Uncertainty Quantification in Synthetic Controls with Staggered Treatment Adoption*

Matias D. Cattaneo†   Yingjie Feng‡   Filippo Palomba§   Rocío Titiunik¶

February 1, 2025

Abstract

We propose principled prediction intervals to quantify the uncertainty of a large class of synthetic control predictions (or estimators) in settings with staggered treatment adoption, offering precise non-asymptotic coverage probability guarantees. From a methodological perspective, we provide a detailed discussion of different causal quantities to be predicted, which we call *causal predictands*, allowing for multiple treated units with treatment adoption at possibly different points in time. From a theoretical perspective, our uncertainty quantification methods improve on prior literature by (i) covering a large class of causal predictands in staggered adoption settings, (ii) allowing for synthetic control methods with possibly nonlinear constraints, (iii) proposing scalable robust conic optimization methods and principled data-driven tuning parameter selection, and (iv) offering valid uniform inference across post-treatment periods. We illustrate our methodology with an empirical application studying the effects of economic liberalization on real GDP per capita for Sub-Saharan African countries. Companion software packages are provided in `Python`, `R`, and `Stata`.

Keywords: causal inference, synthetic controls, staggered treatment adoption, prediction intervals, non-asymptotic inference.

†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

1   Introduction

The synthetic control (SC) method was introduced by Abadie and Gardeazabal (2003), and since then many extensions and generalizations have been proposed (see Abadie, 2021, and references therein). The method is now part of the standard toolkit for program evaluation and treatment effect analysis (Abadie and Cattaneo, 2018), offering a complement to traditional difference-in-differences, event studies, and other panel data approaches for causal inference with longitudinal aggregate data and few treated units. Most of the synthetic control literature concentrates on identification, as well as on prediction or point estimation of treatment effects, under different causal inference frameworks and algorithmic implementations. In contrast, principled uncertainty quantification of synthetic control predictions or estimators in general settings remains mostly unexplored, despite its importance for empirical work.

Following Cattaneo, Feng and Titiunik (2021), who proposed non-asymptotic prediction intervals for synthetic controls with a single-treated-unit, we employ a causal inference framework where potential outcomes are assumed to be random and develop novel prediction intervals to quantify the uncertainty of a large class of synthetic control predictions (or estimators) in settings with staggered treatment adoption and multiple treated units. Our contributions include establishing precise non-asymptotic coverage probability guarantees for our novel prediction intervals, introducing scalable robust optimization implementations for possibly nonlinear constraints in the synthetic controls construction, developing principled tuning parameter selection, and proposing valid joint inference methods across time. Inferential procedures with non-asymptotic probability guarantees are valuable because synthetic control applications often have small sample sizes, impeding the applicability of asymptotic approximations. Conceptually, the proposed prediction intervals capture two sources of uncertainty: one coming from the construction of the synthetic control weights with pre-treatment data, and the other generated by the irreducible sampling variability introduced by the post-treatment outcomes. Our prediction intervals also take into account potential misspecification errors explicitly and enjoy other robustness properties due to their non-asymptotic, generic construction. For example, our methods allow for nonlinear regularization in the synthetic controls construction, which accommodates L2 penalization, L1/L2-combined penalization, and other variants thereof. These nonlinear penalization schemes are better suited for application where the

covariates exhibit collinearity/codependency, as opposed to L1 penalization schemes, which are better suited for covariate shrinkage/selection.

To motivate our methodological work, we begin by introducing an empirical application in Section 2, and we use this application throughout to illustrate our contributions. Following Billmeier and Nannicini (2013), we investigate the effect of economic liberalization in the 1990s on real GDP per capita for Sub-Saharan African countries. This application includes multiple adoption times and multiple treated units. Furthermore, we consider the special case of outcome-only synthetic controls to improve the exposition and carefully account for the fact that the outcome variable in this application is non-stationary and co-integrated. We also discuss more complex empirical analyses in the supplemental appendix.

Our first contribution is methodological in nature due to the complexity added by the staggered treatment adoption setup, which allows for (but does not require) the existence of multiple treated units changing from control to treatment status at possibly different points in time. In Section 3, we introduce a basic causal inference framework that is motivated by our empirical application and specifically tailored to synthetic control methods with staggered treatment adoption. Using this framework, we define different causal quantities to be predicted in the context of synthetic controls, which we refer to as *causal predictands*, and explain how prediction methods are implemented.

Section 4 then discusses uncertainty quantification in the context of our empirical application and basic causal inference framework. Motivated by our empirical application and recent advances in the synthetic controls literature (see Abadie, 2021, and references therein), our analysis focuses exclusively on incorporating in-sample and out-of-sample uncertainty quantification for outcome-only synthetic control methods with non-stationary data and non-linear constraints. In this setting, we present novel prediction intervals with precise non-asymptotic guarantees for synthetic controls with staggered adoption, under easy-to-interpret sufficient conditions. We also discuss scalable, robust conic optimization implementations of our methods (Boyd and Vandenberghe, 2004), data-driven selection of tuning parameters, and valid joint inference across time.

The main empirical results are presented in Section 5. Our findings indicate that the economic liberalization in the 1990s did not have a positive economic impact on most emerging Sub-Saharan African countries. This finding is in line with prior empirical results (Billmeier and Nannicini, 2013). The Supplemental Appendix provides additional empirical evidence supporting our main

findings, including a re-analysis using alternative synthetic control specifications, placebo treatment dates, donor pool constructions, and other related implementations under different assumptions.

While the basic framework developed in Sections 3 and 4 is sufficient for our empirical application, Section 6 offers a substantive generalization of our methods that allows, among others, general misspecification error, multiple covariate features, and cross-equation re-weighting when constructing the synthetic control weights with pre-treatment data. This general framework also provides foundational results for our empirical work, including a theoretical justification for the methods presented in Sections 3 and 4.

1.1 Related Literature

We contribute to developing prediction/estimation and inference methods for synthetic control settings with staggered treatment adoption. Putting aside generic linear factor model or matrix completion methods, Ben-Michael, Feller and Rothstein (2022), Powell (2022), and Shaikh and Toulis (2021) appear to be the only prior papers that have studied staggered treatment adoption for synthetic controls explicitly. The first paper focuses on prediction/estimation in settings where the pre-treatment fit is poor and develops penalization methods to improve the performance of the canonical synthetic control method. Ben-Michael, Feller and Rothstein (2022) also suggest employing a bootstrap method for assessing uncertainty, but no formalization is provided guaranteeing its (asymptotic) validity. Powell (2022) employs a standard parametric panel data model to discuss large-sample estimation and inference methods for a single common treatment effect. Shaikh and Toulis (2021) focus on uncertainty quantification employing a parametric duration model and propose a permutation-based inferential method under a symmetry assumption. Our paper complements these prior contributions by (i) developing a general causal inference framework for synthetic control methods with staggered treatment adoption, and (ii) offering nonparametric inference methods with demonstrable non-asymptotic coverage guarantees and allowing for misspecification in the construction of the synthetic control weights. We also propose novel scalable robust conic optimization implementations, principled tuning parameter selection methods, and a valid joint inference procedure across multiple time periods.

Our quantification of uncertainty via non-asymptotic prediction intervals follows Cattaneo, Feng and Titiunik (2021). These inference methods are motivated by Vovk (2012) and are most closely

related to prior work by Chernozhukov, Wüthrich and Zhu (2021a) and Chernozhukov, Wüthrich and Zhu (2021b) on conformal prediction intervals. Relative to this literature, our contributions include: (i) allowing for a large class of causal predictands in staggered adoption settings (prior work covered only the canonical single treated unit case); (ii) covering a large class of synthetic control predictions with possibly nonlinear constraints (prior work allowed only linear constraints); (iii) developing scalable robust optimization implementations and proposing principled data-driven tuning parameter selection (prior work did not provide guidance on these issues); and (iv) introducing valid uniform inference methods across post-treatment periods (absent in prior work).

There are a few other, conceptually different, recent proposals to quantify uncertainty and conduct inference in the synthetic controls literature. For example, Li (2020) study correctly specified linear factor models, Masini and Medeiros (2021) study high-dimensional penalization methods, Agarwal, Shah, Shen and Song (2021) investigate matrix completion methods, and Shi, Miao, Hu and Tchetgen (2023) develop inference methods using a proximal causal inference framework. All these methods rely on asymptotic approximations, in most cases employing standard Gaussian critical values that assume away misspecification errors and other small sample issues. Our work complements these contributions by providing prediction intervals with non-asymptotic coverage guarantees (Wainwright, 2019). There is also a literature in econometrics on event studies that is loosely related to synthetic controls with staggered adoption: see, for example, Freyaldenhoven et al. (2019, Forthcoming), Miller (2023), and references therein. Finally, all the inferential methods mentioned so far contrast with the original method proposed by Abadie, Diamond and Hainmueller (2010), which relies on design-based permutation of treatment assignment assuming that the potential outcomes are non-random.

1.2 Paper Organization

Section 2 introduces our running empirical application. Section 3 presents the basic causal inference framework for outcome-only synthetic controls with staggered treatment adoption, and Section 4 discusses non-asymptotic uncertainty quantification in that context. Section 5 presents our empirical results. Section 6 gives a self-contained discussion of our most general framework and theoretical results. The Supplemental Appendix includes all proofs, additional empirical results, and other technical details omitted to improve the exposition.

We provide general-purpose software implementing our results in `Python`, `R`, and `Stata`, including detailed documentation and additional replication materials. This software is discussed in our companion article Cattaneo, Feng, Palomba and Titiunik (2025), where we addressed several implementation issues related to numerical optimization and tuning parameter selection. To complement the illustration, in Section S.7 of the Supplemental Appendix, we provide more details on how to prepare the data to analyze staggered treatment adoption using synthetic control methods using our companion software. In addition, in the Supplemental Appendix, Section S.2.3 demonstrates how to obtain tighter bounds for uncertainty quantification when the weighting matrix has a block diagonal structure, and Section S.6 shows how to reformulate the synthetic control problem as a scalable conic optimization problem to improve computational efficiency. For ease of understanding, Table S.1 in the Supplemental Appendix summarizes the notation used throughout the paper.

2 The Effect of Liberalization on GDP for Sub-Saharan African Countries

During the second half of the twentieth century, many countries around the world launched programs of (external) economic liberalization, booming from 22% in 1960 to 73% in the early 2000s (Wacziarg and Welch, 2008). In the last thirty years, scholars have investigated the social and economic consequences of such liberalization programs, often reaching conflicting conclusions (see, e.g., Levine and Renelt, 1992; Sachs, Warner, Åslund and Fischer, 1995; DeJong and Ripoll, 2006).

The impact of liberalization policies on economic welfare has been traditionally investigated with cross-country analyses (e.g. Sachs, Warner, Åslund and Fischer, 1995) and individual case studies (e.g. Bhagwati and Srinivasan, 2001). More recently, scholars have turned to synthetic control methods in the hope of employing a causal inference methodology that allows for the presence of time-varying unobservable confounders. Employing the synthetic control framework originally developed in Abadie and Gardeazabal (2003), Billmeier and Nannicini (2013) analyzed the effects of liberalization in four continents: Africa, Asia, North America, and South America. They used a pre-existing dataset of economic variables (previously used in Giavazzi and Tabellini, 2005) which includes 180 countries, covers the period 1963–2000 and contains an indicator for economic liberalization originally defined in Sachs, Warner, Åslund and Fischer (1995) and updated

in Wacziarg and Welch (2008) (hereafter, the Sachs-Warner indicator). More details on the data and the definition of economic liberalization can be found in Section S.7 of the Supplemental Appendix.

Billmeier and Nannicini (2013) studied the effect of economic liberalization—as measured by the Sachs-Warner indicator—on GDP per capita, to understand whether the adoption of liberalization programs affected economic growth. Building on their study of Sub-Saharan Africa, we illustrate how our formal inference framework can be an essential tool to aggregate results and draw general conclusions from synthetic control analyses under staggered treatment adoption by multiple treated units. The ambitious scope of the analysis in Billmeier and Nannicini (2013) resulted in a large number of synthetic control results. In Sub-Saharan Africa alone, they studied sixteen episodes of liberalization that occurred at ten different periods. The results exhibited considerable heterogeneity, which the authors summarized by grouping the effects into four categories according to two dimensions: whether the effect was positive and whether there was evidence that the effect was robust rather than "coincidental" (a notion of statistical significance). The four categories were (a) countries with a positive and strongly robust effect, (b) countries with a positive and somewhat robust effect, (c) countries with a positive but non-robust effect, and (d) countries with a null effect. They concluded that "Botswana is the only country clearly in the first group, offering a truly convincing success story" (p. 995).

Billmeier and Nannicini (2013) adjudicated the robustness or statistical significance of each of the sixteen effects with the Fisherian-type placebo test developed by Abadie et al. (2010), which was the main inference tool available at the time. Our new inference framework allows us to reconsider the evidence for Sub-Saharan Africa by providing formal tools to perform joint inference and draw a general conclusion about the effects of liberalization by considering the evidence altogether rather than piecemeal.

The data starts in the 1960s, which marks the beginning of Africa's post-colonial era. Since independence from their colonial rulers until the late 1980s, many countries in Sub-Saharan Africa adopted neo-patrimonial political regimes based on the concentration of power in a single individual and the resulting cultivation of personalistic politics, widespread clientelistic networks, and the use of the resources of the state to achieve political legitimacy (Bratton and Van de Walle, 1997).

The 1980s put high stress on these regimes, as economic conditions deteriorated. Negative economic growth, increased debt services as a proportion of exports, currency overvaluation, low

6

commodity prices, low investment, and limited access to credit created an adverse economic environment that these non-democratic regimes were ill-equipped to handle. International finance organizations such as the World Bank and the International Monetary Fund incentivized programs of structural economic adjustment, with disputed success (Callaghy and Ravenhill, 1994).

The adverse economic conditions led to popular protests in the early 1990s, which marked the beginning of a wave of transitions to democracy throughout the continent. Starting in the early 1990s, many countries started a process of political liberalization, partly influenced by the fall of the Berlin Wall in November of 1989 and the subsequent collapse of the communist regimes of the Soviet Union (Bratton and Van de Walle, 1997).

As illustrated in Figure 1, most countries in Sub-Saharan Africa adopted the economic liberalization treatment between the late 1980s and the early or mid-1990s which, as just explained, was a period of major political and economic transition in Africa and beyond. This means that, for many countries in our sample, the economic liberalization treatment occurred nearly simultaneously with other major political and economic shocks that influenced the World's geo-political equilibrium. Given these potential confounders, it is crucial to exploit the staggered adoption of the treatment to make inferences. While the worldwide political changes between the late 1980s and early 1990s affected all countries at the same time, the concrete measures of economic liberalization captured by the Sachs index were introduced in different years for different countries. Our proposed synthetic control methods allow us to define causal predictands and perform joint inferences that leverage the staggered introduction of the treatment, while controlling for worldwide shocks affecting all units.

3   Basic Framework

We set up a basic synthetic control framework that matches the empirical application in Section 2 and suffices to communicate key ideas of our proposed method. A more general and technically more involved framework is deferred to Section 6.

