



Synthetic Control Method for Dutch Policy Evaluation

Nils Verheuv¹ · Joost Witteman¹ · Marilou Vlaanderen¹

Accepted: 25 November 2022 / Published online: 2 January 2023
© Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

We apply the synthetic control method to a case study of the Dutch State Treasury Agency's funding policy. We study empirically what effect a change in funding strategy by the Dutch treasury had on market conditions. First, our results suggest that introducing more uncertainty to the funding policy, by means of a target range for capital market issuances, does not lead to a higher risk premium on Dutch government debt. Second, our paper shows that the synthetic control method, or the related constrained regression, is more suitable than a difference-in-differences method for this particular case study. The synthetic control method and constrained regression only include control units that are similar to the Netherlands. The difference-in-differences estimator includes all control countries, even ones that are not similar to the Netherlands. The difference-in-differences estimates incorrectly suggest that introducing more uncertainty in the funding policy leads to a higher risk premium. This shows that synthetic control and constrained regression are in some cases more suitable than a standard difference-in-differences. Furthermore, the synthetic control and constrained regression results hold when time and unit placebo tests are applied, whereas difference-in-differences results are not robust for this case study.

Keywords Policy evaluation · Sovereign debt management · Synthetic control method · Risk premia · Comparative methods

JEL Classification C80 · E60 · H63

1 Introduction

Governments have to fund their debt. In the Netherlands, this is the task of the Dutch State Treasury Agency (DSTA; Dutch: Agentschap van de Generale Thesaurie). To obtain funding at the lowest cost possible, with an acceptable risk to the budget, the DSTA has a funding policy in place. Funding policy choices have an impact on

✉ Nils Verheuv
n.verheuv@seo.nl

¹ SEO Amsterdam Economics, Amsterdam, Netherlands

market outcomes. Uncertainty surrounding the funding policy can result in higher risk premium on sovereign debt. In light of the *Regeling Periodiek Evaluatieonderzoek* (RPE),¹ which states that Dutch government policy has to be evaluated upon its effectiveness and efficiency at least every seven years, Hers et al. (2019) evaluated the funding policy and interest rate risk framework of the DSTA. Hers et al. (2019) applied the synthetic control method as in Abadie et al. (2010) for the evaluation of the funding policy, which was not used before in RPE-evaluations. This paper takes the case study of the DSTA funding policy to further analyze how the synthetic control method compares to more standard econometric methods.

The synthetic control method is an econometric tool to measure the effect of policy interventions (Abadie & Gardeazabal, 2003; Abadie et al., 2010). The synthetic control method assigns weights to several control units to closely match the treatment unit. The synthetic control method is particularly useful when variation in treatment is small, or even when there is only one treatment unit. Furthermore, even with relatively limited time series data, an effect can be estimated. This separates the synthetic control method from the ‘conventional’ econometric estimation models such as difference-in-differences and matching, for which applicability relies on a sufficient number of treated and untreated units. As such, the synthetic control method provides a drastic shift in our understanding of the (potential) effects of events that are unique or rarely observed. Athey and Imbens (2017, p.9) call the synthetic control method: “arguable the most important innovation in the policy evaluation literature in the last 15 years”.

The difference-in-differences method depends on the validity of the common trend assumption. The levels of the outcome variable may differ for the treatment and the control group, but they should have a common trend. The hypothesis is that if the treatment and control group follow a similar pattern in the pre-treatment period, they would also follow a similar pattern in the post-treatment period, if treatment were absent. This requires that no other events occurred to the control group that would affect this trend. Comparing the treatment and control group in the post-treatment period, while controlling for the trends, gives the treatment effect. The synthetic control method is more flexible as it only requires a subset of the control group to have a common trend with the treatment group.

The case studied in this paper concerns a change in the funding policy of the DSTA in 2016. The DSTA is responsible for obtaining funding to finance the Dutch government debt. Each year, the DSTA publishes an outlook with the amount of capital to be raised in the following year. Before 2016, the DSTA would publish an exact estimate of the amount to be raised on the capital market for the following year.² This changed for the outlook of 2016, published in late 2015. For the first time, DSTA announced a *target range* for the amount of capital to be raised, instead of an exact estimate. This introduced uncertainty to investors as they are now less sure which amount will be raised during the year. This provided the DSTA with more flexibility, but possibly at the cost of higher

¹ <https://wetten.overheid.nl/BWBR0040754/2018-03-27>.

² Excluding loans and bonds with a maturity shorter than one year.

uncertainty. Investors had less certainty whether they could fulfill their demand or not. As investors generally dislike uncertainty, they may require a higher risk premium on Dutch debt. This would become visible in an increased yield spread.

This paper has a twofold contribution. First, it yields practical insights for policy makers and sovereign debt managers. Second, it analyzes how the *reliability* of the synthetic control method differs from the reliability of the above-mentioned ‘conventional’ methods. These contributions derive from the following two main research questions.

Has the switch to a target range for capital market issuance increased the risk premium on Dutch sovereign debt?

How does applying the synthetic control method result in different outcomes compared to other, more standard, policy evaluation methods?

We answer these questions by applying various econometric methods. We select the difference-in-differences method, the standard synthetic control method, and a constrained regression. The constrained regression is the synthetic control method without any covariates. The synthetic control method and the constrained regression show that introducing more uncertainty in the DSTA’s funding policy did not lead to a significantly higher yield (spread).

