Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

Online Appendix
The role of parallel trends in event study settings: An application to environmental economics

Michelle Marcus
Vanderbilt University

Pedro H. C. Sant'Anna
Vanderbilt University

September 3, 2020

Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

A Appendix: Additional details about GMM

In Section 4 of the main text, we introduced the GMM estimator that exploits all the moment restrictions imposed by the model. In doing so, we have used the stylized example to clarify the calculation details and have mentioned to important quantities: $g_a(W_i)$, which provides the summands for all (linearly independent) moment conditions, and the "selection matrix" A that is used to get the $ATT(g,t)$'s. We what follows, we provide a detailed characterization of these measures. First, in the context of our stylized example,

$$g_a(W_i) = \begin{pmatrix} \dfrac{(1-D_{i3})(Y_{i3}-Y_{i2})}{a_C^{prop}+a_4^{prop}} + \dfrac{G_{i3}Y_{i2}}{a_3^{prop}} - a_{3,3}(0) \\ \dfrac{(1-D_{i4})(Y_{i3}-Y_{i2})}{a_C^{prop}} + \dfrac{G_{i3}Y_{i2}}{a_3^{prop}} - a_{3,3}(0) \\ \dfrac{(1-D_{i4})(Y_{i4}-Y_{i3})}{a_C^{prop}} + a_{3,3}(0) - a_{3,4}(0) \\ \dfrac{(1-D_{i4})(Y_{i4}-Y_{i3})}{a_C^{prop}} + \dfrac{G_{i4}Y_{i3}}{a_4^{prop}} - a_{4,4}(0) \\ \dfrac{G_{i3}Y_{i3}}{a_3^{prop}} - a_{3,3}(1) \\ \dfrac{G_{i3}Y_{i4}}{a_3^{prop}} - a_{3,4}(1) \\ \dfrac{G_{i4}Y_{i4}}{a_4^{prop}} - a_{4,4}(1) \\ C_i - a_C^{prop} \\ G_{i3} - a_3^{prop} \\ G_{i4} - a_4^{prop} \end{pmatrix}. \tag{A.1}$$

Note that the (sample) moment restrictions (A.1) exploit the fact that $\mathbb{E}[1-D_3] = \alpha_C^{prop} + \alpha_4^{prop}$ and $\mathbb{E}[1-D_4] = \alpha_C^{prop}$.

Finally, the "selection matrix" A is given by

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

B Appendix: Bootstrap Implementation

In what section, we describe how one can implement the bootstrap procedure mentioned in Section 5. The main advantages of adopting it to conduct asymptotically valid inference are (a) it is robust against multiple-testing problems, (b) it is computationally simple and does not requires re-estimating all the parameters in each bootstrap draw, (c) it does not run the risk of having "empty" bootstrapped groups in some draws, and (d) it can be easily modified to conduct cluster-robust inference. The implementation of this bootstrap procedure follows closely the discussion in Callaway and Sant'Anna (2020).

First, denote the sample-analogue of $\phi_{ny+}(W_i; g, t)$ by

$$\widehat{\phi}_{ny+}(W_i; g, t) = \left(\frac{G_{ig}}{\mathbb{E}_n[G_g]} \left((Y_{it} - Y_{ig-1}) - \frac{\mathbb{E}_n[G_g \cdot (Y_t - Y_{g-1})}{\mathbb{E}_n[G_g]} \right) \right.$$
$$\left. - \sum_{s=g}^{t} \frac{(1-D_{is})(1-G_{ig})}{\mathbb{E}_n[(1-D_s)(1-G_g)]} \left(\Delta Y_{is} - \frac{\mathbb{E}_n[(1-D_s)(1-G_g)\cdot \Delta Y_s]}{\mathbb{E}_n[(1-D_s)(1-G_g)]} \right) \right),$$

where, for a generic random variable Z, $\mathbb{E}_n[Z] \equiv n^{-1}\sum_{i=1}^{n} Z_i$. Let $\widehat{\Phi}_{ny+}(W_i; t \geq g)$ denote the collection of $\widehat{\phi}_{ny+}(W_i; g, t)$ across all periods t and groups g such that $t \geq g$. 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 $\{W_i\}_{i=1}^{n}$. A popular example involves *iid* Bernoulli random variables $\{V_i\}$ with $P(V=1) = P(V=-1) = 0.5$.

