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

Carolina Caetano* Brantly Callaway†

September 9, 2024

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. Appendix SA covers the two period case. Appendix SB covers the multiple period case. Appendix SC provides additional clarifications and details about some issues mentioned in the main text. Finally, Appendix SD provides some supplementary results and discussion for the application about stand-your-ground laws in the main text.

## SA Additional Theoretical Results with Two Periods

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

### SA.1 Identification Results

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

**Proposition S1.** *Under Assumptions 1 to 3,*

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

*Proof.* Notice that

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

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

## SA.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

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

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

***Proof of Lemma 2.*** Notice that

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

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

□

## SA.3 Alternative Conditions on the Propensity Score for Interpreting TWFE Regressions

This section considers alternative conditions on the propensity score that can rationalize interpreting $\alpha$ in Equation (6) 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 S2.** *Under Assumptions 1 to 3 and Assumption PS,*

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

*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 $\alpha$ in Proposition A2 from the Appendix of the main text as a starting point to show the result. Under Assumption 3, the term in Equation (24) is equal to $\mathbb{E}[w(\Delta X_{t^*})ATT(X_{t^*}, X_{t^*-1}, Z)|D=1]$. Next, write the numerator of the term in Equation (25) as

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

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

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

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 $\Delta X_{t^*}$), the second equality also holds by the law of iterated expectations, and the third equality holds by Assumption PS. Next, notice that

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

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 re-writes 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 that 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, and, therefore, it immediately follows that the weights cannot be negative in this case. □

**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), Słoczyński (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 under conditions that (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 $\Delta X_{t^*}$, and (iii) the propensity score conditional on $\Delta X_{t^*}$ 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. Linearity of the propensity score is commonly used to rationalize interpreting the coefficient on a treatment variable as a weighted average of conditional average treatment effects in a setting with cross-sectional data and under the assumption of unconfoundedness (e.g., Angrist (1998) and Aronow and Samii (2016)). In those cases, it 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, it seems particularly implausible as (i) it requires the propensity score to only depend on changes in covariates over time, and (ii) 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 a model for a binary outcome relative to simple nonlinear models such as logit or probit.

## SA.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. In particular, The next proposition shows that the implicit regression weights in Section 4 of the main text balance the mean of $\Delta X_{t^*}$ between the treated and untreated group.

**Proposition S3.** *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:

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

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 $\Delta X_{t^*}$, 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 $\Delta X_{t^*}$ which is orthogonal to $\Delta X_{t^*}$. □

While Proposition S3 shows that the implicit regression weights balance $\Delta X_{t^*}$ 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

[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 in cases 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.

other functions of time-invariant and/or time-varying covariates). As leading examples, notice that

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

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

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

which holds by using the same arguments as in the proof of Proposition S3. 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.

## SA.5 More Details about the Properties of Implicit AIPW Weights

This section contains the proof of Lemma 3, 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.

***Proof of Lemma 3.*** Recall that we defined $X=(X_{t^*},X_{t^*-1},Z)$. Then, notice that

$$\begin{aligned}\mathbb{E}\Big[\mathrm{L}_0(\Delta Y_{t^*}|X)\Big|D=1\Big]&=\mathbb{E}\Big[X'\mathbb{E}[XX'|D=0]^{-1}\mathbb{E}[X\Delta Y_{t^*}|D=0]\Big|D=1\Big]\\&=\mathbb{E}\Big[\mathbb{E}[X'|D=1]\mathbb{E}[XX'|D=0]^{-1}X\Delta Y_{t^*}\Big|D=0\Big]\\&=\frac{(1-\pi)}{\pi}\mathbb{E}\left[\underbrace{\mathbb{E}\left[\frac{p(X)}{1-p(X)}X'\Big|D=0\right]\mathbb{E}[XX'|D=0]^{-1}}_{\gamma_0'}X\Delta Y_{t^*}\Big|D=0\right]\\&=\frac{(1-\pi)}{\pi}\mathbb{E}[\gamma_0'X\Delta Y_{t^*}|D=0]\\&=\mathbb{E}\Big[\vartheta_0^{\mathrm{L}_0}\Delta Y_{t^*}\Big|D=0\Big]\end{aligned}\tag{S1}$$