Suppose that we observe N countries ("units") for T years ("periods"). Countries are indexed by $i = 1, \ldots, N$, and years are indexed by $t = 1, \ldots, T$. For each country i in year t, we can observe the GDP per capita Y_{it} ("outcome") and a time T_i that indicates when country i adopted economic

Figure 1: *Staggered Treatment Adoption in Africa.*
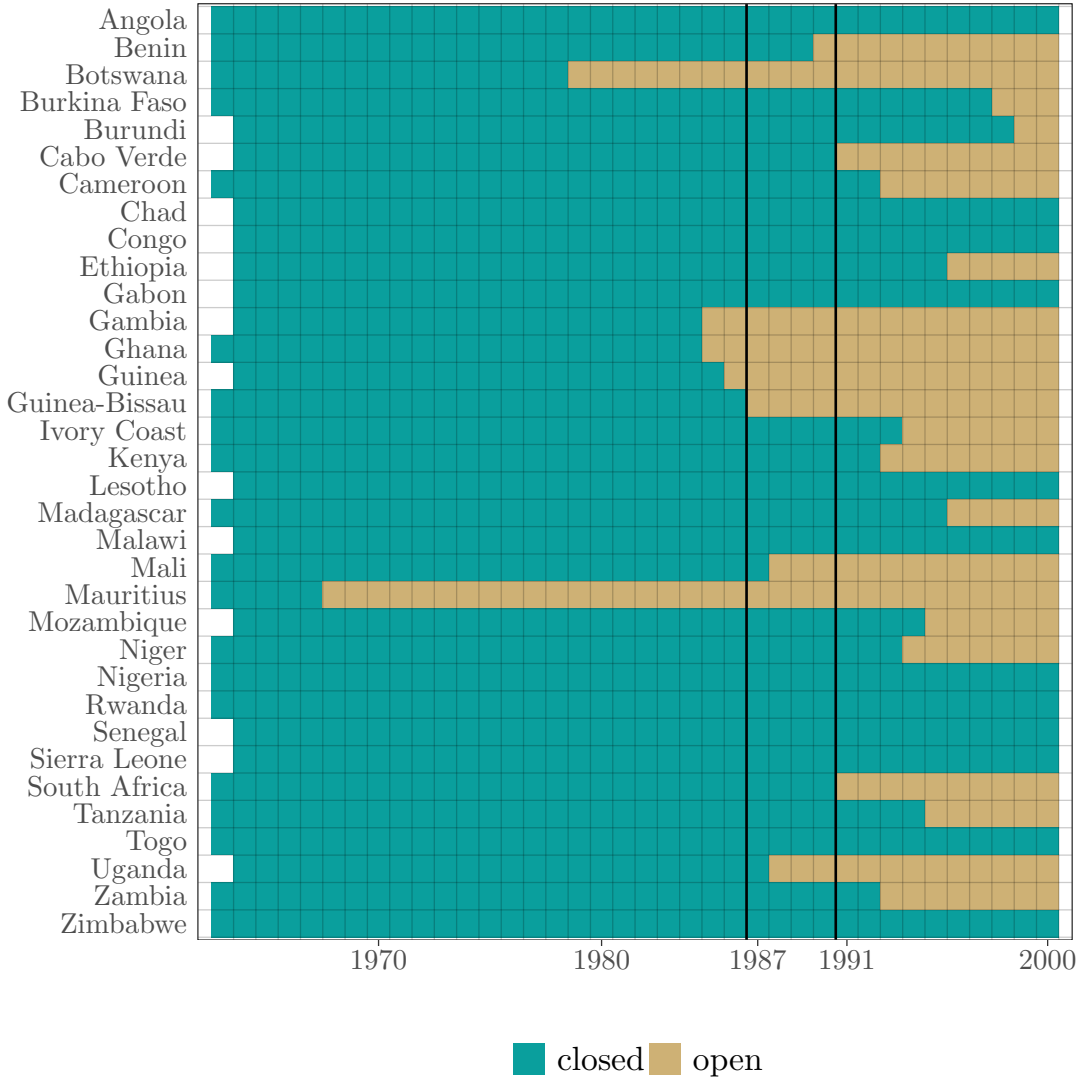
Notes: A country i in year t is defined to be "closed" or "open" using to the Sachs index. Thick vertical lines–corresponding to 1987 and 1991–delimit the three main liberalization waves of Billmeier and Nannicini (2013).

liberalization ("treatment"). Assume that a country remains economically liberalized whenever $t \geq T_i$. (In our empirical application this is always verified.) Without loss of generality, countries are ordered in the adoption times so that $1 \leq T_1 \leq T_2 \leq \cdots \leq T_N \leq \infty$, with $T_i = \infty$ denoting that country i remains untreated throughout the observation period. Let $\mathcal{N} = \{i : T_i = \infty\}$ be the group of "never-liberalized" countries and $\mathcal{E} = \{i : T_i < \infty\}$ the group of "ever-liberalized" countries. Define $J_0 = |\mathcal{N}|$ and $J_1 = |\mathcal{E}|$, where we use $|\mathcal{A}|$ to denote the number of elements in $\mathcal{A}$ for any finite set $\mathcal{A}$.

We are interested in the effect of economic liberalization on a country's GDP per capita. Let $Y_{it}(s)$ denote the potential GDP per capita of country i in year t that would be observed had country i adopted the economic liberalization treatment in year s, for $s = 1, \ldots, T, \infty$, and assume $Y_{it}(s) = Y_{it}(\infty)$ whenever $t < s$. Implicitly, these simplifications impose two standard assumptions: no spillovers (the potential outcomes of country i depend only on i's adoption time) and no anticipation (a country's potential outcomes prior to the treatment are equal to the outcomes it would have had if it had never been treated). The observed GDP per capita Y_{it} can be written as

$$Y_{it} = Y_{it}(\infty)\mathbb{1}(t < T_i) + Y_{it}(T_i)\mathbb{1}(t \geq T_i).$$

That is, whenever the economic liberalization has not been adopted, we always observe the potential GDP per capita under the "never-treated" status; otherwise the potential GDP per capita corresponding to the adoption time T_i is observed.

Our primary goal is to make a prediction of the missing counterfactual GDP per capita $Y_{it}(\infty)$ for an ever-liberalized country in the post-treatment period; we then use this unit-level predictand as the basis of a variety of other aggregate predictands.

For simplicity, we take all never-treated units in $\mathcal{N}$ as "donors" for SC prediction, and let all years prior to the economic liberalization, $\{t : 1 \leq t \leq T_i - 1\}$, be the training period used to obtain the SC weights for each ever-liberalized country $i \in \mathcal{E}$. The idea of SC is to find a vector of weights on the donor countries such that the weighted average of pre-treatment GDP per capita of donor countries matches that of the treated country as closely as possible, subject to some regularization constraints. Let $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the usual L1 and L2 vector norms respectively, and $(v_j : j \in \mathcal{A})$ denote a vector consisting of all v_j's with $j \in \mathcal{A}$ for a set $\mathcal{A}$. In our empirical analysis below, the SC weights are obtained via an L1-L2 constrained regression:

$$\widehat{\mathbf{w}}^{[i]} = \underset{\mathbf{w} \in \mathcal{W}^{[i]}}{\arg\min} \sum_{t=1}^{T_i - 1} \left(Y_{it} - \mathbf{Y}_{\mathcal{N}t}'\mathbf{w}\right)^2 \quad \text{for each } i \in \mathcal{E}, \tag{3.1}$$

where $\mathbf{Y}_{\mathcal{N}t} = (Y_{jt} : j \in \mathcal{N})$ is the vector of GDP per capita for donor countries at time t, and $\mathcal{W}^{[i]} = \{\mathbf{w} \in \mathbb{R}_+^{J_0} : \|\mathbf{w}\|_1 = 1, \|\mathbf{w}\|_2 \leq Q^{[i]}\}$ for some constant $Q^{[i]} > 0$. The choice $Q^{[i]}$ is discussed in Section S.7.2 of the Supplemental Appendix. The prediction of country i's counterfactual GDP

per capita k periods after the economic liberalization is given by

$$\widehat{Y}_{i(T_i+k)}(\infty) = \mathbf{Y}'_{\mathcal{N}(T_i+k)}\widehat{\mathbf{w}}^{[i]}, \quad \text{for each } i \in \mathcal{E} \text{ and } 0 \le k \le T - T_i.$$

In what follows, we define a variety of causal predictands and construct their SC predictions accordingly.

Example TSUS (Time-specific unit-specific predictand). The first predictand we consider is the effect of the intervention for a specific unit in a specific time period, which is the primary causal predictand of interest in classical SC analysis with only one treated unit.

For each ever-treated unit $i \in \mathcal{E}$, we can define the TSUS predictand in each post-treatment period:

$$\tau_{ik} := Y_{i(T_i+k)}(T_i) - Y_{i(T_i+k)}(\infty), \quad 0 \le k \le T - T_i.$$

In our empirical application, this predictand captures the effect of economic liberalization on a specific country, measured k years after the adoption of the liberalization policy. Given the prediction $\widehat{Y}_{i(T_i+k)}(\infty)$, we predict τ_{ik} by

$$\widehat{\tau}_{ik} := Y_{i(T_i+k)}(T_i) - \widehat{Y}_{i(T_i+k)}(\infty).$$

See Figure 2(a) for a graphical representation of $\widehat{\tau}_{ik}$. ⌐

Example TAUS (Time-averaged unit-specific predictand). When there are multiple post-treatment periods, scholars may be interested in the effect on a treated unit averaged across all periods. This defines the TAUS predictand, which takes the average over time of the TSUS predictands for an ever-treated unit i:

$$\tau_{i\cdot} := \frac{1}{T - T_i + 1} \sum_{k=0}^{T-T_i} \tau_{ik}.$$

In our empirical application, this is the effect of economic liberalization on a specific country, averaged over the entire post-liberalization period. Given the TSUS prediction for country i in

period k, we predict $\tau_{i\cdot}$ by

$$\widehat{\tau}_{i\cdot} := \frac{1}{T - T_i + 1} \sum_{k=0}^{T-T_i} \widehat{\tau}_{ik}.$$

See Figure 2(b) for a graphical representation of $\widehat{\tau}_{i\cdot}$. ⌐

Example TSUA (Time-specific unit-averaged predictand).   When there are multiple treated units, it is also of interest to define a predictand that captures the average effect of the intervention across a given group of units, at a single period in time. Let $\mathcal{Q} \subseteq \{1, 2, \cdots, T\}$ be a set of adoption times. The TSUA predictand is defined as

$$\tau_{\mathcal{Q}k} := \frac{1}{Q} \sum_{i:T_i \in \mathcal{Q}} \tau_{ik}, \qquad Q = |\{i : T_i \in \mathcal{Q}\}|.$$

In our empirical application, this is the average effect of economic liberalization on the group of countries with adoption times in $\mathcal{Q}$, measured k years after liberalization. In later empirical analysis we consider the treatment effect on countries that liberalized in three different waves—before 1987, between 1987 and 1991, and after 1991. Given the TSUS predictions, we predict $\tau_{\mathcal{Q}k}$ by

$$\widehat{\tau}_{\mathcal{Q}k} := \frac{1}{Q} \sum_{i:T_i \in \mathcal{Q}} \widehat{\tau}_{ik}.$$

See Figure 2(c) for a graphical representation of $\widehat{\tau}_{\mathcal{Q}k}$. ⌐

Example TAUA (Time-averaged unit-averaged predictand).   Finally, when there are multiple treated units and multiple post-treatment periods, we may be interested in a predictand that captures the overall average effect of the intervention.

We define the TAUA predictand as the average of the TSUS predictands across all treated units and over L post-treatment periods:

$$\tau_{\cdot\cdot} := \frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i \in \mathcal{E}} \tau_{ik}.$$

In our empirical application, this is the average effect of economic liberalization on all ever-liberalized countries over L years after the policy adoption. We assume all ever-treated units are observed at least L periods after the adoption for some $L \geq 0$. However, since the observation

11

ends at time T, some ever-treated units may have to be excluded in this definition as L varies. Given the TSUS predictions, we predict $\tau_{..}$ by

$$\widehat{\tau}_{..} := \frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i \in \mathcal{E}} \widehat{\tau}_{ik}.$$

⌐

We conclude this section with a final note on the nature of the predictands defined above. The potential outcomes, treatment adoption times, and individual effects are all viewed as *random* quantities. We assume that there is only a fixed (possibly small) number of ever-treated units and time periods, which is often the case in synthetic control analysis and accommodates our empirical application. Thus, the various predictands defined above are also random quantities in general, which is why we prefer referring to them as "predictands" rather than as treatment "effects". However, we do occasionally use the term "effect" or "predicted effect" to emphasize its randomness and to maintain consistency with the term used for analogous quantities defined in the SC literature under a fixed, non-random potential outcomes framework. In classical large-sample causal analysis, target parameters are often probability or ergodic (non-random) limits of the average effects above as $Q \to \infty$, $J_1 \to \infty$, and/or $T - T_i \to \infty$. Although our results are also valid in such large-sample settings, we develop statistical inference methods based on prediction intervals that describe a region where a new realization of a random causal predictand of interest is likely to be observed, rather than the usual confidence intervals giving a region in the parameter space for a non-random parameter of interest.

4   Uncertainty Quantification

Let τ denote any of the four causal predictands defined above. Our goal is to construct a (random) prediction interval $\mathcal{I}$ such that, with some high probability $(1 - \pi)$ over a conditioning set $\mathscr{H}$, $\mathcal{I}$ covers τ with a pre-specified probability $(1 - \alpha)$ given $\mathscr{H}$:

$$\mathbb{P}\Big\{ \mathbb{P}\big[\tau \in \mathcal{I} \,\big|\, \mathscr{H} \big] \geq 1 - \alpha \Big\} \geq 1 - \pi. \tag{4.1}$$

Figure 2: *Graphical Representation of the Predictands.*
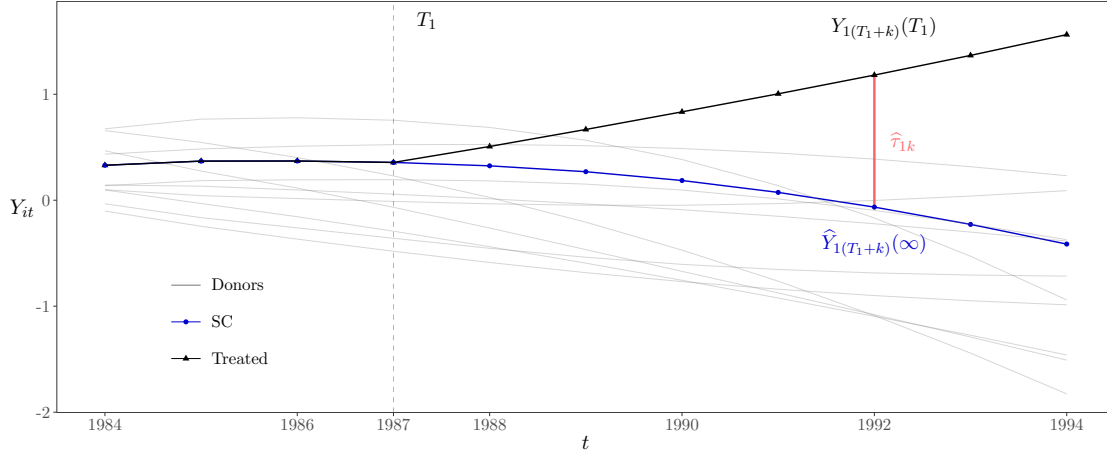
(a) *Time-specific unit-specific (TSUS) predictand.*
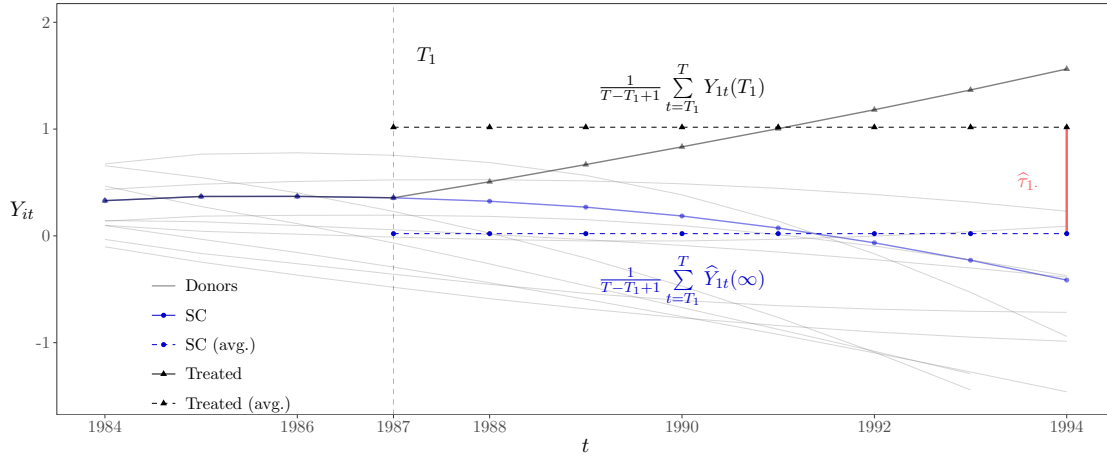
(b) *Time-averaged unit-specific (TAUS) predictand.*
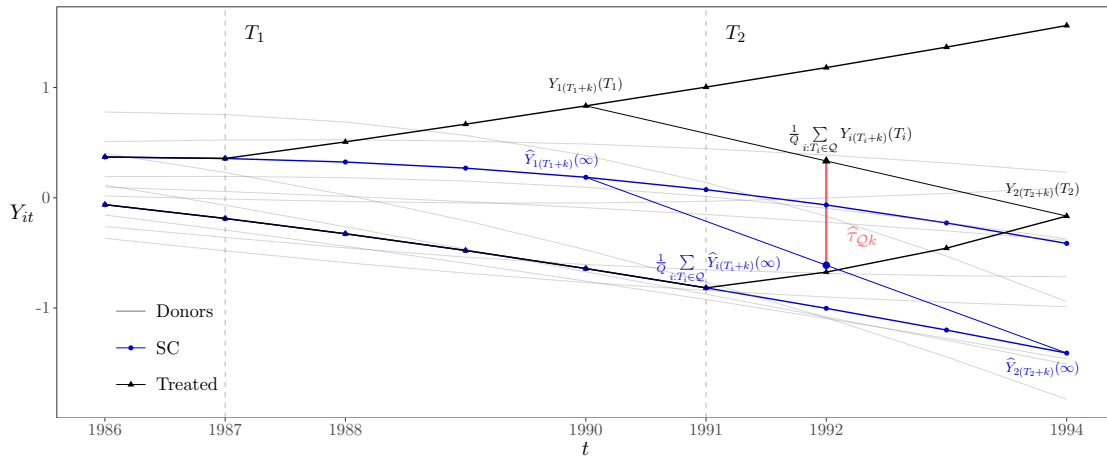
(c) *Time-specific unit-averaged (TSUA) predictand.*

