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

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

May 4, 2023

This supplement contains all proofs, additional results, other technical details about data preparation and estimation, and the code used to reproduce the results.

[*]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.

Complete Contents

S.1 Other Strategies for Uncertainty Quantification

S.1.1 Out-of-Sample Error

In Section 4 we discuss the approach for quantifying the uncertainty of $e_t^{[i]}$ based on the non-asymptotic bounds. We briefly describe another two strategies below.

- *Location-scale model.* Suppose that $e_t^{[i]} = \mathbb{E}[e_t^{[i]}|\mathscr{H}] + (\mathbb{V}[e_t^{[i]}|\mathscr{H}])^{1/2}\varepsilon_t^{[i]}$ with $\varepsilon_t^{[i]}$ statistically independent of $\mathscr{H}$. The bounds on $e_t^{[i]}$ can now be set as $M_{2,\mathtt{L}} = \mathbb{E}[e_t^{[i]}|\mathscr{H}]+(\mathbb{V}[e_t^{[i]}|\mathscr{H}])^{1/2}\mathfrak{c}_\varepsilon(\alpha_2/2)$ and $M_{2,\mathtt{U}} = \mathbb{E}[e_t^{[i]}|\mathscr{H}] + (\mathbb{V}[e_t^{[i]}|\mathscr{H}])^{1/2}\mathfrak{c}_\varepsilon(1-\alpha_2/2)$ where $\mathfrak{c}_\varepsilon(\alpha_2/2)$ and $\mathfrak{c}_\varepsilon(1-\alpha_2/2)$ are $\alpha_2/2$ and $(1-\alpha_2/2)$ quantiles of $\varepsilon_t^{[i]}$, respectively, and α_2 is the desired pre-specified level.

- *Quantile regression.* We can determine the $\alpha_2/2$ and $(1-\alpha_2/2)$ conditional quantiles of $e_t^{[i]}|\mathscr{H}$. Consequently, another possibility is to employ quantile regression methods to estimate those quantities using pre-treament data.

S.1.2 Simultaneous Prediction Intervals

Section 4.2 constructs prediction intervals with simultaneous coverage. We briefly describe two other common approaches below.

- *Bonferroni-type correction.* There has been 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 $k = 0, \cdots, L$, use any strategy described in Section 4 to construct a prediction interval for $e_{T_i+k}^{[i]}$ that has coverage probability at least $1-(\alpha_2/(L+1))$. Then, the simultaneous coverage probability of the $L+1$ prediction intervals $\{\tilde{\mathcal{I}}_k : 0 \le k \le L\}$ is at least $1-\alpha_2$. 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

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

  Then, one can construct the prediction intervals $\tilde{\mathcal{I}}_k$ with corresponding coverage probabilities p_k and $p_{k,k+1}$ such that $1-\sum_{k=0}^{L} p_k + \sum_{k=0}^{L-1} p_{k,k+1} \ge 1-\alpha_2$. Such bounds are usually sharper, but their implementation requires the modelling of the dependence of $(e_t^{[i]}, e_{t+1}^{[i]})$ 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 $(e_{T_i}^{[i]}, \cdots, e_{T_i+L}^{[i]})'$ jointly follows a conditional Gaussian distribution with mean zero and variance $\Sigma_{\mathscr{H}}$. Then,

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

  where χ^2_{L+1} is χ^2 distribution with $L+1$ degrees of freedom. The sequence of prediction intervals $\tilde{\mathcal{I}}_k = \Big[-\sigma_{\mathscr{H},kk}\sqrt{\chi^2_{L+1}(1-\alpha_2)},\ \sigma_{\mathscr{H},kk}\sqrt{\chi^2_{L+1}(1-\alpha_2)}\Big]$ have the simultaneous coverage probability at least $1-\alpha_2$, where $\sigma^2_{\mathscr{H},kk}$ is the kth diagonal element of $\Sigma_{\mathscr{H}}$ and $\chi^2_{L+1}(1-\alpha_2)$ is the $(1-\alpha_2)$-quantile of χ^2 distribution with $L+1$ degrees of freedom.

S.2 Proofs

S.2.1 Proof of Theorem 1

Proof. Let

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

Accordingly, define

$$\tilde{\varsigma}^\star_{\mathtt{U}} = \sup\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\},$$
$$\tilde{\varsigma}^{\mathsf{int}}_{\mathtt{U}} = \sup\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\},$$
$$\tilde{\varsigma}_{\mathtt{U}} = \sup\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell(\boldsymbol{\delta}) \le 0\right\}.$$

For any $\alpha_0 \in [0,1]$, let $\tilde{\mathfrak{c}}^\star_{\mathtt{U}}(\alpha_0)$ be the α_0-quantile of $\tilde{\varsigma}^\star_{\mathtt{U}}$ conditional on the data and $\tilde{\mathfrak{c}}_{\mathtt{U}}(\alpha_0)$ be the α_0-quantile of $\tilde{\varsigma}_{\mathtt{U}}$ conditional on $\mathscr{H}$. Similarly, define

$$\tilde{\varsigma}^\star_{\mathtt{L}} = \inf\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta^\star,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\},$$
$$\tilde{\varsigma}^{\mathsf{int}}_{\mathtt{L}} = \inf\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\},$$
$$\tilde{\varsigma}_{\mathtt{L}} = \inf\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell(\boldsymbol{\delta}) \le 0\right\}.$$

Let $\tilde{\mathfrak{c}}^\star_{\mathtt{L}}(\alpha_0)$ be the α_0-quantile of $\tilde{\varsigma}^\star_{\mathtt{L}}$ conditional on the data and $\tilde{\mathfrak{c}}_{\mathtt{L}}(\alpha_0)$ be the α_0-quantile of $\tilde{\varsigma}_{\mathtt{L}}$ 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 $1-\pi^\star_\gamma$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_{\gamma,2}$, the Kullback-Leibler divergence $\mathbb{KL}(\mathbb{P}_1, \mathbb{P}_2) \le 2(\epsilon^\star_{\gamma,1})^2$, and by Pinsker's inequality, this implies that for any $\kappa' \le \kappa$,

$$|\mathbb{P}^\star(\tilde{\varsigma}^{\mathsf{int}}_{\mathtt{U}} \le \kappa) - \mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{U}} \le \kappa)| \le \epsilon^\star_{\gamma,1} \quad \text{and} \quad |\mathbb{P}^\star(\tilde{\varsigma}^{\mathsf{int}}_{\mathtt{L}} \ge \kappa') - \mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{L}} \ge \kappa')| \le \epsilon^\star_{\gamma,1}.$$

On the other hand, note that by condition (iii), on an event with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi^\star_\Delta$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1-\epsilon^\star_\Delta$, the event $\{\tilde{\varsigma}^\star_{\mathtt{U}} \le \kappa, \tilde{\varsigma}^\star_{\mathtt{L}} \ge \kappa'\}$ implies that

$$\begin{aligned}\sup\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\} &\le \kappa + \|\mathbf{p}_\tau'\|_* \sup_{\mathbf{a}\in\Delta\cap\mathcal{B}(0,\varpi^\star_\delta)} \mathrm{dist}(\mathbf{a}, \Delta^\star)\\ &\le \kappa + \|\mathbf{p}_\tau'\|_*\varpi^\star_\Delta =: \kappa + \varepsilon_\Delta\end{aligned}$$

and

$$\begin{aligned}\inf\left\{\mathbf{p}_\tau'\boldsymbol{\delta} : \boldsymbol{\delta} \in \Delta,\, \|\boldsymbol{\delta}\| \le \varpi^\star_\delta,\, \ell^\star(\boldsymbol{\delta}) \le 0\right\} &\ge \kappa' - \|\mathbf{p}_\tau'\|_* \sup_{\mathbf{a}\in\Delta\cap\mathcal{B}(0,\varpi^\star_\delta)} \mathrm{dist}(\mathbf{a}, \Delta^\star)\\ &\ge \kappa - \varepsilon_\Delta.\end{aligned}$$

Therefore,

$$\mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{U}}^\star \le \kappa) \le \mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{U}}^{\mathsf{int}} \le \kappa + \varepsilon_\Delta) \quad \text{and} \quad \mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{L}}^\star \ge \kappa') \ge \mathbb{P}^\star(\tilde{\varsigma}_{\mathtt{L}}^{\mathsf{int}} \ge \kappa' - \varepsilon_\Delta).$$

Then, 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

$$\begin{aligned}
&1-\alpha_1/2 \le \mathbb{P}^\star\Big(\tilde{\varsigma}_{\mathtt{U}}^\star \le \mathfrak{c}^\star(1-\alpha_1/2)\Big) \le \mathbb{P}^\star\Big(\tilde{\varsigma}_{\mathtt{U}} \le \mathfrak{c}^\star(1-\alpha_1/2)+\varepsilon_\Delta\Big) + \epsilon_{\gamma,1}^\star \quad \text{and}\\
&1-\alpha_1/2 \le \mathbb{P}^\star\Big(\tilde{\varsigma}_{\mathtt{L}}^\star \ge \mathfrak{c}^\star(\alpha_1/2)\Big) \le \mathbb{P}^\star\Big(\tilde{\varsigma}_{\mathtt{L}} \ge \mathfrak{c}^\star(\alpha_1/2)-\varepsilon_\Delta\Big) + \epsilon_{\gamma,1}^\star.
\end{aligned}$$

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

$$\tilde{\mathfrak{c}}_{\mathtt{U}}(1-\alpha_1/2-\epsilon_{\gamma,1}^\star) \ge \mathfrak{c}_{\mathtt{U}}(1-\alpha_1/2-\epsilon_{\gamma,1}^\star-\epsilon_\delta^\star) \quad \text{and} \quad \tilde{\mathfrak{c}}_{\mathtt{L}}(\alpha_1/2+\epsilon_{\gamma,1}^\star) \le \mathfrak{c}_{\mathtt{L}}(\alpha_1/2+\epsilon_{\gamma,1}^\star+\epsilon_\delta^\star).$$

Using conditions (i) and (ii), we have with $\mathbb{P}$-probability over $\mathscr{H}$ at least $1-\pi_\gamma-\pi_\gamma^\star-\pi_\Delta^\star-\pi_\delta^\star$,

$$\begin{aligned}
&\mathbb{P}\Big(\mathfrak{c}_{\mathtt{L}}^\star(\alpha_1/2)-\varepsilon_\Delta \le \mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \le \mathfrak{c}_{\mathtt{U}}^\star(1-\alpha_1/2)+\varepsilon_\Delta\Big|\mathscr{H}\Big)\\
\ge&\,\mathbb{P}\Big(\tilde{\mathfrak{c}}_{\mathtt{L}}(\alpha_1/2+\epsilon_{\gamma,1}^\star) \le \mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \le \tilde{\mathfrak{c}}_{\mathtt{U}}(1-\alpha_1/2-\epsilon_{\gamma,1}^\star)\Big|\mathscr{H}\Big) - \epsilon_{\gamma,2}^\star - \epsilon_\Delta^\star\\
\ge&\,\mathbb{P}\Big(\mathfrak{c}_{\mathtt{L}}(\alpha_1/2+\epsilon_{\gamma,1}^\star+\epsilon_\delta^\star) \le \mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \le \mathfrak{c}_{\mathtt{U}}(1-\alpha_1/2-\epsilon_{\gamma,1}^\star-\epsilon_\delta^\star)\Big|\mathscr{H}\Big) - \epsilon_{\gamma,2}^\star - \epsilon_\Delta^\star\\
\ge&\,1-\alpha_1-2\epsilon_{\gamma,1}^\star-2\epsilon_\delta^\star-\epsilon_\gamma-\epsilon_{\gamma,2}^\star-\epsilon_\Delta^\star.
\end{aligned}$$

Finally, by condition (v), we immediately have $\mathbb{P}(M_{2,\mathtt{L}} \le e_\tau \le M_{2,\mathtt{U}}) \ge 1-\alpha_2$. Then the proof is complete. ∎

S.2.2 Verification of Condition (i) in Theorem 1

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

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

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 1 indeed requires that $\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ can be approximated by a Gaussian vector $\mathbf{G}$. In general, this can be shown if we assume $(u_{t,1}^{[1]},\cdots,u_{t,M}^{[1]},\cdots,u_{t,1}^{[N_1]},u_{t,M}^{[N_1]})'$ is independent or weakly dependent conditional on $\mathscr{H}$. In the following we verify condition (i) by imposing a conditional independence assumption. The extension that allows for weakly dependent errors can be established using the idea of Theorem A in Cattaneo et al. (2021).

Lemma S.1. *Assume $\mathcal{W}$ and $\mathcal{R}$ are convex, $\widehat{\boldsymbol{\beta}}$ in Equation (3.1) and $\boldsymbol{\beta}_0$ in Equation (4.2) exist, and $\mathscr{H}=\sigma(\mathbf{B},\mathbf{C},\mathbf{p}_\tau)$. In addition, for some finite nonnegative constants, the following conditions hold:*

(i) $\mathbf{u}_t = (u_{t,1}^{[1]},\cdots,u_{t,M}^{[1]},\cdots,u_{t,1}^{[N_1]},\cdots,u_{t,M}^{[N_1]})$ *is independent over t conditional on $\mathscr{H}$;*

(ii) $\mathbb{P}(\sum_{t=1}^{T_0}\mathbb{E}[\|\sum_{j=1}^{N_1}\sum_{l=1}^{M}\tilde{\mathbf{z}}_{t,l}^{[j]}(u_{t,l}^{[j]}-\mathbb{E}[u_{t,l}^{[j]}|\mathscr{H}])\|^3|\mathscr{H}] \ge \epsilon_\gamma(84(d^{1/4}+16))^{-1}) \ge 1-\pi_\gamma$ *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 $1-\pi_\gamma$,

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

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, $\mathcal{A}_\kappa := \{\boldsymbol{\xi} \in \mathbb{R}^d : \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\xi}} \mathbf{p}_\tau'\boldsymbol{\delta} \leq \kappa\}$ and $\mathcal{A}_\kappa' = \{\boldsymbol{\xi} \in \mathbb{R}^d : \inf_{\boldsymbol{\delta}\in\mathcal{M}_{\xi}} \mathbf{p}_\tau'\boldsymbol{\delta} \geq \kappa\}$ are convex. By Berry-Esseen Theorem for convex sets Raič (2019),

$$|\mathbb{P}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma} \in \mathcal{A}_\kappa|\mathscr{H}) - \mathbb{P}(\mathbf{G} \in \mathcal{A}_\kappa|\mathscr{H})| \leq 42(d^{1/4}+16)\sum_{t=1}^{T_0}\mathbb{E}\Big[\Big\|\sum_{j=1}^{N_1}\sum_{l=1}^{M}\tilde{\mathbf{z}}_{t,l}^{[j]}\tilde{u}_{t,l}^{[j]}\Big\|^3|\mathscr{H}\Big],$$

where $\tilde{u}_{t,l}^{[j]} = u_{t,l}^{[j]} - \mathbb{E}[u_{t,l}^{[j]}|\mathscr{H}]$. 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 $1-\pi_\gamma$,

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

Then the proof is complete. □

S.2.3 Proof of Lemma 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 $\boldsymbol{m}_{\mathtt{eq}}(\boldsymbol{\beta}_0) = \mathbf{0}$ and $\boldsymbol{m}_{\mathtt{in}}(\boldsymbol{\beta}_0) \leq \mathbf{0}$. If the jth inequality constraint is binding, i.e., $m_{\mathtt{in},j}(\boldsymbol{\beta}_0) = 0$, then $m_{\mathtt{in},j}(\widehat{\boldsymbol{\beta}}) = \frac{\partial}{\partial\boldsymbol{\beta}'}m_{\mathtt{in},j}(\tilde{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ for some $\tilde{\boldsymbol{\beta}}$ between $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$. Then, for $\mathfrak{c} := \max_{1\leq j\leq d_{\mathtt{in}}}\sup_{\boldsymbol{\beta}\in\mathcal{B}(\boldsymbol{\beta}_0,\varpi_\delta^\star)}\|\frac{\partial}{\partial\boldsymbol{\beta}}m_{\mathtt{in},j}(\boldsymbol{\beta})\|_*$, we have $\max_{1\leq j\leq d_{\mathtt{in}}}|m_{\mathtt{in},j}(\widehat{\boldsymbol{\beta}})| \leq \mathfrak{c}\varpi_\delta^\star$ 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 $1-\pi_\Delta^\star$. By the condition imposed on the tuning parameter ϱ_js, 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. Then, for $\varepsilon \leq \varpi_\delta^\star$,

$$\Delta^\star \cap \mathcal{B}(0,\varepsilon) = \Big\{\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}} \in \mathcal{B}(0,\varepsilon) : \boldsymbol{m}_{\mathtt{eq}}(\boldsymbol{\beta}) = \mathbf{0}, \boldsymbol{m}_{\mathtt{in},j}(\boldsymbol{\beta}) \leq \boldsymbol{m}_{\mathtt{in},j}(\widehat{\boldsymbol{\beta}}) \text{ for } j \in \mathcal{A}\Big\}.$$

Without loss of generality, we assume $\boldsymbol{m}_{\mathtt{in}}(\boldsymbol{\beta}_0) = \mathbf{0}$ from now on. Otherwise, the non-binding constraints can be dropped and the proof can proceed the same way described below.

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

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

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, $\|\Gamma^0 - \Gamma^\star\| \le \mathfrak{C}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|$ with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1 - \epsilon^\star_\Delta$, on an event with $\mathbb{P}$-probability at least $1 - \pi^\star_\Delta$.