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 $p(X)/\big(1-p(X)\big)$ 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. As a first step, given the above definition of $\gamma_0$, we have that

$$\frac{p(X)}{\big(1-p(X)\big)}=\gamma_0'X+e^{odds}\tag{S2}$$

where $e^{odds}$ is the projection error from projection $p(X)/\big(1-p(X)\big)$ on $X$ among the untreated group. The projection error satisfies $\mathbb{E}[e^{odds}|D=0]=0$ by the orthogonality of projection. Then, we have that

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

where the first equality holds by the definition of $\vartheta_0^{\mathrm{L}_0}$, the second equality holds from Equation (S2), the third equality holds by the orthogonality of the projection, the fourth equality holds by repeatedly applying the law of iterated expectations and canceling terms, and the last equality holds because $\mathbb{E}[p(X)]=\pi$ by the definition of the propensity score.

Finally, one can see 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 $p(X)/\big(1-p(X)\big)$ on $X$ is negative. □

**Remark S1** (Implicit regression adjustment weights)**.** An immediate implication of Lemma 3 is that regression adjustment estimators can be re-formulated as weighting estimators. In particular, define

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

then it immediately follow from Lemma 3

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

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.

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

**Lemma S1.** *To conserve on 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 $\vartheta_0^{aipw}$ 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 $\vartheta_0^{aipw}$, the second equality holds by the definition of $\tilde{w}_0^{aipw}$ (and canceling the terms involving $\pi$ 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. □

# SB Additional Theoretical 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.

## SB.1 Proofs of Results with Multiple Periods

### SB.1.1 Identification Results

We start by proving the result on the identification of $ATT(g,t)$ in Proposition 2. Toward this end, we first provide a useful lemma and then prove the main result.

**Lemma S2.** *Under Assumptions MP-1 to MP-5 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-5, and the last equality holds by canceling all the terms involving $\mathbb{E}[Y_s(0)|\mathbf{X}, Z, U = 1]$ for $s = g, \ldots, t-1$. □

***Proof of Proposition 2.*** For any group $g \in \bar{\mathcal{G}}$ and $t \geq g$ (i.e., post-treatment periods for group $g$), we have that

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

where the first equality holds by the definition of $ATT_{g,t}(\mathbf{X}, Z)$, 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 S2 and by writing potential outcomes in terms of their observed counterparts. This proves the first part of the proposition. The second part of the result holds immediately by applying the law of iterated expectations to the expression for $ATT_{g,t}(\mathbf{X}, Z)$. □

### SB.1.2 TWFE Regressions with Multiple Periods

Next, we prove the results on interpreting TWFE regressions with multiple periods and variation in treatment timing. The first result collects several useful properties of double-demeaned random variables.

**Lemma S3.** *Let $A_{it}$ denote a random variable that can vary across units and time periods, let $\ddot{A}_{it}$ denote the double-demeaned version of $A_{it}$, let $B_i$ denote a random variable that does not vary over time, and let $\zeta_t$ denote a non-random variable that can change values across time periods. The following properties of double-demeaned random variables hold:*

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

The results in Lemma S3 are well known. We provide the proofs here for completeness.

*Proof.* For Part (1), notice that

$$\begin{aligned}\mathbb{E}[\ddot{A}_t] &= \mathbb{E}\left[A_t - \bar{A} - \mathbb{E}[A_t] + \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[A_s]\right] \\ &= \mathbb{E}[A_t] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}A_t\right] - \mathbb{E}[A_t] + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[A_t] \\ &= 0\end{aligned}$$

where the first equality holds by the definition of $\ddot{A}_{it}$, the second equality holds by passing the expectation through the sums/differences in the previous line and by the definition of $\bar{A}_i$, and the last equality holds by changing the order of the expectation and the sum for the second term and then canceling. For Part (2), notice that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\ddot{A}_{it} &= \frac{1}{T}\sum_{t=1}^{T}\left(A_{it} - \bar{A}_i - \mathbb{E}[A_t] + \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[A_s]\right) \\ &= \frac{1}{T}\sum_{t=1}^{T}A_{it} - \bar{A}_i - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[A_t] + \frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[A_s] \\ &= 0\end{aligned}$$

