Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Document de Treball 2020/05 1/39 pág.
Working Paper 2020/05 1/39 pág.

"Ensuring Stability, Accuracy and Meaningfulness in Synthetic Control Methods: The Regularized SHAP-Distance Method"

Daniel Albalate, Germà Bel and Ferran A. Mazaira-Font

UBIREA Institut de Recerca en Economia Aplicada Regional i Pública

Abstract

The synthetic control method (SCM) has been increasingly adopted to evaluate causal effects under quasi-experimental designs. However, SCM suffers from sound weaknesses that compromise its accuracy, stability and meaningfulness. The Regularized SHAP-distance synthetic control method (RSD-SCM) is proposed as solution. We evaluate the economic effect of the government formation deadlock in Spain, 2016. The deadlock did not negatively affect economic growth, as the economy grew 1.58% more without full government; standard SCM method overestimates the effect by 0.23 pp. We show that RSD-SCM offers higher accuracy and stability, while ensuring the economic meaningfulness of covariates used in building the counterfactual.

JEL classification: C32, E65, H11

Keywords: Synthetic control, SHAP, Regularization, Quasiexperiments, Causality, Government

Daniel Albalate: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: albalate@ub.edu

Germà Bel: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: gbel@ub.edu

Ferran A. Mazaira-Font: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: fmazaifo7@alumnes.ub.edu

Ensuring Stability, Accuracy and Meaningfulness in Synthetic Control Methods: The Regularized SHAP-Distance Method

By DANIEL ALBALATE, GERMÀ BEL AND FERRAN A. MAZAIRA-FONT*

The synthetic control method (SCM) has been increasingly adopted to evaluate causal effects under quasi-experimental designs. However, SCM suffers from sound weaknesses that compromise its accuracy, stability and meaningfulness. The Regularized SHAP-distance synthetic control method (RSD-SCM) is proposed as solution. We evaluate the economic effect of the government formation deadlock in Spain, 2016. The deadlock did not negatively affect economic growth, as the economy grew 1.58% more without full government; standard SCM method overestimates the effect by 0.23 pp. We show that RSD-SCM offers higher accuracy and stability, while ensuring the economic meaningfulness of covariates used in building the counterfactual.

JEL: C32, E65, H11

Keywords: Synthetic control; SHAP; Regularization; Quasi-experiments; Causality; Government.

Since the seminal works of Abadie and Gardeazábal (2003) and Abadie, Diamond and Hainmueller (2010), the synthetic control method (SCM) has been increasingly adopted as a technique to evaluate causal effects under quasi-experimental design. The method provides a practical solution to the evaluation of case studies in which either only a single – or very few – aggregate units are treated (countries, regions, cities, etc.) and is considered one of the most influential recent contributions to empirical policy evaluation. Athey and Imbens (2017, p.9), for instance, describe it as "arguably the most important innovation in the policy evaluation literature in the last 15 years". Thus, it is no surprise that the SCM has become very popular as a policy evaluation tool, especially where the use of other methods – randomized trials (RT), propensity score matching

* Albalate: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit), Universitat de Barcelona, John Keynes 1-11, 08034 Barcelona, Spain (e-mail: albalate@ub.edu); Bel: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit), Universitat de Barcelona, John Keynes 1-11, 08034 Barcelona, Spain (e-mail: gbel@ub.edu) ; Mazaira-Font: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit), Universitat de Barcelona, John Keynes 1-11, 08034 Barcelona, Spain (e-mail: ferran@arcvi.io)

(PSM), difference-in-differences (DiD), etc. – is not feasible or relies on overly strong assumptions. Unlike other regression-based quasi-experimental methods, the SCM creates a hypothetical counterfactual (the synthetic unit) by taking the weighted average of pre-intervention outcomes from selected donors (control units). The impact of treatment is quantified by the simple difference between the treated unit and its synthetic cohort after treatment (post-treatment period).

As discussed in a series of papers by its pioneering authors (see Abadie and Gardeazábal 2003; Abadie, Diamond and Hainmueller 2010 and 2015; and Abadie and L'Hour, 2019), the SCM has two main advantages over other methods, such as regression-based counterfactuals or nearest neighbor matching. First, by being constrained to non-negative weights that need to sum one, (unlike nearest neighbor matching) it does not impose a fixed number of matches and ensures sparsity, while avoiding negative weights or weights greater than one that otherwise would imply an unchecked extrapolation outside the support of the data as well as complicating the interpretation of the estimate. Second, weights are calculated to minimize the discrepancies between the treated unit and the synthetic control in the outcome and the values of certain matching variables or covariates. Thus, the SCM ensures that the synthetic unit reproduces the control unit not only in terms of the outcome, but also in terms of the drivers that explain the evolution of the outcome of the treated unit before treatment. These two properties mean the SCM has become an important framework for the evaluation of the effects of treatment in many settings (see, for example, Montalvo, 2011; Billmeier and Nannicini, 2013; Cavallo et al., 2013; Kleven, Landais and Saez, 2013; Bohn, Lofstrom and Raphael, 2014; Percoco, 2015; Acemoglu et al., 2016; Grier and Maynard, 2016; Kreif et al., 2016; Goin, Rudolph and Ahern, 2017; among many others).

In spite of the influential contribution made by the SCM, we argue that the method suffers from a number of critical weaknesses that erode the reliability and robustness of its causal estimates and, consequently, of its policy implications. The contribution of this paper is, therefore, twofold. First, we formally prove these weaknesses and explain their sources and consequences. Second, we propose a modified version of the SCM, coined as the regularized SHAP-distance synthetic control method (RSD-SCM), that overcomes these limitations without increasing complexity. In fact, our approach is simpler and more operational, since it breaks down the NP-hard problem of nested optimization into two independent problems of quadratic optimization with linear constraints. We show that current quasi-experimental methods for estimating covariate importance under the SCM are extremely unstable and highly dependent on the donor pool, even if we add or remove donors with no weight in the optimal synthetic control. Consequently, weight estimation is also affected. Moreover, we show that the estimation of covariate importance is not built on sound economic foundations, but rather it is the result of interpolation biases in order to minimize balance error on covariates. Therefore, it compromises the main spirit of the SCM: to build a benchmark that behaves as the treated unit not only in terms of the outcome but also in terms of the underlying drivers that explain this outcome.

The method we propose ensures robustness of the estimation of both covariate importance and the weights. By decoupling the estimation of covariate importance from that of weights, the RSD-SCM minimizes interpolation biases and guarantees economic sense. To estimate covariate importance, we use a new methodology for estimating feature importance suggested by Lundberg and Lee (2017, 2019): SHapley Additive exPlanation (SHAP) Values. This method allows us to analyze the marginal effects and average contribution of the different features of a model, even in the case of non-parametric models. Thus, we can obtain sound estimates for each unit of the relation of the different covariates with the outcome, and define a distance between the donor pool and the treated unit in terms of how covariates influence the outcome. To estimate weights, the procedure we use minimizes quadratic error in the pre-treatment outcome, introducing a pairwise distance penalization on the covariates, similar to the penalization parameter suggested for disaggregated data by Abadie and L'Hour (2019). Roughly speaking, we obtain a synthetic control that is the benchmark that best reproduces the pre-treatment outcome and whose behavior is explained by the same factors that explain the treated unit.

To illustrate the main advantages of the proposed RSD-SCM, we apply both methods to an evaluation of the causal economic effects of the ten-month long government formation impasse in Spain, after the December 2015 elections. In line with the approach taken by Albalate and Bel (2020) for the 18-month government formation deadlock in Belgium (the longest such impasse to date), we take the GDP per capita growth rate as the best indicator of overall economic performance. We use the SCM to build an appropriate counterfactual to identify and isolate the gap between Spain's actual GDP per capita growth rate and the rate at which it would have grown without a government formation deadlock. Our results indicate that the growth rate was not affected by government deadlock, ruling out any damage to the economy attributable to the institutional impasse. More importantly, the case study allows us to highlight the main differences between the SCM and RSD-SCM and the advantages of employing the latter.

The rest of this paper is organized as follows. First, we describe the standard SCM and we evaluate its stability, consistency and economic meaningfulness. In light of the limitations identified, in section II we propose a new regularized SHAP-distance synthetic control method (RSD-SCM) that overcomes the limitations of the standard SCM. In section III we apply both methods to empirically illustrate their differences and to highlight the advantages of RSD-SCM. The case study involves an estimation of the causal economic effects of a long government formation deadlock in Spain between December 2015 and October 2016. We discuss these findings, focusing, first, on the magnitude of the differences between the two methods (SCM vs. RSD-SCM) and, second, on the economic implications of the impasse. In section IV, we offer our main conclusions and considerations about the new method proposed.

I. The synthetic control method: An evaluation of its stability, consistency and meaningfulness.

A. The working of the SCM

The synthetic control method is a powerful approach to assessing causal treatment effects for aggregate units uniquely, or among very few, treated by a policy. This approach builds a counterfactual of a specific treated unit as a weighted average of a number of control units (the so-called donor pool), to reproduce what would have been its performance if it had not been exposed to the treatment and to identify, by its difference with respect to reality, the causal effect of the policy. The weights are computed by minimizing a vector distance between the treated unit and the donors over a series of pre-intervention covariates. In this section, we first describe the main features of the SCM and then we evaluate its consistency and stability.

The SCM assumes there are J control units and observations during T periods. Let X_{TU} be a $(K \times 1)$ vector of the outcome growth predictors of the treated unit (the covariates). Let $X = (X_1, ..., X_J)$ be a (KJ) matrix which contains the values of the same variables for the J possible control units. Both X_{TU} and X could include pre-treatment observations of the dependent variable. Let V be a diagonal matrix with non-negative components reflecting the relative importance of the different growth predictors. Let Y_{TU} be a $(T \times 1)$ vector whose elements are the values of the outcome of the treated unit for the T periods, and $Y = (Y_1, ..., Y_J)$ a $(T \times J)$ matrix whose elements are the values of the outcome of the control units. Then, the counterfactual is built as YW^*, where $W^* = (w_1^*, ..., w_2^*)$ is a $(K \times 1)$ vector whose elements are the weights of the control units in the counterfactual.

