Supplement to "Uncertainty Quantification in Synthetic Controls with Staggered Treatment Adoption"

Matias D. Cattaneo[*] Yingjie Feng[†] Filippo Palomba[‡] Rocío Titiunik[§]

March 8, 2025

This supplement contains a general framework encompassing the results presented in the paper, technical proofs, additional methodological results, other details about data preparation and implementation, and further empirical evidence.

[*]Department of Operations Research and Financial Engineering, Princeton University.
[†]School of Economics and Management, Tsinghua University.
[‡]Department of Economics, Princeton University.
[§]Department of Politics, Princeton University.

Contents

S.1 Notation

We summarize the notation used throughout the paper in the following table.

Table S.1: *Summary of Notation*

Quantity	**Description**
τ_{ik}	time-specific unit-specific predictand
$\tau_{i\cdot}$	time-averaged unit-specific predictand
$\tau_{\mathcal{Q}k}$	time-specific unit-averaged predictand
$\tau_{\cdot\cdot}$	time-averaged unit-averaged predictand
$\mathcal{N}$	set of never-treated units
$\mathcal{E}$	set of ever-treated units
$\mathcal{W}^{[i]}$ or $\mathcal{W}$	constraint set imposed in SC construction
J_0	number of never-treated units
J_1	number of ever-treated units
Q	number of units in $\mathcal{Q}$
Y_{it}	real GDP per capita of country i in year t
$\mathbf{Y}_{\mathcal{N}t}$	a vector of outcomes of never-treated (donor) units
u_{it}	out-of-sample error in decomposition of $\hat{Y}_{it}$
$\varrho_\ell^{[i]}$	tuning parameter used to check if the ℓ-th constraint is binding
$\mathcal{M}_{\widehat{\gamma}}^{[i]}$ or $\mathcal{M}_{\widehat{\gamma}}$	feasible set for SC weights deduced from optimization
$\mathcal{M}_{\mathbf{G}}^{[i]\star}$ or $\mathcal{M}_{\mathbf{G}}^{\star}$	feasible set used for in-sample uncertainty quantification
$\Delta^{[i]}$ or Δ	centered constraint set
$\mathrm{dist}(\mathbf{a},\mathcal{A})$	distance between a point $\mathbf{a}$ and a set $\mathcal{A}$, i.e., $\inf_{\boldsymbol{\lambda}\in\mathcal{A}}\|\mathbf{a}-\boldsymbol{\lambda}\|$
$\mathcal{A}_\varepsilon$	ε-enlargement of $\mathcal{A}$, i.e., $\{\mathbf{a} : \mathrm{dist}(\mathbf{a},\mathcal{A})\leq\varepsilon\}$
$\lvert\mathcal{A}\rvert$	number of elements in a set $\mathcal{A}$
$\times_{i\in\mathcal{S}}\mathcal{A}_i$	Cartesian product of sets $\mathcal{A}_i$'s for $i\in\mathcal{S}$
$\|\cdot\|_1$, $\|\cdot\|_2$	L1 and L2 vector norms
$s_{\min}(\mathbf{A})$, $s_{\max}(\mathbf{A})$	minimum and maximum singular values of a matrix $\mathbf{A}$
$(v_j : j\in\mathcal{A})$	a vector consisting of all v_j's with $j\in\mathcal{A}$
$\otimes$	Kronecker product
subscripts `in`, `out`	"in-sample", "out-of-sample"
superscript $[i]$	indicate a quantity is related to a treated unit i
lower bar, upper bar	"lower bound", "upper bound"
$\star$	indicate a quantity is related to simulation

S.2 Theoretical Foundations

This section presents a general framework that can accommodate more flexible specifications in SC analysis. As in the paper, we still consider the case with J_0 never-treated units and J_1 ever-treated units that adopt a treatment at possibly different times. However, now we assume that a user may want to obtain SC weights by matching on M features (denoted by a subscript $l = 1, \cdots, M$ below) with additional covariates adjustment, rather than relying solely on pre-treatment outcomes.

Specifically, let $\mathbf{A}_l^{[i]} = (a_{1,l}^{[i]}, \cdots, a_{T_{i0},l}^{[i]})' \in \mathbb{R}^{T_{i0}}$ be the l-th feature of the treated unit i measured in T_{i0} (user-specified) pre-treatment periods. For each feature l and each treated unit i, there exist $J_0 + K$ variables that are used to predict or match the T_{i0}-dimensional vector $\mathbf{A}_l^{[i]}$. These $J_0 + K$ variables are separated into two groups denoted by $\mathbf{B}_l^{[i]} = (\mathbf{B}_{1,l}^{[i]}, \mathbf{B}_{2,l}^{[i]}, \cdots, \mathbf{B}_{J_0,l}^{[i]}) \in \mathbb{R}^{T_{i0}\times J_0}$ and $\mathbf{C}_l^{[i]} = (\mathbf{C}_{1,l}^{[i]}, \cdots, \mathbf{C}_{K,l}^{[i]}) \in \mathbb{R}^{T_{i0}\times K}$, respectively. More precisely, for each $j = 1, \ldots, J_0$, $\mathbf{B}_{j,l}^{[i]} =$

$(b^{[i]}_{j1,l}, \cdots, b^{[i]}_{jT_{i0},l})'$ corresponds to the l-th feature of the j-th unit in the donor pool measured in T_{i0} pre-treatment periods, and for each $k = 1, \dots, K$, $\mathbf{C}^{[i]}_{k,l} = (c^{[i]}_{k1,l}, \cdots, c^{[i]}_{kT_{i0},l})'$ is another vector of control variables used to predict $\mathbf{A}^{[i]}_l$ over the same pre-intervention time span. Stacking the M equations (corresponding to M features) for each treated unit, we define

$$\underbrace{\mathbf{A}^{[i]}}_{T_{i0}\cdot M\times 1} = \begin{bmatrix}\mathbf{A}^{[i]}_1\\ \vdots \\ \mathbf{A}^{[i]}_M\end{bmatrix}, \quad \underbrace{\mathbf{B}^{[i]}}_{T_{i0}\cdot M\times J_0} = \begin{bmatrix}\mathbf{B}^{[i]}_1\\ \vdots \\ \mathbf{B}^{[i]}_M\end{bmatrix}, \quad \underbrace{\mathbf{C}^{[i]}}_{T_{i0}\cdot M\times K\cdot M} = \begin{bmatrix}\mathbf{C}^{[i]}_1 & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{[i]}_2 & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}^{[i]}_M\end{bmatrix}.$$

For instance, in Section S.9 we revisit our empirical application, where $\mathbf{A}^{[i]}$ contains the (log) GDP per capita and the investment-to-GDP ratio ($M = 2$) of an ever-liberalized economy i during the pre-liberalization period, and $\mathbf{B}^{[i]}$ contains the same two features of the donor economies used to match $\mathbf{A}^{[i]}$. For each feature $l = 1, 2$, $\mathbf{C}^{[i]}_l$ contains an intercept and a linear time trend ($K = 2$).

We search for a vector of weights $\mathbf{w} = (\mathbf{w}^{[1]\prime}, \cdots, \mathbf{w}^{[J_1]\prime})' \in \mathcal{W} \subseteq \mathbb{R}^{J_0 J_1}$, which is common across the M features, and a vector of coefficients $\mathbf{r} = (\mathbf{r}^{[1]\prime}, \cdots, \mathbf{r}^{[J_1]\prime})' \in \mathcal{R} \subseteq \mathbb{R}^{KMJ_1}$, such that the linear combination of $\mathbf{B}^{[i]}$ and $\mathbf{C}^{[i]}$ matches $\mathbf{A}^{[i]}$ as closely as possible, for all $i \in \mathcal{E}$. The feasibility sets $\mathcal{W}$ and $\mathcal{R}$ capture the restrictions imposed. Analogously to Equation (3.1) in the paper, such SC weights are obtained via the following optimization problem: for some $(\tilde{T}\cdot M)\times(\tilde{T}\cdot M)$ symmetric weighting matrix $\mathbf{V}$ with $\tilde{T} = \sum_{i=1}^{J_1} T_{0i}$,

$$\widehat{\boldsymbol{\beta}} := (\widehat{\mathbf{w}}', \ \widehat{\mathbf{r}}')' \in \underset{\mathbf{w}\in\mathcal{W},\,\mathbf{r}\in\mathcal{R}}{\arg\min}\ (\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r})'\mathbf{V}(\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r}) \tag{S.2.1}$$

where

$$\underbrace{\mathbf{A}}_{\tilde{T}\cdot M\times 1} = \begin{bmatrix}\mathbf{A}^{[1]}\\ \vdots \\ \mathbf{A}^{[J_1]}\end{bmatrix}, \quad \underbrace{\mathbf{B}}_{\tilde{T}\cdot M\times J_0\cdot J_1} = \begin{bmatrix}\mathbf{B}^{[1]} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{B}^{[2]} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}^{[J_1]}\end{bmatrix}, \quad \underbrace{\mathbf{C}}_{\tilde{T}\cdot M\times K\cdot M\cdot J_1} = \begin{bmatrix}\mathbf{C}^{[1]} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^{[2]} & \cdots & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{C}^{[J_1]}\end{bmatrix}.$$

Accordingly, we write $\widehat{\mathbf{w}} = (\widehat{\mathbf{w}}^{[1]\prime}, \cdots, \widehat{\mathbf{w}}^{[J_1]\prime})'$ where each $\widehat{\mathbf{w}}^{[i]} = (\widehat{w}^{[i]}_1, \cdots, \widehat{w}^{[i]}_{J_0})'$ is the SC weights on J_0 donor units that are used to predict the counterfactual of the treated unit i. Similarly, write $\widehat{\mathbf{r}} = (\widehat{\mathbf{r}}^{[1]\prime}, \cdots, \widehat{\mathbf{r}}^{[J_1]\prime})'$ and $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}^{[1]\prime}, \cdots, \widehat{\boldsymbol{\beta}}^{[J_1]\prime})'$.

Remark 1 (Weighting Matrix)**.** As pointed out by Ben-Michael et al. (2022), with multiple treated units, the SC weights could be constructed in two ways: (i) optimizing the separate fit for each treated unit; and (ii) optimizing the pooled fit for the average of the treated units. These ideas can be accommodated by choosing a proper weighting matrix $\mathbf{V}$. For example, assume $T_{i0} = T_0$ for simplicity. Taking $\mathbf{V} = \mathbf{I}_{T_0 M J_1}$ yields

$$\widehat{\boldsymbol{\beta}} = \underset{\mathbf{w}\in\mathcal{W},\mathbf{r}\in\mathcal{R}}{\arg\min} \sum_{i=1}^{J_1}\sum_{l=1}^{M}\sum_{t=1}^{T_0}\left(a^{[i]}_{t,l} - \mathbf{b}^{[i]\prime}_{t,l}\mathbf{w}^{[i]}_l - \mathbf{c}^{[i]\prime}_{t,l}\mathbf{r}^{[i]}_l\right)^2,$$

where $\mathbf{B}^{[i]}_l := (\mathbf{b}^{[i]}_{1,l}, \cdots, \mathbf{b}^{[i]}_{T_0,l})'$ is the l-th feature of the J_0 donor units, and $\mathbf{C}^{[i]}_l := (\mathbf{c}^{[i]}_{1,l}, \cdots, \mathbf{c}^{[i]}_{T_0,l})'$ is the additional K variables used to predict $\mathbf{A}^{[i]}_l$. The objective above is equivalent to minimizing the sum of squared errors of the pre-treatment fit for *each* treated unit and thus is termed "separate fit". By contrast, consider the following weighting matrix: $\mathbf{V} = \frac{1}{J_1^2}\mathbf{1}_{J_1}\mathbf{1}'_{J_1} \otimes \mathbf{I}_{T_0 M}$ where $\otimes$ denotes the Kronecker product operator. Then,

$$\widehat{\boldsymbol{\beta}} = \underset{\mathbf{w}\in\mathcal{W},\mathbf{r}\in\mathcal{R}}{\arg\min} \sum_{l=1}^{M}\sum_{t=1}^{T_0}\left[\frac{1}{J_1}\sum_{i=1}^{J_1}\left(a_{t,l}^{[i]} - \mathbf{b}_{t,l}^{[i]\prime}\mathbf{w}^{[i]} - \mathbf{c}_{t,l}^{[i]\prime}\mathbf{r}_l^{[i]}\right)\right]^2.$$

In this case, the goal is to minimize the sum of squared *averaged* errors across all treated units, which is usually termed "pooled fit". ⌟

Given the SC weights, the predicted counterfactual outcome of each treated unit $i \in \mathcal{E}$ is

$$\widehat{Y}_{i(T_i+k)}(\infty) := \mathbf{x}_{T_i+k}^{[i]\prime}\widehat{\mathbf{w}}^{[i]} + \mathbf{g}_{T_i+k}^{[i]\prime}\widehat{\mathbf{r}}^{[i]} = \mathbf{p}_{T_i+k}^{[i]\prime}\widehat{\boldsymbol{\beta}}^{[i]}, \quad \mathbf{p}_{T_i+k}^{[i]} = (\mathbf{x}_{T_i+k}^{[i]\prime},\ \mathbf{g}_{T_i+k}^{[i]\prime})', \quad k \geq 0,$$

where $\mathbf{x}_{T_i+k}^{[i]}$ is a vector of predictors of the donor units used to predict the counterfactual of the treated unit i measured k periods after the treatment, and $\mathbf{g}_{T_i+k}^{[i]}$ is a vector of predictors that correspond to the additional control variables specified in $\mathbf{C}^{[i]}$. In general, the variables included in $\mathbf{x}_{T_i+k}^{[i]}$ and $\mathbf{g}_{T_i+k}^{[i]}$ need not be the same as those in $\mathbf{B}^{[i]}$ and $\mathbf{C}^{[i]}$.

Again, let τ denote any of the four causal predictands defined in the paper. Then, the prediction of τ can be constructed accordingly and uniformly expressed as

$$\widehat{\tau} = \mathsf{L}(\{Y_{it}\}) - \mathbf{p}_\tau'\widehat{\boldsymbol{\beta}}, \tag{S.2.2}$$

with $\mathsf{L}(\{Y_{it}\})$ some linear combination of observed post-treatment outcomes and $\mathbf{p}_\tau'\widehat{\boldsymbol{\beta}}$ the prediction of the corresponding counterfactual. $\mathbf{p}_\tau$ here denotes a predictor vector associated with the predictand τ, whose specific expression in each case is as follows:

$$\begin{aligned}
\mathbf{p}_{\tau_{ik}} &= (\underbrace{\mathbf{0}_{J_0+KM}',\cdots,\mathbf{0}_{J_0+KM}'}_{(i-1)\text{ vectors}},\ \mathbf{p}_{T_i+k}^{[i]\prime},\ \underbrace{\mathbf{0}_{J_0+KM}',\cdots,\mathbf{0}_{J_0+KM}'}_{(J_1-i)\text{ vectors}})',\\
\mathbf{p}_{\tau_{i\cdot}} &= \Big(\underbrace{\mathbf{0}_{J_0+KM}',\cdots,\mathbf{0}_{J_0+KM}'}_{(i-1)\text{ vectors}},\ \frac{1}{T-T_i+1}\sum_{t\geq T_i}\mathbf{p}_t^{[i]\prime},\ \underbrace{\mathbf{0}_{J_0+KM}',\cdots,\mathbf{0}_{J_0+KM}'}_{(J_1-i)\text{ vectors}}\Big)',\\
\mathbf{p}_{\tau_{\mathcal{Q}k}} &= \Big(\frac{1}{Q}\mathbf{p}_{T_1+k}^{[1]\prime}\mathbb{1}(T_1\in\mathcal{Q}), \frac{1}{Q}\mathbf{p}_{T_2+k}^{[2]\prime}\mathbb{1}(T_2\in\mathcal{Q}),\cdots,\frac{1}{Q}\mathbf{p}_{T_{J_1}+k}^{[J_1]\prime}\mathbb{1}(T_{J_1}\in\mathcal{Q})\Big)', \quad \text{and}\\
\mathbf{p}_{\tau_{\cdot\cdot}} &= \Big(\frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_1+k}^{[1]\prime}, \frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_2+k}^{[2]\prime},\cdots,\frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_{J_1}+k}^{[J_1]\prime}\Big)',
\end{aligned}$$

where $\mathbf{0}_{J_0+KM}$ denotes a (J_0+KM)-dimensional vector of zeros.

Analogously to Equation (4.2) in the paper, the pseudo-true value of SC weights in this framework is defined by

$$\boldsymbol{\beta}_0 := (\mathbf{w}_0',\ \mathbf{r}_0')' = \underset{\mathbf{w}\in\mathcal{W},\,\mathbf{r}\in\mathcal{R}}{\arg\min}\ \mathbb{E}\Big[(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r})'\mathbf{V}(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r})\Big|\mathscr{H}\Big], \tag{S.2.3}$$

and then we can write

$$\mathbf{A} = \mathbf{B}\mathbf{w}_0 + \mathbf{C}\mathbf{r}_0 + \mathbf{U}, \qquad \mathbf{w}_0\in\mathcal{W}, \qquad \mathbf{r}_0\in\mathcal{R}, \tag{S.2.4}$$

where $\mathbf{U} = (\mathbf{u}^{[1]\prime},\cdots,\mathbf{u}^{[J_1]\prime})' \in \mathbb{R}^{\tilde{T}M}$ is the corresponding pseudo-true residual relative to the σ-field $\mathscr{H} = \{\mathbf{B},\mathbf{C},\mathbf{p}_\tau\}$.

As before, we differentiate the contribution of the in-sample error $\mathsf{InErr}(\tau)$ and the out-of-sample error $\mathsf{OutErr}(\tau)$ to the uncertainty of SC prediction of τ. For the in-sample uncertainty, the optimization bounds used in the paper can be generalized to this setup. Specifically, let $d_\beta = J_0 + KM$, $\mathbf{Z} = (\mathbf{B}, \mathbf{C})$, $\widehat{\mathbf{Q}} = \mathbf{Z}'\mathbf{V}\mathbf{Z}$, $\widehat{\boldsymbol{\gamma}}' = \mathbf{U}'\mathbf{V}\mathbf{Z}$, $\boldsymbol{\gamma} = \mathbb{E}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$, and $\Delta = \{\boldsymbol{\delta} \in \mathbb{R}^{d_\beta} : \boldsymbol{\delta} + \boldsymbol{\beta}_0 \in \mathcal{W} \times \mathcal{R}\}$. It follows from the optimality of $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$ and the convexity of $\mathcal{W}$ and $\mathcal{R}$ that $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \in \Delta$ and $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\widehat{\mathbf{Q}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq 0$. Also, given the expression (S.2.2), $\mathsf{InErr}(\tau)$ can be generally expressed as $-\mathbf{p}_\tau(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Then, a valid, though infeasible, prediction interval for the in-sample error $\mathsf{InErr}(\tau)$ is $[\underline{\mathfrak{c}}(\alpha_{\mathtt{in}}/2),\ \overline{\mathfrak{c}}(1 - \alpha_{\mathtt{in}}/2)]$, where $\underline{\mathfrak{c}}(\alpha)$ denotes the α-quantile of $\inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta}$ and $\overline{\mathfrak{c}}(\alpha)$ denotes α-quantile of $\sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta}$, conditional on $\mathscr{H}$, for any $\alpha \in (0,1)$, with $\mathcal{M}_{\mathbf{G}} = \{\boldsymbol{\delta} \in \Delta : \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2\mathbf{G}'\boldsymbol{\delta} \leq 0\}$, $\mathbf{G}|\mathscr{H} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$. Analogously to Equation (4.4) in the paper, once a feasible variance estimator $\widehat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ and a feasible constraint set $\Delta^\star$ that contains the original Δ in the small neighborhood of zero (as described by condition (iii) in Theorem S.1 below) are available, we can set

$$\underline{M}_{\mathtt{in}}(\tau) = \underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2) \quad \text{and} \quad \overline{M}_{\mathtt{in}}(\tau) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2) \tag{S.2.5}$$

where $\underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2)$ is the $(\alpha_{\mathtt{in}}/2)$-quantile of $\inf_{\boldsymbol{\delta} \in \mathcal{M}^\star_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta}$, and $\overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2)$ is the $(1 - \alpha_{\mathtt{in}}/2)$-quantile of $\sup_{\boldsymbol{\delta} \in \mathcal{M}^\star_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta}$ conditional on the data, with $\mathcal{M}^\star_{\mathbf{G}} = \{\boldsymbol{\delta} \in \Delta^\star :\ \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2(\mathbf{G}^\star)'\boldsymbol{\delta} \leq 0\}$ and $\mathbf{G}^\star|\mathsf{Data} \sim \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$.

For the out-of-sample error, analogously to Equation (4.6) in the paper, if $\mathsf{OutErr}(\tau)$ is assumed to be sub-Gaussian conditional on $\mathscr{H}$ with parameter $\sigma_{\mathscr{H}}$, then we can set

$$\begin{aligned} \underline{M}_{\mathtt{out}}(\tau) &= \mathbb{E}[\mathsf{OutErr}(\tau)|\mathscr{H}] - \sqrt{2\sigma^2_{\mathscr{H}} \log(2/\alpha_{\mathtt{out}})} \quad \text{and} \\ \overline{M}_{\mathtt{out}}(\tau) &= \mathbb{E}[\mathsf{OutErr}(\tau)|\mathscr{H}] + \sqrt{2\sigma^2_{\mathscr{H}} \log(2/\alpha_{\mathtt{out}})}. \end{aligned} \tag{S.2.6}$$

We present a general theorem that justifies the above method under high-level conditions and covers the results given in the paper as a special case. Let $\|\cdot\|_{\mathtt{F}}$ be the Frobenius matrix norm, $\|\cdot\|$ any ℓ_p vector norm with $p \geq 1$, and $\mathcal{B}(\mathbf{0}, c) = \{\boldsymbol{\lambda} \in \mathbb{R}^{d_\beta} : \|\boldsymbol{\lambda}\| \leq c\}$ a c-neighborhood of zero for any $c > 0$.

Theorem S.1. *Assume $\mathcal{W}$ and $\mathcal{R}$ are convex, $\widehat{\boldsymbol{\beta}}$ in (S.2.1) and $\boldsymbol{\beta}_0$ in (S.2.3) exist, $\mathscr{H} = \sigma(\mathbf{B}, \mathbf{C}, \mathbf{p}_\tau)$, and $\underline{M}_{\mathtt{in}}(\tau)$, $\overline{M}_{\mathtt{in}}(\tau)$, $\underline{M}_{\mathtt{out}}(\tau)$ and $\overline{M}_{\mathtt{out}}(\tau)$ are specified as in (S.2.5) and (S.2.6). In addition, for some finite non-negative constants ϵ_γ, π_γ, $\varpi^\star_\delta$, $\epsilon^\star_\delta$, $\pi^\star_\delta$, $\epsilon^\star_\Delta$, $\pi^\star_\Delta$, $\epsilon^\star_{\gamma,1}$, $\epsilon^\star_{\gamma,2}$ and $\pi^\star_\gamma$, the following conditions hold:*

(i) $\mathbb{P}[\mathbb{P}(-\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \in [\underline{\mathfrak{c}}(\alpha_0), \overline{\mathfrak{c}}(1 - \alpha_0)]|\mathscr{H}) \geq 1 - 2\alpha_0 - \epsilon_\gamma] \geq 1 - \pi_\gamma$ *for any* $\alpha_0 \in (0,1)$*;*
(ii) $\mathbb{P}[\mathbb{P}(\sup\{\|\boldsymbol{\delta}\| : \boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}\} \leq \varpi^\star_\delta|\mathscr{H}) \geq 1 - \epsilon^\star_\delta] \geq 1 - \pi^\star_\delta$*;*
(iii) $\mathbb{P}[\mathbb{P}(\Delta \cap \mathcal{B}(\mathbf{0}, \varpi^\star_\delta) \subseteq \Delta^\star\ |\mathscr{H}) \geq 1 - \epsilon^\star_\Delta] \geq 1 - \pi^\star_\Delta$*;*
(iv) $\mathbb{P}[\mathbb{P}(\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_{d_\beta}\|_{\mathtt{F}} \leq 2\epsilon^\star_{\gamma,1}|\mathscr{H}) \geq 1 - \epsilon^\star_{\gamma,2}] \geq 1 - \pi^\star_\gamma$*;*
(v) $\mathsf{OutErr}(\tau) - \mathbb{E}[\mathsf{OutErr}(\tau)|\mathscr{H}]$ *is sub-Gaussian conditional on $\mathscr{H}$ with parameter $\sigma_{\mathscr{H}}$.*

Then, for $\epsilon^\star_{\gamma,1} \in [0, 1/4]$,

$$\mathbb{P}\Big\{\mathbb{P}\big(\tau \in [\widehat{\tau} - \overline{M}_{\mathtt{in}}(\tau) - \overline{M}_{\mathtt{out}}(\tau),\ \widehat{\tau} - \underline{M}_{\mathtt{in}}(\tau) - \underline{M}_{\mathtt{out}}(\tau)]\big|\mathscr{H}\big) \geq 1 - \alpha_{\mathtt{in}} - \alpha_{\mathtt{out}} - \epsilon\Big\} \geq 1 - \pi,$$

where $\epsilon = \epsilon_\gamma + 2\epsilon^\star_{\gamma,1} + \epsilon^\star_{\gamma,2} + 2\epsilon^\star_\delta + \epsilon^\star_\Delta$ and $\pi = \pi_\gamma + \pi^\star_\gamma + \pi^\star_\delta + \pi^\star_\Delta$.

Assumptions (i)–(iv) imposed in Theorem S.1 are high-level, which are used for in-sample uncertainty quantification and can be verified in many practically relevant scenarios such as the cointegrated data considered in the paper. We give more detailed discussion of each condition in Section S.2.1 below. Assumption (v), as we emphasized before, is a moment condition used to showcase our out-of-sample uncertainty quantification strategy and can be relaxed by utilizing other appropriate concentration inequalities.

S.2.1 Discussion of Conditions (i)–(iv)

In this section, we discuss the justification of the high-level conditions (i)–(iv) in more detail.

- **Condition (i)**. This condition formalizes the idea of distributional approximation of $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ by a Gaussian vector $\mathbf{G}$. It can be verified under different primitive conditions, such as Assumption 1 in the paper, that accommodates non-stationary data and is applicable to our empirical application. Lemma S.2 provides a general way to verify condition (i) by assuming the the pseudo-true residuals in $\mathbf{U}$ are independent over t conditional on $\mathscr{H}$. In fact, (i) also holds when the errors are only weakly dependent (e.g., β-mixing) conditional on $\mathscr{H}$.

- **Condition (ii)**. This is a mild condition on the concentration of $\boldsymbol{\delta}\in\mathcal{M}_{\mathbf{G}}$. The requirement $\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2\mathbf{G}'\boldsymbol{\delta}\leq 0$ is usually known as the *basic inequality* in regression analysis; see, for example, Wainwright (2019, Chapter 7) for the case of lasso. The vector $\mathbf{G}$ is (conditionally) Gaussian by construction, making condition (ii) easy to verify based on well-known bounds for Gaussian distributions. This condition holds in a variety of empirically relevant settings, including outcomes-only regression with i.i.d. data, multi-equation regression with weakly dependent data, and settings with cointegrated outcomes and features.