In contrast, the difference-in-differences analysis points towards a significant and positive effect on the yield spread. However, the results do not seem to be fully robust when subjected to two types of placebo tests: time placebo tests and unit placebo tests. That is, the time placebo test shows that results are as well significant for the period in which treatment did not yet take place. Similarly, the unit placebo test shows that results are also significant for an untreated unit. As such, the difference-in-difference results do not seem robust. Both the synthetic control method and constrained regression yield insignificant results, thereby passing both placebo tests.

The results between the methods differ because the selection of control group countries differs. The synthetic control method and constrained regression give the highest weight to Finland, France, Germany, and Austria as control group, whereas other countries receive a small weight. These countries have a yield (spread) that is relatively close to the Dutch yield (spread). However, the difference-in-differences method also gives an equal weight to Belgium, Italy, Portugal, and Spain as control units, which have a higher yield (spread) than the Netherlands. The yield (spreads) for these countries declined more than the Dutch yield (spread) because they stood at substantially higher levels in the pre-treatment period. This explains the difference in the results. When the sample size is limited, each control unit has a large impact on the estimated coefficient. Therefore, control units that are similar to the treatment unit should be chosen. That is the major feature of the synthetic control method and the constrained regression.

The synthetic control method has a wide range of applications. In its first application, Abadie and Gardeazabal (2003) analyzed the economic costs of conflict using the Basque country as case study. Another classic application of the synthetic control method concerns the effect of anti-smoking legislation in California on smoking per capita (Abadie et al., 2010). The economic effects of the 1990 German reunification have also been studied using the synthetic control method (Abadie et al., 2014).

Following the seminal work of Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2014), the synthetic control method has been applied to various other case studies. Three general fields of research are natural disasters, terrorist attacks, and economic policies. Specifically, the synthetic control method has been used to study the effect of earthquakes on economic growth in Italy (Barone & Moncetti, 2014); the effect of hurricane Iniki on economic growth in Hawaii (Coffman & Noy, 2011); the effect of the 1928 Great Mississippi Flood on electoral outcomes (Heersink et al., 2017); and the effect of catastrophic natural disasters on economic growth (Cavallo et al., 2013). Another part of literature applies the synthetic control method to measure the effects of terrorism, as in Abadie and Gardeazabal (2003). Gautier et al. (2009) examined the effect of the Theo van Gogh murder on house prices in Amsterdam using a synthetic control method. The effect of the 2005 London bombings on labor markets and housing markets is studied by Ratcliffe et al. (2013). Furthermore, the synthetic control method is also applied in other crime-related fields (Saunders et al., 2015).

Lastly, many different economic policies are evaluated using the synthetic control method. Amongst others are: the effect of European integration on per capita income and labor productivity (Campos et al., 2014); the effect of youth minimum wages on employment in the United States (Powell, 2017); the effect of Right-to-Work laws on unionization, employment rate, and sectoral wages in the United States (Eren & Ozbelik, 2017); the effect of a VAT hike for Swedish restaurants and catering services on turnover, wages, profit, employment, and net entry (Falkenhall et al., 2020); the effect of a refugee wave on local labor markets, revisiting the Mariel boatlift using a synthetic control method (Peri & Yasenov, 2019); and the effect of the Brexit vote on stock prices and government bond yields (Opatrny, 2020).

This paper fits in the last strand of literature, namely the application of the synthetic control method in policy evaluation, especially for the Netherlands. Furthermore, this paper is the first to use the synthetic control method to study the funding policy of the government debt agency.

The structure of the paper is as follows. Section 2 describes the main methodology. Section 3 describes the DSTA case study and explains why this is an interesting case study for the synthetic control method. Section 4 shows the main results of the various methods and Sect. 5 concludes.

2 Methodology

This section describes the methodology for the econometric methods. The description of the methodology is based on Doudchenko and Imbens (2017; henceforth: D&I). We provide an understanding of the most salient features of the methods and further refer to D&I for methodological details.³

³ D&I further discuss issues regarding the objective function, estimation, and regularization. Based on the general form in Eq. (1), other methods can be derived as well, such as elastic net. We focus on DID, SC, and CR, because they are the most common applications.

We seek to estimate any causal effect of the change in the Dutch funding policy on Dutch sovereign debt funding outcomes, specifically the effect of the introduction of a capital issuance target range instead of a capital issuance point estimate on bond yield spreads relative to German 10-year yields. As in D&I, this corresponds to the panel data case where there are $N + 1$ cross-sectional units (one of which is treated) that are observed for T periods (both pre-treatment and post-treatment). Every cross-sectional unit has two associated potential outcomes for every period, namely the outcome given treatment and given non-treatment / control, or $Y_{it}(1)$ and $Y_{it}(0)$ respectively. The causal effect of treatment at the unit-time index then is the difference between potential outcomes. The empirical issue is that post-treatment the potential outcome given control is not observed for the treated – who are treated after all. In terms of our case study, we do not observe the (counterfactual) Dutch yield spreads post the introduction of the capital issuance target range if the capital issuance target range had not been introduced. The empirical problem is the imputation of the (unobserved) potential outcome given control for the treated post-treatment (D&I). There are different models available to this end.

