Supplementary Appendix: Difference-in-Differences when Parallel Trends Holds Conditional on Covariates

Carolina Caetano[*] Brantly Callaway[†]

March 21, 2026

The Supplementary Appendix provides proofs for many of the results in the main text, formalizes results for some of the claims in the main text, and provides additional details on topics that were only briefly mentioned in the main text. Section SA1 covers the two period case. Section SA2 covers the multiple period case. Section SA3 provides additional clarifications and details about some issues mentioned in the main text. Section SA4 provides some supplementary results and discussion for the application about stand-your-ground laws in the main text. Finally, Section SA5 contains a systematic review of empirical difference-in-differences papers and how covariates are used in those papers.

SA1 Additional Results with Two Periods

This section contains additional results and proofs related to the setting with two time periods in the main text.

SA1.1 Identification Results

To start, we show that the ATT is identified under the conditions discussed in the main text.

Proposition SA1. *Under Assumptions 1 to 3,*

$$\mathrm{ATT} = \mathbb{E}[\Delta Y_{t^*}|D=1] - \mathbb{E}\Big[\mathbb{E}[\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z, D=0]\big|D=1\Big].$$

Proof. Notice that

$$\begin{aligned}
\mathrm{ATT} &= \mathbb{E}[Y_{t^*}(1) - Y_{t^*}(0)|D=1] \\
&= \mathbb{E}[Y_{t^*}(1) - Y_{t^*-1}(0)|D=1] - \mathbb{E}[Y_{t^*}(0) - Y_{t^*-1}(0)|D=1] \\
&= \mathbb{E}[Y_{t^*}(1) - Y_{t^*-1}(0)|D=1] - \mathbb{E}\Big[\mathbb{E}[\Delta Y_{t^*}(0)|X_{t^*}, X_{t^*-1}, Z, D=1]\Big|D=1\Big] \\
&= \mathbb{E}[\Delta Y_{t^*}|D=1] - \mathbb{E}\Big[\mathbb{E}[\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z, D=0]\Big|D=1\Big],
\end{aligned}$$

where the first equality holds by the definition of ATT, the second equality holds by adding and subtracting $\mathbb{E}[Y_{t^*-1}(0)|D=1]$, the third equality holds by the law of iterated expectations, and the last equality holds by Assumptions 2 and 3 and replaces potential outcomes with their observed counterparts. ☐

[*]John Munro Godfrey, Sr. Department of Economics, University of Georgia. carol.caetano@uga.edu

[†]John Munro Godfrey, Sr. Department of Economics, University of Georgia. brantly.callaway@uga.edu

SA1.2 Proofs of Results from Main Text

Proof of Lemma 1. Starting with the left-hand side of the expression in the lemma, we have that

$$\begin{aligned}\mathbb{E}\Big[\mathrm{L}(D|\Delta X_{t^*})\mathrm{L}_d(\Delta Y_{t^*}|\Delta X_{t^*})\Big|D=d\Big] &= \gamma'\mathbb{E}\Big[\Delta X_{t^*}\Delta Y_{t^*}\Big|D=d\Big] - \gamma'\mathbb{E}\Big[\Delta X_{t^*}(\Delta Y_{t^*}-\mathrm{L}_d(\Delta Y_{t^*}|\Delta X_{t^*}))\Big|D=d\Big] \\ &= \mathbb{E}\Big[\mathrm{L}(D|\Delta X_{t^*})\Delta Y_{t^*}|D=d\Big],\end{aligned}$$

where the first equality holds by the definition of $\mathrm{L}(D|\Delta X_{t^*})$ and by adding and subtracting the term $\gamma'\mathbb{E}[\Delta X_{t^*}\Delta Y_{t^*}|D=d]$, and the second equality holds because $\gamma'\Delta X_{t^*} = \mathrm{L}(D|\Delta X_{t^*})$ (for the first term) and because ΔX_{t^*} is uncorrelated with the projection error $\Delta Y_{t^*} - \mathrm{L}_d(\Delta Y_{t^*}|\Delta X_{t^*})$ conditional on $D=d$ (for the second term). □

Proof of Lemma 2. Notice that

$$\begin{aligned}\mathbb{E}\Big[(D-\mathrm{L}(D|\Delta X_{t^*}))^2\Big] &= \mathbb{E}\Big[(D-\mathrm{L}(D|\Delta X_{t^*}))D\Big] \\ &= \mathbb{E}\Big[1-\mathrm{L}(D|\Delta X_{t^*})\Big|D=1\Big]\pi,\end{aligned}$$

where the first equality holds because $\mathrm{L}(D|\Delta X_{t^*}) = \Delta X_{t*}'\gamma$ is uncorrelated with the projection error $D - \mathrm{L}(D|\Delta X_{t^*})$, and the second equality holds by the law of iterated expectations. □

Proof of Proposition 1. To conserve on notation, let $X = (X_{t^*}, X_{t^*-1}, Z)$. Then, we can re-write $\widetilde{\mathrm{ATT}}$ as

$$\widetilde{\mathrm{ATT}} = \mathbb{E}\left[\Delta Y_{t^*} - \mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=1\right] - \mathbb{E}\left[\tilde{w}_0^{aipw}(\Delta Y_{t^*} - \mathrm{L}_0(\Delta Y_{t^*}|X))\Big|D=0\right].$$

Re-arranging terms, we have that

$$\begin{aligned}\widetilde{\mathrm{ATT}} &= \mathbb{E}\left[\Delta Y_{t^*}\Big|D=1\right] - \mathbb{E}\left[\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=1\right] - \mathbb{E}\left[\tilde{w}_0^{aipw}\Delta Y_{t^*}\Big|D=0\right] + \mathbb{E}\left[\tilde{w}_0^{aipw}\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=0\right] \\ &=: A - B - C + D. \qquad \text{(SA1)}\end{aligned}$$

Terms A and C are straightforward to deal with as they are already weighted averages of ΔY_{t^*}. For Term B, from Lemma SA1 we have that

$$B = \mathbb{E}\left[\frac{(1-\pi)}{\pi}\gamma_0' X\Delta Y_{t^*}\Big|D=0\right],$$

where γ_0 is the projection coefficient from projecting $p(X)/(1-p(X))$ on X among the untreated group. Furthermore, notice that

$$\mathbb{E}\left[\frac{(1-\pi)}{\pi}\gamma_0' X\Big|D=0\right] = \mathbb{E}\left[\frac{(1-\pi)}{\pi}\frac{p(X)}{(1-p(X))}\Big|D=0\right] = 1,$$

where the first equality holds because $\gamma_0' X$ is the linear projection of $p(X)/(1-p(X))$ on X among the untreated group, and, hence, the corresponding projection errors have mean zero for the untreated group, and the second equality holds by repeated application of the law of iterated expectations. Thus, we have that

$$B = \mathbb{E}\left[\vartheta_{0,B}^{aipw}\Delta Y_{t^*}\Big|D=0\right] \quad \text{where} \quad \vartheta_{0,B}^{aipw} = \frac{\gamma_0' X}{\mathbb{E}[\gamma_0' X|D=0]}. \tag{SA2}$$

Next, turning to the "numerator" of Term D, notice that

$$\begin{aligned}
\mathbb{E}\left[\tilde{\varpi}_0^{aipw}\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=0\right] &= \frac{(1-\pi)}{\pi}\mathbb{E}\left[\frac{\tilde{p}(X)}{(1-\tilde{p}(X))}X'\mathbb{E}[XX'|D=0]^{-1}\mathbb{E}[X\Delta Y_{t^*}|D=0]\Big|D=0\right] \\
&= \frac{(1-\pi)}{\pi}\underbrace{\mathbb{E}\left[\frac{\tilde{p}(X)}{(1-\tilde{p}(X))}X'\Big|D=0\right]\mathbb{E}[XX'|D=0]^{-1}}_{\tilde{\gamma}_0'}\mathbb{E}[X\Delta Y_{t^*}|D=0] \\
&= \mathbb{E}\left[\frac{(1-\pi)}{\pi}\tilde{\gamma}_0' X\Delta Y_{t^*}\Big|D=0\right], \qquad \text{(SA3)}
\end{aligned}$$

where the first equality holds by the definitions of $\tilde{\varpi}_0^{aipw}$ and $\mathrm{L}_0(\Delta Y_{t^*}|X)$, the second equality removes the nonrandom terms from the expectation, and the last equality holds by the definition of $\tilde{\gamma}_0$. Furthermore, for the "denominator" of Term D, notice that

$$\begin{aligned}
\mathbb{E}\left[\tilde{\varpi}_0^{aipw}\Big|D=0\right] &= \frac{(1-\pi)}{\pi}\mathbb{E}\left[\frac{\tilde{p}(X)}{(1-\tilde{p}(X))}\Big|D=0\right] \\
&= \mathbb{E}\left[\frac{(1-\pi)}{\pi}\tilde{\gamma}_0' X\Big|D=0\right], \qquad \text{(SA4)}
\end{aligned}$$

where the first equality holds by the definition of $\tilde{\varpi}_0^{aipw}$, and the second equality holds because $\tilde{\gamma}_0' X$ is the linear projection of $\tilde{p}(X)/(1-\tilde{p}(X))$ on X among the untreated group which has mean zero projection errors conditional on $D=0$. Combining the expressions in Equations (SA3) and (SA4), we have that

$$D = \mathbb{E}\left[\vartheta_{0,D}^{aipw}\Delta Y_{t^*}\Big|D=0\right] \quad \text{where} \quad \vartheta_{0,D}^{aipw} = \frac{\tilde{\gamma}_0' X}{\mathbb{E}[\tilde{\gamma}_0' X|D=0]}. \tag{SA5}$$

Combining Equations (SA2) and (SA5) with Equation (SA1) yields the first part of the result. That the weights have mean one follows because (i) $\vartheta_1^{aipw} = 1$ and (ii) each of the components of ϑ_0^{aipw} has mean one; there are three of these terms, two are added and one is subtracted as in Equation (SA1). Furthermore, given that two of the components of ϑ_0^{aipw} involve linear projections, the weights can be negative. Finally, we show that the weights balance the covariates in Lemma SA2. □

SA1.3 Conditions on the Propensity Score for Interpreting TWFE

This section considers alternative conditions on the propensity score that can rationalize interpreting α in Equation (3) as a weighted average of conditional average treatment effects, as was discussed in Remark 1 in Section 3 in the main text. These are alternative conditions that can eliminate the misspecification bias terms in Theorem 1. Consider the following assumption:

Assumption PS (Linearity of the Propensity Score)**.** $p(X_{t^*}, X_{t^*-1}, Z) = \mathrm{L}(D|\Delta X_{t^*})$.

Proposition SA2. *Under Assumptions 1 to 3 and Assumption PS,*

$$\alpha = \mathbb{E}\Big[w(\Delta X_{t^*})\text{ATT}(X_{t^*}, X_{t^*-1}, Z)\Big|D=1\Big],$$

where $w(\Delta X_{t^*})$ are the same as the weights defined in Theorem 1. In this case, the weights are guaranteed to be non-negative.

Proof. We use the decomposition of α in Proposition A2 from the appendix to the main text as a starting point. Under Assumption 3, the term in Equation (15) is equal to $\mathbb{E}[w(\Delta X_{t^*})\text{ATT}(X_{t^*}, X_{t^*-1}, Z)|D=1]$. Next, write the numerator of the term in Equation (16) as

$$\begin{aligned}\mathbb{E}\left[(1-\text{L}(D|\Delta X_{t^*}))\mathbb{E}[\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z, D=0]|D=1\right]\\ - \mathbb{E}\left[(1-\text{L}(D|\Delta X_{t^*}))\text{L}_0(\Delta Y_{t^*}|\Delta X_{t^*})|D=1\right] =: A - B,\end{aligned}$$

and we consider each of these terms in turn. First, notice that

$$\begin{aligned}A &= \mathbb{E}\left[\frac{p(X_{t^*}, X_{t^*-1}, Z)(1-\pi)}{(1-p(X_{t^*}, X_{t^*-1}, Z))\pi}(1-\text{L}(D|\Delta X_{t^*}))\mathbb{E}[\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z, D=0]\Big|D=0\right]\\ &= \mathbb{E}\left[\frac{p(X_{t^*}, X_{t^*-1}, Z)(1-\pi)}{(1-p(X_{t^*}, X_{t^*-1}, Z))\pi}(1-\text{L}(D|\Delta X_{t^*}))\Delta Y_{t^*}\Big|D=0\right]\\ &= \mathbb{E}\left[\frac{p(X_{t^*}, X_{t^*-1}, Z)(1-\pi)}{\pi}\Delta Y_{t^*}\Big|D=0\right],\end{aligned}$$

where the first equality holds by the law of iterated expectations (and it is worth mentioning that the outside expectation is over the joint distribution of (X_{t^*}, X_{t^*-1}, Z) which accounts for the propensity score depending on all three of these rather than, say, only ΔX_{t^*}), the second equality also holds by the law of iterated expectations, and the third equality holds by Assumption PS. Next, notice that

$$\begin{aligned}B &= \mathbb{E}\left[\frac{p(X_{t^*}, X_{t^*-1}, Z)(1-\pi)}{(1-p(X_{t^*}, X_{t^*-1}, Z))\pi}(1-\text{L}(D|\Delta X_{t^*}))\text{L}_0(\Delta Y_{t^*}|\Delta X_{t^*})\Big|D=0\right]\\ &= \mathbb{E}\left[\frac{1-\pi}{\pi}\text{L}(D|\Delta X_{t^*})\text{L}_0(\Delta Y_{t^*}|\Delta X_{t^*})\Big|D=0\right]\\ &= \mathbb{E}\left[\frac{1-\pi}{\pi}\text{L}(D|\Delta X_{t^*})\Delta Y_{t^*}\Big|D=0\right]\\ &= \mathbb{E}\left[\frac{p(X_{t^*}, X_{t^*-1}, Z)(1-\pi)}{\pi}\Delta Y_{t^*}\Big|D=0\right],\end{aligned}$$

where the first equality holds by the law of iterated expectations, the second equality holds by Assumption PS (it cancels the terms involving one minus the propensity score and the one minus the linear projection and then rewrites the propensity score in the numerator as the linear projection in Assumption PS), the third equality holds by Lemma 1, and the last equality holds by Assumption PS. That $A = B$ implies the first part of the proposition. That the weights are non-negative under Assumption PS holds because $p(X_{t^*}, X_{t^*-1}, Z)$ is uniformly bounded below 1. □

Discussion

Imposing conditions on the propensity score is very common in the literature on interpreting regressions under unconfoundedness with cross-sectional data (e.g., Angrist (1998), Aronow and Samii (2016), Sloczynski (2022), and Ishimaru (2024)). However, in our setting (and as mentioned in Remark 1 in the main text), the conditions required for Assumption PS to hold in applications are likely to be quite strong. Assumption PS says that the probability of being treated conditional on time-varying and time-invariant characteristics is equal to the linear projection of the treatment on the change in the time-varying covariates over time. This condition can be rationalized if (i) the propensity score does not depend on time-invariant covariates, (ii) the propensity score conditional on X_{t^*} and X_{t^*-1} only depends on ΔX_{t^*}, and (iii) the propensity score conditional on ΔX_{t^*} is linear. In addition to the main result in Theorem 2 holding in this case, the weights in Theorem 2 are guaranteed to be non-negative under this alternative assumption. In cross-sectional cases, linearity of the propensity score sometimes holds by construction (e.g., when the covariates are all discrete and a full set of interactions is included in the model). In our case, though, this seems particularly implausible because it requires the propensity score to only depend on changes in covariates over time and, even with fully interacted discrete regressors, the propensity score is unlikely to be linear in changes in the regressors over time.[1] Finally, unlike linear models for the outcome, imposing a linear model for the propensity score is not usually as natural relative to simple nonlinear models such as logit or probit.

SA1.4 More Details about the Properties Implicit TWFE Weights

This section proves the claims about the balancing properties of the implicit TWFE regression weights mentioned in Section 4.1 in the main text. The next proposition shows that the implicit TWFE regression weights balance the mean of ΔX_{t^*} between the treated and untreated groups.

