Eastern Journal of European Studies
DOI: 10.47743/ejes-2024-0101 ● JUNE 2024 ● VOLUME 15, ISSUE 1

Assessing the benefits of European integration: a comparative and algorithmic approach

Sara Casagrande[a], Bruno Dallago[a]
[a] University of Trento, Italy

Abstract
This article investigates the economic consequences of EU membership for the countries that established the EU in 1992. The Synthetic Control method is used in the frame of an algorithmic methodology that aims to guide the choice of donor pool countries and predictors by testing different combinations. The algorithm allows to judge the appropriateness of the research design by computing a set of synthetic countries able to improve the performance of the benchmark case (i.e., a synthetic country computed using an initial large basket of donor pool units and predictors). According to the results, the algorithm has been able to improve the counterfactual scenarios' precision and robustness for all tested countries. It shows that the economic effect of the EU membership has significantly varied among countries. Results suggest that the European integration process has not prevented persistent divergence and heterogeneity of growth paths among member countries.

Keywords: algorithm, comparative economic studies, European integration, European Union, synthetic control method

Introduction

The second World War made European integration fundamental in the process of building a future of peace and prosperity. Since the 1950s, European countries have stipulated treaties that confirmed their desire to deepen integration and cooperation. The Economic and Monetary Union (EMU) represented a cornerstone of the European project and the premise for the creation of a fiscal and political union. It is not easy to take stock of the economic advantages gained by EU member countries so far. Among the various counterfactual methodologies used for this type of investigation, the synthetic control (SC) method is becoming a widespread approach.

Despite the dominance of other well-established approaches (e.g., the regression discontinuity design and the difference-in-differences approaches), SC is gaining popularity especially among researchers engaged with macroeconomic

✉Sara Casagrande, Post-Doc Research Fellow at Faculty of Law, University of Trento, Italy; email: sara.casagrande@unitn.it.

issues, economic comparisons and policy evaluation analyses (Adhikari, 2022; Cerulli, 2022; Gilchrist et al., 2023). Indeed, SC enables us to judge whether a particular policy (or event) has been beneficial or not for a country by comparing its performance with that of a synthetic country, not affected by the policy (or event). More formally, the SC method "is a data-driven research design that provides a systematic way of constructing a synthetic unit from the weighted average of units from the comparison units, such that the constructed synthetic unit matches the path of the outcome variable as well as all the important predictors of the outcome variable for many periods before the policy's implementation" (Adhikari, 2022, p. 47). Various contributions to the literature have attempted to analyse the benefits of European integration by using the SC method. Most contributions study the impact of European membership (or the adoption of the euro) by comparing the GDP per capita of the OECD countries that joined the EU (or the euro) with a counterfactual represented by a weighted average of a combination of OECD countries that did not join the EU (the synthetic country or economy).

The reliability of the SC results can be ensured by various robustness checks. The literature is concerned with improving the accuracy of the solution provided by the SC method, and SC properties and reliability have been investigated (Ferman & Pinto, 2021). However, SC results may be influenced by the algorithm used for the construction of the counterfactual (Kuosmanen et al., 2021), the choice of units in the donor pool (Greathouse et al., 2023) and predictors (Abadie, 2021; Vives-i-Bastida, 2023). As a consequence, a growing literature is currently addressing how the SC method can be improved and extended (Ben-Michael et al., 2021; McClelland & Mucciolo, 2022); moreover, a more general classes of synthetic control estimators have been proposed (Doudchenko & Imbens, 2016). New estimators that build on insights deriving from the SC method, together with other methods (e.g., difference-in-differences) have been proposed (Arkhangelsky et al., 2021). All these recent developments, aimed to overcome the limits and extend the SC method have made its application less trivial and more prudent although many variations need to be carefully evaluated and compared.

In this article, we analyze the benefits of European integration through an algorithmic procedure. The algorithmic procedure aims to evaluate the research design as a whole, and compare a set of SCs generated by alternative combinations of predictors and donor pools units. We start from the original version of the SC to present the potentials of the algorithm and we leave an evaluation of the role of more sophisticated SC versions to future research. Indeed, an advantage of this algorithm is that it can be applied in combination with different algorithms for the computation of the counterfactual. In this article, we considered the most widely used in the literature, i.e. the *Synth* software[1], but the algorithm can be adapted to the use of

[1] We use the MATLAB algorithm at the base of the software *Synth* that implements synthetic control methods. This is probably the most used algorithm for implementing the synthetic

different types of SC (e.g., the penalized synthetic control estimator of Abadie & L'Hour, 2021). Different methods for the identification of predictors can be also implemented (in this article, we consider LASSO). As a consequence, the algorithm and the results should be considered as starting points for further research. It is worth remembering that this approach has the disadvantage of requiring computational time and capacity.

In detail, the algorithm starts from an initial basket of predictors and countries (i.e., a standard research design, with the countries and indicators most widely used in the literature) and verifies the presence of redundant predictors and identifies those countries which are poorly compatible with the tested one. Starting from this "benchmark case", the algorithm removes predictors and donor pool countries by testing different combinations. After this, it searches the best combination (or set of combinations) of predictors and donor pool countries that allow to improve counterfactual precision with respect to the benchmark case. The selection of the final set of predictors and donor pool countries is, therefore, the result of a data-driven approach. Before proceeding with the empirical analysis, we use simulations under a variety of data generating processes coherently with the approach used by Abadie and Vives-i-Bastida (2022) in order to demonstrate that the algorithm represents a valid tool to improve the performance of the SC with respect to a benchmark case.

We investigate the benefit of European integration for the 12 countries that established the EU in 1992 (treated units) and consider the 13 OECD countries that are not members of the EU as potential donor pool units. We start with a standard research design, most used in the literature, composed by an initial basket of variables and countries. The algorithm is able to improve the robustness of the analysis and the counterfactual scenarios' precision for all treated countries by selecting relevant predictors and countries to be used in SCs. It also offers useful insight about the appropriateness of the research design and robustness of the outcomes. The results indicate that looking for "winners" and "losers" may be misleading for a correct assessment of the effects of European integration, because the countries' performance is also influenced by national and community factors. In particular, the results indicate that the economic effect of EU membership has been very different among member countries, and the integration process has not solved divergences and heterogeneity among their growth paths, which is in line with the insights of a growing literature on this topic.

The article proceeds as follows. Section 1 presents a brief literature review about the SC method and its role in comparative European studies. In section 2, the

control method and it is available at https://web.stanford.edu/~jhain/synthpage.html. It is the companion software developed by Abadie et al. (2010; 2015). However, this algorithm can be tested with different software for calculating the counterfactual, such as the one proposed by Kuosmanen et al. (2021).

algorithmic methodology aimed at improving SC performance is explained and tested through simulations. In section 3, the algorithm is applied to the evaluation of the EU membership benefits and the results are presented and commented. The last section concludes.

1. The synthetic control method and European studies

1.1. Counterfactual analysis and the SC: an overview

Counterfactual analysis has attracted the attention of many scholars in recent years, especially of those interested in macroeconomic issues and case-study research. According to Mahoney and Barrenechea (2019, p. 306), "counterfactual analysis is intended to help analysts evaluate the effect of an actual world event by considering what would have happened if the event did not occur or occurred differently. Typically, these evaluations involve the formulation of 'what-if' arguments that rerun history with the counterfactual antecedent in place". Different methodologies exist for constructing counterfactual scenarios, which are used in particular in the evaluation of government policy impact: control-group analysis, cost-benefit analysis, regression analysis, selection and assistance modelling, self-assessment approach using survey techniques and shift-share analysis (Lenihan & Hart, 2004). However, these approaches add to the larger and better-known set of methodologies used in micro and macro analyses in the field of causal inference. Angrist and Pischke (2015) introduced the term "furious five" to identify the five most frequently used methods of causal inference: difference-in-differences method, instrumental variables, regression (OLS), regression discontinuity design, random assignment.

The synthetic control (SC) method is an alternative approach pioneered by Abadie and Gardeazabal (2003) and Abadie et al. (2010; 2015). It became popular in recent years and started to be applied in different research fields. According to Athey and Imbens (2017, p. 9), the SC method is "arguably the most important innovation in the policy evaluation literature in the last 15 years". More recently, the SC method has been described as a valid device "in undergraduate econometrics or capstone courses to estimate the impact of economic policies" (Adhikari, 2022, p. 46) and it "has proved to be very successful in addressing relevant policy evaluation analyses" (Cerulli, 2022, p. 339). It started to be used in different research fields such as comparative economic history (Gilchrist et al., 2023).

The purpose of the SC is to investigate what a particular unit's performance (e.g., a country) would have been in the absence of a particular intervention or event defined in general as the treatment (e.g., a policy, a treaty, a war, a crisis etc.). The SC method compares the performance of the unit (treated) with a weighted combination of unaffected units that have not been exposed to the treatment (donor pool units). The SC is the weighted average of the units in the donor pool. In the original version of the SC method, weights should be non-negative and sum to one. In this way, the weights are sparse: only some units in the donor pool contribute to estimate the counterfactual. For this reason, the SC has been considered a transparent data-driven methodology. Weights are chosen so that the resulting SC best resembles the pre-intervention values of predictors of an outcome variable of the treated unit (e.g., GDP if the units are countries).

Following Abadie (2021), consider to have data for J+1 units (j=1, …, J+1) and assume that the first unit j=1 is the treated unit, which is the only one affected by the treatment. The other units, j=2, …, J+1, are a collection of untreated units (i.e., not affected by the treatment) that can be part of the set of units to be used for the comparison. This set is called the donor pool.

Since we are considering panel data, the T periods are divided between a pre-treatment ($t< T_0$) and post-treatment period ($t>T_0$). The treatment corresponds to time $t=T_0$. For each unit j at time t, we observe the outcome Y_{jt} and a set of *k* outcome predictors X_{1j}, …., X_{kj}. The matrix $\mathbf{X_0}$ (k x J) collects all the predictors for all units.

If Y^I_{1t} is the potential response of the outcome under treatment (which is observed in data in period $t>T_0$) and Y^N_{jt} is the potential response of the outcome without treatment, the effect of the intervention for the affected unit j=1 in period *t* (with $t>T_0$) is:

$$\tau_{1t} = Y^I_{1t} - Y^N_{1t} \tag{1}$$

The challenge is to estimate and reproduce Y^N_{1t} for $t>T_0$, which is the counterfactual outcome and corresponds to the estimation of how the outcome of the affected unit would have evolved in time without the treatment, given that the evolution of the outcome in the presence of the intervention is observed ($Y^I_{1t}= Y_{1t}$).