where the first equality holds by the definition of $\ddot{A}_{it}$, the second equality holds by passing the average through the sums/differences (and because $\bar{A}_i$ and $\frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[A_s]$ do not vary over time), and the last equality holds because the first term is equal to $\bar{A}_i$ and because the last two terms cancel. Part (3) is an immediate implication of Part (1). In particular,

$$\mathbb{E}\left[\ddot{A}_t\zeta_t\right] = \mathbb{E}[\ddot{A}_t]\zeta_t = 0$$

which holds by Part (1). Part (4) is an immediate implication of Part (2). In particular,

$$\frac{1}{T}\sum_{t=1}^{T}\ddot{A}_{it}B_i = \left(\frac{1}{T}\sum_{t=1}^{T}\ddot{A}_{it}\right)B_i = 0$$

which holds by Part (2). Part (5) also immediately follows from the previous results. In particular,

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\ddot{A}_tB\right] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\ddot{A}_t\right)B\right] = 0$$

which holds by Part (2). □

**Lemma S4.** *Under Assumptions MP-1, MP-3 and MP-4, the numerator in the expression for $\alpha$ in Equation (14) can be expressed as*

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

*Proof.* First, notice that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\ddot{D}_t\ddot{Y}_t\Big] &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_tY_t\big] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_t\bar{Y}\big] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_t\mathbb{E}[Y_t]\big] + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_t\big]\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_t] \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_tY_t\big] \end{aligned} \tag{S3}$$

where the first equality holds by the definition of $\ddot{Y}_t$, and the second equality holds by applying Lemma S3.5, Lemma S3.3, and Lemma S3.1 to the second, third, and fourth terms in the previous line, respectively. Next, notice that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\big(\ddot{X}_t'\Gamma\big)\ddot{Y}_t\Big] = \Gamma'\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\ddot{X}_t\ddot{Y}_t\Big] = \Gamma'\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\ddot{X}_tY_t\Big] \\ = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\big(\ddot{X}_t'\Gamma\big)Y_t\Big]\end{aligned} \tag{S4}$$

where the first equality holds by rearranging terms, the second equality holds by the same arguments as for $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\ddot{D}_t\ddot{Y}_t\big]$ above, and the last equality holds by rearranging terms again. The result holds by combining the expressions in Equations (S3) and (S4). □

**Lemma S5.** *Under Assumptions MP-1, MP-3 and MP-4, the following result holds*

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

*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

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

which holds by Lemma S3.5 because $Y_{G-1}$ is time-invariant. Next, notice that

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

where the first equality holds by changing the order of the expectation and summation, and the second equality holds by Lemma S3.2. Combining the expressions from the two previous displays completes the proof. □

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

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

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

$$\Lambda_0 := \left( \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big[ \ddot{X}_t^{\dagger} \ddot{X}_t^{\dagger'} \Big| U = 1 \Big] \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big[ \ddot{X}_t^{\dagger} \ddot{Y}_t^{\dagger} \Big| U = 1 \Big] \tag{S5}$$

and additionally recall that in the main text, we defined

$$\lambda_t := \mathbb{E}[Y_t - X_t' \Lambda_0 | U = 1] \quad \text{and} \quad \bar{\lambda} := \mathbb{E}[\bar{Y} - \bar{X}' \Lambda_0 | U = 1] \tag{S6}$$

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

$$\lambda_{G_i - 1} := \sum_{s=1}^{T} \lambda_s \mathbb{1}\{s = G_i - 1\}$$

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

**Lemma S6.** *Under Assumptions MP-1, MP-3 and MP-4, the following results hold*

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

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

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

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

*Proof.* For Part (A), notice that

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

where the first equality holds by rearranging terms, and the second equality holds by Lemma S3.1. For Part (B),

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

where the first equality holds by the definition of $\lambda_{G_i - 1}$, the second equality holds by re-arranging the sums and expectation, and the last equality holds by Lemma S3.2. Next, for Part (C), notice that

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

where the first equality holds using the same sort of argument (in reverse) as in Lemma S3, 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

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

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 S3.2. □

**Lemma S7.** *Under Assumptions MP-1, MP-3 and MP-4, the denominator in the expression for $\alpha$ in Equation (14) 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

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

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 S4, the third equality holds by the law of iterated expectations and the definition of $h(g,t)$ and by Assumption MP-1 (so that $D_t = \mathbb{1}\{t \geq G\}$), 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 $\mathcal{G}$). □

Next, we provide a proposition that delivers a useful decomposition for $\alpha$ in Equation (1) in the case with multiple periods and variation in treatment timing considered in Section 5.