Proposition SA3. *Under Assumptions 1 and 2,*

$$\mathbb{E}\Big[w_1(\Delta X_{t^*})\Delta X_{t^*}\Big|D=1\Big] = \mathbb{E}\Big[w_0(\Delta X_{t^*})\Delta X_{t^*}\Big|D=0\Big].$$

Proof. Consider the difference between the numerators of each term:

$$\begin{aligned}
\mathbb{E}\left[\pi(1-\mathrm{L}(D|\Delta X_{t^*}))\Delta X_{t^*}\Big|D=1\right] &- \mathbb{E}\left[(1-\pi)\mathrm{L}(D|\Delta X_{t^*})\Delta X_{t^*}\Big|D=0\right] \\
&= \mathbb{E}\Big[D(1-\mathrm{L}(D|\Delta X_{t^*}))\Delta X_{t^*} - (1-D)\mathrm{L}(D|\Delta X_{t^*})\Delta X_{t^*}\Big] \\
&= \mathbb{E}\Big[\Delta X_{t^*}\Big(D - D\mathrm{L}(D|\Delta X_{t^*}) - \mathrm{L}(D|\Delta X_{t^*}) + D\mathrm{L}(D|\Delta X_{t^*})\Big)\Big] \\
&= \mathbb{E}\Big[\Delta X_{t^*}\Big(D - \mathrm{L}(D|\Delta X_{t^*})\Big)\Big] = 0,
\end{aligned}$$

where the first equality holds by the law of iterated expectations and by combining terms, the second equality expands both terms from the previous line and factors out ΔX_{t^*}, the third equality holds by canceling terms, and the last equality holds because $(D - \mathrm{L}(D|\Delta X_{t^*})$ is the projection error of D on

[1] For example, suppose that the only covariate is binary. In the cross-sectional case considered by other papers mentioned above, the propensity score would be linear by construction. However, the change in the covariate over time would be a single variable that can take the values -1, 0, or 1. Moreover, the change in a binary covariate over time is equal to 0 when the covariate is equal to 1 in both periods or when the covariate is equal to 0 in both periods. This suggests that the propensity score would not be linear (at least not by construction) in the change in covariates over time, even in this very simple case.

ΔX_{t^*}, which is orthogonal to ΔX_{t^*}. $\square$

While Proposition SA3 shows that the implicit regression weights balance ΔX_{t^*} for the treated group relative to the untreated group, the proof is also instructive for seeing that the implicit regression weights do not necessarily balance time-invariant covariates or levels of time-varying covariates (or other functions of time-invariant and/or time-varying covariates). As leading examples, notice that

$$\mathbb{E}\Big[w_1(\Delta X_{t^*})X_{t^*}\Big|D=1\Big] - \mathbb{E}\Big[w_0(\Delta X_{t^*})X_{t^*}\Big|D=0\Big] = \frac{\mathbb{E}\Big[X_{t^*}(D-\mathrm{L}(D|\Delta X_{t^*}))\Big]}{\mathbb{E}\Big[\big(D-\mathrm{L}(D|\Delta X_{t^*})\big)^2\Big]} \neq 0$$

$$\mathbb{E}\Big[w_1(\Delta X_{t^*})X_{t^*-1}\Big|D=1\Big] - \mathbb{E}\Big[w_0(\Delta X_{t^*})X_{t^*-1}\Big|D=0\Big] = \frac{\mathbb{E}\Big[X_{t^*-1}(D-\mathrm{L}(D|\Delta X_{t^*}))\Big]}{\mathbb{E}\Big[\big(D-\mathrm{L}(D|\Delta X_{t^*})\big)^2\Big]} \neq 0$$

$$\mathbb{E}\Big[w_1(\Delta X_{t^*})Z\Big|D=1\Big] - \mathbb{E}\Big[w_0(\Delta X_{t^*})Z\Big|D=0\Big] = \frac{\mathbb{E}\Big[Z(D-\mathrm{L}(D|\Delta X_{t^*}))\Big]}{\mathbb{E}\Big[\big(D-\mathrm{L}(D|\Delta X_{t^*})\big)^2\Big]} \neq 0,$$

which holds by using the same arguments as in the proof of Proposition SA3. This shows that, in general, the implicit regression weights do not balance time-invariant covariates or the levels of time-varying covariates between the treated group and the untreated group.

Remark SA1 (Covariate balance profile for TWFE)**.** It is also worth pointing out that, although the weights balance the mean of ΔX_{t^*}, they do it with respect to a different covariate profile than the one for the ATT. For example, the correct weighting scheme for the ATT would have $w_1(\Delta X_{t^*}) = 1$, and $w_0(\Delta X_{t^*})$ such that applying them to the untreated group would re-weight it to have the same distribution of ΔX_{t^*} as for the treated group. Neither of these holds, and this suggests that, even if the implicit TWFE weights did balance the distribution of covariates, it would still not be sufficient for α to be equal to the ATT. See Chattopadhyay and Zubizarreta (2023) for an extensive discussion of this property of the weights, and an explicit expression for the covariate profile for which the weights balance.

Remark SA2 (TWFE with time-invariant covariates)**.** It is straightforward to extend the results in Theorems 1 and 2 and Proposition SA3 to a modified TWFE regression that includes time-invariant covariates with time-varying coefficients:

$$Y_{it} = \theta_t + \eta_i + \alpha D_{it} + X_{it}'\beta + Z_i'\gamma_t + e_{it}.$$

After taking first differences, the estimating equation becomes

$$\Delta Y_{it^*} = \Delta\theta_{t^*} + \alpha D_{it^*} + \Delta X_{it^*}'\beta + Z_i'\tilde{\gamma} + \Delta e_{it^*}.$$

where $\tilde{\gamma} := \gamma_{t^*} - \gamma_{t^*-1}$. From this equation, it follows that Z_i has the same properties as ΔX_{it^*} that we emphasize in the paper: (i) using this specification eliminates the bias in Term (A) of Theorem 1 and implies that there is no hidden linearity bias with respect to the time-invariant covariate, and (ii) it does not fix the weighting issues that were discussed after Theorem 2 in the main text.

SA1.5 More Details about the Properties of Implicit AIPW Weights

This section contains Lemma SA1, which is related to interpreting regression adjustment as reweighting, and the proof of the balancing properties of the implicit AIPW weights, which is a part of Proposition 1 in the main text.

Lemma SA1. *To conserve on notation, let $X = (X_{t^*}, X_{t^*-1}, Z)$. Under Assumptions 1 and 2,*

$$\mathbb{E}\Big[\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=1\Big] = \mathbb{E}\Big[\vartheta_0^{\mathrm{L}_0}\Delta Y_{t^*}\Big|D=0\Big],$$

where $\vartheta_0^{\mathrm{L}_0}$ are weights defined as

$$\vartheta_0^{\mathrm{L}_0} := \frac{(1-\pi)}{\pi}\gamma_0' X.$$

Here, γ_0 denotes the linear projection coefficient from projecting $p(X)/(1-p(X))$ on X among the untreated group. In addition, $\mathbb{E}[\vartheta_0^{\mathrm{L}_0}|D=0]=1$ (i.e., the weights have mean one).

Proof. Recall that we defined $X = (X_{t^*}, X_{t^*-1}, Z)$. Then, notice that

$$\begin{aligned}
\mathbb{E}\Big[\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=1\Big] &= \mathbb{E}\Big[X'\mathbb{E}[XX'|D=0]^{-1}\mathbb{E}[X\Delta Y_{t^*}|D=0]\Big|D=1\Big] \\
&= \mathbb{E}\Big[\mathbb{E}[X'|D=1]\mathbb{E}[XX'|D=0]^{-1}X\Delta Y_{t^*}\Big|D=0\Big] \qquad \text{(SA6)} \\
&= \frac{(1-\pi)}{\pi}\mathbb{E}\left[\underbrace{\mathbb{E}\left[\frac{p(X)}{1-p(X)}X'\Big|D=0\right]\mathbb{E}[XX'|D=0]^{-1}}_{\gamma_0'}X\Delta Y_{t^*}\Big|D=0\right] \\
&= \frac{(1-\pi)}{\pi}\mathbb{E}[\gamma_0' X\Delta Y_{t^*}|D=0] \\
&= \mathbb{E}\Big[\vartheta_0^{\mathrm{L}_0}\Delta Y_{t^*}\Big|D=0\Big],
\end{aligned}$$

where the first equality holds by the definition of $\mathrm{L}_0(\Delta Y_{t^*}|X)$, the second equality holds by rearranging the terms inside the expectation, the third equality holds by re-weighting the distribution of X for the untreated group to match the treated group (which itself follows from repeatedly applying the law of iterated expectations), the fourth equality holds by noticing that the underlined term in the previous line is equal to the projection coefficient from projecting $p(X)/(1-p(X))$ on X for the untreated group, and the last line holds by the definition of $\vartheta_0^{\mathrm{L}_0}$.

Next, we show that the weights have mean one. Notice that

$$\begin{aligned}
\mathbb{E}\left[\vartheta_0^{\mathrm{L}_0}\Big|D=0\right] &= \frac{(1-\pi)}{\pi}\mathbb{E}[\gamma_0' X|D=0] \\
&= \frac{(1-\pi)}{\pi}\mathbb{E}\left[\frac{p(X)}{1-p(X)}\Big|D=0\right] \\
&= \frac{1}{\pi}\mathbb{E}[p(X)] = 1,
\end{aligned}$$

where the first equality holds by the definition of $\vartheta_0^{\mathrm{L}_0}$, the second equality holds because the projection error from projecting $p(X)/(1-p(X))$ on X among the untreated group has mean zero, and the third and fourth equalities hold by repeatedly applying the law of iterated expectations. □

Note that it is possible that $\vartheta_0^{\mathrm{L}_0}$ can be negative for any values of the covariates among the untreated group such that the linear projection of $p(X)/(1-p(X))$ on X is negative.

Remark SA3 (Straightforward estimation of weights)**.** All the components ϑ_0^{aipw} can be calculated easily. To see this, note that the first and third term only depend on $\tilde{p}(X_{t^*}, X_{t^*-1}, Z)$, which comes from the proposed model for the propensity score. The second term is less obvious, as it depends on $p(X_{t^*}, X_{t^*-1}, Z)$, the actual (unknown) propensity score. However, notice that $\gamma_0 = \dfrac{\pi}{1-\pi}\mathbb{E}[XX'|D = 0]^{-1}\mathbb{E}[X|D = 1]$, see in particular Equation (SA6) in the proof of Lemma SA1, which can be directly estimated.

Remark SA4 (Implicit regression adjustment weights)**.** An immediate implication of Lemma SA1 is that regression adjustment estimators can be re-formulated as weighting estimators. In particular, define

$$\widetilde{\mathrm{ATT}}^{ra} := \mathbb{E}\left[\Delta Y_{t^*} - \mathrm{L}_0(\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z)\Big|D = 1\right],$$

then it immediately follows from Lemma SA1 that

$$\widetilde{\mathrm{ATT}}^{ra} := \mathbb{E}\Big[\vartheta_1^{\mathrm{L}_0}\Delta Y_{t^*}|D = 1\Big] - \mathbb{E}\Big[\vartheta_0^{\mathrm{L}_0}\Delta Y_{t^*}|D = 0\Big],$$

where $\vartheta_1^{\mathrm{L}_0} := 1$ and $\vartheta_0^{\mathrm{L}_0}$ is defined as above. This result is similar to the one in Kline (2011, Proposition 2). Our result is in the context of DiD rather than for cross-sectional settings, and we express the weights in a slightly different way that involves the linear projection of the odds ratio rather than the odds ratio itself.

Remark SA5 (More special cases of Proposition 1)**.** Remark 3 in the main text pointed out that regression adjustment and IPW estimands could be re-formulated as weighting estimands as a special case of the result for AIPW in Proposition 1. That remark can be further generalized. Suppose that one includes the covariates $X^{or} := C^{or}(X_{t^*}, X_{t^*-1}, Z)$ in the working outcome regression model, and the covariates $X^{ps} := C^{ps}(X_{t^*}, X_{t^*-1}, Z)$ in the working model for the propensity score (where C^{or} and C^{ps} are functions that could include subsets of the covariates, higher order terms, interactions, etc.), then the results in Proposition 1 continue to apply with $\tilde{p}(X^{ps})$ replacing $\tilde{p}(X_{t^*}, X_{t^*-1}, Z)$, $p(X^{or})$ replacing $p(X_{t^*}, X_{t^*-1}, Z)$, and all linear projections being on X^{or} rather than X. Although including different covariates in the working models for the outcome regression and the propensity score is not common in practice, this sort of result is potentially useful in applications where the number of treated units is relatively small. In this case, it is difficult to include many covariates in the working model for the propensity score, but it is feasible to include more covariates in the working model for the outcome regression. This is particularly relevant for staggered adoption designs, where the number of treated units in a particular timing group can be small.

Remark SA6 (Comparison to Chattopadhyay and Zubizarreta (2023))**.** As noted in the main text, our AIPW decomposition in Proposition 1 is mechanically similar to the ones in Chattopadhyay and Zubizarreta (2023). That paper considers the case with cross-sectional data under unconfoundedness, and provides a general result on re-formulating AIPW estimators as weighting estimators in that context (Theorem 3). Besides differences related to DiD relative to cross-sectional settings, there are still some differences between our results and their results. They mainly focus on ATE rather than ATT, though they briefly mention how their results apply to the ATT in their Supplementary Appendix (see, in particular, the discussion on p. 4). That said, their expression for the weights differs conceptually from ours, as our results depend to a large extent on linear projections of odds ratios rather than adjusted differences in means of the covariates across groups. In addition, our weights are slightly numerically different from the weights that we get when we use the `lmw` R package (Chattopadhyay et al. (2023)) in settings where the results are comparable. Finally, we provide a direct proof delivering the implicit AIPW weights for the ATT rather than deriving them as a byproduct of a general result.

Next, we prove the balancing properties of the implicit AIPW weights mentioned in Proposition 1 in the main text.

Lemma SA2. *To simplify notation, let $X = (X_{t^*}, X_{t^*-1}, Z)$. Under Assumptions 1 and 2,*

$$\mathbb{E}[\vartheta_0^{aipw} X | D = 0] = \mathbb{E}[X | D = 1].$$

Proof. Recalling the definition of ϑ_0^{aipw} from Proposition 1 in the main text, we have that

$$\begin{aligned}
\mathbb{E}\left[\vartheta_0^{aipw} X \middle| D = 0\right] &= \mathbb{E}\left[\left(\tilde{w}_0^{aipw} + \frac{\gamma_0' X}{\mathbb{E}[\gamma_0' X | D = 0]} - \frac{\tilde{\gamma}_0' X}{\mathbb{E}[\tilde{\gamma}_0' X | D = 0]}\right) X \middle| D = 0\right] \\
&= \mathbb{E}\left[\left(\frac{\frac{\tilde{p}(X)}{(1-\tilde{p}(X))}}{\mathbb{E}\left[\frac{\tilde{p}(X)}{(1-\tilde{p}(X))}\middle| D = 0\right]} + \frac{\gamma_0' X}{\mathbb{E}[\gamma_0' X | D = 0]} - \frac{\tilde{\gamma}_0' X}{\mathbb{E}[\tilde{\gamma}_0' X | D = 0]}\right) X \middle| D = 0\right] \\
&= \mathbb{E}\left[\frac{\gamma_0' X}{\mathbb{E}[\gamma_0' X | D = 0]} X \middle| D = 0\right] \\
&= \frac{1-\pi}{\pi}\mathbb{E}\left[\frac{p(X)}{1-p(X)} X \middle| D = 0\right] \\
&= \mathbb{E}[X | D = 1],
\end{aligned}$$

where the first equality holds by the definition of ϑ_0^{aipw}; the second equality holds by the definition of $\tilde{w}_0^{aipw}$ (and canceling the terms involving π that are common to the numerator and denominator); the third equality holds because (i) $\tilde{\gamma}_0' X$ is the linear projection of $\tilde{p}(X)/(1-\tilde{p}(X))$ on X among the untreated group and its projection error is orthogonal to X conditional on $D = 0$, and (ii) replacing $\tilde{\gamma}_0' X$ in the numerator and denominator of the third term in the previous line results in it canceling with the first term; the fourth equality holds by (i) multiplying the numerator and denominator by $\pi/(1-\pi)$ (after this multiplication the denominator is equal to 1), and (ii) $\gamma_0' X$ is the linear projection of $p(X)/(1-p(X))$ on X among the untreated group and its projection error is orthogonal to X conditional on $D = 0$; and the last equality holds by repeatedly applying the law of iterated expectations. □

SA2 Additional Results with Multiple Periods

This section contains proofs of all of our results involving multiple periods from the main text. In addition, it provides formal results for some claims in the main text regarding implicit TWFE and AIPW weights.

SA2.1 Assumptions with Multiple Periods

This section formalizes the assumptions discussed in Section 5 in the main text.

Assumption MP-1 (Staggered Treatment Adoption)**.** *For all units and time periods* $t = 2, \ldots, T$, $D_{it-1} = 1 \implies D_{it} = 1$.

Assumption MP-2 (No-Anticipation)**.** *For* $t < G_i$ *(i.e., pre-treatment periods for unit* i*),* $Y_{it} = Y_{it}(0)$.

Assumption MP-1 formalizes the notion of staggered treatment adoption discussed in the main text. It says that units can be treated at different points in time, but once a unit becomes treated, it remains treated in subsequent periods. Assumption MP-2 says that, in periods before a unit is treated, its observed outcomes are untreated potential outcomes. This rules out that the treatment affects outcomes in periods before the treatment actually occurs.

Next, define X_{it} to be a $k \times 1$ vector of time-varying covariates, and let $\mathbf{X}_i := (X_{i1}', X_{i2}', \ldots, X_{i\mathrm{T}}')'$ denote the $Tk \times 1$ vector that stacks the time-varying covariates across periods. Finally, we continue to use Z_i to denote an $l \times 1$ vector of time-invariant covariates.