D&I consider the class of models where the unobserved potential outcome is interpolated as a linear model. Suppose that only unit 0 is treated and we consider time t' a post-treatment period. The treatment effect is $\tau_{0t'} = Y_{0t'}(1) - Y_{0t'}(0) = Y_{0t'}^{obs} - Y_{0t'}(0)$, where the last equation follows from the fact that for the treated, we observe the realized post-treatment outcome. D&I remark that many estimators in the literature impute $Y_{0t'}(0)$ as:

$$\hat{Y}_{0,t'}(0) = \mu + \sum_{i=1}^N \omega_i Y_{i,t'}^{obs} \tag{1}$$

As such, the imputed control outcome for the treated is a linear combination of control units. One way to identify the parameters μ and ω_i in Eq. (1) is to estimate them with ordinary least squares over the full panel. D&I highlight that in practice this estimation may be infeasible if there are more control units than pre-treatment periods, or imprecise depending on the relative magnitude of the number of units and the number of periods. In practice then, researcher must impose restrictions on parameters μ and ω_i . D&I remark that different restrictions on the parameters imply different estimation strategies. Specifically, D&I note that by considering the following constraints:

1. NO INTERCEPT: $\mu = 0$. The model does not have an intercept.
2. ADDING UP: $\sum_{i=1}^N \omega_i = 1$. The weights on the control units add up to 1.
3. NON-NEGATIVITY: $\omega_i \geq 0, i = 1, \dots, N$. The weights on the control units are non-negative.
4. CONSTANT WEIGHTS: $\omega_i = \bar{\omega}, i = 1, \dots, N$. The weights are the same for all control units.

As a result, we can consider (at least) differences-in-difference (DID), synthetic control (SC) and constrained regression (CR) as part of the same class of models.

Table 1 Overview methods and constraints. Source: Doudchenko and Imbens (2017)

Binding constraints	No intercept	Adding-up	Non-negativity	Constant weights
Difference-in-differences	No	Yes	Yes	Yes
Synthetic control	Yes	Yes	Yes	No
Constrained regression	No	Yes	Yes	No

CR is a form of SC, but without any covariates. Specifically, Table 1 notes which combination of constraints corresponds to which estimator:

- 1 *No intercept.* This constraint rules out that the treatment unit has systematically larger outcomes than the control units by a constant amount. This constraint is implausible if the treatment unit is an outlier with respect to the control units (D&I). This constraint applies to SC.
- 2 *Adding up.* The weights on the individual units should add up to 1. When the weights have to sum up to one and the treatment unit is an outlier, a synthetic control unit cannot be constructed, as the synthetic control unit is a weighted average of other control units, and the treatment unit has systematically smaller or larger outcomes. This constraint applies to DID, SC, and CR
- 3 *Non-negativity.* This constraint in SC helps for regularization and ensures only a few control units have non-zero weights. If the partial correlations between the outcomes for a control unit and the treatment unit are negative, a negative weight may be more appropriate and improve the fit of the outcome. This constraint applies to DID, SC, and CR.
- 4 *Constant weights.* This method ensures the weights do not differ per control unit; all control units have an equal weight. This constraint applies to DID.

D&I show for four classic case studies⁴ that the choice of estimator significantly influences the estimated coefficients. The three methods (DID, SC, and CR) are applied to the case study regarding public debt management by the DSTA. The case study is described in chapter 3.

D&I argue that neither difference-in-differences, synthetic control, or constrained regression may be universally optimal. Instead, depending on the relative magnitude of the number of control units and the number of pre-treatment periods, different combinations of the constraints mentioned above may produce more or less credible estimates. Indeed, as per Abadie and Gardeazabal (2003) and Abadie et al. (2010) synthetic control can be a suitable policy evaluation method when the variation in treatment is small, or when there is only one treatment unit (Abadie & Gardeazabal, 2003; Abadie et al., 2010). Furthermore, with relatively limited time series data, an

⁴ The Mariel boatlift study (Card, 1990), the New-Jersey Pennsylvania minimum wage study (Card & Krueger, 1994), the California smoking legislation study (Abadie et al., 2010), and the German reunification study (Abadie et al., 2014).

effect can be estimated. When there is a limited number of control units available in the data, the synthetic control method still produces reliable results. In our case study below, we compare the estimates produced by difference-in-differences, synthetic control, and constrained regression to highlight the effect of the choice of estimation method in a macro-economic policy evaluation context.

3 Case study

This paper applies the abovementioned methods to the case study regarding DSTA's funding policy. Section 3.1 gives a short description of funding policy in public debt management and describes the specific policy change, namely the introduction of a target range for capital market issuances. It also argues why this case study is suitable for the comparison of the various methods. Section 3.2 gives the empirical specification and Sect. 3.3 discusses the descriptive statistics.

3.1 DSTA's funding policy

This paper applies the discussed methods to a case study regarding the public debt management by the Dutch State Treasury Agency (DSTA). The DSTA has the task to ensure that the government's financing needs, and its payment obligations are met at the lowest possible cost over the medium to long run, consistent with a prudent degree of risk. In its funding policy, the DSTA has three guiding principles: consistency, transparency, and liquidity. Consistent and transparent debt management reduces uncertainty for investors, arguably leading to a lower risk premium on the sovereign debt. This contributes to the objective of lowest funding cost possible. Liquidity ensures Dutch sovereign bonds are attractive to investors, and that the Dutch government can raise enough capital if necessary. Furthermore, investors are willing to pay a liquidity premium for liquid instruments, leading to a lower yield on the debt and thus lower funding costs.