**Proposition S4.** *Under Assumptions MP-1, MP-3 and MP-4, $\alpha$ 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 $w_{g,t}^{twfe}(\ddot{X}_t)$ is defined in Proposition 3 in the main text.*

*Proof.* First, we consider the numerator in the expression for $\alpha$ in Equation (14). Notice that

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

where the first equality holds by Lemma S4, the second equality holds by Lemma S5, the third equality holds by Lemma S6, 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 S10 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 (14) from Lemma S7 and by the definition of $w_{g,t}^{twfe}(\ddot{X}_t)$. □

**Proposition S5.** *Under Assumptions MP-1, MP-3 and MP-4, $\alpha$ 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 $w_{g,t}^{twfe}(\ddot{X}_t)$ is defined in Proposition 3 in the main text.*

*Proof.* The result holds immediately by applying the law of iterated expectations to both terms in the expression for $\alpha$ from Proposition S4. □

***Proof of Proposition 3.*** Starting from the expression for $\alpha$ in Proposition S5, the first part of the result holds by adding and subtracting

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

Next, we prove the properties of the weights. That the weights sum to one across post-treatment periods holds immediately by the definition of the weights. We show that the weights sum to negative one across pre-treatment periods in Lemma S11 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'Haultfœuille (2020), which can be negative. This completes the proof for the additional properties of the weights from the proposition. □

***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-5. For $g \in \bar{\mathcal{G}}$ and $t \geq g$ (i.e., post-treatment periods for group $g$), we have that

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

which holds by Proposition 2. Plugging these expressions into Proposition 3 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-6** (Additional Assumptions to Rule Out Bias Terms in TWFE Regression under Staggered Treatment Adoption)**.** *The following conditions hold for all time periods $t = 2, \ldots, T$:*[2]

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

[2] 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 $\Delta Y_t$ on $\Delta X_t$ 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 S8.** *Under Assumptions MP-1 to MP-5 and MP-6,*

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

*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-6(5), and the last equality holds by the definition of $\lambda_t$ in the main text. □

***Proof of Theorem 4.*** Notice that, from Theorem 3, the first result will hold if, for any $g \in \bar{\mathcal{G}}$ and for any $t \in \{1,\dots,T\}$, $\xi_{t,g-1}(\mathbf{X}, Z) = 0$ where

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

Consider the case where $t \geq g$ (and note 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 $\mathbb{E}[Y_s|\mathbf{X}, Z, U=1] - \big(\lambda_s + X_s'\Lambda_0\big)$ for all values of $s = g, \dots, t-1$, the second equality holds from applying Assumption MP-6(1)-(4), and the last equality holds by Assumption MP-6(5) and by Lemma S8.

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

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

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

where the second equality holds by property (i) of the weights from Proposition 3. □

## SB.2 Results on Implicit TWFE Weights with Multiple Periods

In this section, we prove the claim in Equation (18) 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 then we prove the claim from the main text.

**Lemma S9.** *Under Assumptions MP-1, MP-3 and MP-4,*

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

*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 S3.1, 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 S6.** *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 $\alpha$ in Equation (14), we have that

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

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

where the first equality holds by Lemma S4, the second equality holds as an implication of Lemma S3, the third equality holds by the law of iterated expectation, 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 S9, 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 $\alpha$ in Equation (14) and from the definition of $\bar{w}^{twfe}(g,t)$. □