Assumption MP-3 (Multi-Period Sampling)**.** *The observed data* $\{Y_{i1}, \ldots, Y_{iT}, X_{i1}, \ldots, X_{iT}, Z_i, G_i\}_{i=1}^n$ *are i.i.d.*

Assumption MP-4 (Multi-Period Overlap)**.** $\mathrm{P}(G = g) > \epsilon$ *and* $\mathrm{P}(U = 1|\mathbf{X}, Z) > \epsilon$ *for all* $g \in \mathcal{G}$, *and some* $\epsilon > 0$.

Assumptions MP-3 and MP-4 extend Assumptions 1 and 2 to a setting with more than two time periods and variation in treatment timing. The setup and assumptions considered here are standard in the DiD literature, up to paying close attention to time-varying and time-invariant covariates.

SA2.1.1 Comments/Clarifications on Assumptions MP-PT and MP-1 to MP-4

This section contains some additional discussion about Assumptions MP-PT and MP-1 to MP-4, particularly regarding some ways that these conditions can be relaxed in settings that commonly occur in applications. All the issues discussed here are often relevant for empirical work, and solutions proposed in existing work apply immediately to our framework.

Remark SA7 (Staggered treatment adoption)**.** Assumption MP-1, about staggered treatment adoption, is common in the econometrics literature and covers a large number of empirical applications, but it does not cover all empirical applications that use DiD identification strategies. See Callaway (2023) for additional discussion of this assumption and de Chaisemartin and D'Haultfoeuille (2024) and Yanagi (2022) for recent work on relaxing it.

Remark SA8 (No never-treated units)**.** It is without loss of generality to suppose a never-treated group exists. In applications, if all units are eventually treated, the setting considered in the main text implicitly drops periods where all units are treated. There is no comparison group for those periods, and difference-in-differences identification strategies are not useful for recovering treatment effect parameters in those periods (except, possibly, under additional assumptions that we do not consider here). In this case, we would set T to be the last period with available untreated units.

Remark SA9 (Already-treated units)**.** In the main text, we dropped units that were already treated in the first period. This is because DiD identification strategies are not useful for recovering treatment effect parameters for this group without imposing extra assumptions, nor is this group useful for recovering the path of untreated potential outcomes for other groups. There is one interesting assumption that allows us not to drop these units: after a group has been exposed to the treatment for "long enough", it can re-enter the comparison group. This may be an attractive assumption for some applications, as it increases the size of the comparison group, but, to be clear, it is an extra (and possibly strong) assumption. See Cengiz et al. (2019) for an example of using this type of extra condition.

Remark SA10 (Anticipation)**.** The no-anticipation assumption, Assumption MP-2, is widely used in the econometrics literature on DiD (see, for example, Callaway and Sant'Anna (2021), Sun and Abraham (2021), and Callaway (2023)). That said, it may be a strong assumption in many applications. For example, many DiD applications involve policies that are voted on in one period but not implemented until a later period. In this case, it seems likely that intelligent units (such as people or firms) would be likely to respond in the intermediate period, violating the no-anticipation condition. Although violations of no-anticipation are possible in many applications, it turns out that it is straightforward to relax no-anticipation to some version of limited-anticipation, where observed outcomes are equal to untreated potential outcomes "far enough before" the treatment occurs, by "backing up" the entire analysis so that the base period is the most recent period before anticipation effects start (rather than using $(g-1)$ as the base period of the analysis). We do not pursue relaxing no-anticipation here, as what "far enough before" means is application-specific, and, therefore, no-anticipation is the natural baseline case.

SA2.2 Supplementary Identification Results with Multiple Periods

This section provides a formal identification result for $\text{ATT}(g,t)$ under staggered treatment adoption.

Lemma SA3. *Under Assumptions MP-PT and MP-1 to MP-4 and for any group $g \in \bar{\mathcal{G}}$ and $t \geq g$ (i.e., post-treatment periods for group g)*

$$\mathbb{E}[Y_t(0) - Y_{g-1}(0)|\mathbf{X}, Z, G = g] = \mathbb{E}[Y_t(0) - Y_{g-1}(0)|\mathbf{X}, Z, U = 1].$$

Proof. Notice that

$$\begin{aligned}
\mathbb{E}[Y_t(0) - Y_{g-1}(0)|\mathbf{X}, Z, G = g] &= \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|\mathbf{X}, Z, G = g] \\
&= \sum_{s=g}^{t} \mathbb{E}[Y_s(0) - Y_{s-1}(0)|\mathbf{X}, Z, U = 1] \\
&= \mathbb{E}[Y_t(0) - Y_{g-1}(0)|\mathbf{X}, Z, U = 1],
\end{aligned}$$

where the first equality holds by adding and subtracting $\mathbb{E}[Y_s(0)|\mathbf{X}, Z, G = g]$ for $s = g, \ldots, t-1$, the second equality holds by Assumption MP-PT, and the last equality holds by canceling all the terms involving $\mathbb{E}[Y_s(0)|\mathbf{X}, Z, U=1]$ for $s = g, \ldots, t-1$. $\square$

Proposition SA4. *Under Assumptions MP-PT and MP-1 to MP-4, for $t \geq g$,*

$$\text{ATT}_{g,t}(\mathbf{X}, Z) = \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, G=g] - \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, U=1]$$

and
$$\begin{aligned}\text{ATT}(g,t) &= \mathbb{E}[\text{ATT}_{g,t}(\mathbf{X}, Z)|G=g] \\ &= \mathbb{E}[Y_t - Y_{g-1}|G=g] - \mathbb{E}\Big[\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, U=1]\Big|G=g\Big].\end{aligned}$$

Proof. For any group $g \in \bar{\mathcal{G}}$ and $t \geq g$ (i.e., post-treatment periods for group g), we have that

$$\begin{aligned}\text{ATT}_{g,t}(\mathbf{X}, Z) &= \mathbb{E}[Y_t(g) - Y_t(0)|\mathbf{X}, Z, G=g] \\ &= \mathbb{E}[Y_t(g) - Y_{g-1}(0)|\mathbf{X}, Z, G=g] - \mathbb{E}[Y_t(0) - Y_{g-1}(0)|\mathbf{X}, Z, G=g] \\ &= \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, G=g] - \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, U=1],\end{aligned}$$

where the first equality holds by the definition of $\text{ATT}_{g,t}(\mathbf{X}, Z)$, the second equality holds by adding and subtracting $\mathbb{E}[Y_{g-1}(0)|\mathbf{X}, Z, G=g]$, and the third equality holds by Lemma SA3, and by writing potential outcomes in terms of their observed counterparts. This proves the first part of the proposition. The second part of the proposition holds immediately by applying the law of iterated expectations to the expression for $\text{ATT}_{g,t}(\mathbf{X}, Z)$. $\square$

Proposition SA4 shows that $\text{ATT}_{g,t}(\mathbf{X}, Z)$ and $\text{ATT}(g,t)$ are identified.[2] This is a generalization of Proposition SA1 from the setting with two periods to a setting with staggered treatment adoption. These expressions involve the trend in outcomes from period $(g-1)$, which is the period right before treatment starts for group g, to period t.

SA2.3 Supplementary Results for TWFE with Multiple Periods

This section collects additional results on decomposing and interpreting TWFE regressions with multiple periods and staggered treatment adoption. It also provides proofs of the results in Section 5 in the main text. An important first step for many of our results below is to rewrite α from Equation (1) as

$$\alpha = \frac{\frac{1}{\text{T}}\sum_{t=1}^{\text{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\ddot{Y}_t\Big]}{\frac{1}{\text{T}}\sum_{t=1}^{\text{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)^2\Big]}, \tag{SA7}$$

which holds using Frisch-Waugh-Lovell arguments. We start by providing a decomposition of α. The first result collects several useful properties of double-demeaned random variables.

Lemma SA4. *Let A_{it} denote a random variable that can vary across units and time periods, $\ddot{A}_{it}$ denote the double-demeaned version of A_{it}, B_i denote a random variable that does not vary over time, and ζ_t*

[2]As discussed in the main text, the same set of assumptions also rationalize using different comparison groups such as the not-yet-treated group (i.e., where the comparison group is defined by $D_t = 0$ rather than $U = 1$).

denote a non-random variable that can change values across time periods. The following properties of double-demeaned random variables hold:

$$\text{(1) } \mathbb{E}\Big[\ddot{A}_t\Big] = 0; \quad \text{(2) } \frac{1}{T}\sum_{t=1}^{T} \ddot{A}_{it} = 0; \quad \text{(3) } \mathbb{E}\Big[\ddot{A}_t \zeta_t\Big] = 0; \quad \text{(4) } \frac{1}{T}\sum_{t=1}^{T} \ddot{A}_{it} B_i = 0; \quad \text{(5) } \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[\ddot{A}_t B\Big] = 0.$$

The results in Lemma SA4 are well known, so we omit their proof.

Lemma SA5. *Under Assumptions MP-1, MP-3 and MP-4, the numerator in the expression for α in Equation (SA7) can be expressed as*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\ddot{Y}_t\Big] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma) Y_t\Big].$$

Proof. First, notice that

$$\begin{aligned} \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[\ddot{D}_t \ddot{Y}_t\Big] &= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t Y_t] - \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t \bar{Y}] - \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t \mathbb{E}[Y_t]] + \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t] \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[Y_t] \\ &= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t Y_t], \end{aligned} \tag{SA8}$$

where the first equality holds by the definition of $\ddot{Y}_t$, and the second equality holds by applying Lemma SA4.5, Lemma SA4.3, and Lemma SA4.1 to the second, third, and fourth terms in the previous line, respectively. Next, notice that

$$\begin{aligned} \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{X}_t'\Gamma)\ddot{Y}_t\Big] &= \Gamma' \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[\ddot{X}_t \ddot{Y}_t\Big] = \Gamma' \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[\ddot{X}_t Y_t\Big] \\ &= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{X}_t'\Gamma) Y_t\Big], \end{aligned} \tag{SA9}$$

where the first equality holds by rearranging terms, the second equality holds by the same arguments as for $\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}[\ddot{D}_t \ddot{Y}_t]$ above, and the last equality holds by rearranging terms again. The result holds by combining the expressions in Equations (SA8) and (SA9). □

Lemma SA6. *Under Assumptions MP-1, MP-3 and MP-4, the following result holds*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma) Y_{G-1}\Big] = 0,$$

where Y_{iG_i-1} is the outcome for unit i in the time period right before it becomes treated (for never-treated units, it is Y_{iT}, i.e., their outcome in the last period).

Proof. Notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[\ddot{D}_t Y_{G-1}\Big] = 0,$$

which holds by Lemma SA4.5 because Y_{G-1} is time-invariant. Next, notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{X}_t'\Gamma)Y_{G-1}\Big] = \Gamma'\mathbb{E}\left[\left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\ddot{X}_t\right)Y_{G-1}\right] = 0,$$

where the first equality holds by changing the order of the expectation and summation, and the second equality holds by Lemma SA4.2. Combining the two previous expressions above completes the proof. □

For the next result, we introduce some new notation. For a time-varying random variable A_{it}, define

$$\ddot{A}_{it}^{\dagger} := A_{it} - \bar{A}_i - \mathbb{E}[A_t|U=1] + \frac{1}{\mathrm{T}}\sum_{s=1}^{\mathrm{T}}\mathbb{E}[A_s|U=1],$$

which double demeans A_{it} with respect to the untreated group. Now, consider the linear projection of $\ddot{Y}_{it}^{\dagger}$ on $\ddot{X}_{it}^{\dagger}$ using the untreated group. The linear projection coefficient is given by

$$\Lambda_0 := \left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[\ddot{X}_t^{\dagger}\ddot{X}_t^{\dagger'}\Big|U=1\Big]\right)^{-1}\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[\ddot{X}_t^{\dagger}\ddot{Y}_t^{\dagger}\Big|U=1\Big], \tag{SA10}$$

and additionally recall that, in the main text, we defined

$$\lambda_t := \mathbb{E}[Y_t - X_t'\Lambda_0|U=1] \quad \text{and} \quad \bar{\lambda} := \mathbb{E}[\bar{Y} - \bar{X}'\Lambda_0|U=1]. \tag{SA11}$$

Next, we define λ_{G_i-1} to be λ_{g-1} after setting $g = G_i$, unit i's actual group. It is useful below to express this as

$$\lambda_{G_i-1} := \sum_{s=1}^{\mathrm{T}}\lambda_s\mathbf{1}\{s = G_i - 1\}.$$

The terms above show up in the misspecification bias terms in the decomposition of α from the TWFE regression. The next result provides some properties of these terms that we use below.

Lemma SA7. *Under Assumptions MP-1, MP-3 and MP-4, the following results hold*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\lambda_t\Big] = 0; \tag{A}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\lambda_{G-1}\Big] = 0; \tag{B}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\ddot{D}_t - \ddot{X}_t'\Gamma\right)X_t'\Lambda_0\right] = 0; \tag{C}$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\ddot{D}_t - \ddot{X}_t'\Gamma\right)X_{G-1}'\Lambda_0\right] = 0. \tag{D}$$

Proof. For Part (A), notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\lambda_t\Big]=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\Big\{\mathbb{E}[\ddot{D}_t]\lambda_t-\Gamma'\mathbb{E}[\ddot{X}_t]\lambda_t\Big\}=0,$$

where the first equality holds by rearranging terms, and the second equality holds by Lemma SA4.1. For Part (B),

$$\begin{aligned}\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\lambda_{G-1}\Big]&=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[(\ddot{D}_t-\ddot{X}_t'\Gamma)\left(\sum_{s=1}^{\mathrm{T}}\lambda_s\mathbf{1}\{s=G\text{-}1\}\right)\right]\\&=\sum_{s=1}^{\mathrm{T}}\lambda_s\mathbb{E}\left[\mathbf{1}\{s=G\text{-}1\}\left\{\left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\ddot{D}_t\right)-\left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\ddot{X}_t\right)'\Gamma\right\}\right]\\&=0,\end{aligned}$$

where the first equality holds by the definition of λ_{G_i-1}, the second equality holds by re-arranging the sums and expectation, and the last equality holds by Lemma SA4.2. Next, for Part (C), notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\left(\ddot{D}_t-\ddot{X}_t'\Gamma\right)X_t'\Lambda_0\right]=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\left(\ddot{D}_t-\ddot{X}_t'\Gamma\right)\ddot{X}_t'\right]\Lambda_0=0,$$

where the first equality holds using the same sort of argument (in reverse) as in Lemma SA4, and the second equality holds because $(\ddot{D}_t-\ddot{X}_t'\Gamma)$ is the projection error from projecting $\ddot{D}_t$ on $\ddot{X}_t$, which is uncorrelated with $\ddot{X}_t$. Finally, for Part (D), notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\left(\ddot{D}_t-\ddot{X}_t'\Gamma\right)X_{G-1}'\Lambda_0\right]=\mathbb{E}\left[\left\{\left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\ddot{D}_t\right)-\left(\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\ddot{X}_t\right)'\Gamma\right\}X_{G-1}'\Lambda_0\right]=0,$$

where the first equality holds by swapping the order of the summation and expectation (and because $\ddot{D}_t$ and $\ddot{X}_t$ are the only two terms that depend on t), and the second equality holds by Lemma SA4.2. □

Lemma SA8. *Under Assumptions MP-1, MP-3 and MP-4, the denominator in the expression for* α *in Equation* (SA7) *can be expressed as*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)^2\Big]=\sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\mathbb{E}\left[\frac{(h(g,t)-\ddot{X}_t'\Gamma)\pi_g}{T}\middle|G=g\right].$$

Proof. Notice that

$$\begin{aligned}\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)^2\Big]&=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\ddot{D}_t\Big]\\&=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)D_t\Big]\\&=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\mathcal{G}}\mathbb{E}\Big[(h(g,t)-\ddot{X}_t'\Gamma)\mathbf{1}\{t\geq g\}\Big|G=g\Big]\pi_g\end{aligned}$$

$$= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{\mathrm{T}} \mathbb{E}\left[\frac{(h(g,t) - \ddot{X}_t'\Gamma)\pi_g}{\mathrm{T}} \middle| G = g \right],$$

where the first equality holds because $(\ddot{D}_t - \ddot{X}_t'\Gamma)$ is the projection error from projecting $\ddot{D}_t$ on $\ddot{X}_t$, and is therefore uncorrelated with $\ddot{X}_t$; the second equality holds by an analogous argument to the one in Lemma SA5; the third equality holds by the law of iterated expectations and the definition of $h(g,t)$, and by Assumption MP-1 (so that $D_t = \mathbf{1}\{t \geq G\}$); and the last equality holds by combining terms and discarding terms that are equal to 0 (also notice that there are no post-treatment periods for the never-treated group, which implies that we can sum across groups in $\bar{\mathcal{G}}$ rather than all groups in $\mathcal{G}$). $\square$

The next proposition delivers a useful decomposition for α in Equation (1) in the case with multiple periods and variation in treatment timing considered in Section 5.

Proposition SA5. *Under Assumptions MP-1, MP-3 and MP-4, α from the regression in Equation (1) can be expressed as*

$$\alpha = \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{T} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \Big| G = g \Big]$$
$$+ \sum_{g \in \bar{\mathcal{G}}} \sum_{t=1}^{g-1} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \Big| G = g \Big],$$

where the weights are the same as in Theorem 3 in the main text and satisfy the same properties.