We define $\widehat{ATT}_{ny+}(t \geq g)^*$, a bootstrap draw of $\widehat{ATT}_{ny+}(t \geq g)$, via

$$\widehat{ATT}_{ny+}(t \geq g)^* = \widehat{ATT}_{ny+}(t \geq g) + \frac{1}{n}\sum_{i=1}^{n} V_i \cdot \widehat{\Phi}_{ny+}(W_i; t \geq g)). \tag{B.1}$$

We now describe a practical bootstrap algorithm to compute studentized confidence bands that cover $ATT(g,t)$ simultaneously over all $t \geq g$ with a prespecified probability $1-\alpha$ in large samples. This is similar to the bootstrap procedure used in C&S.

Algorithm B.1. (1) *Draw a realization of* $\{V_i\}_{i=1}^{n}$. (2) *Compute* $\widehat{ATT}_{ny+}(t \geq g)^*$ *as in (B.1), denote its* (g,t)*-element as* $\widehat{ATT}^*_{ny+}(g,t)$, *and form a bootstrap draw of its limiting distribution as* $\hat{R}^*(g,t) = \sqrt{n}\left(\widehat{ATT}^*_{ny+}(g,t) - \widehat{ATT}_{ny+}(g,t)\right)$. (3) *Repeat steps 1-2 B times.* (4) *Compute a bootstrap estimator of the main diagonal of* $\Omega^{1/2}$ *such as the bootstrap interquartile range normalized by the interquartile range of the standard normal distribution,* $\widehat{\Omega}^{1/2}_{boot}(g,t) = (q_{0.75}(g,t) - q_{0.25}(g,t))/(z_{0.75} - z_{0.25})$, *where* $q_p(g,t)$ *is the pth sample quantile of the* $\hat{R}^*(g,t)$ *in the B draws, and* z_p *is the pth quantile of the standard normal distribution.* (5) *For each bootstrap draw, compute* $t-test_{t\geq g} = \max_{(g,t)}\left|\hat{R}^*(g,t)\right|\widehat{\Omega}_{boot}(g,t)^{-1/2}$. (6) *Construct* $\widehat{c}_{1-\alpha}$ *as the empirical* $(1-a)$*-quantile of the B bootstrap draws of* $t-test_{t\geq g}$. (7) *Construct the bootstrapped simultaneous confidence band for* $ATT(g,t)$, $t \geq g$, *as* $\widehat{C}_{bot}(g,t) = [\widehat{ATT}_{ny+}(g,t) \pm \widehat{c}_{1-\alpha}\widehat{\Omega}_{boot}(g,t)^{-1/2}/\sqrt{n}]$.

The next corollary states that the simultaneous confidence band for $ATT(g,t)$ described in Algorithm B.1 has correct asymptotic coverage.

Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

Corollary B.1. *Under the Assumptions of Proposition 3, for any $0 < \alpha < 1$, as $n \to \infty$,*

$$P\left(ATT\left(g,t\right) \in \widehat{C}_{boot}\left(g,t\right), \forall g \leq t\right) \to 1-\alpha,$$

where $\widehat{C}_{boot}\left(g,t\right)$ is as defined in Algorithm B.1.

Remark B.1. *We note that the multiplier bootstrap described above can be easily adapted to account for clustering, provided that the number of clusters is "large." In those cases, instead of drawing unit-specific V's to construct the bootstrapped estimator, one simply need to draw cluster-specific V's; see, e.g., Sherman and Le Cessie (2007), Kline and Santos (2012), Cheng et al. (2013).*

C Appendix: Proof of main results

Proof of Proposition 1: Part (i) follows directly from Theorem 3.4 and Theorem 4.5 in Newey and McFadden (1994). Part (ii) directly follows from Chamberlain (1987).□

