Supplementary Appendix: Difference-in-Differences with Multiple Time Periods

Brantly Callaway* Pedro H. C. Sant'Anna†

March 1, 2019

This supplementary appendix contains (a) the proofs of the results stated in the main text; (b) results for the case where a researcher has access to repeated cross sections data rather than panel data; (c) extensions of our main results when using "not yet treated" observations as a control group; (d) additional details on group-time average treatment effects under an unconditional parallel trends assumption, paying particular attention to the possibilities of using regressions to estimate group-time average treatment effects; and (e) additional details about the spatial distribution of state-level minimum wage policy changes in our sample.

Appendix A: Proofs of Main Results

We provide the proofs of our results in this appendix. Before proceeding, we first state and prove several auxiliary lemmas that help us proving our main theorems.

Let

$$ATT_X(g,t) = \mathbb{E}[Y_t(1) - Y_t(0)|X, G_g = 1].$$

Lemma A.1. *Under Assumptions 1-4, and for* $2 \leq g \leq t \leq \mathcal{T}$,

$$ATT_X(g,t) = \mathbb{E}[Y_t - Y_{g-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|X, C = 1] \ a.s..$$

Proof of Lemma A.1: In what follows, take all equalities to hold almost surely (a.s.). Notice that for identifying $ATT_X(g,t)$, the key term is $E[Y_t(0)|X, G_g = 1]$. And notice that for $h > s$, $E[Y_s(0)|X, G_h = 1] = E[Y_s|X, G_h = 1]$, which holds because in time periods before an individual is first treated, their untreated potential outcomes are observed outcomes. Also, note that, for $2 \leq g \leq t \leq \mathcal{T}$,

$$\mathbb{E}[Y_t(0)|X, G_g = 1] = \mathbb{E}[\Delta Y_t(0)|X, G_g = 1] + \mathbb{E}[Y_{t-1}(0)|X, G_g = 1] \quad \text{(A.1)}$$

*Department of Economics, Temple University. Email: brantly.callaway@temple.edu
†Department of Economics, Vanderbilt University. Email: pedro.h.santanna@vanderbilt.edu

$$= \mathbb{E}[\Delta Y_t|X, C = 1] + \mathbb{E}[Y_{t-1}(0)|X, G_g = 1],$$

where the first equality holds by adding and subtracting $E[Y_{t-1}(0)|X, G_g = 1]$ and the second equality holds by Assumption 2. If $g = t-1$, then the last term in the final equation is identified; otherwise, one can continue recursively in similar way to (A.1) but starting with $\mathbb{E}[Y_{t-1}(0)|X, G_g = 1]$. As a result,

$$\begin{aligned} \mathbb{E}[Y_t(0)|X, G_g = 1] &= \sum_{j=0}^{t-g} \mathbb{E}[\Delta Y_{t-j}|X, C = 1] + \mathbb{E}[Y_{g-1}|X, G_g = 1] \\ &= \mathbb{E}[Y_t - Y_{g-1}|X, C = 1] + \mathbb{E}[Y_{g-1}|X, G_g = 1]. \end{aligned} \tag{A.2}$$

Combining (A.2) with the fact that, for all $g \leq t$, $\mathbb{E}[Y_t(1)|X, G_g = 1] = \mathbb{E}[Y_t|X, G_g = 1]$ (which holds because observed outcomes for group g in period t with $g \leq t$ are treated potential outcomes), implies the result. $\square$

Next, recall that

$$\hat{\pi}_g = \arg\max_{\pi} \sum_{i:G_{ig}+C_i=1} G_{ig} \ln\left(p_g\left(X_i'\pi\right)\right) + (1 - G_{ig}) \ln\left(1 - p_g\left(X_i'\pi\right)\right),$$

$\dot{p}_g = \partial p_g(u)/\partial u$, $\dot{p}_g(X) = \dot{p}_g\left(X'\pi_g^0\right)$, and π_g^0 is the true, unknown vector of parameter indexing the generalized propensity score $p_g(X) = \mathbb{E}\left[G_g|X, G_g + C = 1\right]$.

Lemma A.2. *Under Assumption 5,*

$$\sqrt{n}\left(\hat{\pi}_g - \pi_g^0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_g^{\pi}\left(\mathcal{W}_i\right) + o_p(1),$$

where

$$\xi_g^{\pi}(\mathcal{W}) = \mathbb{E}\left[\frac{(G_g + C)\,\dot{p}_g(X)^2}{p_g(X)\left(1 - p_g(X)\right)} XX'\right]^{-1} X \frac{(G_g + C)\left(G_g - p_g(X)\right)\dot{p}_g(X)}{p_g(X)\left(1 - p_g(X)\right)}.$$

Proof of Lemma A.2: Let $n_{gc} = \sum_{i=1}^{n}(C_i + G_{ig})$. Under Assumption 5, from Theorem 5.39 and Example 5.40 in van der Vaart (1998), we have

$$\begin{aligned} &\sqrt{n_{gc}}\left(\hat{\pi}_g - \pi_g^0\right) \\ &= \frac{1}{\sqrt{n_{gc}}} \sum_{i:G_{ig}+C_i=1} \left(\mathbb{E}\left[\frac{\dot{p}_g(X)^2}{p_g(X)\left(1 - p_g(X)\right)} XX' \Big| G_g + C = 1\right]^{-1} X_i \frac{\left(G_{ig} - p_g(X_i)\right)\dot{p}_g(X_i)}{p_g(X_i)\left(1 - p_g(X_i)\right)}\right) + o_p(1) \\ &= \frac{\mathbb{E}\left[G_g + C\right]}{\sqrt{n_{gc}}} \sum_{i=1}^{n} \left(\mathbb{E}\left[\frac{(G_g + C)\,\dot{p}_g(X)^2}{p_g(X)\left(1 - p_g(X)\right)} XX'\right]^{-1} X_i \frac{(G_{ig} + C_i)\left(G_{ig} - p_g(X_i)\right)\dot{p}_g(X_i)}{p_g(X_i)\left(1 - p_g(X_i)\right)}\right) + o_p(1) \end{aligned}$$

$$= \frac{\mathbb{E}_n\left[G_g + C\right]}{\sqrt{n_{gc}}} \sum_{i=1}^{n} \xi_g^{\pi}\left(\mathcal{W}_i\right) + o_p\left(1\right)$$
$$= \frac{\sqrt{n_{gc}}}{n} \sum_{i=1}^{n} \xi_g^{\pi}\left(\mathcal{W}_i\right) + o_p\left(1\right).$$

Thus,

$$\begin{aligned} \sqrt{n}\left(\hat{\pi}_g - \pi_g^0\right) &= \frac{\sqrt{n}}{\sqrt{n_{gc}}} \sqrt{n_{gc}}\left(\hat{\pi}_g - \pi_g^0\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_g^{\pi}\left(\mathcal{W}_i\right) + o_p\left(1\right), \end{aligned}$$

and the proof is complete. □

For an arbitrary π, let $p_g\left(x;\pi\right) = p_g\left(x'\pi\right),\ \dot{p}_g\left(x;\pi\right) = \dot{p}_g\left(x'\pi\right)$, for all $g = 2, \ldots, \mathcal{T}$. Define the classes of functions,

$$\begin{aligned} \mathcal{H}_{1,g} &= \left\{ (x,c) \mapsto c \frac{p_g\left(x;\pi\right)}{1 - p_g\left(x;\pi\right)} : \pi \in \Pi_g \right\}, \\ \mathcal{H}_{2,g} &= \left\{ (x,c,y_t,y_{g-1}) \mapsto c \frac{p_g\left(x;\pi\right)\left(y_t - y_{g-1}\right)}{1 - p_g\left(x;\pi\right)} : \pi \in \Pi_g \right\} \\ \mathcal{H}_{3,g} &= \left\{ (x,c,y_t,y_{g-1}) \mapsto x \frac{c\ \dot{p}_g\left(x;\pi\right)\left(y_t - y_{g-1}\right)}{\left(1 - p_g\left(x;\pi\right)\right)^2} : \pi \in \Pi_g \right\}, \\ \mathcal{H}_{4,g} &= \left\{ (x,c) \mapsto x \frac{c\ \dot{p}_g\left(x;\pi\right)}{\left(1 - p_g\left(x;\pi\right)\right)^2} : \pi \in \Pi_g \right\}, \\ \mathcal{H}_{5,g} &= \left\{ (x,c,g_g) \mapsto x \frac{\left(g_g + c\right)\left(g_g - p_g\left(x;\pi\right)\right)\dot{p}_g\left(x;\pi\right)}{p_g\left(x;\pi\right)\left(1 - p_g\left(x;\pi\right)\right)} : \pi \in \Pi_g \right\}. \end{aligned}$$

Lemma A.3. *Under Assumptions 1 and 5, for all $g = 2, \ldots, \mathcal{T}$, $t = 2, \ldots, \mathcal{T}$, the classes of functions $\mathcal{H}_{j,g}$, $j = \{1, 2 \ldots, 5\}$, are Donsker.*

Proof of Lemma A.3: This follows from Example 19.7 in van der Vaart (1998).

Lemma A.4. *Under Assumptions 1 and 5, the null hypothesis*

$$H_0 : \mathbb{E}[Y_t - Y_{t-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{t-1}|X, C = 1] = 0 \ a.s. for\ all\ 2 \leq t < g \leq \mathcal{T},$$

can be equivalently written as

$$H_0 : \mathbb{E}\left[\left. \left(\frac{G_g}{\mathbb{E}\left[G_g\right]} - \frac{\dfrac{p_g\left(X\right) C}{1 - p_g\left(X\right)}}{\mathbb{E}\left[\dfrac{p_g\left(X\right) C}{1 - p_g\left(X\right)}\right]} \right) \left(Y_t - Y_{t-1}\right) \right| X \right] = 0 \ a.s. for\ all\ 2 \leq t < g \leq \mathcal{T}.$$

Proof of Lemma A.4: First note that

$$\mathbb{E}[Y_t - Y_{t-1}|X, G_g = 1] = \mathbb{E}[G_g (Y_t - Y_{t-1}) |X, G_g = 1] \\ = \mathbb{E}\left[\frac{G_g}{\mathbb{E}[G_g|X]} (Y_t - Y_{t-1}) \middle| X \right].$$

Analogously,

$$\mathbb{E}[Y_t - Y_{t-1}|X, C = 1] = \mathbb{E}\left[\frac{C}{\mathbb{E}[C|X]} (Y_t - Y_{t-1}) \middle| X \right],$$

implying that

$$\mathbb{E}[Y_t - Y_{t-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{t-1}|X, C = 1] = 0 \ \ a.s.\text{for all } 2 \leq t < g \leq \mathcal{T}.$$

$$\Longleftrightarrow$$

$$\mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g|X]} - \frac{C}{\mathbb{E}[C|X]} \right) (Y_t - Y_{t-1}) \middle| X \right] = 0 \ a.s.\text{for all } 2 \leq t < g \leq \mathcal{T}.$$

Given that under Assumptions 4 and 5, $\mathbb{E}[G_g + C|X] > 0$ $a.s.$, we have that

$$\mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g|X]} - \frac{C}{\mathbb{E}[C|X]} \right) (Y_t - Y_{t-1}) \middle| X \right] = 0 \ a.s.\text{for all } 2 \leq t < g \leq \mathcal{T}$$

if and only if

$$\mathbb{E}\left[\mathbb{E}[G_g + C|X] \left(\frac{G_g}{\mathbb{E}[G_g|X]} - \frac{C}{\mathbb{E}[C|X]} \right) (Y_t - Y_{t-1}) \middle| X \right] = 0 \ a.s.\text{for all } 2 \leq t < g \leq \mathcal{T}. \quad \text{(A.3)}$$

By noticing that

$$p_g(X) = \frac{\mathbb{E}[G_g|X]}{\mathbb{E}[G_g + C|X]}, \ 1 - p_g(X) = \frac{\mathbb{E}[C|X]}{\mathbb{E}[G_g + C|X]},$$

and that both of these are bounded away from zero under Assumption 5, we can rewrite (A.3) as

$$\mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\dfrac{p_g(X)C}{1 - p_g(X)}}{\mathbb{E}\left[\dfrac{p_g(X)C}{1 - p_g(X)}\right]} \right) (Y_t - Y_{t-1}) \middle| X \right] = 0 \ a.s.\text{for all } 2 \leq t < g \leq \mathcal{T},$$

since

$$\mathbb{E}\left[\frac{p_g(X)C}{(1 - p_g(X))}\right] = \mathbb{E}\left[\frac{\mathbb{E}[G_g|X, C + G_g = 1]C}{\mathbb{E}[C|X, C + G_g = 1]}\right] \\ = \mathbb{E}\left[\frac{\mathbb{E}[G_g|X]C}{\mathbb{E}[C|X]}\right] \\ = \mathbb{E}\left[\frac{\mathbb{E}[G_g|X]\mathbb{E}[C|X]}{\mathbb{E}[C|X]}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{E}\left[G_g|X\right]\right] \\
&= \mathbb{E}\left[G_g\right]. \qquad \text{(A.4)}
\end{aligned}
$$

This completes the proof.□

Now, we are ready to proceed with the proofs of our main theorems.