- **Condition (iii)**. This is a high-level requirement on the constraint set $\Delta^\star$ used in the simulation. Intuitively, to obtain valid bounds on the in-sample error, the supremum or infimum of $-\mathbf{p}_\tau'\boldsymbol{\delta}$ should be searched for over a set of $\boldsymbol{\delta}$ values that contains Δ, or at least the portion of it within a small neighborhood of zero. One simple example that always satisfies condition (iii) is $\Delta^\star=\{(\mathbf{w}_1'-\mathbf{w}_2',\mathbf{r}_1'-\mathbf{r}_2')':(\mathbf{w}_1,\mathbf{r}_1),(\mathbf{w}_2,\mathbf{r}_2)\in\mathcal{W}\times\mathcal{R}\}$. However, this set is typically large, leading to overly conservative bounds. Thus, we provide an improved, general strategy to construct $\Delta^\star$ in this setting. It can be shown to satisfy condition (iii) if the constraints specified in $\mathcal{W}$ and $\mathcal{R}$ are formed by smooth functions. Suppose that

$$\mathcal{W}\times\mathcal{R}=\Big\{\boldsymbol{\beta}\in\mathbb{R}^{d_\beta}:\boldsymbol{m}_=(\boldsymbol{\beta})=\mathbf{0},\boldsymbol{m}_\leq(\boldsymbol{\beta})\leq\mathbf{0}\Big\},$$

  where $\boldsymbol{m}_=(\cdot)\in\mathbb{R}^{d_=}$ and $\boldsymbol{m}_\leq(\cdot)\in\mathbb{R}^{d_\leq}$ and $d_=$ and $d_\leq$ denote the number of equality and inequality constraints in $\mathcal{W}\times\mathcal{R}$, respectively. Let the ℓ-th constraint in $\boldsymbol{m}_\leq(\cdot)$ be $m_{\leq,\ell}(\cdot)$. Given tuning parameters $\varrho_\ell>0$, $\ell=1,\cdots,d_\leq$, let $\mathcal{A}=\{\ell_1,\cdots,\ell_k\}$ denote the set of indices for the inequality constraints such that $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}})>-\varrho_\ell$. We define

$$\widehat{\Delta}=\Big\{\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}:\boldsymbol{m}_=(\boldsymbol{\beta})=\mathbf{0},\ m_{\leq,\ell}(\boldsymbol{\beta})\leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}})\mathbb{1}(\ell\in\mathcal{A}),\ \ell=1,\cdots,d_\leq\Big\}. \qquad \text{(S.2.7)}$$

  Then, let $\Delta^\star=\widehat{\Delta}_\varepsilon:=\{\boldsymbol{\delta}:\mathrm{dist}(\boldsymbol{\delta},\widehat{\Delta})\leq\varepsilon\}$, where $\mathrm{dist}(\boldsymbol{\delta},\widehat{\Delta})=\inf_{\boldsymbol{\lambda}\in\widehat{\Delta}}\|\boldsymbol{\delta}-\boldsymbol{\lambda}\|_2$. That is, $\Delta^\star$ is an "ε-enlargement" of $\widehat{\Delta}$ for some $\varepsilon\geq 0$. The following lemma verifies condition (iii) for this $\Delta^\star$.

  Lemma S.1. *Assume that with probability over $\mathscr{H}$ at least $1-\pi^\star_\Delta$, the following conditions hold: (i) $\mathbb{P}(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2\leq\varpi^\star_\delta|\mathscr{H})\geq 1-\epsilon^\star_\Delta$; (ii) $\boldsymbol{m}(\cdot)=(\boldsymbol{m}_=(\cdot)',\boldsymbol{m}_\leq(\cdot)')'$ is twice continuously differentiable on $\mathcal{B}(\boldsymbol{\beta}_0,\varpi^\star_\delta)$ with $\inf_{\boldsymbol{\beta}\in\mathcal{B}(\boldsymbol{\beta}_0,\varpi^\star_\delta)}s_{\min}(\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{m}(\boldsymbol{\beta}))\geq c_{\min}$ for some constant $c_{\min}>0$; and (iii) for all $1\leq\ell\leq d_\leq$ and some $\mathfrak{c}>0$ specified in the proof, $\varrho_\ell>\mathfrak{c}\varpi^\star_\delta$, and $\varrho_\ell<$*

$|m_{\leq,\ell}(\boldsymbol{\beta}_0)| - \mathfrak{c}\varpi_\delta^\star$ *if* $m_{\leq,\ell}(\boldsymbol{\beta}_0) \neq 0$. *Then, condition (iii) of Theorem S.1 holds for* $\Delta^\star = \widehat{\Delta}_\varepsilon$ *with* $\varepsilon \geq \mathfrak{C}(\varpi_\delta^\star)^2$ *for some constant* $\mathfrak{C} > 0$. *If* $\boldsymbol{m}(\boldsymbol{\beta})$ *is linear in* $\boldsymbol{\beta}$*, condition (iii) holds for* $\Delta^\star = \widehat{\Delta}$.

The strategy described in the paper is an application of this theoretical result, as detailed in S.6. In this lemma, the tuning parameters ϱ_ℓ's are introduced to guarantee that, with high probability, we can correctly differentiate the binding inequality constraints from the other non-binding ones. Section S.2.2 below provides more practical details about choosing ϱ_ℓ. Also, the concentration requirement for $\widehat{\boldsymbol{\beta}}$ specified in this lemma is mild. Since $\widehat{\boldsymbol{\beta}}$ satisfies the basic inequality $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\widehat{\mathbf{Q}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq 0$, the concentration of $\widehat{\boldsymbol{\beta}}$ can be shown by combining a distributional approximation of $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ by a Gaussian vector $\mathbf{G}$ and the idea outlined in the previous discussion about condition (ii).

- **Condition (iv)**. This is a requirement that $\widehat{\boldsymbol{\Sigma}}$ be a "good" approximation of the unknown covariance matrix $\boldsymbol{\Sigma}$. Many standard covariance estimation strategies such as the family of well-known heteroskedasticity-consistent estimators can be utilized.

S.2.2 Defining Constraints in Simulation

We propose a feasible strategy to construct the constraint set $\Delta^\star$ used in the simulation. We focus on the most common case in practice: each constraint only restricts the SC weights $\widehat{\boldsymbol{\beta}}^{[i]}$ for one treated unit i, so there is no "cross-treated-unit" constraint.

First, we introduce the tuning parameters ϱ_ℓ, $\ell = 1, \cdots, d_{\leq}$, to determine which inequality constraints are binding. We define ϱ_ℓ as a high-probability bound on $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]})$. By Taylor expansion, if the constraint $m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq 0$ is binding (i.e., $m_{\leq,\ell}(\boldsymbol{\beta}_0^{[i]}) = 0$), then $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) \approx \frac{\partial}{\partial \boldsymbol{\beta}'} m_{\leq,\ell}(\boldsymbol{\beta}_0^{[i]})(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]})$. Therefore, given a high-probability bound $\varrho^{[i]}$ on $\|\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}\|_2$, we set

$$\varrho_\ell = \Big\| \frac{\partial}{\partial \boldsymbol{\beta}} m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) \Big\|_2 \times \varrho^{[i]}. \tag{S.2.8}$$

Next, to select $\varrho^{[i]}$, we employ the basic inequality $(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]})'\widehat{\mathbf{Q}}(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}) - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}) \leq 0$. As mentioned before, it always holds by optimality of $\widehat{\boldsymbol{\beta}}$, which, combined with the Gaussian approximation of $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ by $\mathbf{G}$, implies a high-probability deviation bound for $\widehat{\boldsymbol{\beta}}^{[i]}$: $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \leq 2\|\mathbf{G}\|_2/s_{\min}(\widehat{\mathbf{Q}})$. Motivated by this fact, we propose to set $\varrho^{[i]} = \mathcal{C}T_0^{-1/2}$, where $\mathcal{C}$ is defined as

$$\mathcal{C} = \frac{\sqrt{d_{\beta,0} \log(d_\beta) \log(T_0)} \max_{1 \leq j \leq J_0} \widehat{\sigma}_{b_j} \widehat{\sigma}_u}{\min_{1 \leq j \leq J_0} \widehat{\sigma}_{b_j}^2}, \tag{S.2.9}$$

where $d_{\beta,0}$ denotes the number of nonzeros in $\widehat{\boldsymbol{\beta}}$, and $\widehat{\sigma}_u$ and $\widehat{\sigma}_{b_j}$ are the estimated (unconditional) standard deviation of $\mathbf{u}^{[i]}$ and the j-th column of $\mathbf{B}^{[i]}$, respectively. If the SC weights are constructed by matching on both stationary and non-stationary features, the precision of the estimation is primarily determined by the non-stationary components. In such cases, one may disregard the stationary components when determining $\mathcal{C}$ using (S.2.9). See Section S.6.1 for more explanation. Moreover, the constant in (S.2.9) is a "rule-of-thumb" choice that can be rationalized under specific conditions and at least have the correct order of magnitude for $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$. Details are provided in Section S.6.1.

Then, in the simulation, we impose

$$m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) + \frac{1}{2} s_{\max}\Big(\frac{\partial}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]})\Big) \times (\varrho^{[i]})^2 \tag{S.2.10}$$

if $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) > -\varrho_\ell$ ("binding"), and retain $m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq 0$ otherwise ("non-binding"). This proposed adjusted constraint is motivated by the characterization of the distance between Δ and $\widehat{\Delta}$ in Lemma S.1, which typically depends on the second-order expansion of the constraint function around the true values. See Section S.6.2.2 for more discussion. If $m_{\leq,\ell}(\cdot)$ is linear, the second-order derivative of $m_{\leq,\ell}(\cdot)$ is exactly zero, leading to the constraint $m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]})$. Thus, the adjustment due to the second term on the right-hand side of (S.2.10) is only necessary for nonlinear constraints. This coincides with the result in Lemma S.1. In the special case of the L2 constraint used in our empirical application, (S.2.10) implies that we impose $\|\boldsymbol{\beta}^{[i]}\|_2^2 \leq \|\widehat{\boldsymbol{\beta}}^{[i]}\|_2^2 + (\varrho^{[i]})^2$ in the simulation if $m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq 0$ is determined to be a binding constraint.

S.3 Other Strategies for Uncertainty Quantification

S.3.1 Out-of-Sample Error

In Section 4.2 we discuss the approach for quantifying the out-of-sample uncertainty based on the non-asymptotic bounds. We briefly describe two strategies below.

- *Location-scale model.* Suppose that $u_{it} = \mathbb{E}[u_{it}|\mathscr{H}] + (\mathbb{V}[u_{it}|\mathscr{H}])^{1/2}\nu_{it}$ with ν_{it} statistically independent of $\mathscr{H}$. The bounds on u_{it} can now be set to $\underline{M}_{\texttt{out}} = \mathbb{E}[u_{it}|\mathscr{H}] + (\mathbb{V}[u_{it}|\mathscr{H}])^{1/2}\mathfrak{c}_\nu(\alpha_{\texttt{out}}/2)$ and $\overline{M}_{\texttt{out}} = \mathbb{E}[u_{it}|\mathscr{H}] + (\mathbb{V}[u_{it}|\mathscr{H}])^{1/2}\mathfrak{c}_\nu(1-\alpha_{\texttt{out}}/2)$ where $\mathfrak{c}_\nu(\alpha_{\texttt{out}}/2)$ and $\mathfrak{c}_\nu(1-\alpha_{\texttt{out}}/2)$ are $\alpha_{\texttt{out}}/2$ and $(1-\alpha_{\texttt{out}}/2)$ quantiles of ν_{it}, respectively, and $\alpha_{\texttt{out}}$ is the desired pre-specified level.
- *Quantile regression.* We can determine the $\alpha_{\texttt{out}}/2$ and $(1-\alpha_{\texttt{out}}/2)$ conditional quantiles of $u_{it}|\mathscr{H}$. Consequently, another possibility is to employ quantile regression methods to estimate those quantities using pre-treatment data.

S.3.2 Simultaneous Prediction Intervals

Section 4.4 constructs prediction intervals with simultaneous coverage. We briefly describe two other common approaches below.

- *Bonferroni-type correction.* There is a large literature on Bonferroni corrections that can be used to construct multiple prediction intervals with simultaneous coverage. For example, consider a simple correction strategy: for each $k = 0, \dots, L$, use any strategy described in Section 4.2 or Section S.3.1 to construct a prediction interval for $u_{i(T_i+k)}$ that has a coverage probability at least $1-(\alpha_{\texttt{out}}/(L+1))$. Then, the simultaneous coverage probability of the $L+1$ prediction intervals $\{\tilde{\mathcal{I}}_k : 0 \leq k \leq L\}$ is at least $1-\alpha_{\texttt{out}}$. Some other more sophisticated corrections are also available in the literature (see, e.g., Ravishanker et al., 1987). For instance, the second-order Bonferroni-type bound implies that

$$\mathbb{P}\big[u_{i(T_i+k)} \in \tilde{\mathcal{I}}_k \text{ for all } 0 \leq k \leq L \,\big|\, \mathscr{H}\big] \geq 1 - \sum_{k=0}^{L} p_k + \sum_{k=0}^{L-1} p_{k,k+1}, \quad \text{where}$$
$$p_k = \mathbb{P}(u_{i(T_i+k)} \in \tilde{\mathcal{I}}_k|\mathscr{H}), \quad p_{k,k+1} = \mathbb{P}(u_{i(T_i+k)} \in \tilde{\mathcal{I}}_k,\ u_{i(T_i+k+1)} \in \tilde{\mathcal{I}}_{k+1}|\mathscr{H}).$$

  Then, one can construct the prediction intervals $\tilde{\mathcal{I}}_k$'s with corresponding coverage probabilities p_k and $p_{k,k+1}$ such that $1-\sum_{k=0}^{L} p_k + \sum_{k=0}^{L-1} p_{k,k+1} \geq 1-\alpha_{\texttt{out}}$. Such bounds are usually sharper, but their implementation requires the modeling of the dependence of $(u_{it}, u_{i(t+1)})$ conditional on $\mathscr{H}$ and is computationally more burdensome.

- *Scheffé-type intervals.* An alternative approach is to construct Scheffé-type simultaneous prediction intervals, though stronger distributional assumptions need to be made. For instance, assume that $(u_{iT_i}, \cdots, u_{i(T_i+L)})'$ jointly follows a conditional Gaussian distribution with mean zero and variance $\Sigma_{\mathscr{H}}$. Then,

$$(u_{iT_i}, \cdots, u_{i(T_i+L)})\Sigma_{\mathscr{H}}^{-1}(u_{iT_i}, \cdots, u_{i(T_i+L)})' \sim \chi^2_{L+1},$$

where χ^2_{L+1} is χ^2 distribution with $L+1$ degrees of freedom. The sequence of prediction intervals $\tilde{\mathcal{I}}_k = \Big[-\sigma_{\mathscr{H},kk}\sqrt{\chi^2_{L+1}(1-\alpha_{\mathtt{out}})},\ \sigma_{\mathscr{H},kk}\sqrt{\chi^2_{L+1}(1-\alpha_{\mathtt{out}})}\Big]$ have the simultaneous coverage probability at least $1-\alpha_{\mathtt{out}}$, where $\sigma^2_{\mathscr{H},kk}$ is the k-th diagonal element of $\Sigma_{\mathscr{H}}$ and $\chi^2_{L+1}(1-\alpha_{\mathtt{out}})$ is the $(1-\alpha_{\mathtt{out}})$-quantile of χ^2 distribution with $L+1$ degrees of freedom.

S.3.3 Alternative Bounds for In-sample Error

When the causal predictand of interest depends on treatment effects on *multiple* treated units, such as the time-specific unit-averaged predictand $\tau_{\mathcal{Q}k}$, there is an alternative method for in-sample uncertainty quantification that may yield tighter prediction intervals. It relies on the fact that if the SC weights $\widehat{\boldsymbol{\beta}}^{[i]}$ for each ever-treated unit i are obtained through a *separate* optimization process using data $\mathbf{A}^{[i]}$, $\mathbf{B}^{[i]}$ and $\mathbf{C}^{[i]}$, then, as outlined in Remark 1, a sequence of restrictions must be obeyed: for each $i \in \mathcal{E}$,

$$(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]})'\widehat{\mathbf{Q}}^{[i]}(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}) - 2(\widehat{\boldsymbol{\gamma}}^{[i]} - \boldsymbol{\gamma}^{[i]})'(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}) \le 0,\ \widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]} \in \Delta^{[i]}. \tag{S.3.1}$$

Then, analogously to (S.2.5), for any predictand τ defined before, we can set

$$\begin{aligned} \underline{M}_{\mathtt{in}}(\tau) &= (\alpha_{\mathtt{in}}/2)\text{-quantile of } \inf_{\boldsymbol{\delta}\in\tilde{\mathcal{M}}^\star_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta} \quad \text{and} \\ \overline{M}_{\mathtt{in}}(\tau) &= (1-\alpha_{\mathtt{in}}/2)\text{-quantile of } \sup_{\boldsymbol{\delta}\in\tilde{\mathcal{M}}^\star_{\mathbf{G}}} -\mathbf{p}_\tau'\boldsymbol{\delta} \end{aligned} \tag{S.3.2}$$

conditional on the data, with $\tilde{\mathcal{M}}^\star_{\mathbf{G}} = \{\boldsymbol{\delta} = (\boldsymbol{\delta}^{[1]\prime}, \cdots, \boldsymbol{\delta}^{[J_1]\prime})' :\ \boldsymbol{\delta}^{[i]} \in \Delta^{[i]\star},\ \boldsymbol{\delta}^{[i]\prime}\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta}^{[i]} - 2(\mathbf{G}^{[i]\star})'\boldsymbol{\delta}^{[i]} \le 0,\ i \in \mathcal{E}\}$ and $\mathbf{G}^\star = (\mathbf{G}^{[1]\prime}, \cdots, \mathbf{G}^{[J_1]\prime})'|\mathsf{Data} \sim \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$. Again, $\Delta^{[i]}$ is replaced with its feasible (enlarged) version $\Delta^{[i]\star}$, and $\widehat{\boldsymbol{\gamma}}^{[i]} - \boldsymbol{\gamma}^{[i]}$ is replaced with the approximating Gaussian vector $\mathbf{G}^{[i]\star}$. As a reminder, this strategy is not equivalent to doing in-sample uncertainty quantification for each time-specific unit-specific prediction $\widehat{\tau}_{ik}$ and then combining their bounds together, since the whole vector $\mathbf{G}^\star$ has the same covariance as $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\gamma}}^{[1]\prime}, \cdots, \widehat{\boldsymbol{\gamma}}^{[J_1]\prime})'$, thus maintaining the correlation structure among different treated units.

Technically, the above optimization procedure for constructing SC weights is equivalent to a special case of (S.2.1), where the weighting matrix $\mathbf{V}$ is block diagonal, taking the form $\mathbf{V} = \mathsf{diag}(\mathbf{V}^{[1]}, \cdots, \mathbf{V}^{[J_1]})$, and the feasible set $\Delta = \mathcal{W} \times \mathcal{R}$ is the Cartesian product of $\Delta^{[i]} = \mathcal{W}^{[i]} \times \mathcal{R}^{[i]}$ for each subvector $\boldsymbol{\beta}^{[i]}$ (hence, there is no cross-treated-unit constraint). The general in-sample uncertainty quantification strategy presented earlier in (S.2.5) relies on the fact that $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\widehat{\mathbf{Q}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sum_{i\in\mathcal{E}}[(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]})'\widehat{\mathbf{Q}}^{[i]}(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]}) - 2(\widehat{\boldsymbol{\gamma}}^{[i]} - \boldsymbol{\gamma}^{[i]})'(\widehat{\boldsymbol{\beta}}^{[i]} - \boldsymbol{\beta}_0^{[i]})] \le 0$ and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \in \Delta = \times_{i\in\mathcal{E}}\Delta^{[i]}$, which are immediately implied by (S.3.1). In this sense, the restriction (S.3.1) is stricter, making the alternative bounds in (S.3.2) tighter (or at least no looser) than those in (S.2.5). Table S.2 quantifies the gains from this restriction in terms of the length of the prediction intervals in our leading empirical application.

Due to the complexity of the feasibility set $\tilde{\mathcal{M}}^\star_{\mathbf{G}}$, the alternative bounds (S.3.2) may not be

Table S.2: *Achieved reduction in the prediction intervals length with separate optimization.*

$\mathcal{W}$	*Ridge*		*Simplex*		*L1-L2*	
Predictand	$M=1$	$M=2$	$M=1$	$M=2$	$M=1$	$M=2$
All treated units						
τ_{ik}, Figure 3(a)	41.98 [16.80; 92.28]	55.05 [33.15; 80.68]	59.68 [17.70; 93.56]	60.63 [36.61; 92.64]	60.08 [11.66; 92.92]	60.38 [37.44; 92.97]
$\tau_{i\cdot}$, Figure 4(a)	42.78 [21.08; 87.87]	56.54 [35.36; 73.31]	61.32 [22.20; 93.66]	61.13 [40.56; 81.27]	60.41 [15.28; 91.55]	60.17 [42.20; 81.02]
$\tau_{\cdot k}$, Figure 6(a)	94.21 [94.06; 94.24]	94.09 [94.04; 94.16]	94.49 [94.40; 95]	94.12 [94.07; 94.25]	95.03 [94.77; 95.11]	94.33 [94.3; 94.41]
Countries Liberalized Before 1987						
$\tau_{\mathcal{Q}_1k}$, Figure 5(a)	76.65 [76.55; 76.92]	78.91 [77.61; 79.84]	78.04 [77.64; 78.92]	78.04 [77.73; 78.72]	79.92 [79.78; 80.13]	76.64 [76.45; 77.03]
Countries Liberalized in 1987-1991						
$\tau_{\mathcal{Q}_2k}$, Figure 5(c)	84.86 [84.47; 85.30]	84.76 [84.41; 84.94]	85.74 [85.45; 86.44]	86.28 [86.08; 86.54]	85.51 [84.86; 85.92]	85.12 [84.98; 85.23]
Countries Liberalized After 1991						
$\tau_{\mathcal{Q}_3k}$, Figure 5(e)	76.65 [76.55; 76.92]	78.91 [77.61; 79.84]	78.04 [77.64; 78.92]	78.04 [77.73; 78.72]	79.92 [79.78; 80.13]	76.64 [76.45; 77.03]

Notes: For each target predictand we report the median percentage change in the length of the prediction intervals, whereas in brackets we report the minimum and maximum change, respectively. The reported statistics are computed across horizons and/or treated units. We computed the percentage change of the prediction intervals without out-of-sample uncertainty.

theoretically justified using the same argument for (S.2.5). However, other techniques, such as strong approximations, can be employed to establish the validity of (S.3.2), albeit at the expense of additional technical complexity. We do not pursue further justification of (S.3.2) and leave it for future research.

S.4 Linear Factor Model Justification

There are different ways to justify the synthetic control method, including the cointegrated system motivated by our empirical application. In this section, we briefly discuss an alternative justification that assumes the data are generated through a linear factor model:

$$Y_{it}(\infty) = \lambda_t \mu_i + \nu_{it}, \qquad 1 \leq i \leq N,\ 1 \leq t \leq T.$$

For simplicity, assume that only the first unit is treated, the common factor λ_t is i.i.d. over $1 \leq t \leq T$, ν_{it} is i.i.d. across $1 \leq i \leq N$ and over $1 \leq t \leq T$, and $\{\lambda_t : 1 \leq t \leq T\}$ and $\{\nu_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ are independent of each other. In addition, we assume that $\mathbb{E}[\nu_{it}] = 0$, the factor loadings $\{\mu_i : 1 \leq i \leq N\}$ are fixed, and only the pre-intervention outcomes are used in the SC construction. Then, $\mathscr{H} = \{(Y_{2t}(\infty), \cdots, Y_{Nt}(\infty)) : 1 \leq t \leq T\}$. Accordingly, $\mathbf{w}_0$ is given by the following expression:

$$\mathbf{w}_0 = \underset{\mathbf{w}\in\mathcal{W}}{\arg\min}\, \frac{1}{T_0} \sum_{t=1}^{T_0} \mathbb{E}\Big[\Big((\mu_1 - \boldsymbol{\mu}_c'\mathbf{w})\lambda_t + (\nu_{1t} - \boldsymbol{\nu}_{t,c}'\mathbf{w})\Big)^2 \Big| \mathscr{H}\Big],$$

where $\boldsymbol{\mu}_c = (\mu_2, \cdots, \mu_N)'$ and $\boldsymbol{\nu}_{t,c} = (\nu_{2t}, \cdots, \nu_{Nt})'$, assuming the expectation (exists and) is finite. Define $\mathscr{H}_t = \{Y_{2t}(\infty), \cdots, Y_{Nt}(\infty)\}$. Then, $\mathbf{w}_0$ can be further written as

$$\mathbf{w}_0 = \underset{\mathbf{w}\in\mathcal{W}}{\arg\min} \Big\{\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\lambda_t^2|\mathscr{H}_t](\mu_1 - \boldsymbol{\mu}_c'\mathbf{w})^2 + \mathbb{E}[\nu_{11}^2] + \mathbf{w}'\Big(\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\boldsymbol{\nu}_{t,c}\boldsymbol{\nu}_{t,c}'|\mathscr{H}_t]\Big)\mathbf{w}$$
$$- 2(\mu_1 - \boldsymbol{\mu}_c'\mathbf{w})\Big(\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\lambda_t\boldsymbol{\nu}_{t,c}'|\mathscr{H}_t]\Big)\mathbf{w}\Big\},$$

assuming these expectations are finite. Given our assumptions on λ_t and ν_{it}, we expect that

$$\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\lambda_t^2|\mathscr{H}_t] \approx \mathbb{E}[\lambda_t^2], \quad \frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\boldsymbol{\nu}_{t,c}\boldsymbol{\nu}_{t,c}'|\mathscr{H}_t] \approx \mathbb{E}[\nu_{1t}^2]\mathbf{I}_{N-1}, \quad \text{and} \quad \frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}[\lambda_t\boldsymbol{\nu}_{t,c}'|\mathscr{H}_t] \approx \mathbf{0}$$

with high probability when T_0 is large, which can be shown under additional moment conditions. As a consequence, our expression for $\mathbf{w}_0$ is similar to that in Ferman and Pinto (2021).

Then, the in-sample error for the prediction of the TSUS effect τ_{1k} is given by

$$\mathsf{InErr}(\tau_{1k}) = -\mathbf{Y}_{\mathcal{N}t}'(\widehat{\mathbf{w}} - \mathbf{w}_0),$$

which represents the error in estimating the weights. The out-of-sample error is given by

$$\mathsf{OutErr}(\tau_{1k}) = Y_{1(T_1+k)} - \mathbf{Y}_{\mathcal{N}(T_1+k)}'\mathbf{w}_0 = \nu_{1(T_1+k)} - \boldsymbol{\nu}_{T_1+k,c}'\mathbf{w}_0 + (\mu_1 - \boldsymbol{\mu}_c'\mathbf{w}_0)\lambda_{T_1+k}.$$

The first term $\nu_{1(T_1+k)}$ represents the treated unit's innovation in the post-treatment period T_1+k; the second term $\boldsymbol{\nu}_{T_1+k,c}'\mathbf{w}_0$ represents the weighted average of the innovations for the donor units in the post-treatment period T_1+k; and the third term $(\mu_1 - \boldsymbol{\mu}_c'\mathbf{w}_0)\lambda_{T_1+k}$ can be thought of as the impact of the "bias" of the SC weights in T_1+k, if one thinks the target weight in this context is the one recovering the true factor loading μ_1 of the treated unit. Given the conditions imposed, our method is still applicable to this scenario, and importantly, the "bias" due to $(\mu_1 - \boldsymbol{\mu}_c'\mathbf{w}_0)\lambda_{T_1+k}$ is taken into account.

S.5 Proofs

S.5.1 Proof of Corollary 1

Proof. We only need to verify the conditions in Theorem S.1. The proof is divided into several steps.

Step 0: We first give several useful facts. The SC construction considered in this corollary is a special case of that considered in Theorem S.1, where the weighting matrix is an identity matrix, and there is only one feature to be matched (GDP per capita) and no additional covariates. As shown in the main paper, for any $\widehat{\tau} \in \{\widehat{\tau}_{ik}, \widehat{\tau}_{i\cdot}, \widehat{\tau}_{\mathcal{Q}k}, \widehat{\tau}_{\cdot\cdot}\}$, the in-sample error can always be expressed as $\mathbf{p}_\tau'(\widehat{\mathbf{w}}-\mathbf{w}_0)$. Note that $\Delta^{[i]}$ and $\Delta^{[i]\star}$ are convex in this case, and thus $\times_{i\in\mathcal{S}}\Delta^{[i]}$ and $\times_{i\in\mathcal{S}}\Delta^{[i]\star}$ (for some set of treated units $\mathcal{S}$) are all convex as well. For any κ, define $\mathcal{A}_\kappa = \{\boldsymbol{\xi} : \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\boldsymbol{\xi}}} \mathbf{p}_\tau'\boldsymbol{\delta} \le \kappa\}$, which is convex by Lemma 2 of Cattaneo et al. (2021). Let $d_\beta = J_0J_1$, and $\mathfrak{C}_1$, $\mathfrak{C}_2$, $\cdots$ denote some constant independent of T_0. Throughout this proof, $\|\cdot\|$ denotes the L2 norm for vectors and the operator norm for matrices.

Step 1: We want to verify condition (i) in Theorem S.1. Without loss of generality, we only consider the upper bound in this step, and the lower bound follows similarly. So our goal is to bound $|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H})-\mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})|$.

Note that by Assumption 1, u_{it} is independent conditional on $\mathscr{H}$. To simplify expressions, assume in this proof that $\boldsymbol{\gamma}=0$, and define $\tilde{u}_{it}=u_{it}$ if $t\leq T_i-1$ and $\tilde{u}_{it}=0$ if $t>T_i-1$. So we can write $\widehat{\boldsymbol{\gamma}}=(\sum_{t=1}^{T_1-1}\mathbf{Y}'_{\mathcal{N}t}u_{1t},\cdots,\sum_{t=1}^{T_{J_1}-1}\mathbf{Y}'_{\mathcal{N}t}u_{J_1t})'=\sum_{t=1}^{T_{J_1}}(\mathbf{I}_{J_1}\otimes\mathbf{Y}_{\mathcal{N}t})\tilde{u}_t$ where $\tilde{u}_t=(u_{1t},\cdots,u_{J_1t})'$.