Proof. First, we consider the numerator in the expression for α in Equation (SA7). Notice that

$$\frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\ddot{Y}_t \Big]$$
$$= \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)Y_t \Big]$$
$$= \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)(Y_t - Y_{G-1}) \Big]$$
$$= \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma) \Big\{ (Y_t - Y_{G-1}) - \Big(\lambda_t - \lambda_{G-1} + (X_t - X_{G-1})'\Lambda_0 \Big) \Big\} \Big]$$
$$= \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \sum_{g \in \mathcal{G}} \mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma) \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \Big| G = g \Big] \pi_g$$
$$= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\left[\frac{(h(g,t) - \ddot{X}_t'\Gamma)\pi_g}{\mathrm{T}} \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \middle| G = g \right]$$
$$= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{\mathrm{T}} \mathbb{E}\left[\frac{(h(g,t) - \ddot{X}_t'\Gamma)\pi_g}{\mathrm{T}} \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \middle| G = g \right]$$
$$+ \sum_{g \in \bar{\mathcal{G}}} \sum_{t=1}^{g-1} \mathbb{E}\left[\frac{(h(g,t) - \ddot{X}_t'\Gamma)\pi_g}{\mathrm{T}} \Big\{ (Y_t - Y_{g-1}) - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0 \Big) \Big\} \middle| G = g \right],$$

where the first equality holds by Lemma SA5, the second equality holds by Lemma SA6, the third equality holds by Lemma SA7, the fourth equality holds by the law of iterated expectations and by the definition of $h(g,t)$, the fifth equality holds by rearranging terms and from Lemma SA11 below (which shows that the sum across time periods of the conditional expectations for the never-treated group is equal to zero), and the last equality holds by splitting the summation into pre- and post-treatment periods. Then, the result holds by combining the last expression above with the expression for the denominator in Equation (SA7) from Lemma SA8 and by the definition of $w_{g,t}^{twfe}(\ddot{X}_t)$. □

Proposition SA6. *Under Assumptions MP-1, MP-3 and MP-4, α from the regression in Equation (1) can be expressed as*

$$\alpha = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big\{\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X},Z,G=g] - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0\Big)\Big\}\Big|G=g\Big]$$
$$+ \sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big\{\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X},Z,G=g] - \Big(\lambda_t - \lambda_{g-1} + (X_t - X_{g-1})'\Lambda_0\Big)\Big\}\Big|G=g\Big],$$

where $w_{g,t}^{twfe}(\ddot{X}_t)$ are defined in Theorem 3 in the main text and satisfy the same properties in post-treatment periods.

Proof. The result holds immediately by applying the law of iterated expectations to both terms in the expression for α from Proposition SA5. □

Proposition SA7. *Under Assumptions MP-1, MP-3 and MP-4,*

$$\alpha = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big(\mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,G{=}g] - \mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,U{=}1]\Big)\Big|G=g\Big] \tag{A}$$
$$+ \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big\{\mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,U{=}1] - \Big((\lambda_t - \lambda_{g-1}) + (X_t\text{-}X_{g-1})'\Lambda_0\Big)\Big\}\Big|G=g\Big] \tag{B}$$
$$+ \sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big(\mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,G{=}g] - \mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,U{=}1]\Big)\Big|G=g\Big] \tag{C}$$
$$+ \sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big\{\mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,U{=}1] - \Big((\lambda_t - \lambda_{g-1}) + (X_t\text{-}X_{g-1})'\Lambda_0\Big)\Big\}\Big|G=g\Big], \tag{D}$$

where the weights are the same as in Theorem 3 and Proposition SA6, and satisfy the same properties.

Proof. Starting from the expression for α in Proposition SA6, the first part of the result holds by adding and subtracting

$$\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X},Z,U=1]\Big|G=g\Big].$$

Next, we prove the properties of the weights. By the definition of the weights, it follows immediately that they sum to one across post-treatment periods. We show that the weights sum to negative one across pre-treatment periods in Lemma SA12 below. That the weights can be negative holds because these are linear projection-type weights. To give a concrete example, suppose that $\Gamma = 0$. Then, the weights are the same as in de Chaisemartin and D'Haultfoeuille (2020), which can be negative. □

Proposition SA7 is an important intermediate result for understanding α from a TWFE regression with multiple periods and staggered treatment adoption. It provides a decomposition of α from Equation (1) in that setting. It is a decomposition in the sense that it only relies on regularity assumptions and does not invoke identification assumptions such as parallel trends or no-anticipation. Notice that Terms (A)-(D) differ along two dimensions. First, Terms (A) and (B) involve post-treatment periods only, while Terms (C) and (D) involve only pre-treatment periods. Second, Terms (A) and (C) involve differences between conditional expectations of paths of outcomes between group g and the never-treated group. Once we invoke parallel trends and no-anticipation, the underlying expressions in Term (A) will become group-time average treatment effects, while the underlying expressions in Term (C) will be equal to zero (more details below). Term (C) involves violations of parallel trends in pre-treatment periods, which, as can be seen from the proposition, will impact the estimate of α. Terms (B) and (D) are post-treatment and pre-treatment misspecification bias terms, respectively. These terms include hidden linearity bias terms similar to the ones we emphasized in the two-period case. We consider these in substantially more detail below. That the weights on Term (B) sum to one (as opposed to, say, zero) indicates that the importance/magnitude of misspecification bias is on par with the magnitude of the treatment effects themselves. That the weights on Terms (C) and (D) are negative arises because these are pre-treatment periods and, hence, "comparison periods". That these weights sum to negative one indicates that pre-treatment violations of parallel trends and pre-treatment misspecification bias are as important for the resulting estimate of α as the treatment effects themselves.

Proof of Theorem 3. For $g \in \bar{\mathcal{G}}$ and $t < g$ (i.e., pre-treatment periods for group g),

$$\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, G = g] - \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, U = 1] = 0$$

under Assumption MP-PT. For $g \in \bar{\mathcal{G}}$ and $t \geq g$ (i.e., post-treatment periods for group g), we have that

$$\mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, G = g] - \mathbb{E}[Y_t - Y_{g-1}|\mathbf{X}, Z, U = 1] = \text{ATT}_{g,t}(\mathbf{X}, Z),$$

which holds by Proposition SA4. Plugging these expressions into Proposition SA7 implies the result. □

Next, we state the formal versions of the conditions discussed in the main text to rule out the misspecification bias term in Theorem 3.

Assumption MP-5 (Additional Assumptions to Rule Out Bias Terms in TWFE Regression under Staggered Treatment Adoption)**.** *The following conditions hold for all time periods $t = 2, \dots, T$:*[3]

(1) $\mathbb{E}[\Delta Y_t(0)|\mathbf{X}, Z, U = 1] = \mathbb{E}[\Delta Y_t(0)|\mathbf{X}, U = 1]$;

(2) $\mathbb{E}[\Delta Y_t(0)|\mathbf{X}, U = 1] = \mathbb{E}[\Delta Y_t(0)|X_t, X_{t-1}, U = 1]$;

(3) $\mathbb{E}[\Delta Y_t(0)|X_t, X_{t-1}, U = 1] = \mathbb{E}[\Delta Y_t(0)|\Delta X_t, U = 1]$;

[3] The assumption needs slightly more notation for linear projections than was used in the main text. Here, $\lambda_{0,t,t-1}$ and $\Lambda_{0,t,t-1}$ denote the intercept and slope coefficients from the linear projection of ΔY_t on ΔX_t for the never-treated group.

(4) $\mathbb{E}[\Delta Y_t(0)|\Delta X_t, U=1] = \lambda_{0,t,t-1} + \Delta X_t'\Lambda_{0,t,t-1}$*;*

(5) $\Lambda_{0,t,t-1} = \Lambda_0$*.*

Lemma SA9. *Under Assumptions MP-PT and MP-1 to MP-5,*

$$\lambda_{0,t,t-1} = \lambda_t - \lambda_{t-1}.$$

Proof. Notice that

$$\begin{aligned}\lambda_{0,t,t-1} &= \mathbb{E}[Y_t - Y_{t-1}|U=1] - \mathbb{E}[(X_t - X_{t-1})|U=1]'\Lambda_{0,t,t-1} \\ &= \mathbb{E}[Y_t - X_t'\Lambda_0|U=1] - \mathbb{E}[Y_{t-1} - X_{t-1}'\Lambda_0|U=1] \\ &= \lambda_t - \lambda_{t-1},\end{aligned}$$

where the first equality holds by the definition of $\lambda_{0,t,t-1}$, the second equality holds by Assumption MP-5(5), and the last equality holds by the definition of λ_t in the main text. □

Theorem SA1. *Under Assumptions MP-PT and MP-1 to MP-5,*

$$\alpha = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\text{ATT}_{g,t}(\mathbf{X},Z)\Big|G=g\Big],$$

where the weights are the same as in Theorem 3 and Proposition SA6 and satisfy the same properties. If, in addition, $\text{ATT}_{g,t}(\mathbf{X},Z) = \text{ATT}(g,t)$*, then*

$$\alpha = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\Big\{\mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t)\Big|G=g\Big]\text{ATT}(g,t)\Big\}.$$

If, in addition, $\text{ATT}_{g,t}(\mathbf{X},Z) = \text{ATT}$*, then*

$$\alpha = \text{ATT}.$$

Proof of Theorem SA1. Notice that, from Theorem 3, the first result will hold if, for any $g \in \bar{\mathcal{G}}$ and for any $t \in \{1,\ldots,\text{T}\}$,

$$\xi_{t,g-1}(\mathbf{X},Z) := \mathbb{E}[Y_t\text{-}Y_{g-1}|\mathbf{X},Z,U{=}1] - \Big((\lambda_t - \lambda_{g-1}) + (X_t\text{-}X_{g-1})'\Lambda_0\Big) = 0.$$

Consider the case where $t \geq g$, noting that $\xi_{g-1,g-1}(\mathbf{X},Z) = 0$ by construction, and the same sort of argument as we consider here can be used for the case where $t < g$. Then, we have that

$$\begin{aligned}\xi_{t,g-1}(\mathbf{X},Z) &= \sum_{s=g}^{t}\Big\{\mathbb{E}[\Delta Y_s|\mathbf{X},Z,U=1] - \Big(\Delta\lambda_s + \Delta X_s'\Lambda_0\Big)\Big\} \\ &= \sum_{s=g}^{t}\Big\{\lambda_{0,s,s-1} + \Delta X_s'\Lambda_{0,s,s-1} - \Big(\Delta\lambda_s + \Delta X_s'\Lambda_0\Big)\Big\} = 0,\end{aligned}$$

where the first equality holds by adding and subtracting $\mathbb{E}[Y_s|\mathbf{X},Z,U=1] - (\lambda_s + X_s'\Lambda_0)$ for all values of

$s = g, \ldots, t-1$, the second equality holds from applying Assumption MP-5(1)-(4), and the last equality holds by Assumption MP-5(5) and by Lemma SA9.

The second part of the result, i.e., under the additional condition that $\text{ATT}_{g,t}(\mathbf{X}, Z) = \text{ATT}(g,t)$, follows immediately by taking $\text{ATT}(g,t)$ outside of the expectation in the expression in the first result.

The third part of the result, i.e., under the additional condition that $\text{ATT}_{g,t}(\mathbf{X}, Z) = \text{ATT}$, holds by noticing that, in this case,

$$\alpha = \text{ATT} \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{\mathcal{T}} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] = \text{ATT},$$

where the second equality holds by property (i) of the weights from Proposition SA7. □

This result is the multiple period version of Theorem 2 from the earlier case with two time periods. The first part shows that, when we add the five conditions discussed above, α from the TWFE regression can be interpreted as a weighted average of conditional average treatment effects. However, even if these conditions hold, the weights on $\text{ATT}_{g,t}(\mathbf{X}, Z)$s are non-transparent and inherited from the estimation method, and it is possible for them to be negative (see de Chaisemartin and D'Haultfoeuille (2020)). The second and third parts layer on additional restrictions on treatment effect heterogeneity. For the second part, even if $\text{ATT}_{g,t}(\mathbf{X}, Z)$ does not vary across covariates (a strong additional condition), one will still recover weighted averages of $\text{ATT}(g,t)$, the weights will still be difficult to interpret, and the weights can still be negative. Finally, if we fully rule out any forms of systematic treatment effect heterogeneity, then α will be equal to the ATT.

Comparison of Multi-Period TWFE Decomposition to Other Papers

The results in Theorems SA1 and 3 are related to several other results in the literature. Both de Chaisemartin and D'Haultfoeuille (2020) and Goodman-Bacon (2021) include some results for TWFE regressions that include time-varying covariates, although they mainly consider interpreting TWFE regressions with multiple periods and variation in treatment timing in a setting without covariates in the parallel trends assumption. Some of our results, particularly the first part of Theorem SA1, are closely related to Theorem S4 in Online Appendix 3.3 of de Chaisemartin and D'Haultfoeuille (2020). In that theorem, de Chaisemartin and D'Haultfoeuille (2020) essentially take as a starting point the combination of our Assumptions MP-PT and MP-5 and show that, under a conditional parallel trends assumption that involves only changes in observed covariates and linearity assumptions, their main results related to multiple periods and variation in treatment timing essentially continue to apply. Our weights in Theorem SA1 are the same as in that paper, though we expand it in important ways by allowing for levels of time-varying covariates and time-invariant covariates to be in the parallel trends assumption and by providing conditions under which the expression for α can be simplified. Our results in Theorem 3, which provide possible sources of misspecification bias, are also new to the literature. Goodman-Bacon (2021, Section 5.2) provides a different type of decomposition where α is decomposed into a "within" component and "between" component. The between-component arises due to variation in treatment timing and can be expressed as an adjusted-by-covariates 2x2 difference-in-differences comparison. The within component comes from variation in the covariates within a particular group and is, therefore, related to our expression for α in the setting with only two time periods. Relative to Goodman-Bacon (2021), we further decompose this type of term into several more primitive objects that highlight that researchers should be careful in interpreting "within" components as averages of causal effects unless they are willing to invoke extra assumptions. In work first made publicly available after the first version of our paper, Lin and Zhang (2022) build on some of our results and the event study decomposition in Sun and Abraham (2021) and show that an additional bias term can arise in event study regressions that include time-varying covariates. Finally, Ishimaru (2022, Section 2.2), like de Chaisemartin and D'Haultfoeuille (2020), provides conditions under which TWFE regressions that include covariates can be interpreted as weighted averages of underlying treatment effect parameters. These include a version of conditional parallel trends that holds conditional on the change in covariates over time, an assumption on the linearity of the propensity score conditional on changes in covariates over time, and a decomposition of a modified TWFE regression that additionally includes time-varying coefficients on time-varying covariates. Also related is Hazlett and Xu (2018), who use a re-weighting approach in panel data settings, though their focus is on balancing pre-treatment outcome trajectories rather than covariates.

SA2.4 Covariate Balance Diagnostics with Multiple Periods

Next, we discuss how to extend our TWFE and AIPW diagnostics in Section 4 to settings with multiple periods and variation in treatment timing. Similar to the case with two periods, the goal in this section is to show that (i) α from the TWFE regression and $\widetilde{\text{ATT}}^{aipw,o}$ from our AIPW estimator can be recast as weighting estimators, and then (ii) the implicit weights can be applied to levels of the time-varying covariates and the time-invariant covariates in order to understand how well each of these balances covariates for treated groups relative to the never-treated group. As above, this provides a way to assess the sensitivity of the TWFE regression to hidden linearity bias.

Toward this end (and like for the case with two periods), notice that if we could find balancing weights that, for a particular group g in time period t, balance the distribution of the time-varying and time-invariant covariates for the never-treated group relative to group g, then we could recover $\text{ATT}(g,t)$ by applying these weights to the path of outcomes for the never-treated group. In particular, for some group $g \in \bar{\mathcal{G}}$ and post-treatment time period $t \geq g$, let $\vartheta_{g,t}(\mathbf{X}, Z)$ denote balancing weights that re-weight the never-treated group so that, after applying the weights, it has the same distribution of $(\mathbf{X}, Z)$ as group g. Given these balancing weights and using very similar arguments as in Section 4, one can show that

$$\text{ATT}(g,t) = \mathbb{E}[Y_t - Y_{g-1}|G=g] - \mathbb{E}[\vartheta_{g,t}(\mathbf{X}, Z)(Y_t - Y_{g-1})|U=1] \tag{SA12}$$

and

$$\text{ATT}^o = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{\mathcal{T}} \left\{\mathbb{E}\left[Y_t - Y_{g-1}\middle|G=g\right] - \mathbb{E}\left[\vartheta_{g,t}(\mathbf{X}, Z)(Y_t - Y_{g-1})\middle|U=1\right]\right\} w^o(g,t). \tag{SA13}$$

Equations (SA12) and (SA13) show that group-time average treatment effects and ATT^o can, under conditional parallel trends, be recovered by re-weighting the never-treated group to have the same distribution of $\mathbf{X}$ and Z as each group.