Generally, the choice of the conditioning set $\mathscr{H}$ determines the uncertainty that would not be taken into account by the prediction interval. If we do not condition on anything, technically, $\mathscr{H}$ is a trivial σ-field, and $\mathcal{I}$ reduces to an unconditional prediction interval. Since we regard synthetic controls as a regression-based out-of-sample forecasting problem, where it is standard to condition on the "covariates" (outcomes of donor countries in this context), we focus on the uncertainty from the ever-treated countries. Thus, we let $\mathscr{H} = \{Y_{it} : i \in \mathcal{N}, 1 \leq t \leq T\}$. Also, in practice one needs to set a desired (conditional) coverage level $(1 - \alpha)$, say 90%, whereas the probability loss π over $\mathscr{H}$ is a "small" number that only needs to be theoretically characterized. In this paper, all results are valid if the training period is long enough ($T_i - 1$ is large), with the associated probability $1 - \pi$ characterized precisely. Thus, we say that the conditional prediction interval offers finite-sample probability guarantees. Our results imply that $\pi \to 0$ as $T_i \to \infty$, but no limits or asymptotic arguments are used in this paper.

To better understand the sources of uncertainty of SC predictions, we define the target quantity of the SC weights (conditional on $\mathscr{H}$) that is analogous to (3.1):

$$\mathbf{w}_0^{[i]} = \underset{\mathbf{w} \in \mathcal{W}^{[i]}}{\arg \min} \; \mathbb{E} \left[\sum_{t=1}^{T_i - 1} \left(Y_{it} - \mathbf{Y}'_{\mathcal{N}t} \mathbf{w} \right)^2 \bigg| \mathscr{H} \right]. \tag{4.2}$$

Thus, we can write

$$Y_{it}(\infty) = \mathbf{Y}'_{\mathcal{N}t} \mathbf{w}_0^{[i]} + u_{it}, \quad i \in \mathcal{E}, \; 1 \leq t \leq T, \tag{4.3}$$

where u_{it} is the corresponding pseudo-true residual relative to the conditioning set $\mathscr{H}$. Then, for each ever-treated country $i \in \mathcal{E}$, we have the following decomposition of the counterfactual outcome prediction:

$$\widehat{Y}_{i(T_i+k)}(\infty) - Y_{i(T_i+k)}(\infty) = \mathbf{Y}'_{\mathcal{N}(T_i+k)} (\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}) - u_{i(T_i+k)},$$

where $\mathbf{Y}'_{\mathcal{N}(T_i+k)} (\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]})$ captures the *in-sample uncertainty* from the SC weights construction using the pre-treatment information, and $u_{i(T_i+k)}$ captures the *out-of-sample uncertainty* from the stochastic error in a specific post-treatment period. Notice that the in-sample uncertainty quantification is necessary in this scenario since the conditioning set $Y_{it} \notin \mathscr{H}$ for $i \in \mathcal{E}$.

Accordingly, a similar decomposition can be performed for the prediction $\widehat{\tau}$ of τ:

$$\widehat{\tau} - \tau = \mathsf{InErr}(\tau) + \mathsf{OutErr}(\tau),$$

where $\mathsf{InErr}(\tau)$ and $\mathsf{OutErr}(\tau)$ denote the in-sample error and the out-of-sample error associated with the prediction of τ respectively. Specific expressions of such errors for each treatment effect prediction are given below:

$$\mathsf{InErr}(\tau_{ik}) = -\mathbf{Y}'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}), \qquad\qquad \mathsf{OutErr}(\tau_{ik}) = u_{i(T_i+k)},$$

$$\mathsf{InErr}(\tau_{i\cdot}) = -\frac{1}{T - T_i + 1}\sum_{k=0}^{T-T_i}\mathbf{Y}'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}), \qquad \mathsf{OutErr}(\tau_{i\cdot}) = \frac{1}{T - T_i + 1}\sum_{k=0}^{T-T_i}u_{i(T_i+k)},$$

$$\mathsf{InErr}(\tau_{\mathcal{Q}k}) = -\frac{1}{Q}\sum_{i:T_i\in\mathcal{Q}}\mathbf{Y}'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}), \qquad \mathsf{OutErr}(\tau_{\mathcal{Q}k}) = \frac{1}{Q}\sum_{i:T_i\in\mathcal{Q}}u_{i(T_i+k)},$$

$$\mathsf{InErr}(\tau_{\cdot\cdot}) = -\frac{1}{LJ_1}\sum_{k=1}^{L}\sum_{i\in\mathcal{E}}\mathbf{Y}'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}), \qquad \mathsf{OutErr}(\tau_{\cdot\cdot}) = \frac{1}{LJ_1}\sum_{k=1}^{L}\sum_{i\in\mathcal{E}}u_{i(T_i+k)}.$$

The target SC weights $\mathbf{w}_0^{[i]}$ and the residual u_{it} need to be understood in context. In principle, the weights $\mathbf{w}_0^{[i]}$ represent a "stable" cross-sectional relationship among treated and donor units that can be learned in the training period and translated to the post-treatment period, which is the common feature of all SC methods. In our empirical application, the outcome Y_{it} is GDP per capita, which is well known to be a non-stationary time series. Then, the idea of SC appears to be more applicable when GDP per capita sequences of different countries are cointegrated, where the "stable relationship" is given by the cointegrating vector, and the remainder u_{it} is a stationary series. Thus, we make the following assumption on the data generating process:

Assumption 1 (Data generating process)**.** *Assume that for each $j \in \mathcal{N}$, $Y_{jt} = Y_{j(t-1)} + v_{jt}$, for each $i \in \mathcal{E}$, $Y_{it} = \mathbf{Y}'_{\mathcal{N}t}\mathbf{w}_0^{[i]} + u_{it}$, and $(\mathbf{u}'_t, \mathbf{v}'_t)'$ is i.i.d. over t, where $\mathbf{u}_t = (u_{1t}, \cdots u_{J_1 t})'$ and $\mathbf{v}_t = (v_{(J_1+1)t}, \cdots, v_{Nt})'$. Assume $\mathbf{v}_t$ is sub-Gaussian, and u_{it} is sub-Gaussian conditional on $\mathscr{H}$ with parameter σ_{it} for each $i \in \mathcal{E}$.*

The outcomes of donor units are assumed to follow the simple unit root process with sub-Gaussian errors; other non-stationary patterns can also be accommodated at the cost of more technicalities. The conditional sub-Gaussianity of u_{it}, as precisely defined in Section 4.2, enables us to characterize

the tail probability of u_{it} exceeding a given bound, which is useful for quantifying the out-of-sample uncertainty, but can also be replaced with weaker moment conditions. We emphasize that Assumption 1 is motivated by our empirical application and only used to illustrate the key idea of our method; our theory is more general and can be applied to other types of data, e.g., weakly dependent time series satisfying certain mixing conditions while allowing for a certain degree of model misspecification. See more discussion in Sections 4.1 and 6 below. Also, we view (4.3) as a *predictive* model, and the cointegrated system is merely one way to justify the SC approach; in the Supplemental Appendix Section S.3, we briefly discuss an alternative justification that assumes a linear factor model, and explain how to interpret the in-sample and out-of-sample uncertainty in that scenario.

We employ an intuitive strategy to construct prediction intervals for any causal predictand τ defined before. Specifically, if we can find random intervals $[\underline{M}_{\mathsf{in}}(\tau), \overline{M}_{\mathsf{in}}(\tau)]$ and $[\underline{M}_{\mathsf{out}}(\tau), \overline{M}_{\mathsf{out}}(\tau)]$ that (conditionally) cover the in-sample error $\mathtt{InErr}(\tau)$ and the out-of-sample error $\mathtt{OutErr}(\tau)$, respectively, with certain probabilities, i.e.,

$$\mathbb{P}\Big\{\mathbb{P}\big[\underline{M}_{\mathsf{in}}(\tau) \leq \mathtt{InErr}(\tau) \leq \overline{M}_{\mathsf{in}}(\tau) \mid \mathscr{H}\big] \geq 1 - \alpha_{\mathsf{in}}\Big\} \geq 1 - \pi_{\mathsf{in}} \qquad \text{and}$$

$$\mathbb{P}\Big\{\mathbb{P}\big[\underline{M}_{\mathsf{out}}(\tau) \leq \mathtt{OutErr}(\tau) \leq \overline{M}_{\mathsf{out}}(\tau) \mid \mathscr{H}\big] \geq 1 - \alpha_{\mathsf{out}}\Big\} \geq 1 - \pi_{\mathsf{out}},$$

then, by a union bound we have

$$\mathbb{P}\Big\{\mathbb{P}\big[\widehat{\tau} - \overline{M}_{\mathsf{in}}(\tau) - \overline{M}_{\mathsf{out}}(\tau) \leq \tau \leq \widehat{\tau} - \underline{M}_{\mathsf{in}}(\tau) - \underline{M}_{\mathsf{out}}(\tau)\big|\mathscr{H}\big] \geq 1 - \alpha_{\mathsf{in}} - \alpha_{\mathsf{out}}\Big\} \geq 1 - \pi_{\mathsf{in}} - \pi_{\mathsf{out}}.$$

That is, the prediction interval $\mathcal{I}(\tau) := [\widehat{\tau} - \overline{M}_{\mathsf{in}}(\tau) - \overline{M}_{\mathsf{out}}(\tau), \ \widehat{\tau} - \underline{M}_{\mathsf{in}}(\tau) - \underline{M}_{\mathsf{out}}(\tau)]$ achieves $(1 - \alpha_{\mathsf{in}} - \alpha_{\mathsf{out}})$ coverage probability conditional on $\mathscr{H}$, which holds with probability at least $(1 - \pi_{\mathsf{in}} - \pi_{\mathsf{out}})$ over $\mathscr{H}$. In practice, we can set, for example, $\alpha_{\mathsf{in}} = \alpha_{\mathsf{out}} = \alpha/2$ to achieve the desired coverage level $1 - \alpha$, and our theory will precisely characterize $\pi_{\mathsf{in}} + \pi_{\mathsf{out}}$ and ensure that it is small at least when the training period is long enough. Bounding two errors separately makes the resultant prediction interval for τ conservative, but it clearly differentiates the contribution of the SC training procedure and the out-of-sample prediction procedure to the final inference.

4.1 In-Sample Uncertainty Quantification

We begin with the in-sample uncertainty quantification for the SC prediction of the time-specific counterfactual outcome $Y_{i(T_i+k)}(\infty)$ for each ever-treated unit, and then extend it to the four causal predictands defined before. Importantly, we take the decomposition (4.3) as a predictive model with no structural meaning, which does not have to be a "correctly specified" mean regression model. That is, we allow the outcomes of the donor countries $\mathbf{Y}_{\mathcal{N}t}$ to be possibly correlated with the SC residual u_{it}, i.e., $\boldsymbol{\gamma}^{[i]} := \mathbb{E}[\mathbf{Y}_{\mathcal{N}t} u_{it} | \mathscr{H}] \neq \mathbf{0}$, which violates the standard identification assumption for mean regression models. Accommodating this possibility is important, especially in high-dimensional scenarios, since SC constraints in $\mathcal{W}^{[i]}$ are usually imposed to make SC predictions feasible, and probably more stable, rather than being grounded in prior knowledge of the true data generating process.

Our proposed uncertainty quantification strategy only makes use of three basic facts: (i) $\widehat{\mathbf{w}}^{[i]}$ is the optimizer of the sample-based SC problem (3.1); (ii) $\mathbf{w}_0^{[i]}$ is the optimizer of the population-based SC problem (4.2); and (iii) the L1-L2 constraint set $\mathcal{W}^{[i]}$ used in our empirical application is convex. Actually, (i) and (ii) are true by definition of SC, and (iii) are true for all commonly used constraints, such as simplex, L1 constraint, L2 constraint, or some combinations thereof. Consequently, it can be shown that the following inequality holds (deterministically):

$$(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]})'\widehat{\mathbf{Q}}^{[i]}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}) - 2(\widehat{\boldsymbol{\gamma}}^{[i]} - \boldsymbol{\gamma}^{[i]})'(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}) \leq 0 \quad \text{for each } i \in \mathcal{E},$$

where $\widehat{\mathbf{Q}}^{[i]} = \sum_{t=1}^{T_i-1} \mathbf{Y}_{\mathcal{N}t}\mathbf{Y}_{\mathcal{N}t}'$ and $\widehat{\boldsymbol{\gamma}}^{[i]} = \sum_{t=1}^{T_i-1} \mathbf{Y}_{\mathcal{N}t} u_{it}$. Therefore, the following bounds on the in-sample error of the prediction $\widehat{Y}_{i(T_i+k)}(\infty)$ hold:

$$\inf_{\boldsymbol{\delta} \in \mathcal{M}_{\widehat{\gamma}}^{[i]}} \mathbf{Y}_{\mathcal{N}(T_i+k)}'\boldsymbol{\delta} \leq \mathbf{Y}_{\mathcal{N}(T_i+k)}'(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]}) \leq \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\widehat{\gamma}}^{[i]}} \mathbf{Y}_{\mathcal{N}(T_i+k)}'\boldsymbol{\delta},$$

where $\mathcal{M}_{\widehat{\gamma}}^{[i]} = \{\boldsymbol{\delta} \in \Delta^{[i]} : \boldsymbol{\delta}'\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta} - 2(\widehat{\boldsymbol{\gamma}}^{[i]} - \boldsymbol{\gamma}^{[i]})'\boldsymbol{\delta} \leq 0\}$ and $\Delta^{[i]} = \{\boldsymbol{\delta} : \boldsymbol{\delta} + \mathbf{w}_0^{[i]} \in \mathcal{W}^{[i]}\}$. Conditional on $\mathscr{H}$, $\widehat{\boldsymbol{\gamma}}^{[i]}$ is random, making the feasible set $\mathcal{M}_{\widehat{\gamma}}^{[i]}$, and thus the resulting upper and lower bounds, stochastic as well. We can employ a normal distributional approximation of $\widehat{\boldsymbol{\gamma}}^{[i]}$ and, for instance, set the lower and upper bounds to $(\alpha_{\texttt{in}}/2)$-quantile of $\inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]}} \mathbf{Y}_{\mathcal{N}(T_i+k)}'\boldsymbol{\delta}$ and $(1-\alpha_{\texttt{in}}/2)$-quantile

17

of $\sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}$, respectively, conditional on $\mathscr{H}$, where $\mathcal{M}_{\mathbf{G}}^{[i]} = \{\boldsymbol{\delta} \in \Delta^{[i]} : \boldsymbol{\delta}' \widehat{\mathbf{Q}}^{[i]} \boldsymbol{\delta} - 2\mathbf{G}^{[i]\prime} \boldsymbol{\delta} \leq 0\}$, $\mathbf{G}^{[i]} | \mathscr{H} \sim \mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma}^{[i]})$ and $\boldsymbol{\Sigma}^{[i]} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}^{[i]} | \mathscr{H}]$. It can be shown that, with high probability over $\mathscr{H}$, such an interval has at least $(1 - \alpha_{\mathtt{in}} - \frac{c(\log T_0)^2}{\sqrt{T_0}})$ conditional coverage probability for the in-sample error $\mathbf{Y}'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}_0^{[i]})$ for some constant $c > 0$, where $\frac{c(\log T_0)^2}{\sqrt{T_0}}$ is some probability loss due to the normal approximation.

