Prediction Intervals for Synthetic Control Methods*
Supplemental Appendix

Matias D. Cattaneo[†] Yingjie Feng[‡] Rocio Titiunik[§]

August 8, 2021

Abstract

This supplement contains theoretical proofs of the results discussed in the main paper and additional numerical evidence. Section SA-1 provides further details of the three examples discussed in the paper. Section SA-2 presents all the proofs. Section SA-3 gives additional simulation evidence for the finite-sample performance of the proposed prediction intervals. Section SA-4 provides another empirical application studying the economic impact of 1990 German reunification on West Germany.

*Cattaneo and Titiunik gratefully acknowledges financial support from the National Science Foundation (SES-2019432).

[†]Department of Operations Research and Financial Engineering, Princeton University.

[‡]School of Economics and Management, Tsinghua University.

[§]Department of Politics, Princeton University.

Contents

SA-1 Details on Examples

This section provides more details of the three examples discussed in the main paper. Recall that we consider the standard synthetic control constraint: $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}_+^N : \|\mathbf{w}\|_1 = 1\}$ and $\mathcal{R} = \mathbb{R}^{KM}$. For simulation-based inference, we define explicitly a relaxed constraint set based on the original estimated coefficients $\widehat{\boldsymbol{\beta}}$: $\Delta^\star = \{\mathbf{D}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^\star) : \boldsymbol{\beta} = (\mathbf{w}', \mathbf{r}')', \mathbf{w} \in \mathbb{R}_+^N, \ \|\mathbf{w}\|_1 = \|\widehat{\mathbf{w}}^\star\|_1\}$, where $\widehat{\boldsymbol{\beta}}^\star = (\widehat{\mathbf{w}}^{\star\prime}, \widehat{\mathbf{r}}')'$, $\widehat{\mathbf{w}}^\star = (\widehat{\omega}_2^\star, \cdots, \widehat{\omega}_{N+1}^\star)'$, $\widehat{\omega}_j^\star = \widehat{\omega}_j \mathbb{1}(|\widehat{\omega}_j| > \varrho)$, and ϱ is a tuning parameter that ensures the constraint set in the simulation world preserves the local geometry of Δ. Moreover, we set $\mathbf{x}_T = (Y_{2T}(0), \cdots, Y_{(N+1)T}(0))'$ as it is common in the SC literature. Finally, we let $\mathfrak{C}$, $\mathfrak{C}^\star$ and $\mathfrak{c}$, with various sub-indexes, denote non-negative finite constants not depending on T_0. In simple cases, we give the exact expression of these constants, while in other cases they can be characterized from the proofs of the results.

SA-1.1 Outcomes-only

In this example the SC weights are constructed based on past outcomes only, and the model allows for an intercept. Thus, the working model simplifies to

$$a_t = \mathbf{b}_t'\mathbf{w}_0 + r_0 + u_t, \qquad t = 1, \cdots, T_0,$$

where $a_t := Y_{1t}(0)$, $\mathbf{b}_t := (Y_{2t}(0), Y_{3t}(0), \ldots, Y_{(N+1)t}(0))'$, and with $M = 1$, $K = 1$, and $d = N + 1$. Let $\mathbf{z}_t = (\mathbf{b}_t', 1)'$, $\boldsymbol{\beta}_0 = (\mathbf{w}_0', r_0)'$. We further assume independent sampling across time, and thus set $\mathbf{D} = T_0^{1/2}\mathbf{I}_d$. A natural variance estimator in this setting is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T_0}\sum_{t=1}^{T_0} \mathbf{z}_t\mathbf{z}_t'(\widehat{u}_t - \widehat{\mathbb{E}}[u_t|\mathbf{b}_t])^2,$$

where $\widehat{u}_t = a_t - \mathbf{z}_t'\widehat{\boldsymbol{\beta}}$, and $\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]$ denotes some estimate of the conditional mean of the pseudo-residuals.

The following theorem gives precise primitive conditions and verifies the high-level conditions of Theorems 1 and 2 in the main paper. Recall that $\mathbf{w}_0 = (w_{0,1}, w_{0,2}, \ldots, w_{0,J})'$ is defined in Section 2 of the paper, and $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ are the minimum and the maximum eigenvalues of a generic square matrix $\mathbf{M}$.

Theorem SA-1 (Outcomes-only). *Let* $\{\mathbf{z}_t, u_t\}_{t=1}^T$ *be i.i.d over* $t = 1, \cdots, T_0$.

Assume that, for finite non-negative constants $\bar{\eta}_1$, $\bar{\eta}_2$, $\underline{\eta}_1$ *and* $\underline{\eta}_2$, *the following conditions hold:*

(SA-1.i) $\max_{1\leq t\leq T_0} \mathbb{E}[|u_t|^3|\mathbf{B}] \leq \bar{\eta}_1$ *a.s. on* $\sigma(\mathbf{B})$ *and* $\mathbb{E}[\|\mathbf{z}_t\|^6] \leq \bar{\eta}_2$.

(SA-1.ii) $\min_{1\leq t\leq T_0} \mathbb{V}[u_t|\mathbf{B}] \geq \underline{\eta}_1$ *a.s. on* $\sigma(\mathbf{B})$ *and* $\lambda_{\min}(\mathbb{E}[\mathbf{z}_t\mathbf{z}_t']) \geq \underline{\eta}_2$.

Then, the conditions of Theorem 1 hold with $\pi_\gamma = \mathfrak{C}_\pi T_0^{-1}$ *and* $\epsilon_\gamma = \mathfrak{C}_\epsilon T_0^{-1/2}$, *where* $\mathfrak{C}_\pi = \frac{d}{\bar{\eta}_2} + \frac{4d^4\bar{\eta}_2}{\underline{\eta}_2^2}$ *and* $\mathfrak{C}_\epsilon = 42(d^{1/4} + 16)\frac{2^{5/2}d^{3/2}\bar{\eta}_1\bar{\eta}_2}{(\underline{\eta}_1\underline{\eta}_2)^{3/2}}$. *Therefore, condition* (T2.i) *of Theorem 2 also holds, while condition* (T2.ii) *holds with* $\pi_\delta^\star = \mathfrak{C}_\pi^\star T_0^{-1}$, $\epsilon_\delta^\star = \mathfrak{C}_\epsilon^\star T_0^{-1}$, *and* $\varpi_\delta^\star = \mathfrak{C}_\varpi^\star\sqrt{\log T_0}$, *where* $\mathfrak{C}_\pi^\star = \mathfrak{C}_\pi$, $\mathfrak{C}_\epsilon^\star = 2d$, *and* $\mathfrak{C}_\varpi^\star = 8\sqrt{d\bar{\eta}_1\bar{\eta}_2}/\underline{\eta}_2$.

Assume in addition that, for finite non-negative constant $\pi_w^\star$, *the following condition holds:*

(SA-1.iii) $\varrho = \varpi_\delta^\star/\sqrt{T_0}$ *and* $\mathbb{P}(\min\{|w_{0,j}| : w_{0,j} \neq 0\} \geq \varrho) \geq 1 - \pi_w^\star$.

Then, condition (T2.iii) *of Theorem 2 holds with* $\pi_\Delta^\star = \pi_\delta^\star + \pi_w^\star + \pi_\gamma$ *and* $\epsilon_\Delta^\star = \epsilon_\delta^\star + \epsilon_\gamma$.

Finally, also assume that, for finite non-negative constants $\varpi_u^\star$, $\epsilon_u^\star$ *and* $\pi_u^\star$, *the following conditions hold:*

(SA-1.iv) $\max_{1\leq t\leq T_0} \mathbb{E}[|u_t|^4|\mathbf{B}] \leq \bar{\eta}_1$ *a.s. on* $\sigma(\mathbf{B})$ *and* $\mathbb{E}[\|\mathbf{z}_t\|^{12}] \leq \bar{\eta}_2$.

(SA-1.v) $\mathbb{P}[\mathbb{P}(\max_{1\leq t\leq T_0} |\widehat{\mathbb{E}}[u_t|\mathbf{b}_t] - \mathbb{E}[u_t|\mathbf{b}_t]| \leq \varpi_u^\star|\mathscr{H}) \geq 1 - \epsilon_u^\star] \geq 1 - \pi_u^\star$.

Then, condition (T2.iv) *of Theorem 2 holds with* $\pi_\gamma^\star = \mathfrak{C}_{\pi,1}^\star T_0^{-1} + \pi_\delta^\star + \pi_u^\star + \pi_\gamma$, $\epsilon_{\gamma,1}^\star = \mathfrak{C}_{\epsilon,1}^\star(T_0^{\upsilon-1/2} + \varpi_\delta^\star T_0^{-1/2} + \varpi_u^\star)$, $\epsilon_{\gamma,2}^\star = \mathfrak{C}_{\epsilon,2}^\star T_0^{-2\upsilon} + \mathfrak{C}_{\epsilon,3}^\star T_0^{-1} + \epsilon_\delta^\star + \epsilon_u^\star + \epsilon_\gamma$ *for any* $\upsilon \in (0, 1/2)$, *and non-negative constants* $\mathfrak{C}_{\pi,1}^\star, \mathfrak{C}_{\epsilon,1}^\star, \mathfrak{C}_{\epsilon,2}^\star$ *and* $\mathfrak{C}_{\epsilon,3}^\star$, *which are characterized in the proof.*

SA-1.2 Multi-equation with Weakly Dependent Data

In our second example, we incorporate pre-intervention covariates in the construction of the SC weights and allow for stationary weakly dependent time series data. We let $M = 2$ (two features) and $K = 0$ (no additional controls) only for simplicity, which gives the working model

$$a_{t,1} = \sum_{j=1}^{J} b_{jt,1}w_{0,j} + u_{t,1},$$
$$a_{t,2} = \sum_{j=1}^{J} b_{jt,2}w_{0,j} + u_{t,2},$$

$t = 1, \cdots, T_0$. The first equation could naturally correspond to pre-intervention outcomes as in the previous example, i.e., $a_{t,1} := Y_{1t}(0)$ and $\mathbf{b}_{t,1} := (Y_{2t}(0), Y_{3t}(0), \ldots, Y_{(N+1)t}(0))'$, while the second equation could correspond to some other covariate also used to construct $\widehat{\mathbf{w}}$ as described in Section 2 of the paper. Let $\mathbf{b}_{t,l} = (b_{1t,l}, \cdots, b_{Jt,l})'$, for $l = 1, 2$. To provide interpretable primitive conditions, we also assume $\mathbf{u}_t = (u_{t,1}, u_{t,2})'$ and $\mathbf{b}_t = (\mathbf{b}'_{t,1}, \mathbf{b}'_{t,2})'$ follow independent first-order stationary autoregressive (AR) processes:

$$\mathbf{u}_t = \mathbf{H}_u \mathbf{u}_{t-1} + \boldsymbol{\zeta}_{t,u}, \qquad \mathbf{H}_u = \mathrm{diag}(\rho_{1,u}, \rho_{2,u}),$$
$$\mathbf{b}_t = \mathbf{H}_b \mathbf{b}_{t-1} + \boldsymbol{\zeta}_{t,b}, \qquad \mathbf{H}_b = \mathrm{diag}(\rho_{1,b}, \rho_{2,b}, \cdots, \rho_{J,b}),$$

where $\boldsymbol{\zeta}_{t,u}$ and $\boldsymbol{\zeta}_{t,b}$ are i.i.d. over t, independent of each other, and $\mathrm{diag}(\cdot)$ denotes a diagonal matrix with the function arguments as the corresponding diagonal elements. Let $\mathbf{D} = T_0^{1/2}\mathbf{I}_d$ and recall that $\mathbf{U} = (u_{1,1}, \cdots, u_{T_0,1}, u_{1,2}, \cdots, u_{T_0,2})'$ in this case. A natural, generic variance estimator is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T_0}\mathbf{Z}'\widehat{\mathbb{V}}[\mathbf{U}|\mathscr{H}]\mathbf{Z},$$

where $\widehat{\mathbb{V}}[\mathbf{U}|\mathscr{H}]$ is an estimate of $\mathbb{V}[\mathbf{U}|\mathscr{H}]$. In this example, $\boldsymbol{\Sigma}$ corresponds to the (conditional) long-run variance, and naturally $\widehat{\boldsymbol{\Sigma}}$ can be chosen to be any standard estimator thereof.

Because of the time dependence in this example, the following theorem gives primitive conditions that verify the high-level conditions of Theorem A in the appendix (instead of Theorem 1) in the paper, as well as the high-level conditions of Theorem 2 for implementation.

Theorem SA-2 (Multi-equation with Weakly Dependent Data). *Let $\{\boldsymbol{\zeta}_{t,u}\}_{t=1}^T$ and $\{\boldsymbol{\zeta}_{t,b}\}_{t=1}^T$ be i.i.d over $t = 1, \cdots, T_0$ with mean zero, finite variance, and independent of each other.*

Assume that, for finite positive constants φ and φ', and finite non-negative constants $\bar{\eta}_0, \bar{\eta}_1, \bar{\eta}_2$, $\underline{\eta}_1$ and $\underline{\eta}_2$, the following conditions hold:

(SA-2.i) $\|\mathbf{H}_b\| < 1$, $\|\mathbf{H}_u\| < 1$, *and $\boldsymbol{\zeta}_{t,u}$ and $\boldsymbol{\zeta}_{t,b}$ have densities $f_u(\cdot)$ and $f_b(\cdot)$ satisfying* $\int \|\mathbf{x}\|^{\varphi} f_u(\mathbf{x})d\mathbf{x} < \infty$, $\int \|\mathbf{x}\|^{\varphi} f_b(\mathbf{x})d\mathbf{x} < \infty$, $\int |f_u(\mathbf{x}) - f_u(\mathbf{x}-\boldsymbol{\theta})|d\mathbf{x} \leq \bar{\eta}_0\|\boldsymbol{\theta}\|^{\varphi'}$ *for all* $\boldsymbol{\theta} \in \mathbb{R}^2$, *and* $\int |f_b(\mathbf{x}) - f_b(\mathbf{x}-\boldsymbol{\theta})|d\mathbf{x} \leq \bar{\eta}_0\|\boldsymbol{\theta}\|^{\varphi'}$ *for all* $\boldsymbol{\theta} \in \mathbb{R}^J$.

(SA-2.ii) $\max_{1\leq t\leq T_0} \mathbb{E}[\|\mathbf{u}_t\|^4|\mathbf{B}] \leq \bar{\eta}_1$ *a.s. on* $\sigma(\mathbf{B})$ *and* $\mathbb{E}[\|\mathbf{b}_t\|^6] \leq \bar{\eta}_2$.

(SA-2.iii) $\min_{1\leq t\leq T_0} \lambda_{\min}(\mathbb{V}[\mathbf{U}|\mathbf{B}]) \geq \underline{\eta}_1$ *a.s. on* $\sigma(\mathbf{B})$ *and* $\min_{l=1,2} \lambda_{\min}(\mathbb{E}[\mathbf{b}_{t,l}\mathbf{b}'_{t,l}]) \geq \underline{\eta}_2$.

