DOI: 10.1111/joes.12493

JOURNAL OF ECONOMIC SURVEYS WILEY

ARTICLE

Synthetic Control Method: A tool for comparative case studies in economic history

David Gilchrist[1] | **Thomas Emery**[2] | **Nuno Garoupa**[3] | **Rok Spruk**[4]

[1] Business School, University of Western Australia, Crawley, WA, Australia

[2] University of Western Australia, Crawley, Australia

[3] George Mason University, Fairfax, VA, USA

[4] University of Ljubljana, Ljubljana, Slovenia

Correspondence
David Gilchrist, Business School, University of Western Australia, 8716 Hackett Dr, Crawley WA 6009, Australia.
Email: david.gilchrist@uwa.edu.au

Abstract

The Synthetic Control Method (SCM) has become a widely used tool in both identifying and estimating the causal impact of policies, shocks, and interventions of interest on economic and social outcomes. The technique has become particularly popular in estimating the effect of these shocks on a single treated unit. As a transparent and data-driven statistical technique, the goal of the SCM is to construct an artificial control group for the treated unit that has similar pretreatment characteristics but has not undergone the treatment itself thus developing a plausible counterfactual against which impacts resulting from structural changes can be evaluated as part of a historical investigation. The method works well when the control group balances pre-intervention outcomes and auxiliary covariates as much as possible. In spite of its widespread adoption, the use of the SCM in comparative economic history has lagged behind other areas of economics. In this article, we critically review the properties of the SCM and discuss the necessary conditions for a plausible application of the technique to comparative economic history in support of research designed to answer some of the long-running historical questions and demonstrate the potential to use SCM in comparative economic history studies by estimating the impact of the oil discovery in the 1920s on Venezuela's long-term economic growth.

KEYWORDS
comparative economics, economic history, Synthetic Control Method

JEL CLASSIFICATION
C12, C21, C33, N00, O43

1 INTRODUCTION

"Arguably the most important innovation in the evaluation literature in the last 15 years is the synthetic control approach developed by Abadie et al. (2010, 2015) and Abadie and Gardeazabal (2003). This method builds on difference-in-differences estimation, but uses arguably more attractive comparisons to get causal effects."

Susan Athey and Imbens paper Guido Imbens (Athey & Imbens, 2017**)**

Using the Synthetic Control Method (SCM) to evaluate the impact of a particular intervention (say, an event) on particular outcomes of interest (say, GDP growth – the "treated unit") has become a widely used tool in policy evaluation literature (Abadie & Gardeazabal, 2003; Abadie et al., 2010, 2015; Ben-Michael et al., 2021). In a broader sense, the method relies on pretreatment outcome matching (Imbens & Wooldridge, 2009) and uses pre-intervention trends in outcomes and auxiliary covariates to simulate a counterfactual outcome scenario reflective of the hypothetical absence of the intervention itself. The general thrust of the method is to construct the outcome of interest from an artificial control group of units that have never experienced the intervention, but which have similar outcome dynamics and covariate levels as those of the treated unit in the pre-intervention period. This is achieved by constructing a weighted average of control units that match the treated unit's outcome and auxiliary covariate values prior to the intervention. This is termed the "artificial unit," "synthetic unit," or "synthetic control group." If the past outcomes and covariates are reasonably well-balanced between the treated unit and its control group, the difference in the outcome between the treated unit and its artificial (i.e., synthetic) control group plausibly captures the impact of the policy or intervention in question.[1] In this respect, the idea of the SCM is to compute the gap in the given outcome of interest for the treated unit by assuming that the intervention or policy of interest had never happened thereby exploiting pre-intervention trends in the outcomes and covariates between the treated unit and its control group (McClelland & Gault, 2017). In this way, a researcher can obtain a meaningful representation of the intervention impact relative to the evolution of the outcome in the synthetic control group. Of course, the normal caveats on the use of counterfactuals in historiography apply. However, this method does provide a sound basis for identifying structural and other economic changes that may then be the focus of historical study.[2]

As such, the SCM can yield a plausible assessment of the impact of the policy or intervention in question provided that the imbalance in the pre-intervention outcomes and auxiliary covariates is minimized. The key limitation of the method pertains to the difficulty of achieving an exact balance between the treated and control units without producing bias (Ferman & Pinto, 2021; Ben Michael et al. 2021). Several empirical strategies to address the bias that arises from poor pre-intervention matching have been proposed (Abadie & Imbens, 2011; Doudchenko & Imbens, 2016; Li, 2020), including intending to achieve exact matching such as in the case of using an outcome model with a large pre-intervention period (Garoupa & Spruk, 2019), the imposition of negative weighting of the synthetic control group through calibrated propensity scores (Wang & Zubizarreta, 2020), the use of root-finding constrained optimization and the interior point method (Vanderbei, 1999), nonparametric construction of weights (Cerulli, 2019), the construction of prediction intervals to deal with the uncertainty of synthetic control estimates (Cattaneo et al., 2021), and the construction of empirical rejection probabilities to reject the null hypothesis (Firpo & Possebom, 2018).

In spite of its wide popularity in policy and impact evaluation studies, the use of the SCM to identify research directions in pursuit of answers to both short- and long-run questions in economic and institutional history has been scarce. Perhaps the most well-known example of applying the method to economic history was undertaken by Abadie et al. (2015). In this case, the authors study the impact of German reunification (i.e., Wiedervereinigung) on the economic growth of West Germany. Their approach is to exploit pre-unification trends in economic growth and its determinants to construct the synthetic control group of jurisdictions that best reproduces the pre-unification trajectory of West Germany but that did not experience unification. They use this to construct a counterfactual trajectory of economic growth in the hypothetical absence of reunification. Their results suggest that, in the absence of reunification, West Germany would most likely have achieved significantly higher ongoing economic growth possibly driven by the lack of the large fiscal redistribution from West Germany to the formerly East German territories in the hypothetical case.

Another famous, and perhaps more illuminative, example is that of Billmeier and Nannicini (2013). The authors study the effect of economic liberalization on economic growth in a large sample of countries. By comparing post-liberalization growth trajectories of the affected countries to the weighted trajectories of similar but untreated countries, they uncover the temporally and spatially heterogeneous impact of liberalization on growth. That is, liberalization episodes, which occurred early in a country's history, had a significant positive growth impact whereas more recent episodes did not.

However, to fill the void in the literature, we discuss both the importance and potential of the SCM for comparative case studies in economic history. To this end, we exploit the discovery of oil in Venezuela in the 1920s to estimate the impact of natural resources on the long-term economic performance. Drawing on a large dataset of countries from 1870 onwards, we are able to estimate the counterfactual scenario in the hypothetical absence of the oil discovery. In this respect, our aim is to replicate several existing studies on the resource curse and Dutch disease associated with the decline of manufacturing after the discovery of large-scale natural resources, and the subsequent effect on economic growth. Some of the studies in the resource curse literature include Boschini et al. (2007), Papyrakis and Gerlach (2004), Sachs and Warner (1995), and Brunnschweiler (2008), among many others (see Busse & Gröning, 2013).

Our evidence successfully replicates the findings advocating the fundamental importance of both institutional quality and strength in determining the balance between the blessing or curse posited by the discovery of large-scale natural resources (Robinson et al., 2006). By making use of the SCM, our results show that the discovery of oil in Venezuela yielded a largely positive but temporary deviation of the growth trajectory from the long-run equilibrium. In spite of the large and significant rapid growth acceleration, the growth premium associated with the discovery of oil began to flatten out commencing with the large-scale nationalization of an oil industry, and further deteriorated during the Chavez and Maduro years. Down to the present day, Venezuela's growth trajectory converged to the level of its synthetic control group where a similar oil shock is not perceivable.

As such, in this paper, we review the SCM and discuss both the scope and context of the applicability of the method in comparative economic history to identify and posit answers to some long-running historical questions. In doing so, we discuss the theoretical and empirical plausibility of the assumptions underlying the validity of the SCM, critically evaluate the state of the literature, and identify research opportunities we consider extant in economic history. We also comment on the weaknesses inherent in the approach.

The paper is organized as follows: Section 2 discusses the prior literature; Section 3 provides the key analytical narratives of the synthetic control analysis; Section 4 discusses the applications of the method in comparative economic history, and adds the case of the oil discovery of Venezuela to the extant literature, including the acknowledgment of limitations inherent in the method; Section 5 provides general discussion, and Section 6 concludes.

2 | PRIOR LITERATURE

While the application of the SCM in comparative economic history has been minimal, the SCM has gained significant momentum in the scholarly literature and has been applied in several different settings in a wide range of applications in economics and political science. These applications include evaluating the impact of: terrorism, civil wars, and political risk (Abadie & Gardeazabal, 2003; Bove et al., 2017; Montalvo, 2011; Yu & Wang, 2013); natural disasters (Barone & Mocetti, 2014; Cavallo et al., 2013; Coffman & Noy, 2012; Smith, 2015); economic and trade liberalization (Billmeier & Nannicni, 2013; Gathani et al., 2013; Hosny, 2012); health policy (Abadie et al., 2010; Bauhoff, 2014; Kreif et al., 2016; Spruk & Kovac, 2020); organized crime (Becker & Klößner, 2017; Donohue et al., 2019; Pinotti, 2015; Rydberg et al., 2018; Saunders et al., 2015), political reforms and regime changes (Abadie et al., 2015; Billmeier & Nannicini, 2009; Campos et al., 2019; Carrasco et al., 2014; Garcia Ribeiro et al., 2013; Spruk, 2019); social and political connections (Acemoglu et al., 2016); labor (Bohn et al., 2014; Calderon, 2014; De Souza, 2014); and local development policies (Ando, 2015; Gobillon & Magnac, 2016; Kirkpatrick & Bennear, 2014; Possebom, 2017), among several others.

As described above, the SCM has been designed for comparative case studies in small and moderately sized samples as a transparent and data-driven way of constructing a synthetic control group for use in comparing the outcome of interest in an affected unit with the outcome of that control group without the direct exposure of that control group to the intervention of interest (Becker & Klößner, 2018). Data-driven processes of constructing the synthetic control group are based on selecting weights of past outcomes and auxiliary covariates through the diagonal matrix comparison that determines which comparison units' outcome process characteristics are within the convex hull of the affected unit to best reproduce the outcome trajectory for the treated unit in the hypothetical absence of the intervention (Kaul et al., 2021)

In relation to the minimal application of the SCM to comparative economic history, examples include Campos et al. (2019) who examine the contribution of institutional integration to economic growth and use the SCM to assess the growth effects of European Union membership for non-founding states. They find heterogeneous and largely positive effects of EU membership on growth. By simulating the economic growth trajectory without EU membership, they use a sample of non-EU countries as a donor pool to construct the country-level synthetic control groups that best capture pre-EU growth trajectories of non-founding EU members. The reported and somewhat positive effect of being an EU member on growth appears to be robust across more than 10,000 randomly generated donor samples. Somewhat intriguingly, they find that Greece is the only exception to the positive effects since its likely growth trajectory without EU membership appears to be better than the actual growth trajectory with the EU membership, although no clear consensus has been reached in the literature (Garoupa & Spruk, 2020).

Another prominent example of the use of the SCM to answer important historical questions comes from Grier and Maynard (2016) who examine the economic growth and development of the Venezuelan economy during the Chavez administration. A controversial figure, Chavez won three democratic elections, rewrote the Venezuelan constitution, restructured its Supreme Court, and survived numerous coup and recall attempts. The proponents of *Chavismo* argue that poverty rates dropped massively during his term in office alongside reduced inequality and the rapid expansion of access to health and education for the poor. Opposing this, critics argue that Venezuela under Chavez underwent a rampant deterioration of institutional quality that paved the way for an increase in crony capitalism with numerous industries nationalized, the expropriation of private property, and the gross mismanagement of Venezuela's largest oil company amidst the introduction of rigid food and price control policies. Empirically, the two antithetic views open an intriguing question: Was Chavez the hero of the poor, supporting them with the beneficial economic growth impact of his policies, or simply the precursor of the resource disease and Venezuela's current disaster. Grier and Maynard (2016) apply the SCM to study the economic and social impact of Chavez's policies. In particular, they simulate the trajectories of various economic and social outcomes in the hypothetical absence of Chavez using other countries without his influence as a control group. Their results indicate no discernible improvement in poverty, health, and inequality outcomes during the Chavez administration compared to the control group of countries that best capture Venezuela's pre-Chavez economic, health-, and inequality-related outcomes. In other words, the counterfactual supports the second proposition that the Chavez administration did in fact reduce the economic opportunity of Venezuela – it adds weight to this proposition and works with existing evidence to reinforce the position.

In a related work, Absher et al. (2020) extend the study and set to examine the economic effects of the left-wing populist leaders across Latin America. They find evidence of substantially harmed economic growth that is not offset by the decreases in infant mortality and inequality. Their estimates suggest that at the end of left-wing populist leadership, the affected countries appear to be about 20% poorer relative to the synthetic counterfactuals with notable heterogeneity between Venezuela, Bolivia, and Nicaragua on one hand (i.e., where deterioration is most rampant) and Ecuador on the other (i.e., which keeps pace with its synthetic counterfactual).

Similarly, the question whether the long-standing socialist rule of Fidel Castro helped or hurt the Cuban economy has also received much attention in the scholarly debate. Deploying a reasonably large sample of countries for the period 1920–2000, Jales et al. (2018) find a moderate negative impact of the Castro regime on the economic growth trajectory of Cuba relative to the synthetic control group, which appears to be robust to various alterations of the original synthetic control estimator. Using this method, we can create models to assess likely impact on other social considerations too. Moreover, Bologna Pavlik and Geloso (2018) employ the SCM to test the hypothesis as to whether the policies of the Castro regime led to lower mortality rates. They use the set of Latin American countries as a control sample to build the counterfactual distribution of mortality rates as an approximation of the Castro regime's effect on infant mortality. Contrary to the popular narrative, they find that relative to the synthetic control group, Cuban infant mortality rates increased. The general narrative of these experiments is to describe the nature of an institutional shock with respect to its impact on economic growth and development and to build a counterfactual that has a higher quantitative validity.