The vector of weights W^* is chosen to minimize the following objective function:

$$D(W) = (X_{TU} - XW)'V(X_{TU} - XW) \tag{1}$$

subject to $w_i \geq 0$ and $\sum_{i=1}^{K} w_i = 1$.
Matrix V is chosen as

$$V = \underset{V \in \mathscr{V}}{\operatorname{argmin}} (Y_{TU} - YW^*(V))'(Y_{TU} - YW^*(V)) \tag{2}$$

where $\mathscr{V}$ is the set of all non-negative diagonal $(K \times K)$ matrices, whose Euclidean norm is one.

Several contributions have recently been made aimed at extending the scope of use of the SCM and improving its accuracy and robustness. As regards the former, Powell (2018) suggested a way to estimate policy effects when the outcomes of the treated unit lies outside the convex hull of the outcomes of the other units. Since the treated unit may be part of a synthetic control for a non-treated unit, the post-treatment outcome differences for these units are informative of the policy effect. In recent studies, the SCM has been extended to contexts with disaggregated data, where samples contain large numbers of treated and untreated units, and interest lies in the average effect of the treatment among the treated (see Abadie and L'Hour, 2019). Building synthetic controls for each of the treated units as opposed to a synthetic control for the average treated unit has been proposed in order to minimize interpolation biases.

To increase SCM accuracy and robustness, studies have addressed three issues: the role of covariates, the estimation of weights and the best way to gauge the uncertainty of the estimated treatment effect. As regards the first of these, Doudchenko and Imbens (2016), Gobillon and Magnac (2016) and Kaul et al. (2015) showed that high accuracy can only be achieved if the outcome is also included as a covariate (for example, by including some lagged outcomes). However, by so doing, other covariates may become irrelevant. Studies of the estimation of weights have proposed different strategies to reduce interpolation biases. Abadie et al. (2015) suggested restricting the size of the donor pool and considering only units similar to the treated unit, while other studies introduced penalization parameters for the weights. Hastie et al. (2009, 2015), combined a Lasso and Ridge regularization to capture a preference for a small number of non-zero weights, as well as a preference for smaller weights. Likewise, Abadie et al. (2019) introduced a penalization parameter that trades off pairwise matching discrepancies with respect to the characteristics of each unit in the synthetic control against matching discrepancies with respect to the characteristics of the synthetic control unit as a whole. This reduces interpolation biases by prioritizing inclusion of units in the synthetic control that are close to the treated units and it ensures uniqueness, in the context of a large number of units in the donor pool, when single untreated units or different combinations of them may provide close matches to the treated units.

Finally, to gauge the uncertainty of the estimated treatment effect, the SCM compares the estimated treatment effect with the "effects" estimated from placebo tests in which the treatment is randomly assigned to a control unit (see Abadie and Gardeazábal, 2003). Building multiple synthetic controls by leaving countries out of the optimal control has also been proposed (Abadie et al., 2010). In this regard, Xu (2017) proposed a parametric bootstrap procedure to obtain confidence intervals of the estimates of the treatment effect.

Despite the interesting advances made by this literature, the works cited do not address the concerns about the SCM that we highlight and evaluate in the next subsection and which are critical for the quality of its outcomes. Our main concerns are related to SCM stability, consistency and meaningfulness.

B. An evaluation of the SCM as a bilevel problem

The SCM is characterized as a bilevel problem. Such problems are optimization problems (upper-level) that contain another optimization problem as a constraint (lower-level).

Definition 1. *For the upper-level objective function $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and lower-level objective function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ the bilevel problem is given by*

$$\begin{aligned}
&\min_{x_u \in X_u, x_l \in X_L} F(x_u, x_l) \\
&subject \;\; to \\
&x_l \in \underset{x_l \in X_L}{\operatorname{argmin}} \{ f(x_u, x_l) : g_j(x_u, x_l) \leq 0, \;\; j = 1, \dots, J \} \\
&G_k(x_u, x_l) \leq 0, \;\; k = 1, \dots, K
\end{aligned}$$

where $G_k : X_U \times X_L \to \mathbb{R}$, $k = 1, \dots, K$ denote the upper level constraints, and $g_j : X_U \times X_L \to \mathbb{R}$ represent the lower level constraints, respectively. Equality constraints may also exist that have been avoided for brevity.

Figure 1 illustrates a general bilevel problem solving structure involving interlinked optimization at both levels. Given a x_u vector, x_l^* is the optimal lower level vector for the lower level optimization. But, as seen in the figure, the solution (x_l^*, x_u) is not optimal for the upper level optimization given x_l^*.

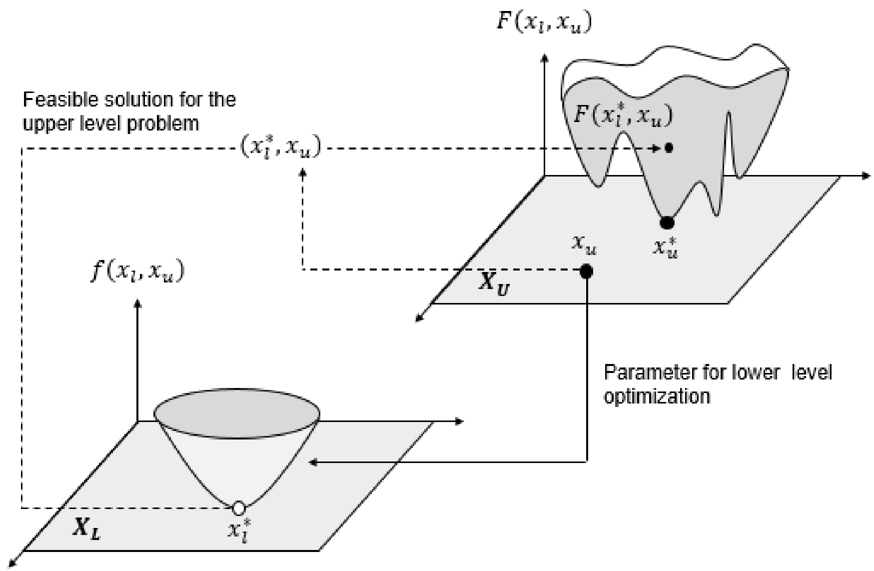


Figure 1. : **A general sketch of a bilevel problem.**

The SCM proposed by Abadie et Gardeazabal is a bilevel optimization problem of the form:

$$\min_{V,W} \ (Y_{TU} - WY)'(Y_{TU} - WY)$$

$$subject \ \ to$$

$$W \in argmin_W \{(X_{TU} - WX)'V(X_{TU} - WX) : 0 \leq w_i, \ \ j = 1, ..., K, \ \ \sum_{i=1}^{K} w_i = 1\}$$

$$v_j \leq 0, \ \ j = 1, ..., J; \sum_{i=1}^{J} v_j = 1$$

Bilevel programming is known to be strongly NP-hard (Hansen, Jaumar and Savard, 1992), and it has been proven that merely evaluating a solution for optimality is also a NP-hard task (Vicente, Savard and Judice, 1994). Moreover, the hierarchical structure may introduce difficulties such as non-convexity and disconnectedness even for simpler instances of bilevel optimization, which may cause solutions to be highly unstable to small perturbations and the algorithm to converge to different local optima.

In the particular case of the SCM method, the flaws of bilevel optimization imply that the solution V can be completely arbitrary and highly unstable to small perturbations. As a result, weights are also unstable and V does not offer reliable insights in terms of economic meaningfulness since it can be driven by interpolation biases.

Let us illustrate the aforementioned flaws with two simple examples.

Example 1: V completely arbitrary

Let us note by $W(V)$ the set of W that are solution of the lower-level problem, given V. Let us note by $\Omega = \{(W^1,V^1)..,(W^s,V^s)\}$ the set of weights and feature importance matrix such that $X^i_{TU} = W^jX^i$ for at least some covariate i but not all of them, and that W^j is a solution of the lower-level optimization, that is, $W^j \in W(V^j)$. Let us consider that there exists W^* such that $X_{TU} = W^*X$. Then:

i) For any V, $W^* \in W(V)$

ii) The optimal solution (or solutions) of the SCM problem is the solution of the problem

$$\min_{(W,V)\in\Omega\cup(W^*,V^*)} (Y_{TU} - WY)'(Y_{TU} - WY)$$

In particular, if W^* adjust better Y_{TU} than any $W^j \in \Omega$, the solution of the SCM admits any arbitrary V.

Proof

i) Given that $X_{TU} = W^*X$, we have that $D(W^*) \equiv (X_{TU} - W^*X)'V'(X_{TU} - W^*X) = 0$ for any V. Since the lower level function D is non-negative for any positive semi-definite matrix V, $W^* \in W(V)$ for any positive semi-definite V.

ii) Let us consider $V \notin \{V^1,...,V^s\}$. Let us assume that there exists $\hat{W} \neq W^*$ such that $\hat{W} \in W(V)$. Since $D(W^*) = 0$ and $W \in W(V)$, then $D(\hat{W}) = 0$. But since V is positive semi-definite, there has to exists at least one covariate i such that $X^i_{TU} = \hat{W}X^i$, and that is impossible because $V \notin \{V^1,...,V^s\}$. Thus, for any $W \in W(V)$ with $V \notin \Omega$, $W = W^*$. Therefore, the solution of the bilevel problem has to be pair $(W,V) \in \Omega \cup (W^*,V)$ that minimizes the upper level problem.