Let

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

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

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

Thus, $\Gamma^0_{\mathtt{eq}}\boldsymbol{\lambda}^0 = \mathbf{0}$, $\Gamma^0_{\mathtt{in}}\boldsymbol{\lambda}^0 \le \mathbf{0}$. Note that by Taylor's expansion,

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

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, $\boldsymbol{m}_{\mathtt{eq}}(\tilde{\boldsymbol{\beta}}) = \mathbf{0}$ and $\boldsymbol{m}_{\mathtt{in}}(\tilde{\boldsymbol{\beta}}) \le \boldsymbol{m}_{\mathtt{in}}(\widehat{\boldsymbol{\beta}})$.

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. $\square$

S.3 Second Order Cone Programming

For the sake of completeness, at the beginning of this section we define again the three types of convex optimization problems presented in the main text in Section 5.1. Second, we illustrate the link between these three families of convex problems. Third, we note that the optimization problems underlying the prediction/estimation and uncertainty quantification problems for SC presented in Section 3 are QCQPs and QCLPs, respectively, and show how to represent them as SOCPs. Finally, we show how to write the Lasso-type constraint in conic form.

S.3.1 Families of convex optimization problems

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

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{x}'\mathbf{P}_0\mathbf{x} + \mathbf{q}_0'\mathbf{x} + w && \text{(S.3.1)} \\ \text{subject to} \quad & \mathbf{x}'\mathbf{P}_j\mathbf{x} + \mathbf{q}_j'\mathbf{x} + r_j \leq 0, \quad j = 1, \ldots, m && \text{(Quadratic inequality constraint)} \\ & \mathbf{F}\mathbf{x} = \mathbf{g}, && \text{(Linear equality constraint)} \end{aligned}$$

where $\mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_m \in \mathcal{M}_{n\times n}(\mathbb{R})$, $\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_m \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{F} \in \mathcal{M}_{m\times n}(\mathbb{R})$, $\mathbf{g} \in \mathbb{R}^m$, and $r_0, r_1, \ldots, r_m, w \in \mathbb{R}$. If all the matrices $\mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_m$ 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 which 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 defition 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

$\mathcal{K}_l := \{(k_0, \mathbf{k}_1) \in \mathbb{R} \times \mathbb{R}^l : ||\mathbf{k}_1||_2 \leq k_0\}, l = 1, \dots, L$. 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

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}'\mathbf{x}, && \text{(S.3.2)} \\ \text{subject to} \quad & \mathbf{G}\mathbf{x} \preceq_{\mathcal{K}} \mathbf{h}, && \text{(Second-order cone constraint)} \\ & \mathbf{A}\mathbf{x} = \mathbf{b}. && \text{(Linear equality constraint)} \end{aligned}$$

S.3.2 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.3.1) in the form of (S.3.2). 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.3.1) 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 $\mathbf{RR} = \mathbf{R}'\mathbf{R} = \mathbf{P}$. 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

$$||\mathbf{P}^{1/2}\mathbf{x}||_2^2 \leq y \iff \left\| \begin{bmatrix} 1 - y \\ 2\mathbf{P}^{1/2}\mathbf{x} \end{bmatrix} \right\|_2 \leq 1 + y, \tag{S.3.3}$$

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.3.3) 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 $\mathbf{P}^{1/2}\mathbf{x} \preceq_{\mathcal{C}} y$.

SC problem with lasso-type $\mathcal{W}$ as a SOCP. 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:

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{r}} \quad (\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r})'\mathbf{V}(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r}) && \text{(S.3.4)} \\ &\text{subject to} \quad ||\mathbf{w}^{[i]}||_1 \leq Q_1^{[i]}, \quad i=1,\ldots,N_1 && \text{(L1 inequality constraints)} \end{aligned}$$

We can write the optimization problem in (S.3.4) as a SOCP of the following form

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{r},\{\mathbf{z}_i\}_{i=1}^{N_1},v} \quad v \\ &\text{subject to} \quad \begin{bmatrix} 1-v \\ 2\mathbf{V}^{1/2}(\mathbf{A}-\mathbf{B}\mathbf{w}-\mathbf{C}\mathbf{r}) \end{bmatrix} \preceq_{\mathcal{C}_1} 1+v, && \text{(cone in } \mathbb{R}^{2+T_0\cdot M\cdot N_1}) \\ &\qquad \mathbf{1}'\mathbf{z}_i \preceq_{\mathcal{C}_2} Q_1^{[i]}, \quad i=1,\ldots,N_1 && (N_1 \text{ cones in } \mathbb{R}) \\ &\qquad -\mathbf{w} \preceq_{\mathcal{C}_3} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J\cdot N_1}) \\ &\qquad \mathbf{w} \preceq_{\mathcal{C}_4} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J\cdot N_1}) \end{aligned}$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2^{N_1} \times \mathcal{C}_3 \times \mathcal{C}_4 = \mathcal{K}_{1+T_0\cdot M\cdot N_1} \times \mathbb{R}_+^{N_1} \times \mathbb{R}_+^{J\cdot N_1} \times \mathbb{R}_+^{J\cdot N_1}$ is the conic constraint for this program and $\mathbf{z} := (\mathbf{z}_1', \ldots, \mathbf{z}_{N_1}')'$.

For uncertainty quantification, we need to solve the optimization problem underlying (4.4). Here we discuss the lower bound only for brevity. Recalling that $\boldsymbol{\beta} = (\mathbf{w}', \mathbf{r}')'$, we have

$$\begin{aligned} &\inf_{\boldsymbol{\beta}=(\mathbf{w}',\mathbf{r}')'} \quad \mathbf{p}_\tau'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) && \text{(S.3.5)} \\ &\text{subject to} \quad ||\mathbf{w}^{[i]}||_1 \leq Q_1^{[i]} + \varrho^{[i]}, \quad i=1,\ldots,N_1 && \text{(L1 inequality constraints)} \\ &\qquad (\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})'\widehat{\mathbf{Q}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) - 2(\mathbf{G}^\star)'(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}) \leq 0, && \text{(constrained least squares)} \end{aligned}$$

where the scalars $\varrho_1^{[i]}, i=1,\ldots,N_1$ are regularization parameters used to relax Δ to $\Delta^\star$.

We can cast the SC optimization problem in (S.3.5) in conic form as follows:

$$\begin{aligned} &\min_{\mathbf{w},\mathbf{r},\{\mathbf{z}_i\}_{i=1}^{N_1},t} \quad \mathbf{p}_\tau'\boldsymbol{\beta} \\ &\text{subject to} \quad t + \mathbf{a}'\boldsymbol{\beta} + f \preceq_{\mathcal{C}_1} 0, && \text{(cone in } \mathbb{R}) \\ &\qquad \mathbf{1}'\mathbf{z}^{[i]} \preceq_{\mathcal{C}_2} Q_1^{[i]} + \varrho^{[i]}, \quad i=1,\ldots,N_1 && (N_1 \text{ cones in } \mathbb{R}) \\ &\qquad -\mathbf{w} \preceq_{\mathcal{C}_3} \mathbf{z}, && \text{(cone in } \mathbb{R}^{J\cdot N_1}) \\ &\qquad \mathbf{w} \preceq_{\mathcal{C}_4} \mathbf{z}. && \text{(cone in } \mathbb{R}^{J\cdot N_1}) \\ &\qquad \begin{bmatrix} 1-t \\ 2\mathbf{Q}^{1/2}\boldsymbol{\beta} \end{bmatrix} \preceq_{\mathcal{C}_5} 1+t, && \text{(cone in } \mathbb{R}^{2+(J+KM)\cdot N_1}) \end{aligned}$$

where $\mathcal{K} = \mathcal{C}_1 \times \mathcal{C}_2^{N_1} \times \mathcal{C}_3 \times \mathcal{C}_4 \times \mathcal{C}_5 = \mathbb{R}_+ \times \mathbb{R}_+^{N_1} \times \mathbb{R}_+^{J \cdot N_1} \times \mathbb{R}_+^{J \cdot N_1} \times \mathcal{K}_{1+(J+KM)\cdot N_1}$ 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.4 Data Preparation and Software Implementation

In this section, we first describe the variables in the Billmeier and Nannicini (2013) (BN, henceforth) dataset and explain all the changes and additions we made. Then, we illustrate how to use the companion `R` package `scpi` to prepare the data for the analysis and we briefly explain how to replicate the results.

S.4.1 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 condition 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 two variables because of missing data. Specifically, in the case of Europe - analyzed in the main article - at least one of these variables is completely missing before 1990 for Albania, Bulgaria, Romania, Slovak Republic, and Slovenia. Moreover, the original schooling variable has missing data for our countries of interest. Thus we replace it with the percentage of complete secondary schooling attained in population from the Barro-Lee dataset on educational attainment.[2]

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 data cleaning criteria to select the countries to be included in our final dataset:

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

[2] *source*: `http://www.barrolee.com/`

1. At least 35 data entries for real GDP per capita from 1963 to 2000. This excludes the following economies (number of periods when real GDP per capita is observed in parenthesis): Armenia (8), Azerbaijan (8), Belarus (8), Dominican Republic (21), Ecuador (31), El Salvador (30), Estonia (8), Georgia (8), Kazakhstan (8), Kyrgyzstan (8), Latvia (8), Lithuania (8), Moldova (8), Russia (8), Tajikistan (8), Turkmenistan (8), Ukraine (8), and Uzbekistan (8).

2. At least 15 pre-treatment periods, that is, the country is observed for at least 15 periods before it embarks on the liberalization program in the Sachs-Warner. This excludes the following economies (year in which the country's economy was first declared liberalized in parenthesis): Australia (1964), Austria ($\leq$ 1963), Barbados (1966), Belgium ($\leq$ 1963), Bolivia (1964), Canada ($\leq$ 1963), Chile (1976), Cyprus (1964), Denmark (1966), Finland ($\leq$ 1963), France ($\leq$ 1963), Germany ($\leq$ 1963), Greece ($\leq$ 1963), Hong Kong ($\leq$ 1963), Indonesia (1970), Ireland (1966), Italy ($\leq$ 1963), Jamaica ($\leq$ 1963), Japan (1964), Jordan (1965), Kenya (1965), Luxembourg ($\leq$ 1963), Malaysia ($\leq$ 1963), Mauritius (1968), Morocco ($\leq$ 1963), Netherlands ($\leq$ 1963), Norway ($\leq$ 1963), Peru ($\leq$ 1963), Portugal ($\leq$ 1963), Singapore (1965), South Korea (1968), Spain ($\leq$ 1963), Sri Lanka (1977), Sweden ($\leq$ 1963), Switzerland ($\leq$ 1963), Syria ($\leq$ 1963), Taiwan ($\leq$ 1963), Thailand ($\leq$ 1963), United Kingdom ($\leq$ 1963), United States ($\leq$ 1963), Yemen (1965).

3. At least one entry for the percentage of complete secondary schooling attained in population. This excludes the following economies: Angola, Burkina Faso, Chad, Ethiopia, Guinea, Guinea-Bissau, Madagascar, Nigeria, and North Macedonia.

4. The liberalization is an absorbing state, that is, whenever the country embarks on a liberalization program, it remains so for the whole remaining time this country is observed. This excludes Venezuela.

Table S.1 shows the final set of countries we select, together with their treatment date.

Table S.1: *List of all countries included in our analysis*

Africa				Asia			
Algeria	∞	Madagascar	1996	Bangladesh	1996	Nepal	1991
Angola	∞	Malawi	∞	China	∞	Pakistan	∞
Benin	1990	Mali	1988	India	∞	Philippines	1988
Botswana	1979	Mauritania	1995	Iran	∞	Turkey	1989
Burkina Faso	1998	Mozambique	1995	Israel	1985		
Burundi	1999	Niger	1994				
Cameroon	1993	Nigeria	∞				
Chad	∞	Rwanda	∞				
Congo	∞	Senegal	∞	Europe			
Egypt	1995	Sierra Leone	∞	Albania	1992	Malta	∞
Ethiopia	1996	South Africa	1991	Bulgaria	1991	North Macedonia	1994
Gabon	∞	Tanzania	1995	Croatia	∞	Poland	1990
Gambia	1985	Togo	∞	Czech Republic	1991	Romania	1992
Ghana	1985	Tunisia	1989	Hungary	1990	Slovak Republic	1991
Guinea	1986	Uganda	1988	Iceland	∞	Slovenia	1991
Guinea-Bissau	1987	Zambia	1993				
Ivory Coast	1994	Zimbabwe	∞				
Lesotho	∞						
North America				South America			
Costa Rica	1986	Mexico	1986	Argentina	1991	Guyana	1988
Guatemala	1988	Nicaragua	1991	Brazil	1991	Paraguay	1989
Haiti	∞	Panama	1996	Colombia	1986	Uruguay	1990
Honduras	1991						

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

To be precise, not all treated units in Table S.1 are analyzed in our work; we only study our target causal predictands for those units having at least 10 post-treatment periods. In other words, units treated after 1992 are not analyzed. Despite this fact, we still include units treated after 1992 because the staggered nature of this design, together with the fact that we focus on effects up to 10 years after liberalization events, make it possible to use such economies as donor units. To better illustrate this concept, take the case of Botswana that gets the treatment in 1979 and for which we study the effects of such treatment up to 1988 (see Figure S.1c). This allows us to use all countries treated from 1989 (included) onwards as donor units. In the same spirit, all countries treated after 1994 can be used as donor units for Israel (see Figure S.4c).

S.4.2 Code

In what follows we briefly explain how to replicate the results presented in the main paper and in this supplemental appendix using our `R` software package `scpi`. In particular, we focus on the two functions `scdataMulti()` and `scplotMulti()`, which have to be considered as the multiple treated units counterpart of `scdata()` and `scplot()` that we introduce and thoroughly describe in Cattaneo, Feng, Palomba and Titiunik (2023).

The next snippet shows how to prepare the data to predict the individual treatment effect τ_{it} and the average post-treatment effect $\tau_{i\cdot}$. To accomplish this task, we leverage the function `scdataMulti()`, which receives as an input an object of class `DataFrame` and outputs the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{P}$ together with other useful objects that are descried in detail in the help files.

```
# Load packages and set seed
```

```r
library(scpi)
library(haven)
set.seed(8894)

# Load dataset
data <- haven::read_dta(paste0(path.data, "final_data.dta"))
data <- subset(data, restricted2 == 1)
data$lgdp <- log(data$rgdppp)

# Features to be matched
features <- list(c("lgdp","lsc"))

# Covariates for adjustment
covs.adj <- list("lgdp" = c("constant", "trend"),
         "lsc" = c("constant", "trend"))

# Select treated units to be analyzed
units <- unique(subset(data, continent == "Europe" & treated == 1 & trDate <= 1992)$
    countryname)
units <- units[!(units %in% c("Czech Republic", "Poland"))]

# Set number of post-estimation periods
post.est <- 10

# Prepare data for Individual treatment effect
df <- scdataMulti(data, id.var = "countryname", outcome.var = "lgdp",
            treatment.var = "liberalization", time.var = "year",
            constant = TRUE, cointegrated.data = TRUE, post.est = post.est,
            units.est = units, features = features, cov.adj = covs.adj, anticipation = 1)

# Prepare data for Average post-treatment effect
df.unit <- scdataMulti(data, id.var = "countryname", outcome.var = "lgdp", effect = "unit",
             treatment.var = "liberalization", time.var = "year", constant = TRUE,
             cointegrated.data = TRUE, post.est = post.est, units.est = units,
             features = features, cov.adj = covs.adj, anticipation = 1)

# Prepare data for Average treatment effect on the treated
df.time <- scdataMulti(data, id.var = "countryname", outcome.var = "lgdp", effect = "time",
             treatment.var = "liberalization", time.var = "year",
             constant = TRUE, cointegrated.data = TRUE, post.est = post.est,
             units.est = units[[cont]], features = features, cov.adj = covs.adj,
             anticipation = 1)
```

First of all, it is important to notice the presence of the option `effect = "unit"` when it comes to predict the average post-treatment effect and of the option `effect = "time"` when the predictand of interest is the average treatment effect on the treated. Second, the most important input variable for `scdataMulti()` is the treatment variable `treatment.var`. In our case, `treatment.var` is the Sachs-Warner indicator which *de facto* has 1 as the absorbing state (once a country is described as open in our data, it remains so for all the period of analysis). Third, `scdataMulti()` has two options that allow users to choose which treated units to study and how many post-treatment periods should be considered in the analysis. These options are `units.est` and `post.est`, respectively. Finally, all the other options—`cointegrated.data`, `constant`, `anticipation`, `features`, `cov.adj`, `id.var` and `time.var`—are identical to those of `scdata()`.