Finally, and extending the narrative, Garoupa and Spruk (2019) propose a taxonomy of institutional changes distinguishing between gradual institutional changes that help improve long-run growth without a major deviation, institutional changes imposed by a shock resulting in a temporary deviation of the growth path from its long-run equilibrium that eventually fizzles out, and structural breaks with a major and permanent deviation of long-run growth equilibrium. This approach allows for a clear empirical distinction between temporary deviations of growth equilibrium and structural breaks whilst alternative empirical techniques tend to blur the two phenomena. The presence or absence of a structural break in the post-intervention period is clearly of particular interest to the practitioners of the SCM.

3 | SYNTHETIC CONTROL METHOD ITSELF – ANALYSIS

3.1 | Framework

In its simplest form, the synthetic control setup to study the impact of a certain intervention on the outcome of interest involves the set of $i = 1, 2, \dots J + 1$ affected units, which may be either countries, regions, cities, firms, or households, which are exposed to the intervention of interest. Without a loss of generality, we assume that only one unit is exposed to the intervention while the remaining J number of units are potential candidates in the donor pool to be used to construct the synthetic control group against which to evaluate the impact of the intervention (Ferman, 2021a, b).

Let $\ln y_{i,t}^N$ be the outcome of interest for the ith unit in the absence of the intervention at time t within the discrete-time horizon $t = 1, 2, \dots T$ and let T_0 denote the number of pre-intervention periods from the discrete-time horizon such that $1 \leq T_0 < T$. Suppose that the outcome of interest in the presence of the intervention is denoted as $\ln y_{i,t}^I$, and assume that the period of intervention lasts from $T_0 + 1$ to T.

Two standard assumptions underline the validity of the SCM in providing a plausible interpretation of the impact of the intervention. First, the intervention of interest at the time of its implementation is independent of the outcome of interest in the pre-intervention period (Xu, 2017).

This implies that if the intervention of interest invokes anticipatory effects prior to the timing of the intervention, it is not likely to yield a plausible representation of its impact. This assumption in turn implies that the intervention of interest should have no prior impact on the outcome of interest. That said, if the intervention of interest is anticipated, the outcome prior to the intervention may react to it, which most likely violates the assumption. Second, the units exposed to the intervention should not interfere in the intervention itself suggesting that potentially interfering units should be excluded from the donor pool in order to isolate the impact of the intervention on the treated unit (Cao & Dowd, 2019).[3] For instance, suppose the researcher is interested in the impact of democratization on the growth of regions in an affected country. To estimate and isolate the impact of the transition to democracy on regional growth, and ensure that both assumptions are not violated, all potentially interfering regions must be excluded from the donor pool. This implies that the treated region in question should be matched with the set of countries or regions that have not been affected by democratization.[4]

Mathematically, this can be described as follows. Let $\lambda_1 = \ln y^I_{i,t} - \ln y^N_{i,t}$ describe the effect of the intervention of interest for country i at time t where $\lambda_1 = (\lambda_{1,T_0+1}, \dots, \lambda_{1,T})$ captures the full set of the effects of intervention in the post-treatment period. Further assume that $D_{i,t} = 1 \cdot [(i \in J+1) \to \{0,1\}]$ is a simple linear indicator function that takes the value of 1 if the ith country is exposed to the intervention at time T_0, and 0 otherwise. Hence, the outcome of interest for country i at time t is given by:

$$\ln y_{i,t} = \ln y^N_{i,t} + \lambda_{i,t} \cdot D_{i,t} \tag{1}$$

The practitioners of the SCM often aim to estimate the effect of the given intervention on the outcome of interest. The general thrust of Equation (1) is that the level of the outcome of interest in the absence of the intervention is, by default, unobserved to the econometrician. Under standard conditions, the vector of post-treatment effects of the intervention of interest $\lambda_1 = (\lambda_{1,T_0+1}, \dots, \lambda_{1,T})$ can only be estimated for the period $t > T_0$, which implies that $\lambda_{1t} = \ln y^I_{1,t} - \ln y^N_{1,t} = \ln y_{1,t} - \ln y^N_{1,t}$. Since $y^N_{1,t}$ is unobserved to the econometrician, Abadie et al. (2010) advocate the use of the latent factor model where observed components of the outcome together with the matrix of observed and unobserved common factors are used to approximate $y^N_{1,t}$ in the pre-intervention period. The latent factor model provides a valid counterfactual trajectory of the outcome provided that transitory shocks exhibit a zero conditional mean independence assumption.

Since $y^N_{1,t}$ can only be estimated, a researcher using the synthetic control estimator must rely on the behavior of the outcome of interest in the donor pool not being exposed to the intervention of interest. This allows for the construction of the synthetic control group for the treated unit where observed covariates of y are set to match the treated unit on the set of observable time-varying and time-invariant characteristics prior to the timing of the intervention. Matching the treated unit with the control group unaffected by the intervention ensures that unobserved heterogeneity bias is not projected out of the counterfactual outcome trajectory since common time factors are matched automatically between the treated unit and its control group.

Further, suppose that the donor pool comprises $J-1$ units excluding the affected unit. By relying on the behavior of the outcome in the synthetic control group, $J \times 1$ vector of weights allows us to reweight the behavior of the outcome in the control group such that its outcome of interest will mimic the characteristics of the treated unit as much as possible given the similarities in covariates and past outcome realizations. The vector of weights may be described by $\mathbf{W} = (w_2, w_3, \dots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \dots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Notice that each particular value of the vector W represents a potential control unit with which to construct the synthetic control group. Under these circumstances, the synthetic control group is a weighted average of control countries sharing similar pre-intervention characteristics captured by observed covariates and pre-intervention outcomes. A standard theorem underlying the ability of the synthetic control estimator to reproduce the outcome of the treated unit in the absence of the intervention suggests that the discrepancy in the outcome of interest between the treated unit and its synthetic control group will disappear provided that the pre-intervention period is sufficiently large. This implies that under the standard conditions, the synthetic control group may provide a plausible characterization of the missing counterfactual scenario. An approximately unbiased estimator of $\lambda_{1,t}$ is then given by the underlying difference between the observed outcome and synthetic control group that holds the scale of transitory shocks constant provided that pre-intervention characteristics of the treated unit can be matched with its control group through a data-generating process.

One of the standard issues in synthetic control analysis arises from the potential lack of fit in the counterfactual outcome trajectory relative to the observed outcome of the affected unit prior to the intervention (Ferman & Pinto, 2021). On several occasions, the fit may be poor due to interpolation biases being large relative to the sample size. The traditional approach advocated by Abadie et al. (2010) and Cavallo et al. (2013) is to adjust the underlying specification with the appropriate covariates set to avoid a poorly fit synthetic control unit, or remove the observations with a pre-intervention root mean square prediction error (RMSE) of greater than $\sqrt{3}$ multiplied by average pre-intervention RMSE (Acemoglu et al., 2016). More recently, Ferman et al. (2020) propose several specification-searching empirical strategies based on the prediction error of each specification and placebo estimations.

An important question regarding the ability of the synthetic control estimator to capture the effect of the intervention of interest concerns the composition of the control group for the treated unit. Under the most plausible circumstances, if the assumption on conditional independence of the intervention and prior outcomes holds, the affected unit should exhibit similar trends in past outcomes and covariates to serve as a meaningful representation of the unobserved counterfactual scenario. To avoid excessively large counterfactuals, numerous adjustments of the RMSE have been proposed to allow for large but plausible treatment effects without an artificially inflated counterfactual scenario arising from the lack of fit (Adhikari & Alm, 2016; Dube & Zipperer, 2015; Ferman et al., 2020).

In the absence of the similarity in the trend of covariates and past outcomes used in the synthetic control specifications, the estimated counterfactual scenario is most likely plagued by the lack of fit. On the contrary, the synthetic control estimator allows the researcher to find an error-minimizing combination of weights of covariates and past outcomes that lie within the convex hull of the treated unit before the intervention. Recall that $\mathbf{W}$ is a $J \times 1$ vector of non-negative weights such that $\mathbf{W} = (w_2, \ldots, w_{J+1})'$ for $j = 2, \ldots J + 1$ where $w_2 + \cdots + w_{J+1} = 1$. Each value from W represents the weighted average of the control unit's covariates that lie within the convex hull of the treated unit and serves as a synthetic control group facilitating the assessment of the effect of the intervention. The convexity of the combinations from the untreated units ensures that the weights are additive by themselves.

The key question pertains regarding the similarity of covariate-level characteristics between the affected unit and its control units unaffected by the intervention. The standard approach to minimize the pre-intervention distance between the affected and unaffected units, denoted as $\|\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}\|$, is to consider Abadie and Gardeazabal (2003) and Abadie et al. (2010) semi-definite fully symmetric $r \times r$ matrix V that captures the characteristics-based distance between the affected unit and unafected ones, viz:

$$\left\| \mathbf{X}_{\mathbf{1},j} - \mathbf{X}_{\mathbf{0},i}\mathbf{W} \right\|_{\mathbf{V}} = \sqrt{\left(\mathbf{X}_{\mathbf{1},j} - \mathbf{X}_{\mathbf{0},i}\mathbf{W}\right)' \mathbf{V} \left(\mathbf{X}_{\mathbf{1},j} - \mathbf{X}_{\mathbf{0},i}\mathbf{W}\right)} \tag{2}$$

where W is the distance-minimizing vector of non-negative weights used to match the treated and untreated units in terms of the covariate characteristics before the intervention, where $\mathbf{X}_{\mathbf{1},j}$ is the covariate-level vector for the treated unit, $\mathbf{X}_{\mathbf{0},i}$ is the covariate-level vector of the unaffected unit, and V is the positive semi-definite weighing matrix that captures the degree of similarity in terms of covariate values prior to the intervention. By extracting the similarity in terms of covariates and past outcomes, one is able to determine the quality of the fit between the treated unit and its unaffected counterparts and to gauge whether the synthetic control group is a plausible representation of the treated unit's outcome trajectory in the hypothetical absence of the intervention.

3.2 | Inference

The effect of the intervention of interest on the outcome in question also hinges on the statistical significance of the outcome gap between the treated and untreated unit as an approximation of the counterfactual scenario. The standard approach used to tackle the statistical significance of the outcome gap is to run a series of (i) in-space and (ii) in-time placebo tests. More specifically, in-space placebo tests allow the researcher to determine whether the effect of the intervention appears to be specific to the treated unit or whether it is also perceivable in the control sample. For example, Abadie et al. (2010) perform a series of placebo checks to determine the impact of Proposition 99 (i.e., a large-scale anti-tobacco legislation introduced in California in 1989) on the prevalence of smoking. They assign Proposition 99 to all other states that did not implement Proposition 99 and thereby shift California from the treatment sample to the control sample. The underlying intuition of this approach is straightforward. If the effect of Proposition 99 appears to be specific to California, the effect of Proposition 99 on the smoking rates in other states should be either imperceptible or driven by the lack of fit. Under these circumstances, the evidence would lend support to the argument that a significant drop in smoking rates occurred in California in the post-Proposition period. On the other hand, if the gap in smoking rates in California in the post-intervention period is similar, then the analysis most likely does not provide evidence of the significant impact of the Proposition on the smoking rates. Hence, if the distribution of placebo effects of the intervention yields many effects as large as the baseline estimated, then the estimated impact of the intervention is most likely observed by chance and clearly not driven by the intervention in question. By default, such a nonparametric test does not impose any distribution of the random error term. Abadie et al. (2010) extend this approach further and also consider the ratio of post-intervention and pre-intervention RMSE to judge the uniqueness of the effect.

Further, suppose the effect of the intervention of interest is described by $\hat{\lambda}_{1t}$ and that the distribution of the in-space placebo effects is given by $\hat{\lambda}_{1t}^{Placebo} = \{\hat{\lambda}_{jt} : j \neq 1\}$. The two-tailed p-value on the effect of the intervention of interest may be computed as $P = Prob\left(\left|\hat{\lambda}_{1t}^{Placebo}\right| \geq \left|\hat{\lambda}_{1t}\right|\right) = \left(\sum_{j\neq 1} 1 \cdot \left|\hat{\lambda}_{jt}\right| \geq \left|\hat{\lambda}_{1t}\right|\right)^{-\left(\frac{1}{J}\right)}$ whereas one-tailed p-values for strictly positive effects are given by$P = Prob\left(\left|\hat{\lambda}_{1t}^{Placebo}\right| \geq \left|\hat{\lambda}_{1t}\right|\right) = \left(\sum_{j\neq 1} 1 \cdot \left|\hat{\lambda}_{jt}\right| \geq \left|\hat{\lambda}_{1t}\right|\right)$. Notice that since the intervention in most cases is not randomly distributed across the sample, the placebo distribution serves as a typical randomization inference. Since the p-values on the effect of the intervention of interest are derived non-parametrically, the probabilities represent the proportion of control units that have the estimated impact of the intervention at least as large as the treated unit (Galiani & Quistorff, 2017). One caveat should be that of the case of states. Specifically, the placebo effects may be relatively large if the treated and control units are not well-matched in the period preceding the intervention for reasons other than the lack of fit. The standard approach to partially mitigate the dissimilarity issue is to adjust the use of placebo coefficients $\hat{\lambda}_{1t}$ for the quality of pre-intervention match in two steps. In the first step, the quality of the match is adjusted by the multiple of the placebo effects to consider only those placebo gaps that match reasonably well. Abadie et al. (2010) employ this restriction and iteratively exclude those placebos that are between five times and two times the size of California's pre-intervention RMSE and show that the impact remains largely unchanged, even under the least lenient RMSE restrictions. In the second step, the placebo effects are divided by the pre-shock match quality parameter to obtain the distribution of pseudo t-statistics and compute the relevant p-values, which allows us to conduct the statistical inference on the intervention of interest.