Proof of Proposition 2: Notice that for identifying $ATT(g,t)$, $t \geq g$, the key term is $E[Y_t(0)|G_g = 1]$, as $\mathbb{E}[Y_t(1)|G_g = 1] = \mathbb{E}[Y_t|G_g = 1]$ for all $t \geq g$. Hence, once we identify $E[Y_t(0)|G_g = 1]$, the proof is complete.
Towards this end, Note that, for $2 \leq g \leq t \leq \mathcal{T}$,

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

where the first equality holds by adding and subtracting $E[Y_{t-1}(0)|G_g = 1]$, the second equality holds by Assumption 8, and the third equality by Assumption 3 and the fact that for $t \geq g$, $D_t = 0$ implies $G_g = 0$, so conditioning on $G_g = 0$ is redundant when $t \geq g$.
The key now is to identify $\mathbb{E}[Y_{t-1}(0)|G_g = 1]$. Towards this end, notice that, if $t = g$, $\mathbb{E}[Y_{t-1}(0)|G_g = 1] = \mathbb{E}[Y_{g-1}|G_g = 1]$ by Assumption 3, and therefore $\mathbb{E}[Y_{t-1}(0)|G_g = 1]$ is identified. When $t > g$, we can continue recursively in to (C.1) but starting with $\mathbb{E}[Y_{t-1}(0)|G_g = 1]$. As a result,

$$\mathbb{E}[Y_t(0)|G_g = 1] = \sum_{s=g}^{t} \mathbb{E}[\Delta Y_s|D_s = 0, G_g = 0] + \mathbb{E}[Y_{g-1}|G_g = 1], \tag{C.2}$$

concluding the proof, as there is no potential outcome in the right-hand side of (C.2).□

Proof of Proposition 3: Notice that, for all $t \geq g$,

$$\sqrt{n}\left(\widehat{ATT}_{ny+} - ATT\right)(g,t) \tag{C.3}$$

Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

$$
\begin{aligned}
&= \sqrt{n}\left(\frac{\mathbb{E}_n[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}_n[G_g]} - \frac{\mathbb{E}_n[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}_n[G_g]}\right) \\
&\quad - \sum_{s=g}^{t} \sqrt{n}\left(\frac{\mathbb{E}_n[(1-D_s)(1-G_g)\cdot \Delta Y_s]}{\mathbb{E}_n[(1-D_s)(1-G_g)]} - \frac{\mathbb{E}[(1-D_s)(1-G_g)\cdot \Delta Y_s]}{\mathbb{E}[(1-D_s)(1-G_g)]}\right) \\
&= \sqrt{n}A_n(g,t) - \sum_{s=g}^{t}\sqrt{n}B_n(g,s,t), \qquad \text{(C.4)}
\end{aligned}
$$

where, for a generic random variable Z, $\mathbb{E}_n[Z] = n^{-1}\sum_{i=1}^{n} Z_i$ denote the sample mean of Z. From simple algebra (first equality) and the continuous mapping theorem (second equality), it follows that

$$
\begin{aligned}
\sqrt{n}A_n(g,t) &= \frac{1}{\mathbb{E}[G_g]}\sqrt{n}\mathbb{E}_n\left[G_g \cdot (Y_t - Y_{g-1}) - \mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]\right] \\
&\quad - \frac{\mathbb{E}_n[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}_n[G_g]\mathbb{E}[G_g]}\sqrt{n}\mathbb{E}_n[G_g - \mathbb{E}[G_g]] \\
&= \frac{1}{\mathbb{E}[G_g]}\sqrt{n}\mathbb{E}_n\left[G_g \cdot (Y_t - Y_{g-1}) - \mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]\right] \\
&\quad - \frac{1}{\mathbb{E}[G_g]}\frac{\mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}[G_g]}\sqrt{n}\mathbb{E}_n[G_g - \mathbb{E}[G_g]] + o_p(1).
\end{aligned}
$$

