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

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

December 1, 2020

This supplementary appendix contains (a) the proofs for the results for repeated cross-section cases; (b) additional details about our application on the minimum wage; (c) Monte Carlo simulations illustrating the finite sample properties of our estimation and inference procedures.

Appendix SA: Proofs of Repeated Cross Sections Results

Before proving Theorem B.1 we introduce two auxiliary lemmas. Recall that

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

Lemma SA.1. *Let Assumption 1, 3, 4, 6, and B.1 hold. Then, for all g and t such that $(g-\delta) \in \mathcal{G}$, $t \in \{2, \dots \mathcal{T} - \delta\}$ and $t \geq g - \delta$,*

$$ATT_X(g,t) = \left(m_{g,t}^{rc,treat}(X) - m_{g,g-\delta-1}^{rc,treat}(X)\right) - \left(m_{c,t}^{rc,nev}(X) - m_{c,g-\delta-1}^{rc,nev}(X)\right) \ a.s..$$

Proof of Lemma SA.1: In what follows, take all equalities to hold almost surely (a.s.).

Then, we have that

$$\begin{aligned}
&ATT_X(g,t) \\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \mathbb{E}[Y_t(0) - Y_{g-\delta-1}(0)|X, G_g = 1] \\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \sum_{\ell=0}^{t-g-\delta} \mathbb{E}[\Delta Y_{t-\ell}(0)|X, G_g = 1] \\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \sum_{\ell=0}^{t-g-\delta} \mathbb{E}[\Delta Y_{t-\ell}(0)|X, C = 1] \\
&= (\mathbb{E}[Y_t(g)|X, G_g = 1] - \mathbb{E}[Y_{g-\delta-1}(0)|X, G_g = 1]) \\
&\qquad - (\mathbb{E}[Y_t(0)|X, C = 1] - \mathbb{E}[Y_{g-\delta-1}(0)|X, C = 1]) \\
&= \left(m_{g,t}^{rc,treat}(X) - m_{g,g-\delta-1}^{rc,treat}(X)\right) - \left(m_{c,t}^{rc,nev}(X) - m_{c,g-\delta-1}^{rc,nev}(X)\right)
\end{aligned}$$

[*]Department of Economics, University of Georgia. Email: brantly.callaway@uga.edu
[†]Department of Economics, Vanderbilt University. Email: pedro.h.santanna@vanderbilt.edu

where the first equality follows from adding and subtracting $\mathbb{E}[Y_{g-\delta-1}(0)|X, G_g = 1]$, the second equality from simple algebra, the third equality by Assumption 4, the fourth equality by simple algebra and the linearity of expectations, and the last equality from (2.1), Assumption 3, and Assumption B.1. □

Lemma SA.2. *Let Assumptions 1, 3, 5, 6, and B.1 hold. Then, for all g and t such that $g \in \mathcal{G}_\delta$, $t \in \{2, \ldots \mathcal{T} - \delta\}$ with $t \geq g - \delta$*

$$ATT_X(g,t) = \left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,ny}_{t+\delta,t}(X) - m^{rc,ny}_{t+\delta,g-\delta-1}(X)\right) \ a.s..$$

Proof of Lemma SA.2: The proof follows similar steps as the proof of Lemma SA.1. Taking all equalities to hold almost surely (a.s.), we have that

$$\begin{aligned}
&ATT_X(g,t)\\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \mathbb{E}[Y_t(0) - Y_{g-\delta-1}(0)|X, G_g = 1]\\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \sum_{\ell=0}^{t-g-\delta} \mathbb{E}[\Delta Y_{t-\ell}(0)|X, G_g = 1]\\
&= \mathbb{E}[Y_t(g) - Y_{g-\delta-1}(0)|X, G_g = 1] - \sum_{\ell=0}^{t-g-\delta} \mathbb{E}[\Delta Y_{t-\ell}(0)|X, D_{t+\delta} = 0, G_g = 0]\\
&= (\mathbb{E}[Y_t(g)|X, G_g = 1] - \mathbb{E}[Y_{g-\delta-1}(0)|X, G_g = 1])\\
&\qquad - (\mathbb{E}[Y_t(0)|X, D_{t+\delta} = 0, G_g = 0] - \mathbb{E}[Y_{g-\delta-1}(0)|X, D_{t+\delta} = 0, G_g = 0])\\
&= \left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,ny}_{t+\delta,t}(X) - m^{rc,ny}_{t+\delta,g-\delta-1}(X)\right)
\end{aligned}$$

where the first equality follows from adding and subtracting $\mathbb{E}[Y_{g-\delta-1}(0)|X, G_g = 1]$, the second equality from simple algebra, the third equality by Assumption 5 with $s = t + \delta$, the fourth equality by simple algebra and linearity of expectations, and the last equality from (2.1), Assumption 3 and Assumption B.1. □

Proof of Theorem B.1:

Part 1: Identification when Assumption 4 is invoked.

In this case, given the result in Lemma SA.1 and the stationarity condition in Assumption B.1, for all g and t such that $g \in \mathcal{G}_\delta$, $t \in \{2, \ldots \mathcal{T} - \delta\}$ and $t \geq g - \delta$, it follows that

$$\begin{aligned}
&ATT(g,t)\\
&= \mathbb{E}_M[ATT_X(g,t)|G_g = 1, T_t = 1]\\
&= \mathbb{E}_M[ATT_X(g,t)|G_g = 1]\\
&= \mathbb{E}_M\left[\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,nev}_{c,t}(X) - m^{rc,nev}_{c,g-\delta-1}(X)\right)\middle| G_g = 1\right]\\
&= \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M[G_g]}\left(\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,nev}_{c,t}(X) - m^{rc,nev}_{c,g-\delta-1}(X)\right)\right)\right]\\
&= ATT^{nev}_{or,rc}(g,t;\delta), \qquad\qquad \text{(SA.1)}
\end{aligned}$$

where the second equality follows from the fact that $ATT_X(g,t)$ is a function of X, which, by Assumption B.1, is invariant to T.

Next, we show that $ATT(g,t) = ATT^{nev}_{ipw,rc}(g,t;\delta)$. By the law of iterated expectations, Assumption B.1 and Assumption 1 for all g and t such that $g \in \mathcal{G}_\delta$, $t \in \{2, \ldots \mathcal{T} - \delta\}$ and $t \geq g - \delta$,

$$\begin{aligned}
\mathbb{E}_M\left[w^{treat}(g,t) \cdot Y\right] &= \frac{\mathbb{E}_M\left[T_t G_g \cdot Y\right]}{\mathbb{E}_M\left[T_t G_g\right]} \\
&= \frac{\mathbb{E}_M\left[G_g \cdot Y | T_t = 1\right]}{\mathbb{E}_M\left[G_g | T_t = 1\right]} \\
&= \mathbb{E}_M\left[Y | T_t = 1, G_g = 1\right] \\
&= \mathbb{E}_M\left[m^{rc,treat}_{g,t}(X) | T_t = 1, G_g = 1\right] \\
&= \mathbb{E}_M\left[m^{rc,treat}_{g,t}(X) | G_g = 1\right] \\
&= \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]} m^{rc,treat}_{g,t}(X)\right].
\end{aligned} \tag{SA.2}$$

Analogously,

$$\mathbb{E}_M\left[w^{treat}(g, g-\delta-1) \cdot Y\right] = \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]} m^{rc,treat}_{g,g-\delta-1}(X)\right]. \tag{SA.3}$$