To increase the consistency and transparency of its funding policy, the DSTA issues an Annual Outlook at the end of the year in which it announces how much funding it will obtain by issuing certain bonds and bills in the upcoming year. However, since 2016 there has been a significant change in the policy. Before 2016, the DSTA announced a fixed amount it would raise on the capital market (debt instruments with a maturity of more than one year) in the following year. From 2016 onwards, however, the DSTA did not announce a fixed amount to be raised on the capital market, but it announced a target range instead. The target range has a width of 4 to 6 billion euros on average. For 2020 the announced target range was set at € 21 to € 26 billion.⁵

The switch to a target range marked a change from a more consistent and transparent policy to a more flexible policy. The DSTA (2015) stated "*Given the*

⁵ Which in the end turned out to be too low, as the Covid-19 stimulus package led to a large increase in the financing need.

circumstances, a bit more flexibility in the split between the call on the capital market and the money market is deemed desirable.” The previous year the DSTA had a lower funding need than expected, which meant that with a fixed amount to be raised on the capital market, there was a disbalance with money market issuances. The DSTA therefore introduced more flexibility by using a target range for capital market issuances instead. They did not take countermeasures to control for this, apart from their usual focus on consistency in the funding policy, which remained unchanged.

Since 2016, investors face more uncertainty regarding the amount to be issued, as they only have a target range as an indication. As investors face more uncertainty, they may demand a larger risk premium on the Dutch government bonds, increasing the yield on the Dutch sovereign debt. Anecdotal evidence by primary dealers indicates that they indeed perceived the change as increasing uncertainty in the funding policy. Specifically, investors face more uncertainty whether they can fulfill their demand for Dutch government bonds and serve their clients, or whether they fall short of their demand. Another possibility is that they have to take up more government bonds than planned, leading to adjustments in their portfolio. This uncertainty could have potentially negative effects on the yields of the bonds. The expectation was that the change would not fundamentally change the risk profile of Dutch government bonds, because on other aspects the DSTA’s funding policy scores high on consistency and transparency (Hers et al., 2019). However, a small impact was expected.

This leads to our hypothesis that the shift away from estimates and towards targets was accompanied by a higher risk premium. To empirically test this hypothesis, we use the yield spread between Dutch and German 10-year government bonds as a proxy for the risk premium for Dutch government bonds. Since German bonds are considered to be close to risk free, the yield spread against German bonds is a measure of the risk premium. Figure 1 sets off the yield spread for the countries in the sample for the period 2012 to 2019. The Netherlands has the lowest yield spread in general, although the Finnish yield spread is also very low, and sometimes lower than the Dutch yield spread. Notice that the scale in Panel B is differs from the scale in Panel A, as Italy, Portugal, and Spain had substantially higher yield spreads.

The case study is suitable to test the synthetic control method because it fits the four minimum requirements, as described by Abadie et al. (2010).

1. The treatment happened to only one unit (the Netherlands). Other countries did not implement a similar policy before or after the treatment.⁶
2. The treatment is a one-time event, i.e., it happened only once (2016Q1) and then remained unchanged.
3. There are sufficient time series data available to empirically test the effect of the change to a target range on the risk premium on Dutch sovereign debt.

⁶ Abadie (2021) states that modern applications of the synthetic control method can account for multiple treated units as well.

4. The number of control units is small, which is exactly the context for which synthetic control method was designed for.

3.2 Empirical specification

3.2.1 Data and control variables

We empirically estimate the risk premium as the Dutch 10-year yield spread versus Germany. The yield spread is a good measure of the risk premium. Using the yield spread excludes the possibility of including Germany as control unit, therefore we also present analysis using the yield, not the spread.

Other variables can also influence the yield (spread). For control variables, it is important that they vary per country. Variables that influence the yield (spread) but do not vary per country, such as the monetary policy rate, do not give any information about which set of countries closely replicates the Netherlands. These relevant control variables to base the matching on are:

- S&P sovereign credit rating, transformed into numerical values
- Industrial production to measure economic growth
- Government debt as percentage of GDP.
- Government balance as percentage of GDP.

These control variables are relevant because they are expected to influence the yield (spread). A higher credit rating signifies higher creditworthiness and thus a lower risk premium on the government debt. A higher growth rate of industrial production is positively associated with economic growth and ensures that the government has a larger future stream of income to repay the loan, thus reducing risk. A country with a higher government debt ratio has to repay a larger amount of debt, relative to the earning capacity of the economy, meaning the debt funding is riskier (Afonso et al., 2015). A country with a larger government deficit needs to attract more capital on the market, raising the yield on sovereign bonds. Additionally, a consistent deficit deteriorates the government's ability to repay the debts, raising the yield spread.