Example 2: V unstable and driven by interpolation biases

Let us consider a donor pool formed by 15 units with 3 covariates each one: $X_i =$

(C_i^1, C_i^2, C_i^3), where C_i^j is the average value of the covariate j for the unit i during the time period. Let us consider that the covariates are distributed as:

$$\begin{aligned} C_i^1 &\sim Uniform(0,1) \\ C_i^2 &\sim 0.5C_i^1 + Uniform(0,1) + N(0,0.1) \\ C_i^3 &\sim 0.5C_i^1 + Uniform(0,1) + N(0,0.1) \end{aligned}$$

Let us consider that the growth rate of X_i for the period t, $\gamma_{i,t}$ is defined as:

$$\gamma_{i,t} = \frac{0.2 + 0.2\hat{C}_i^1 + 0.5\hat{C}_i^2 + 0.3\hat{C}_i^3}{100} + \varepsilon_t$$

where $\hat{C}_i^j = \frac{C_i^j - min(C^j)}{max(C^j) - min(C^j)}$ and $\varepsilon_t \sim N(0, 0.02)$.
Thus, we are ensuring that covariates are related with output, since growth rates depend on covariates. All covariates are given the same output value at $t = 0$, that for simplicity we take as 1. Therefore,

$$Y_i^t = \prod_{r=1}^{t} (1 + \gamma_{i,r}), \qquad t = 1, \dots, T$$

Let us defined the treated unit as:

$$Y_{TU}^t = \frac{Y_i^t + Y_j^t}{2} + N(0, 0.01)$$

$$C_{TU}^k = \frac{C_i^k + C_j^k}{2} + N(0, 0.05)$$

where X_i, X_j are two randomly selected donors such that $C_1^{i,j} < 0.5$, $C_2^{i,j} < 0.5$, and the correlation between X_i and X_j is higher than 0.7. Note that the treated unit is related with donors in terms of output and covariates.
Figure 2 shows 4 examples of simulated results. In each graph it is represented the R^2 value of the synthetic unit in the z axis when feature importance is $V = (x, y, 1 - x - y)$. As can be seen, the upper-level problem (the sum of squares, which is a linear transformation of the R^2) is highly non-convex and there are multiple local optima. Moreover, small variations in V can lead to huge changes in R^2. For example, in the first figure, the maximum R^2 is 0.948 and corresponds to $x = 0$ and $y = 1$. However, a small perturbation

can lead to the lowest R^2, -0.212, at $x = 0.05$ and $y = 0.95$.

In section III, we will present an empirical assessment of the numerical instability for

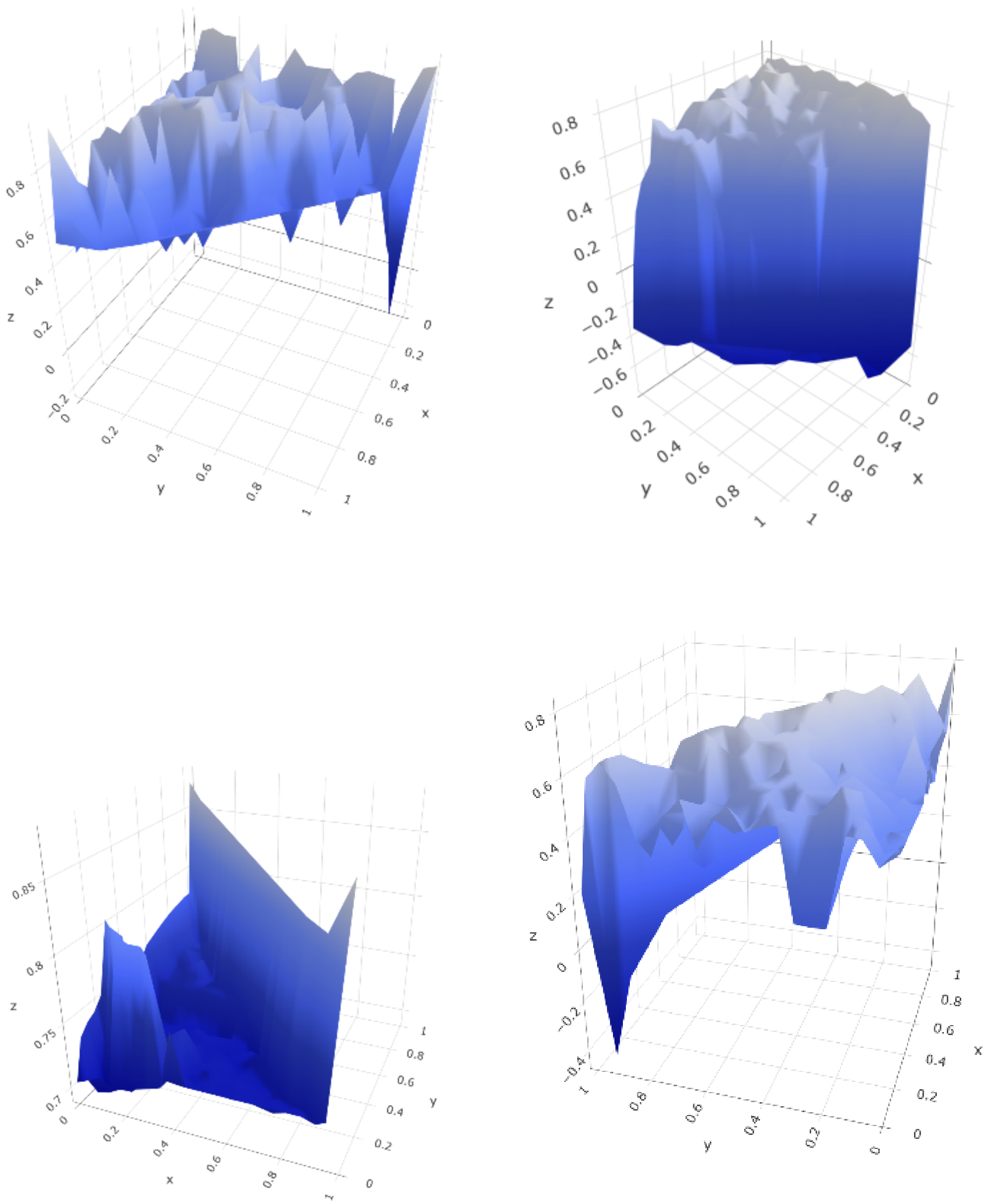


Figure 2. : **Four illustrations of goodness of fit in pre-treatment outcomes given V**

the case study of Spain's government deadlock. As we will see, current implementation of the SCM method can lead to very unstable results just by removing or adding to the donor pool units that are given no weights in the synthetic control. Although this is clearly counter-intuitive and should not be possible, it is due to the fact that the implementation of the optimization problem is done using an interior point method (Abadie et al., 2011). That is, weights can be given values close to zero, but not zero. Thus, although the final result is presented as 0, the real value for the algorithm could be of the order of 10^{-7} or 10^{-8} (depending on the margin parameter given to the function). Therefore, removing units with *zero* weight in the solution is equivalent to introduce a very small perturbation, which, as we have showed, can be devastating in terms of optimal parameters and goodness of fit.

II. The Regularized SHAP-distance based synthetic control method

The aim of this section is to propose and present a modification of the SCM that can guarantee economic meaningfulness and the stability of feature importance, at the same time as it increases the robustness of the estimation of weights and treatment effect. Our proposal is coined as the regularized SHAP-distance synthetic control method (RSD-SCM) and is designed as an operational alternative to the use of the SCM that involves less complexity than the standard approach due to the NP-hard nature of bilevel optimization.

In the previous section we showed that the minimization problem of SCM is defined over covariates and that feature importance estimation is nested to weights that can lead to considerable instability and a lack of economic meaningfulness. Therefore, we propose decoupling feature importance from weight estimation by defining the optimization problem of the SCM as a minimization of the error in the pre-treatment outcome adjustment, conditional to using units that are as similar as possible to the treatment unit. Moreover, we also present a concrete methodology for feature estimation and ensuring unit similarity that guarantees economic sense and stability.

A. *Modified optimization function*

Let us note by $d(X_{TU}, X_i)$ a distance between the the treated unit and the unit i, dependant on their respective vector of covariates X_{TU}, X_i. The vector of weights in our modified method is chosen to minimize

$$D(W) = \frac{\sum_{t=1}^{T}(Y_{TU}^{t} - WY^{t})^{2}}{\sum_{t=1}^{T}(Y_{TU}^{t} - \bar{Y}_{TU})^{2}} + \lambda \sum_{i=1}^{J} I(w_i) d(X_{TU}, X_i) \tag{3}$$

where $I(x)$ equals 1 if $x > 0$ and 0 otherwise, and $\lambda > 0$ is a regularization parameter that balances the potential trade-off between a pure minimization of the adjustment error and ensuring the similarity between units considered for the control group and the treated unit. The choice of λ is not uniquely determined and depends on several conditionings. For example, the stronger the relation between the covariates and output evolution, the more sense it makes to choose a higher value of λ. In the next section (II.B), we present a method for assessing the importance of λ and for choosing an adequate value.

It is important to note that the distance function is not linked to the weights, which in the SCM contributed to increasing instability and reducing economic meaningfulness. We should also stress that we consider the pairwise distance of each unit with respect to the treated unit and not only with respect to the synthetic unit. In other words, we require all the units entering the synthetic control to be similar to the treated unit so as to minimize interpolation biases (as suggested also in Abadie and L'Hour, 2019). Roughly speaking, this change is equivalent to saying that, for example, what most resembles a medium-size house is not the average of a small and a big house, but the average of two medium-size houses. In the next section we present a distance function especially suited for building the synthetic control.

B. SHAP-based distance

Intuitively, we would like to consider that a unit is similar to the treated unit if their outcomes evolved in a similar way before the treatment and for similar reasons. For example, a 99% correlation in the evolution of GDP per capita between two countries would tell us nothing about their similarity if one has an economy based on natural resources that grew because of a hike in petrol prices whereas the other's growth was attributable to manufacturing exports. In short, to define a distance between units it is critical we understand the relationship between their outcome and their covariates. To do so, we propose the following methodology. First, we build a model of the evolution of the outcome using covariates as explanatory variables. Second, we use one of the newest and most popular methods of model interpretation to estimate the average marginal contribution of each feature to each prediction of the model: the SHapley Additive exPlanation or

SHAP values. Finally, by estimating the SHAP values, we are able to define a distance based on feature importance and average contributions to outcome evolution.