Applying the Berry-Esseen theorem for convex sets (Raič, 2019),

$$|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H})-\mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})|\leq 42(d_\beta^{1/4}+16)\sqrt{J_1}\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\|^3\sum_{t=1}^{T_{J_1}}\mathbb{E}\Big[\sum_{i=1}^{J_1}\Big\|\mathbf{Y}_{\mathcal{N}t}u_{it}\Big\|^3\Big|\mathscr{H}\Big].$$

Given condition (i) in the corollary, we only need to bound $\sum_{t=1}^{T_{J_1}}\|\mathbf{Y}_{\mathcal{N}t}\|^3$ and $\|\boldsymbol{\Sigma}^{-1}\|$. First note that

$$\frac{1}{T_0^3}\sum_{t=1}^{T_{J_1}}\|\mathbf{Y}_{\mathcal{N}t}\|^3\leq\frac{\sqrt{J_0}}{T_0^{3/2}}\sum_{t=1}^{T_{J_1}}\sum_{j\in\mathcal{N}}|Y_{jt}/\sqrt{T_0}|^3,$$

By Assumption 1, each $\mathbf{Y}_{\mathcal{N}t}$ can be understood as a multivariate partial sum process indexed by t. By strong approximation of partial sum processes (e.g., Lemma 2.2 of Chang et al., 2006),

$$\mathbb{P}\Big(\max_{1\leq t\leq T_0}\|T_0^{-1/2}\mathbf{Y}_{\mathcal{N}t}-\tilde{\mathbf{G}}(t/T_0)\|\geq T_0^{-0.1}\Big)\leq\mathfrak{C}_1T_0^{-1},$$

where $\tilde{\mathbf{G}}(\cdot)$ is a J_0-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 $1\leq j\leq J_0$, for any $m>0$,

$$\mathbb{P}(\max_{0\leq r\leq 1}|\tilde{G}_j(r)|>m)\leq 2\mathbb{P}(\max_{0\leq r\leq 1}\tilde{G}_j(r)>m)=2\mathbb{P}(|G_j(1)|>m),$$

where $\tilde{G}_j(\cdot)$ is the j-th element of $\mathbf{G}(\cdot)$. Using the tail bound for Gaussian distributions, we can set $m=\sqrt{2\log(2J_0T_0)\sigma^2_{\max}}$ where $\sigma^2_{\max}$ is the largest variance of $\{v_{jt}:1\leq j\leq J_0\}$, which leads to $\max_{0\leq r\leq 1}|G_j(r)|\leq m$ with probability over $\mathscr{H}$ at least $1-(J_0T_0)^{-1}$. Therefore,

$$\frac{1}{T_0^3}\sum_{t=1}^{T_0}\|\mathbf{Y}_{\mathcal{N}t}\|^3\leq J_0^{3/2}T_0^{-1/2}\Big(\sqrt{2\log(2J_0T_0)\sigma^2_{\max}}+T_0^{-0.1}\Big)^3$$

with probability over $\mathscr{H}$ at least $1-T_0^{-1}-\mathfrak{C}_1T_0^{-1}$. In addition, note that $\frac{1}{T_0^3}\sum_{t=T_0+1}^{T_{J_1}}\|\mathbf{Y}_{\mathcal{N}t}\|^3\leq\frac{4}{T_0^3}\sum_{t=T_0+1}^{T_{J_1}}(\|\mathbf{Y}_{\mathcal{N}T_0}\|^3+\|\sum_{s=T_0+1}^{t}\mathbf{v}_s\|^3)$. Then, by the previous result and sub-Gaussianity of $\mathbf{v}_t$, we conclude that with probability at least $1-\mathfrak{C}_\pi T_0^{-1}$ for some constant $\mathfrak{C}_\pi>0$,

$$\frac{1}{T_0^3}\sum_{t=1}^{T_{J_1}}\|\mathbf{Y}_{\mathcal{N}t}\|^3\leq J_0^{3/2}(T_0^{-1/2}+\mathfrak{C}_2T_0^{-3/2}+\mathfrak{C}_3)\Big(\sqrt{2\log(2J_0T_0)\sigma^2_{\max}}+T_0^{-0.1}\Big)^3.$$

Finally, we consider $\boldsymbol{\Sigma}$. By assumption in the corollary, $s_{\min}(\boldsymbol{\Sigma})\geq\eta\widehat{\mathbf{Q}}$. Recall that $\widehat{\mathbf{Q}}=\mathrm{diag}(\widehat{\mathbf{Q}}^{[1]},\cdots,\widehat{\mathbf{Q}}^{[J_1]})$. Again, by strong approximation used previously, with probability over $\mathscr{H}$ at least $1-\mathfrak{C}_\pi T_0^{-1}$,

$$\left\|\frac{1}{T_0^2}\sum_{t=1}^{T_0}\mathbf{Y}_{\mathcal{N}t}\mathbf{Y}_{\mathcal{N}t}'-\frac{1}{T_0}\sum_{t=1}^{T_0}\tilde{\mathbf{G}}\Big(\frac{t}{T_0}\Big)\tilde{\mathbf{G}}\Big(\frac{t}{T_0}\Big)'\right\|\leq 2\sqrt{J_0}T_0^{-0.1}(m+T_0^{-0.1}).$$

Then, by the condition in the corollary, with probability over $\mathscr{H}$ at least $1-\pi_0-\mathfrak{C}_\pi T_0^{-1}$, $s_{\min}(\widehat{\mathbf{Q}}^{[1]})\geq\frac{2}{3}(\log T_0)^{-1/5}T_0^2$ for T_0 large enough. For other blocks $\widehat{\mathbf{Q}}^{[j]}$ with $j\neq 1$, using the previous results about the bound on $\mathbf{Y}_{\mathcal{N}t}$, we have with probability $1-\mathfrak{C}_\pi T_0^{-1}$, $\|\widehat{\mathbf{Q}}^{[j]}-\widehat{\mathbf{Q}}^{[1]}\|\leq\mathfrak{C}_4(m+T_0^{-0.1})^2T_0$. Therefore, we can conclude that with probability over $\mathscr{H}$ at least $1-\pi_0-\mathfrak{C}_\pi T_0^{-1}$, $s_{\min}(\widehat{\mathbf{Q}})\geq(\log T_0)^{-1/5}T_0^2/2$ for T_0 large enough.

Therefore, we can take $\pi_\gamma=\pi_0+\mathfrak{C}_\pi T_0^{-1}$ and $\epsilon_\gamma=\mathfrak{C}_\epsilon(\log T_0)^2T_0^{-1/2}$ for some non-negative finite constant $\mathfrak{C}_\epsilon$ implied by the previous calculations.

Step 2: Consider condition (ii) in Theorem S.1. We have the following basic inequality hold:

$$\lambda_{\min}(\mathfrak{D}_T^{-1}\widehat{\mathbf{Q}}\mathfrak{D}_T^{-1})\|\mathfrak{D}_T\boldsymbol{\delta}\|^2\leq\boldsymbol{\delta}\widehat{\mathbf{Q}}\boldsymbol{\delta}\leq 2\mathbf{G}'\boldsymbol{\delta}\leq 2\|\mathfrak{D}_T^{-1}\mathbf{G}\|\|\mathfrak{D}_T\boldsymbol{\delta}\|.$$

where $\mathfrak{D}_T=\mathrm{diag}((T_1-1)\mathbf{I}_{T_1-1},\cdots,(T_{J_1}-1)\mathbf{I}_{T_{J_1}-1})$. By condition (ii) imposed in the corollary and the argument given in Step 1, $s_{\min}(\mathfrak{D}_T^{-1}\widehat{\mathbf{Q}}\mathfrak{D}_T^{-1})\geq(\log T_0)^{-1/5}/2$ and $\lambda_{\max}(\mathfrak{D}_T^{-1}\widehat{\mathbf{Q}}\mathfrak{D}_T^{-1})\leq 2(\log T_0)^{1/5}$, with probability over $\mathscr{H}$ at least $1-\pi_\gamma$. Then, by the Gaussian tail bound, we can take $\varpi_\delta^\star=\mathfrak{C}_5(\log T_0)^{0.9}/T_0$, $\pi_\delta^\star=\pi_\gamma$, and $\epsilon_\delta^\star=T_0^{-1}$.

Step 3: Consider condition (iii) in Theorem S.1. Given the specific choices of $\varrho_\ell^{[i]}$'s and the argument in Step 2, the conditions in Lemma S.1 are satisfied for T_0 large enough, and we can differentiate the binding and nonbinding constraints with high probability. For each treated unit i, an L1-L2 constraint is imposed on the corresponding SC weights. The set $\Delta^{[i]}$ then can be written as $\Delta^{[i]}=\Delta_{\mathtt{l}}^{[i]}\cap\Delta_{\mathtt{nl}}^{[i]}$, where $\Delta_{\mathtt{l}}^{[i]}$ and $\Delta_{\mathtt{nl}}^{[i]}$ denote the feasibility sets defined by the L1 (simplex) and L2 (ridge) constraints respectively. Similarly, $\widehat{\Delta}$ in (S.2.7) for this special case can be written as $\widehat{\Delta}^{[i]}=\widehat{\Delta}_{\mathtt{l}}^{[i]}\cap\widehat{\Delta}_{\mathtt{nl}}^{[i]}$. The L1 constraint is linear, and thus by Lemma S.1, $\Delta_{\mathtt{l}}^{[i]}\cap\mathcal{B}(\mathbf{0},\varpi_\delta^\star)\subseteq\widehat{\Delta}_{\mathtt{l}}^{[i]}$. For the L2 constraint, Lemma S.1 implies that $\Delta_{\mathtt{nl}}^{[i]}\cap\mathcal{B}(\mathbf{0},\varpi_\delta^\star)\subseteq\{\boldsymbol{\delta}:\mathrm{dist}(\boldsymbol{\delta},\widehat{\Delta}_{\mathtt{nl}}^{[i]}\leq\mathfrak{C}(\varpi_\delta^\star)^2)\}$ for some constant $\mathfrak{C}>0$. Then, by the specified adjustment for the L2 constraint, condition (iii) holds for every $\Delta^{[i]\star}$ as well as $\times_{i\in\mathcal{S}}\Delta^{[i]\star}$, for $\mathcal{S}=\mathcal{Q}$ or $\mathcal{E}$. In this case, we can set $\pi_\Delta^\star=\pi_\gamma$ and $\epsilon_\Delta^\star=T_0^{-1}+\epsilon_\gamma$.

Step 4: Finally, consider condition (iv) in Theorem S.1. Note that

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

Since $\lambda_{\min}(\boldsymbol{\Sigma})\geq\eta\lambda_{\min}(\widehat{\mathbf{Q}})\geq\eta(\log T_0)^{-1/5}T_0^2/2$, with probability over $\mathscr{H}$ at least $1-\pi_\delta^\star$, it follows that $\pi_\gamma^\star=\pi_\gamma$, $\epsilon_{\gamma,2}^\star=\epsilon_{\Sigma,2}^\star$ and $\epsilon_{\gamma,1}^\star=2\sqrt{d_\beta}\epsilon_{\Sigma,1}^\star(\log T_0)^{1/5}/(T_0^2\eta)$.

Given all results above, we finally set $\pi=\pi_\gamma$ and $\epsilon=2\mathfrak{C}_\epsilon(\log T_0)^2T_0^{-1/2}+4\sqrt{d_\beta}\epsilon_{\Sigma,1}^\star(\log T_0)^{1/5}/(T_0^2\eta)+\epsilon_{\Sigma,2}^\star+3T_0^{-1}$. Then, the proof is complete. □

S.5.2 Proof of Theorem S.1

Proof. Let

$$\ell(\boldsymbol{\delta})=\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2\mathbf{G}'\boldsymbol{\delta}\quad\text{with }\mathbf{G}|\mathscr{H}\sim\mathsf{N}(\mathbf{0},\boldsymbol{\Sigma}),$$

$$\ell^\star(\boldsymbol{\delta}) = \boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta} - 2(\mathbf{G}^\star)'\boldsymbol{\delta} \quad \text{with } \mathbf{G}^\star|\mathsf{Data} \sim \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}).$$

Accordingly, define

$$\begin{aligned}
\bar{\varsigma}^\star &= \sup\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \\
\bar{\varsigma}_{\mathtt{r}}^{\mathsf{int}} &= \sup\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\}, \\
\bar{\varsigma}_{\mathtt{r}} &= \sup\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell(\boldsymbol{\delta}) \le 0\Big\}.
\end{aligned}$$

The subscript $\mathtt{r}$ indicates the quantity is a supremum over a further restricted region for $\boldsymbol{\delta}$ (due to the constraint $\|\boldsymbol{\delta}\| \le \varpi_\delta^\star$). For any $\alpha_0 \in [0,1]$, let $\bar{\mathfrak{c}}_{\mathtt{r}}(\alpha_0)$ be the α_0-quantile of $\bar{\varsigma}_{\mathtt{r}}$ conditional on $\mathscr{H}$. Similarly, define

$$\begin{aligned}
\underline{\varsigma}^\star &:= \inf\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \\
\underline{\varsigma}_{\mathtt{r}}^{\mathsf{int}} &= \inf\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\}, \\
\underline{\varsigma}_{\mathtt{r}} &= \inf\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell(\boldsymbol{\delta}) \le 0\Big\}.
\end{aligned}$$

Let $\underline{\mathfrak{c}}_{\mathtt{r}}(\alpha_0)$ be the α_0-quantile of $\underline{\varsigma}_{\mathtt{r}}$ conditional on $\mathscr{H}$.

Let $\mathbb{P}_1 = \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ and $\mathbb{P}_2 = \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$. By condition (iv), on an event with $\mathbb{P}$-probability at least $1-\pi_\gamma^\star$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_{\gamma,2}^\star$, the Kullback-Leibler divergence $\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) \le 2(\epsilon_{\gamma,1}^\star)^2$, and by Pinsker's inequality, this implies that for any $\kappa' \le \kappa$,

$$|\mathbb{P}^\star(\bar{\varsigma}_{\mathtt{r}}^{\mathsf{int}} \le \kappa) - \mathbb{P}^\star(\bar{\varsigma}_{\mathtt{r}} \le \kappa)| \le \epsilon_{\gamma,1}^\star \quad \text{and} \quad |\mathbb{P}^\star(\underline{\varsigma}_{\mathtt{r}}^{\mathsf{int}} \ge \kappa') - \mathbb{P}^\star(\underline{\varsigma}_{\mathtt{r}} \ge \kappa')| \le \epsilon_{\gamma,1}^\star.$$

On the other hand, note that by condition (iii), on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\Delta^\star$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_\Delta^\star$, the event $\{\bar{\varsigma}^\star \le \kappa\}$ implies that

$$\sup\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \le \sup\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \le \kappa$$

and $\{\underline{\varsigma}^\star \ge \kappa'\}$ implies that

$$\inf\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi_\delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \ge \inf\Big\{ -\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \ell^\star(\boldsymbol{\delta}) \le 0\Big\} \ge \kappa'.$$

Therefore,

$$\mathbb{P}^\star(\bar{\varsigma}^\star \le \kappa) \le \mathbb{P}^\star(\bar{\varsigma}_{\mathtt{r}}^{\mathsf{int}} \le \kappa) \quad \text{and} \quad \mathbb{P}^\star(\underline{\varsigma}^\star \ge \kappa') \le \mathbb{P}^\star(\underline{\varsigma}_{\mathtt{r}}^{\mathsf{int}} \ge \kappa').$$

Then, by definitions of $\bar{\varsigma}^\star$ and $\underline{\varsigma}^\star$, on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma^\star-\pi_\Delta^\star$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_{\gamma,2}^\star-\epsilon_\Delta^\star$, we have

$$\begin{aligned}
1-\alpha_{\mathtt{in}}/2 &\le \mathbb{P}^\star\Big(\bar{\varsigma}^\star \le \overline{M}_{\mathtt{in}}(\tau)\Big) \le \mathbb{P}^\star\Big(\bar{\varsigma}_{\mathtt{r}} \le \overline{M}_{\mathtt{in}}(\tau)\Big) + \epsilon_{\gamma,1}^\star \quad \text{and} \\
1-\alpha_{\mathtt{in}}/2 &\le \mathbb{P}^\star\Big(\underline{\varsigma}^\star \ge \underline{M}_{\mathtt{in}}(\tau)\Big) \le \mathbb{P}^\star\Big(\underline{\varsigma}_{\mathtt{r}} \ge \underline{M}_{\mathtt{in}}(\tau)\Big) + \epsilon_{\gamma,1}^\star.
\end{aligned}$$

Also, by condition (ii), we have with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\delta^\star$,

$$\bar{\mathfrak{c}}_{\mathtt{r}}(1-\alpha_{\mathtt{in}}/2-\epsilon_{\gamma,1}^\star) \ge \bar{\mathfrak{c}}(1-\alpha_{\mathtt{in}}/2-\epsilon_{\gamma,1}^\star-\epsilon_\delta^\star) \quad \text{and} \quad \underline{\mathfrak{c}}_{\mathtt{r}}(\alpha_{\mathtt{in}}/2+\epsilon_{\gamma,1}^\star) \le \underline{\mathfrak{c}}(\alpha_{\mathtt{in}}/2+\epsilon_{\gamma,1}^\star+\epsilon_\delta^\star).$$

Using condition (i) and all results above, we conclude that with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma-\pi_\gamma^\star-\pi_\Delta^\star-\pi_\delta^\star$,

$$\begin{aligned}
&\mathbb{P}\Big(\underline{M}_{\texttt{in}}(\tau)\leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq \overline{M}_{\texttt{in}}(\tau)\Big|\mathscr{H}\Big)\\
\geq&\,\mathbb{P}\Big(\underline{\mathfrak{c}}_{\mathtt{r}}(\alpha_{\texttt{in}}/2+\epsilon_{\gamma,1}^\star)\leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq \bar{\mathfrak{c}}_{\mathtt{r}}(1-\alpha_{\texttt{in}}/2-\epsilon_{\gamma,1}^\star)\Big|\mathscr{H}\Big)-\epsilon_{\gamma,2}^\star-\epsilon_\Delta^\star\\
\geq&\,\mathbb{P}\Big(\underline{\mathfrak{c}}(\alpha_{\texttt{in}}/2+\epsilon_{\gamma,1}^\star+\epsilon_\delta^\star)\leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq \bar{\mathfrak{c}}(1-\alpha_{\texttt{in}}/2-\epsilon_{\gamma,1}^\star-\epsilon_\delta^\star)\Big|\mathscr{H}\Big)-\epsilon_{\gamma,2}^\star-\epsilon_\Delta^\star\\
\geq&\,1-\alpha_{\texttt{in}}-2\epsilon_{\gamma,1}^\star-2\epsilon_\delta^\star-\epsilon_\gamma-\epsilon_{\gamma,2}^\star-\epsilon_\Delta^\star.
\end{aligned}$$

Finally, by condition (v), we immediately have $\mathbb{P}(\underline{M}_{\texttt{out}}(\tau)\leq \mathsf{OutErr}(\tau)\leq \overline{M}_{\texttt{out}}(\tau))\geq 1-\alpha_{\texttt{out}}$. Then the proof is complete. □

S.5.3 Verification of Condition (i) in Theorem S.1

As explained in the main paper, by convexity of the constraint set $\mathcal{W}\times\mathcal{R}$ and the optimality of $\widehat{\boldsymbol{\beta}}$,

$$\inf_{\boldsymbol{\delta}\in\mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}}}-\mathbf{p}_\tau'\boldsymbol{\delta}\leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}}}-\mathbf{p}_\tau'\boldsymbol{\delta},$$

where $\mathcal{M}_{\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}}=\{\boldsymbol{\delta}\in\Delta:\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\delta}\}$. Thus, condition (i) in Theorem S.1 indeed requires that $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ can be approximated by a Gaussian vector $\mathbf{G}$. Corollary 1 in the paper provided a verification of this condition in the special case of cointegrated data. In this section, we provide a more general way to verify condition (i) by imposing a conditional independence assumption on the pseudo-true residuals. The extension that allows for weakly dependent errors can be established using the idea of Theorem A in Cattaneo et al. (2021). For simplicity, we assume that only T_0 pre-treatment periods are used to obtain the weights. Also, we write $\mathbf{U}^{[i]}=(u_{it,1},\cdots,u_{it,M})'$, which is the vector of pseudo-true residuals corresponding to the treated unit i.

Lemma S.2. *Assume $\mathcal{W}$ and $\mathcal{R}$ are convex, $\widehat{\boldsymbol{\beta}}$ in Equation (S.2.1) and $\boldsymbol{\beta}_0$ in Equation (S.2.3) exist, and $\mathscr{H}=\sigma(\mathbf{B},\mathbf{C},\mathbf{p}_\tau)$. In addition, for some finite nonnegative constants, the following conditions hold:*

(i) $\mathbf{u}_t=(u_{1t,1},\cdots,u_{1t,M},\cdots,u_{J_1t,1},\cdots,u_{J_1t,M})$ *is independent over t conditional on $\mathscr{H}$;*

(ii) $\mathbb{P}(\sum_{t=1}^{T_0}\mathbb{E}[\|\sum_{j=1}^{J_1}\sum_{l=1}^{M}\tilde{\mathbf{z}}_{t,l}^{[j]}(u_{jt,l}-\mathbb{E}[u_{jt,l}|\mathscr{H}])\|_2^3|\mathscr{H}]\geq\epsilon_\gamma(84(d_\beta^{1/4}+16))^{-1})\geq 1-\pi_\gamma$ *where $\tilde{\mathbf{z}}_{t,l}^{[j]}$ is the $((j-1)T_0M+(l-1)T_0+t)$-th column of $\boldsymbol{\Sigma}^{-1/2}\mathbf{Z}'$.*

Then, with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma$,

$$\mathbb{P}\Big(\underline{\mathfrak{c}}(\alpha_0)\leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\leq \bar{\mathfrak{c}}(1-\alpha_0)|\mathscr{H}\Big)\geq 1-2\alpha_0-\epsilon_\gamma.$$

Proof. Define $\mathcal{M}_{\boldsymbol{\xi}}=\{\boldsymbol{\delta}\in\Delta:\boldsymbol{\delta}'\widehat{\mathbf{Q}}\boldsymbol{\delta}-2\boldsymbol{\xi}'\boldsymbol{\delta}\}$. Fix $\widehat{\mathbf{Q}}$ and $\mathbf{p}_\tau$. By Lemma 2 of Cattaneo et al. (2021), for any κ, $\mathcal{A}_\kappa:=\{\boldsymbol{\xi}\in\mathbb{R}^{d_\beta}:\sup_{\boldsymbol{\delta}\in\mathcal{M}_\xi}-\mathbf{p}_\tau'\boldsymbol{\delta}\leq\kappa\}$ and $\mathcal{A}_\kappa'=\{\boldsymbol{\xi}\in\mathbb{R}^{d_\beta}:\inf_{\boldsymbol{\delta}\in\mathcal{M}_\xi}-\mathbf{p}_\tau'\boldsymbol{\delta}\geq\kappa\}$ are convex. By Berry-Esseen Theorem for convex sets Raič (2019),

$$|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\in\mathcal{A}_\kappa|\mathscr{H})-\mathbb{P}(\mathbf{G}\in\mathcal{A}_\kappa|\mathscr{H})|\leq 42(d_\beta^{1/4}+16)\sum_{t=1}^{T_0}\mathbb{E}\Big[\Big\|\sum_{j=1}^{J_1}\sum_{l=1}^{M}\tilde{\mathbf{z}}_{t,l}^{[j]}\tilde{u}_{jt,l}\Big\|^3|\mathscr{H}\Big],$$

where $\tilde{u}_{jt,l} = u_{jt,l} - \mathbb{E}[u_{jt,l}|\mathscr{H}]$. By condition (ii), with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma$,

$$|\mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H}) - \mathbb{P}(\mathbf{G} \in \mathcal{A}_\kappa|\mathscr{H})| \leq \epsilon_\gamma/2.$$

Then, for any κ, with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma$,

$$\mathbb{P}(-\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq \kappa|\mathscr{H}) \geq \mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H}) \geq \mathbb{P}(\mathbf{G} \in \mathcal{A}_\kappa|\mathscr{H}) - \epsilon_\gamma/2.$$

Similarly, we can show for any κ,

$$\mathbb{P}(-\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \geq \kappa|\mathscr{H}) \geq \mathbb{P}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \in \mathcal{A}_\kappa'|\mathscr{H}) \geq \mathbb{P}(\mathbf{G} \in \mathcal{A}_\kappa'|\mathscr{H}) - \epsilon_\gamma/2.$$

Therefore, with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma$,

$$\mathbb{P}\Big(\underline{\mathfrak{c}}(\alpha_0) \leq -\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \leq \bar{\mathfrak{c}}(1-\alpha_0)\Big|\mathscr{H}\Big) \geq 1 - 2\alpha_0 - \epsilon_\gamma.$$

Then the proof is complete. □

S.5.4 Proof of Lemma S.1

Proof. In this proof, the constant $\mathfrak{C} > 0$ is a generic constant that is independent of T_0 and may be different in different uses.

Note that $\boldsymbol{m}_=(\boldsymbol{\beta}_0) = \mathbf{0}$ and $\boldsymbol{m}_\leq(\boldsymbol{\beta}_0) \leq \mathbf{0}$. For $\mathfrak{c} := \max_{1\leq\ell\leq d_\leq} \sup_{\boldsymbol{\beta}\in\mathcal{B}(\boldsymbol{\beta}_0,\varpi_\delta^\star)} \|\frac{\partial}{\partial\boldsymbol{\beta}} m_{\leq,\ell}(\boldsymbol{\beta})\|$, we have $\max_{1\leq\ell\leq d_\leq} |m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}) - m_{\leq,\ell}(\boldsymbol{\beta}_0)| \leq \mathfrak{c}\varpi_\delta^\star$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_\Delta^\star$, on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\Delta^\star$. Note that if the ℓ-th inequality constraint is binding, i.e., $m_{\leq,\ell}(\boldsymbol{\beta}_0) = 0$, then $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}) = \frac{\partial}{\partial\boldsymbol{\beta}'} m_{\leq,\ell}(\tilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ for some $\tilde{\boldsymbol{\beta}}$ between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. By the condition imposed on the tuning parameters ϱ_ℓ's, on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\Delta^\star$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_\Delta^\star$, $\mathcal{A}$ coincides with the set of indices for the binding inequality constraints. Without loss of generality, we assume $\boldsymbol{m}_\leq(\boldsymbol{\beta}_0) = \mathbf{0}$ hereafter. Otherwise, the non-binding constraints can be dropped, and the proof can proceed the same way as described below.

Define $\Gamma_=(\boldsymbol{\beta}) = \frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{m}_=(\boldsymbol{\beta})$ and $\Gamma_\leq(\boldsymbol{\beta}) = \frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{m}_\leq(\boldsymbol{\beta})$. Let

$$\Gamma(\boldsymbol{\beta}) = \Big(\Gamma_='(\boldsymbol{\beta}), \Gamma_\leq'(\boldsymbol{\beta}), \Gamma_c'(\boldsymbol{\beta}_0)\Big)', \quad \Gamma^0 = \Gamma(\boldsymbol{\beta}_0), \quad \Gamma^\star = \Gamma(\widehat{\boldsymbol{\beta}}),$$

where $\Gamma_c(\boldsymbol{\beta}_0)$ is chosen such that $\Gamma(\boldsymbol{\beta}_0)$ is non-degenerate. By conditions (i) and (ii) imposed in the lemma, $\|\Gamma^0 - \Gamma^\star\| \leq \mathfrak{C}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon_\Delta^\star$, on an event with $\mathbb{P}$-probability at least $1-\pi_\Delta^\star$.

Let

$$\boldsymbol{m}^+(\cdot) = \Big(\boldsymbol{m}_='(\cdot),\ \boldsymbol{m}_\leq'(\cdot),\ (\cdot - \boldsymbol{\beta}_0)' \times \Gamma_c(\boldsymbol{\beta}_0)'\Big)'.$$

Then, $\boldsymbol{m}^+(\boldsymbol{\beta}_0) = \mathbf{0}$. For each $\boldsymbol{\beta}$ in the neighborhood around $\boldsymbol{\beta}_0$ such that $\boldsymbol{\beta} - \boldsymbol{\beta}_0 \in \Delta \cap \mathcal{B}(0, \varpi_\delta^\star)$, define

$$\boldsymbol{\lambda}^0 = (\Gamma^0)^{-1}\Big(\boldsymbol{m}^+(\boldsymbol{\beta}) - \boldsymbol{m}^+(\boldsymbol{\beta}_0)\Big).$$

Thus, $\Gamma_=^0\boldsymbol{\lambda}^0 = \mathbf{0}$, $\Gamma_\leq^0\boldsymbol{\lambda}^0 \leq \mathbf{0}$. Note that by Taylor's expansion,

$$\|\boldsymbol{\lambda}^0 - (\Gamma^0)^{-1}\Gamma^0(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \leq \mathfrak{C}\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2,$$

implying that $\|\boldsymbol{\lambda}^0 - (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| \le \mathfrak{C}(\varpi^\star_\delta)^2$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_\Delta$, on an event with $\mathbb{P}$-probability at least $1-\pi^\star_\Delta$.