Then, the conditions of Theorem A hold with $\pi_\gamma = \mathfrak{C}_\pi T_0^{-\mathfrak{c}_\pi}$ *and* $\epsilon_\gamma = \mathfrak{C}_\epsilon T_0^{-\mathfrak{c}_\epsilon}$ *for non-negative constants* $\mathfrak{C}_\pi$ *and* $\mathfrak{C}_\epsilon$*, and some positive constants* $\mathfrak{c}_\pi$ *and* $\mathfrak{c}_\epsilon$*, which are characterized in the proof. Therefore, condition* (T2.i) *of Theorem 2 also holds, while condition* (T2.ii) *holds with* $\pi_\delta^\star = \mathfrak{C}_\pi^\star T_0^{-\mathfrak{c}_\pi^\star}$, $\epsilon_\delta^\star = \mathfrak{C}_\epsilon^\star T_0^{-1}$, $\varpi_\delta^\star = \mathfrak{C}_\varpi^\star \sqrt{\log T_0}$ *for non-negative constants* $\mathfrak{C}_\pi^\star$, $\mathfrak{C}_\epsilon^\star$ *and* $\mathfrak{C}_\varpi^\star$*, and positive constant* $\mathfrak{c}_\pi^\star$*, which are also characterized in the proof.*

Assume in addition that, for finite non-negative constant $\pi_w^\star$*, the following condition holds:*

(SA-2.iv) $\varrho = \varpi_\delta^\star/\sqrt{T_0}$ *and* $\mathbb{P}(\min\{|w_{0,j}| : w_{0,j} \neq 0\} \geq \varrho) \geq 1 - \pi_w^\star$.

Then, condition (T2.iii) *of Theorem 2 holds with* $\pi_\Delta^\star = \pi_\delta^\star + \pi_w^\star + \pi_\gamma$ *and* $\epsilon_\Delta^\star = \epsilon_\delta^\star + \epsilon_\gamma$.

Finally, also assume that, for finite non-negative constants $\epsilon_{\Sigma,1}^\star$, $\epsilon_{\Sigma,2}^\star$ *and* $\pi_\Sigma^\star$*, the following condition holds:*

(SA-2.v) $\mathbb{P}(\mathbb{P}(\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \leq \epsilon_{\Sigma,1}^\star | \mathscr{H}) \geq 1 - \epsilon_{\Sigma,2}^\star) \geq 1 - \pi_\Sigma^\star$.

Then, condition (T2.iv) *of Theorem 2 holds with* $\epsilon_{\gamma,1}^\star = \sqrt{d}\epsilon_{\Sigma,1}^\star/(2\underline{\eta}_1\underline{\eta}_2)$, $\epsilon_{\gamma,2}^\star = \epsilon_{\Sigma,2}^\star$ *and* $\pi_\gamma^\star = \pi_\Sigma^\star + \pi_\delta^\star$.

SA-1.3 Cointegration

Our third and final example illustrates how non-stationary data can also be handled within our framework. Suppose that for each $1 \leq l \leq M$, $\{a_{t,l}\}_{t=1}^T$, $\{b_{1t,l}\}_{t=1}^T, \cdots, \{b_{Jt,l}\}_{t=1}^T$ are $I(1)$ processes, and $\{c_{1t,l}\}_{t=1}^T, \cdots, \{c_{Kt,l}\}_{t=1}^T$ and $\{u_{t,l}\}_{t=1}^T$ are $I(0)$ processes. Therefore, $\mathbf{A}$ and $\mathbf{B}$ form a cointegrated system. For simplicity, consider the following example: for each $l = 1, \cdots, M$ and $j = 1, \cdots, J$,

$$a_{t,l} = \sum_{j=1}^{J} b_{jt,l} w_{0,j} + \sum_{k=1}^{K} c_{kt,l} r_{0,k,l} + u_{t,l},$$

$$b_{jt,l} = b_{j(t-1),l} + v_{jt,l},$$

where $u_{t,l}$ and $v_{jt,l}$ are stationary unobserved disturbances. In this scenario, $(1, -\mathbf{w}_0')'$ plays the role of a cointegrating vector such that the linear combination of $\mathbf{A}$ and $\mathbf{B}$ is stationary. The normalizing matrix $\mathbf{D} = \mathrm{diag}\{T_0, \cdots, T_0, \sqrt{T_0}, \cdots, \sqrt{T_0}\}$, where the first J elements are T_0 and the remaining ones are $\sqrt{T_0}$. Let $\check{\mathbf{Z}}_t = (\check{\mathbf{z}}_{t,1}, \cdots, \check{\mathbf{z}}_{t,M})$ where $\check{\mathbf{z}}_{t,l} = (\check{\mathbf{z}}_{t,\triangleleft,l}', \check{\mathbf{z}}_{t,\triangleright,l}')'$ be the $((l-1)T_0 + t)$th column of $\text{diag}\{T_0^{-1/2}\mathbf{I}_J, \mathbf{I}_{KM}\}\mathbf{Z}'$. We use $\check{\mathbf{z}}_{t,\lhd,l}$ to denote the vector of the first J elements of $\check{\mathbf{z}}_t$, which corresponds to the nonstationary components. The remaining vector is denoted by $\check{\mathbf{z}}_{t,\rhd,l}$, which corresponds to the stationary components. Recall $\mathbf{u}_t = (u_{t,1}, \cdots, u_{t,M})'$. Write $\mathbf{v}_{t,l} = (v_{1t,l}, \cdots, v_{Jt,l})'$, $\mathbf{v}_t = (\mathbf{v}'_{t,1}, \cdots, \mathbf{v}'_{t,M})'$, $\mathbf{c}_{t,l} = (c_{1t,l}, \cdots, c_{kt,l})'$, $\mathbf{C}_t = (\mathbf{c}_{t,1}, \cdots, \mathbf{c}_{t,M})$. We allow some elements in $\mathbf{v}_t$ to be used in $\mathbf{C}_t$. Let $\mathbf{q}_t$ collect all distinct variables in $\mathbf{u}_t$, $\mathbf{v}_t$, $\mathbf{c}_{t,1}$, $\cdots$, $\mathbf{c}_{t,M}$. Moreover, define

$$\mathbf{Q}_{\lhd} = \frac{1}{T_0}\sum_{l=1}^{M}\sum_{t=1}^{T_0}\mathbf{G}_l(t/T_0)\mathbf{G}_l(t/T_0)',$$

where $\mathbf{G} = (\mathbf{G}'_1, \cdots, \mathbf{G}'_M)'$ is a mean-zero Brownian motion on $[0,1]$ with variance $\mathbb{E}[\mathbf{v}_t\mathbf{v}'_t]$.

As in the precious example, we consider the generic variance estimator

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T_0}\sum_{t=1}^{T_0}\check{\mathbf{Z}}_t\widehat{\mathbb{V}}[\mathbf{u}_t|\mathscr{H}]\check{\mathbf{Z}}'_t,$$

where $\widehat{\mathbb{V}}[\mathbf{u}_t|\mathscr{H}]$ is an estimate of $\mathbb{V}[\mathbf{u}_t|\mathscr{H}]$.

The following theorem gives more primitive conditions and verifies the high-level conditions of Theorems 1 and 2 in the paper for the cointegration scenario.

Theorem SA-3 (Cointegration)**.** *Assume that* $\{\mathbf{q}_t\}_{t=1}^{T}$ *is i.i.d over* $t = 1, \cdots, T_0$ *with mean zero and finite variance.*

Assume that, for $\psi \geq 3$ *and finite non-negative constants* $\bar{\eta}_1$, $\bar{\eta}_2$, $\underline{\eta}_1$, $\underline{\eta}_2$ *and* $\mathfrak{c}_Q$, $\pi_{Q,1}$, *finite positive constants* $\mathfrak{C}_Q$ *and* $\pi_{Q,2}$, *and constant* $\nu_Q \in (0, 1/2)$, *the following conditions hold:*

(SA-3.i) $\max_{1\leq t\leq T_0}\mathbb{E}[\|\mathbf{u}_t\|^{\psi}|\mathbf{B},\mathbf{C}] \leq \bar{\eta}_1$ *a.s. on* $\sigma(\mathbf{B},\mathbf{C})$ *and* $\mathbb{E}[\|\mathbf{q}_t\|^{\psi}] \leq \bar{\eta}_2$.

(SA-3.ii) $\min_{1\leq t\leq T_0}\lambda_{\min}(\mathbb{V}[\mathbf{u}_t|\mathbf{B},\mathbf{C}]) \geq \underline{\eta}_1$ *a.s. on* $\sigma(\mathbf{B},\mathbf{C})$ *and* $\lambda_{\min}(\mathbb{E}[\mathbf{q}_t\mathbf{q}'_t]) \geq \underline{\eta}_2$.

(SA-3.iii) $\mathbb{P}\big[(\log T_0)^{-\mathfrak{c}_Q} \leq \lambda_{\min}(\mathbf{Q}_{\lhd}) \leq \lambda_{\max}(\mathbf{Q}_{\lhd}) \leq (\log T_0)^{\mathfrak{c}_Q}\big] \geq 1 - \pi_{Q,1}$.

(SA-3.iv) $\mathbb{P}\big[\|\frac{1}{T_0}\sum_{l=1}^{M}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\lhd,l}\check{\mathbf{z}}'_{t,\rhd,l}\| \leq \mathfrak{C}_Q T_0^{-1/2+\nu_Q}\big] \geq 1 - \pi_{Q,2}$.

Then, conditions (T1.i) *and* (T1.ii) *of Theorem 1 hold with* $\pi_\gamma = \mathfrak{C}_{\pi,1}T_0^{-\psi\nu} + \mathfrak{C}_{\pi,2}T_0^{-1} + \pi_{Q,1} + \pi_{Q,2}$ *and* $\epsilon_\gamma = \mathfrak{C}_\epsilon(\log T_0)^{\frac{3}{2}(1+\mathfrak{c}_Q)}T_0^{-1/2}$ *for finite non-negative constants* $\mathfrak{C}_{\pi,1}$, $\mathfrak{C}_{\pi,2}$ *and* $\mathfrak{C}_\epsilon$, *which are characterized in the proof, and for any* $\nu \in (0, 1/2 - 1/\psi)$. *Therefore, condition* (T2.i) *of Theorem 2 also holds, while condition* (T2.ii) *holds with* $\varpi^\star_\delta = \mathfrak{C}^\star_\varpi(\log T_0)^{2\mathfrak{c}_Q+1/2}$, $\epsilon^\star_\delta = \mathfrak{C}^\star_\epsilon T_0^{-1}$, *and* $\pi^\star_\delta = \pi_\gamma$, *for finite non-negative constants* $\mathfrak{C}^\star_\varpi$ *and* $\mathfrak{C}^\star_\epsilon$, *which are also characterized in the proof.*

Assume in addition that, for finite non-negative constant $\pi^\star_w$, the following condition holds:

(SA-3.v) $\varrho = \varpi^\star_\delta / T_0$ *and* $\mathbb{P}(\min\{|w_{0,j}| : w_{0,j} \neq 0\} \geq \varrho) \geq 1 - \pi^\star_w$.

Then, condition (T2.iii) *of Theorem 2 holds with* $\pi^\star_\Delta = \pi^\star_\delta + \pi^\star_w + \pi_\gamma$ *and* $\epsilon^\star_\Delta = \epsilon^\star_\delta + \epsilon_\gamma$.

Finally, also assume that, for finite non-negative constants $\epsilon^\star_{\Sigma,1}$, $\epsilon^\star_{\Sigma,2}$ and $\pi^\star_\Sigma$, the following condition holds:

(SA-3.vi) $\mathbb{P}(\mathbb{P}(\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \leq \epsilon^\star_{\Sigma,1} | \mathscr{H}) \geq 1 - \epsilon^\star_{\Sigma,2}) \geq 1 - \pi^\star_\Sigma$.

Then, condition (T2.iv) *of Theorem 2 holds with* $\epsilon^\star_{\gamma,1} = \mathfrak{C}^\star_{\epsilon,1}(\log T_0)^{\mathfrak{c}_Q}\epsilon^\star_{\Sigma,1}$, $\epsilon^\star_{\gamma,2} = \epsilon^\star_{\Sigma,2}$, *and* $\pi^\star_\gamma = \pi^\star_\Sigma + \pi^\star_\delta$, *for finite non-negative constant* $\mathfrak{C}^\star_{\epsilon,1}$, *which is characterized in the proof.*

SA-2 Proofs

SA-2.1 Proof of Lemma 1

Let $\mathcal{E}$ be the event on which

$$\mathbb{P}\Big[M_{1,\mathtt{L}} \leq \mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) \leq M_{1,\mathtt{U}} \;\Big|\; \mathscr{H}\Big] \geq 1 - \alpha_1, \qquad \text{and}$$
$$\mathbb{P}\Big[M_{2,\mathtt{L}} \leq e_T \leq M_{2,\mathtt{U}} \;\Big|\; \mathscr{H}\Big] \geq 1 - \alpha_2.$$

By assumption, $\mathbb{P}(\mathcal{E}) \geq 1 - \pi_1 - \pi_2$. On $\mathcal{E}$, we have that

$$\begin{aligned}
&\mathbb{P}\Big[M_{1,\mathtt{L}} + M_{2,\mathtt{L}} \leq \mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) + e_T \leq M_{1,\mathtt{U}} + M_{2,\mathtt{U}} \Big| \mathscr{H}\Big] \\
&= 1 - \mathbb{P}\Big[\Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) + e_T > M_{1,\mathtt{U}} + M_{2,\mathtt{U}}\Big\} \cup \\
&\qquad \Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) + e_T < M_{1,\mathtt{L}} + M_{2,\mathtt{L}}\Big\} \Big| \mathscr{H}\Big] \\
&\geq 1 - \mathbb{P}\Big[\Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) > M_{1,\mathtt{U}}\Big\} \cup \Big\{e_T > M_{2,\mathtt{U}}\Big\} \cup \\
&\qquad \Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) < M_{1,\mathtt{L}}\Big\} \cup \Big\{e_T < M_{2,\mathtt{L}}\Big\} \Big| \mathscr{H}\Big] \\
&\geq 1 - \mathbb{P}\Big[\Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) > M_{1,\mathtt{U}}\Big\} \cup \Big\{\mathbf{p}_T'(\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}) < M_{1,\mathtt{L}}\Big\} \Big| \mathscr{H}\Big] - \\
&\qquad \mathbb{P}\Big[\Big\{e_T > M_{2,\mathtt{U}}\Big\} \cup \Big\{e_T < M_{2,\mathtt{L}}\Big\} \Big| \mathscr{H}\Big].
\end{aligned}$$

Then, the result directly follows. □

SA-2.2 Proof of Lemma 2

By definition of $\widehat{\boldsymbol{\beta}}$, $(\mathbf{A}-\mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{A}-\mathbf{Z}\widehat{\boldsymbol{\beta}}) \leq (\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_0)'(\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_0)$, implying that $\widehat{\boldsymbol{\delta}}'\widehat{\mathbf{Q}}\widehat{\boldsymbol{\delta}} \leq 2\widehat{\boldsymbol{\gamma}}'\widehat{\boldsymbol{\delta}} = 2\boldsymbol{\gamma}'\widehat{\boldsymbol{\delta}}+2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\widehat{\boldsymbol{\delta}}$. On the other hand, for any $\boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_0+\alpha(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$, $\alpha \in (0,1]$, since $\mathcal{W}\times\mathcal{R}$ is convex, $\boldsymbol{\beta}_\alpha \in \mathcal{W}\times\mathcal{R}$. By definition of $\mathbf{w}_0$, $\mathbb{E}_{\widehat{\boldsymbol{\beta}}}[(\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_0)'(\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_0)|\mathscr{H}] \leq \mathbb{E}_{\widehat{\boldsymbol{\beta}}}[(\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_\alpha)'(\mathbf{A}-\mathbf{Z}\boldsymbol{\beta}_\alpha)|\mathscr{H}]$, where $\mathbb{E}_{\widehat{\boldsymbol{\beta}}}[\cdot|\mathscr{H}]$ denotes the expectation against the conditional distribution of $(\mathbf{A},\mathbf{Z})$ given $\mathscr{H}$ with $\widehat{\boldsymbol{\beta}}$ treated as fixed. Then, we have $\alpha\widehat{\boldsymbol{\delta}}'\widehat{\mathbf{Q}}\widehat{\boldsymbol{\delta}} \geq 2\boldsymbol{\gamma}'\widehat{\boldsymbol{\delta}}$. Since it holds for any $\alpha \in (0,1]$, $\boldsymbol{\gamma}'\widehat{\boldsymbol{\delta}} \leq 0$. Consequently, we have $\widehat{\boldsymbol{\delta}}'\widehat{\mathbf{Q}}\widehat{\boldsymbol{\delta}} \leq 2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\widehat{\boldsymbol{\delta}}$.

For any $\boldsymbol{\xi}_1,\boldsymbol{\xi}_2 \in \mathbb{R}^d$ such that $\sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}_1}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa$ and $\sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}_2}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa$. Let $\tilde{\boldsymbol{\xi}} = \alpha\boldsymbol{\xi}_1+(1-\alpha)\boldsymbol{\xi}_2$ for $\alpha\in[0,1]$. Consider $\mathcal{M}_{\tilde{\boldsymbol{\xi}}} = \{\boldsymbol{\delta}\in\Delta : \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2\tilde{\boldsymbol{\xi}}'\boldsymbol{\delta} \leq 0\}$. For any $\boldsymbol{\delta}\in\mathcal{M}_{\tilde{\boldsymbol{\xi}}}$,

$$\Big(\alpha\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2\alpha\boldsymbol{\xi}_1'\boldsymbol{\delta}\Big) + \Big((1-\alpha)\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2(1-\alpha)\boldsymbol{\xi}_2'\boldsymbol{\delta}\Big) \leq 0.$$

It immediately follows that either $\alpha\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2\alpha\boldsymbol{\xi}_1'\boldsymbol{\delta} \leq 0$ or $((1-\alpha)\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2(1-\alpha)\boldsymbol{\xi}_2'\boldsymbol{\delta}) \leq 0$, which implies that $\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}_1}$ or $\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}_2}$. In either case, $\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa$. Therefore, $\{\boldsymbol{\xi} : \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa\}$ is convex. The proof for $\{\boldsymbol{\xi} : \inf_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \geq \kappa\}$ is similar. □

SA-2.3 Proof of Theorem 1

We write $\tilde{u}_{t,l} = u_{t,l} - \mathbb{E}[u_{t,l}|\mathscr{H}]$. Fix $\widehat{\mathbf{Q}}$ and $\mathbf{p}_T$, define $\mathcal{A}_\kappa = \{\boldsymbol{\xi}\in\mathbb{R}^d : \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa\}$ for $\kappa\in\mathbb{R}$. $\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa\} \subseteq \{\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \leq \kappa\}$ for any κ. By Berry-Esseen Theorem for convext sets (Raič, 2019),

$$\Big|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H}) - \mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})\Big| \leq 42(d^{1/4}+16)\sum_{t=1}^{T_0}\mathbb{E}\Big[\Big\|\sum_{l=1}^{M}\tilde{\mathbf{z}}_{t,l}\tilde{u}_{t,l}\Big\|^3|\mathscr{H}\Big].$$

Recall that $\tilde{\mathbf{z}}_{t,l}$ is the $(t+(l-1)T_0)$th column of $\boldsymbol{\Sigma}^{-1/2}\mathbf{D}^{-1}\mathbf{Z}'$, $\mathbf{G}|\mathscr{H}\sim\mathsf{N}(0,\boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$.

Then, by condition (T1.ii), with probability at least $1-\pi_\gamma$ over $\mathscr{H}$,

$$\Big|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H}) - \mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})\Big| \leq \epsilon_\gamma.$$

It yields the desired result: for all κ,

$$\mathbb{P}(\mathbf{p}_T'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq\kappa|\mathscr{H})\geq\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H})\geq\mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})-\epsilon_\gamma=\mathbb{P}(\varsigma_{\mathtt{U}}^\dagger\leq\kappa|\mathscr{H})-\epsilon_\gamma.$$

□

Remark SA-2.1 (Lower Bound for Theorem 1)**.** The lower bound for $\mathbf{p}_T'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ follows similarly. Specifically, under the same conditions of Theorem 1, we have

$$\mathbb{P}\Big[\mathbb{P}(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}}\geq\mathfrak{c}^\dagger(\alpha)|\mathscr{H})\geq 1-(\alpha+\epsilon_\gamma)\Big]\geq 1-\pi_\gamma,$$

where $\mathfrak{c}^\dagger(\alpha)$ denotes the α-quantile of $\varsigma_{\mathtt{L}}^\dagger=\inf\{\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta}:\boldsymbol{\delta}\in\mathcal{M}_{\mathbf{G}}\}$ conditional on $\mathscr{H}$, with $\mathcal{M}_{\mathbf{G}}=\{\boldsymbol{\delta}\in\Delta:\ell^\dagger(\boldsymbol{\delta})\leq 0\}$, $\ell^\dagger(\boldsymbol{\delta})=\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2\mathbf{G}'\boldsymbol{\delta}$, and $\mathbf{G}|\mathscr{H}\sim\mathsf{N}(\mathbf{0},\boldsymbol{\Sigma})$. ⌟

SA-2.4 Proof of Theorem 2

We first introduce some notation. Define $\mathcal{A}_\kappa=\{\boldsymbol{\xi}\in\mathbb{R}^d:\sup_{\boldsymbol{\delta}\in\tilde{\mathcal{M}}^\star_{\boldsymbol{\xi}}}\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta}\leq\kappa\}$ where

$$\tilde{\mathcal{M}}^\star_{\boldsymbol{\xi}}=\Big\{\boldsymbol{\delta}\in\Delta^\star:\|\boldsymbol{\delta}\|\leq\varpi^\star_\delta,\ \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2\boldsymbol{\xi}'\boldsymbol{\delta}\leq 0\Big\}.$$

Accordingly, define

$$\tilde{\varsigma}^\star_{\mathtt{U}}=\sup\Big\{\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta}:\boldsymbol{\delta}\in\Delta^\star,\ \|\boldsymbol{\delta}\|\leq\varpi^\star_\delta,\ \ell^\star(\boldsymbol{\delta})\leq 0\Big\}\text{ and}$$
$$\tilde{\varsigma}^\dagger_{\mathtt{U}}=\sup\Big\{\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta}:\boldsymbol{\delta}\in\Delta,\ \|\boldsymbol{\delta}\|\leq\varpi^\star_\delta,\ \ell^\dagger(\boldsymbol{\delta})\leq 0\Big\}.$$

Let $\tilde{\mathfrak{c}}^\star(1-\alpha)$ be the $(1-\alpha)$-quantile of $\tilde{\varsigma}^\star_{\mathtt{U}}$ conditional on the data. Similarly, $\tilde{\mathfrak{c}}^\dagger(1-\alpha)$ is the $(1-\alpha)$-quantile of $\tilde{\varsigma}^\dagger_{\mathtt{U}}$ conditional on $\mathscr{H}$.

Let $\mathbb{P}_1=\mathsf{N}(0,\widehat{\boldsymbol{\Sigma}})$ and $\mathbb{P}_2=\mathsf{N}(0,\boldsymbol{\Sigma})$. Then, by Pinsker's inequality, for any κ,

$$\Big|\mathbb{P}(\mathbf{G}^\star\in\mathcal{A}_\kappa|\mathscr{H},\mathbf{A})-\mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})\Big|\leq\sqrt{\mathbb{KL}(\mathbb{P}_1,\mathbb{P}_2)/2},$$

where $\mathbb{KL}(\cdot,\cdot)$ is the Kullback-Leibler divergence. Define $\boldsymbol{\Gamma} = \boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2}$. Note that

$$\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) = -0.5\log\{\det(\boldsymbol{\Gamma})\} + 0.5\,\mathrm{tr}(\boldsymbol{\Gamma} - \mathbf{I}) = 0.5\sum_{j=1}^{d}(\lambda_j - \log(\lambda_j + 1)),$$

where λ_j is the jth largest eigenvalue of $\boldsymbol{\Gamma} - \mathbf{I}$. By condition (T2.iv), with probability over $\mathscr{H}$ at least $1 - \pi^\star_\gamma$, it holds that with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1 - \epsilon^\star_{\gamma,2}$, $|\lambda_1| \leq 0.5$ and

$$\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) \leq 0.5\sum_{j=1}^{d}\lambda_j^2 \leq (2\epsilon^\star_{\gamma,1})^2/2.$$

In view of condition (T2.iii), for any κ,

$$\mathbb{P}\Big\{\mathbb{P}\Big(|\mathbb{P}^\star(\tilde{\varsigma}^\star_{\mathtt{U}} \leq \kappa) - \mathbb{P}^\star(\tilde{\varsigma}^\dagger_{\mathtt{U}} \leq \kappa)| \leq \epsilon^\star_{\gamma,1}\Big|\mathscr{H}\Big) \geq 1 - \epsilon^\star_{\gamma,2} - \epsilon^\star_\Delta\Big\} \geq 1 - \pi^\star_\gamma - \pi^\star_\Delta.$$

Note that by construction, $\mathfrak{c}^\star(1-\alpha) \geq \tilde{\mathfrak{c}}^\star(1-\alpha)$. Then, we have

$$\mathbb{P}\Big\{\mathbb{P}\Big(\mathfrak{c}^\star(1-\alpha) \geq \tilde{\mathfrak{c}}^\dagger(1-\alpha-\epsilon^\star_{\gamma,1})\Big|\mathscr{H}\Big) \geq 1 - \epsilon^\star_{\gamma,2} - \epsilon^\star_\Delta\Big\} \geq 1 - \pi^\star_\gamma - \pi^\star_\Delta. \tag{SA-2.1}$$

By condition (T2.ii), with probability over $\mathscr{H}$ at least $1 - \pi^\star_\delta$,

$$\mathbb{P}\Big(\mathcal{M}_{\mathbf{G}} = \{\boldsymbol{\delta} \in \Delta : \|\boldsymbol{\delta}\| \leq \varpi^\star_\delta, \ell^\dagger(\boldsymbol{\delta}) \leq 0\}\Big|\mathscr{H}\Big) \geq 1 - \epsilon^\star_\delta,$$

which implies that for any κ, $\mathbb{P}(\tilde{\varsigma}^\dagger_{\mathtt{U}} \leq \kappa|\mathscr{H}) \leq \mathbb{P}(\varsigma^\dagger_{\mathtt{U}} \leq \kappa|\mathscr{H}) + \epsilon^\star_\delta$. Thus,

$$\mathbb{P}\Big\{\mathbb{P}\Big(\tilde{\mathfrak{c}}^\dagger(1-\alpha-\epsilon^\star_{\gamma,1}) \geq \mathfrak{c}^\dagger(1-\alpha-\epsilon^\star_{\gamma,1}-\epsilon^\star_\delta)\Big|\mathscr{H}\Big)\Big\} \geq 1 - \pi^\star_\delta. \tag{SA-2.2}$$

Therefore, on an event $\mathcal{A} \in \mathscr{H}$ with $\mathbb{P}(\mathcal{A}) \geq 1 - \pi_\gamma - \pi^\star_\gamma - \pi^\star_\Delta - \pi^\star_\delta$,

$$\begin{aligned}
\mathbb{P}\Big(\mathbf{p}_T'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq \mathfrak{c}^\star(1-\alpha)\Big|\mathscr{H}\Big) &\geq \mathbb{P}\Big(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \leq \tilde{\mathfrak{c}}^\dagger(1-\alpha-\epsilon^\star_{\gamma,1})\Big|\mathscr{H}\Big) - \epsilon^\star_{\gamma,2} - \epsilon^\star_\Delta \\
&\geq \mathbb{P}\Big(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \leq \mathfrak{c}^\dagger(1-\alpha-\epsilon^\star_{\gamma,1}-\epsilon^\star_\delta)\Big|\mathscr{H}\Big) - \epsilon^\star_{\gamma,2} - \epsilon^\star_\Delta \\
&\geq 1 - \alpha - \epsilon^\star_{\gamma,1} - \epsilon^\star_\delta - \epsilon_\gamma - \epsilon^\star_{\gamma,2} - \epsilon^\star_\Delta,
\end{aligned}$$

where the first line follows by inequality (SA-2.1), the second by inequality (SA-2.2), and the third by condition (T2.i). Then the proof is complete. $\square$

Remark SA-2.2 (Lower Bound for Theorem 2)**.** The lower bound for $\mathbf{p}_T'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ follows similarly. Specifically, we replace condition (T2.i) in Theorem 2 with $\mathbb{P}[\mathbb{P}(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \geq \mathfrak{c}^{\dagger}(\alpha)|\mathscr{H}) \geq 1-\alpha-\epsilon_\gamma] \geq 1-\pi_\gamma$, and the other conditions in Theorem 2 remain the same. Under these conditions, we have for $\epsilon_\gamma \in [0, 0.25]$,

$$\mathbb{P}\Big[\mathbb{P}(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \geq \mathfrak{c}^{\star}(\alpha)|\mathscr{H}) \geq 1-\alpha-\epsilon\Big] \geq 1-\pi,$$

where $\mathfrak{c}^{\star}(\alpha)$ denotes the α-quantile of $\varsigma_{\mathtt{L}}^{\star} = \inf\{\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^{\star}, \ell^{\star}(\boldsymbol{\delta}) \leq 0\}$ conditional on the data. π and ϵ are defined as in Theorem 2. ⌟

SA-2.5 Proof of Theorem A

Throughout this proof, we write $\mathfrak{b}(\cdot) := \mathfrak{b}(\cdot;\mathscr{H})$ and assume $\mathbf{u}_t$ is correctly centered, i.e., $\mathbb{E}[\mathbf{u}_t|\mathscr{H}] = 0$, only for ease of notation. Let $\|\mathbf{v}\|_\infty$ denote the sup-norm of a generic vector $\mathbf{v}$. As in the proof of Theorem 1, fix $\widehat{\mathbf{Q}}$ and $\mathbf{p}_T$, define $\mathcal{A}_\kappa = \{\boldsymbol{\xi} \in \mathbb{R}^d : \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}}} \mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} \leq \kappa\}$ for each $\kappa \in \mathbb{R}$. Since $\{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \in \mathcal{A}_\kappa\} \subseteq \{\mathbf{p}_T'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \leq \kappa\}$ for any κ,

$$\mathbb{P}(\mathbf{p}_T'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \leq \kappa|\mathscr{H}) \geq \mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H}).$$

It suffices to approximate $\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H})$.

According to the blocking design, let $\mathbf{h}_k = (\mathbf{u}_t)_{t\in\mathcal{J}_k}$ and $\boldsymbol{\hbar}_k = (\mathbf{u}_t)_{t\in\mathcal{J}_k'}$ be vectors that contain all $\mathbf{u}_t$'s with t in $\mathcal{J}_k$ and $\mathcal{J}_k'$ respectively. By Berbee's coupling lemma (Berbee, 1987), on an enlarged probability space, there exists $\{\mathbf{h}_k^*\}_{k=1}^m$ that is independent over k conditional on $\mathscr{H}$ such that for each k, $\mathbf{h}_k$ and $\mathbf{h}_k^*$ are identically distributed (conditional on $\mathscr{H}$), and $\mathbb{P}(\mathbf{h}_k \neq \mathbf{h}_k^*|\mathscr{H}) \leq \mathfrak{b}(v)$. Similarly, on an enlarged probability space, there exists $\{\boldsymbol{\hbar}_k^*\}_{k=1}^m$ that is independent over k conditional on $\mathscr{H}$ such that for each k, $\boldsymbol{\hbar}_k$ and $\boldsymbol{\hbar}_k^*$ are identically distributed (conditional on $\mathscr{H}$) and $\mathbb{P}(\boldsymbol{\hbar}_k \neq \boldsymbol{\hbar}_k^*|\mathscr{H}) \leq \mathfrak{b}(q)$. Let $\mathbf{u}_t^*$ be the vector in $\mathbf{h}_k^*$ and $\boldsymbol{\hbar}_k^*$ that corresponds to $\mathbf{u}_t$. Accordingly, $\mathbf{s}_t^*$ is the same as $\mathbf{s}_t$ except that $\mathbf{u}_t$ is replaced by $\mathbf{u}_t^*$. Also, $\mathbf{S}_{k,\square}^* = \sum_{t\in\mathcal{J}_k}\mathbf{s}_t^*$ and $\mathbf{S}_{k,\diamond}^* = \sum_{j\in\mathcal{J}_k'}\mathbf{s}_t^*$.

By the coupling argument, we have

$$\mathbb{P}\Big(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H}\Big) \leq \mathbb{P}\Big(\Big\|\sum_{k=1}^m \mathbf{S}_{k,\diamond}\Big\|_\infty \geq \xi_1\Big|\mathscr{H}\Big) + \mathbb{P}\Big(\Big\|\mathbf{S}_{m+1,1}\Big\|_\infty \geq \xi_2\Big|\mathscr{H}\Big)+$$

$$
\begin{aligned}
&\mathbb{P}\Big(\sum_{k=1}^{m}\mathbf{S}_k \in \mathcal{A}_\kappa^{\xi_1+\xi_2}\Big|\mathscr{H}\Big)\\
\le\ &\mathbb{P}\Big(\Big\|\sum_{k=1}^{m}\mathbf{S}_{k,\diamond}^*\Big\|_\infty \ge \xi_1\Big|\mathscr{H}\Big) + \mathbb{P}\Big(\Big\|\mathbf{S}_{m+1,1}\Big\|_\infty \ge \xi_2\Big|\mathscr{H}\Big)\\
&\mathbb{P}\Big(\sum_{k=1}^{m}\mathbf{S}_{k,\square}^* \in \mathcal{A}_\kappa^{\xi_1+\xi_2}\Big|\mathscr{H}\Big) + m\mathfrak{b}(v) + m\mathfrak{b}(q)\\
=:\ & I + II + III + m\mathfrak{b}(v) + m\mathfrak{b}(q),
\end{aligned}
$$

where $\mathcal{A}_\kappa^{\xi_1+\xi_2} = \{\mathbf{v} : \rho(\mathbf{v},\mathcal{A}_\kappa) \le \xi_1 + \xi_2\}$ is the $(\xi_1+\xi_2)$-enlargement of $\mathcal{A}_\kappa$ for $\xi_1, \xi_2 \ge 0$ and $\rho(\mathbf{v},\mathcal{A}_\kappa) = \inf_{\mathbf{v}'\in\mathcal{A}_\kappa}\|\mathbf{v}-\mathbf{v}'\|$.

For I, by Markov's inequality, condition (TA.ii), and Lemma 8 of Chernozhukov, Chetverikov and Kato (2015), for some absolute constant $\mathfrak{C}_1 > 0$, for any $\xi_1 > 0$,

$$
\mathbb{P}\Big(\Big\|\sum_{k=1}^{m}\mathbf{S}_{k,\diamond}^*\Big\|_\infty \ge \xi_1\Big|\mathscr{H}\Big) \le \mathfrak{C}_1\xi_1^{-1}\Big(\sqrt{mv\bar{\sigma}^2\log d} + \eta_1^{1/\psi}\log d\Big),
$$

where we use the fact that

$$
\begin{aligned}
(\mathbb{E}[(\max_{1\le k\le m}\|\mathbf{S}_{k,\diamond}^*\|_\infty)^2|\mathscr{H}])^{1/2} &\le (\mathbb{E}[\max_{1\le j\le d}\max_{1\le k\le m}|S_{jk,\diamond}^*|^\psi|\mathscr{H}])^{1/\psi}\\
&\le \Big(\mathbb{E}\Big[\sum_{j=1}^{d}\sum_{k=1}^{m}|S_{jk,\diamond}|^\psi\Big|\mathscr{H}\Big]\Big)^{1/\psi} \le \eta_1^{1/\psi},
\end{aligned}
$$

which holds with probability over $\mathscr{H}$ at least $1-\pi_{\gamma,1}$. Taking $\xi_1 = \mathfrak{C}_1\eta_6^{-1}\Big(\sqrt{mv\bar{\sigma}^2\log d}+\eta_1^{1/\psi}\log d\Big)$, we have $I \le \eta_6$.

For II, take $\xi_2 = (d\eta_2)^{1/\psi}\eta_6^{-1/\psi}$. By Markov's inequality and condition (TA.iii), with probability over $\mathscr{H}$ at least $1-\pi_{\gamma,2}$,

$$
II \le d\max_{1\le j\le d}\mathbb{P}\Big(\Big|\sum_{t\in\mathcal{J}_{m+1}} s_{jt}\Big| \ge \xi_2|\mathscr{H}\Big) \le \max_{1\le j\le d}\frac{\mathbb{E}[|\sum_{t\in\mathcal{J}_{m+1}} s_{jt}|^\psi|\mathscr{H}]\eta_6}{\eta_2} \le \eta_6.
$$

For III, since $\mathcal{A}_\kappa^{\xi_1+\xi_2}$ is also convex and $\{\mathbf{S}_{k,\square}^*\}$ is independent over k, repeating the argument in the proof of Theorem 1 and using condition (TA.iv), we have that for any $\xi_1 > 0$ and $\xi_2 > 0$, with probability over $\mathscr{H}$ at least $1-\pi_{\gamma,3}$,

$$\Big|III - \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa^{\xi_1+\xi_2}|\mathscr{H})\Big| \le 42(d^{1/4}+16)\sum_{k=1}^m \mathbb{E}\Big[\Big\|\mathbf{\Sigma}_\square^{-1/2}\mathbf{S}_{k,\square}\Big\|^3\Big|\mathscr{H}\Big] \le \eta_3$$

for $\mathbf{G}_\square|\mathscr{H} \sim \mathsf{N}(0,\mathbf{\Sigma}_\square)$.

Combining the results above, we have with probability over $\mathscr{H}$ at least $1-\pi_{\gamma,1}-\pi_{\gamma,2}-\pi_{\gamma,3}$,

$$\mathbb{P}\Big(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H}\Big) \le \mathbb{P}\Big(\mathbf{G}_\square \in \mathcal{A}_\kappa^{\xi_1+\xi_2}\Big|\mathscr{H}\Big) + 5\eta_6.$$

On the other hand, for any $\xi>0$, define

$$\mathcal{A}_\kappa^{-\xi} = \Big\{\mathbf{h} : \mathcal{B}(\mathbf{h},\xi)\subseteq\mathcal{A}_\kappa\Big\}.$$

Suppose that $\mathcal{A}_\kappa^{-\xi_1-\xi_2}\neq\emptyset$. Note that for any $\mathbf{h}_1,\mathbf{h}_2\in\mathcal{A}_\kappa^{-\xi}$ and $\omega\in[0,1]$, define $\mathbf{h}'=\omega\mathbf{h}_1+(1-\omega)\mathbf{h}_2$. For any $\mathbf{h}''\in\mathcal{B}(\mathbf{h}',\xi)$,

$$\mathbf{h}'' = \mathbf{h}'+(\mathbf{h}''-\mathbf{h}') = \omega(\mathbf{h}_1+(\mathbf{h}''-\mathbf{h}')) + (1-\omega)(\mathbf{h}_2+(\mathbf{h}''-\mathbf{h}')).$$

Since $\|\mathbf{h}''-\mathbf{h}'\|\le\xi$, $\mathbf{h}_1+(\mathbf{h}''-\mathbf{h}')\in\mathcal{B}(\mathbf{h}_1,\xi)\subseteq\mathcal{A}_\kappa$. Similarly, $\mathbf{h}_2+(\mathbf{h}''-\mathbf{h}')\in\mathcal{A}_\kappa$. Therefore, $\mathbf{h}''\in\mathcal{A}_\kappa$. Since it holds for arbitrary $\mathbf{h}''\in\mathcal{B}(\mathbf{h}',\xi)$, $\mathcal{A}_\kappa^{-\xi}$ is convex. Then, repeat the coupling argument and apply the Berry-Esseen inequality again. We have the following inequalities hold with probability over $\mathscr{H}$ at least $1-\pi_{\gamma,3}$:

$$\begin{aligned}
&\mathbb{P}(\mathbf{G}_\square\in\mathcal{A}_\kappa^{-(\xi_1+\xi_2)}|\mathscr{H}) - \eta_3 \le \mathbb{P}\Big(\sum_{k=1}^m \mathbf{S}_{k,\square}^*\in\mathcal{A}_\kappa^{-(\xi_1+\xi_2)}\Big|\mathscr{H}\Big)\\
\le&\,\mathbb{P}\Big(\sum_{k=1}^m \mathbf{S}_{k,\square}\in\mathcal{A}_\kappa^{-(\xi_1+\xi_2)}\Big|\mathscr{H}\Big) + m\mathfrak{b}(v)\\
\le&\,\mathbb{P}\Big(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa\Big|\mathscr{H}\Big) + \mathbb{P}\Big(\Big\|\sum_{k=1}^m\mathbf{S}_{k,\diamond}\Big\|_\infty\ge\xi_1\Big|\mathscr{H}\Big) + \mathbb{P}\Big(\Big\|\mathbf{S}_{m+1,\diamond}\Big\|_\infty\ge\xi_2\Big|\mathscr{H}\Big) + m\mathfrak{b}(v).
\end{aligned}$$

Using the bounds obtained previously, the above implies that with probability over $\mathscr{H}$ at least

$1 - \pi_{\gamma,1} - \pi_{\gamma,2} - \pi_{\gamma,3}$,

$$\mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa | \mathscr{H}) \geq \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa^{-(\xi_1 - \xi_2)} | \mathscr{H}) - 5\eta_6.$$

By condition (TA.v) and Anti-Concentration of the Gaussian measure for convex sets (see, e.g., Lemma A.2 of Chernozhukov, Chetverikov, Kato et al. (2017)), for some absolute constant $\mathfrak{C}_2 > 0$, with probability over $\mathscr{H}$ at least $1 - \pi_{\gamma,4}$,

$$\mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa^{\xi_1+\xi_2} | \mathscr{H}) \leq \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa | \mathscr{H}) + \mathfrak{C}_2 d\eta_4(\xi_1 + \xi_2), \quad \text{and}$$
$$\mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa^{-(\xi_1+\xi_2)} | \mathscr{H}) \geq \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa) - \mathfrak{C}_2 d\eta_4(\xi_1 + \xi_2).$$

The result for $\mathcal{A}_\kappa^{-(\xi_1+\xi_2)} = \emptyset$ trivially follows. Then, we have with probability over $\mathscr{H}$ at least $1 - \pi_{\gamma,1} - \pi_{\gamma,2} - \pi_{\gamma,3} - \pi_{\gamma,4}$,

$$\left| \mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa | \mathscr{H}) - \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa | \mathscr{H}) \right| \leq 5\eta_6 + \mathfrak{C}_2 d\eta_4(\xi_1 + \xi_2) \leq \mathfrak{C}_3 \eta_6$$

for some constant $\mathfrak{C}_3 > 0$.

Finally, as in the proof of Theorem 2, letting $\mathbb{P}_1 = \mathsf{N}(0, \boldsymbol{\Sigma})$ and $\mathbb{P}_2 = \mathsf{N}(0, \boldsymbol{\Sigma}_\square)$, by Pinsker's inequality, for any κ,

$$\left| \mathbb{P}(\mathbf{G} \in \mathcal{A}_\kappa | \mathscr{H}) - \mathbb{P}(\mathbf{G}_\square \in \mathcal{A}_\kappa | \mathscr{H}) \right| \leq \sqrt{\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2)/2},$$

where

$$\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) = -0.5 \log\{\det(\boldsymbol{\varGamma})\} + 0.5 \operatorname{tr}(\boldsymbol{\varGamma} - \mathbf{I}) = 0.5 \sum_{j=1}^{d} (\lambda_j - \log(\lambda_j + 1)),$$

where $\boldsymbol{\varGamma} = \boldsymbol{\Sigma}_\square^{-1/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}_\square^{-1/2}$ and λ_j is the jth largest eigenvalue of $\boldsymbol{\varGamma} - \mathbf{I}$. By condition (TA.vi), with probability over $\mathscr{H}$ at least $1 - \pi_{\gamma,5}$, $|\lambda_1| \leq 0.5$ and

$$\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) \leq 0.5 \sum_{j=1}^{d} \lambda_j^2 \leq (2\eta_5)^2/2.$$

Then the desired result follows. $\square$

Remark SA-2.3 (Lower Bound for Theorem A)**.** The lower bound for $\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}}$ follows similarly. Specifically, under the same conditions of Theorem A, we have for $\eta_5 \in [0, 0.25]$

$$\mathbb{P}\Big[\mathbb{P}(\mathbf{p}_T'\mathbf{D}^{-1}\widehat{\boldsymbol{\delta}} \geq \mathfrak{c}^\dagger(\alpha)|\mathscr{H}) \geq 1-\alpha-\epsilon_\gamma\Big] \geq 1-\pi_\gamma,$$

where $\mathfrak{c}^\dagger(\alpha)$ denotes the α-quantile of $\varsigma_{\mathtt{L}}^\dagger = \inf\{\mathbf{p}_T'\mathbf{D}^{-1}\boldsymbol{\delta} : \boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}\}$ conditional on $\mathscr{H}$, with $\mathcal{M}_{\mathbf{G}} = \{\boldsymbol{\delta} \in \Delta : \ell^\dagger(\boldsymbol{\delta}) \leq 0\}$, $\ell^\dagger(\boldsymbol{\delta}) = \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2\mathbf{G}'\boldsymbol{\delta}$, and $\mathbf{G}|\mathscr{H} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$. ϵ_γ and π_γ are defined as in Theorem A. ⌟

SA-2.6 Proof of Theorem SA-1

Note that given the assumption, conditional on $\mathbf{B}$, $\mathbf{b}_T$ is independent of the pre-treatment data. Therefore, when we verify the conditions of Theorems 1 and 2, the analysis conditional on $\mathscr{H}$ is the same as that conditional on $\mathbf{B}$ only.