Next, by noticing that

$$p_g(X) = \frac{\mathbb{E}_M\left[G_g | X\right]}{\mathbb{E}_M\left[G_g + C | X\right]},\ 1 - p_g(X) = \frac{\mathbb{E}_M\left[C | X\right]}{\mathbb{E}_M\left[G_g + C | X\right]}, \tag{SA.4}$$

it follows from the law of iterated expectations, the total law of probability, and Assumption B.1,

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

Thus,

$$\begin{aligned}
\mathbb{E}_M\left[w^{comp}_{nev}(g,t) \cdot Y\right] &= \frac{\mathbb{E}_M\left[T_t \cdot \frac{\mathbb{E}_M\left[G_g | X\right] C}{\mathbb{E}_M\left[C | X\right]} Y\right]}{\mathbb{E}_M\left[G_g\right] \cdot \lambda_t} \\
&= \frac{\mathbb{E}_M\left[\left.\frac{\mathbb{E}_M\left[G_g | X\right] C}{\mathbb{E}_M\left[C | X\right]} Y \right| T_t = 1\right] \cdot \lambda_t}{\mathbb{E}_M\left[G_g\right] \cdot \lambda_t} \\
&= \frac{\mathbb{E}_M\left[\left.\frac{\mathbb{E}_M\left[G_g | X\right]}{\mathbb{E}_M\left[C | X\right]} \mathbb{E}_M\left[C \cdot Y | X, T_t = 1\right] \right| T_t = 1\right]}{\mathbb{E}_M\left[G_g\right]} \\
&= \frac{\mathbb{E}_M\left[\mathbb{E}_M\left[G_g | X\right] \cdot \mathbb{E}_M\left[Y | X, C = 1, T_t = 1\right] \middle| T_t = 1\right]}{\mathbb{E}_M\left[G_g\right]}
\end{aligned}$$

$$= \frac{\mathbb{E}_M\left[G_g \cdot m^{rc,nev}_{c,t}(X)\right]}{\mathbb{E}_M\left[G_g\right]}. \tag{SA.5}$$

Following the same steps, we get

$$\mathbb{E}_M\left[w^{comp}_{nev}(g, g-\delta-1)\cdot Y\right] = \frac{\mathbb{E}_M\left[G_g \cdot m^{rc,nev}_{c,g-\delta-1}\right]}{\mathbb{E}_M\left[G_g\right]}. \tag{SA.6}$$

Thus, by combining (SA.2), (SA.3), (SA.5) and (SA.6), and exploiting (SA.1), we have that for all g and t such that $g \in \mathcal{G}_\delta$, $t \in \{2, \ldots \mathcal{T} - \delta\}$ and $t \geq g - \delta$,

$$\begin{aligned}
&ATT^{nev}_{ipw,rc}(g, t; \delta) \\
&= \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]}\left(\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,nev}_{c,t}(X) - m^{rc,nev}_{c,g-\delta-1}(X)\right)\right)\right] \\
&= ATT(g, t).
\end{aligned}$$

Finally, we show that $ATT(g,t) = ATT^{nev}_{dr,rc}(g, t; \delta)$. Towards this end, we rewrite $ATT^{nev}_{dr,rc}(g, t; \delta)$ as

$$\begin{aligned}
&ATT^{nev}_{dr,rc}(g, t; \delta) \\
&= \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]}\left(\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,nev}_{c,t}(X) - m^{rc,nev}_{c,g-\delta-1}(X)\right)\right)\right] \\
&+ \mathbb{E}_M\left[w^{treat}(g, t)\left(Y - m^{rc,treat}_{g,t}(X)\right) - w^{treat}(g, g-\delta-1)\left(Y - m^{rc,treat}_{g,g-\delta-1}(X)\right)\right] \\
&- \mathbb{E}_M\left[w^{comp}_{nev}(g, t)\left(Y - m^{rc,nev}_{c,t}(X)\right) - w^{comp}_{nev}(g, g-\delta-1)\left(Y - m^{rc,nev}_{c,g-\delta-1}(X)\right)\right] \\
&= ATT^{nev}_{or,rc}(g, t; \delta) + ATT^{nev}_{ipw,rc}(g, t; \delta) \\
&- \left(\mathbb{E}_M\left[w^{treat}(g, t)\, m^{rc,treat}_{g,t}(X) - w^{treat}(g, g-\delta-1)\, m^{rc,treat}_{g,g-\delta-1}(X)\right]\right. \\
&\qquad \left. -\mathbb{E}_M\left[w^{comp}_{nev}(g, t)\, m^{rc,nev}_{c,t}(X) - w^{comp}_{nev}(g, g-\delta-1)\, m^{rc,nev}_{c,g-\delta-1}(X)\right]\right).
\end{aligned} \tag{SA.7}$$

Thus, since we have already established that $ATT(g,t) = ATT^{nev}_{or,rc}(g, t; \delta) = ATT^{nev}_{ipw,rc}(g, t; \delta)$, it suffices to establish

$$\begin{aligned}
ATT(g,t) = &\ \mathbb{E}_M\left[w^{treat}(g, t)\, m^{rc,treat}_{g,t}(X) - w^{treat}(g, g-\delta-1)\, m^{rc,treat}_{g,g-\delta-1}(X)\right] \\
&- \mathbb{E}_M\left[w^{comp}_{nev}(g, t)\, m^{rc,nev}_{c,t}(X) - w^{comp}_{nev}(g, g-\delta-1)\, m^{rc,nev}_{c,g-\delta-1}(X)\right].
\end{aligned}$$

Following the same steps we used to establish (SA.2), (SA.3), (SA.5) (SA.6), we have that

$$\begin{aligned}
&\mathbb{E}_M\left[w^{treat}(g, t)\, m^{rc,treat}_{g,t}(X) - w^{treat}(g, g-\delta-1)\, m^{rc,treat}_{g,g-\delta-1}(X)\right] \\
&- \mathbb{E}_M\left[w^{comp}_{nev}(g, t)\, m^{rc,nev}_{c,t}(X) - w^{comp}_{nev}(g, g-\delta-1)\, m^{rc,nev}_{c,g-\delta-1}(X)\right] \\
&= \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]}\left(\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,nev}_{c,t}(X) - m^{rc,nev}_{c,g-\delta-1}(X)\right)\right)\right] \\
&= ATT(g, t),
\end{aligned}$$

where the last step follows from (SA.1).

Part 2: Identification when Assumption 5 is invoked.

In this case, given the result in Lemma SA.1 and the stationarity condition in Assumption B.1, for all g and t such that $g \in \mathcal{G}_\delta$, $t \in \{2, \dots \mathcal{T} - \delta\}$ and $t \geq g - \delta$, it follows that

$$\begin{aligned}
&ATT(g,t) \\
&= \mathbb{E}_M[ATT_X(g,t)|G_g = 1, T_t = 1] \\
&= \mathbb{E}_M[ATT_X(g,t)|G_g = 1] \\
&= \mathbb{E}_M\left[\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,ny}_{t+\delta,t}(X) - m^{rc,ny}_{t+\delta,g-\delta-1}(X)\right)\middle| G_g = 1\right] \\
&= \mathbb{E}\left[\frac{G_g}{\mathbb{E}[G_g]}\left(\left(m^{rc,treat}_{g,t}(X) - m^{rc,treat}_{g,g-\delta-1}(X)\right) - \left(m^{rc,ny}_{t+\delta,t}(X) - m^{rc,ny}_{t+\delta,g-\delta-1}(X)\right)\right)\right] \\
&= ATT^{ny}_{or,rc}(g,t;\delta),
\end{aligned} \tag{SA.8}$$

where the second equality follows from the fact that $ATT_X(g,t)$ is a function of X, which, by Assumption B.1, is invariant to T.