Proof of Theorem 1: Given the result in Lemma A.1,

$$
\begin{aligned}
ATT(g,t) &= \mathbb{E}[ATT_X(g,t)|G_g = 1] \\
&= \mathbb{E}\left[\mathbb{E}[Y_t - Y_{g-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|X, C = 1]\Big|G_g = 1\right] \\
&:= \mathbb{E}[A_X|G_g = 1] - \mathbb{E}[B_X|G_g = 1],
\end{aligned}
$$

and we consider each term separately. For the first term

$$
\begin{aligned}
\mathbb{E}[A_X|G_g = 1] &= \mathbb{E}[Y_t - Y_{g-1}|G_g = 1] \\
&= \mathbb{E}\left[\frac{G_g}{\mathbb{E}\left[G_g\right]}(Y_t - Y_{g-1})\right]. \qquad \text{(A.5)}
\end{aligned}
$$

For the second term, by repetition of the law of iterated expectations, we have

$$
\begin{aligned}
\mathbb{E}[B_X|G_g = 1] &= \mathbb{E}\left[\mathbb{E}[Y_t - Y_{g-1}|X, C = 1]\Big|G_g = 1\right] \\
&= \mathbb{E}\left[G_g\mathbb{E}[C(Y_t - Y_{g-1})|X, C = 1]\Big|G_g = 1\right] \\
&= \mathbb{E}\left[G_g\mathbb{E}\left[\frac{C}{(1 - p_g(X))}(Y_t - Y_{g-1})\big|X, G_g + C = 1\right]\Big|G_g = 1\right] \\
&= \frac{\mathbb{E}\left[G_g\mathbb{E}\left[\frac{C}{(1 - p_g(X))}(Y_t - Y_{g-1})\big|X, G_g + C = 1\right]\Big|G_g + C = 1\right]}{\mathbb{E}\left[G_g|G_g + C = 1\right]} \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}(Y_t - Y_{g-1})\big|X, G_g + C = 1\right]\Big|G_g + C = 1\right]}{\mathbb{E}\left[G_g|G_g + C = 1\right]} \\
&= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\left[\mathbb{E}\left[G_g + C|X\right]\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}(Y_t - Y_{g-1})\big|X, G_g + C = 1\right]\right] \\
&= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\left[\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}(Y_t - Y_{g-1})\big|X\right]\right] \\
&= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}(Y_t - Y_{g-1})\right] \\
&= \frac{\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}(Y_t - Y_{g-1})\right]}{\mathbb{E}\left[\frac{p_g\left(X\right)C}{(1 - p_g(X))}\right]}, \qquad \text{(A.6)}
\end{aligned}
$$

where (A.6) follows from (A.4). The proof is completed by combining (A.5) and (A.6). □

Proof of Theorem 2: Remember that

$$\widehat{ATT}(g,t) = \mathbb{E}_n\left[\frac{G_g}{\mathbb{E}_n[G_g]}(Y_t - Y_{g-1})\right] - \mathbb{E}_n\left[\frac{\dfrac{\hat{p}_g(X)C}{1-\hat{p}_g(X)}}{\mathbb{E}_n\left[\dfrac{\hat{p}_g(X)C}{1-\hat{p}_g(X)}\right]}(Y_t - Y_{g-1})\right],$$
$$:= \widehat{ATT}_g(g,t) - \widehat{ATT}_C(g,t),$$

and

$$ATT(g,t) = \mathbb{E}\left[\frac{G_g}{\mathbb{E}[G_g]}(Y_t - Y_{g-1})\right] - \mathbb{E}\left[\frac{\dfrac{p_g(X)C}{1-p_g(X)}}{\mathbb{E}\left[\dfrac{p_g(X)C}{1-p_g(X)}\right]}(Y_t - Y_{g-1})\right]$$
$$:= ATT_g(g,t) - ATT_C(g,t).$$

In what follows we will separately show that, for $2 \leq g \leq t \leq \mathcal{T}$,

$$\sqrt{n}\left(\widehat{ATT}_g(g,t) - ATT_g(g,t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{gt}^{G}(\mathcal{W}_i) + o_p(1), \tag{A.7}$$

and

$$\sqrt{n}\left(\widehat{ATT}_C(g,t) - ATT_C(g,t)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{gt}^{C}(\mathcal{W}_i) + o_p(1). \tag{A.8}$$

Then,

$$\sqrt{n}(\widehat{ATT}(g,t) - ATT(g,t)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{gt}(\mathcal{W}_i) + o_p(1)$$

hold from (A.7) and (A.8), and the asymptotic normality result follows from the application of the multivariate central limit theorem.

Let $\beta_g = \mathbb{E}[G_g]$ and $\widehat{\beta}_g = \mathbb{E}_n[G_g]$, and note that

$$\sqrt{n}\left(\widehat{\beta}_g - \beta_g\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(G_{ig} - \mathbb{E}[G_g]).$$

Then, for all $2 \leq g \leq t \leq \mathcal{T}$, by the continuous mapping theorem,

$$\sqrt{n}\left(\widehat{ATT}_g(g,t) - ATT_g(g,t)\right) = \frac{1}{\widehat{\beta}_g}\sqrt{n}\left(\mathbb{E}_n[G_g(Y_t - Y_{g-1})] - \mathbb{E}[G_g(Y_t - Y_{g-1})]\right)$$
$$- \mathbb{E}[G_g(Y_t - Y_{g-1})]\sqrt{n}\left(\frac{1}{\beta_g} - \frac{1}{\widehat{\beta}_g}\right)$$

$$
\begin{aligned}
&= \frac{1}{\beta_g}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(G_{ig}\left(Y_{it}-Y_{ig-1}\right)-\mathbb{E}\left[G_g\left(Y_t-Y_{g-1}\right)\right]\right)\\
&- \frac{\mathbb{E}\left[G_g\left(Y_t-Y_{g-1}\right)\right]}{\beta_g^2}\sqrt{n}\left(\widehat{\beta}_g-\beta_g\right)+o_p\left(1\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{G_{ig}\left(Y_{it}-Y_{ig-1}\right)}{\beta_g}-\frac{G_{ig}\mathbb{E}\left[G_g\left(Y_t-Y_{g-1}\right)\right]}{\beta_g^2}\right)+o_p\left(1\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{G_{ig}\left(\left(Y_{it}-Y_{ig-1}\right)-ATT_g(g,t)\right)}{\beta_g}+o_p\left(1\right)\\
&:= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{gt}^{G}(\mathcal{W}_i)+o_p\left(1\right),
\end{aligned}
$$

concluding the proof of (A.7).

Next we focus on (A.8). For an arbitrary function g, let

$$
w\left(g\right)=\frac{g\left(X\right)C}{1-g\left(X\right)},
$$

and note that

$$
\begin{aligned}
\sqrt{n}\left(\widehat{ATT}_C(g,t)-ATT_C(g,t)\right) &= \frac{1}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}\sqrt{n}\left(\mathbb{E}_n\left[w\left(\hat{p}_g\right)\left(Y_t-Y_{g-1}\right)\right]-\mathbb{E}\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)\right]\right)\\
&- \frac{\mathbb{E}\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)\right]}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]\mathbb{E}\left[w\left(p_g\right)\right]}\sqrt{n}\left(\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]-\mathbb{E}\left[w\left(p_g\right)\right]\right)\\
&:= \frac{1}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}\cdot\sqrt{n}A_n\left(\hat{p}_g\right)-\frac{ATT_C(g,t)}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}\cdot\sqrt{n}B_n\left(\hat{p}_g\right).
\end{aligned}
$$

From Assumption 5, Lemmas A.2 and A.3, and the continuous mapping theorem,

$$
\begin{aligned}
\frac{1}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]} &= \frac{1}{\mathbb{E}\left[w\left(p_g\right)\right]}+o_p\left(1\right),\\
\frac{ATT_C(g,t)}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]} &= \frac{ATT_C(g,t)}{\mathbb{E}\left[w\left(p_g\right)\right]}+o_p\left(1\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\sqrt{n}\left(\widehat{ATT}_C(g,t)-ATT_C(g,t)\right) &= \frac{1}{\mathbb{E}\left[w\left(p_g\right)\right]}\cdot\sqrt{n}A_n\left(\hat{p}_g\right)\\
&\qquad - \frac{ATT_C(g,t)}{\mathbb{E}\left[w\left(p_g\right)\right]}\cdot\sqrt{n}B_n\left(\hat{p}_g\right)+o_p\left(1\right) \qquad \text{(A.9)}
\end{aligned}
$$

Applying a classical mean value theorem argument,

$$
A_n\left(\hat{p}_g\right)=\mathbb{E}_n\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)\right]-\mathbb{E}\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)\right]
$$

$$+\mathbb{E}_n\left[X\left(\frac{C}{1-p_g\left(X;\bar{\pi}_g\right)}\right)^2\dot{p}_g\left(X;\bar{\pi}_g\right)\left(Y_{it}-Y_{ig-1}\right)\right]'\left(\hat{\pi}_g-\pi_g^0\right),$$

where $\bar{\pi}$ is an intermediate point that satisfies $\left|\bar{\pi}_g-\pi_g^0\right|\leq\left|\hat{\pi}_g-\pi_g^0\right|$ $a.s.$ Thus, by Assumption 5, Lemmas A.2 and A.3, and the Classical Glivenko-Cantelli's theorem,

$$A_n\left(\hat{p}_g\right)=\mathbb{E}_n\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)-\mathbb{E}\left[w\left(p_g\right)\left(Y_t-Y_{g-1}\right)\right]\right] \tag{A.10}$$

$$+\mathbb{E}\left[X\left(\frac{C}{1-p_g\left(X\right)}\right)^2\dot{p}_g\left(X\right)\left(Y_{it}-Y_{ig-1}\right)\right]'\left(\hat{\pi}_g-\pi_g^0\right)+o_p\left(n^{-1/2}\right). \tag{A.11}$$

Analogously,

$$B_n\left(\hat{p}_g\right)=\mathbb{E}_n\left[w\left(p_g\right)-\mathbb{E}\left[w\left(p_g\right)\right]\right]$$
$$+\mathbb{E}\left[X\left(\frac{C}{1-p_g\left(X\right)}\right)^2\dot{p}_g\left(X\right)\right]'\left(\hat{\pi}_g-\pi_g^0\right)+o_p\left(n^{-1/2}\right). \tag{A.12}$$

Then, (A.9), (A.10), (A.12) and Lemma A.2 yield (A.8), concluding the proof.□

Proof of Theorem 3: Note that, by the conditional multiplier central limit theorem, see Lemma 2.9.5 in van der Vaart and Wellner (1996), as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\Psi_{g\leq t}(\mathcal{W}_i)\xrightarrow{d}N(0,\Sigma), \tag{A.13}$$

where $\Sigma=\mathbb{E}[\Psi_{g\leq t}(\mathcal{W})\Psi_{g\leq t}(\mathcal{W})']$. Thus, to conclude the proof that

$$\sqrt{n}\left(\widehat{ATT}^*_{g\leq t}-\widehat{ATT}_{g\leq t}\right)\underset{*}{\xrightarrow{d}}N(0,\Sigma),$$

it suffices to show that, for all $2\leq g\leq t\leq\mathcal{T}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\widehat{\psi}_{gt}(\mathcal{W}_i)-\psi_{gt}(\mathcal{W}_i)\right]=o_{p^*}\left(1\right).$$