By contrast, an in-time placebo test is based on the assignment of the intervention of interest to a deliberately false date and is akin to the preprogram test approach advocated by Heckman and Hotz (1989). The underlying intuition is both simple and straightforward. If the effect of the intervention of interest is perceivable in the given year, shifting the intervention period to a wrong date should yield implausible effects that do not arise from the year of the true treatment. In a much-debated example, Abadie et al. (2015) study the economic growth impact of the reunification of West and East Germany in 1990. They conduct a simple in-time placebo test and shift the year of the unification from 1990 to 1970 as the mid-range of the sample period. Their findings suggest that the trajectory of real West Germany and its synthetic counterpart are identical before and after 1970 as a deliberately wrong year of the intervention.[5]

4 | APPLICATIONS IN COMPARATIVE ECONOMIC HISTORY: SCOPE

The framework of the SCM posits a vast potential for the application of the method to identify key questions and to assist in answering some of the long-running historical questions. Before these questions can be answered, a few pressing issues concerning research design and the composition of the samples remain. Given the scope of the analysis in estimating plausible counterfactual scenarios, the question that perhaps warrants heightened attention is: How should a valid synthetic control analysis of the important historical questions look?

4.1 | Treatment scope

The scope of treatment used to study the impact of historical shocks, policy-related or institutional changes on the economic outcomes of interest should satisfy three specific criteria. First, the shock assigned to the affected unit should not be easily anticipated by prior economic conditions. If these conditions predict the timing of the treatment, the conditional independence assumption on the treatment of interest most likely fails and masks the estimated counterfactual scenario with the pre-shock outcome dynamics that are not necessarily attributed to the treatment of interest. A good example of a well-defined and unanticipated treatment is that of a natural disaster. For instance, Barone and Mocetti (2014) consider two large-scale earthquakes that occurred in two different Italian regions: one in 1976 (in Friuli Venezia Giulia) and the other in 1980 (in Irpinia) to estimate the impact of natural disasters on within-country economic growth. The use of natural disasters does not fall short of the conditional independence assumption since the timing cannot be anticipated. As a treatment itself, it provides a plausible shock that mimics the characteristics of the random assignment. Apart from providing strength to the feasibility of the assumption, natural disasters may provide an in-depth perspective of how local institutions and economic actors react to such shocks, and whether such disasters have either temporary or long-term economic impact. The authors find an almost zero short-term effect of the earthquake in both cases. In the long term, they find opposite results.[6] The authors argue that the positive nature of the shock in Friuli Venezia Giulia and the negative nature of the shock in Irpinia reflect the strength of local institutions in combating the economic effects of the earthquake. In the latter case, the construction of better infrastructure to replace that damage might increase the potential output and enhance the long-term growth trajectory. In the former, the construction of better infrastructure is hindered by the diversion and misallocation of public funds by rent-seeking behavior that distorts markets and reduces the potential output, hence, inflicting a permanent economic growth penalty from the earthquake. In this respect, the interaction between large-scale natural disasters and the subsequent institutional changes should not be neglected.

In a further example, Cavallo et al. (2013) examine the causal impact of catastrophic natural disasters on economic growth by leveraging both between- and within-country growth variation in 196 countries for the period 1970–2008. By applying the synthetic control estimator to the countries undergoing large-scale natural disasters, they are able to construct a counterfactual growth scenario in the hypothetical absence of the disaster. By controlling for the intensity of disaster, they show that large-scale disasters have a permanent economic growth effect only if followed by the necessary radical political revolution that fundamentally reshapes the institutional structure of the society. They construct two comparative case studies of such disasters, namely the 1972 earthquake in Nicaragua and the 1978 earthquake in Iran, to support their argument.

An important caveat to keep in mind concerns multiple cases of the same treatment. For instance, several scholars have tried to estimate the economic growth effects of institutional integration (Campos et al., 2019; Garoupa & Spruk, 2020; Maseland & Spruk, 2020). One such example relates to membership of the European Union and the effect of the institutional bonus of the EU-wide institutional framework on within-country economic growth. To isolate the effect of EU membership on growth, Campos et al. (2019) and Garoupa and Spruk (2020) set up the synthetic control designed to include only one treated country and several non-treated ones. This implies that a single EU country is used in the treatment sample whilst the control samples include only non-EU units to capture the effect of EU membership on the economic growth trajectory. If other EU countries were included in the donor sample, the estimated treatment effect of EU membership would not be valid.[7] Furthermore, Gobillon and Magnac (2016) use time-varying regional data to study the impact of enterprise zone policy on local unemployment in France in the 1990s and pool difference-in-differences with SCMs into an interactive effect model. Interestingly, they find only small short-run effects of the zone policy on the local employment rate (Gobillon et al., 2012).

The general thrust of the treatment selection is its uniqueness. If the treatment is experienced by multiple units, the units undergoing the treatment at the same time must be excluded from the donor pool to provide a valid and plausible counterfactual scenario. For instance, Castillo et al. (2017) investigate the impact of tourism policy on the employment rate using a comprehensive Tourism Development Policy implemented in the province of Salta in Argentina during the period 2003–2010. By adopting the synthetic control estimator to estimate the counterfactual employment rate in the absence of the policy, they first exclude the provinces that have undertaken a similar policy initiative in their sample period, namely, Buenos Aires City, Buenos Aires Province, Córdoba, and Río Negro.[8]

4.2 | Placebo analysis

4.2.1 | In-time placebo analysis

One of the key questions arising from the estimated counterfactual scenario is the assessment of whether the gap in the outcome between the actual treated unit and its synthetic control group is not driven by pretreatment trends. We have already introduced the work of Abadie et al. (2015) where they study the economic growth impact of German re-unification and use a deliberately wrong date of the unification by pushing the 1990 treatment year to the middle of the pretreatment period. They find that the growth trajectories of actual West Germany and its synthetic counterpart under a falsely assigned unification year are almost identical, which gives empirical support to the argument on the negative economic growth effect of re-unification on the West German economy. If the pre-existing trends are perceived when the treatment is applied to a false year, the influence of pretreatment events and shocks cannot be excluded and most likely taints the counterfactual scenario with the influence of these changes. Under these circumstances, the adjustment of the time period to better capture the impact of the treatment of interest is warranted alongside the control sample that does not contain any notion of the treatment of interest.[9] Rigorous empirical evidence using the synthetic control estimator to examine the causal impact of the intervention of interest should not be able to reject the null hypothesis of in-time placebo effects in the wrongly assigned year of the intervention and should be able to confirm the underlying post-treatment effect that begins to unfold in the year of the intervention.

4.2.2 | In-space placebo analysis

Another important question concerns the statistical significance of the effect of the intervention of interest, which is subject to extensive discussion in the current literature (Abadie, 2021).[10] Since, in most observational settings, the underlying intervention in question is not randomly assigned per se, two criteria are highlighted: (i) availability of a well-defined procedure to select the comparison unit, and (ii) the notion that permutation methods do not approximate the sampling distribution of test statistics, which implies that inference on the post-intervention effect should be interpreted under an “as-if-random” assumption. Whilst an in-time placebo study is able to rule out the presence or absence of pre-intervention outcome trends, it does not suggest any notion of statistical significance behind the estimated effect. For instance, how likely is it that the estimated effect of the intervention is obtained by chance? The standard way to evaluate the significance of synthetic control estimates is to ask whether the results are randomly driven (Abadie et al., 2010; Ferman & Pinto, 2017). More specifically, how likely would we obtain an effect of similar size if we were to choose a unit other than the treated unit?

To address this issue, Bertrand et al. (2004) advocate the use of in-space placebo tests. Such tests involve the application of the synthetic control estimator to the units that did not implement the intervention of interest. In their famous study of the impact of terrorism on economic growth, Abadie and Gardeazabal (2003) perform a series of in-space placebo tests by applying the synthetic control estimator to a region from the synthetic control group for the Basque Country that was not exposed to terrorism. For this purpose, they select Catalonia and show that it remains unaffected by terrorist activity compared to the Basque Country. Moreover, Abadie et al. (2010) then extend this approach by advocating the iterative application of the synthetic control estimator to all control units from the donor pool that were not exposed to the intervention of interest. The underlying thrust of the placebo studies concerns the similarity of the outcome gaps in the post-intervention period between the treated unit and all control units. If placebo analysis creates outcome gaps similar to the one for the actual treated unit, then the synthetic control analysis is quite unlikely to provide evidence of the significant impact of the intervention. On the other hand, if placebo analysis uncovers an outcome gap for the treated unit that is unusually large compared to the gaps for the control units, then the evidence of a significant effect becomes plausible. This implies that the ratio of the post- and pre-intervention prediction errors should be large for the treated unit and gradually smaller for the untreated units.

Furthermore, Galiani and Quistorff (2017) automate the computation of the placebo gaps along with a nonparametric set of *p*-values indicating how likely it is that the effect is driven by chance across the full range of post-intervention years. As suggested by Ferman and Pinto (2021), the inference on the placebo gaps is tainted by the size distortions even when the pre-/post-intervention prediction error ratio has the same marginal distribution for all placebo runs. Since such *p*-values are nonparametric and indicate the proportion of control units with the effect as large as the effect on the treated unit, Garoupa and Spruk (2019) perform a difference-in-differences test of the uniqueness of the placebo gaps that can be applied to single or multiple treated units, and show that the test performs reasonably well in large samples.

4.2.3 | Empirical challenges

Both in-space and in-time placebo analyses can be used to evaluate the statistical significance and uniqueness of the underlying historical, natural, or policy-related shocks and interventions. The general thrust of the placebo analysis is the permutation of the treatment to the units in the donor pool that have not undergone the intervention of interest, or the assignment of the intervention to a deliberately false date. Nevertheless, several empirical challenges remain. First, although in-space placebo analyses may indicate both size and direction of effect not perceivable in the donor pool, the underlying effect may be driven by substantial uncertainty. Against this backdrop, Firpo and Possebom (2018) propose parametric weights for the *p*-value on the treatment effect that includes the Abadie et al. (2010) equal-weights benchmark to analyze the sensitivity of the test result to deviations from the benchmark. More specifically, by inverting the test statistic to modify RMSE, the scholar is able to test any sharp null hypothesis and construct empirical confidence intervals on the treatment effect. These confidence intervals may potentially unravel the uncertainty behind the effect of the intervention in question on the outcome of interest more fully. Additionally, the choice of the placebo year may be ambiguous and, at times, arbitrary, which renders the inference on the uniqueness of the intervention somewhat more difficult. To fill the void in the literature, Hahn and Shi (2017) consider the performance of the permutation test in the context of the SCM. They show that symmetry assumption, as a crucial condition for the validity of the permutation test, may be violated. In turn, the violation of the assumption can distort the size of the permutation test. They recommend the application of Andrews (2003) stability test of the no treatment hypothesis as the equivalent of stationarity of the outcome time series, which appears to be asymptotically valid. Third, the placebo analysis may be affected by the lack of solid pretreatment fit, which brings the inference on the estimated effect into question. Several empirical strategies have been proposed to deal with imperfect pretreatment fit. For instance, Abadie and L'Hour (2021) review the original synthetic control estimator and show that a unique solution in finding the control unit that best reproduces the outcome of interest for the treated unit before the intervention might not exist, and propose a data-driven choice of penalized synthetic control estimator to further minimize the discrepancy between the treated unit and its control group.[11] More recently, Amjad et al. (2018) and Athey et al. (2021) propose matrix completion methods for estimating causal effects in the panel data where exposure to treatment is not constant after the intervention.[12]

4.3 | Leave-one-out analysis

The next question concerns the effect of ambiguity that arises from the intervention of interest. Klößner et al. (2018) review the findings by Abadie et al. (2015) on the economic growth effect of German re-unification and address the variation in the post-re-unification economic growth gap by performing a series of effect simulations with a different composition of the donor pool. More specifically, they find that the economic growth gap between West Germany and its synthetic control group is strongly influenced by the US data, and show that excluding the United States of America from the donor pool renders the estimated growth gap much smaller and no longer significant in the in-space placebo analysis. The composition of the donor pool matters greatly for the size of the underlying effect and has, by default, immediate implications for statistical significance.[13]

To address this issue, the authors undertake a "leave-one-out" analysis by iteratively excluding the countries with the positive weight share from the control group for each treated country to enforce sparsity on the underlying control groups. They show that the covariate balance remains unaffected since the exclusion does not alter the prewar prediction error.

Leave-one-out analyses can be further extended by considering a more rigorous exclusion set such as the exclusion of the entire synthetic control group from the donor pool to check whether the size and significance of the effect are stable or not. Such analysis is contingent on the sample size such that it is sufficiently large to permit such exclusions without invoking an under-powered research design where the inference is typically driven by a few control units.