Next, by exploiting that sample mean are linear operators, it follows that

$$
\begin{aligned}
&\sqrt{n}A_n(g,t) \\
&= \sqrt{n}\mathbb{E}_n\left[\frac{G_g \cdot (Y_t - Y_{g-1})}{\mathbb{E}[G_g]} - \frac{\mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}[G_g]} - \frac{\mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}[G_g]}\frac{(G_g - \mathbb{E}[G_g])}{\mathbb{E}[G_g]}\right] + o_p(1) \\
&= \sqrt{n}\mathbb{E}_n\left[\frac{G_g}{\mathbb{E}[G_g]}\left((Y_t - Y_{g-1}) - \frac{\mathbb{E}[G_g \cdot (Y_t - Y_{g-1})]}{\mathbb{E}[G_g]}\right)\right] + o_p(1) \\
&= \sqrt{n}\mathbb{E}_n\left[\phi^{g}_{ny+}(W;g,t)\right] + o_p(1). \qquad \text{(C.5)}
\end{aligned}
$$

Following exactly the same steps, it follows that

$$
\begin{aligned}
\sqrt{n}B_n(g,s,t) &= \mathbb{E}_n\left[\frac{(1-D_s)(1-G_g)}{\mathbb{E}[(1-D_s)(1-G_g)]}\left(\Delta Y_t - \frac{\mathbb{E}[(1-D_s)(1-G_g)\cdot \Delta Y_t]}{\mathbb{E}[(1-D_s)(1-G_g)]}\right)\right] + o_p(1) \\
&= \mathbb{E}_n\left[\phi^{s}_{ny+}(W;g,t)\right] + o_p(1). \qquad \text{(C.6)}
\end{aligned}
$$

By combining (C.4), (C.5) and (C.6), it then follows that

$$
\begin{aligned}
\sqrt{n}\left(\widehat{ATT}_{ny+} - ATT\right)(g,t) &= \sqrt{n}\mathbb{E}_n\left[\phi^{g}_{ny+}(W;g,t) - \sum_{s=g}^{t}\phi^{s}_{ny+}(W;g,t)\right] + o_p(1) \\
&= \sqrt{n}\mathbb{E}_n\left[\phi_{ny+}(W;g,t)\right] + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{ny+}(W_i;g,t) + o_p(1),
\end{aligned}
$$

concluding the proof of (36). The proof of (37) now follows from the application of the

Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

Lindeberg–Lévy Central Limit Theorem. □

D Appendix: Additional details about application

Compliance data on inspections, violations, and enforcement actions comes from the National Enforcement and Compliance History Online (ECHO) database from 1976 to 2008.[1] These data were originally housed in the Permit Compliance System (PCS) database and have since been transitioned to the newer Integrated Compliance Information System, National Pollutant Discharge Elimination system (ICIS-NPDES). This paper uses data from the ICIS-NPDES, rather than the PCS. Because of this transition, the format of the data and available variables differ somewhat from Grooms (2015), which may account for some small discrepancies in the replication.

Following Grooms (2015), we construct binary indicators at the facility level, which equal 1 if any inspection, violation, or enforcement action occurs at the facility in each year. These indicators are then summed across all facilities in a state-year and normalized by the total number of facilities in the state. We define the total number of facilities in the state as the number of unique facilities ever observed in the state. These measures can be interpreted as the fraction of total facilities with at least one inspection, violation, or enforcement action in a state and year.

State authorization data comes from the EPA, which records the date when each state was authorized to perform each phase of the CWA.[2] As many states receive authorization for the first four phases in the same year, we define the year of authorization as the year in which the state was authorized to perform the first phase of the program, administering individual NPDES permits.

Finally, corruption data come from the "Report to Congress on the Activities and Operations of the Public Integrity Section" (PIN data), which records the federal public corruption convictions in each state and year. We average convictions per capita within a state across years and define states above the median as "corrupt" states. Here, to compute the median, we use all available states in the data including those "never-treated" and "always-treated." This way, our definition of a "corrupt state" matches that used by Grooms (2015).

[1]National Pollutant Elimination Discharge System (NPDES) data downloads available at https://echo.epa.gov/tools/data-downloads#downloads.