The dataset contains quarterly data over 2012Q4 until 2019Q4. The treatment occurred at 2016Q1. As such, the dataset includes sufficient data before the treatment to construct the synthetic control group, and also includes sufficient data after the treatment to observe any treatment effects. The dataset consists of observations on nine euro area countries: Austria, Belgium, Finland, France, Germany, Italy, Netherlands, Portugal, and Spain. The data ends at 2019Q4 so the Covid-period does not influence the results. Additionally, four years of post-treatment data is likely sufficient to detect any effects.

The training period is from 2012Q4 to 2015Q4. The financial crisis and eurozone crisis do not influence the country weights as such. As shown below, the Netherlands and the control group follow a similar trend for the yield spread from 2012Q4 to 2015Q4 onwards. Before 2012Q4, the Dutch yield spread had a significantly

different trend, making that period less suitable for comparison. A wider time range, for example from 2006Q1 onwards was analyzed, but did not improve the results. The pre-treatment matching thus is based on data from 2012Q4 to 2015Q4.

Data on 10-year yields is obtained from OECD, data on government debt from Eurostat, data on industrial production from IMF and the credit ratings are based on S&P data.

3.2.2 Difference-in-differences specification

This paragraph discusses a classic difference-in-differences model for the yield spread. DID controls for differences in initial levels of the yield spread. It examines how the Dutch yield spread has developed relative to the development of the control groups' yield spreads. The main assumption is that before the treatment took place (2016Q1), the treatment unit (Netherlands) and the control group show a common trend in the yield spread. In its basic form, the difference-in-differences regression looks as follows:

$$\text{DID} : y_{it} = \beta_0 + \beta_1 NL_i + \beta_2 \text{Post2016Q1}_t + \beta_3 (NL \times \text{Post2016Q1})_{it} + \varepsilon_t \quad (2)$$

y_{it} is the yield spread for country i at quarter t . NL is a dummy for the Netherlands, the treatment unit. Post2016Q1 is a dummy that indicates the post-treatment period, and the interaction term is a dummy that takes on a value of one if the country is the Netherlands and the time is in the post-treatment period, and zero otherwise. β_3 is the DID estimator, as it tells how the development of the Dutch yield spread compares to the development of the control group yield spreads in the post-treatment period. This basic form can be extended by including time fixed effects to better capture the trend in yield spreads, and by including the macroeconomic control variables to capture variation difference between countries. The models include heteroskedasticity consistent standard errors. The standard errors are not clustered due to the low number of clusters this would yield.

To test the common trend in the pre-treatment period, we regressed the yield spread on the macroeconomic control variables, a dummy for the Netherlands, time dummies, and an interaction parameter between the Netherlands and the time dummies. If the Netherlands has a common trend with the control group countries, the coefficient on the interaction effect should be statistically insignificant. For the period 2010Q1–2012Q3, the interaction effect was statistically significant in most time periods. This indicates that the Netherlands followed a different trend than the control group countries. However, from 2012Q4 onwards, the coefficient on the interaction effect was statistically insignificant, indicating no different trend for the Netherlands and the other countries. Hence, the pre-treatment period is set at 2012Q4–2015Q4, as the Netherlands and the other countries had a similar trend in the yield spread.

After 2016, the economic fundamentals moved in the same direction for most countries, as shown by Tables 13 and 14 in the Appendix. Almost all countries saw an improvement in their government balance, including the Netherlands. The countries that have high S&P credit ratings pre-treatment, continue to do so

Table 2 Cross-country comparison yield spreads. Source: Own calculations, based on OECD data for yields

Country	Pre-treatment 2012Q4–2015Q4	Post-treatment 2016Q1–2019Q4	Change post-treatment (%) 2016Q1–2019Q4 versus 2012Q4–2015Q4
	10-year yield spread versus Germany (in percentage points)		
Netherlands	0.29	0.19	– 34%
Austria	0.35	0.29***	– 17%
Belgium	0.60***	0.55***	– 8%
Finland	0.27	0.27***	0%
France	0.51***	0.41***	– 20%
Italy	2.01***	1.90***	– 5%
Portugal	3.35***	2.07***	– 38%
Spain	2.10***	1.12***	– 47%
All except Netherlands	1.31***	0.92***	– 30%
All except Netherlands, Italy, Portugal, Spain	0.43***	0.34***	– 21%

*, **, ***: mean yield spread differs from Dutch mean at 10, 5, and 1% level, respectively

post-treatment. That being said, the Netherlands is the only country within this group for which the credit rating increases even further. S&P increased the rating for the Netherlands from AA+ to AAA in 2015Q4, so before treatment took place. Because changes in credit ratings are small in both directions, the extent to which countries resemble the Netherlands in terms of credit rating remains largely unaffected. For example, Finland maintained the same credit rating (AA+) from 2014Q4 until 2019Q4. So both the Netherlands and Finland, which are similar in terms of the yield (spread), no post-treatment change in the credit rating occurred.

For industrial production the Netherlands saw a slight decrease, whereas and other countries experienced an increase in the post-treatment period compared to the pre-treatment period. Nevertheless, the data shows that the trends in macro factors do not diverge much. Additionally, we have found no evidence of a similar funding policy change in the control group countries. Hence, the forward-looking nature of the common trend seems to hold as well.