4.4 | Specification search

One of the key steps to undertake synthetic control analysis is to identify the covariates of the outcome variable. In the ideal setup, these covariates are able to account for the bulk of the outcome variance in the pre-intervention years. The natural question to ask is how to select the covariates to build both comprehensive and compact synthetic control specification. As a rule of thumb, Hahn and Shi (2017) advocate for a large number of covariates compared to the size of the donor pool as a way to improve the choice of weights assigned to each control unit given the set of predictor weights. Since both time-varying and time-invariant covariates can be considered, one potentially appealing candidate is the lagged outcome variable. Apart from addressing the standard endogeneity concerns inherent in within/between-type of outcome variation, the inclusion of the lagged outcome variable largely ameliorates the standard omitted variable bias. Furthermore, Athey and Imbens (2006) argue that including covariates other than the lagged outcome variable implies that these covariates rarely matter.

Against this backdrop, Klößner et al. (2018) and Kaul et al. (2021) suggest that researchers should consider using several combined variants of the values of outcome variables in the pre-intervention years in their empirical strategies. In addition, Ferman et al. (2018) show that the lack of clear theoretical guidance on the selection of covariates creates numerous specification-searching problems. They recommend the consideration of different sets of lags and covariates and report them all whilst discarding specifications with the average of pretreatment outcome since it fails to exploit the pre-intervention outcome dynamics. To successfully exploit the specification-searching opportunities, McClelland and Gault (2017) recommend the choice of a small number of outcome variable lags that follow the outcome trend in the pre-intervention period.

4.5 | Differential trend analysis

Last, the ability of the synthetic control estimator to estimate the causal impact of the intervention of interest depends on the researcher's ability to determine whether the outcome in the treated unit and its synthetic control group follows a different trend in the post-intervention period. Ideally, the outcome trends in the treated unit and its synthetic control group should be indistinguishable whilst being statistically distinguishable in the post-treatment period. To address this issue, Spruk and Kovac (2020) test the similarity of outcome trends between the treated unit and its synthetic control group before and after the intervention. Their study examines the causal impact on cardiovascular diseases and obesity rates of the ban passed by Denmark on trans fats in 2003. More specifically, they construct 95% confidence intervals for the trend slopes of mortality rates in Denmark and its synthetic control group before and after the ban. By comparing the change in the slope of actual Denmark and its synthetic counterpart in the pre- versus post-intervention period. They undertake a triple difference test to determine whether the policy intervention induced the structural break in the mortality trajectory. By embedding a simple Chow test into the synthetic control setup, they reject the null hypothesis on the absence of differential trends with a parametric *p*-value. Such analysis can potentially uncover the presence or absence of differential trends in the treated unit and its synthetic control group before and after the intervention using the triple difference test. The ability to empirically defend the differential trend assumption most likely lends further credence and support to the validity and plausibility of the synthetic control analysis in estimating the causal impact of the intervention of interest.

4.6 | Limitations and prospective mitigations

Several caveats should be stated concerning the evaluation of the impact of policies or institutional changes on economic outcomes when the SCM is used. Despite many strengths, the synthetic control estimate of the outcome gap between treated and control units is subject to several limitations that impede rigorous causal inference. First, the estimated outcome gap in the post-treatment period can be affected by the low quality of the fit. If the treated and control units' outcome trajectory exhibits a poor fit in the pretreatment period, the post-treatment outcome gap might be driven either by omitted variables or by alternative shocks or policy changes that invoke pre-existing trends and render the estimated gap problematic. To address this issue, the researchers using the method should build treatment and control samples with diligent care to allow for the consistent similarity of outcome trends and dynamics prior to the postulated shock. In the absence of trend similarity in the pretreatment stage, the estimated gaps are most likely artificially inflated or driven by sample selection bias that renders the identification of the causal impact of the intervention of interest questionable. In this respect, as the parallel trend assumption is not specifically tested in the synthetic control setup, our recommendation to scholars is to cross-validate the estimated gaps by comparing them with a traditional difference-in-differences analysis where the parallel trend assumption between the treated unit and its control sample can be tested directly (Muralidharan & Prakash, 2017). A similarity of the synthetic control estimate and the difference-in-differences estimate might serve as an important robustness check on the susceptibility of the synthetic control approach to the pretreatment parallel trend assumption, which typically allows for the causal identification of the policy and institutional changes using within-unit outcome variation over time.

Additionally, the use of the SCMs is intended to demonstrate plausible counterfactuals in order to identify prospective areas for further historical analysis. The counterfactuals themselves are not intended to be necessarily explanative of the historical phenomenon but, rather, to provide a confirmation as to the structural break's likely existence and then to allow for the assessment of historical evidence in the context of confirming the outcomes the break had caused.

4.7 | Example application: The long-term effects of the oil discovery in Venezuela

4.7.1 | Background and data

We demonstrate the use of the SCM for comparative case studies in economic history by exploiting a unique quasi-natural experiment. Namely, we exploit the discovery of oil in Venezuela at the beginning of the 20th century to examine the contribution of a large-scale discovery of natural resources to the long-term economic performance, which has been the subject of intense scholarly discussion. Our strategy is to estimate the missing counterfactual scenario in response to the discovery of oil by making use of the SCM compared to the more traditional parametric approach that is common in the literature.

In spite of the centuries of knowledge of existing oil reserves, Venezuela's oil industry began to develop in 1908 when president Juan Vicente Gómez began granting several concessions to explore, produce, and refine oil to a close circle of friends who then passed them on to foreign oil companies. In 1914, after the completion of Zumaque-I oil well, Mene Grande was discovered as the first major important oil field in the Maracaibo Basin. After several more oil fields were discovered, by the end of 1918, Venezuela's oil exports totaled 21,194 metric tons, and appeared for the first time in the export statistics. In the subsequent year, Venezuela became one of the largest per-capita producers of oil worldwide. In 1922, the blowout of the Barroso No. 2 well in Cabimas in present-day Zulia state marked the onset of Venezuela becoming a major oil producer. In the subsequent decades, dramatic increases in oil production and exports began to dominate other sectors with a rapid decrease in agricultural production, which fell from one third in 1920 to 10% in 1950. Becoming a major oil producer raises questions concerning the long-term effects of the oil discovery on economic performance. With the exception of Hartwell et al. (2019, 2021) who study the effect of the oil discovery on income inequality, the effects of oil discoveries on long-term economic performance remain an open question.

To this end, our sample comprises the dataset of growth and development trajectories and characteristics for 56 countries[14] for the period 1870–2015 and is based on the prior work of Garoupa and Spruk (2019). Our dependent variable is GDP per capita whereas the covariates comprise the set of growth and development variables used to capture the similarities of Venezuela with the rest of the world prior to the discovery of oil. To parse out the effect of the oil discovery, major oil producers where oil merchandise exceeds 20% of the exports are excluded from the donor pool[15] to prevent the estimated counterfactual scenario with a confounding influence of the oil discovery and exports perceivable in other oil producers that are hypothetically similar to Venezuela in terms of growth and development characteristics and do not provide a variation from which valid counterfactual scenarios could be properly estimated.

Our battery of covariates in the synthetic control specification is divided into four blocks that combined: (i) prediscovery GDP per capita dynamics with auxiliary covariates related to; (ii) physical geography; (iii) institutional quality and legal history; and (iv) demographic, human capital, and historical genetic diversity. To capture flexible pretreatment outcome dynamics, we follow Ferman et al. (2020), and consider a distinctive variety of pretreatment outcome combinations to fully capture the dynamics prior to the outcome which, importantly, has been shown to affect the outcome gap prior to the treatment (Abadie et al., 2015; Doudchenko & Imbens, 2016). Thus, our specification considers the level of initial GDP per capita, first lag of GDP per capita, and 10-year averaged GDP per capita. The latter effectively ensures that, due to the potential AR(1) process, pretreatment GDP per capita variation does not swallow the importance of auxiliary covariates associated with the physical geographic and institutional similarity between Venezuela and its control group. Physical geographic covariates were identified by Nunn and Puga (2012) and capture time-invariant characteristics such as latitude, longitude, soil quality, and access to coastline among several others. The institutional quality covariates comprise the V-DEM indices of electoral and liberal democracy identified by Teorel et al. (2019), common law versus civil law indicator variables (La Porta et al., 2008), and data on the presence of armed domestic conflict (Brecke, 2001). Demographic covariates comprise time-varying variables such as population size, density, and growth (Maddison, 2007) that are updated with the US Census ICP data, and time-invariant ones such as the share of European descendants (Easterly & Levine, 2016), literacy rate in the year 1900 (Madsen et al., 2015), and ancestry-adjusted genetic diversity and its squared term (Ashraf & Galor, 2013).

Table 1 reports the prediscovery covariate balance between the actual Venezuela and its synthetic control group. Balancing prediscovery GDP per capita dynamics and the auxiliary covariates indicates a reasonable fit of Venezuela's growth trajectory with its control group. The size of the root mean-squared prediction error is 0.074, which does not invoke the possibility of pre-existing trends tainting the per capita GDP gap associated with the discovery of oil. In particular, the real Venezuela has a very similar per capita income level than its synthetic control group along with similar geographic characteristics such as the size of the land area, latitude and longitude coordinates, soil quality, and access to the coastline. Institutional quality between Venezuela and its synthetic control group is synthetically well-matched with similar levels of electoral and liberal democracy in the pretreatment period along with a very similar fraction of time where domestic armed conflict had been present. At the same time, both demographic structures, human capital, and genetic diversity of the synthetic control group tend to capture the characteristics of the real Venezuela reasonably well.

TABLE 1 Prediscovery covariate balance

	Real Venezuela	Synthetic Venezuela
RMSE	**0.074**	
Panel A: Prediscovery GDP per capita dynamics		
log GDP per capita in 1870	6.35	6.39
log 10y GDP per capita (1870–1880)	6.41	6.46
log 10y GDP per capita (1880–1890)	6.68	6.60
log 10y GDP per capita (1890–1900)	6.74	6.73
log 10y GDP per capita (1900–1910)	6.74	6.79
log 10y GDP per capita (1910–1920)	6.91	6.94
log 10y GDP per capita (1870–1880)	7.67	6.97
log GDP per capita ($t-1$)	6.35	6.39
Panel B: Physical geography covariates		
Island	0	0.02
Landlocked	0	0.03
Log land area	13.69	13.64
Terrain ruggedness	0.63	1.86
Latitude	7.13	15.14
Longitude	−66.15	22.96
Soil quality	21.57	28.09
Desert	0	2.42
Tropical	96.75	38.60
Distance to coast	0.31	0.24
Fraction of area within 100 km coast	25.25	39.20
Panel C: Institutional quality and legal history covariates		
British common law	0	0.51
Civil law	1	0.49
V-DEM Polyarchy index	0.09	0.12
V-DEM Liberal democracy index	0.07	0.10
Armed conflict	0.08	0.07
Panel D: Demographic, human capital and genetic covariates		
Log population density	1.29	2.34
Log population size	14.66	15.85
Population growth	0.01	0.02
Share of European descendants	0.55	0.20
Literacy rate in 1900	0.20	0.28
Ancestry-adjusted predicted genetic diversity	0.69	0.69
Ancestry-adjusted squared predicted genetic diversity	0.48	0.48

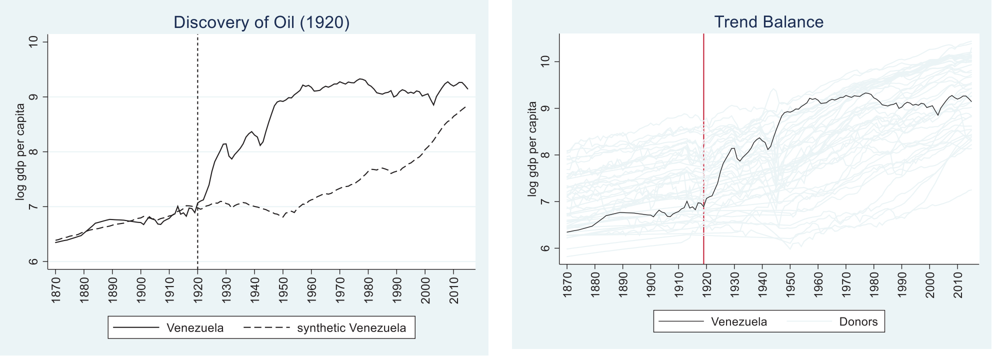

FIGURE 1 Long-term effect of oil discovery on Venezuela's economic growth [Colour figure can be viewed at wileyonlinelibrary.com]

4.7.2 | Results

Figure 1 presents the long-term effect of the oil discovery on Venezuela's economic growth. The evidence indicates an immediate and rapid acceleration of economic growth in the years after the discovery of oil. The size of the effect is quantitatively large and indicates the doubling of per capita income in 10 years after the discovery of oil, and the quadrupling of per capita income until the year 1950. The right part of the figure posits Venezuela's observed growth trajectory in the comparative perspective, and shows that, by the late 1950s, the per capita income level reached the top of the sampled world income distribution. Afterwards, the growth trajectory exhibits a pattern of a diminishing rate of increase along with a clearly perceptible growth stagnation in the 1970s followed by persistent economic decline down to the present day. By the last years of our post-treatment period, our estimates indicate a rapid disappearance of the per capita income gap that appears to be triggered by the discovery of oil in the 1920s. Thus, our estimates highlight a bold and substantial evidence in support of the notion of long-term economic decline driven by the large-scale discovery of oil, which is consistent with the extant literature on the adverse effects of natural resource discoveries being a curse rather than a cure for economic growth.