Towards this, note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\widehat{\psi}_{gt}(\mathcal{W}_i)-\psi_{gt}(\mathcal{W}_i)\right]=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\widehat{\psi}^G_{gt}(\mathcal{W}_i)-\psi^G_{gt}(\mathcal{W}_i)\right] \tag{A.14}$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\cdot\left[\widehat{\psi}^C_{gt}(\mathcal{W}_i)-\psi^C_{gt}(\mathcal{W}_i)\right],$$

where

$$\widehat{\psi}^G_{gt}(\mathcal{W})=\frac{G_g}{\mathbb{E}_n\left[G_g\right]}\left[\left(Y_t-Y_{g-1}\right)-\widehat{ATT}_g(g,t)\right],$$

and

$$\widehat{\psi}^{C}_{gt}(\mathcal{W}) = \frac{w\left(\hat{p}_g\right)}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}\left[\left(Y_{it}-Y_{ig-1}\right)-\widehat{ATT}_C(g,t)\right] + \widehat{M}_{gt}'\,\widehat{\xi}^{\pi}_{g}\left(\mathcal{W}\right),$$

with

$$\begin{aligned}
w\left(\hat{p}_g\right) &= \frac{\hat{p}_g\left(X\right)C}{1-\hat{p}_g\left(X\right)},\\
\widehat{M}_{gt} &= \frac{\mathbb{E}_n\left[X\left(\dfrac{C}{1-\hat{p}_g\left(X\right)}\right)^2 \dot{\hat{p}}_g\left(X\right)\left[\left(Y_{it}-Y_{ig-1}\right)-\widehat{ATT}_g(g,t)\right]\right]}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]},\\
\widehat{\xi}^{\pi}_{g}\left(\mathcal{W}\right) &= \mathbb{E}_n\left[\frac{\left(G_g+C\right)\dot{\hat{p}}_g\left(X\right)^2}{\hat{p}_g\left(X\right)\left(1-\hat{p}_g\left(X\right)\right)}XX'\right]^{-1} X\frac{\left(G_g+C\right)\left(G_g-\hat{p}_g\left(X\right)\right)\dot{\hat{p}}_g\left(X\right)}{\hat{p}_g\left(X\right)\left(1-\hat{p}_g\left(X\right)\right)}.
\end{aligned}$$

We will show that each term in (A.14) is $o_{p^*}(1)$. For the first term in (A.14), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot\left[\widehat{\psi}^{G}_{gt}(\mathcal{W}_i)-\psi^{G}_{gt}(\mathcal{W}_i)\right]\\
&= \left[\frac{1}{\mathbb{E}_n\left[G_g\right]}-\frac{1}{\mathbb{E}\left[G_g\right]}\right]\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot G_{ig}\left(Y_{it}-Y_{ig-1}\right)\\
&- \left[\widehat{ATT}_g(g,t)-ATT_g(g,t)\right]\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot G_{ig},\\
&= o_{p^*}\left(1\right), \qquad\qquad (A.15)
\end{aligned}$$

where the last equality follows from the results in Theorem 1, together with the law of large numbers, continuous mapping theorem, and Lemma 2.9.5 in van der Vaart and Wellner (1996).

For the second term in (A.14), we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot\left[\widehat{\psi}^{C}_{gt}(\mathcal{W}_i)-\psi^{C}_{gt}(\mathcal{W}_i)\right]\\
&= \frac{1}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot\left(w_i\left(\hat{p}_g\right)-w_i\left(p_g\right)\right)\left(Y_{it}-Y_{ig-1}\right)\\
&+ \left(\frac{1}{\mathbb{E}_n\left[w\left(\hat{p}_g\right)\right]}-\frac{1}{\mathbb{E}\left[w\left(p_g\right)\right]}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot w_i\left(p_g\right)\left(Y_{it}-Y_{ig-1}\right)\\
&+ \left(\widehat{M}_{gt}-M_{gt}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot \xi^{\pi}_{g}\left(\mathcal{W}_i\right).\\
&+ \widehat{M}_{gt}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i\cdot\left(\widehat{\xi}^{\pi}_{g}\left(\mathcal{W}_i\right)-\xi^{\pi}_{g}\left(\mathcal{W}_i\right)\right)\\
&:= A_{1n}+A_{2n}+A_{3n}+A_{4n}.
\end{aligned}$$

From Lemma A.3, we have that $\mathcal{H}_{1,g}$, $\mathcal{H}_{2,g}$, $\mathcal{H}_{3,g}$ and $\mathcal{H}_{5,g}$ are Donsker, and by Assumption 5, $\mathbb{E}\left[w\left(p_g\right)\right]$ it is bounded away from zero. Thus, by a stochastic equicontinuity argument, Glivenko-Cantelli's theorem, continuous mapping theorem, and Theorem 2.9.6 in van der Vaart and Wellner (1996),

$$A_{1n} = o_{p^*}\left(1\right), \; A_{2n} = o_{p^*}\left(1\right), \; A_{3n} = o_{p^*}\left(1\right), \text{ and } A_{4n} = o_{p^*}\left(1\right),$$

implying that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i \cdot \left[\widehat{\psi}_{gt}^{C}(\mathcal{W}_i) - \psi_{gt}^{C}(\mathcal{W}_i)\right] = o_{p^*}\left(1\right). \tag{A.16}$$

From (A.13)-(A.16), it follows that

$$\sqrt{n}\left(\widehat{ATT}^{*}_{g\leq t} - \widehat{ATT}_{g\leq t}\right) \overset{d}{\underset{*}{\to}} N(0,\Sigma).$$

Finally, by the continuous mapping theorem, see e.g. Theorem 10.8 in Kosorok (2008), for any continuous functional $\Gamma(\cdot)$

$$\Gamma\left(\sqrt{n}\left(\widehat{ATT}^{*}_{g\leq t} - \widehat{ATT}_{g\leq t}\right)\right) \overset{d}{\underset{*}{\to}} \Gamma\left(N(0,V)\right),$$

concluding our proof. □

Proof of Theorem 4: In order to prove the first part of Theorem 4, we first show that, under H_0, for all $2 \leq t < g \leq \mathcal{T}$,

$$\widehat{J}(u,g,t,\hat{p}_g) = \mathbb{E}_n\left[\psi_{ugt}^{test}(\mathcal{W}_i)\right] + o_p\left(n^{-1/2}\right),$$

Towards this end, we write

$$\begin{aligned}\widehat{J}(u,g,t,\hat{p}_g) &= \mathbb{E}_n\left[\frac{G_g}{\mathbb{E}_n\left[G_g\right]}1(X\leq u)\left(Y_t - Y_{t-1}\right)\right] \\ &\quad - \mathbb{E}_n\left[\frac{\dfrac{\hat{p}_g\left(X\right)C}{1-\hat{p}_g\left(X\right)}}{\mathbb{E}_n\left[\dfrac{\hat{p}_g\left(X\right)C}{1-\hat{p}_g\left(X\right)}\right]}1(X\leq u)\left(Y_t - Y_{t-1}\right)\right] \\ &:= \widehat{J}_G(u,g,t,\hat{p}_g) - \widehat{J}_C(u,g,t,\hat{p}_g),\end{aligned}$$

and analyze each term separately.

As in the proof of Theorem 1, let $\beta_g = \mathbb{E}\left[G_g\right]$ and $\widehat{\beta}_g = \mathbb{E}_n\left[G_g\right]$. Applying a classical mean value theorem argument, uniformly in $u \in \mathcal{X}$,

$$\begin{aligned}\widehat{J}_G(u,g,t,\hat{p}_g) &= \mathbb{E}_n\left[\frac{G_g}{\beta_g}1(X\leq u)\left(Y_t - Y_{t-1}\right)\right] \\ &\quad - \frac{\mathbb{E}_n\left[G_g 1(X\leq u)\left(Y_t - Y_{t-1}\right)\right]}{\bar{\beta}_g^2} \cdot \mathbb{E}_n\left[G_g - \mathbb{E}\left[G_g\right]\right].\end{aligned}$$

where $\bar{\beta}_g$ is an intermediate point that satisfies $\left|\bar{\beta}_g - \beta_g\right| \leq \left|\widehat{\beta}_g - \beta_g\right|$ *a.s.*. Define the class of functions

$$\mathcal{H}_{6,g} = \left\{(x, g_g, y_t, y_{t-1}) \mapsto g_g\left(y_t - y_{t-1}\right) 1\left\{x \leq u\right\} : u \in \mathcal{X}\right\}.$$

By Example 19.11 in van der Vaart (1998), $\mathcal{H}_{6,g}$ is Donsker under Assumption 5. Furthermore,

$$\mathbb{E}_n\left[G_g - \mathbb{E}\left[G_g\right]\right] = O_p\left(n^{-1/2}\right).$$

Thus, by the Glivenko-Cantelli's theorem and the continuous mapping theorem, uniformly in $u \in \mathcal{X}$,

$$\begin{aligned}\widehat{J}_G(u, g, t, \hat{p}_g) &= \mathbb{E}_n\left[\frac{G_g}{\mathbb{E}\left[G_g\right]} 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right] \\ &\quad - \frac{J_G(u, g, t, p_g)}{\mathbb{E}\left[G_g\right]} \cdot \mathbb{E}_n\left[G_g - \mathbb{E}\left[G_g\right]\right] + o_p\left(n^{-1/2}\right) \\ &= \mathbb{E}_n\left[w_g^G\left(\left(Y_t - Y_{t-1}\right) 1(X \leq u) - \mathbb{E}\left[w_g^G 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right]\right)\right] + J_G(u, g, t, p_g) \\ &\quad + o_p\left(n^{-1/2}\right),\end{aligned} \tag{A.17}$$

where

$$J_G(u, g, t, p_g) = \mathbb{E}\left[\frac{G_g}{\mathbb{E}\left[G_g\right]} 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right].$$

We analyze $\widehat{J}_C(u, g, t, \hat{p}_g)$ next. Applying a classical mean value theorem argument, uniformly in $u \in \mathcal{X}$,

$$\begin{aligned}\widehat{J}_C(u, g, t, \hat{p}_g) &= \widehat{J}_C(u, g, t, p_g) \\ &\quad + \frac{\mathbb{E}_n\left[X \dfrac{C\, \dot{p}_g\left(X; \bar{\pi}_g\right)}{\left(1 - p_g\left(X; \bar{\pi}_g\right)\right)^2} 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right]'}{\mathbb{E}_n\left[\dfrac{p_g\left(X; \bar{\pi}_g\right) C}{1 - p_g\left(X; \bar{\pi}_g\right)}\right]}\left(\hat{\pi}_g - \pi_g^0\right) \\ &\quad - \frac{\mathbb{E}_n\left[X \dfrac{C\, \dot{p}_g\left(X; \bar{\pi}_g\right)}{\left(1 - p_g\left(X; \bar{\pi}_g\right)\right)^2}\right]'}{\mathbb{E}_n\left[\dfrac{p_g\left(X; \bar{\pi}_g\right) C}{1 - p_g\left(X; \bar{\pi}_g\right)}\right]} \frac{\mathbb{E}_n\left[\dfrac{p_g\left(X; \bar{\pi}_g\right) C}{1 - p_g\left(X; \bar{\pi}_g\right)} 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right]}{\mathbb{E}_n\left[\dfrac{p_g\left(X; \bar{\pi}_g\right) C}{1 - p_g\left(X; \bar{\pi}_g\right)}\right]}\left(\hat{\pi}_g - \pi_g^0\right)\end{aligned}$$

where $\bar{\pi}$ is an intermediate point that satisfies $\left|\bar{\pi}_g - \pi_g^0\right| \leq \left|\hat{\pi}_g - \pi_g^0\right|$ *a.s*, and

$$\widehat{J}_C(u, g, t, p_g) = \frac{\mathbb{E}_n\left[\dfrac{p_g\left(X\right) C}{1 - p_g\left(X\right)} 1(X \leq u)\left(Y_t - Y_{t-1}\right)\right]}{\mathbb{E}_n\left[\dfrac{p_g\left(X\right) C}{1 - p_g\left(X\right)}\right]}.$$

Define the classes of functions

$$\mathcal{H}_{7,g} = \left\{ (x, c, y_t, y_{t-1}) \mapsto \frac{p_g(x;\pi)}{1 - p_g(x;\pi)} c (y_t - y_{t-1}) 1\{x \le u\} : \pi \in \Pi_g, u \in \mathcal{X} \right\},$$

$$\mathcal{H}_{8,g} = \left\{ (x, c, y_t, y_{t-1}) \mapsto x \frac{\dot{p}_g(x;\pi) c (y_t - y_{t-1}) 1\{x \le u\}}{(1 - p_g(x;\pi))^2} : \pi \in \Pi_g, u \in \mathcal{X} \right\},$$

$$\mathcal{H}_{9,g} = \left\{ (x, c) \mapsto \frac{c p_g(x;\pi)}{1 - p_g(x;\pi)} : \pi \in \Pi_g, \right\},$$

$$\mathcal{H}_{10,g} = \left\{ (x, c) \mapsto x \frac{\dot{p}_g(x;\pi) c}{(1 - p_g(x;\pi))^2} : \pi \in \Pi_g \right\}.$$

By Examples 19.7, 19.11, and 19.20 in van der Vaart (1998), all these classes of functions are Donsker under Assumption 5. Thus, by the Glivenko-Cantelli's theorem, continuous mapping theorem, and Lemma A.2, uniformly in $u \in \mathcal{X}$,

$$\widehat{J}_C(u, g, t, \hat{p}_g) = \widehat{J}_C(u, g, t, p_g) + M_{ugt}^{test\,\prime} \left(\hat{\pi}_g - \pi_g^0 \right) + o_p\left(n^{-1/2}\right), \tag{A.18}$$

for every g, t.