SA2.4.1 Results on Implicit TWFE Weights with Multiple Periods

In this section, we prove the claim in Equation (11) in the main text about implicit TWFE weights with multiple periods. Towards this end, we first provide a supporting lemma, then provide a similar result where the first period is used as the base period, and finally we prove the claim from the main text.

Lemma SA10. *Under Assumptions MP-1, MP-3 and MP-4,*

$$\mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma)\Big|U=1\Big]\pi_0 = -\sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma)\Big|G=g\Big]\pi_g.$$

Proof. Notice that

$$\begin{aligned}
0 &= \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\Big] \\
&= \sum_{g\in\mathcal{G}}\mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma)\Big|G=g\Big]\pi_g \\
&= \sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma)\Big|G=g\Big]\pi_g + \mathbb{E}\Big[(h(g,t) - \ddot{X}_t'\Gamma)\Big|U=1\Big]\pi_0,
\end{aligned}$$

where the first equality holds by Lemma SA4.1, the second equality holds by the law of iterated expectations, and the third equality holds by pulling the untreated group out of the summation. Then, the result holds by rearranging terms. □

Proposition SA8. *Under Assumptions MP-1, MP-3 and MP-4,*

$$\alpha = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{T}\bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t - Y_{i1})\Big|G=g\right] - \mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t - Y_{i1})\Big|U=1\right]\right\}.$$

Proof. Starting with the numerator of α in Equation (SA7), we have that

$$\begin{aligned}
&\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\ddot{Y}_t\Big] \\
&\quad= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)Y_t\Big] \\
&\quad= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)(Y_t - Y_{i1})\Big] \\
&\quad= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\mathcal{G}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)(Y_t - Y_{i1})\Big|G=g\Big]\pi_g \\
&\quad= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)(Y_t - Y_{i1})\Big|G=g\Big]\pi_g + \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)(Y_t - Y_{i1})\Big|U=1\Big]\pi_0 \\
&\quad= \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\left[\frac{(\ddot{D}_t - \ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t - \ddot{X}_t'\Gamma)|G=g]}(Y_t - Y_{i1})\Big|G=g\right]\mathbb{E}[(\ddot{D}_t - \ddot{X}_t'\Gamma)|G=g]\pi_g
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\frac{(\ddot{D}_t-\ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|U=1]}(Y_t-Y_{i1})\Big|U=1\right]\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|U=1]\pi_0\\
&=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\left[\frac{(\ddot{D}_t-\ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]}(Y_t-Y_{i1})\Big|G=g\right]\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]\pi_g\\
&\quad-\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\sum_{g\in\bar{\mathcal{G}}}\mathbb{E}\left[\frac{(\ddot{D}_t-\ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|U=1]}(Y_t-Y_{i1})\Big|U=1\right]\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]\pi_g\\
&=\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]\frac{\pi_g}{\mathrm{T}}\left\{\mathbb{E}\left[\frac{(\ddot{D}_t-\ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]}(Y_t-Y_{i1})\Big|G=g\right]\right.\\
&\qquad\qquad\left.-\mathbb{E}\left[\frac{(\ddot{D}_t-\ddot{X}_t'\Gamma)}{\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|U=1]}(Y_t-Y_{i1})\Big|U=1\right]\right\}\\
&=\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]\frac{\pi_g}{\mathrm{T}}\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{i1})\Big|G=g\right]\right.\\
&\qquad\qquad\left.-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{i1})\Big|U=1\right]\right\},
\end{aligned}
$$

where the first equality holds by Lemma SA5, the second equality holds as an implication of Lemma SA4, the third equality holds by the law of iterated expectations, the fourth equality holds by separating the never-treated group from the other groups, the fifth equality holds by multiplying and dividing by $\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|G=g]$ and by $\mathbb{E}[(\ddot{D}_t-\ddot{X}_t'\Gamma)|U=1]$, the sixth equality holds by Lemma SA10, the seventh equality holds by combining the summations and rearranging terms, and the last equality holds by the definition of $w_{g,t}^{1,twfe}$ and $w_{g,t}^{0,twfe}$. Then, the main claim of the proposition holds by dividing the previous expression by the denominator of α in Equation (SA7) and from the definition of $\bar{w}^{twfe}(g,t)$. □

Proposition SA9. *Under Assumptions MP-1, MP-3 and MP-4,*

$$
\begin{aligned}
\alpha=&\sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}+\\
&+\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1}\bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}+r,
\end{aligned}
$$

where

$$
r=-\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{T}\bar{w}^{twfe}(g,t)\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_{g-1}-Y_{i1})\Big|U=1\right].
$$

In addition,

$$
\mathbb{E}[w_{g,t}^{1,twfe}(\mathbf{X},Z)|G=g]=\mathbb{E}[w_{g,t}^{0,twfe}(\mathbf{X},Z)|U=1]=1.
$$

Proof. Starting from the result in Proposition SA8, we have that

$$\begin{aligned}
\alpha &= \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}\\
&+\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}\\
&+\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_{g-1}-Y_{i1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_{g-1}-Y_{i1})\Big|U=1\right]\right\}\\
&=\sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}\\
&+\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{g-1} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}+r,
\end{aligned}$$

where the first equality holds by adding and subtracting

$$\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\left\{\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)Y_{g-1}\Big|G=g\right]-\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)Y_{g-1}\Big|U=1\right]\right\}$$

to the result of Proposition SA8 and rearranging terms, and the second equality holds because $\displaystyle\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_{g-1}-Y_{i1})\Big|G=g\right]=0$[4] and by the definition of the remainder term r. That $w_{g,t}^{1,twfe}(\mathbf{X},Z)$ and $w_{g,t}^{0,twfe}(\mathbf{X},Z)$ have mean one follows immediately from their definitions. □

Remark SA11 (Comments about the remainder term)**.** We note here that having a remainder in the expression for α in Proposition SA9 is undesirable. As discussed in the main text, it is a byproduct of using $g-1$ as the base period. Notice that there is no remainder term when one uses the first period as the base period, as in Proposition SA8. In our application, when we compute these remainder terms across different specifications, they are uniformly negligible. We conjecture that the remainder will likely be small in most applications for four reasons. First, the weights sum to zero rather than one, that is, $\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{\mathrm{T}} \bar{w}^{twfe}(g,t)\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)\Big|U=1\right]=0$. Second, this term equals zero if the distribution of $(\mathbf{X},Z)$ is the same for all groups. Third, this term equals zero if $\mathbb{E}[Y_t|\mathbf{X},Z,U=1]$ is constant across time. Fourth, this term equals zero if, for $t=2,\dots,\mathrm{T}$, $\mathbb{E}[\Delta Y_t|\mathbf{X},Z,U=1]=\mathbb{E}[\Delta Y_t|U=1]$. While none of the second, third, or fourth conditions are necessarily likely to hold exactly in particular applications, the remainder term will be small when these terms are small. Taken together, all four of these reasons suggest that r should be small, often very small, in most applications.

Remark SA12 (Related strategies in empirical work)**.** The ideas presented in this section are broadly similar to the idea of using covariates as outcomes to assess balance, which is relatively common in empirical work in economics (see Pei et al. (2019) for a detailed discussion of this strategy). This

[4]To see this, notice that the numerator of $\bar{w}^{twfe}(g,t)$ cancels with the denominator of $w_{g,t}^{1,twfe}(\mathbf{X},Z)$, and then the only time-varying terms remaining in the numerator of this expression are $\ddot{D}_t$ and $\ddot{X}_t$, which sum to zero by Lemma SA4.2.

approach is not feasible for assessing balance with respect to covariates that do not vary over time or for time-varying covariates that are included in the TWFE regression. Alternatively, some papers check for balance in terms of pre-treatment characteristics (see, for example, Goodman-Bacon and Cunningham (2019, Table 3)). The working paper version of Goodman-Bacon (2021) discusses comparing the averages of time-varying covariates (including levels) for early, late, and never-treated groups (see also Almond et al. (2011) and Bailey and Goodman-Bacon (2015) as examples of empirical work using this sort of strategy). Relative to these strategies, a main advantage of the weighting strategy discussed in this section is that one can directly use the implicit regression weights from a main TWFE specification used in a particular application and assess balance for functions of covariates that are included in the model.

SA2.4.2 Results on Implicit AIPW Weights with Multiple Periods

In this section, we prove the claim in Equation (12) in the main text about implicit AIPW weights with multiple periods.

Proposition SA10. *Under Assumptions MP-1, MP-3 and MP-4,*

$$\widetilde{\mathrm{ATT}}^{aipw,o} = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T} w^o(g,t)\left\{\mathbb{E}\left[\vartheta^{1,aipw}_{g,t}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right] - \mathbb{E}\left[\vartheta^{0,aipw}_{g,t}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right]\right\}.$$

Proof. Recall from Equation (9) in the main text that

$$\begin{aligned}\widetilde{\mathrm{ATT}}^{aipw}(g,t) &= \mathbb{E}\left[(Y_t\text{-}Y_{g-1})-\tilde{\mathrm{L}}^0_{g,t}(Y_t\text{-}Y_{g-1}|\mathbf{X},Z)\Big|G=g\right]\\ &\quad-\mathbb{E}\left[\tilde{w}^{0,aipw}_{g,t}(\mathbf{X},Z)\big((Y_t\text{-}Y_{g-1})-\tilde{\mathrm{L}}^0_{g,t}(Y_t\text{-}Y_{g-1}|\mathbf{X},Z)\big)\Big|U=1\right].\end{aligned}$$

Considering the subgroup such that $\mathbf{1}\{G=g\}+U=1$ (that is, either group g or the never-treated group) and using the same argument as in Proposition 1 from the case with two periods and two groups (up to differences about the base period and that we use covariates across all time periods rather than just two periods), it follows that

$$\widetilde{\mathrm{ATT}}^{aipw}(g,t) = \mathbb{E}\left[\vartheta^{1,aipw}_{g,t}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|G=g\right] - \mathbb{E}\left[\vartheta^{0,aipw}_{g,t}(\mathbf{X},Z)(Y_t-Y_{g-1})\Big|U=1\right] \qquad \text{(SA14)}$$

for any $g\in\bar{\mathcal{G}}$ and $t\geq g$. Next, recall that, by Equation (10) in the main text, we have that

$$\widetilde{\mathrm{ATT}}^{aipw,o} = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{\mathrm{T}}\widetilde{\mathrm{ATT}}^{aipw}(g,t)w^o(g,t). \qquad \text{(SA15)}$$

Plugging the expression for $\widetilde{\mathrm{ATT}}^{aipw}(g,t)$ from Equation (SA14) into Equation (SA15) completes the proof. ☐

SA2.5 Additional Supporting Results

This section contains supporting results for proving the main results.

Lemma SA11. *Under Assumptions MP-1, MP-3 and MP-4,*

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\Big\{(Y_t-Y_T)-\Big(\lambda_t-\lambda_T+(X_t-X_T)'\Lambda_0\Big)\Big\}\Big|U=1\Big]=0.$$

Proof. To start with, notice that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\Big\{(Y_t-Y_{\mathrm{T}})-\Big(\lambda_t-\lambda_{\mathrm{T}}+(X_t-X_{\mathrm{T}})'\Lambda_0\Big)\Big\}\Big|U=1\Big]$$
$$=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[(\ddot{D}_t-\ddot{X}_t'\Gamma)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\Big], \quad \text{(SA16)}$$

which holds from the properties of double-demeaned random variables in Lemma SA4. Notice that, given the definitions of λ_t, $\bar{\lambda}$, and Λ_0, $\Big((Y_t-\bar{Y})-((\lambda_t-\bar{\lambda})+(X_t-\bar{X})'\Lambda_0)\Big)$ is the projection error from projecting $(Y_t-\bar{Y})$ on $(X_t-\bar{X})$ and time fixed effects.

Next, notice that, for some non-random time-varying variable ζ_t, we have that

$$\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[\zeta_t\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\Big]$$
$$=\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\zeta_t\Big\{\mathbb{E}\Big[(Y_t-\bar{Y})-(X_t-\bar{X})'\Lambda_0\Big|U=1\Big]-(\lambda_t-\bar{\lambda})\Big\}=0, \quad \text{(SA17)}$$

where the first equality holds by rearranging terms, and the second equality holds by the definitions of λ_t and $\bar{\lambda}$. This shows that the mean of the projection error multiplied by any time-varying, non-random variable is equal to zero. We use this result below.

Recall that $\ddot{D}_{it}=D_{it}-\bar{D}_i-\mathbb{E}[D_t]+\frac{1}{\mathrm{T}}\sum_{s=1}^{\mathrm{T}}\mathbb{E}[D_s]$. For units in the never-treated group, we have that $\ddot{D}_{it}=-\mathbb{E}[D_t]+\frac{1}{\mathrm{T}}\sum_{s=1}^{\mathrm{T}}\mathbb{E}[D_s]$, which holds because D_{it} and $\bar{D}_i$ both are equal to zero for units in this group. Notice that this term is time-varying but non-random.

Then, we can decompose the expression in Equation (SA16) as

$$(SA16)=-\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\left(\mathbb{E}[D_t]-\frac{1}{\mathrm{T}}\sum_{s=1}^{\mathrm{T}}\mathbb{E}[D_s]\right)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\middle|U=1\right] \quad \text{(SA18)}$$
$$-\Gamma'\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\Big[\Big(X_t-\bar{X}\Big)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\Big] \quad \text{(SA19)}$$
$$+\Gamma'\frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}\mathbb{E}\left[\left(\mathbb{E}[X_t]-\frac{1}{\mathrm{T}}\sum_{s=1}^{\mathrm{T}}\mathbb{E}[X_s]\right)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\middle|U=1\right] \quad \text{(SA20)}$$
$$=0,$$

where the result holds because (i) Equations (SA18) and (SA20) involve means of time-varying, non-random variables multiplied by the projection error discussed above, which are equal to zero from the argument in Equation (SA17), and (ii) Equation (SA19) is equal to zero because the $(X_t-\bar{X})$ term is orthogonal to the projection error term, $\Big((Y_t - \bar{Y}) - ((\lambda_t - \bar{\lambda}) + (X_t - \bar{X})'\Lambda_0)\Big)$. This completes the proof. □

Lemma SA12. *Under Assumptions MP-1, MP-3 and MP-4, the weights in Proposition SA7 sum to negative one across pre-treatment periods. That is,*

$$\sum_{g \in \mathcal{G}} \sum_{t=1}^{g-1} \mathbb{E}\left[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \right] = -1.$$

Proof. Notice that

$$\begin{aligned}
0 = \frac{\frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\Big]}{\frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)^2\Big]} &= \sum_{g \in \mathcal{G}} \sum_{t=1}^{\mathrm{T}} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] \\
&= \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{\mathrm{T}} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] + \sum_{g \in \mathcal{G}} \sum_{t=1}^{g-1} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] \\
\implies \sum_{g \in \mathcal{G}} \sum_{t=1}^{g-1} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] &= - \sum_{g \in \bar{\mathcal{G}}} \sum_{t=g}^{\mathrm{T}} \mathbb{E}\Big[w_{g,t}^{twfe}(\ddot{X}_t) \Big| G = g \Big] = -1,
\end{aligned}$$

where the first term holds immediately by applying the results in Lemma SA4 to the numerator, the second equality holds by the law of iterated expectations and the definition of $w_{g,t}^{twfe}(\ddot{X}_t)$, and the third equality holds by splitting the summation (and for the first term, noticing that there are no post-treatment periods for the untreated group, so that the summation can be over $\bar{\mathcal{G}}$ rather than $\mathcal{G}$). The last line holds because the sum of the post-treatment weights equals one. □

Remark SA13 (Clarification on groups included in summation)**.** One subtle point worth mentioning is that the pre-treatment sum in the decomposition of α in Proposition SA7 excludes the never-treated group, while the sum of the pre-treatment weights includes the never-treated group. For the pre-treatment weights to sum to negative one, they must include the never-treated group. This difference can be explained in the following way: if one includes the never-treated group in the decomposition of α (i.e., where the pre-treatment sum is over $\mathcal{G}$ rather than $\bar{\mathcal{G}}$), the extra term that this introduces is equal to zero (see Lemma SA11 as well as the proof of Proposition SA5). Therefore, α effectively includes positive weight on a term equal to zero by construction.