As the common trend assumption likely holds, the DID regressions can be executed. There are four main models:

1. Basic DID (as described above)
2. Basic DID + time fixed effects
3. Basic DID + macroeconomic control variables
4. Basic DID + macroeconomic control variables + time fixed effects

The regression is performed on the same set of countries and the same time period as SC and CR, to allow comparability between the results. Again, Germany

is excluded as it has a zero yield spread by default. This is in line with SC and CR, where Germany is also excluded from the estimation. The macroeconomic control variables are the numerical S&P rating, the log of industrial production, the government balance and the government debt ratio. Additionally, as SC and CR implicitly control for each time period, we include time fixed effects in the DID models as well.

3.3 Descriptive statistics

Table 2 shows that Finland and the Netherlands are very similar in the pre-treatment period. It is thus expected that Finland receives a large weight in SC. Most other countries have a significantly higher pre-treatment yield spread. Austria, Belgium, and France have a pre-treatment yield spread that is 1.2 to 2 times higher than the Dutch yield spread. However, the Italian, Portuguese, and Spanish yield spreads are 7 to 12 times higher. The average yield spread for the control group in the pre-treatment period is 1.31 percentage points, substantially higher than the Dutch yield spread. When Italy, Portugal, and Spain are excluded, the average is 0.43% points, which is still roughly 50 percent higher, but closer to the Dutch yield spread. This indicates that the weighted average of all countries does not represent the Netherlands well, hence SC can be used.

The hypothesis is that the yield spread on Dutch government debt increased as a result of the change towards a more flexible funding policy. The descriptive statistics above show that the yield spread is even lower in the post-treatment period (Table 2).⁷ This shows that the yield spread did not rise. However, almost all countries show a downward-sloping trend in the yield spread. It could be that the Dutch yield spread only decreased because of this common trend, but that the decrease was smaller than in other countries. The final column shows that the Dutch yield spread declined 34 percent compared to 2012–2015. Portugal and Spain experienced a larger decline, but the decline was substantially smaller for Austria, Belgium, Finland, France, and Italy. This comparison does not show evidence in favor of the hypothesis that the change in the funding policy led to a significantly higher yield spread. However, no conclusions can be drawn on the basis of these summary statistics. Instead, the formal models must point out whether there is a significant effect or not.

In addition to matching based on yields or yield spreads, the synthetic control group is also constructed using matching on several macroeconomic control variables: S&P credit rating, industrial production, and the government debt ratio (Table 13 in the Appendix). The Netherlands had an average S&P credit rating of 10.4 in the pre-treatment period, which is between AA and AAA, with the balance shifted towards AA. This is a very high credit rating, which translates into a low risk premium. Germany had the highest credit rating, namely AAA at any point in time in the post-treatment period. France and Belgium had a similar

⁷ Descriptive statistics for the yield are given in Table 13 the Appendix.

credit rating as the Netherlands, but slightly lower, namely between A and AA. The Netherlands had a medium ranking for industrial production index, similar to Finland and France. Apart from Finland, the Netherlands had the lowest government debt ratio.

A statistical test on the means in the pre-treatment period shows that the Netherlands differs from all countries on some variables. The control variables for the Netherlands are most similar to Finland. However, there is not one country that exactly matches the Netherlands on all aspects, which is why a combination of countries is required to construct the synthetic control group.

4 Results

This chapter discusses the main differences between the various methods. In Sect. 4.1 the country weights that determine the composition of the (synthetic) control group are compared. In Sect. 4.2, the predictor balance shows how well the (synthetic) control group matches the pre-treatment data for the Netherlands. If the (synthetic) control group matches the data for the Netherlands well, a treatment effect can be estimated. Finally, placebo tests in Sect. 4.3 show whether the results are robust.

The following methods are tested in this chapter:

1. Yield: SC
2. Yield: CR
3. Yield: CR relaxing ADDING UP
4. Yield spread: SC
5. Yield spread: CR
6. Yield spread: CR relaxing ADDING UP
7. Yield spread: DiD
8. Yield spread: DiD excluding periphery countries

Chapter 2 listed the three constraints that are at play for SC and CR: ADDING UP, NON-NEGATIVITY, and NO CONSTANT. The ADDING UP constraint can be relaxed using a simple trick, to show that this leads to significantly different results. Recall that as the yield spread is calculated against German yields, Germany does not appear as a control group for the yield spread, since the yield spread is always zero. Exactly this feature can be used to relax ADDING UP. ADDING UP means that the sum of all weights must be equal to 1. One could surpass this constraint by including Germany as control group, which always has a yield spread of 0. Thus, the weight given to Germany does not show up in the synthetic yield spread. Effectively, the sum of the weights is then one minus the weight on Germany. For example, if France and Germany both receive a weight of 0.5, the synthetic yield spread is calculated as $(0.5 \times 0 + 0.5 \times \text{yieldspread}_{FR})$, which reduces to $(0.5 \times \text{yieldspread}_{FR})$. Effectively, the weights sum up to only