Denote

$$\hat{\beta}_g^C = \mathbb{E}_n\left[\frac{p_g(X) C}{1 - p_g(X)}\right], \ \beta_g^C = \mathbb{E}\left[\frac{p_g(X) C}{1 - p_g(X)}\right].$$

Applying a classical mean value theorem argument, we have

$$\widehat{J}_C(u, g, t, p_g) = \frac{\mathbb{E}_n\left[\frac{p_g(X) C}{1 - p_g(X)} 1(X \le u) (Y_t - Y_{t-1})\right]}{\mathbb{E}\left[\frac{p_g(X) C}{1 - p_g(X)}\right]} - \frac{\mathbb{E}_n\left[\frac{p_g(X) C}{1 - p_g(X)} 1(X \le u) (Y_t - Y_{t-1})\right]}{\left(\bar{\beta}_g^C\right)^2} \cdot \mathbb{E}_n\left[\frac{p_g(X) C}{1 - p_g(X)} - \mathbb{E}\left[\frac{p_g(X) C}{1 - p_g(X)}\right]\right]$$

where $\bar{\beta}_g^C$ is an intermediate point that satisfies $\left|\bar{\beta}_g^C - \beta_g^C\right| \le \left|\widehat{\beta}_g^C - \beta_g^C\right|$ *a.s.*. Since $\mathcal{H}_{7,g}$ is a Donsker Class of functions and

$$\mathbb{E}_n\left[\frac{p_g(X) C}{1 - p_g(X)} - \mathbb{E}\left[\frac{p_g(X) C}{1 - p_g(X)}\right]\right] = O_p\left(n^{-1/2}\right),$$

we have that, by the Glivenko-Cantelli's theorem and the continuous mapping theorem, uniformly in $u \in \mathcal{X}$,

$$\widehat{J}_C(u, g, t, p_g) = \mathbb{E}_n\left[w_g^C \ (Y_t - Y_{t-1}) 1(X \le u)\right]$$

$$
\begin{aligned}
&- \frac{\mathbb{E}\left[w_g^C\ \left(Y_t - Y_{t-1}\right) 1(X \le u)\right]}{\mathbb{E}\left[\dfrac{p_g\left(X\right)C}{1-p_g\left(X\right)}\right]} \cdot \mathbb{E}_n\left[\frac{p_g\left(X\right)C}{1-p_g\left(X\right)} - \mathbb{E}\left[\frac{p_g\left(X\right)C}{1-p_g\left(X\right)}\right]\right] + o_p\left(n^{-1/2}\right) \\
&= \mathbb{E}_n\left[w_g^C\ \left(\left(Y_t - Y_{t-1}\right) 1(X \le u) - \mathbb{E}\left[w_g^C 1(X \le u)\left(Y_t - Y_{t-1}\right)\right]\right)\right] + J_C(u,g,t,p_g) \qquad \text{(A.19)} \\
&+ o_p\left(n^{-1/2}\right).
\end{aligned}
$$

Hence, from (A.17), (A.18), (A.19), and the asymptotic linear representation of $\left(\hat{\pi}_g - \pi_g^0\right)$ in Lemma A.2, for every g, t,

$$
\widehat{J}(u,g,t,\hat{p}_g) = \mathbb{E}_n\left[\psi_{ugt}^{test}(\mathcal{W})\right] + \left(J_G(u,g,t,p_g) - J_C(u,g,t,p_g)\right) + o_p\left(n^{-1/2}\right) \qquad \text{(A.20)}
$$

By noticing that under H_0, $J_G(u,g,t,p_g) = J_C(u,g,t,p_g)$ for all $u \in \mathcal{X}$, (g,t) such that $2 \le t < g \le \mathcal{T}$, we have that, under H_0, uniformly in $u \in \mathcal{X}$, for all $2 \le t < g \le \mathcal{T}$

$$
\widehat{J}(u,g,t,\hat{p}_g) = \mathbb{E}_n\left[\psi_{ugt}^{test}(\mathcal{W}_i)\right] + o_p\left(n^{-1/2}\right).
$$

In order to show that $\sqrt{n}\widehat{J}_{g>t}\left(u\right) \Rightarrow \mathbb{G}(u)\ in\ l^{\infty}\left(\mathcal{X}\right)$, it suffices to show that the class of functions

$$
\mathcal{H}_{10} = \left\{\left(x, g_g, c, y_t, y_{t-1}\right) \mapsto \psi_{ugt}^{test} : u \in \mathcal{X},\ 2 \le t < g \le \mathcal{T}\right\}
$$

is Donsker. This follows straightforwardly from the previously discussed Donsker results and Example 19.20 in van der Vaart (1998). Finally,

$$
CvM_n \xrightarrow{d} \int_{\mathcal{X}} \left|\mathbb{G}(u)\right|_M^2 F_X\left(du\right)
$$

follows from the continuous mapping theorem,

$$
\sup_{u \in \mathcal{X}} \left|F_{n,X}\left(u\right) - F_X\left(u\right)\right| = o_{a.s.}\left(1\right),
$$

and the Helly-Bray Theorem.

Next, we study the behavior of CvM_n under H_1. First, note that under H_1, for some $u \in \mathcal{X}$, and some (g,t), $2 \le t < g \le \mathcal{T}$,

$$
J(u,g,t,p_g) \neq 0.
$$

Thus, from (A.20), under H_1, uniformly in $u \in \mathcal{X}$,

$$
\sqrt{n}\widehat{J}_{g>t}\left(u\right) = O_p\left(n^{1/2}\right),
$$

implying that CvM_n diverges to infinity under H_1. Because $c_{\alpha}^{CvM} = O\left(1\right)$ a.s., as $n \to \infty$,

$$
P\left(CvM_n > c_{\alpha}^{CvM}\right) \to 1,
$$

concluding the proof of Theorem 4. □

Proof of Theorem 5: In the proof of Theorem 4, we have shown that

$$\mathcal{H}_{11} = \left\{ (x, g_g, c, y_t, y_{t-1}) \mapsto \psi_{ugt}^{test} : u \in \mathcal{X},\ 2 \leq t < g \leq \mathcal{T} \right\}$$

is a Donsker class of functions. Then, by the conditional multiplier functional central limit theorem, see Theorem 2.9.6, in van der Vaart and Wellner (1996), as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \cdot \Psi_{g>t}^{test}(\mathcal{W}_i) \underset{*}{\Rightarrow} \mathbb{G}(u)\ in\ l^{\infty}\left(\mathcal{X}\right),$$

where $\mathbb{G}(u)\ in\ l^{\infty}\left(\mathcal{X}\right)$ is the same Gaussian process of Theorem 4 and $\underset{*}{\Rightarrow}$ indicates weak convergence in probability under the bootstrap law. Thus, to conclude the proof it suffices to show that, for all $2 \leq t < g \leq \mathcal{T}$, uniformly in $u \in \mathcal{X}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \cdot \left[\widehat{\psi}_{ugt}^{test}(\mathcal{W}_i) - \psi_{ugt}^{test}(\mathcal{W}_i) \right] = o_{p^*}\left(1\right). \tag{A.21}$$

The proof of (A.21) follows exactly the same steps as the proof of Theorem 3, and is therefore omitted.□

Appendix B: Additional Results for Repeated Cross Sections

In this section we extend our results to the case with repeated cross sections data instead of panel data. Here we assume that for each individual in the pooled sample, we observe $(Y, G_1, \ldots, G_{\mathcal{T}}, C, T, X)$ where $T \in \{1, \ldots, \mathcal{T}\}$ denotes the time period when that individual is observed. Let $T_t = 1$ if an observation is observed at time t, and zero otherwise.

We assume that random samples are available for each time period.

Assumption B.1. *Conditional of $T = t$, the data are independent and identically distributed from the distribution of $(Y_t, G_1, \ldots, G_{\mathcal{T}}, C, X)$, for all $t = 1, \ldots, \mathcal{T}$.*

Assumption B.1 implies that our sample consists of random draws from the mixture distribution

$$F_M(y, g_1, \ldots, g_{\mathcal{T}}, c, t, x) = \sum_{t=1}^{\mathcal{T}} \lambda_t \cdot F_{Y,G_1,\ldots,G_{\mathcal{T}},C,X|T}(y, g_1, \ldots, g_{\mathcal{T}}, c, x|t),$$

where $\lambda_t = P(T_t = 1)$. Notice that, once one conditions on the time period, then expectations under the mixture distribution correspond to population expectations. Also, because X, G_g, and C are observed for all individuals, one can use draws from the mixture distribution to estimate the generalized propensity score. With some abuse of notation, we then use $p_g(X)$ as a short notation for $\mathbb{E}_M[G_g|X, G_g + C = 1]$, where $\mathbb{E}_M[\cdot]$ denotes expectations with respect to $F_M(\cdot)$.

Define the stabilized weights

$$w_{treat}(a,b) = T_b \cdot G_a / \mathbb{E}_M[T_b \cdot G_a],$$
$$w_{cont}(a,b) = \frac{T_b \cdot p_a(X) C}{1 - p_a(X)} \Big/ \mathbb{E}_M\left[\frac{T_b \cdot p_a(X) C}{1 - p_a(X)}\right],$$

where $a, b = 1, 2, \ldots \mathcal{T}$.

Theorem B.1. *Under Assumption B.1 and Assumptions 2-4 in the main text, for $2 \le g \le t \le \mathcal{T}$, the group-time average treatment effect for group g in period t is nonparametrically identified, and given by*

$$ATT(g,t) = \mathbb{E}_M[(w_{treat}(g,t) - w_{treat}(g,g-1)) \cdot Y] - \mathbb{E}_M[(w_{cont}(g,t) - w_{cont}(g,g-1)) \cdot Y].$$

Proof of Theorem B.1: By the law of iterated expectations, Assumption B.1 and Assumption 3 in the main text, for all $2 \le g \le t \le \mathcal{T}$,

$$\begin{aligned}\mathbb{E}_M[w_{treat}(g,t) \cdot Y] &= \frac{\mathbb{E}_M[T_t G_g \cdot Y]}{\mathbb{E}_M[T_t G_g]} \\ &= \frac{\mathbb{E}[G_g \cdot Y | T_t = 1]}{\mathbb{E}[G_g | T_t = 1]} \\ &= \mathbb{E}[Y | T_t = 1, G_g = 1] \\ &= \mathbb{E}[Y_t(1) | G_g = 1].\end{aligned}$$

To complete the proof of Theorem B.1, we must show that

$$\mathbb{E}_M[(w_{treat}(g, g-1,) + w_{cont}(g,t) - w_{cont}(g,g-1)) \cdot Y] = \mathbb{E}[Y_t(0) | G_g = 1]. \quad \text{(B.1)}$$

Towards this, from Assumption B.1 and proceeding as in Lemma A.1, we get

$$\begin{aligned}\mathbb{E}[Y_t(0)|X, G_g = 1] =& \mathbb{E}[Y(0)|X, G_g = 1, T_t = 1] \\ &= \mathbb{E}[Y|X, G_g = 1, T_{g-1} = 1] \\ &+ \mathbb{E}[Y|X, C = 1, T_t = 1] - \mathbb{E}[Y|X, C = 1, T_{g-1} = 1].\end{aligned} \quad \text{(B.2)}$$