SC aims to reproduce Y^N_{1t} using a combination of unaffected units which share similar characteristics of j=1. The synthetic control is a weighted average of the units in the donor pool and can be represented as $\mathbf{W} = (w_2, \dots, w_{J+1})^T$, so that the following SC estimator can be computed:

$$\hat{Y}^N_{1t} = \sum_{j=2}^{J+1} w_j\, Y_{jt} \tag{2}$$

Consequently, the estimate of the treatment effect can be computed as:

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j\, Y_{jt} \tag{3}$$

As demonstrated by Abadie et al. (2010), the SC estimator is asymptotically unbiased. The task is to compute **W**. Abadie and Gardeazabal (2003) and Abadie et al. (2010) suggest to choose the synthetic control **W*** that minimizes:

$$||\mathbf{X}_1 - \mathbf{X}_0\mathbf{W}|| = \left(\sum_{h=1}^{k} v_h \left(X_{h1} - w_2 X_{h2} - \cdots - w_{J+1} X_{hJ+1}\right)^2\right)^{1/2} \tag{4}$$

For each potential choice of a vector of positive constraints **V** = (v_1, …, v_k), it is possible to produce a synthetic control **W(V)**, which can be determined by minimizing the above equation through constrained quadratic optimization. Abadie and Gardeazabal (2003) and Abadie et al. (2010) choose **V** so as to minimize the mean squared prediction error (MSPE) of the synthetic control with respect to Y^N_{1t} (with $\Gamma_0 \subseteq \{1, 2, \ldots, T_0\}$):

$$\sum_{t \in \Gamma_0} \left(Y_{1t} - w_2(\boldsymbol{V})Y_{2t} - \cdots - w_{J+1}(\boldsymbol{V})Y_{J+1t}\right)^2 \tag{5}$$

As noted by Abadie (2021, p. 401), “the ability of a synthetic control to reproduce the trajectory of the outcome variable for the treated unit over an extended period of time […] provides an indication of low bias”. This ability can be measured using the root mean square prediction error (RMSPE). A small pre-treatment RMSPE implies a good fit of the SC. This ease in reading and interpreting the results and the transparency of the SC counterfactual estimate are surely two important strengths of this methodology.

The SC results are subject to different types of robustness checks and placebo test (see Abadie, 2021). In-time placebo test, or backdating, consists in reassigning the treatment period to a different year with the purpose to rule out anticipation effects. In-space placebo test reassigns the treatment to donor pool units with the purpose to rule out that the treatment had an effect on a donor pool unit. Indeed, donor pool units should not be influenced by the treatment, i.e., no spillover effects should be present. Other tests verify the robustness of the results with respect to changes in the research design. Leave-one-out analysis checks whether results depend crucially on the presence of a particular unit in the donor pool or a particular predictor. Finally, results can be subjected to statistical significance test. The intuition at the basis of this test is that while a small pre-treatment RMSPE implies good SC fit, a large post-treatment RMSPE implies a relevant impact of the treatment. Consequently, a larger ratio (post/pre RMSPE) for the treated country with respect to the majority of the placebo-treated units (i.e., donor pool units treated in the in-space placebo), implies a significant treatment effect.

Recently, the SC method has undergone important developments in methodological terms, with the aim of overcoming some of its limitations. Improvements or extensions to the original SC method have been proposed (see McClelland &Mucciolo, 2022). Doudchenko and Imbens (2016) present a more general class of SC estimators while new estimators that build on insights deriving from the SC method together with other methods (e.g., difference-in-differences) have been proposed by Arkhangelsky et al. (2021). More recently, Zheng and Chen (2023) propose a dynamic synthetic control method for evaluating treatment effects in auto-regressive processes. Vives-i-Bastida (2023) proposes a data-driven penalized synthetic control method that allows to automatically select the most important predictors given a large set of potentially relevant predictors. Similarly, Greathouse et al. (2023) suggest algorithmic methods that allow to identify relevant donor pool units.

1.2. European studies and the role of the SC

Within comparative economic studies, a recurring question concerns the impact of the European integration process on the economic performance of member countries.

Forecasts about the success of the Eurozone project were contradictory. Alesina and Barro (2002) and Rose (2000) underlined the positive impact of currency unions on trade. According to Frankel and Rose (2002), trade would have been the main growth channel within a monetary union and predicted for Denmark, UK and Sweden a 20% increase in income per capita by adopting the euro. Similarly, other authors tried to quantify the economic benefits of monetary unions (e.g., Carré & Collard, 2003; Devereux et al., 2003) and underlined the convenience for countries such as UK and Sweden to adopt the euro (e.g., Ferreira-Lopes, 2010; Pesaran et al., 2007).

The European integration process demonstrated, however, that different European countries react differently to EU policies and shocks, and follow different development and recovery paths (Hancké, 2012; Hassel, 2014). The European sovereign debt crisis made evident and accelerated the divergence between resilient and vulnerable Eurozone countries, challenging the implementation of common policies (Dallago, 2016).

Estimating the impact of EU membership and euro adoption is not a trivial exercise, especially considering the EU countries heterogeneity (Crespo et al., 2008) and the potential dependency of results on the data and methodologies used. The positive trade effects have been emphasized by Barr et al. (2003), Flam and Nordström (2006) and Baldwin et al. (2008). Other scholars, however, have not identified significant effects of the euro on trade (e.g., Berger & Nitsch, 2008; Mancini-Griffoli & Pauwels, 2006; Santos Silva & Tenreyro, 2010). Drake and Mills (2010) found that the GDP growth trend diminished after the euro introduction, while according to Giannone et al. (2010) the euro had no clear impact on the per capita GDP growth.

Given this varied theoretical background, it is not surprising that a growing strand of literature is considering SC for evaluating the impact of the European project on the economic performance of member countries. For example, Hope (2016) used SC for estimating the impact of EMU on the current account balances of member states and found evidence about EMU responsibility for the divergence in current account balances. Mäkelä (2016) used SC for investigating the impact of EMU on its members' long-term government bond yields and concluded that "from the viewpoint of sovereign borrowing, it would be beneficial for a country to maintain its own currency and monetary policy" (p. 4510). Campos et al. (2019) used the SC to evaluate the impact of EU membership on the GDP per capita and labor productivity for the countries that joined the EU from 1973 to 2004. Their results indicate positive effects which differ across countries and over time and which are negative only for Greece. Some scholars analyzed the impact of the euro or EU membership on the GDP per capita by using similar SC research designs (see for example Fernández & Garcia-Perea, 2015; Verstegen et al., 2017; Puzzello & Gomis-Porqueras, 2018; Gasparotti & Kullas, 2019; Gabriel & Pessoa, 2020). It is worth noting that despite some convergence in identifying Italy as a "loser" and Ireland as a "winner", the results are rather heterogeneous despite the very similar research designs. Since in most cases the analyses using the SC method are concerned with providing valid robustness tests, the heterogeneity is not necessarily caused by a fallacious research design but by the results' sensitivity to the data and algorithm used.

Ferman et al. (2020, p. 511) claim that "an important limitation of the SC method, however, is that there is no consensus on the choice of predictor variables". As confirmed by Abadie (2021), the result of the SC method is influenced by the choice of predictors, the selection of units in the donor pool, and the quality of data. Despite some recent attempts to mitigate this issue, there are not, unfortunately, any well-established strategies aimed to guide the selection process of units and predictors. For example, Ferman et al. (2020, p. 510) "provide recommendations to limit the possibilities for specification searching in the SC method". Vives-i-Bastida (2023) and Greathouse et al. (2023) propose strategies to limit discretion in the selection of predictors and donor pool units. All these valuable contributions should probably be combined to identify a robust strategy that guides the construction of a reliable SC. However, it would also remain useful to develop new strategies that allow comparing sets of SCs generated by various combinations of predictors and donor pool units for a broader assessment of the adequacy of the research design. This would also allow scholars to overcome a tendency to identify "winners" and "losers" in the evaluation of the impact of a policy or event thanks to the comparison of a set of SCs. The purpose of this article is precisely to propose an algorithmic procedure that allows to do this.

2. An algorithmic approach to SC: rationale and methodology

2.1. Improving SC potentialities: the strategy

The intuition at the basis of our algorithm is that the counterfactual can be improved if the variables chosen are significant predictors of the outcome variable and if the donor pool units' characteristics are compatible with those of the treated unit. Starting from this hypothesis, coherent with the basic principle of comparative economics, the task is to verify the presence of redundant predictors and identify donor pool units that are not compatible with the treated unit and remove them by testing different combinations. After this, it is necessary to search the best combination (or set of combinations) of predictors and donor pool units that allow to improve counterfactual precision. The selection of the final set of predictors and donor pool units is therefore the result of a data-driven approach.

2.2. The benchmark case

The first step consists in the computation of the first SC in which the whole initial basket of donor pool units and predictors is used. This is the *benchmark case* and represents the standard research design. The robustness checks described in section 1.1 help to clarify whether the results obtained in the benchmark case can be improved by refining the research design. If there are good reasons to suppose that results can be improved, redundant predictors and donor pool units that are not very compatible with the treated unit are identified by using LASSO. It is worth remembering that, in addition to LASSO, other methodological approaches can be implemented within the algorithm, and these alternative methods should be tested in future research.

2.3. Identification of redundant predictors and poorly compatible donor pool units

The literature agrees that the identification of redundant predictors and unsuitable donor pool units is important for improving the quality of the analysis. Several authors suggest using the LASSO (Least Absolute Shrinkage and Selection Operator) for this task, within different methodological approaches. For example, with reference to the estimation of average treatment effects with panel data, Li and Bell (2017, p. 66) propose "using the LASSO method to select control units and show via simulations that it dominates many conventional methods". LASSO is commonly considered superior with respect to other methods such as stepwise regression (criticized for example in Smith, 2018). LASSO is also used within SC literature (e.g., Vives-i-Bastida, 2023).

Intuitively, LASSO can be interpreted as a constrained form of Ordinary Least Squares (OLS) regression. Indeed, while OLS selects those beta coefficients that minimize the residual sum of squares (RSS), LASSO adds a penalty, which is equal to the sum of the beta coefficients (in absolute values) multiplied by a non-negative regularization parameter λ that modulates the entity of the penalty. The regularization parameter represents the shrinkage that allows to eliminate redundant predictors and promotes sparse models.

In its general form, coherently with Tibshirani (1996) and Hastie et al. (2009), it considers a sample of *N* number of observations with *p* covariates (or predictors) and a single output variable. y_i is the outcome at observation *i*, while x_{ij} is the predictor *j* at observation *i*. The purpose of LASSO is to solve the following problem for a given value of λ, a non-negative regularization parameter:

$$\min_{\beta_0,\beta}\left\{\sum_{i=1}^{N}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2\right\}\ sub.to\ \sum_{j=1}^{p}\left|\beta_j\right|\leq\lambda \qquad (6)$$

This is a quadratic programming problem with linear inequality constraints that can also be expressed in the Lagrangian form:

$$\min_{\beta_0,\beta}\ \left(\frac{1}{2}\sum_{i=1}^{N}(y_i-\beta_0-\sum_{j=1}^{p}x_{ij}\beta_j)^2+\lambda\sum_{j=1}^{p}\left|\beta_j\right|\right) \qquad (7)$$

It is clear that, according to the value of λ, the least relevant beta coefficients for the output variable are decreased to zero and, consequently, removed from the model. This implies that, for computing the LASSO solution, it is necessary to use an algorithm able to calculate how solutions vary according to the value of λ (Hastie et al., 2009). The task is to find the optimal λ able to provide the most predictive model. One strategy is to test different λ and choose the optimal one using cross-validation. Cross-validation can be defined as a resampling method in which data are separated between a training set and a test set. In this article, the LASSO fit has been constructed by using 10-fold cross-validation; other solutions should be tested in future research. The model is built in the training set so as to estimate the outcome in the test set. Then, the mean squared error (MSE) is calculated in the test set. In general, the λ where the minimum cross-validated MSE is observed is selected, but other criteria can be used for selecting the optimum λ. In this article, those λ where the minimum cross-validated MSE is observed and also those λ for which the error is within one standard error of the minimum MSE are identified. The algorithm classifies as relevant those predictors that are confirmed by both criteria and as potentially relevant those confirmed at least by one criterion.