SA2.6 Formalizing no "bad controls"

In the main text, we noted that we implicitly ruled out "bad controls", which means that we rule out that the time-varying covariates are affected by the treatment. To formalize this, consider a setting with T time periods, a single treated group, denoted by $D = 1$, an untreated group, denoted by $D = 0$, and suppose that the treated group becomes treated in period t^*. Let $X_t(0)$ and $X_t(1)$ denote treated and untreated "potential covariates", i.e., the covariates that would realize in period t if a unit were untreated or treated. We also maintain no anticipation for the covariates. This notation explicitly allows for the possibility that the treatment could affect the covariates. Given this notation, we can write a version of the parallel trends assumption in Assumption MP-PT as:

$$\mathbb{E}[\Delta Y_t(0) \mid X_1(0), \ldots, X_{\mathrm{T}}(0), Z, D = 1] = \mathbb{E}[\Delta Y_t(0) \mid X_1(0), \ldots, X_{\mathrm{T}}(0), Z, D = 0], \tag{SA21}$$

which says that, absent participating in the treatment, the trend in untreated potential outcomes would be the same for the treated group as for the untreated group after conditioning on covariates in all periods. In addition, we need one more assumption that rules out feedback from the treatment to the covariates:

$$\big(X_{t^*}(1), \ldots, X_{\mathrm{T}}(1)\big) \mid X_1, \ldots X_{t^*-1}, Z, D = 1 \sim \big(X_{t^*}(0), \ldots, X_{\mathrm{T}}(0)\big) \mid X_1, \ldots X_{t^*-1}, Z, D = 1, \tag{SA22}$$

i.e., that the distribution of treated potential covariates in post-treatment periods is the same as the distribution of untreated potential covariates in the same periods for the treated group. This is an assumption that rules out X_t being a bad control. Caetano et al. (2022) call this assumption *covariate exogeneity*. A simple (though stronger) sufficient condition for covariate exogeneity is that $X_t(0) = X_t(1)$, which says that participating in the treatment does not affect the covariates at all. Under these assumptions, it is straightforward to show that, for $t \geq t^*$,

$$\mathrm{ATT}_t = \mathbb{E}[Y_t - Y_{t^*-1}|D = 1] - \mathbb{E}\big[\mathbb{E}[Y_t - Y_{t^*-1}|X_1, \ldots, X_{\mathrm{T}}, Z, D = 0]\big|D = 1\big],$$

which corresponds to our identification result in Proposition SA4 (specialized to the case with a single treated group).

Given that most empirical researchers are very careful when it comes to "bad controls", our sense is that they typically have this sort of covariate exogeneity condition in mind when they use TWFE regressions that include time-varying covariates. Covariate exogeneity rules out feedback between the treatment and the covariates, but it is also worth pointing out that this condition also (effectively) rules out feedback from the outcomes to the covariates as well, e.g., this condition would be violated if the treatment affects the outcome, which then affects the covariate.

SA2.7 Alternative versions of parallel trends with multiple periods

In this section, we consider alternative versions of Assumption MP-PT. For clarity, we continue to focus on the same setting as in Section SA2.6, where there are T time periods and a single treated group that becomes treated in period t^*. In this context, Assumption MP-PT becomes

$$\mathbb{E}[\Delta Y_t(0)|X_1, \ldots, X_{\mathrm{T}}, Z, D = 1] = \mathbb{E}[\Delta Y_t(0)|X_1, \ldots, X_{\mathrm{T}}, Z, D = 0] \tag{SA23}$$

for all $t = 2, \ldots, \mathrm{T}$. We take this as the natural version of conditional parallel trends with multiple periods for several reasons. First, there is a strong argument that this is the version of parallel trends that most empirical work uses, as TWFE regressions typically invoke strict exogeneity assumptions that involve conditioning on covariates in all periods. Second, relative to other versions of conditional parallel trends that condition on time-varying covariates in fewer periods, it is a favorable assumption for the TWFE regression. For example, Proposition SA6 (which is an important intermediate decomposition of the TWFE regression that our main results depend on) does not hold under versions of parallel trends that do not include the full covariate history. In contrast, one can adapt our regression adjustment and/or AIPW approaches under alternative versions of parallel trends that we consider below. Probably the leading alternative version of conditional parallel trends is to assume:

$$\mathbb{E}[\Delta Y_t(0)|X_t, X_{t-1}, Z, D=1] = \mathbb{E}[\Delta Y_t(0)|X_t, X_{t-1}, Z, D=0], \tag{SA24}$$

which says that the trend in untreated potential outcomes is the same for the treated and untreated group, conditional on time-invariant covariates and time-varying covariates only in periods t and $t-1$. This assumption is non-nested with the one in Equation (SA23), though the empirical content of these assumptions will be very similar in most applications. As noted above, this assumption does not work for many arguments related to TWFE, as they often rely on conditioning on the entire covariate history. It also creates some complications relative to the identification results that we emphasized in the main text. Notice that Lemma SA3 does not hold in this case. To see the implications of this, consider identifying ATT_{t^*+1} under this assumption:

$$\begin{aligned}\text{ATT}_{t^*+1} &= \mathbb{E}[Y_{t^*+1} - Y_{t^*-1}|D=1] - \mathbb{E}[\Delta Y_{t^*+1}(0)|D=1] - \mathbb{E}[\Delta Y_{t^*}(0)|D=1] \\ &= \mathbb{E}[Y_{t^*+1} - Y_{t^*-1}|D=1] - \mathbb{E}\big[\mathbb{E}[\Delta Y_{t^*+1}|X_{t^*+1}, X_{t^*}, Z, D=0]\big|D=1\big] \\ &\qquad - \mathbb{E}\big[\mathbb{E}[\Delta Y_{t^*}|X_{t^*}, X_{t^*-1}, Z, D=0]\big|D=1\big].\end{aligned}$$

This term is identified, but the terms in the counterfactual trend of untreated potential outcomes do not combine, leading to substantially more terms to deal with.

An interesting intermediate case arises from making the assumption in Equation (SA23) and then making the dimension reduction assumption that

$$\mathbb{E}[\Delta Y_t(0)|\mathbf{X}, Z, D=d] = \mathbb{E}[\Delta Y_t(0)|X_t, X_{t-1}, Z, D=d]. \tag{SA25}$$

This is not a parallel trends assumption. Rather, it says that trends in untreated potential outcomes do not depend on time-varying covariates in other periods after conditioning on the covariates in the current periods. Assumptions like this are in line with our discussion about dimension reduction in the main text.

One more alternative parallel trends assumption that is worth considering is the following:

$$\mathbb{E}[\Delta Y_t(0)|X_{t-1}, Z, D=1] = \mathbb{E}[\Delta Y_t(0)|X_{t-1}, Z, D=0]. \tag{SA26}$$

An advantage of this assumption is that it allows X_t to be a bad control.[5] That being said, it is a different identification assumption from the original one, and, at least to a certain extent, it does not correspond to the implicit identification strategy of the TWFE regression, which uses contemporaneous covariates. That being said, there are some very interesting connections that can be made here. Notice that using

[5]See Bonhomme and Sauder (2011), Lechner (2011), and Caetano et al. (2022) for more discussion. See also Renson et al. (2023) and Shahn et al. (2025) for approaches that extend ideas from the literature on time-varying treatments with panel data (Robins (1986), Robins et al. (2000), and Hernán and Robins (2020)) from a setting where identification is based on sequential unconfoundedness to a DiD setting.

this assumption leads to the following estimand:

$$\text{ATT}_{t^*} = \mathbb{E}[\Delta Y_{t^*}|D=1] - \mathbb{E}\big[\mathbb{E}[\Delta Y_{t^*}|X_{t^*-1}, Z, D=0]\big|D=1\big]. \tag{SA27}$$

This is the same estimand Callaway and Sant'Anna (2021) use with time-varying covariates, and we report some results in the application section that use this estimand as well. But, interestingly, one arrives at the same estimand by using the original parallel trends in Equation (SA23) combined with the following dimension reduction assumption:

$$\mathbb{E}[\Delta Y_t(0)|\mathbf{X}, Z, D=1] = \mathbb{E}[\Delta Y_t(0)|X_{t-1}, Z, D=1]. \tag{SA28}$$

Finally, it is worth briefly mentioning the conditions that rationalize the estimand in Equation (SA27) while allowing for X_t to be a bad control. In particular, consider the assumption that

$$(X_t(0), X_{t+1}(0), \ldots, X_{\mathrm{T}}(0)) \perp\!\!\!\perp D|X_1, \ldots, X_{t-1}, Z. \tag{SA29}$$

Caetano et al. (2022) refer to this condition as *covariate unconfoundedness*. It says that the covariates would evolve similarly for the treated and untreated groups, absent participating in the treatment. To contrast covariate unconfoundedness with covariate exogeneity, consider our application from the main text about stand-your-ground laws, and suppose that the only time-varying covariate is the state's population. Covariate exogeneity is plausible if stand-your-ground laws do not affect states' populations, but they allow for populations to evolve differently for the treated and untreated groups. Covariate unconfoundedness allows for stand-your-ground laws to affect states' populations, but it requires that the populations would evolve similarly for the treated and untreated groups, absent participating in the treatment. Most likely, researchers have the covariate exogeneity condition in mind when they consider covariates like a state's population.

Caetano et al. (2022) show that, together, the assumptions in Equations (SA21) and (SA29) imply that ATT_t is identified as in Equation (SA27). See Caetano et al. (2022) for more details about the covariate unconfoundedness assumption in Equation (SA29), as well as related approaches to dealing with bad controls.

SA3 Miscellaneous Additional Results/Details

This section contains further details about several issues that were briefly mentioned in the main text.

SA3.1 Models that Rationalize Parallel Trends and a Comparison to Other Approaches

In this section, we consider linear models for untreated potential outcomes. These models (i) provide a natural way to rationalize the conditional parallel trends assumption (see, e.g., Blundell and Costa Dias (2009), Gardner et al. (2023), and Borusyak et al. (2024)), (ii) clarify the assumptions required to eliminate TWFE misspecification bias, and (iii) facilitate comparison of our approach with heterogeneity-robust DiD estimators in the literature.

Consider the following linear model for untreated potential outcomes:[6]

$$Y_{it}(0) = \theta_t + \eta_i + Z_i'\delta_t + X_{it}'\beta_t + e_{it}. \tag{SA30}$$

Taking first differences implies that

$$\Delta Y_{it^*}(0) = \tilde{\theta}_{t^*} + Z_i'\tilde{\delta}_t + \Delta X_{it^*}'\beta_{t^*} + X_{it^*-1}'\tilde{\beta}_{t^*} + \Delta e_{it^*}, \tag{SA31}$$

where $\tilde{\beta}_{t^*} := (\beta_{t^*} - \beta_{t^*-1})$. In this model, the path of untreated potential outcomes can depend on time-invariant covariates, the level of time-varying covariates, and how time-varying covariates change over time. The conditional parallel trends assumption holds in this model for untreated potential outcomes under the condition that $\mathbb{E}[\Delta e_{t^*}|X_{t^*}, X_{t^*-1}, Z, D = 1] = \mathbb{E}[\Delta e_{t^*}|X_{t^*}, X_{t^*-1}, Z, D = 0]$.

Assumption 4 provided three conditions that were used in Theorem 2 to eliminate the misspecification bias terms that underly the TWFE estimand. These conditions can be seen as restrictions on the model for untreated potential outcomes in Equation (SA30). Condition (A) is satisfied if $\delta_t = \delta$. Conditions (B) and (C) are satisfied if, additionally, $\beta_t = \beta$. These are strong extra conditions that rule out time-invariant covariates affecting the path of untreated potential outcomes at all, and impose that the only way that time-varying covariates can affect the path of untreated potential outcomes is by how they change over time (not how the effects of the covariates change over time).

It is also interesting to compare the assumptions used above to rationalize the TWFE regression to other estimation strategies recently proposed in the econometrics literature to address the limitations of TWFE regressions (a number of which provide ways to include covariates).

First, Gardner et al. (2023), Borusyak et al. (2024), and Liu et al. (2024) propose "imputation" estimation strategies. The idea is basically to (i) split the data into the set of treated observations and untreated observations; (ii) using the untreated observations, estimate a model that includes time fixed-effects, unit fixed-effects, and covariates; and (iii) given the estimated parameters from this model, impute untreated potential outcomes for treated observations. An estimate of the ATT arises from comparing the average observed outcome for treated observations to the average imputed outcome for these observations. The particular version of imputation proposed by Gardner et al. (2023) and Borusyak et al. (2024) relies on estimating the following regression for untreated observations:

$$Y_{it}(0) = \theta_t + \eta_i + X_{it}'\beta + e_{it} \tag{SA32}$$

using untreated observations.

Specialized to the case with two time periods (besides the parallel trends assumption), the key condition to rationalize this approach is to assume that

$$\mathbb{E}[\Delta Y_{t^*}(0)|X_{t^*}, X_{t^*-1}, Z, D = 0] = \mathrm{L}_0(\Delta Y_{t^*}|\Delta X_{t^*}).$$

Thus, like the result on interpreting α in Theorem 2, the imputation estimators discussed here also im-

[6] To be clear, in this section, we use θ_t and η_i (and similar notation) as generic notation for time and unit fixed effects, and these are not the same as the corresponding terms in Equation (1).

plicitly rely on all three parts of Assumption 4. Or, relative to the model in Equation (SA30), these estimation strategies rely on β_t and δ_t in Equation (SA30) being constant across time periods, which implies that effects of time-varying and time-invariant covariates on untreated potential outcomes are constant over time. These are strong extra conditions that researchers ought to weigh carefully in applications.[7] However, relative to the regression in Equation (3), under exactly the same conditions, the imputation estimators directly target the ATT rather than recover a hard-to-interpret weighted average of conditional ATT's.

Next, Callaway and Sant'Anna (2021) propose propensity score re-weighting, regression adjustment, and doubly robust estimation strategies. While the doubly robust estimation strategy offers some additional advantages, the regression adjustment estimation strategy is immediately comparable to the discussion here.[8] Specialized to the case with two periods (in addition to parallel trends), their regression adjustment strategy imposes that

$$\mathbb{E}[\Delta Y_{t^*}(0)|X_{t^*}, X_{t^*-1}, Z, D = 0] = \mathrm{L}_0(\Delta Y_{t^*}|X_{t^*-1}, Z).$$

This condition requires that (i) the path of untreated potential outcomes conditional on time-varying and time-invariant covariates only depends on the time-varying covariates in the pre-treatment period (not in the post-treatment periods) and time-invariant covariates, and (ii) a linearity condition. The first condition is different from the one in Assumption 4, but it is in line with a number of papers in the econometrics literature on difference-in-differences that include time-invariant covariates and pre-treatment time-varying covariates (which are subsequently effectively treated as time-invariant covariates) in the parallel trends assumption (see, for example, Heckman et al. (1998), Abadie (2005), and Bonhomme and Sauder (2011)).

As in Callaway and Sant'Anna (2021), in the main text we proposed a doubly robust AIPW estimator. Specialized to its regression adjustment version, our approach amounts to assuming that

$$\mathbb{E}[\Delta Y_{t^*}(0)|X_{t^*}, X_{t^*-1}, Z, D = 0] = \mathrm{L}_0(\Delta Y_{t^*}|\Delta X_{t^*}, X_{t^*-1}, Z).$$

This is a linearity condition, but it inherits the advantages of both the imputation estimation strategies in Gardner et al. (2023) and Borusyak et al. (2024) and the regression adjustment strategies proposed in Callaway and Sant'Anna (2021): the path of untreated potential outcomes can depend on (i) the levels of time-varying covariates, (ii) the change in time-varying covariates over time, and (iii) time-invariant

[7] Alternatively, the approaches proposed in de Chaisemartin and D'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2024) impose a local version of linearity that results in the path of untreated potential outcomes only depending on the change in covariates over time, but allow for *how* the change in covariates over time affect the path of untreated potential outcomes to vary across time (those papers also discuss how to include time-invariant covariates in this framework). See, in particular, Assumption S4 in the Supplementary Appendix of de Chaisemartin and D'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2024, Eq. (24)). In the setting with two periods, this is a distinction without difference, but with multiple periods, this is a weaker condition, as it would not require Assumption MP-5(5) to hold, while the one-shot imputation estimators do need it.

[8] Regression adjustment is very similar in spirit to the imputation estimators discussed above. It is possible to view the regression adjustment estimators discussed here as imputation estimators; see Callaway (2023) for additional related discussion.

covariates. Moreover, like those approaches (but unlike α from the TWFE regression), this approach directly targets ATT.

SA3.2 Miscellaneous Additional Comments/Clarifications

This section contains several miscellaneous additional comments, clarifications, and details for some of the statements and claims made in the main text.

Remark SA14 (Sampling weights)**.** Many DiD applications include sampling weights. These weights are often used in applications with aggregate data where the number of individual units varies across the observed aggregate units. Researchers include sampling weights to give more weight to larger aggregate units, thereby adjusting the target parameter (see Pfeffermann (1993) and Solon et al. (2015)). For instance, our application uses state-level data, but the population size differs in each state. Many of the results in Cheng and Hoekstra (2013) are weighted by the population size of each state, aiming to interpret the ATT as being representative of the average effect on individuals rather than states. All of the arguments presented in the paper remain valid with sampling weights, with expectations replaced by weighted expectations. Furthermore, the code provided in our two companion software packages supports the use of sampling weights.