Once the data matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{P}$ have been prepared, uncertainty quantification proceeds in the standard way. Specifically, the user has to call the function `scpi()` with the appropriate input object.

```r
res <- scpi(df, sims = sims, cores = cores, w.constr = list("name" = "L1-L2"),
          u.order = 1, u.lags = 1)

res.unit <- scpi(df.unit, sims = sims, cores = cores, w.constr = list("name" = "L1-L2"),
             u.order = 1, u.lags = 1)
```

```
res.time <- scpi(df.time, sims = sims, cores = cores, w.constr = list("name" = "L1-L2"),
             u.order = 1, u.lags = 1)
```

Finally, it is possible to visualize the results through the function `scplotMulti()`. This function allows the user to either plot the synthetic control and the actual time series (`type = "series"`) or directly the treatment effect time series (`type = "treatment"`). This latter is simply the difference between the actual time series and the synthetic control one. In addition, simultaneous prediction intervals are readily available through the option `joint = TRUE`.

```
scplotMulti(res[, type = "series", joint = TRUE, ncols = 2)
scplotMulti(res.unit[, type = "series", joint = TRUE, ncols = 2)
scplotMulti(res.time[, type = "series", joint = TRUE, ncols = 2)

scplotMulti(res[, type = "treatment", joint = TRUE, ncols = 2)
scplotMulti(res.unit[, type = "treatment", joint = TRUE, ncols = 2)
scplotMulti(res.time[, type = "treatment", joint = TRUE, ncols = 2)
```

The next and final snippet showcases how to prepare the data for the third causal quantity analyzed in this paper: the average treatment effect on the treated at s_0.

```
# Load packages and set seed
library(scpi)
library(haven)
set.seed(8894)

data <- haven::read_dta(paste0(path.data, "final_data.dta"))
data <- subset(data, restricted2 == 1)
data$lgdp <- log(data$rgdppp)
post.est <- 10

# One Feature
eu.tr.91 <- unique(subset(data, trDate %in% c(1990,1991,1992) & continent == "Europe")$
    countryname)

data.co <- subset(data, !(countryname %in% eu.tr.91))
data.tr <- subset(data, countryname %in% eu.tr.91)

data.tr.agg <- aggregate(data.tr[c("lgdp", "liberalization")],
                         by=list(data.tr$year), FUN=mean, na.rm = TRUE)
names(data.tr.agg) <- c("year","lgdp","liberalization")
data.tr.agg$countryname <- "Average Unit"
data.tr.agg$liberalization <- 1*(data.tr.agg$liberalization > 0)
data.agg <- rbind(data.tr.agg, data.co[names(data.tr.agg)])

features <- list(c("lgdp"))
covs.adj <- list("lgdp" = c("constant", "trend"))

df.agg <- scdataMulti(data.agg, id.var = "countryname", outcome.var = "lgdp",
                      treatment.var = "liberalization", time.var = "year",
            constant = FALSE, cointegrated.data = T, post.est = 10,
            units.est = "Average Unit", features=features, cov.adj=covs.adj)

res.avg.l1l2 <- scpi(df.agg, sims = sims, cores = 1, w.constr = list("name" = "L1-L2"),
                     u.order = 1, u.lags = 1)
```

Data preparation follows exactly the lines of Example 3.3, and thus the selected treated units are aggregated into a single one—termed "*ave*"—whose potential outcomes are simply

$$Y_t^{\mathsf{ave}}(s) := \frac{1}{N_{s_0}} \sum_{i:T_i=s_0} Y_{it}(s), \quad t = 1, \cdots, T.$$

The predictand of interest can be computed using "ave" as the only treated unit and control units from pre-treatment periods. Then, everything else proceeds as in the individual treatment effect case.

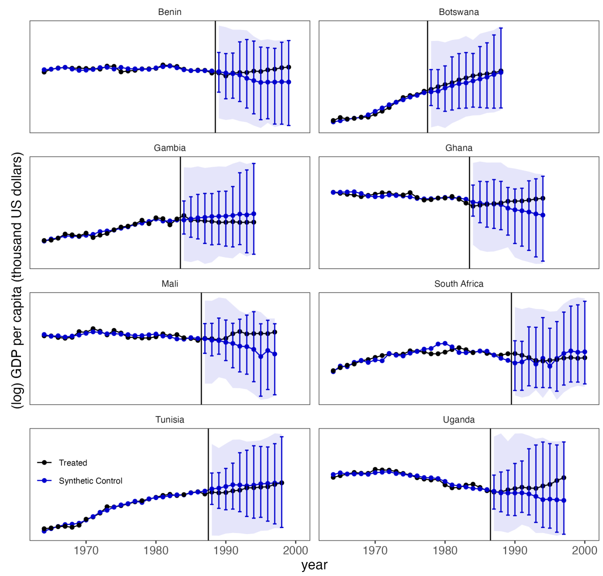
S.5 Results with Simplex Constraint

In this Section we report the results for Africa, Asia, Europe, North America, and South America using the simplex constraint in place of the L1-L2 constraint used in the main analysis, that is,

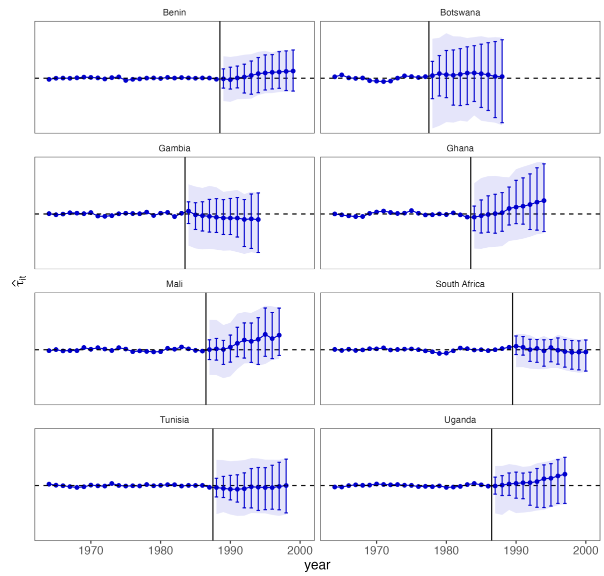
$$\mathcal{W} = \bigtimes_{i=1}^{N_1} \left\{ \mathbf{w}^{[i]} \in \mathbb{R}_+^J : ||\mathbf{w}^{[i]}||_1 = 1 \right\}.$$

S.5.1 Africa

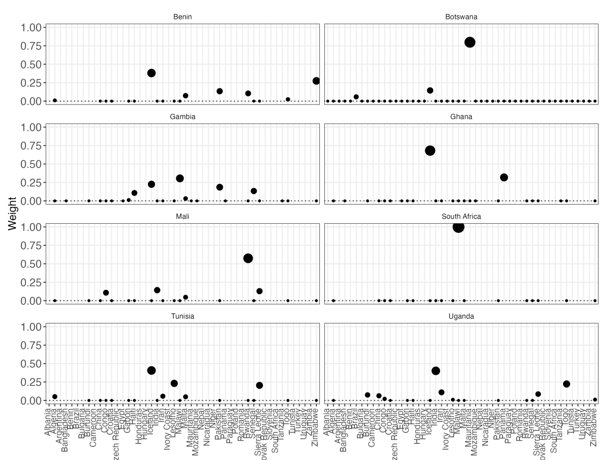
Figure S.1: *Individual Treatment Effects* $\widehat{\tau}_{it}$.



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



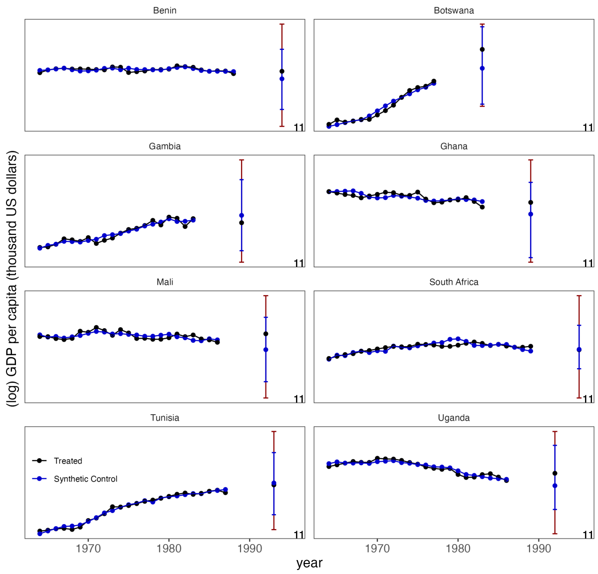
(b) $\widehat{\tau}_{it}$



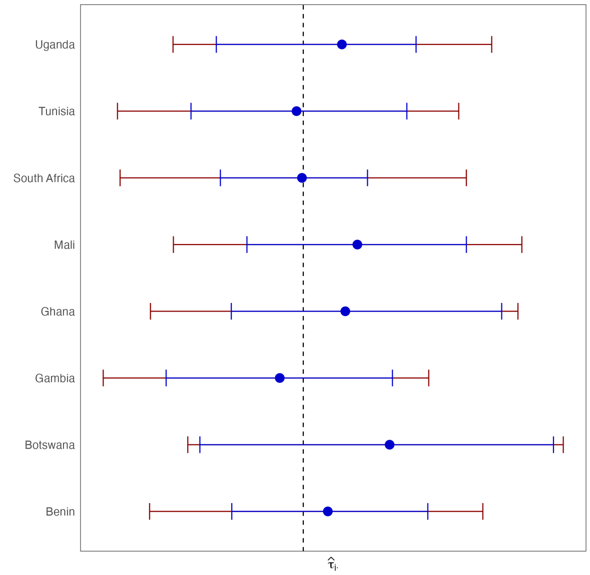
(c) $\widehat{w}_i$

Notes: 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.

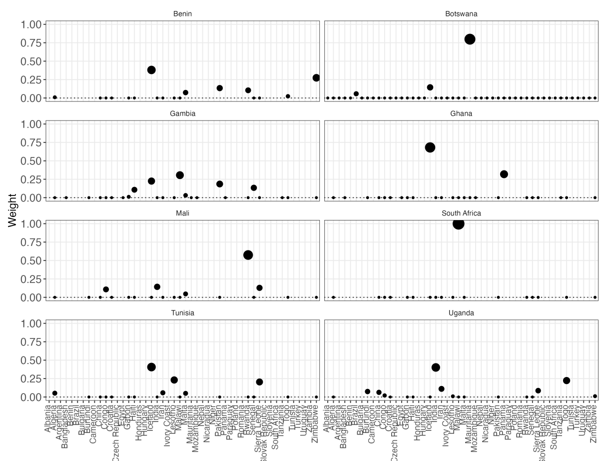
Figure S.2: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



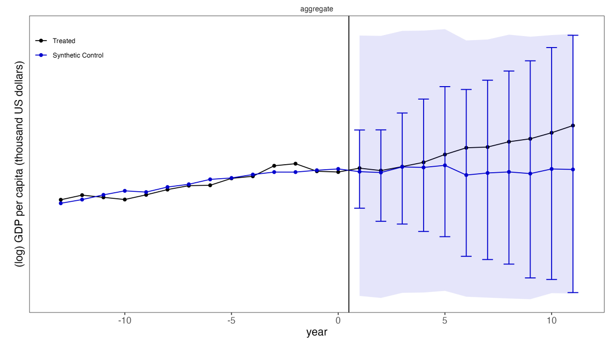
(b) $\widehat{\tau}_{i\cdot}$



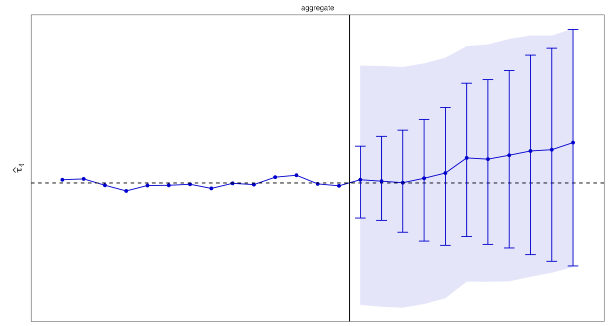
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

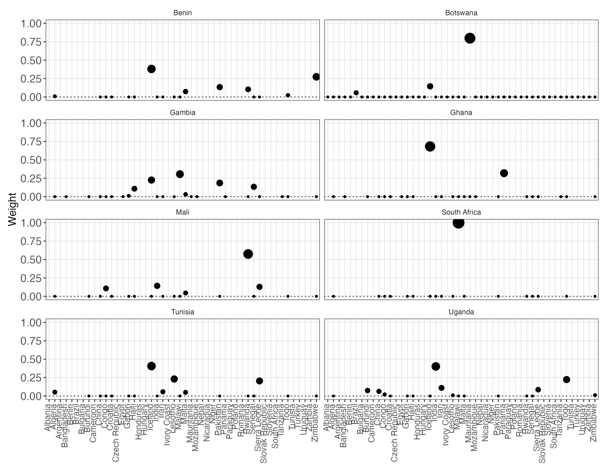
Figure S.3: *Average Treatment Effects on the Treated $\tau_{\cdot k}$.*



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



(b) $\widehat{\tau}_{\cdot k}$

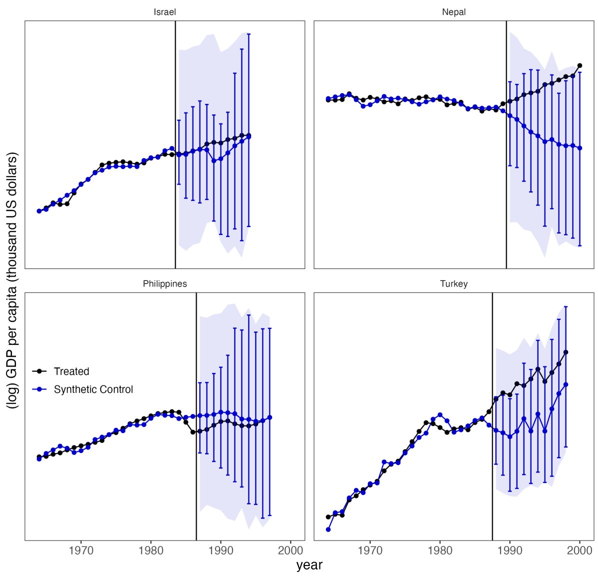


(c) $\widehat{w}_i$

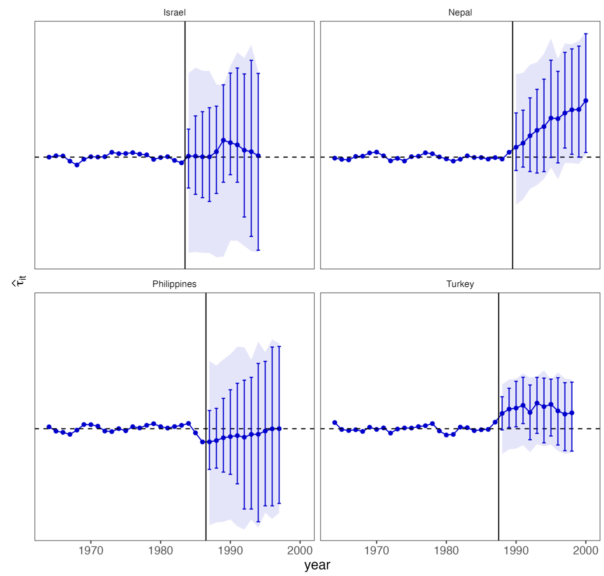
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.5.2 Asia

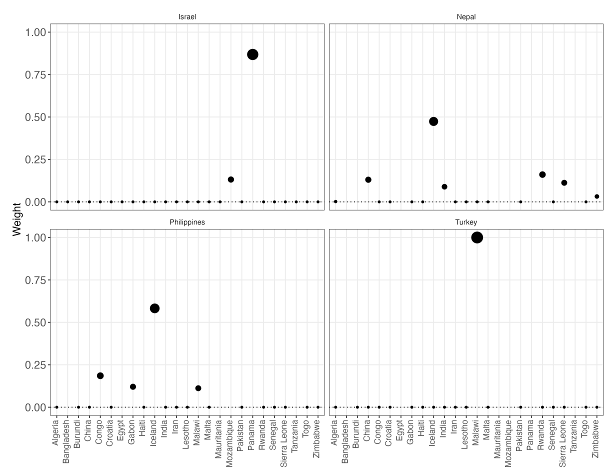
Figure S.4: *Individual Treatment Effects* $\widehat{\tau}_{it}$.



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



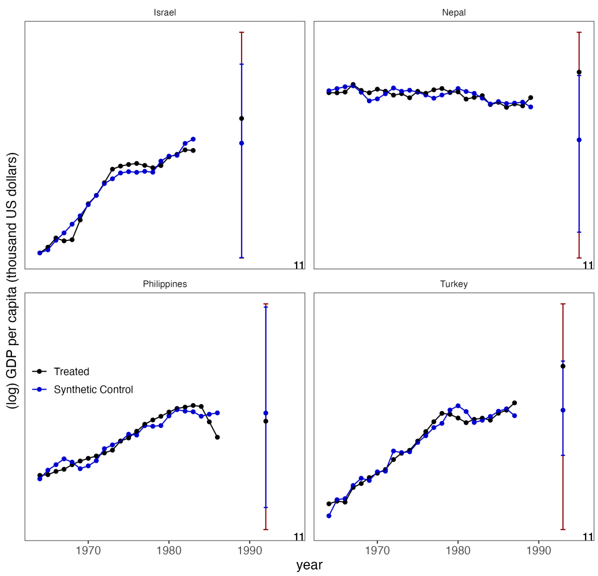
(b) $\widehat{\tau}_{it}$



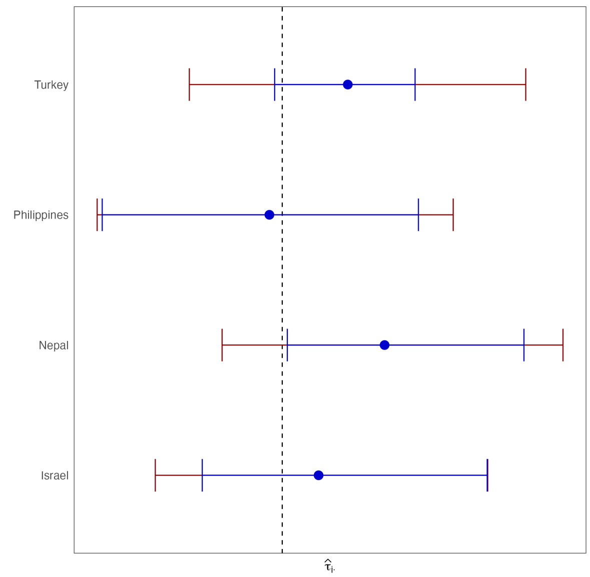
(c) $\widehat{w}_i$

Notes: 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.