Through LASSO, the algorithm selects a set of non-redundant predictors (i.e., those predictors not excluded by LASSO) for each treated unit. Similarly, the algorithm also identifies a set of non-redundant predictors for each unit of the donor pool. Since it is more probable that SC improves its performance if there is a certain degree of compatibility between the treated and the donor pool units, we compare the non-redundant predictors of the treated unit with those of the donor pool units. We also suppose that donor pool units sharing a similar set of non-redundant predictors compared to the treated unit should be more compatible (i.e., they should share the same determinants of the outcome variable). Consequently, for each treated unit, the algorithm identifies a set of potentially excludable donor pool units, i.e., those units with a set of very different non-redundant predictors with respect to the tested unit. Initially, the donor pool units are ranked according to the number of predictors that are also non-redundant for the treated unit. Then, the number of donor pool units which will actually be tested to verify their exclusion is determined by the number of redundant predictors of the treated unit. Indeed, for performing SC, it is necessary to guarantee that the number of units in each donor pool is not inferior to the number of predictors considered as non-redundant for the treated unit. This means that at least a number of donor pool units equal to the number of predictors will be preserved from the test, and these units will be those in the top zone of the ranking.

2.4. Computing combinations to test

At this point, for each treated unit, there is a set of potentially redundant predictors and a group of units that are poorly compatible with the one treated. To check whether it is possible to improve the accuracy of the counterfactual by reducing the RMSPE, the algorithm tests all possible combinations by removing one or more of these predictors in combination with the elimination of one or more of these potentially unfit units in the donor pool. This step implies the computation of a potentially large number of combinations. Consider a number of redundant predictors equal to *n* (for *i*=1,…,*n*) and a number of redundant donor pool units equal to *m* (for *j*=1,…,*m*). We have to test if it is possible to improve the RMSPE by removing one or more predictors in combination with one or more donor pool units. Consequently, we have to remove them in groups. It is possible to construct groups whose size can be from 1 to *n* (for predictors) or *m* (for units). For each group size, we have a set of combinations to test. For example, imagine 10 predictors of which the first 3 are potentially redundant (predictors 1, 2 and 3) so that *n*=3. We have to test the removing of the following set of combinations of predictors: (1), (2), (3), (1,2), (1,3), (2,3), (1,2,3). Consequently, we have groups of size 1 (k_1 =1), groups of size 2 (k_2 =*2*), and a group of size 3 (k_3 =*3*). Thus, it is possible to compute the binomial coefficient of *n* and k_i as in the following formula, which correspond to the number of combinations of *n* predictors (*m* units) taken k_i at a time:

$$comb_i = \frac{n!}{k_i!(n-k_i)!} \qquad for\ i = 1, \dots, n \tag{8}$$

$$comb_j = \frac{m!}{k_j!(m-k_j)!} \qquad for\ j = 1, \dots, m \tag{9}$$

According to the previous example; *comb*$_1$*=3*, *comb*$_2$*=3; comb*$_3$*=1* so that with 3 redundant predictors (1,2 and 3), we have 7 combinations to test. The same procedure should be repeated for redundant units. At the end, we know how many combinations should be tested for predictors and units. Imagine a number *P* of combinations to be tested for predictors (in the example *P*=7) and a number *C* of combinations to be tested for units. Since we are interested in the computation of all the possible combinations of predictors and donor pool units to test, we have to consider the Cartesian product between *P* and *C*. The Cartesian product provides a starting number θ of combinations to be tested. From this set those combinations in which the remaining number of predictors is greater that the remaining number of donor pool units are excluded; indeed, in these cases, the SC cannot be computed. Since the order of predictors and donor pool units may influence the result, for each combination, a case in which predictors, donor pool units, both or none of them have a random order has been randomly considered. This implies that the total number of combinations to be tested is θ minus those that cannot be computed (so as to obtain Φ).

2.5. Apply the SC and the search algorithm

The SC that minimizes the RMSPE is the best solution to which the most proper set of predictors and donor pool units correspond. Clearly, the total number of combinations to test may be extremely large as the number of predictors and donor pool units to test increases. Consequently, the identification of the best solution may become impossible or extremely expensive in terms of computational time. In order to verify the appropriateness of the research design, it may be sufficient to identify the set of combinations which is closest to the best solution. This allows us to understand which predictors and donor pool units should be effectively excluded from the analysis. Indeed, it should not be forgotten that a very high number of combinations to test indicates that too many donor pool units and predictors are redundant, and this should induce a review of the research design. However, also in these cases, it is possible to approach the best set of solutions and find the best solution among them. This is possible through a search algorithm that should progressively reduce the number of combinations to test. This can be done by excluding one at a time the predictor and donor pool country (among those identified by LASSO as potentially excludable and consequently to be tested) most present in the “best solutions”. The “best solutions” are those with a lower RMSPE with respect to the benchmark case, which we call “SC cloud”. This part of the algorithm works as follows:

1. Test a starting number δ of combinations selected randomly from Φ. Compute for each combination the SC and the RMSPE.

2. Identify the combination that allows to minimize the RMSPE (i.e., the "best solution") and identify those combinations that produce a lower RMSPE with respect to the benchmark case (i.e., the "SC cloud"). If it is not the case, repeat point 1 until all combinations are tested, or until a lower RMSPE with respect to the benchmark case is found.
3. Check which predictor and which donor pool unit among those tested in the "SC cloud" are the most significant (i.e., the predictor which is most present and the country with the highest average weight). Exclude them from the set of predictors and donor pool units to test.
4. Compute the new set of combinations Φ. Test a starting number δ of combinations selected randomly and compute for each combination the SC and the RMSPE.
5. Identify the combination that allows to minimize the RMSPE (i.e., the new "best solution") and the new "SC cloud".
6. If the new "best solution" has a RMSPE lower with respect to the previous one, carry on the algorithm from step 3, otherwise stop.

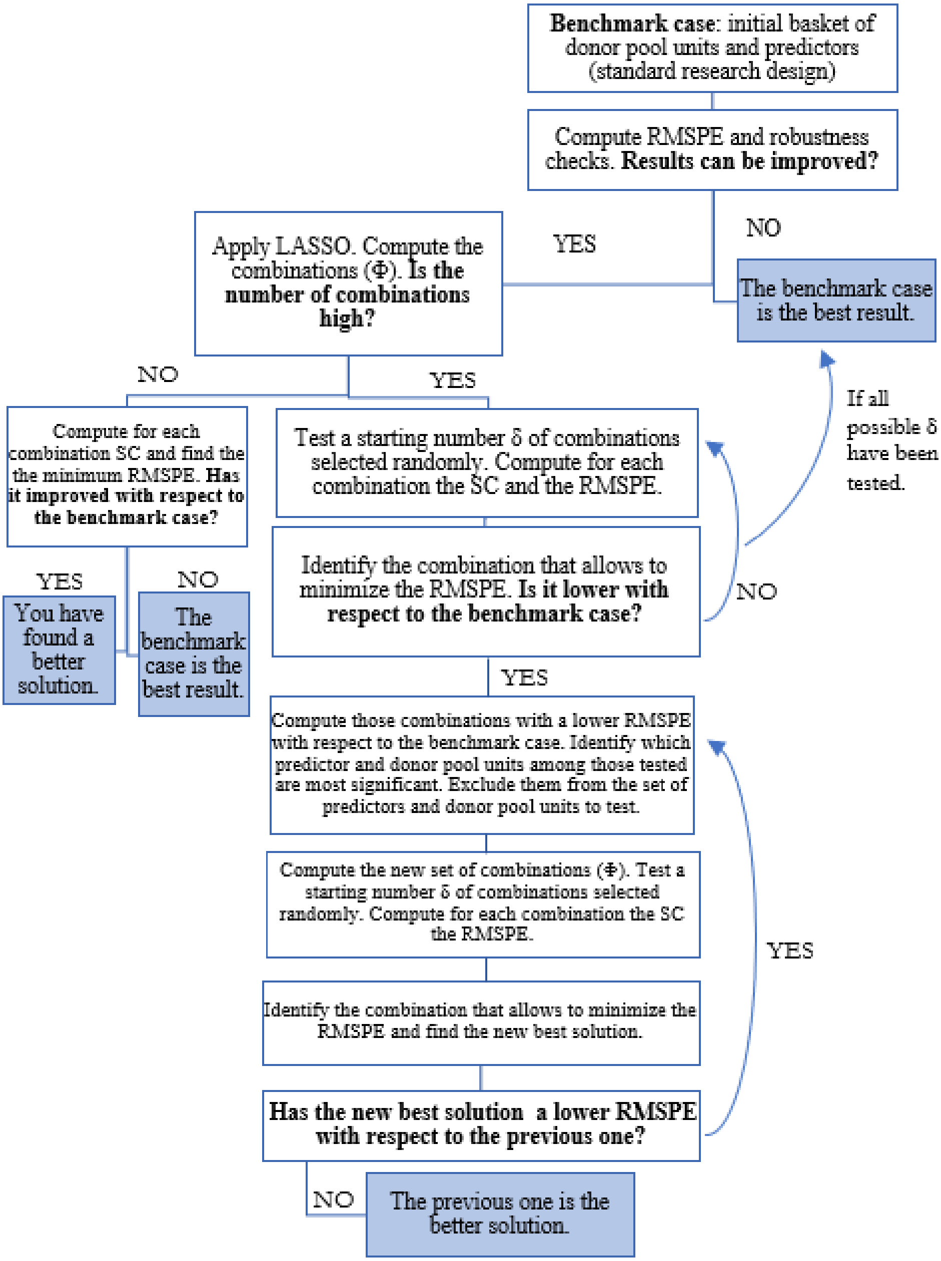
It is worth noting that the starting number of combinations selected randomly (δ) is fixed exogenously but since the procedure iterates until there is a gain in terms of RMSPE reduction, the real total number of combinations selected randomly results to be data-driven.

2.6. Summary: the algorithm

All the steps described from section 2.2 on are a sequence of operations part of a unique algorithm (implemented in MATLAB). The algorithm starts from a "benchmark case" with an initial basket of predictors and donor pool units. The algorithm is able to discard redundant predictors and poorly compatible units (if present), thus providing a counterfactual with a lower RMSPE. All the steps have been graphically represented in Figure 1.

This algorithm provides various information which, if properly interpreted, offers more articulate answers to the initial research question. Indeed, if many predictors are redundant and there are many units that are not compatible with the treated one, this indicates that the research design should be probably revised. However, the search algorithm may offer a better solution among a random number of combinations, which grows as long as the algorithm manages to obtain a better counterfactual. If the number of redundant predictors and poorly compatible units is low, it is possible to test all combinations and, maybe, find the best solution. If this solution significantly increases the quality of SC and, also, the "SC cloud" offers conclusions consistent with the best solution, we can trust the robustness of the result.