Next, define $\tilde{\boldsymbol{m}}(\cdot) = \boldsymbol{m}^+(\widehat{\boldsymbol{\beta}} + \cdot) - \boldsymbol{m}^+(\widehat{\boldsymbol{\beta}})$ and $\tilde{\boldsymbol{\beta}} = \boldsymbol{\phi}^\star + \widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\phi}^\star$ defined below:

$$\boldsymbol{\phi}^\star := \tilde{\boldsymbol{m}}^{-1}\Big(\Gamma^\star(\boldsymbol{\lambda}^0 - (\Gamma^\star)^{-1}(\Gamma^\star - \Gamma^0)\boldsymbol{\lambda}^0)\Big).$$

By Taylor's expansion,

$$\begin{aligned}\boldsymbol{\phi}^\star &= \tilde{\boldsymbol{m}}^{-1}(\mathbf{0}) + \Big[\frac{\partial}{\partial \boldsymbol{\phi}'}\tilde{\boldsymbol{m}}(\mathbf{0})\Big]^{-1}\Gamma^\star\Big(\boldsymbol{\lambda}^0 - (\Gamma^\star)^{-1}(\Gamma^\star - \Gamma^0)\boldsymbol{\lambda}^0\Big) + \mathfrak{Re}\\ &= \boldsymbol{\lambda}^0 - (\Gamma^\star)^{-1}(\Gamma^\star - \Gamma^0)\boldsymbol{\lambda}^0 + \mathfrak{Re},\end{aligned}$$

where $\|(\Gamma^\star)^{-1}(\Gamma^\star - \Gamma^0)\boldsymbol{\lambda}^0 + \mathfrak{Re}\| \le \mathfrak{C}\|\boldsymbol{\lambda}^0\|^2$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_\Delta$, on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi^\star_\Delta$. That is, we actually find $\tilde{\boldsymbol{\beta}}$ such that $\|(\tilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}) - \boldsymbol{\lambda}^0\| \le \mathfrak{C}(\varpi^\star_\delta)^2$. Note that

$$\boldsymbol{m}^+(\widehat{\boldsymbol{\beta}} + \boldsymbol{\phi}^\star) - \boldsymbol{m}^+(\widehat{\boldsymbol{\beta}}) = \tilde{\boldsymbol{m}}(\boldsymbol{\phi}^\star) = \tilde{\boldsymbol{m}}\Big(\tilde{\boldsymbol{m}}^{-1}\Big(\Gamma^\star(\boldsymbol{\lambda}^0 - (\Gamma^\star)^{-1}(\Gamma^\star - \Gamma^0)\boldsymbol{\lambda}^0)\Big)\Big).$$

Thus, $\boldsymbol{m}_=(\tilde{\boldsymbol{\beta}}) = \mathbf{0}$ and $\boldsymbol{m}_\le(\tilde{\boldsymbol{\beta}}) \le \boldsymbol{m}_\le(\widehat{\boldsymbol{\beta}})$, i.e., $\boldsymbol{\phi}^\star \in \widehat{\Delta}$. This shows that $\mathrm{dist}(\Delta \cap \mathcal{B}(\mathbf{0}, \varpi^\star_\delta), \widehat{\Delta}) \le \varpi^\star_\Delta$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_\Delta$, on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi^\star_\Delta$. Then, the desired conclusion holds by definition of the ε-enlargement $\widehat{\Delta}_\varepsilon$. Note that when constraints are linear, $\Gamma^0 = \Gamma^\star$ and the second-order derivative of $\boldsymbol{m}^+(\cdot)$ is exactly zero. So the above calculation implies that $\boldsymbol{\beta} - \boldsymbol{\beta}_0 = \boldsymbol{\lambda}_0 = \boldsymbol{\phi}^\star$.

In the above, we make use of the fact that $\Gamma^\star$ is non-degenerate, i.e., its smallest eigenvalue is bounded away from zero. Note that by assumptions on the constraints, it is feasible to construct Γ_c such that Γ^0 is non-degenerate (with high probability over $\mathscr{H}$). Then, by Weyl's inequality,

$$s_{\min}(\Gamma^\star(\Gamma^\star)') \ge s_{\min}(\Gamma^0(\Gamma^0)') - \mathfrak{C}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|,$$

implying $s_{\min}(\Gamma^\star) \ge s_{\min}(\Gamma^0) - \mathfrak{C}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| > 0$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_\Delta$, on an event with $\mathbb{P}$-probability at least $1-\pi^\star_\Delta$, where $s_{\min}(\cdot)$ here denotes the smallest eigenvalue of the symmetric matrix inside. Then, the proof is complete. □

S.6 Discussion on Tuning Parameters

S.6.1 Determining Binding Constraints

Our proposed strategy in Section S.2.2 for constructing the constraint set in the simulation relies on the tuning parameters ϱ_ℓ's, which are used to differentiate the binding constraints from the non-binding ones. The recommended choice described in (S.2.8) and (S.2.9) can be rationalized as follows.

For ease of notation, suppose that there is only one treated unit, and thus we can omit the superscript $[i]$ for quantities like $\boldsymbol{\beta}_0^{[i]}$ or $\varrho^{[i]}$ with no confusion. Assume the ℓ-th constraint $m_{\le,\ell}(\boldsymbol{\beta}) \le 0$ is binding ($m_{\le,\ell}(\boldsymbol{\beta}_0) = 0$), and $m_{\le,\ell}(\cdot)$ is sufficiently smooth. By the mean value theorem, $|m_{\le,\ell}(\widehat{\boldsymbol{\beta}})| = |\frac{\partial}{\partial \boldsymbol{\beta}'} m_{\le,\ell}(\check{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)| \le \|\frac{\partial}{\partial \boldsymbol{\beta}} m_{\le,\ell}(\check{\boldsymbol{\beta}})\|_2 \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$, where $\check{\boldsymbol{\beta}}$ is some point between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. The parameters $\varrho_\ell^{[i]}$ are used to characterize this upper bound. For $\frac{\partial}{\partial \boldsymbol{\beta}} m_{\le,\ell}(\check{\boldsymbol{\beta}})$, a natural "estimator" would be $\frac{\partial}{\partial \boldsymbol{\beta}} m_{\le,\ell}(\widehat{\boldsymbol{\beta}})$. The difference between the two is bounded by $c_1\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$ for some constant $c_1 > 0$ (with high probability). Then, it remains to characterize $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2$.

Note that the basic inequality below always holds (deterministically) by optimality of $\widehat{\boldsymbol{\beta}}$:

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \leq \frac{2\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_2/\sqrt{T_0}}{s_{\min}(\widehat{\mathbf{Q}})/T_0} \times T_0^{-1/2} =: \mathcal{C}T_0^{-1/2}. \tag{S.6.1}$$

We first discuss the denominator. Recall that $\widehat{\mathbf{Q}} = \mathbf{Z}'\mathbf{V}\mathbf{Z}$. Since an (unrestricted) intercept is often included in the SC construction, we assume the predictor variables included in $\mathbf{Z}$ are de-meaned (and the constant term is excluded from $\mathbf{Z}$). If $\mathbf{V}$ is an identity matrix and the predictor variables are (approximately) orthogonal to each other, then $s_{\min}(\widehat{\mathbf{Q}}/T_0)$ is the minimum diagonal entry of $\widehat{\mathbf{Q}}/T_0$, i.e., $\min_{1\leq j\leq J_0} \widehat{\sigma}^2_{b_j}$, which motivates the choice of the denominator in (S.2.9). On the other hand, recall that the numerator $\widehat{\boldsymbol{\gamma}} = \mathbf{Z}'\mathbf{u}$. As discussed in the paper, we can (conditionally) approximate $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$ by the Gaussian vector $\mathbf{G}$. Then, by Gaussian tail bound, it can be shown that, for any large $\lambda > 0$, $T_0^{-1/2}\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_\infty \leq (\sqrt{2\log d_\beta} + \lambda) \times \sqrt{\max_{1\leq j\leq J_0} \mathbb{V}[\widehat{\gamma}_j|\mathscr{H}]/T_0}$ with high probability. When $\mathbf{u}$ is independent of $\mathscr{H}$ and its components are independent over t, $\max_{1\leq j\leq J_0} \mathbb{V}[\widehat{\gamma}_j|\mathscr{H}]/T_0 = \sigma_u^2 \times \max_{1\leq j\leq J_0} \widehat{\sigma}^2_{b_j}$. Therefore, we can set

$$\mathcal{C} = \mathcal{C}_1 := \frac{2\sqrt{d_\beta}(\sqrt{2\log d_\beta} + \lambda)\max_{1\leq j\leq J_0} \widehat{\sigma}_{b_j}\widehat{\sigma}_u}{\min_{1\leq j\leq J_0} \widehat{\sigma}^2_{b_j}}$$

for any large λ. In most applications, however, the SC weights are sparse due to the simplex- or lasso-type constraints. It is known from the sparse linear regression literature that the bound $\mathcal{C}_1$ can be further improved by replacing the factor $\sqrt{d_\beta}$ by $c_3\sqrt{\|\boldsymbol{\beta}_0\|_0}$, where $\|\cdot\|_0$ denotes the number of nonzeros in a vector and $c_3 > 0$ is some absolute constant (see, e.g., Wainwright, 2019, Theorem 7.13). Assuming $\sqrt{\log T_0} \geq 4\sqrt{2}c_3$, we take $\lambda = \sqrt{\log d_\beta \log T_0}/(4c_3)$ and use $\|\widehat{\boldsymbol{\beta}}\|_0$ as a proxy for $\|\boldsymbol{\beta}_0\|_0$, which yields our recommended choice of $\mathcal{C}$ given in (S.2.9):

$$\mathcal{C} = \mathcal{C}_2 := \frac{\sqrt{\|\widehat{\boldsymbol{\beta}}\|_0 \log d_\beta \log T_0}\max_{1\leq j\leq J_0} \widehat{\sigma}_{b_j}\widehat{\sigma}_u}{\min_{1\leq j\leq J_0} \widehat{\sigma}^2_{b_j}}.$$

Furthermore, when $\widehat{\sigma}^2_{b_j}$ is roughly the same across j, then $\mathcal{C}_2$ can be further simplified to

$$\mathcal{C} = \mathcal{C}_3 = \frac{\sqrt{\|\widehat{\boldsymbol{\beta}}\|_0 \log d_\beta \log T_0}\,\widehat{\sigma}_u}{\min_{1\leq j\leq J_0} \widehat{\sigma}_{b_j}}.$$

While these choices are only rules of thumb justified under specific assumptions on the data generating process, they at least have the correct order of magnitude and are valid at least when T_0 is large. For instance, for the cointegrated data considered in our basic setup, it can be shown that, for sufficiently small $c_4 > 0$, $\widehat{\sigma}^2_{b_j} \geq c_4 T_0$ with high probability, and $\widehat{\sigma}^2_u$ is bounded with high probability. Therefore, the suggested ϱ based on $\mathcal{C}_2$ has the order of T_0^{-1}, which is the well-known rate of convergence for cointegrated regression.

When SC weights are obtained by matching on both stationary and non-stationary features, the contribution of the stationary components to the concentration of $\widehat{\boldsymbol{\beta}}$ is negligible. In other words, the order of magnitude for the bound on $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2$ is primarily determined by the non-stationary components. To see this, assume that we match on two features, a non-stationary $\mathbf{B}_1 = (\mathbf{b}_{1,1}, \cdots, \mathbf{b}_{T_0,1})' \in \mathbb{R}^{T_0\times J_0}$ and a stationary $\mathbf{B}_2 = (\mathbf{b}_{1,2}, \cdots, \mathbf{b}_{T_0,2})' \in \mathbb{R}^{T_0\times J_0}$. Consider the basic inequality (S.6.1) again. $\widehat{\mathbf{Q}}/T_0$ can be written as $\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{b}_{t,1}\mathbf{b}_{t,1}' + \frac{1}{T_0}\sum_{t=1}^{T_0}\mathbf{b}_{t,2}\mathbf{b}_{t,2}' =: \mathrm{I}+\mathrm{II}$. As explained previously, with high probability, $s_{\min}(\mathrm{I}/T_0)$ is bounded from below by T_0, up to a sufficiently small constant, whereas II/T_0 is much smaller—under mild conditions, it is bounded with high probability. For the numerator $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})/T_0$, we can similarly decompose it into a non-stationary part and a stationary part. By the argument described before, it follows that the stationary part is much smaller than the non-stationary part, leading to a deviation bound for $\widehat{\boldsymbol{\beta}}$ on the order of T_0^{-1}. Therefore, we recommend applying the above formulas using the non-stationary features only in practice.

Finally, as a reminder, other choices of ϱ can be proposed based on a slightly different basic inequality: $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2 \leq 2\|\widehat{\boldsymbol{\gamma}}\|_2/s_{\min}(\widehat{\mathbf{Q}})$. It also holds deterministically by optimality of $\widehat{\boldsymbol{\beta}}$. The denominator is the same as before, but the numerator $\widehat{\boldsymbol{\gamma}}$ is not demeaned. When the model is correctly specified ($\boldsymbol{\gamma}=\mathbf{0}$), the numerator may be small; however, when the model is misspecified, $\widehat{\boldsymbol{\gamma}}$ is not necessarily small, and it may be characterized by its sample analogue. Thus, we can set

$$\varrho = \frac{2\sqrt{d_\beta}\max_{1\leq j\leq J_o}\widehat{\sigma}^2_{b_j u}}{\min_{1\leq j\leq J_0}\widehat{\sigma}^2_{b_j}}$$

where $\widehat{\sigma}^2_{b_j u}$ is the estimated covariance between the pseudo-true residual $\mathbf{u}$ and the j-th column of $\mathbf{B}$ (the features of the j-th control unit). However, compared to the alternatives described in the paper, this bound is generally loose and thus is not recommended.

S.6.2 Adjustment for Nonlinear Constraints

As described in the main paper, we have an additional adjustment to nonlinear constraints in the simulation (the second term on the right-hand side of (S.2.10)). In this section, we briefly discuss the necessity of this adjustment and the justification of the proposed strategy.

S.6.2.1 Necessity of Adjustment

The constraints allowed in this paper is more general than those considered in Cattaneo et al. (2021). Specifically, condition (T2.iii) in Cattaneo et al. (2021) requires the constraint set used in the simulation for in-sample uncertainty quantification be locally *equal to* the original constraint set in the synthetic control problem. In the notation of this paper, that condition means that, with some high probability, $\Delta^\star \cap \mathcal{B}(\mathbf{0}, \varpi_\delta^\star) = \Delta \cap \mathcal{B}(\mathbf{0}, \varpi_\delta^\star)$ for some (small) $\varpi_\delta^\star$-neighborhood of zero. This is generally true when constraints are formed by linear functions. To see this, suppose that there are two donor units, an L1 constraint $\mathcal{W} = \{(w_1, w_2) : |w_1| + |w_2| \leq 1\}$ is imposed, and the pseudo-true value is $\mathbf{w}_0 = (1, 0)$. The constraint set $\Delta^\star$ used in the simulation relies on the estimated weights $\widehat{\mathbf{w}}$, which are generally close (but not exactly equal) to $\mathbf{w}_0$ with high probability. For instance, let $\widehat{\mathbf{w}} = (0.99, 0.01)$, and applying the suggested strategy described in Section S.2.2, one correctly finds that two linear constraints among those derived from decomposing the original L1 constraint, $w_1 + w_2 \leq 1$ and $w_1 - w_2 \leq 1$, are binding. Thus, one defines $\Delta^\star = \{(\delta_1, \delta_2) : (\delta_1 + 0.99) + (\delta_2 + 0.01) \leq 0.99 + 0.01, (\delta_1 + 0.99) - (\delta_2 + 0.01) \leq 0.99 - 0.01\}$, which is exactly the same as $\Delta = \mathcal{W} - \mathbf{w}_0$ locally around $(0, 0)$.

However, such an exact equality is generally not possible when constraints are formed by nonlinear functions. Still consider the same example, but an L2 constraint $\mathcal{W} = \{(w_1, w_2) : w_1^2 + w_2^2 \leq 1\}$ is imposed instead. Applying the same strategy, one correctly finds that the constraint should be binding and thus lets $\Delta^\star = \{(\delta_1, \delta_2) : (\delta_1 + 0.99)^2 + (\delta_2 + 0.01)^2 \leq 0.99^2 + 0.01^2\}$, whereas the (centered) original constraint set is $\Delta = \{(\delta_1, \delta_2) : (\delta_1 + 1)^2 + \delta_2^2 = 1\}$. In other words, Δ and $\Delta^\star$ are two "circles" passing through $(0,0)$. Although they are locally close near $(0,0)$, their boundaries have different tangents at that point, and hence the two sets are not locally identical, no matter how small the neighborhood $\mathcal{B}(\mathbf{0},\varpi^\star)$ is. In this sense, the results in Cattaneo et al. (2021) are more applicable to synthetic controls with linear constraints (such as simplex and Lasso). By contrast, the local approximate equality in condition (iii) of Theorem S.1 is much weaker and makes the results in this paper applicable to more general cases with possibly nonlinear constraints. Given the fact described above, when there are (binding) nonlinear constraints, we propose to enlarge the feasibility set defined by nonlinear constraints to satisfy the sufficient condition (iii) of Theorem S.1.

S.6.2.2 Enlarging Nonlinear Constraint Sets

In this section we briefly discuss the justification for the adjustment in (S.2.10). Our strategy is motivated by the proof of Lemma S.1, which characterizes the distance between $\Delta\cap\mathcal{B}(\mathbf{0},\varpi^\star_\delta)$ and $\widehat{\Delta}$. For simplicity, assume that there is only one treated unit (so the superscript $[i]$ can be omitted with no confusion).

Suppose that we only have binding inequality constraints $\boldsymbol{m}(\boldsymbol{\beta})\leq\mathbf{0}$ (so $\boldsymbol{m}(\boldsymbol{\beta}_0)=\mathbf{0}$), where $\boldsymbol{m}(\cdot)$ is a $d_\leq$-vector of sufficiently smooth functions. For any feasible $\boldsymbol{\beta}\in\mathbb{R}^{d_\beta}$ close to $\boldsymbol{\beta}_0$ such that $\boldsymbol{\beta}-\boldsymbol{\beta}_0\in\Delta\cap\mathcal{B}(\mathbf{0},\varpi^\star_\delta)$,

$$\boldsymbol{m}(\boldsymbol{\beta})=\boldsymbol{m}(\boldsymbol{\beta})-\boldsymbol{m}(\boldsymbol{\beta}_0)=\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{m}(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\Big[\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{m}(\check{\boldsymbol{\beta}})\Big](\boldsymbol{\beta}-\boldsymbol{\beta}_0)$$

for some point $\check{\boldsymbol{\beta}}$ between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$, where $\partial^2\boldsymbol{m}(\check{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is a $d_\beta\times d_\beta\times d_\leq$ array, with each sheet the second-order derivative matrix for one constraint function. Thus, $(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'[\partial^2\boldsymbol{m}(\check{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'](\boldsymbol{\beta}-\boldsymbol{\beta}_0)$ is a $d_\leq$-vector, with each element corresponding to $(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'[\partial^2 m_\ell(\check{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'](\boldsymbol{\beta}-\boldsymbol{\beta}_0)$ for $1\leq\ell\leq d_\leq$.

Let $\Gamma_0=\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{m}(\boldsymbol{\beta}_0)$ and $\widehat{\Gamma}=\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{m}(\widehat{\boldsymbol{\beta}})$. Assume Γ_0 is invertible in the neighborhood of $\boldsymbol{\beta}_0$. (When $d_\leq<d_\beta$, we can complement Γ_0 with additional rows and construct an invertible one, which is formalized in the proof of Lemma S.1). Then,

$$\Gamma_0^{-1}\boldsymbol{m}(\boldsymbol{\beta})-(\boldsymbol{\beta}-\boldsymbol{\beta}_0)=\frac{1}{2}\Gamma_0^{-1}\Big[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{m}(\check{\boldsymbol{\beta}})(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Big]=:\mathscr{L}$$

It can be shown that $\boldsymbol{\lambda}:=\Gamma_0^{-1}\boldsymbol{m}(\boldsymbol{\beta})$ is (approximately) in $\widehat{\Delta}$. To see this,

$$\boldsymbol{m}(\widehat{\boldsymbol{\beta}}+\boldsymbol{\lambda})\approx\boldsymbol{m}(\widehat{\boldsymbol{\beta}})+\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{m}(\boldsymbol{\beta}_0)\boldsymbol{\lambda}=\boldsymbol{m}(\widehat{\boldsymbol{\beta}})+(\boldsymbol{m}(\boldsymbol{\beta})-\boldsymbol{m}(\boldsymbol{\beta}_0))\leq\boldsymbol{m}(\widehat{\boldsymbol{\beta}}).$$

This is formalized in the proof of Lemma S.1. For the moment, assume $\boldsymbol{\lambda}\in\widehat{\Delta}$. Then,

$$\begin{aligned}\boldsymbol{m}(\boldsymbol{\beta}-\boldsymbol{\beta}_0+\widehat{\boldsymbol{\beta}})-\boldsymbol{m}(\widehat{\boldsymbol{\beta}})&=\boldsymbol{m}(\widehat{\boldsymbol{\beta}}+\boldsymbol{\lambda}-\mathscr{L})-\boldsymbol{m}(\widehat{\boldsymbol{\beta}})\\&\approx\boldsymbol{m}(\widehat{\boldsymbol{\beta}}+\boldsymbol{\lambda})-\Gamma_0\mathscr{L}-\boldsymbol{m}(\widehat{\boldsymbol{\beta}})\\&\leq-\Gamma_0\mathscr{L}\\&\approx-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{m}(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0).\end{aligned}$$

The errors introduced by "≈" are of smaller order under mild conditions. The above calculation implies that any $\boldsymbol{\beta}-\boldsymbol{\beta}_0\in\Delta\cap\mathcal{B}(\mathbf{0},\varpi^\star_\delta)$ (approximately) belongs to the following adjusted version of $\widehat{\Delta}$ in (S.2.7):

$$\Big\{\boldsymbol{\delta}:\boldsymbol{m}(\widehat{\boldsymbol{\beta}}+\boldsymbol{\delta})\leq\boldsymbol{m}(\widehat{\boldsymbol{\beta}})+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{m}(\boldsymbol{\beta}_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Big\}.$$

Then, for each inequality constraint $m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}+\boldsymbol{\delta})\leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}})$, the "adjustment" to the upper bound is at most

$$\frac{1}{2}s_{\max}\Big(\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}m_{\leq,\ell}(\boldsymbol{\beta}_0)\Big)\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|_2^2.$$

Given the high-probability bound ϱ for $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2$, we propose the adjustment in (S.2.10).

Notes: Our enlargement strategy used to define $\Delta^\star$ in Section S.2.2 can be conceptually understood as follows. A constraint set Δ can be generally written as $\Delta=\cap_{j=1}^{L}A_j$ for a sequence of convex sets A_j. Each A_j is defined by one inequality. Accordingly, $\widehat{\Delta}$ in (S.2.7) (with no enlargement) can be written as $\widehat{\Delta}=\cap_{j=1}^{L}\widehat{A}_j$.

Let $\widehat{A}_{j,\varepsilon_j}$ be an ε_j-enlargement of $\widehat{A}_j$. We can apply Lemma S.1 to each pair of A_j and $\widehat{A}_j$, and then $A_j\subseteq\widehat{A}_{j,\varepsilon_j}$ for an appropriate (small) ε_j. For any point $g\in\Delta$, $g\in A_j$ for every j. So $g\in\widehat{A}_{j,\varepsilon_j}$. Then, $g\in\cap_{j=1}^{L}\widehat{A}_{j,\varepsilon_j}$. Since it holds for any g, we have $\Delta\subseteq\Delta^\star:=\cap_{j=1}^{L}\widehat{A}_{j,\varepsilon_j}$. The sufficient condition (iii) in Theorem S.1 holds.

S.7 Second Order Cone Programming

In this section, we first define three types of convex optimization problems that we will be relying on (for background knowledge and technical details, see Boyd and Vandenberghe (2004)). Second, we illustrate the link between these three families of convex problems. Third, we show that the optimization problems underlying the prediction/estimation and uncertainty quantification problems for SC presented in Section S.2 are quadratically constrained quadratic program (QCQP) and quadratically constrained linear problem (QCLP), respectively, and show how to represent them as second-order cone program (SOCP). Finally, we provide two examples by showing how to write the L1-L2-type and Lasso-type constraints in conic form. These approaches are implemented in our companion general-purpose software (Cattaneo, Feng, Palomba and Titiunik, 2025), where we show that they lead to remarkable speed and scalability improvements.

S.7.1 Families of Convex Optimization Problems

QCQPs and QCLPs. A quadratically constrained quadratic program is an optimization problem of with the following form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}'\mathbf{P}_0\mathbf{x} + \mathbf{q}_0'\mathbf{x} + w && \text{(S.7.1)} \\ \text{subject to} \quad & \mathbf{x}'\mathbf{P}_j\mathbf{x} + \mathbf{q}_j'\mathbf{x} + r_j \leq 0, \quad j = 1, \dots, m, && \text{(Quadratic inequality constraint)} \\ & \mathbf{F}\mathbf{x} = \mathbf{g}, && \text{(Linear equality constraint)} \end{aligned}$$

where $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_m \in \mathcal{M}_{n\times n}(\mathbb{R})$, $\mathbf{q}_0, \mathbf{q}_1, \dots, \mathbf{q}_m \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{F} \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathbf{g} \in \mathbb{R}^m$, and $r_0, r_1, \dots, r_m, w \in \mathbb{R}$. If all the matrices $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_m$ are positive semi-definite the QCQP is convex. Moreover, if $\mathbf{P}_0 = \mathbf{0}$ the QCQP becomes a QCLP. For this reason, in what follows we will restrict our attention to QCQPs as they naturally embed QCLPs.

SOCPs. To define a SOCP, it is necessary to first give the definition of a second-order cone and then introduce the notion of associated *generalized inequality.*

Second-order cone definition. A set $\mathcal{C}$ is called a *cone* if for every $\mathbf{x} \in \mathcal{C}$ and $\alpha \geq 0$ we have $\alpha\mathbf{x} \in \mathcal{C}$. A set $\mathcal{C}$ is a *convex cone* if it is convex and a cone, i.e. if $\forall\, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$,and $\forall\, \alpha_1, \alpha_2 \geq 0$, we have

$$\alpha_1\mathbf{x}_1 + \alpha\mathbf{x}_2 \in \mathcal{C}.$$

Now, consider any norm $||\cdot||$ defined on the Euclidean space $\mathbb{R}^n$. The *norm cone* associated with the norm $||\cdot||$ is defined to be the set

$$\mathcal{C} = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} : ||\mathbf{x}|| \leq t\}$$

and it is a convex cone by the standard properties of the norms. A *second-order cone* is the associated norm cone for the Euclidean norm and it is typically defined as

$$\mathcal{C} = \left\{(\mathbf{x}, t) \in \mathbb{R}^{n+1} : \|\mathbf{x}\|_2 \leq t\right\} = \left\{ \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} : \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix}' \begin{bmatrix} \mathbf{I} & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ t \end{bmatrix} \leq 0,\ t \geq 0 \right\}$$

Generalized inequality. A cone $\mathcal{C}$ is *solid* if it has non-empty interior and it is **pointed** if $\mathbf{x} \in \mathcal{C}, -\mathbf{x} \in \mathcal{C}$ implies that $\mathbf{x} = 0$. We say that a cone $\mathcal{C}$ is *proper* if is it convex, closed, solid, and pointed. Proper cones in the Euclidean space $\mathbb{R}^n$ are useful because they induce a partial ordering that enjoys almost all the properties of the basic one in $\mathbb{R}$. Therefore, given a cone $\mathcal{C} \subseteq \mathbb{R}^n$, we can define the generalized inequality $\preceq_{\mathcal{C}}$ for any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbf{x} \preceq_{\mathcal{C}} \mathbf{y} \iff \mathbf{y} - \mathbf{x} \in \mathcal{C}.$$

From this definition we can see that quadratic constraints such as $||\mathbf{x}||_2 \leq h$ can be re-written as second-order cone constraints of the form $\mathbf{x} \preceq_{\mathcal{C}} h$ for some second-order cone $\mathcal{C}$. Note that if $\mathcal{C} = \mathbb{R}^m_+$ then if $m = 1$, $\preceq_{\mathcal{C}}$ is the standard inequality $\leq$ in $\mathbb{R}$, whereas if $m > 1$, $\preceq_{\mathcal{C}}$ is the component-wise inequality in $\mathbb{R}^m$.