Next, we show that $ATT(g,t) = ATT^{ny}_{ipw,rc}(g,t;\delta)$. From (SA.2) and (SA.3), we have that

$$\mathbb{E}_M\left[w^{treat}(g,t)\cdot Y\right] = \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M[G_g]} m^{rc,treat}_{g,t}(X)\right], \tag{SA.9}$$

$$\mathbb{E}_M\left[w^{treat}(g,g-\delta-1)\cdot Y\right] = \mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M[G_g]} m^{rc,treat}_{g,g-\delta-1}(X)\right]. \tag{SA.10}$$

Next, by noticing that

$$p_{g,t+\delta}(X) = \frac{\mathbb{E}_M[G_g|X]}{\mathbb{E}_M[G_g + (1-D_{t+\delta})(1-G_g)|X]}, \quad 1 - p_{g,t+\delta}(X) = \frac{\mathbb{E}_M[(1-D_{t+\delta})(1-G_g)|X]}{\mathbb{E}_M[G_g + (1-D_{t+\delta})(1-G_g)|X]}, \tag{SA.11}$$

it follows from the law of iterated expectations, the total law of probability, and Assumption B.1 that

$$\begin{aligned}
&\mathbb{E}_M\left[\frac{T_t \cdot p_{g,t+\delta}(X)(1-D_{t+\delta})(1-G_g)}{(1-p_{g,t+\delta}(X))}\right] \\
&= \mathbb{E}_M\left[T_t \cdot \frac{\mathbb{E}_M[G_g|X](1-D_{t+\delta})(1-G_g)}{\mathbb{E}_M[(1-D_{t+\delta})(1-G_g)|X]}\right] \\
&= \mathbb{E}_M\left[\frac{\mathbb{E}_M[G_g|X]\,\mathbb{E}_M[(1-D_{t+\delta})(1-G_g)|X]}{\mathbb{E}_M[(1-D_{t+\delta})(1-G_g)|X]}\middle| T_t = 1\right]\cdot\lambda_t \\
&= \mathbb{E}_M[\mathbb{E}_M[G_g|X]|\, T_t = 1]\cdot\lambda_t \\
&= \mathbb{E}_M[G_g]\cdot\lambda_t.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{E}_M\left[w^{comp}_{ny}(g,t,t+\delta)\cdot Y\right] \\
&= \frac{\mathbb{E}_M\left[T_t \cdot \dfrac{\mathbb{E}_M[G_g|X](1-D_{t+\delta})(1-G_g)}{\mathbb{E}_M[(1-D_{t+\delta})(1-G_g)|X]} Y\right]}{\mathbb{E}_M[G_g]\cdot\lambda_t}
\end{aligned}$$

$$= \frac{\mathbb{E}_M\left[\left.\frac{\mathbb{E}_M\left[G_g|X\right]\left(1-D_{t+\delta}\right)\left(1-G_g\right)}{\mathbb{E}_M\left[\left(1-D_{t+\delta}\right)\left(1-G_g\right)|X\right]}Y\right|T_t=1\right]\cdot\lambda_t}{\mathbb{E}_M\left[G_g\right]\cdot\lambda_t}$$
$$= \frac{\mathbb{E}_M\left[\mathbb{E}_M\left[G_g|X\right]\cdot\mathbb{E}_M\left[Y|\,X,D_{t+\delta}=0,G_g=0T_t=1\right]|\,T_t=1\right]}{\mathbb{E}_M\left[G_g\right]}$$
$$= \frac{\mathbb{E}_M\left[G_g\cdot m^{rc,ny}_{t+\delta,t}\left(X\right)\right]}{\mathbb{E}_M\left[G_g\right]}. \tag{SA.12}$$

Following the same steps, we get

$$\mathbb{E}_M\left[w^{comp}_{ny}\left(g,g-\delta-1,t+\delta\right)\cdot Y\right]=\frac{\mathbb{E}_M\left[G_g\cdot m^{rc,ny}_{t+\delta,g-\delta-1}\right]}{\mathbb{E}_M\left[G_g\right]}. \tag{SA.13}$$

Thus, by combining (SA.9), (SA.10), (SA.12) and (SA.13), and exploiting (SA.8), we have that for all g and t such that $g\in\mathcal{G}_\delta$, $t\in\{2,\dots\mathcal{T}-\delta\}$ and $t\geq g-\delta$,

$$ATT^{nev}_{ipw,rc}\left(g,t;\delta\right)=\mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]}\left(\left(m^{rc,treat}_{g,t}\left(X\right)-m^{rc,treat}_{g,g-\delta-1}\left(X\right)\right)-\left(m^{rc,ny}_{t+\delta,t}\left(X\right)-m^{rc,ny}_{t+\delta,g-\delta-1}\left(X\right)\right)\right)\right]$$
$$=ATT\left(g,t\right).$$

Finally, we show that $ATT(g,t)=ATT^{ny}_{dr,rc}\left(g,t;\delta\right)$. Towards this end, from analogous steps to (SA.7), we rewrite $ATT^{ny}_{dr,rc}\left(g,t;\delta\right)$ as

$$\begin{aligned}
&ATT^{nev}_{dr,rc}\left(g,t;\delta\right)\\
&=2\cdot ATT\left(g,t\right)\\
&-\Big(\mathbb{E}_M\left[w^{treat}\left(g,t\right)m^{rc,treat}_{g,t}\left(X\right)-w^{treat}\left(g,g-\delta-1\right)m^{rc,treat}_{g,g-\delta-1}\left(X\right)\right]\\
&\qquad-\mathbb{E}_M\left[w^{comp}_{ny}\left(g,t,t+\delta\right)m^{rc,ny}_{t+\delta,t}\left(X\right)-w^{comp}_{ny}\left(g,g-\delta-1,t+\delta\right)m^{rc,ny}_{t+\delta,g-\delta-1}\left(X\right)\right]\Big).
\end{aligned}$$

Thus, it suffices to show that

$$\begin{aligned}
ATT(g,t)=&\,\mathbb{E}_M\left[w^{treat}\left(g,t\right)m^{rc,treat}_{g,t}\left(X\right)-w^{treat}\left(g,g-\delta-1\right)m^{rc,treat}_{g,g-\delta-1}\left(X\right)\right]\\
&-\mathbb{E}_M\left[w^{comp}_{ny}\left(g,t,t+\delta\right)m^{rc,ny}_{t+\delta,t}\left(X\right)-w^{comp}_{ny}\left(g,g-\delta-1,t+\delta\right)m^{rc,ny}_{t+\delta,g-\delta-1}\left(X\right)\right].
\end{aligned}$$

Following the same steps we used to establish (SA.9), (SA.10), (SA.12) (SA.13), we have that

$$\begin{aligned}
&\mathbb{E}_M\left[w^{treat}\left(g,t\right)m^{rc,treat}_{g,t}\left(X\right)-w^{treat}\left(g,g-\delta-1\right)m^{rc,treat}_{g,g-\delta-1}\left(X\right)\right]\\
&-\mathbb{E}_M\left[w^{comp}_{ny}\left(g,t,t+\delta\right)m^{rc,ny}_{t+\delta,t}\left(X\right)-w^{comp}_{ny}\left(g,g-\delta-1,t+\delta\right)m^{rc,ny}_{t+\delta,g-\delta-1}\left(X\right)\right]\\
&=\mathbb{E}_M\left[\frac{G_g}{\mathbb{E}_M\left[G_g\right]}\left(\left(m^{rc,treat}_{g,t}\left(X\right)-m^{rc,treat}_{g,g-\delta-1}\left(X\right)\right)-\left(m^{rc,ny}_{t+\delta,t}-m^{rc,ny}_{t+\delta,g-\delta-1}\left(X\right)\right)\right)\right]\\
&=ATT\left(g,t\right),
\end{aligned}$$

where the last step follows from (SA.8).