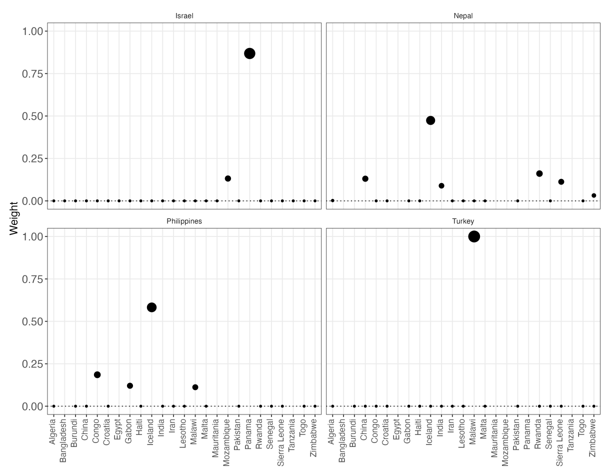
Figure S.5: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



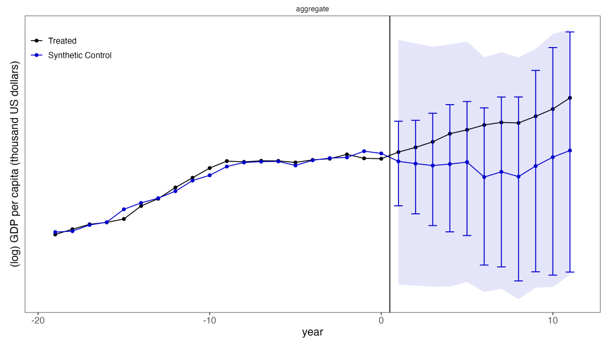
(b) $\widehat{\tau}_{i\cdot}$



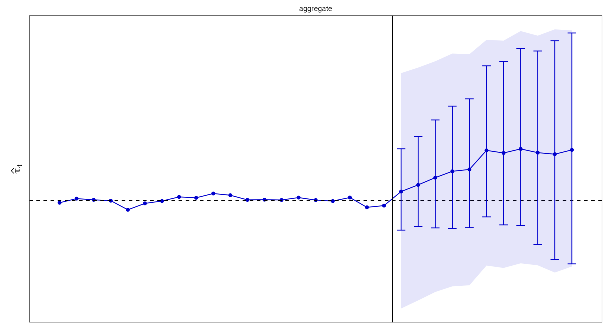
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

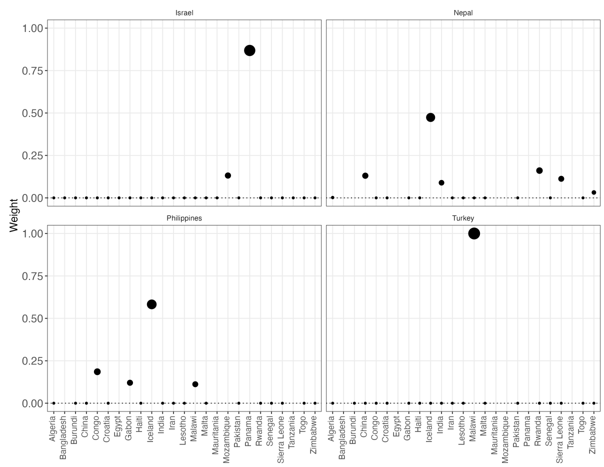
Figure S.6: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$

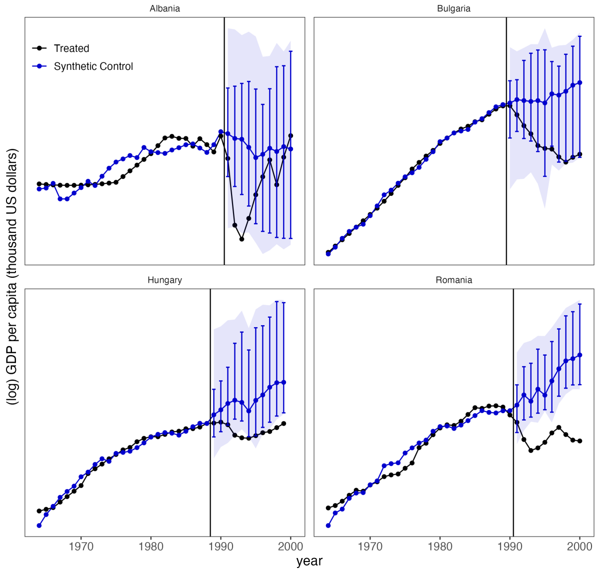


(c) $\widehat{w}_i$

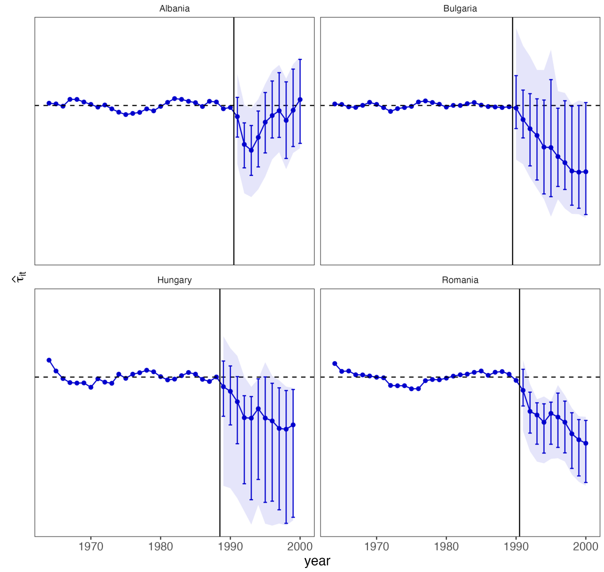
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.5.3 Europe

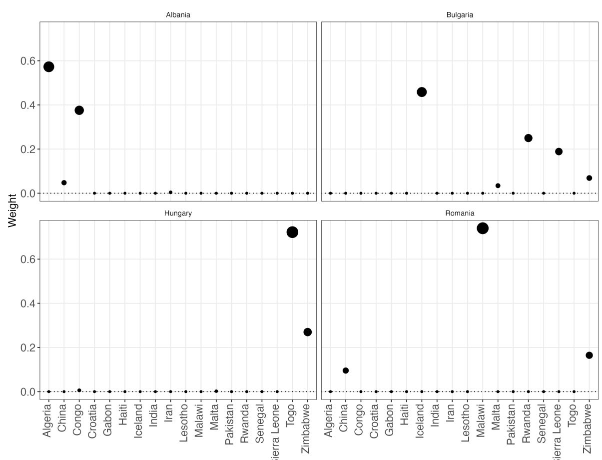
Figure S.7: *Individual Treatment Effects* $\widehat{\tau}_{it}$.



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



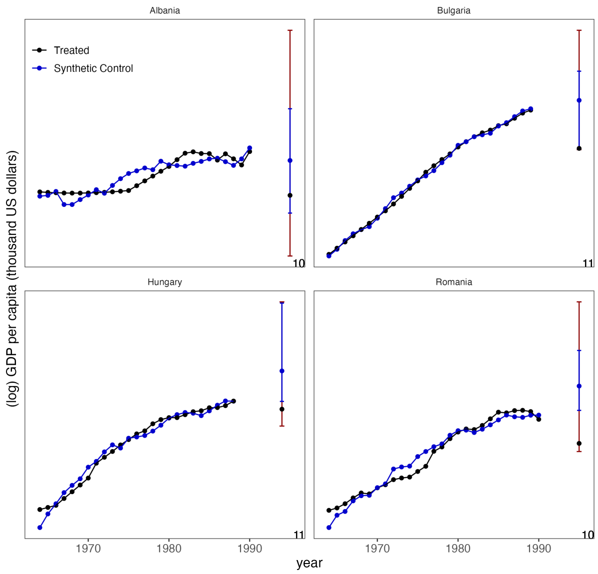
(b) $\widehat{\tau}_{it}$



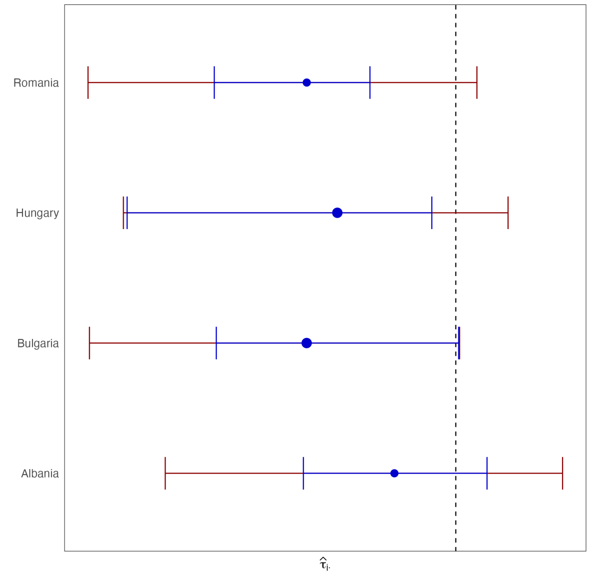
(c) $\widehat{w}_i$

Notes: 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.

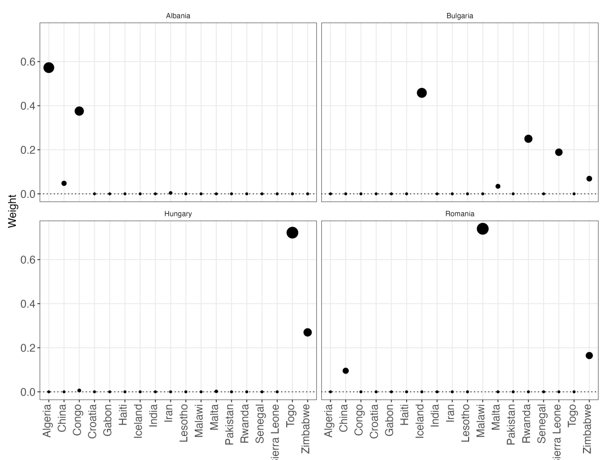
Figure S.8: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



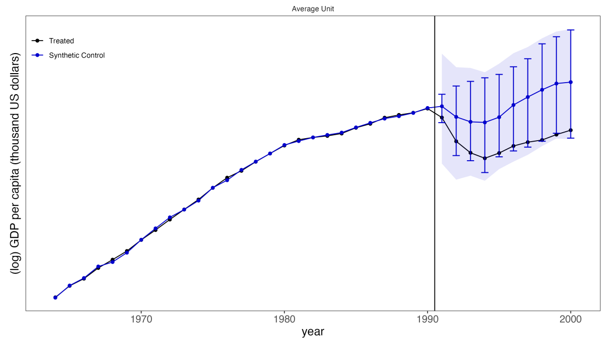
(b) $\widehat{\tau}_{i\cdot}$



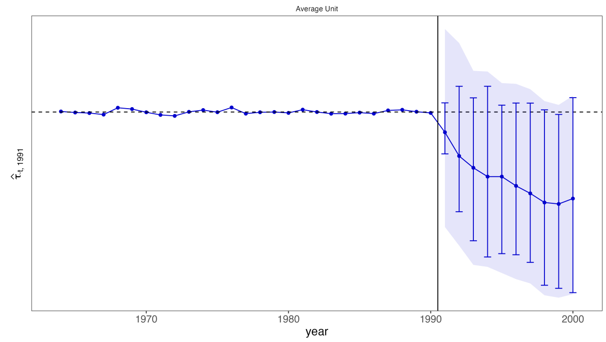
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

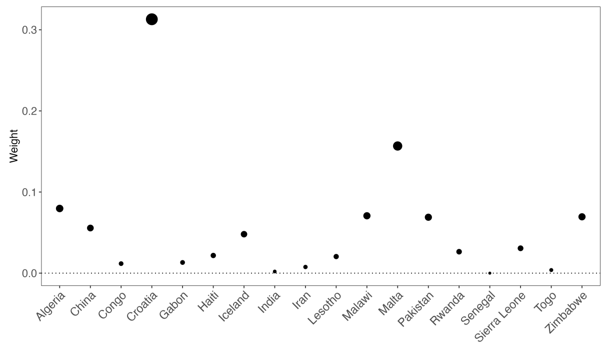
Figure S.9: *Average Treatment Effect on the Treated in 1990 and 1991.*



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



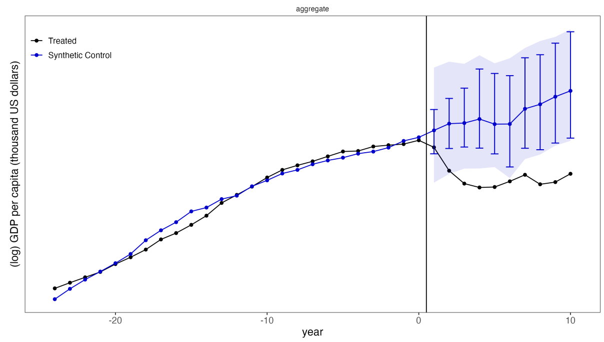
(b) $\widehat{\tau}_{it}$



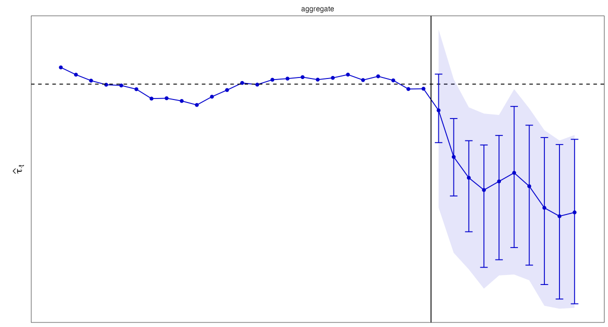
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 500 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

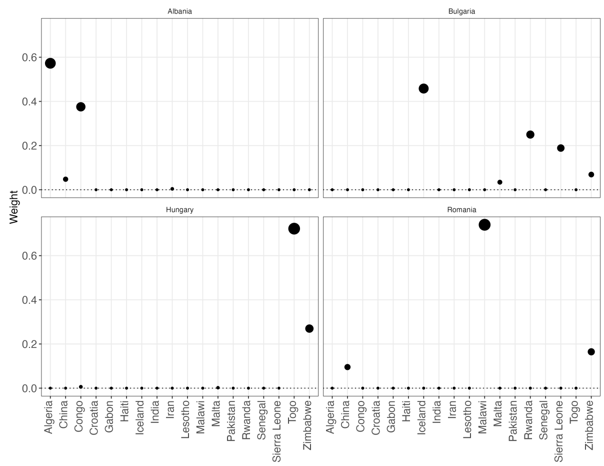
Figure S.10: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$

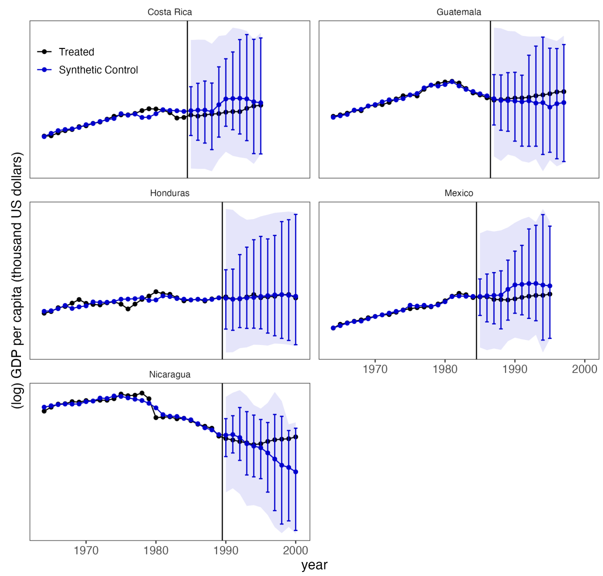


(c) $\widehat{w}_i$

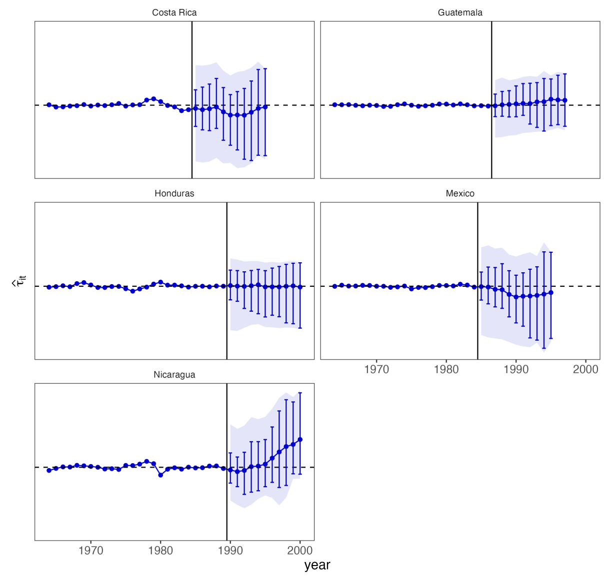
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.5.4 North America

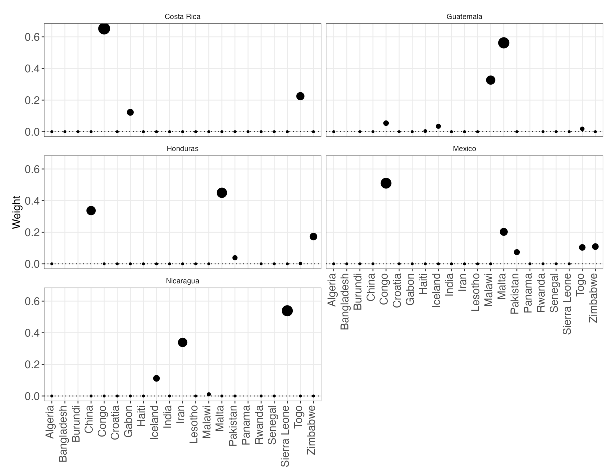
Figure S.11: *Individual Treatment Effects $\widehat{\tau}_{it}$.*



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



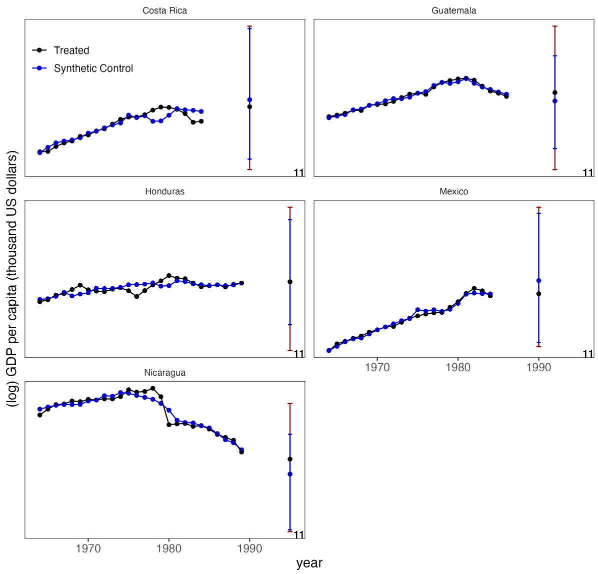
(b) $\widehat{\tau}_{it}$



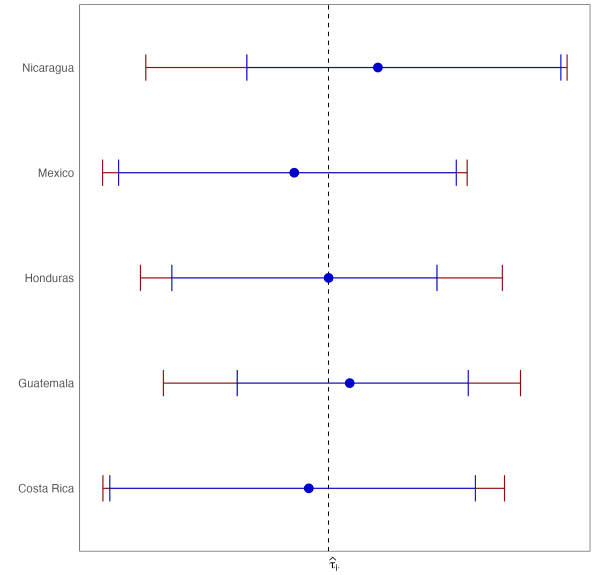
(c) $\widehat{w}_i$

Notes: 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.