Remark SA15 (Event studies)**.** Our discussion in the main text mainly focused on ATT^o, but event studies are very commonly estimated and reported in empirical work. The event study parameter is given by

$$ATT^{es}(e) := \mathbb{E}[Y_{G+e} - Y_{G+e}(0)|G \in \mathcal{G}_e],$$

where e indexes event-time (the number of periods since the treatment started) and $\mathcal{G}_e = \{g \in \bar{\mathcal{G}} | g + e \in [2, \mathrm{T}]\}$, which is the set of groups that (i) ever-participate in the treatment and (ii) are observed to have been treated for e periods in some observed period. Thus, $\mathrm{ATT}^{es}(e)$ is the average treatment effect when units have been treated for exactly e periods. Like ATT^o, $\mathrm{ATT}^{es}(e)$ is a weighted average of group-time average treatment effects. In particular,

$$\mathrm{ATT}^{es}(e) := \sum_{g \in \mathcal{G}_e} w^{es}(g,e)\mathrm{ATT}(g, g+e),$$

where $w^{es}(g,e) := \mathrm{P}(G = g|G \in \mathcal{G}_e)$, which is the relative size of group g among groups that are observed to have been exposed to the treatment for e periods in some observed period. The expression for $\mathrm{ATT}^{es}(e)$ above indicates that, if we can identify/estimate group-time average treatment effects, then we can aggregate them into an event study. Although we did not emphasize event studies in the main text, they are computed by default in our accompanying `ptetools` R package. See Callaway and Sant'Anna (2021) for other parameters that may be of interest in DiD applications with multiple periods and variation in treatment timing.

Remark SA16 (Clarification on calculating implicit AIPW weights)**.** The weights $\vartheta_0^{\mathrm{L}_0}$ (from Lemma SA1 and Proposition 1) look challenging to estimate in practice because γ_0 involves the unknown propensity score $p(X)$. However, notice that an alternative expression for the weights can be found in Equation (SA6) in the proof of Lemma SA1. In particular, it also holds that $\vartheta_0^{\mathrm{L}_0} = \mathbb{E}[X'|D=1]\mathbb{E}[XX'|D=0]^{-1}X$, which can be directly estimated.

Remark SA17 (Non-normalized AIPW weights)**.** Proposition 1 was derived for the AIPW estimator that uses normalized weights, which often delivers better performance in finite samples (see, e.g., Busso et al. (2014)). Without normalized weights (i.e., if we use $\tilde{\varpi}_0^{aipw}$ instead of $\tilde{w}_0^{aipw}$), the claims of Proposition 1 still hold; namely, the weights ϑ_0^{aipw} still have mean one and can be negative. To see this, in this case, $\vartheta_0^{aipw} = \frac{1-\pi}{\pi}\left(\frac{\tilde{p}(X)}{(1-\tilde{p}(X))} + \gamma_0'X - \tilde{\gamma}_0'X\right)$, and $\mathbb{E}[\vartheta_0^{aipw}|D=0] = \frac{(1-\pi)}{\pi}\mathbb{E}[\gamma_0'X|D=0] = 1$, where we cancel the first and third terms in the expression for ϑ_0^{aipw} by the orthogonality of the projection errors. This argument is slightly different from the one in the proof of Proposition 1 because the mean of the first and third terms may not be equal to one, yet they are equal to each other and cancel out in the expression for ϑ_0^{aipw}.

Remark SA18 (Effective sample size calculation)**.** We calculate the effective sample size for the AIPW estimator in the case with multiple periods in the following way.[9]

$$\widehat{ESS}_0^o = \left(\sum_{g\in\bar{G}}\sum_{t=g}^{\mathrm{T}} \hat{w}^o(g,t) \times \frac{n_0}{\widehat{\mathrm{Var}}_0(\vartheta_{g,t}^{0,aipw}(\mathbf{X},Z))+1}\right)\mathrm{N}_{post},$$

where $\widehat{\mathrm{Var}}_0(\vartheta_{g,t}^{0,aipw}(\mathbf{X},Z))$ denotes the sample variance of the AIPW weights among untreated units, and

$$n_0 := \sum_{i=1}^n U_i \qquad \text{and} \qquad \mathrm{N}_{post} := \sum_{g\in\bar{G}}\sum_{t=g}^{\mathrm{T}} 1,$$

so that n_0 is the number of never-treated units and N_{post} is the cumulative number of post-treatment time periods across all groups. Reporting the effective sample size complements our covariate balance diagnostics, as it introduces a tradeoff when considering adding additional functions of the covariates to the model: adding more functions of the covariates tends to improve balance but may also lead to more extreme weights and, therefore, a smaller effective sample size. We are unaware of any existing definitions of effective sample size for a staggered treatment adoption setting. Future work could consider alternative, possibly better, definitions of effective sample size, but this version satisfies some natural properties. First, if $\widehat{\mathrm{Var}}_0(\vartheta_{g,t}^{0,aipw}(\mathbf{X},Z)) = 0$ for all g and t (this means that these weights are all equal to each other and equal to n_0^{-1}), then $\widehat{ESS}_0^o = n_0 \times \mathrm{N}_{post}$.[10] Second, for larger values of $\widehat{\mathrm{Var}}_0(\vartheta_{g,t}^{0,aipw}(\mathbf{X},Z))$ (indicating that the weights are concentrating more on fewer units), $\widehat{ESS}_0^o$ decreases. This is a desirable and standard property of notions of effective sample size. Finally, one can compute a similar measure of effective sample size for the TWFE regression by replacing $\hat{w}^o(g,t)$ with $\bar{w}^{twfe}(g,t)$ from the main text and $\vartheta_{g,t}^{0,aipw}(\mathbf{X},Z)$

[9] We report effective sample sizes only using untreated observations, as, at least for our AIPW estimators, the effective sample size for the treated group is just equal to the actual sample size. This holds because our weights on treated units are all equal to one.

[10] It is perhaps debatable whether one should define $\widehat{ESS}_0^o$ so that it is equal to n_0 or $n_0 \times \mathrm{N}_{post}$ in this case. We favor the second choice as, by virtue of aggregating, we effectively use more information for estimating ATT^o than for, say, $\mathrm{ATT}(g,t)$. If we aim for n_0 instead, then the measure of effective sample size would not acknowledge that we have more information for ATT^o than for $\mathrm{ATT}(g,t)$.

with $w_{g,t}^{0,twfe}(\mathbf{X}, Z)$ from the main text.

SA4 Additional Details/Results from the Application

SA4.1 More Details about the Setup in Cheng and Hoekstra (2013)

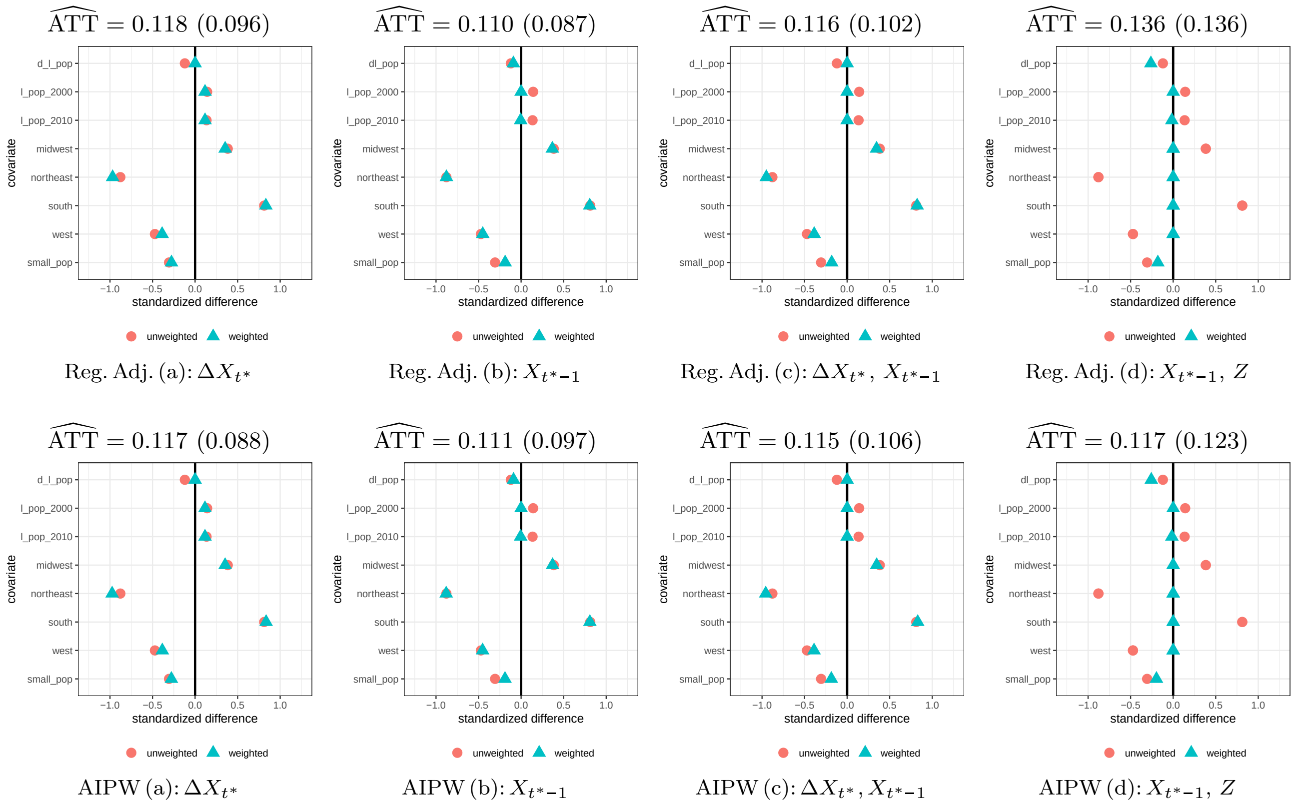
Cheng and Hoekstra (2013)'s main results come from a TWFE regression similar to Equation (1). However, it is worth clarifying a few additional differences relative to the setting we considered in the main text. First, many of their TWFE regressions include region-by-year fixed effects. In the main text, we mainly used region as an example of a time-invariant covariate that is not included in the TWFE regression, but we provide additional results where region-by-year fixed effects are included in the TWFE regression below. Second, for policies implemented during the middle of a year, they code the treatment as the percentage of the year during which the policy was implemented. By contrast, we set the treatment variable equal to one if the policy was implemented in any part of a particular year. In total, Cheng and Hoekstra (2013) consider six different TWFE specifications of increasing complexity, ranging from TWFE with no additional controls to including region-by-year fixed effects, time-varying covariates, contemporaneous crime rates (they argue that these are possibly endogenous and so mainly include them as a robustness check), state-specific linear time trends, and combinations of these. Besides that, Cheng and Hoekstra (2013) provide results for several additional outcomes besides just homicides. Notice that the covariate balancing properties of the TWFE regression (or alternative approaches that we proposed) are invariant to the outcome, i.e., all of the covariate balance figures reported in the main text are precisely the same if one uses a different outcome. That said, of course, changing the outcome changes the value of $\hat{\alpha}$ from the TWFE regression or $\widehat{\text{ATT}}$ from our approaches.

SA4.2 Additional Results from the Application

In this section, we provide three sets of additional results that were briefly mentioned in the main text. First, Figure SA1 provides covariate balance statistics for eight different specifications which come from (i) either using regression adjustment or AIPW or (ii) varying the covariates included in the estimations among the following four sets of covariates: (a) the change in log population only, (b) the level of log population in 2000 only, (c) both the change in log population and the level of log population in 2000, and (d) the level of log population in 2000 and region. Two of the eight specifications above are especially worth emphasizing. Specification (a), particularly in the regression adjustment case, corresponds to the specification used in imputation strategies that linearly include X_t. Specification (d), in either the regression adjustment or AIPW cases, corresponds to the default way to include covariates in Callaway and Sant'Anna (2021).

There are two main takeaways from this figure. First, in terms of covariate balance, regression adjustment and AIPW are very similar in all cases. Second, the regression adjustment specification (a), which only includes the change in log population, does not perform well in balancing the level of log population, especially compared to specifications that directly include the level of log population for at least one of the periods. On the other hand, Specification (d), which includes both the level of log population in 2000 and region, balances the covariates well, almost as well as our approach in Figure 1.

Figure SA1: Additional Results for Two Period Covariate Balance using Reg. Adj. and AIPW



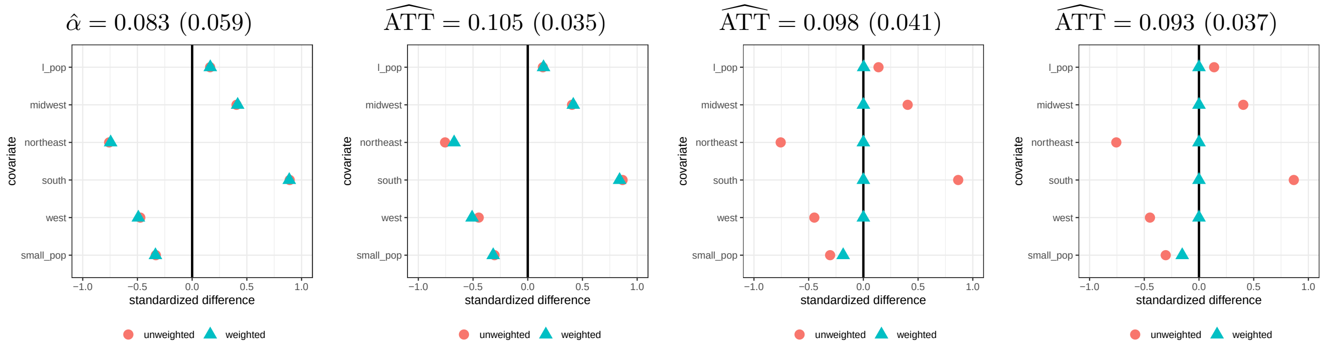
Notes: The figure reports estimates of the effects of stand-your-ground laws on homicides and covariate balance statistics. See Figure 1 for details about the covariates and interpreting standardized differences. The top row provides regression adjustment results with different covariate specifications indicated in each panel. The bottom row includes analogous results from AIPW estimation with different covariate specifications indicated in each panel. In each case, the same covariate specification is used for the propensity score as for the outcome regression model.

Second, we consider the short specification that only includes the log of population as a covariate, but move from the setting with two periods to using data from all years from 2000 to 2010. The results are reported in Figure SA2. The estimates of the effect of stand-your-ground laws on homicides are quite similar to the results with two periods that were reported in Figures 1 and SA1 except that the standard errors are notably smaller here. In terms of covariate balance, like the multiple-period results presented in the main text, we describe covariate balance in terms of how well each implicit weighting scheme balances the average of each covariate. The results are qualitatively similar to earlier results. The implicit TWFE weights (Panel (a)) essentially do not affect covariate balance relative to the raw data. Regression adjustment that only includes the change in time-varying covariates (Panel (b)) also does not improve covariate balance. In Panel (c), we control for the change in time-varying covariates, the pre-treatment level of time-varying covariates, and time-invariant covariates, which was our main "simple" suggested approach in the main text. Note that it is not "by construction" that this approach balances the average of the time-varying covariates.[11] However, despite that, it still performs well in terms of balancing the covariates: the standardized difference of average log population is 0.138 in the raw data, and it is reduced to 0.003 after applying the implicit regression adjustment weights. Finally, Panel (d) provides results where we use the first two principal components of log population as covariates. Once

[11] To be clear, it does balance the time-invariant covariates exactly by construction, but it does not balance the average of the time-varying covariates (here: log population) by construction.

again, covariate balance is not "by construction" equal to zero here, but, using the principal components, further reduces the standardized difference of the average of log population to 0.0000001 after applying the implicit weights.

Figure SA2: Covariate Balance with Multiple Periods

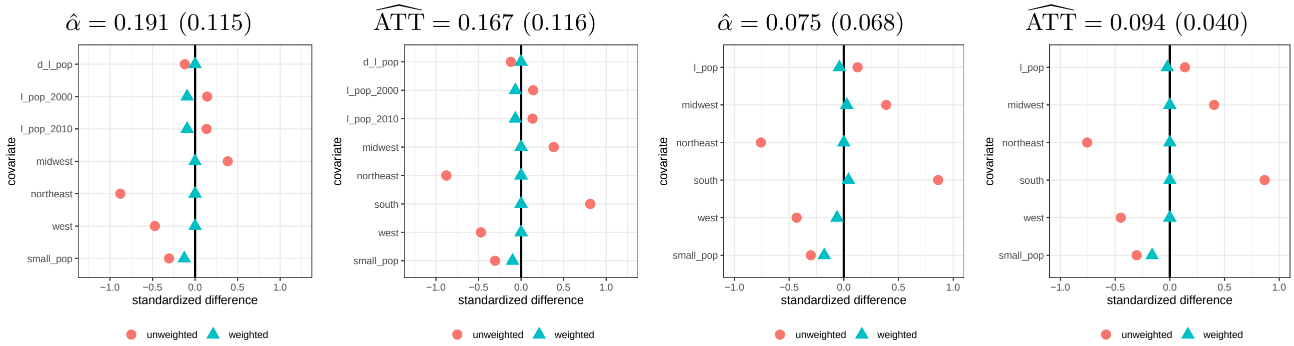


(a) TWFE (b) Reg. Adj: ΔX (c) Reg. Adj.: $\Delta X, X_{g-1}, Z$ (d) Reg. Adj: PCA

Notes: The figure reports estimates of the effects of stand-your-ground laws on homicides and covariate balance statistics using all available data from 2000-2010. The balance statistics are invariant to the outcome. Different covariates are displayed along the y-axis. `l_pop` is the average of the log of population for a particular state from 2000 to 2010; `midwest`, `northeast`, `south`, `west` are indicators of Census region; `small_pop` is an indicator for log(population) < 15;. The x-axis reports standardized differences for the mean of each covariate between the treated group and untreated group that come from our multi-period diagnostics for TWFE and regression adjustment/AIPW discussed in the main text. The red circles provide the standardized difference for the raw difference, and the blue triangles show the standardized difference after applying the implicit weighting scheme from each estimation method. Panel (a) provides results from a TWFE regression that includes D_t and X_t as regressors. Panels (b)-(d) come from regression adjustment estimators using different sets of covariates.