□

Appendix SB: Additional Details about Empirical Application

In our empirical application, we study the effect of the minimum wage on teen employment. This section provides some additional details about the data that we use and some additional results from our application.

Additional Details on Data Construction

In September 1997, the federal minimum wage was increased from \$4.75 per hour to \$5.15 per hour. It remained flat until July 2007 when it was increased again to \$5.85 per hour. For forty states, the federal minimum wage was the binding minimum wage in Q2 of 2000.[1] We omit the other ten states as we do not observe their untreated potential outcomes in 2001. We drop seven other states for lack of data on teen employment. We also drop four other states in the Northern census region because there are no states in the Northern census region that had not increased their minimum wage by 2007,[2] and census region is an important control variable in the minimum wage literature. Our outcome is quarterly employment among teenagers in the first quarter of each year from 2001 to 2007. We use first quarter employment because it is further away from the federal minimum wage increase in Q3 of 2007. Our final sample includes county level teen employment for 29 states matched with county characteristics.

Our strategy is to divide the observations based on the timing of when a state increased its minimum wage above the federal minimum wage. States that did not raise their minimum wage during this period form the untreated group. We also have groups of states that increased their minimum wage during 2004, 2006, and 2007. Before 2004, Illinois did not have a state minimum wage. In Q1 of 2004, Illinois set a state minimum wage of \$5.50 which was 35 cents higher than the federal minimum wage. In Q1 of 2005, Illinois increased its minimum wage to \$6.50 where it stayed for the remainder of the period that we consider. No other states changed their minimum wage policy by the first quarter of 2005. In the second quarter of 2005, Florida and Wisconsin set a state minimum wage above the federal minimum wage. In Q3 of 2005, Minnesota also set a state minimum wage. Florida and Wisconsin each gradually increased their minimum wages over time, while Minnesota's was flat over the rest of the period. These three states constitute the treated group for 2006. West Virginia increased its minimum wage in Q3 of 2006; Michigan and Nevada increased their minimum wages in Q4 of 2006; Colorado, Maryland, Missouri, Montana, North Carolina, and Ohio increased their state minimum wages in Q1 of 2007. These states form the 2007 treated group. Table SB.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.

Among these states there is some heterogeneity in the size of the minimum wage increase. For example, West Virginia had the smallest minimum wage increase to \$5.85. At the other extreme, Michigan increased its minimum wage to \$6.95 and then to \$7.15 by Q2 of 2007. Across groups (defined by the year when they increased their minimum wage), the average minimum wage increase was to \$6.50 for $g = 2004$, to \$6.44 for $g = 2006$, and to \$6.42 for $g = 2007$.

[1] To be precise, we use only employment data from the first quarter of each year. A state is considered to raise its minimum wage in year y if it raised its minimum wage in Q2, Q3, or Q4 of year $y - 1$ or in Q1 of year y.

[2] It would be possible to identify some group-time average treatment effects for some of these states as we could use the "late-treated" states as the comparison group in the earlier periods for the "early-treated" states, but we do not pursue this here.

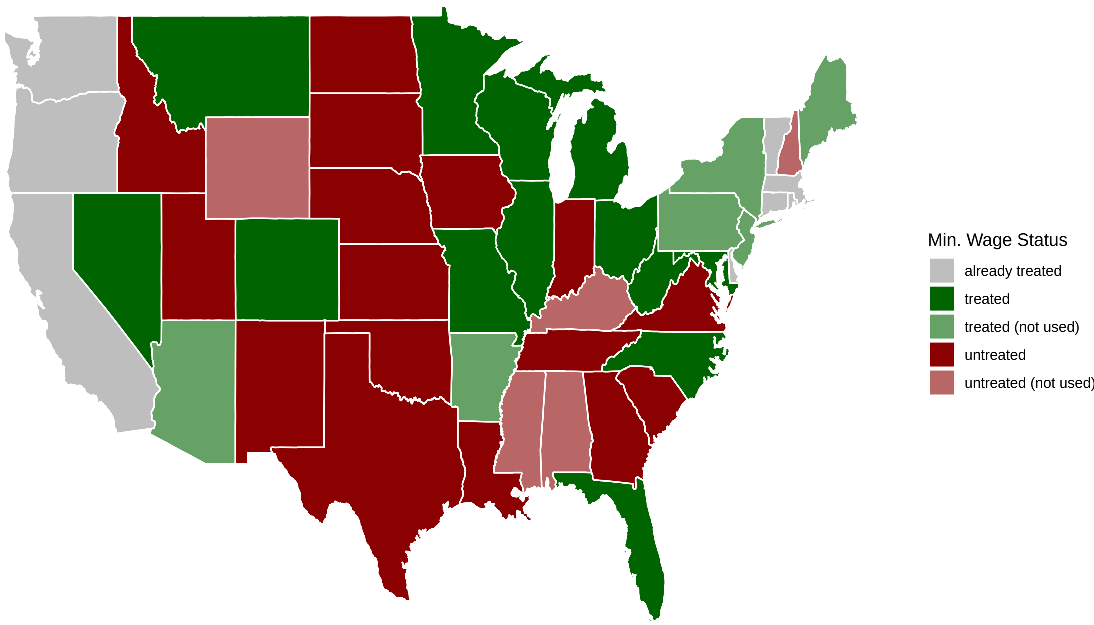
Table SB.1: Timing of States Raising Minimum Wage

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 Q3 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 *. There are 29 states included in the final sample.

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 SB.1 contains the spatial distribution of state-level minimum wage policy changes in our sample. Indeed, Figure SB.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 SB.1: The Spatial Distribution of Minimum Wage Policies



Notes: Gray states had minimum wages higher than the federal minimum wage in Q1 of 2001 and are not used in our analysis. Green states increased their state minimum wage between Q2 of 2001 and Q1 of 2007 and constitute different treated groups in the paper. 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 2001 and 2007 with available data (Delaware has missing data); these omitted states are in light green in the figure. See Table SB.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 2001 to 2007 and form the "never-treated" group. Some of these also have missing data; these states are in light red in the figure.