Table 2 reports the composition of the synthetic control groups for Venezuela. Notice that the control group provides the set of countries whose prediscovery growth paths best reproduce Venezuela's growth trajectory to facilitate the counterfactual scenario in the hypothetical absence of the oil discovery. The country-level weights denote the extent to which each country in the donor pool is able to reproduce the synthetic version of Venezuela, mimicking the actual Venezuela as much as possible for the given set of prediscovery outcomes and auxiliary covariates. The set of positive weights ensures that the differences in the prediscovery GDP per capita dynamics and the covariates should be as small as possible. The weights are obtained in two steps. In the first step, we rely on the cross-validation technique introduced by Abadie et al. (2015) and further discussed by Klößner et al. (2018). Namely, the prediscovery period is divided into a training and validation period. For any given pre-T0 outcome and covariate, training donor weights are obtained to minimize the predictive discrepancy between the actual Venezuela and the donor pool to the degree possible. In this step, covariate-level weights are obtained through the cross-validation by making use of the data on the per capita GDP as a dependent variable to determine the optimal weights. Finally, main donor weights and post-discovery values of the GDP per capita compute the counterfactual. As indicated in Table 2, the growth trajectory of the actual

TABLE 2 Composition of synthetic control group

Argentina	0	Mexico	0.25
Australia	0	Morocco	0
Austria	0	Nepal	0
Belgium	0	Netherlands	0
Brazil	0.05	New Zealand	0
Burma	0.51	Philippines	0.02
Chile	0.10	Poland	0
China	0	Portugal	0
Czech Republic	0	Slovenia	0
Denmark	0	South Africa	0
Egypt	0	South Korea	0.05
Finland	0	Spain	0
France	0	Sri Lanka	0
Germany	0	Sweden	0
Greece	0	Switzerland	0
Hong Kong	0	Syria	0
Hungary	0	Thailand	0
India	0	Tunisia	0
Ireland	0	Turkey	0
Italy	0	United Kingdom	0
Jamaica	0	United States	0
Japan	0	Uruguay	0
Jordan	0.03	Vietnam	0
Lebanon	0	**RMSE**	0.074

Venezuela is best synthesized as a convex combination of the growth and development characteristics of Burma (51%), Mexico (25%), Chile (10%), Brazil (5%), South Korea (5%), Jordan (3%), and Philippines (2%), respectively. This particular combination of convex characteristics that best mimics the actual Venezuelan growth trajectory yields a RMSE of 7.4%, which seems to be consistent with the quality of fit measure originally advocated by Adhikari and Alm (2016).

4.7.3 | In-space placebo analysis

The evidence so far indicates a large-scale and long, but temporary, effect of the oil discovery on Venezuela's economic growth, which is consistent with the notion of the structural change imposed by a shock with a long-running but temporary deviation of the growth path from its long-run equilibrium (Garoupa & Spruk, 2018). The obvious question to ask is: how do we evaluate the statistical significance of the effect over time? To answer this question, we ask whether our result is entirely driven by chance alone following Abadie et al. (2010)'s inferential techniques. Namely, how often would we obtain the estimate of a similar magnitude if we choose a random

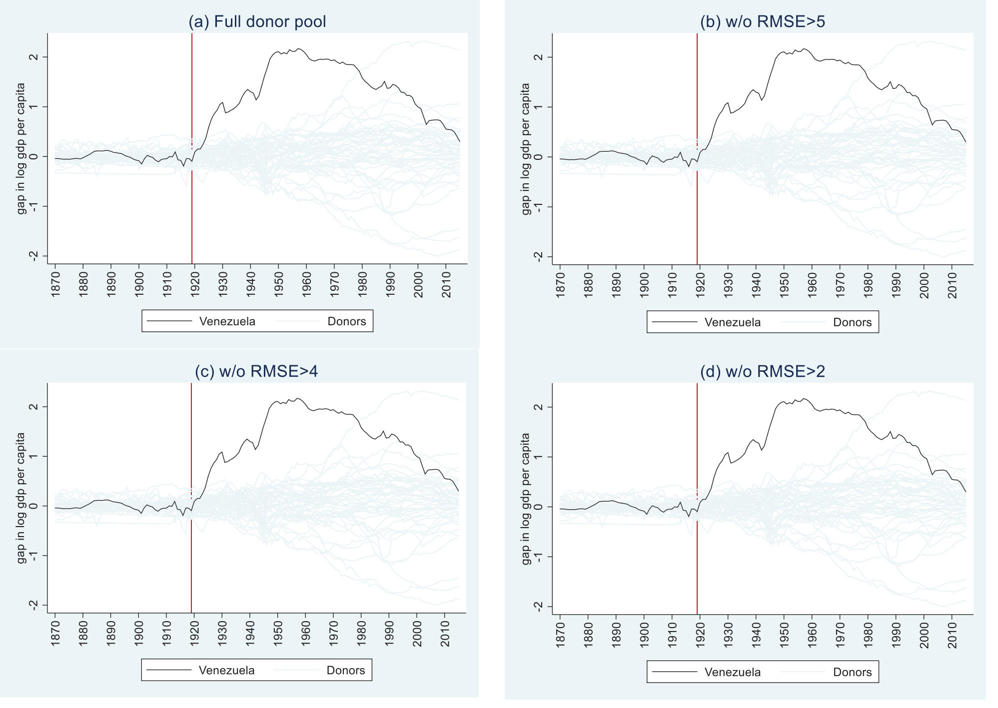

FIGURE 2 Placebo analysis of the oil discovery growth effect [Colour figure can be viewed at wileyonlinelibrary.com]

country instead of Venezuela? By relying on the in-space placebo analyses, pioneered by Abadie and Gardeazabal (2003) and Bertrand et al. (2004), we perform a series of placebo simulations where the synthetic control estimator is iteratively applied to all the countries in the donor pool that did not undergo a large-scale oil discovery during the period of our investigation.

The intuition behind the placebo simulations is straightforward. If permuting the discovery of oil to all countries in the donor pool creates per capita income gaps of magnitudes similar to the one estimated for Venezuela, then the analysis likely does not corroborate the notion of a significant temporary impact of the oil discovery. By contrast, if the permutation of the oil discovery indicates the gap identified for Venezuela as being unusually large, unique, and easily perceivable relative to the gaps in placebo simulations, then the analysis may provide some evidence supporting the notion of the significant temporary effect of the oil discovery on Venezuela's economic growth. Our placebo simulations are based on the iterative application of the synthetic control estimator to every country in the donor pool, effectively shifting Venezuela to the donor pool which treats other countries as if they underwent the oil discovery. For each placebo simulation, we compute the estimated effect that provides the distribution of the per capita income gaps for the countries without having discovered oil.

Figure 2 presents the in-space placebo analysis of the oil discovery-driven growth effect. The synthetic control estimator provides a reasonably good fit of the economic growth trajectory of Venezuela prior to the oil discovery. Prediscovery RMSE (which denotes the average of squared discrepancies between actual Venezuela and its synthetic version) is about 0.07, which implies that the per capita income gap between Venezuela and its synthetic control group is not likely to be artificially created by the lack of fit and, instead, may be driven by the oil discovery itself.

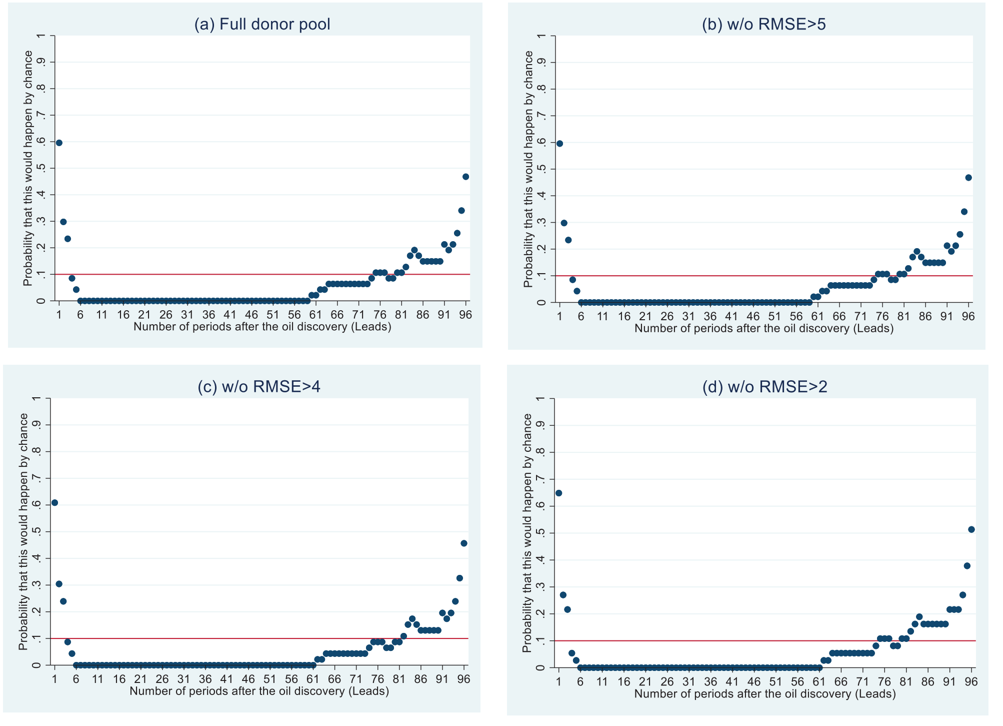

FIGURE 3 Randomization inference on the growth effect of oil discovery [Colour figure can be viewed at wileyonlinelibrary.com]

As advocated by Abadie et al. (2010, 2015), placebo simulations with high prediscovery RMSE may fail to provide the evidence on the rarity of estimating a large post-discovery of the per capita GDP gap for a country that was reasonably well fitted prior to the discovery. To address this particular concern, we present several versions of placebo assessments excluding countries beyond a certain level of prediscovery RMSE. In Figure 2 (b) we exclude countries with the prediscovery RMSE more than five times that of Venezuela, dropping the countries with large values of prediscovery RMSE. Among the full set of placebos, the gap for Venezuela appears to be unusually and temporarily large whilst disappearing down to the present day. In Panel (c), a less lenient cutoff is used, discarding the countries with RMSE four times that of the original simulation. The temporary uniqueness of the Venezuela gap is both perceivable and detectable compared to the earlier specification. Panel (d) presents the most conservative placebo simulation where we focus only on those countries that fit almost as well as Venezuela with a prediscovery proposition not higher than twice the level of Venezuela's RMSE. Comparing Venezuela's gap against the restricted distribution renders the gap highly unusual, unique, and perceivable.

In Figure 3, we present the intertemporal distribution of p-values from a random permutation test across the placebo simulation. The p-values indicate the proportion of countries with the placebo gap as large as that of Venezuela, which we interpret as evidence of the statistical significance of the effect of the oil discovery under conventional significance tests. The key strength of the test is the ability to gauge the evolution of statistical significance associated with the gap over time. The evidence suggests an immediate drop in the p-value to the zero threshold in the sixth year after the discovery of oil. The p-values remain flat for 56 years in the post-treatment period,

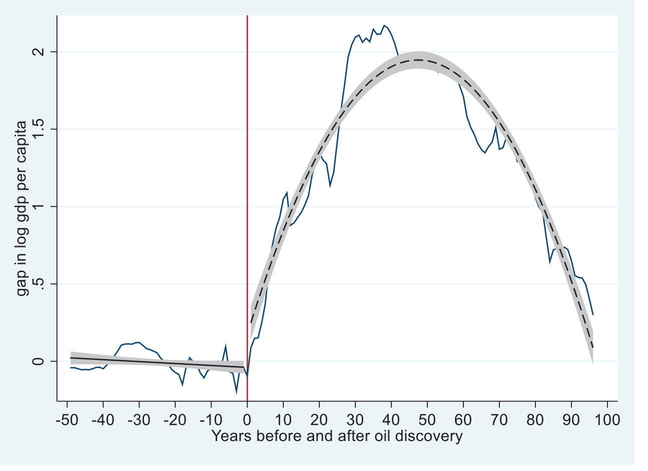

FIGURE 4 Differential trend analysis [Colour figure can be viewed at wileyonlinelibrary.com]

which corresponds to the year 1976, when the oil industry underwent large-scale nationalization. Afterwards, the *p*-values exhibit a marked. and gradual increase, indicating the temporary nature of the oil discovery impacting economic growth. In the 80th post-treatment year (i.e., 2000), the *p*-value tends to increase above the 10% threshold and keeps increasing thereafter, offering further support for our argument.

4.7.4 | Differential trend analysis

Another potential caveat invokes the ability to perceive a differential trend in the trajectory of economic growth in response to the oil discovery between Venezuela and its synthetic control group. Figure 4 compares the time trend of the GDP per capita gap of Venezuela along with the 95% confidence interval before and after the intervention. To assess the perceptibility of differential trends, we compute the linear time trend in the prediscovery period and the quadratic trend in the post-discovery period. The evidence consistently suggests a marked distinction in the time trend before and after the discovery. In the prediscovery period, the slope of the time trend is zero, which corresponds to the placebo simulations in Figure 3 reasonably well. In the post-discovery period, the time trend associated with the gap exhibits a hump-shaped pattern with relatively thin confidence intervals that invoke little aggregate uncertainty about the effect itself. The evidence consistently points out and confirms the temporary nature of the oil discovery in shaping the trajectory of economic growth. The Chow test statistic on the structural break test is above 2300 in absolute terms with *p*-value = 0.000, respectively.

4.7.5 | Leave-one-out analysis

Another concern arises from the disproportionate influence of the countries in the synthetic control group on the per capita GDP gap associated with the oil discovery. Notice that the synthetic control group for Venezuela is dominated by Burma, which represents about one-half of the total size of the control group followed by Mexico, which comprises one-fourth of the control group. Strongly concentrated control groups may innately invoke a disproportionate influence of high-

FIGURE 5 Leave-one-out analysis of the long-term effect of oil discovery [Colour figure can be viewed at wileyonlinelibrary.com]

leverage observations that may overwhelm the counterfactual scenario with a bulwark of country-specific policies and idiosyncratic influences that could pose a concern for the estimated growth gap. To address these concerns, we stepwise exclude the base set of countries from the control group as well as the entire original control group to check for potential discrepancies and relevance of high-leveraged observations for the counterfactual scenario. A similar approach has been advocated by Klößner et al. (2018) as a check on the stability of the counterfactual scenario to the influence of potentially outlying observations.