**Proposition S7.** *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 S6, we have that

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

where the first equality holds by adding and subtracting

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

to the result of Proposition S6 and rearranging terms, and the second equality holds because $\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{T} \bar{w}^{twfe}(g,t)\mathbb{E}\left[w_{g,t}^{1,twfe}(\mathbf{X},Z)(Y_{g-1}-Y_{i1})\Big|G=g\right]=0$[3] 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 S2** (Comments about remainder term)**.** We note here that having a remainder in the expression for $\alpha$ in Proposition S7 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 S6. 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, $\sum_{g\in\bar{\mathcal{G}}}\sum_{t=1}^{T}\bar{w}^{twfe}(g,t)\mathbb{E}\left[w_{g,t}^{0,twfe}(\mathbf{X},Z)\Big|U=1\right]=0$. 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 is equal to zero if, for $t=2,\dots,T$, $\mathbb{E}[\Delta Y_t|\mathbf{X},Z,U=1]=\mathbb{E}[\Delta Y_t|U=1]$. 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.

### SB.3 Results on Implicit AIPW Weights with Multiple Periods

In this section, we prove the claim in Equation (21) in the main text about implicit AIPW weights with multiple periods.

[3] 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 S3.2.

**Proposition S8.** *Under Assumptions MP-1, MP-3 and MP-4,*

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

*Proof.* Recall from Equation (19) in the main text that

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

Considering the subgroup such that $\mathbb{1}\{G=g\}+U=1$ (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

$$\widetilde{ATT}^{aipw}(g,t) = \mathbb{E}\left[\vartheta^{1,aipw}_{g,t}(\mathbf{X},Z)(Y_t - Y_{g-1})\Big|G=g\right] - \mathbb{E}\left[\vartheta^{0,aipw}_{g,t}(\mathbf{X},Z)(Y_t - Y_{g-1})\Big|U=1\right] \tag{S7}$$

for any $g\in\bar{\mathcal{G}}$ and $t\geq g$. Next, recall that, by Equation (20) in the main text, we have that

$$\widetilde{ATT}^{aipw,o} = \sum_{g\in\bar{\mathcal{G}}}\sum_{t=g}^{T}\widetilde{ATT}^{aipw}(g,t)w^o(g,t) \tag{S8}$$

Plugging the expression for $\widetilde{ATT}^{aipw}(g,t)$ from Equation (S7) into Equation (S8) completes the proof. □

## SB.4 Additional Supporting Results

This section contains supporting results for proving the main results.

**Lemma S10.** *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

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}&\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\Big\{(Y_t - Y_T) - \Big(\lambda_t - \lambda_T + (X_t - X_T)'\Lambda_0\Big)\Big\}\Big|U=1\Big]\\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[(\ddot{D}_t - \ddot{X}_t'\Gamma)\Big\{(Y_t - \bar{Y}) - \Big(\lambda_t - \bar{\lambda} + (X_t - \bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\Big]\end{aligned} \tag{S9}$$

which holds from the properties of double-demeaned random variables in Lemma S3. Notice that, given the definitions of $\lambda_t$, $\bar{\lambda}$, and $\Lambda_0$, $\Big((Y_t-\bar{Y}) - \big((\lambda_t-\bar{\lambda}) + (X_t-\bar{X})'\Lambda_0\big)\Big)$ 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 $\zeta_t$, we have that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[\zeta_t\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\Big] \\
&\qquad\qquad = \frac{1}{T}\sum_{t=1}^{T}\zeta_t\left\{\mathbb{E}\Big[(Y_t-\bar{Y})-(X_t-\bar{X})'\Lambda_0\Big|U=1\Big]-(\lambda_t-\bar{\lambda})\right\} \\
&\qquad\qquad = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(S10)}
\end{aligned}$$

where the first equality holds by rearranging terms, and the second equality holds by the definitions of $\lambda_t$ 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.