Finally, we provide results from TWFE regressions and regression adjustment that include region-by-year fixed effects in addition to the (transformed) log of population. These results are intermediate cases relative to the ones reported in the main text, where we provided results using short specifications that only included the log of population as a covariate, or results using long specifications that included time-invariant region, along with a large number of time-varying covariates. We provide results both for the case with two periods and with all time periods in Figure SA3. In terms of covariate balance, among the specifications that inherit transformed values of the time-varying covariates, these are the ones that perform the best. Notice that, in Panels (a), (b) and (d), the specifications balance region by construction, but all specifications here do well at balancing all the covariates being considered; this is especially true for regression adjustment. This suggests that, under the assumption that parallel trends holds conditional on log population and region, the regression adjustment approach that includes the change in log population and region does well for balancing region and the level of log population. In other words, it seems unlikely to be sensitive to hidden linearity bias. Still, we emphasize that there is value to explicitly checking covariate balance. As discussed in the main text, with the longer covariate specification (see Figure 2 in the main text), including only region and the changes in the covariates over time does not balance the levels of the same covariates well, indicating that these specifications could be sensitive to hidden linearity bias.

Figure SA3: Two Period Covariate Balance using TWFE and AIPW



(a) TWFE: $\Delta X_{t^*}, Z$ two periods

(b) Reg. Adj.: $\Delta X_{t^*}, Z$ two periods

(c) TWFE: $\Delta X, Z$ all periods

(d) Reg. Adj.: $\Delta X, Z$ all periods

Notes: The figure reports estimates of the effects of stand-your-ground laws on homicides and covariate balance statistics. Panels (a) and (b) use the two-period data discussed in the main text, while Panels (c) and (d) use all available data from 2000-2010. The balance statistics are invariant to the outcome. Different covariates are displayed along the y-axis. In the first two panels, `d_l_pop` is the change in the log of state-level population from 2000 to 2010; `l_pop_2000` and `l_pop_2010` are the level of the log of state-level population in 2000 and 2010, respectively. In the second two panels, `l_pop` is the average of the log of population for a particular state from 2000 to 2010. In all panels, `midwest`, `northeast`, `south`, `west` are indicators of Census region and `small_pop` is an indicator for log(population) < 15. The x-axis reports standardized differences for the mean of each covariate between the treated group and untreated group that come from our multi-period diagnostics for TWFE and regression adjustment/AIPW discussed in the main text. The red circles provide the standardized difference for the raw difference, and the blue triangles show the standardized difference after applying the implicit weighting scheme from each estimation method. All the results in the figure include region as a covariate.

SA5 Review of Empirical Applications

This section reviews 25 empirical difference-in-differences papers, focusing on how these papers use covariates in their analysis. The papers we review come from two sources. First, we use the twelve papers discussed in Roth (2022). Second, we use thirteen papers that were the source of empirical applications in either well-known difference-in-differences papers (de Chaisemartin and D'Haultfoeuille (2020), Goodman-Bacon (2021), Callaway and Sant'Anna (2021), Sun and Abraham (2021), Gardner et al. (2023), Borusyak et al. (2024), and Dube et al. (2025)) or in several difference-in-differences review papers (Baker et al. (2022), de Chaisemartin and D'Haultfoeuille (2023), and Baker et al. (2025)).

The results of our review are provided in Table SA1. As discussed in the main text, by far the most common way to include covariates (at least in the papers we reviewed) was to use the TWFE regression in Equation (1) that we emphasized in the main text. Of the 18 papers that included covariates in their main specifications, 15 included covariates as in Equation (1), 2 included baseline versions of the covariates (i.e., used the value of the time-varying covariates in the first period), and 1 added baseline values of some covariates to the TWFE regression in Equation (1).

Table SA1: Use of Covariates in Selected Papers

Paper	$\mathbf{X}'_{it}\boldsymbol{\beta}$	**Baseline Covs.**	**No Covs.**
Acemoglu et al. (2019)[a]			✓
Bailey and Goodman-Bacon (2015)	✓	✓	
Beck et al. (2010)	✓		
Black et al. (2022)	✓		
Bosch and Campos-Vazquez (2014)		✓	
Broda and Parker (2014)			✓
Cheng and Hoekstra (2013)	✓		
Deryugina (2017)		✓	
Deschenes et al. (2017)	✓		
Dobkin et al. (2018)[b]			✓
Dube et al. (2010)	✓		
Fauver et al. (2017)	✓		
Fitzpatrick and Lovenheim (2014)	✓		
Gallagher (2014)			✓
Gentzkow et al. (2011)	✓		
He and Wang (2017)	✓		
Kuziemko et al. (2018)	✓		
Lafortune et al. (2018)			✓
Leblebicioğlu and Weinberger (2020)	✓		
Markevich and Zhuravskaya (2018)[c]			✓
Stevenson and Wolfers (2006)	✓		
Tewari (2014)	✓		
Ujhelyi (2014)	✓		
Vella and Verbeek (1998)	✓		
Wolfers (2006)			✓

Notes: The table describes how different empirical papers use covariates. The column labeled $X'_{it}\beta$ indicates that the paper introduced covariates as in the TWFE regression in Equation (1) in the main text, the version of the TWFE regression that was considered in detail in the paper. The column labeled "Baseline Covs." indicates papers that used the value of time-varying covariates in a baseline period as covariates. Finally, the column labeled "No Covs." indicates papers that did not include covariates. There were several papers that did not exactly fit any category (see discussion in footnotes below); we used "No Covs." as the default in those cases.

[a] Some specifications include time-varying covariates, but the main results come from dynamic panel data models without covariates.

[b] The only covariate included is a control for survey wave.

[c] Includes only time-invariant covariates interacted with an indicator for post-treatment periods.

References

Abadie, Alberto (2005). "Semiparametric difference-in-differences estimators". *The Review of Economic Studies* 72.1, pp. 1–19.

Acemoglu, Daron, Suresh Naidu, Pascual Restrepo, and James A Robinson (2019). "Democracy does cause growth". *Journal of Political Economy* 127.1, pp. 47–100.

Almond, Douglas, Hilary W Hoynes, and Diane Whitmore Schanzenbach (2011). "Inside the war on poverty: The impact of food stamps on birth outcomes". *The Review of Economics and Statistics* 93.2, pp. 387–403.

Angrist, Joshua (1998). "Estimating the labor market impact of voluntary military service using Social Security data on military applicants". *Econometrica* 66.2, pp. 249–288.

Aronow, Peter and Cyrus Samii (2016). "Does regression produce representative estimates of causal

effects?" *American Journal of Political Science* 60.1, pp. 250–267.
Bailey, Martha J and Andrew Goodman-Bacon (2015). "The War on Poverty's experiment in public medicine: Community health centers and the mortality of older Americans". *American Economic Review* 105.3, pp. 1067–1104.
Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro H.C. Sant'Anna (2025). "Difference-in-differences designs: A practitioner's guide". *Journal of Economic Literature.* Forthcoming.
Baker, Andrew C, David F Larcker, and Charles CY Wang (2022). "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics* 144.2, pp. 370–395.
Beck, Thorsten, Ross Levine, and Alexey Levkov (2010). "Big bad banks? The winners and losers from bank deregulation in the United States". *Journal of Finance* 65.5, pp. 1637–1667.
Black, Bernard, Alex Hollingsworth, Leticia Nunes, and Kosali Simon (2022). "Simulated power analyses for observational studies: An application to the Affordable Care Act Medicaid expansion". *Journal of Public Economics* 213, p. 104713.
Blundell, Richard and Monica Costa Dias (2009). "Alternative approaches to evaluation in empirical microeconomics". *Journal of Human Resources* 44.3, pp. 565–640.
Bonhomme, Stephane and Ulrich Sauder (2011). "Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling". *Review of Economics and Statistics* 93.2, pp. 479–494.
Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024). "Revisiting event-study designs: Robust and efficient estimation". *Review of Economic Studies*, rdae007.
Bosch, Mariano and Raymundo M Campos-Vazquez (2014). "The trade-offs of welfare policies in labor markets with informal jobs: The case of the "Seguro Popular" program in Mexico". *American Economic Journal: Economic Policy* 6.4, pp. 71–99.
Broda, Christian and Jonathan A Parker (2014). "The economic stimulus payments of 2008 and the aggregate demand for consumption". *Journal of Monetary Economics* 68, S20–S36.
Busso, Matias, John DiNardo, and Justin McCrary (2014). "New evidence on the finite sample properties of propensity score reweighting and matching estimators". *Review of Economics and Statistics* 96.5, pp. 885–897.
Caetano, Carolina, Brantly Callaway, Robert Payne, and Hugo Sant'Anna (2022). "Difference in differences with time-varying covariates". Working Paper.
Callaway, Brantly (2023). "Difference-in-differences for policy evaluation". *Handbook of Labor, Human Resources and Population Economics.* Ed. by Zimmermann, Klaus F. Springer International Publishing, pp. 1–61.
Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2, pp. 200–230.
Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019). "The effect of minimum wages on low-wage jobs". *The Quarterly Journal of Economics* 134.3, pp. 1405–1454.
Chattopadhyay, Ambarish, Noah Greifer, and Jose R Zubizarreta (2023). "lmw: Linear model weights for causal inference". Working Paper.
Chattopadhyay, Ambarish and José R Zubizarreta (2023). "On the implied weights of linear regression for causal inference". *Biometrika* 110.3, pp. 615–629.
Cheng, Cheng and Mark Hoekstra (2013). "Does strengthening self-defense law deter crime or escalate violence? Evidence from expansions to Castle Doctrine". *Journal of Human Resources* 48.3, pp. 821–854.
de Chaisemartin, Clement and Xavier D'Haultfoeuille (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". *American Economic Review* 110.9, pp. 2964–2996.
de Chaisemartin, Clément and Xavier D'Haultfoeuille (2023). "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey". *The Econometrics Journal* 26.3, pp. C1–C30.

de Chaisemartin, Clément and Xavier D'Haultfoeuille (2024). "Difference-in-differences estimators of intertemporal treatment effects". *Review of Economics and Statistics*, pp. 1–45.

Deryugina, Tatyana (2017). "The fiscal cost of hurricanes: Disaster aid versus social insurance". *American Economic Journal: Economic Policy* 9.3, pp. 168–198.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro (2017). "Defensive investments and the demand for air quality: Evidence from the NOx budget program". *American Economic Review* 107.10, pp. 2958–2989.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J Notowidigdo (2018). "The economic consequences of hospital admissions". *American Economic Review* 108.2, pp. 308–352.

Dube, Arindrajit, Daniele Girardi, Òscar Jordà, and Alan M Taylor (2025). "A local projections approach to difference-in-differences". *Journal of Applied Econometrics* 40.7, pp. 741–758.

Dube, Arindrajit, T William Lester, and Michael Reich (2010). "Minimum wage effects across state borders: Estimates using contiguous counties". *The Review of Economics and Statistics* 92.4, pp. 945–964.

Fauver, Larry, Mingyi Hung, Xi Li, and Alvaro G Taboada (2017). "Board reforms and firm value: Worldwide evidence". *Journal of Financial Economics* 125.1, pp. 120–142.

Fitzpatrick, Maria D and Michael F Lovenheim (2014). "Early retirement incentives and student achievement". *American Economic Journal: Economic Policy* 6.3, pp. 120–154.

Gallagher, Justin (2014). "Learning about an infrequent event: Evidence from flood insurance take-up in the United States". *American Economic Journal: Applied Economics*, pp. 206–233.

Gardner, John, Neil Thakral, Linh T Tô, and Luther Yap (2023). "Two-stage differences in differences". Working Paper.

Gentzkow, Matthew, Jesse M Shapiro, and Michael Sinkinson (2011). "The effect of newspaper entry and exit on electoral politics". *American Economic Review* 101.7, pp. 2980–3018.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, pp. 254–277.

Goodman-Bacon, Andrew and Jamein P Cunningham (2019). "Changes in family structure and welfare participation since the 1960s: The role of legal services". Working Paper.

Hazlett, Chad and Yiqing Xu (2018). "Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data". Working Paper.

He, Guojun and Shaoda Wang (2017). "Do college graduates serving as village officials help rural China?" *American Economic Journal: Applied Economics* 9.4, pp. 186–215.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd (1998). "Characterizing selection bias using experimental data". *Econometrica* 66.5, pp. 1017–1098.

Hernán, Miguel A. and James M. Robins (2020). *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC.

Ishimaru, Shoya (2022). "What do we get from a two-way fixed effects estimator? Implications from a general numerical equivalence". Working Paper.

— (2024). "Empirical decomposition of the IV-OLS gap with heterogeneous and nonlinear effects". *Review of Economics and Statistics*, pp. 1–16.

Kline, Patrick (2011). "Oaxaca-Blinder as a reweighting estimator". *American Economic Review* 101.3, pp. 532–37.

Kuziemko, Ilyana, Katherine Meckel, and Maya Rossin-Slater (2018). "Does managed care widen infant health disparities? Evidence from Texas Medicaid". *American Economic Journal: Economic Policy* 10.3, pp. 255–283.

Lafortune, Julien, Jesse Rothstein, and Diane Whitmore Schanzenbach (2018). "School finance reform and the distribution of student achievement". *American Economic Journal: Applied Economics* 10.2, pp. 1–26.

Leblebicioğlu, Aslı and Ariel Weinberger (2020). "Credit and the labour share: Evidence from US states". *The Economic Journal* 130.630, pp. 1782–1816.

Lechner, Michael (2011). "The estimation of causal effects by difference-in-difference methods". *Foundations and Trends in Econometrics* 4.3, pp. 165–224.
Lin, Lihua and Zhengyu Zhang (2022). "Interpreting the coefficients in dynamic two-way fixed effects regressions with time-varying covariates". *Economics Letters* 216, p. 110604.
Liu, Licheng, Ye Wang, and Yiqing Xu (2024). "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data". *American Journal of Political Science* 68.1, pp. 160–176.
Markevich, Andrei and Ekaterina Zhuravskaya (2018). "The economic effects of the abolition of serfdom: Evidence from the Russian Empire". *American Economic Review* 108.4-5, pp. 1074–1117.
Pei, Zhuan, Jörn-Steffen Pischke, and Hannes Schwandt (2019). "Poorly measured confounders are more useful on the left than on the right". *Journal of Business & Economic Statistics* 37.2, pp. 205–216.
Pfeffermann, Danny (1993). "The role of sampling weights when modeling survey data". *International Statistical Review/Revue Internationale de Statistique*, pp. 317–337.
Renson, Audrey, Michael G Hudgens, Alexander P Keil, Paul N Zivich, and Allison E Aiello (2023). "Identifying and estimating effects of sustained interventions under parallel trends assumptions". *Biometrics* 79.4, pp. 2998–3009.
Robins, James (1986). "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect". *Mathematical Modelling* 7.9-12, pp. 1393–1512.
Robins, James M, Miguel Angel Hernán, and Babette Brumback (2000). "Marginal structural models and causal inference in Epidemiology". *Epidemiology* 11.5, p. 551.
Roth, Jonathan (2022). "Pretest with caution: Event-study estimates after testing for parallel trends". *American Economic Review: Insights* 4.3, pp. 305–322.
Shahn, Zach, Oliver Dukes, Meghana Shamsunder, David Richardson, Eric Tchetgen Tchetgen, and James Robins (2025). "Structural nested mean models under parallel trends assumptions". Working Paper.
Sloczynski, Tymon (2022). "Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights". *The Review of Economics and Statistics* 104.3, pp. 501–509.
Solon, Gary, Steven J Haider, and Jeffrey M Wooldridge (2015). "What are we weighting for?" *Journal of Human Resources* 50.2, pp. 301–316.
Stevenson, Betsey and Justin Wolfers (2006). "Bargaining in the shadow of the law: Divorce laws and family distress". *The Quarterly Journal of Economics* 121.1, pp. 267–288.
Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". *Journal of Econometrics* 225.2, pp. 175–199.
Tewari, Ishani (2014). "The distributive impacts of financial development: Evidence from mortgage markets during us bank branch deregulation". *American Economic Journal: Applied Economics* 6.4, pp. 175–196.
Ujhelyi, Gergely (2014). "Civil service rules and policy choices: evidence from US state governments". *American Economic Journal: Economic Policy* 6.2, pp. 338–380.
Vella, Francis and Marno Verbeek (1998). "Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men". *Journal of Applied Econometrics* 13.2, pp. 163–183.
Wolfers, Justin (2006). "Did unilateral divorce laws raise divorce rates? A reconciliation and new results". *American Economic Review* 96.5, pp. 1802–1820.
Yanagi, Takahide (2022). "Doubly robust difference-in-differences with general treatment patterns". Working Paper.