However, the above bounds cannot be directly implemented in practice, since the (centered) constraint set $\Delta^{[i]}$ depends on the unknown pseudo-true value $\mathbf{w}_0^{[i]}$, and the normal vector $\mathbf{G}^{[i]}$ depends on the unknown covariance matrix $\boldsymbol{\Sigma}^{[i]}$. We thus propose a feasible simulation-based strategy allowing for *possibly nonlinear* constraints, with unknown quantities replaced with plug-in approximations thereof.

First, we need a feasible constraint set $\Delta^{[i]\star}$ for simulation that is "similar" to $\Delta^{[i]}$ (near the origin). A general data-dependent strategy for constructing $\Delta^{[i]\star}$ is discussed in Section 6.2. We now focus on the L1-L2 constraint in our empirical application, where $\Delta^{[i]}$ consists of one equality constraint ($\sum_{j \in \mathcal{N}} w_j = 1$), a sequence of linear inequality constraints ($w_j \geq 0$ for $j \in \mathcal{N}$), and one *nonlinear* inequality constraint ($\sum_{j \in \mathcal{N}} w_j^2 \leq (Q^{[i]})^2$) on the weights vector $\mathbf{w} = (w_j : j \in \mathcal{N})$. The equality constraint is maintained without adjustment in the simulation. For inequality constraints, we introduce a tuning parameter $\varrho^{[i]}$ to determine whether they are binding and then make adjustments accordingly. Specifically, we set $\varrho^{[i]} = C^{[i]} \frac{\log T_0}{T_0}$ for some constant $C^{[i]} > 0$, which can be viewed as a high-probability bound on the deviation of $\widehat{\mathbf{w}}^{[i]}$ from $\mathbf{w}_0^{[i]}$. A data-dependent way to choose $\varrho^{[i]}$ is described in Section 6.2. Let $\widehat{\mathbf{w}} = (\widehat{w}_j : j \in \mathcal{N})$. For each linear constraint $w_j \geq 0$, we consider it binding and use $w_j \geq \widehat{w}_j^{[i]}$ in the simulation if $\widehat{w}_j^{[i]} < \varrho^{[i]}$; otherwise, the original constraint, $w_j \geq 0$, is retained. For the nonlinear constraint $\|\mathbf{w}\|_2^2 \leq (Q^{[i]})^2$, impose $\|\mathbf{w}\|_2^2 \leq \|\widehat{\mathbf{w}}^{[i]}\|_2^2 + (\varrho^{[i]})^2$ in the simulation if $(Q^{[i]})^2 - \|\widehat{\mathbf{w}}^{[i]}\|_2^2 < 2\|\widehat{\mathbf{w}}\|_2 \varrho^{[i]}$, and retain $\|\mathbf{w}\|_2^2 \leq (Q^{[i]})^2$ otherwise. Then, $\Delta^{[i]\star}$ is defined as the set of vectors $\mathbf{w} - \widehat{\mathbf{w}} \in \mathbb{R}^{J_0}$ satisfying all these restrictions.

It can be shown that, with high probability, all points in $\Delta^{[i]}$ (near the origin) are contained in $\Delta^{[i]\star}$. In fact, if all constraints are linear in parameters, $\Delta^{[i]}$ and $\Delta^{[i]\star}$ are (locally) identical. With nonlinear constraints, the two sets are typically not identical but still remain close so that the "enlargement" from $\Delta^{[i]}$ to $\Delta^{[i]\star}$ is modest. See Section S.5.2 of the Supplemental Appendix for details. Given this fact, searching for the supremum or infimum of $\mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}$ over this (slightly) larger set $\Delta^{[i]\star}$ yields valid bounds on the in-sample error.

Second, we need an estimator $\widehat{\boldsymbol{\Sigma}}^{[i]}$ of the covariance matrix $\boldsymbol{\Sigma}^{[i]}$. A variety of well-established heteroskedasticity/serial-correlation-robust estimators can be used. We require $\widehat{\boldsymbol{\Sigma}}^{[i]}$ to be a "good" approximation of $\boldsymbol{\Sigma}^{[i]}$ in the sense of condition (iii) in Corollary 1 below. This allows us to approximate the infeasible normal distribution $\mathsf{N}(\mathbf{0}, \boldsymbol{\Sigma}^{[i]})$ by $\mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}^{[i]})$, which can be simulated using the data.

Once $\Delta^{[i]\star}$ and $\widehat{\boldsymbol{\Sigma}}^{[i]}$ are available, we can simply draw random vectors from $\mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}^{[i]})$ conditional on the data, and then set

$$\underline{M}_{\tt in} = \underline{\mathfrak{c}}^\star(\alpha_{\tt in}/2) \quad \text{and} \quad \overline{M}_{\tt in} = \overline{\mathfrak{c}}^\star(1 - \alpha_{\tt in}/2), \tag{4.4}$$

where

$$\underline{\mathfrak{c}}^\star(\alpha_{\tt in}/2) = (\alpha_{\tt in}/2)\text{-quantile of} \inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]\star}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta},$$

$$\overline{\mathfrak{c}}^\star(1 - \alpha_{\tt in}/2) = (1 - \alpha_{\tt in}/2)\text{-quantile of} \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]\star}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta},$$

conditional on the data, with $\mathcal{M}_{\mathbf{G}}^{[i]\star} = \{\boldsymbol{\delta} \in \Delta^{[i]\star} : \boldsymbol{\delta}'\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta} - 2(\mathbf{G}^{[i]\star})'\boldsymbol{\delta} \leq 0\}$ and $\mathbf{G}^{[i]\star}|\text{Data} \sim \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}}^{[i]})$. Under some mild regularity conditions, $[\underline{M}_{\tt in}, \overline{M}_{\tt in}]$ is a feasible prediction interval that achieves the desired conditional coverage of the counterfactual outcome $Y_{i(T_i+k)}$.

We are ready to construct bounds on the in-sample error for each treatment effect prediction.

Example TSUS (Time-specific unit-specific predictand, continued). Since the in-sample error of $\widehat{\tau}_{ik}$ is the opposite of that of $\widehat{Y}_{i(T_i+k)}$, we simply set $\underline{M}_{\tt in}(\tau_{ik}) = -\overline{M}_{\tt in}$ and $\overline{M}_{\tt in}(\tau_{ik}) = -\underline{M}_{\tt in}$. $\quad\lrcorner$

Example TAUS (Time-averaged unit-specific predictand, continued). Since the in-sample error of $\widehat{\tau}_{i\cdot}$ is (the opposite of) the average in-sample error of $\widehat{Y}_{i(T_i+k)}$ over time, we can set $\underline{M}_{\tt in}(\tau_{i\cdot}) = \underline{\mathfrak{c}}^\star(\alpha_{\tt in}/2)$ and $\overline{M}_{\tt in}(\tau_{i\cdot}) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\tt in}/2)$ where

$$\underline{\mathfrak{c}}^\star(\alpha_{\tt in}/2) = (\alpha_{\tt in}/2)\text{-quantile of} \inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]\star}} -\frac{1}{T - T_i + 1} \sum_{k=0}^{T-T_i} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta},$$

$$\overline{\mathfrak{c}}^\star(1 - \alpha_{\tt in}/2) = (1 - \alpha_{\tt in}/2)\text{-quantile of} \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^{[i]\star}} -\frac{1}{T - T_i + 1} \sum_{k=0}^{T-T_i} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta},$$

conditional on the data. ⌋

Example TSUA (Time-specific unit-averaged predictand, continued). Since the in-sample error of $\widehat{\tau}_{Qk}$ is (the opposite of) the average in-sample error of $\widehat{Y}_{i(T_i+k)}$ across multiple treated units, we can set $\underline{M}_{\mathtt{in}}(\tau_{Qk}) = \underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2)$ and $\overline{M}_{\mathtt{in}}(\tau_{Qk}) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2)$ where

$$\underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2) = (\alpha_{\mathtt{in}}/2)\text{-quantile of} \inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^*} -\frac{1}{Q} \sum_{i:T_i \in \mathcal{Q}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}^{[i]},$$

$$\overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2) = (1 - \alpha_{\mathtt{in}}/2)\text{-quantile of} \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^*} -\frac{1}{Q} \sum_{i:T_i \in \mathcal{Q}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}^{[i]},$$

conditional on the data, and $\mathcal{M}_{\mathbf{G}}^\star = \{\boldsymbol{\delta} \in \times_{i:T_i \in \mathcal{Q}} \Delta^{[i]\star} : \sum_{i:T_i \in \mathcal{Q}}[\boldsymbol{\delta}^{[i]\prime}\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta}^{[i]\prime} - 2(\mathbf{G}^{[i]\star})'\boldsymbol{\delta}^{[i]\prime}] \leq 0\}$ with $\boldsymbol{\delta} = (\boldsymbol{\delta}^{[i]} : T_i \in \mathcal{Q}) \in \mathbb{R}^{QJ_0}$. In this case, the in-sample uncertainty depends on SC weights for multiple treated units, so we aggregate the previously described quadratic constraint for each $\boldsymbol{\delta}^{[i]}$ into one single quadratic constraint for the entire vector $\boldsymbol{\delta}$. With multiple treated units, we should draw the *long* vector $\mathbf{G}^\star := (\mathbf{G}^{[i]\star} : T_i \in \mathcal{Q})$ conditional on the data from $\mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$ with $\widehat{\boldsymbol{\Sigma}}$ an estimate of $\boldsymbol{\Sigma} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$ for $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\gamma}}^{[i]} : T_i \in \mathcal{Q})$. Therefore, the correlation structure among different treated units is implicitly captured. ⌋

Example TAUA (Time-averaged unit-averaged predictand, continued). Since the in-sample error of $\widehat{\tau}_{..}$ is (the opposite of) the average in-sample error of $\widehat{Y}_{i(T_i+k)}$ across multiple treated units over time, we set $\underline{M}_{\mathtt{in}}(\tau_{..}) = \underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2)$ and $\overline{M}_{\mathtt{in}}(\tau_{..}) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2)$ where

$$\underline{M}_{\mathtt{in}}(\tau_{..}) = (\alpha_{\mathtt{in}}/2)\text{-quantile of} \inf_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^*} -\frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i \in \mathcal{E}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}^{[i]},$$

$$\overline{M}_{\mathtt{in}}(\tau_{..}) = (1 - \alpha_{\mathtt{in}}/2)\text{-quantile of} \sup_{\boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}^*} -\frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i \in \mathcal{E}} \mathbf{Y}'_{\mathcal{N}(T_i+k)} \boldsymbol{\delta}^{[i]},$$

conditional on the data, and $\mathcal{M}_{\mathbf{G}}^\star = \{\boldsymbol{\delta} \in \times_{i:i \in \mathcal{E}} \Delta^{[i]\star} : \sum_{i \in \mathcal{E}}[\boldsymbol{\delta}^{[i]\prime}\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta}^{[i]\prime} - 2(\mathbf{G}^{[i]\star})'\boldsymbol{\delta}^{[i]\prime}] \leq 0\}$. As in the previous example, we draw random vectors $\mathbf{G}^\star := (\mathbf{G}^{[i]\star} : i \in \mathcal{E})$ conditional on the data from $\mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$ with $\widehat{\boldsymbol{\Sigma}}$ an estimate of $\boldsymbol{\Sigma} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$ for $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\gamma}}^{[i]} : i \in \mathcal{E})$. ⌋

Remark 1 (Alternative Bounds). For the unit-averaged TSUA and TAUA predictands, we combine the quadratic constraints from optimization into a single constraint on the entire (centered) SC

weights vector $\boldsymbol{\delta}$. This "aggregation" strategy is consistent with the view of taking (3.1) as a multiple-equation regression problem, as detailed in Section 6. By contrast, we could also retain the individual quadratic constraints $\boldsymbol{\delta}^{[i]\prime}\widehat{\mathbf{Q}}^{[i]}\boldsymbol{\delta}^{[i]\prime} - 2(\mathbf{G}^{[i]\star})'\boldsymbol{\delta}^{[i]\prime} \leq 0$, for $T_i \in \mathcal{Q}$ or $i \in \mathcal{E}$, each restricting a subvector $\boldsymbol{\delta}^{[i]}$ of $\boldsymbol{\delta}$. This alternative strategy applies to the special case of SC analysis in which SC weights for each treated unit are constructed separately, yielding possibly tighter bounds on in-sample errors; see more detailed discussion in the Supplemental Appendix Section S.2.3. In Table S.2 of the Supplemental Appendix, we also illustrate the potential improvement of this method in our empirical application in terms of interval length. ⌙

Remark 2 (Scalable Optimization Implementations)**.** The proposed bounds on in-sample errors are suprema and infima of linear functions subject to some linear or quadratic constraints on parameters. In Section S.6 of the Supplemental Appendix, we show that such problems can be recast as conic optimization programs, which provide massive speed improvements in practice. ⌙

4.2 Out-of-Sample Uncertainty Quantification

To bound the out-of-sample error, we propose an easy-to-implement approach based on non-asymptotic concentration inequalities. Recall that by Assumption 1, $u_{it} - \mathbb{E}[u_{it}|\mathscr{H}]$ is conditional-on-$\mathscr{H}$ sub-Gaussian with parameter σ_{it}, which implies that for any $\lambda > 0$,

$$\mathbb{P}\Big(|u_{it} - \mathbb{E}[u_{it}|\mathscr{H}]| \geq \lambda \Big| \mathscr{H}\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2\sigma_{it}^2}\Big). \tag{4.5}$$

Consequently, for the out-of-sample error $-u_{i(T_i+k)}$ of the counterfactual outcome prediction $\widehat{Y}_{i(T_i+k)}$, we can set

$$
\begin{aligned}
\underline{M}_{\mathsf{out}} &= -\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - \sqrt{2\sigma_{i(T_i+k)}^2 \log(2/\alpha_{\mathsf{out}})}, \quad \text{and} \\
\overline{M}_{\mathsf{out}} &= -\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + \sqrt{2\sigma_{i(T_i+k)}^2 \log(2/\alpha_{\mathsf{out}})},
\end{aligned}
\tag{4.6}
$$

which yields a prediction interval $[\underline{M}_{\mathsf{out}}, \overline{M}_{\mathsf{out}}]$ that covers $-u_{i(T_i+k)}$ with at least $(1-\alpha_{\mathsf{out}})$ conditional coverage probability. This bound does not rely on the exact distribution of u_{it} and is valid as long as u_{it} has tails no fatter than those specified in (4.5). We emphasize that the sub-Gaussianity assumption is one of many possibilities. The above strategy could be applied using other concen-

tration inequalities requiring weaker moment conditions, though the resulting prediction intervals may be wider.

In practice, one could first construct pre-treatment residuals $\widehat{u}_{it} = Y_{it} - \mathbf{Y}'_{\mathcal{N}t}\widehat{\mathbf{w}}^{[i]}$ for each $i \in \mathcal{E}$ and $t < T_i$, and then estimate the conditional moments of u_{it} employing various parametric or nonparametric regression of $\widehat{u}_{it}$. Such estimates can then be translated into the necessary estimates of $\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}]$ and $\sigma^2_{i(T_i+k)}$ for constructing $\underline{M}_{\text{out}}$ and $\overline{M}_{\text{out}}$. The smallest parameter σ^2_{it} that makes the tail probability bound (4.5) hold is sometimes termed the optimal variance proxy, which is known to be the same as $\mathbb{V}[u_{it}|\mathscr{H}]$ if u_{it} is *strictly* sub-Gaussian conditional on $\mathscr{H}$ (Arbel et al., 2020; Bobkov et al., 2024). In such cases estimating the first two conditional moments of u_{it} suffices to implement the bounds in (4.6). Moreover, the unknown conditional moments could also be set using external information, or tabulated across different values to assess the sensitivity of the resulting prediction intervals.

We can apply the same idea to bound the out-of-sample errors of the four treatment effect predictions.