Figure 5 presents the results of a leave-one-out analysis of the long-term effect of the oil discovery. In each iteration of the base synthetic control specification, each country that enters the control group for Venezuela is excluded from the donor pool whilst re-estimating the core long-term growth specification using the restricted donor pool. Table 3 reports the composition of the corresponding synthetic control groups for each iteration of the specification. The evidence almost

TABLE 3 Composition of leave-one-out synthetic control groups

	w/o Burma	w/o Mexico	w/o Chile	w/o Brazil and South Korea	w/o countries with $w < 0.05$	w/o full-set donors
RMSE	0.082	0.069	0.084	0.075	0.070	0.088
Argentina	0	0	0	0	0	0
Australia	0	0	0	0	0	0
Austria	0	0	0	0	0	0
Belgium	0	0	0	0	0	0
Brazil	0.17	0.09	0.05	–	–	–
Burma	–	0.38	0.36	0.58	0.60	–
Chile	0.06	0.28	–	0.09	0.10	–
China	0	0	0	0	0	0
Czech Republic	0	0	0	0	0	0
Denmark	0	0	0	0	0	0
Egypt	0	0	0	0	0	0
Finland	0	0	0	0	0	0
France	0	0	0	0	0	0
Germany	0	0	0	0	0	0
Greece	0	0	0	0	0	0
Hong Kong	0	0	0	0	0	0
Hungary	0	0	0	0	0	0
India	0	0	0	0	0	0
Ireland	0	0	0	0	0	0
Italy	0	0	0	0	0	0
Jamaica	0	0	0	0	0	0
Japan	0	0	0	0	0	0
Jordan	0.05	0	0.07	0.03	–	0
Lebanon	0	0	0	0	0	0
Mexico	0.28	–	0.36	0.26	0.23	–
Morocco	0	0	0	0	0	0
Nepal	0	0	0	0	0	0
Netherlands	0	0	0	0	–	0
New Zealand	0	0	0	0	0	0.11
Philippines	0	0.04	0.09	0.05	0	–
Poland	0	0	0	0	0	0
Portugal	0	0	0	0	0	0
Slovenia	0	0	0	0	0	0
South Africa	0	0	0	0	0	0
South Korea	0.17	0.21	0	–	–	–
Spain	0	0	0	0	0	0
Sri Lanka	0	0	0	0	0	0

(Continues)

TABLE 3 (Continued)

	w/o Burma	w/o Mexico	w/o Chile	w/o Brazil and South Korea	w/o countries with $w < 0.05$	w/o full-set donors
Sweden	0	0	0	0	0	0
Switzerland	0	0	0	0	0	0
Syria	0	0	0	0	0	0
Thailand	0.27	0	0.08	0	0.08	0.16
Tunisia	0	0	0	0	0	0
Turkey	0	0	0	0	0	0
United Kingdom	0	0	0	0	0	0
United States	0	0	0	0	0	0
Uruguay	0	0	0	0	0	0
Vietnam	0	0	0	0	0	0.73

unequivocally suggests that the growth trajectories of the actual Venezuela and its synthetic version tend to converge down to the present day, confirming the notion of the long-running but temporary growth effect of the oil discovery. The magnitude of the estimated gap exhibits a marked increase in the early years after the discovery and the gradual disappearance of the growth premium of the oil discovery that begins to unfold in the follow-up of the large-scale nationalization of the oil industry in 1976. When the two sets of high-leverage observations (i.e., Burma and Mexico) are excluded from the donor pool, the gap in GDP per capita between actual Venezuela and its synthetic control group tends to flatten out almost completely with the estimated difference that is indistinguishable from zero. Quantitatively similar results are reached when either Chile or Brazil and South Korea are dropped off the donor pool, respectively. Even though the estimated gap between actual Venezuela and its synthetic version is still positive, the significance of the gap disappears entirely (p-value = 0.623). In the same vein, we find a negligible and insignificant gap in the specification where the full set of the base synthetic control group is excluded from the donor pool. In the presence of well-reproduced growth trajectory prior to the discovery, the growth premium induced by the oil discovery gradually disappears. In this particular iteration of the analysis, the growth trajectory of Venezuela prior to the discovery is best reproduced as a convex combination of growth and development characteristics of Vietnam (73%), Thailand (16%), and New Zealand (11%), respectively.

4.7.6 | In-time placebo analysis

An additional concern regarding the inference on the long-term growth effect of oil discovery arises from the possibility of alternative shocks and institutional changes being driving forces behind the counterfactual growth trajectory. By considering such a limitation, the long-term growth effect of the oil discovery may be driven by the subsequent institutional changes such as democratization in the 1950s, oil nationalization in 1970, and the tide towards the socialist revolution that began under Chavez in the early 2000s. To address these caveats, alternative years of structural changes ought to be considered.

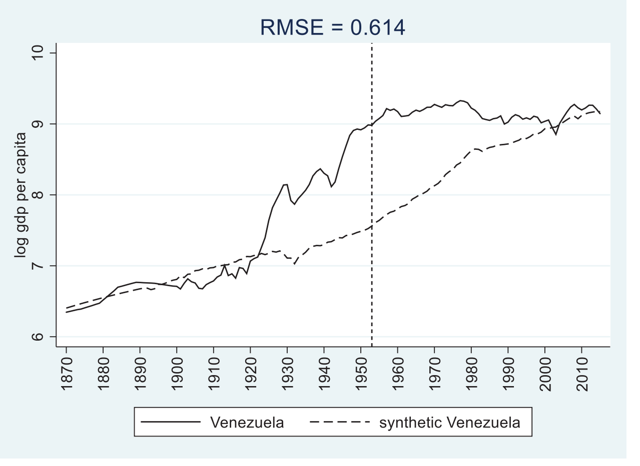

FIGURE 6 Falsely assigned year of oil discovery [Colour figure can be viewed at wileyonlinelibrary.com]

To avoid arbitrary judgment and selection of years, we employ the structural break test analysis proposed by Zivot and Andrews (2002), which may indicate the years in which structural breaks in the underlying time series are perceptible. The output of the Zivot-Andrews test indicates the year 1952 as the year in which another structural break with the highest breakpoint *t*-statistics is likely. The 1950s in Venezuela coincide with the large-scale institutional turnover deposing the Perez Jimenez dictatorship and facilitating the 1958 coup d'etat that paved the way towards democratization. Hence, we use the year 1952 in our synthetic control specification as a treatment year. Such deliberately false assignment of the year of structural change may be viewed as in-time placebo analysis.

Figure 6 reports the per capita GDP gap in response to the falsely assigned year of the structural change. The figure provides no evidence of a structural break that could be comparable to the oil discovery in 1920. Apart from a lack of evidence of well-matched and reproduced growth trajectories between the actual Venezuela and its synthetic version prior to 1952, we find no evidence in favor of the alternative changes and trends behind the counterfactual scenario, which, at best, does not cast significant doubt on whether our baseline result is plausible.

4.7.7 | Covariate sensitivity analysis

Last, we address the ambiguity concerning the choice of covariates in the synthetic control specification given that some covariates may have better predictive capacity than others in synthesizing Venezuela's growth trajectory compared to the rest of the world. Given the temporal dependence and the likely AR(1) process, it is not superfluous to suspect that prediscovery per capita income dynamics may have greater predictive power than most auxiliary covariates. Our approach to tackle the ambiguity arising from the choice of covariates is based on restricting the battery of covariates to those that match best in the pretreatment training and validation period. Although the choice of the restriction threshold may be arbitrary to some extent, we consider several plausible alternatives. We begin by extracting the normalized covariate-level weight shares from the covariate similarity matrix to detect the covariates that matter most for synthetically matching

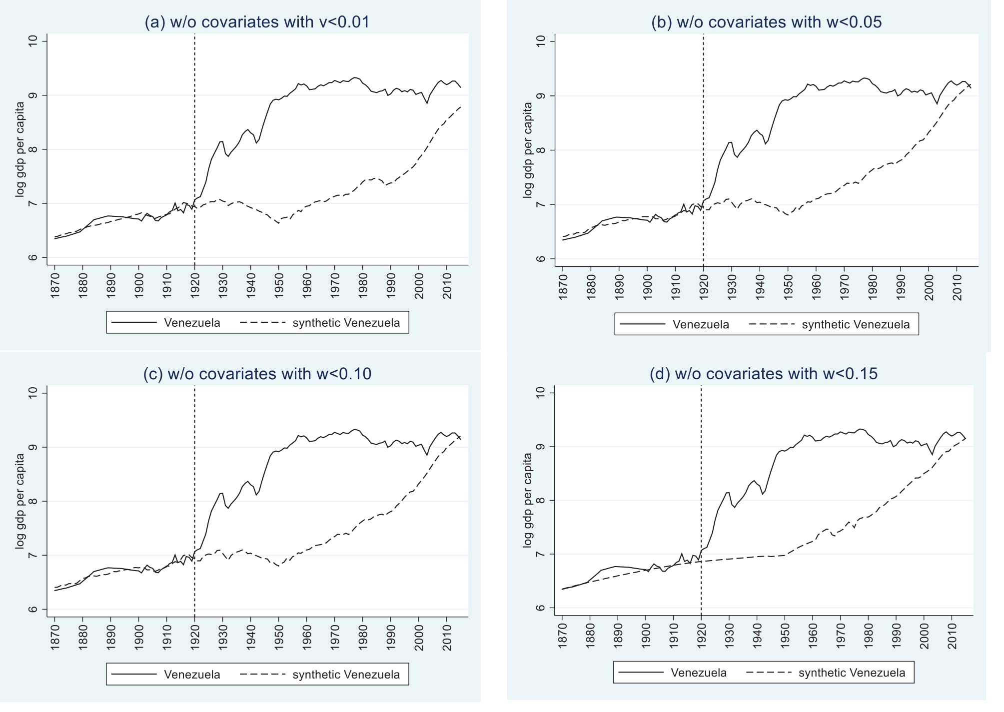

FIGURE 7 Covariate sensitivity analysis [Colour figure can be viewed at wileyonlinelibrary.com]

Venezuela's growth trajectory with the rest of the world. We then stepwise exclude the covariates with less than 1%, 5%, 10%, and 15%weight rate in the covariate convex hull, and re-estimate the synthetic control specification using the restricted sets of covariates.

Figure 7 presents the sensitivity analysis of the oil discovery-driven long-term per capita income gap with respect to the choice of covariates. In Panel (a), we exclude covariates with less than 1%weight rate, which mainly includes time-invariant physical geographic covariates that do not exhibit substantial predictive power. The evidence confirms a reasonably strong similarity of the estimated per capita income gap compared to the baseline evidence. In Panel (b), we heighten the restriction cutoff to 5%. The evidence highlights an almost uniform and long-running gradual convergence of the actual Venezuelan growth trajectory with the counterfactual Venezuela that would have left oil in the ground. Similar evidence is perceivable in Panel (c) where the covariates with less than 10% weight rate are excluded from the batter. The most conservative simulation of the synthetic control model specification is presented in Panel (d) where the covariates with less than 15%weight share are dropped off the covariate list. This leaves us with only those covariates that are best able to capture the Venezuelan prediscovery growth trajectory, which consist of the per capita GDP dynamics and historical genetic diversity. Even in the presence of few covariates with well-performing predictive properties, the evidence appears to reiterate the baseline notion of the very long but temporary structural change where the economic growth opportunities emanating from the discovery of oil production appear to be wasted in a gradual fashion.

Figure 8 reports the composition of covariate-restricted synthetic control groups for Venezuela. In the specification where weakly performing covariates (i.e., $v < 0.01$) are excluded from the battery, the prediscovery growth trajectory of actual Venezuela can be best reproduced as a convex combination of the growth and development characteristics of Burma (64%), Mexico (22%), Chile

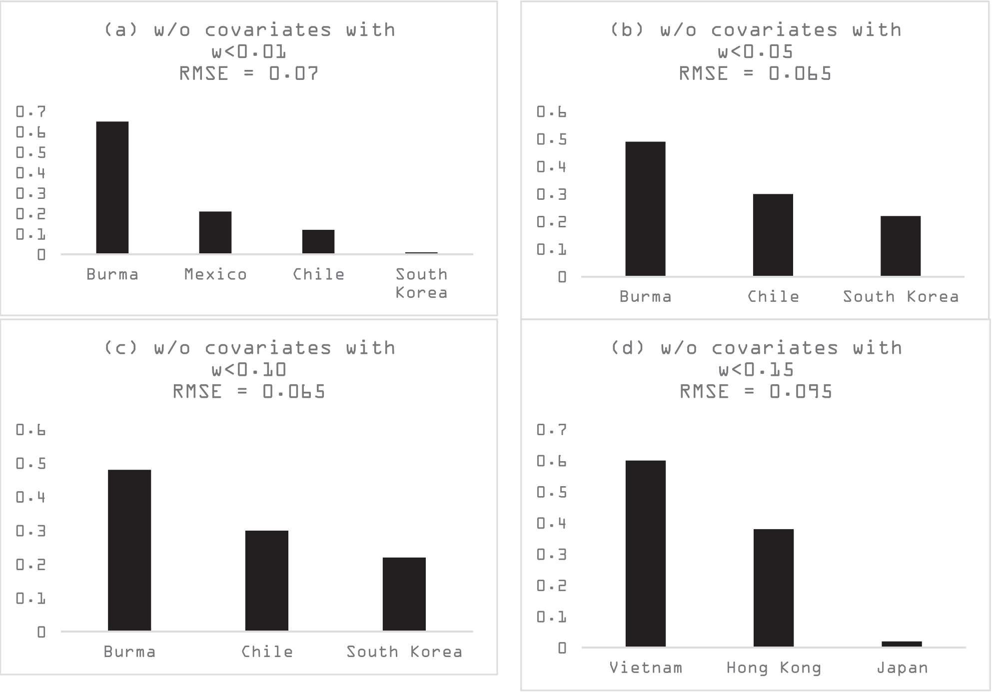

FIGURE 8 Composition of covariate-restricted synthetic control groups

(12%), and South Korea (2%), respectively. A very similar composition of the synthetic control group is apparent in alternate specifications with a more restrictive cutoff with no discernible differences in RMSE across these specifications, which, at least empirically confirms our argument.