Figure 1. The whole algorithm


Benchmark case: initial basket of donor pool units and predictors (standard research design)
Compute RMSPE and robustness checks. Results can be improved?
YES
NO
The benchmark case is the best result.
Apply LASSO. Compute the combinations (Φ). Is the number of combinations high?
NO
YES
Compute for each combination SC and find the the minimum RMSPE. Has it improved with respect to the benchmark case?
YES
NO
You have found a better solution.
The benchmark case is the best result.
Test a starting number δ of combinations selected randomly. Compute for each combination the SC and the RMSPE.
If all possible δ have been tested.
Identify the combination that allows to minimize the RMSPE. Is it lower with respect to the benchmark case?
NO
YES
Compute those combinations with a lower RMSPE with respect to the benchmark case. Identify which predictor and donor pool units among those tested are most significant. Exclude them from the set of predictors and donor pool units to test.
Compute the new set of combinations (Ŧ). Test a starting number δ of combinations selected randomly. Compute for each combination the SC the RMSPE.
YES
Identify the combination that allows to minimize the RMSPE and find the new best solution.
Has the new best solution a lower RMSPE with respect to the previous one?
NO
The previous one is the better solution.

Source: authors' representation

2.7. Testing the algorithm: simulation results

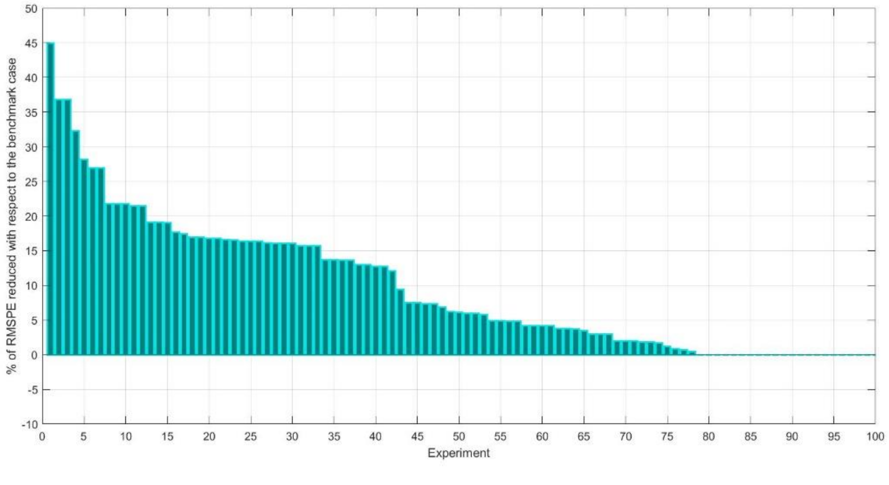
In order to demonstrate that the algorithm represents a valid tool to improve the performance of SC, we used simulations under a variety of data generating processes coherently with the approach used by Abadie and Vives-i-Bastida (2022). In particular, we have used their generative model:

$$Y_{it}^{N} = \delta_t + \lambda_{f(i)t} + \varepsilon_{it} \qquad with\ \varepsilon_{it} \sim i.i.d.N(0,\sigma^2) \tag{10}$$

There are *F* common factors and each unit *i* loads exclusively on a single factor *f(i)*. Each common factor $\lambda_{f(i)t}$ follows a Gaussian *AR(1)* process with ρ as autoregressive coefficient and standard innovations. Since the number of units must be greater with respect to the number of common factors, some units will have the same common factor. It will be randomly determined which units will have the same common factor, regardless of whether it is a treated unit or not.

We have generated data for 10 units, 30 periods with 7 common factors, and random values for ρ (between 0,1 and 0,9) and σ^2 (between 0,1 and 2). The date of intervention for the treated unit is fixed after 20 periods. To demonstrate the effectiveness of the algorithm, we consider a low starting number of combinations to be extracted randomly (δ=10), given that, in this type of simulations, there are thousands of combinations to be tested (Φ). We have repeated this experiment 100 times and the results are reported in Figure 2.

Figure 2. Reduction of RMSPE with respect to the benchmark case in percentage values



Source: authors' calculations

Experiments have been ordered according to the percentage gain in terms of reduction of RMSPE with respect to the benchmark case. Despite the extremely low number of combinations extracted, the algorithm is able to find a better solution with respect to the benchmark case in the 79% of the experiments.

3. Assessing the benefits of European integration

3.1. Research design: aims and data

The purpose of our analysis is to test the potentialities of the algorithm in the assessment of the benefit of EU membership. We start from a standard research design. Consequently, in the *benchmark case*, we use donor pool countries and predictors frequently used in the literature and compatible with our analysis (e.g., Fernández & Garcia-Perea, 2015; Gabriel & Pessoa, 2020; Gasparotti & Kullas, 2019; Puzzello & Gomis-Porqueras, 2018; Verstegen et al., 2017). It should be remembered that the selection of variables and countries according to previous empirical literature is a strategy followed by other scholars that use the SC method (e.g., Gabriel & Pessoa, 2020). In our case, we consider this choice acceptable for the benchmark case. A more in-depth analysis about the choice of the initial basket of predictors and countries, the impact of the different databases and the methodological implications will be part of future research. Similarly, a more detailed analysis of the importance of also considering institutional indicators among the predictors is left to future research.

We use the GDP per capita as the outcome variable. The GDP per capita is a fundamental index used to assess the economic success of a country and its economic policies. We consider the signing of the Maastricht Treaty as the "treatment" and 1992 as the treatment date. Indeed, the Maastricht Treaty has strengthened and restricted the constraints on economic policy choices, influencing long-term strategies, also of those member countries that have not adopted the euro. It is worth remembering that, even if we consider the Maastricht Treaty as a the "treatment", nothing prevents us from assessing the impact of events that have come later such as the introduction of the euro.

Among the 38 OECD member countries, the 12 countries that established the EU in 1992 are the treated ones: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, and UK. The 13 OECD countries that have not joined the EU are all potential donor pool countries: Australia, Canada, Chile, Iceland, Israel, Japan, Korea, Mexico, New Zealand, Norway, Switzerland, Turkey, United States (Colombia and Costa Rica have been excluded since they only joined OECD in 2020 and 2021 respectively). The comparison between the OECD countries that have joined or not joined the EU allows us to compare countries that, despite differences, share long-term development and cooperation objectives. This is an important aspect because by introducing developing countries characterized by significantly different economic and social structures, “results are condemned to be biased” (Gabriel & Pessoa, 2020, p. 6). For this reason, we have only considered OECD countries. However, it remains undeniable that important differences exist among these countries, and that the algorithm should help to find the most compatible group starting from this initial standard basket.

The treatment date is 1992, the year when the Maastricht Treaty was signed. Data from 1970 to 2019 have been collected in order to ensure a suitable number of pre-treatment years. We stop at 2019 to avoid taking into account the socio-economic shocks caused by the Covid-19 pandemic and the geopolitical tensions that followed the invasion of Ukraine in February 2022. It is worth remembering that, although the fulfilling of the requirements of the Maastricht treaty and the *acquis communautaire* imply that candidate countries are somehow “treated” before joining the EU, we could expect that this treatment is not such as to invalidate our analysis because we focus on those countries that established the EU in 1992. The application of our analysis to the Eastern transition countries which joined the EU more recently would be more controversial.

The following set of predictors (excluding the first of the list that is the outcome variable) has been selected for the period 1970-2019: the real GDP per capita, balance of trade, private consumption expenditure, general government final consumption expenditure, gross capital formation, resident patent applications, employment share, age dependency ratio, labour productivity, inflation, human capital, general government debt (See Appendix for a detailed description of these variables). Most data come from Penn World Table (PWT), while others come from the World Intellectual Property Organization (WIPO), OECD databases and the Global Debt Database.

3.2. The benchmark case: consider all donor pool countries and predictors

We consider, as benchmark case, the SC applied to the 12 treated countries by using all predictors and donor pool countries. The weights associated to each donor pool country for each treated country are reported in Table 1. Weights represent the contribution of each donor pool country for the construction of each SC (see section 1.1. for the interpretation of weights). The last row presents the RMSPEs which give an idea of the quality of the SCs in terms of their ability to reproduce the trajectory of the outcome variable for the treated unit.

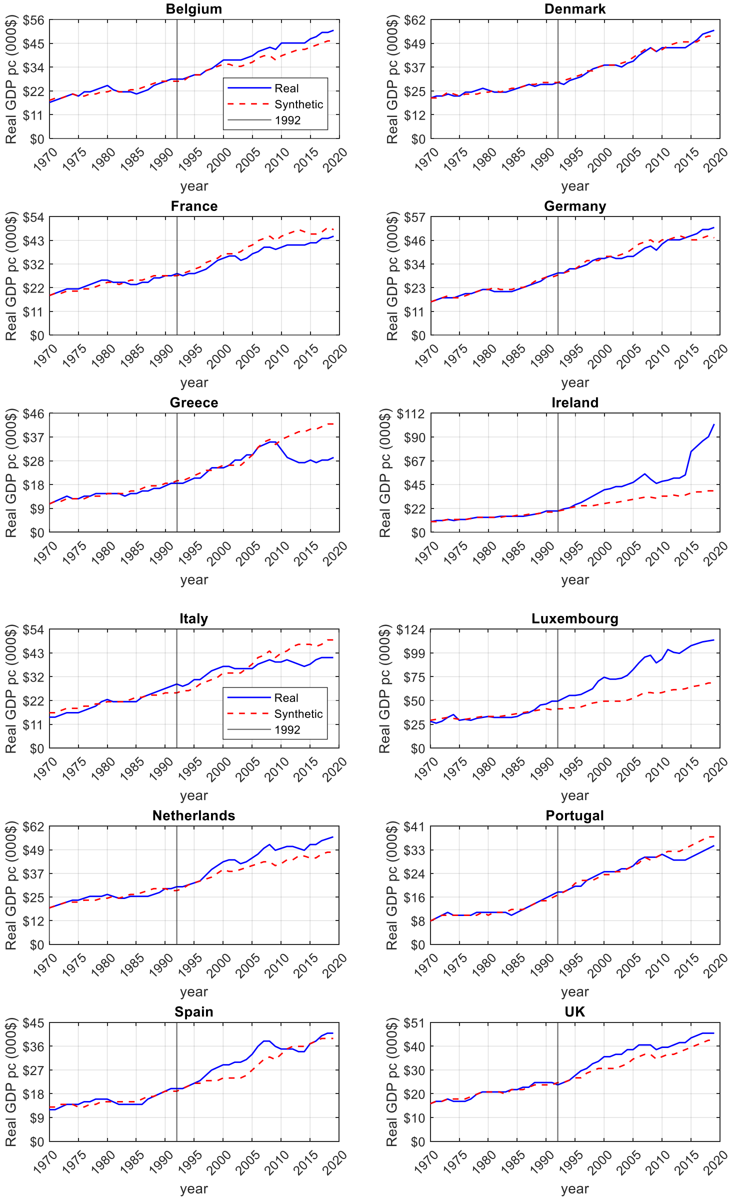
Table 1. Benchmark case: donor pool weights and RMSPEs