Outcome evolution model

Let us note the growth rate of unit i by

$$g_i^t = \frac{Y_i^t - Y_i^{t-1}}{Y_i^{t-1}}$$

where $i \in \{1, ..., J, TU\}$. Recall that Y_{TU} is the $(T \times 1)$ vector containing the values of the outcome for the treated unit, and $Y = (Y_1, ...Y_J)$ the $(T \times J)$ matrix with values of the outcome for the control units.
Let us consider $F(X_i^s | s \in \{1, ..., t\})$ a model for g^t, that is

$$F(X_i^s | t \in \{1, ..., t\}) = g_i^t + \varepsilon_t$$

where ε_t is the error term at time t.
F could be a parametric linear model, such as linear regressions, or a more complex machine learning model, such as a gradient boosting tree. It may also include past information from covariates.

Shapley Additive Explanation Values

SHAP values have been proposed as a unified framework for assigning feature importance to parametric and non-parametric models (Lundberg and Lee, 2017 and 2019). The main idea underpinning SHAP values – a method developed in coalitional game theory – is that a prediction can be explained by assuming that each feature value of the instance is a "player" in a game where the prediction is the payout. Thus, the problem of feature importance is equivalent to distributing the "payout" fairly among the features. To calculate SHAP values for a data instance, we need to estimate the average marginal contribution of a feature value across all possible "coalitions" of the other features, compared to the average prediction for the dataset. In order to do so, a general method for estimating SHAP values is suggested by Lundberg and Lee (2017, 2019), known as Tree-SHAP. Roughly speaking, given an instance x, the method calculates the marginal effect of a feature i estimating how the concrete value of feature i on x modifies the output of the model with respect other instances that share some of the features with x but not i.

Formally, let us consider the subset S of the set of input variables V and $F_x(S) =$

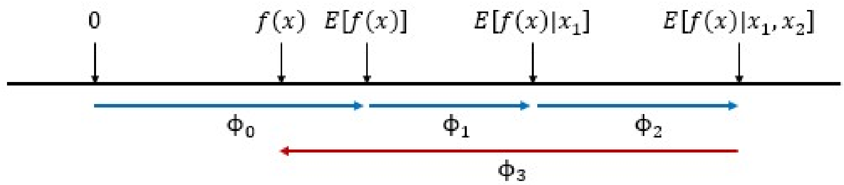


Figure 3. : **SHAP values explain the output of a function f as a sum of the effects ϕ_i of each feature being introduced into a conditional expectation.**

$E[F(x)|x_S]$ the expected value of the model F conditioned on the subset of input features S. Then, SHAP values are the combination of these conditional expectations:

$$\phi_i(x) = \sum_{S \subset V \setminus \{i\}} \frac{|S|!(|V|-|S|-1)!}{|V|!}[F_x(S \cup \{i\}) - F_x(S)]$$

where the combinations are needed because for non-linear functions the order in which features are introduced matters. For a linear model $F(x) = \sum_{i=1}^{k} \alpha_i x_i$, the SHAP value is straightforward:

$$\phi_j(x) = \alpha_j(\hat{x}_j - E(x_j))$$

Figure 3 shows how the SHAP values explain the output of a function f as a sum of the effects ψ_i of each feature being introduced into a conditional expectation.

Feature importance and SHAP-based distance

Let us note by $\phi_m(X_1^t)$ the SHAP value of the covariate m for the treated unit at time t. Then, we can estimate the relative importance of the covariate m in the outcome evolution of the treated unit, RI_m, as:

$$RI_m = \frac{\sum_{t=1}^{T} |\phi_m(X_1^t)|}{\sum_{j=1}^{J} \sum_{t=1}^{T} |\phi_j(X_1^t)|}$$

where J is the total number of covariates and T the total number of observations.
Therefore, we can define V as the diagonal matrix such that $V_{jj} = RI_j$. This matrix has economic sense, because it is exactly estimating the importance of each covariate on the outcome evolution of the treated unit before the treatment. It is also stable, in the sense that it relies on the stability of parameter estimation or model inclusion of the different variables. Thus, features whose relation with outcome is less robust will tend to not be

considered (for example, discarded in a linear model if their p.value is lower than 0.1 or 0.05) or be assigned lower relevance.

Having estimated V, let define AC_m^i the average contribution of feature m in outcome evolution for unit i:

$$AC_m^i = \frac{\sum_{t=1}^{T} \phi_m(X_i^t)}{T}$$

Then, we can define the distance between non-treated unit U_i and the treated unit TU as:

$$d(U_i, TU) = (AC^i - AC^{TU})' V (AC^i - AC^{TU}) \tag{4}$$

where $AC^i = (AC_1^i, ..., AC_J^i)$ and $AC^{TU} = (AC_1^{TU}, ..., AC_J^{TU})$ are the vectors containing the average contributions of the covariates for unit U_i and the treated unit TU.

Choice of regularization parameter

Let us consider $\lambda_0 > 0$, $\varepsilon > 0$, and $\lambda_1 = \lambda_0 + \varepsilon$. Let us note by $W(\lambda)$ the weights estimated given . Note that for ε small enough, $W(\lambda_0) = W(\lambda_1)$, since the I function in Equation 3 is binary.

Let us define a critical value λ, λ_i^{CV}, as the minimum value of λ such that the synthetic control is modified for any $\lambda_{i-1}^{CV} < \lambda < \lambda_i^{CV}$. We define $\lambda_0^{CV} = 0$. Note that $W(\lambda_i^{CV} + \varepsilon) = W(\lambda_i^{CV})$ for $\varepsilon \in [0, \lambda_{i+1}^{CV} - \lambda_i^{CV})$.

Let us consider $R^2(\lambda)$ as the R^2 of the synthetic control $Y(\lambda) = W(\lambda)Y$, and

$$d(\lambda) = \sum_{i=1}^{J} I(w_i(\lambda)) d(X_{TU}, X_i) / \sum_{i=1}^{J} I(w_i(\lambda))$$

the average distance of the countries with positives weights $W(\lambda)$.

It is clear that $R^2(0) \geq R^2(\lambda)$ and $d(0) \geq d(\lambda)$ for all $\lambda > 0$. Let us define also the error loss of λ as the ratio

$$EL(\lambda) = \frac{1 - R^2(\lambda)}{1 - R^2(0)}$$

and the similarity gain as

$$SG(\lambda) = \frac{d(0) - d(\lambda)}{d(0)}$$

The lower the EL the better, since this means the goodness of fit is near to the maximum possible. Likewise, the higher the SG the better, since this means that the countries in the synthetic control are closer to the treated unit. Therefore, we propose finding the optimal value of the regularization parameter, λ, that minimizes the EL:SG ratio

That is:

$$\lambda^* = \min_{\lambda \in CV_\nu} \frac{EL(\lambda)}{SG(\lambda)}$$

where CV_ν is the set of critical values of λ such that $R^2(\lambda) > \nu R^2(0)$. That is, critical values that ensure at less a certain level of goodness of fit, to prevent degenerated cases where the distance is almost no related with outcome. We recommend using $\nu = 0.9$ or 0.95.

For $\lambda = 0$, the EL:SG ratio is defined as ∞.

III. Illustration: The economic effects of the government formation deadlock in Spain, 2016.

Although hardly new, lengthy government formation processes in parliamentary regimes after a general election are becoming more usual in Europe. In the last two decades, there have been seven cases of government formation deadlocks lasting more than three months (see Table 1 for chronological list and specific details of each impasse).

Contrary to widespread claims that government deadlocks and the associated political instability harm a country's growth by disrupting economic policies that might otherwise promote better performance (Alesina et al., 1996; Angelopoulos and Economides, 2008; Aisen and Veiga, 2013), studies of recent impasses provide evidence that this might not always be the case. Using the SCM to build an appropriate counterfactual to reproduce Belgium's economic growth if it had had a full-powered government, Albalate and Bel (2020) reported a non-negative effect on economic growth during the 18 months of government deadlock in that country following the June 2010 election. The study suggests that certain characteristics peculiar to Belgium could be behind this (perhaps)

Table 1—: **Government formation deadlocks longer than three months in Europe.**

Country	**General Elections**	**Deadlock's duration**
Belgium	June, 2007	6 months
Belgium	June, 2010	18 months
Spain	December, 2015	10 months
Netherlands	March, 2017	7 months
Germany	September, 2017	6 months
Italy	March, 2017	3 months
Spain	April, 2019	10 months
Belgium	May, 2019	ongoing

counter-intuitive result: First, the country's highly decentralized multi-level governance, which assigns a considerable number of functions and powers to the communities and regions, at the same time as the European Union's institutions have absorbed some of the core functions performed by conventional Member States (Bouckaert and Brans, 2012; Hooghe, 2012); second, the existence of robust, efficient institutions, outside government, that played a positive role in protecting the economy from the difficulties of the impasse; and, third, the delay in fiscal consolidation that could have caused higher short-term economic growth than might otherwise have been expected.

A. *Spain's political deadlock.*

The general election held in Spain on December 20th, 2015 resulted in a fragmented political landscape following the emergence of two new political parties: Podemos (left-wing) and Ciudadanos (Cs) (right-wing). Table 2 records the names of the parties elected and the percentage of votes received by each.

In spite of winning the election, the Partido Popular (PP) (right-wing), who ruled Spain

Table 2—: **General elections results, Spain 2016**

	Nov. 2011	**Dec. 2015**		**June 2016**	
Political Party	Seats	% vote	Seats	% vote	Seats
Partido Popular (PP)	186	28.7	123	33.0	137
Partido Socialista (PSOE)	120	22.0	90	22.7	85
Podemos	-	20.7	69	21.1	71
Ciudadanos (Cs)	-	13.9	40	13.1	32
Izquierda Unida (IU)	9	3.7	2	-	-
Esquerra Republicana (ERC)	3	2.4	9	2.6	9
Democràcia Llibertat (DiL)/(CDC)	16	2.3	8	2.0	8
Partido Nacionalista Vasco (PNV)	5	1.2	6	1.2	5
Amaiur/ Euskal Herria (EHBildu)	7	0.9	2	0.8	2
Coalición Canaria (CC)	2	0.3	1	0.3	1

Note: Podemos was the main party of a left-wing coalition including regional parties En Comú Podem (Catalonia), Compromís (Valencia) and En Marea (Galicia)] in 2015. In the election 2016 included as well Izquierda Unida.