5 | CONCLUSION

In this paper, we review the SCM, analyze some of its most pressing problems, propose solutions that scholars and practitioners might effectively deploy, and identify opportunities for the application of the method to comparative economic history in order to uncover the long-term effects of historical shocks and institutional changes for further historical investigation. The key question concerns the appropriate estimate of the counterfactual scenario that is ex ante unobserved to the econometrician, in response to a historical shock or change to provide plausible evidence of the impact.

Our discussion suggests a few discernible and easily tractable guidelines for practitioners in comparative economic history in undertaking synthetic control analysis. More specifically, we outline the contours of treatment selection and the necessity of undetectable treatment in the donor pool and the synthetic control group. If the treatment is present in the control group for the treated unit, the application of the synthetic control estimator to the treated unit most likely yields misleading results. By reviewing a couple of established case studies, we show that the unanticipated treatments or interventions of interest tend to have the largest potential to uncover the causal effect of interest provided that the control group is not directly tainted by the same shock. Moreover, we advocate a combined spatial and temporal placebo analysis of the underlying intervention. This implies that to gauge the significance of the effect, researchers are advised to proceed in two distinctive directions. First, they are advised to assign the intervention of interest to the treated unit into a deliberately false year to gauge whether pre-existing trends affect the counterfactual scenario. And second, they are advised to apply the synthetic control estimator to all untreated units and determine whether the effect appears to be unusually large for the treated unit. Furthermore, we discuss additional considerations pertaining to the ambiguity of the effect, specification search guidelines, and the detection of differential trends between the treated unit and its synthetic control group before and after the intervention.

To fill the void in the literature, we exploit the discovery of oil in Venezuela in the early 20th century to estimate the impact of natural resources on long-term economic performance. Our analysis serves two goals, namely: (i) to replicate the findings of some papers from the literature on resource curse and economic growth; and (ii) to present an opportunity in comparative economic history to elicit the impact of large-scale institutional and structural change on economic outcomes. By drawing on a large dataset of countries for the period 1870–2016, we are able to estimate the missing counterfactual scenario of Venezuela's economic growth trajectory in the hypothetical absence of the oil discovery. We show that the oil discovery triggered a large-scale but temporary deviation of growth trajectory from its long-run equilibrium that lasted from 1920 to the early 1970s when Venezuela attained one of the highest levels of per capita income worldwide. Starting with the nationalization of the oil industry in 1970, a massive growth premium of the oil discovery began to disappear and further deteriorated during the years of Chavez and Maduro political rule. Down to the present day, the growth trajectory of actual Venezuela has stagnated to the level implied by its synthetic control group, indicating the absence of a strong and robust high-quality institutional environment to prevent resource blessing from becoming a resource curse.

In essence, the SCM offers numerous opportunities to tackle the effects of historical shocks by estimating the counterfactual scenario. Such scenarios allow the identification of historical shocks either as gradual, temporary, or permanent changes of the long-run economic equilibrium (Garoupa & Spruk, 2019). These scenarios also permit a detailed investigation into the size and significance of the effect. Moreover, the method should not be confined to estimating the long-term growth impact of these events. For example, in principle, using comparative constitutional data, the method can be applied to discern the impact of specific events, both political and nonpolitical, on de facto constitutional protection by going back before the 19th century. By estimating the counterfactual scenario, the differences in the temporary and permanent effect of specific events on constitutional protection can be discussed with the reasonably clear empirical evidence at hand. Many other applications are possible as long as data are available. Armed with an additional battery of placebo tests, when appropriately applied without ex ante ideological or other hindsight bias, the synthetic control estimator can reveal insightful evidence on the long-term economic and social consequences of the institutional shocks that advance the state of causal inference both in economics and economic history. Such revelations can, and will, act as pointers to inform further and more traditional historical investigation.

6 ACKNOWLEDGEMENTS

The authors gratefully acknowledge Les Oxley for excellent editorial guidance and the anonymous referee for excellent comments which greatly improved the paper. The authors declare no conflict of interest and the usual disclaimer applies.

ORCID

David Gilchrist https://orcid.org/0000-0003-0937-8934
Rok Spruk https://orcid.org/0000-0002-7269-2339

Notes

[1] Botosaru and Ferman (2019) show that even in the absence of perfect balance assumption on both auxiliary covariates and pretreatment outcomes, it is possible to derive the bounds on the bias of synthetic control estimator since the presence of perfect balance does not per se imply an approximate balance for all covariates included in the specification.

[2] It should be noted that the application of the SCM per se is not limited to economic growth or other macroeconomic or microeconomic variables of interest. Although the focus of our investigation is economic growth over the long historical period, the usefulness of the method for non-economic variables such as demographic changes, election outcomes, institutional and political stability, and human capital stock among many others should not be downgraded.

[3] Against this backdrop, Di Stefano and Mellace (2020) introduce the inclusive synthetic control method as a modification of the originally proposed method by Abadie and Gardeazabal (2003). The inclusive synthetic control method allows the inclusion of units potentially affected directly and indirectly by the intervention in the donor pool, which is particularly suitable for the settings with multiple treated units where some of the units might be contaminated by the spillover effect of the intervention.

[4] One such example is highlighted by Melcarne and Spruk (2019). They examine the impact of the institutional transition from monarchy to democracy on the economic growth of Italian regions in the 100-year period from 1870 to 1970. To exclude potentially interfering regions, they effectively match every Italian region with the set of other countries that have not undergone the transition to democracy to prevent the impact of democratization from being contaminated by the presence of democracy in the potential donor countries.

[5] More recently, Gharehgozli (2017) estimates the effect of recent international sanctions on Iran's real GDP by comparing Iran with a sample of OPEC members where no sanctions were imposed, and performs a series of in-time placebo simulations where the timing of international sanctions is shifted to deliberately wrong years to evaluate the credibility of the effects. Moreover, Bilgel and Karhasan (2019) attempt to estimate the causal effect of separatist terrorism (i.e., proxied by the terrorist attacks of Kurdistan Workers Party) on economic development in Turkey over the last three decades. In a similar vein, they undertake a series of in-time placebo analyses where the dates of terrorist outbreak are assigned to the false years to determine whether synthetic control estimates reflect the effect of terrorism on economic development. Other examples of using in-time placebo analysis to evaluate the significance of synthetic control estimates also include the economic effects of Arab Spring (Echevarría and García-Enríquez, 2019, 2020), impact of export restrictions (Garcia-Lembergman et al. 2018), effect of natural resource discovery on income inequality (Hartwell et al. 2021), the economic impact of Hong Kong's sovereignty transfer (Li et al. 2021), effect of federalism on welfare spending using the case of Belgium (Arnold and Stadelmann-Steffen 2017) among several others.

[6] More specifically, 20 years after the earthquake, they find that the observed per capita income in Friuli Venezia Giulia is 23% higher than in the synthetic control group, while the observed per capita income of Irpinia is 12% lower than in the synthetic control group.

[7] Relatedly, Possebom (2017) applies the SCM to Brazilian city-level data for the period 1920–1999 to study the impact of the Free Trade Zone of Manaus, and compares Manaus against 48 MCAs (i.e., Minimum Comparable Areas) in the Brazilian Northern Region to isolate the impact of free-trade zone on a range of economic outcomes.

[8] In a similar vein, Ando (2015) uses the synthetic control estimator to estimate the impact of establishing nuclear power facilities in Japan in 1970s and 1980s on local per capita income. To establish a plausible counterfactual scenario, the author compares a single municipality where such a facility was established against a sample of non-nuclear municipalities.

[9] In a similar example. Emery and Spruk (2020) examine the causal impact of factional politics on long-run development by exploiting the within-country growth variation using the 1958 civil conflict in Lebanon. They perform an in-time placebo study and assign the 1958 conflict to 1975 civil war and find that the structural break between actual Lebanon and its synthetic version does not occur in the year of the civil war, but rather is apparent with the 1958 conflict.

[10] In particular, Abadie (2021) shows that placebo analysis is based on the permutation distribution of treatment effect obtained by iteratively reassigning the treatment to the donor pool and estimate the placebo effects. A complication may arise if synthetic control estimator is able to closely fit the trajectory of outcome before the intervention, the same may not be true for all units in the donor pool. As a partial remedy, the ratio of post-intervention and pre-intervention fit is proposed as test statistics as well as the use of permutation distribution for inference.

[11] To correct the biases possibly produced by synthetic control estimator, several empirical strategies have been proposed in the literature including the residualization of the outcomes with respect to covariates before computing synthetic control estimates (Doudchenko and Imbens 2016), reduction of regularization biases in inference methods for regression-based variants of synthetic control estimator (Arkhangelsky et al. 2021), use of resampling procedure to test for the effect of intervention that does not depend on post-intervention asymptotics (Masini and Medeiros 2021), and use of techniques allowing for the extrapolation outside the convex hull such as the use of elastic net penalties (Doudchenko and Imbens 2016), estimators with non-negative weight regularization (Hsiao et al. 2012, Li et al. (2019)), estimators using negative weights to redress bias-producing imperfect fit before the intervention (Ben-Michael et al. 2021)

[12] In the broadest sense, matrix completion estimation uses the observed elements of the matrix of outcomes in the control group to impute the missing elements of outcomes corresponding to the treated units. The estimator is particularly suitable in settings where both cross-sectional and temporal dimension of the data are large, and does not seem to be asymptotically biased, and also allows for the nonzero correlation with time-varying unobserved confounders (Ferman, 2021a, b).

[13] Extending this discussion with another example, Spruk and Keseljevic (2020) examine the causal impact of the Yugoslav war on the economic growth trajectory of former Yugoslav republics. They confront a significant variation in the composition of synthetic control groups across the treated countries. For instance, given the covariate weights, the economic growth trajectory of Slovenia before the Yugoslav war can be best synthesized by Switzerland (37%), Malta (28%), Iceland (12%), Japan (9%), and Oman (8%). On the other hand, Kosovo's prewar growth trajectory can be best reproduced by Cape Verde (32%), Malta (26%), Romania (23%), Mozambique (10%), and a few others with minor weight shares.

[14] Algeria, Argentina, Australia, Austria, Belgium, Brazil, Burma, Canada, Chile, China, Colombia, Czech Republic, Denmark, Egypt, Finland, France, Germany, Greece, Hong Kong, Hungary, India, Indonesia, Iran, Iraq, Ireland, Italy, Jamaica, Japan, Jordan, Lebanon, Malaysia, Mexico, Morocco, Nepal, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Slovenia, South Africa, South Korea, Spain, Sri Lanka, Sweden, Switzerland, Syria, Thailand, Tunisia, Turkey, United Kingdom, United States, Uruguay, Venezuela, Vietnam.

[15] These countries include Algeria, Canada, Colombia, Indonesia, Iran, Iraq, Malaysia, and Norway.

REFERENCES

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391–425.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, *93*(1), 113–132.

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*, *29*(1), 1–11.

Abadie, A., Diamond, A., & 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*(490), 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the Synthetic Control Method. *American Journal of Political Science*, *59*(2), 495–510.

Abadie, A., & L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, *116*(536), 1817–1834.

Absher, S., Grier, K., & Grier, R. (2020). The economic consequences of durable left-populist regimes in Latin America. *Journal of Economic Behavior & Organization*, *177*, 787–817.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., & Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics*, *121*(2), 368–391.

Adhikari, B., & Alm, J. (2016). Evaluating the economic effects of flat tax reforms using Synthetic Control Methods. *Southern Economic Journal*, *83*(2), 437–463.

Amjad, M., Shah, D., & Shen, D. (2018). Robust synthetic control. *The Journal of Machine Learning Research*, *19*(1), 802–852.

Ando, M. (2015). Dreams of urbanization: Quantitative case studies on the local impacts of nuclear power facilities using the Synthetic Control Method. *Journal of Urban Economics*, *85*, 68–85.

Andrews, D. W. (2003). End-of-sample instability tests. *Econometrica*, *71*(6), 1661–1694.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, *111*(12), 4088–4118

Arnold, T., & Stadelmann-Steffen, I. (2017). How federalism influences welfare spending: Belgium federalism reform through the perspective of the synthetic control method. *European Journal of Political Research*, *56*(3), 680–702.

Ashraf, Q., & Galor, O. (2013). The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development. *American Economic Review*, *103*(1), 1–46.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, *74*(2), 431–497.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, *116*(536), 1716–1730.

Barone, G., & Mocetti, S. (2014). Natural disasters, growth and institutions: A tale of two earthquakes. *Journal of Urban Economics*, *84*, 52–66.

Bauhoff, S. (2014). The effect of school district nutrition policies on dietary intake and overweight: A synthetic control approach. *Economics & Human Biology*, *12*, 45–55.

Becker, M., & Klößner, S. (2017). Estimating the economic costs of organized crime by synthetic control methods. *Journal of Applied Econometrics*, *32*(7), 1367–1369.