Example TSUS (Time-specific unit-specific predictand, continued). Since the out-of-sample error in this case is u_{it}, we simply set $\underline{M}_{\text{out}}(\tau_{ik}) = \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - [2\sigma^2_{i(T_i+k)}\log(2/\alpha_{\text{out}})]^{1/2}$ and $\overline{M}_{\text{out}}(\tau_{ik}) = \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + [2\sigma^2_{i(T_i+k)}\log(2/\alpha_{\text{out}})]^{1/2}$. ⌟

Example TAUS (Time-averaged unit-specific predictand, continued). Since $u_{it} - \mathbb{E}[u_{it}|\mathscr{H}]$ is condition-on-$\mathscr{H}$ sub-Gaussian with parameter σ_{it}, it can be shown that the out-of-sample error $\texttt{OutErr}(\tau_{i\cdot})$, as the average of u_{it} over time, is also sub-Gaussian, satisfying that for any $\lambda > 0$,

$$\mathbb{P}\Big(|\texttt{OutErr}(\tau_{i\cdot}) - \mathbb{E}[\texttt{OutErr}(\tau_{i\cdot})|\mathscr{H}]| \geq \lambda|\mathscr{H}\Big) \leq 2\exp\Big(-\frac{\lambda^2}{2\sigma^2_{i\cdot}}\Big), \quad \sigma_{i\cdot} := \frac{1}{T - T_i + 1}\sum_{t=T_i}^{T}\sigma_{it}.$$

Therefore, we can set

$$\underline{M}_{\text{out}}(\tau_{i\cdot}) = \frac{1}{T - T_i + 1}\sum_{k=0}^{T-T_i}\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - \sqrt{2\sigma^2_{i\cdot}\log(2/\alpha_{\text{out}})}, \quad \text{and}$$

$$\overline{M}_{\text{out}}(\tau_{i\cdot}) = \frac{1}{T - T_i + 1}\sum_{k=0}^{T-T_i}\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + \sqrt{2\sigma^2_{i\cdot}\log(2/\alpha_{\text{out}})}.$$

This result holds regardless of the dependence structure of u_{it}, but if u_{it} is indeed independent over

t, the same result holds with $\sigma_{i\cdot} := \frac{1}{T-T_i+1} \left(\sum_{t=T_i}^{T} \sigma_{it}^2 \right)^{1/2}$, leading to an improved bound. ⌟

Example TSUA (Time-specific unit-averaged predictand, continued). In this scenario, the out-of-sample error $\mathtt{OutErr}(\tau_{\mathcal{Q}k})$ is a cross-sectional average of u_{it} at different times. The uncertainty quantification strategy outlined previously in Example TAUS can still be applied, with the caveat that it is uncommon in SC analysis to assume u_{it} is stationary and/or independent over i. By contrast, it is reasonable to assume u_{it} is stationary and/or independent (at least weakly dependent) over time. Therefore, we employ the concentration inequality that holds under a general dependence structure and set

$$\underline{M}_{\mathsf{out}}(\tau_{\mathcal{Q}k}) = \frac{1}{Q} \sum_{i:T_i\in\mathcal{Q}} \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - \sqrt{2\sigma_{\mathcal{Q}k}^2 \log(2/\alpha_{\mathsf{out}})}, \quad \text{and}$$

$$\overline{M}_{\mathsf{out}}(\tau_{\mathcal{Q}k}) = \frac{1}{Q} \sum_{i:T_i\in\mathcal{Q}} \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + \sqrt{2\sigma_{\mathcal{Q}k}^2 \log(2/\alpha_{\mathsf{out}})}$$

with $\sigma_{\mathcal{Q}k} := \frac{1}{Q} \sum_{i:T_i\in\mathcal{Q}} \sigma_{i(T_i+k)}$. ⌟

Example TAUA (Time-averaged unit-averaged predictand, continued). Since the adoption time T_i may be heterogeneous across i, the out-of-sample error $\mathtt{OutErr}(\tau_{\cdot\cdot})$ is an average of out-of-sample errors of different units in different periods, which is (conditionally) sub-Gaussian as well. Then, we set

$$\underline{M}_{\mathsf{out}}(\tau_{\cdot\cdot}) = \frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i\in\mathcal{E}} \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - \sqrt{2\sigma_{\cdot\cdot}^2 \log(2/\alpha_{\mathsf{out}})}, \quad \text{and}$$

$$\overline{M}_{\mathsf{out}}(\tau_{\cdot\cdot}) = \frac{1}{LJ_1} \sum_{k=1}^{L} \sum_{i\in\mathcal{E}} \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + \sqrt{2\sigma_{\cdot\cdot}^2 \log(2/\alpha_{\mathsf{out}})}$$

with $\sigma_{\cdot\cdot} := \frac{1}{LJ_1} \sum_{k=1}^{K} \sum_{i\in\mathcal{E}} \sigma_{i(T_i+k)}$. ⌟

In addition to the concentration-based approach described above, other strategies, including location-scale models and quantile regression, were proposed in Cattaneo, Feng and Titiunik (2021) for out-of-sample uncertainty quantification. We briefly review them in the Supplemental Appendix, Section S.2.1.

4.3 Algorithmic Implementation

We summarize our methodology for the particular case illustrated in this section, focusing on the L1-L2 constraint and using τ_{ik} (TSUS) as the target predictand. A more detailed discussion of the procedures, including recommended rules of thumb for implementation and other practical regularization choices, can be found in Section S.7.2 of the Supplemental Appendix and in our companion software article Cattaneo, Feng, Palomba and Titiunik (2025, Section 3.1 and Section 4.3). In particular, the required optimization programs are implemented using modern, robust, and scalable conic programming methods (see Remark 2 for details). Finally, all our proposed methods are readily available for implementation via our general-purpose software package `scpi` in `Python`, `R`, and `Stata`.

4.4 Theoretical Justification

The following corollary, as a special case of our more general results presented in Section 6, provides a theoretical justification of the proposed method, which closely matches our empirical application. Let $T_0 = T_1 - 1$, and $m_\ell^{[i]}(\mathbf{w}) \leq 0$ be the ℓ-th inequality constraint among $-w_j \leq 0$, $j \in \mathcal{N}$, and $\|\mathbf{w}\|_2^2 \leq (Q^{[i]})^2$. Define $\tilde{\mathbf{Q}} = \frac{1}{T_0} \sum_{t=1}^{T_0} \tilde{\mathbf{G}}(t/T_0) \tilde{\mathbf{G}}(t/T_0)'$ where $\tilde{\mathbf{G}}$ is a mean-zero Brownian motion on $[0,1]$ with variance $\mathbb{E}[\mathbf{v}_t \mathbf{v}_t']$. We use $s_{\min}(\mathbf{A})$ and $s_{\max}(\mathbf{A})$ to denote the minimum and maximum singular values of a matrix $\mathbf{A}$, respectively. Also, for simplicity, in the following corollary we define $\boldsymbol{\Sigma} = \mathbb{V}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$ for $\widehat{\boldsymbol{\gamma}} = (\widehat{\boldsymbol{\gamma}}^{[i]} : i \in \mathcal{E})$ and let $\widehat{\boldsymbol{\Sigma}}$ be an estimator of $\boldsymbol{\Sigma}$, as in Example TAUA.

Corollary 1. *Let Assumption 1 hold. In addition, suppose that with probability over $\mathscr{H}$ at least $1 - \pi_0$, the following conditions hold:*

(i) $\min_{1 \leq t \leq T} s_{\min}(\mathbb{E}[\mathbf{u}_t \mathbf{u}_t'|\mathscr{H}]) > \eta$ *for some constant $\eta > 0$;*

(ii) $(\log T_0)^{-1/5} \leq s_{\min}(\tilde{\mathbf{Q}}) \leq s_{\max}(\tilde{\mathbf{Q}}) \leq (\log T_0)^{1/5}$;

(iii) $\mathbb{P}(\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\| \leq \epsilon_{\Sigma,1}^\star)|\mathscr{H}) \geq 1 - \epsilon_{\Sigma,2}^\star$ *with* $\epsilon_{\Sigma,1}^\star \leq \frac{T_0^2 \eta}{8\sqrt{d}(\log T_0)^{1/5}}$;

(iv) *For all $1 \leq \ell \leq J_0 + 1$, $\varrho_\ell^{[i]} < |m_\ell^{[i]}(\mathbf{w}_0^{[i]})| - (\log T_0)T_0^{-1}$ if $m_\ell^{[i]}(\mathbf{w}_0^{[i]}) \neq 0$.*

Then, for any causal prediction $\widehat{\tau} \in \{\widehat{\tau}_{ik}, \widehat{\tau}_{i\cdot}, \widehat{\tau}_{\mathcal{Q}k}, \widehat{\tau}_{\cdot\cdot}\}$, when T_0 is large enough,

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

Algorithm 1 Uncertainty Quantification (Outcome-only Non-stationary Case)

Require: data $\{(Y_{it}, T_i)_{i=1}^N : t = 1, \ldots, T\}$, sets of donors $\mathcal{N}$ and treated units $\mathcal{E}$, confidence levels $\alpha_{\texttt{in}}, \alpha_{\texttt{out}} \in (0, 1)$, and the number of simulations $S \in \mathbb{N}$.

Step 1: Computation of $\widehat{\mathbf{w}}$

1: tune $Q^{[i]}$, for $i \in \mathcal{E}$ (see Section S.7.2 of the Supplemental Appendix)
2: solve the optimization problem in (3.1) for each $i \in \mathcal{E}$

Step 2: In-sample Uncertainty Quantification

3: form feasible constraint set $\Delta^{[i]\star}$ for each $i \in \mathcal{E}$
    a: tune $\varrho^{[i]}$ using (6.9)
    b: define

$$\Delta^{[i]\star} = \Big\{ \mathbf{w} - \widehat{\mathbf{w}}^{[i]} : \|\mathbf{w}\|_1 = 1, w_j \geq \widehat{w}_j^{[i]} \mathbb{1}(\widehat{w}_j^{[i]} < \varrho^{[i]}), \; j = 1, \ldots, J_0,$$

$$\|\mathbf{w}\|_2^2 - (Q^{[i]})^2 \leq (\|\widehat{\mathbf{w}}^{[i]}\|_2^2 - (Q^{[i]})^2) \mathbb{1}(\|\mathbf{w}^{[i]}\|_2^2 - (Q^{[i]})^2 > -2\|\widehat{\mathbf{w}}^{[i]}\|_2 \varrho^{[i]}) + (\varrho^{[i]})^2 \Big\}$$

4: estimate $\boldsymbol{\Sigma}^{[i]}$ for each $i \in \mathcal{E}$
    a: estimate pre-treatment residuals as $\widehat{u}_{it} = Y_{it} - \mathbf{Y}_{\mathcal{N}t}' \widehat{\mathbf{w}}^{[i]}$ for $t < T_i$
    b: estimate $\mathbb{E}[u_{it} \mid \mathscr{H}]$ with a linear regression for $t < T_i$
    c: estimate $\mathbb{V}[u_{it} \mid \mathscr{H}]$ with $\widehat{\mathbb{V}}[u_{it} \mid \mathscr{H}] = (\widehat{u}_{it} - \widehat{\mathbb{E}}[u_{it} \mid \mathscr{H}])^2$ for $t < T_i$
    d: compute $\widehat{\boldsymbol{\Sigma}}^{[i]} = \mathbf{Y}_{\mathcal{N}}' \widehat{\mathbb{V}}[\mathbf{u}^{[i]}|\mathscr{H}]\mathbf{Y}_{\mathcal{N}})$ where $\mathbf{Y}_{\mathcal{N}} = [\mathbf{Y}_{\mathcal{N}1} \cdots \mathbf{Y}_{\mathcal{N}(T_i-1)}]'$
5: **for** $s \leftarrow 1$ to S **do:**
    a: draw $\mathbf{G}^\star \mid \mathscr{H} \sim \mathsf{N}(\mathbf{0}, \widehat{\boldsymbol{\Sigma}})$
    b: define $\ell^\star(\boldsymbol{\delta}) = \boldsymbol{\delta}' \widehat{\mathbf{Q}} \boldsymbol{\delta} - 2(\mathbf{G}^\star)' \boldsymbol{\delta}$ with $\widehat{\mathbf{Q}} = \text{diag}(\widehat{\mathbf{Q}}^{[i]}, i \in \mathcal{E}), \widehat{\mathbf{Q}}^{[i]} = \mathbf{Y}_{\mathcal{N}}' \mathbf{Y}_{\mathcal{N}}$
    c: solve

$$l_{(s)} := \inf_{\substack{\boldsymbol{\delta} \in \Delta^{[i]\star} \\ \ell^\star(\boldsymbol{\delta}) \leq 0}} \mathbf{Y}_{\mathcal{N}(T_i+k)}' \boldsymbol{\delta} \qquad \text{and} \qquad u_{(s)} := \sup_{\substack{\boldsymbol{\delta} \in \Delta^{[i]\star} \\ \ell^\star(\boldsymbol{\delta}) \leq 0}} \mathbf{Y}_{\mathcal{N}(T_i+k)}' \boldsymbol{\delta}$$

6: get $\underline{\mathfrak{c}}^\star$ as the $(\alpha_{\texttt{in}}/2)$−quantile of $\{l_{(s)}\}_{s=1}^S$ and $\overline{\mathfrak{c}}^\star$ as the $(1 - \alpha_{\texttt{in}}/2)$−quantile of $\{u_{(s)}\}_{s=1}^S$
7: get $\underline{M}_{\texttt{in}} = \underline{\mathfrak{c}}^\star, \overline{M}_{\texttt{in}} = \overline{\mathfrak{c}}^\star$, and so $[\underline{M}_{\texttt{in}}(\tau_{ik}); \overline{M}_{\texttt{in}}(\tau_{ik})] = [-\overline{M}_{\texttt{in}}; -\underline{M}_{\texttt{in}}]$

Step 3: Out-of-sample Uncertainty Quantification

8: use pre-treatment residuals to estimate $\mathbb{E}[u_{i(T_i+k)}|\mathscr{H}]$ and $\sigma_{i(T_i+k)}^2$
9: get $[\underline{M}_{\texttt{out}}(\tau_{ik}); \overline{M}_{\texttt{out}}(\tau_{ik})] = \widehat{\mathbb{E}}\left[u_{i(T_i+k)}|\mathscr{H}\right] \pm \sqrt{2\widehat{\sigma}_{i(T_i+k)}^2 \log(2/\alpha_{\texttt{out}})}$

Step 4: Compute Prediction Intervals

10: form an interval as $[\widehat{\tau}_{ik} - \overline{M}_{\texttt{in}}(\tau_{ik}) - \overline{M}_{\texttt{out}}(\tau_{ik}); \widehat{\tau}_{ik} - \underline{M}_{\texttt{in}}(\tau_{ik}) - \underline{M}_{\texttt{out}}(\tau_{ik})]$

---

where $\epsilon = 2\mathfrak{C}_\epsilon (\log T_0)^2 T_0^{-1/2} + 4\epsilon_{\Sigma,1}^\star \sqrt{d}(\log T_0)^{1/5} T_0^{-2} \eta^{-1} + \epsilon_{\Sigma,2}^\star + 3T_0^{-1}$, $\pi = \pi_0 + \frac{\mathfrak{C}_\pi}{T_0}$, and $\mathfrak{C}_\epsilon$ and $\mathfrak{C}_\pi$ are some constants characterized in the proof.

    The additional technical conditions imposed in this corollary are intuitive: (i) is a mild restriction used to guarantee the variance matrix $\boldsymbol{\Sigma}$ is non-degenerate; (ii) is used to characterize the upper and

lower bounds on the matrices $\widehat{\mathbf{Q}}^{[i]}$ and can be shown to hold with high probability when T_0 is large; (iii) is a high-level condition on the convergence of the variance estimator $\widehat{\boldsymbol{\Sigma}}$ and can be verified on a case-by-case basis; and (iv) guarantees that the non-binding constraints can be differentiated from the binding ones under our thresholding rule, making the inclusion restriction on the constraint sets imposed in Theorem 1 satisfied. If these conditions hold with high probability (π_0 is small), then this corollary shows that the proposed prediction intervals can achieve approximately $(1-\alpha_{\mathtt{in}}-\alpha_{\mathtt{out}})$ conditional coverage probability, with the probability loss ϵ decreasing as T_0 grows and $\epsilon^\star_{\Sigma,1}$ and $\epsilon^\star_{\Sigma,2}$ get small (i.e., the variance estimator $\widehat{\boldsymbol{\Sigma}}$ becomes more "precise").