From the above result, it follows that

$$\begin{aligned}\mathbb{E}[Y_t(0)|X, G_g = 1] =& \mathbb{E}[\mathbb{E}[Y|X, G_g = 1, T_{g-1} = 1] | G_g = 1, T_{g-1} = 1] \\ &+ \mathbb{E}[\mathbb{E}[Y|X, C = 1, T_t = 1] | G_g = 1, T_t = 1]\end{aligned} \quad \text{(B.3)}$$

$$- \mathbb{E}\left[\mathbb{E}[Y|X, C=1, T_{g-1}=1]\middle| G_g=1, T_{g-1}=1\right].$$

We consider each term separately. For the first term of (B.3),

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[Y|X, G_g=1, T_{g-1}=1]\middle| G_g=1, T_{g-1}=1\right] &= \mathbb{E}[Y|G_g=1, T_{g-1}=1] \\ &= \mathbb{E}_M\left[w_{treat}\left(t, g\right) \cdot Y\right]. \end{aligned} \tag{B.4}$$

Let $\mathbb{E}[Y|X, C=1, T_t=1] = A_{C=1,T_t=1}(X)$, and note that, by repeated application of the law of iterated expectations as in the proof of Theorem 1, we have that for the second term of (B.3),

$$\begin{aligned}\mathbb{E}\left[A_{C=1,T_t=1}(X)\middle| G_g=1, T_t=1\right] &= \mathbb{E}\left[G_g\middle| T_t=1\right]^{-1}\mathbb{E}\left[\frac{p_g(X)\,C}{(1-p_g(X))}Y\middle| T_t=1\right] \\ &= \mathbb{E}_M\left[G_g\cdot T_t\right]^{-1}\mathbb{E}_M\left[\frac{T_t\cdot p_g(X)\,C}{(1-p_g(X))}Y\right] \\ &= \mathbb{E}_M\left[w_{cont}\left(g,t\right)\cdot Y\right], \end{aligned} \tag{B.5}$$

where the last equality follows from $p_g(X) := \mathbb{E}_M\left[G_g|X, G_g + C = 1\right]$, and

$$\begin{aligned}\mathbb{E}_M\left[\frac{T_t\cdot p_g(X)\,C}{(1-p_g(X))}\right] &= \mathbb{E}_M\left[T_t\cdot\frac{\mathbb{E}_M\left[G_g|X, C+G_g=1\right]C}{\mathbb{E}_M\left[C|X, C+G_g=1\right]}\right] \\ &= \mathbb{E}_M\left[T_t\cdot\frac{\mathbb{E}_M\left[G_g|X\right]C}{\mathbb{E}_M\left[C|X\right]}\right] \\ &= \mathbb{E}_M\left[\frac{\mathbb{E}_M\left[G_g|X\right]\mathbb{E}_M\left[C|X\right]}{\mathbb{E}_M\left[C|X\right]}\right] \\ &= \mathbb{E}_M\left[T_t\cdot\mathbb{E}\left[G_g|X\right]\right] \\ &= \mathbb{E}_M\left[T_t\cdot G_g\right]. \end{aligned}$$

Following analogous steps, we get that, for the third term of (B.3),

$$\mathbb{E}\left[A_{C=1,T_{g-1}=1}(X)\middle| G_g=1, T_{g-1}=1\right] = \mathbb{E}_M\left[w_{cont}\left(g, g-1\right)\cdot Y\right]. \tag{B.6}$$

Then, (B.1) follows by combining (B.4), (B.5) and (B.6). The proof of Theorem B.1 is therefore completed. □

The identification results in Theorem B.1 suggest a simple two-step estimation procedure for the $ATT(g,t)$ with repeated cross-section data. Similar to the panel data case discussed in the main text, we propose to estimate $ATT(g,t)$ by

$$\begin{aligned}\widehat{ATT}(g,t) = \mathbb{E}_n\left[\left(\widehat{w}_{treat}\left(g,t\right) - \widehat{w}_{treat}\left(g, g-1\right)\right)\cdot Y\right] - \\ \mathbb{E}_n\left[\left(\widehat{w}_{cont}\left(g,t;\hat{p}\right) - \widehat{w}_{cont}\left(g, g-1;\hat{p}\right)\right)\cdot Y\right]. \end{aligned}$$

where $\hat{p}_g(\cdot)$ is an estimate of $p_g(\cdot)$, and for $a, b = 1, 2, \ldots \mathcal{T}$,

$$\widehat{w}_{treat}(a,b) = T_b \cdot G_a / \mathbb{E}_n [T_b \cdot G_a],$$
$$\widehat{w}_{cont}(a,b;\hat{p}) = \frac{T_b \cdot \hat{p}_a(X) C}{1-\hat{p}_a(X)} \bigg/ \mathbb{E}_n \left[\frac{T_b \cdot \hat{p}_a(X) C}{1-\hat{p}_a(X)} \right].$$

Next, we show that $\widehat{ATT}(g,t)$ is $\sqrt{n}$-consistent, admits an asymptotically linear representation, and is asymptotically normal. These results are analogous to Theorem 2 in the main text. Let $ATT_{g\le t}$ and $\widehat{ATT}_{g\le t}$ denote the vector of $ATT(g,t)$ and $\widehat{ATT}(g,t)$, respectively, for all $g = 2, \ldots, \mathcal{T}$ and $t = 2, \ldots, \mathcal{T}$ with $g \le t$. Define

$$\psi_{g,t}^{rc}(\mathcal{W}_i) = \left(\psi_{g,t}^{rc,G}(\mathcal{W}_i) - \psi_{g,g-1}^{rc,G}(\mathcal{W}_i)\right) - \left(\psi_{g,t}^{rc,C}(\mathcal{W}_i) - \psi_{g,g-1}^{rc,C}(\mathcal{W}_i)\right),$$

where, for $g, t = 1, 2, \ldots \mathcal{T}$,

$$\psi_{g,t}^{rc,G}(\mathcal{W}) = w_{treat}(g,t) \;\; [Y - \mathbb{E}_M [w_{treat}(g,t) \cdot Y]],$$
$$\psi_{g,t}^{rc,C}(\mathcal{W}) = w_{cont}(g,t) \;\; [Y - \mathbb{E}_M [w_{cont}(g,t) \cdot Y]] + M_{g,t}^{rc\,\prime} \, \xi_g^{\pi}(\mathcal{W}),$$

and

$$M_{g,t}^{rc} = \frac{\mathbb{E}_M \left[X \left(\frac{T_t \cdot C}{1 - p_g(X)} \right)^2 \dot{p}_g(X) \cdot [Y - \mathbb{E}[w_{cont}(g,t) \cdot Y]] \right]}{\mathbb{E}_M \left[\frac{T_t \cdot p_g(X) C}{1 - p_g(X)} \right]},$$

which is a $k \times 1$ vector, with k the dimension of X, and $\xi_g^{\pi}(\mathcal{W})$ is as defined in (3.1) in the main text. Finally, let $\Psi_{g\le t}^{rc}$ denote the collection of $\psi_{g,t}^{rc}$ across all periods t and groups g such that $g \le t$.

Theorem B.2. *Under Assumption B.1 and Assumptions 2-5 in the main text, for* $2 \le g \le t \le \mathcal{T}$,

$$\sqrt{n}(\widehat{ATT}(g,t) - ATT(g,t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{g,t}^{rc}(\mathcal{W}_i) + o_p(1).$$

Furthermore,

$$\sqrt{n}(\widehat{ATT}_{g\le t} - ATT_{g\le t}) \xrightarrow{d} N(0, \Sigma^{rc})$$

where $\Sigma^{rc} = \mathbb{E}_M[\Psi_{g\le t}^{rc}(\mathcal{W})\Psi_{g\le t}^{rc}(\mathcal{W})']$.

Proof of Theorem B.2: The proof of Theorem B.2 follows the same steps as the Proof of Theorem 2. From Theorem B.1, for each $2 \le g \le t \le \mathcal{T}$ we can write

$$\sqrt{n}(\widehat{ATT}(g,t) - ATT(g,t))$$
$$= \sqrt{n} \left(\mathbb{E}_n [\widehat{w}_{treat}(g,t) \cdot Y] - \mathbb{E}_M [w_{treat}(g,t) \cdot Y] \right)$$

$$
\begin{aligned}
&- \sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{treat}\left(g, g-1\right) \cdot Y\right] - \mathbb{E}_M\left[w_{treat}\left(g, g-1\right) \cdot Y\right]\right) \\
&- \sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}\left(g, t; \hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]\right) \\
&+ \sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}\left(g, g-1; \hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{cont}\left(g, g-1\right) \cdot Y\right]\right).
\end{aligned} \tag{B.7}
$$

We analyze each term separately. First, note that, for each $2 \leq g \leq t \leq \mathcal{T}$,

$$
\sqrt{n}\left(\mathbb{E}_n\left[T_t \cdot G_g\right] - \mathbb{E}_M\left[T_t \cdot G_g\right]\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(T_{it} \cdot G_{ig} - \mathbb{E}\left[T_t \cdot G_g\right]\right).
$$

Then, by the continuous mapping theorem,

$$
\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{treat}\left(g, t\right) \cdot Y\right] - \mathbb{E}_M\left[w_{treat}\left(g, t\right) \cdot Y\right]\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{g,t}^{rc,G}(\mathcal{W}_i) + o_p\left(1\right). \tag{B.8}
$$

Analogously,

$$
\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{treat}\left(g, g-1\right) \cdot Y\right] - \mathbb{E}_M\left[w_{treat}\left(g, g-1\right) \cdot Y\right]\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{g,g-1}^{rc,G}(\mathcal{W}_i) + o_p\left(1\right). \tag{B.9}
$$

Next we focus on $\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}\left(g, t; \hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]\right)$. To simplify notation, write

$$
w_{a,b}\left(p\right) = \frac{T_b \cdot p_a\left(X\right) C}{1 - p_a\left(X\right)},
$$

and note that $\widehat{w}_{cont}\left(g, t; \hat{p}\right) = w_{g,t}\left(\hat{p}\right) / \mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]$ and $w_{cont}\left(g, t; p\right) = w_{g,t}\left(p\right) / \mathbb{E}_M\left[w_{g,t}\left(p\right)\right]$. Then,

$$
\begin{aligned}
&\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}\left(g, t; \hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]\right) \\
&= \frac{1}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]} \sqrt{n}\left(\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{g,t}\left(p\right) \cdot Y\right]\right) \\
&- \frac{\mathbb{E}_M\left[w_{g,t}\left(p\right) \cdot Y\right]}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right] \mathbb{E}_M\left[w_{g,t}\left(p\right)\right]} \sqrt{n}\left(\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right] - \mathbb{E}_M\left[w_{g,t}\left(p\right)\right]\right) \\
&:= \frac{1}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]} \cdot \sqrt{n} A_{n,\ g,t}^{rc}\left(\hat{p}_g\right) - \frac{\mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]} \cdot \sqrt{n} B_{n,\ g,t}^{rc}\left(\hat{p}_g\right).
\end{aligned}
$$

From Assumption 5, Lemmas A.2 and A.3, and the continuous mapping theorem,

$$
\begin{aligned}
\frac{1}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]} &= \frac{1}{\mathbb{E}_M\left[w_{g,t}\left(p\right)\right]} + o_p\left(1\right), \\
\frac{\mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]}{\mathbb{E}_n\left[w_{g,t}\left(\hat{p}\right)\right]} &= \frac{\mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]}{\mathbb{E}_M\left[w_{g,t}\left(p\right)\right]} + o_p\left(1\right).
\end{aligned}
$$

Thus,

$$
\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}\left(g, t; \hat{p}\right) \cdot Y\right] - \mathbb{E}_M\left[w_{cont}\left(g, t\right) \cdot Y\right]\right)
$$

$$= \frac{1}{\mathbb{E}_M\left[w_{g,t}(p)\right]} \cdot \sqrt{n} A^{rc}_{n,\ g,t}\left(\hat{p}_g\right) - \frac{\mathbb{E}_M\left[w_{cont}(g,t)\cdot Y\right]}{\mathbb{E}_M\left[w_{g,t}(p)\right]} \cdot \sqrt{n} B^{rc}_{n,\ g,t}\left(\hat{p}_g\right) + o_p(1) \quad \text{(B.10)}$$