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

Then, we can decompose the expression in Equation (S9) as

$$\begin{aligned}
(S9) = &-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\mathbb{E}[D_t]-\frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[D_s]\right)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\right] && \text{(S11)} \\
&-\Gamma'\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[(X_t-\bar{X})\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\right] && \text{(S12)} \\
&+\Gamma'\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(\mathbb{E}[X_t]-\frac{1}{T}\sum_{s=1}^{T}\mathbb{E}[X_s]\right)\Big\{(Y_t-\bar{Y})-\Big(\lambda_t-\bar{\lambda}+(X_t-\bar{X})'\Lambda_0\Big)\Big\}\Big|U=1\right] && \text{(S13)} \\
&= 0
\end{aligned}$$

where the result holds because (i) Equations (S11) and (S13) involve means of time-varying, nonrandom variables multiplied by the projection error discussed above which are equal to zero from the argument in Equation (S10), and (ii) Equation (S12) is equal to zero because the $(X_t - \bar{X})$ term is orthogonal to the projection error term, $\Big((Y_t-\bar{Y})-\big((\lambda_t-\bar{\lambda})+(X_t-\bar{X})'\Lambda_0\big)\Big)$. This completes the proof. □

**Lemma S11.** *Under Assumptions MP-1, MP-3 and MP-4, the weights in Proposition 3 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

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

where the first term holds immediately by applying the results in Lemma S3 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 S3** (Clarification on groups included in summation)**.** One subtle point worth mentioning is that the pre-treatment sum in the decomposition of $\alpha$ in Proposition 3 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 $\alpha$ (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 S10 as well as the proof of Proposition S4). Therefore, $\alpha$ effectively includes positive weight on a term equal to zero by construction.

# SC Miscellaneous Additional Results/Details

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

## SC.1 Detailed Comparison to Other Estimation Strategies

This section provides a more detailed discussion of how the conditions in Assumption 4, used to eliminate the misspecification bias terms in Theorem 1 in the main text, are related to implicit assumptions related to the covariates (implicit because they arise due to functional form choices for how covariates enter certain models) for several recently "heterogeneity-robust" approaches to DiD. We discussed this briefly in the main text in Remark 2.

First, Gardner, Thakral, Tô, and Yap (2023), Borusyak, Jaravel, and Spiess (2024), and Liu, Wang, and Xu (2024) propose "imputation" estimation strategies. The basic idea is to (i) split the data into the set of treated observations and untreated observations (by construction this will be an unbalanced panel dataset as it includes units that do not participate in the treatment in any period as well as data from pre-treatment periods for units that are eventually treated); (ii) to estimate a model that includes time fixed-effects, unit fixed-effects, and covariates; and (iii) given the estimated parameters from this model, to impute untreated potential outcomes for treated observations, and 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, Thakral, Tô, and Yap (2023) and Borusyak, Jaravel, and Spiess (2024) relies on estimating the following regression for untreated observations

$$Y_{it}(0) = \theta_t + \eta_i + X_{it}'\beta + e_{it} \tag{S14}$$

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 $\alpha$ in Theorem 2, the imputation estimators discussed here also implicitly rely on all three parts of Assumption 4. Or, relative to the discussion in Section 2, these estimation strategies rely on $\beta_t$ and $\delta_t$ in Equation (3) 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.[4] However, relative to the regression in Equation (6), 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.[5] 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 imposes 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 econo-

[4] Alternatively, the approaches proposed in de Chaisemartin and D'Haultfœuille (2020) and de Chaisemartin and d'Haultfœuille (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'Haultfœuille (2020) and de Chaisemartin and d'Haultfœuille (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-6(5) to hold while the one-shot imputation estimators do need it.

[5] 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.

metrics 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, Ichimura, Smith, and Todd (1998), Abadie (2005), and Bonhomme and Sauder (2011)).[6]

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, Thakral, Tô, and Yap (2023) and Borusyak, Jaravel, and Spiess (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 covariates. Moreover, like those approaches (but unlike $\alpha$ from the TWFE regression), this approach directly targets $ATT$.

## SC.2 Comments/Clarifications on Assumptions MP-1 to MP-5

This section contains some additional discussion about Assumptions MP-1 to MP-5, particularly concerning 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 S4** (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'Haultfœuille (2024) and Yanagi (2022) for recent work on relaxing this assumption.

**Remark S5** (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 S6** (Already-treated units)**.** In the main text, we dropped units that were already treated in the first periods. 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. We note that one interesting assumption, resulting in not necessarily dropping these units, is to assume that, after a group has been

[6] Under certain conditions, conditioning only on pre-treatment time-varying covariates can allow for the time-varying covariates themselves to be affected by the treatment. See Bonhomme and Sauder (2011), Lechner (2011), and Caetano, Callaway, Payne, and Sant'Anna (2022) for more discussion.