4.5   Simultaneous Prediction Intervals

So far we have focused on constructing prediction intervals that have high coverage of the desired predictands or treatment effects, in particular, the TSUS predictand, which captures the effect of the intervention for a given unit in *each* post-treatment period. In some applications, it might be appealing to construct prediction intervals that have high *simultaneous* coverage in multiple post-treatment periods, usually termed *simultaneous prediction intervals*. They can be employed to test, for example, whether the largest (or smallest) treatment effect across different periods is significantly different from zero.

Specifically, for a particular treated unit $i \in \mathcal{E}$, we aim to construct a sequence of intervals $\mathcal{I}_k$ for $0 \leq k \leq L$ for some $L \leq T - T_i$ such that

$$\mathbb{P}\Big\{\mathbb{P}\big[\tau_{ik} \in \mathcal{I}_k, \text{ for all } 0 \leq k \leq L \,\big|\, \mathscr{H}\big] \geq 1 - \alpha\Big\} \geq 1 - \pi.$$

As described before, the uncertainty of the predicted TSUS effect $\widehat{\tau}_{ik}$ comes from the in-sample error $\mathsf{InErr}(\tau_{ik}) = -Y'_{\mathcal{N}(T_i+k)}(\widehat{\mathbf{w}}^{[i]} - \mathbf{w}^{[i]}_0)$ and the out-of-sample error $\mathsf{OutErr}(\tau_{ik}) = u_{i(T_i+k)}$.

Regarding the in-sample error, the following is an immediate generalization of the bound described in Section 4.1, which enjoys simultaneous coverage in multiple periods:

$$\underline{M}_{\mathtt{in}}(\tau_{ik}) = \underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2) \qquad \text{and} \qquad \overline{M}_{\mathtt{in}}(\tau_{ik}) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2), \tag{4.7}$$

26

where

$$\underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2) = (\alpha_{\mathtt{in}}/2)\text{-quantile of} \inf_{\boldsymbol{\delta}\in\mathcal{M}_{\mathbf{G}}^{[i]\star},\, 0\leq k\leq L} -\mathbf{Y}'_{\mathcal{N}(T_i+k)}\boldsymbol{\delta},$$

$$\overline{\mathfrak{c}}^\star(1-\alpha_{\mathtt{in}}/2) = (1-\alpha_{\mathtt{in}}/2)\text{-quantile of} \sup_{\boldsymbol{\delta}\in\mathcal{M}_{\mathbf{G}}^{[i]\star},\, 0\leq k\leq L} -\mathbf{Y}'_{\mathcal{N}(T_i+k)}\boldsymbol{\delta},$$

conditional on the data. It can shown that this generally guarantees that $[\underline{M}_{\mathtt{in}}(\tau_{ik}), \overline{M}_{\mathtt{in}}(\tau_{ik})]$ can simultaneously cover the in-sample error $\mathsf{InErr}(\tau_{ik})$ for all $0 \leq k \leq L$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1 - \alpha_{\mathtt{in}}$, up to some small loss.

Regarding the out-of-sample error, an easy-to-implement strategy analogous to that described in Section 4.2 is to adjust the bounds on the out-of-sample error $\mathsf{OutErr}(\tau_{ik}) = u_{i(T_i+k)}$ based on maximal inequalities. Recall that each $u_{i(T_i+k)} - \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}]$, $0 \leq k \leq L$, is conditionally sub-Gaussian with parameter $\sigma_{i(T_i+k)}$ (but is not necessarily independent over k). Then, for any $\lambda > 0$,

$$\mathbb{P}\Big(\max_{0\leq k\leq L} |u_{i(T_i+k)} - \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}]| \geq \lambda \Big|\mathscr{H}\Big) \leq 2\sum_{k=0}^{L} \exp\Big(-\frac{\lambda^2}{2\sigma_{i(T_i+k)}^2}\Big).$$

If $\sigma_{i(T_i+k)} \leq \sigma_i$ for all $0 \leq k \leq T - T_i$, then one can set $\underline{M}_{\mathtt{out}}(\tau_{ik}) = \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] - \lambda$ and $\overline{M}_{\mathtt{out}}(\tau_{ik}) = \mathbb{E}[u_{i(T_i+k)}|\mathscr{H}] + \lambda$ with $\lambda = \sqrt{2\sigma_i^2 \log(2(L+1)/\alpha_2)}$. It can be shown that $[\underline{M}_{\mathtt{out}}(\tau_{ik}), \overline{M}_{\mathtt{out}}(\tau_{ik})]$ can simultaneously cover $\mathsf{OutErr}(\tau_{ik})$ for all $0 \leq k \leq L$, with $\mathbb{P}(\cdot|\mathscr{H})$-probability at least $1 - \alpha_{\mathtt{out}}$. Compared with prediction intervals with validity for each period constructed similarly, these simultaneous prediction intervals are slightly wider due to the additional factor $\sqrt{\log(L+1)}$. In practice, assuming u_{it}'s are strictly sub-Gaussian, one only needs to estimate the conditional mean and variance of u_{it} using the pre-treatment residuals to implement the proposed bounds; flexible parametric or non-parametric estimation methods can be used.

Again, the sub-Gaussianity assumption can be relaxed using other concentration inequalities requiring weaker moment conditions, though the resulting simultaneous prediction intervals may be wider. Also, there are other strategies to construct prediction intervals that simultaneously cover multiple out-of-sample errors, though they are computationally more cumbersome and usually require more stringent conditions. See Section S.2.2 of the Supplemental Appendix for a brief discussion.

The idea outlined above to achieve simultaneous coverage is general and can also be used to,

for example, construct prediction intervals that simultaneously cover the TSUS predictands for multiple treated units rather than for multiple post-treatment periods. In our empirical application, we construct simultaneous prediction intervals for time-averaged effects across different economies; see details in Section 5.

5   Empirical Results

We use our framework to evaluate the effect of economic liberalization on (log) real GDP per capita in the sixteen countries that adopted economic liberalization in Sub-Saharan Africa: Benin, Botswana, Cabo Verde, Cameroon, Gambia, Ghana, Guinea, Guinea, Guinea-Bissau, Ivory Coast, Mali, Mauritius, Niger, South Africa, Uganda, and Zambia.

We emphasize that when the target predictand does not involve an average across countries (e.g., TSUS and TAUS), we include not-yet-treated countries in the donor pool. Including not-yet-treated units in other scenarios would raise issues, as their pre-treatment features would appear both in the conditioning set $\mathscr{H}$ and in the outcome vector, thereby invalidating the attempt of quantifying their contribution to prediction uncertainty.

Overall, our point predictions suggest that (external) liberalization episodes in Sub-Saharan Africa had no impact on GDP per capita. Although we find that the point prediction for Botswana is large and positive, our predictands and uncertainty characterization indicate that this predicted effect cannot be distinguished from zero. Our new evidence thus suggests that Botswana's success story was possibly the cause rather than the consequence of economic liberalization. In what follows, we report the predictands described in Section 3, Examples TSUS, TAUS, and TSUA.

In all our results, we compute the weights using an L1-L2 constraint and use pre-treatment GDP as the only feature. In the Supplemental Appendix Section S.8, we also report the results using (i) a simplex-type constraint, (ii) a ridge-type constraint, (iii) more than one feature, (iv) using placebo treatment dates, and (v) leaving one of the donors out at a time. All the implementation details are reported in Supplemental Appendix Section S.7.

TSUS predicted effects in every period after liberalization (τ_{ik}). We first analyze the predicted treatment effect for every individual country that adopts liberalization in each post-liberalization period (up to five years after adoption), which is an example of the TSUS predictand. In Figure 3,

we show the predicted synthetic control outcomes (panel (a)) with the corresponding 90% prediction intervals, and the estimated weights $\widehat{\mathbf{w}}^{[i]}, i \in \mathcal{E}$ (panel (b)) for our sixteen countries.

Figure 3: *Time-specific unit-specific (TSUS) predicted effects in every period, $\widehat{\tau}_{ik}$.*
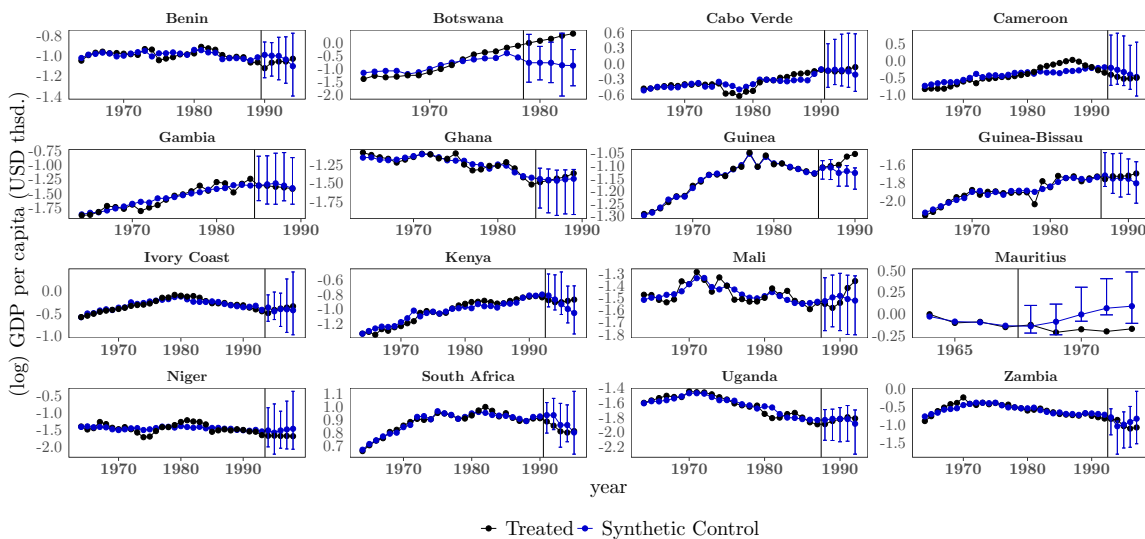
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$
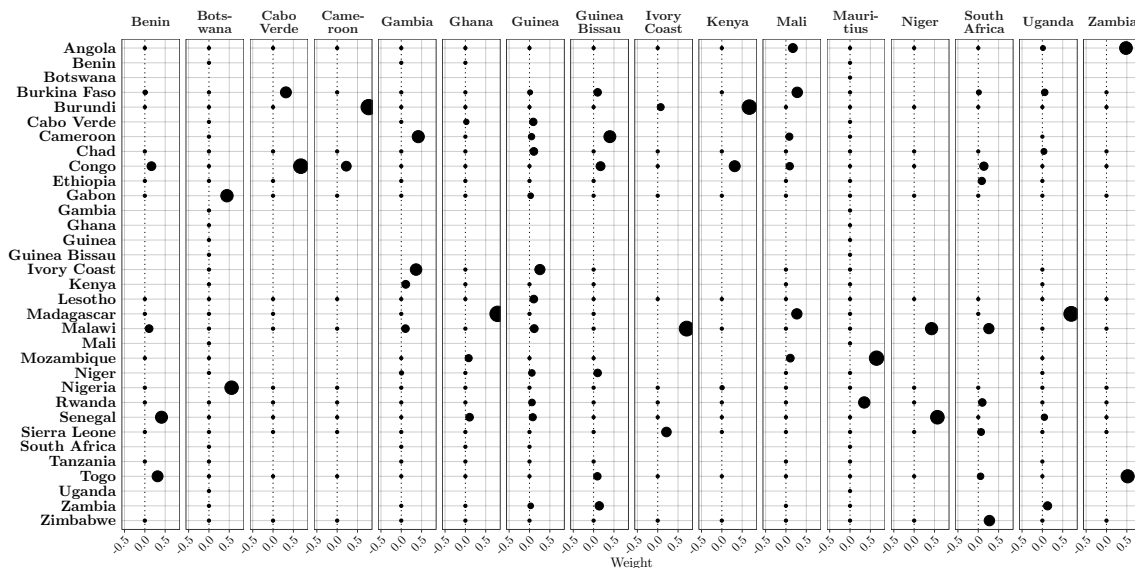
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

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.

In most countries, the realized trajectory of GDP per capita (black lines) lies roughly on top of the synthetic one (blue lines), suggesting that in the absence of the liberalization event, real income per capita would have been approximately equal. The 90% prediction intervals (blue vertical bars) indicate that, in almost all cases, the distance between the actual GDP series and the counterfactual one is not different from zero with high probability for almost all units and periods.

For Botswana, Guinea, and Mauritius the realized trajectories of GDP per capita (black lines) lay above the synthetic ones, suggesting a positive treatment effect. The 90% prediction intervals show that this effect can be distinguished from zero with high probability after three years from the liberalization (with the exception of Botswana after four years), whereas it cannot in the first two years. We stress that the pre-treatment fit for Botswana is poorer than the one for Mauritius and Guinea. In Section S.8 of the Supplemental Appendix, we show that these findings are robust when a simplex-type constraint is employed, but not with a ridge-type constraint or when a second feature (the investment-to-GDP ratio) is included.

TAUS predicted effects, averaged over five years $(\tau_{i\cdot})$. The second causal predictand of interest is the effect for each of the sixteen African countries we study, averaged over the five periods following the liberalization treatment in each country (up to the year 2000). This is an example of the TAUS predictand. Figure 4 shows that in almost all countries the liberalization episode seems to have a negligible effect on real GDP per capita.

By looking at the prediction intervals, we can see that 13 treated units show an average effect that cannot be distinguished from zero with high probability. On the other hand, for Botswana, Guinea, and Mauritius we find a positive TAUS which is also statistically significant with high probability. Again, we stress the poor pre-treatment fit in interpreting the results for Botswana. Moreover, in Section S.8 of the Supplemental Appendix, we show that employing a simplex-type constraint yields similar results, although these results are not robust when a second feature (investment-to-GDP ratio) is introduced or when a ridge-type constraint is used.