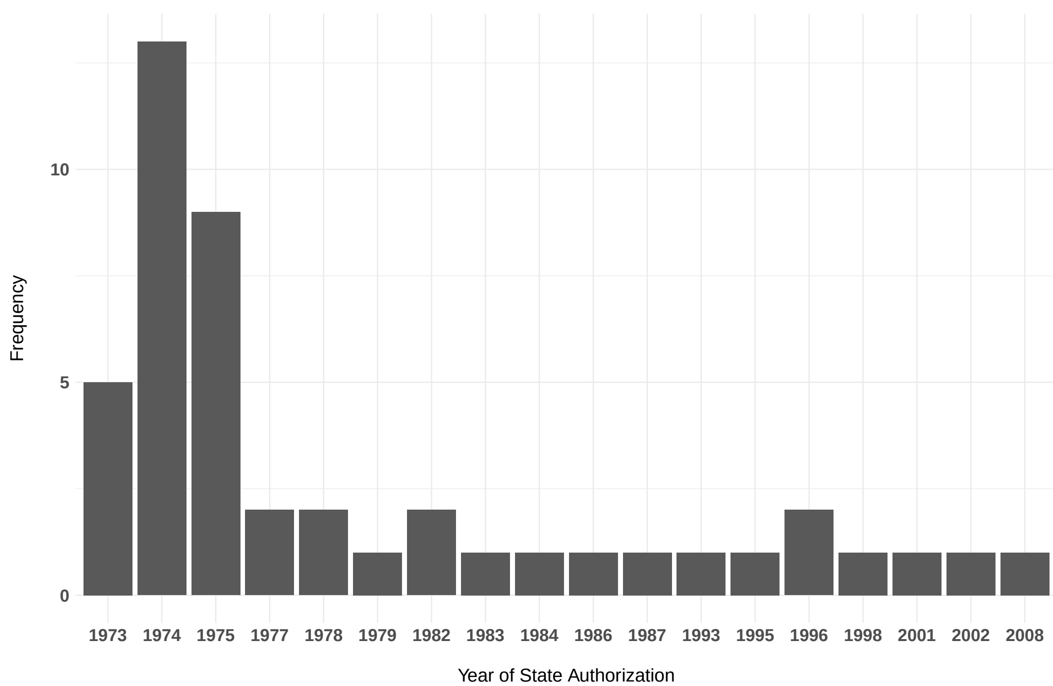
[2]These data can be accessed at https://www.epa.gov/npdes/npdes-state-program-information.

Supplemental Material (not copyedited or formatted) for: Michelle Marcus, Pedro H. C. Sant'Anna. 2021.
"The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics."
Journal of the Association of Environmental and Resource Economists 8(2). DOI: https://doi.org/10.1086/711509.

Table D.1: Summary Statistics Replication

	Grooms	Marcus & Sant'Anna	
		All	Drop "always-treated"
	(1)	(2)	(3)
Inspection Rate	0.067	0.058	0.053
	(0.067)	(0.066)	(0.059)
Violation Rate	0.028	0.024	0.028
	(0.054)	(0.049)	(0.050)
Enforcement Rate	0.013	0.0098	0.012
	(0.020)	(0.018)	(0.020)
Total Facilities	8,180	8,303	9,693
	(10,049)	(8,952)	(10,588)
Corruption Score	0.292	0.289	0.324
	(0.299)	(0.133)	(0.137)
Observations	1,650	1,650	759

Figure D.1: Time-distribution of State Authorization



Notes: Shows the distribution of state authorization over time. State authorization is defined as the year in which the state was authorized to administer individual NPDES permits.

References

Callaway, Brantly and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *arXiv preprint arXiv:1803.09015*, 2020.

Chamberlain, Gary, "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 1987, *34* (3).

Cheng, Guang, Zhuqing Yu, and Jianhua Z. Huang, "The cluster bootstrap consistency in generalized estimating equations," *Journal of Multivariate Analysis*, 2013, *115.*

Grooms, Katherine K, "Enforcing the Clean Water Act: The effect of state-level corruption on compliance," *Journal of Environmental Economics and Management*, 2015, *73.*

Kline, Patrick and Andres Santos, "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods*, 2012, *1* (1).

Newey, Whitney K. and Daniel McFadden, "Large sample estimation and hypothesis testing," in "Handbook of Econometrics," Vol. 4, Amsterdam: North-Holland: Elsevier, 1994, chapter 36.

Sherman, Michael and Saskia Le Cessie, "A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models," *Communications in Statistics - Simulation and Computation*, 2007, *26* (3).