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, Dube, Lindner, and Zipperer (2019) for an example of using this type of extra condition.

**Remark S7** (Anticipation)**.** The no-anticipation assumption, Assumption MP-2, is widely used in the econometrics literature on difference-in-differences (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.

## SC.3 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 S8** (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, Haider, and Wooldridge (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 S9** (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 $\mathcal{G}_e = \{g \in \bar{\mathcal{G}} | g+e \in [2,T]\}$, 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, $ATT^{es}(e)$ is the average treatment effect when units have been treated for exactly $e$ periods. Like $ATT^o$, $ATT^{es}(e)$ is a weighted average of group-time average treatment effects. In particular,

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

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 $ATT^{es}(e)$ in the previous display 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 `pte` 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 S10** (Clarification on calculating implicit AIPW weights)**.** The weights $\vartheta_0^{\mathrm{Lo}}$ (from Lemma 3 and Proposition 1) look challenging to estimate in practice because $\gamma_0$ involves the unknown propensity score $p(X)$. However, notice that an alternative expression for the weights can be found in Equation (S1) in the proof of Lemma 3; in particular, it also holds that $\vartheta_0^{\mathrm{Lo}} = \mathbb{E}[X'|D=1]\mathbb{E}[XX'|D=0]^{-1}X$, which can be directly estimated.

**Remark S11** (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, DiNardo, and McCrary (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 $\vartheta_0^{aipw}$ 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 $\vartheta_0^{aipw}$ 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 both be equal to one, yet they are equal to each other and cancel out in the expression for $\vartheta_0^{aipw}$.

**Remark S12** (Effective sample size calculation)**.** We calculate the effective sample size for the AIPW estimator in the case with multiple periods in the following way.[7]

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

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

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

[7] 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.

so that $n_0$ is the number of never-treated units and $\mathrm{N}_{post}$ is the cumulative number of post-treatment time periods across all groups. 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 $\widehat{\mathrm{Var}}_0\big(\tilde{w}_{g,t}^{0,aipw}(\mathbf{X},Z)\big) = 0$ for all $g$ and $t$ (this means that these weights are all equal to each other and equal to $n_0^{-1}$), then $\widehat{ESS}_0^o = n_0 \times \mathrm{N}_{post}$.[8] Second, for larger values of $\widehat{\mathrm{Var}}_0\big(\tilde{w}_{g,t}^{0,aipw}(\mathbf{X},Z)\big)$—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 $\tilde{w}_{g,t}^{0,aipw}(\mathbf{X},Z)$ with $w_{g,t}^{0,twfe}(\mathbf{X},Z)$ from the main text.

# SD Additional Details and Results from the Application

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

Cheng and Hoekstra (2013)'s main results come from a similar TWFE regression to the one in Equation (1). However, it is worth clarifying a few additional differences relative to the setting that 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. Below, we provide additional results where region-by-year fixed effects are included in the TWFE regression. Second, for policies implemented during the middle of a year, they code the treatment as the percentage of the year that 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 additionally including region-by-year fixed effects, time-varying covariates, additional 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 $\widehat{ATT}$ from our approaches.

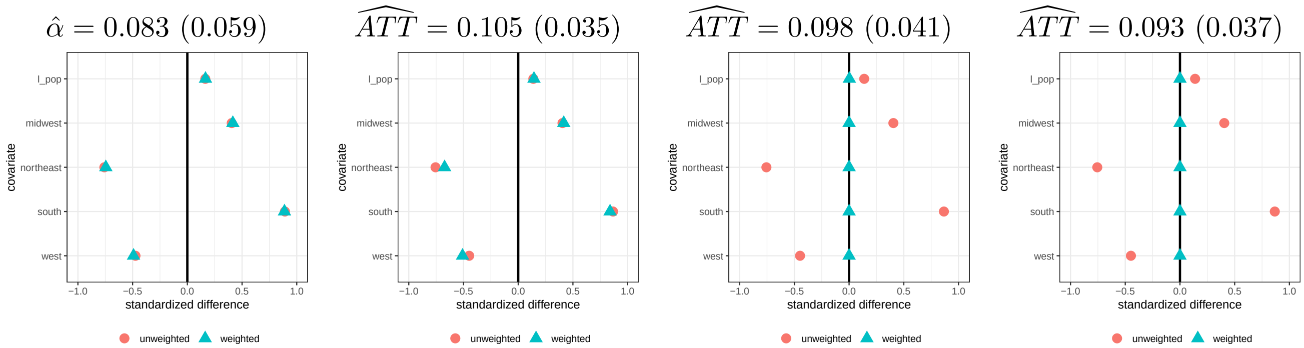
## SD.2 Additional Results from the Application