TSUA predicted effects, averaged over countries that liberalized in each of three waves: before 1987, between 1987 and 1991, and after 1991 $(\tau_{\mathcal{Q}_1 k}, \tau_{\mathcal{Q}_2 k}, \tau_{\mathcal{Q}_3 k})$. In interpreting the evidence for each individual country, Billmeier and Nannicini (2013) consider the hypothesis that liberalization only led to economic growth for countries that liberalized early. Our framework allows us to group

Figure 4: *Time-averaged unit-specific (TAUS) predicted effects, averaged over five years, $\hat{\tau}_{i\cdot}$.*
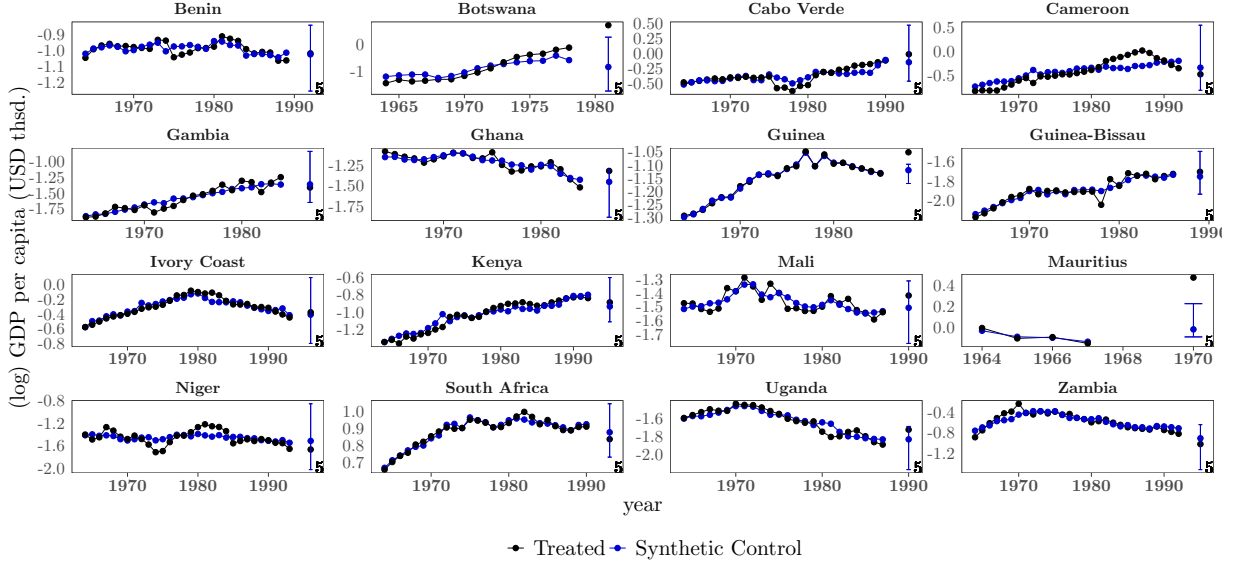
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$
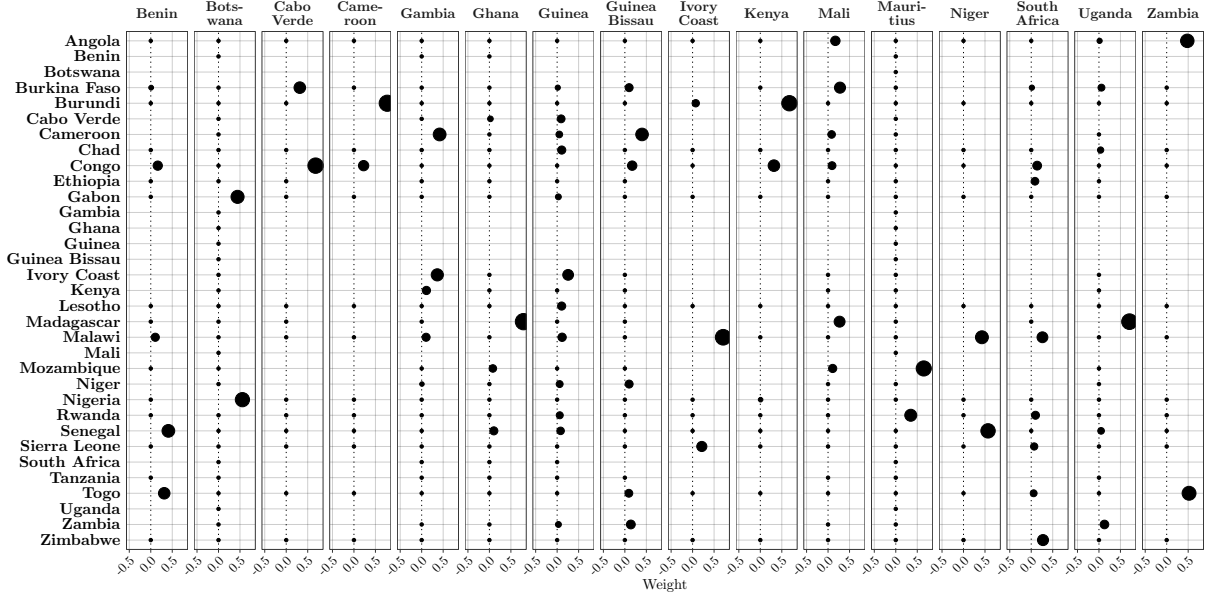
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

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.

countries according to the era in which they liberalized and consider their joint trajectory in a formal way. To do so, we use the three waves considered by Billmeier and Nannicini (2013) and study three different predictands that average over all countries that liberalized during each wave: the TSUA effect for all countries that liberalized before 1986 (Botswana, Gambia, Ghana, and Guinea), $\tau_{Q_1 k}$ with $Q_1 = \{t : t < 1987\}$; the TSUA effect for all countries that liberalized between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), $\tau_{Q_2 k}$ with $Q_2 = \{t : 1987 \leq t \leq 1991\}$; and the TSUA effect for all countries that liberalized between 1992 and 1994 (Burkina Faso, Burundi, Cameroon, Ethiopia, Ivory Coast, Mozambique, Niger, Tanzania, and Zambia), $\tau_{Q_3 k}$ with $Q_3 = \{t : 1991 < t \leq 1994\}$. The predicted effect is calculated for every post-liberalization period (up to five years), each representing a specific example of the TSUA predictand.

The results, presented in Figure 5, show that, with high probability and when joint prediction intervals are considered, we can conclude that countries that liberalized before 1987 had a positive TSUA predicted effect. We caution the reader in interpreting this evidence because of the poor pre-treatment fit in Figure 5(a). The quality of the pre-treatment fit follows from the small cardinality of the donor group for this predictand, given the exclusion of not-yet-treated countries. These findings are robust to the inclusion of a second feature and if a simplex-type constraint is employed (see Section S.8 of the Supplemental Appendix). However, relying on a ridge-type constraint (Section S.8.1 of the Supplemental Appendix), which greatly improves the pre-treatment fit and hence is our preferred specification for this predictand, shows that, with high probability, we can conclude that countries that liberalized enjoyed a positive TSUA effect after 5 years in all three waves, thus not supporting the original hypothesis of Billmeier and Nannicini (2013).

TSUA predicted effects, averaged over all liberalized countries ($\tau_{\mathcal{E} k}$). In this last and fourth exercise, we focus on a popular causal predictand: the effect in every period after treatment averaged over all the treated countries. This is yet another example of the TSUA predictand, with $Q = \mathcal{E}$. This predictand thus averages over all countries that liberalized, regardless of when they did so. We report this predictand in every year after the adoption of liberalization (up to five years) which occurs at different times for different countries. Figure 6 reports the results. To compute this predictand, we first construct a synthetic control for each treated unit (see panel (b)) and then we

Figure 5: *Time-specific unit-averaged (TSUA) predicted effects in each period, averaged over three groups of countries.*

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

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

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

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

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

(f) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$
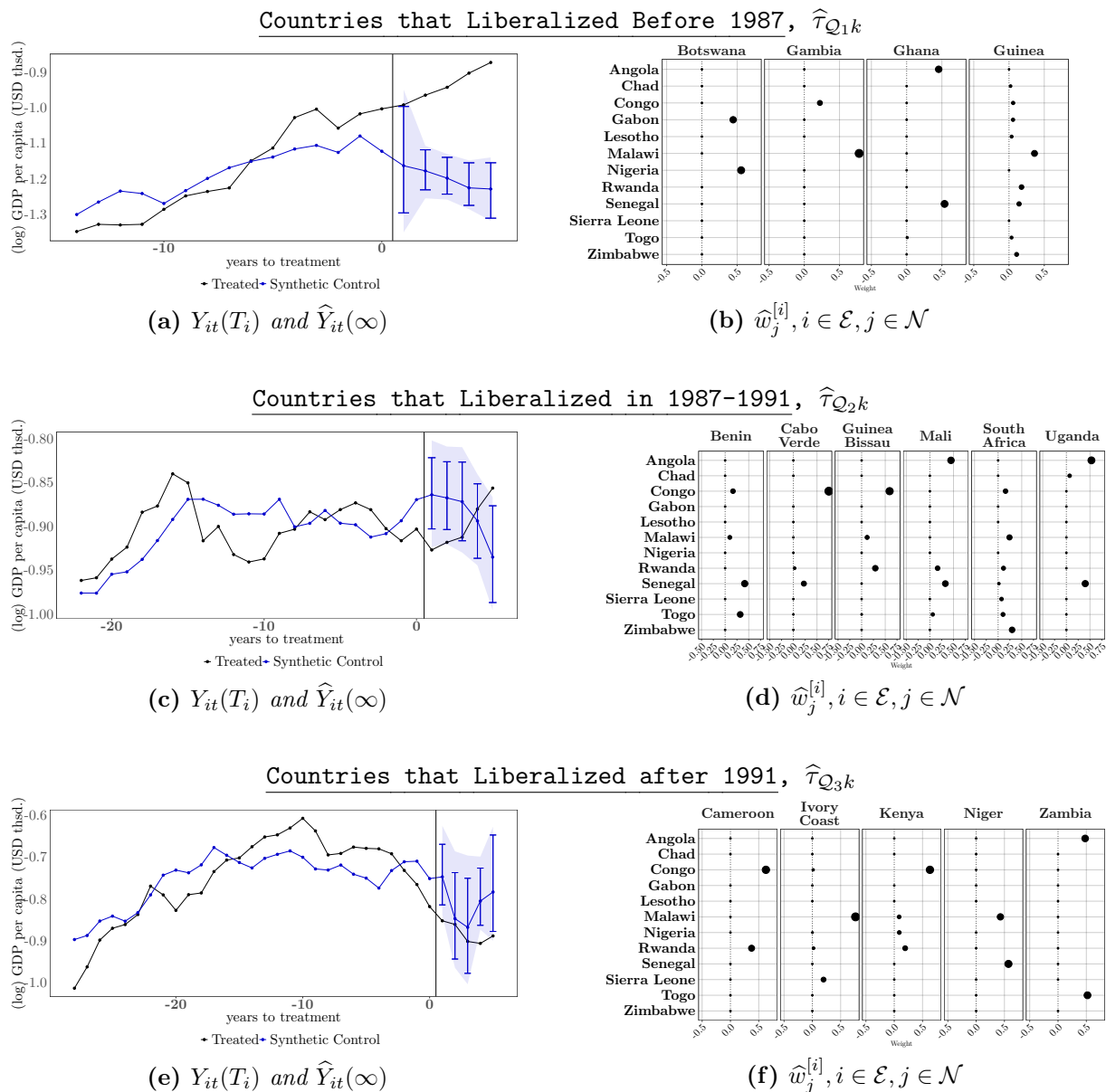
Notes: TSUA prediction in every period after treatment (up to five years), averaged over all countries that liberalized in each of three waves: before 1987 (Botswana, Gambia, Ghana, and Guinea), between 1987 and 1991 (Benin, Cabo Verde, Guinea-Bissau, Mali, South Africa, and Uganda), and after 1991 (Burkina 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.

pool all the synthetic controls together in every post-treatment period to get a single prediction and a single prediction interval. Panel (a) and panel (b) show that pooling across the 16 African countries that embarked on liberalization programs helps reduce the uncertainty surrounding the synthetic trajectory. This TSUA predictand shows that, with high probability, the liberalization had a statistically significant impact on GDP-per-capita after two years. This result is robust to both simplex- and ridge-type constraints (this latter also with a better pre-treatment fit), but not to the inclusion of a second feature (see Section S.8 of the Supplemental Appendix).

6   Theoretical Foundations

This section presents a general framework that can accommodate more flexible specifications in SC analysis. As in Section 3, 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 $j = 1, \ldots, J_0$, $\mathbf{B}_{j,l}^{[i]} = (b_{j1,l}^{[i]}, \cdots, b_{jT_{i0},l}^{[i]})'$ 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 $k = 1, \ldots, K$, $\mathbf{C}_{k,l}^{[i]} = (c_{k1,l}^{[i]}, \cdots, c_{kT_{i0},l}^{[i]})'$ is another vector of control variables used to predict $\mathbf{A}_l^{[i]}$ over the same pre-intervention time span. Stacking the M equations (corresponding to M features) for each treated unit, we define

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

For instance, in the Supplemental Appendix 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

Figure 6: *Time-specific unit-averaged (TSUA) predicted effect, averaged over all treated units,* $\widehat{\tau}_{\mathcal{E}k}$.
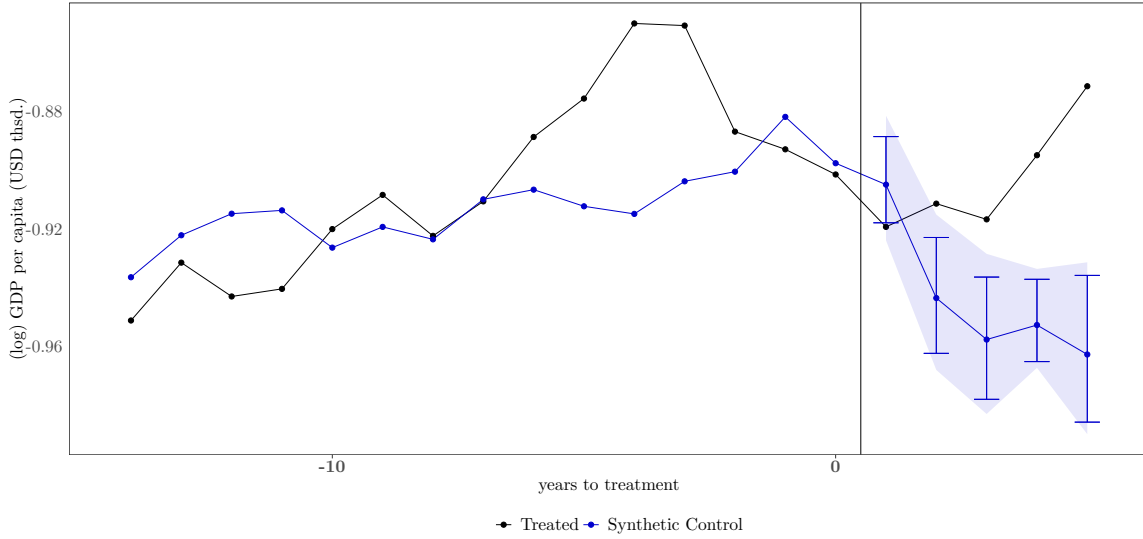
(a) $Y_{it}(T_i)$ *and* $\widehat{Y}_{it}(\infty)$
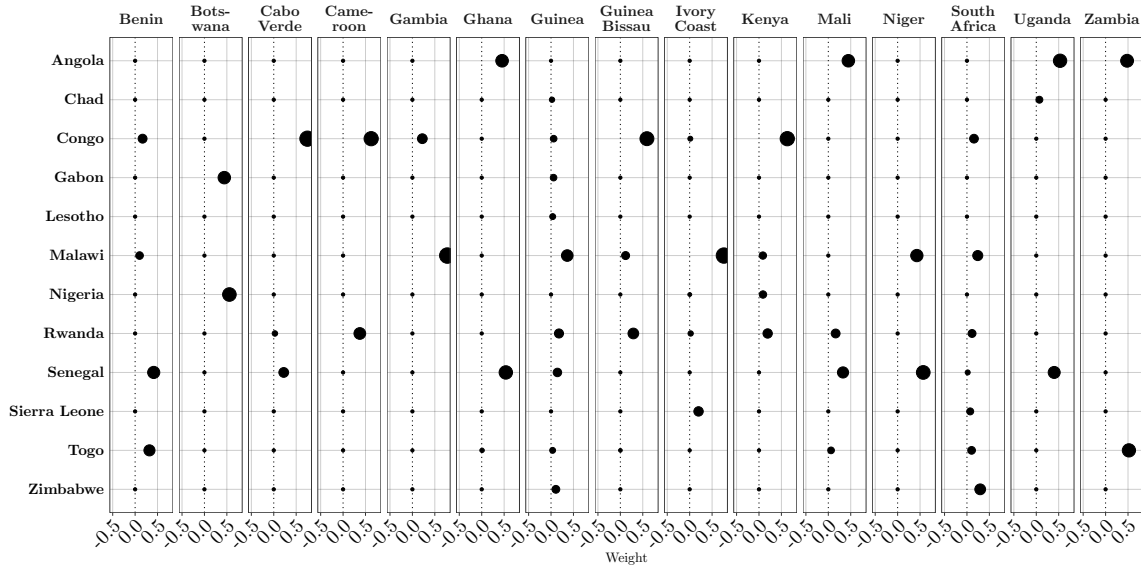
(b) $\widehat{w}_j^{[i]}, i \in \mathcal{E}, j \in \mathcal{N}$

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.