Table 3 Country weights differ per method. Source: Own calculations

Variable	Yield (in percentage points)			Yield spread (in percentage points)				
	SC	CR	CR no ADDING UP	SC	CR	CR no ADDING UP	DID	DID (excl. periphery)
Country								
Austria	0.7	6.6	10.0	35.3	1.0	6.0	14.2	25.0
Belgium	7.7	5.9	9.5	0.0	0.6	5.3	14.2	25.0
Finland	71.0	6.9	10.2	64.6	96.8	6.3	14.2	25.0
France	2.4	6.1	9.7	0.0	0.7	5.5	14.2	25.0
Germany	16.6	66.6	10.8			68.8		
Italy	0.7	3.2	7.0	0.0	0.3	3.2	14.2	0.0
Portugal	0.3	1.6	5.2	0.0	0.2	1.9	14.2	0.0
Spain	0.7	3.1	6.9	0.0	0.3	3.1	14.2	0.0
Zero			30.8					
Mean squared prediction error	0.3	0.7	2.8	0.3	0.4	0.7		

0.5. The ADDING UP constraint is relaxed. The same analysis can be done for the yield. Although Germany does not have a zero yield (at least not over the entire sample), a new unit has to be created. This unit is called ‘Zero’ and always has a zero yield.

The various models are tested on the same data, so any difference in results does not stem from differences in data. The pre-treatment period is 2012Q4–2015Q4, using quarterly data. The post-treatment period is 2016Q1–2019Q4. The models do not include any special predictors, as those did not add to the fit of the models. The synthetic control models include industrial production, S&P credit rating, and the government debt to GDP ratio as control variables.⁸ The significance of the results is analyzed using time and unit placebo tests in Sect. 4.4.

Country weights.

The synthetic control group for the Netherlands consists mainly of Finland and Germany, such that other countries receive a smaller weight (Table 3).⁹ When the yield spread is used, the synthetic control group weighs heavily on Finland (65 to 98%). However, when ADDING UP is relaxed, Germany has the largest weight (69%). South-European countries such as Italy, Portugal, and Spain generally receive a low weight, below 7 percent. This is as expected, as the descriptive statistics showed that these countries are not similar to the Netherlands. Austria receives

⁸ The long-term government debt to total debt ratio was also tested but received a zero weight in all models and are thus excluded from the models.

⁹ With SC, the control variables also receive weights. CR includes only 1 control variable, the dependent variable, which has a weight of 1. The weights for the control variables in SC with the yield are 0.911 for the yield, 0.033 for the S&P credit rating, 0 for industrial production, and 0.057 for the government debt ratio. The weights for the control variables in SC with the yield spread are 0.967 for the yield spread, 0.027 for the S&P credit rating, 0.006 for industrial production, and 0 for the government debt ratio.

Table 4 V-weights for control variables in SC and CR models. Source: Own calculations of synthetic control, constrained regression, and constrained regression with relaxing the ADDING UP constraint for the yield spread

Dependent Variable	Yield			Yield spread		
	SC	CR	CR no ADD- ING UP	SC	CR	CR no ADDING UP
Yield (spread)	55.1%	100.0%	100.0%	68.3%	100.0%	100.0%
S&P rating	26.7%			3.3%		
Industrial production	0.0%			0.0%		
Government debt ratio	15.2%			28.4%		
Government balance	0.0%			0.0%		

a high weight in the synthetic control group for the yield. Belgium also does not receive a weight larger than 10%. Concluding, the weights per method differ per method but are similar in terms of relative size.

For replicating the yield, SC has the best fit (because it has the lowest MSPE). For the yield spread however, SC and CR perform similarly. Relaxing the ADDING UP constraint does not improve the fit of the models. However, for the yield spread, the MSPE is still small. That being said, because relaxing the ADDING UP constraint can be relevant as the Dutch yield spread is generally the lowest in the sample, it is included in the main discussions.

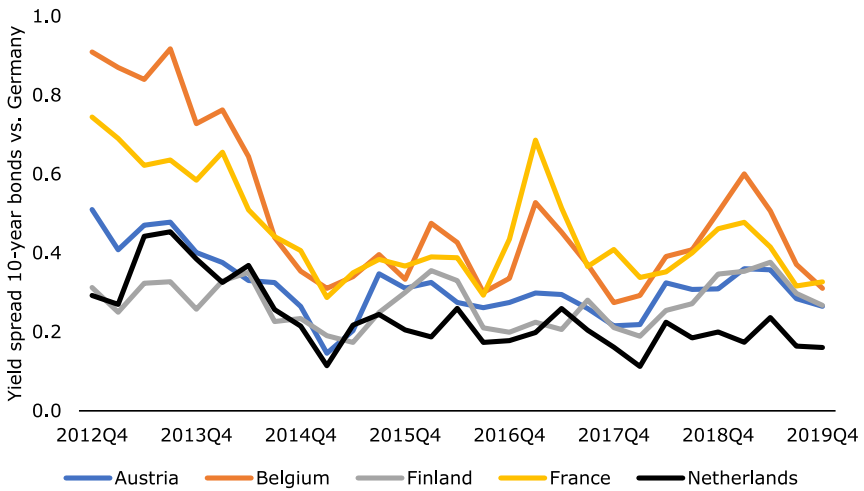
For DID, all countries receive the same weight. As Germany is excluded since it does not have a yield spread, there are seven countries in the control group, meaning each country receives a 14.2% weight. As Spain, Italy and Portugal are very different from the Netherlands (they receive a zero or small weight under SC and CR), a DID based on Austria, Belgium, Finland, and France may prove to be a better model, as they are more similar to the Netherlands. In this case, those countries all receive a 25 percent weight.