In this section, we provide two sets of additional results that were briefly mentioned in the main text. First, we consider the short-specification that only includes the log of population as a covariate,

[8] 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, $ATT(g,t)$. 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 $ATT(g,t)$.

but move from the setting with two periods to using data from all years from 2000 to 2010. The results are reported in Figure S1.

Figure S1: Covariate Balance with Multiple Periods




(a) TWFE (b) Reg. Adj: $\Delta X$ (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; and **`midwest`**, **`northeast`**, **`south`**, **`west`** are indicators of Census region. 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.

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 2 in the main text except for 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 those reported in the main text. 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.[9] 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 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.

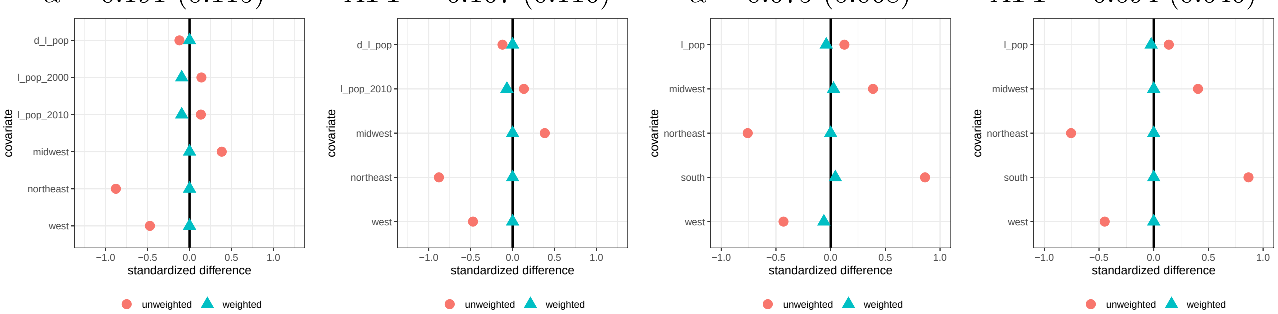
Finally, we provide results from TWFE regressions and regression adjustment that include region-

[9] 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.

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 S2. 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 3 in the main text), including only region and the changes in the covariates over time does not do well at balancing the levels of the same covariates, indicating that these specifications could be sensitive to hidden linearity bias.

Figure S2: Two Period Covariate Balance using TWFE and AIPW

$\hat{\alpha} = 0.191\ (0.115)$ | $\widehat{ATT} = 0.167\ (0.116)$ | $\hat{\alpha} = 0.075\ (0.068)$ | $\widehat{ATT} = 0.094\ (0.040)$




(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; and in all panels, **`midwest`**, **`northeast`**, **`south`**, **`west`** are indicators of Census region. 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.

# References


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

Angrist, Joshua D (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 M and Cyrus Samii (2016). "Does regression produce representative estimates of causal effects?" *American Journal of Political Science* 60.1, pp. 250–267.

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.

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.

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'Haultfœuille (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". *American Economic Review* 110.9, pp. 2964–2996.

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

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

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

Ishimaru, Shoya (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.

Lechner, Michael (2011). "The estimation of causal effects by difference-in-difference methods". *Foundations and Trends in Econometrics* 4.3, pp. 165–224.

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.

Pfeffermann, Danny (1993). "The role of sampling weights when modeling survey data". *International Statistical Review/Revue Internationale de Statistique*, pp. 317–337.

Słoczyński, 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.

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.

Yanagi, Takahide (2022). "Doubly robust difference-in-differences with general treatment patterns". Working Paper.