Becker, M., & Klößner, S. (2018). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics*, *5*, 1–19.

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789–1803.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Bilgel, F., & Karahasan, B. C. (2019). Thirty years of conflict and economic growth in Turkey: A synthetic control approach. *Defence and Peace Economics*, *30*(5), 609–631.

Billmeier, A., & Nannicini, T. (2009). Trade openness and growth: Pursuing empirical glasnost. *IMF Staff Papers*, *56*(3), 447–475.

Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, *95*(3), 983–1001.

Bohn, S., Lofstrom, M., & Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, *96*(2), 258–269.

Bologna Pavlik, J., & Geloso, V. (2018). *The Cuban revolution and infant mortality: A synthetic control approach* [Working Paper No. 3201194]. Social Science Research Network.

Boschini, A. D., Pettersson, J., & Roine, J. (2007). Resource curse or not: A question of appropriability. *Scandinavian Journal of Economics*, *109*(3), 593–617.

Botosaru, I., & Ferman, B. (2019). On the role of covariates in the synthetic control method. *The Econometrics Journal*, *22*(2), 117–130.

Bove, V., Elia, L., & Smith, R. P. (2017). On the heterogeneous consequences of civil war. *Oxford Economic Papers*, *69*(3), 550–568.

Brecke, P. (2001). Long-term patterns of violent conflict. In *Proceedings of the Uppsala Conflict Data Conference*.

Brunnschweiler, C. N. (2008). Cursing the blessings? Natural resource abundance, institutions, and economic growth. *World Development*, *36*(3), 399–419.

Busse, M., & Gröning, S. (2013). The resource curse revisited: Governance and natural resources. *Public Choice*, *154*(1–2), 1–20.

Calderon, G. (2014). *The effects of child care provision in Mexico* [Working Paper No. 2014-07]. Banco de Mexico.

Campos, N. F., Coricelli, F., & Moretti, L. (2019). Institutional integration and economic growth in Europe. *Journal of Monetary Economics*, *103*, 88–104.

Cao, J. & Dowd, C. (2019). Estimation and inference for synthetic control methods with spillover effects. arXiv preprint No. 1902.07343, arXiv.

Carrasco, V., De Mello, J. M. P., & Duarte, I. (2014). A decáda perdida: 2003–2013. Texto para Discussão. Departamento de Economia, Pontifíca Universidade Católica.

Castillo, V., Garone, L. F., Maffioli, A., & Salazar, L. (2017). The causal effects of regional industrial policies on employment: A synthetic control approach. *Regional Science and Urban Economics*, *67*, 25–41.

Cattaneo, M. D., Feng, Y., & Titiunik, R. (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, *116*(536), 1865–1880.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, *95*(5), 1549–1561.

Cerulli, G. (2019). A flexible Synthetic Control Method for modeling policy evaluation. *Economics Letters*, *182*, 40–44.

Coffman, M., & Noy, I. (2012). Hurricane Iniki: Measuring the long-term economic impact of a natural disaster using synthetic control. *Environment and Development Economics*, *17*(2), 187–205.

De Souza, F. F. A. (2014). *Tax evasion and inflation: Evidence from the nota fiscal paulista program*. [Master thesis, Pontifícia Universidade Católica].

Di Stefano, R., & Mellace, G. (2020). *The inclusive synthetic control method* [Discussion Paper on Business and Economics No. 14]. University of Southern Denmark.

Doudchenko, N., & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and Synthetic Control Methods: A synthesis* [Working Paper No. 22791]. National Bureau of Economic Research.

Donohue, J. J., Aneja, A., & Weber, K. D. (2019). Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, *16*(2), 198–247.

Dube, A., & Zipperer, B. (2015). *Pooling multiple case studies using synthetic controls: An application to minimum wage policies* [Discussion Paper No. 8944]. Institute for the Study of Labor.

Easterly, W., & Levine, R. (2016). The European origins of economic development. *Journal of Economic Growth*, *21*(3), 225–257.

Echevarría, C. A., & García-Enríquez, J. (2019). The economic consequences of the Libyan Spring: A synthetic control analysis. *Defence and Peace Economics*, *30*(5), 592–608.

Echevarría, C. A., & García-Enríquez, J. (2020). The economic cost of the Arab Spring: The case of the Egyptian revolution. *Empirical Economics*, *59*(3), 1453–1473.

Emery, T., & Spruk, R. (2020). *Long-term economic consequences of factious tensions: Evidence from Lebanon* [Working Paper No. 3538578]. Social Science Research Network.

Ferman, B. (2021a). On the properties of synthetic control estimator with many periods and many controls. *Journal of American Statistical Association*, *116*(536), 1764–1772.+

Ferman, B. (2021b). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association*, *116*(536), 1764–1772.

Ferman, B., & Pinto, C. (2017). *Placebo tests for synthetic controls* [Working Paper No. 78079]. Munich Personal Repec Archive.

Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, *39*(2), 510–532.

Ferman, B., & Pinto, C. (2021). Synthetic controls with imperfect fit. *Quantitative Economics*, *12*(4), 1197–1221.

Ferman, B., Pinto, C., & Possebom, V. (2018). *Cherry picking with synthetic controls* [Working Paper No. 85138]. Munich Personal Repec Archive.

Firpo, S., & Possebom, V. (2018). Synthetic Control Method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, *6*(2), 1–26.

Galiani, S., & Quistorff, B. (2017). The synth_runner package: Utilities to automate synthetic control estimation using synth. *The Stata Journal*, *17*(4), 834–849.

Garcia-Lembergman, E., Rossi, M. A., & Stucchi, R. (2018). The impact of export restrictions on production: A synthetic control approach. *Economía*, *18*(2), 147–173.

Garcia Ribeiro, F., Stein, G., & Kang, T. (2013). The Cuban experiment: Measuring the role of the 1959 revolution on economic performance using synthetic control. In *Proceedings of the 47th Annual Conference of the CEA*. HEC.

Garoupa, N., & Spruk, R. (2018). *Lone wolves, sleeping elephants or predator hawks? Institutional shocks and counterfactual long-run development paths* [Working Paper No. 3181899]. Social Science Research Network.

Garoupa, N., & Spruk, R. (2019). *Using Synthetic Control Method to estimate the long-run effects of institutional change. Seminar presentation*. University of Southern Denmark, May 14.

Garoupa, N., & Spruk, R. (2020). *Does European Union matter for long-run growth and evelopment? An exploratory analysis based on synthetic controls* [Working Paper No. 3568134]. Social Science Research Network.

Gathani, S., Santini, M., & Stoelinga, D. (2013). Innovative techniques to evaluate the impacts of private sector developments reforms: An application to Rwanda and 11 other countries. Presented at: In *Proceedings of the MPSA Annual Conference*, April 11–13.

Gharehgozli, O. (2017). An estimation of the economic cost of recent sanctions on Iran using the synthetic control method. *Economics Letters*, *157*, 141–144.

Gobillon, L., Magnac, T., & Selod, H. (2012). Do unemployed workers benefit from enterprise zones? The French experience. *Journal of Public Economics*, *96*(9–10), 881–892.

Gobillon, L., & Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, *98*(3), 535–551.

Grier, K., & Maynard, N. (2016). The economic consequences of Hugo Chavez: A synthetic control analysis. *Journal of Economic Behavior & Organization*, *125*, 1–21.

Hahn, J., & Shi, R. (2017). Synthetic control and inference. *Econometrics*, *5*(4), 1–12.

Hartwell, C. A., Horvath, R., Horvathova, E., & Popova, O. (2019). Democratic institutions, natural resources, and income inequality. *Comparative Economic Studies*, *61*(4), 531–550.

Hartwell, C. A., Horvath, R., Horvathova, E., & Popova, O. (2021). Natural resources and income inequality in developed countries: Synthetic control method evidence. *Empirical Economics*, *62*(2), 297–338.

Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, *84*(408), 862–874.

Hsiao, C., Steve Ching, H., & Ki Wan, S. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics*, *27*(5), 705–740.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic literature*, *47*(1), 5–86.

Hosny, A. S. (2012). Algeria's trade with GAFTA countries: A synthetic control approach. *Transition Studies Review*, *19*(1), 35–42.

Jales, H., Kang, T. H., Stein, G., & Garcia Ribeiro, F. (2018). Measuring the role of the 1959 revolution on Cuba's economic performance. *The World Economy*, *41*(8), 2243–2274.

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2021). Standard Synthetic Control Methods: The case of using all pre-intervention outcomes together with covariates. *Journal of Business & Economic Statistics*. Forthcoming.

Kirkpatrick, A. J., & Bennear, L. S. (2014). Promoting clean energy investment: An empirical analysis of property assessed clean energy. *Journal of Environmental Economics and Management*, *68*(2), 357–375.

Klößner, S., Kaul, A., Pfeifer, G., & Schieler, M. (2018). Comparative politics and the Synthetic Control Method revisited: A note on Abadie et al.(2015). *Swiss Journal of Sconomics and Statistics*, *154*(1), 1–11.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., & Sutton, M. (2016). Examination of the Synthetic Control Method for evaluating health policies with multiple treated units. *Health Economics*, *25*(12), 1514–1528.

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The economic consequences of legal origins. *Journal of Economic Literature*, *46*(2), 285–332.

Li, F., Thomas, L. E., & Li, F. (2019). Addressing extreme propensity scores via the overlap weights. *American Journal of Epidemiology*, *188*(1), 250–257.

Li, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, *115*(532), 2068–2083.

Li, J., Sangal, S., & Shao, L. (2021). Assessing economic impact of sovereignty transfer over Hong Kong: A synthetic control approach. *Applied Economics*, *53*(30), 3499–3514.

Maddison, A. (2007). *The world economy. Volume 1: A millennial perspective. Volume 2: Historical statistics*. Organization for Economic Cooperation and Development.

Madsen, J. B., Raschky, P. A., & Skali, A. (2015). Does democracy drive income in the world, 1500–2000?. *European Economic Review*, *78*, 175–195.

Maseland, R., & Spruk, R. (2020). *American exceptionalism and the benefits of statehood: An analysis of the growth effects of joining the United States of America* [Working Paper No. 3526844]. Social Science Research Network.

Masini, R., & Medeiros, M. C. (2021). Counterfactual analysis with artificial controls: Inference, high dimensions and nonstationarity. *Journal of the American Statistical Association*, *116*(536), 1773–1788.

Melcarne, A., & Spruk, R. (2019). In vincit qui se vincit: Institutional dependence and regional economic divergence in postwar Italy. In *New Frontiers in Law and Economics, Ringvorlesung*. Bucerius Law School. June 11.

McClelland, R., & Gault, S. (2017). *The Synthetic Control Method as a tool to understand state policy* (Research Report). Urban Institute.

Montalvo, J. G. (2011). Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. *Review of Economics and Statistics*, *93*(4), 1146–1154.

Muralidharan, K., & Prakash, N. (2017). Cycling to school: Increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*, *9*(3), 321–350.

Nunn, N., & Puga, D. (2012). Ruggedness: The blessing of bad geography in Africa. *Review of Economics and Statistics*, *94*(1), 20–36.

Papyrakis, E., & Gerlagh, R. (2004). The resource curse hypothesis and its transmission channels. *Journal of Comparative Economics*, *32*(1), 181–193.

Pinotti, P. (2015). The economic costs of organised crime: Evidence from Southern Italy. *The Economic Journal*, *125*(586), F203–F232.

Possebom, V. (2017). Free trade zone of Manaus: An impact evaluation using the Synthetic Control Method. *Revista Brasileira de Economia*, *71*(2), 217–231.

Robinson, J. A., Torvik, R., & Verdier, T. (2006). Political foundations of the resource curse. *Journal of Development Economics*, *79*(2), 447–468.

Rydberg, J., McGarrell, E. F., Norris, A., & Circo, G. (2018). A quasi-experimental synthetic control evaluation of a place-based police-directed patrol intervention on violent crime. *Journal of Experimental Criminology*, *14*(1), 83–109.

Sachs, J. D., & Warner, A. (1995). *Natural resource abundance and economic growth* [Working Paper No. 5398]. National Bureau of Economic Research.

Saunders, J., Lundberg, R., Braga, A. A., Ridgeway, G., & Miles, J. (2015). A synthetic control approach to evaluating place-based crime interventions. *Journal of Quantitative Criminology*, *31*(3), 413–434.

Smith, B. (2015). The resource curse exorcised: Evidence from a panel of countries. *Journal of Development Economics*, *116*, 57–73.

Spruk, R. (2019). The rise and fall of Argentina. *Latin American Economic Review*, *28*, 1–40.

Spruk, R., & Keseljevic, A. (2020). *Long-term effects of Yugoslav war*. Mimeo.

Spruk, R., & Kovac, M. (2020). Does a ban on trans fats improve public health: Synthetic control evidence from Denmark. *Swiss Journal of Economics and Statistics*, *156*(1), 1–32.

Vanderbei, R. J. (1999). LOQO: An interior point code for quadratic programming. *Optimization Methods and Software*, *11*(1–4), 451–484.

Wang, Y., & Zubizarreta, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, *107*(1), 93–105.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*(1), 57–76.

Yu, J., & Wang, C. (2013). Political risk and economic development: A case study of China. *Economic Research*, *26*(2), 35–50.

Zivot, E., & Andrews, D. W. K. (2002). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, *20*(1), 25–44.

How to cite this article: Gilchrist, D., Emery, T., Garoupa, N., & Spruk, R. (2023). Synthetic control method: A tool for comparative case studies in economic history. *Journal of Economic Surveys*, *37,* 409–445. https://doi.org/10.1111/joes.12493