Second-order cone program. Let $\mathcal{K}$ be a cone such that $\mathcal{K} = \mathbb{R}^m_+ \times \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_L$ where $\mathcal{K}_l := \{(k_0, \mathbf{k}_1) \in \mathbb{R} \times \mathbb{R}^l : ||\mathbf{k}_1||_2 \leq k_0\}, l = 1, \ldots, L$. Let $\preceq_{\mathcal{K}}$ be the *generalized inequality* associated with the cone $\mathcal{K}$. An optimization problem is called *second-order cone program* if it has the following form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}'\mathbf{x}, && \text{(S.7.2)} \\ \text{subject to} \quad & \mathbf{G}\mathbf{x} \preceq_{\mathcal{K}} \mathbf{h}, && \text{(Second-order cone constraint)} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. && \text{(Linear equality constraint)} \end{aligned}$$

S.7.2 Link Between QCQP and SOCPs

Any QCQP can be converted to a SOCP (Boyd and Vandenberghe, 2004). In other words, we can always rewrite an optimization problem such as (S.7.1) in the form of (S.7.2). First, we present the general result and then we explain all the necessary steps to reformulate QCQPs as SOCPs. Without loss of generality, assume that $w = 0$ in (S.7.1) and, to ease notation, let $m = 1$ so that there is only a single quadratic inequality constraint. Moreover, given any positive semi-definite matrix $\mathbf{P}$, let $\mathbf{P}^{1/2}$ be the square root of $\mathbf{P}$, that is the unique symmetric positive semi-definite matrix $\mathbf{R}$ such that $\mathbf{R}\mathbf{R} = \mathbf{R}'\mathbf{R} = \mathbf{P}$. Then for any QCQP the following two formulations are equivalent

QCQP

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}'\mathbf{P}_0\mathbf{x} + \mathbf{q}_0'\mathbf{x} \\ \text{subject to} \quad & \mathbf{F}\mathbf{x} = \mathbf{g}, \\ & \mathbf{x}'\mathbf{P}_1\mathbf{x} + \mathbf{q}_1'\mathbf{x} + r_1 \leq 0. \end{aligned}$$

SOCP

$$\begin{aligned} \min_{\mathbf{x}, v, t, s} \quad & v + \mathbf{q}_0'\mathbf{x} \\ \text{subject to} \quad & \mathbf{F}\mathbf{x} = \mathbf{g}, \\ & t + \mathbf{q}_1'\mathbf{x} + r_1 \preceq_{\mathbb{R}_+} 0, \\ & \mathbf{P}_0^{1/2}\mathbf{x} \preceq_{K_{1+n}} v, \\ & \mathbf{P}_1^{1/2}\mathbf{x} \preceq_{K_{1+n}} s. \end{aligned}$$

We can see that the logic beneath the conversion of a QCQP into a SOCP is to “linearize” all the non-linear terms appearing either in the objective function or in the inequality constraints. The “linearization” step does come at a cost, as it requires the introduction of a slack variable every time we rely on it. Indeed, above we linearized the objective function and the quadratic inequality constraint by introducing two auxiliary slack variables.

More formally, let $\mathbf{x}'\mathbf{P}\mathbf{x}$ be a symmetric positive semi-definite quadratic form and consider the constraint $\mathbf{x}'\mathbf{P}\mathbf{x} \leq y$. Then

(i) Since $\mathbf{P}$ is symmetric positive semi-definite the epigraph $\mathbf{x}'\mathbf{P}\mathbf{x} \leq y$ is a convex set and $\mathbf{P}^{1/2}$ is well-defined.

(ii) Write the inequality constraint as a constraint involving the Euclidean norm $\|\cdot\|_2$

$$y \geq \mathbf{x}'\mathbf{P}\mathbf{x} = \mathbf{x}'\mathbf{P}^{1/2}\mathbf{P}^{1/2}\mathbf{x} = ||\mathbf{P}^{1/2}\mathbf{x}||_2^2.$$

(iii) Note that

$$||\mathbf{P}^{1/2}\mathbf{x}||_2^2 \leq y \iff \left\|\begin{bmatrix} 1-y \\ 2\mathbf{P}^{1/2}\mathbf{x} \end{bmatrix}\right\|_2 \leq 1+y, \tag{S.7.3}$$

which can be verified by squaring the two sides of the last inequality and expand the norm.

(iv) More is true, as the right-most inequality in (S.7.3) defines the following second-order cone for given $\mathbf{P}^{1/2}$

$$\mathcal{C} = \left\{\left(1-y, 2\mathbf{P}^{1/2}\mathbf{x}, 1+y\right) : \left\|\begin{bmatrix} 1-y \\ 2\mathbf{P}^{1/2}\mathbf{x} \end{bmatrix}\right\|_2 \leq 1+y\right\},$$

which in turn induces the generalized inequality $\mathbf{P}^{1/2}\mathbf{x} \preceq_{\mathcal{C}} y$.

S.7.3 Specific Synthetic Control Problems as SOCPs

We illustrate the approach described above for the L1-L2 constraint and the Lasso constraint. To ease notation, we do not consider the regularization to the local geometry of Δ. Note that simplex, ridge, or least squares are particular cases of L1-L2. Throughout this section, we assume that $T_i \equiv T_0$ for all $i \in \mathcal{E}$.

L1-L2-type $\mathcal{W}$. Consider first the prediction/estimation SC optimization problem, which relies on the following program:

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{r}} \quad & (\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r})'\mathbf{V}(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r}) & \text{(S.7.4)}\\
\text{subject to} \quad & ||\mathbf{w}^{[i]}||_1 = 1, \quad i=1,\ldots,J_1, & \text{(L1 equality constraints)}\\
& ||\mathbf{w}^{[i]}||_2 \leq Q^{[i]}, \quad i=1,\ldots,J_1, & \text{(L2 inequality constraints)}\\
& \mathbf{w} \geq \mathbf{0}, & \text{(non-negativity constraint)}
\end{aligned}$$

where, as always, $\geq$ is understood as a component-wise inequality for vectors ($\mathbf{w} \in \mathbb{R}^{J_0 \cdot J_1}$). First, notice that the non-convex constraints $||\mathbf{w}^{[i]}||_1 = 1, i = 1, \ldots, J_1$ can be replaced with the convex constraints $\mathbf{1}'\mathbf{w}^{[i]} = 1, i = 1, \ldots, J_1$ because of the non-negativity constraint on the elements of $\mathbf{w}$. Then, we can cast (S.7.4) as a SOCP as follows

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{r},v,\{s_i\}_{i=1}^{J_1}} \quad & v & \\
\text{subject to} \quad & \mathbf{1}'\mathbf{w}^{[i]} = 1, & \text{(L1 equality constraints)}\\
& -\mathbf{w} \preceq_{\mathcal{C}_1} \mathbf{0}, & \text{(cone in } \mathbb{R}^{J_0\cdot J_1})\\
& \begin{bmatrix} 1-v \\ 2\mathbf{V}^{1/2}(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r}) \end{bmatrix} \preceq_{\mathcal{C}_2} 1+v, & \text{(cone in } \mathbb{R}^{2+\tilde{T}\cdot M})\\
& s_i \preceq_{\mathcal{C}_3} Q^{[i]}, \quad i=1,\ldots,J_1, & (J_1 \text{ cones in } \mathbb{R})
\end{aligned}$$

$$\begin{bmatrix} 1-s_i \\ 2\mathbf{w}^{[i]} \end{bmatrix} \preceq_{\mathcal{C}_4} 1+s_i, \quad i=1,\dots,J_1, \qquad (J_1 \text{ cones in } \mathbb{R}^{2+J_0})$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{C}_3^{J_1} \times \mathcal{C}_4^{J_1} = \mathbb{R}_+^{J_0 \cdot J_1} \times \mathcal{K}_{\tilde{T}\cdot M+1} \times \mathbb{R}_+^{J_1} \times \mathcal{K}_{J_0+1}^{J_1}$ is the conic constraint for this program.

For uncertainty quantification, we need to solve the optimization problem underlying (S.2.5). We discuss the lower bound only. Recalling that $\boldsymbol{\beta} = (\mathbf{w}', \mathbf{r}')'$, we have

$$\begin{aligned}
\inf_{\boldsymbol{\beta}=(\mathbf{w}',\mathbf{r}')'} \quad & \mathbf{p}_\tau'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) && \text{(S.7.5)} \\
\text{subject to} \quad & ||\mathbf{w}^{[i]}||_1 = 1, \quad i=1,\dots,J_1, && \text{(L1 equality constraints)} \\
& ||\mathbf{w}^{[i]}||_2 \leq Q^{[i]}, \quad i=1,\dots,J_1, && \text{(L2 inequality constraints)} \\
& \mathbf{w} \geq \mathbf{0}, && \text{(non-negativity constraint)} \\
& (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'\widehat{\mathbf{Q}}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) - 2(\mathbf{G}^\star)'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \leq 0. && \text{(constrained least squares)}
\end{aligned}$$

We can cast the SC optimization problem in (S.7.5) in conic form as follows:

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{r},\{s_i\}_{i=1}^{J_1},t} \quad & \mathbf{p}_\tau'\boldsymbol{\beta} \\
\text{subject to} \quad & \mathbf{1}'\mathbf{w}^{[i]} = 1, \quad i=1,\dots,J_1, && \text{(L1 equality constraints)} \\
& t + \mathbf{a}'\boldsymbol{\beta} + f \preceq_{\mathcal{C}_1} 0, && \text{(cone in } \mathbb{R}) \\
& -\mathbf{w} \preceq_{\mathcal{C}_2} \mathbf{0}, && \text{(cone in } \mathbb{R}^{J_0\cdot J_1}) \\
& s_i \preceq_{\mathcal{C}_3} Q^{[i]}, \quad i=1,\dots,J_1, && (J_1 \text{ cones in } \mathbb{R}) \\
& \begin{bmatrix} 1-s_i \\ 2\mathbf{w}^{[i]} \end{bmatrix} \preceq_{\mathcal{C}_4} 1+s_i, \quad i=1,\dots,J_1, && (J_1 \text{ cones in } \mathbb{R}^{2+J_0}) \\
& \begin{bmatrix} 1-t \\ 2\mathbf{Q}^{1/2}\boldsymbol{\beta} \end{bmatrix} \preceq_{\mathcal{C}_5} 1+t, && \text{(cone in } \mathbb{R}^{2+(J_0+KM)\cdot J_1})
\end{aligned}$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2 \times \mathcal{C}_3 \times \mathcal{C}_4 \times \mathcal{C}_5 = \mathbb{R}_+ \times \mathbb{R}_+^{J_0\cdot J_1} \times \mathbb{R}_+^{J_1} \times \mathcal{K}_{1+J_0}^{J_1} \times \mathcal{K}_{1+(J_0+KM)\cdot J_1}$ is the conic constraint for this program, $\mathbf{a} = -2('\mathbf{Q}\widehat{\boldsymbol{\beta}} + \mathbf{G}^\star)'$, and $f = \widehat{\boldsymbol{\beta}}'\mathbf{Q}\widehat{\boldsymbol{\beta}} + 2\mathbf{G}^\star\widehat{\boldsymbol{\beta}}$.

Lasso-type $\mathcal{W}$. We show how to write the QCQP as a SOCP when $\mathcal{W}$ has a lasso-type constraint. In this case, the SC weight construction (3.1) has the form:

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{r}} \quad & (\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r})'\mathbf{V}(\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r}) && \text{(S.7.6)} \\
\text{subject to} \quad & ||\mathbf{w}^{[i]}||_1 \leq 1, \quad i=1,\dots,J_1. && \text{(L1 inequality constraints)}
\end{aligned}$$

We can write the optimization problem in (S.7.6) as a SOCP of the following form

$$\begin{aligned}
\min_{\mathbf{w},\mathbf{r},\{\mathbf{z}_i\}_{i=1}^{J_1},v} \quad & v \\
\text{subject to} \quad & \begin{bmatrix} 1-v \\ 2\mathbf{V}^{1/2}(\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r}) \end{bmatrix} \preceq_{\mathcal{C}_1} 1+v, && \text{(cone in } \mathbb{R}^{2+T_0\cdot M\cdot J_1}) \\
& \mathbf{1}'\mathbf{z}_i \preceq_{\mathcal{C}_2} 1, \quad i=1,\dots,J_1, && (J_1 \text{ cones in } \mathbb{R}) \\
& -\mathbf{w} \preceq_{\mathcal{C}_3} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J_0\cdot J_1})
\end{aligned}$$

$$\mathbf{w} \preceq_{\mathcal{C}_4} \mathbf{z}, \qquad \text{(cone in } \mathbb{R}^{J_0 \cdot J_1})$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2^{J_1} \times \mathcal{C}_3 \times \mathcal{C}_4 = \mathcal{K}_{1+T_0 \cdot M \cdot J_1} \times \mathbb{R}_+^{J_1} \times \mathbb{R}_+^{J_0 \cdot J_1} \times \mathbb{R}_+^{J_0 \cdot J_1}$ is the conic constraint for this program and $\mathbf{z} := (\mathbf{z}_1', \ldots, \mathbf{z}_{J_1}')'$.

For uncertainty quantification, we need to solve the optimization problem underlying (S.2.5). Here we discuss the lower bound only for brevity. Recalling that $\boldsymbol{\beta} = (\mathbf{w}', \mathbf{r}')'$, we have

$$\begin{aligned}
&\inf_{\boldsymbol{\beta}=(\mathbf{w}',\mathbf{r}')'} \quad \mathbf{p}_\tau'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) && \text{(S.7.7)} \\
&\text{subject to} \quad ||\mathbf{w}^{[i]}||_1 \leq 1, \quad i = 1, \ldots, J_1, && \text{(L1 inequality constraints)} \\
&\qquad (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'\widehat{\mathbf{Q}}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) - 2(\mathbf{G}^\star)'(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \leq 0. && \text{(constrained least squares)}
\end{aligned}$$

We can cast the SC optimization problem in (S.7.7) in conic form as follows:

$$\begin{aligned}
&\min_{\mathbf{w},\mathbf{r},\{\mathbf{z}_i\}_{i=1}^{J_1},t} \quad \mathbf{p}_\tau'\boldsymbol{\beta} \\
&\text{subject to} \quad t + \mathbf{a}'\boldsymbol{\beta} + f \preceq_{\mathcal{C}_1} 0, && \text{(cone in } \mathbb{R}) \\
&\qquad \mathbf{1}'\mathbf{z}^{[i]} \preceq_{\mathcal{C}_2} 1, \quad i = 1, \ldots, J_1, && (J_1 \text{ cones in } \mathbb{R}) \\
&\qquad -\mathbf{w} \preceq_{\mathcal{C}_3} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J_0 \cdot J_1}) \\
&\qquad \mathbf{w} \preceq_{\mathcal{C}_4} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J_0 \cdot J_1}) \\
&\qquad \begin{bmatrix} 1 - t \\ 2\mathbf{Q}^{1/2}\boldsymbol{\beta} \end{bmatrix} \preceq_{\mathcal{C}_5} 1 + t, && \text{(cone in } \mathbb{R}^{2+(J_0+KM)\cdot J_1})
\end{aligned}$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2^{J_1} \times \mathcal{C}_3 \times \mathcal{C}_4 \times \mathcal{C}_5 = \mathbb{R}_+ \times \mathbb{R}_+^{J_1} \times \mathbb{R}_+^{J_0 \cdot J_1} \times \mathbb{R}_+^{J_0 \cdot J_1} \times \mathcal{K}_{1+(J_0+KM)\cdot J_1}$ is the conic constraint for this program, $\mathbf{a} = -2('\mathbf{Q}\widehat{\boldsymbol{\beta}} + \mathbf{G}^\star)'$, and $f = \widehat{\boldsymbol{\beta}}'\mathbf{Q}\widehat{\boldsymbol{\beta}} + 2\mathbf{G}^\star\widehat{\boldsymbol{\beta}}$.

S.8 Data Preparation and Software Implementation

In this section, we first describe the variables in the Billmeier and Nannicini (2013) (BN, henceforth) dataset and then go through the details of our empirical specification.

S.8.1 Data Description

The original BN dataset contains data on some economic and political variables for 180 countries, over a period of time spanning from 1960 to 2005.[1] In detail, the variables available in the dataset are:

- real GDP per capita in 2002 US dollars.
- enrollment rate in secondary schooling.
- population growth.
- yearly inflation rate.
- the investment ratio (the investment of a country as a percentage of GDP).
- an indicator that captures whether the country is a democracy (1) or not (0).
- an indicator that captures whether the economy of the country is considered closed (0) or not (1) as developed in Sachs, Warner, Åslund and Fischer (1995) (henceforth, Sachs-Warner indicator). In particular, the indicator takes value 0 if any of the following conditions is verified:
  - i) the average tariff is above 40%;
  - ii) non-tariff barriers are imposed on a volume of imports larger than 40%;
  - iii) the country has a socialist economic system;
  - iv) the exchange rate black market premium is above 20%;
  - v) state monopolies control most of the country exports.

Despite having six candidate variables to match on, we end up matching on at most two variables because of missing data. Specifically, we always match on real GDP per-capita, whereas in the robustness exercise in Supplemental Appendix S.9 we also match on investment-to-GDP ratio.

A second difference of our final dataset from the original one used in BN is the final pool of countries - treated and donors - on which we conduct the analysis. In particular, we adopt the following criteria to select the countries to be included in our final dataset as either donors or treated units:

1. We restrict the analysis to countries in Sub-Saharan Africa. We define this group by excluding from the analysis North African countries according to the United Nations (UN) classification (Algeria, Egypt, Libya, Morocco, Sudan, Tunisia, and Western Sahara) as well as any additional country that is not covered by the UN classification but is a member of the Arab League (Djibouti, Mauritania, Somalia).

[1] We downloaded the dataset from the Harvard Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28699.

2. As in Bratton and Van de Walle (1997), we exclude Namibia because the Sachs-Werner indicator is missing for such a country.

Table S.3 shows the final set of countries we select, together with their treatment date.

Table S.3: *List of all countries included in our analysis*

Angola	∞	Lesotho	∞
Benin	1990	Madagascar	1996
Botswana	1979	Malawi	∞
Burkina Faso	1998	Mali	1988
Burundi	1999	Mauritius	1968
Cabo Verde	1991	Mozambique	1995
Cameroon	1993	Niger	1994
Chad	∞	Nigeria	∞
Congo	∞	Rwanda	∞
Ethiopia	1996	Senegal	∞
Gabon	∞	Sierra Leone	∞
Gambia	1985	South Africa	1991
Ghana	1985	Tanzania	1995
Guinea	1986	Togo	∞
Guinea-Bissau	1987	Uganda	1988
Ivory Coast	1994	Zambia	1993
Kenya	1993	Zimbabwe	∞

Notes: ∞ denotes that a country has never experienced liberalization during the observed time span.

S.8.2 Implementation Details

In this section, we describe all the details of the empirical application presented in the main text.

Constraint type. Our preferred specification uses the L1-L2 constraint, i.e.,

$$\mathcal{W}_{\mathtt{L1-L2}} = \bigtimes_{i=1}^{J_1} \left\{ \mathbf{w}^{[i]} \in \mathbb{R}_+^{J_0} : ||\mathbf{w}^{[i]}||_1 = 1, ||\mathbf{w}^{[i]}||_2 \leq Q^{[i]} \right\},$$

whereas the results in the Supplemental Appendix Section S.9.2 and S.9.1 use simplex- and Ridge-type constraints of the form

$$\mathcal{W}_{\mathtt{S}} = \bigtimes_{i=1}^{J_1} \left\{ \mathbf{w}^{[i]} \in \mathbb{R}_+^{J_0} : ||\mathbf{w}^{[i]}||_1 = 1 \right\}, \qquad \mathcal{W}_{\mathtt{R}} = \bigtimes_{i=1}^{J_1} \left\{ \mathbf{w}^{[i]} \in \mathbb{R}^{J_0} : ||\mathbf{w}^{[i]}||_2 \leq Q^{[i]} \right\}.$$

Table S.4 shows the effective values for $Q^{[i]}, i = 1, \ldots, J_1$ that we compute in our empirical application. Further below we explain in greater detail how these regularization parameters are computed in practice.

Selected features. Our main specification uses only one feature ($M = 1$)–the logarithm of real GDP per-capita–, uses the identify weighting matrix, and includes a constant term, that is

$$\mathbf{B}^{[i]} = [\mathbf{Y}_1 \quad \cdots \quad \mathbf{Y}_{J_0}], \quad \mathbf{C}^{[i]} = \mathbf{1}_{T_i}, \quad \mathcal{R} = \mathbb{R}^{J_1}, \quad \mathbf{V}^{[i]} = \mathbf{I}_{T_i}, \quad i = 1, \ldots, J_1,$$

where $\mathbf{Y}_j = (Y_{j1}, \ldots, Y_{jT_i})', j = 1, \ldots, J_0$ is the pre-treatment log-GDP per-capita of the j-th donor, $\mathbf{1}_{T_i}$ is a $T_i \times 1$ vector of ones, and $\mathbf{I}_{T_i}$ is the $T_i \times T_i$ identify matrix.

In Supplemental Appendix Section S.9.3, we present results using two features ($M = 2$), where we also match on the investment-to-GDP ratio and control for a feature-specific constant term and linear trend, i.e.

$$\mathbf{B}^{[i]} = \begin{bmatrix} \mathbf{Y}_1 & \cdots & \mathbf{Y}_{J_0} \\ \mathbf{IR}_1 & \cdots & \mathbf{IR}_{J_0} \end{bmatrix}, \quad \mathbf{C}^{[i]} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T_i \end{bmatrix}, \quad \mathcal{R} = \mathbb{R}^{2 \cdot J_1}, \quad \mathbf{V}^{[i]} = \begin{bmatrix} \mathbf{S}_Y^{[i]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{IR}^{[i]} \end{bmatrix}, \quad i = 1, \ldots, J_1,$$

where $\mathbf{IR}_j = (IR_{j1}, \ldots, IR_{jT_i})', j = 1, \ldots, J_0$ is the pre-treatment investment-to-GDP ratio of the j-th donor, $\mathbf{S}_Y^{[i]} = \text{diag}(\widehat{\sigma}_{Y,1}^{-1}, \ldots, \widehat{\sigma}_{Y,T_i}^{-1})$, $\mathbf{S}_{IR}^{[i]} = \text{diag}(\widehat{\sigma}_{IR,1}^{-1}, \ldots, \widehat{\sigma}_{IR,T_i}^{-1})$, with

$$\widehat{\sigma}_{W,t} = \left(\frac{1}{J_0 - 1} \sum_{j=1}^{J_0} (W_{jt} - \overline{W}_t)^2 \right)^{1/2}, \quad \overline{W}_t = \frac{1}{J_0} \sum_{i=1}^{J_0} W_{jt}, \quad t = 1, \ldots, T_i, \quad W \in \{Y, IR\},$$

and $\text{diag}(\mathbf{x})$ yields a square diagonal matrix with the elements of $\mathbf{x}$ on its main diagonal.

Tuning parameters. Regarding the choice of $Q^{[i]}$, it is well-established that the Ridge regression problem can be equivalently expressed both as an unconstrained penalized optimization problem and as a constrained optimization problem. For simplicity, assume $\mathbf{C}$ is not included and $M = 1$. The two Ridge-type optimization formulations are as follows:

$$\widehat{\mathbf{w}}^{[i]} = \arg \min_{\mathbf{w}^{[i]} \in \mathcal{W}} (\mathbf{A}^{[i]} - \mathbf{B}^{[i]}\mathbf{w}^{[i]})' \mathbf{V}^{[i]} (\mathbf{A}^{[i]} - \mathbf{B}^{[i]}\mathbf{w}^{[i]}) + \lambda^{[i]} ||\mathbf{w}^{[i]}||_2^2,$$

where $\lambda^{[i]} \geq 0$ is a regularization parameter, and

$$\widehat{\mathbf{w}}^{[i]} = \arg \min_{\mathbf{w}^{[i]} \in \mathcal{W}, ||\mathbf{w}^{[i]}||_2^2 \leq \left(Q^{[i]}\right)^2} (\mathbf{A}^{[i]} - \mathbf{B}^{[i]}\mathbf{w}^{[i]})' \mathbf{V}^{[i]} (\mathbf{A}^{[i]} - \mathbf{B}^{[i]}\mathbf{w}^{[i]}),$$

where $Q^{[i]} \geq 0$ is an explicit upper bound on the norm of $\mathbf{w}^{[i]}$. Under the assumption of Gaussian errors, an optimal choice of the regularization parameter $\lambda^{[i]}$ for risk minimization, as suggested by Hoerl et al. (1975), is:

$$\lambda^{[i]} = \frac{J_0 (\widehat{\sigma}_{\text{OLS}}^{[i]})^2}{||\widehat{\mathbf{w}}_{\text{OLS}}^{[i]}||_2^2},$$

where $(\widehat{\sigma}_{\text{OLS}}^{[i]})^2$ and $\widehat{\mathbf{w}}_{\text{OLS}}^{[i]}$ are the estimates of the residual variance and the coefficients from the ordinary least squares (OLS) regression of $\mathbf{A}^{[i]}$ onto $\mathbf{B}^{[i]}$, respectively. Given the two optimization problems above, there exists a one-to-one correspondence between $\lambda^{[i]}$ and $Q^{[i]}$. For example, assuming the columns of $\mathbf{B}^{[i]}$ are orthonormal, the closed-form solution for the Ridge estimator is:

$$\widehat{\mathbf{w}}^{[i]} = (\mathbf{I} + \lambda^{[i]}\mathbf{I})^{-1} \widehat{\mathbf{w}}_{\text{OLS}}^{[i]},$$

and if the constraint on the ℓ_2-norm is active, we have $Q^{[i]} = \|\widehat{\mathbf{w}}^{[i]}\|_2 = \|\widehat{\mathbf{w}}_{\text{OLS}}^{[i]}\|_2 / (1 + \lambda^{[i]})$. When more than one feature is considered (i.e., $M > 1$), we compute the constraint size $Q_\ell^{[i]}$ for each feature $\ell = 1, \ldots, M$, and then choose $Q^{[i]}$ as the most restrictive constraint to promote shrinkage of $\mathbf{w}^{[i]}$:

$$Q^{[i]} := \min_{\ell = 1, \ldots, M} Q_\ell^{[i]}.$$

Table S.4: *Pre-treatment length and values of the regularization parameters* $Q^{[i]}, i = 1, \ldots, J_1$.

Treated Unit	$Q^{[i]}, M = 1$	$Q^{[i]}, M = 2$	$\mathcal{T}_i$	$\lvert\mathcal{T}_i\rvert$
Benin	2.041	0.699	1963 - 1989	27
Botswana	2.119	0.771	1963 - 1978	16
Cabo Verde	0.877	0.5	1963 - 1990	28
Cameroon	2.336	0.786	1963 - 1992	30
Gambia	1.401	1.035	1963 - 1984	22
Ghana	0.858	0.972	1963 - 1984	22
Guinea	0.591	0.65	1963 - 1985	23
Guinea-Bissau	0.942	0.718	1963 - 1986	24
Ivory Coast	1.212	0.838	1963 - 1993	31
Kenya	0.959	0.569	1963 - 1992	30
Mali	1.445	0.65	1963 - 1987	25
Mauritius	1.705	1.145	1963 - 1967	5
Niger	2.252	1.793	1963 - 1993	31
South Africa	1.169	0.5	1963 - 1990	28
Uganda	1.797	0.5	1963 - 1987	25
Zambia	2.554	0.5	1963 - 1992	30

Notes: the rule of thumb to compute $Q^{[i]}$ is the same for $L1$-$L2$ and Ridge. Moreover, the values for $Q^{[i]}$ reported are the ones obtained when predicting τ_{it} and $\tau_{i\cdot}$ which are identical by construction. Finally, $Q^{[i]}$ is lower-bounded at 0.5 to avoid excessive shrinkage. More details on the rules of thumb used can be found in Cattaneo, Feng, Palomba and Titiunik (2025), Section 3.1.