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 $l = 1, 2$, $\mathbf{C}_l^{[i]}$ 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 $\mathbf{r} = (\mathbf{r}^{[1]\prime}, \cdots, \mathbf{r}^{[J_1]\prime})' \in \mathcal{R} \subseteq \mathbb{R}^{KM J_1}$, 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 (3.1), 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 $\tilde{T} = \sum_{i=1}^{J_1} T_{0i}$,

$$\widehat{\boldsymbol{\beta}} := (\widehat{\mathbf{w}}', \, \widehat{\mathbf{r}}')' \in \underset{\mathbf{w} \in \mathcal{W}, \, \mathbf{r} \in \mathcal{R}}{\arg\min} \, (\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r})' \mathbf{V} (\mathbf{A} - \mathbf{B}\mathbf{w} - \mathbf{C}\mathbf{r}) \tag{6.1}$$

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 $\widehat{\mathbf{w}}^{[i]} = (\widehat{w}_1^{[i]}, \cdots, \widehat{w}_{J_0}^{[i]})'$ 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 3 (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

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

where $\mathbf{B}_l^{[i]} := (\mathbf{b}_{1,l}^{[i]}, \cdots, \mathbf{b}_{T_0,l}^{[i]})'$ is the l-th feature of the J_0 donor units, and $\mathbf{C}_l^{[i]} := (\mathbf{c}_{1,l}^{[i]}, \cdots, \mathbf{c}_{T_0,l}^{[i]})'$ is the additional K variables used to predict $\mathbf{A}_l^{[i]}$. 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,

$$\widehat{\boldsymbol{\beta}} = \underset{\mathbf{w} \in \mathcal{W}, \mathbf{r} \in \mathcal{R}}{\arg\min} \sum_{l=1}^{M}\sum_{t=1}^{T_0}\left[\frac{1}{J_1}\sum_{i=1}^{J_1}\left(a_{t,l}^{[i]} - \mathbf{b}_{t,l}^{[i]'}\mathbf{w}^{[i]} - \mathbf{c}_{t,l}^{[i]'}\mathbf{r}_l^{[i]}\right)\right]^2.$$

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

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

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 Section 3. Then, the prediction of τ can be constructed accordingly and uniformly expressed as

$$\widehat{\tau} = \mathsf{L}(\{Y_{it}\}) - \mathbf{p}'_\tau\widehat{\boldsymbol{\beta}}, \tag{6.2}$$

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:

$$\mathbf{p}_{\tau_{ik}} = (\underbrace{\mathbf{0}'_{J_0+KM},\cdots,\mathbf{0}'_{J_0+KM}}_{(i-1)\text{ vectors}}, \mathbf{p}_{T_i+k}^{[i]'}, \underbrace{\mathbf{0}'_{J_0+KM},\cdots,\mathbf{0}'_{J_0+KM}}_{(J_1-i)\text{ vectors}})',$$

$$\mathbf{p}_{\tau_{i\cdot}} = \left(\underbrace{\mathbf{0}'_{J_0+KM},\cdots,\mathbf{0}'_{J_0+KM}}_{(i-1)\text{ vectors}}, \frac{1}{T-T_i+1}\sum_{t\geq T_i}\mathbf{p}_t^{[i]'}, \underbrace{\mathbf{0}'_{J_0+KM},\cdots,\mathbf{0}'_{J_0+KM}}_{(J_1-i)\text{ vectors}}\right)',$$

$$\mathbf{p}_{\tau_{\mathcal{Q}k}} = \left(\frac{1}{Q}\mathbf{p}_{T_1+k}^{[1]'}\mathbb{1}(T_1 \in \mathcal{Q}), \frac{1}{Q}\mathbf{p}_{T_2+k}^{[2]'}\mathbb{1}(T_2 \in \mathcal{Q}), \cdots, \frac{1}{Q}\mathbf{p}_{T_{J_1}+k}^{[J_1]'}\mathbb{1}(T_{J_1} \in \mathcal{Q})\right)', \quad \text{and}$$

$$\mathbf{p}_{\tau_{\cdot\cdot}} = \left(\frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_1+k}^{[1]'}, \frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_2+k}^{[2]'}, \cdots, \frac{1}{LJ_1}\sum_{k=1}^{L}\mathbf{p}_{T_{J_1}+k}^{[J_1]'}\right)',$$

37

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

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

$$\boldsymbol{\beta}_0 := (\mathbf{w}_0', \, \mathbf{r}_0')' = \underset{\mathbf{w}\in\mathcal{W}, \, \mathbf{r}\in\mathcal{R}}{\arg\min} \, \mathbb{E}\Big[(\mathbf{A} - \mathbf{Bw} - \mathbf{Cr})'\mathbf{V}(\mathbf{A} - \mathbf{Bw} - \mathbf{Cr})\Big| \mathscr{H}\Big], \qquad (6.3)$$

and then we can write

$$\mathbf{A} = \mathbf{Bw}_0 + \mathbf{Cr}_0 + \mathbf{U}, \qquad \mathbf{w}_0 \in \mathcal{W}, \qquad \mathbf{r}_0 \in \mathcal{R}, \qquad (6.4)$$

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 Section 4.1 can be generalized to this setup. Specifically, let $d_\beta = J_0 + KM$, $\mathbf{Z} = (\mathbf{B}, \mathbf{C})$, $\widehat{\mathbf{Q}} = \mathbf{Z}'\mathbf{VZ}$, $\widehat{\boldsymbol{\gamma}}' = \mathbf{U}'\mathbf{VZ}$, $\boldsymbol{\gamma} = \mathbb{E}[\widehat{\boldsymbol{\gamma}}|\mathscr{H}]$, 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 (6.2), $\mathsf{InErr}(\tau)$ 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 (4.4), 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 1 below) are available, we can set

$$\underline{M}_{\mathtt{in}}(\tau) = \underline{\mathfrak{c}}^\star(\alpha_{\mathtt{in}}/2) \quad \text{and} \quad \overline{M}_{\mathtt{in}}(\tau) = \overline{\mathfrak{c}}^\star(1 - \alpha_{\mathtt{in}}/2) \qquad (6.5)$$

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

For the out-of-sample error, analogously to (4.6), if $\mathsf{OutErr}(\tau)$ is assumed to be sub-Gaussian

38

conditional on $\mathscr{H}$ with parameter $\sigma_{\mathscr{H}}$, then we can set

$$\underline{M}_{\text{out}}(\tau) = \mathbb{E}[\text{OutErr}(\tau)|\mathscr{H}] - \sqrt{2\sigma_{\mathscr{H}}^2 \log(2/\alpha_{\text{out}})} \quad \text{and}$$
$$\overline{M}_{\text{out}}(\tau) = \mathbb{E}[\text{OutErr}(\tau)|\mathscr{H}] + \sqrt{2\sigma_{\mathscr{H}}^2 \log(2/\alpha_{\text{out}})}. \tag{6.6}$$

We present a general theorem that justifies the above method under high-level conditions and covers the results given in Section 4 as a special case. Let $\|\cdot\|_{\text{F}}$ 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 1. *Assume $\mathcal{W}$ and $\mathcal{R}$ are convex, $\widehat{\boldsymbol{\beta}}$ in (6.1) and $\boldsymbol{\beta}_0$ in (6.3) exist, $\mathscr{H} = \sigma(\mathbf{B}, \mathbf{C}, \mathbf{p}_\tau)$, and $\underline{M}_{\text{in}}(\tau), \overline{M}_{\text{in}}(\tau), \underline{M}_{\text{out}}(\tau)$ and $\overline{M}_{\text{out}}(\tau)$ are specified as in (6.5) and (6.6). In addition, for some finite non-negative constants ϵ_γ, π_γ, $\varpi_\delta^\star$, $\epsilon_\delta^\star$, $\pi_\delta^\star$, $\epsilon_\Delta^\star$, $\pi_\Delta^\star$, $\epsilon_{\gamma,1}^\star$, $\epsilon_{\gamma,2}^\star$ and $\pi_\gamma^\star$, the following conditions hold:*

(i) $\mathbb{P}[\mathbb{P}(-\mathbf{p}_\tau'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \in [\mathfrak{c}(\alpha_0), \bar{\mathfrak{c}}(1-\alpha_0)]|\mathscr{H}) \geq 1 - 2\alpha_0 - \epsilon_\gamma] \geq 1 - \pi_\gamma$ *for any* $\alpha_0 \in (0,1)$;

(ii) $\mathbb{P}[\mathbb{P}(\sup\{\|\boldsymbol{\delta}\| : \boldsymbol{\delta} \in \mathcal{M}_{\mathbf{G}}\} \leq \varpi_\delta^\star|\mathscr{H}) \geq 1 - \epsilon_\delta^\star] \geq 1 - \pi_\delta^\star$;

(iii) $\mathbb{P}[\mathbb{P}(\Delta \cap \mathcal{B}(\mathbf{0}, \varpi_\delta^\star) \subseteq \Delta^\star |\mathscr{H}) \geq 1 - \epsilon_\Delta^\star] \geq 1 - \pi_\Delta^\star$;

(iv) $\mathbb{P}[\mathbb{P}(\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_{d_\beta}\|_{\text{F}} \leq 2\epsilon_{\gamma,1}^\star|\mathscr{H}) \geq 1 - \epsilon_{\gamma,2}^\star] \geq 1 - \pi_\gamma^\star$;

(v) $\text{OutErr}(\tau) - \mathbb{E}[\text{OutErr}(\tau)|\mathscr{H}]$ *is sub-Gaussian conditional on $\mathscr{H}$ with parameter $\sigma_{\mathscr{H}}$.*

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

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

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

Assumptions (i)–(iv) imposed in Theorem 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 Section 4. We give more detailed discussion of each condition in Section 6.1 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.

6.1 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 that accommodates non-stationary data and is applicable to our empirical application. Lemma S.1 in the Supplemental Appendix 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

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

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

$$\widehat{\Delta} = \left\{ \boldsymbol{\beta} - \widehat{\boldsymbol{\beta}} : \boldsymbol{m}_=(\boldsymbol{\beta}) = \mathbf{0},\ m_{\leq,\ell}(\boldsymbol{\beta}) \leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}})\mathbb{1}(\ell \in \mathcal{A}),\ \ell = 1,\cdots,d_\leq \right\}. \qquad (6.7)$$

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 1. *Assume that with probability over $\mathscr{H}$ at least $1 - \pi_\Delta^\star$, the following conditions hold: (i)* $\mathbb{P}(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_2 \leq \varpi_\delta^\star | \mathscr{H}) \geq 1 - \epsilon_\Delta^\star$; *(ii)* $\boldsymbol{m}(\cdot) = (\boldsymbol{m}_=(\cdot)', \boldsymbol{m}_\leq(\cdot)')'$ *is twice continuously differentiable on* $\mathcal{B}(\boldsymbol{\beta}_0, \varpi_\delta^\star)$ *with* $\inf_{\boldsymbol{\beta} \in \mathcal{B}(\boldsymbol{\beta}_0, \varpi_\delta^\star)} s_{\min}(\frac{\partial}{\partial \boldsymbol{\beta}'} \boldsymbol{m}(\boldsymbol{\beta})) \geq c_{\min}$ *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,* $\varrho_\ell > \mathfrak{c}\varpi_\delta^\star$, *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 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 Section 4.1 is an application of this theoretical result, as detailed in Section S.5.2.2 of the Supplemental Appendix. 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 6.2 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.

6.2 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

$$\varrho_\ell = \left\| \frac{\partial}{\partial \boldsymbol{\beta}} m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) \right\|_2 \times \varrho^{[i]}. \tag{6.8}$$

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

$$\mathcal{C} = \frac{\sqrt{d_{\beta,0} \log(d_\beta) \log(T_0)} \max_{1 \leq j \leq J_0} \widehat{\sigma}_{b_j} \widehat{\sigma}_u}{\min_{1 \leq j \leq J_0} \widehat{\sigma}_{b_j}^2}, \tag{6.9}$$

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 (6.9). See Supplemental Appendix Section S.5.1 for more explanation. Moreover, the constant in (6.9) 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.5.1 of the Supplemental Appendix.

Then, in the simulation, we impose

$$m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) + \frac{1}{2} s_{\max}\left(\frac{\partial}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]}) \right) \times (\varrho^{[i]})^2 \tag{6.10}$$

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 1, which typically depends on the second-order expansion of the constraint function around the true values. See Section S.5.2.2 of the Supplemental Appendix 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

42

$m_{\leq,\ell}(\boldsymbol{\beta}^{[i]}) \leq m_{\leq,\ell}(\widehat{\boldsymbol{\beta}}^{[i]})$. Thus, the adjustment due to the second term on the right-hand side of (6.10) is only necessary for nonlinear constraints. This coincides with the result in Lemma 1. In the special case of the L2 constraint used in our empirical application, (6.10) 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.

7   Conclusion

We developed prediction intervals to quantify the uncertainty of a large class of synthetic control predictions (or estimators) in settings with staggered treatment adoption. Because many synthetic control applications have a limited number of observations, our inference procedures are based on non-asymptotic concentration arguments. The construction of our prediction intervals is designed to capture two sources of uncertainty: the first is the construction or estimation of the synthetic control weights with pre-treatment data, and the second is the variability of the post-treatment outcomes. By combining both sources in a prediction interval, our procedure offers precise non-asymptotic coverage probability guarantees and allows researchers to implement sensitivity analyses to assess how robust the conclusions of the analysis are to various levels of uncertainty. Our framework is general, allowing for one or multiple treated units, simultaneous or staggered treatment adoption, linear or non-linear constraints, and stationary or non-stationary data. To enhance implementation, we also showed how to recast the methods as conic optimization programs and how to choose the necessary tuning parameters in a principled data-driven way. We illustrated our methods with an empirical application studying the effect of economic liberalization on real GDP per capita in Sub-Saharan African countries, motivated by the work of Billmeier and Nannicini (2013).

All our methods are implemented in `Python`, `R`, and `Stata` software, which is publicly available (https://nppackages.github.io/scpi/), and discussed in detail in our companion article Cattaneo, Feng, Palomba and Titiunik (2025) and in Section S.7 of the Supplemental Appendix.

References

Abadie, A. (2021), "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, 59, 391–425.

Abadie, A., and Cattaneo, M. D. (2018), "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505.

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132.

Agarwal, A., Shah, D., Shen, D., and Song, D. (2021), "On Robustness of Principal Component Regression," *Journal of the American Statistical Association*, 116, 1731–1745.

Arbel, J., Marchal, O., and Nguyen, H. D. (2020), "On Strict Sub-Gaussianity, Optimal Proxy Variance and Symmetry for Bounded Random Variables," *ESAIM: Probability and Statistics*, 24, 39–55.

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.

Bhagwati, J., and Srinivasan, T. N. (2001), "Outward-Orientation and Development: Are Revisionists Right?" in *Trade, Development and Political Economy*, Springer, pp. 3–26.

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

Bobkov, S., Chistyakov, G., and Götze, F. (2024), "Strictly Subgaussian Probability Distributions," *Electronic Journal of Probability*, 29, 1–28.

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.

Callaghy, T. M., and Ravenhill, J. (1994), *Hemmed In: Responses to Africa's Economic Decline*, Columbia 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.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021a), "Distributional Conformal Prediction," *Proceedings of the National Academy of Sciences*, 118, e2107794118.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021b), "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 116, 1849–1864.

DeJong, D. N., and Ripoll, M. (2006), "Tariffs and Growth: An Empirical Exploration of Contingent Relationships," *Review of Economics and Statistics*, 88, 625–640.

Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (Forthcoming), "Visualization, identification, and estimation in the linear panel event-study design," *Advances in Economics and Econometrics: Twelfth World Congress*.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019), "Pre-event trends in the panel event-study design," *American Economic Review*, 109, 3307–3338.

Giavazzi, F., and Tabellini, G. (2005), "Economic and Political Liberalizations," *Journal of Monetary Economics*, 52, 1297–1330.

Levine, R., and Renelt, D. (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, 82, 942–963.

Li, K. T. (2020), "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 115, 2068–2083.

Masini, R., and Medeiros, M. C. (2021), "Counterfactual Analysis with Artificial Controls: Inference, High Dimensions and Nonstationarity," *Journal of the American Statistical Association*, 116, 1773–1788.

Miller, D. L. (2023), "An Introductory Guide to Event Study Models," *Journal of Economic Perspectives*, 37, 203–230.

Powell, D. (2022), "Synthetic Control Estimation Beyond Comparative Case Studies: Does the Minimum Wage Reduce Employment?" *Journal of Business & Economic Statistics*, 40, 1302–1314.

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.

Shaikh, A. M., and Toulis, P. (2021), "Randomization Tests in Observational Studies with Staggered Adoption of Treatment," *Journal of the American Statistical Association*, 116, 1835–1848.

Shi, X., Miao, W., Hu, M., and Tchetgen, E. T. (2023), "Theory for Identification and Inference with Synthetic Controls: A Proximal Causal Inference Framework," arXiv:2108.13935.

Vovk, V. (2012), "Conditional Validity of Inductive Conformal Predictors," in *JMLR: Workshop and Conference Proceedings 25, Asian Conference on Machine Learning*, pp. 475–490.

Wacziarg, R., and Welch, K. H. (2008), "Trade Liberalization and Growth: New Evidence," *World Bank Economic Review*, 22, 187–231.

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