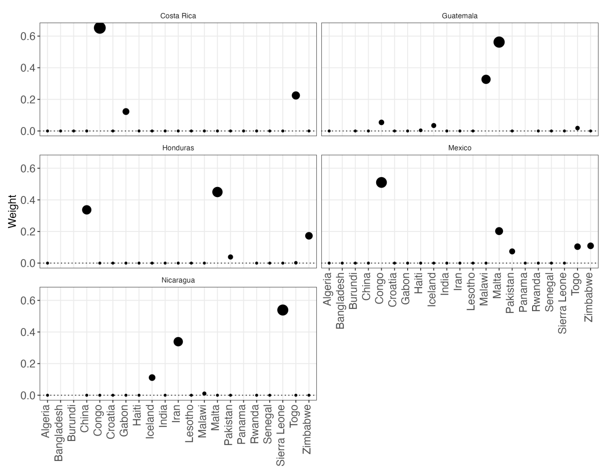
Figure S.12: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



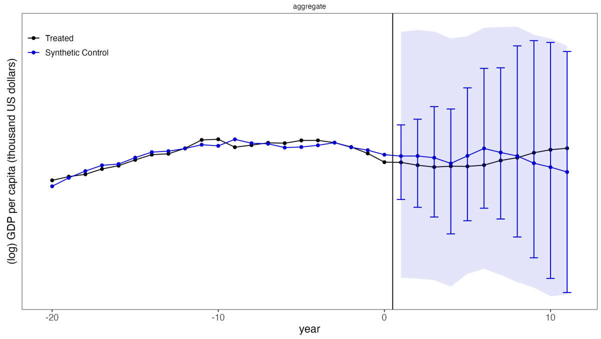
(b) $\widehat{\tau}_{i\cdot}$



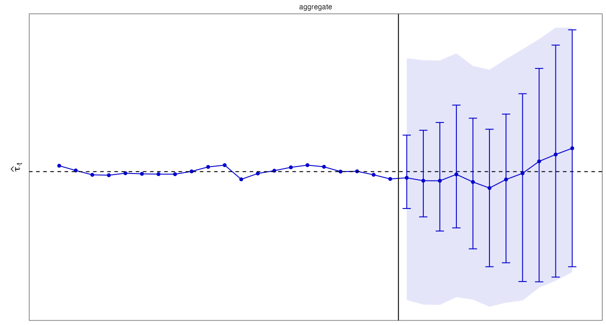
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

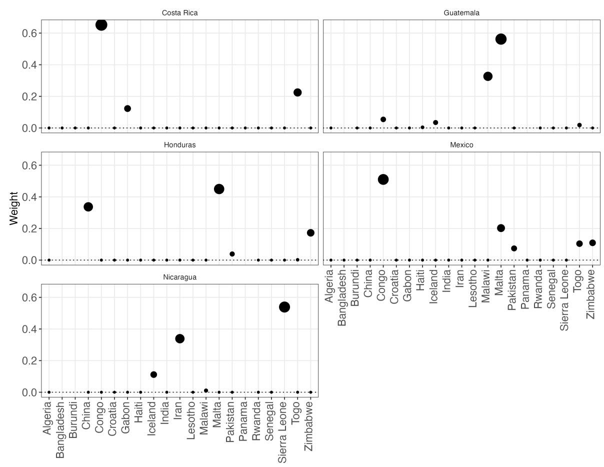
Figure S.13: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$

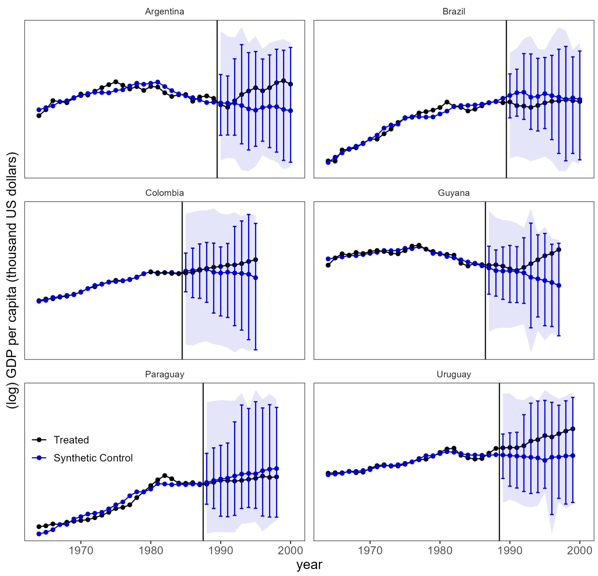


(c) $\widehat{w}_i$

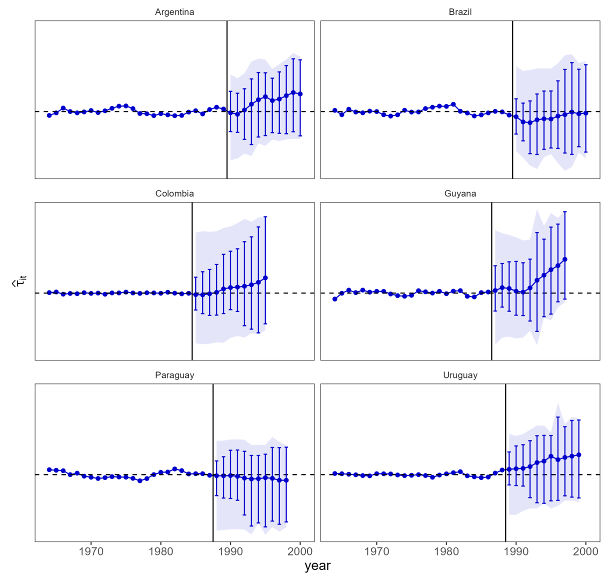
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.5.5 South America

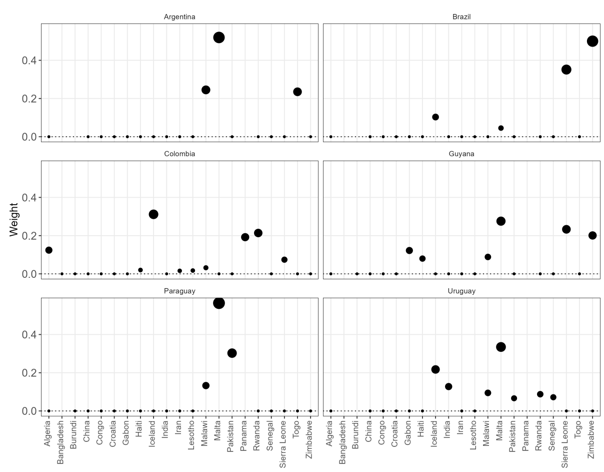
Figure S.14: *Individual Treatment Effects* $\widehat{\tau}_{it}$.



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



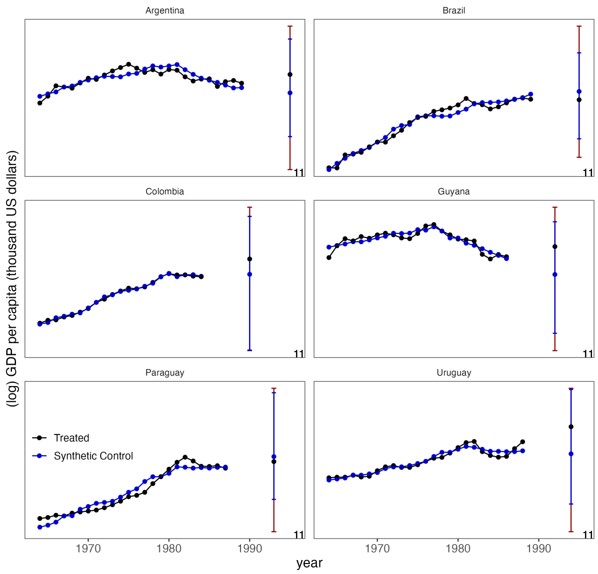
(b) $\widehat{\tau}_{it}$



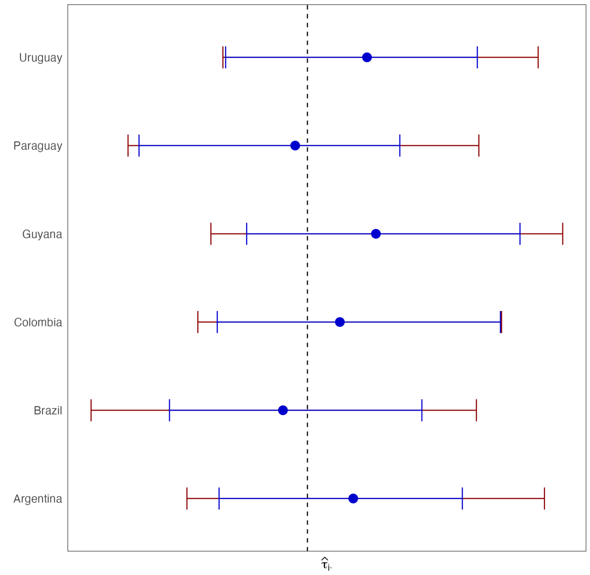
(c) $\widehat{w}_i$

Notes: 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.

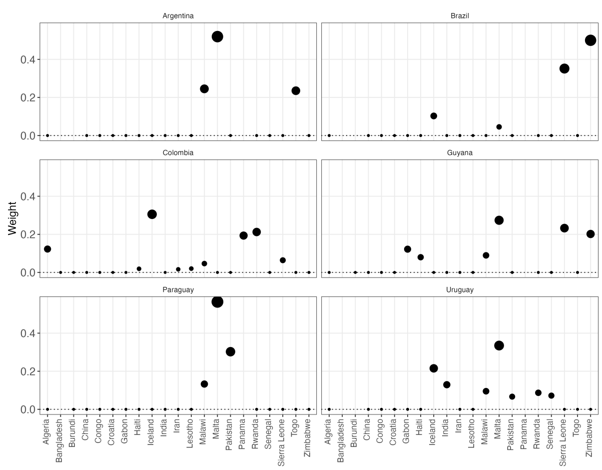
Figure S.15: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



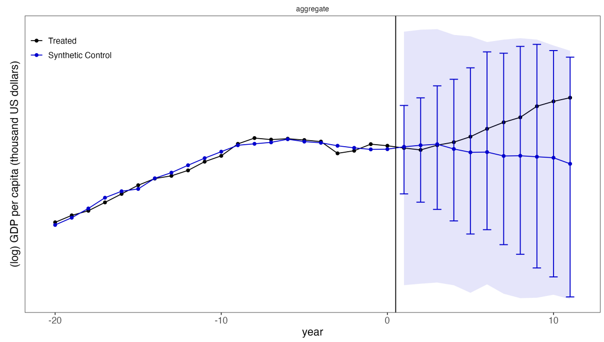
(b) $\widehat{\tau}_{i\cdot}$



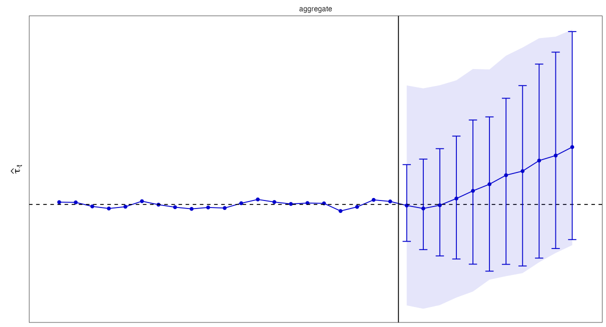
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

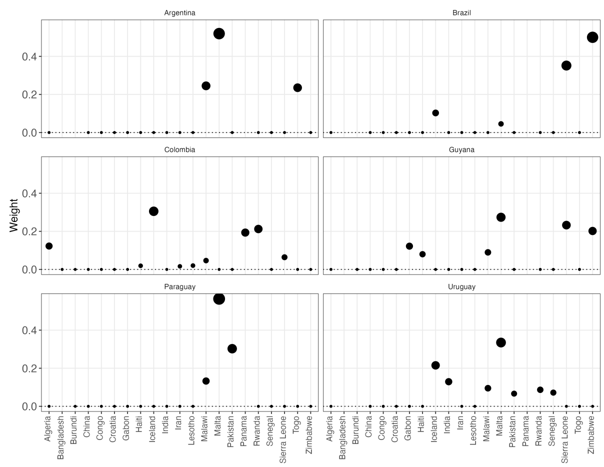
Figure S.16: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$



(c) $\widehat{w}_i$

Notes: 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 using sub-Gaussian bounds.

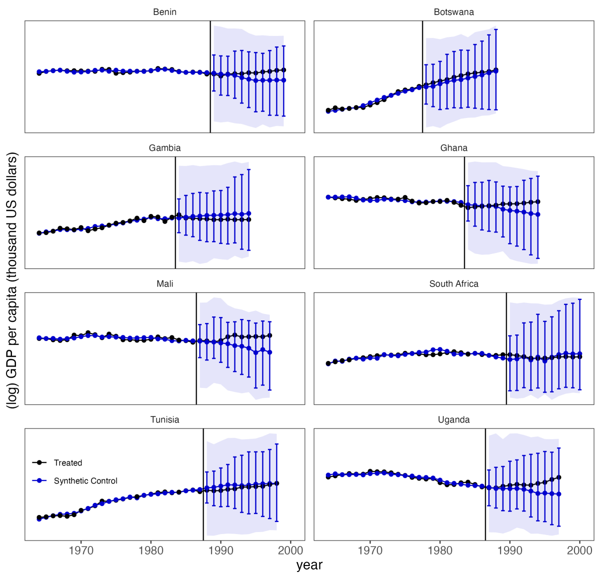
S.6 Results with L1-L2, Other Continents

In this Section we report the results for Africa, Asia, North America, and South America using the L1-L2 constraint as in the main analysis for Europe, that is,

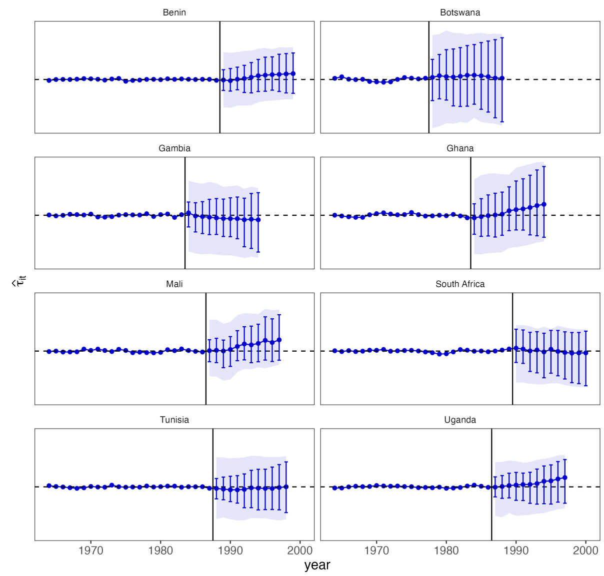
$$\mathcal{W} = \bigtimes_{i=1}^{N_1} \left\{ \mathbf{w}^{[i]} \in \mathbb{R}^J : ||\mathbf{w}^{[i]}||_1 = 1, ||\mathbf{w}^{[i]}||_2 \leq Q_2^{[i]} \right\}.$$