In-sample Uncertainty. In order to quantify the in-sample uncertainty from estimating the SC weights, we need to construct the bounds $\underline{M}_{\texttt{in}}$ and $\overline{M}_{\texttt{in}}$ on $\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. The following strategy is adopted. First, we treat the synthetic control weights as possibly misspecified, thus estimating both the first and second conditional moments of the pseudo-true residuals $\mathbf{u}$. The conditional first moment $\mathbb{E}[\mathbf{u} \,|\, \mathscr{H}]$ is estimated feature-by-feature using a linear-in-parameters regression of the residual $\widehat{\mathbf{u}} = \mathbf{A} - \mathbf{B}\widehat{\mathbf{w}} - \mathbf{C}\widehat{\mathbf{r}}$ on $\mathbf{B}$ and the first lag of $\mathbf{B}$, whereas the conditional second moment $\mathbb{V}[\mathbf{u} \,|\, \mathscr{H}]$ is estimated with an HC1-type estimator. We then draw $S = 200$ i.i.d. random vectors from the Gaussian distribution $\mathsf{N}(0, \widehat{\boldsymbol{\Sigma}})$, conditional on the data, to simulate the criterion function $\ell^\star_{(s)}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) := (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\widehat{\mathbf{Q}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) - 2\mathbf{G}'_{(s)}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $s = 1, \ldots, 200$, and solve the following optimization problems

$$l_{(s)} := \inf_{\substack{\boldsymbol{\beta}-\boldsymbol{\beta}_0 \in \Delta^\star, \\ \ell^\star_{(s)}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \leq 0}} \mathbf{p}_\tau'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \qquad \text{and} \qquad u_{(s)} := \sup_{\substack{\boldsymbol{\beta}-\boldsymbol{\beta}_0 \in \Delta^\star, \\ \ell^\star_{(s)}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) \leq 0}} \mathbf{p}_\tau'(\boldsymbol{\beta} - \boldsymbol{\beta}_0),$$

where $\Delta^\star$ is constructed as explained in Section S.2.1. Finally, $\underline{M}_{\texttt{in}}$ is the $(\alpha_1/2)-$quantile of $\{l_{(s)}\}_{s=1}^S$ and $\overline{M}_{\texttt{in}}$ is the $(1 - \alpha_1/2)-$quantile of $\{u_{(s)}\}_{s=1}^S$, where α_1 is set to 0.05.

Out-of-sample Uncertainty. In order to quantify the out-of-sample uncertainty from the stochastic error in the post-treatment period, we need to construct the bounds $\underline{M}_{\texttt{out}}$ and $\overline{M}_{\texttt{out}}$ on the out-of-sample error e_τ (associated with the τ prediction). We employ the non-asymptotic bounds described in (S.2.6), assuming that $e_\tau - \mathbb{E}[e_\tau|\mathscr{H}]$ is sub-Gaussian conditional on $\mathscr{H}$. Then, we take

$$\underline{M}_{\texttt{out}} := \mathbb{E}[e_\tau|\mathscr{H}] - \sqrt{2\sigma^2_{\mathscr{H}} \log(2/\alpha_2)} \qquad \text{and} \qquad \overline{M}_{\texttt{out}} := \mathbb{E}[e_\tau|\mathscr{H}] + \sqrt{2\sigma^2_{\mathscr{H}} \log(2/\alpha_2)},$$

We set $\alpha_2 = 0.05$, and the conditional mean $\mathbb{E}[e_\tau|\mathscr{H}]$ and the sub-Gaussian parameter $\sigma_{\mathscr{H}}$ are parametrized and estimated by a linear-in-parameters regression of the pre-treatment residuals on $\mathbf{B}$.

Finally, the prediction intervals for the counterfactual outcome and the treatment effect of interest are given by

$$\left[\mathbf{p}_{\tau'}'\widehat{\boldsymbol{\beta}} - \overline{M}_{\mathtt{in}} + \underline{M}_{\mathtt{out}};\ \mathbf{p}_{\tau'}'\widehat{\boldsymbol{\beta}} - \underline{M}_{\mathtt{in}} + \overline{M}_{\mathtt{out}}\right] \quad \text{and} \quad \left[\widehat{\tau} + \underline{M}_{\mathtt{in}} - \overline{M}_{\mathtt{out}};\ \widehat{\tau} + \overline{M}_{\mathtt{in}} - \underline{M}_{\mathtt{out}}\right],$$

respectively.

Other assumptions. Throughout all our specifications, we maintain the assumptions that (i) there is no anticipation of the treatment and (ii) $\mathbf{A}$ and $\mathbf{B}$ form a cointegrated system. When (i) is relaxed to allow for anticipation results remain qualitatively the same.

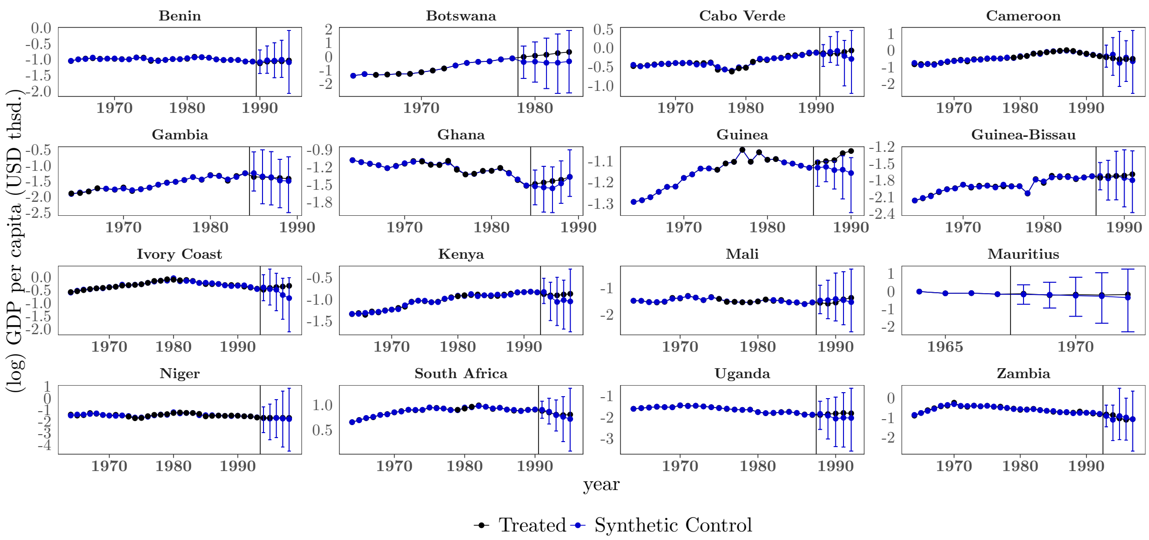
S.9 Robustness Checks

In this Section, we redo the empirical analysis presented in the main paper with five variations: (*i*) using a simplex-type constraint in place of the L1-L2 constraint (Section S.9.1); (*ii*) using a Ridge-type constraint in place of the L1-L2 constraint (Section S.9.2; (*iii*) using the investment-to-GDP ratio as an additional feature (Section S.9.3); (*iv*) using a placebo treatment date (Section S.9.4; and (*v*) leaving one donor out at the time (Section S.9.5).

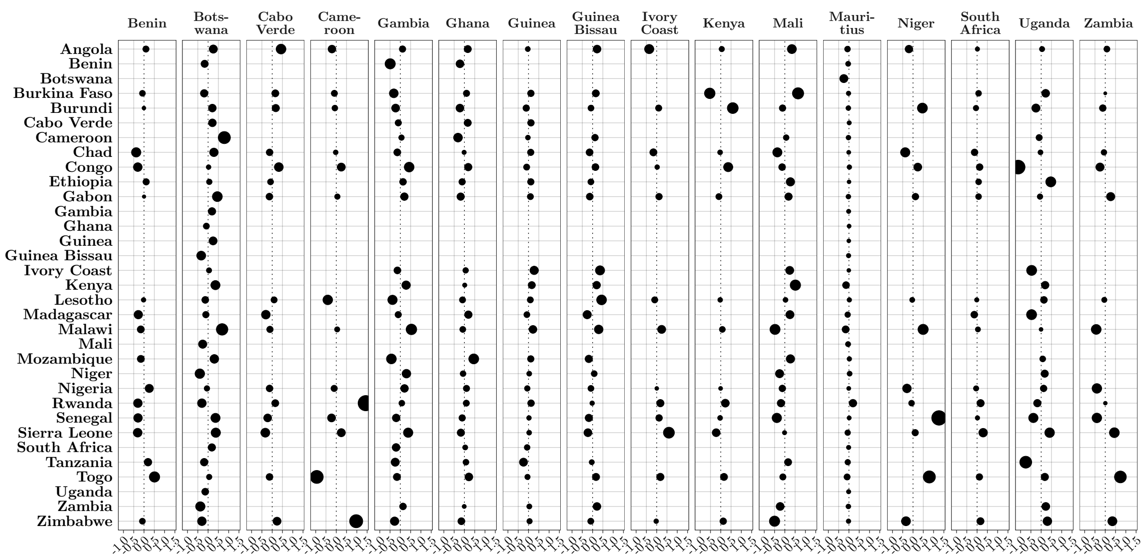
S.9.1 Ridge-Type Constraint

TSUS predicted effects in every period after liberalization (τ_{ik}).

Figure S.1: *Time-specific unit-specific (TSUS) predicted effects in every period,* $\widehat{\tau}_{ik}$*.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

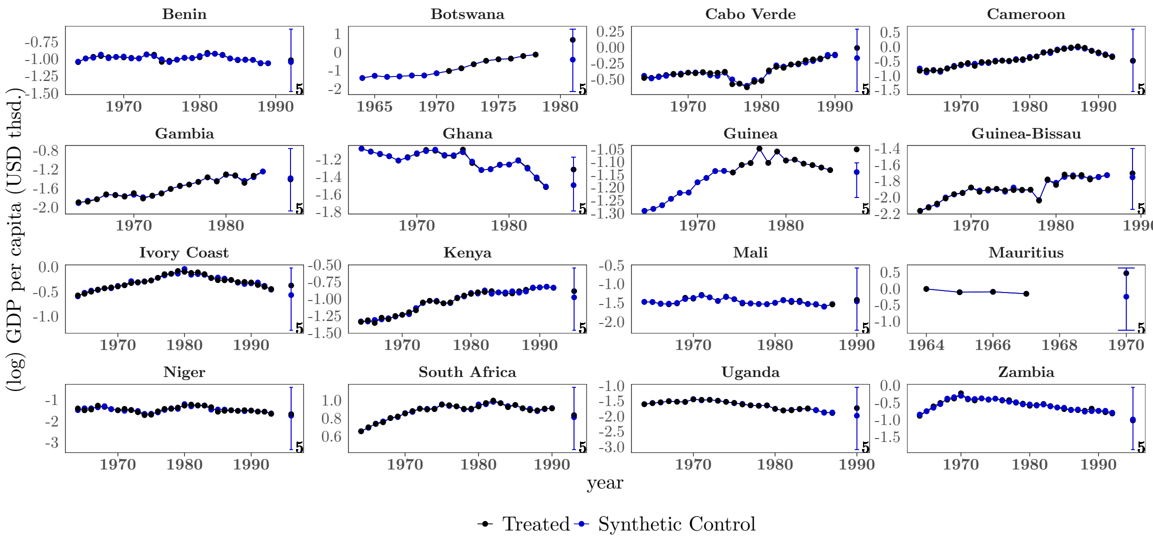


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

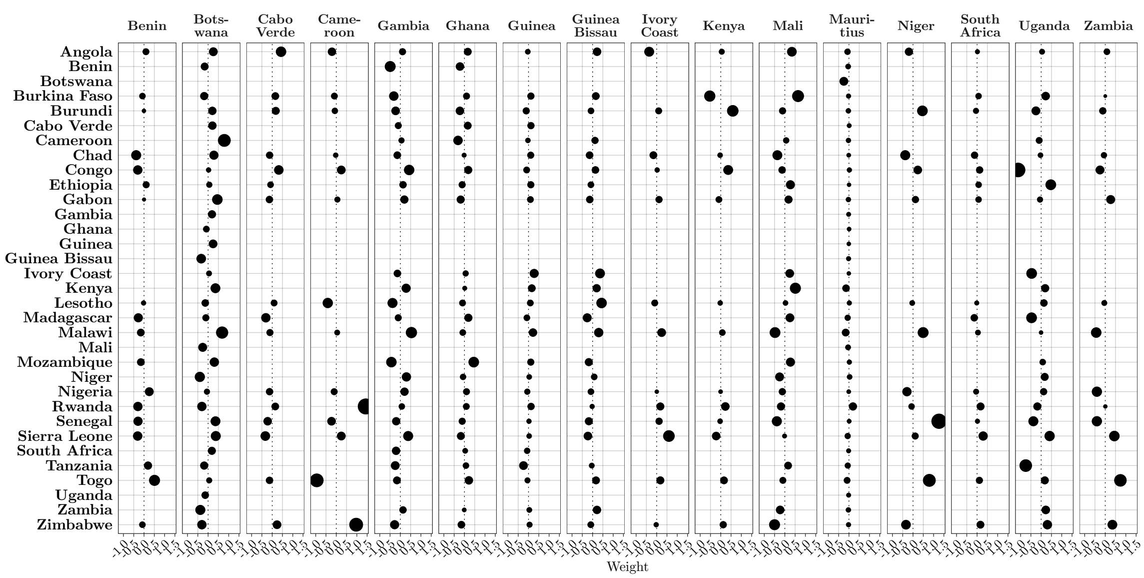
Notes: Panel (a): TSUS prediction for every country in each of five periods after treatment. Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TAUS predicted effects, averaged over five years ($\tau_{i\cdot}$).

Figure S.2: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years, $\widehat{\tau}_{i\cdot}$.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



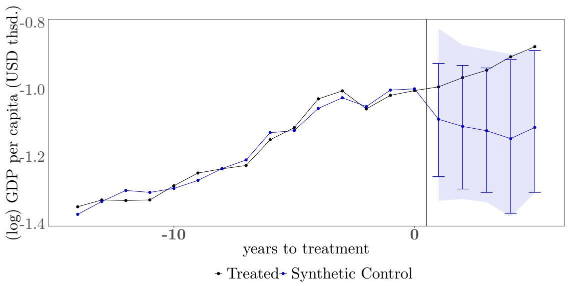
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TAUS prediction for every country averaged over the five periods following treatment (up to the year 2000). Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

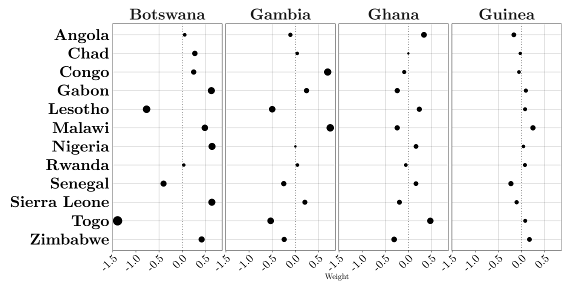
TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 (τ_{Q_1k}, τ_{Q_2k}, τ_{Q_3k}).

Figure S.3: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*

Countries that Liberalized Before 1987, $\widehat{\tau}_{Q_1k}$

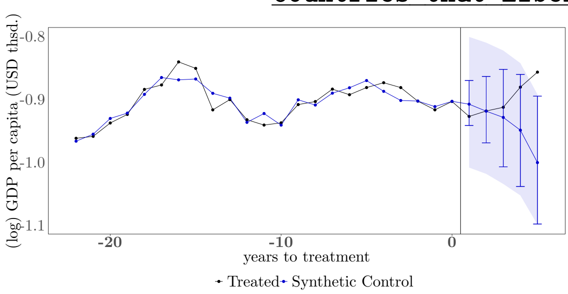


(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

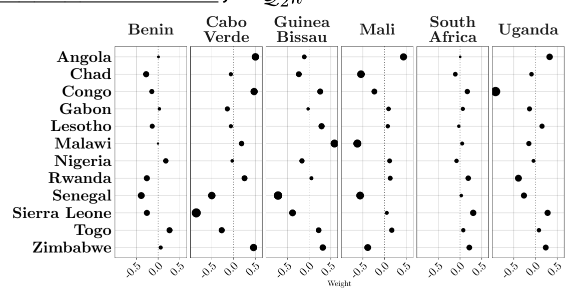


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized in 1987-1991, $\widehat{\tau}_{Q_2k}$

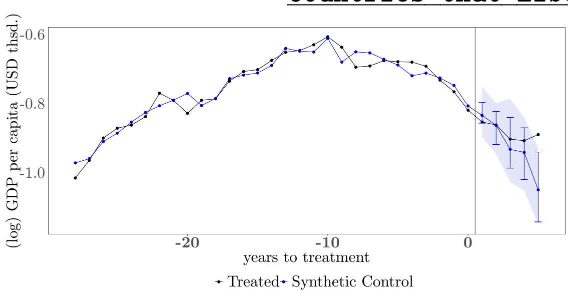


(c) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

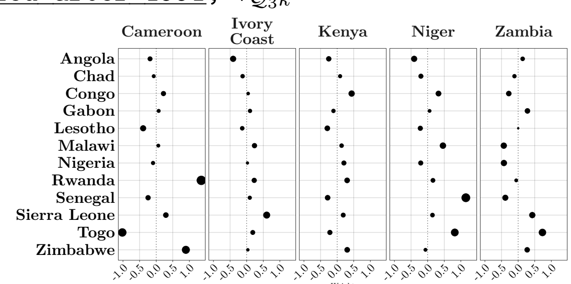


(d) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized after 1991, $\widehat{\tau}_{Q_3k}$



(e) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

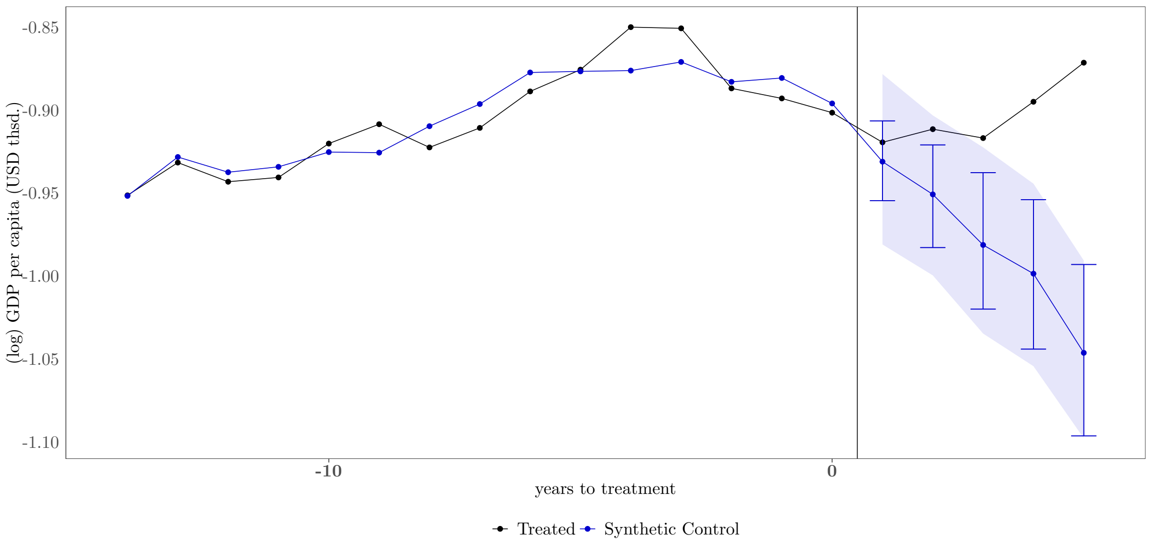


(f) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

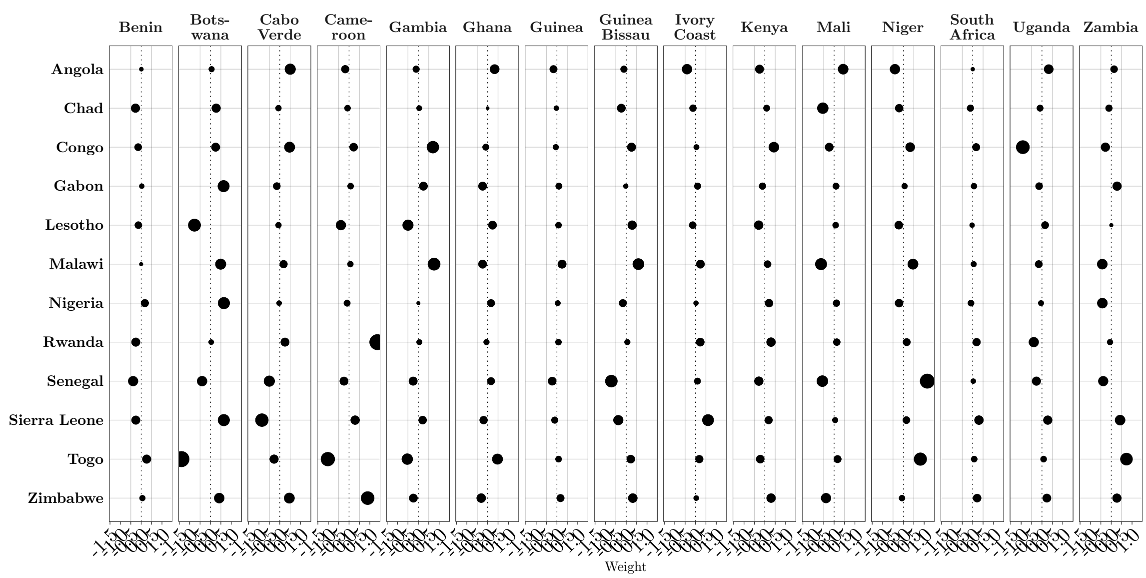
Notes: TSUA prediction in every period after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia). Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E}k}$).

Figure S.4: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units,* $\widehat{\tau}_{\mathcal{E}k}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



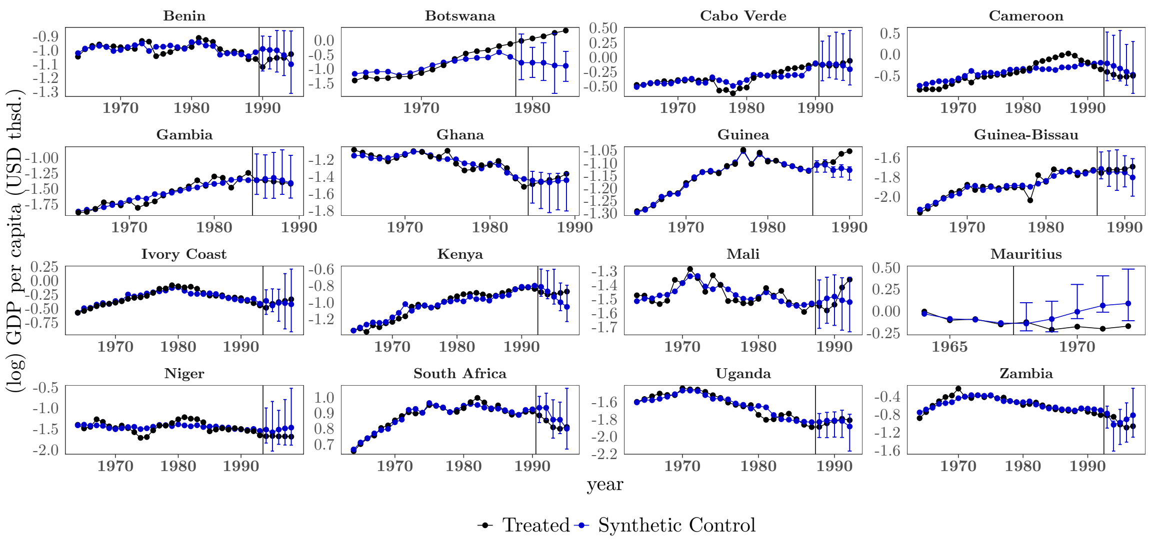
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TSUA prediction in every period after treatment averaged over all the treated countries. Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

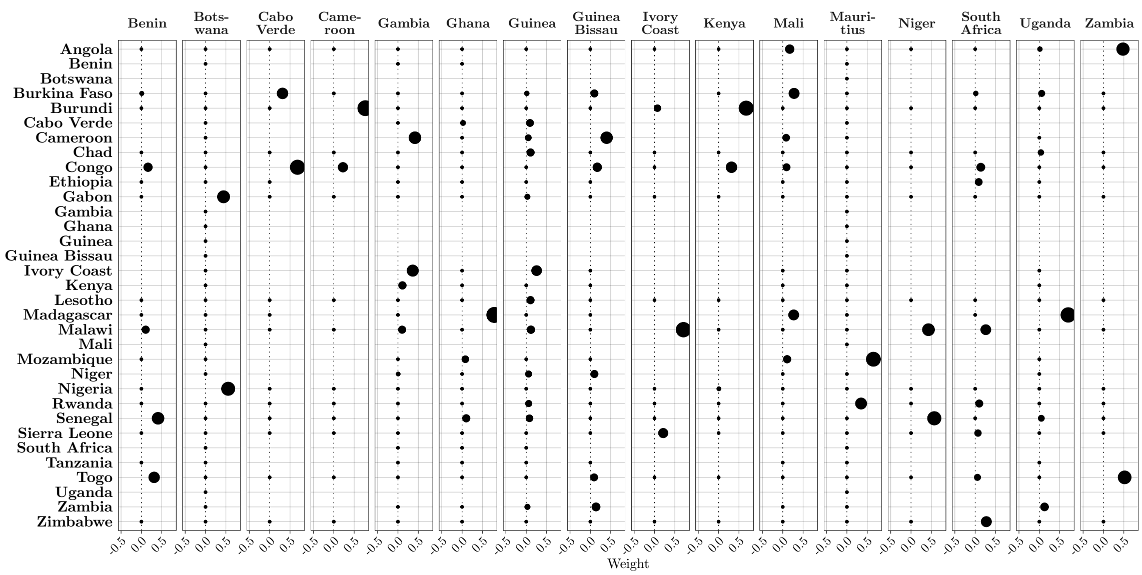
S.9.2 Simplex-Type Constraint

<u>TSUS predicted effects in every period after liberalization (τ_{ik}).</u>

Figure S.5: *Time-specific unit-specific (TSUS) predicted effects in every period,* $\widehat{\tau}_{ik}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

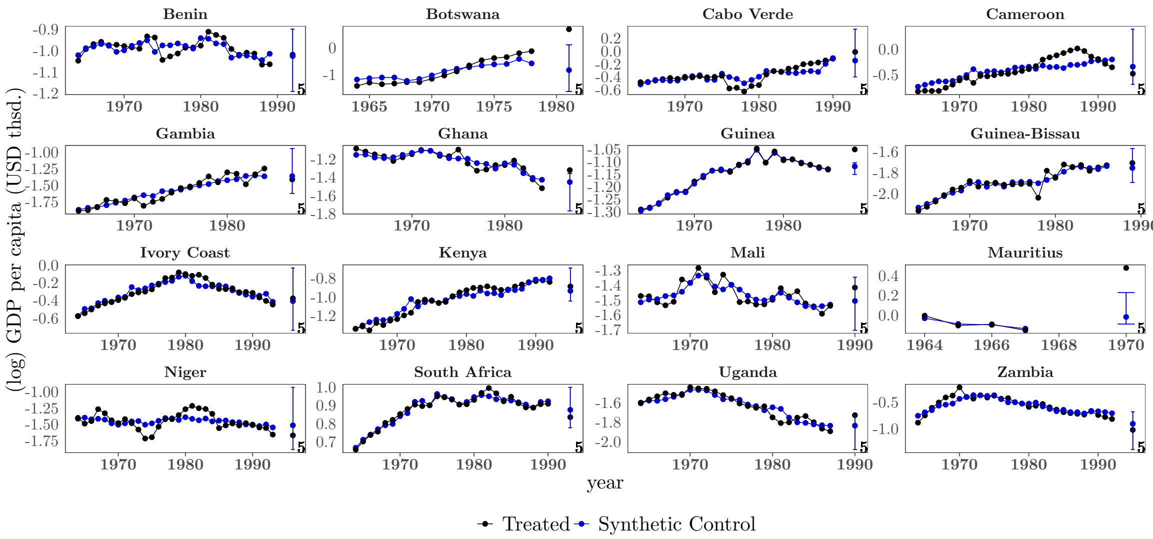


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

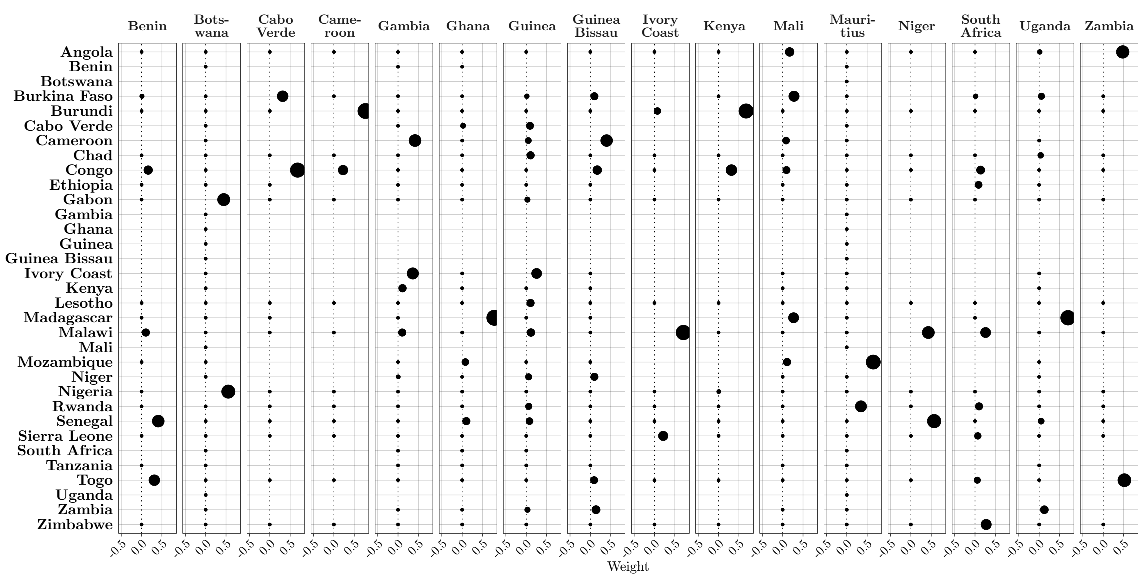
Notes: Panel (a): TSUS prediction for every country in each of five periods after treatment. Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TAUS predicted effects, averaged over five years ($\tau_{i\cdot}$).