Donor pool/Treated countries	Belgium	Denmark	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	UK
Australia	0.00	0.00	0.42	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Canada	0.34	0.08	0.21	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.00
Chile	0.21	0.00	0.00	0.05	0.10	0.00	0.00	0.00	0.00	0.19	0.28	0.28
Iceland	0.07	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.17
Israel	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.05	0.00	0.03
Japan	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Korea	0.00	0.00	0.00	0.00	0.23	0.43	0.00	0.00	0.00	0.42	0.19	0.00
Mexico	0.00	0.00	0.16	0.00	0.00	0.30	0.00	0.00	0.00	0.06	0.09	0.18
New Zealand	0.00	0.49	0.00	0.00	0.00	0.07	0.00	0.00	0.28	0.28	0.00	0.00
Norway	0.00	0.20	0.21	0.25	0.14	0.00	0.31	0.00	0.11	0.00	0.00	0.00
Switzerland	0.13	0.18	0.00	0.00	0.11	0.00	0.00	0.57	0.00	0.00	0.26	0.00
Turkey	0.00	0.00	0.00	0.00	0.34	0.00	0.39	0.00	0.00	0.00	0.18	0.00
US	0.00	0.05	0.00	0.00	0.00	0.21	0.30	0.43	0.00	0.00	0.00	0.35
RMSPE bench	1245	819	1171	697	871	455	1881	3406	1364	572	1084	619

Source: authors' calculations

Figure 3 introduces a graphical comparison of the dynamics of real GDP per capita for each treated country (line) versus the performance of each SC (dashed line). If the trend of the two lines is similar before 1992, this implies a good fit of the SC (low RMSPE). The gap after 1992 should correspond to the GDP gain (or loss) induced mainly by the EU membership. It is interesting to observe the strong growth of the real GDP per capita in Ireland and Luxembourg compared to their SC. These results are coherent with other findings in literature. Ireland's positive gap is the consequence of its strong economic development in the 1990s. Ireland was a lot poorer than most industrial countries but it became the Celtic Tiger thanks to FDI, low taxes and important EU contributions. Luxembourg experienced a strong growth from the 1980s thanks to the strong expansion of the financial sector, EU expenditure and the presence of some important European institutions based in Luxembourg. As confirmed by Puzzello and Gomis-Porqueras (2018), Luxembourg has been one of the richest countries in the last decades. These circumstances make Ireland and Luxembourg two particular cases, well-known in the literature, and our analysis is in line with these results.

Figure 3. Comparing counterfactual scenarios for all treated countries in the benchmark case



Note: Dynamics of the real GDP per capita for each treated country (bold blue line) versus the performance of the synthetic control of the benchmark case (dashed red line). The vertical line corresponds to the treatment date 1992.
Source: authors' representation

Robustness checks confirm that the counterfactuals of Ireland and Luxembourg are the most reliable and do not need to be considered in the further steps of the algorithm. This is evident by observing the results of the placebo tests (see Figure A1-A5 in the Appendix). Indeed, these two countries clearly outperform with respect to the others in all tests. In the Appendix, in Figure A1, we report in-time placebo test. We reassign the treatment period to 1980: we do not observe the presence of relevant estimated effects in the pre-treatment period. In-space placebo tests are reported in Figure A2. We reassign the treatment to a donor pool country with the purpose to rule out that the treatment has an effect on donor pool countries. In Figure A3 and A4, we consider leave-one-out analysis for verifying if the results depend on the presence of a particular country in the donor pool or on a particular predictor. According to results, Ireland and Luxembourg have the most reliable counterfactuals. This result is also confirmed by the statistical significance test reported in Figure A5. This implies that these two countries perform significantly better than the other countries whatever counterfactual is considered, and the application of our algorithm could not add anything to this conclusion.

3.3. Applying the algorithm: selecting predictors and donor pool countries

For each treated country, the algorithm has selected those predictors that can be excluded according to LASSO (E in Table 2) and the set of donor pool countries less compatible with the treated one (X in Table 3).

Table 2. Predictors excludable according to LASSO (E)

Country/ Indicator	Private consumption	Government consumption	Gross capital formation	Employment share	Human capital	Age dependency ratio	Labour productivity	Inflation	Patents application	Trade	Government debt	Total
Belgium	E				E						E	3
Denmark	E	E					E	E		E	E	6
France	E	E	E				E	E				5
Germany			E									1
Greece	E	E	E					E	E		E	6
Italy			E							E	E	3
Netherlands			E					E		E		3
Portugal	E			E								2
Spain			E									1
UK		E			E	E				E		4
Total	5	4	6	1	2	1	2	4	1	4	4	

Source: authors' calculation

Table 3. Donor pool countries excludable (X)

Donor pool/Treated countries	Belgium	Denmark	France	Germany	Greece	Italy	Netherlands	Portugal	Spain	UK	Total
Australia	X		X	X	X	X		X	X		7
Canada		X	X		X		X			X	5
Chile		X	X		X		X			X	5
Iceland	X			X		X	X	X	X	X	7
Israel		X	X		X		X			X	5
Japan		X	X		X		X			X	5
Korea	X	X						X		X	4
Mexico		X	X		X	X	X	X			6
New Zealand		X	X		X		X			X	5
Norway		X	X		X		X			X	5
Switzerland	X		X		X			X			4
Turkey	X	X	X	X	X	X	X	X	X		9
United States						X	X			X	3
Total	5	9	10	3	10	5	10	6	3	9	

Source: authors' calculations

The total number of combinations Φ to test can be different for each treated country as reported in Table 4. Since, in some cases, the number of combinations becomes so high to results impossible or extremely time consuming to be computed, in these cases, we have calculated δ=100. These combinations have been taken randomly among all those possible ones for a given country treated (i.e., Denmark, France, Greece, the Netherlands, and the UK). This number can increase until the search algorithm is able to find a better counterfactual, as explained in section 2.5. The number of combinations to be calculated can give an idea of the quality of the research design. Indeed, a large number of combinations means that many predictors and donor pool countries are probably poorly significant.

Table 4. Number of combinations to be computed

Treated countries	Belgium	Denmark	France	Germany	Greece	Italy	Netherlands	Portugal	Spain	UK
Number of combinations	196	22710	16297	7	39084	196	2317	138	7	4035

Source: authors' calculations

3.4. The final results: analysis and comments

In Table 5, it is possible to observe the donor pool countries selected by the algorithm for each treated country and their weights. An empty cell indicates that the donor pool country is not present in the counterfactual while a number equal to 0.00 indicates that the country is in the donor pool but its weight is equal to zero. Below, the benchmark RMSPEs have been reported for comparison with the new RMSPEs. The last row indicates that the algorithm has been able to identify a better SC for all treated countries so as to obtain a reduction of RMSPE with respect to the benchmark case.

Table 5. Final results after the algorithm: weights and RMSPEs

Donor pool/Treated countries	Belgium	Denmark	France	Germany	Greece	Italy	Netherlands	Portugal	Spain	UK
Australia	0.00	0.00	0.00		0.00	0.00	0.00		0.00	0.10
Canada	0.18	0.02	0.32	0.00		0.00	0.21	0.00	0.18	0.19
Chile	0.00	0.00		0.00	0.26	0.00	0.00	0.41	0.49	0.25
Iceland		0.00	0.03		0.00	0.24				0.06
Israel	0.26		0.01	0.00	0.00	0.00	0.12	0.02	0.00	0.02
Japan	0.21	0.00	0.02	0.35	0.22	0.45	0.08	0.29	0.00	0.12
Korea	0.00	0.00	0.00	0.25	0.00	0.18	0.00	0.17	0.16	0.00
Mexico	0.19		0.38	0.00		0.00	0.26		0.03	0.00
New Zealand	0.00	0.50	0.00	0.00		0.00	0.00	0.00	0.00	
Norway	0.00	0.23		0.14	0.11	0.13		0.00	0.00	0.25
Switzerland	0.17	0.18	0.24	0.26	0.08	0.00	0.33	0.00	0.14	0.00
Turkey		0.00		0.00	0.33			0.12		0.00
US	0.00	0.07	0.00	0.00	0.00	0.00		0.00	0.00	
RMSPE bench	1245	819	1171	697	871	1881	1364	572	1084	619
RMSPE	1110	817	928	625	507	537	1101	401	1076	568
improvement	135	2	243	72	364	1344	263	171	8	51

Source: authors' calculations

In Figure 4, there is a graphical comparison between the dynamics of the real GDP per capita for each treated country (bold line), the SC of the benchmark case (dotted line) and the performance of the best SC (dashed line) that corresponds to the best combination (i.e., the one able to minimize the RMSPE). SCs that have a lower RMSPE than the benchmark case (dotted line), but do not represent the best case, have been also reported in the graph and represent the SC cloud (lines).

The results obtained after the implementation of the algorithm allow a more articulated analysis of the gain/losses induced by the EU membership and an evaluation of the appropriateness of the research design for each treated country.

In the case of Belgium, the quite low number of tested combinations leads us to have confidence in the research design considered. However, the initial high value of the RMSPE and the modest gain obtained through the algorithm (with respect to the average RMSPE obtained using the algorithm equal to 767) do not exclude that better results are possible by adding other predictors or donor pool countries. In addition to this, the results indicate that Belgium has improved its performance relative to its SC especially from the 2000s. This could indicate that Belgium may have been able to benefit from the introduction of the euro. The statistical significance test (see in the Appendix Figure A6) supports this interpretation, together with the fact that Belgium outperforms all counterfactuals, including the cloud.

The case of Denmark is more difficult to interpret. Many predictors and donor pool countries appear to be redundant. The algorithm achieves a negligible improvement of the RMSPE. The standard research design is most likely inadequate for this treated country. The dynamics of the curves in Figure 4 are very similar. This could mean, among other things, that the impact of membership has been negligible for a country that subsequently did not adopt the euro (the statistical significance test supports this interpretation - see Figure A6 in the Appendix). However, for the reasons set out above, these findings should be taken with caution and subjected to further research.

Also, in the French case, the research design presents a good number of potentially redundant predictors and various countries that are not clearly compatible. However, the good reduction of RMSPE and the dense cloud of SC below the benchmark case indicate that the algorithm allows to improve the quality of the SC. Compared to the benchmark case, the French performance is not much worse than its counterfactual, but it seems beyond doubt that France has lost ground over time, especially in recent years. However, the strong dispersion of the cloud induces caution on these conclusions, as the statistical significance test also indicates.

The German case is decidedly different. The research design seems appropriate as confirmed by a good initial RMSPE for the benchmark case (if compared with the average value 1032), and the few combinations to test. The algorithm further improves the quality of the SC and reverses the conclusion that may have been drawn by observing the benchmark case. Germany seems to have lost ground with respect to its counterfactual from the 2000s, and only recently did it manage to recover. The small cloud and the statistical significance test seem to confirm the robustness of these observations.

