *American Political Science Review* 108(2):235–251.

Wellek, Stefan. 2010. *Testing Statistical Hypotheses of Equivalence and Noninferiority*. Chapman and Hall/CRC.

Xu, Yiqing. 2017. “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models.” *Political Analysis* 25(1):57–76.