Figure S.6: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years, $\widehat{\tau}_{i\cdot}$.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



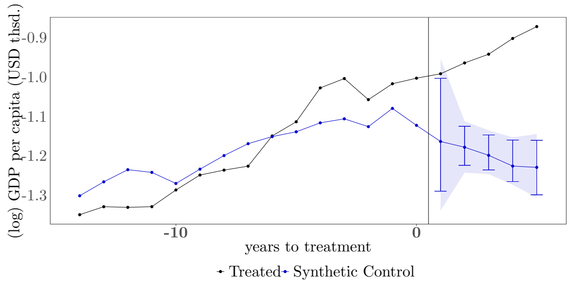
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TAUS prediction for every country averaged over the five periods following treatment (up to the year 2000). Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

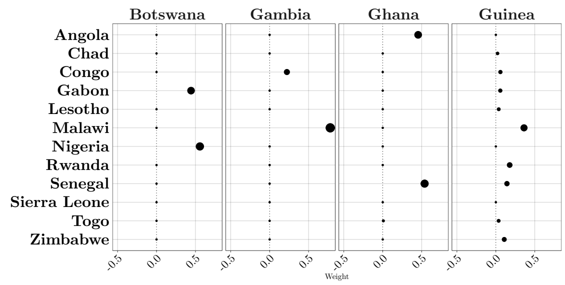
TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 (τ_{Q_1k}, τ_{Q_2k}, τ_{Q_3k}).

Figure S.7: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*

Countries that Liberalized Before 1987, $\widehat{\tau}_{Q_1k}$

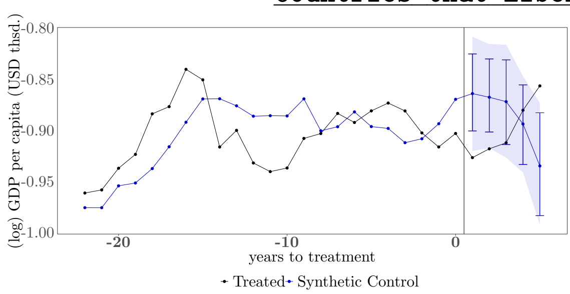


(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

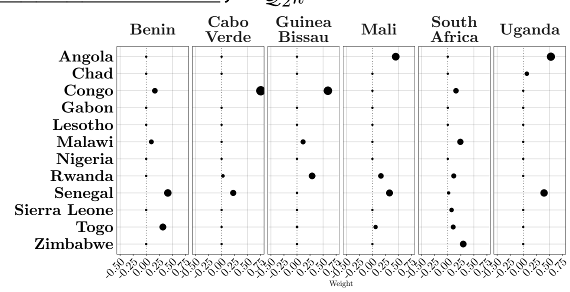


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized in 1987-1991, $\widehat{\tau}_{Q_2k}$

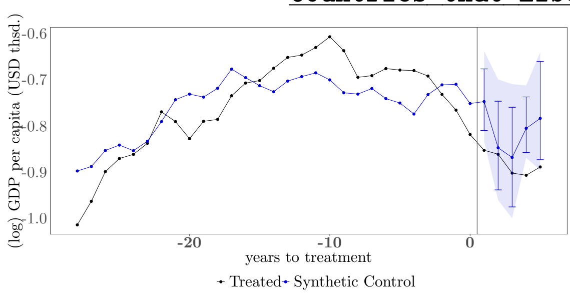


(c) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

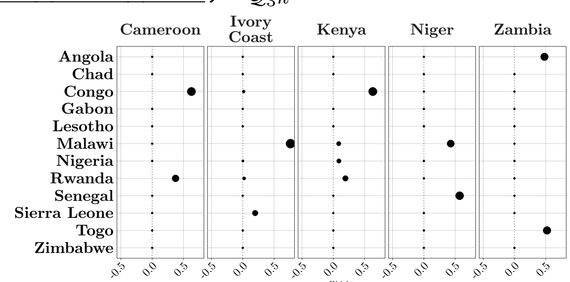


(d) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized after 1991, $\widehat{\tau}_{Q_3k}$



(e) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

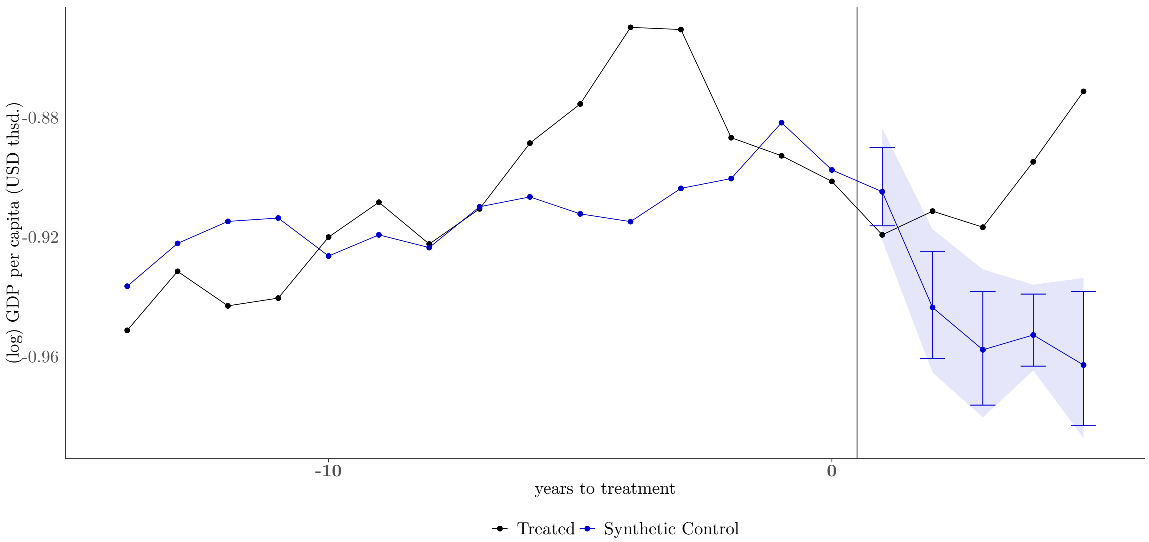


(f) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

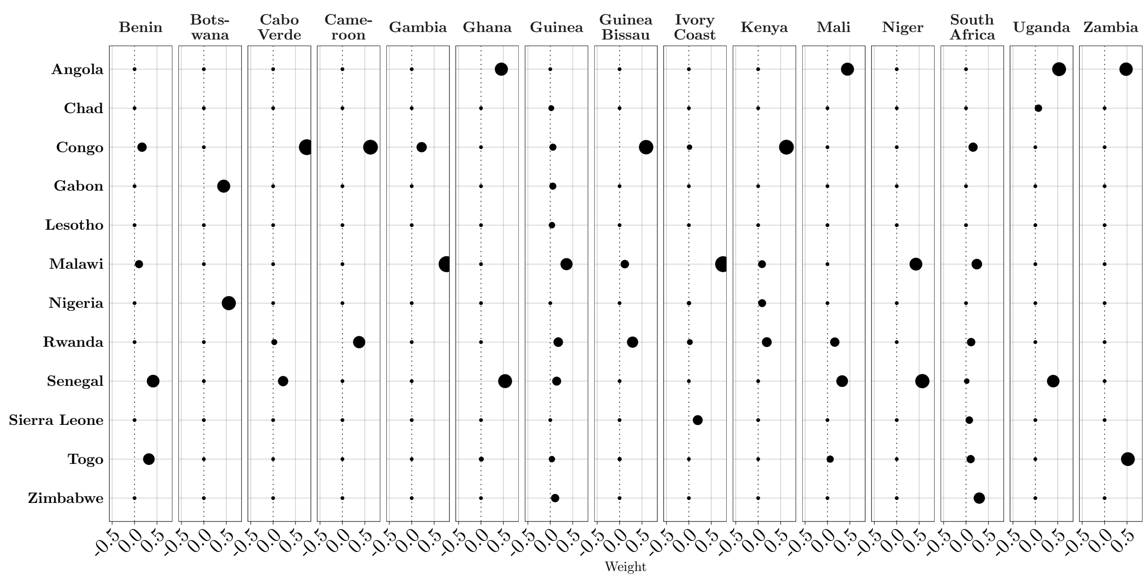
Notes: TSUA prediction in every period after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia). Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E}k}$).

Figure S.8: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units,* $\widehat{\tau}_{\mathcal{E}k}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



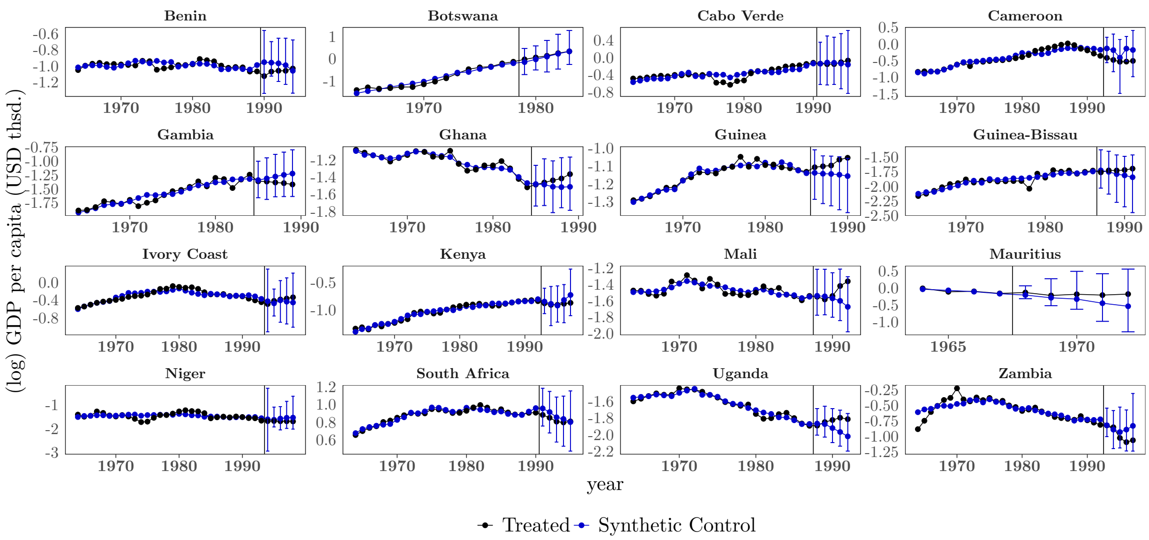
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TSUA prediction in every period after treatment averaged over all the treated countries. Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

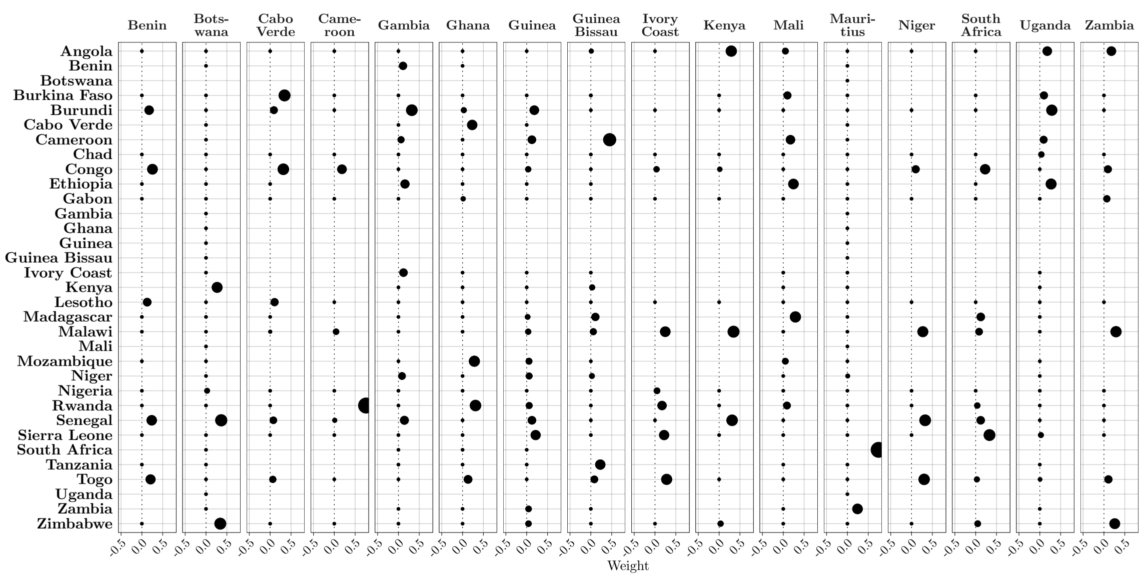
S.9.3 Multiple Features

<u>TSUS predicted effects in every period after liberalization (τ_{ik}).</u>

Figure S.9: *Time-specific unit-specific (TSUS) predicted effects in every period,* $\widehat{\tau}_{ik}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

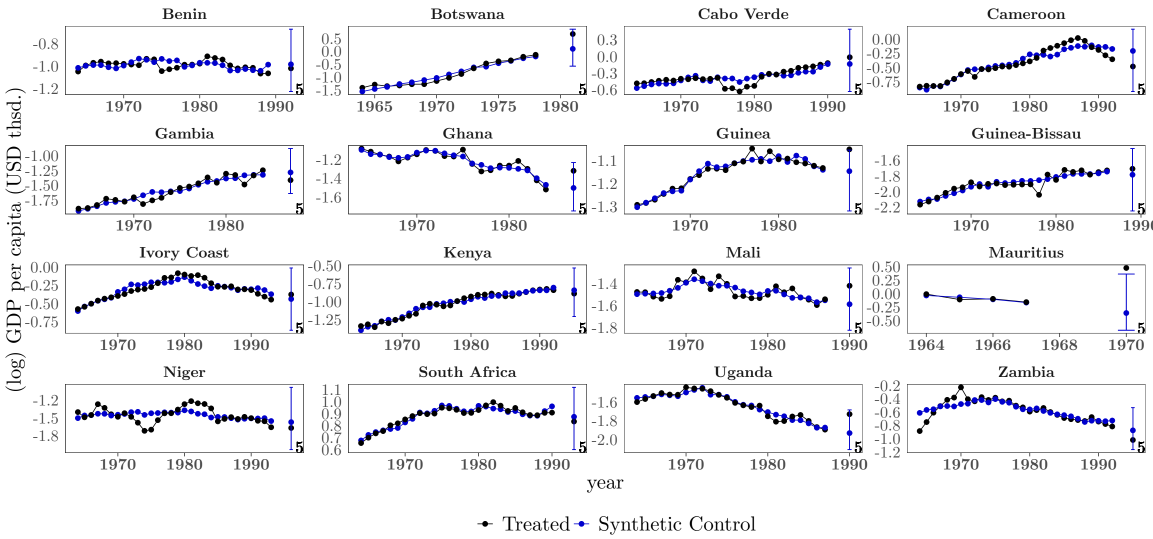


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

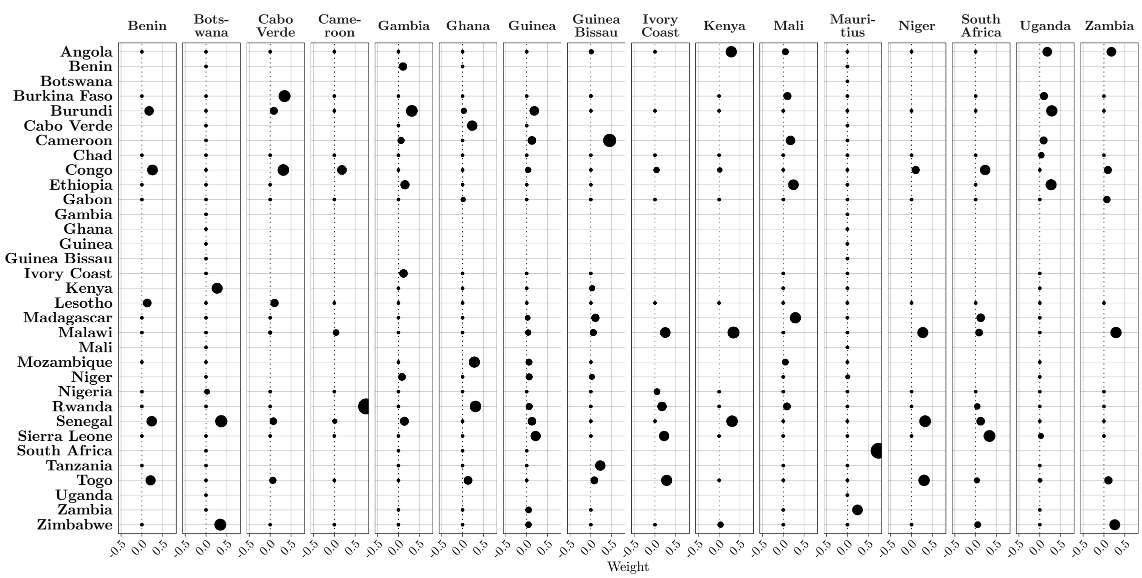
Notes: Panel (a): TSUS prediction for every country in each of five periods after treatment. Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TAUS predicted effects, averaged over five years ($\tau_{i\cdot}$).

Figure S.10: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years,* $\widehat{\tau}_{i\cdot}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



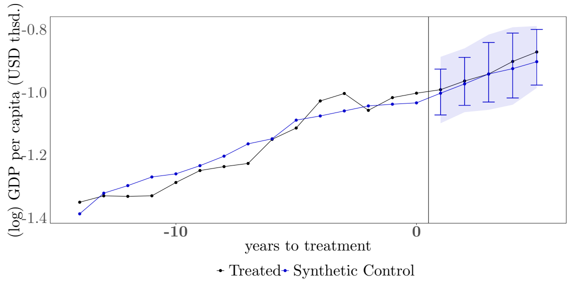
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TAUS prediction for every country averaged over the five periods following treatment (up to the year 2000). Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

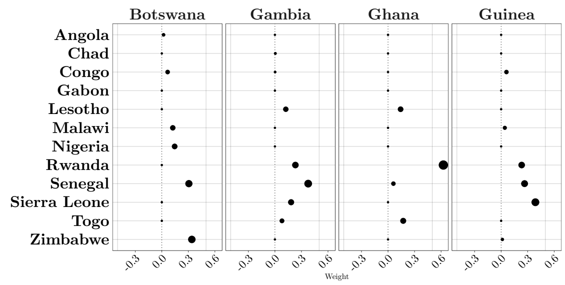
TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 ($\tau_{\mathcal{Q}_1k}$, $\tau_{\mathcal{Q}_2k}$, $\tau_{\mathcal{Q}_3k}$).

Figure S.11: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*

Countries that Liberalized Before 1987, $\widehat{\tau}_{\mathcal{Q}_1k}$

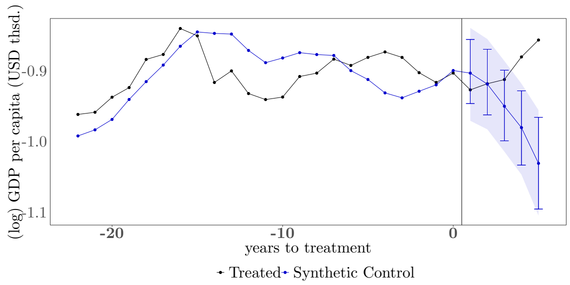


(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

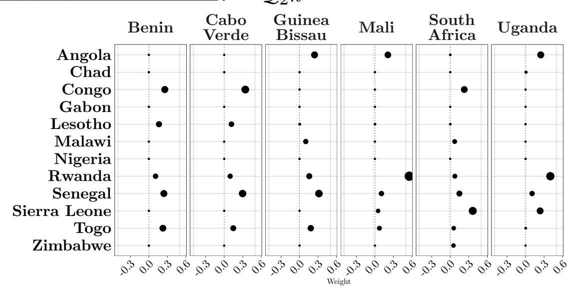


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized in 1987-1991, $\widehat{\tau}_{\mathcal{Q}_2k}$

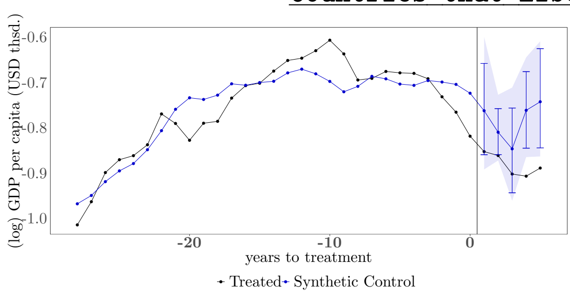


(c) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

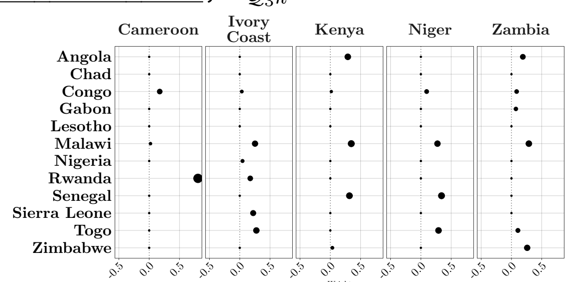


(d) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized after 1991, $\widehat{\tau}_{\mathcal{Q}_3k}$



(e) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

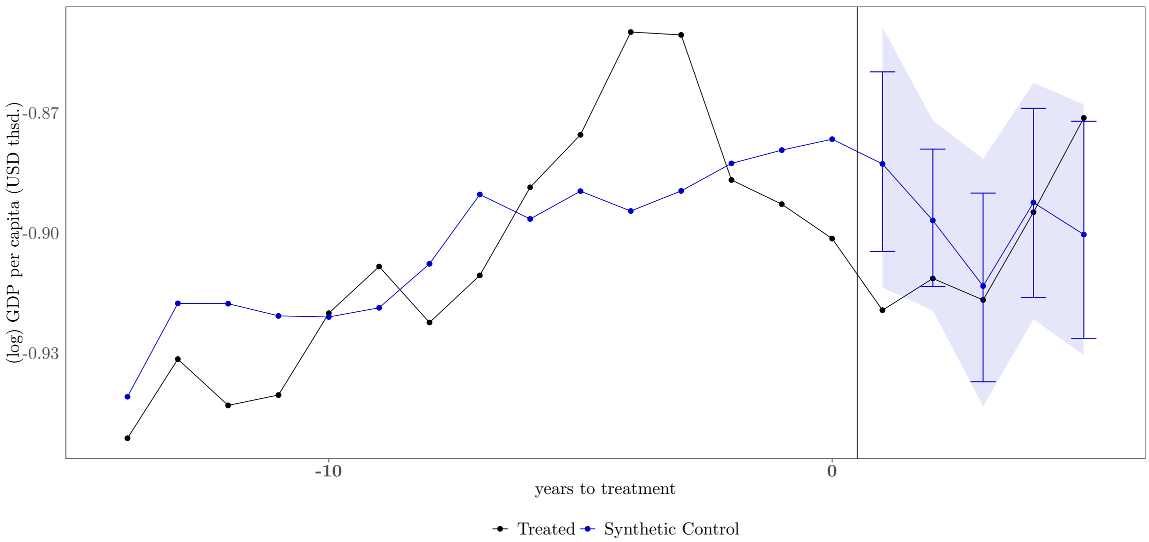


(f) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

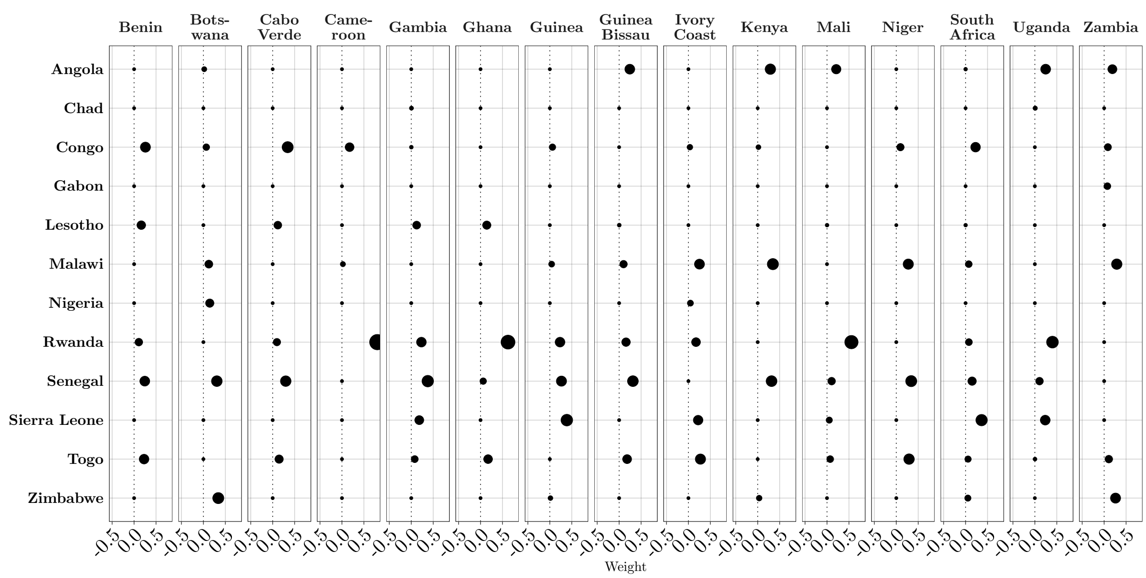
Notes: TSUA prediction in every period after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia). Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E}k}$).

Figure S.12: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units, $\widehat{\tau}_{\mathcal{E}k}$.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



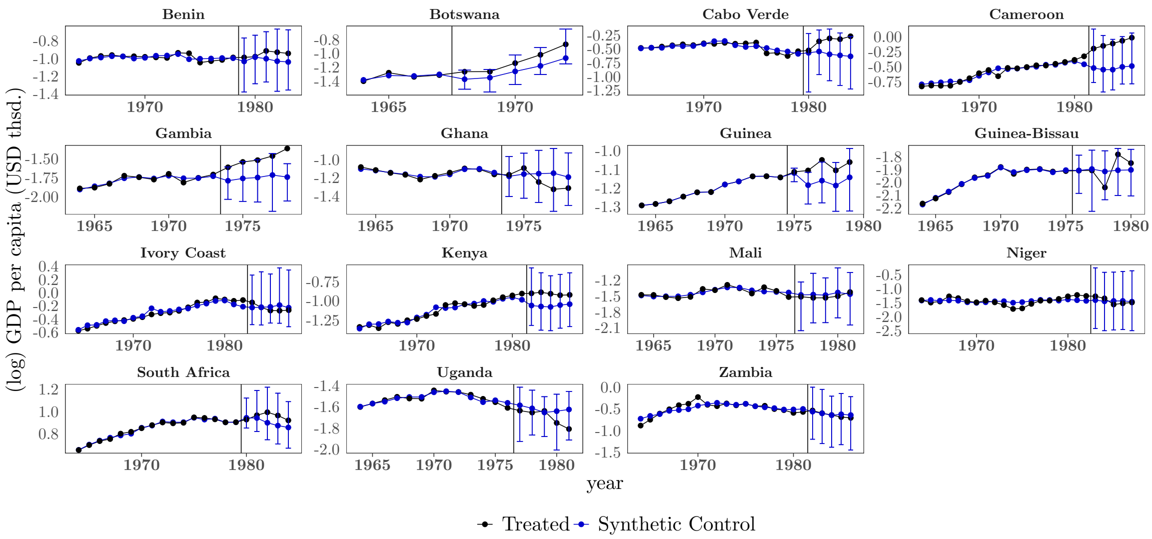
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TSUA prediction in every period after treatment averaged over all the treated countries. Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

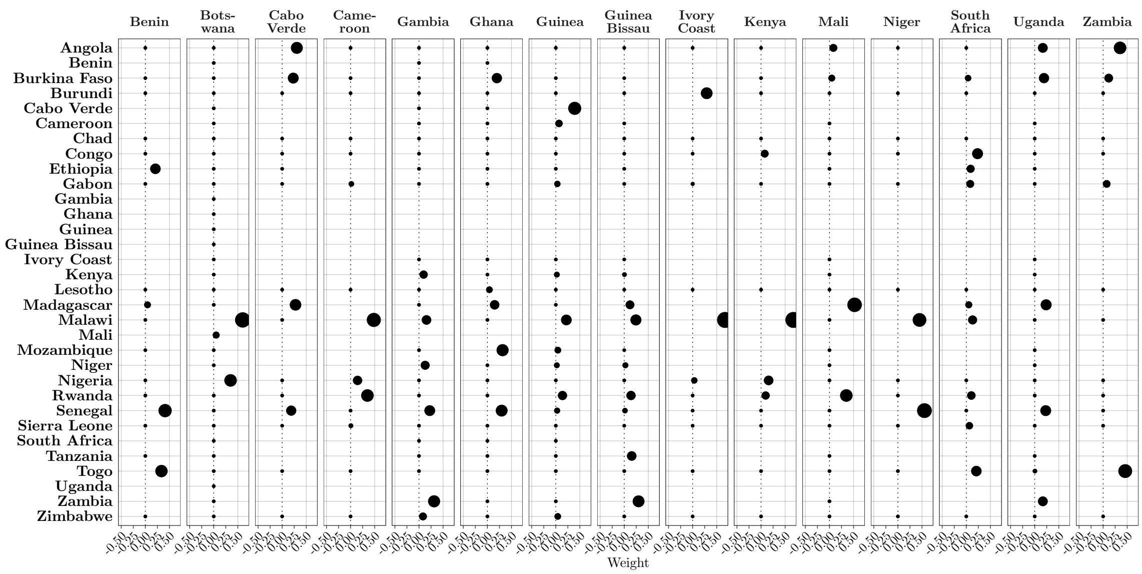
S.9.4 Placebo Treatment Date

In this exercise, we set $\widetilde{T}_i = T_i - 11$ and conduct the same exercise as in the main text.
TSUS predicted effects in every period after liberalization (τ_{ik}).