Applying a classical mean value theorem argument,

$$A^{rc}_{n,\ g,t}\left(\hat{p}_g\right) = \mathbb{E}_n\left[w_{g,t}(p)\cdot Y\right] - \mathbb{E}_M\left[w_{g,t}(p)\cdot Y\right] + \mathbb{E}_n\left[X\left(\frac{T_t\cdot C}{1-p_g\left(X;\bar{\pi}_g\right)}\right)^2 \dot{p}_g\left(X;\bar{\pi}_g\right)\cdot Y\right]'\left(\hat{\pi}_g - \pi^0_g\right),$$

where $\bar{\pi}$ is an intermediate point that satisfies $\left|\bar{\pi}_g - \pi^0_g\right| \leq \left|\hat{\pi}_g - \pi^0_g\right|$ $a.s.$ Thus, by Assumption 5, Lemmas A.2 and A.3, and the Glivenko-Cantelli's theorem,

$$A^{rc}_{n,\ g,t}\left(\hat{p}_g\right) = \mathbb{E}_n\left[w_{g,t}(p)\,Y - \mathbb{E}_M\left[w_{g,t}(p)\,Y\right]\right] + \mathbb{E}_M\left[X\left(\frac{T_t\cdot C}{1-p_g(X)}\right)^2 \dot{p}_g(X)\cdot Y\right]'\left(\hat{\pi}_g - \pi^0_g\right) + o_p\left(n^{-1/2}\right). \quad \text{(B.11)}$$

Analogously,

$$B_n\left(\hat{p}_g\right) = \mathbb{E}_n\left[w_{g,t}(p) - \mathbb{E}_M\left[w_{g,t}(p)\right]\right] + \mathbb{E}_M\left[X\left(\frac{T_t\cdot C}{1-p_g(X)}\right)^2 \dot{p}_g(X)\right]'\left(\hat{\pi}_g - \pi^0_g\right) + o_p\left(n^{-1/2}\right). \quad \text{(B.12)}$$

Combining (B.10), (B.11), (B.12) with Lemma A.2 yield

$$\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}(g,t;\hat{p})\cdot Y\right] - \mathbb{E}_M\left[w_{cont}(g,t)\cdot Y\right]\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi^{rc,C}_{g,t}(\mathcal{W}_i) + o_p(1). \quad \text{(B.13)}$$

Using the same arguments, we conclude that

$$\sqrt{n}\left(\mathbb{E}_n\left[\widehat{w}_{cont}(g,g-1;\hat{p})\cdot Y\right] - \mathbb{E}_M\left[w_{cont}(g,g-1)\cdot Y\right]\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi^{rc,C}_{g,g-1}(\mathcal{W}_i) + o_p(1). \quad \text{(B.14)}$$

Hence, from (B.7), (B.8), (B.9), (B.13) and (B.14), we conclude that, for each $2 \leq g \leq t \leq \mathcal{T}$,

$$\sqrt{n}(\widehat{ATT}(g,t) - ATT(g,t)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi^{rc}_{g,t}(\mathcal{W}_i) + o_p(1).$$

The proof is then completed by applying the multivariate central limit theorem.□

Based on the above results, one can conclude that estimation and inference procedures for $ATT(g,t)$ in the case of repeated cross sections is similar to what we did in the case with panel data. In fact, one simply needs to adjust the weights slightly. In order to conduct asymptotically valid simultaneous inference, one can leverage the asymptotic linear representation in Theorem B.2, and use a multiplier bootstrap procedure analogous to the one in Theorem 3. The proof of the bootstrap validity in the repeated cross section case follows exactly the same steps as in Theorem 3 and is therefore omitted.

Appendix C: Analysis with "Not yet Treated" as a Control Group

In this appendix, we discuss the case where one considers the "not yet treated" instead of the "never treated" as a control group. This case is particularly relevant in applications when eventually (almost) all units are treated, though the timing of the treatment differs across groups. To carry out this analysis, we make the following assumptions.

Assumption C.1. $\{Y_{i1}, Y_{i2}, \ldots Y_{i\mathcal{T}}, X_i, D_{i1}, D_{i2}, \ldots, D_{i\mathcal{T}}\}_{i=1}^{n}$ *is independent and identically distributed (iid).*

Assumption C.2. *For all* $t = 2, \ldots, \mathcal{T}$, $g = 2, \ldots, \mathcal{T}$ *such that* $g \leq t$,

$$\mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = \mathbb{E}[Y_t(0) - Y_{t-1}(0)|X, D_t = 0] \ a.s..$$

Assumption C.3. *For* $t = 2, \ldots, \mathcal{T}$,

$$D_{t-1} = 1 \text{ implies that } D_t = 1$$

Assumption C.4. *For all* $t = 2, \ldots, \mathcal{T}$, $g = 2, \ldots, \mathcal{T}$, $P(G_g = 1) > 0$ *and* $P(D_t = 1|X) < 1$ *a.s..*

Assumptions C.1 and C.3 are the same as Assumptions 1 and 3 in the main text. Assumptions C.2 and C.4 are the analogues of Assumptions 2 and 4, but using those "not yet treated" ($D_t = 0$) as a control group instead of the "never treated" ($C = 0$ or $D_{\mathcal{T}} = 0$). Note that Assumption C.4 rules out the case in which eventually everyone is treated; in these time periods, there is no "control group" available, and therefore the data itself is not informative about the average treatment effect when $D_t = 1$ *a.s.*. In these cases, one should concentrate their attention only to the time periods such that $P(D_t = 1|X) < 1$ *a.s.*.

It is worth emphasizing that our parallel trend assumption (PTA) C.2 does not restrict pre-trends and is strictly weaker than the PTA imposed by Abraham and Sun (2018), for example.

To see this, note that, using our notation, Abraham and Sun (2018)'s parallel trends assumption (PTA) (Assumption 1 on page 7 of their manuscript dated September 6, 2018) can be written as[1]

Assumption (Parallel trends, Abraham and Sun, 2018): For all $t, s = 1, 2, \ldots, \mathcal{T}$, $g = 2, \ldots, \mathcal{T}$, $\mathbb{E}\left[Y_{i,t}(0) - Y_{i,s}(0)\,|G_g = 1\right] = \mathbb{E}\left[Y_{i,t}(0) - Y_{i,s}(0)\,|C = 1\right] = \mathbb{E}\left[Y_{i,t}(0) - Y_{i,s}(0)\right]$.

Note that Abraham and Sun (2018)'s PTA is unconditional, whereas our is conditional (the unconditional case is a special case of our conditional parallel trends assumption when X only includes an intercept).[2] In addition, their PTA holds for every time period t, s and requires common trends for all groups g. Our PTA, C.2, on the other hand, only holds for the post-treatment $g \leq t$ (i.e., we do not restrict pre-trends), and it only requires parallel trends between the group treated at time g and the "supergroup" formed by averaging all groups not yet treated by time g. Hence, one can clearly see that our PTA is strictly weaker than the one in Abraham and Sun (2018). In fact, we note that the proofs of most propositions in Abraham and Sun (2018) seem to exploit this stronger version of the PTA, implying that this subtle difference is not innocuous. The same caveats apply to de Chaisemartin and D'Haultfœuille (2018) when one specializes their more general treatment selection setup to staggered adoptions designs, though this is not the main focus of their analysis.

Next, we focus our attention on identification. Remember that

$$ATT_X(g,t) = \mathbb{E}[Y_t(1) - Y_t(0)|X, G_g = 1].$$

The next lemma states that, under Assumptions C.1-C.4, we can identify $ATT_X(g,t)$ for $2 \leq g \leq t \leq \mathcal{T}$. This is the analogue of Lemma A.1. In fact, Lemma C.1 extends the $W^*_{TC}(1,0,t)$ estimand discussed in Section 1.2 of the Supplemental Appendix of de Chaisemartin and D'Haultfœuille (2017) in two dimensions: it allows for covariates and distinguishes between treatment adoption time and time since treatment. This latter feature is particularly useful for highlighting treatment effect heterogeneity and studying treatment effect dynamics.

Lemma C.1. *Under Assumptions C.1-C.4, and for* $2 \leq g \leq t \leq \mathcal{T}$*,*

$$ATT_X(g,t) = \mathbb{E}[Y_t - Y_{g-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|X, D_t = 0] \ a.s..$$

Proof of Lemma C.1: In what follows, take all equalities to hold almost surely (a.s.). Notice that for identifying $ATT_X(g,t)$, the key term is $E[Y_t(0)|X, G_g = 1]$. And notice that for $h > s$,

[1]Let E be the time one is treated and $Y^{\infty}_{i,t}$ denote the potential outcome for individual i at time t under no treatment. Then, Assumption 1 of Abraham and Sun (2018) states that $\mathbb{E}\left[Y^{\infty}_{i,t} - Y^{\infty}_{i,s}|E_i = e\right]$ is the same for all $e \in supp\,(E_i)$ and for all s, t and is equal to $\mathbb{E}\left[Y^{\infty}_{i,t} - Y^{\infty}_{i,s}\right]$.

[2]In their Appendix F, Abraham and Sun (2018) also introduce a conditional parallel trend assumption - see their Assumption 7, on page 54. Nonetheless, they do not have any proof of their identification result, and all the other caveats discussed in this paragraph still apply.

$E[Y_s(0)|X, G_s = 1] = E[Y_s|X, G_h = 1]$, which holds because in time periods before an individual is first treated, their untreated potential outcomes are observed outcomes. Also, note that, for $2 \leq g \leq t \leq \mathcal{T}$,

$$\begin{aligned} \mathbb{E}[Y_t(0)|X, G_g = 1] &= \mathbb{E}[\Delta Y_t(0)|X, G_g = 1] + \mathbb{E}[Y_{t-1}(0)|X, G_g = 1] \\ &= \mathbb{E}[\Delta Y_t|X, D_t = 0] + \mathbb{E}[Y_{t-1}(0)|X, G_g = 1], \end{aligned} \tag{C.1}$$

where the first equality holds by adding and subtracting $E[Y_{t-1}(0)|X, G_g = 1]$ and the second equality holds by Assumption C.2. If $g = t-1$, then the last term in the final equation is identified; otherwise, one can continue recursively in similar way to (C.1) but starting with $\mathbb{E}[Y_{t-1}(0)|X, G_g = 1]$. As a result,

$$\begin{aligned} \mathbb{E}[Y_t(0)|X, G_g = 1] &= \sum_{j=0}^{t-g} \mathbb{E}[\Delta Y_{t-j}|X, D_t = 0] + \mathbb{E}[Y_{g-1}|X, G_g = 1] \\ &= \mathbb{E}[Y_t - Y_{g-1}|X, D_t = 0] + \mathbb{E}[Y_{g-1}|X, G_g = 1]. \end{aligned} \tag{C.2}$$

Combining (C.2) with the fact that, for all $g \leq t$, $\mathbb{E}[Y_t(1)|X, G_g = 1] = \mathbb{E}[Y_t|X, G_g = 1]$ (which holds because observed outcomes for group g in period t with $g \leq t$ are treated potential outcomes), implies the result. $\square$

With the result of Lemma C.1 in hand, we proceed to show that $ATT(g,t)$ is nonparametrically identified under Assumptions C.1 - C.4 and for $2 \leq g \leq t \leq \mathcal{T}$. The following theorem is the analogue of Theorem 1. Let $p_{g,t}(X) := P(G_g = 1|X, (G_g = 1 \cup D_t = 0))$ denote the generalized propensity score that uses all the information about those individuals first treated at time g and those not yet treated at time $t \geq g$.