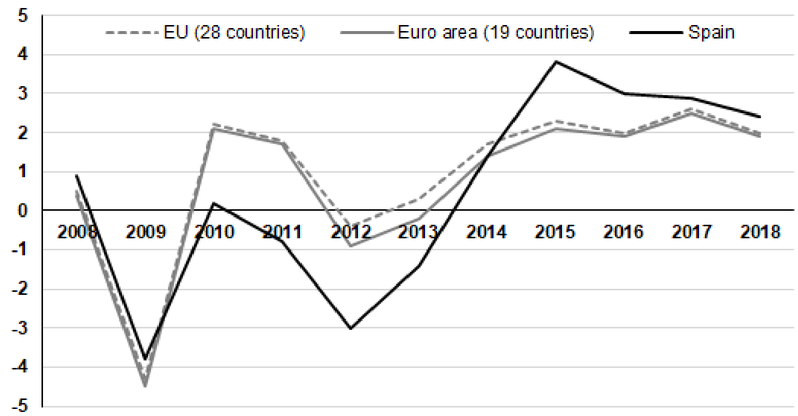
with an absolute majority between 2011 and 2015, lost 63 seats and with them its majority. Due to the numerous corruption cases in which leading members of the PP were then embroiled, the other main right-wing party, Cs, refused to facilitate a right-wing government and offered their votes to the Partido Socialista Obrero Española (PSOE).

Together the two parties controlled 130 of the chamber's 350 seats and needed either the support of Podemos or the abstention of the PP to achieve a sufficient majority to form a government. Neither of the two requirements was met and fresh elections were held in June 2016. For the first time since the restitution of Democracy in 1978, the failure of the legislature to form a government had resulted in a new call to the polls. As can be seen in Table 2, the 2016 election results reinforced the position of the PP, but these 14 additional seats were still not enough to form a government. After two months of negotiations clouded by huge uncertainty, Cs (along with the Coalición Canaria, a right-wing regional party in the Canary Islands) announced their support for Mariano Rajoy, PP's candidate for the Presidency. With 170 votes and the controversial, and what would prove decisive, abstention of the PSOE, Rajoy was re-elected President of Spain on October 29th, 2016, ending a ten-month long deadlock.
Despite this period of impasse and the limited powers of a caretaker government, Spain's economic performance did not appear to suffer greatly, despite claims from politicians and the media to the contrary as they pressed for a political deal. Indeed, even the Spanish Central Bank (Banco de España) published an article in 2017 estimating the negative effect of the political uncertainty of the previous year at just 0.1% of GDP, although this result was not statistically significant (see Gil, Pérez and Urtasun, 2017). If we observe the GDP growth rate (Figure 4), Spain's performance during 2016 was slightly higher than the EU average, and better than the euro area average, as it had been in 2015. However, as Albalate and Bel (2020) discuss in their evaluation of the 18-month government deadlock in Belgium, this comparison tells us only how Spain's performance compared to that of the other countries of Europe, but it offers no causal insights as to how it might have performed had it had a full-powered government. Thus, what we need is to build a counterfactual to reproduce how Spain would have performed in the absence of its government formation deadlock.

B. Results with the standard synthetic control method

To evaluate the robustness and meaningfulness of the SCM and the advantages of implementing our proposed RSD-SCM alternative, we compare the estimates provided by the two methods of the causal effects of this political deadlock. First, we apply the standard SCM, with and without outcome lags in the covariates, to show that in both cases covariate importance is highly unstable, highly dependent on the donor pool and lacking in economic meaningfulness. Second, we implement our proposed SHAP-distance

Figure 4. : **Real Gross Domestic Product per capita growth rate (2008-2016)**



based synthetic control method to show how this approach addresses and avoids the main weaknesses of SCM, providing more stable, accurate and meaningful estimates.

The donor pool used in the comparison includes a sample of the EU-28 countries. Malta and Luxembourg had to be excluded given the amount of missing data for some of the key predictors used in the analysis. Belgium was excluded since it was also affected by a lengthy government deadlock between 2010 and 2011, and Ireland because of the marked change in GDP pc in 2014 (26.3% growth rate) due to the reallocation of the intellectual property of large multinational firms.

Tables 3 and 4 report the pretreatment values of several variables typically associated with a country's growth potential and used as covariates, as well as their relative importance, for the case without and with lagged outcomes. Table 5 presents the weight matrix for the donor pool, where the synthetic weight is the country weight assigned to each country. When the lagged outcomes are not included, the synthetic Spain is made up of the four main contributors: Portugal (33.5%), France (30.7%), Greece (23.3%), and Italy (12.0%). Finland also plays a role, but only a minor one (0.4%). The rest of the countries do not contribute to the synthetic unit. When using this control to predict Spain's GDP per capita from 2001 to 2015, R2 is 92.60% and the mean absolute percentage error (MAPE) is 0.64%. When initial and final outcomes are included, the results are quite similar. The main contributors remain the same, although their relative importance changes. The minor role played by Finland disappears, and, instead, Denmark (3.7%)

and Sweden (2.6%) enter the synthetic control. The goodness of fit improves slightly (R^2 = 93.44%) and the MAPE remains unchanged at 0.64 Figures 5 and 6 show the GDP

Table 3—: **Covariate means and importance without including lagged outcome**

Predictor	**Spain**	**Synthetic**	**Pool**	**Importance**
Openness	57.03	60.13	98.00	54.1%
Low education	50.12	48.12	28.04	26.0%
Trade surplus	-1.37	-3.74	-0.51	12.8%
Unemployment	15.90	10.78	9.21	4.2%
High education	27.28	18.46	21.44	1.9%
Investment	24.02	20.81	22.31	1.1%
Debt	61.41	98.88	56.68	0.0%

Note: Average 2001-2015

Table 4—: **Covariate means and importance including initial and final outcome**

Predictor	**Spain**	**Synthetic**	**Pool**	**Importance**
GDP per capita^{2001}	22190.00	22120.28	19306.36	57.2%
GDP per capita^{2015}	23080.00	22523.79	22645.00	30.1%
Low education	50.12	48.83	28.04	7.5%
Openness	57.03	62.88	98.00	3.7%
Unemployment	15.90	10.81	9.21	0.8%
Trade surplus	-1.37	-3.79	-0.51	0.4%
Investment	24.02	20.75	22.31	0.1%
Debt	61.41	97.39	56.68	0.1%
High education	27.28	18.41	21.44	0.0%

Note: Average 2001-2015

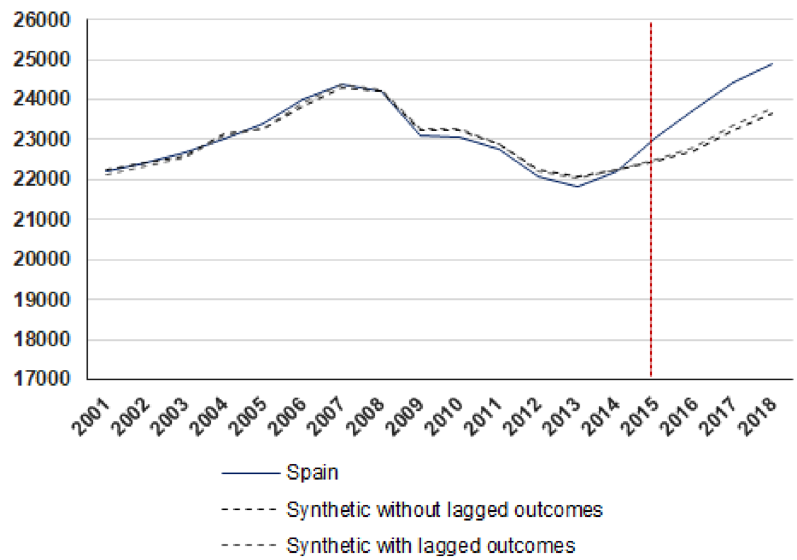
per capita evolution of the real and synthetic Spain, as well as their difference in GDP growth. In both cases, the growth rate during the 2016 deadlock was around 1.8 percentage points (p.p.) higher than expected, while in 2017 and 2018 the gap was reduced to 0.4 and 0.1 p.p. respectively.

To evaluate the robustness of the SCM, two placebo tests have been widely used: in-time and in-space. In the former, the SCM is applied considering that the treatment occurred in an earlier timeframe (i.e. the treatment is reassigned to occur during the pretreatment period) and so the control is built using observations up to this new moment in time. Thus, this test examines the uncertainty associated with making a prediction after the

Table 5—: **Percentage weight vector with and without including lagged outcome**

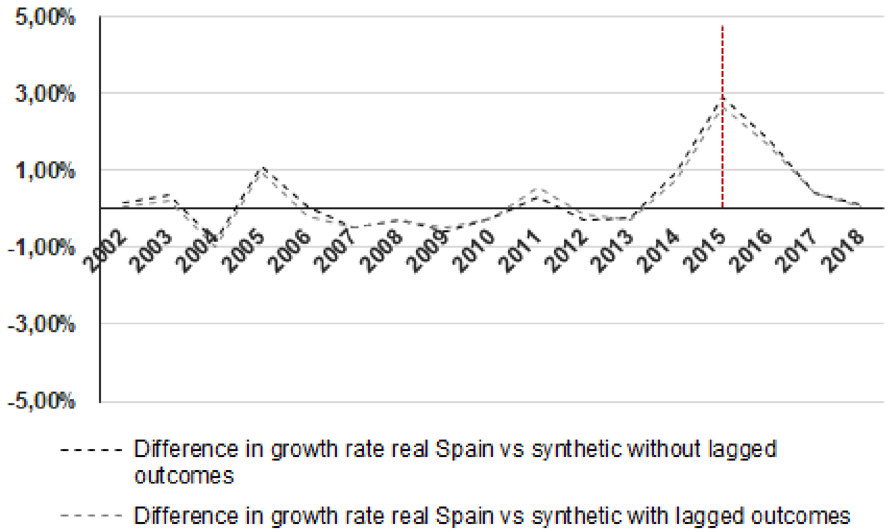
Country	Weight without lagged outcome	Weight with lagged outcome	Country	Weight without lagged outcome	Weight with lagged outcome
Austria	0	0	Hungary	0	0
Bulgaria	0	0	Italy	12.0	8.8
Croatia	0	0	Latvia	0	0
Cyprus	0	0	Lithuania	0	0
Czechia	0	0	Netherlands	0	0
Denmark	0	3.7	Poland	0	0
Estonia	0	0	Portugal	33.5	37.5
Finland	0.4	0	Slovakia	0	0
France	30.7	21.8	Slovenia	0	0
Germany	0	0	Sweden	0	2.6
Greece	23.3	25.4	UK	0	0

Figure 5. : **GDP per capita evolution: real vs synthetic Spain**



last observation considered for the estimation. In the second placebo test, the SCM is applied to the control units as if they too had been treated at the same moment of time as the treated unit. Hence, it tests the uncertainty associated with the volatility of outcomes

Figure 6. : **Difference in GDP per capita growth rate: real minus synthetic Spain**



of the control units during the treatment.