S.6.1 Africa

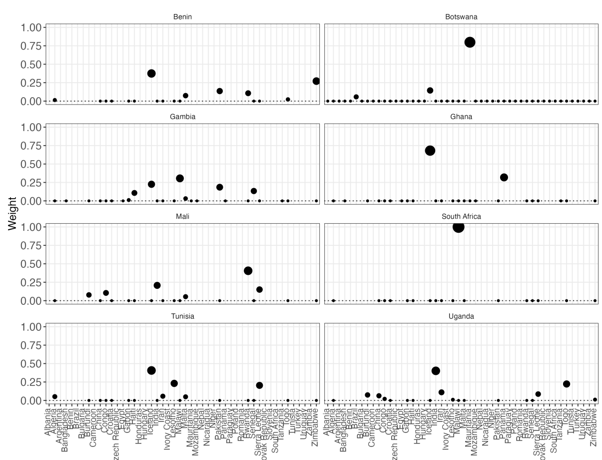
Figure S.17: *Individual Treatment Effects $\widehat{\tau}_{it}$.*



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



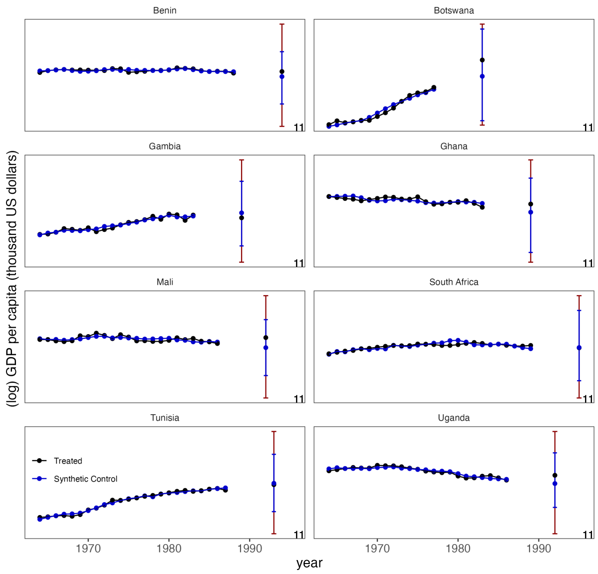
(b) $\widehat{\tau}_{it}$



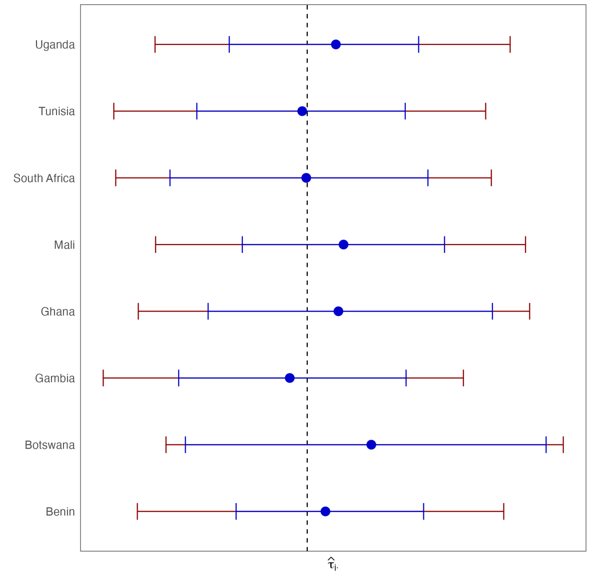
(c) $\widehat{w}_i$

Notes: 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.

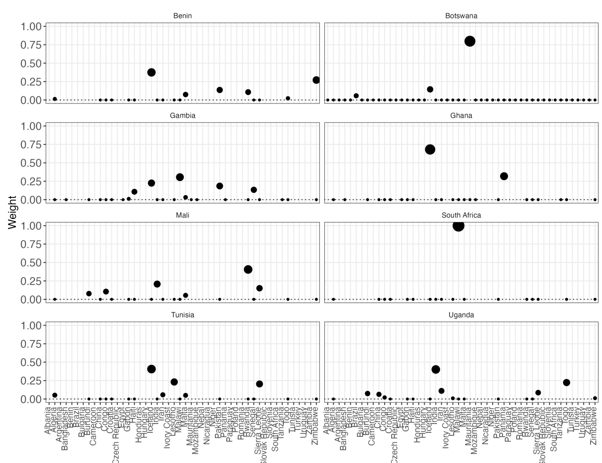
Figure S.18: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



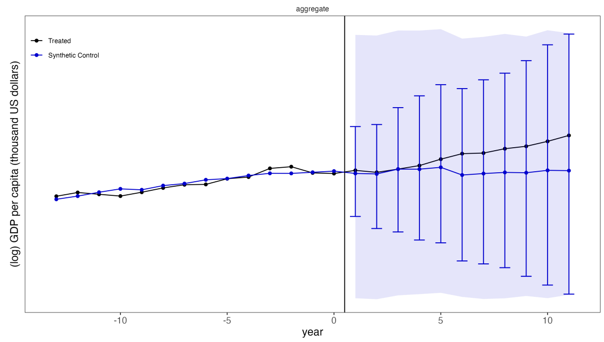
(b) $\widehat{\tau}_{i\cdot}$



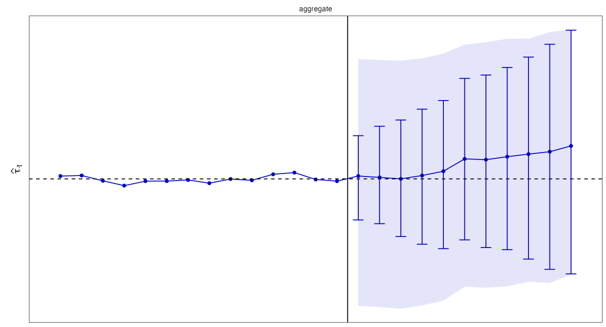
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

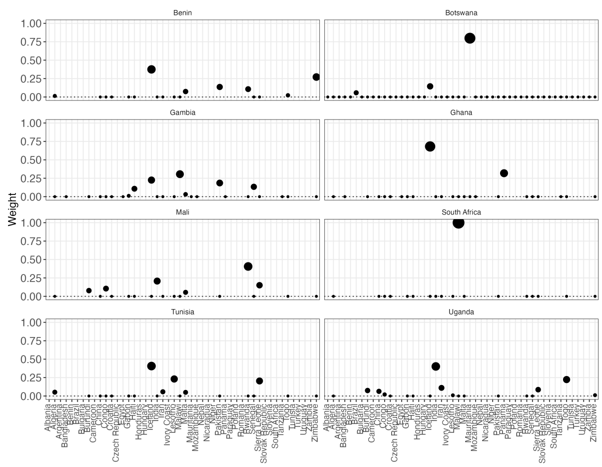
Figure S.19: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$

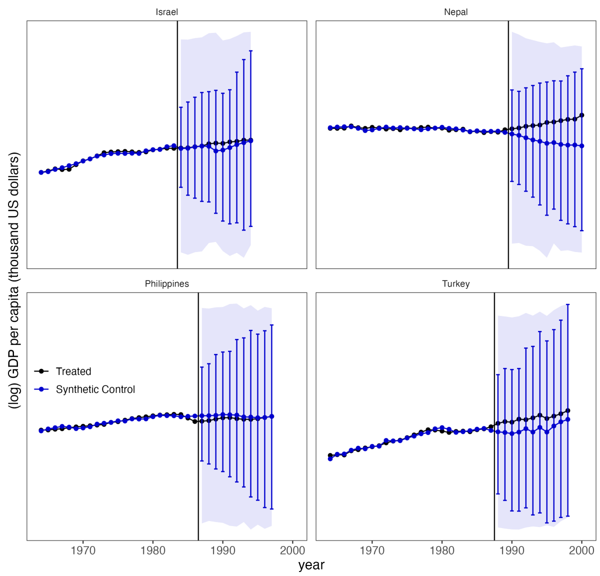


(c) $\widehat{w}_i$

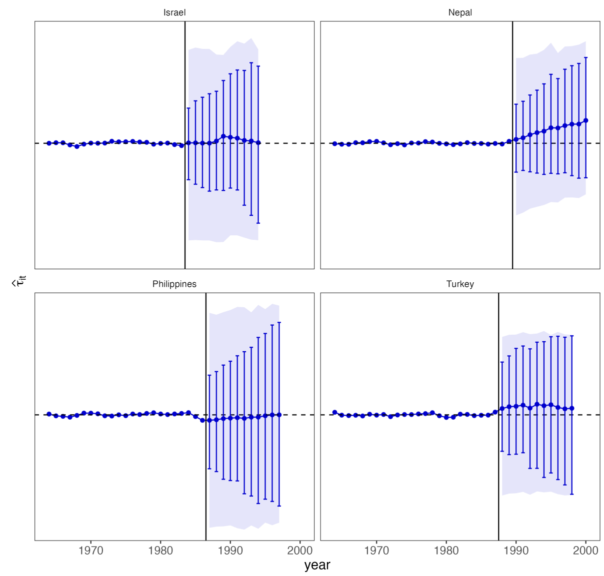
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.6.2 Asia

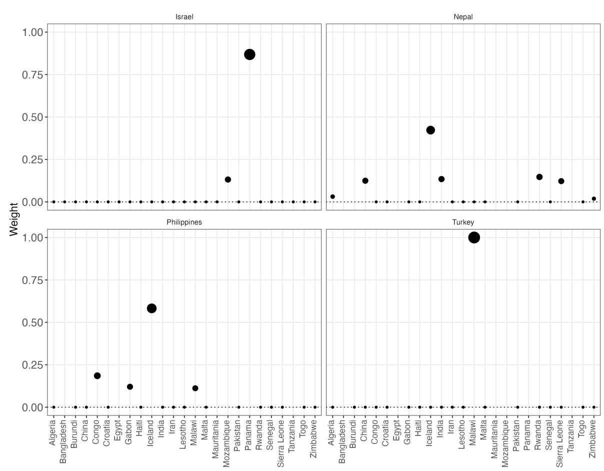
Figure S.20: *Individual Treatment Effects $\widehat{\tau}_{it}$.*



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



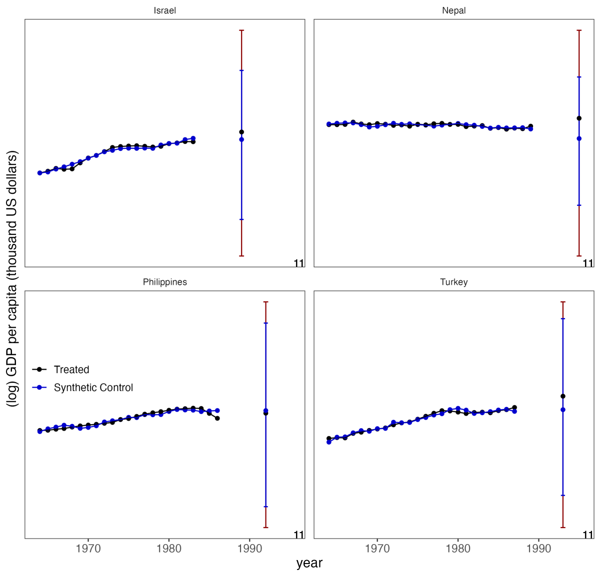
(b) $\widehat{\tau}_{it}$



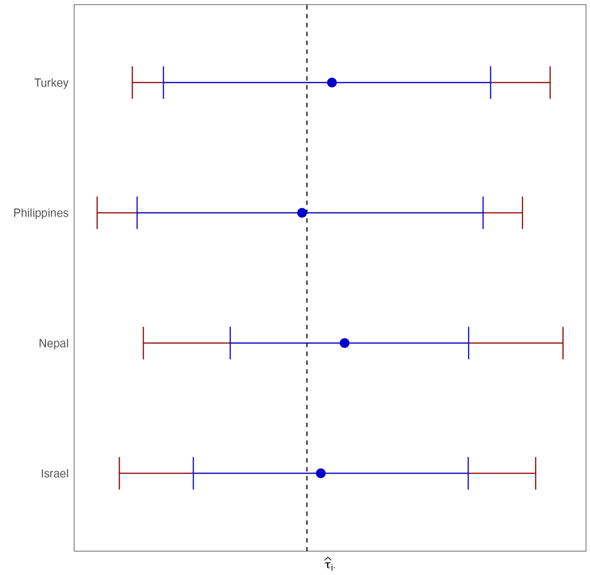
(c) $\widehat{w}_i$

Notes: 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.

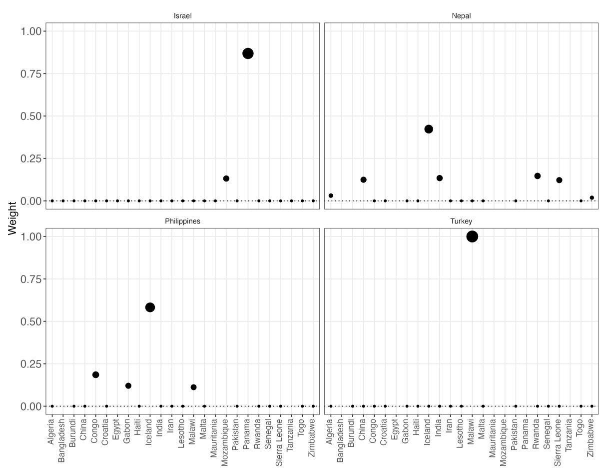
Figure S.21: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



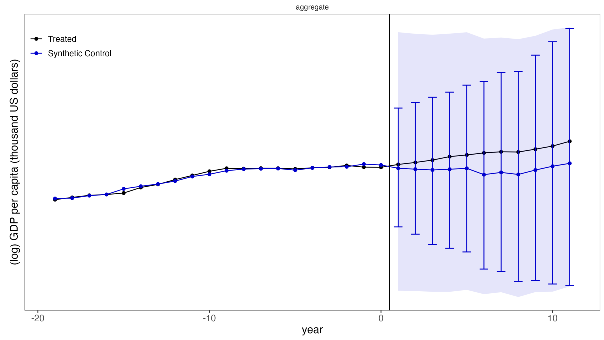
(b) $\widehat{\tau}_{i\cdot}$



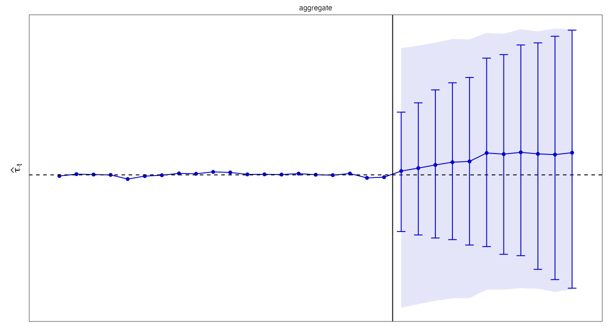
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

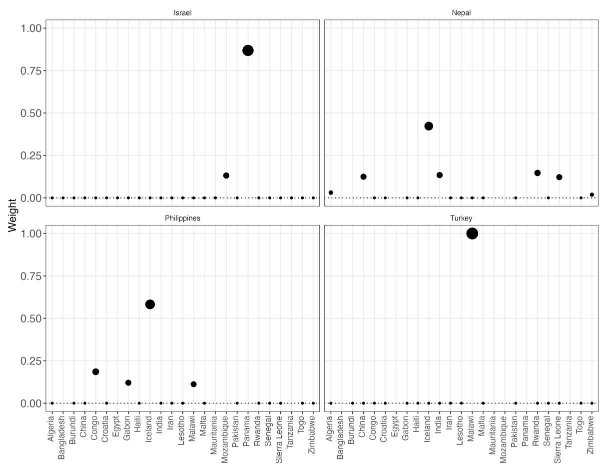
Figure S.22: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$

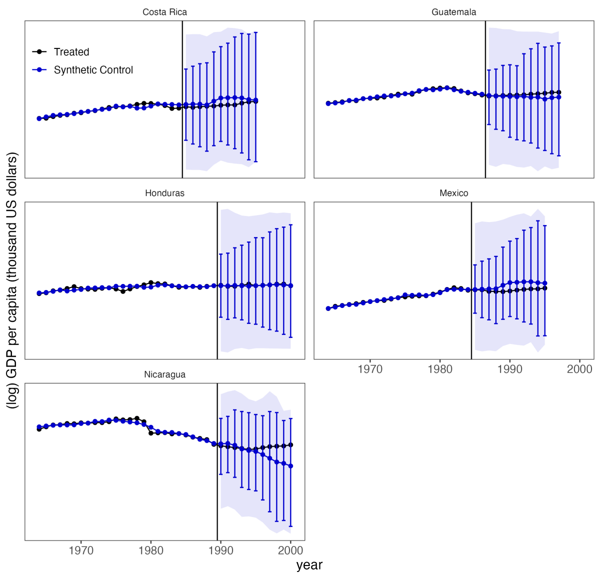


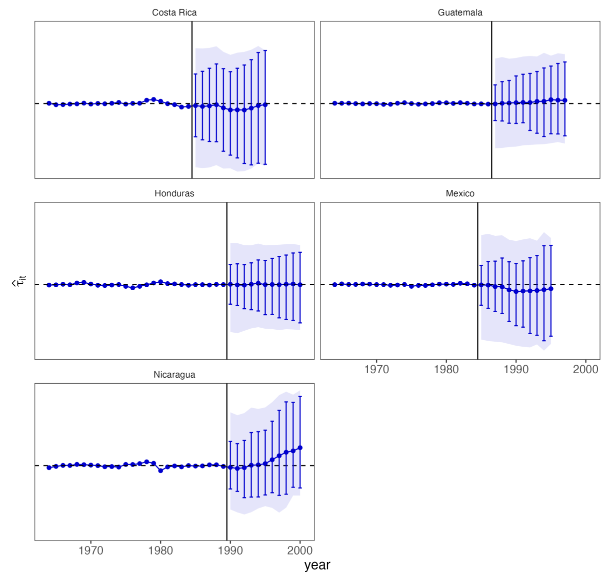
(c) $\widehat{w}_i$

Notes: 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 using sub-Gaussian bounds.