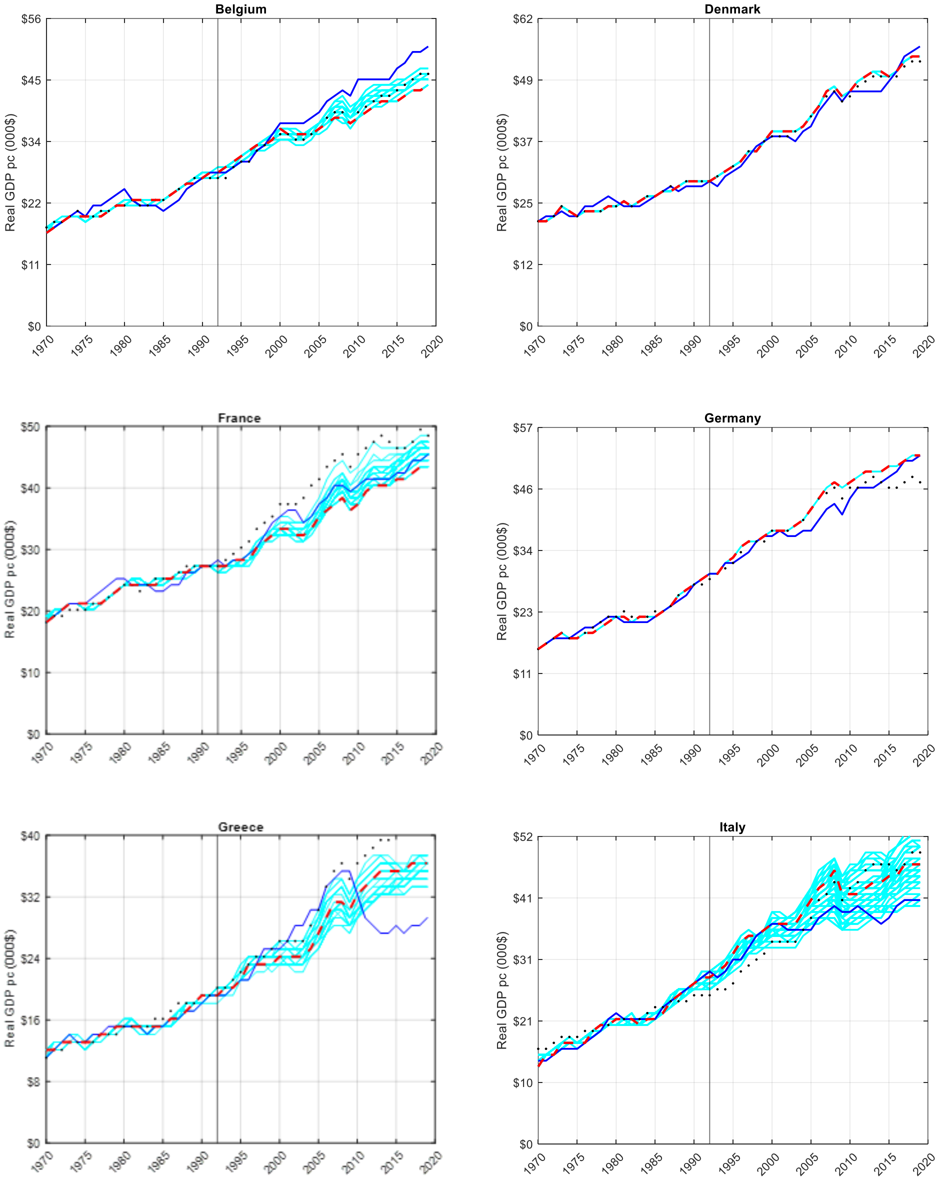
As for Greece, although it seems clear that the research design can be improved, it seems difficult to obtain results capable of contradicting what can already be concluded by observing Figure 4. The algorithm has made it possible to significantly improve the quality of the SC compared to the benchmark case and the statistical significance test confirms the robustness of the result. For a while, Greece performed better than all its counterfactuals and, most likely, this was especially favoured by the introduction of the euro and the subsequent euphoric phase which ended with the financial crisis, as seen in Figure 4. However, there is no doubt that the outbreak of the sovereign debt crisis led Greece to experience an unprecedented phase of crisis in whatever counterfactual, as confirmed by observing the cloud.

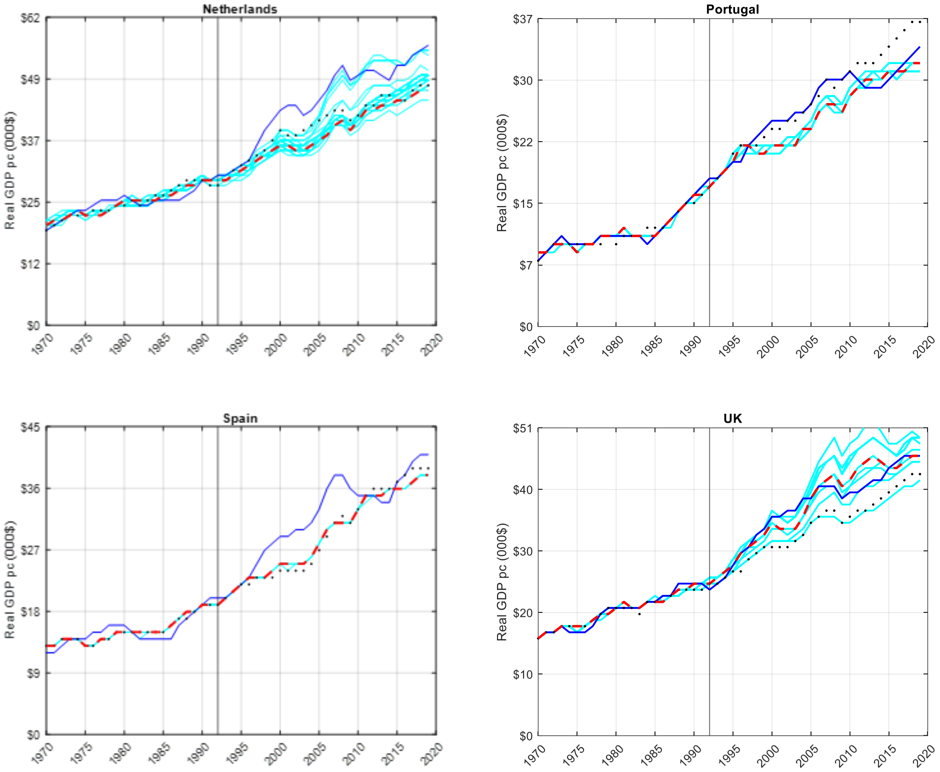
The Italian case differs from the previous ones. The research design looks good as confirmed by the low number of combinations to be calculated. However, the presence of some predictors or donor pool countries was particularly damaging to the quality of the SC in the benchmark case. This is demonstrated by the strong reduction of RMSPE and the statistical significance test. Italy has undoubtedly lost ground compared to almost all its counterfactuals after joining the euro. The European membership was not a real game changer; Italian productivity stopped growing in 2001. A plausible explanation for this phenomenon can be traced back to the consequences for the Italian economy of having given up monetary sovereignty and, consequently, of the impossibility to use monetary devaluation. However, it is worth underlying how monetary devaluation was a weak and controversial tool to face issues that needed to be addressed through structural and fiscal reforms and investments to improve competitiveness and productivity. The comparison of the Italian case with Belgium is interesting. In the period, Belgium’s debt/GDP ratio was higher than Italy’s. In Belgium, the game changer was not the reduction of public expenditure, but the GDP growth driven by strong labour productivity.

For the Netherlands, the algorithm improved the quality of the SC, as confirmed by the statistical significance test despite the RMSPE remaining quite high compared to the average. The results suggest that the Netherlands continued to benefit from the positive gap induced by the Dutch ‘miracle’ started in the mid-1990s (Albers & Langedijk, 2004).

In the Portuguese case, starting from a good research design, the algorithm improves the quality of the SC and allows to draw different conclusions compared to the benchmark case. The Portuguese economy outperformed its counterfactual but, after the sovereign debt crisis, its performance became more similar to its counterfactuals. Portugal experienced a very positive phase in the 1990s, which was also thanks to the considerable fall in interest rates driven by the prospect to enter EMU (Abreu, 2006) and EU financial support. However, Portugal has suffered the impact of the European sovereign debt crisis. Also, in this case, the impact of the EU membership cannot be trivially evaluated despite the statistical significance test supporting this interpretation (see Figure A6 in the Appendix).

Figure 4. Final results after the algorithm: comparing counterfactual scenarios for all treated countries





Note: Dynamics of the real GDP per capita for each treated country (bold line) versus the performance of the best SC able to minimize RMSPE (dashed line). The dotted lines correspond to the SC of the benchmark case while light blue lines correspond to the SCs of the cloud.
Source: authors' representation

The research design for Spain seems adequate as indicated by the extremely low number of combinations to be tested. However, the RMSPE is quite high compared to the average, and the algorithm does not allow to improve significantly the SC, as confirmed by the statistical significance test. Despite this, the Spanish economy experienced a period of prosperity until the sovereign debt crisis greater than any tested counterfactual. Spain has been one of the most dynamic economies in the 1990s: it benefited from the introduction of the euro but appears to have suffered the consequences of the sovereign debt crisis. According to the results, Spain has overcome the crisis better than others, with a significantly better performance than its counterfactual in recent years. However, in light of previous observations, we remain cautious in estimating how well Spain has managed to outperform its counterfactual.

For the UK case, the algorithm obtained a modest improvement in terms of RMSPE and completely changed the conclusion that may be drawn by observing the benchmark case. The cloud indicates that there are good reasons to believe that the British economy did not actually outperform so starkly its counterfactual and, instead, it has been losing ground since 2005 and has only recovered in recent years.

In general, the results indicate that the economic effect of EU membership has significantly varied among countries, and that disparities persist among them. This conclusion is coherent with other findings in the literature, such as Camagni et al. (2020), who confirm that the widening and deepening of the EU have exacerbated intraregional disparities. Disparities change over time and space and especially concern the more peripheral countries (Monastiriotis et al., 2017). However, rather than classifying countries according to "winners" and "losers", it is worth noting how the European integration process seems to have favoured a cluster-based regional convergence, which is reflected in the heterogeneity of growth paths (Cutrini, 2019; Iammarino et al., 2019; Monfort et al., 2013).

Conclusions

The synthetic control (SC) method is a promising methodology for comparative economic studies. Recently, the SC method has undergone important developments in methodological terms, with the aim of overcoming some of its limitations and provide improvements or extensions. One of the purposes is to limit discretion in the research design so as to construct more reliable SCs. However, it would also be useful to develop new strategies that allow to compare sets of SCs generated by various combinations of predictors and donor pool units for a broader assessment of the appropriateness of the research design. The presence of a set of SCs should help to avoid the tendency to identify "winners" and "losers" in the evaluation of the impact of a policy or event, especially within the economic field. The purpose of this article has been precisely to propose an algorithmic procedure that allows to do this.

The algorithm has been tested through simulation and used to investigate a well-known issue within comparative economic studies: the evaluation of benefits of European integration. Starting from a standard research design, the algorithm has improved the robustness and the precision of counterfactual scenarios for all countries through the removal of redundant predictors and poorly compatible donor pool units. The algorithm has provided useful insight about the robustness of results and the appropriateness of the research design also through an intuitive graphic representation of the outcomes. It is worth remembering that we leave to future research the evaluation of the potential of the algorithm with more sophisticated SC versions. This implies that the algorithm and the results presented in this article should be considered as starting points for further research.