However, neither of these two tests evaluates the stability of covariate importance and weights, and their economic foundations, which are the main ideas on which the SCM is based. Even if the methodology passes the in-time and in-space placebo tests, it would be difficult to rely on its results if, for example, V had no relation with economic theory. Moreover, if a placebo test fails to confirm robustness, it is unable to tell us whether it is because the treatment had no significant effect or because the methodology was not properly applied and its accuracy could be improved (for example, by adding new covariates).

Here, therefore, we analyze the stability and economic meaningfulness of covariate importance and weights and show that SCM does not guarantee any of them, even in those cases when the methodology passes the placebo tests.

As Figures 5 and 6 show, in this particular case, including lagged outcomes has almost no impact in terms of the goodness of fit and the estimation of the treatment effect. All it achieves is to slightly modify the relative weights of the donor pool, especially as regards France, whose contribution falls from 30.7 to 21.8%, and Sweden and Denmark, which enter the synthetic Spain with weights of around 3% (Table 5). However, as recognized elsewhere, including Doudchenko and Imbens (2016), Gobillon and Magnac (2016) and Kaul et al. (2015), variable importance is greatly affected, and the covariates become almost irrelevant (Tables 3 and 4).

This might not, however, be important in terms of the consistency and robustness of V. If the covariates reflect the economy drivers, they could also be influencing lagged outcomes in the sense that countries with more similar values for these covariates would also have similar outcomes. This means that the importance of the covariates could be hidden behind the lagged outcomes; the only problem being that the inclusion of lagged outcomes makes it almost impossible to gain any economic insights from V and to judge whether the estimations are the result of an interpolation bias.

However, here, an analysis of V shows that its estimation is neither consistent with the economic foundations nor is it stable. First, if we turn our attention to the SCM without lagged outcomes, the most important variables are openness (54.1%) and low education (26.0%), while unemployment, investment and debt have almost no influence (4.2, 1.1 and 0%, respectively). The Spanish economy's cumulative growth per capita in real terms from 2000 to 2007 was 2.8 p.p. higher than that of the euro area, driven mainly by exceptionally high levels of investment due to the housing bubble (Akin et al., 2014). Total investment in Spain averaged 27.7% of GDP from 2000 to 2007, 5.2 p.p. higher than in the euro area. Once the crisis began, investment dropped significantly, reaching a minimum of 17.4% in 2013, almost 13 p.p. lower than its maximum in 2006. In the euro area the fall in investment was much lower: from a maximum of 23.4% in 2007 to a minimum of 19.7% in 2013. Unemployment more than tripled, from 8.2% in 2007 to a maximum of 27% in the first quarter of 2013, the highest level in the euro area. As a result, Spain's public debt almost tripled, growing from 35.8% of GDP in 2007 to 99.3% in 2015. In the euro area, however, the increase was much lower, from 65.9 to 90.8%.

Thus, quite simply it is meaningless to devise a similarity measure with respect to Spain that assigns no importance to debt, unemployment and investment, while at the same time assigning almost 70% of the importance to the degree of openness and the percentage of the population with a low education. Openness, for example, remained largely stable before and during the crisis. In the period 2000–2007, imports and exports accounted for 56.2% of GDP, while in 2008–2015 they represented 58.2%. In conclusion, openness and low education levels are not assigned a high level of importance because they are the main drivers of Spain's economy, but rather because a number of countries whose real GDP per capita evolution correlated highly with Spain's presented similar levels of openness and low education (see Appendix I).

Secondly, because of the interpolation bias, covariate importance, weights and goodness of fit are highly unstable and dependent on the donor pool. Table 6 shows the average importance and standard deviation for 100 simulations after removing three countries from the donor pool that were assigned no weights in the synthetic Spain. The standard deviation is higher than 50% of the average importance estimation for almost all covariates, both with and without lagged outcomes. As a result, the distance between the synthetic and real Spain is modified and, so, the weights are adjusted accordingly. As can be seen in Table 7, the weights are also highly unstable, except for Greece and Portugal (though, in the case of the latter, only when the lagged outcomes are included). Yet, weight instability may not necessarily compromise the SCM. Indeed, it might just be the result of the fact that the donors are so similar to each other that a small perturbation in V modifies the selection of one of them into the control. However, as Table 8 shows, the goodness of fit is significantly affected for the SCM without lagged outcomes and slightly affected for SCM with lagged outcomes. Given the high instability of the goodness of fit for

Table 6—: **Variable importance stability**

	No lagged outcomes		**With lagged outcomes**	
Predictor	**Mean**	**St. Dev**	**Mean**	**St.Dev**
Debt	2.40 %	4.14%	0.28 %	0.53%
Unemployment	5.57%	2.83%	1.86%	1.85%
Openness	41.3%	22.54%	11.40%	12.16%
Investment	3.38%	2.85%	0.73%	1.18%
Trade	14.36%	15.52%	3.51%	4.91%
Low education	29.12%	18.06%	12.30%	13.57%
High education	3.85%	3.13%	0.87%	2.02%
GDP per capita ^{2001}			40.99%	22.11%
GDP per capita ^{2015}			28.03%	10.50%

Note: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool

the SCM without lagged outcomes, an in-time placebo test using 2012 as the treatment year should identify this lack of consistency and robustness. However, the same does not hold true for the SCM with lagged outcomes (Figure 7). Yet, in both cases, the placebo test fails to provide any information as to why the methodology works properly or not, or whether it can be improved.

In conclusion, we have shown, first, that covariate importance may not be consistent with economic theory and provide no meaningful insights; second, that this lack of meaning is due to interpolation biases that make estimations highly unstable and dependent on irrelevant countries (i.e. countries with no weight) in the donor pool; and, third, that although

Table 7—: **Unit weights stability**

	No lagged outcomes		**With lagged outcomes**	
Country	**Mean**	**St. Dev**	**Mean**	**St.Dev**
Austria	0	0	0	0
Bulgaria	0	0	0	0
Croatia	1.17	4.04	1.44	3.27
Cyprus	0	0	0	0
Czechia	0	0	0	0
Denmark	0	0	1.11	3.47
Estonia	0.47	1.95	0	0
Finland	12.14	14.15	0.35	1.36
France	22.49	16.97	24.16	14.27
Germany	0	0.01	0.09	0.69
Greece	23.38	5.38	22.65	5.68
Hungary	0	0	0	0
Italy	3.90	7.27	11.89	11.84
Latvia	0	0.02	0	0
Lithuania	0	0	0	0
Netherlands	0.05	0.41	0	0
Poland	0.05	0.41	0.03	0.24
Portugal	36.33	10.04	34.07	4.64
Slovakia	0.02	0.13	0.47	2.11
Slovenia	0	0	0	0
Sweden	0	0.02	3.13	4.86
United Kingdom	0	0	0.60	3.45

Note: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool.

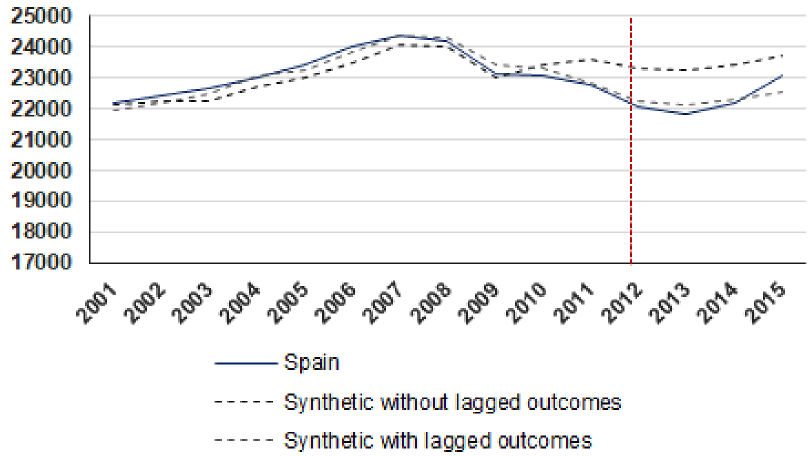
Table 8—: **Goodness of fit stability**

	No lagged outcomes		**With lagged outcomes**	
Measure	**Mean**	**St. Dev**	**Mean**	**St.Dev**
R^2	81.42%	14.03%	91.17%	2.84%
MAPE	1.16%	0.43%	0.77%	0.17%

Note: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool.

including lagged outcomes may make the results more robust in terms of goodness of fit, it does not solve the problem of meaning and stability of covariate importance; moreover, it also tends to make the other covariates irrelevant, thus compromising the main idea behind the SCM. Finally, we have also shown that standard robustness checks, such as the in-time placebo test, may be unable to identify these flaws and to suggest any strategy to improve the results.