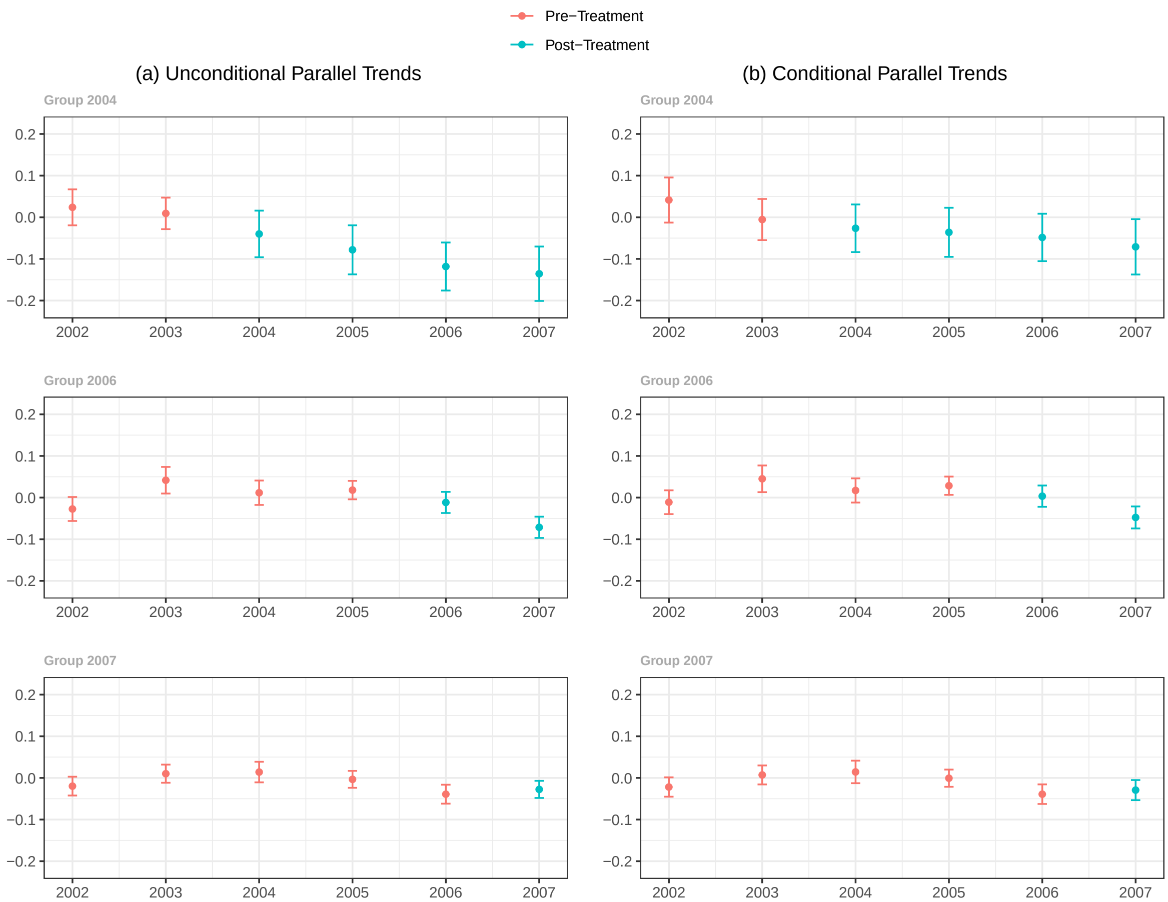
Additional Results

Next, we provides some additional results from our application. These correspond to the results from the main text under different conditions: (i) using the not-yet-treated (instead of the never-treated) as the comparison group, and (ii) allowing for treatment anticipation (with never-treated as comparison group).

Results with Not-Yet-Treated Comparison Group

First, we consider analogous results to those in the main text but using the not-yet-treated group as the comparison group. This comparison group is larger than the never-treated group; in this case, late-treated units are included in the comparison group for early-treated units in early periods. These results are available in Figure SB.2 and Table SB.2. These results are very similar to the results using the never-treated group in the main text.

Figure SB.2: Minimum Wage Group-Time Average Treatment Effects using Not-Yet-Treated Comparison Group



Notes: The effect of the minimum wage on teen employment estimated under the unconditional parallel trends assumption (Panel (a)) and the conditional parallel trends assumption (Panel (b)) using the not-yet-treated group as the comparison group (this is in contrast to the main text that used the never-treated group as the comparison group). Red lines give point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the county level. Under the null hypothesis of the parallel trends assumption holding in all periods, these should be equal to 0. Blue lines provide point estimates and uniform 95% confidence bands for the treatment effect of increasing the minimum wage allowing for clustering at the county level. The top row includes states that increased their minimum wage in 2004, the middle row includes states that increased their minimum wage in 2006, and the bottom row includes states that increased their minimum wage in 2007. No states raised their minimum wages in other years prior to 2007.

Table SB.2: Minimum Wage Aggregated Treatment Effect Estimates using Not-Yet-Treated Comparison Group

(a) Unconditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.037
					(0.006)
Simple Weighted Average					-0.050
					(0.007)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.093	-0.042	-0.028		-0.038
	(0.020)	(0.008)	(0.007)		(0.006)
Event Study	e=0	e=1	e=2	e=3	
	-0.025	-0.073	-0.118	-0.136	-0.088
	(0.005)	(0.009)	(0.020)	(0.024)	(0.012)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.040	-0.078	-0.065	-0.078	-0.065
	(0.022)	(0.022)	(0.010)	(0.006)	(0.012)
Event Study	e=0	e=1			
w/ Balanced Groups	-0.020	-0.073			-0.047
	(0.009)	(0.009)			(0.008)

(b) Conditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.008
					(0.006)
Simple Weighted Average					-0.032
					(0.007)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.045	-0.022	-0.029		-0.029
	(0.019)	(0.008)	(0.008)		(0.006)
Event Study	e=0	e=1	e=2	e=3	
	-0.021	-0.044	-0.048	-0.071	-0.046
	(0.006)	(0.009)	(0.021)	(0.025)	(0.012)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.026	-0.036	-0.023	-0.049	-0.034
	(0.022)	(0.020)	(0.009)	(0.007)	(0.012)
Event Study	e=0	e=1			
w/ Balanced Groups	-0.006	-0.044			-0.025
	(0.009)	(0.009)			(0.008)

Notes: The table reports aggregated treatment effect parameters under the unconditional and conditional parallel trends assumptions using not-yet-treated units as the comparison group and with clustering at the county level. The row 'TWFE' reports the coefficient on a post-treatment dummy variable from a two-way fixed effects regression. The row 'Simple Weighted Average' reports the weighted average (by group size) of all available group-time average treatment effects as in Equation (3.10). The row 'Group-Specific Effects' summarizes average treatment effects by the timing of the minimum wage increase; here, g indexes the year that a county is first treated. The row 'Event Study' reports average treatment effects by the length of exposure to the minimum wage increase; here, e indexes the length of exposure to the treatment. The row 'Calendar Time Effects' reports average treatment effects by year; here, t indexes the year. The row 'Event Study w/ Balanced Groups' reports average treatment effects by length of exposure using a fixed set of groups at all lengths of exposure; here, e indexes the length of exposure and the sample consists of counties that have at least two years of exposure to minimum wage increases. The column 'Single Parameters' represents a further aggregation of each type of parameter, as discussed in the text.

Results with Treatment Anticipation

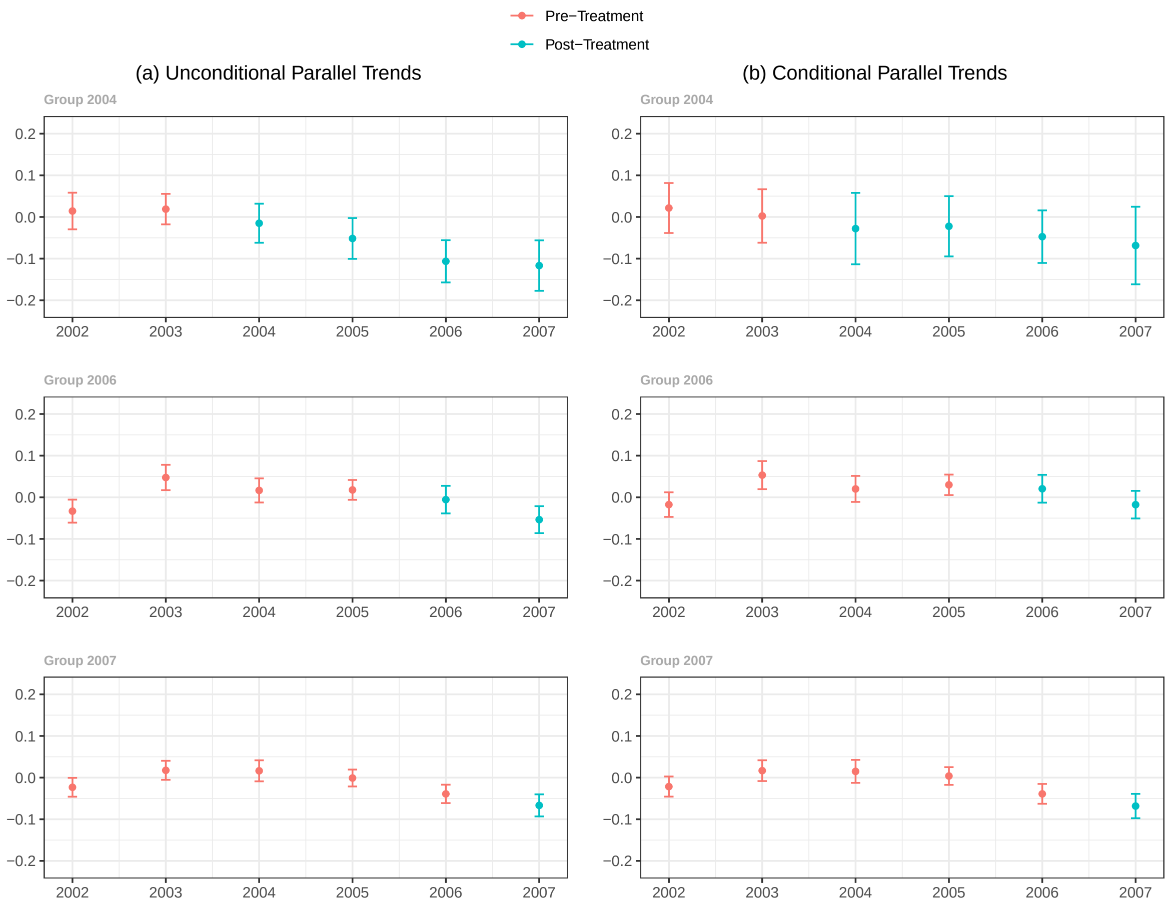
Finally, we provide results allowing for treatment anticipation. These results are available in Figure SB.3 and Table SB.3. The key difference between this case and the one without anticipation that is considered in the main text is that the "reference" period is adjusted from the year immediately preceding the minimum wage increase in a particular state to two years before the minimum wage increase. In other words, for $g = 2004$, the "reference" period is 2002 instead of 2003; for $g = 2006$, it is 2004 instead of 2005; for $g = 2007$, it is 2005 instead of 2006.

The discussion in the text suggests that group-time average treatment effects, under one year anticipation, are only identified in years up to 2006 and for groups that first increase their minimum wage in 2006 or earlier. We decided to show results through 2007 and for the group of states that increased their minimum wage in 2007 for two reasons.[3] First, the group first treated in 2007 is seemingly the one with the most evidence of anticipation effects – see Figure 1 in the main text, and for this reason we thought it was helpful to show these results. Second, under the additional condition that the anticipation effects are not positive, then the group-time average treatment effects in 2007 will be attenuated. This holds because, under this extra condition, the trend in teen employment for the comparison states would be less than or equal to the trend in the absence of the minimum wage change due to anticipation effects among the comparison states. This sort of extra condition would hold if, in the year right before the minimum wage increase, teen employment either does not change or is negatively affected by the impending minimum wage increase. This is likely to be a reasonable extra condition in our particular application.

Allowing for anticipation, the results are quite similar to the results in the main text. The most notable difference among the group-time average treatment effects is in 2007 for the group first treated in 2007. In that case, under unconditional parallel trends, the point estimate changes from a 2.8% decrease in teen employment (in the case without anticipation) to a 6.7% decrease in teen employment (with anticipation). The mechanical reason for this is that teen employment seems to have decreased already (under parallel trends) in 2006 for the group first treated in 2007. Without anticipation, 2006 is the base period and so the additional decrease (under parallel trends) is somewhat mitigated. On the other hand, when the base period is 2005, the full decrease (under parallel trends) from 2005 to 2007 is ascribed to the minimum wage increase in 2007 resulting in substantially larger estimates of the effect of the minimum wage increase. If there are actually anticipation effects, then this is the appropriate comparison to make.

[3] This essentially imposes a no-anticipation condition for the never-treated group. If one does not find this extra-condition reasonable, it suffices to ignore all the treatment effects for $t = 2007$.

Figure SB.3: Minimum Wage Group-Time Average Treatment Effects with 1 Year Anticipation



Notes: The effect of the minimum wage on teen employment estimated under the unconditional parallel trends assumption (Panel (a)) and the conditional parallel trends assumption (Panel (b)) using the never-treated group as the comparison group and allowing for 1 year anticipation of the minimum wage increase. Relative to the case without anticipation, the main difference is that 2002 (for $g = 2004$), 2004 (for $g = 2006$), and 2005 (for $g = 2007$) are used as the "reference" periods instead of 2003, 2005, and 2006, respectively, due to anticipation. Red lines give point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the county level. Under the null hypothesis of the parallel trends assumption holding in all periods, these should be equal to 0. Blue lines provide point estimates and uniform 95% confidence bands for the treatment effect of increasing the minimum wage allowing for clustering at the county level. The top row includes states that increased their minimum wage in 2004, the middle row includes states that increased their minimum wage in 2006, and the bottom row includes states that increased their minimum wage in 2007. No states raised their minimum wages in other years prior to 2007.

Table SB.3: Minimum Wage Aggregated Treatment Effect Estimates with 1 Year Anticipation

(a) Unconditional Parallel Trends

	Partially Aggregated				Single Parameters
Simple Weighted Average					-0.057
					(0.007)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.072	-0.030	-0.067		-0.058
	(0.017)	(0.010)	(0.009)		(0.007)
Event Study	e=0	e=1	e=2	e=3	
	-0.046	-0.053	-0.106	-0.117	-0.080
	(0.007)	(0.009)	(0.017)	(0.022)	(0.011)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.015	-0.052	-0.056	-0.079	-0.050
	(0.017)	(0.018)	(0.010)	(0.008)	(0.010)
Event Study	e=0	e=1			
w/ Balanced Groups	-0.009	-0.053			-0.031
	(0.009)	(0.009)			(0.009)

(b) Conditional Parallel Trends

	Partially Aggregated				Single Parameters
Simple Weighted Average					-0.039
					(0.009)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.041	0.001	-0.068		-0.048
	(0.024)	(0.010)	(0.009)		(0.007)
Event Study	e=0	e=1	e=2	e=3	
	-0.042	-0.019	-0.047	-0.069	-0.044
	(0.008)	(0.010)	(0.020)	(0.030)	(0.015)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.028	-0.022	-0.012	-0.052	-0.029
	(0.029)	(0.025)	(0.011)	(0.008)	(0.015)
Event Study	e=0	e=1			
w/ Balanced Groups	0.005	-0.019			-0.007
	(0.012)	(0.011)			(0.010)

Notes: The table reports aggregated treatment effect parameters under the unconditional and conditional parallel trends assumptions allowing for one year anticipation of of participating in the treatment and with clustering at the county level. The row 'Simple Weighted Average' reports the weighted average (by group size) of all available group-time average treatment effects as in Equation (3.10). The row 'Group-Specific Effects' summarizes average treatment effects by the timing of the minimum wage increase; here, g indexes the year that a county is first treated. The row 'Event Study' reports average treatment effects by the length of exposure to the minimum wage increase; here, e indexes the length of exposure to the treatment. The row 'Calendar Time Effects' reports average treatment effects by year; here, t indexes the year. The row 'Event Study w/ Balanced Groups' reports average treatment effects by length of exposure using a fixed set of groups at all lengths of exposure; here, e indexes the length of exposure and the sample consists of counties that have at least two years of exposure to minimum wage increases. The column 'Single Parameters' represents a further aggregation of each type of parameter, as discussed in the text.

Appendix SC: Monte Carlo Simulations

For this section, we consider the finite sample properties of our estimators of group-time average treatment effects and dynamic effects (i.e., event studies). For all of the simulations below, we consider the case with a single covariate $X \sim N(0,1)$, assume that there is no treatment effect anticipation, i.e., $\delta = 0$, and restrict our attention to the case where panel data are available. We allow the number of time periods, $\mathcal{T}$, the group structure, $\mathcal{G} = \{0, 2, 3, \ldots, G\}$, and the sample size n to vary across designs. We compare the performance of our proposed DiD estimators based on the regression (REG), inverse probability weighted (IPW), and doubly-robust (DR) estimands discussed in the main text in terms of average bias, root mean-squared error (RMSE), empirical pointwise 95% rejection probability (Rej.P), and average length of the pointwise 95% confidence interval. For completeness, we also consider the unconditional DiD estimand introduced in equation (2.7) of the main text where the parallel trends is assumed to hold unconditionally. For simplicity, we only report simulation results using the "not-yet-treated" as comparison groups. All results are based on 2,000 Monte Carlo simulations.

We consider the following data generating process. First, we set the probability of belonging to group $g \in \mathcal{G}$ as

$$P(G = g|X) = \frac{\exp(X'\gamma_g)}{\sum_{g \in \mathcal{G}} \exp(X'\gamma_g)}$$

where $\gamma_g = 0.5\, g/G$. We define untreated potential outcomes by

$$Y_{it}(0) = \delta_t + \eta_i + X'\beta_t + u_{it}$$

where $\delta_t = t$, $\eta|G, X \sim N(G, 1)$, $\beta_t = t$, and $u_t|G, X \sim N(0, 1)$ in all time periods. Finally, for units in group $g \neq 0$ and in their post-treatment periods (i.e., outcomes when $t \geq g$), we set

$$\begin{aligned} Y_{it}(g) &= \delta_t + \eta_i + X'\beta_t + \delta_e + v_{it} \\ &= Y_{it}(0) + \delta_e + (v_{it} - u_{it}) \end{aligned}$$

where $\delta_e = e + 1 = t - g + 1$ and $v_t|X, G, u_t \sim N(0, 1)$, with $e = t - g$ denoting the length of exposure to the treatment for a unit in period t. This setup implies that, for all groups and all post-treatment periods, $ATT(g, t) = t - g + 1 = e + 1$ so that there are dynamic effects of participating in the treatment that are equal to the number of time periods that a group has been exposed to the treatment. In all simulation exercises, we consider a logistic generalized propensity score and a linear outcome regression working model where X enters the working models linearly. In this case, both the generalized propensity score and the outcome regression models are correctly specified. This means that all of the estimands (DR, REG, and IPW) that we proposed in the main text are correctly specified. However, parallel trends are only satisfied after conditioning on X, so the unconditional DiD estimands are not equal to the $ATT(g, t)$. For the outcome regressions, we estimate the unknown parameters using ordinary least squares, whereas we use maximum likelihood to estimate the unknown generalized propensity score parameters.

We consider three different setups: (a) $\mathcal{T} = G = 4$, (b) $\mathcal{T} = 20, G = 4$, and (c) $\mathcal{T} = G = 20$. For each setup, we consider three different sample sizes, $n = 50, 200, 1000$. In all cases we report results for $ATT(2, 2)$ and $\theta_e(0)$. For the cases where $\mathcal{T} = 20$, we also report estimates for $\theta_e(10)$. Setup (a) involves a "large" number of units but small number of number of time periods; i.e., this is a textbook short-panel case. Setup (b) involves many time periods but the number of groups is still small. Setup (c) involves a larger number of groups and time periods. Our theoretical results suggest that, when the sample size n is relatively large when compared to the number of groups G, our proposed conditional DiD estimators should perform well. It is less clear how well our estimators will perform when n and G are of similar magnitude. For estimating group-time average treatment effects, the number of units in a particular group can be very small in this setup which suggests that our asymptotic approximations may not work well. The dynamic effects estimators may work better as they effectively pool from different groups with the same event time which makes the effective sample size larger for these cases. This is likely to be especially true when event time is equal to 0 (all groups that are ever treated can be pooled together to estimate $\theta_e(0)$). Considering $\theta_e(10)$ is an intermediate case as it involves pooling groups that participate in the treatment for at least 10 periods, and there are fewer of these groups than there are for estimating $\theta_e(0)$. We also note that with $n = 50$ and $\mathcal{T} = G = 20$, it is infeasible to use a never treated group as the control group as, in many simulations, the number of never treated units is very small or even zero.

The results are provided in Tables SC.1 - SC.3. To summarize these results, as expected, whenever the number of units per group is moderate (all cases with either 4 groups or $n = 1000$), all our conditional DiD estimators for the considered functionals have good statistical properties. When the number of units per group is small (the case in Table SC.3 with $n = 50$), the performance of our estimators of $ATT(2,2)$ is relatively poor – which is not surprising as the number of units in group 2 can be quite small.[4] But, even in this case, our event-study-type estimators perform relatively well, especially when one focuses on instantaneous effects, $\theta_e(0)$; the size distortions associated with the DiD estimators for $\theta_e(10)$ are relatively moderate, which we find encouraging. We also note that, as expected, ignoring the important role of covariates leads to unreliable and imprecise inference procedures.

All in all, these simple simulations highlight that (i) our different estimators perform well when the theory suggests that they should; and (ii) even though our theoretical results do not formally consider settings with n and $\mathcal{T}$ (and G) of similar magnitudes, the statistical properties of our proposed event-study estimators can be relatively stable.

Table SC.1: Monte Carlo simulations based on 2,000 repetitions. Setup with $\mathcal{T} = 4$ and $G = 4$.

		$ATT(2,2)$				$\theta_e(0)$				$\theta_e(10)$			
	n	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL
DR	50	0.007	0.476	0.067	1.868	0.010	0.321	0.048	1.240	-	-	-	-
REG	50	0.006	0.474	0.066	1.868	0.009	0.317	0.046	1.249	-	-	-	-
IPW	50	0.006	0.477	0.067	1.880	0.021	0.326	0.053	1.258	-	-	-	-
UNC	50	-0.037	0.574	0.072	2.220	0.200	0.416	0.089	1.425	-	-	-	-
DR	200	0.006	0.231	0.050	0.918	0.007	0.155	0.057	0.599	-	-	-	-
REG	200	0.006	0.231	0.051	0.917	0.008	0.155	0.055	0.598	-	-	-	-
IPW	200	0.006	0.231	0.048	0.919	0.009	0.158	0.049	0.613	-	-	-	-
UNC	200	-0.036	0.286	0.058	1.113	0.200	0.270	0.193	0.708	-	-	-	-
DR	1000	-0.003	0.106	0.049	0.407	-0.001	0.070	0.059	0.265	-	-	-	-
REG	1000	-0.003	0.106	0.047	0.408	-0.001	0.070	0.059	0.265	-	-	-	-
IPW	1000	-0.002	0.106	0.052	0.409	-0.002	0.072	0.060	0.270	-	-	-	-
UNC	1000	-0.039	0.135	0.066	0.497	0.191	0.208	0.653	0.315	-	-	-	-

Notes: Rows labeled 'DR', 'REG', and 'IPW' use the doubly robust, regression, and inverse probability weighting estimators discussed in the main text, respectively. Rows labeled 'UNC' uses the unconditional DiD estimator that ignores covariates. Columns labeled '$ATT(2,2)$' are based on estimates for the group-time average treatment effect for group 2 in time period 2; columns labeled '$\theta_e(0)$' are based on estimates of the (aggregated) dynamic effects estimator for event time $e = 0$ and '$\theta_e(10)$' for event time $e = 10$. Finally, "Bias","RMSE", "Rej.P", and "CIL', stand for the average simulated bias, simulated root mean-squared errors, average of the simulated error for the 95% coverage probability, and 95% confidence interval length, respectively. Standard errors are computed using the multiplier-bootstrap proposed in the paper, with 1,000 bootstrap draws. Pointwise asymptotic-based critical values are used to compute 95% confidence intervals.

[4] In fact, in the setup of this simulation, there are some cases where $ATT(2,2)$ cannot be computed (e.g., some simulation runs do not include any units in group 2) and we omit these results.

Table SC.2: Monte Carlo simulations based on 2,000 repetitions. Setup with $\mathcal{T} = 20$ and $G = 4$.

		$ATT(2,2)$				$\theta_e(0)$				$\theta_e(10)$			
	n	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL
DR	50	0.014	0.468	0.050	1.850	0.002	0.327	0.066	1.243	-0.005	0.370	0.072	1.375
REG	50	0.014	0.466	0.055	1.852	0.002	0.324	0.063	1.254	-0.005	0.365	0.067	1.396
IPW	50	0.017	0.471	0.049	1.872	0.017	0.329	0.058	1.262	0.298	1.667	0.057	5.079
UNC	50	-0.039	0.565	0.063	2.224	0.191	0.413	0.086	1.435	3.897	5.311	0.227	13.757
DR	200	0.010	0.235	0.056	0.918	-0.002	0.156	0.056	0.602	-0.003	0.169	0.050	0.675
REG	200	0.010	0.235	0.055	0.917	-0.002	0.154	0.056	0.600	-0.003	0.168	0.054	0.673
IPW	200	0.010	0.235	0.055	0.920	-0.001	0.158	0.056	0.613	0.020	0.676	0.032	2.369
UNC	200	-0.025	0.285	0.054	1.114	0.192	0.266	0.192	0.708	3.987	4.375	0.624	6.939
DR	1000	0.003	0.104	0.053	0.407	0.001	0.069	0.061	0.265	-0.002	0.074	0.043	0.296
REG	1000	0.003	0.104	0.050	0.407	0.001	0.069	0.062	0.264	-0.002	0.074	0.041	0.296
IPW	1000	0.003	0.104	0.051	0.407	0.000	0.070	0.061	0.270	-0.002	0.261	0.050	1.019
UNC	1000	-0.035	0.134	0.070	0.496	0.192	0.209	0.669	0.315	3.988	4.065	0.999	3.095

Notes: Rows labeled 'DR', 'REG', and 'IPW' use the doubly robust, regression, and inverse probability weighting estimators discussed in the main text, respectively. Rows labeled 'UNC' uses the unconditional DiD estimator that ignores covariates. Columns labeled '$ATT(2,2)$' are based on estimates for the group-time average treatment effect for group 2 in time period 2; columns labeled '$\theta_e(0)$' are based on estimates of the (aggregated) dynamic effects estimator for event time $e = 0$ and '$\theta_e(10)$' for event time $e = 10$. Finally, "Bias","RMSE", "Rej.P", and "CIL', stand for the average simulated bias, simulated root mean-squared errors, average of the simulated error for the 95% coverage probability, and 95% confidence interval length, respectively. Standard errors are computed using the multiplier-bootstrap proposed in the paper, with 1,000 bootstrap draws. Pointwise asymptotic-based critical values are used to compute 95% confidence intervals.

Table SC.3: Monte Carlo simulations based on 2,000 repetitions. Setup with $\mathcal{T} = 20$ and $G = 20$.

		$ATT(2,2)$				$\theta_e(0)$				$\theta_e(10)$			
	n	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL	Bias	RMSE	Rej.P	CIL
DR	50	-0.082	1.091	0.346	2.319	-0.005	0.301	0.053	1.088	-0.007	0.583	0.062	1.827
REG	50	-0.078	1.084	0.336	2.324	-0.031	0.509	0.060	1.249	0.043	2.415	0.063	2.156
IPW	50	-0.081	1.094	0.342	2.355	0.004	0.249	0.044	0.995	-0.034	1.078	0.012	4.831
UNC	50	-0.302	1.308	0.365	2.753	-0.051	0.315	0.076	1.097	-1.801	4.285	0.111	14.294
DR	200	0.005	0.489	0.079	1.829	0.001	0.115	0.048	0.457	0.009	0.180	0.055	0.701
REG	200	0.005	0.489	0.082	1.826	0.001	0.118	0.049	0.460	0.008	0.180	0.051	0.703
IPW	200	0.007	0.489	0.079	1.827	0.006	0.115	0.043	0.459	0.062	0.391	0.020	1.599
UNC	200	-0.220	0.629	0.110	2.203	-0.034	0.142	0.055	0.544	-1.720	2.584	0.158	7.461
DR	1000	0.004	0.206	0.057	0.806	0.000	0.050	0.049	0.199	-0.004	0.077	0.050	0.303
REG	1000	0.004	0.206	0.058	0.805	0.000	0.050	0.045	0.199	-0.004	0.077	0.045	0.303
IPW	1000	0.004	0.206	0.056	0.806	0.000	0.050	0.046	0.200	0.014	0.110	0.033	0.459
UNC	1000	-0.217	0.331	0.139	0.980	-0.039	0.071	0.091	0.241	-1.760	1.956	0.550	3.343

Notes: Rows labeled 'DR', 'REG', and 'IPW' use the doubly robust, regression, and inverse probability weighting estimators discussed in the main text, respectively. Rows labeled 'UNC' uses the unconditional DiD estimator that ignores covariates. Columns labeled '$ATT(2,2)$' are based on estimates for the group-time average treatment effect for group 2 in time period 2; columns labeled '$\theta_e(0)$' are based on estimates of the (aggregated) dynamic effects estimator for event time $e = 0$ and '$\theta_e(10)$' for event time $e = 10$. Finally, "Bias","RMSE", "Rej.P", and "CIL', stand for the average simulated bias, simulated root mean-squared errors, average of the simulated error for the 95% coverage probability, and 95% confidence interval length, respectively. Standard errors are computed using the multiplier-bootstrap proposed in the paper, with 1,000 bootstrap draws. Pointwise asymptotic-based critical values are used to compute 95% confidence intervals.

References

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.