The results indicate that the economic effect of the EU membership is very different among countries. It is difficult to establish whether the EU membership has been a game changer on its own. Although the European membership in strong connection with the adoption of the euro have influenced the way in which countries have reacted to the crisis or have designed their polices, it would be reductive to divide European countries between “winners” and “losers” as a simple consequence of their decision to join the EU or the euro. Indeed, national, community and chance factors have influenced the performance of member countries. These results offer interesting insights about the implications of the European institutional variety for the European integration project and suggest that disparities among EU member countries tend to persist. The European integration process seems to have favoured a cluster-based regional convergence, which is reflected in the heterogeneity of growth paths.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391-425. https://doi.org/10.1257/jel.20191450

Abadie, A. & Vives-i-Bastida, J. (2022). *Synthetic Controls in Action*. *arXiv preprint arXiv:2203.06279*. https://doi.org/10.48550/arXiv.2203.06279

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. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science, 59*(2), 495-510. https://doi.org/10.1111/ajps.12116

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, *93*(1), 113-132. https://doi.org/10.1257/000282803321455188

Abadie, A., & L’Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association, 116*(536), 1817-1834. https://doi.org/10.1080/01621459.2021.1971535

Abreu, O. (2006). Portugal’s boom and bust: Lessons for euro newcomers. *ECFIN Country Focus, 3*(16), 22-12.

Adhikari, B. (2022). A Guide to Using the Synthetic Control Method to Quantify the Effects of Shocks, Policies, and Shocking Policies. *The American Economist, 67*(1), 46-63. https://doi.org/10.1177/05694345211019714

Albers, R., & Langedijk, S. (2004). The Netherlands: From Riches to Rags. *EFCIN Country Focus*, *1*(13). https://ec.europa.eu/economy_finance/publications/pages/publication1425_en.pdf

Alesina, A., & Barro, R. J. (2002). Currency unions. *The Quarterly Journal of Economics, 117*(2), 409-436. https://doi.org/10.1162/003355302753650283

Angrist, J. D., & Pischke, J. S. (2015). *Mastering metrics: The path from cause to effect*. Princeton (NJ): Princeton University Press.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, *111*(12), 4088-4118. https://doi.org/10.1257/aer.20190159

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives, 31*(2), 3-32. http://www.jstor.org/stable/44234997

Baldwin, R., DiNino, V., Fontagné, L., De Santis, R. A., & Taglioni, D. (2008). *Study of the Impact of the Euro on Trade and Foreign Direct Investment* (European Economic and Monetary Union Working Paper No. 321). http://dx.doi.org/10.2139/ssrn.1163774

Barr, D., Breedon, F., & Miles, D. (2003). Life on the outside: economic conditions and prospects outside euroland. *Economic Policy*, *18*(37), 573-613. https://doi.org/10.1111/1468-0327.00116_1

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789-1803. https://doi.org/10.1080/01621459.2021.1929245

Berger, H. & Nitsch, V. (2008). Zooming out: The trade effect of the euro in historical perspective. *Journal of International Money and Finance*, *27*(8), 1244-1260. https://doi.org/10.1016/j.jimonfin.2008.07.005

Camagni, R., Capello R., Cerisola, S. & Ugo Fratesi, U. (2020). Fighting Gravity: Institutional Changes and Regional Disparities in the EU. *Economic Geography*, *96*(2), 108-136. https://doi.org/10.1080/00130095.2020.1717943

Campos, N. F., Coricelli, F., & Moretti, L. (2019). Institutional integration and economic growth in Europe. *Journal of Monetary Economics, 103*, 88-104. https://doi.org/10.1016/j.jmoneco.2018.08.001

Carré, M., & Collard, F. (2003). Monetary union: A welfare based approach. *European Economic Review*, *47*(3), 521-552. https://doi.org/10.1016/S0014-2921(01)00170-2

Cerulli, G. (2022). *Econometric evaluation of socio-economic programs. Theory and applications*. Second Edition. Berlin: Springer. https://doi.org/10.1007/978-3-662-65945-8

Crespo Cuaresma, J., Ritzberger-Grünwald, D., & Silgoner, M. A. (2008). Growth, convergence and EU membership. *Applied Economics*, *40*(5), 643-656. https://doi.org/10.1080/00036840600749524

Cutrini, E. (2019). Economic integration, structural change, and uneven development in the European Union. *Structural Change and Economic Dynamics*, *50*, 102-113. https://doi.org/10.1016/j.strueco.2019.06.007

Dallago, B. (2016). *One Currency, Two Europes. Towards a Dual Eurozone*. Singapore: World Scientific. https://doi.org/10.1142/9949

Devereux, M., Engel, C., & Tille, C. (2003). Exchange rate pass-through and the welfare effects of the euro. *International Economic Review, 44*(1), 223-242. https://doi.org/10.1111/1468-2354.t01-1-00068

Doudchenko, N., & Imbens, G. W. (2016). *Balancing, regression, difference-in-differences and synthetic control methods: A synthesis*. NBER Working Paper Series No. 22791, National Bureau of Economic Research. https://doi.org/10.3386/w22791

Drake, L. & Mills, T. C. (2010). Trends and cycles in Euro area real GDP. *Applied Economics*, *42*(11), 1397-1401. https://doi.org/10.1080/00036840701721372

Feenstra, R. C., R. Inklaar & Timmer M. P. (2015), The Next Generation of the Penn World Table. *American Economic Review*, *105*(10), 3150-3182. http://www.jstor.org/stable/43821370

Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, *39*(2), 510-532. https://doi.org/10.1002/pam.22206

Ferman, B., & Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, *12*(4), 1197-1221. https://doi.org/10.3982/QE1596

Fernández, C. & Garcia-Perea, P. (2015). *The Impact of the Euro on Euro Area GDP Per Capita*. Working Paper No. 1530, Banco de España. https://dx.doi.org/10.2139/ssrn.2690211

Ferreira-Lopes, A., (2010). In or out? The welfare costs of EMU membership. *Economic Modelling 27*(2), 585-594. https://doi.org/10.1016/j.econmod.2009.11.013

Firpo, S., & Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference, 6*(2), 1-26. https://doi.org/10.1515/jci-2016-0026

Flam, H. & Nordström, H. (2006). *Trade volume effects of the euro: aggregate and sector estimates*. Seminar Papers 746, Stockholm University, Institute for International Economic Studies.

Frankel, J., & Rose, A. (2002). An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics*, *117*(2), 437-466. https://doi.org/10.1162/003355302753650292

Gabriel, R. D. & Pessoa, A. S. (2020). *Adopting the Euro: A Synthetic Control Approach*. MPRA Paper No. 99391. Munich Personal RePEc Archive.. https://dx.doi.org/10.2139/ssrn.3563044

Gasparotti, A., & Kullas, M. (2019). *20 Years of the Euro: Winners and Losers*. Centre for European Policy (CEP), 25, Freiburg.

Giannone, D., Lenza, M., & Reichlin, L. (2010). *Business Cycles in the Euro Area*. CEPR Discussion Papers 7124. Centre for Economic Policy Research. https://cepr.org/publications/dp7124

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*(2), 409-445. https://doi.org/10.1111/joes.12493

Greathouse, J. A., Bayani, M., & Coupet, J. (2023). Splash! Robustifying Donor Pools for Policy Studies. *arXiv preprint arXiv:2308.13688*. https://doi.org/10.48550/arXiv.2308.13688

Hancké, B. (2012). *Worlds apart? Labour Unions, Wages and Monetary Integration in Continental Europe*. Policy Paper No.128, IHS Political Science Series. http://aei.pitt.edu/id/eprint/33636

Hassel, A. (2014). *Adjustments in the Eurozone: Varieties of Capitalism and the Crisis in Southern Europe*. (LEQS Discussion Paper Series No.76.). LSE European Institute, London School of Economics. http://dx.doi.org/10.2139/ssrn.2436454

Hastie, T. J., Tibshirani, R. J., & d Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Ed. New York: Springer-Verlag. https://doi.org/10.1007/978-0-387-84858-7

Hope, D. (2016). Estimating the effect of the EMU on current account balances: a synthetic control approach. *European Journal of Political Economy*, *44*, 20-40. https://doi.org/10.1016/j.ejpoleco.2016.05.002

Iammarino, S., Rodriguez-Pose, A., and Storper, M. (2019). Regional inequality in Europe: evidence, theory and policy implications. *Journal of Economic Geography*, *19*(2), 273-298. https://doi.org/10.1093/jeg/lby021

Kuosmanen, T., Zhou, X., Eskelinen, J., & Malo, P. (2021). *Design Flaw of the Synthetic Control Method.* MPRA Paper No. 106390. Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/106390/

Jalles, J. T. (2010). How to measure innovation? New evidence of the technology-growth linkage. *Research in Economics*, *64*(2), 81-96. https://doi.org/10.1016/j.rie.2009.10.007

Lenihan, H., & Hart, M. (2004). The use of counterfactual scenarios as a means to assess policy deadweight: an Irish case study. *Environment and planning C: government and policy*, *22*(6), 817-839. https://doi.org/10.1068/c0413

Li, K. T., & Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics, 197*(1), 65-75. https://doi.org/10.1016/j.jeconom.2016.01.011

Mahoney, J., & Barrenechea, R. (2019). The logic of counterfactual analysis in case-study explanation. *The British Journal of Sociology*, *70*(1), 306-338. https://doi.org/10.1111/1468-4446.12340

Mäkelä, E. (2016). The price of the euro: evidence from sovereign debt markets. *Applied Economics, 48*(47), 4510-4525. https://doi.org/10.1080/00036846.2016.1161714

Mancini-Griffoli, T. & Pauwels, L. L. (2006). *Is There a Euro Effect on Trade? An Application of End-of-Sample Structural Break Tests for Panel Data*. IHEIDWorking Papers No. 04-2006, Economics Section, The Graduate Institute of International Studies.

McClelland, R., & Mucciolo, L. (2022). *An update on the synthetic control method as a tool to understand state policy.* Tax Policy Center. Urban Institute and Brookings Institution. https://www.taxpolicycenter.org/publications/update-synthetic-control-method-tool-understand-state-policy/full

Monastiriotis, V., Kallioras, D. & Petrakos, G. (2017). The regional impact of European Union association agreements: an event-analysis approach to the case of Central and Eastern Europe, *Regional Studies*, *51*(10), 1454-1468. https://doi.org/10.1080/00343404.2016.1198472.