S.6.3 North America

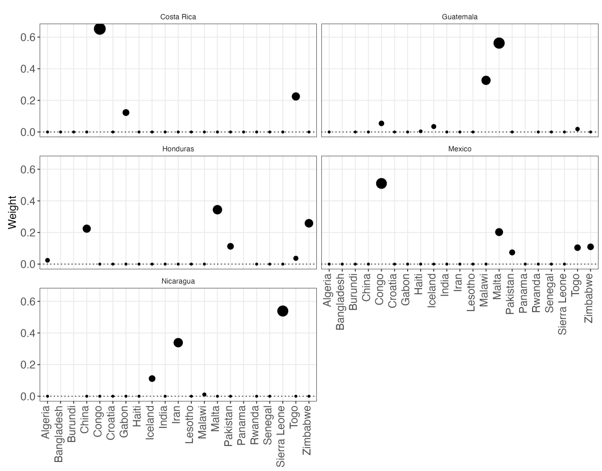
Figure S.23: *Individual Treatment Effects* $\widehat{\tau}_{it}$.





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

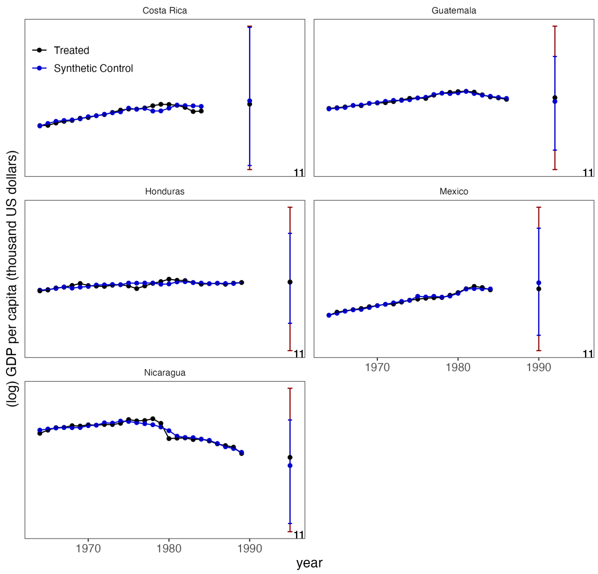
(b) $\widehat{\tau}_{it}$



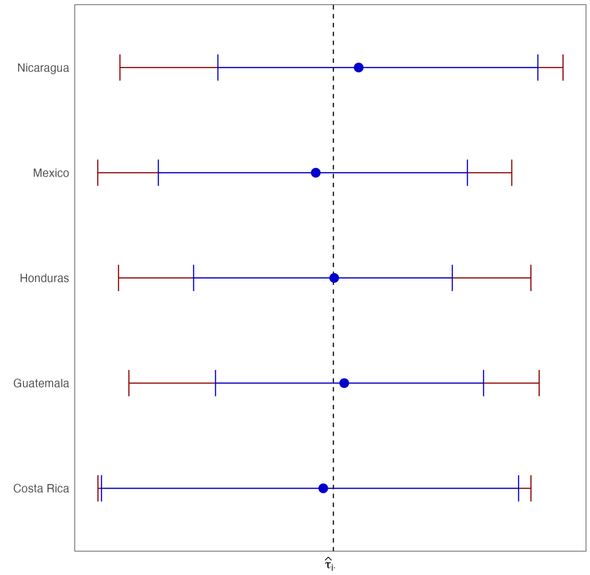
(c) $\widehat{w}_i$

Notes: 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.

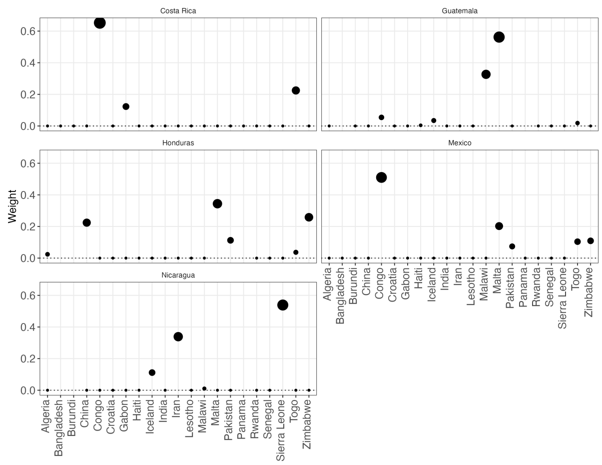
Figure S.24: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



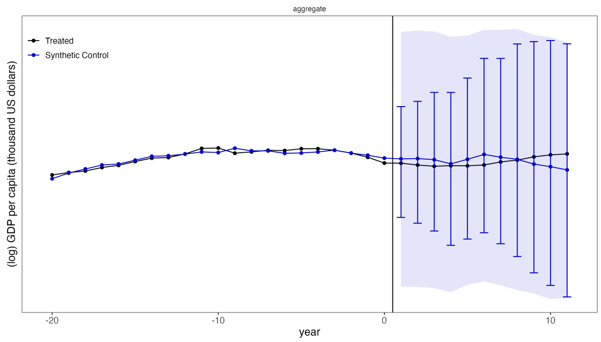
(b) $\widehat{\tau}_{i\cdot}$



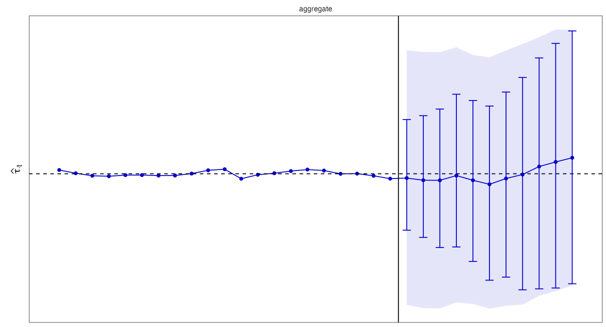
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

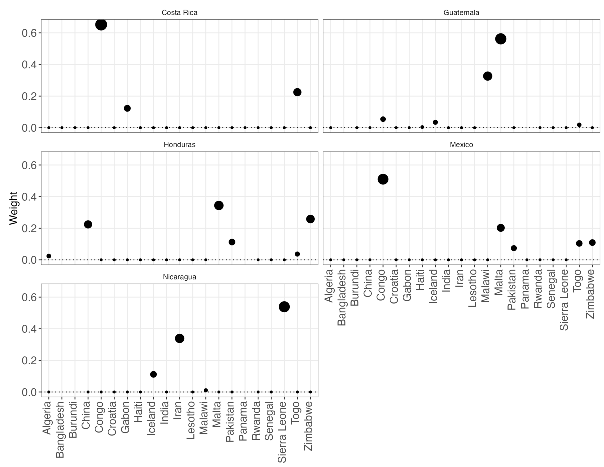
Figure S.25: *Average Treatment Effects on the Treated $\tau_{\cdot k}$.*



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



(b) $\widehat{\tau}_{\cdot k}$

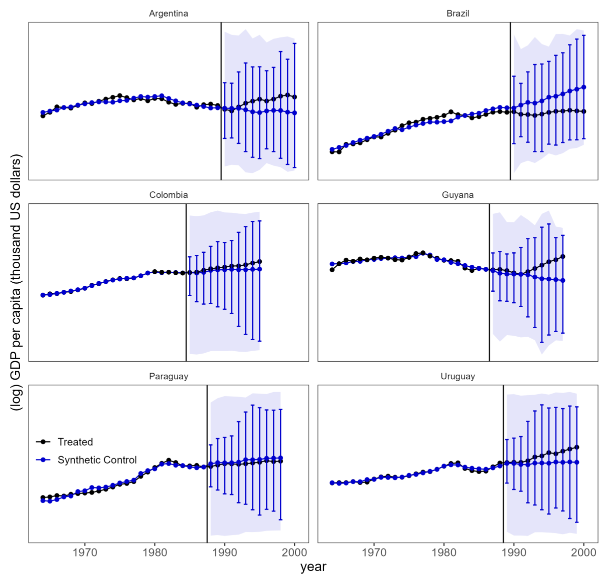


(c) $\widehat{w}_i$

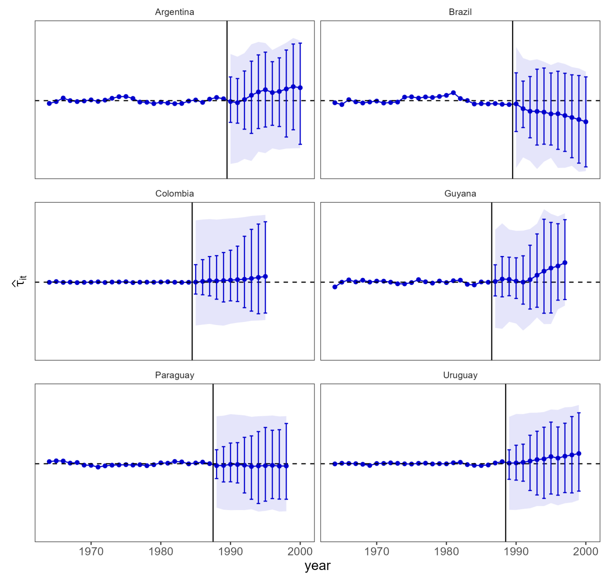
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.6.4 South America

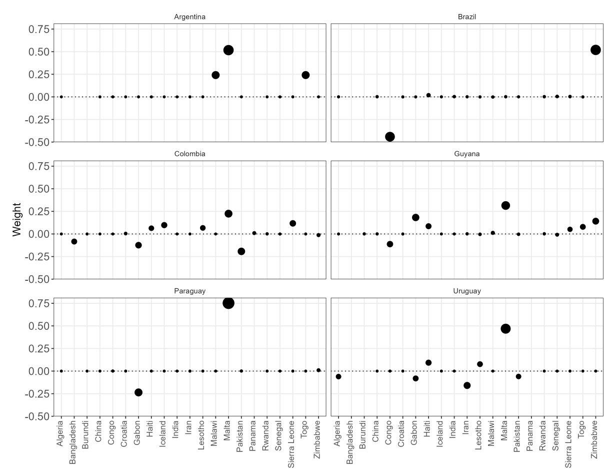
Figure S.26: *Individual Treatment Effects $\widehat{\tau}_{it}$.*



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



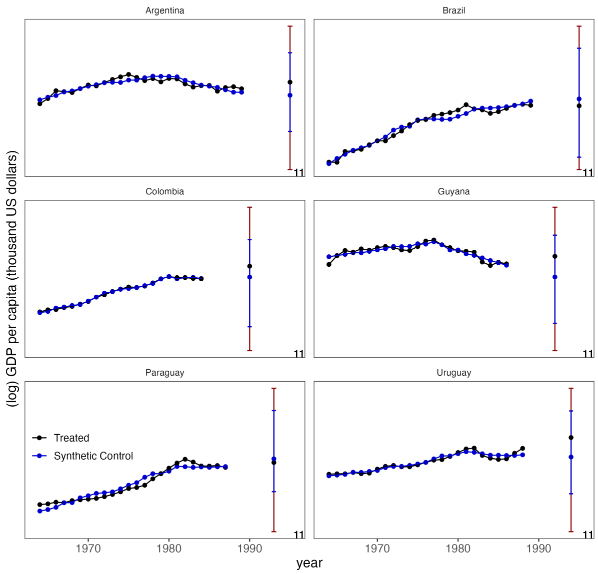
(b) $\widehat{\tau}_{it}$



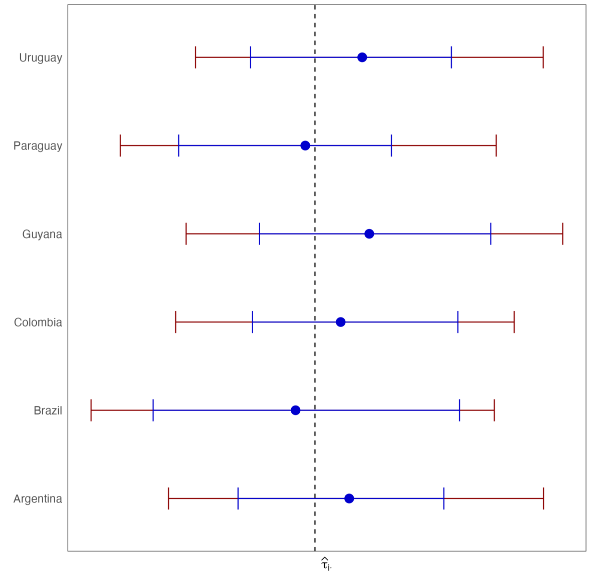
(c) $\widehat{w}_i$

Notes: 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.

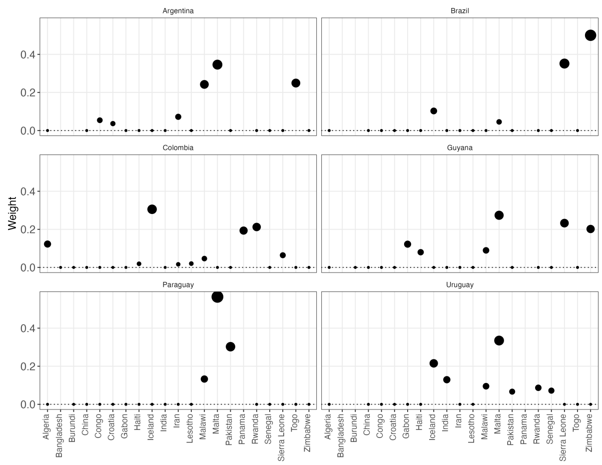
Figure S.27: *Average Post-Treatment Effects* $\widehat{\tau}_{i\cdot}$.



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



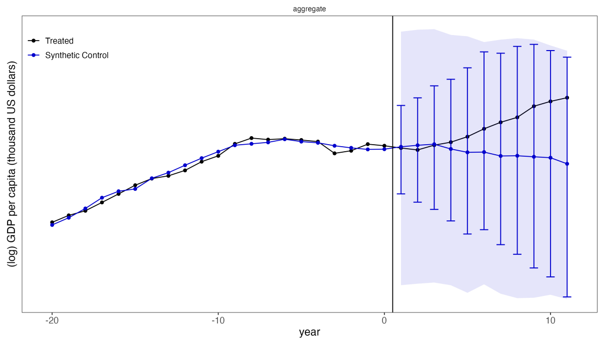
(b) $\widehat{\tau}_{i\cdot}$



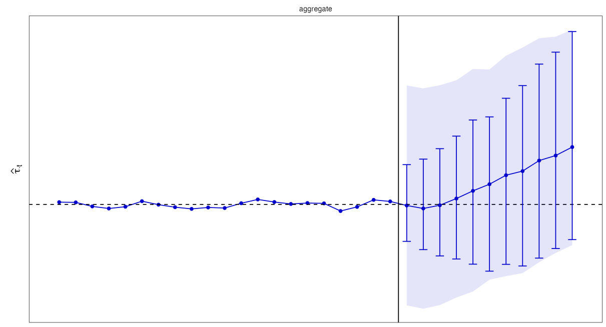
(c) $\widehat{w}_i$

Notes: Blue bars report 90% prediction intervals, whereas red bars report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty is quantified using sub-Gaussian bounds.

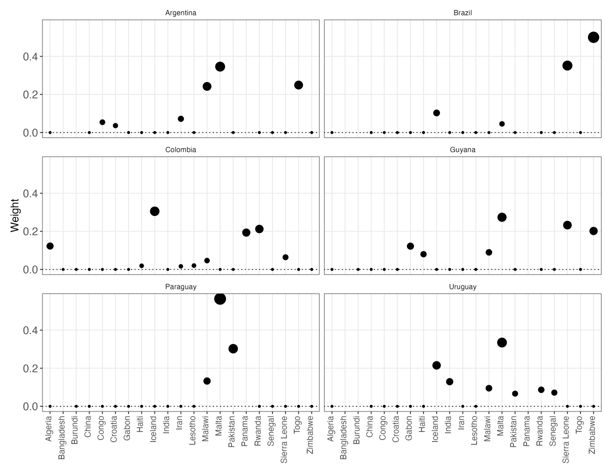
Figure S.28: *Average Treatment Effects on the Treated* $\tau_{\cdot k}$.



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



(b) $\widehat{\tau}_{\cdot k}$



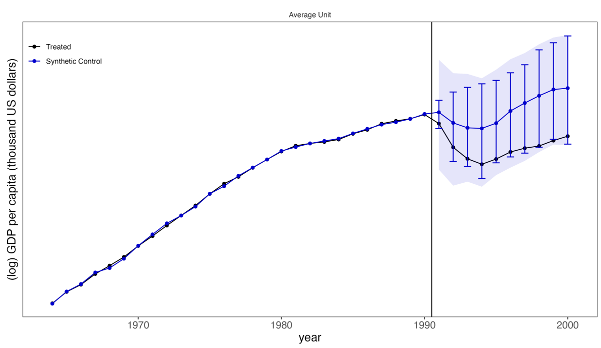
(c) $\widehat{w}_i$

Notes: 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 using sub-Gaussian bounds.