Theorem C.1. *Under Assumptions C.1-C.4 and for $2 \leq g \leq t \leq \mathcal{T}$, the group-time average treatment effect for group g in period t is nonparametrically identified, and given by*

$$ATT(g,t) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\dfrac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}}{\mathbb{E}\left[\dfrac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}\right]}\right)(Y_t - Y_{g-1})\right]. \tag{C.3}$$

Proof of Theorem C.1: Given the result in Lemma C.1,

$$\begin{aligned} ATT(g,t) &= \mathbb{E}[ATT_X(g,t)|G_g = 1] \\ &= \mathbb{E}\left[\mathbb{E}[Y_t - Y_{g-1}|X, G_g = 1] - \mathbb{E}[Y_t - Y_{g-1}|X, D_t = 0]\middle|G_g = 1\right] \\ &:= \mathbb{E}[A_X|G_g = 1] - \mathbb{E}[B_X^{n.yet}|G_g = 1], \end{aligned}$$

and we consider each term separately. For the first term

$$\begin{aligned}\mathbb{E}[A_X|G_g=1] &= \mathbb{E}[Y_t - Y_{g-1}|G_g=1]\\ &= \mathbb{E}\left[\frac{G_g}{\mathbb{E}\left[G_g\right]}(Y_t - Y_{g-1})\right]. \end{aligned} \tag{C.4}$$

For the second term, note that $G_g = 1$ implies that $D_t = 1$ for all $t \geq g$. Then, by repetition of the law of iterated expectations, we have

$$\begin{aligned}\mathbb{E}[B_X^{n.yet}|G_g=1] &= \mathbb{E}\left[\mathbb{E}[Y_t - Y_{g-1}|X, D_t=0]\Big|G_g=1\right]\\ &= \mathbb{E}\left[G_g\mathbb{E}[(1-D_t)\,(Y_t - Y_{g-1})|X, D_t=0]\Big|G_g=1\right]\\ &= \mathbb{E}\left[G_g\mathbb{E}\left[\frac{1-D_t}{1-p_{g,t}(X)}(Y_t - Y_{g-1})\big|X,(G_g=1\cup D_t=0)\right]\Big|G_g=1\right]\\ &= \frac{\mathbb{E}\left[G_g\mathbb{E}\left[\frac{1-D_t}{1-p_{g,t}(X)}(Y_t - Y_{g-1})\big|X,(G_g=1\cup D_t=0)\right]\Big|G_g=1\cup D_t=0\right]}{\mathbb{E}\left[G_g|G_g=1\cup D_t=0\right]}\\ &= \frac{\mathbb{E}\left[\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{1-p_{g,t}(X)}(Y_t - Y_{g-1})\big|X,(G_g=1\cup D_t=0)\right]\Big|G_g=1\cup D_t=0\right]}{\mathbb{E}\left[G_g|G_g=1\cup D_t=0\right]}\\ &= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\Big[\mathbb{E}\left[1\left\{G_g=1\cup D_t=0\right\}|X\right]\cdot\\ &\qquad\qquad\cdot\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{1-p_{g,t}(X)}(Y_t - Y_{g-1})\big|X,(G_g=1\cup D_t=0)\right]\Big]\\ &= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\left[\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{(1-p_{g,t}(X))}(Y_t - Y_{g-1})\big|X\right]\right]\\ &= \mathbb{E}\left[G_g\right]^{-1}\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{(1-p_{g,t}(X))}(Y_t - Y_{g-1})\right]\\ &= \frac{\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{(1-p_{g,t}(X))}(Y_t - Y_{g-1})\right]}{\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{(1-p_{g,t}(X))}\right]}, \end{aligned} \tag{C.5}$$

where (C.5) follows from

$$\begin{aligned}\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{(1-p_{g,t}(X))}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{E}\left[G_g|X,(G_g=1\cup D_t=0)\right](1-D_t)}{\mathbb{E}\left[1-D_t|X,(G_g=1\cup D_t=0)\right]}|X\right]\right]\\ &= \mathbb{E}\left[\frac{\mathbb{E}\left(G_g|X\right)}{\mathbb{E}\left[(1-D_t)\,|X\right]}\left(1-D_t\right)\right]\\ &= \mathbb{E}\left[\frac{\mathbb{E}\left(G_g|X\right)}{\mathbb{E}\left[(1-D_t)\,|X\right]}\mathbb{E}\left[(1-D_t)\,|X\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[G_g|X\right]\right]\end{aligned}$$

$$= \mathbb{E}\left[G_g\right].$$

The proof is completed by combining (C.4) and (C.5). □

Once we have established nonparametric identification of $ATT(g,t)$, we can follow a similar two-step estimation strategy as described in Section 3. More precisely, under Assumptions C.1 - C.4 and for $2 \leq g \leq t \leq \mathcal{T}$, one can estimate $ATT(g,t)$ by

$$\widehat{ATT}_{n.yet}(g,t) = \mathbb{E}_n\left[\left(\frac{G_g}{\mathbb{E}_n\left[G_g\right]} - \frac{\dfrac{\hat{p}_{g,t}\left(X\right)\left(1-D_t\right)}{1-\hat{p}_{g,t}\left(X\right)}}{\mathbb{E}_n\left[\dfrac{\hat{p}_{g,t}\left(X\right)\left(1-D_t\right)}{1-\hat{p}_{g,t}\left(X\right)}\right]}\right)\left(Y_t - Y_{g-1}\right)\right],$$

where $\hat{p}_{g,t}\left(X\right)$ is an estimate of $P\left(G_g = 1 | X, \left(G_g = 1 \cup D_t = 0\right)\right)$. Note that here we need to estimate different propensity scores for different pairs (g,t) periods, whereas the case analyzed in the main text we estimate different propensity scores for different treatment groups g only. From a computational perspective, this is the only difference between $\widehat{ATT}_{n.yet}(g,t)$ and $\widehat{ATT}(g,t)$.

Next we establish the asymptotic properties of $\widehat{ATT}_{n.yet}(g,t)$. These results are very similar to those in Lemma A.2, and Theorems 2 and 3. In what follows, we rely on the following assumption about the propensity score $p_{g,t}\left(X\right)$.

Assumption C.5. *For all $t = 2, \ldots, \mathcal{T}$, $g = 2, \ldots, \mathcal{T}$, $t \geq g$ (i) there exists a known function $\Lambda : \mathbb{R} \to [0,1]$ such that $p_{g,t}(X) = P(G_g = 1 | X, G_g = 1 \cup D_t = 0) = \Lambda(X'\pi^0_{g,t})$; (ii) $\pi^0_{g,t} \in \text{int}(\Pi)$, where Π is a compact subset of $\mathbb{R}^k$; (iii) the support of X, $\mathcal{X}$, is a subset of a compact set S, and $\mathbb{E}[XX' | G_g = 1 \cup D_t = 0]$ is positive definite; (iv) let $\mathcal{U} = \left\{x'\pi : x \in \mathcal{X}, \pi \in \Pi\right\}$; $\forall\ u \in \mathcal{U}$, $\exists \varepsilon > 0$ such that $\Lambda\left(u\right) \in \left[\varepsilon, 1-\varepsilon\right]$, $\Lambda\left(u\right)$ is strictly increasing and twice continuously differentiable with first derivatives bounded away from zero and infinity, and bounded second derivative; (vi) $\mathbb{E}\left[Y_t^2\right] < \infty$ for all $t = 1, \ldots, \mathcal{T}$.*

Assumption C.5 is the analogue of Assumption 5 in the main text. Under Assumption C.5, we can consistently estimate $\pi^0_{g,t}$ using maximum likelihood, i.e.,

$$\hat{\pi}_{g,t} = \arg\max_{\pi} \sum_{i : G_{ig} = 1 \cup D_{it} = 0} G_{ig} \ln\left(p_{g,t}\left(X_i'\pi\right)\right) + \left(1 - G_{ig}\right) \ln\left(1 - p_{g,t}\left(X_i'\pi\right)\right).$$

Let $\dot{p}_{g,t} = \partial p_{g,t}\left(u\right) / \partial u$, and $\dot{p}_{g,t}\left(X\right) = \dot{p}_{g,t}\left(X'\pi^0_{g,t}\right)$.

Lemma C.2. *Under Assumption C.1 and C.5,*

$$\sqrt{n}\left(\hat{\pi}_{g,t} - \pi^0_{g,t}\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi^{\pi}_{g,t}\left(\mathcal{W}_i\right) + o_p\left(1\right),$$

where

$$\xi^{\pi}_{g,t}\left(\mathcal{W}\right)=\mathbb{E}\left[\frac{1\left\{G_g=1\cup D_t=0\right\}\dot{p}_g\left(X\right)^2}{p_{g,t}\left(X\right)\left(1-p_{g,t}\left(X\right)\right)}XX'\right]^{-1}X\frac{1\left\{G_g=1\cup D_t=0\right\}\left(G_g-p_{g,t}\left(X\right)\right)\dot{p}_{g,t}\left(X\right)}{p_{g,t}\left(X\right)\left(1-p_{g,t}\left(X\right)\right)}. \tag{C.6}$$

Proof of Lemma C.2: The proof follows the same steps as the Proof of Lemma A.2 in a routine fashion. Details are omitted for brevity.□

Let $ATT_{g\leq t}$ and $\widehat{ATT}^{n.yet}_{g\leq t}$ denote the vector of $ATT(g,t)$ and $\widehat{ATT}_{n.yet}(g,t)$, respectively, for all $g=2,\dots,\mathcal{T}$ and $t=2,\dots,\mathcal{T}$ with $g\leq t$. Next, we show that $\sqrt{n}\left(\widehat{ATT}^{n.yet}_{g\leq t}-ATT_{g\leq t}\right)$ admits an asymptotically linear representation, and establish its joint limiting distribution. Let

$$w^G_g=\frac{G_g}{\mathbb{E}\left[G_g\right]},\qquad w^{n.yet}_{g,t}=\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{1-p_{g,t}\left(X\right)}\bigg/\mathbb{E}\left[\frac{p_{g,t}\left(X\right)\left(1-D_t\right)}{1-p_{g,t}\left(X\right)}\right],$$

and define

$$\psi^{n.yet}_{g,t}(\mathcal{W}_i)=\psi^G_g(\mathcal{W}_i)-\psi^{cont}_{g,t}(\mathcal{W}_i),$$

where

$$\begin{aligned}\psi^G_g(\mathcal{W})&=w^G_g\;\left[\left(Y_t-Y_{g-1}\right)-\mathbb{E}\left[w^G_g\left(Y_t-Y_{g-1}\right)\right]\right],\\ \psi^{cont}_{g,t}(\mathcal{W})&=w^{n.yet}_{g,t}\;\left[\left(Y_t-Y_{g-1}\right)-\mathbb{E}\left[w^{n.yet}_{g,t}\left(Y_t-Y_{g-1}\right)\right]\right]+M^{n.yet\,\prime}_{g,t}\,\xi^{\pi}_{g,t}\left(\mathcal{W}\right),\end{aligned}$$

and

$$M^{n.yet}_{g,t}=\frac{\mathbb{E}\left[X\left(\dfrac{1-D_t}{1-p_{g,t}\left(X\right)}\right)^2\dot{p}_{g,t}\left(X\right)\left[\left(Y_{it}-Y_{ig-1}\right)-\mathbb{E}\left[w^{n.yet}_{g,t}\left(Y_t-Y_{g-1}\right)\right]\right]\right]}{\mathbb{E}\left[\dfrac{p_{g,t}\left(X\right)\left(1-D_t\right)}{1-p_{g,t}\left(X\right)}\right]}$$

which is a $k\times1$ vector, with k the dimension of X, and $\xi^{\pi}_{g,t}\left(\mathcal{W}\right)$ is as defined in (C.6). Finally, let $\Psi^{n.yet}_{g\leq t}$ denote the collection of $\psi^{n.yet}_{g,t}$ across all periods t and groups g such that $g\leq t$.

Theorem C.2. *Under Assumptions C.1-C.5, for* $2\leq g\leq t\leq\mathcal{T}$,

$$\sqrt{n}\left(\widehat{ATT}^{n.yet}_{g\leq t}-ATT_{g\leq t}\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^{n.yet}_{gt}(\mathcal{W}_i)+o_p(1).$$

Furthermore,

$$\sqrt{n}\left(\widehat{ATT}^{n.yet}_{g\leq t}-ATT_{g\leq t}\right)\xrightarrow{d}N(0,\Sigma_{n.yet})$$

where $\Sigma_{n.yet}=\mathbb{E}[\Psi^{n.yet}_{g\leq t}(\mathcal{W})\Psi^{n.yet}_{g\leq t}(\mathcal{W})']$.

Proof of Theorem C.2: The proof follows the same steps as the proof of Theorem 2 in a routine fashion, where $(1-D_t)$ plays the role of C, $\widehat{\pi}_{g,t}$ plays the role of $\widehat{\pi}_g$, $\xi^{\pi}_{g,t}(\mathcal{W}_i)$ plays the role of $\xi^{\pi}_{g}(\mathcal{W}_i)$, and $M^{n.yet}_{gt}$ plays the role of M_{gt}. We omit the details for brevity.□

Next, we establish the asymptotic validity of a simple multiplier bootstrap procedure. To proceed, let $\widehat{\Psi}^{n.yet}_{g\leq t}(\mathcal{W})$ denote the sample-analogue of $\Psi^{n.yet}_{g\leq t}(\mathcal{W})$, where population expectations are replaced by their empirical analogue, and the true generalized propensity score, $p_{g,t}$, and its derivatives, $\dot{p}_{g,t}$, are replaced by their MLE estimates, $\hat{p}_{g,t}$ and $\widehat{\dot{p}}_{g,t}$, respectively. Let $\{V_i\}_{i=1}^n$ be a sequence of *iid* random variables with zero mean, unit variance and bounded support, independent of the original sample $\{\mathcal{W}_i\}_{i=1}^n$.