Figure 7. : **Placebo test for Spain**



C. Results with the regularized SHAP-distance synthetic control

In this subsection we build the synthetic Spain adhering to the strategy described in Section III; that is, we build a model of GDP per capita growth, define a distance using SHAP values, select a regularization parameter and estimate optimal weights. We considered a linear model of the GDP per capita growth rate from 2001 to 2015, using as our explanatory variables the covariates used in the previous subsection and all the countries in the donor pool including Spain. The results are presented in Table 9 (variables with gr indicate growth rates of the covariate). Note that while the covariates are able to explain 73.33% of the variation in economic growth, around 25% of the variation remains unexplained. Thus, a synthetic control that relies solely on the covariates would not be sufficiently accurate or robust. Table 14 shows the feature importance of the different covariates of economic growth in Spain and the donor pool. According to the RSD-SCM results, Spain's economic evolution has been characterized primarily by high levels of unemployment and debt growth. Conditional convergence has had a much lower impact on Spain than it has had on the donor pool, mainly because Spain's GDP was already very close to the average of the selected countries (a 6% difference, on average, during

Table 9—: **Real GDP per capita growth model (OLS)**

Variable name	**Estimate**	**Std. Error**	**t value**	**Pr($>\|t\|$)**
Intercept	3.722e-02	(4.128e-03)	9.016	$<$ 2e-16
invest_gr	3.569e-03	(1.052e-03)	3.391	0.00078
openness_gr	1.692e-03	(1.704e-04)	9.930	$<$ 2e-16
gdp_pc_lag1	-9.058e-07	(9.574e-08)	-9.461	$<$ 2e-16
debt_gr	-1.705e-03	(2.070e-04)	-8.236	4.24e-15
unemp_gr	-6.046e-03	(8.767e-04)	-6.897	2.73e-11
educlow	-1.945e-04	(6.552e-05)	-2.969	0.00321
educhigh	2.786e-04	(1.331e-04)	2.093	0.03709
Residual standard error: 62.56 on 329 degrees of freedom Multiple R-squared: 0.7333, Adjusted R-squared: 0.7277 F-statistic: 129.2 on 7 and 329 DF, p-value: $<$ 2.2e-16				

Note: Error weighted by population

the period).

Using covariate importance, we define the distance between countries in the donor pool

Table 10—: **Covariate importance**

Predictor	**Importance Spain**	**Importance pool**
Unemployment	30.03	17.43
Debt	27.78	15.74
Openness	11.94	21.48
Investment	11.48	11.09
Low education	11.13	4.50
High education	4.19	4.39
Conditional convergence (GDP_lag1)	3.44	25.37

Note: SHAP feature importance during 2001-2015

and Spain as in Equation 4, but normalizing to be between 0 and 1. The corresponding results are presented in Table 13.

We applied the EL:SG ratio to select the optimal λ (as described in Section II.B). Results are presented in Table 12. It is worth pointing out that even in the case of no regularization and using a distance that was independent from Spain's pre-treatment adjustment, the average distance of countries in the synthetic control is much lower than the average distance of those in the donor pool (0.21 vs. 0.33). This means that the more

Table 11—: **Donor pool ordered by distance (normalized to 100)**

Country	**Distance**	**Country**	**Distance**
Cyprus	3.22	Hungary	26.22
Greece	8.53	Germany	31.71
Italy	9.78	Sweden	39.03
France	11.98	Latvia	40.05
Portugal	14.52	Czechia	40.15
United Kingdom	16.76	Denmark	51.24
Slovenia	18.61	Estonia	52.98
Croatia	22.15	Lithuania	54.20
Austria	22.76	Poland	54.82
Finland	24.18	Slovakia	64.58
Netherlands	25.25	Bulgaria	100.00

similar countries are to Spain, the more likely they are to be selected. Furthermore, we can see that normalizing by distance with the optimal λ, $\lambda^* = 0.1$, reduces the number of countries from six to four (parsimony) and almost halves the distance, with a loss of only 0.45 p.p. in R^2.

When using the modified synthetic method with $\lambda = 0.1$, we obtain a counterfactual

Table 12—: **Choice of regularization parameter**

Critical values	$\mathbf{R^2}$	**Distance**	**EL:SG**	**Countries**
3.14	92.63	0.11	4.95	4
0.10	96.39	0.12	2.75	4
0	96.84	0.21	∞	6

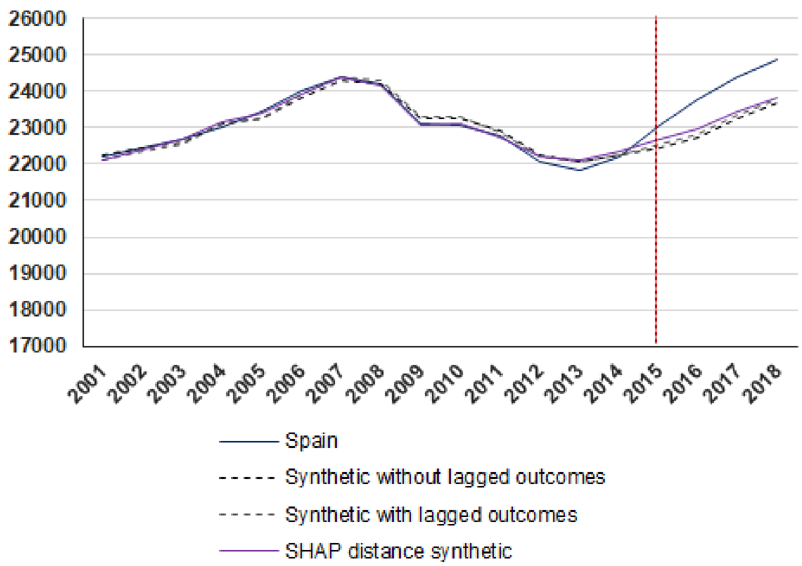
with four countries, R^2 = 96.39% and MAPE = 0.42%. Table 13 presents the estimated weights. Note that this counterfactual uses fewer countries than the standard method and obtains between 3 and 4 additional percentage points in R^2.

Figures 8 and 9 show that the different counterfactuals we have built estimate similar government deadlock impacts. In all cases, growth in 2016 was higher than expected, lying in a range between 1.58 p.p. (RSD-SCM) and 1.81 p.p. (SCM without lagged outcomes), a gap between the two estimates of 0.23 p.p. Thus, SCM overestimates the positive effect of the Spanish deadlock by a significant magnitude. As can be seen, the RSD-SCM ensures greater economic meaningfulness of feature importance, achieving

Table 13—: **Percentage weight vector using the SHAP distance SCM (only positive weights)**

Country	**Weight**	**Country**	**Weight**
Greece	18.2	Portugal	34.5
Italy	19.4	UK	27.9

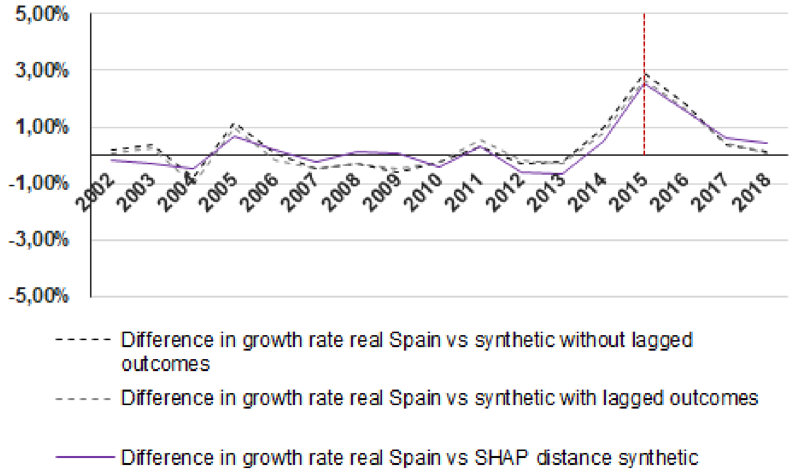
Figure 8. : **GDP per capita evolution: real vs synthetic Spain**



better results while reducing the number of parameters. However, our primary goal was to provide a more robust method. As Table 14 shows, the covariate importance estimates are highly stable, with standard deviations of 2 p.p., while thanks to the stability of the feature importance, distance is also stable. Moreover, in all simulations, the same countries are selected and assigned the same weights. Therefore, goodness of fit has no variation

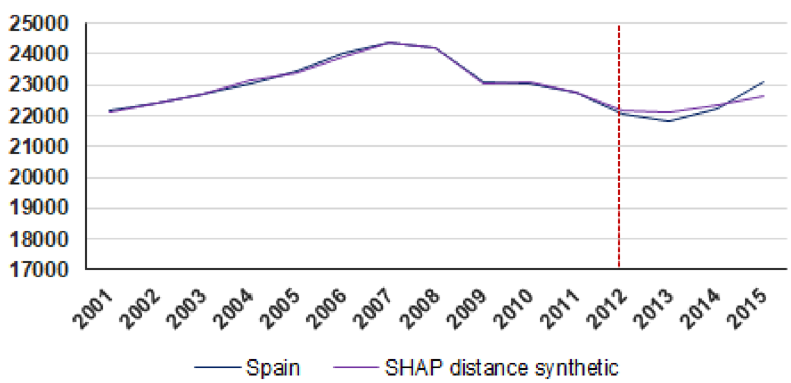
Finally, Figure 10 shows the results of the in-time placebo test. In the case of the in-space placebo test, we excluded countries whose MAPE for 2001–2015 was three times higher than Spain's. Thus, countries with an MAPE greater than 1.2% were excluded when we compared the base model to the best placebos (with eight countries surviving). The comparison showed a difference in the average treatment effect in 2016 for placebo

Figure 9. : **Difference in GDP per capita growth rate: real minus synthetic Spain**



countries of -0.006 p.p. in the growth rate and of 0.92 p.p. in the standard deviation. The treatment effect for Spain is estimated at 1.58, which is higher than 0 at a 7.8% confidence level, assuming a normal distribution of the placebo estimates.

Figure 10. : **Placebo test in time for SHAP distance synthetic Spain**



IV. Conclusion

The synthetic control method has been an influential innovation in quasi-experimental design, combining as it does elements of matching and difference-in-differences, and

Table 14—: **Feature importance stability**

Predictor	**Mean**	**St. Dev**
Unemployment	29.81%	(2.42%)
Debt	28.29%	(1.43%)
Openness	11.59%	(1.49%)
Investment	11.30%	(1.86%)
Low education	10.90%	(1.38%)
High education	4.61%	(0.47%)
Conditional convergence (GDP_lag1)	3.49%	(2.11%)

Note: Results over 100 simulations removing 3 countries with no weight in the synthetic Spain built with all the donor pool

providing a systematic approach to building a counterfactual. Similarly, it offers new opportunities for evaluating causal treatment effects in single – or in very few – aggregate units of interest. The method's impact on the empirical policy evaluation literature has been far-reaching and continues to grow, with its application in an increasing number of disciplines, including Economics, Political Science, Epidemiology, Transportation, Engineering, etc.