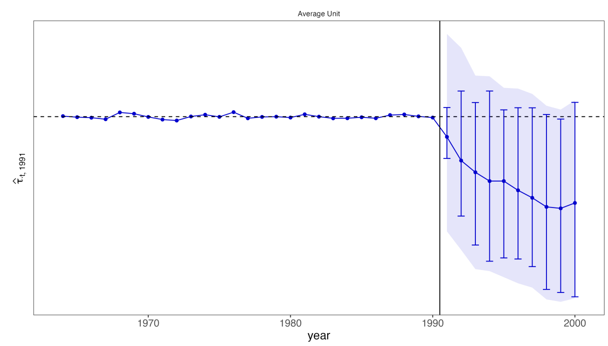
S.7 Average treatment effect on countries liberalized in 1991

In this section we present alternative predictions of the average treatment effect on those European countries that underwent a liberalization program in 1991. First, we replicate the main analysis presented in Section 6 using a simplex-type constraint, and then we conduct the same exercise using two features: the logarithm of GDP per capita and the percentage of complete secondary schooling attained in population.

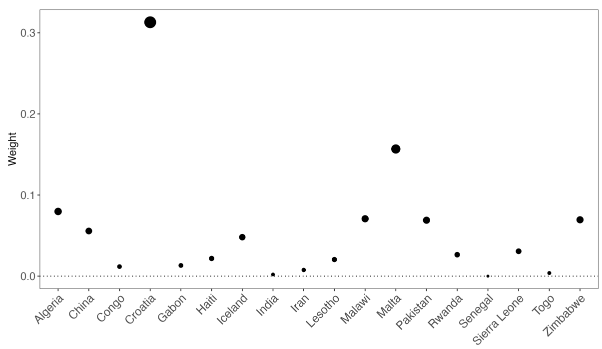
Figure S.29: *Average Treatment Effect on the Treated in 1991: Simplex constraint,* $M = 1$.



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



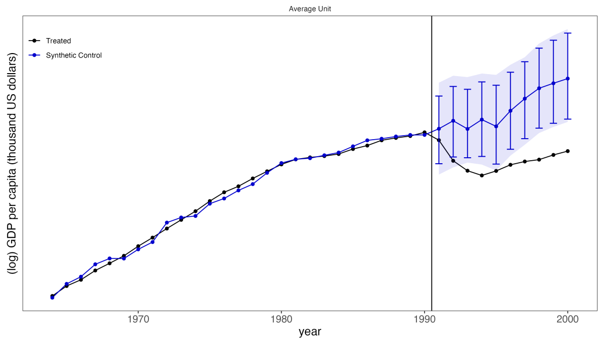
(b) $\widehat{\tau}_{\cdot t,1991}$



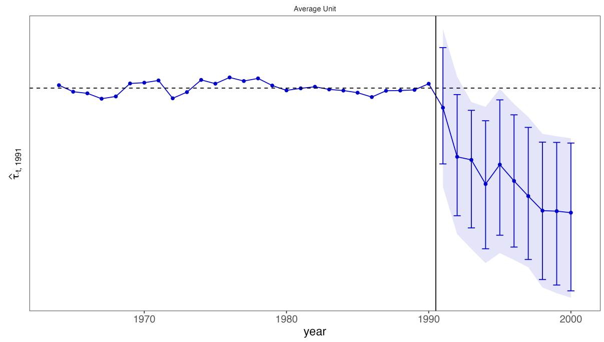
(c) $\widehat{w}_i$

Notes: 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 using sub-Gaussian bounds.

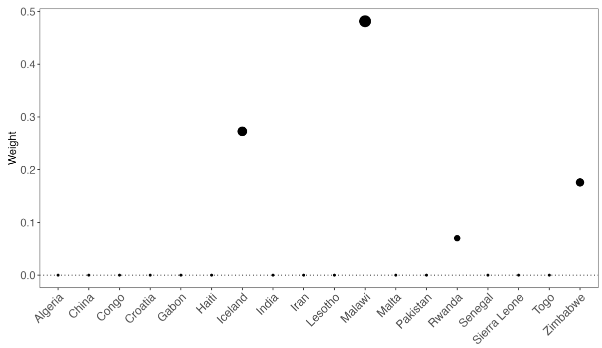
Figure S.30: *Average Treatment Effect on the Treated in 1991: L1-L2 constraint, $M = 1$.*



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



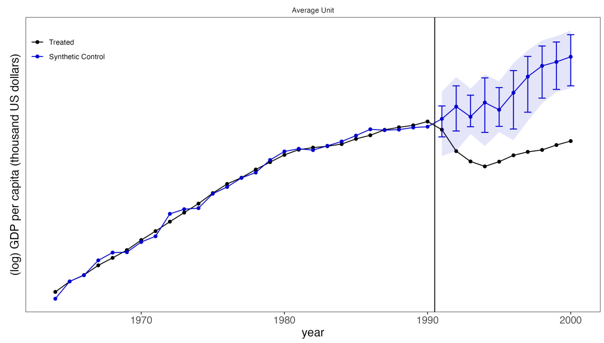
(b) $\widehat{\tau}_{\cdot t,1991}$



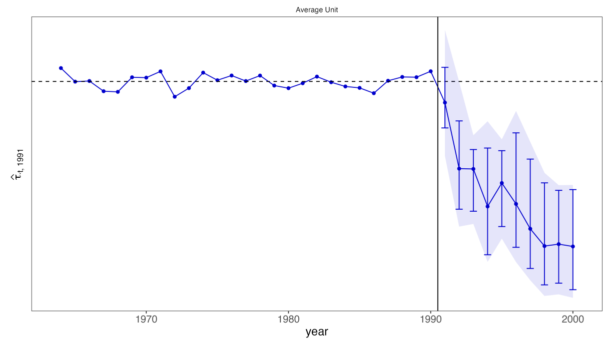
(c) $\widehat{w}_i$

Notes: 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 using sub-Gaussian bounds.

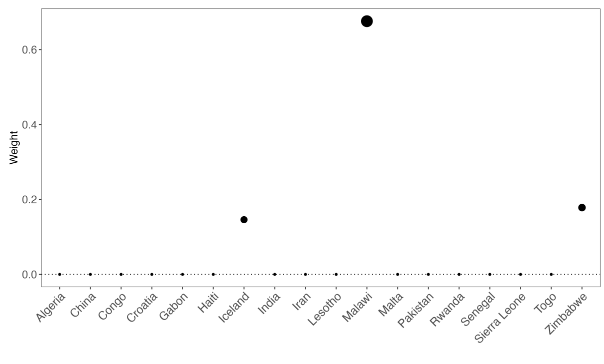
Figure S.31: *Average Treatment Effect on the Treated in 1991: Simplex constraint, $M = 2$.*



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



(b) $\widehat{\tau}_{\cdot t,1991}$



(c) $\widehat{w}_i$

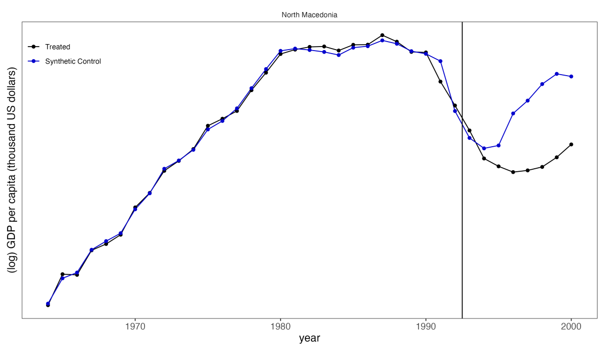
Notes: Blue bars report 90% prediction intervals, whereas blue shaded areas report 90% simultaneous prediction intervals. In-sample uncertainty is quantified using 200 simulations, whereas out-of-sample uncertainty using sub-Gaussian bounds.

S.8 Excluded Countries

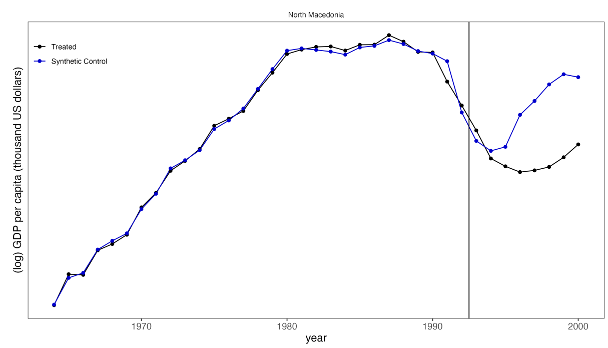
In this section we present point predictions of individual treatment effects for all those treated units that have been excluded from the main analysis. Three different reasons lead us to remove a country from our study: (i) missing additional features; (ii) poor pre-treatment fit; (iii) treatment date between 1993 and 2000.

Case I: Missing Additional Features. In the case of North Macedonia, we only have data on real GDP per capita, as we do not observe the percentage of complete secondary schooling attained in population. Figure S.32 shows the estimated synthetic control for North Macedonia when matching exclusively on real income.

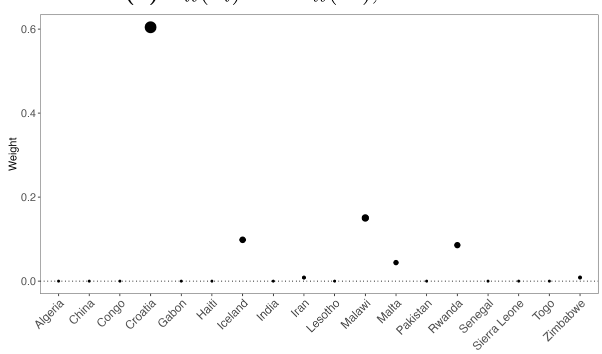
Figure S.32: *Individual Treatment Effect* $\widehat{\tau}_{it}$*: North Macedonia.*



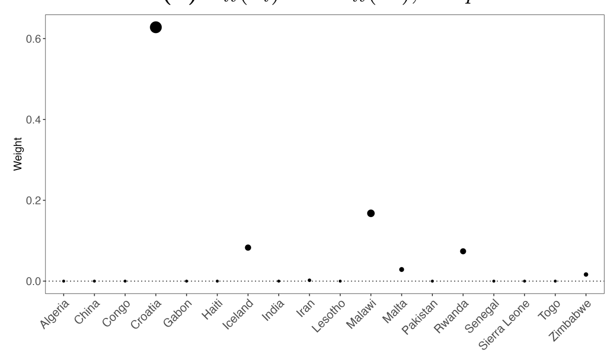
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, L1-L2*



(b) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, simplex*



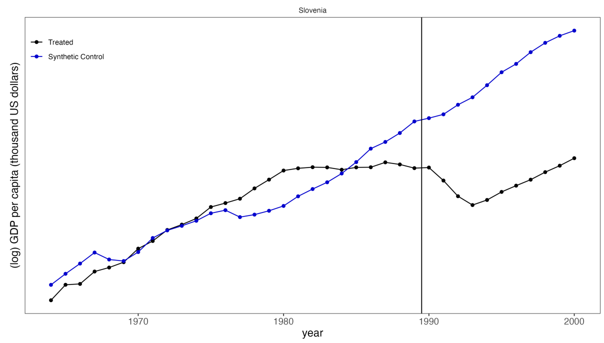
(c) $\widehat{w}_i$*, L1-L2*



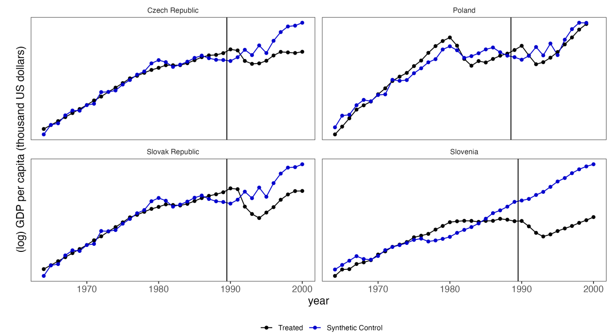
(d) $\widehat{w}_i$*, simplex*

Case II: Poor pre-treatment fit. We excluded Slovenia from the main analysis that uses the L1-L2 constraint. This decision is motivated by an evident discrepancy between the realized series for real income and the synthetic control (Figure S.33a). When we use a simplex-type constraint (Figure S.33b), the number of excluded countries rises to 4— Czech Republic, Poland, Slovak Republic, and Slovenia. This happens because the synthetic control obtained under the simplex constraint is a convex combinations of unexposed units, i.e., each weight is forced to be within the closed interval $[0, 1]$ (Figure S.33d). The L1-L2 constraint relies on the same constraint but on top of that adds a ridge-type constraint. In this case, it is evident that this leads to a better pre-treatment matching for at least three countries (Czech Republic, Poland, and Slovak Republic).

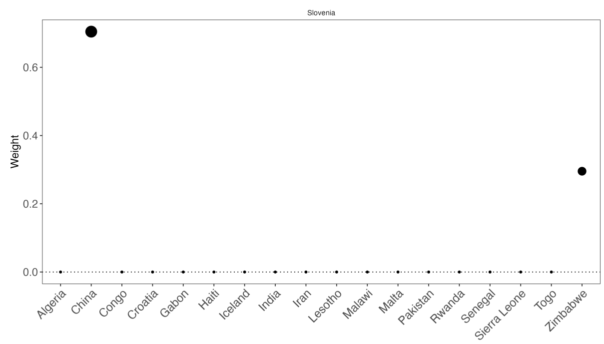
Figure S.33: *Individual Treatment Effect $\widehat{\tau}_{it}$: Excluded European countries.*



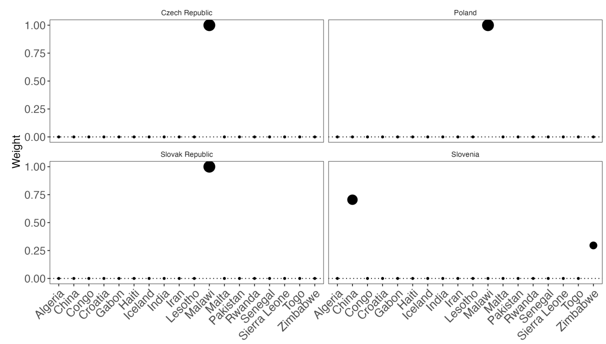
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, L1-L2*



(b) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, simplex*



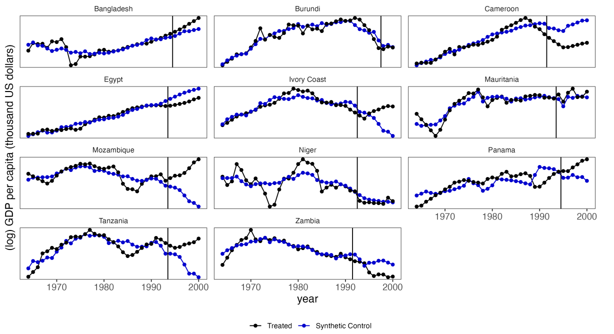
(c) $\widehat{w}_i$*, L1-L2*



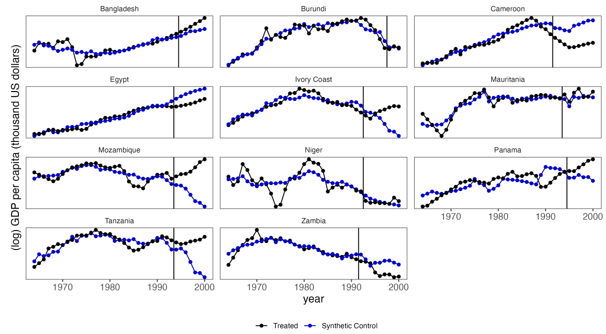
(d) $\widehat{w}_i$*, simplex*

<u>Case III: treatment date after 1992</u>. Since our main analysis requires at least 10 post-treatment periods, our dataset ranges until the year 2000, and we consider one year of possible anticipation effects, units treated in 1993 or afterwards are not included. Indeed, such units are observed for less than 10 years in the treated status. Figure S.34 reports the point predictions for all these countries.

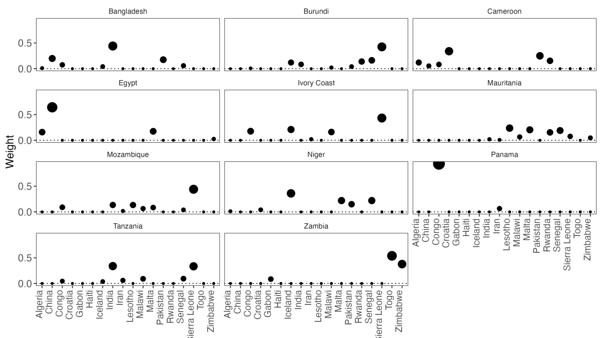
Figure S.34: *Individual Treatment Effect $\widehat{\tau}_{it}$: Treated after 1992.*



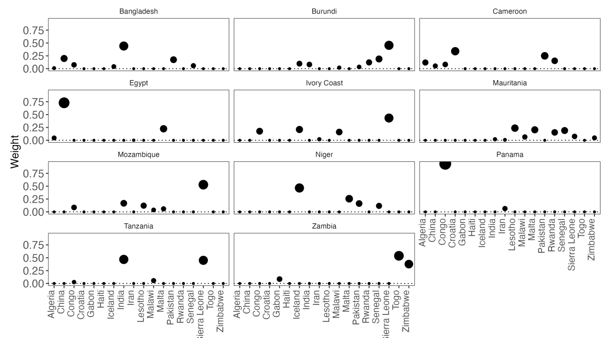
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, L1-L2*



(b) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$*, simplex*



(c) $\widehat{w}_i$*, L1-L2*



(d) $\widehat{w}_i$*, simplex*

Supplement References

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.

Cattaneo, M. D., Feng, Y., Palomba, F., and Titiunik, R. (2023), "`scpi`: Uncertainty Quantification for Synthetic Control Methods," arXiv:2202.05984.

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

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.