We define $\widehat{ATT}^{n.yet,*}_{g\leq t}$, a bootstrap draw of $\widehat{ATT}^{n.yet}_{g\leq t}$, via

$$\widehat{ATT}^{n.yet,*}_{g\leq t} = \widehat{ATT}^{n.yet}_{g\leq t} + \mathbb{E}_n\left[V\cdot\widehat{\Psi}^{n.yet}_{g\leq t}(\mathcal{W})\right].$$

Theorem C.3. *Under Assumptions C.1-C.5,*

$$\sqrt{n}\left(\widehat{ATT}^{*}_{g\leq t} - \widehat{ATT}^{n.yet}_{g\leq t}\right) \overset{d}{\underset{*}{\to}} N(0,\Sigma_{n.yet}),$$

where $\Sigma_{n.yet}$ as in Theorem C.2, and $\overset{d}{\underset{*}{\to}}$ denotes weak convergence (convergence in distribution) of the bootstrap law in probability, i.e., conditional on the original sample $\{\mathcal{W}_i\}_{i=1}^n$. Additionally, for any continuous functional $\Gamma(\cdot)$

$$\Gamma\left(\sqrt{n}\left(\widehat{ATT}^{*}_{g\leq t} - \widehat{ATT}^{n.yet}_{g\leq t}\right)\right) \overset{d}{\underset{*}{\to}} \Gamma\left(N(0,\Sigma_{n.yet})\right).$$

Proof of Theorem C.3: The proof follows exactly the same steps as the proof of Theorem 3 in a routine fashion. We omit the details for brevity.□

Given the results in Theorems C.2 and C.3 and their similarity with Theorems 2 and 3 in the main text, it is clear that one can estimate simultaneous confidence intervals by adapting Algorithm 1 to the current setup in a straightforward manner.

Appendix D: Additional Results for the Case without Covariates

Panel Data

The case where the DID assumption holds without conditioning on covariates is of particular interest. In this appendix, we briefly consider whether or not it is possible to obtain $ATT(g,t)$

using a regression approach like the two period - two group case. A natural starting point is the model

$$Y_{igt} = \alpha_t + c_i + \gamma_{gt} G_{igt} + u_{igt}$$

where α_t is a vector of time period fixed effects (we normalize α_1 to be equal to zero and γ_{g1} to be equal to 1), c_i is time invariant unobserved heterogeneity that can be distributed differently across groups, and G_{igt} is is a dummy variable indicating whether or not individual i is a member group g and the time period is t. Differencing the model across time periods results in

$$\Delta Y_{igt} = \tilde{\alpha}_t + \gamma_{gt} G_{igt} + \Delta u_{igt},$$

where $\tilde{\alpha}_t = \alpha_t - \alpha_{t-1}$. Notice that this is a fully saturated model in group and time effects. It is straightforward to show that

$$\gamma_{gt} = E[\Delta Y_t | G_g = 1] - E[\Delta Y_t | C = 1].$$

When $g = t$, this is exactly the DID estimator. Under the augmented unconditional version of the parallel trends assumption, γ_{gt} should be equal to 0 for all $g > t$, and it is straightforward to test this using output from standard regression software (e.g. Wald test). For $t > g$, the long difference estimate of $ATT(g,t)$ can be constructed by

$$\begin{aligned} ATT(g,t) &= E[Y_t - Y_{g-1} | G_g = 1] - E[Y_t - Y_{g-1} | C = 1] \\ &= \sum_{s=g}^{t} \left(E[\Delta Y_s | G_g = 1] - E[\Delta Y_s | C = 1] \right) \\ &= \sum_{s=g}^{t} \gamma_{gs} \end{aligned}$$

This implies that, under the (augmented) unconditional parallel trends assumption, $ATT(g,t)$ can be recovered using a regression approach. However, combining the estimates of the parameters in this way does not seem to offer much convenience relative to simply computing the estimates directly using the main approach suggested in the paper. Thus, unlike the 2-period case, it does not appear that there is as exact of a mapping from a regression coefficient to a group-time average treatment effect.

Common Approaches to Pre-Testing in the Unconditional Case

Finally in this section, we consider the most common approach to pre-testing the augmented unconditional version of the parallel trends assumption, that is, to run the following regression (see Autor et al. (2007) and Angrist and Pischke (2009)).

$$Y_{it} = \alpha_t + \theta_g + \beta_0 D_{it} + \sum_{j=1}^{q} \beta_j \Delta D_{it,t+j} + u_{it} \tag{D.1}$$

where D_{it} is a dummy variable for whether or not individual i is treated in period t (notice that this is not whether they are *first treated* in period t but whether or not they are treated at all; it is a post-treatment dummy variable), $\Delta D_{it,t+j}$ is a j period lead for individual i who is first treated in period $t+j$. For example, when $t=2$, $\Delta D_{i2,4} = 1$ (for $j=2$) for individuals who are first treated in period 4, which indicates that the group of individuals first treated in period 4 will be treated 2 periods from period t.

Then, one can pre-test the unconditional parallel trends assumption by testing if $\beta_j = 0$ for $j = 1, \ldots, q$. Under the Unconditional DID Assumption, each β_j will be 0. One advantage of this approach is that it allows simple graphs of pre-treatment trends. However, it is possible for this approach to miss departures from the unconditional parallel trends assumption that our test would not miss.

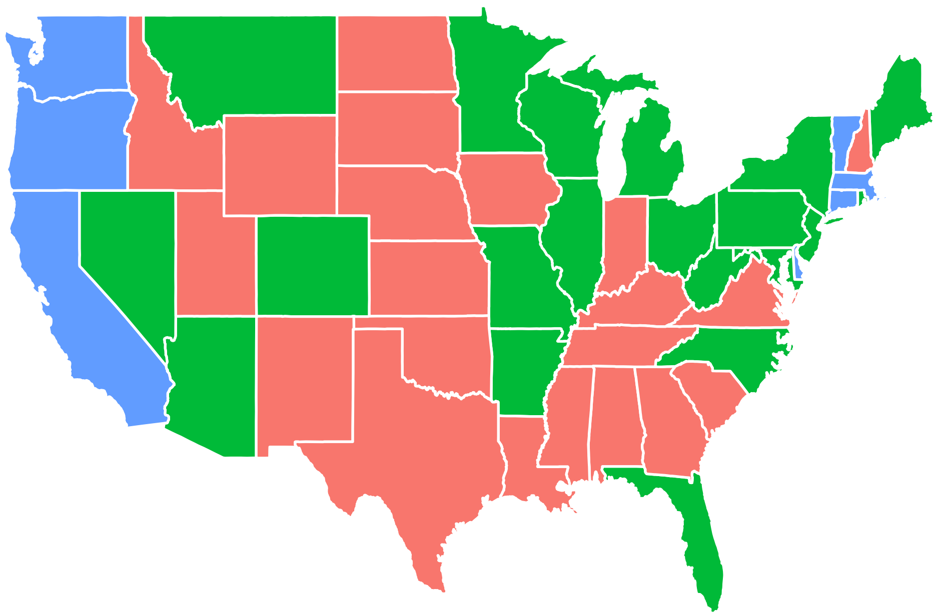
Consider the case with four periods and three groups – the control group, a group first treated in period 4, and a group first treated in period 3. Also, consider the case with $q = 1$. It is easy to show that $\beta_1 = \mathbb{E}[\Delta Y_3 | G_4 = 1] - \mathbb{E}[\Delta Y_3 | C = 1]$ and $\beta_1 = \mathbb{E}[\Delta Y_2 | G_3 = 1] - \mathbb{E}[\Delta Y_1 | C = 1]$ so that the estimate of β_1 will be a weighted average of these two pre-trends. Thus, the unconditional augmented parallel trends assumption could be violated in ways that offset each other leading to β_1 being equal to 0. Even more importantly, the weights associated with the regression coefficient β_1 may not be convex; see Propositions 3 and 7 in Abraham and Sun (2018) for detailed arguments. As a consequence, tests for pre-trends based on (D.1) may not be reliable under treatment effect heterogeneity. Our approach described in Remark 6 in the main text, on the other hand, does not suffer from this potential drawback.

Appendix E: Additional Details about Empirical Application

In our empirical application, we study the effect of the minimum wage on teen employment. Our treatment is if the state-level minimum wage is above the federal-level minimum wage, and we exploit variation in the state-level minimum wage policy changes. Table 1 contains the complete details of the exact date when a state changed its minimum wage as well as which states are used in our analysis.

Dube et al. (2010) argue that differential trends in employment rates across regions bias estimates of the effect of changes in state-level minimum wages. Figure 1 contains the spatial distribution of state-level minimum wage policy changes in our sample. Indeed, Figure 1 shows that states in the Southeast are less likely to increase their minimum wage between 2001 and 2007 than states in the Northeast or Midwest, corroborating the argument in Dube et al. (2010).

Figure 1: The Spatial Distribution of States by Minimum Wage Policy



Notes: Blue states had minimum wages higher than the federal minimum wage in Q1 of 2000. Green states increased their state minimum wage between Q2 of 2000 and Q1 of 2007. Some of these states are omitted from the main dataset either due to missing data or being located in the Northern census region where there are no states that did not raise their minimum wage between 2000 and 2007 with available data. Otherwise, the green states constitute the treated group. See Table 1 for exact timing of each state's change in the minimum wage. Red states did not increase their minimum wage over the period from 2000 to 2007.

Table 1: Timing of States Raising Minimum Wage

State	Year-Quarter Raised MW	State	Year-Quarter Raised MW
Alabama	Never Increased	Montana*	2007-1
Alaska	Always Above	Nebraska*	Never Increased
Arizona	2007-1	Nevada*	2006-4
Arkansas	2006-4	New Hampshire	Never Increased
California	Always Above	New Jersey	2005-4
Colorado*	2007-1	New Mexico*	Never Increased
Connecticut	Always Above	New York	2005-1
Delaware	1999-2	North Carolina*	2007-1
Florida*	2005-2	North Dakota*	Never Increased
Georgia*	Never Increased	Ohio*	2007-1
Hawaii	Always Above	Oklahoma*	Never Increased
Idaho*	Never Increased	Oregon	Always Above
Illinois*	2004-1	Pennsylvania	2007-1
Indiana*	Never Increased	Rhode Island	1999-3
Iowa*	2007-2	South Carolina*	Never Increased
Kansas*	Never Increased	South Dakota*	Never Increased
Kentucky	Never Increased	Tennessee*	Never Increased
Louisiana*	Never Increased	Texas*	Never Increased
Maine	2002-1	Utah*	Never Increased
Maryland*	2007-1	Vermont	Always Above
Massachusetts	Always Above	Virginia*	Never Increased
Michigan*	2006-4	Washington	1999-1
Minnesota*	2005-3	West Virginia*	2006-3
Mississippi	Never Increased	Wisconsin*	2005-2
Missouri*	2007-1	Wyoming	Never Increased

Notes: The timing of states increasing their minimum wage above the federal minimum wage of $5.15 per hour which was set in Q4 of 1997 and did not change again until it increased in Q3 of 2007. States that are ultimately included in the main sample are denoted with a *. States that had minimum wages higher than the federal minimum wage at the beginning of the period are excluded. We also exclude some states who raised their minimum wage very soon after the federal minimum wage increase, some others due to lack of data availability, and those in the Northern Census region. There are 29 states ultimately included in the sample.

References

Abraham, S., and Sun, L. (2018), "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Working Paper*, .

Angrist, J. D., and Pischke, J.-S. (2009), *Mostly Harmless Econometrics: An Empiricist ' s Companion*, Princeton, NJ: Princeton University Press.

Autor, D. H., Kerr, W. R., and Kugler, A. D. (2007), "Do employment protections reduce productivity? Evidence from U.S. states," *The Economic Journal*, 117, 189–217.

de Chaisemartin, C., and D'Haultfœuille, X. (2017), "Fuzzy differences-in-differences," *The Review of Economic Studies*, (February), 1–30.

de Chaisemartin, C., and D'Haultfœuille, X. (2018), "Two-way fixed effects estimators with heterogeneous treatment effects," *Working Paper*, .

Dube, A., Lester, T. W., and Reich, M. (2010), "Minimum wage effects across state borders: Estimates using contiguous counties," *Review of Economics and Statistics*, 92(4), 945–964.

Kosorok, M. R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, New York, NY: Springer.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.