Figure S.13: *Time-specific unit-specific (TSUS) predicted effects in every period, $\widehat{\tau}_{ik}$.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

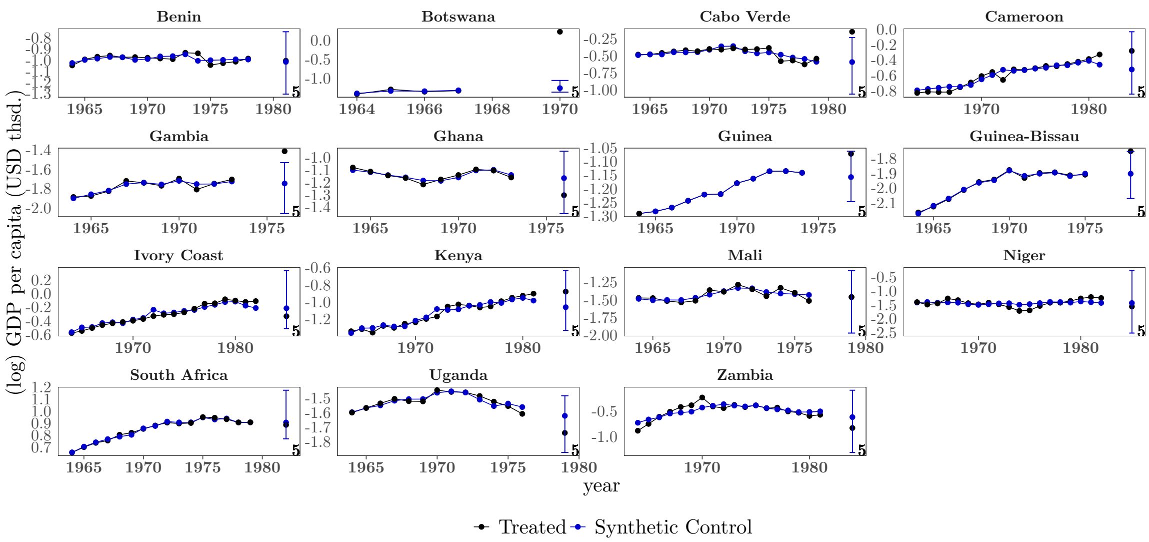


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

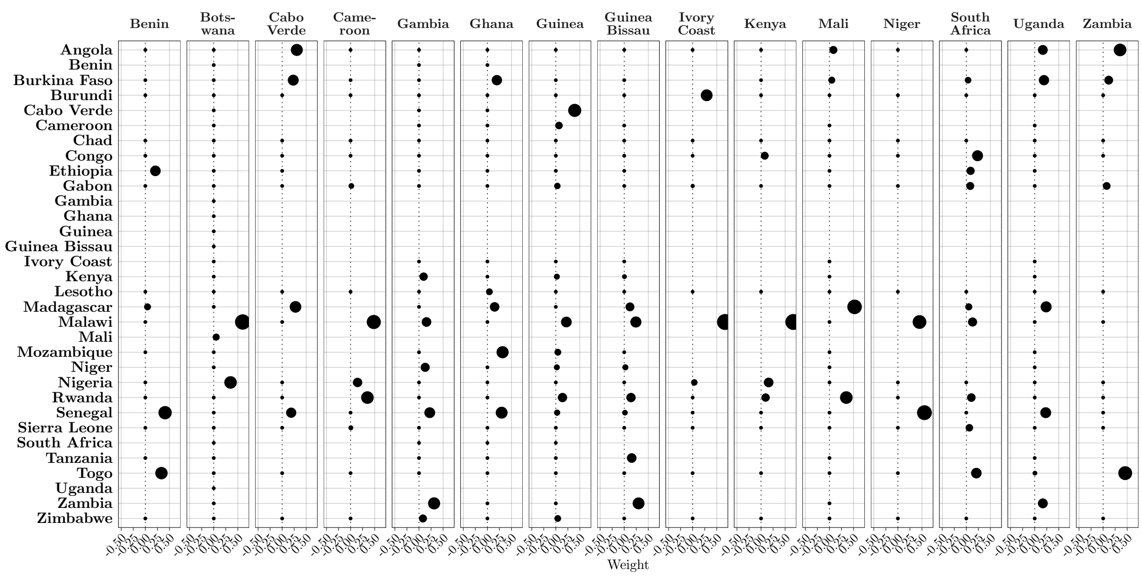
Notes: Panel (a): TSUS prediction for every country in each of five periods after treatment. Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TAUS predicted effects, averaged over five years ($\tau_{i\cdot}$).

Figure S.14: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years,* $\widehat{\tau}_{i\cdot}$*.*



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



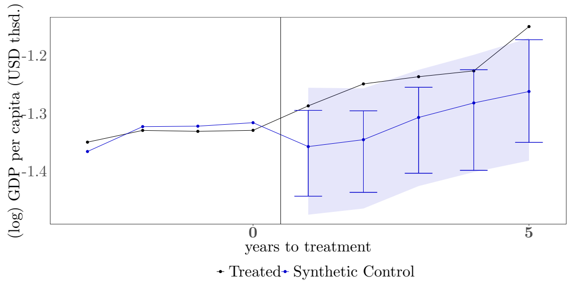
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Notes: Panel (a): TAUS prediction for every country averaged over the five periods following treatment (up to the year 2000). Blue bars report 90% prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

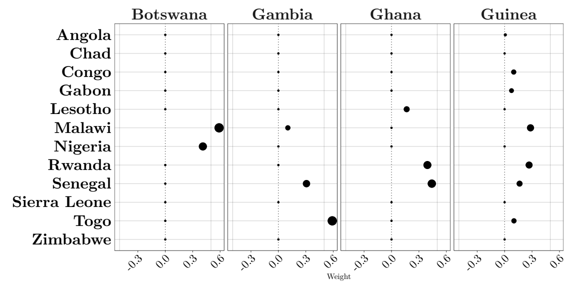
TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 ($\tau_{Q_1 k}$, $\tau_{Q_2 k}$, $\tau_{Q_3 k}$).

Figure S.15: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*

Countries that Liberalized Before 1987, $\widehat{\tau}_{Q_1 k}$

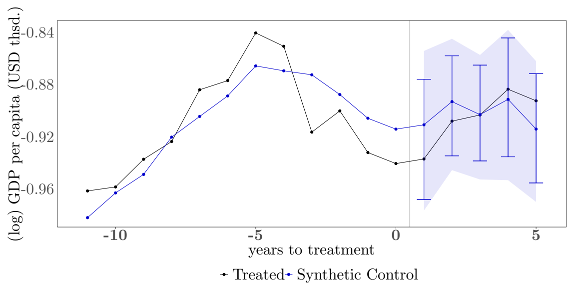


(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

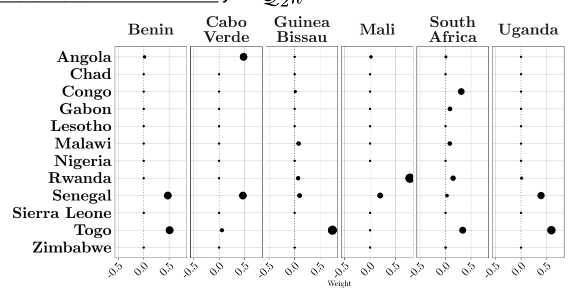


(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized in 1987-1991, $\widehat{\tau}_{Q_2 k}$

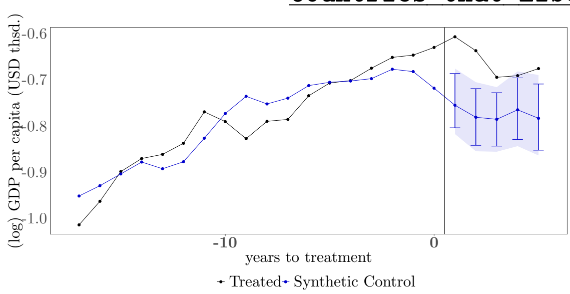


(c) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

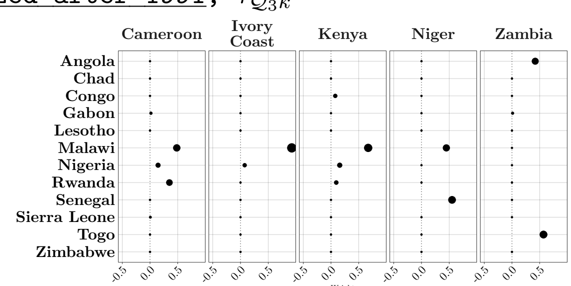


(d) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

Countries that Liberalized after 1991, $\widehat{\tau}_{Q_3 k}$



(e) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

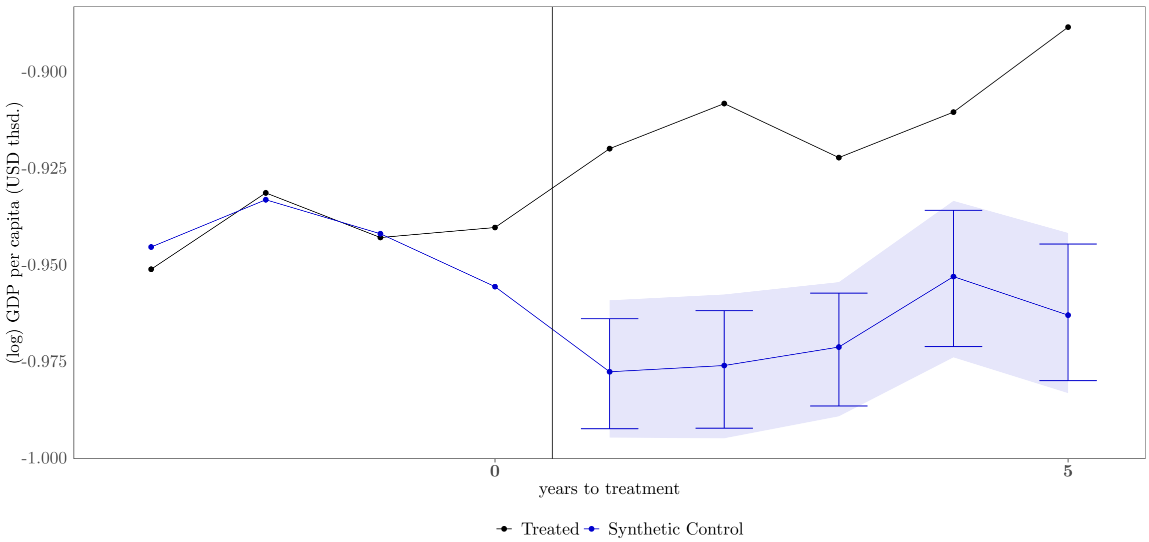


(f) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

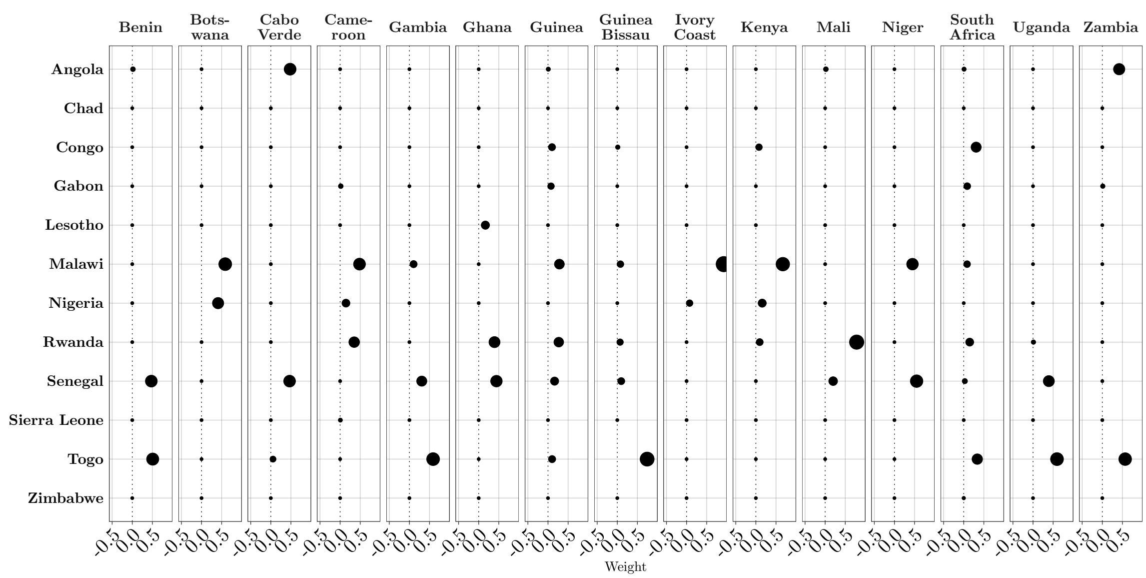
Notes: TSUA prediction in every period after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia). Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E}k}$).

Figure S.16: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units,* $\widehat{\tau}_{\mathcal{E}k}$.



(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$



(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

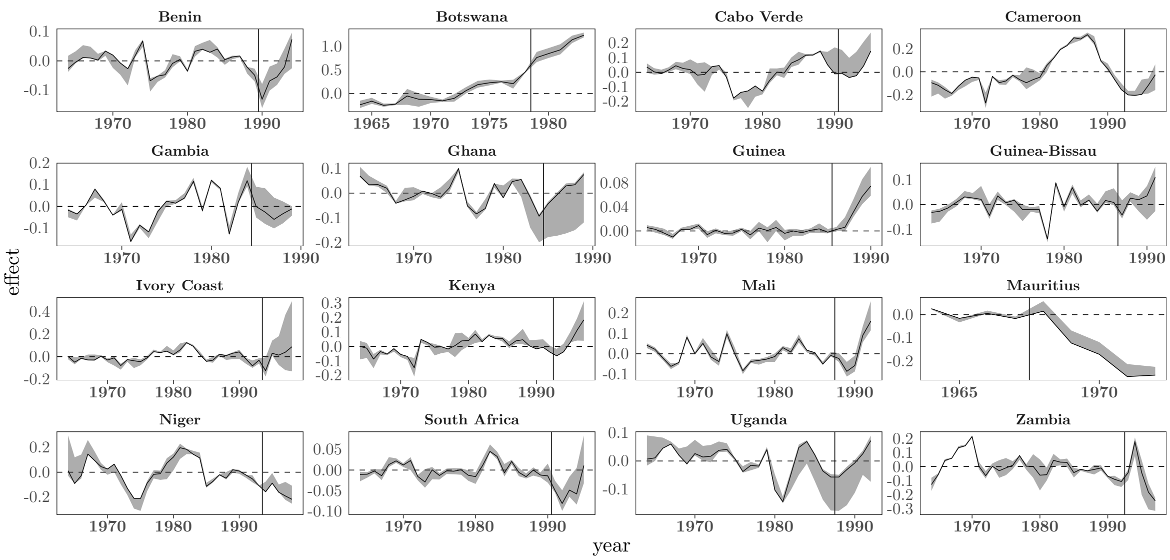
Notes: Panel (a): TSUA prediction in every period after treatment averaged over all the treated countries. Blue bars report 90% prediction intervals, whereas blue-shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds. Panel (b): each dot represents the weight that the donor (row) gets in forming the synthetic control for the treated unit (column). When there is no dot, it means that the unit was not part of the donor pool for the treated unit in question.

S.9.5 Leave-One-Donor-Out

In this exercise, we remove with replacement one donor at a time from the donor pool and recompute our six predictands of interest.

TSUS predicted effects in every period after liberalization (τ_{ik}).

Figure S.17: *Time-specific unit-specific (TSUS) predicted effects in every period,* $\widehat{\tau}_{ik}$.



Notes: The black solid line depicts the TSUS prediction for every country in the pre-treatment period and each of the five periods after treatment. Gray-shaded areas highlight the area between the lower and upper value of $\widehat{\tau}_{it}$ when leaving one of the donors out at a time.

TAUS predicted effects, averaged over five years ($\tau_{i\cdot}$).

Figure S.18: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years,* $\widehat{\tau}_{i\cdot}$.

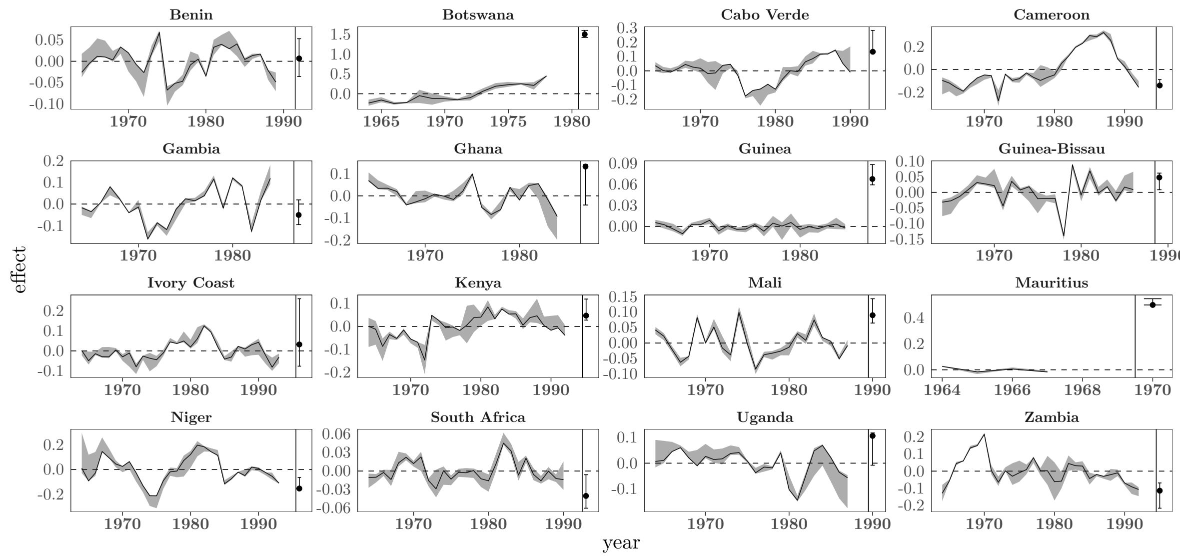
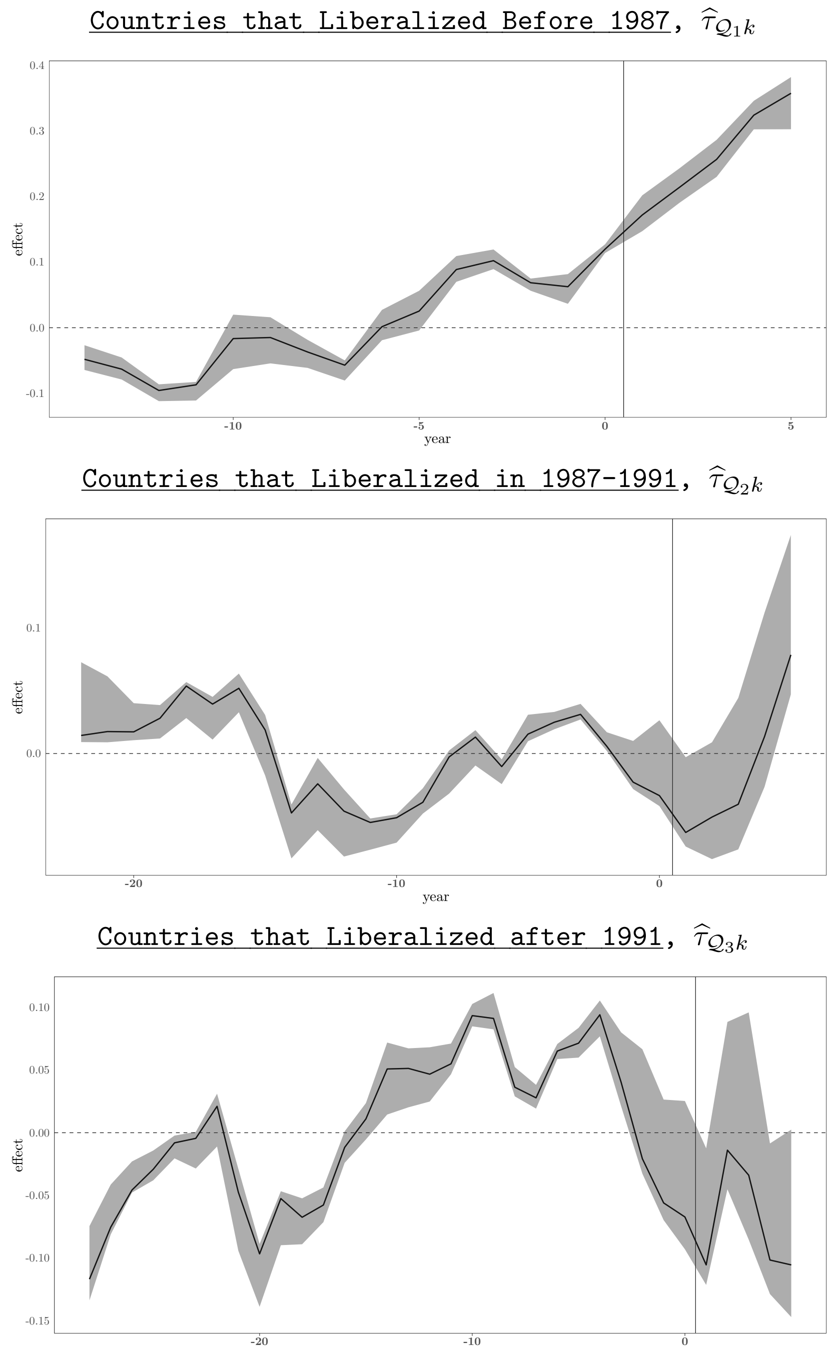


Figure S.19: $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$

Notes: The black solid line depicts the TSUS prediction for every country in the pre-treatment period and each of the five periods after treatment. Black vertical bars highlight the area between the lower and upper value of $\widehat{\tau}_{\cdot t}$ when leaving one of the donors out at a time.

TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 ($\tau_{Q_1 k}$, $\tau_{Q_2 k}$, $\tau_{Q_3 k}$).

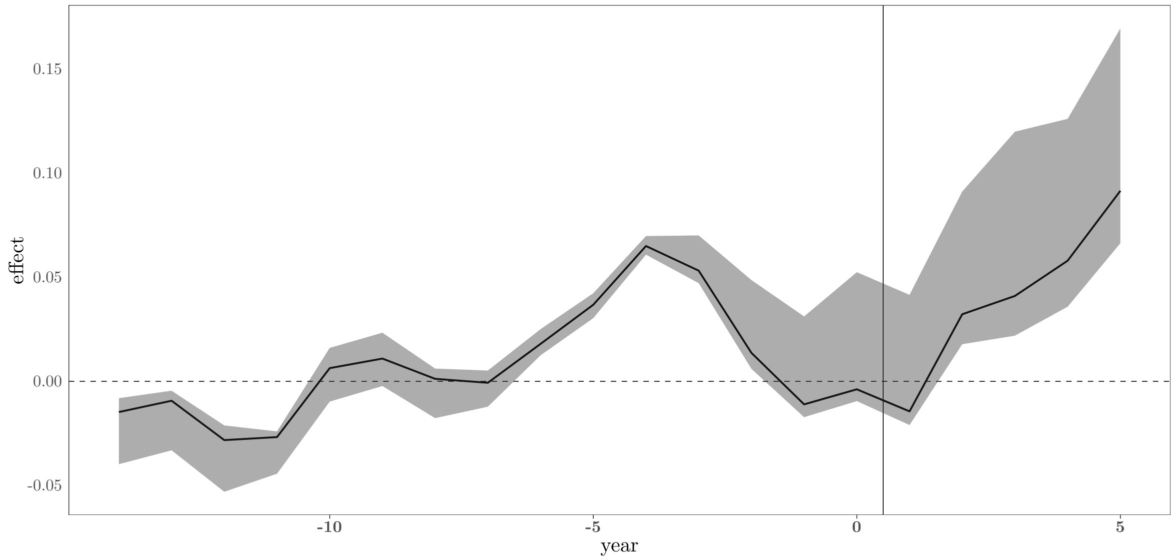
Figure S.20: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*



Notes: The black solid line depicts the TSUA prediction in every period before and after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia). Gray-shaded areas highlight the area between the lower and upper value of $\widehat{\tau}_{\mathcal{Q}_j t}, j = 1, 2, 3$, when leaving one of the donors out at a time.

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E}k}$).

Figure S.21: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units,* $\widehat{\tau}_{\mathcal{E}k}$.



Notes: The black solid line depicts the TSUA prediction in every period before and after treatment averaged over all the treated countries. Gray-shaded areas highlight the area between the lower and upper value of $\widehat{\tau}_{\mathcal{E}t}$ when leaving one of the donors out at a time.

References

Ben-Michael, E., Feller, A., and Rothstein, J. (2022), "Synthetic Controls with Staggered Adoption," *Journal of the Royal Statistical Society, Series B*, 84, 351–381.

Billmeier, A., and Nannicini, T. (2013), "Assessing Economic Liberalization Episodes: A Synthetic Control Approach," *Review of Economics and Statistics*, 95, 983–1001.

Boyd, S., and Vandenberghe, L. (2004), *Convex Optimization*, Cambridge University Press.

Bratton, M., and Van de Walle, N. (1997), *Democratic Experiments in Africa: Regime Transitions in Comparative Perspective*, Cambridge university press.

Cattaneo, M. D., Feng, Y., Palomba, F., and Titiunik, R. (2025), "`scpi`: Uncertainty Quantification for Synthetic Control Methods," *Journal of Statitistical Software*.

Cattaneo, M. D., Feng, Y., and Titiunik, R. (2021), "Prediction Intervals for Synthetic Control Methods," *Journal of the American Statistical Association*, 116, 1865–1880.

Chang, Y., Park, J. Y., and Song, K. (2006), "Bootstrapping Cointegrating Regressions," *Journal of Econometrics*, 133, 703–739.

Ferman, B., and Pinto, C. (2021), "Synthetic Controls with Imperfect Pre-Treatment Fit," *Quantitative Economics*.

Hoerl, A. E., Kannard, R. W., and Baldwin, K. F. (1975), "Ridge Regression: Some Simulations," *Communications in Statistics-Theory and Methods*, 4, 105–123.

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

Ravishanker, N., Hochberg, Y., and Melnick, E. L. (1987), "Approximate Simultaneous Prediction Intervals for Multiple Forecasts," *Technometrics*, 29, 371–376.

Sachs, J. D., Warner, A., Åslund, A., and Fischer, S. (1995), "Economic Reform and the Process of Global Integration," *Brookings Papers on Economic Activity*, 1995, 1–118.

Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press.