The SCM is credited with many advantages, including its transparency, sparsity, and the safeguard it provides against specification searches. Nevertheless, we have shown in this paper that the SCM also suffers from a number of critical drawbacks and limitations, some of them directly derived from its being characterized as a bilevel problem. In short, we have shown that (1) the covariate importance may not be consistent with economic theory, thus eroding the model's meaningfulness; (2) estimates are unstable – due to the interpolation bias and the nested nature of the optimization problem – and overly dependent on irrelevant countries in the donor pool; and, (3) including lagged outcomes does not solve the problem of meaning and the stability of covariate importance – even if the goodness of fit improves – but rather it makes other covariates irrelevant, compromising the main idea underpinning the SCM.

As an alternative to the SCM, we have proposed the Regularized SHAP-distance synthetic control method (RSD-SCM), which overcomes the main limitations of the standard method by decoupling feature importance from weight estimation and by providing a new methodology for feature estimation and unit similarity that ensure meaningfulness and stability.

Here, both methods were used to evaluate the effects on GDP growth of a ten-month government formation deadlock in Spain and their results compared. We provide evidence, consistent with Albalate and Bel (2020), refuting the negative economic effects of lengthy impasses in government formation. Thus, not only did Spain's economy not suffer any damage, it actually benefited by 1.58 p.p.; however, and more importantly in the context of this paper, the SCM overestimates these causal effects by 0.23 p.p. with respect to the results obtained using the RSD-SCM. Moreover, we have demonstrated that the RSD-SCM is a more stable, accurate and meaningful method than the standard SCM.

REFERENCES

Abadie, Alberto and Javier Gardeazábal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." American Economic Review 93 (1):113-32.

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." Journal of the America Statistical Association 105 (490): 493-505.

Abadie, Alberto, Alexis Diamond, Jens Hainmueller. 2011. "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies." Journal of Statistical Software 42 (13):1-17.

Abadie, Alberto, Alexis Diamond and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." American Journal of Political Science 59 (2): 495-510.

Abadie, Alberto and Jérémy L'Hour. 2019. "A penalized synthetic control estimator for disaggregated data." http://economics.mit.edu/files/18642 Accessed 2020-02-18

Acemoglu, Daron, Simon Johson, amir Kermani, James Kwak and Todd Mitton. 2016. "The value of connections in turbulent times: Evidence from the United States." Journal of Financial Economics 121 (2): 368-91

Aisen, Ari and Francisco J. Veiga. 2013. "How does political instability affect economic growth?" European Journal of Political Economy 29: 151-167

Akin,Ozlem, José García-Montalvo, Jaume García-Villar, Peydró, José-Luis, Josep M. Raya. 2014. "The real estate and credit bubble: evidence from Spain." SERIEs: Journal of the Spanish Economic Association 5: 223-43.

Albalate, Daniel and Germà Bel. 2020. "Do government formation deadlocks really damage economic growth? Evidence from history's longest period of government formation impasse." Governance 33 (1): 155-71.

Alesina, Alberto, Sule Ozler, Nouriel Roubini and Phillip Swagel. 1999. “Political instability and economic growth.” Journal of Economic Growth 1 (2), 189-211.

Angelopoulos, Konstantinos and George Economides. 2008. “Fiscal policy, rent seeking, and growth under electoral uncertainty: theory and evidence from the OECD.” Canadian Journal of Economics 41 (4): 1375-405.

Athey, Susan and Guido W Imbens. 2017. “The State of Applied Econometrics: Causality and Policy Evaluation.” Journal of Economic Perspectives 31 (2): 3-32.

Billmeier, Andreas ad Tommaso Nannicini. 2013. “Assessing economic liberalization episodes: a synthetic control approach.” The Review of Economics and Statistics 95 (3): 983–1001.

Bohn, Sara, magnus Lofstrom and Steven Raphael. 2014. “Did the 2007 legal Arizona workers act reduce the State’s unauthorized immigrant population?” The Review of Economics and Statistics 96 (2):258-69.

Bouckaert Geert and Marleen Brans. 2012. “Governing without Government: Lessons from Belgium’s Caretaker Government.” Governance 25 (2), 173-76.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy and Juan Pantano. 2013. “Catastrophic natural disasters and Economic Growth.” The Review of Economics and Statistics 95(5): 1549-61.

Doudchenko, Nikolay and Guido W. Imbens. 2016. “Balancing, regression, difference-in-differences and synthetic control methods: A synthesis.” NBER Working Papers, 22791.

Gobillon, Laurent and Thierry Magnac. 2016. “Regional policy evaluation: Interactive fixed effects and synthetic controls.” Review of Economics and Statistics 98 (3): 535–51.

Goin, Dana E., Kara and Jennifer Ahern. 2017. “Impact of drought on crime in California: A synthetic control approach.” PLoS ONE 12(10): e0185629.

Grier, Kevin and Norman Maynard. 2016. “The economic consequences of Hugo Chavez: A synthetic control analysis.” Journal of Economic Behavior Organization 125, 1-21.

de Jong, Jaspr, Marien Ferdinandusse, Josip Funda and Igor Vetlov. 2017. “The effect of public investment in Europe: a model-based assessment.” ECB - Working Paper Series, No. 2021.

Hansen, Pierre, Brigitte Jaumard and Gilles Savard. 1992. “New branch-and-bound rules for linear bilevel programming." SIAM Journal on Scientific and Statistical Computing 13 (5): 1194-217.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2009. The Elements of Statistical Learning: Data mining, inference, and Prediction. Second Edition. New York: Springer-Verlag.

Hastie, Trevor, Robert Tibshirani and Martin Wainwright. 2015. Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press

Hooghe, Marc. 2012. "Does Multi-Level Governance Reduce the Need for National Government?" European Political Science 11 (1): 90-5

Kleven, Henrik. J., Camille Landais and Emmanuel Saez. 2013. "Taxation and international migration of superstars: Evidence from the European football market." American Economic Review 103 (5), 1892–1924.

Kreif, Noémi, Richard Grieve, Dominic Hangartner, Alex Turner, Silviya Nikolova and Matt Sutton. 2016. "Examination of the synthetic control method for evaluating health policies with multiple treated units." Health Economics, 25 (12), 1514–528.

Lundberg, Scott. M., and Su-In Lee. 2017. "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems. NIPS proceedings 30, 1-10.

Lundberg, Scott and Su-In Lee. 2019. "Consistent Individualized Feature Attribution for Tree Ensembles." https://arxiv.org/abs/1802.03888v3. Accessed 2020-02-15

Montalvo, José 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): 1146–54.

Percoco, Marco. 2015. "Heterogeneity in the reaction of traffic flows to road pricing: a synthetic control approach applied to Milan." Transportation 42: 1063-79.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer and Manuel Schieler. 2015. "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates." MPRA Paper 83790, University Library of Munich, Germany.

Vicente, L., G. Savardand and J. Júdice. 1994. "Descent approaches for quadratic bilevel programming." Journal of Optimization Theory and Applications 81 (2): 379-99.

Xu, Yiquing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." Political Analysis 25(1): 57-76.

APPENDIX I: DESCRIPTION OF THE DATA

Country	GDP pc	Investment	Low education	High education	Trade surplus	Openness	Unemployment	Debt	Correl. w/ Spain[1]
Austria	34799.3	23.0	23.2	17.8	3.3	96.9	5.0	74.1	32.2
Bulgaria	4684.0	23.3	28.2	19.6	-8.2	107.5	11.3	26.8	24.7
Croatia	10212.7	24.3	26.1	15.0	-5.5	81.9	13.6	53.3	73.5
Cyprus	22330.0	20.0	31.9	30.3	-1.9	117.0	7.5	69.8	88.2
Czechia	14212.7	27.6	15.6	13.6	2.7	126.4	6.8	33.3	42.7
Denmark	44235.3	20.5	28.2	27.7	5.8	94.7	5.8	41.3	68.7
Estonia	11397.3	29.0	19.1	28.8	-1.8	140.9	9.6	6.6	47.5
Finland	34439.3	22.6	24.9	30.6	3.1	75.5	8.3	45.3	70.4
France	30627.3	22.0	33.0	25.3	-0.3	56.0	9.0	76.3	43.5
Germany	31506.7	19.9	21.8	21.9	5.3	76.0	7.6	70.1	7.1
Greece	19633.3	19.6	39.3	19.9	-7.9	56.0	14.7	131.8	86.8
Hungary	9894.7	22.6	26.4	16.3	2.0	147.7	8.1	69.0	51.9
Italy	27232.7	19.9	48.4	12.0	0.4	51.9	8.9	115.6	55.8
Latvia	8695.3	26.4	21.7	20.7	-9.0	103.4	12.1	25.8	43.2
Lithuania	8906.7	21.1	18.6	25.1	-4.7	123.2	11.4	27.5	26.7
Netherlands	37693.3	20.6	32.5	26.2	8.2	131.5	5.2	55.7	47.7
Poland	8680.7	20.1	19.9	17.0	-1.5	78.9	12.7	48.1	5.6
Portugal	16602.7	20.9	68.3	13.4	-5.6	69.5	10.4	91.3	88.5
Slovakia	11454.0	23.8	17.7	13.4	-0.2	155.0	14.6	42.6	21.0
Slovenia	17049.3	23.7	22.0	19.1	1.4	126.8	7.1	40.7	63.8
Spain	22962.7	24.0	50.1	27.3	-1.4	57.0	15.9	61.4	100.0
Sweden	38762.0	23.1	23.7	27.5	5.0	83.7	7.3	43.6	36.5
UK	30013.3	16.9	26.4	30.4	-1.8	55.5	6.1	58.1	55.8

1. Correlation between Spain's GDP and each country

Note: Average 2001-2015

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona