Supplementary Material

Difference-in-differences with as few as two cross-sectional units – A new perspective to the democracy-growth debate

Gilles Boevi Koumou Emmanuel Selorm Tsyawo

The supplementary material provides detailed proofs and extensions of the theoretical results presented in the main text. Specifically, Appendix S.1 contains proofs of all theoretical results in the main text. Appendix S.2 generalises the baseline model in (3.2) to accommodate both stochastic and deterministic trends, and extends the two-unit framework to settings with multiple treated units. Supporting lemmata and propositions referenced in the main text are provided in Appendices S.3 and S.4, respectively. Further discussions on related literature, pre-trends testing adapted to the baseline setting of this paper, and the empirical setting are presented in Appendix S.5, while simulation results appear in Appendix S.6.

S.1 Proofs of main results

S.1.1 Proof of Proposition 1

Proposition 1. *Let $\varphi(\cdot)$ be some possibly unbounded function $\varphi : [-\mathcal{T}] \cup [T] \to \mathbb{R}$ and $\eta > 1/2$, then untreated potential outcomes following (2.3) satisfy Assumption 1.*

Proof:
The assumption $\mathbb{E}[e_t \mid D = d] = 0,\ d \in \{0,1\}$ and (2.3) imply

$$\mathbb{E}[Y_t(0) \mid D = d] = \alpha_0 + (\alpha_1 - \alpha_0)d + \varphi(t) + \nu_0(t, d; \eta).$$

Thus,

$$\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D = d] = \varphi(t) - \varphi(\tau) + \nu_0(t, d; \eta) - \nu_0(\tau, d; \eta)$$

for any pair $(\tau, t) \in [-\mathcal{T}] \times [T]$, and

$$\begin{aligned}\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D = 1] - \mathbb{E}[Y_t(0) - Y_\tau(0) \mid D = 0] &= \big(\nu_0(t,1;\eta) - \nu_0(\tau,1;\eta)\big) - \big(\nu_0(t,0;\eta) - \nu_0(\tau,0;\eta)\big) \\ &= \frac{1}{4}\big|1 + 0.5\sin(t)\big|t^{-\eta} + \frac{1}{4}\big|1 + 0.5\sin(-\tau)\big|(-\tau)^{-\eta}.\end{aligned}$$

It then follows that

$$\begin{aligned}&\frac{1}{\mathcal{T}T}\sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\big(\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D = 1] - \mathbb{E}[Y_t(0) - Y_\tau(0) \mid D = 0]\big) \\ &\qquad = \frac{1}{4T}\sum_{t=1}^{T}\big|1 + 0.5\sin(t)\big|t^{-\eta} + \frac{1}{4\mathcal{T}}\sum_{-\tau=1}^{\mathcal{T}}\big|1 + 0.5\sin(-\tau)\big|(-\tau)^{-\eta}.\end{aligned}$$

Since $\left|1+0.5\sin(t)\right| \leq (3/2)\,\forall\, t \geq 1$, then for $\eta > 0$,

$$\begin{aligned}\left|\frac{1}{T}\sum_{t=1}^{T}\left|1+0.5\sin(t)\right|t^{-\eta}\right| &\leq \frac{3}{2T}\sum_{t=1}^{T}t^{-\eta}\\ &\leq \frac{3}{2T}\Big(1+\int_{1}^{T}t^{-\eta}dt\Big)\\ &= \frac{3}{2}\begin{cases}T^{-1}\Big(1-\frac{1}{1-\eta}\Big)+T^{-\eta}\frac{1}{(1-\eta)}, & \eta\neq 1\\ T^{-1}+T^{-1}\log(T), & \eta=1\end{cases}.\end{aligned}$$

By the same token,

$$\left|\frac{1}{\mathcal{T}}\sum_{-\tau=1}^{\mathcal{T}}\left|1+0.5\sin(-\tau)\right|(-\tau)^{-\eta}\right| \leq \frac{3}{2}\begin{cases}\mathcal{T}^{-1}\Big(1-\frac{1}{1-\eta}\Big)+\mathcal{T}^{-\eta}\frac{1}{(1-\eta)}, & \eta\neq 1\\ \mathcal{T}^{-1}+\mathcal{T}^{-1}\log(\mathcal{T}), & \eta=1\end{cases}$$

Since $T^{-1}\log(T) = o(T^{-\bar{\eta}})\,\forall\,\bar{\eta} < 1$ and $\eta > 1/2$, it follows by the triangle inequality that

$$\frac{1}{\mathcal{T}T}\sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\big(\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]\big) = \mathcal{O}\Big((\mathcal{T}\wedge T)^{-(1/2+(\dot{\eta}-1/2))}\Big),\ \dot{\eta}\in(1/2,1).$$

Set $\gamma = \dot{\eta}-(1/2)$ and observe that $\gamma \in (0,1/2)$, which implies $\gamma > 0$. This completes the proof. □

S.1.2 Proof of Theorem 1

Theorem 1. *Let Assumptions 1 to 3 hold, then*

(a) $\sup_{\{w,\psi\}\in\mathcal{W}^2}|ATT_{\omega,T}-\widetilde{ATT}_n^{w,\psi}| = \mathcal{O}\big(n^{-(1/2+\gamma\wedge\delta)}\big)$ *and*

(b) $\sup_{\{\psi,\phi\}\in\mathcal{W}^2}|\widetilde{ATT}_n^{w,\psi}-\widetilde{ATT}_n^{w,\phi}| = \mathcal{O}\big(n^{-(1/2+\gamma\wedge\delta)}\big)$ *uniformly in* $w\in\mathcal{W}$*, where the constants* $\gamma > 0$ *and* $\delta > 0$ *are defined in Assumptions 1 and 2, respectively.*

Proof:

Part (a): Recall $R_{\tau,t} = \mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]-\mathbb{E}[Y_\tau(1)-Y_\tau(0)\mid D=1]$, then for any pair $(\tau,t)\in[-\mathcal{T}]\times[T]$,

$$\begin{aligned}ATT(t) =&\mathbb{E}[Y_t(1)-Y_t(0)\mid D=1]\\ =&\mathbb{E}[Y_t(1)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]\\ =&\mathbb{E}[Y_t(1)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]\\ &-\big(\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]\big)\\ =&\mathbb{E}[Y_t(1)-Y_\tau(1)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]\\ &-\big\{\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]-\mathbb{E}[Y_\tau(1)-Y_\tau(0)\mid D=1]\big\}\\ =&\mathbb{E}[Y_{1,t}-Y_{1,\tau}]-\mathbb{E}[Y_{0,t}-Y_{0,\tau}]\\ &-\big(\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=1]-\mathbb{E}[Y_t(0)-Y_\tau(0)\mid D=0]-\mathbb{E}[Y_\tau(1)-Y_\tau(0)\mid D=1]\big)\\ =:&\,ATT_{\tau,t}-R_{\tau,t}.\end{aligned}\tag{S.1.1}$$

The first equality follows by definition; the second subtracts and adds $\mathbb{E}[Y_\tau(0) \mid D=1]$; the third subtracts and adds $\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D=0]$; the fourth subtracts and adds $\mathbb{E}[Y_\tau(1) \mid D=1]$; the fifth uses the fact that treated potential outcomes are observed for the treated unit and untreated potential outcomes are observed for the untreated unit; and the last line follows by definition.

Since pre-treatment and post-treatment weights each sum to one by construction, it follows from the above decomposition of $ATT(t)$ that

$$\begin{aligned} ATT_{\omega,T} :&= \sum_{t=1}^{T} w_T(t) ATT(t) = \Big(\sum_{-\tau=1}^{\mathcal{T}} \psi_{\mathcal{T}}(-\tau)\Big) \sum_{t=1}^{T} w_T(t) ATT(t) \\ &= \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) ATT_{\tau,t} - \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) R_{\tau,t}. \end{aligned}$$

By the triangle inequality and Assumption 1,

$$\begin{aligned} &\Big| \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) \Big(\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D=1] - \mathbb{E}[Y_t(0) - Y_\tau(0) \mid D=0] \Big) \Big| \\ &\le \Big(T \sup_{w \in \mathcal{W}} \max_{t \in [T]} w_T(t) \Big) \times \Big(\mathcal{T} \sup_{\psi \in \mathcal{W}} \max_{\tau \in [-\mathcal{T}]} \psi_{\mathcal{T}}(-\tau) \Big) \\ &\quad \times \Big| \frac{1}{\mathcal{T} T} \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \Big(\mathbb{E}[Y_t(0) - Y_\tau(0) \mid D=1] - \mathbb{E}[Y_t(0) - Y_\tau(0) \mid D=0] \Big) \Big| \\ &= \mathcal{O}\big((\mathcal{T} \wedge T)^{-(1/2+\gamma)} \big). \end{aligned}$$

In a similar vein,

$$\begin{aligned} \Big| \sum_{-\tau=1}^{\mathcal{T}} \psi_{\mathcal{T}}(-\tau) \big(\mathbb{E}[Y_\tau(1) - Y_\tau(0) \mid D=1] \big) \Big| &\le \Big(\mathcal{T} \sup_{\psi \in \mathcal{W}} \max_{\tau \in [-\mathcal{T}]} \psi_{\mathcal{T}}(-\tau) \Big) \Big| \frac{1}{\mathcal{T}} \sum_{-\tau=1}^{\mathcal{T}} \big(\mathbb{E}[Y_\tau(1) - Y_\tau(0) \mid D=1] \big) \Big| \\ &= \mathcal{O}(\mathcal{T}^{-(1/2+\delta)}) \end{aligned}$$

holds by Assumption 2. Thanks to the above and Assumption 3,

$$\begin{aligned} R_n^{w,\psi} := \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) R_{\tau,t} &= \mathcal{O}\big((\mathcal{T} \wedge T)^{-(1/2+\gamma)} \big) + \mathcal{O}(\mathcal{T}^{-(1/2+\delta)}) \\ &= \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)}) \text{ uniformly in } \{w,\psi\} \in \mathcal{W}^2. \qquad \text{(S.1.2)} \end{aligned}$$

To see why (S.1.2) holds, note that

$$\begin{aligned} \mathcal{O}\big((\mathcal{T} \wedge T)^{-(1/2+\gamma)} \big) + \mathcal{O}(\mathcal{T}^{-(1/2+\delta)}) &\asymp \frac{1}{(\mathcal{T} \wedge T)^{(1/2+\gamma)}} + \frac{1}{\mathcal{T}^{(1/2+\delta)}} \\ &\lesssim \frac{1}{\mathcal{T}^{(1/2+\gamma\wedge\delta)} \wedge T^{(1/2+\gamma)}} \lesssim \frac{1}{\mathcal{T}^{(1/2+\gamma\wedge\delta)} \wedge T^{(1/2+\gamma\wedge\delta)}} \\ &= \Big(\frac{1}{\mathcal{T} \wedge T} \Big)^{(1/2+\gamma\wedge\delta)} = \Big(\frac{n^{-1}}{(1-\lambda_n) \wedge \lambda_n} \Big)^{(1/2+\gamma\wedge\delta)} \\ &\asymp n^{-(1/2+\gamma\wedge\delta)} \end{aligned}$$

uniformly in n subject to Assumption 3.

Averaging both sides of (S.1.1) and applying the foregoing gives

$$\begin{aligned}\sum_{t=1}^{T} w_T(t)ATT(t) &= \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\psi_{\mathcal{T}}(-\tau)w_T(t)ATT_{\tau,t} - \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\psi_{\mathcal{T}}(-\tau)w_T(t)R_{\tau,t}\\ &= \widetilde{ATT}_n^{w,\psi} - R_n^{w,\psi}\\ &= \widetilde{ATT}_n^{w,\psi} + \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)}) \qquad \text{(S.1.3)}\end{aligned}$$

uniformly in $\{w,\psi\}\in\mathcal{W}^2$. $\widetilde{ATT}_n^{w,\psi}$ is identified from the data sampling process. Identification therefore holds *modulo* $R_n^{w,\psi} = \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)})$.

Part (b): Fix $w\in\mathcal{W}$ and observe that since pre-treatment weights must sum to one and post-treatment weights must also sum to one,

$$\begin{aligned}\sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\psi_{\mathcal{T}}(-\tau)w_T(t)ATT(t) &= \Big(\sum_{-\tau=1}^{\mathcal{T}}\psi_{\mathcal{T}}(-\tau)\Big)\sum_{t=1}^{T}w_T(t)ATT(t)\\ &= \sum_{t=1}^{T}w_T(t)ATT(t) =: ATT_{\omega,T}\end{aligned}$$

, i.e., $ATT_{\omega,T}$ does not depend on $\psi\in\mathcal{W}$ for $\tau\leq -1$. Thus, for any $\{\psi,\phi\}\in\mathcal{W}^2$,

$$\begin{aligned}\widetilde{ATT}_n^{w,\psi} - \widetilde{ATT}_n^{w,\phi} &= \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}w_T(t)\big(\psi_{\mathcal{T}}(-\tau)ATT_{\tau,t} - ATT(t) + ATT(t) - \phi_{\mathcal{T}}(-\tau)ATT_{\tau,t}\big)\\ &= \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\psi_{\mathcal{T}}(-\tau)w_T(t)\big(ATT_{\tau,t} - ATT(t)\big)\\ &\quad - \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\phi_{\mathcal{T}}(-\tau)w_T(t)\big(ATT_{\tau,t} - ATT(t)\big)\\ &= \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\psi_{\mathcal{T}}(-\tau)w_T(t)R_{\tau,t} - \sum_{-\tau=1}^{\mathcal{T}}\sum_{t=1}^{T}\phi_{\mathcal{T}}(-\tau)w_T(t)R_{\tau,t}\\ &= R_n^{w,\psi} - R_n^{w,\phi}.\end{aligned}$$

The conclusion follows from the triangle inequality and (S.1.2) in Part (a) above.

□

S.1.3 Proof of Theorem 2

Theorem 2. *Under Assumptions 1 to 7,* $s_{\omega,n}^{-1}(\widehat{ATT}_{\omega,T} - ATT_{\omega,T}) \xrightarrow{d} \mathcal{N}(0,1)$ *uniformly in* $\mathcal{W}^2$.

Proof:

Under the conditions of Theorem 1, (S.1.1), and (S.1.3),

$$\begin{aligned} ATT_{\omega,T} &= \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) \Big(\mathbb{E}[Y_{1,t} - Y_{1,\tau}] - \mathbb{E}[Y_{0,t} - Y_{0,\tau}] \Big) - \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) R_{\tau,t} \\ &= \sum_{t=-\mathcal{T}}^{T} \omega_n(t) \mathbb{E}[X_t] - \sum_{-\tau=1}^{\mathcal{T}} \sum_{t=1}^{T} \psi_{\mathcal{T}}(-\tau) w_T(t) R_{\tau,t} \\ &= \sum_{t=-\mathcal{T}}^{T} \omega_n(t) \mathbb{E}[X_t] + \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)}) \end{aligned}$$

uniformly in $\mathcal{W}^2$. In addition to (4.1),

$$\widehat{ATT}_{\omega,T} - ATT_{\omega,T} = \sum_{t=-\mathcal{T}}^{T} \omega_n(t)(X_t - \mathbb{E}[X_t]) + \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)}) \tag{S.1.4}$$

uniformly in $\mathcal{W}^2$. Further,

$$\begin{aligned} s_{\omega,n}^{-1}(\widehat{ATT}_{\omega,T} - ATT_{\omega,T}) &= s_{\omega,n}^{-1} \sum_{t=-\mathcal{T}}^{T} \omega_n(t)(X_t - \mathbb{E}[X_t]) + \mathcal{O}((\sqrt{n} s_{\omega,n})^{-1} n^{-(\gamma\wedge\delta)}) \\ &= \sum_{t=-\mathcal{T}}^{T} X_{nt} + \mathcal{O}(n^{-(\gamma\wedge\delta)}) \end{aligned}$$

by Assumption 5 uniformly in $\mathcal{W}^2$. The conclusion follows from Lemma S.3.2.

□

S.1.4 Proof of Corollary 1

Corollary 1. *Suppose Assumptions 1 to 4, 6 and 7 hold for each of the J unique treated-control pairs. Suppose further that Assumption 5-Ext holds, then $(\tau_J' S_{\omega,n} \tau_J)^{-1/2} \tau_J' (\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) \xrightarrow{d} \mathcal{N}(0,1)$ uniformly in $\mathcal{W}^2$ for all $\tau_J \in \mathbb{S}^J$.*

Proof: The following expansion follows from (S.1.4) uniformly in $\mathcal{W}^2$:

$$\tau_J'(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) = \sum_{t=-\mathcal{T}}^{T} \omega_n(t) \tau_J' (X_{\cdot t} - \mathbb{E}[X_{\cdot t}]) + \mathcal{O}(n^{-(1/2+\gamma\wedge\delta)})$$

where $X_{\cdot t} := (X_{1t}, \ldots, X_{Jt})'$ and $X_{jt},\ j \in [J]$ is the difference between the outcomes of the treated unit and the j'th control unit. Further,

$$\begin{aligned}
&(\tau_J' S_{\omega,n}\tau_J)^{-1/2}\tau_J'(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) \\
&= (\tau_J' S_{\omega,n}\tau_J)^{-1/2}\sum_{t=-\mathcal{T}}^{T}\omega_n(t)\tau_J'(X_{\cdot t} - \mathbb{E}[X_{\cdot t}]) + (n\tau_J' S_{\omega,n}\tau_J)^{-1/2}\times\mathcal{O}(n^{-(\gamma\wedge\delta)}) \\
&= (\tau_J' S_{\omega,n}\tau_J)^{-1/2}\sum_{t=-\mathcal{T}}^{T}\omega_n(t)\tau_J'(X_{\cdot t} - \mathbb{E}[X_{\cdot t}]) + \mathcal{O}(n^{-(\gamma\wedge\delta)})
\end{aligned}$$

since $\inf_{\tau_J\in\mathbb{S}^J} n\tau_J' S_{\omega,n}\tau_J =: \rho_{\min}(nS_{\omega,n}) \geq \epsilon^2$ by Assumption 5-Ext. The conclusion follows similarly from the proof of Lemma S.3.2 and is therefore omitted. □

S.1.5 Proof of Proposition 2

Proposition 2. *Under the assumptions of Corollary 1, $h_J^* := S_{\omega,n}^{-1}\mathbb{1}_J(\mathbb{1}_J' S_{\omega,n}^{-1}\mathbb{1}_J)^{-1}$ delivers the most efficient estimator of $ATT_{\omega,T}$ among all $h_J \in \mathbb{H}^J$.*

Proof: Consider the classical minimum distance problem (4.2) (see, e.g., Wooldridge (2010, Sect. 14.5)) with an arbitrary positive definite weight matrix H_n. The solution is given by Equation (4.2). Under the conditions of Corollary 1,

$$\mathbb{V}[\widehat{ATT}^{h_J}_{\omega,T}]^{-1/2}(\widehat{ATT}^{h_J}_{\omega,T} - ATT_{\omega,T}) = (n\mathbb{V}[\widehat{ATT}^{h_J}_{\omega,T}])^{-1/2}\sqrt{n}(\widehat{ATT}^{h_J}_{\omega,T} - ATT_{\omega,T}) \xrightarrow{d} \mathcal{N}(0,1)$$

where $\mathbb{V}[\widehat{ATT}^{h_J}_{\omega,T}] = (\mathbb{1}_J' H_n\mathbb{1}_J)^{-2}\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J$. Since $P_n := S_{\omega,n}^{1/2}H_n\mathbb{1}_J(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}\mathbb{1}_J' H_n S_{\omega,n}^{1/2}$ is idempotent,

$$\begin{aligned}
&\mathbb{1}_J' S_{\omega,n}^{-1}\mathbb{1}_J - (\mathbb{1}_J' H_n\mathbb{1}_J)(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}(\mathbb{1}_J' H_n\mathbb{1}_J) \\
&= \mathbb{1}_J'(S_{\omega,n}^{-1} - H_n\mathbb{1}_J(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}\mathbb{1}_J' H_n)\mathbb{1}_J \\
&= \mathbb{1}_J' S_{\omega,n}^{-1/2}S_{\omega,n}^{1/2}(S_{\omega,n}^{-1} - H_n\mathbb{1}_J(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}\mathbb{1}_J' H_n)S_{\omega,n}^{1/2}S_{\omega,n}^{-1/2}\mathbb{1}_J \\
&= \mathbb{1}_J' S_{\omega,n}^{-1/2}(\mathrm{I} - S_{\omega,n}^{1/2}H_n\mathbb{1}_J(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}\mathbb{1}_J' H_n S_{\omega,n}^{1/2})S_{\omega,n}^{-1/2}\mathbb{1}_J \\
&= ||(\mathrm{I} - P_n)S_{\omega,n}^{-1/2}\mathbb{1}_J||^2.
\end{aligned}$$

From the foregoing,

$$\Big(\mathbb{1}_J' S_{\omega,n}^{-1}\mathbb{1}_J - (\mathbb{1}_J' H_n\mathbb{1}_J)(\mathbb{1}_J' H_n S_{\omega,n} H_n\mathbb{1}_J)^{-1}(\mathbb{1}_J' H_n\mathbb{1}_J)\Big)/n = ||(\mathrm{I} - P_n)(nS_{\omega,n})^{-1/2}\mathbb{1}_J||^2 \geq 0.$$

Thus, the optimal choice of H_n is $S_{\omega,n}^{-1}$, which yields the optimal linear combination $h_J^* := \frac{1}{\mathbb{1}_J' S_{\omega,n}^{-1}\mathbb{1}_J}S_{\omega,n}^{-1}\mathbb{1}_J$ in $\mathbb{H}^J$. This proves the assertion as claimed. □

S.1.6 Proof of Theorem 3

Theorem 3. *Let Assumptions 3 to 8 hold, then (a) under $\mathbb{H}_o$, $\widehat{Q}_{\omega,n} \xrightarrow{d} \chi^2_{J-1}$; (b) under $\mathbb{H}_{an}$ and if $F_n^\omega \notin \mathcal{F}_o^\omega$, $\widehat{Q}_{\omega,n} \xrightarrow{d} \chi^2_{J-1}(\theta)$; and (c) under $\mathbb{H}_a$ and if $F_n^\omega \notin \mathcal{F}_o^\omega$, $\widehat{Q}_{\omega,n} \to \infty$ uniformly in $\mathcal{W}^2$.*

Proof: Recalling $h_J'\mathbb{1}_J = 1$ for all $h_J \in \mathbb{H}^J$, the following expansion holds:

$$\begin{aligned}\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J\widehat{ATT}^*_{\omega,T} &= (\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) - \mathbb{1}_J(\widehat{ATT}^*_{\omega,T} - ATT_{\omega,T})\\ &= (\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) - \mathbb{1}_J\hat{h}_J^{*\prime}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T})\\ &= [\mathrm{I}_J - \mathbb{1}_J\hat{h}_J^{*\prime}](\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T})\\ &= [\mathrm{I}_J - \mathbb{1}_J(\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}](\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T})\\ &= [\widehat{S}_{\omega,n}^{1/2} - \mathbb{1}_J(\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1/2}]\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T})\end{aligned}$$

where $\mathbb{H}^J \ni h_J^* := \widehat{S}_{\omega,n}^{-1}\mathbb{1}_J(\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}$. Pre-multiplying the above by $\widehat{S}_{\omega,n}^{-1/2}$, it follows that

$$\begin{aligned}\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J\widehat{ATT}^*_{\omega,T}) &= [\mathrm{I}_J - \widehat{S}_{\omega,n}^{-1/2}\mathbb{1}_J(\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1/2}]\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T})\\ &=: (\mathrm{I}_J - \widehat{P}_n)\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}).\end{aligned}$$

Observe that $\widehat{P}_n := \widehat{S}_{\omega,n}^{-1/2}\mathbb{1}_J(\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}\mathbb{1}_J'\widehat{S}_{\omega,n}^{-1/2}$ is a projection matrix, thus $\mathrm{I}_J - \widehat{P}_n$ is idempotent. Denote its probability limit, in view of Assumption 8 and the continuous mapping theorem, as: $\widetilde{P}_n := \widetilde{S}_{\omega,n}^{-1/2}\mathbb{1}_J(\mathbb{1}_J'\widetilde{S}_{\omega,n}^{-1}\mathbb{1}_J)^{-1}\mathbb{1}_J'\widetilde{S}_{\omega,n}^{-1/2}$. Moreover, since $\mathrm{rank}[\widetilde{S}_{\omega,n}^{-1/2}\mathbb{1}_J] = 1$, $\mathrm{rank}[\mathrm{I}_J - \widetilde{P}_n] = J - 1$. Also, denoting $\widehat{\mathrm{I}}_n := \widehat{S}_{\omega,n}^{-1/2}\widetilde{S}_{\omega,n}^{1/2}$, the following expansion follows from (S.1.4):

$$\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) = \widehat{\mathrm{I}}_n\widetilde{S}_{\omega,n}^{-1/2}\sum_{t=-\mathcal{T}}^{T}\omega_n(t)(X_{\cdot t} - \mathbb{E}[X_{\cdot t}]) + \widehat{\mathrm{I}}_n(n\widetilde{S}_{\omega,n})^{-1/2}\sqrt{n}R_{\cdot n}.$$

Under Assumption 8, $||\widehat{\mathrm{I}}_n - \mathrm{I}_J|| \xrightarrow{p} 0$. In addition to the conditions of Lemma S.3.2, the continuous mapping theorem, and the Cramér-Wold device,

$$\widehat{\mathrm{I}}_n\widetilde{S}_{\omega,n}^{-1/2}\sum_{t=-\mathcal{T}}^{T}\omega_n(t)(X_{\cdot t} - \mathbb{E}[X_{\cdot t}]) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_J).$$

Part (a): Under $\mathbb{H}_o$, $||\widehat{\mathrm{I}}_n\widetilde{S}_{\omega,n}^{-1/2}R_{\cdot n}^{w,\psi}|| \xrightarrow{p} 0$ thus $\widehat{S}_{\omega,n}^{-1/2}(\widehat{ATT}_{\omega,\cdot T} - \mathbb{1}_J ATT_{\omega,T}) \xrightarrow{d} \mathcal{N}(0, \mathrm{I}_J)$ and the conclusion follows.

Part (b): Under $\mathbb{H}_{an}$,

$$\widehat{\mathrm{I}}_n(n\widetilde{S}_{\omega,n})^{-1/2}\sqrt{n}R_{\cdot n}^{w,\psi} = \widehat{\mathrm{I}}_n(n\widetilde{S}_{\omega,n})^{-1/2}\tau_J^R||\sqrt{n}R_{\cdot n}^{w,\psi}||$$

where $\tau_J^R = \sqrt{n}R_{\cdot n}^{w,\psi}||\sqrt{n}R_{\cdot n}^{w,\psi}||^{-1}$. Under the condition that τ_J^R does not lie in the null space of $(n\widetilde{S}_{\omega,n})^{-1/2}[\mathrm{I}_J - \widetilde{P}_n](n\widetilde{S}_{\omega,n})^{-1/2}$ as $n \to \infty$, the conclusion for this part follows.

Part (c): Under $\mathbb{H}_a$ and the condition that τ_J^R does not lie in the null space of $(n\widetilde{S}_{\omega,n})^{-1/2}[\mathrm{I}_J - \widetilde{P}_n](n\widetilde{S}_{\omega,n})^{-1/2}$ as $n \to \infty$, $||[\mathrm{I}_J - \widetilde{P}_n](n\widetilde{S}_{\omega,n})^{-1/2}\tau_J^R||^2 \cdot ||\sqrt{n}R_{\cdot n}^{w,\psi}||^2 \to \infty$ as $n \to \infty$.

□

S.2 Extensions

S.2.1 Extension – Unit roots and high persistence

In spite of the relatively weak conditions, e.g., for identification (Assumptions 1 and 2), and sampling (Assumptions 6 and 7) that are allowed under the baseline model, unit roots and high persistence in X_t can lead to a violation of the dominance condition in Assumption 4 and therefore hamper inference using Theorem 2. To this end, this paper extends the baseline theory to accommodate unit root processes and deterministic time trends.

S.2.1.1 Unit roots in X_t

The incidence of unit-root non-stationary processes is quite common in economics. Consider the simple linear model (3.2). X_t becomes a unit-root process if $\beta_1 = 1$ in $U_t = \beta_1 X_{t-1} + \mathcal{E}_t$ and $\mathcal{E}_t$ is some white-noise process. In this first instance, i.e., $X_t = \beta_0 + \widetilde{ATT}_n^{w,\psi} \mathbb{1}\{t \geq 1\} + X_{t-1} + \mathcal{E}_t$, applying the first difference removes the unit root, namely, $\Delta X_t = \beta_0 + \widetilde{ATT}_n^{w,\psi} \mathbb{1}\{t \geq 1\} + \mathcal{E}_t$ where $\Delta X_t := X_t - X_{t-1}$. Thus, handling unit roots in the T-DiD context is straightforward.

S.2.1.2 Persistence in X_t

Economic time series data, e.g., GDP data, are often highly persistent. Very high $|\beta_1| < 1$ in $U_t = \beta_1 X_{t-1} + \mathcal{E}_t$ can pose a problem for inference, especially in small samples. Thus, following, e.g., Hansen (2021, Sect. 16.15) and Baillie et al. (2025), one may consider substituting in the term U_t into (3.2):

$$X_t = \beta_0 + \widetilde{ATT}_n^{w,\psi} \mathbb{1}\{t \geq 1\} + \beta_1 X_{t-1} + \mathcal{E}_t. \tag{S.2.1}$$

$\widetilde{ATT}_n^{w,\psi}$ in the above model is an average (contemporaneous) treatment effect on the treated. Including the lagged X_t in (S.2.1) ensures the resulting residuals are less serially correlated, thereby improving the efficiency of the estimator (Baillie et al., 2025).

Quite importantly, the presence of auto-correlation also has ramifications from an identification perspective. Suppose the true data-generating process followed (S.2.1), then for $\mu := \beta_0/(1-\beta_1)$ where $|\beta_1| < 1$, X_t can be written in the form

$$X_t = \mu + \beta_1^{t+\mathcal{T}+1}\big(X_{-\mathcal{T}-1} - \mu\big) + \underbrace{\widetilde{ATT}_n^{w,\psi}\Big(\frac{1-\beta_1^t}{1-\beta_1}\Big)}_{\widetilde{A}_n(t)} \mathbb{1}\{t \geq 1\} + \underbrace{\sum_{s=-\mathcal{T}}^{t} \beta_1^{t-s} \mathcal{E}_s}_{\widetilde{\mathcal{E}}_t}.$$

From the above, ignoring persistence in X_t has consequences for identification; the probability limit of the estimator is the multiplier effect $\dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \widetilde{A}_n(t) = \widetilde{ATT}_n^{w,\psi}\Big(1 - \dfrac{\beta_1(1-\beta_1^T)}{T(1-\beta_1)}\Big)\dfrac{1}{1-\beta_1} \approx \widetilde{ATT}_n^{w,\psi} + \widetilde{ATT}_n^{w,\psi}\dfrac{\beta_1}{1-\beta_1}$ where the effective estimand $\widetilde{ATT}_n^{w,\psi}$ is properly viewed as the average contemporaneous effect of treatment over the post-treatment window (under the conditions of Theorem 1) and the extra term $\widetilde{ATT}_n^{w,\psi}\frac{\beta_1}{1-\beta_1}$ occurs via the dynamic propagation through X_{t-1}. The resulting error term $\widetilde{\mathcal{E}}_t := \sum_{s=-\mathcal{T}}^{t} \beta_1^{t-s}\mathcal{E}_s$ is non-trivially serially correlated, with implications for CLT-based inference and efficiency. For example, including moving average terms to reduce the serial correlation increases the number of parameters to estimate and reduces the precision of the ATT estimate.

S.2.1.3 Serially Correlated U_t

Suppose that U_t follows an MA(q) process, namely $U_t = \sum_{s=0}^{q} \beta_s \mathcal{E}_{t-s}$. Ignoring this form of temporal dependence in U_t when estimating (3.2) does not lead to inconsistency of the estimator. Moreover, HAC estimators of the standard errors are consistent, although there is a loss in efficiency in not explicitly accounting for this dependence structure. In the context of the SC or factor models, Gonçalves and Ng (2024) demonstrates improved coverage rates and reduced bias from controlling for lagged errors in imputing counterfactual outcomes under stationarity conditions.

S.2.2 Extension – Deterministic time trend

Remark 2.1, for example, shows that common and possibly unbounded time trends get differenced away in $X_t := Y_{1,t} - Y_{0,t}$ and do not pose a problem for the asymptotic identification and inference on $ATT_{\omega,\mathcal{T}}$. However, if time trends differ between the treated and untreated units, the T-DiD estimate captures these differing trends, which is undesirable from an identification point of view. This subsection follows Li (2020) and Hamilton (1994, Chap. 16) in dealing with deterministic time trends.

Consider the leading case of a deterministic trend in X_t with $U_t = \beta_1 \tilde{t}(t) + \mathcal{E}_t$ in (3.2) namely,

$$\begin{aligned} X_t &= \beta_0 + \widetilde{ATT}_n^{w,\psi} \mathbb{1}\{t \geq 1\} + \beta_1 \tilde{t}(t) + \mathcal{E}_t \\ &= \mathcal{X}_t \widetilde{\beta}_n^{w,\psi} + \mathcal{E}_t \\ &= \mathcal{X}_t \Gamma_n^{-1/2} \Gamma_n^{1/2} \widetilde{\beta}_n^{w,\psi} + \mathcal{E}_t \\ &= \mathcal{X}_{\Gamma,t} \widetilde{\beta}_{\Gamma,n}^{w,\psi} + \mathcal{E}_t \end{aligned} \tag{S.2.2}$$

where $\tilde{t}(t) := \mathcal{T} + t + 1$, $\mathcal{X}_t := \big(1, \mathbb{1}\{t \geq 1\}, \tilde{t}(t)\big)$, and $\mathcal{X}_{\Gamma,t} := \mathcal{X}_t \Gamma_n^{-1/2}$. To simplify the exposition, the following sampling assumption is imposed on $\mathcal{E}_t$ in (S.2.2).

Assumption 6-Ext. *$\mathcal{E}_t$ in (S.2.2) is a square-integrable martingale difference sequence with $\mathbb{E}[\mathcal{E}_t|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\mathcal{E}_t^2|\mathcal{F}_{t-1}] = \sigma_\varepsilon^2 > 0$ where $\mathcal{F}_{t-1} := \sigma\big(X_{-\mathcal{T}}, \ldots, X_{t-1}\big)$ denotes the natural filtration.*

$\Gamma_n^{-1/2}$ is a diagonal matrix such that

$$\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) \mathcal{X}_{\Gamma,t}' \mathcal{X}_{\Gamma,t} = Q_{\widetilde{w},n} + \mathcal{O}(n^{-1}),$$

and $Q_{\widetilde{w},n}$ is a positive-definite matrix under Assumption 3. For concreteness and ease of exposition, consider the uniform (regression) weighting scheme $\widetilde{w}(t) = T^{-1}\mathbb{1}\{t \geq 1\} + \mathcal{T}^{-1}\mathbb{1}\{t \leq -1\}$. The appropriate scaling matrix is $\Gamma_{\omega,n}^{1/2} = \text{diag}(1, 1, n) =: \Gamma_n^{1/2}$. The following is a linear representation of the ordinary least squares estimator of $\widetilde{\beta}_{\Gamma,n}^{w,\psi}$ in (S.2.2):

$$(\widehat{\beta}_{\Gamma,n}^{w,\psi} - \widetilde{\beta}_{\Gamma,n}^{w,\psi}) := \Big(\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) \mathcal{X}_{\Gamma,t}' \mathcal{X}_{\Gamma,t} \Big)^{-1} \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) \mathcal{X}_{\Gamma,t}' \mathcal{E}_t.$$

Define $Q(\lambda_n) := Q_A(\lambda_n)^{-1} Q_B(\lambda_n) Q_A(\lambda_n)^{-1}$ where

$$Q_A(\lambda_n) := \begin{bmatrix} 2 & 1 & \left(\frac{3-2\lambda_n}{2}\right) \\ 1 & 1 & \left(\frac{2-\lambda_n}{2}\right) \\ \left(\frac{3-2\lambda_n}{2}\right) & \left(\frac{2-\lambda_n}{2}\right) & \frac{1}{3}(2\lambda_n^2 - 5\lambda_n + 4) \end{bmatrix}, \tag{S.2.3}$$

$$Q_B(\lambda_n) := \sigma_\varepsilon^2 \begin{bmatrix} \frac{1}{\lambda_n(1-\lambda_n)} & \frac{1}{\lambda_n} & \frac{1}{\lambda_n} \\ \frac{1}{\lambda_n} & \frac{1}{\lambda_n} & \frac{1}{2}\left(\frac{2-\lambda_n}{\lambda_n}\right) \\ \frac{1}{\lambda_n} & \frac{1}{2}\left(\frac{2-\lambda_n}{\lambda_n}\right) & \frac{2}{\lambda_n}(3-2\lambda_n) \end{bmatrix}, \tag{S.2.4}$$

and $\sigma_\varepsilon^2 := \mathbb{V}[\mathcal{E}_t]$ under Assumption 6-Ext. Part (c) of Lemma S.3.3 shows that $Q_A(\lambda)$, $Q_B(\lambda)$, and $Q(\lambda)$ are positive-definite uniformly in $\lambda \in [\epsilon,\ 1-\epsilon] \subset (0,1)$.[20] Recall $\mathbb{S}^J$ is the space of all $J \times 1$ vectors with unit Euclidean norm. Let $e_2 := (0,1,0)'$ denote a standard basis vector, $\widetilde{e}_2 := ||Q_B(\lambda_n)^{1/2}e_2||^{-1}Q_B(\lambda_n)^{1/2}e_2$, and $Q(\lambda_n)^{1/2} := Q_A(\lambda_n)^{-1}Q_B(\lambda_n)^{1/2}$. The following states an asymptotic normality result under deterministic time trends.

Theorem 4. *Suppose Assumptions 1 to 3 and Assumption 6-Ext hold, then*

(a)

$$\tau_3' Q(\lambda_n)^{-1/2}\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n}) \xrightarrow{d} \mathcal{N}(0,1)$$

uniformly in $\tau_3 \in \mathbb{S}^3$, *and*

(b)

$$\frac{\sqrt{n}(\widehat{ATT}_{\omega,T} - ATT_{\omega,T})}{\sqrt{e_2'Q(\lambda_n)e_2}} = \widetilde{e}_2' Q(\lambda_n)^{-1/2}\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n}) + o(1) \xrightarrow{d} \mathcal{N}(0,1)$$

as $n \to \infty$ *under the uniform weighting scheme.*

See Appendix S.2.4 for proof.

For the general weighting scheme, one has

$$\begin{aligned}
&\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathcal{X}_{\Gamma,t}'\mathcal{X}_{\Gamma,t} \\
&= \Gamma_n^{-1/2}\sum_{t=-\mathcal{T}}^{T}\begin{bmatrix} \widetilde{w}_n(t) & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\tilde{t}(t) \\ \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\tilde{t}(t) \\ \widetilde{w}_n(t)\tilde{t}(t) & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\tilde{t}(t) & \widetilde{w}_n(t)\tilde{t}(t)^2 \end{bmatrix}\Gamma_n^{-1/2} \\
&= \begin{bmatrix} 2 & 1 & \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\left(\frac{\tilde{t}(t)}{n}\right) \\ 1 & 1 & \sum_{t=1}^{T}\widetilde{w}_n(t)\left(\frac{\tilde{t}(t)}{n}\right) \\ \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\left(\frac{\tilde{t}(t)}{n}\right) & \sum_{t=1}^{T}\widetilde{w}_n(t)\left(\frac{\tilde{t}(t)}{n}\right) & \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\left(\frac{\tilde{t}(t)}{n}\right)^2 \end{bmatrix}
\end{aligned}$$

where the constants in the above matrix follow from the requirement that weights for pre-treatment and post-treatment periods each sum to one and $\tilde{t}(t) := \mathcal{T} + t + 1$. Thus, generally, the above matrix is well behaved; it is positive definite and all entries are bounded and bounded away from zero under Assumption 3.[21]

[20] Also see Figure S.5.

[21] $\mathcal{X}_t\Gamma_n^{-1/2} = (1, \mathbb{1}\{t \geq 1\}, \tilde{t}(t)/n)$ has full column rank, the entry $\tilde{t}(t)/n \in (0,1]$ uniformly in $\tilde{t}(\cdot) \in [n]$, and

S.2.3 Multiple treated units

The index i is included in the subscripts for emphasis and to render the respective quantities specific to a treated unit $i \in [I]$. Define the $I \times 1$ vector $\widehat{ATT}^*_{\omega,\cdot T} := (\widehat{ATT}^*_{\omega,1T}, \dots, \widehat{ATT}^*_{\omega,IT})'$, the $I \times 1$ vector $ATT_{\omega,\cdot T} := (ATT_{\omega,1T}, \dots, ATT_{\omega,IT})'$, and the $I \times I$ matrix $S^*_{\omega,n} := \mathbb{E}[(\widehat{ATT}^*_{\omega,\cdot T} - ATT_{\omega,\cdot T})(\widehat{ATT}^*_{\omega,\cdot T} - ATT_{\omega,\cdot T})']$.

Corollary 2. *Let Assumptions 1 to 4, 6 and 7 for each unique treated-control pair (i,j) such that $j \in [J_i]$ and $i \in [I]$ hold. In addition to Assumption 5-Ext, (a)*

$$(s^*_{\omega,in})^{-1}(\widehat{ATT}^*_{\omega,iT} - ATT_{\omega,iT}) \xrightarrow{d} \mathcal{N}(0,1)$$

and (b)

$$(\tau_I' S^*_{\omega,n} \tau_I)^{-1} \tau_I' (\widehat{ATT}^*_{\omega,\cdot T} - ATT_{\omega,\cdot T}) \xrightarrow{d} \mathcal{N}(0,1)$$

for all $\tau_I \in \mathbb{S}^I$ uniformly in $\mathcal{W}^2$.

Proof. **Part (a):** The proof follows from that of Corollary 1 given the specific linear combination in $\widehat{ATT}^*_{iT} = h^*_{iJ}{}' \widehat{ATT}_{i\cdot T} = (\mathbb{1}_J' S_{in}^{-1} \mathbb{1}_J)^{-1} \mathbb{1}_J' S_{in}^{-1} \widehat{ATT}_{i\cdot T}$, set $\tau_J = h^*_{iJ} ||h^*_{iJ}||^{-1}$.

Part (b): The proof of this part is similar to Corollary 1 and is hence omitted. □

Remark S.2.1. *Part (b) of Corollary 2 is based on efficient estimators of $ATT_{\omega,iT}$. This, however, need not be the case. More general weighting schemes are admissible, including one that simply selects a single control unit out of the set $[J_i]$ of valid controls for unit i.*

S.2.4 Proof of Theorem 4

Theorem 4. *Suppose Assumptions 1 to 3 and Assumption 6-Ext hold, then*

(a)

$$\tau_3' Q(\lambda_n)^{-1/2} \sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n}) \xrightarrow{d} \mathcal{N}(0,1)$$

uniformly in $\tau_3 \in \mathbb{S}^3$, and

(b)

$$\frac{\sqrt{n}(\widehat{ATT}_{\omega,T} - ATT_{\omega,T})}{\sqrt{e_2' Q(\lambda_n) e_2}} = \widetilde{e}_2' Q(\lambda_n)^{-1/2} \sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n}) + o(1) \xrightarrow{d} \mathcal{N}(0,1)$$

as $n \to \infty$ under the uniform weighting scheme.

Proof:

$\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) \mathbb{1}\{t \le -1\} = \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) \mathbb{1}\{t \ge 1\} = 1.$

Part (a): Under the conditions of Lemma S.3.3, the continuous mapping theorem and the continuity of the inverse at a non-singular matrix,

$$\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n}) = \Big(\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{X}_{\Gamma,t}\Big)^{-1}\sqrt{n}\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t$$
$$= Q_A(\lambda_n)^{-1}\sqrt{n}\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t + \mathcal{O}_p(n^{-1}).$$

Thus, under the conditions of Lemma S.3.3,

$$\mathbb{V}\Big[\tau_3'\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n} - \widetilde{\beta}^{w,\psi}_{\Gamma,n})\Big] = \tau_3'Q(\lambda_n)\tau_3 + \mathcal{O}(n^{-1}).$$

The rest of the proof is conducted in the following steps.

First, the scalar-valued random variable $\Upsilon_t := \sqrt{n}\tau_3'\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t$ constitutes a square-integrable martingale difference sequence under Assumption 6-Ext since the following hold: (1) $\mathbb{E}[\Upsilon_t|\mathcal{F}_{\Upsilon,t-1}] = 0$ where $\mathcal{F}_{\Upsilon,t} := \sigma(\{\Upsilon_{t'}, t' \leq t\})$ is the natural filtration and (2)

$$\begin{aligned}
\sigma^2_{\Upsilon,t} := \mathbb{E}[\Upsilon_t^2|\mathcal{F}_{t-1}] &= \sigma^2_\varepsilon n\widetilde{w}_n(t)^2\tau_3'\mathcal{X}'_{\Gamma,t}\mathcal{X}_{\Gamma,t}\tau_3\\
&= n\sigma^2_\varepsilon||\widetilde{w}_n(t)\mathcal{X}_{\Gamma,t}\tau_3||^2\\
&\leq n\sigma^2_\varepsilon\frac{1}{\mathcal{T}^2\wedge T^2}||\mathcal{X}_{\Gamma,t}||^2\\
&= n\sigma^2_\varepsilon\frac{1}{\mathcal{T}^2\wedge T^2}||(1, \mathbb{1}\{t\geq 1\}, \tilde{t}(t)/n)||^2\\
&\leq 3n\sigma^2_\varepsilon\frac{1}{\mathcal{T}^2\wedge T^2}\\
&= 3\sigma^2_\varepsilon\frac{1}{n}\frac{1}{(1-\lambda_n)^2\wedge\lambda_n^2} =: c^2_{\Upsilon,n}
\end{aligned}$$

Observe that $c^2_{\Upsilon,n}$ does not vary with t. First, the above implies $\{\Upsilon_t^2/c^2_{\Upsilon,n}\}$ is uniformly integrable.

Second, define

$$s^2_{\Upsilon,n} := \sum_{t=-\mathcal{T}}^{T}\mathbb{E}[\Upsilon_t^2] = n\sigma^2_\varepsilon\sum_{t=-\mathcal{T}}^{T}||\widetilde{w}_n(t)\mathcal{X}_{\Gamma,t}\tau_3||^2 = ||Q_B(\lambda_n)^{1/2}\tau_3||^2 + \mathcal{O}(n^{-1}).$$

From the above, the lower bound can be characterised as

$$s^2_{\Upsilon,n} = n\sigma^2_\varepsilon\sum_{t=-\mathcal{T}}^{T}||\widetilde{w}_n(t)\mathcal{X}_{\Gamma,t}\tau_3||^2 \geq \sigma^2_\varepsilon\frac{1}{(1-\lambda_n)^2\vee\lambda_n^2}\frac{1}{n}\sum_{t=-\mathcal{T}}^{T}||\mathcal{X}_{\Gamma,t}||^2 \geq \sigma^2_\varepsilon\frac{1}{(1-\lambda_n)^2\vee\lambda_n^2}.$$

Thus,

$$\sup_n n\frac{c^2_{\Upsilon,n}}{s^2_{\Upsilon,n}} \leq 3\frac{(1-\lambda_n)^2\vee\lambda_n^2}{(1-\lambda_n)^2\wedge\lambda_n^2} = 3\Big(\frac{(1-\lambda_n)\vee\lambda_n}{(1-\lambda_n)\wedge\lambda_n}\Big)^2 < 3\Big(\frac{1-\epsilon}{\epsilon}\Big)^2 < \infty$$

under Assumption 3.

Taken together, it follows from Davidson (2021, Theorem 25.4) that conditions 25.3(a) and 25.3(b) of Davidson (2021, Theorem 25.3) are satisfied for $\Upsilon_{nt} := \Upsilon_t/s_{\Upsilon,n}$. It follows from

Davidson (2021, Theorem 25.3) that $\sum_{t=-\mathcal{T}}^{T}\Upsilon_{nt}\xrightarrow{d}\mathcal{N}(0,1)$ as $n\to\infty$ under Assumption 3.

Noticing that $\mathbb{S}^3\ni\widetilde{\tau}_3:=||Q_B(\lambda_n)^{1/2}\tau_3||^{-1}Q_B(\lambda_n)^{1/2}\tau_3$, for $\tau_3\in\mathbb{S}^3$,

$$\begin{aligned}
&\widetilde{\tau}_3'Q(\lambda_n)^{-1/2}\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n}-\widetilde{\beta}^{w,\psi}_{\Gamma,n})=\widetilde{\tau}_3'Q(\lambda_n)^{-1/2}Q_A(\lambda_n)^{-1}\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t+\mathcal{O}_p(n^{-1})\\
&=\widetilde{\tau}_3'Q(\lambda_n)^{-1/2}\underbrace{Q_A(\lambda_n)^{-1}Q_B(\lambda_n)^{1/2}}_{=:Q(\lambda_n)^{1/2}}\Big(Q_B(\lambda_n)^{-1/2}\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t\Big)+\mathcal{O}_p(n^{-1})\\
&=\widetilde{\tau}_3'Q_B(\lambda_n)^{-1/2}\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t+\mathcal{O}_p(n^{-1})\\
&:=||Q_B(\lambda_n)^{1/2}\tau_3||^{-1}\tau_3'\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t+\mathcal{O}_p(n^{-1})\\
&=:\sum_{t=-\mathcal{T}}^{T}\Upsilon_{nt}+\mathcal{O}_p(n^{-1})\xrightarrow{d}\mathcal{N}(0,1).
\end{aligned}\tag{S.2.5}$$

Part (b): Noticing that $\mathbb{S}^3\ni\widetilde{e}_2:=||Q_B(\lambda_n)^{1/2}e_2||^{-1}Q_B(\lambda_n)^{1/2}e_2$, for $(0,1,0)'=:e_2\in\mathbb{S}^3$, it follows from (S.2.5) that

$$\begin{aligned}
\widetilde{e}_2'Q(\lambda_n)^{-1/2}\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n}-\widetilde{\beta}^{w,\psi}_{\Gamma,n})&=||Q_B(\lambda_n)^{1/2}e_2||^{-1}e_2'\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t+\mathcal{O}_p(n^{-1})\\
&=\frac{\sqrt{n}(\widehat{ATT}_{\omega,T}-\widetilde{ATT}^{w,\phi}_n)}{\sqrt{e_2'Q(\lambda_n)e_2}}\\
&=\frac{\sqrt{n}(\widehat{ATT}_{\omega,T}-ATT_{\omega,T})}{\sqrt{e_2'Q(\lambda_n)e_2}}+\frac{\sqrt{n}R^{w,\phi}_n}{\sqrt{e_2'Q(\lambda_n)e_2}}\\
&=\frac{\sqrt{n}(\widehat{ATT}_{\omega,T}-ATT_{\omega,T})}{\sqrt{e_2'Q(\lambda_n)e_2}}+o(1).
\end{aligned}$$

The last equality follows from the conditions of Theorem 1 and the continuous mapping theorem. Thus, invoking part (a) above,

$$\frac{\sqrt{n}(\widehat{ATT}_{\omega,T}-ATT_{\omega,T})}{\sqrt{e_2'Q(\lambda_n)e_2}}=\widetilde{e}_2'Q(\lambda_n)^{-1/2}\sqrt{n}(\widehat{\beta}^{w,\psi}_{\Gamma,n}-\widetilde{\beta}^{w,\psi}_{\Gamma,n})+o(1)\xrightarrow{d}\mathcal{N}(0,1),$$

and this proves the assertion as claimed. □

S.3 Supporting Lemmata

Lemma S.3.1. *Suppose Assumptions 3 to 5 hold, then* $\dfrac{||X_{nt}||_r}{c_{nt}}\lesssim 1$ *uniformly in* (n,t) *and* $\mathcal{W}^2$ *where* $r>2$.

Proof.

$$
\begin{aligned}
&||\omega_n(t)(X_t - \mathbb{E}[X_t])||_r = w_T(t)||X_t - \mathbb{E}[X_t]||_r \mathbb{1}\{t \geq 1\} + \psi_{\mathcal{T}}(-t)||X_t - \mathbb{E}[X_t]||_r \mathbb{1}\{t \leq -1\} \\
&= w_T(t)||(Y_{1,t} - Y_{0,t}) - \mathbb{E}[(Y_{1,t} - Y_{0,t})]||_r \mathbb{1}\{t \geq 1\} + \psi_{\mathcal{T}}(-t)||(Y_{1,t} - Y_{0,t}) - \mathbb{E}[(Y_{1,t} - Y_{0,t})]||_r \mathbb{1}\{t \leq -1\} \\
&\leq C\big(w_T(t)\mathbb{1}\{t \geq 1\} + \psi_{\mathcal{T}}(-t)\mathbb{1}\{t \leq -1\}\big) \\
&\leq \frac{C}{\lambda_n n} \sup_{\phi \in \mathcal{W}} \Big\{T \max_{t \in [T]} \phi_T(t)\Big\} \mathbb{1}\{t \geq 1\} + \frac{C}{(1-\lambda_n)n} \sup_{\phi \in \mathcal{W}} \Big\{\mathcal{T} \max_{\tau \in [-\mathcal{T}]} \phi_{\mathcal{T}}(-\tau)\Big\} \mathbb{1}\{t \leq -1\} \\
&\lesssim n^{-1}.
\end{aligned}
$$

The first inequality follows from Assumption 4. The last line uses the definition of $\mathcal{W}$ in (2.2) and Assumption 3. It follows from the above and Assumption 5 that

$$
\frac{||X_{nt}||_r}{c_{nt}} = \frac{||\omega_n(t)(X_t - \mathbb{E}[X_t])||_r}{s_{\omega,n}(\sigma_{nt} \vee s_{\omega,n})} \lesssim \frac{1}{\sqrt{n}s_{\omega,n}(\sqrt{n}\sigma_{nt} \vee \sqrt{n}s_{\omega,n})} \leq \frac{1}{\epsilon\sqrt{n}\sigma_{nt} \vee \epsilon^2} = \frac{1}{\epsilon\sqrt{n}\sigma_{nt}} \wedge \frac{1}{\epsilon^2} \leq \frac{1}{\epsilon^2}
$$

uniformly in (n,t) and $\mathcal{W}^2$. □

Lemma S.3.2. *Suppose Assumptions 3 to 7 hold, then* $\sum_{t=-\mathcal{T}}^{T} X_{nt} \xrightarrow{d} \mathcal{N}(0,1)$ *uniformly in* $\mathcal{W}^2$.

Proof. The proof proceeds by the following steps.

(a)

$$
\Big|\Big| \sum_{t=-\mathcal{T}}^{T} X_{nt} \Big|\Big|_2 = s_{\omega,n}^{-1} \Big|\Big| \sum_{t=-\mathcal{T}}^{T} \omega_n(t)(X_t - \mathbb{E}[X_t]) \Big|\Big|_2 = 1.
$$

(b) By Lemma S.3.1,

$$
\Big|\Big| \frac{X_{nt}}{c_{nt}} \Big|\Big|_r = \Big|\Big| \frac{X_{nt}}{\sigma_{nt} \vee s_{\omega,n}} \Big|\Big|_r \lesssim 1
$$

uniformly in (n,t) and $\mathcal{W}^2$ hence, $\frac{X_{nt}}{c_{nt}}$ is L_r-bounded uniformly in (n,t) and $\mathcal{W}^2$ under Assumptions 3 to 5.

(c) The conditions of Lemma S.3.1 imply $\frac{d_{nt}}{c_{nt}} \leq \bar{d}\frac{||X_{nt}||_2}{\sigma_{nt} \vee s_{\omega,n}} \lesssim 1$ uniformly in (n,t) and $\mathcal{W}^2$.

Steps (a)-(c) above and Assumptions 6 and 7 verify the conditions of Davidson (2021, Theorem 25.12) thus $\sum_{t=-\mathcal{T}}^{T} X_{nt} \xrightarrow{d} \mathcal{N}(0,1)$ uniformly in $\mathcal{W}^2$ as claimed. □

Lemma S.3.3. *Fix* $\widetilde{w}_n(t) = T^{-1}\mathbb{1}\{t \geq 1\} + \mathcal{T}^{-1}\mathbb{1}\{t \leq -1\}$, *i.e., the uniform regression weighting scheme, then under Assumption 3 and Assumption 6-Ext,*

(a)

$$
\sum_{t=-\mathcal{T}}^{T} \widetilde{w}(t)\mathcal{X}'_{\Gamma,t}\mathcal{X}_{\Gamma,t} = Q_A(\lambda_n) + \mathcal{O}(n^{-1});
$$

(b)

$$\mathbb{V}\Big[\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t\Big] = Q_B(\lambda_n) + \mathcal{O}(n^{-1})$$

where the expressions of $Q_A(\lambda_n)$ and $Q_B(\lambda_n)$ are given in (S.2.3) and (S.2.4), respectively; and

(c) $Q_A(\lambda)$, $Q_B(\lambda)$, and $Q(\lambda) := Q_A(\lambda)^{-1}Q_B(\lambda)Q_A(\lambda)^{-1}$ are (each) positive definite uniformly in $\lambda \in [\epsilon,\ 1-\epsilon] \subset (0,1)$.

Proof. **Part (a)**

As pre- and post-treatment weights must sum to one, one has, using the uniform regression weighting scheme $\widetilde{w}_n(t) = T^{-1}\mathbb{1}\{t \geq 1\} + \mathcal{T}^{-1}\mathbb{1}\{t \leq -1\}$ that,

$$\begin{aligned}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}(t)\mathcal{X}'_t\mathcal{X}_t &= \sum_{t=-\mathcal{T}}^{T}\begin{bmatrix}\widetilde{w}_n(t) & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\tilde{t}(t)\\ \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\tilde{t}(t)\\ \widetilde{w}_n(t)\tilde{t}(t) & \widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\tilde{t}(t) & \widetilde{w}_n(t)\tilde{t}(t)^2\end{bmatrix}\\ &= \begin{bmatrix}2 & 1 & 1+n\Big(\frac{3-2\lambda_n}{2}\Big)\\ 1 & 1 & \frac{1}{2}+n\Big(\frac{2-\lambda_n}{2}\Big)\\ 1+n\Big(\frac{3-2\lambda_n}{2}\Big) & \frac{1}{2}+n\Big(\frac{2-\lambda_n}{2}\Big) & q^*_{A,n}\end{bmatrix}\end{aligned}$$

where

$$q^*_{A,n} = \frac{1}{6\lambda_n}\Big(2\big((2\lambda_n-1)(1-\lambda_n)^2+1\big)n^2 + 3\big((1-\lambda_n)(2\lambda_n-1)+1\big)n + 2\lambda_n\Big)$$

which is well-defined for $\lambda_n > 0$ thanks to Assumption 3.

The non-trivial entries of the above matrix are computed in the following steps. First,

$$\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\tilde{t}(t) = \frac{1}{T}\sum_{t=\mathcal{T}+1}^{n} t = \frac{1}{T}\frac{T}{2}(\mathcal{T}+1+n) = \frac{1}{2}(\mathcal{T}+1+n) = \frac{1}{2} + n\Big(\frac{2-\lambda_n}{2}\Big).$$

Second,

$$\begin{aligned}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\tilde{t}(t) &= \frac{1}{\mathcal{T}}\sum_{t=1}^{\mathcal{T}} t + \frac{1}{T}\sum_{t=\mathcal{T}+1}^{n} t\\ &= \frac{1}{2}(\mathcal{T}+1) + \frac{1}{2}(\mathcal{T}+1+n) = 1 + n\Big(\frac{3-2\lambda_n}{2}\Big).\end{aligned}$$

Third,

$$\begin{aligned}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\tilde{t}(t)^2 &= \frac{1}{\mathcal{T}}\sum_{t=1}^{\mathcal{T}}t^2 + \frac{1}{T}\sum_{t=\mathcal{T}+1}^{n}t^2 \\ &= \frac{1}{\mathcal{T}}\sum_{t=1}^{\mathcal{T}}t^2 + \frac{1}{T}\Big(\sum_{t=1}^{n}t^2 - \sum_{t=1}^{\mathcal{T}}t^2\Big) \\ &= \frac{1}{T}\sum_{t=1}^{n}t^2 + \Big(\frac{1}{\mathcal{T}} - \frac{1}{T}\Big)\sum_{t=1}^{\mathcal{T}}t^2 \\ &= \frac{1}{T}\Big(\frac{1}{6}n(n+1)(2n+1)\Big) + \Big(\frac{1}{\mathcal{T}} - \frac{1}{T}\Big)\Big(\frac{1}{6}\mathcal{T}(\mathcal{T}+1)(2\mathcal{T}+1)\Big) \\ &= \frac{1}{6\lambda_n}\Big(2\big((2\lambda_n-1)(1-\lambda_n)^2+1\big)n^2 + 3((1-\lambda_n)(2\lambda_n-1)+1)n + 2\lambda_n\Big) \\ &=: q^*_{A,n}.\end{aligned}$$

Next, fix $\Gamma_{\omega,n}^{-1/2} = \mathrm{diag}(1, 1, n^{-1})$, then,

$$\begin{aligned}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}(t)\mathcal{X}'_{\Gamma,t}\mathcal{X}_{\Gamma,t} &= \begin{bmatrix} 2 & 1 & \left(\frac{3-2\lambda_n}{2}\right) \\ 1 & 1 & \left(\frac{2-\lambda_n}{2}\right) \\ \left(\frac{3-2\lambda_n}{2}\right) & \left(\frac{2-\lambda_n}{2}\right) & \frac{1}{3}(2\lambda_n^2 - 5\lambda_n + 4)\end{bmatrix} + \mathcal{O}(n^{-1}) \\ &=: Q_A(\lambda_n) + \mathcal{O}(n^{-1}).\end{aligned}$$

Part (b)

Now consider

$$\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t = \begin{bmatrix}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{E}_t \\ \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}\mathcal{E}_t \\ \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)(\tilde{t}(t)/n)\mathcal{E}_t\end{bmatrix}.$$

Under Assumption 6-Ext on $\mathcal{E}_t$,

$$\begin{aligned}
\mathbb{V}[\sqrt{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathcal{X}'_{\Gamma,t}\mathcal{E}_t] &= \sigma_\varepsilon^2 n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\mathcal{X}'_{\Gamma,t}\mathcal{X}_{\Gamma,t} \\
&= \sigma_\varepsilon^2 n\sum_{t=-\mathcal{T}}^{T}\begin{bmatrix} \widetilde{w}_n(t)^2 & \widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)^2\tilde{t}(t)/n \\ \widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\} & \widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\}\tilde{t}(t)/n \\ \widetilde{w}_n(t)^2\tilde{t}(t)/n & \widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\}\tilde{t}(t)/n & \widetilde{w}_n(t)^2\left(\tilde{t}(t)/n\right)^2 \end{bmatrix} \\
&= \sigma_\varepsilon^2\begin{bmatrix} \frac{1}{\lambda_n(1-\lambda_n)} & \frac{1}{\lambda_n} & \frac{1}{\lambda_n} \\ \frac{1}{\lambda_n} & \frac{1}{\lambda_n} & \frac{1}{2}\left(\frac{2-\lambda_n}{\lambda_n}\right) \\ \frac{1}{\lambda_n} & \frac{1}{2}\left(\frac{2-\lambda_n}{\lambda_n}\right) & \frac{2}{\lambda_n}(3-2\lambda_n) \end{bmatrix} + \mathcal{O}(n^{-1}) \\
&= Q_B(\lambda_n) + \mathcal{O}(n^{-1}).
\end{aligned}$$

The entries of the above matrix are computed in the following steps.

First,

$$n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\} = \frac{n}{T^2}\sum_{t=-\mathcal{T}}^{T}\mathbb{1}\{t\geq 1\} = \frac{n}{T} = \frac{1}{\lambda_n}.$$

Second,

$$n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2 = \frac{n}{\mathcal{T}^2}\sum_{t=-\mathcal{T}}^{T}\mathbb{1}\{t\leq -1\} + \frac{n}{T^2}\sum_{t=-\mathcal{T}}^{T}\mathbb{1}\{t\geq 1\} = \frac{n}{\mathcal{T}} + \frac{n}{T} = \frac{1}{\lambda_n(1-\lambda_n)}.$$

Third,

$$\begin{aligned}
n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\}\tilde{t}(t)/n &= \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\mathbb{1}\{t\geq 1\}\tilde{t}(t) \\
&= \frac{1}{T^2}\sum_{t=\mathcal{T}+1}^{n} t = \frac{1}{2T}(\mathcal{T}+1+n) \\
&= \frac{1}{2}\frac{(2-\lambda_n)}{\lambda_n} + \frac{1}{2\lambda_n n} \\
&= \frac{1}{2}\frac{(2-\lambda_n)}{\lambda_n} + \mathcal{O}(n^{-1})
\end{aligned}$$

thanks to Assumption 3.

Fourth,

$$
\begin{aligned}
n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\tilde{t}(t)/n &= \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\tilde{t}(t) = \frac{1}{\mathcal{T}^2}\sum_{t=1}^{\mathcal{T}} t + \frac{1}{T^2}\sum_{t=\mathcal{T}+1}^{n} t \\
&= \frac{1}{2\mathcal{T}}(\mathcal{T}+1) + \frac{1}{2T}(\mathcal{T}+1+n) \\
&= \frac{1}{2}\Big(1+\frac{2-\lambda_n}{\lambda_n}\Big) + \frac{1}{2n}\Big(\frac{1}{(1-\lambda_n)} + \frac{1}{\lambda_n}\Big) \\
&= \frac{1}{2}\Big(1+\frac{2-\lambda_n}{\lambda_n}\Big) + \mathcal{O}(n^{-1}) \\
&= \frac{1}{\lambda_n} + \mathcal{O}(n^{-1})
\end{aligned}
$$

thanks to Assumption 3.

Fifth,

$$
\begin{aligned}
& n\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2(\tilde{t}(t)/n)^2 = \frac{1}{n}\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)^2\tilde{t}(t) \\
&= \frac{1}{n}\Big(\frac{1}{\mathcal{T}^2}\sum_{t=1}^{\mathcal{T}} t^2 + \frac{1}{T^2}\sum_{t=\mathcal{T}+1}^{n} t^2\Big) \\
&= \frac{1}{n}\Big(\frac{1}{T^2}\sum_{t=1}^{n} t^2 + \Big(\frac{1}{\mathcal{T}^2} - \frac{1}{T^2}\Big)\sum_{t=1}^{\mathcal{T}} t^2\Big) \\
&= \frac{1}{6nT^2}n(n+1)(2n+1) + \frac{1}{6n}\Big(\frac{T^2-\mathcal{T}^2}{\mathcal{T}^2T^2}\Big)\big(\mathcal{T}(\mathcal{T}+1)(2\mathcal{T}+1)\big) \\
&= \frac{1}{\lambda_n^2 n^3}\Big(2n^3+3n^2+n\Big) + \Big(\frac{2\lambda_n-1}{\lambda_n^2(1-\lambda_n)^2n^3}\Big)\big(2(1-\lambda_n)^3n^3 + 3(1-\lambda_n)^2n^2 + (1-\lambda_n)n\big) \\
&= \frac{2}{\lambda_n^2} + \frac{2}{\lambda_n^2}(2\lambda_n-1)(1-\lambda_n) + \frac{1}{\lambda_n^2}\Big(\frac{3}{n}+\frac{1}{n^2}\Big) + \Big(\frac{2\lambda_n-1}{\lambda_n^2(1-\lambda_n)^2}\Big)\Big(\frac{3(1-\lambda_n)^2}{n} + \frac{1-\lambda_n}{n^2}\Big) \\
&= \frac{2}{\lambda_n^2}\big(1+(2\lambda_n-1)(1-\lambda_n)\big) + \mathcal{O}(n^{-1}) \\
&= \frac{2}{\lambda_n}\big(3-2\lambda_n\big) + \mathcal{O}(n^{-1}).
\end{aligned}
$$

All the above entries are well-defined thanks to Assumption 3.

Part (c)

Let

$$
Q_A(\lambda) = \begin{bmatrix} 2 & 1 & \left(\frac{3-2\lambda}{2}\right) \\ 1 & 1 & \left(\frac{2-\lambda}{2}\right) \\ \left(\frac{3-2\lambda}{2}\right) & \left(\frac{2-\lambda}{2}\right) & \frac{1}{3}(2\lambda^2-5\lambda+4) \end{bmatrix} =: \begin{bmatrix} A_{1:2,1:2}(\lambda) & A_{1:2,3}(\lambda) \\ A_{1:2,3}(\lambda)' & A_{3,3}(\lambda) \end{bmatrix},
$$

where $A_{1:2,1:2}(\lambda) := \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ is the 2×2 leading principal sub-matrix. The eigenvalues of

$A_{1:2,1:2}(\lambda)$ are $\big((3+\sqrt{5})/2,\,(3-\sqrt{5})/2\big)'$, thus $A_{1:2,1:2}(\lambda)$ is positive definite uniformly in $\lambda \in [\epsilon,\,1-\epsilon]$. Further, consider the Schur complement of $A_{1:2,1:2}(\lambda)$ in $Q_A(\lambda)$, namely

$$\begin{aligned}\mathrm{SC}_A(\lambda) :&= A_{3,3}(\lambda) - A_{1:2,3}(\lambda)' A_{1:2,1:2}^{-1}(\lambda) A_{1:2,3}(\lambda) \\ &= \frac{(2\lambda^2 - 5\lambda + 4)}{3} - \frac{2\lambda^2 - 6\lambda + 5}{4} \\ &= \frac{1}{6}\big(\lambda - \tfrac{1}{2}\big)^2 + \frac{1}{24}.\end{aligned}$$

Over the open interval $(0,1)$, $\mathrm{SC}_A(\lambda)$ is a parabola, convex with a global minimum at $\lambda^* = 1/2$ which equals $\mathrm{SC}_A(1/2) = 1/24 > 0$. It follows from the above and Bernstein (2009, Proposition 8.2.4(v)) that $Q_A(\lambda)$ is positive definite uniformly in $\lambda \in [\epsilon, 1-\epsilon],\ \epsilon \in (0, 1/2]$.

Next, consider

$$\frac{\lambda}{\sigma_\varepsilon^2} Q_B(\lambda) = \begin{bmatrix} \frac{1}{(1-\lambda)} & 1 & 1 \\ 1 & 1 & \frac{(2-\lambda)}{2} \\ 1 & \frac{(2-\lambda)}{2} & 2(3-2\lambda) \end{bmatrix},$$

and notice that since $\sigma_\varepsilon^2 > 0$, $Q_B(\lambda)$ is positive definite in $\lambda \in [\epsilon, 1-\epsilon]$ if and only if $\frac{\lambda}{\sigma_\varepsilon^2} Q_B(\lambda)$ is positive definite.

Let $B_{1:k,1:k}(\lambda)$ denote the upper-left $k \times k$ sub-matrix of $\frac{\lambda}{\sigma_\varepsilon^2} Q_B(\lambda)$. $\det\big(B_{1,1}(\lambda)\big) = \frac{1}{(1-\lambda)} \geq \frac{1}{(1-\epsilon)} > 0$ for all $\lambda \in [\epsilon, 1-\epsilon]$. $\det(B_{1:2,1:2}(\lambda)) = \frac{1}{1-\lambda} - 1 \geq \frac{1}{(1-\epsilon)} - 1 = \frac{\epsilon}{(1-\epsilon)} > 0$. This confirms $B_{1:2,1:2}(\lambda)$ is positive definite uniformly in $\lambda \in [\epsilon, 1-\epsilon]$ by Sylvester's criterion. Compute the Schur complement of $B_{1:2,1:2}(\lambda)$ in $\frac{\lambda}{\sigma_\varepsilon^2} Q_B(\lambda)$:

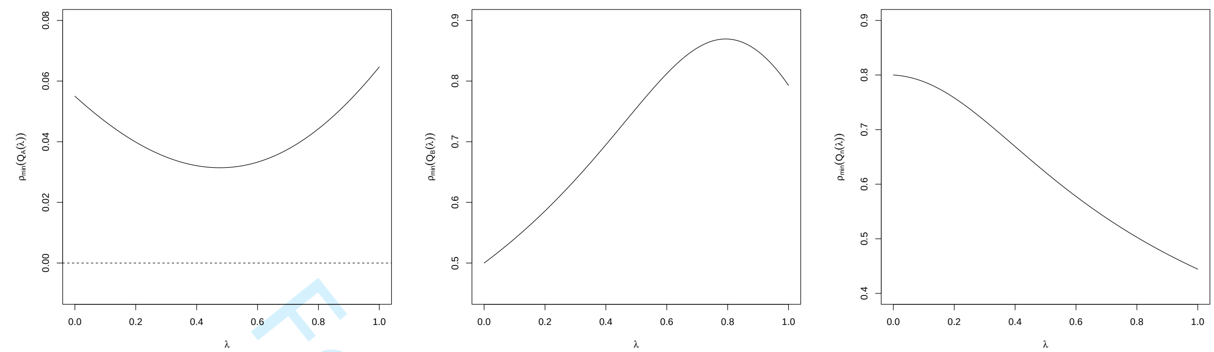
$$\begin{aligned}\mathrm{SC}_B(\lambda) :&= B_{3,3}(\lambda) - B_{1:2,3}(\lambda)' B_{1:2,1:2}^{-1}(\lambda) B_{1:2,3}(\lambda) \\ &= 2(3-2\lambda) - (1-(3/4)\lambda) \\ &= 4^{-1}(20 - 13\lambda)\end{aligned}$$

$\mathrm{SC}_B(\lambda) \geq 4^{-1}(7 + 13\epsilon) > 0$ uniformly in $\lambda \in [\epsilon, 1-\epsilon]$. Again, invoking Bernstein (2009, Proposition 8.2.4(v)), it follows from the above that $Q_B(\lambda)$ is positive definite uniformly in $\lambda \in [\epsilon, 1-\epsilon],\ \epsilon \in (0, 1/2]$.

Lastly, observe that for all $\tau \in \mathbb{S}^3$, $\tau' Q(\lambda) \tau = \tau Q_A(\lambda)^{-1} Q_B(\lambda) Q_A(\lambda)^{-1} \tau = \widetilde{\tau}' Q_B(\lambda) \widetilde{\tau} = ||Q_B^{1/2}(\lambda)\widetilde{\tau}|| > 0$ since $\widetilde{\tau} := Q_A(\lambda)^{-1}\tau \neq 0$ by the positive definiteness of $Q_A(\lambda)$ and $Q_B(\lambda)$ uniformly in $\lambda \in [\epsilon,\,1-\epsilon]$ from the foregoing.

This completes the proof for this part. Figure S.5 illustrates the result using plots of the minimum eigenvalues of $Q_A(\lambda)$, $Q_B(\lambda)$ and $Q(\lambda)$, respectively, as a function of $\lambda \in [\epsilon_\lambda,\ 1-\epsilon_\lambda],\ \epsilon_\lambda = 10^{-6}$.

Figure S.5: Minimum Eigenvalues of $Q_A(\lambda)$, $Q_B(\lambda)$, and $Q(\lambda)$.



(a) $Q_A(\lambda)$ (b) $Q_B(\lambda)$ (c) $Q(\lambda)$

Notes: The continuous lines plot the minimum eigenvalue of the respective 3×3 matrices as a function of $\lambda \in [\epsilon_\lambda,\ 1-\epsilon_\lambda],\ \epsilon_\lambda = 10^{-6}$. The horizontal dashed line in the first plot corresponds to zero.

□

S.4 Useful Propositions

S.4.1 T-DiD as a least-squares estimator

Proposition 3. *The vector of minimisers of the least squares criterion*

$$S_{\widetilde{w},n}(\beta_0, B) := \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)(X_t - \beta_0 - B\mathbb{1}\{t \geq 1\})^2$$

is given by $(\widehat{\beta}_0, \widehat{B})' = \underset{(\beta_0,B)'}{\operatorname{argmin}}\, S_{\widetilde{w},n}(\beta_0, B)$ *where* $\widehat{B} = \widehat{ATT}_{\omega,T}$.

Proof. The first order conditions with respect to β_0 and B are given by

$$\frac{\partial S_{\widetilde{w},n}(\widehat{\beta}_0, \widehat{B})}{\partial \widehat{\beta}_0} := -2 \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)(X_t - \widehat{\beta}_0 - \widehat{B}\mathbb{1}\{t \geq 1\}) = 0; \text{ and} \tag{S.4.1}$$

$$\frac{\partial S_{\widetilde{w},n}(\widehat{\beta}_0, \widehat{B}))}{\partial \widehat{B}} := -2 \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathbb{1}\{t \geq 1\}(X_t - \widehat{\beta}_0 - \widehat{B}\mathbb{1}\{t \geq 1\}) = 0. \tag{S.4.2}$$

Since $\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathbb{1}\{t \leq -1\} = \sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)\mathbb{1}\{t \geq 1\} = 1$ and $\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t) = 2$ by definition, it follows from (S.4.1) that

$$\widehat{\beta}_0 = \frac{1}{2}\Big(\sum_{t=-\mathcal{T}}^{T} \widetilde{w}_n(t)X_t - \widehat{B}\Big)$$

which, substituted into (S.4.2), gives

$$\begin{aligned}\widehat{B} &= 2\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}X_t - \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)X_t \\ &= 2\sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}X_t - \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\big(\mathbb{1}\{t\leq -1\}+\mathbb{1}\{t\geq 1\}\big)X_t \\ &= \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\geq 1\}X_t - \sum_{t=-\mathcal{T}}^{T}\widetilde{w}_n(t)\mathbb{1}\{t\leq -1\}X_t \\ &= \sum_{t=1}^{T} w_T(t)X_t - \sum_{-\tau=1}^{\mathcal{T}}\psi(-\tau)X_t =: \widehat{ATT}_{\omega,T}\end{aligned}$$

by (3.1). □

S.4.2 Deriving the T-DiD from SC-type arguments

Let $\widehat{Y}_{1,t}(0)$, $t\in[-\mathcal{T}]\cup[T]$ denote the synthetic control. Then for the demeaned convex-weighted SC with a single control unit, the SC weight is trivially one:

$$\Big(\widehat{Y}_{1,\tau}(0) - \frac{1}{\mathcal{T}}\sum_{-\tau'=1}^{\mathcal{T}}Y_{1,\tau'}\Big) = \underbrace{1}_{\widehat{b}_{\mathcal{T}}}\times\Big(Y_{0,\tau} - \frac{1}{\mathcal{T}}\sum_{-\tau'=1}^{\mathcal{T}}Y_{0,\tau'}\Big) \text{ whence}$$

$$\widehat{Y}_{1,\tau}(0) = \underbrace{\frac{1}{\mathcal{T}}\sum_{-\tau'=1}^{\mathcal{T}}\big(Y_{1,\tau'}-Y_{0,\tau'}\big)}_{\widehat{a}_{\mathcal{T}}} + \underbrace{1}_{\widehat{b}_{\mathcal{T}}}\times Y_{0,\tau} =: \widehat{a}_{\mathcal{T}} + \widehat{b}_{\mathcal{T}}Y_{0,\tau}$$

for each $\tau\in[-\mathcal{T}]$.

Using the above to recover untreated potential outcomes of the treated unit in the post-treatment period, one has for each $t\in[T]$, $\widehat{Y}_{1,t}(0)=\widehat{a}_{\mathcal{T}}+\widehat{b}_{\mathcal{T}}Y_{0,t}$. Thus, an estimator of $ATT_{\omega,T}$ can be expressed as

$$\widehat{ATT}^{S}_{\omega,T} := \sum_{t=1}^{T}w_T(t)\big(Y_{1,t}(1)-\widehat{Y}_{1,t}(0)\big) = \sum_{t=1}^{T}w_T(t)\big(Y_{1,t}-Y_{0,t}\big) - \frac{1}{\mathcal{T}}\sum_{-\tau'=1}^{\mathcal{T}}\big(Y_{1,\tau'}-Y_{0,\tau'}\big).$$

From the expression of the T-DiD in (3.1), $\widehat{ATT}^{S}_{\omega,T}$ is a special case of $\widehat{ATT}_{\omega,T}$ with uniform weighting in pre-treatment periods, i.e., $\psi_{\mathcal{T}}(-\tau)=1/\mathcal{T}$.

S.5 Discussions

S.5.1 Further discussion of the literature

The canonical DiD derives identifying variation from pooled cross-sectional units. It is one of the most commonly used econometric methods for estimating the causal effects of policy interventions in various social sciences, including economics and political science. For example, Giavazzi and Tabellini (2005), Persson (2005), Persson and Tabellini (2006), Papaioannou and

Siourounis (2008) and Persson and Tabellini (2009) use more or less the canonical DiD (C-DiD hereafter) to estimate the effect of democracy on economic growth. However, the C-DiD and its refinements, e.g., Ferman and Pinto (2019), Callaway and Sant'Anna (2021), Chan and Kwok (2021), Bellégo, Benatia, and Dortet-Bernadet (2024), Galindo-Silva, Some, and Tchuente (2018), and Picchetti, Pinto, and Shinoki (2024) are unsuited for fixed-N large-T settings (such as the one under consideration in this paper) where there may be temporal dependence, unit roots, deterministic trends, or cross-sectional dependence between units arising from, e.g., business cycles, persistent series, and correlated shocks.

Pooled regression analyses often rule out such cross-sectional dependence, which is common in fixed-N large-T settings. For example, it is plausible that common waves of democratisation or de-democratisation affect multiple countries, thereby inducing cross-sectional dependence. Consider, for instance, the wave of democratisation in West Africa in the early 1990s and the Arab Spring uprisings of the early 2010s. Estimating ATT parameters using two-way fixed effects models faces challenges similar to those in pooled regression analyses, where significant heterogeneity in treatment effects is often masked. For these reasons, the literature largely turns to Synthetic Control (SC) methods in such settings – see e.g., Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), Xu (2017), Abadie (2021), Ferman and Pinto (2021), and Sun, Ben-Michael, and Feller (2023). SC formalises case studies, and it is adapted to exploiting the temporal variation in the data to estimate treatment effects.

There are, however, documented impediments to using the SC from a practical perspective. The SC requires a donor pool of multiple control units. For example, bias from the pre-treatment fit when the number of control units is fixed may not shrink to zero as the number of pre-treatment periods increases (Ferman and Pinto, 2021). Additionally, the survey in Appendix A of Lei and Sudijono (2024) indicates that unit-level placebo inference, which several SC methods use for inference and robustness checks, may be impractical at small nominal levels — see also Fry (2024). In fact, the power of such tests is increasing in the number of selected control units, whereas the smallest attainable p-value is decreasing in the number of selected controls. Per-period treatment effects on the treated unit, which are typical target SC estimands, are not consistently estimable under fixed-N settings. Consequently, the resulting confidence or predictive intervals can be wide and uninformative (Chernozhukov, Wuthrich, and Zhu, 2024). Further, a good pre-treatment fit of the treated unit may not carry over to control units in the donor pool thus complicating the use of permutation methods for SC-based inference (Abadie, 2021). The setting of Li (2020), namely several pre- and post-treatment periods with a fixed number of units is close to the one under consideration in this paper. There are important differences to highlight, however. (1) Li (2020) uses the SC framework and inference therein is sub-sampling-based and non-standard, whereas the T-DiD is asymptotically normal. (2) The Li (2020) model requires multiple (albeit a finite number of) controls for the pre-treatment fitting procedure to be meaningful while the baseline (building block) of the proposed T-DiD requires only one control unit. (3) Li (2020, Assumption 1) imposes stationarity, while it is not required in the current paper.

There are drawbacks to using the SC from an econometric viewpoint. (1) The SC requires a perfect pre-treatment fit, which is often hard to achieve or verify, especially when (potential) outcomes are non-stationary. Inadequate pre-treatment fit, which is more common with longer time series, can introduce bias (Ferman and Pinto, 2021; Ben-Michael, Feller, and Rothstein, 2021). When the pre-treatment fit is good, it may be attributable to non-stationary common shocks or time trends – see, e.g., Ferman and Pinto (2021, p. 1216). (2) SC methods draw on cross-sectional dependence typically modelled via linear factor models, e.g., Ferman and Pinto (2021), Fry (2024), and Sun, Ben-Michael, and Feller (2023). Without strong cross-sectional dependence, the SC can suffer identification challenges and pre-treatment fit bias. (3) When there is only one suitable control unit, as in the current running empirical example, the convex-weighted SC *trivially* reduces to the difference between post-treatment treated and control outcomes. The post-treatment control outcome serves as the post-treatment untreated potential outcome for the treated unit. The convex-weighted demeaned SC is algebraically shown in Appendix S.4.2 to be a form of the proposed T-DiD. SC permutation-based inference is, however, infeasible in this setting as there is only one control unit. This implies a stronger identification condition relative to the proposed T-DiD – the (weighted) average of untreated potential outcomes of both the treated and untreated units must be equal. This rules out *characteristically* different means of the untreated potential outcomes of both units. Although the T-DiD estimator requires at least one valid control unit while the SC only needs a donor pool of peers, it is worth emphasising that control validity in the DiD framework can be weaker than that of the SC. For example, the outcome of a valid DiD control unit does not need to be in *any way* dependent on that of the treated unit in pre-treatment periods since a pre-post asymptotic parallel trends condition suffices. Thus, control units that may be individually SC-irrelevant can be DiD-valid.

The preceding paragraphs highlight two important points. First, an empirical setting with as few as a single treated unit and a single control unit easily arises and becomes empirically relevant once a researcher is interested in unit-specific effects *in lieu of* treatment effects from pooled regression analyses where heterogeneity can be substantially masked. Second, existing methods viz. the C-DiD, SC, and factor models are inadequate in this specific setting. Hence, this paper proposes the T-DiD, which exploits temporal variation in the data for consistent estimation and reliable inference in the presence *or absence* of cross-sectional dependence. Drawing on recent advances in the DiD literature on treatment effect heterogeneity, e.g., Callaway and Sant'Anna (2021), de Chaisemartin and D'Haultfœuille (2020), and Borusyak, Jaravel, and Spiess (2024), this paper defines convex-weighted treatment effects on the treated across time and proposes a consistent and asymptotically normal DiD estimator for it.[22] Thus, this paper introduces the DiD into a space within the body of literature that has hitherto been the preserve of Synthetic Control (SC) methods, factor models, and comparative studies. This paper provides easily interpretable asymptotic parallel trends and asymptotically limited anticipation assumptions under which convex-weighted treatment effects on the treated, namely $ATT_{\omega,T} := \sum_{t=1}^{T} \omega_T(t) ATT(t)$, are *asymptotically* identified with as few as a single treated and a single control unit and several pre- and post-treatment periods.

This paper makes a methodological contribution to the literature on the estimation of treatment effects in time series settings. Arkhangelsky et al. (2021) introduces the Synthetic Difference-in-Differences that combines attractive features of both the SC and C-DiD. It is unsuited for the two-unit baseline case considered in this paper, as it requires the number of untreated units to go to infinity – see Assumption 2 therein. Gonçalves and Ng (2024) provides improved predictors of, e.g., SC counterfactuals by exploiting serial correlation in the error. Pesaran and Smith (2016) and Angrist, Jordà, and Kuersteiner (2018) construct counterfactual outcomes from pre-treatment covariates of the treated unit under the identification assumption that the latter are invariant to treatment. A special case of the aforementioned (without pre-treatment covariates) is the before-after (BA) difference in means estimator, see e.g., Carvalho, Masini, and Medeiros (2018, p. 366) for a discussion. The BA uses the average pre-treatment outcome of the treated unit as its post-treatment untreated potential outcome. The BA estimator, like the Pesaran and Smith (2016), Angrist, Jordà, and Kuersteiner (2018), and Botosaru, Giacomini, and Weidner (2023) estimators, cannot disentangle common shocks from treatment effects, especially if these occur post-treatment.

A related strand of literature concerns the difference-in-discontinuities (diff-in-disc) estimator (e.g., Galindo-Silva, Some, and Tchuente, 2018; Picchetti, Pinto, and Shinoki, 2024), which combines the features of regression discontinuity designs and difference-in-differences. This approach is conceptually distinct from the T-DiD, as it identifies local treatment effects defined around a threshold of a running variable (such as time). In contrast, the T-DiD identifies an estimand that pools across all post-treatment periods. Thus, the two estimators are complementary: the diff-in-disc utilises cross-sectional variation (requiring large N) to estimate an ATT within an asymptotically shrinking bandwidth, whereas the T-DiD leverages a fixed number of units with large T to estimate a weighted average of ATTs globally. Consequently, while the identification conditions of the diff-in-disc are local, those of the T-DiD are global.

[22]Period-specific individual treatment effects on the treated are not consistently estimable in fixed-N settings.

S.5.2 Pre-trends testing

While tests of pre-treatment trends are known to only provide *suggestive evidence*, it is instructive to tease out hypotheses they *actually* test. The idea of a pre-test in the current context can be fashioned as follows. Choose a pre-treatment period $\mathcal{T}_o$ such that $\mathcal{T}_o/\mathcal{T} \in (0,1)$ and define $ATT_{\omega,\mathcal{T}_o} := \sum_{-\tau=1}^{\mathcal{T}_o} w(\tau)ATT(\tau)$. Extending the use of notation in an obvious way, one has the decomposition $ATT_{\omega,\mathcal{T}_o} = \widetilde{ATT}_{\mathcal{T}}^{w,\psi} - R_{\mathcal{T}}^{w,\psi}$ where as before, $\widetilde{ATT}_{\mathcal{T}}^{w,\psi}$ is identified but $ATT_{\omega,\mathcal{T}_o}$ is not, and $R_{\mathcal{T}}^{w,\psi}$ is the difference between a pre-treatment trend bias $TB_{\omega,\mathcal{T}}$ and anticipation bias $ATT_{\omega,\mathcal{T}-\mathcal{T}_o}$. Decompose $R_{\mathcal{T}}^{w,\psi}$ further as $R_{\mathcal{T}}^{w,\psi} := TB_{\omega,\mathcal{T}} - ATT_{\omega,\mathcal{T}-\mathcal{T}_o}$. Thus,

$$\widetilde{ATT}_{\mathcal{T}}^{w,\psi} = TB_{\omega,\mathcal{T}} + (ATT_{\omega,\mathcal{T}_o} - ATT_{\mathcal{T}-\mathcal{T}_o}).$$

In line with the convergence rate-based hypotheses adopted in Section 5, a pre-test adapted to the current setting can be formulated using the representation $|\widetilde{ATT}_{\mathcal{T}}^{w,\psi}| = C_{\omega,\mathcal{T}}\mathcal{T}^{1/2-\check{\delta}}$ where $\{C_{\omega,\mathcal{T}} : \mathcal{T} \geq 1\}$ is a sequence of bounded positive constants.

$$\mathbb{H}_o^{pt} : \check{\delta} > 1/2, \quad \mathbb{H}_{an}^{pt} : \check{\delta} = 1/2, \quad \mathbb{H}_a^{pt} : \check{\delta} < 1/2.$$

The pre-test can be implemented using the t-statistic

$$\widehat{t}_{\omega,\mathcal{T}} = \frac{\widehat{ATT}_{\omega,\mathcal{T}_o}}{\widehat{s}_{\omega,\mathcal{T}}} \text{ where } \widehat{ATT}_{\omega,\mathcal{T}_o} = \sum_{-\tau=1}^{\mathcal{T}} \omega_{\mathcal{T}}(\tau)X_\tau,$$

$\omega_{\mathcal{T}}(\tau) := w_{\mathcal{T}_o}(\tau)\mathbb{1}\{\tau > \mathcal{T}_o\} - \psi_{\mathcal{T}-\mathcal{T}_o}(\tau)\mathbb{1}\{\tau < \mathcal{T}_o\}$, and $\widehat{s}_{\omega,\mathcal{T}}$ is consistent for $s_{\omega,\mathcal{T}} := \mathbb{E}[(\widehat{ATT}_{\omega,\mathcal{T}_o} - \mathbb{E}[\widehat{ATT}_{\omega,\mathcal{T}_o}])^2]$ in the sense $\operatorname{plim}_{\mathcal{T}\to\infty} \widehat{s}_{\omega,\mathcal{T}}/s_{\omega,\mathcal{T}} = 1$.

$TB_{\omega,\mathcal{T}} = o(\mathcal{T}^{1/2})$ neither implies nor is implied by Assumption 1. Similarly, the rate condition on anticipation bias-related terms – $(ATT_{\omega,\mathcal{T}_o} - ATT_{\omega,\mathcal{T}-\mathcal{T}_o}) = o(\mathcal{T}^{1/2})$ – neither implies nor is implied by Assumption 2. Thus, parallel trends in pre-treatment periods offer no guarantee of parallel trends after $t = 0$ in the absence of treatment (Kahn-Lang and Lang, 2020; Dette and Schumann, 2024). The foregoing extends the arguments of Kahn-Lang and Lang (2020) on pre-tests in the C-DiD setting to the current setting with a large time dimension, temporal dependence, cross-sectional dependence, and anticipation bias.

Besides possibly distorting inference, pre-tests are documented in the literature to have poor power performance (Freyaldenhoven, Hansen, and Shapiro, 2019; Roth, 2022). Dette and Schumann (2024) innovatively reverses the burden of proof by testing for pre-treatment parallel trends as the alternative hypothesis. Thus, the test proposed therein under the alternative provides power in favour of pre-treatment parallel trends. While this approach addresses the power deficiency of commonly used tests of pre-treatment parallel trends, it does not cure the inherent problem underscored in the preceding paragraph of pre-tests (in general) being uninformative of the validity (or absence thereof) of parallel trends, especially when violations occur in the post-treatment period.

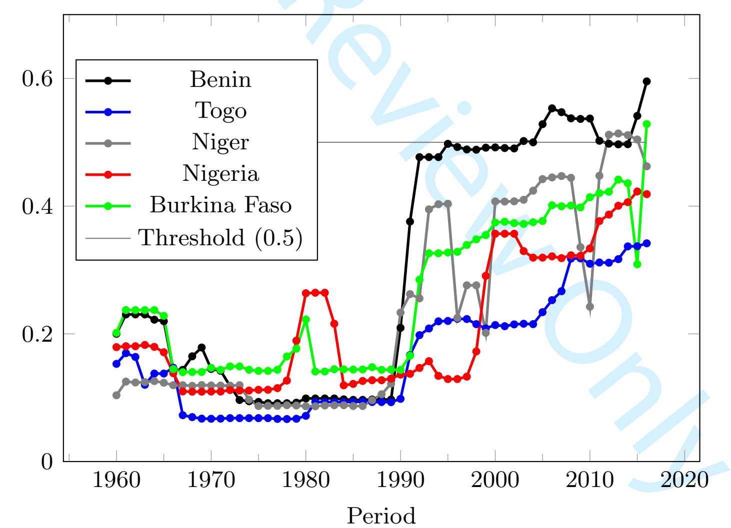
Pre-tests can lead to biased inference if used as a basis for estimating treatment effects on the treated (Roth, 2022). In a similar vein, the foregoing discussion points out why inference on $ATT_{\omega,T}$ based on passing a pre-test risks being misleading because "passing" a pre-test neither implies nor is implied by the identification of $ATT_{\omega,T}$. In contrast, the over-identifying restrictions test in Theorem 3 is valuable as it tests identification save on a space of distributions where the test has no power under the alternative hypotheses. For example, the proposed test can detect identification failure stemming from violations of Assumption 1 or Assumption 2 that might "pass" a pre-test. Moreover, pre-tests are only based on pre-treatment data, whereas the proposed test is based on the entire data set whence its ability to detect violations of

Assumption 1 in post-treatment periods and have more power as evidenced in simulations – see Appendix S.6.4.

S.5.3 Further discussion on the empirical analysis

The empirical application of interest in this paper examines the impact of democracy on economic growth. Given the current global decline in both democracy and satisfaction with democratic systems (Coppedge, Edgell, Knutsen, and Lindberg, 2022; Foa et al., 2020), it is apt to re-examine whether democracy drives economic growth and well-being. Several efforts have been devoted to quantifying the effect of democracy on economic growth – see Doucouliagos and Ulubaşoğlu (2008), Colagrossi, Rossignoli, and Maggioni (2020) and Knutsen (2021) for reviews. However, empirical findings on the relationship between democracy and economic growth are conflicting. Some studies, such as Acemoglu, Naidu, Restrepo, and Robinson (2019), report a positive relationship, while others, like Gerring, Bond, Barndt, and Moreno (2005), find a negative effect, and still others, such as Murtin and Wacziarg (2014), observe no significant effect. As Knutsen (2021) points out, these conflicting findings stem not only from differences in modelling choices and data quality but also from the varying economic performance of regimes with similar levels of democracy, particularly at the autocratic end of the spectrum. For example, Chen and Stengos (2024) finds heterogeneity in effects by regime type, institutional quality, and education levels – see also Acemoglu, Naidu, Restrepo, and Robinson (2019). Thus, one's answer to the democracy-growth question is largely an artefact of the composition of units in the pooled regression analyses.

Figure S.6: Democracy Indices



Notes: The plots above depict democracy indices for West African countries neighbouring Benin from 1960 through 2018.

For the empirical application of the T-DiD, this paper focuses on Benin, a country that experienced strong democratisation in the last three decades. The ideal pool of candidate controls includes neighbouring countries such as Togo, Burkina Faso, Niger, and Nigeria. However, Burkina Faso and Niger have experienced democratic regimes over some periods as shown in Figure S.6 using the democracy index based on the V-Dem project; see section 6.2 for more details. Nigeria has economic characteristics that are dissimilar to those of Benin. For example, Nigeria has an independent monetary policy while Benin does not. In sum, Togo remains the only suitable control unit among the aforementioned candidate controls.

Benin is a former French colony, and Togo is a former French protectorate. Both countries attained independence in 1960. They share a border of almost 651 km and are culturally similar. Both countries are members of the same monetary union – the West African Economic and Monetary Union (UEMOA) – even after independence, and they have the same monetary policy. Politically – see, e.g., Kohnert (2021) – the two countries were autocracies until 1990 when Benin initiated a process of democratisation. The former French president François Mitterand, in the wake of the fall of the Berlin Wall in 1989 and the implosion of the Soviet Union, encouraged francophone African countries to adopt democracy.[23] Benin and Togo have both undergone democratisation processes, but their outcomes differ: Benin becomes a model democracy for the whole of Sub-Saharan Africa, while Togo remains largely perceived as an autocratic country. The choice of Togo as a control unit for Benin weakens the asymptotic parallel trends and limited anticipation identification conditions. Both identification conditions only need to hold conditional not only on the aforementioned observed characteristics but also on unobservable common characteristics.

S.6 Simulations

This section presents simulation results. Section S.6.2 presents the bias and empirical rejection rates of competing estimators, and Section S.6.3 presents power curves corresponding to estimator-based t-tests. The power curves serve to examine the ability of the t-tests to detect significant $ATT_{\omega,T}$. Section Appendix S.6.4 examines the size and power performance of the proposed identification tests, and Section Appendix S.6.4 extends the results to heterogeneous (through time) treatment effects.

S.6.1 DGPs

This section examines the small-sample performance of the T-DiD compared to alternatives such as the SC and the BA under several interesting data-generating processes (DGPs). Consider the following DGP of untreated potential outcomes

$$Y_{d,t}(0) = \alpha_0 + d(\alpha_1 - \alpha_0) + \alpha_2 Y_{d,t-1}(0) + \varphi(t) + d\nu_t + \big(1 - d(1 - 1/\sqrt{2})\big)\big(e_{0,t} + \alpha_3 e_{0,t-1}\big),$$

with the observed outcome generated as

$$Y_{d,t} = Y_{d,t}(0) + d\Big(ATT(t) + \alpha_2\big(Y_{d,t-1} - Y_{d,t-1}(0)\big) + \alpha_4 t + (1/\sqrt{2})\big(e_{1,t} + \alpha_3 e_{1,t-1}\big)\Big) \quad \text{(S.6.1)}$$

where $Y_{d,t}(0)$ denotes the untreated potential outcome of unit $d \in \{0,1\}$. $ATT(t) = 0.0$ for all periods for all DGPs. Variations of $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)'$, $\varphi(t)$, ν_t, and $e_{d,t}$ generate different DGPs described below. $\alpha_1 \neq \alpha_0$ induces a location shift between the untreated potential outcomes of the treated and untreated units; the SC fails under this scenario. $Y_{d,t}$ and $Y_{d,t}(0)$ are first-order auto-regressive AR(1) and first-order moving average MA(1) processes when $|\alpha_2| \in (0,1)$ and $|\alpha_3| \in (0,1)$, respectively. $Y_{d,t}$ and $Y_{d,t}(0)$ become unit root processes when $\alpha_2 = 1$. $\alpha_4 \neq 0$ leads to uncommon trend non-stationarity in $Y_{1,t}$, unlike in $Y_{0,t}$. The errors of the observed and untreated potential outcomes of both the treated and untreated units at each period $t \in [-\mathcal{T}] \cup [T]$ are correlated, thus there is cross-sectional dependence across units with correlation coefficient $\text{cor}[e_{0t}, (1/\sqrt{2})(e_{0t} + e_{1t})] = 1/\sqrt{2} \approx 0.707$. $\varphi(\cdot)$ introduces a common

[23] The statement was made in his speech at Baule during the 16th Conference of Heads of State of Africa and France – see https://www.vie-publique.fr/discours/127621-allocution-de-m-francois-mitterrand-president-de-la-republique-sur-la.

trend while the random variable ν_t introduces violations of the standard parallel trends and no-anticipation assumptions. Quite importantly, the BA fails whenever $\varphi(\cdot)$ in the post-treatment period is not cancelled out (at least asymptotically) by that of the pre-treatment period. For inference, the convex-weighted SC and BA use a straightforward adaptation of the asymptotic normality result in this paper.

The following specifies forms of the common trend $\varphi(t)$ and the error e_{dt}:

$$\varphi(t) = \begin{cases} \varphi_{NT}(t) := 0.0 & \text{no trends} \\ \varphi_{BST}(t) := \sqrt{2}\mathbb{1}\{t \geq 4\} & \text{bounded binary trend} \\ \varphi_{CST}(t) := \cos(t) & \text{bounded continuous trend} \\ \varphi_{NSQT}(t) := t + t^2/500 & \text{unbounded quadratic trend} \end{cases}$$

and

$$e_{dt} = \begin{cases} \varepsilon_{dt}\varepsilon_{dt-1} & \text{MDS} \\ \sigma_{dt}\varepsilon_{dt},\ \sigma_{dt}^2 = 0.4 + 0.3e_{dt-1}^2 + 0.3\sigma_{dt-1}^2 & \text{GARCH(1,1)} \\ \big(\varepsilon_{dt} + \varepsilon_{dt-1}\varepsilon_{dt-2}\big)/\sqrt{2} & \text{White Noise} \end{cases}$$

where $\varepsilon_{0t} \overset{iid}{\sim} (\chi^2 - 1)/\sqrt{2}$ and $\varepsilon_{1t} \overset{iid}{\sim} \mathbb{T}_T/\sqrt{T/(T-2)}$ where $\mathbb{T}_T$ denotes the Student t-distribution with T degrees of freedom. Simulation results are based on $10,000$ Monte Carlo samples. Eleven DGPs are considered in total:

1. DGP SC-BA: $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0$, $\nu_t = 0$, e_{dt} is MDS;
2. DGP BA: $\alpha_1 = -\alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0$, $\nu_t = 0$, e_{dt} is MDS;
3. DGP SC: $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = \varphi_{BST}(t)$, $\nu_t = 0$, e_{dt} is MDS;
4. (A) DGP PT-NA: $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0.0$, $\nu_t \sim \mathcal{N}\big(0.5\mathbb{1}\{t \neq 0\}\text{sgn}(t)|t|^{-0.9}, 1\big)$, e_{dt} is MDS;
4. (B) DGP PT-NA: $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0.0$, $\nu_t \sim \mathcal{N}\big(0.5\mathbb{1}\{t \neq 0\}|t|^{-0.25}, 1\big)$, e_{dt} is MDS;
5. DGP GARCH(1,1): $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0.0$, $\nu_t = 0.0$, e_{dt} is GARCH(1,1);
6. DGP MA(1): $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = 0.0$, $\alpha_3 = 0.25$, $\varphi(t) = 0.0$, $\nu_t = 0.0$, e_{dt} is White Noise;
7. DGP AR(1): $\alpha_0 = \alpha_1 = \alpha_2 = 0.5$, $\alpha_3 = 0.0$, $\varphi(t) = \cos(t)$, $\nu_t = 0.0$, e_{dt} is White Noise;
8. DGP U-R: $\alpha_0 = \alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\alpha_3 = 0.0$, $\varphi(t) = 0.0$, $\nu_t = 0.0$, e_{dt} is White Noise;
9. DGP Q-T: $\alpha_0 = \alpha_1 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = \varphi_{NSQT}(t)$, $\nu_t = 0.0$, e_{dt} is White Noise; and
10. DGP T-T: $\alpha_1 = \alpha_0 = 0.5$, $\alpha_2 = \alpha_3 = 0.0$, $\varphi(t) = 0.0$, $\alpha_4 = 1.0$, $\nu_t = 0$, e_{dt} is MDS.

$\alpha_0 \neq \alpha_1$ in DGP BA, thus the SC estimator is not expected to perform well. Likewise, $\varphi(t) = \varphi_{BST}(t)$ induces a temporal location shift in DGP SC, thus it is not expected to be favourable to the BA estimator. ν_t in DGP (A) PT-NA induces violations of the standard parallel trends and no anticipation assumptions although it satisfies Assumptions 1 and 2. Unlike in DGP PT-NA (A), weaker rates on the violations of Assumptions 1 and 2 are allowable in DGP PT-NA (B) as both biases tend to cancel out asymptotically between the pre-treatment and post-treatment periods. In view of Remark 2.3, Assumptions 1 and 2 are both violated in DGP PT-NA (B) but identification holds because the biases tend to cancel out asymptotically. A common quadratic term via $\varphi(t) = t + t^2/500$ induces trend non-stationarity in DGP Q-T; this DGP is not favourable to the BA estimator. The trend in DGP T-T is linear and affects only the treated unit; the trend leads to identification failure if not removed.

S.6.2 Bias and size control

Table S.1: Simulation - $ATT(t) = 0.0, \ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP SC-BA		MB			MAD			RMSE		
	25	0.002	0.002	0.001	0.14	0.096	0.175	0.216	0.151	0.280
	50	-0.001	0.000	-0.001	0.101	0.07	0.129	0.153	0.107	0.202
	100	0.000	0.000	-0.001	0.072	0.051	0.094	0.108	0.077	0.142
	200	0.001	0.001	0.002	0.051	0.036	0.066	0.076	0.053	0.100
	400	0.000	0.000	0.000	0.037	0.025	0.047	0.054	0.038	0.070
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.012	0.012	0.013	0.057	0.056	0.056	0.116	0.110	0.114
	50	0.012	0.009	0.011	0.052	0.053	0.055	0.105	0.104	0.110
	100	0.008	0.010	0.010	0.050	0.053	0.053	0.101	0.105	0.106
	200	0.008	0.009	0.010	0.049	0.045	0.049	0.100	0.096	0.101
	400	0.011	0.010	0.009	0.049	0.049	0.047	0.099	0.099	0.097
DGP BA		MB			MAD			RMSE		
	25	0.002	-0.998	0.001	0.14	0.997	0.175	0.216	1.009	0.280
	50	-0.001	-1.000	-0.001	0.101	1.000	0.129	0.153	1.006	0.202
	100	0.000	-1.000	-0.001	0.072	1.000	0.094	0.108	1.003	0.142
	200	0.001	-0.999	0.002	0.051	0.999	0.066	0.076	1.001	0.100
	400	0.000	-1.000	0.000	0.037	1.000	0.047	0.054	1.001	0.070
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.012	0.990	0.013	0.057	0.997	0.056	0.116	0.999	0.114
	50	0.012	1.000	0.011	0.052	1.000	0.055	0.105	1.000	0.110
	100	0.008	1.000	0.010	0.050	1.000	0.053	0.101	1.000	0.106
	200	0.008	1.000	0.010	0.049	1.000	0.049	0.100	1.000	0.101
	400	0.011	1.000	0.009	0.049	1.000	0.047	0.099	1.000	0.097
DGP SC		MB			MAD			RMSE		
	25	0.002	0.002	1.246	0.140	0.096	1.248	0.216	0.151	1.277
	50	-0.001	0.000	1.328	0.101	0.070	1.329	0.153	0.107	1.343
	100	0.000	0.000	1.371	0.072	0.051	1.372	0.108	0.077	1.378
	200	0.001	0.001	1.395	0.051	0.036	1.393	0.076	0.053	1.398
	400	0.000	0.000	1.404	0.037	0.025	1.404	0.054	0.038	1.406
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.012	0.012	0.919	0.057	0.056	0.965	0.116	0.110	0.979
	50	0.012	0.009	0.992	0.052	0.053	0.996	0.105	0.104	0.998
	100	0.008	0.010	1.000	0.050	0.053	1.000	0.101	0.105	1.000
	200	0.008	0.009	1.000	0.049	0.045	1.000	0.100	0.096	1.000
	400	0.011	0.010	1.000	0.049	0.049	1.000	0.099	0.099	1.000

Tables S.1 to S.4 present the mean bias (MB), the median absolute deviation (MAD), the

Table S.2: Simulation - $ATT(t) = 0.0,\ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
			MB			MAD			RMSE	
DGP PT-NA (A)	25	0.177	0.089	0.176	0.270	0.179	0.294	0.397	0.265	0.436
	50	0.109	0.055	0.109	0.185	0.127	0.201	0.273	0.187	0.302
	100	0.063	0.030	0.062	0.128	0.087	0.139	0.190	0.129	0.212
	200	0.037	0.018	0.037	0.089	0.061	0.099	0.132	0.090	0.147
	400	0.023	0.011	0.023	0.063	0.043	0.068	0.092	0.064	0.102
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.030	0.029	0.027	0.096	0.086	0.092	0.162	0.142	0.155
	50	0.020	0.021	0.017	0.077	0.074	0.076	0.140	0.130	0.133
	100	0.019	0.018	0.016	0.069	0.059	0.070	0.126	0.111	0.128
	200	0.015	0.012	0.013	0.065	0.056	0.062	0.117	0.110	0.119
	400	0.012	0.012	0.011	0.056	0.054	0.056	0.107	0.105	0.108
			MB			MAD			RMSE	
DGP PT-NA (B)	25	0.001	0.287	0.000	0.241	0.294	0.269	0.356	0.381	0.399
	50	0.001	0.246	0.001	0.168	0.248	0.187	0.250	0.304	0.282
	100	-0.002	0.205	-0.002	0.121	0.206	0.134	0.180	0.241	0.202
	200	-0.001	0.175	-0.001	0.085	0.175	0.095	0.126	0.196	0.142
	400	0.001	0.148	0.001	0.06	0.148	0.067	0.089	0.161	0.099
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.016	0.111	0.016	0.065	0.246	0.061	0.118	0.347	0.119
	50	0.013	0.145	0.011	0.056	0.308	0.052	0.105	0.416	0.107
	100	0.012	0.187	0.010	0.053	0.382	0.057	0.102	0.502	0.108
	200	0.010	0.283	0.011	0.051	0.508	0.049	0.103	0.628	0.102
	400	0.010	0.415	0.009	0.048	0.648	0.047	0.099	0.758	0.101
			MB			MAD			RMSE	
DGP GARCH(1,1)	25	0.004	0.001	-0.003	0.141	0.098	0.177	0.214	0.154	0.275
	50	0.001	0.000	0.002	0.102	0.072	0.130	0.152	0.107	0.199
	100	0.000	0.000	0.000	0.071	0.051	0.092	0.108	0.077	0.141
	200	0.000	0.000	0.000	0.052	0.037	0.066	0.077	0.054	0.099
	400	0.000	0.000	0.000	0.036	0.026	0.047	0.054	0.038	0.071
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.017	0.019	0.014	0.064	0.072	0.058	0.120	0.123	0.111
	50	0.012	0.012	0.012	0.055	0.055	0.055	0.108	0.107	0.106
	100	0.011	0.013	0.009	0.056	0.055	0.051	0.106	0.107	0.100
	200	0.011	0.011	0.009	0.053	0.051	0.046	0.099	0.103	0.096
	400	0.009	0.010	0.010	0.049	0.052	0.048	0.098	0.105	0.100

Table S.3: Simulation - $ATT(t) = 0.0,\ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP MA(1)		MB			MAD			RMSE		
	25	0.005	0.003	-0.002	0.174	0.119	0.218	0.269	0.190	0.347
	50	-0.001	0.000	-0.002	0.128	0.088	0.164	0.192	0.134	0.251
	100	-0.001	0.000	-0.001	0.091	0.065	0.118	0.135	0.096	0.178
	200	0.001	0.001	0.001	0.064	0.045	0.083	0.095	0.067	0.125
	400	0.000	0.000	0.000	0.045	0.032	0.058	0.067	0.048	0.088
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.024	0.022	0.017	0.078	0.074	0.068	0.142	0.131	0.130
	50	0.016	0.014	0.016	0.071	0.063	0.068	0.128	0.121	0.124
	100	0.014	0.013	0.011	0.061	0.061	0.057	0.119	0.116	0.117
	200	0.014	0.012	0.013	0.057	0.058	0.056	0.107	0.109	0.112
	400	0.011	0.012	0.009	0.052	0.050	0.052	0.103	0.104	0.110
DGP AR(1)		MB			MAD			RMSE		
	25	0.003	-0.021	0.003	0.162	0.112	0.208	0.252	0.183	0.324
	50	-0.001	-0.012	0.000	0.111	0.077	0.145	0.168	0.119	0.217
	100	0.000	-0.007	-0.001	0.076	0.054	0.100	0.113	0.082	0.147
	200	0.001	-0.003	0.001	0.053	0.037	0.068	0.078	0.056	0.102
	400	0.000	-0.002	0.000	0.037	0.026	0.047	0.054	0.039	0.071
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.024	0.033	0.021	0.083	0.088	0.082	0.141	0.148	0.143
	50	0.018	0.022	0.02	0.072	0.069	0.071	0.129	0.126	0.133
	100	0.014	0.015	0.014	0.063	0.066	0.061	0.115	0.123	0.117
	200	0.013	0.013	0.015	0.055	0.059	0.062	0.105	0.111	0.114
	400	0.011	0.012	0.012	0.051	0.053	0.059	0.100	0.102	0.116
DGP U-R		MB			MAD			RMSE		
	25	0.003	0.002	-0.002	0.143	0.097	0.178	0.220	0.156	0.286
	50	-0.001	0.000	-0.002	0.103	0.071	0.131	0.155	0.108	0.203
	100	-0.001	0.000	-0.001	0.073	0.052	0.094	0.109	0.078	0.143
	200	0.001	0.000	0.001	0.051	0.036	0.066	0.077	0.054	0.100
	400	0.000	0.000	0.000	0.036	0.026	0.047	0.054	0.038	0.071
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.013	0.015	0.010	0.056	0.056	0.047	0.110	0.105	0.101
	50	0.010	0.010	0.009	0.052	0.048	0.051	0.108	0.099	0.101
	100	0.010	0.010	0.008	0.050	0.052	0.044	0.102	0.103	0.097
	200	0.011	0.009	0.009	0.050	0.052	0.048	0.097	0.100	0.097
	400	0.009	0.010	0.008	0.045	0.047	0.047	0.097	0.096	0.101

Table S.4: Simulation - $ATT(t) = 0.0,\ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP Q-T		MB			MAD			RMSE		
	25	0.003	0.002	25.998	0.141	0.096	25.998	0.217	0.153	25.999
	50	-0.001	0.000	50.998	0.103	0.07	50.998	0.154	0.107	50.998
	100	-0.001	0.000	100.998	0.073	0.052	100.998	0.108	0.077	100.999
	200	0.001	0.001	201.001	0.050	0.036	201	0.076	0.054	201.001
	400	0.000	0.000	401	0.036	0.026	401.001	0.054	0.038	401
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.013	0.015	1.000	0.060	0.058	1.000	0.113	0.105	1.000
	50	0.012	0.011	1.000	0.052	0.049	1.000	0.104	0.100	1.000
	100	0.010	0.010	1.000	0.051	0.051	1.000	0.103	0.103	1.000
	200	0.012	0.010	1.000	0.051	0.051	1.000	0.098	0.100	1.000
	400	0.010	0.010	1.000	0.046	0.047	1.000	0.097	0.096	1.000
DGP T-T		MB			MAD			RMSE		
	25	0.003	25.005	0.008	0.270	25.008	0.344	0.426	25.007	0.571
	50	-0.005	49.995	0.005	0.200	49.997	0.251	0.306	49.996	0.400
	100	-0.002	100.000	-0.002	0.140	100.002	0.183	0.214	100.000	0.282
	200	0.003	200.003	0.002	0.100	200.004	0.130	0.151	200.003	0.199
	400	0.000	400.000	-0.001	0.073	399.999	0.093	0.108	400.000	0.140
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.015	1.000	0.016	0.066	1.000	0.068	0.124	1.000	0.132
	50	0.013	1.000	0.014	0.060	1.000	0.062	0.118	1.000	0.118
	100	0.012	1.000	0.011	0.052	1.000	0.053	0.102	1.000	0.105
	200	0.009	1.000	0.011	0.048	1.000	0.049	0.099	1.000	0.101
	400	0.010	1.000	0.009	0.050	1.000	0.049	0.102	1.000	0.103

root-mean-squared error (RMSE), and the empirical rejection rates of the null hypothesis, $\mathbb{H}_o : ATT_{\omega,T} = 0.0$ at conventional nominal levels 1%, 5%, and 10% for the DiD, SC, and BA estimators across all 11 DGPs. Results are based on 10,000 Monte Carlo replications for each sample size $\mathcal{T} = T \in \{25, 50, 100, 200, 400\}$. Standard errors used throughout the text are heteroskedasticity and auto-correlation robust using the Newey and West (1987) procedure. For all competing estimators, lagged X_t are controlled for in DGP AR(1) while first differences are applied to X_t before estimation in DGP U-R. The uniform weighting scheme is used throughout.

All three estimators perform reasonably under scenarios where they are expected to. The poor performance of the SC in DGP BA confirms its sensitivity to differences in pre-treatment average outcomes, whereas the poor performance of the BA in DGP SC and Q-Trend confirms its sensitivity to common shocks, which the BA cannot disentangle from treatment effects. Although all estimators appear to suffer size distortion at small sample sizes under DGPs PT-NA (A), MA(1), and AR(1), these improve with the sample. The SC only uses data from the post-treatment periods so it fails to harness identifying variation from pre-treatment periods to counter bias occurring in the post-treatment periods. This explains the poor performance of the SC in DGP PT-NA (B) while both the T-DiD and BA perform reasonably. On the whole, one observes the robustness of the proposed T-DiD to several interesting and empirically relevant settings through the simulation results presented in Tables S.1 to S.4.

S.6.3 Detecting significant $ATT_{\omega,T}$s

To ensure that the good size control of the T-DiD is not achieved at the expense of power, the DGP in Appendix S.6 (see (S.6.1)) with $ATT(t) = ATT$, $ATT \in [0.0, 1.0]$ is used. The hypotheses are $\mathbb{H}_o : ATT_{\omega,T} = 0.0$ vs $\mathbb{H}_a : ATT_{\omega,T} \neq 0.0$. Thus, the goal is to ensure that hypothesis tests based on the DiD estimate control size under the null and have the power to detect non-trivial effects of treatment or, generally, deviations away from a posited null hypothesis on the size of treatment effects.

(a) 10% (b) 5% (c) 1%

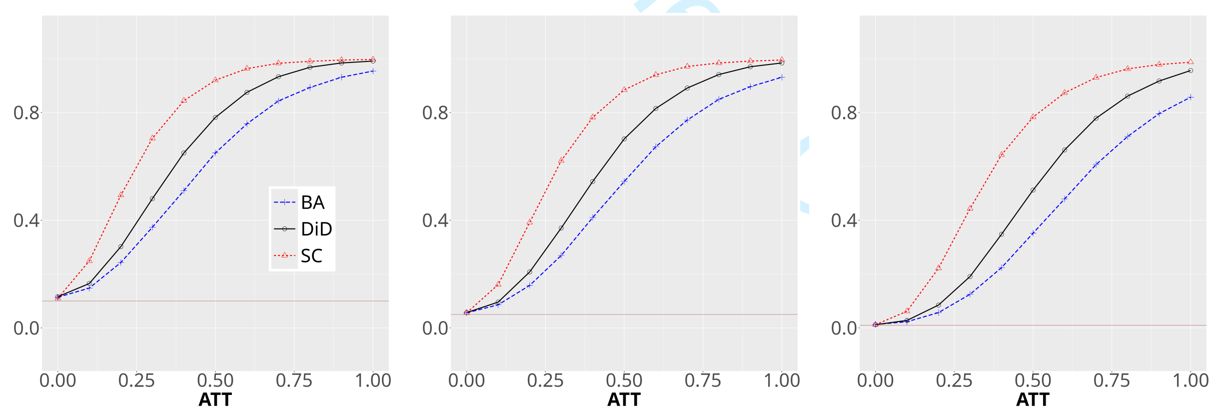


Figure S.7: Power Curves - DGP SC-BA, $\mathcal{T}, T = 25$.

Figures S.7 to S.17 present power curves corresponding to estimator-based t-tests of the hypotheses $\mathbb{H}_o : ATT_{\omega,T} = 0.0$ vs. $\mathbb{H}_a : ATT_{\omega,T} \neq 0.0$. Overall, the power curves continue to reflect the robustness of the T-DiD across interesting settings and its ability to detect significant treatment effects. One also observes the failure of either the SC or BA to meaningfully control size under $\mathbb{H}_o$, have power under $\mathbb{H}_a$, or both whenever any of the stringent conditions under which they hold is violated.

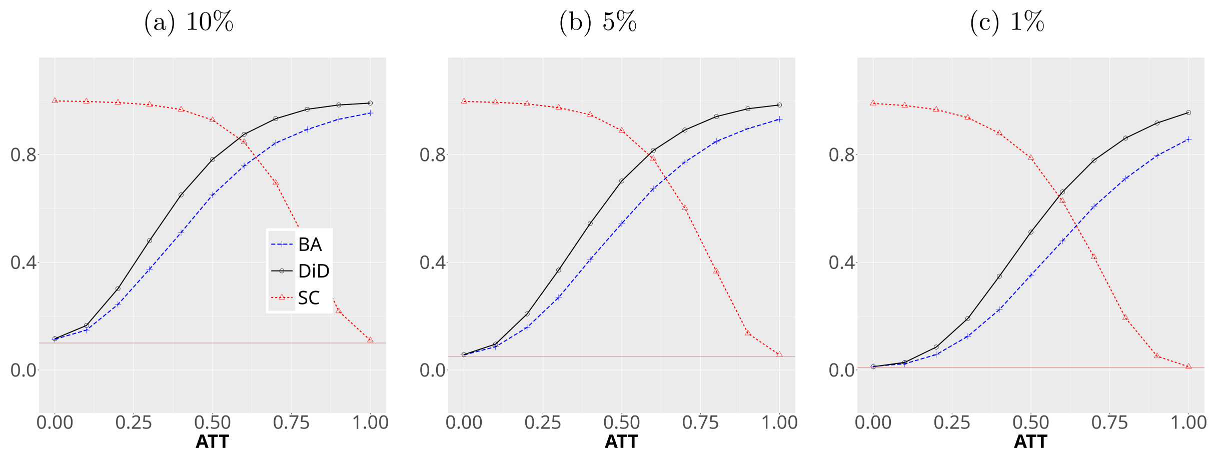


Figure S.8: Power Curves - DGP BA, $\mathcal{T}, T = 25$.

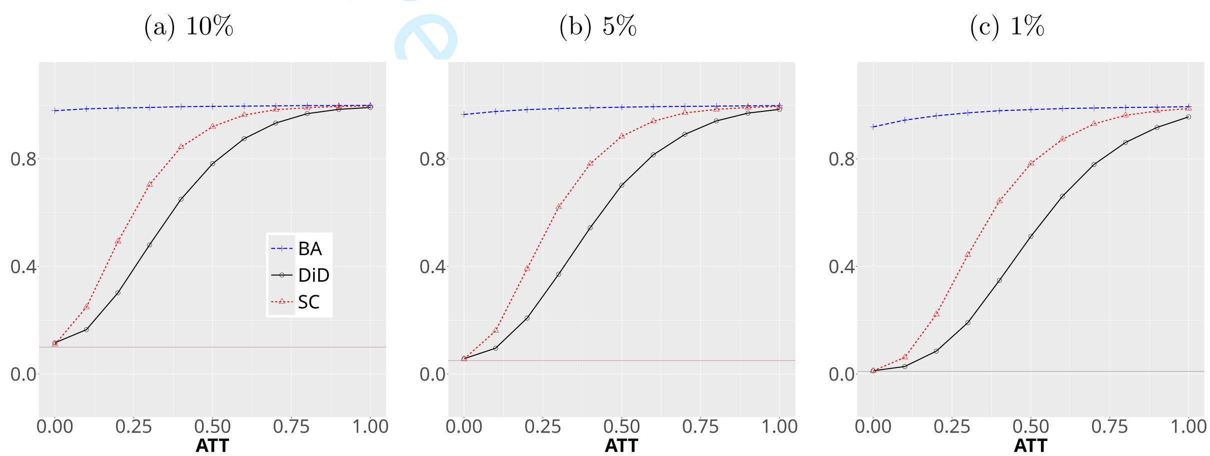


Figure S.9: Power Curves - DGP SC, $\mathcal{T}, T = 25$.

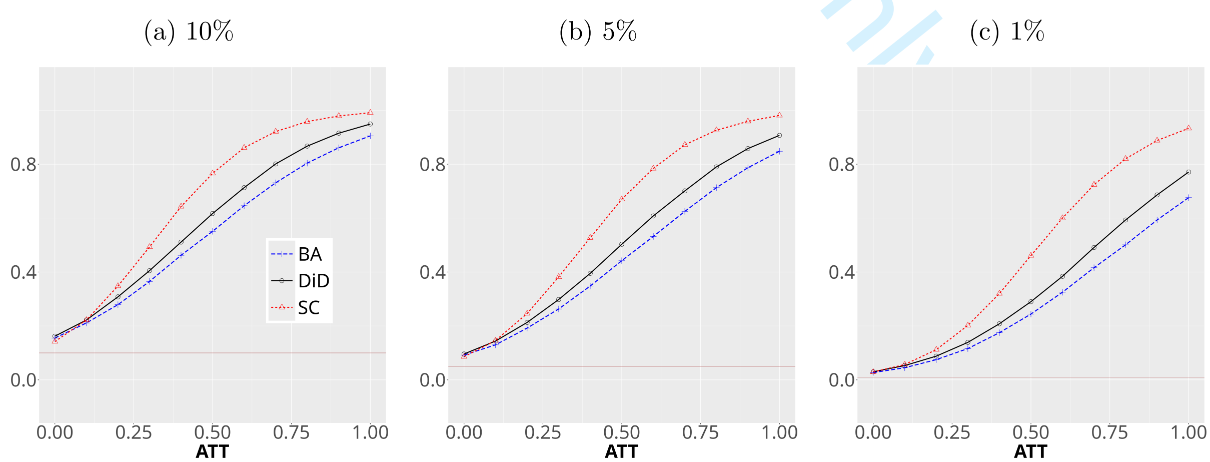


Figure S.10: Power Curves - DGP PT-NA (A), $\mathcal{T}, T = 25$.

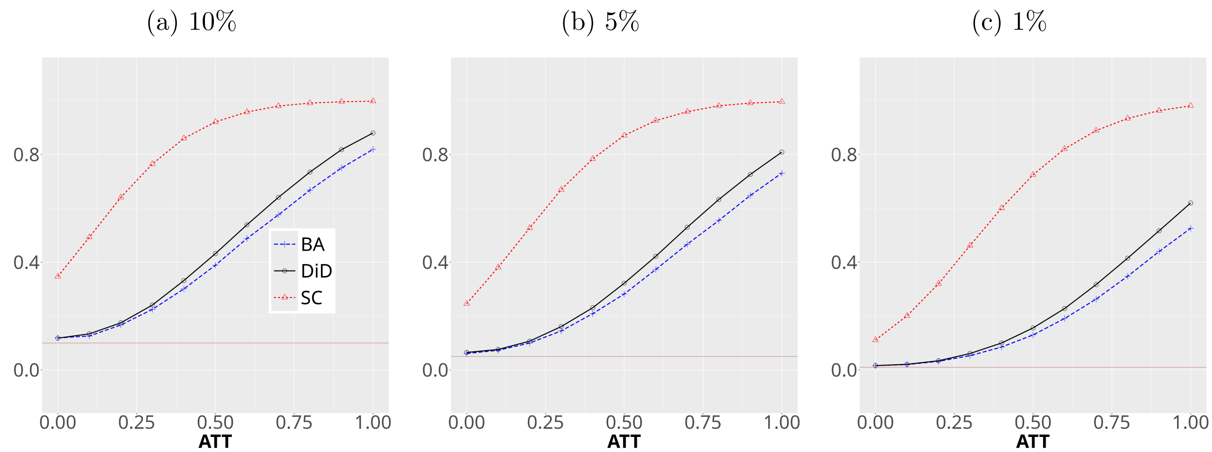


Figure S.11: Power Curves - DGP PT-NA (B), $\mathcal{T}, T = 25$.

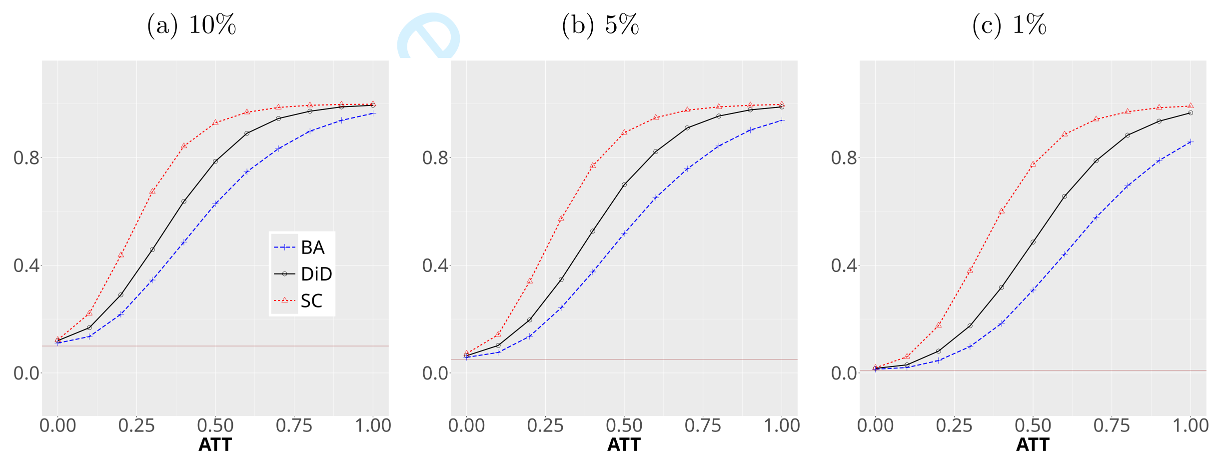


Figure S.12: Power Curves - DGP GARCH(1,1), $\mathcal{T}, T = 25$.

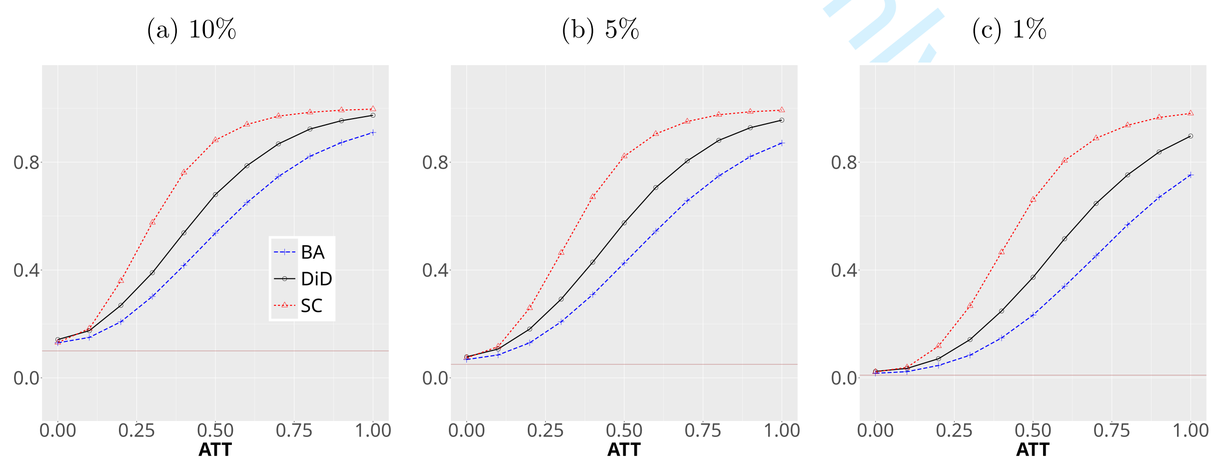


Figure S.13: Power Curves - DGP MA(1), $\mathcal{T}, T = 25$.

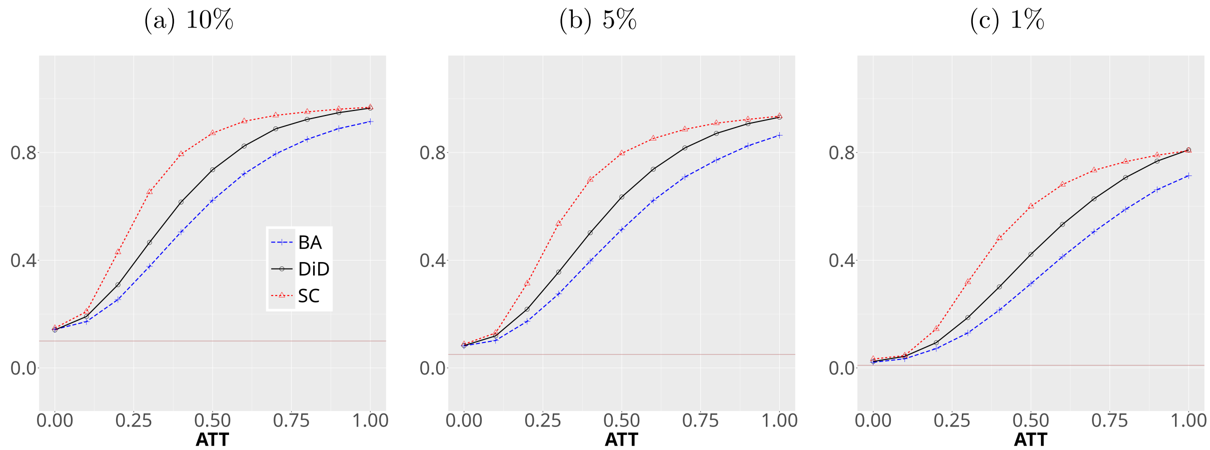


Figure S.14: Power Curves - DGP AR(1), $\mathcal{T}, T = 25$.

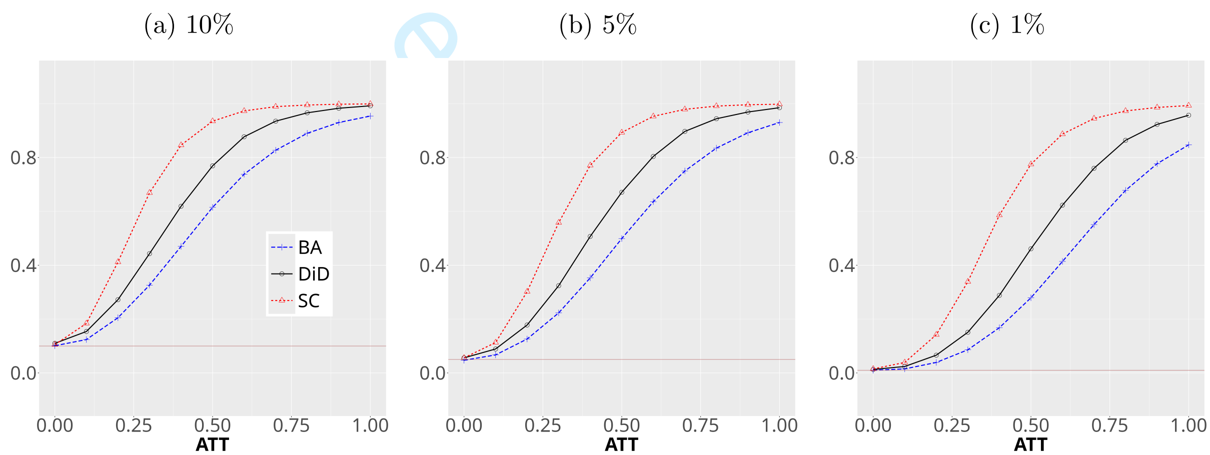


Figure S.15: Power Curves - DGP U-R, $\mathcal{T}, T = 25$.

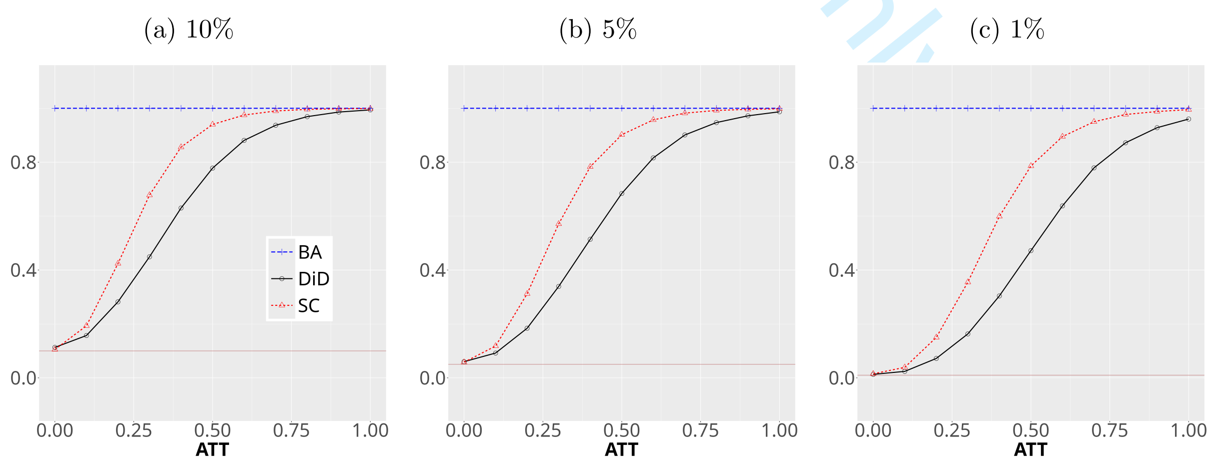


Figure S.16: Power Curves - DGP Q-T, $\mathcal{T}, T = 25$.

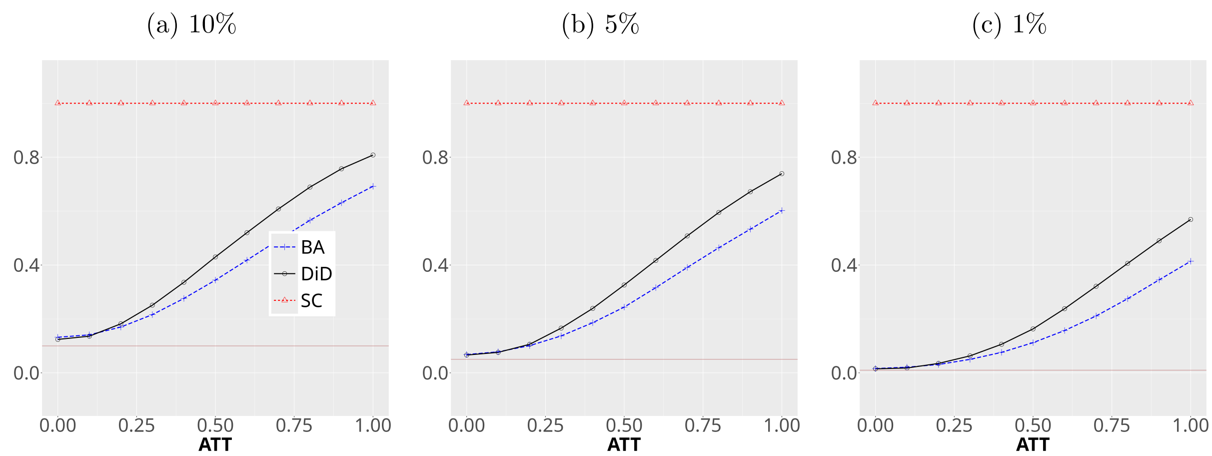


Figure S.17: Power Curves - DGP T-T, $\mathcal{T}, T = 25$.

S.6.4 Testing identification

This section examines the empirical sizes and power performance of the family of over-identifying restrictions and pre-trends tests. As both types of tests are based on estimates of ATT on some (sub-)sample, this gives rise to over-identifying restrictions tests id.BA, id.DiD, and id.SC and pre-tests pt.BA, pt.DiD, and pt.SC, where the prefixes "id" and "pt" denote over-identifying restrictions and pre-trends tests, respectively, using the prefixed estimator SC, DiD, or BA. Pre-trends tests are conducted using approximately half of the pre-treatment periods as a "pre-treatment" period and the other half as the "post-treatment" period.

Modifications of DGPs PT-NA (A) and (B) via the term ν_t provide the following scenarios for examining over-identifying restrictions and pre-trends identification tests using simulations;

1. DGP idTest I : $\nu_t \sim \mathcal{N}\big(2.5h|t|^{-0.25}, 1\big)$ if $t \leq -1$ and $\nu_t = 0$ otherwise, $\alpha_0 = -\alpha_1 = 0.5$;
2. DGP idTest II : $\nu_t \sim \mathcal{N}\big(2.5h|t|^{-0.25}, 1\big)$ if $t \geq 1$ and $\nu_t = 0$ otherwise, $\alpha_0 = \alpha_1 = 0.5$, $\varphi(t) = \sqrt{2}\mathbb{1}\{t \geq 4\}$; and
3. DGP idTest III : $\nu_t \sim \mathcal{N}\big(2.5(1-h(1-\text{sgn}(t))|t|^{-0.25}, 1\big)$, $\alpha_0 = -\alpha_1 = 0.5$, $\varphi(t) = \sqrt{2}\mathbb{1}\{t \geq 4\}$.

$ATT = 0.0$; thus, a "treated" unit is as good as a "control" unit. For the comparison of tests of identification, the t-test version of the over-identifying restrictions test is used since there are only two candidate controls. Thus, both units are candidate controls for the over-identifying restrictions tests; one unit serves as the "treated unit" while the other serves as a "control unit" for pre-trends tests.

In all three DGPs above, identification in the T-DiD context holds when $h = 0$. Both id.DiD and pt.DiD should control size in DGPs idTest I and idTest II while only id.DiD is expected to control size in DGP idTest III. In fact, both Assumptions 1 and 2 are individually violated in DGP idTest III, but DiD-identification holds – see Remark 2.3. For $h \in (0, 1]$ in (1) DGP idTest I, violations of DiD identification occur in the pre-treatment period (which can be picked up by the pt.DiD), (2) DGP idTest II, violations occur in the post-treatment period, which cannot be picked up by the pt.DiD, and (3) violations occur in both pre- and post-treatment periods.

Figures S.18 to S.20 present power curves to compare all six candidate tests of identification. As expected, over-identifying restrictions tests based on the BA and SC continue to bear the drawbacks of the respective estimators. For example, id.BA jointly controls size meaningfully and has non-trivial power in Figure S.18 but not in Figure S.19 whereas the opposite holds for id.SC. All pre-tests have trivial power, fail to control size meaningfully, or both in Figures S.18 to S.20. In sum, this short simulation exercise demonstrates the reliability of the DiD-based over-identifying restrictions test of identification even in scenarios, such as DGP idTest II where

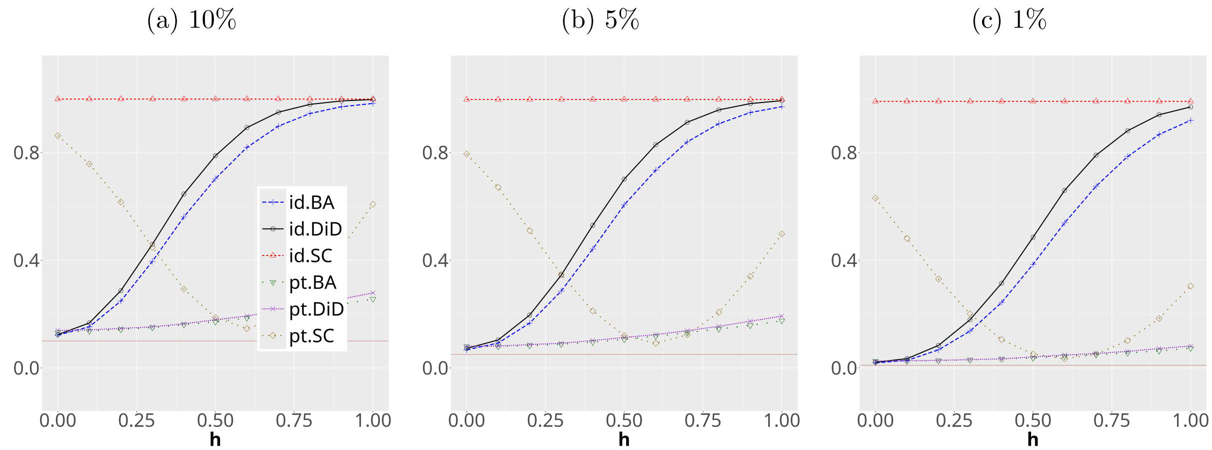


Figure S.18: Power Curves - DGP idTest I, $\mathcal{T}, T = 25$.

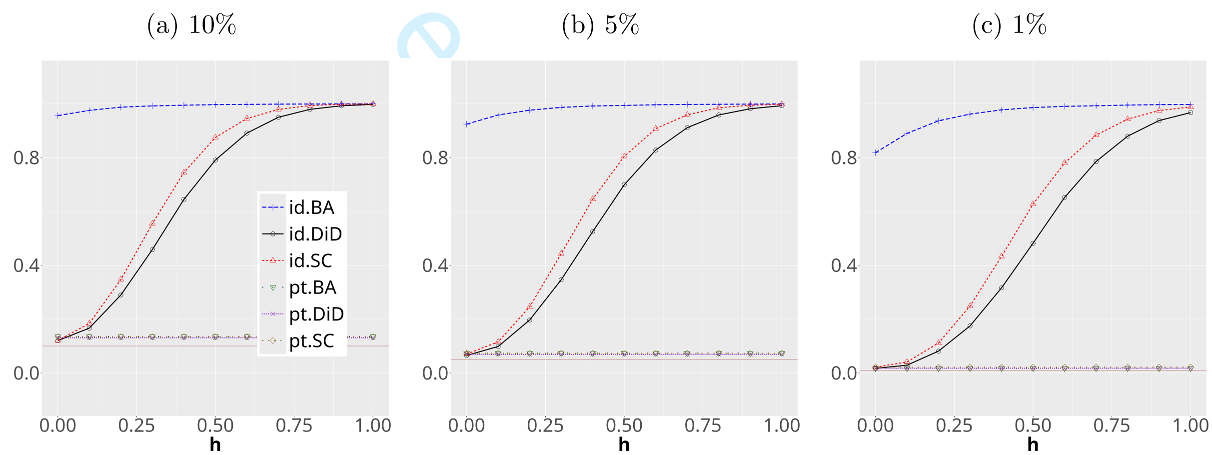


Figure S.19: Power Curves - DGP idTest II, $\mathcal{T}, T = 25$.

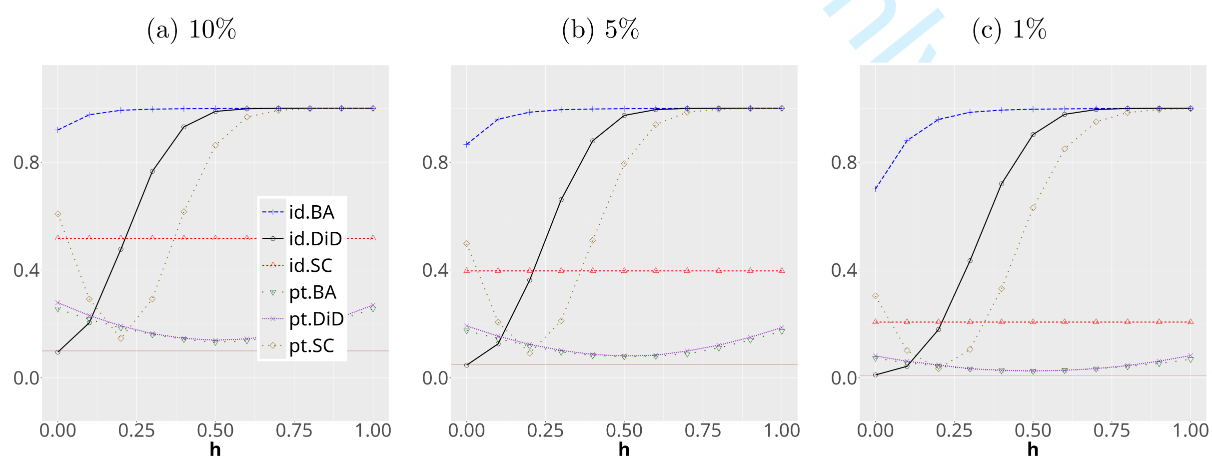


Figure S.20: Power Curves - DGP idTest III, $\mathcal{T}, T = 25$.

pre-tests do not help detect violations of ATT.

S.6.5 Heterogeneous Treatment Effects

This section provides more simulation results. The DGPs remain the same as in Appendix S.6.2, but $ATT(t)$ is heterogeneous in t and $ATT_{\omega,T}$ varies with T:

$$ATT(t) = \begin{cases} \sin(t)/\pi & \text{if } t \geq 1 \\ 0 & \text{otherwise} \end{cases}.$$

With the choice of the uniform weighting scheme, the sequence of ATT parameters (indexed by T) is given by $ATT_{\omega,T} = (\pi T)^{-1} \sum_{t=1}^{T} \sin(t)$. This scenario helps to examine the performance of the DiD when the parameter of interest varies with T.

Generally, Tables S.5 to S.8 confirm simulation results on homogeneous treatment effects. The DiD performs well across all settings considered in the DGPs, whereas the SC and BA are very sensitive to deviations from the strong identification conditions they require. $ATT(t)$ is heterogeneous across post-treatment periods, and the target parameter varies with T and does not impact the performance of the DiD.

S.6.6 Non-constant λ_n

The next set of simulation results concerns the sensitivity to λ_n. To this end, the sequence λ_n is specified as $\lambda_n \in \{0.5, 0.4, 0.6, 0.3, 0.7\}$ varying in $n \in \{50, 100, 200, 400, 800\}$ over the closed interval $\lambda_n \in [0.3, 0.7]$ with $T = \lambda_n n$, $\mathcal{T} = n - T$, and $ATT(t) = 0.0$, $t \geq 1$. The results are presented in Tables S.9 to S.12. Comparing these results to those in Tables S.1 to S.4, one observes little difference in the performance of all estimators. The performance of the T-DiD remains robust across all the DGPs.

Table S.5: Simulation - $ATT(t) = \sin(t)/\pi,\ t \geq 1$

DGP	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP SC-BA		MB			MAD			RMSE		
	25	0.002	0.002	0.001	0.140	0.096	0.175	0.216	0.151	0.280
	50	-0.001	0.000	-0.001	0.101	0.070	0.129	0.153	0.107	0.202
	100	0.000	0.000	-0.001	0.072	0.051	0.094	0.108	0.077	0.142
	200	0.001	0.001	0.002	0.051	0.036	0.066	0.076	0.053	0.100
	400	0.000	0.000	0.000	0.037	0.025	0.047	0.054	0.038	0.070
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.010	0.007	0.011	0.048	0.035	0.050	0.100	0.077	0.103
	50	0.009	0.004	0.009	0.043	0.035	0.049	0.092	0.077	0.102
	100	0.006	0.005	0.008	0.042	0.035	0.048	0.089	0.08	0.096
	200	0.006	0.004	0.008	0.041	0.031	0.044	0.089	0.071	0.091
	400	0.008	0.005	0.008	0.041	0.035	0.043	0.089	0.076	0.090
DGP BA		MB			MAD			RMSE		
	25	0.002	-0.998	0.001	0.140	0.997	0.175	0.216	1.009	0.280
	50	-0.001	-1.000	-0.001	0.101	1.000	0.129	0.153	1.006	0.202
	100	0.000	-1.000	-0.001	0.072	1.000	0.094	0.108	1.003	0.142
	200	0.001	-0.999	0.002	0.051	0.999	0.066	0.076	1.001	0.100
	400	0.000	-1.000	0.000	0.037	1.000	0.047	0.054	1.001	0.070
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.010	0.989	0.011	0.048	0.997	0.050	0.100	0.999	0.103
	50	0.009	1.000	0.009	0.043	1.000	0.049	0.092	1.000	0.102
	100	0.006	1.000	0.008	0.042	1.000	0.048	0.089	1.000	0.096
	200	0.006	1.000	0.008	0.041	1.000	0.044	0.089	1.000	0.091
	400	0.008	1.000	0.008	0.041	1.000	0.043	0.089	1.000	0.090
DGP SC		MB			MAD			RMSE		
	25	0.002	0.002	1.246	0.140	0.096	1.248	0.216	0.151	1.277
	50	-0.001	0.000	1.328	0.101	0.07	1.329	0.153	0.107	1.343
	100	0.000	0.000	1.371	0.072	0.051	1.372	0.108	0.077	1.378
	200	0.001	0.001	1.395	0.051	0.036	1.393	0.076	0.053	1.398
	400	0.000	0.000	1.404	0.037	0.025	1.404	0.054	0.038	1.406
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.010	0.007	0.921	0.048	0.035	0.965	0.100	0.077	0.979
	50	0.009	0.004	0.992	0.043	0.035	0.996	0.092	0.077	0.998
	100	0.006	0.005	0.999	0.042	0.035	1.000	0.089	0.080	1.000
	200	0.006	0.004	1.000	0.041	0.031	1.000	0.089	0.071	1.000
	400	0.008	0.005	1.000	0.041	0.035	1.000	0.089	0.076	1.000

Table S.6: Simulation - $ATT(t) = \sin(t)/\pi, \ t \geq 1$

DGP	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP PT-NA (A)		MB			MAD			RMSE		
	25	0.177	0.089	0.176	0.270	0.179	0.294	0.397	0.265	0.436
	50	0.109	0.055	0.109	0.185	0.127	0.201	0.273	0.187	0.302
	100	0.063	0.03	0.062	0.128	0.087	0.139	0.19	0.129	0.212
	200	0.037	0.018	0.037	0.089	0.061	0.099	0.132	0.090	0.147
	400	0.023	0.011	0.023	0.063	0.043	0.068	0.092	0.064	0.102
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.027	0.025	0.026	0.089	0.075	0.086	0.154	0.132	0.147
	50	0.019	0.017	0.016	0.072	0.067	0.072	0.133	0.12	0.128
	100	0.016	0.014	0.015	0.064	0.052	0.068	0.12	0.099	0.124
	200	0.013	0.010	0.011	0.061	0.048	0.058	0.113	0.100	0.114
	400	0.011	0.010	0.011	0.054	0.047	0.053	0.102	0.096	0.104
DGP PT-NA (B)		MB			MAD			RMSE		
	25	0.001	0.287	0.000	0.241	0.294	0.269	0.356	0.381	0.399
	50	0.001	0.246	0.001	0.168	0.248	0.187	0.25	0.304	0.282
	100	-0.002	0.205	-0.002	0.121	0.206	0.134	0.18	0.241	0.202
	200	-0.001	0.175	-0.001	0.085	0.175	0.095	0.126	0.196	0.142
	400	0.001	0.148	0.001	0.06	0.148	0.067	0.089	0.161	0.099
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.015	0.1	0.014	0.06	0.227	0.058	0.112	0.327	0.115
	50	0.012	0.13	0.01	0.051	0.288	0.05	0.100	0.397	0.105
	100	0.011	0.17	0.009	0.05	0.361	0.055	0.097	0.482	0.105
	200	0.009	0.259	0.01	0.048	0.486	0.045	0.099	0.610	0.096
	400	0.009	0.392	0.008	0.045	0.633	0.045	0.095	0.749	0.098
DGP GARCH(1,1)		MB			MAD			RMSE		
	25	0.004	0.001	-0.003	0.141	0.098	0.177	0.214	0.154	0.275
	50	0.001	0.000	0.002	0.102	0.072	0.13	0.152	0.107	0.199
	100	0.000	0.000	0.000	0.071	0.051	0.092	0.108	0.077	0.141
	200	0.000	0.000	0.000	0.052	0.037	0.066	0.077	0.054	0.099
	400	0.000	0.000	0.000	0.036	0.026	0.047	0.054	0.038	0.071
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.013	0.012	0.010	0.055	0.051	0.05	0.106	0.098	0.102
	50	0.010	0.006	0.009	0.045	0.037	0.048	0.094	0.081	0.097
	100	0.008	0.007	0.008	0.047	0.039	0.046	0.094	0.081	0.091
	200	0.008	0.005	0.007	0.045	0.036	0.041	0.087	0.079	0.088
	400	0.006	0.005	0.008	0.041	0.035	0.042	0.086	0.08	0.092

Table S.7: Simulation - $ATT(t) = \sin(t)/\pi,\ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP MA(1)			MB			MAD			RMSE	
	25	0.005	0.003	-0.002	0.174	0.119	0.218	0.269	0.19	0.347
	50	-0.001	0.000	-0.002	0.128	0.088	0.164	0.192	0.134	0.251
	100	-0.001	0.000	-0.001	0.091	0.065	0.118	0.135	0.096	0.178
	200	0.001	0.001	0.001	0.064	0.045	0.083	0.095	0.067	0.125
	400	0.000	0.000	0.000	0.045	0.032	0.058	0.067	0.048	0.088
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.022	0.018	0.015	0.072	0.061	0.065	0.132	0.113	0.122
	50	0.014	0.009	0.015	0.066	0.051	0.064	0.121	0.102	0.118
	100	0.012	0.010	0.010	0.058	0.052	0.054	0.114	0.103	0.113
	200	0.013	0.010	0.012	0.054	0.051	0.054	0.105	0.101	0.109
	400	0.011	0.011	0.009	0.051	0.049	0.053	0.103	0.100	0.110
DGP AR(1)			MB			MAD			RMSE	
	25	-0.007	-0.029	-0.007	0.160	0.108	0.206	0.248	0.177	0.321
	50	-0.006	-0.016	-0.005	0.110	0.075	0.144	0.166	0.116	0.216
	100	-0.003	-0.009	-0.003	0.076	0.053	0.100	0.112	0.080	0.146
	200	-0.001	-0.004	0.000	0.052	0.036	0.068	0.078	0.055	0.101
	400	-0.001	-0.002	-0.001	0.036	0.026	0.047	0.054	0.038	0.071
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.019	0.021	0.017	0.068	0.065	0.071	0.124	0.115	0.130
	50	0.013	0.013	0.015	0.061	0.049	0.065	0.113	0.096	0.122
	100	0.012	0.010	0.012	0.054	0.046	0.055	0.103	0.093	0.111
	200	0.011	0.009	0.014	0.048	0.043	0.058	0.092	0.087	0.109
	400	0.009	0.007	0.011	0.044	0.038	0.056	0.089	0.078	0.111
DGP U-R			MB			MAD			RMSE	
	25	-0.008	-0.009	-0.013	0.144	0.099	0.178	0.22	0.156	0.286
	50	-0.006	-0.005	-0.007	0.103	0.072	0.131	0.155	0.108	0.203
	100	-0.003	-0.003	-0.004	0.073	0.053	0.094	0.109	0.078	0.143
	200	-0.001	-0.001	-0.001	0.051	0.036	0.066	0.077	0.054	0.100
	400	-0.001	-0.001	-0.001	0.036	0.026	0.047	0.054	0.038	0.071
			Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.011	0.010	0.008	0.049	0.043	0.042	0.097	0.086	0.095
	50	0.008	0.005	0.007	0.045	0.035	0.044	0.096	0.077	0.094
	100	0.007	0.006	0.006	0.043	0.037	0.039	0.091	0.08	0.089
	200	0.009	0.005	0.008	0.042	0.038	0.043	0.086	0.078	0.091
	400	0.007	0.005	0.007	0.039	0.034	0.041	0.085	0.073	0.094

Table S.8: Simulation - $ATT(t) = \sin(t)/\pi,\ t \geq 1$

	$T, \mathcal{T}$	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP Q-T		MB			MAD			RMSE		
	25	0.003	0.002	25.998	0.141	0.096	25.998	0.217	0.153	25.999
	50	-0.001	0.000	50.998	0.103	0.070	50.998	0.154	0.107	50.998
	100	-0.001	0.000	100.998	0.073	0.052	100.998	0.108	0.077	100.999
	200	0.001	0.001	201.001	0.050	0.036	201.000	0.076	0.054	201.001
	400	0.000	0.000	401.000	0.036	0.026	401.001	0.054	0.038	401.000
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.011	0.009	1.000	0.050	0.043	1.000	0.101	0.083	1.000
	50	0.009	0.005	1.000	0.046	0.034	1.000	0.092	0.074	1.000
	100	0.008	0.005	1.000	0.043	0.036	1.000	0.092	0.079	1.000
	200	0.009	0.005	1.000	0.043	0.036	1.000	0.087	0.076	1.000
	400	0.007	0.005	1.000	0.038	0.034	1.000	0.084	0.073	1.000
DGP T-T		MB			MAD			RMSE		
	25	0.075	25.080	0.080	0.271	25.083	0.346	0.432	25.082	0.576
	50	0.032	50.033	0.042	0.202	50.035	0.252	0.308	50.034	0.403
	100	0.017	100.019	0.017	0.140	100.021	0.182	0.215	100.019	0.283
	200	0.012	200.012	0.011	0.101	200.013	0.131	0.151	200.012	0.200
	400	0.002	400.002	0.001	0.073	400.001	0.093	0.108	400.002	0.140
		Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.012	1.000	0.016	0.057	1.000	0.064	0.112	1.000	0.124
	50	0.011	1.000	0.011	0.049	1.000	0.054	0.104	1.000	0.108
	100	0.009	1.000	0.009	0.044	1.000	0.046	0.091	1.000	0.097
	200	0.007	1.000	0.009	0.040	1.000	0.046	0.087	1.000	0.093
	400	0.008	1.000	0.008	0.043	1.000	0.043	0.090	1.000	0.096

Table S.9: Simulation – $ATT(t) = 0.0,\ t \geq 1$, non-constant λ_n

DGP	n	λ_n	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP SC–BA			MB			MAD			RMSE		
	25	0.5	0.002	0.002	0.001	0.140	0.096	0.175	0.216	0.151	0.280
	50	0.4	0.000	0.000	-0.002	0.103	0.079	0.132	0.157	0.121	0.204
	100	0.6	0.001	0.000	-0.002	0.074	0.047	0.095	0.112	0.070	0.147
	200	0.3	-0.001	-0.001	0.000	0.055	0.046	0.071	0.084	0.070	0.109
	400	0.7	0.000	0.000	0.001	0.040	0.022	0.051	0.059	0.032	0.076
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.012	0.012	0.013	0.057	0.056	0.056	0.116	0.110	0.114
	50	0.4	0.011	0.011	0.010	0.056	0.053	0.052	0.111	0.111	0.105
	100	0.6	0.010	0.010	0.011	0.055	0.052	0.055	0.110	0.104	0.110
	200	0.3	0.011	0.010	0.009	0.055	0.053	0.049	0.108	0.104	0.102
	400	0.7	0.011	0.010	0.009	0.048	0.050	0.050	0.094	0.101	0.098
DGP BA			MB			MAD			RMSE		
	25	0.5	0.002	-0.998	0.001	0.140	0.997	0.175	0.216	1.009	0.280
	50	0.4	0.000	-1.000	-0.002	0.103	0.999	0.132	0.157	1.007	0.204
	100	0.6	0.001	-1.000	-0.002	0.074	0.998	0.095	0.112	1.002	0.147
	200	0.3	-0.001	-1.001	0.000	0.055	1.001	0.071	0.084	1.003	0.109
	400	0.7	0.000	-1.000	0.001	0.040	1.000	0.051	0.059	1.001	0.076
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.012	0.990	0.013	0.057	0.997	0.056	0.116	0.999	0.114
	50	0.4	0.011	0.999	0.010	0.056	1.000	0.052	0.111	1.000	0.105
	100	0.6	0.010	1.000	0.011	0.055	1.000	0.055	0.110	1.000	0.110
	200	0.3	0.011	1.000	0.009	0.055	1.000	0.049	0.108	1.000	0.102
	400	0.7	0.011	1.000	0.009	0.048	1.000	0.050	0.094	1.000	0.098
DGP SC			MB			MAD			RMSE		
	25	0.5	0.002	0.002	1.246	0.140	0.096	1.248	0.216	0.151	1.277
	50	0.4	0.000	0.000	1.306	0.103	0.079	1.303	0.157	0.121	1.322
	100	0.6	0.001	0.000	1.377	0.074	0.047	1.380	0.112	0.070	1.385
	200	0.3	-0.001	-0.001	1.379	0.055	0.046	1.377	0.084	0.070	1.383
	400	0.7	0.000	0.000	1.407	0.040	0.022	1.408	0.059	0.032	1.409
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.012	0.012	0.919	0.057	0.056	0.965	0.116	0.110	0.979
	50	0.4	0.011	0.011	0.992	0.056	0.053	0.998	0.111	0.111	0.999
	100	0.6	0.010	0.010	0.999	0.055	0.052	1.000	0.110	0.104	1.000
	200	0.3	0.011	0.010	1.000	0.055	0.053	1.000	0.108	0.104	1.000
	400	0.7	0.011	0.010	1.000	0.048	0.050	1.000	0.094	0.101	1.000

Table S.10: Simulation – $ATT(t) = 0.0,\ t \geq 1$, non-constant λ_n

DGP	n	λ_n	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
			MB			MAD			RMSE		
DGP PT-NA (A)	25	0.5	0.177	0.089	0.176	0.270	0.179	0.294	0.397	0.265	0.436
	50	0.4	0.102	0.064	0.108	0.178	0.145	0.202	0.261	0.211	0.309
	100	0.6	0.075	0.027	0.063	0.150	0.080	0.145	0.225	0.118	0.217
	200	0.3	0.035	0.026	0.042	0.089	0.080	0.107	0.130	0.117	0.159
	400	0.7	0.034	0.008	0.026	0.095	0.036	0.076	0.140	0.054	0.112
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.030	0.029	0.027	0.096	0.086	0.092	0.162	0.142	0.155
	50	0.4	0.024	0.023	0.022	0.080	0.079	0.077	0.141	0.134	0.136
	100	0.6	0.020	0.015	0.017	0.072	0.060	0.067	0.132	0.111	0.127
	200	0.3	0.014	0.013	0.013	0.064	0.059	0.061	0.120	0.110	0.117
	400	0.7	0.012	0.013	0.012	0.060	0.052	0.060	0.113	0.103	0.112
			MB			MAD			RMSE		
DGP PT-NA (B)	25	0.5	0.001	0.287	0.000	0.241	0.294	0.269	0.356	0.381	0.399
	50	0.4	0.023	0.258	0.021	0.175	0.264	0.191	0.259	0.327	0.290
	100	0.6	-0.020	0.198	-0.023	0.124	0.198	0.138	0.187	0.229	0.209
	200	0.3	0.035	0.197	0.036	0.097	0.197	0.106	0.142	0.227	0.158
	400	0.7	-0.031	0.136	-0.031	0.069	0.136	0.076	0.102	0.146	0.114
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.016	0.111	0.016	0.065	0.246	0.061	0.118	0.347	0.119
	50	0.4	0.015	0.133	0.016	0.059	0.284	0.059	0.107	0.394	0.114
	100	0.6	0.016	0.210	0.013	0.060	0.416	0.057	0.113	0.539	0.110
	200	0.3	0.012	0.207	0.012	0.060	0.414	0.058	0.117	0.534	0.113
	400	0.7	0.016	0.494	0.013	0.064	0.728	0.059	0.122	0.817	0.114
			MB			MAD			RMSE		
DGP GARCH(1,1)	25	0.5	0.004	0.001	-0.003	0.141	0.098	0.177	0.214	0.154	0.275
	50	0.4	0.002	0.001	0.001	0.103	0.079	0.131	0.156	0.120	0.203
	100	0.6	0.000	0.000	0.000	0.073	0.046	0.092	0.110	0.070	0.142
	200	0.3	-0.001	-0.001	0.000	0.056	0.047	0.071	0.083	0.070	0.109
	400	0.7	0.000	0.000	0.000	0.039	0.022	0.051	0.059	0.032	0.077
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.017	0.019	0.014	0.064	0.072	0.058	0.120	0.123	0.111
	50	0.4	0.014	0.014	0.013	0.057	0.059	0.056	0.108	0.110	0.105
	100	0.6	0.011	0.010	0.008	0.054	0.053	0.049	0.102	0.105	0.098
	200	0.3	0.011	0.010	0.011	0.051	0.054	0.051	0.103	0.101	0.099
	400	0.7	0.011	0.009	0.011	0.052	0.053	0.050	0.101	0.101	0.102

Table S.11: Simulation – $ATT(t) = 0.0,\ t \geq 1$, non-constant λ_n

DGP	n	λ_n	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP MA(1)				MB			MAD			RMSE	
	25	0.5	0.005	0.003	-0.002	0.174	0.119	0.218	0.269	0.190	0.347
	50	0.4	0.000	0.001	-0.003	0.121	0.097	0.164	0.183	0.150	0.255
	100	0.6	0.001	0.000	-0.002	0.105	0.059	0.118	0.158	0.088	0.182
	200	0.3	-0.001	-0.001	0.000	0.063	0.058	0.089	0.095	0.087	0.136
	400	0.7	0.000	0.000	0.001	0.068	0.028	0.063	0.101	0.040	0.096
				Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.5	0.024	0.022	0.017	0.078	0.074	0.068	0.142	0.131	0.130
	50	0.4	0.017	0.014	0.016	0.070	0.062	0.066	0.130	0.123	0.127
	100	0.6	0.016	0.012	0.014	0.066	0.060	0.063	0.123	0.115	0.121
	200	0.3	0.015	0.012	0.013	0.059	0.060	0.058	0.114	0.113	0.112
	400	0.7	0.013	0.012	0.014	0.059	0.055	0.059	0.107	0.104	0.111
DGP AR(1)				MB			MAD			RMSE	
	25	0.5	0.003	-0.021	0.003	0.162	0.112	0.208	0.252	0.183	0.324
	50	0.4	0.000	-0.014	-0.001	0.115	0.086	0.148	0.172	0.137	0.222
	100	0.6	0.001	-0.005	-0.001	0.077	0.049	0.101	0.116	0.074	0.151
	200	0.3	-0.001	-0.006	0.001	0.058	0.048	0.074	0.086	0.073	0.111
	400	0.7	0.000	-0.001	0.001	0.040	0.022	0.051	0.059	0.033	0.077
				Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.5	0.024	0.033	0.021	0.083	0.088	0.082	0.141	0.148	0.143
	50	0.4	0.018	0.024	0.019	0.071	0.079	0.073	0.131	0.135	0.132
	100	0.6	0.015	0.015	0.016	0.062	0.065	0.066	0.120	0.118	0.124
	200	0.3	0.014	0.015	0.016	0.056	0.063	0.064	0.107	0.115	0.117
	400	0.7	0.011	0.011	0.016	0.053	0.056	0.061	0.102	0.103	0.112
DGP U–R				MB			MAD			RMSE	
	25	0.5	0.003	0.002	-0.002	0.143	0.097	0.178	0.220	0.156	0.286
	50	0.4	0.000	0.001	-0.003	0.106	0.079	0.132	0.160	0.122	0.207
	100	0.6	0.001	0.000	-0.002	0.074	0.048	0.094	0.112	0.071	0.147
	200	0.3	-0.001	-0.001	0.000	0.056	0.047	0.072	0.084	0.070	0.109
	400	0.7	0.000	0.000	0.001	0.039	0.022	0.050	0.059	0.032	0.077
				Rej. 1%			Rej. 5%			Rej. 10%	
	25	0.5	0.013	0.015	0.010	0.056	0.056	0.047	0.110	0.105	0.101
	50	0.4	0.011	0.010	0.011	0.055	0.051	0.052	0.111	0.103	0.102
	100	0.6	0.012	0.009	0.007	0.052	0.052	0.052	0.103	0.103	0.103
	200	0.3	0.012	0.009	0.009	0.052	0.051	0.048	0.100	0.099	0.102
	400	0.7	0.010	0.010	0.011	0.050	0.051	0.051	0.098	0.097	0.097

Table S.12: Simulation – $ATT(t) = 0.0,\ t \geq 1$, non-constant λ_n

	n	λ_n	DiD	SC	BA	DiD	SC	BA	DiD	SC	BA
DGP Q–T			MB			MAD			RMSE		
	25	0.5	0.003	0.002	25.998	0.141	0.096	25.998	0.217	0.153	25.999
	50	0.4	0.000	0.000	49.645	0.106	0.078	49.643	0.159	0.120	49.645
	100	0.6	0.000	0.000	106.371	0.073	0.047	106.373	0.111	0.070	106.371
	200	0.3	-0.001	-0.001	158.174	0.056	0.047	158.169	0.084	0.070	158.174
	400	0.7	0.000	0.000	571.987	0.039	0.022	571.990	0.059	0.032	571.987
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.013	0.015	1.000	0.060	0.058	1.000	0.113	0.105	1.000
	50	0.4	0.010	0.009	1.000	0.053	0.049	1.000	0.109	0.099	1.000
	100	0.6	0.011	0.010	1.000	0.054	0.051	1.000	0.102	0.104	1.000
	200	0.3	0.012	0.009	1.000	0.052	0.052	1.000	0.101	0.100	1.000
	400	0.7	0.009	0.010	1.000	0.048	0.049	1.000	0.098	0.098	1.000
DGP T–T			MB			MAD			RMSE		
	25	0.5	0.003	25.005	0.008	0.270	25.008	0.344	0.426	25.007	0.571
	50	0.4	-0.002	59.995	0.000	0.195	59.998	0.242	0.296	59.995	0.385
	100	0.6	0.004	80.002	-0.005	0.139	80.002	0.176	0.212	80.002	0.275
	200	0.3	-0.004	280.000	-0.001	0.091	280.001	0.119	0.138	280.000	0.179
	400	0.7	0.000	240.000	0.000	0.065	240.000	0.085	0.098	240.000	0.127
			Rej. 1%			Rej. 5%			Rej. 10%		
	25	0.5	0.015	1.000	0.016	0.066	1.000	0.068	0.124	1.000	0.132
	50	0.4	0.012	1.000	0.014	0.060	1.000	0.064	0.117	1.000	0.121
	100	0.6	0.012	1.000	0.010	0.057	1.000	0.056	0.114	1.000	0.108
	200	0.3	0.012	1.000	0.010	0.053	1.000	0.055	0.107	1.000	0.104
	400	0.7	0.011	1.000	0.010	0.053	1.000	0.050	0.108	1.000	0.099

References

[1] Abadie, Alberto. "Using synthetic controls: Feasibility, data requirements, and methodological aspects". *Journal of Economic Literature* 59.2 (2021), pp. 391–425.

[2] Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program". *Journal of the American Statistical Association* 105.490 (2010), pp. 493–505.

[3] Abadie, Alberto and Javier Gardeazabal. "The economic costs of conflict: A case study of the Basque Country". *American Economic Review* 93.1 (2003), pp. 113–132.

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

[5] Angrist, Joshua D, Òscar Jordà, and Guido M Kuersteiner. "Semiparametric estimates of monetary policy effects: String theory revisited". *Journal of Business & Economic Statistics* 36.3 (2018), pp. 371–387.

[6] Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. "Synthetic difference-in-differences". *American Economic Review* 111.12 (2021), pp. 4088–4118.

[7] Baillie, Richard T, Francis X Diebold, George Kapetanios, Kun Ho Kim, and Aaron Mora. "On robust inference in time-series regression". *Econometrics Journal* 28.2 (2025), pp. 131–173.

[8] Bellégo, Christophe, David Benatia, and Vincent Dortet-Bernadet. "The chained difference-in-differences". *Journal of Econometrics* (2024), p. 105783.

[9] Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. "The augmented synthetic control method". *Journal of the American Statistical Association* 116.536 (2021), pp. 1789–1803.

[10] Bernstein, Dennis S. *Matrix mathematics: theory, facts, and formulas.* Princeton university press, 2009.

[11] Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. "Revisiting event-study designs: robust and efficient estimation". *Review of Economic Studies* 91.6 (2024), pp. 3253–3285.

[12] Botosaru, Irene, Raffaella Giacomini, and Martin Weidner. "Forecasted treatment effects". *arXiv preprint arXiv:2309.05639* (2023).

[13] Callaway, Brantly and Pedro HC Sant'Anna. "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2 (2021), pp. 200–230.

[14] Carvalho, Carlos, Ricardo Masini, and Marcelo C Medeiros. "ArCo: An artificial counterfactual approach for high-dimensional panel time-series data". *Journal of Econometrics* 207.2 (2018), pp. 352–380.

[15] Chan, Marc K and Simon S Kwok. "The PCDID approach: Difference-in-differences when trends are potentially unparallel and stochastic". *Journal of Business & Economic Statistics* (2021), pp. 1–18.

[16] Chen, Chaoyi and Thanasis Stengos. "Threshold nonlinearities and the democracy-growth nexus". *The Econometrics Journal* (2024), utae006.

[17] Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. "A t-test for synthetic controls". *arXiv preprint arXiv:1812.10820* (2024).

[18] Colagrossi, Marco, Domenico Rossignoli, and Mario A Maggioni. "Does democracy cause growth? A meta-analysis (of 2000 regressions)". *European Journal of Political Economy* 61 (2020), p. 101824.

[19] Coppedge, Michael, Amanda B Edgell, Carl Henrik Knutsen, and Staffan I Lindberg. *Why Democracies Develop and Decline*. Cambridge University Press, 2022.

[20] Davidson, James. *Stochastic Limit Theory: An Introduction for Econometricians*. Second Edition. Oxford University Press, 2021.

[21] de Chaisemartin, Clement and Xavier D'Haultfœuille. "Two-way fixed effects estimators with heterogeneous treatment effects". *American Economic Review* 110.9 (2020), pp. 2964–2996.

[22] Dette, Holger and Martin Schumann. "Testing for equivalence of pre-trends in Difference-in-Differences estimation". *Journal of Business & Economic Statistics* (2024), pp. 1–13.

[23] Doucouliagos, Hristos and Mehmet Ali Ulubaşoğlu. "Democracy and economic growth: A meta-analysis". *American Journal of Political Science* 52.1 (2008), pp. 61–83.

[24] Ferman, Bruno and Cristine Pinto. "Inference in differences-in-differences with few treated groups and heteroskedasticity". *Review of Economics and Statistics* 101.3 (2019), pp. 452–467.

[25] Ferman, Bruno and Cristine Pinto. "Synthetic controls with imperfect pretreatment fit". *Quantitative Economics* 12.4 (2021), pp. 1197–1221.

[26] Foa, Roberto Stefan, Andrew Klassen, Micheal Slade, Alex Rand, and Rosie Collins. "The global satisfaction with democracy report 2020" (2020).

[27] Freyaldenhoven, Simon, Christian Hansen, and Jesse M Shapiro. "Pre-event trends in the panel event-study design". *American Economic Review* 109.9 (2019), pp. 3307–38.

[28] Fry, Joseph. "A method of moments approach to asymptotically unbiased Synthetic Controls". *Journal of Econometrics* 244.1 (2024), p. 105846.

[29] Galindo-Silva, Hector, Nibene Habib Some, and Guy Tchuente. "Fuzzy difference-in-discontinuities: Identification theory and application to the affordable care act". *arXiv preprint arXiv:1812.06537* (2018).

[30] Gerring, John, Philip Bond, William T Barndt, and Carola Moreno. "Democracy and economic growth: A historical perspective". *World Politics* 57.3 (2005), pp. 323–364.

[31] Giavazzi, Francesco and Guido Tabellini. "Economic and political liberalizations". *Journal of Monetary Economics* 52.7 (2005), pp. 1297–1330.

[32] Gonçalves, Sılvia and Serena Ng. "Imputation of counterfactual outcomes when the errors are predictable". *Journal of Business & Economic Statistics* 42.4 (2024), pp. 1107–1122.

[33] Hamilton, James Douglas. *Time Series Analysis*. Princeton University Press, 1994.

[34] Hansen, Bruce E. *Econometrics*. University of Wisconsin, Dept. of Economics, 2021.

[35] Kahn-Lang, Ariella and Kevin Lang. "The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications". *Journal of Business & Economic Statistics* 38.3 (2020), pp. 613–620.

[36] Knutsen, Carl Henrik. "A business case for democracy: regime type, growth, and growth volatility". *Democratization* 28.8 (2021), pp. 1505–1524.

[37] Kohnert, Dirk. "BTI 2021 - Togo country report: Togo's political and socio-economic development (2019-2021) [author's enhanced version]". *BTI project: Shaping Change - Strategies of Development and Transformation ; Political Economy of Africa*. Gütersloh: Bertelsmann Stiftung, 2021, pp. 1–62.

[38] Lei, Lihua and Timothy Sudijono. "Inference for synthetic controls via refined placebo tests". *arXiv preprint arXiv:2401.07152* (2024).

[39] Li, Kathleen T. "Statistical inference for average treatment effects estimated by synthetic control methods". *Journal of the American Statistical Association* 115.532 (2020), pp. 2068–2083.

[40] Murtin, Fabrice and Romain Wacziarg. "The democratic transition". *Journal of Economic Growth* 19.2 (2014), pp. 141–181.

[41] Newey, WK and KD West. "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix". *Econometrica* 55.3 (1987), pp. 703–708.

[42] Papaioannou, Elias and Gregorios Siourounis. "Democratisation and growth". *The Economic Journal* 118.532 (2008), pp. 1520–1551.

[43] Persson, Torsten. *Forms of democracy, policy and economic development.* 2005.

[44] Persson, Torsten and Guido Tabellini. "Democracy and development: The devil in the details". *American Economic Review* 96.2 (2006), pp. 319–324.

[45] Persson, Torsten and Guido Tabellini. "Democratic capital: The nexus of political and economic change". *American Economic Journal: Macroeconomics* 1.2 (2009), pp. 88–126.

[46] Pesaran, M Hashem and Ron P Smith. "Counterfactual analysis in macroeconometrics: An empirical investigation into the effects of quantitative easing". *Research in Economics* 70.2 (2016), pp. 262–280.

[47] Picchetti, Pedro, Cristine CX Pinto, and Stephanie T Shinoki. "Difference-in-discontinuities: estimation, inference and validity tests". *arXiv preprint arXiv:2405.18531* (2024).

[48] Roth, Jonathan. "Pretest with caution: Event-study estimates after testing for parallel trends". *American Economic Review: Insights* 4.3 (2022), pp. 305–22.

[49] Sun, Liyang, Eli Ben-Michael, and Avi Feller. "Using multiple outcomes to improve the synthetic control method". *arXiv preprint arXiv:2311.16260* (2023).

[50] Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data.* MIT press, 2010.

[51] Xu, Yiqing. "Generalized synthetic control method: Causal inference with interactive fixed effects models". *Political Analysis* 25.1 (2017), pp. 57–76.