Monfort, M., Cuestas, J. C., and Ordonez, J. (2013). Real convergence in Europe: A cluster analysis. *Economic Modelling*, *33*, 689-694. https://doi.org/10.1016/j.econmod.2013.05.015

Pesaran, H. M., Smith, V. L., & Smith, R. P. (2007). What if the UK or Sweden had joined the euro in 1999? An empirical evaluation using a global VAR. *International Journal of Finance and Economics 12*(1), 55-87. https://doi.org/10.1002/ijfe.312

Puzzello, L., & Gomis-Porqueras, P. (2018). Winners and losers from the €uro. *European Economic Review*, *108*, 129-152. https://doi.org/10.1016/j.euroecorev.2018.06.011

Rose, A. K. (2000). One money, one market: the effect of common currencies on trade. *Economic Policy*, *15*(30), 8-45. https://doi.org/10.1111/1468-0327.00056

Santos Silva, J. M. C. & Tenreyro, S. (2010). Currency Unions in Prospect and Retrospect. *Annual Review of Economics*, *2*(1), 51-74. https://doi.org/10.1146/annurev.economics.102308.124508

Smith, G. (2018). Step away from stepwise. *Journal of Big Data*, *5*(1), n.32, 1-12. https://doi.org/10.1186/s40537-018-0143-6

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (58)*, 267-288. http://www.jstor.org/stable/2346178

Verstegen, L., van Groezen, B., & Meijdam, L. (2017). *Benefits of EMU Participation: Estimates using the Synthetic Control Method*. CentER Discussion Paper No. 2017-032. Center for Economic Research. https://dx.doi.org/10.2139/ssrn.3027932

Vives-i-Bastida, J. (2023). Predictor Selection for Synthetic Controls. *arXiv preprint arXiv:2203.11576*. https://doi.org/10.48550/arXiv.2203.11576

Zheng, X., & Chen, S. X. (2023). Dynamic synthetic control method for evaluating treatment effects in auto-regressive processes. *Journal of the Royal Statistical Society Series B*: *Statistical Methodology*, *86*(1), 155-176. https://doi.org/10.1093/jrsssb/qkad103

Appendix

Outcome and predictors description

Outcome

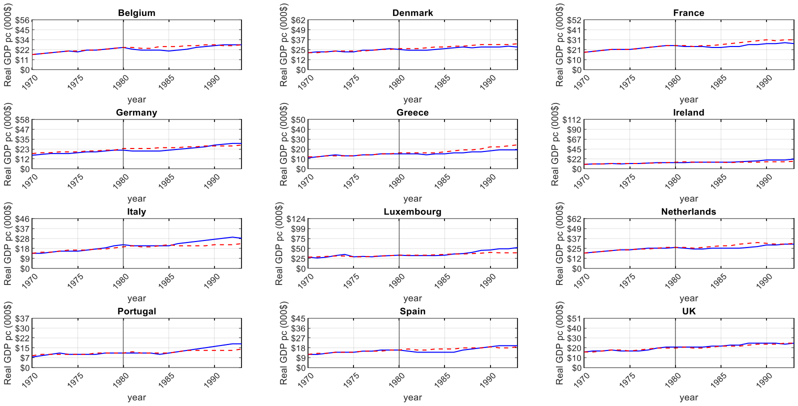
1. Real GDP per capita: Expenditure-side real GDP at chained PPPs (in mil. 2017US$) divided by the total population (in million). Data come from PWT 10.0 (series rgdpe and pop). These data are considered to be the most suitable for comparing living standards across countries and years (Feenstra et al., 2015).

Predictors

2. Balance of trade: exports-imports. Exports of goods and services (share of real GDP) - Data come from PWT 10.0 - Series name: Share of merchandise exports at current PPPs (csh_x). Imports of goods and services (share of real GDP) - Data come from PWT 10.0 - Series name: Share of merchandise imports at current PPPs (csh_m).
3. Private consumption expenditure (share of real GDP) - Data come from PWT 10.0 - Series name: Share of household consumption at current PPPs (csh_c).
4. General government final consumption expenditure (share of real GDP) - Data come from PWT 10.0 - Series name: Share of government consumption at current PPPs (csh_g).
5. Gross capital formation (share of real GDP) - Data come from PWT 10.0 - Series name: Share of gross capital formation at current PPPs (csh_i).
6. Resident patent applications per million population - Data come from WIPO statistics database. Patents are generally considered a good proxy of innovation (Jalles, 2010).
7. Employment share (ratio of total employment to total population) - Data come from PWT 10.0 - Series name: Number of persons engaged (in millions) and Population (in millions) (emp and pop).
8. Age dependency ratio: ratio of dependents (people younger than 15 or older than 64) to the working-age population (those aged 15-64). Data come from OECD.
9. GDP per hour worked (labour productivity) - data come from OECD (series name: gdphrwkd).
10. Inflation: yearly growth rate of a price index (CPI). Data come from OECD (series name: agrwth).
11. Human capital: Human capital index, based on years of schooling and returns to education -Data come from PWT 10.0 (hc).
12. General government debt (% GDP): Public debt as defined in the Maastricht criteria and generally defined as a proxy of the sustainability of government finance - Data come from The Global Debt Database (GDD)[2].

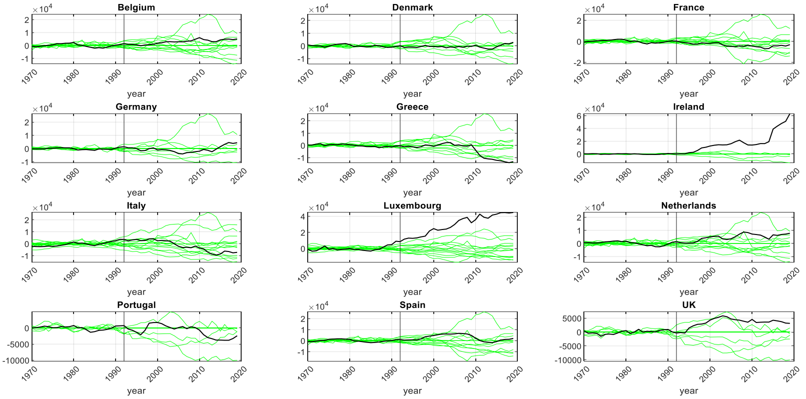
[2] For Mexico, we used public sector debt (percent of GDP), for New Zealand we used Central government debt (percent of GDP) while for all the other countries we used the General government debt (percent of GDP).

Figure A1. Benchmark case: In-time placebo test



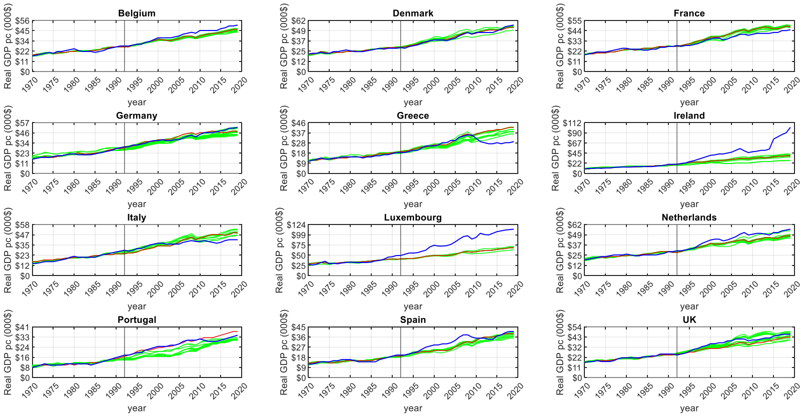
Note: In-time placebo test, or backdating, consists in reassigning the treatment period to a different year with the purpose to rule out anticipation effects. The placebo treatment date is 1980. For all countries, we do not observe the presence of relevant estimated effects in the pre-treatment period.
Source: authors’ calculations

Figure A2. Benchmark case: In-space placebo test



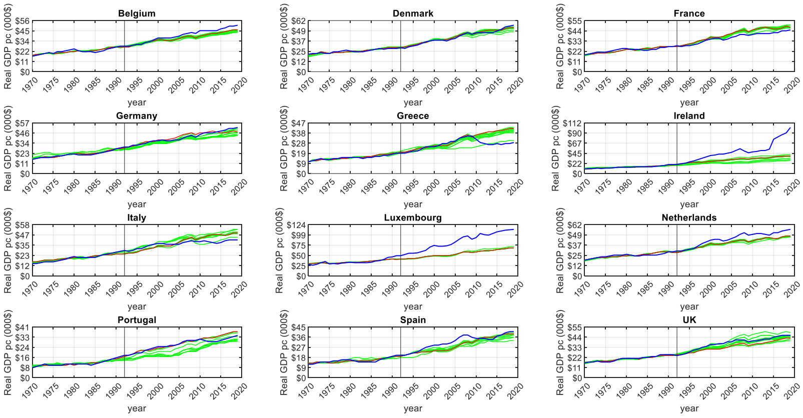
Note: In-space placebo test reassigns the treatment to donor pool countries with the purpose to rule out that the treatment had an effect on a donor pool country. The treatment has been reassigned to donor pool countries (green lines=GDP-SC); the dark line is the treated country (GDP-SC). In the graph, donor pool countries with a sufficient pre-intervention fit are reported (i.e., at most four times greater than the treated country pre-intervention RMSPE, as suggested in Firpo & Possebom (2018) and Gabriel & Pessoa (2020)). The figure indicates that donor pool units have not been significantly influenced by the treatment, i.e., no spillover effects are present.
Source: authors’ calculations

Figure A3. Leave-one-out test (donor pool countries)



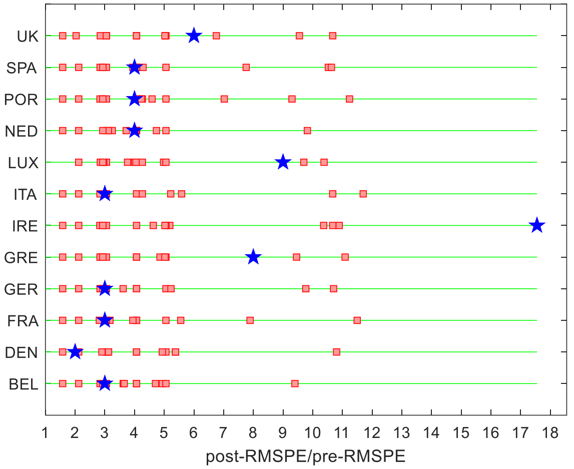
Note: Leave-one-out analysis checks whether results crucially depend on the research design. In the Figure, we analyze whether results depend on the presence of a particular country in the donor pool. The results provide no evidence of this type of dependence.
Source: authors' calculations

Figure A4. Leave-one-out test (predictors)



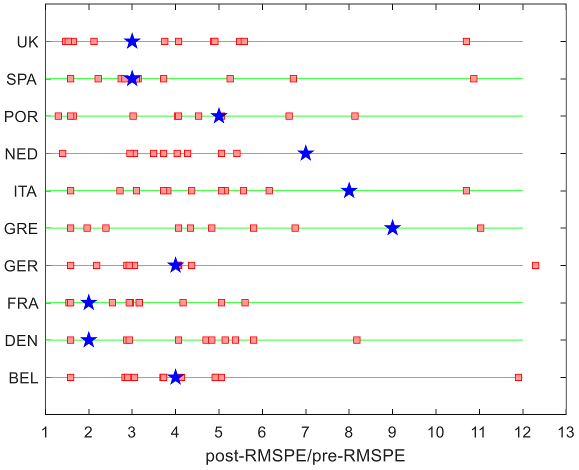
Note: As for Figure A3, applied to predictors, the results provide no evidence of dependence.
Source: authors' calculations

Figure A5. Statistical significance test - benchmark case



Note: A small pre-treatment RMSPE implies good SC fit, a large post- treatment RMSPE implies a relevant intervention impact. A larger ratio for the treated country (blue star) compared to the majority of the placebo-treated countries (i.e., donor pool countries treated in the in-space placebo) implies a significant treatment effect. For Ireland, the real ratio is extremely large (54) and, for graphical reasons, it has been plotted as 18.
Source: authors' calculations

Figure A6. Statistical significance test - post algorithm



As for Figure A5, with the results obtained by applying the algorithm.
Source: authors' calculations