(1) We first verify the conditions of Theorem 1. Condition (T1.i) trivially holds. To show condition (T1.ii), we note by condition (SA-1.i), with probability (over $\mathscr{H}$) one,

$$\mathbb{E}\Big[\|\tilde{\mathbf{z}}_t\tilde{u}_t\|^3\Big|\mathscr{H}\Big] \leq \|\tilde{\mathbf{z}}_t\|^3\mathbb{E}[|\tilde{u}_t|^3|\mathscr{H}] \leq \bar{\eta}_1\|\tilde{\mathbf{z}}_t\|^3.$$

for any t. Then, it suffices to bound

$$\sum_{t=1}^{T_0}\|\tilde{\mathbf{z}}_t\|^3 \leq \|\boldsymbol{\Sigma}^{-1/2}\|^3\Big(\frac{1}{T_0^{3/2}}\sum_{t=1}^{T_0}\|\mathbf{z}_t\|^3\Big).$$

Note that $\frac{1}{T_0}\sum_{t=1}^{T_0}\|\mathbf{z}_t\|^3 \leq \frac{\sqrt{d}}{T_0}\sum_{t=1}^{T_0}\sum_{j=1}^{d}|z_{jt}|^3$ where z_{jt} is the jth element of $\mathbf{z}_t$. By Markov's inequality,

$$\mathbb{P}\Big(\Big|\frac{1}{T_0}\sum_{t=1}^{T_0}|z_{jt}|^3 - \mathbb{E}[|z_{jt}|^3]\Big| \geq \bar{\eta}_2\Big) \leq \frac{1}{T_0\bar{\eta}_2}.$$

Thus, with probability at least $1-d(\bar{\eta}_2T_0)^{-1}$, $\frac{1}{T_0}\sum_{t=1}^{T_0}\sum_{j=1}^{d}|z_{jt}|^3 \leq 2d\bar{\eta}_2$, and then $\frac{1}{T_0}\sum_{t=1}^{T_0}\|\mathbf{z}_t\|^3 \leq 2d^{3/2}\bar{\eta}_2$.

On the other hand, by the lower bound on the conditional variance of u_t, $\lambda_{\min}(\boldsymbol{\Sigma}) \geq \underline{\eta}_1\lambda_{\min}(\widehat{\mathbf{Q}})$.

By Markov's inequality and union bounds,

$$\mathbb{P}\Big(\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t' - \mathbb{E}[\mathbf{z}_t\mathbf{z}_t']\Big\|_{\mathtt{F}} \geq d\underline{\eta}_2/(2d)\Big) \leq \frac{4d^4\bar{\eta}_2}{\underline{\eta}_2^2 T_0}.$$

It follows that $\lambda_{\min}(\mathbf{\Sigma}) \geq \underline{\eta}_1\underline{\eta}_2/2$ with probability over $\mathscr{H}$ at least $1 - \frac{4d^4\bar{\eta}_2}{\underline{\eta}_2^2 T_0}$. Then, condition (T1.ii) holds.

(2) Now, we verify the conditions of Theorem 2. Condition (T2.i) follows by part (1). For conditions (T2.ii) and (T2.iii), note that $\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}$ implies that

$$\lambda_{\min}(\widehat{\mathbf{Q}})\|\boldsymbol{\delta}\|^2 \leq \boldsymbol{\delta}\widehat{\mathbf{Q}}\boldsymbol{\delta} \leq 2\mathbf{G}'\boldsymbol{\delta} \leq 2\|\mathbf{G}\|\|\boldsymbol{\delta}\|. \tag{SA-2.3}$$

Let $\sigma^2_{\max}$ be the largest diagonal element of $\mathbf{\Sigma}$. Note that for Gaussian random variables,

$$\mathbb{P}\Big(\|\mathbf{G}\| \geq \sqrt{2d\log T_0}\sigma_{\max}\Big|\mathscr{H}\Big) \leq 2d\exp\Big(-\frac{2\sigma^2_{\max}\log T_0}{2\sigma^2_{\max}}\Big) \leq 2d/T_0.$$

On the other hand, note that by Markov's inequality and union bounds,

$$\mathbb{P}\Big(\max_{1\leq j\leq d}\Big|\frac{1}{T_0}\sum_{t=1}^{T_0} z_{jt}^2 - \mathbb{E}[z_{jt}^2]\Big| \geq \bar{\eta}_2\Big) \leq \frac{d\bar{\eta}_2}{\bar{\eta}_2^2 T_0}.$$

Thus, $\sigma^2_{\max} \leq 2\bar{\eta}_2\bar{\eta}_1$ with probability over $\mathscr{H}$ at least $1 - \frac{d}{\bar{\eta}_2 T_0}$.

Using the results in part (1), we have for $\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}$, with probability over $\mathscr{H}$ at least $1 - \frac{4d^4\bar{\eta}_2}{\underline{\eta}_2^2 T_0} - \frac{d}{\bar{\eta}_2 T_0}$,

$$\mathbb{P}\Big(\|\boldsymbol{\delta}\| \leq \frac{8\sqrt{d\log T_0}\bar{\eta}_1\bar{\eta}_2}{\underline{\eta}_2}\Big|\mathscr{H}\Big) \geq 1 - \frac{2d}{T_0}.$$

Thus, condition (T2.ii) holds.

For condition (T2.iii), we note that $\widehat{\boldsymbol{\delta}}$ satisfies the same basic inequality (SA-2.3) except $\mathbf{G}$ needs to be replaced by $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$. Noting condition (T1.ii) proved in part (1), we can see the same (upper) bound also holds for $\widehat{\boldsymbol{\delta}}$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1 - 2d/T_0 - \epsilon_\gamma$, with probability over $\mathscr{H}$ at least $1 - \frac{4d^4\bar{\eta}_2}{\underline{\eta}_2^2 T_0} - \frac{d}{\bar{\eta}_2 T_0} - \pi_\gamma$.

Finally, consider the variance estimator $\widehat{\boldsymbol{\Sigma}}$. Since $\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2}-\mathbf{I}\|\leq\lambda_{\min}(\boldsymbol{\Sigma})^{-1}\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|$,

$$\mathrm{tr}\left[(\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2}-\mathbf{I})^2\right]\leq d\lambda_{\min}(\boldsymbol{\Sigma})^{-2}\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|^2.$$

$\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}$ can be decomposed into two parts. First consider $\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_t^2-\boldsymbol{\Sigma}=\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'(\tilde{u}_t^2-\sigma_t^2)$ for $\tilde{u}_t=u_t-\mathbb{E}[u_t|\mathbf{b}_t]$ and $\sigma_t^2=\mathbb{E}[\tilde{u}_t^2|\mathbf{b}_t]$. Applying Markov's inequality again, we have with probability over $\mathscr{H}$ at least $1-\frac{d^2}{\bar{\eta}_2T_0}$, $\max_{1\leq j,j'\leq d}\frac{1}{T_0}\sum_{t=1}^{T_0}z_{jt}^2z_{j't}^2\leq 2\bar{\eta}_2$. Then, it follows by Markov's inequality and condition (SA-1.iv) that with probability over $\mathscr{H}$ at least $1-\frac{d^2}{\bar{\eta}_2T_0}$,

$$\mathbb{P}\Big(\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'(\tilde{u}_t^2-\sigma_t^2)\Big\|_{\mathtt{F}}\geq\frac{d}{T_0^{1/2-v}}\Big|\mathscr{H}\Big)\leq\frac{d^2\bar{\eta}_1T_0^{1-2v}}{T_0}\max_{1\leq j,j'\leq d}\frac{1}{T_0}\sum_{t=1}^{T_0}z_{jt}^2z_{j't}^2\leq\frac{2d^2\bar{\eta}_1\bar{\eta}_2}{T_0^{2v}}.$$

Next, note that

$$\begin{aligned}
&(\widehat{u}_t-\widehat{\mathbb{E}}[u_t|\mathbf{b}_t])^2-\tilde{u}_t^2\\
=&\Big((\widehat{u}_t-u_t)-(\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]-\mathbb{E}[u_t|\mathbf{b}_t])\Big)\Big(\widehat{u}_t+u_t-\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]-\mathbb{E}[u_t|\mathbf{b}_t]\Big)\\
=&\Big(\mathbf{z}_t'(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})-(\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]-\mathbb{E}[u_t|\mathbf{b}_t])\Big)\Big(\mathbf{z}_t'(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})+2u_t-\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]-\mathbb{E}[u_t|\mathbf{b}_t]\Big)\\
=:&\,q_t(q_t+2\tilde{u}_t).
\end{aligned}$$

Write $q_t=q_{t,1}+q_{t,2}$, $q_{t,1}=\mathbf{z}_t'(\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}})$, and $q_{t,2}=\widehat{\mathbb{E}}[u_t|\mathbf{b}_t]-\mathbb{E}[u_t|\mathbf{b}_t]$. Then,

$$\begin{aligned}
\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\Big((\widehat{u}_t-\widehat{\mathbb{E}}[u_t|\mathbf{b}_t])^2-\tilde{u}_t^2\Big)\Big\|&=\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'(q_t^2+2q_t\tilde{u}_t)\Big\|\\
&\leq\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'q_t^2\Big\|+2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_tq_t\Big\|.
\end{aligned}$$

For the first term,

$$\begin{aligned}
\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'q_t^2\Big\|&\leq 2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'q_{t,1}^2\Big\|+2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'q_{t,2}^2\Big\|\\
&\leq 2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\|\mathbf{z}_t\|^2\Big\|\|\boldsymbol{\beta}_0-\widehat{\boldsymbol{\beta}}\|^2+2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\Big\|\max_{1\leq t\leq T_0}q_{t,2}^2.
\end{aligned}$$

Using the bound on $\widehat{\boldsymbol{\delta}}$ explained before and condition (SA-1.v), we have with probability over $\mathscr{H}$

at least $1-\mathfrak{C}_1'/T_0-\pi_\delta^\star-\pi_\gamma-\pi_u^\star$, $\mathbb{P}(\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'q_t^2\|\leq\mathfrak{C}_1((\varpi_\delta^\star)^2/T_0+(\varpi_u^\star)^2)|\mathscr{H})\geq 1-\epsilon_\delta^\star-\epsilon_\gamma-\epsilon_u^\star$ for some constants $\mathfrak{C}_1>0$ and $\mathfrak{C}_1'>0$.

For the second term,

$$
\begin{aligned}
2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_tq_t\Big\| &\leq 2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_tq_{t,1}\Big\|+2\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_tq_{t,2}\Big\| \\
&\leq \Big(\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\|\mathbf{z}_t\|\Big\|+\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\|\mathbf{z}_t\|\tilde{u}_t^2\Big\|\Big)\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\| \\
&\quad+\max_{1\leq t\leq T_0}|q_{t,2}|\Big(\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\Big\|+\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{z}_t\mathbf{z}_t'\tilde{u}_t^2\Big\|\Big).
\end{aligned}
$$

Note that we use the fact that for any number a, $2|a|\leq 1+a^2$. Using the previous argument again, we have with probability over $\mathscr{H}$ at least $1-\mathfrak{C}_2'T_0^{-1}-\pi_\delta^\star-\pi_\gamma-\pi_u^\star$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\mathfrak{C}_2''T_0^{-1}-\epsilon_\delta^\star-\epsilon_\gamma-\epsilon_u^\star$, the above is bounded by $\mathfrak{C}_2(\varpi_\delta^\star/\sqrt{T_0}+\varpi_u^\star)$ for some constants $\mathfrak{C}_2$, $\mathfrak{C}_2'$, and $\mathfrak{C}_2''$. Then, the result follows. □

SA-2.7 Proof of Theorem SA-2

Throughout this proof, $\mathfrak{C}_1$, $\mathfrak{c}_1$, $\cdots$ denote some constants independent of T_0. By assumption, the analysis conditional on $\mathscr{H}$ can be replaced by that conditional on $\mathbf{B}$.

(1) We first verify the conditions of Theorem A. Under condition (T4.i), by Theorem 3.1 of Pham and Tran (1985), $\mathbf{u}_t$ is β-mixing at an exponential rate $\mathfrak{b}(k):=\mathfrak{b}(k;\mathscr{H})=\exp(-\mathfrak{c}_1k)$ (even conditional on $\mathscr{H}$). Similarly, $\mathbf{b}_t$ is also β-mixing at an exponential rate $\exp(-\mathfrak{c}_2k)$.

For condition (TA.ii) of Theorem A, let $\rho=\rho_{u,1}\wedge\rho_{u,2}$ and $\sigma^2=\mathbb{E}[\zeta_{t,u,1}^2]\vee\mathbb{E}[\zeta_{t,u,2}^2]$ where $\boldsymbol{\zeta}_{t,u}=(\zeta_{t,u,1},\zeta_{t,u,2})'$. Note that

$$
\begin{aligned}
\bar{\sigma}^2(v) &= \max_{1\leq j\leq d}\frac{1}{T_0vm}\sum_{k=1}^{m}\mathbb{V}\Big[\sum_{t\in\mathcal{J}_k'}(b_{jt,1}u_{t,1}+b_{jt,2}u_{t,2})\Big|\mathbf{B}\Big] \\
&\leq \max_{1\leq j\leq d}\frac{2}{T_0vm}\sum_{l=1}^{2}\sum_{k=1}^{m}\Big\{\sum_{t\in\mathcal{J}_k'}\frac{\sigma^2}{1-\rho^2}b_{jt,l}^2+\sum_{\ell=1}^{v-1}\Big(\sum_{\substack{t-t'=\ell\\ t,t'\in\mathcal{J}_k'}}b_{jt,l}b_{jt',l}\Big)\frac{\rho^\ell\sigma^2}{1-\rho^2}\Big\} \\
&\leq \max_{1\leq j\leq d}\frac{2}{T_0vm}\sum_{l=1}^{2}\sum_{k=1}^{m}\Big\{\sum_{t\in\mathcal{J}_k'}\frac{\sigma^2}{1-\rho^2}b_{jt,l}^2+2\sum_{\ell=1}^{v-1}\Big(\sum_{t\in\mathcal{J}_k'}b_{jt,l}^2\Big)\frac{|\rho|^\ell\sigma^2}{1-\rho^2}\Big\}.
\end{aligned}
$$

It is bounded by $\mathfrak{C}_1 T_0^{-1}$ with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_3 m^{-1}$, since by the coupling argument used in the proof of Theorem A and Markov's inequality,

$$\mathbb{P}\Big(\Big|\frac{1}{vm}\sum_{k=1}^{m}\sum_{t\in\mathcal{J}_k'}(b_{jt,l}^2-\mathbb{E}[b_{jt,l}^2])\Big|\geq\bar{\eta}_1\Big)\leq\frac{1}{\bar{\eta}_1^2 m^2}\sum_{k=1}^{m}\frac{1}{v^2}\mathbb{E}\Big[\Big(\sum_{t\in\mathcal{J}_k'}(b_{jt,l}-\mathbb{E}[b_{jt,l}])\Big)^2\Big]+m\exp(-\mathfrak{c}_2 q)$$

which is bounded by $\mathfrak{c}_4 m^{-1}$ for some $\mathfrak{c}_4$ large enough.

On the other hand, for $\psi=3$, we have

$$\begin{aligned}
\mathbb{E}\Big[\sum_{j=1}^{d}\sum_{k=1}^{m}|S_{jk,\diamond}|^{\psi}\Big|\mathscr{H}\Big] &= \mathbb{E}\Big[\sum_{j=1}^{d}\sum_{k=1}^{m}\Big|\sum_{t\in\mathcal{J}_k'}s_{jt}\Big|^3\Big|\mathscr{H}\Big]\\
&\leq 4mdT_0^{-3/2}\frac{1}{md}\sum_{j=1}^{d}\sum_{k=1}^{m}\sum_{l=1}^{2}\mathbb{E}\Big[\Big|\sum_{t\in\mathcal{J}_k'}b_{jt,l}u_{t,l}\Big|^3|\mathscr{H}\Big]\\
&\leq 4mdT_0^{-3/2}\frac{1}{md}\sum_{j=1}^{d}\sum_{k=1}^{m}\sum_{l=1}^{2}\max\Big\{\sum_{t\in\mathcal{J}_k'}(\mathbb{E}[|b_{jt,l}u_{t,l}|^4|\mathscr{H}])^{\frac{3}{4}},\\
&\qquad\Big(\sum_{t\in\mathcal{J}_k'}(\mathbb{E}[|b_{jt,l}u_{t,l}|^3|\mathscr{H}])^{\frac{2}{3}}\Big)^{\frac{3}{2}}\Big\}.
\end{aligned}$$

The second line uses Hölder's inequality, and the third line uses Rosenthal inequality for strong mixing sequences (see, e.g., Theorem 2 in Section 1.4 of Doukhan (2012)) and the fact that β-mixing implies strong mixing. Again, by the coupling argument and Markov's inequality, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_5 m^{-1}$, the last line is bounded by $\eta_1:=\mathfrak{C}_2 m(v/T_0)^{3/2}$.

For condition (TA.iii), note that by Rosenthal inequality,

$$\begin{aligned}
\max_{1\leq j\leq d}\mathbb{E}\Big[|S_{j(m+1),\diamond}|^3\Big|\mathscr{H}\Big] &\leq \max_{1\leq j\leq d}4T_0^{-3/2}\sum_{l=1}^{2}\mathbb{E}\Big[\Big|\sum_{t\in\mathcal{J}_{m+1}'}b_{jt,l}u_{t,l}\Big|^3|\mathscr{H}\Big]\\
&\leq \max_{1\leq j\leq d}4T_0^{-3/2}\sum_{l=1}^{2}\max\Big\{\sum_{t\in\mathcal{J}_{m+1}'}(\mathbb{E}[|b_{jt,l}u_{t,l}|^4|\mathscr{H}])^{\frac{3}{4}},\\
&\qquad\Big(\sum_{t\in\mathcal{J}_{m+1}'}(\mathbb{E}[|b_{jt,l}u_{t,l}|^3|\mathscr{H}])^{\frac{2}{3}}\Big)^{\frac{3}{2}}\Big\}.
\end{aligned}$$

Applying Markov's inequality to $\sum_{t\in\mathcal{J}_{m+1}'}|b_{jt,l}|^3$ and $\sum_{t\in\mathcal{J}_{m+1}'}|b_{jt,l}|^2$, by the moment conditions imposed, we have $\max_{1\leq j\leq d}\mathbb{E}[|S_{j(m+1),\diamond}|^3|\mathscr{H}]$ is bounded by $\eta_2:=\mathfrak{C}_3(q/T_0)^{3/2}T_0^{\mathfrak{c}_6}$ with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_7 T_0^{-2\mathfrak{c}_6}$ for any small $\mathfrak{c}_6>0$.

For condition (TA.iv), note that by Hölder's inequality and Rosenthal inequality,

$$\sum_{k=1}^{m}\mathbb{E}[\|\mathbf{S}_{k,\square}\|^3|\mathscr{H}] = \frac{1}{\sqrt{T_0}}\frac{m}{T_0}\frac{1}{m}\sum_{k=1}^{m}\mathbb{E}\Big[\Big(\sum_{j=1}^{d}\Big(\sum_{l=1}^{2}\sum_{t\in\mathcal{J}_k}b_{jt,l}u_{t,l}\Big)^2\Big)^{3/2}\Big|\mathscr{H}\Big]$$
$$\leq \frac{4d^{3/2}}{\sqrt{T_0}}\frac{m}{T_0}\frac{1}{m}\sum_{k=1}^{m}\frac{1}{d}\sum_{j=1}^{d}\sum_{l=1}^{2}\mathbb{E}\Big[\Big|\sum_{t\in\mathcal{J}_k}b_{jt,l}u_{t,l}\Big|^3\Big|\mathscr{H}\Big]$$
$$\leq \frac{4\mathfrak{C}_4 d^{3/2}}{\sqrt{T_0}}\frac{m}{T_0}\frac{1}{m}\sum_{k=1}^{m}\frac{1}{d}\sum_{j=1}^{d}\sum_{l=1}^{2}\max\Big\{\sum_{t\in\mathcal{J}_k}(\mathbb{E}[|b_{jt,l}u_{t,l}|^4|\mathscr{H}])^{\frac{3}{4}},$$
$$\Big(\sum_{t\in\mathcal{J}_k}(\mathbb{E}[|b_{jt,l}u_{t,l}|^3|\mathscr{H}])^{\frac{2}{3}}\Big)^{\frac{3}{2}}\Big\}.$$

Again, by the coupling argument and Markov's inequality, the last term is bounded by $\mathfrak{C}_5 m^{-1/2}$ with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_8 m^{-1}$.

On the other hand,

$$\lambda_{\min}(\mathbf{\Sigma}_\square) = \lambda_{\min}\Big(\sum_{k=1}^{m}\mathbb{V}[\mathbf{S}_{k,\square}|\mathscr{H}]\Big) \geq \underline{\eta}_1\lambda_{\min}\Big(\frac{1}{T_0}\sum_{k=1}^{m}\sum_{l=1}^{2}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}\mathbf{b}_{t,l}'\Big)$$

Note that the minimum eigenvalue of $\mathbb{E}[\frac{1}{q}\sum_{l=1}^{2}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}\mathbf{b}_{t,l}'] = \sum_{l=1}^{2}\mathbb{E}[\mathbf{b}_{t,l}\mathbf{b}_{t,l}']$ is bounded from below by $2\underline{\eta}_2$. Then, by the coupling argument, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_9 m^{-1}$, the last term is bounded from below by $\underline{\eta}_1\underline{\eta}_2$. Therefore, we can set $\eta_3 = \mathfrak{C}_6 m^{-1/2}$. The above argument also shows that condition (TA.v) holds with $\eta_4 = \underline{\eta}_1\underline{\eta}_2$.

For condition (TA.vi), note that

$$\mathbf{\Sigma} = \mathbb{V}\Big[\frac{1}{\sqrt{T_0}}\Big(\sum_{l=1}^{2}\sum_{k=1}^{m}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l} + \sum_{l=1}^{2}\sum_{k=1}^{m+1}\sum_{t\in\mathcal{J}_k'}\mathbf{b}_{t,l}u_{t,l}\Big)\Big|\mathscr{H}\Big]$$
$$= \frac{1}{T_0}\mathbb{V}\Big[\sum_{l=1}^{2}\sum_{k=1}^{m}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big] + \frac{1}{T_0}\mathbb{V}\Big[\sum_{l=1}^{2}\sum_{k=1}^{m+1}\sum_{t\in\mathcal{J}_k'}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big]+$$
$$\frac{2}{T_0}\mathbb{C}\mathrm{ov}\Big[\sum_{l=1}^{2}\sum_{k=1}^{m}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l},\ \sum_{l=1}^{2}\sum_{k=1}^{m+1}\sum_{t\in\mathcal{J}_k'}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big]$$
$$= \mathbf{\Sigma}_\square + \frac{2}{T_0}\sum_{\substack{k,k'\\k<k'}}\mathbb{C}\mathrm{ov}\Big[\sum_{l=1}^{2}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l},\ \sum_{l=1}^{2}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big]+$$

$$\frac{1}{T_0}\mathbb{V}\Big[\sum_{l=1}^{2}\sum_{k=1}^{m+1}\sum_{t\in\mathcal{J}_k'}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big]+\frac{2}{T_0}\mathbb{C}\mathrm{ov}\Big[\sum_{l=1}^{2}\sum_{k=1}^{m}\sum_{t\in\mathcal{J}_k}\mathbf{b}_{t,l}u_{t,l},\ \sum_{l=1}^{2}\sum_{k=1}^{m+1}\sum_{t\in\mathcal{J}_k'}\mathbf{b}_{t,l}u_{t,l}\Big|\mathscr{H}\Big]$$

$$=\mathbf{\Sigma}_\square+2\mathrm{I}+\mathrm{II}+2\mathrm{III}.$$

We consider a generic (j,j')th element in $\mathbf{\Sigma}$. For I, the (j,j')th element can be bounded as follows:

$$\begin{aligned}|\mathrm{I}[j,j']|&:=\Big|\frac{1}{T_0}\sum_{\substack{k,k'\\k<k'}}\sum_{\substack{t\in\mathcal{J}_k,\\t'\in\mathcal{J}_{k'}}}\sum_{l=1}^{2}\sum_{l'=1}^{2}b_{jt,l}b_{j't',l'}\mathbb{C}\mathrm{ov}[u_{t,l},u_{t',l'}|\mathscr{H}]\Big|\\&\le\frac{1}{T_0}\sum_{\substack{k,k'\\k<k'}}\sum_{\substack{t\in\mathcal{J}_k\\t'\in\mathcal{J}_{k'}}}\sum_{l=1}^{2}(b_{jt,l}^2+b_{j't',l}^2)|\rho|^{|t-t'|}\sigma^2/(1-\rho^2)\\&=\frac{\sigma^2}{1-\rho^2}\Big\{\frac{1}{T_0}\sum_{k=1}^{m-1}\sum_{t\in\mathcal{J}_k}\Big(\sum_{l=1}^{2}b_{jt,l}^2\Big)\sum_{k':k'>k}\sum_{t'\in\mathcal{J}_{k'}}|\rho|^{|t-t'|}+\\&\qquad\frac{1}{T_0}\sum_{k'=2}^{m}\sum_{t'\in\mathcal{J}_{k'}}\Big(\sum_{l=1}^{2}b_{j't',l}^2\Big)\sum_{k:k<k'}\sum_{t\in\mathcal{J}_k}|\rho|^{|t-t'|}\Big\}.\end{aligned}$$

Note that by the property of geometric series and the moment condition,

$$\mathbb{E}\Big[\frac{1}{T_0}\sum_{k=1}^{m-1}\sum_{t\in\mathcal{J}_k}\Big(\sum_{l=1}^{2}b_{jt,l}^2\Big)\sum_{k':k'>k}\sum_{t'\in\mathcal{J}_{k'}}|\rho|^{|t-t'|}\Big]\le\mathfrak{C}_7|\rho|^v/q.$$

Also note that by coupling argument, the moment condition, and Markov's inequality,

$$\mathbb{P}\Big(\frac{1}{T_0}\sum_{k=1}^{m-1}\sum_{t\in\mathcal{J}_k}\Big(\sum_{l=1}^{2}(b_{jt,l}^2-\mathbb{E}[b_{jt,l}^2])|\rho|^{t-(k-1)q+v+1}\Big)\ge|\rho|^v/q\Big)\le\mathfrak{c}_{10}m^{-1}.$$

The second term in the decomposition of I can be treated similarly. Therefore, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_{11}m^{-1}$, $\|\mathrm{I}\|\le\mathfrak{C}_8|\rho|^v/q$.

Regarding II, the covariance between blocks can be analyzed exactly the same way as above, and the analysis of the sum of variance of each blocks is similar to that of $\bar{\sigma}^2(v)$. It follows that with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_{12}m^{-1}$, $\|\mathrm{II}\|\le\mathfrak{C}_9(v/q+m^{-1}+|\rho|^q/q)$. The analysis of III is similar to that of I, yielding that with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_{13}m^{-1}$, $\|\mathrm{III}\|\le\mathfrak{C}_{10}q^{-1}$. Using these results, we conclude that $\|\mathbf{\Sigma}-\mathbf{\Sigma}_\square\|\le\mathfrak{C}_{11}(v/q+m^{-1})=:\eta_5$ with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_{14}m^{-1}$. Then, condition (TA.vi) holds. Also note that when q/v and m are sufficiently large, $\eta_5 \leq 0.25$. As specified below, this is true when T_0 is large enough.

Finally, to satisfy condition (TA.vii), we can take $m = \mathfrak{C}_{12} T_0^{2/5}$, $v = \mathfrak{C}_{13} T_0^{\mathfrak{c}_{15}}$, and let $\mathfrak{c}_6$ and $\mathfrak{c}_{15}$ be sufficiently small. Then, we can take $\eta_6 = \mathfrak{C}_{14} T_0^{-3/20-\mathfrak{c}_{16}}$ for sufficiently small $\mathfrak{c}_{16}$. The proof for the first part is complete.

(2) Next, we consider Theorem 2. Condition (T2.i) holds by the argument given in part (1) of this proof. As in the proof of Theorem SA-1, conditions (T2.ii) and (T2.iii) of Theorem 2 can be verified using the basic inequality

$$\lambda_{\min}(\widehat{\mathbf{Q}}) \|\boldsymbol{\delta}\|^2 \leq \boldsymbol{\delta} \widehat{\mathbf{Q}} \boldsymbol{\delta} \leq 2\mathbf{G}'\boldsymbol{\delta} \leq 2\|\mathbf{G}\| \|\boldsymbol{\delta}\|.$$

It has been shown in part (1) that the minimum eigenvalue of $\widehat{\mathbf{Q}}$ is bounded from below by some constant with probability over $\mathscr{H}$ at least $1 - \mathfrak{c}_9 m^{-1}$. On the other hand, let $\sigma^2_{\max}$ be the largest diagonal element of $\boldsymbol{\Sigma}$. Note that for Gaussian random variables,

$$\mathbb{P}\Big(\|\mathbf{G}\| \geq \sqrt{2d \log T_0}\, \sigma_{\max} \Big| \mathscr{H} \Big) \leq 2d \exp\Big(- \frac{2\sigma^2_{\max} \log T_0}{2\sigma^2_{\max}} \Big) \leq 2d/T_0.$$

By the argument given in part (1), we have with probability over $\mathscr{H}$ at least $1 - \mathfrak{c}_9 m^{-1} - \mathfrak{c}_{14} m^{-1}$, $\sigma^2_{\max}$ is bounded by some constant. Then, conditions (T2.ii) and (T2.iii) hold.

For condition (T2.iv), since

$$\mathrm{tr}\left[(\boldsymbol{\Sigma}^{-1/2} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1/2} - \mathbf{I})^2 \right] \leq d \lambda_{\min}(\boldsymbol{\Sigma})^{-2} \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|^2.$$

The result follows by the lower bound on the minimum eigenvalue of $\boldsymbol{\Sigma}$ and the assumption imposed in the theorem. □

SA-2.8 Proof of Theorem SA-3

Throughout this proof, $\mathfrak{C}_1$, $\mathfrak{c}_1, \cdots$ denote some constants independent of T_0. First note that given the assumptions on $\{u_{t,l}\}$ and $\{v_{jt,l}\}$, conditional on $\mathbf{B}$, $\mathbf{p}_T$ is independent of the pre-treatment data. Therefore, when we verify the conditions of Theorems 1 and 2, the analysis conditional on $\mathscr{H}$ is the same as that conditional on $\mathbf{B}$ and $\mathbf{C}$ only.

(1) We consider Theorem 1 first. Condition (T1.i) directly follows from the assumptions on $\mathbf{q}_t$.

For condition (T1.ii), as in the proof of Theorem SA-1, we only need to bound $\frac{1}{T_0^{3/2}}\sum_{t=1}^{T_0}\sum_{l=1}^{M}\|\check{\mathbf{z}}_{t,l}\|^3$ and $\|\mathbf{\Sigma}^{-1}\|$. First consider $\frac{1}{T_0}\sum_{t=1}^{T_0}\sum_{l=1}^{M}\|\check{\mathbf{z}}_{t,l}\|^3$. Note that

$$\frac{1}{T_0}\sum_{t=1}^{T_0}\sum_{l=1}^{M}\|\check{\mathbf{z}}_{t,l}\|^3 \le \frac{\sqrt{d}}{T_0}\sum_{t=1}^{T_0}\sum_{l=1}^{M}\sum_{j=1}^{d}|\check{z}_{jt,l}|^3,$$

where $\check{z}_{jt,l}$ is the jth element of $\check{\mathbf{z}}_{t,l}$.

For the stationary components, i.e., $\{c_{kt,l}\}$, since the third moment is bounded, by Markov's inequality, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_1 T_0^{-1}$, $\frac{1}{T_0}\sum_{l=1}^{M}\sum_{j=J+1}^{d}\sum_{t=1}^{T_0}|\check{z}_{jt,l}|^3 \le \mathfrak{C}_1$.

On the other hand, note that the non-stationary components $\check{\mathbf{z}}_{t,\triangleleft,l}$ can be understood as a multivariate partial sum process indexed by t. Write $\check{\mathbf{Z}}_{\triangleleft}(t) = (\check{\mathbf{z}}'_{t,\triangleleft,1},\cdots,\check{\mathbf{z}}'_{t,\triangleleft,M})'$. By strong approximation of partial sum processes, for any $0<\nu<1/2-1/\psi$,

$$\mathbb{P}\Big(\max_{1\le t\le T_0}\|\check{\mathbf{Z}}_{\triangleleft}(t)-\mathbf{G}(t/T_0)\| \ge T_0^{-\frac{1}{2}+\frac{1}{\psi}+\nu}\Big) \le \mathfrak{C}_2 T_0^{-\psi\nu},$$

where $\mathbf{G}(\cdot) = (\mathbf{G}_1(\cdot)',\cdots,\mathbf{G}_M(\cdot)')'$ is a (JM)-dimensional Brownian motion on $[0,1]$ with the variance $\mathbb{E}[\mathbf{v}_t\mathbf{v}_t']$. On the other hand, it is well known that for each $1\le j\le J$, $1\le l\le M$, for any $m>0$,

$$\mathbb{P}(\max_{0\le r\le 1}|G_{j,l}(r)|>m) \le 2\mathbb{P}(\max_{0\le r\le 1}G_{j,l}(r)>m) = 2\mathbb{P}(|G_{j,l}(1)|>m),$$

where $G_{j,l}(\cdot)$ is the jth element of $\mathbf{G}_l(\cdot)$. Using the tail bound for Gaussian distributions, we can set $m=\sqrt{2\log(2JMT_0)\sigma^2_{\max}}$ where $\sigma^2_{\max}$ is the largest variance of $\{v_{jt,l}: 1\le j\le J, 1\le l\le M\}$, which leads to $\max_{0\le r\le 1}|G_{j,l}(r)|\le m$ with probability over $\mathscr{H}$ at least $1-(JMT_0)^{-1}$. Therefore, we conclude that

$$\frac{1}{T_0}\sum_{t=1}^{T_0}\sum_{l=1}^{M}\sum_{j=1}^{J}|\check{z}_{jt,l}|^3 \le \mathfrak{C}_3\Big(\sqrt{2\log(2JMT_0)\sigma^2_{\max}}+\mathfrak{C}_2 T_0^{\frac{1}{\psi}-\frac{1}{2}+\nu}\Big)^3$$

with probability over $\mathscr{H}$ at least $1-T_0^{-1}-\mathfrak{c}_2 T_0^{-\psi\nu}$.

Finally, we consider $\mathbf{\Sigma}$. By assumption in the theorem, $\lambda_{\min}(\mathbf{\Sigma}) \ge \underline{\eta}_1\widehat{\mathbf{Q}}$. Partition $\widehat{\mathbf{Q}}$ into

$$\begin{bmatrix}\widehat{\mathbf{Q}}_{11} & \widehat{\mathbf{Q}}_{12}\\ \widehat{\mathbf{Q}}_{12} & \widehat{\mathbf{Q}}_{22}\end{bmatrix}.$$

$\widehat{\mathbf{Q}}_{11} \in \mathbb{R}^{J\times J}$ corresponds to the Gram of the non-stationary component:

$$\widehat{\mathbf{Q}}_{11} = \sum_{l=1}^{M}\Big(\frac{1}{T}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\triangleleft,l}\check{\mathbf{z}}_{t,\triangleleft,l}'\Big).$$

Again, by strong approximation used previously, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_2T_0^{-\psi\nu}-T_0^{-1}$,

$$\Big\|\frac{1}{T_0}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\triangleleft,l}\check{\mathbf{z}}_{t,\triangleleft,l}' - \frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{G}_l\Big(\frac{t}{T_0}\Big)\mathbf{G}_l\Big(\frac{t}{T_0}\Big)'\Big\| \leq \mathfrak{C}_4T_0^{\frac{1}{\psi}-\frac{1}{2}+\nu}\sqrt{\log T_0}.$$

Then, by the condition in the theorem, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_2T_0^{-\psi\nu}-T_0^{-1}-\pi_{Q,1}$, $\lambda_{\min}(\widehat{\mathbf{Q}}_{11})\geq(\log T_0)^{-\mathfrak{c}_Q}/2$ for T_0 large enough so that $\mathfrak{C}_4T_0^{\frac{1}{\psi}-\frac{1}{2}+\nu}\sqrt{\log T_0}\leq(\log T_0)^{-\mathfrak{c}_Q}/2$.

Note that $\lambda_{\min}(\sum_{l=1}^{M}\mathbb{E}[\check{\mathbf{z}}_{t,\triangleright,l}\check{\mathbf{z}}_{t,\triangleright,l}])$ is bounded away from zero. Then, by Markov's inequality and the argument used in the proof of Theorem SA-1, $\lambda_{\min}(\widehat{\mathbf{Q}}_{22})\geq\mathfrak{C}_5$ with probability at least $1-\mathfrak{c}_3T_0^{-1}$.

For off-diagonal blocks, note that

$$\widehat{\mathbf{Q}}_{12} = \sum_{l=1}^{M}\Big\{\Big(\frac{1}{T_0}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\triangleleft,l}(\mathbb{E}[\check{\mathbf{z}}_{t,\triangleright,l}])'\Big)+\Big(\frac{1}{T_0}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\triangleleft,l}(\check{\mathbf{z}}_{t,\triangleright,l}-\mathbb{E}[\check{\mathbf{z}}_{t,\triangleright,l}])'\Big)\Big\}.$$

The first term, by assumption, is zero. For the second term, by the condition in the theorem, $\|\frac{1}{T_0}\sum_{t=1}^{T_0}\check{\mathbf{z}}_{t,\triangleleft,l}(\check{\mathbf{z}}_{t,\triangleright,l}-\mathbb{E}[\check{\mathbf{z}}_{t,\triangleright,l}])'\|\leq\mathfrak{C}_6T_0^{-1/2+\nu_Q}$ with probability over $\mathscr{H}$ at least $1-\pi_{Q,2}$. Thus, we have with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_2T_0^{-\psi\nu}-\mathfrak{c}_4T_0^{-1}-\pi_{Q,1}-\pi_{Q,2}$, $\lambda_{\min}(\widehat{\mathbf{Q}})\geq\mathfrak{C}_7(\log T_0)^{-\mathfrak{c}_Q}/2$. Therefore, we can take $\epsilon_\gamma=\mathfrak{C}_\epsilon(\log T_0)^{\frac{3}{2}(1+\mathfrak{c}_Q)}T_0^{-1/2}$ for some non-negative finite constant $\mathfrak{C}_\epsilon$.

(ii) Next, we consider Theorem 2. Condition (T2.i) holds by part (1). As in the proof of Theorem SA-1, condition (T2.ii) can be established using the results given in the previous step, the Gaussian tail bound, and the basic inequality

$$\lambda_{\min}(\widehat{\mathbf{Q}})\|\boldsymbol{\delta}\|^2\leq\boldsymbol{\delta}\widehat{\mathbf{Q}}\boldsymbol{\delta}\leq 2\mathbf{G}'\boldsymbol{\delta}\leq 2\|\mathbf{G}\|\|\boldsymbol{\delta}\|.$$

Recall that $\lambda_{\min}(\widehat{\mathbf{Q}})\geq\mathfrak{C}_7(\log T_0)^{-\mathfrak{c}_Q}$ with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_2T_0^{-\psi\nu}-\mathfrak{c}_4T_0^{-1}-\pi_{Q,1}-\pi_{Q,2}$. On the other hand, by a similar argument, with probability over $\mathscr{H}$ at least $1-\mathfrak{c}_2T_0^{-\psi\nu}-\mathfrak{c}_4T_0^{-1}-\pi_{Q,1}-\pi_{Q,2}$, $\lambda_{\max}(\widehat{\mathbf{Q}})\leq(\log T_0)^{\mathfrak{c}_Q}$. Then, $\varpi_\delta^\star=\mathfrak{C}_8(\log T_0)^{2\mathfrak{c}_Q+0.5}$, $\pi_\delta^\star=\mathfrak{c}_2T_0^{-\psi\nu}-\mathfrak{c}_4T_0^{-1}-$

$\pi_{Q,1} - \pi_{Q,2}$, and $\epsilon^\star_\delta = \mathfrak{C}_9 T_0^{-1}$. Condition (T2.iii) holds by the construction of the thresholding rule and condition (SA-5.v). Finally, for condition (T2.iv), note that

$$\mathrm{tr}\left[(\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2} - \mathbf{I})^2\right] \leq d\lambda_{\min}(\mathbf{\Sigma})^{-2}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|^2.$$

Since $\lambda_{\min}(\mathbf{\Sigma}) \geq \underline{\eta}_1 \lambda_{\min}(\widehat{\mathbf{Q}}) \geq \underline{\eta}_1 \mathfrak{C}_7 (\log T_0)^{-\mathfrak{c}_Q}$ with probability over $\mathscr{H}$ at least $1 - \pi^\star_\delta$, the result follows. □

SA-3 Additional Simulation Evidence

The data generating processes have been described in Section 6.1 of the main paper. We consider models with misspecification error and make use of different methods to estimate the conditional mean, variance and quantiles of u_t given $\mathscr{H}$ whenever needed. Specifically, Tables SA-1 and SA-2 are based on zero-order ("constant") and second-order ("quadratic") polynomial regression methods. The proposed prediction intervals generally perform well with high coverage probability, though they are very conservative in several cases. Note that in models with $\rho = 1$, conditional coverage is more difficult to achieve, since by construction of the evaluation points, we introduce potentially large "shocks" on the control outcomes and thus on the out-of-sample error e_T on purpose. As expected, the constant regression fails to correct the misspecification error, thus leading to very poor coverage. Also, compared with the results based on linear regression reported in the main paper, the prediction intervals based on quadratic regression do not perform well, probably due to the overfitting issue.

For comparison, we also include two other prediction intervals: "PERM" denotes the prediction interval based on the cross-sectional permutation method proposed in Abadie et al. (2010), and "CONF" denotes the prediction interval based on the conformal method developed in Chernozhukov et al. (2021). As expected, the prediction interval "PERM" provides much lower actual coverage probability than the nominal level, since it relies on a cross-sectional permutation and does not apply to the causal inference framework considered in this paper. In contrast, the conformal prediction interval developed in Chernozhukov et al. (2021) performs better, though its actual coverage probability is still lower than the target nominal level in general.

SA-4 Empirical Example: 1990 German Reunification

In this example, we analyze the economic impact on West Germany of 1990 German reunification (see Abadie, 2021, for more details). The key variable of interest is real per capita GDP of West Germany. We first consider the raw data of per capita GDP. Figure SA-1(a) shows the per capita GDP of the synthetic West Germany (dashed blue) and the actual West Germany (solid black). The synthetic control prediction is able to closely approximate the observed trajectory of per capita GDP of West Germany prior to the treatment, which can be expected since the data are non-stationary and may contain (deterministic or stochastic) common trends. After 1990, the synthetic West Germany is above the actual one, suggesting a negative shock on West Germany after reunification. Figure SA-1(b) adds a 95% conservative prediction interval that takes into account the in-sample uncertainty due to the estimated SC weights, and we add the uncertainty associated with the out-of-sample error e_T in Figures SA-1(c)-(e). The constructed PIs for the counterfactual outcome of West Germany are not always separated from the observed sequence. To further assess the robustness of the result, we present the sensitivity analysis of the effect in 1993 in Figure SA-1(f). The constructed PIs cover the observed per capita GDP of West Germany, unless we shrink the estimated (conditional) standard deviation of e_T by a factor of 0.25.

Then, we apply our methods to the (log) GDP growth rate time series. The raw data are transformed by taking the (log) difference operator. Figure SA-2(a) shows that the growth rate of per capita GDP of the synthetic West Germany is above that of the actual West Germany after 1990, suggesting a negative economic shock on West Germany after reunification. The PIs in Figures SA-2(c)-(e) cover the observed sequence for most post-treatment periods, which do not support statistically significant (negative) effects of reunification on West Germany. Figure SA-2(f) shows the sensitivity analysis of the effect in 1993. We can see that, though the PI constructed based on approach 1 in Figure SA-2(c) appears to suggest a significant effect, this result is not very robust. The corresponding PI covers the observed GDP growth rate of West Germany if we inflate the estimated (conditional) standard deviation of e_T by a factor of 1.5.

References

Abadie, A. (2021), "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, forthcoming.

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American statistical Association*, 105, 493–505.

Berbee, H. (1987), "Convergence Rates in the Strong Law for Bounded Mixing Sequences," *Probability Theory and Related Fields*, 74, 255–270.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2015), "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70.

Chernozhukov, V., Chetverikov, D., Kato, K. et al. (2017), "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45, 2309–2352.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021), "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*.

Doukhan, P. (2012), *Mixing: Properties and Examples*, Springer.

Pham, T. D., and Tran, L. T. (1985), "Some Mixing Properties of Time Series Models," *Stochastic Processes and their Applications*, 19, 297–303.

Raič, M. (2019), "A Multivariate Berry–Esseen Theorem with Explicit Constants," *Bernoulli*, 25, 2824–2853.

Table SA-1: Simulation, Misspecification Error, Constant Regression Methods

	M1		**M1-S**		**M2**		**M3**		**PERM**		**CONF**	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
$\rho = 0$												
Cond. 1	0.965	2.211	0.995	2.939	0.987	2.654	0.987	2.654	0.299	0.646	0.864	1.646
2	0.962	2.155	0.994	2.883	0.986	2.598	0.986	2.598	0.270	0.763	0.854	1.642
3	0.961	2.165	0.994	2.893	0.985	2.608	0.985	2.608	0.230	0.891	0.842	1.643
4	0.969	2.260	0.996	2.989	0.988	2.704	0.988	2.704	0.195	1.026	0.830	1.651
5	0.975	2.410	0.997	3.139	0.991	2.853	0.991	2.853	0.162	1.166	0.819	1.665
Uncond.	0.971	2.434	0.995	3.164	0.988	2.881	0.988	2.881	0.411	0.829	0.886	1.687
$\rho = 0.5$												
Cond. 1	0.980	2.493	0.996	3.221	0.993	2.934	0.993	2.934	0.224	0.930	0.854	1.680
2	0.983	2.562	0.997	3.290	0.993	3.003	0.993	3.003	0.270	0.792	0.865	1.691
3	0.985	2.655	0.998	3.383	0.995	3.095	0.995	3.095	0.317	0.678	0.875	1.706
4	0.989	2.768	0.999	3.496	0.995	3.208	0.995	3.208	0.363	0.596	0.887	1.729
5	0.991	2.898	0.999	3.626	0.998	3.338	0.998	3.338	0.408	0.548	0.896	1.756
Uncond.	0.975	2.471	0.996	3.207	0.989	2.920	0.989	2.920	0.467	1.015	0.882	1.695
$\rho = 1$												
Cond. 1	0.990	5.049	1.000	6.474	1.000	6.166	1.000	6.166	1.000	11.253	0.896	3.403
2	1.000	4.842	1.000	6.267	1.000	5.959	1.000	5.959	0.869	11.283	0.985	3.369
3	0.849	4.687	0.989	6.112	0.958	5.804	0.958	5.804	0.004	11.488	0.170	3.345
4	0.041	4.583	0.332	6.008	0.172	5.699	0.172	5.699	0.000	11.934	0.000	3.333
5	0.000	4.545	0.001	5.970	0.000	5.661	0.000	5.661	0.000	12.575	0.000	3.336
Uncond.	0.989	5.632	0.999	7.083	0.995	6.512	0.995	6.512	0.922	17.377	0.895	3.443

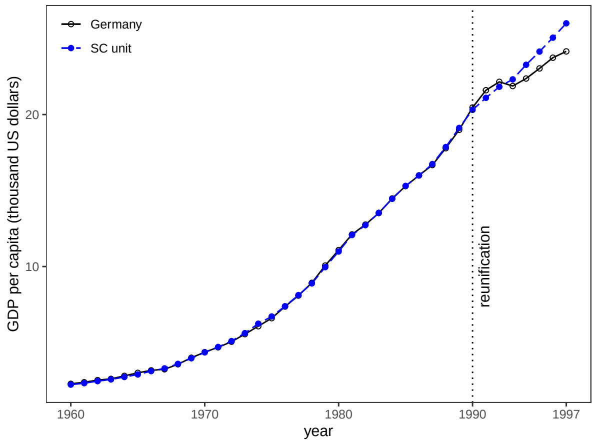
Notes. Conditional mean, variance and quantiles of u_t are estimated based on constant regression methods. CP = coverage probability, AL = average length. "M1": prediction interval for $Y_{1T}(0)$ based on the Gaussian concentration inequality with 90% nominal coverage probability; "M1-S": the same as "M1", but the estimated standard deviation is doubled in the construction; "M2": prediction interval for $Y_{1T}(0)$ based on the location-scale model with 90% nominal coverage probability; "M3": prediction interval for $Y_{1T}(0)$ based on quantile regression with 90% nominal coverage probability; "PERM": prediction interval for $Y_{1T}(0)$ based on the permutation method proposed in Abadie, Diamond and Hainmueller (2010) with 90% nominal coverage probability; "CONF" prediction interval for $Y_{1T}(0)$ based on the conformal method developed in Chernozhukov, Wüthrich and Zhu (2021) with 90% nominal coverage probability.

Table SA-2: Simulation, Misspecification Error, Quadratic Regression Methods

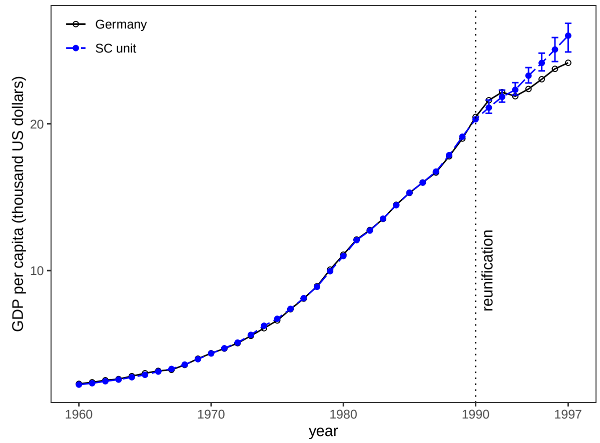
	M1		**M1-S**		**M2**		**M3**		**PERM**		**CONF**	
	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
$\rho = 0$												
Cond. 1	0.929	2.141	0.975	2.866	0.963	2.600	0.951	2.588	0.299	0.646	0.864	1.646
2	0.909	2.083	0.967	2.806	0.947	2.541	0.929	2.495	0.270	0.763	0.854	1.642
3	0.889	2.084	0.956	2.805	0.934	2.542	0.899	2.426	0.230	0.891	0.842	1.643
4	0.881	2.175	0.941	2.899	0.925	2.637	0.863	2.398	0.195	1.026	0.830	1.651
5	0.876	2.342	0.931	3.082	0.914	2.818	0.804	2.384	0.162	1.166	0.819	1.665
Uncond.	0.947	2.350	0.979	3.082	0.971	2.814	0.955	2.728	0.411	0.829	0.886	1.687
$\rho = 0.5$												
Cond. 1	0.926	2.347	0.974	3.054	0.964	2.792	0.942	2.741	0.224	0.930	0.854	1.680
2	0.941	2.402	0.978	3.106	0.968	2.845	0.948	2.789	0.270	0.792	0.865	1.691
3	0.945	2.488	0.980	3.195	0.970	2.934	0.941	2.824	0.317	0.678	0.875	1.706
4	0.947	2.613	0.976	3.331	0.967	3.068	0.918	2.844	0.363	0.596	0.887	1.729
5	0.942	2.788	0.969	3.536	0.961	3.266	0.871	2.845	0.408	0.548	0.896	1.756
Uncond.	0.944	2.356	0.980	3.083	0.968	2.822	0.947	2.760	0.467	1.015	0.882	1.695
$\rho = 1$												
Cond. 1	0.914	3.071	0.946	4.033	0.937	3.685	0.716	2.608	1.000	11.253	0.896	3.403
2	0.950	2.875	0.976	3.791	0.969	3.453	0.900	3.028	0.869	11.283	0.985	3.369
3	0.880	3.284	0.923	4.444	0.907	4.038	0.742	2.891	0.004	11.488	0.170	3.345
4	0.717	6.925	0.768	9.935	0.751	9.157	0.507	2.199	0.000	11.934	0.000	3.333
5	0.581	256.704	0.614	384.612	0.603	392.478	0.374	0.962	0.000	12.575	0.000	3.336
Uncond.	0.962	2.940	0.986	3.713	0.979	3.433	0.959	3.372	0.922	17.377	0.895	3.443

Notes. Conditional mean, variance and quantiles of u_t are estimated based on quadratic regression methods. CP = coverage probability, AL = average length. "M1": prediction interval for $Y_{1T}(0)$ based on the Gaussian concentration inequality with 90% nominal coverage probability; "M1-S": the same as "M1", but the estimated standard deviation is doubled in the construction; "M2": prediction interval for $Y_{1T}(0)$ based on the location-scale model with 90% nominal coverage probability; "M3": prediction interval for $Y_{1T}(0)$ based on quantile regression with 90% nominal coverage probability; "PERM": prediction interval for $Y_{1T}(0)$ based on the permutation method proposed in Abadie, Diamond and Hainmueller (2010) with 90% nominal coverage probability; "CONF" prediction interval for $Y_{1T}(0)$ based on the conformal method developed in Chernozhukov, Wüthrich and Zhu (2021) with 90% nominal coverage probability.

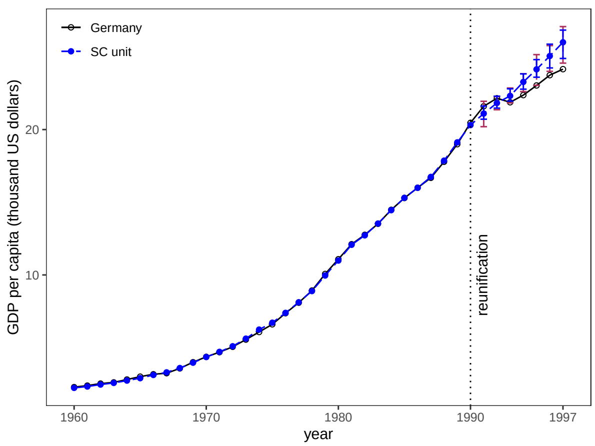
Figure SA-1: 1990 German Reunification: GDP Per Capita.



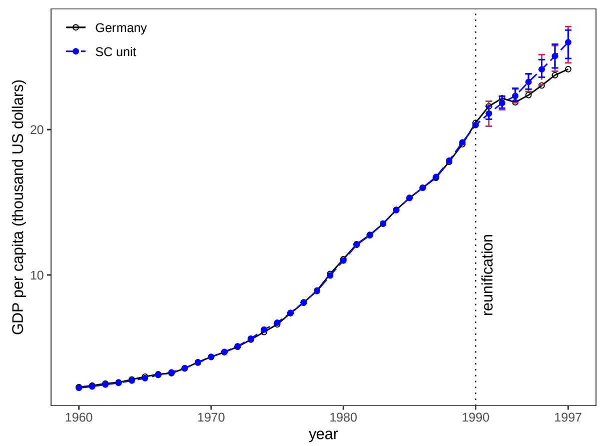
(a) Synthetic West Germany



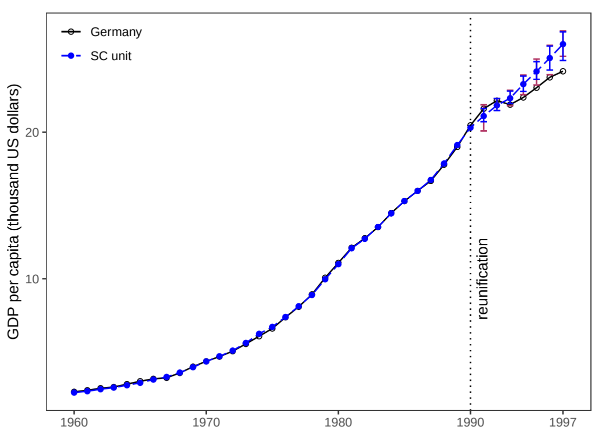
(b) Prediction Interval for $\mathbf{x}_T'\mathbf{w}_0$



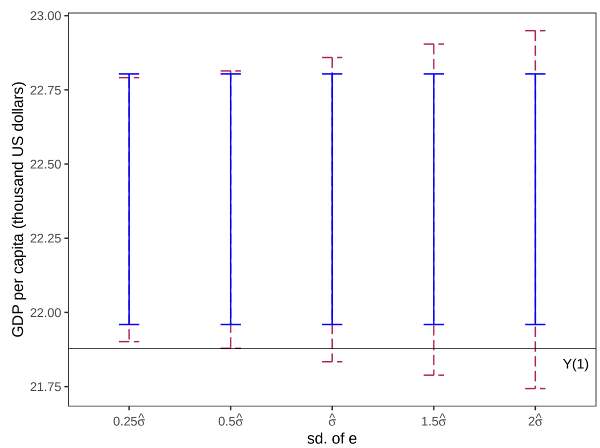
(c) Prediction Interval for $Y_{1T}(0)$, approach 1



(d) Prediction Interval for $Y_{1T}(0)$, approach 2



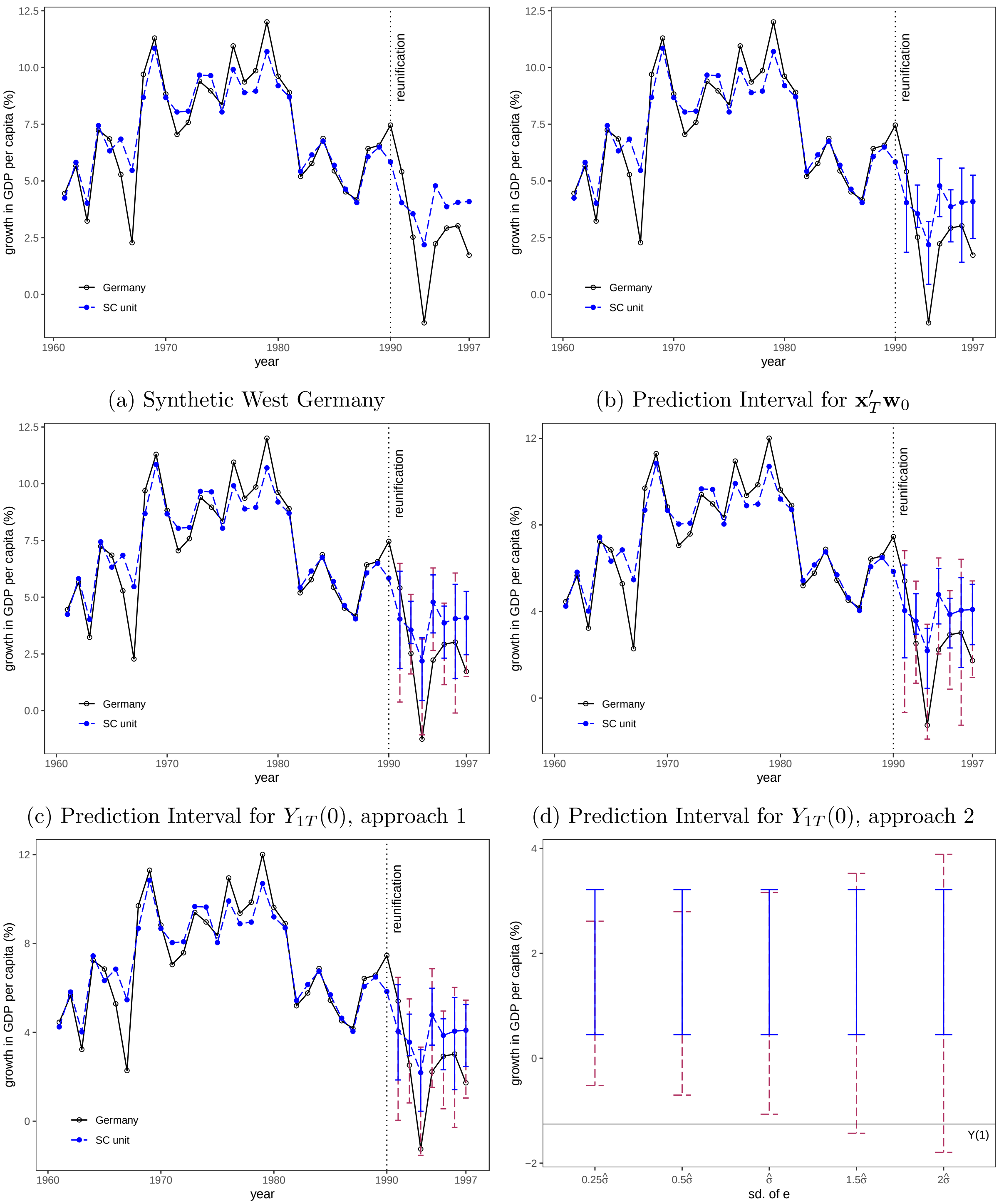
(e) Prediction Interval for $Y_{1T}(0)$, approach 3



(f) Sensitivity Analysis: PIs in 1993

Notes. Panel (a): GDP per capita of West Germany and synthetic West Germany. Panel (b): Prediction interval for synthetic West Germany with at least 95% coverage probability. Panels (c)-(e): Prediction intervals for the counterfactual of West Germany with at least 90% coverage probability based on three methods described in Section 5, respectively. Panel (f): Prediction intervals for the counterfactual West Germany based on approach 1, corresponding to $c \times \sigma_{\mathscr{H}}$, where $c = 0.25, 0.5, 1, 1.5, 2$. The horizontal solid line represents the observed outcome for the treated.

Figure SA-2: 1990 German Reunification: GDP Growth Rate.



(a) Synthetic West Germany

(b) Prediction Interval for $\mathbf{x}_T'\mathbf{w}_0$

(c) Prediction Interval for $Y_{1T}(0)$, approach 1

(d) Prediction Interval for $Y_{1T}(0)$, approach 2

(e) Prediction Interval for $Y_{1T}(0)$, approach 3

(f) Sensitivity Analysis: PIs in 1993

Notes. Panel (a): GDP growth rate of West Germany and synthetic West Germany. Panel (b): Prediction interval for synthetic West Germany with at least 95% coverage probability. Panels (c)-(e): Prediction intervals for the counterfactual of West Germany with at least 90% coverage probability based on three methods described in Section 5, respectively. Panel (f): Prediction intervals for the counterfactual West Germany based on approach 1, corresponding to $c \times \sigma_{\mathscr{H}}$, where $c = 0.25, 0.5, 1, 1.5, 2$. The horizontal solid line represents the observed outcome for the treated.