A key result is that SC and CR have some differences, but the main difference is with respect to DID. DID assigns an equal weight to each control unit. The selection of control units is thus of major importance. This especially holds when the number of control units is small, as in this case study. This results in very different weights than under SC or CR, and in turn significantly impacts the results and the conclusions, as the next section shows.

Next to country weights, so-called w-weights, synthetic control models also calculate control variable weights, so-called v-weights. These v-weights indicate the weight of a control variable in selecting the country weights. A higher v-weight indicates that a control variable helps predicting the outcome variable more than a control variable weigh a lower weight.

For CR, there is only one control variable, namely the dependent variable, so the v-weight for the yield (spread) is 100%. However, SC uses multiple control variables, so has different v-weights for these variables. Table 4 shows that the yield (spread) receives the highest weight, namely 55% for the yield and 68% in the model

Panel A



Panel B

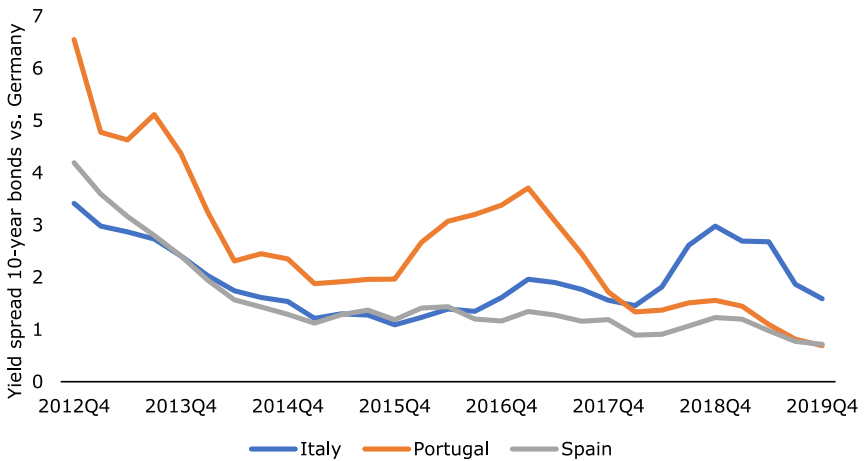


Fig. 1 Yield spread vs. Germany lowest for the Netherlands. Source:OECD. Y-axis denotes the yield spread on 10-year government bonds versus Germany. The X-axis denotes the year and the quarter. Notice that the scale on the Y-axis is different for panel B

Table 5 Predictor balance models yield. Source: Own calculations

Variable	Netherlands Yield	SC	CR
Yield (%)	1.39	1.39	1.39
S&P rating	10.39	10.38	
IP	97.14	98.27	
Government debt ratio (%)	67.25	67.24	
Government balance (%)	-2.80	-2.21	

Table 6 Predictor balance models yield spread (in percentage points). Source: Own calculations. Data over 2012Q4–2015Q4

Variable	Netherlands	SC	CR	CR no ADDING UP	DID	DID (excl. periphery)
	Yield spread					
Yield spread (%)	0.29	0.30	0.29	0.29	1.15	0.35
S&P rating	10.39	10.40			6.94	9.98
IP	97.14	100.20			99.03	102.82
Government debt ratio (%)	67.25	67.21			98.17	84.25
Government balance (%)	-2.80	-2.40			-3.42	-2.31

for the yield spread. For the yield, the S&P rating receives a weight of 27%, and the government debt ratio has a v-weight of 15%. Industrial production and the government balance receive a zero weight, meaning these variables did not affect the selection of the country weights. For the yield spread, the S&P rating has a lower weight, 3%, and the government debt ratio has a higher weight, 28%. Again, industrial production and the government balance received a zero v-weight. Concluding, in predicting yields or yield spreads, the most important control variables are the S&P credit rating and the government debt ratio.

Predictor balance.

Using the country weights, the predictor balance can be calculated. This shows how well the synthetic control group matches the data for the Netherlands. If the matching is done well, there should not be a large difference between the data for the Netherlands and for the synthetic Netherlands.

The Dutch 10-year yield is matched exactly by the synthetic control group in both SC and CR (Table 5). The control variables are also matched closely. This shows that both SC and CR lead to a good synthetic control group, as it is by and large indistinguishable from the Netherlands in the pre-treatment period. Additionally, SC matches the control variables well too. Similar results are shown in Figs. 6 and 7 in the Appendix. In these graphs, the synthetic yield can hardly be distinguished from the Dutch yield in the pre-treatment period, showing that the synthetic control group is a good match for the Netherlands.

The yield spread is also matched closely by SC, CR and CR no ADDING UP (Table 6). This is also shown in Fig. 8 in the Appendix, the synthetic yield spread and the Dutch yield spread show a similar trend from 2012 onwards. The movements are not matched perfectly, but both lines show a declining trend. For SC, most control variables are also matched closely, especially the S&P rating and the government debt ratio. The fit is less accurate for industrial production and the government balance, but these variables also receive a zero v-weight in SC.

The predictor balance for DID models is considerably worse. The average yield spread for the DID control group is four times higher than the average pre-treatment yield spread for the Netherlands. The average S&P rating is also much lower. When limiting the DID sample to Austria, Belgium, Finland, and France, the results are

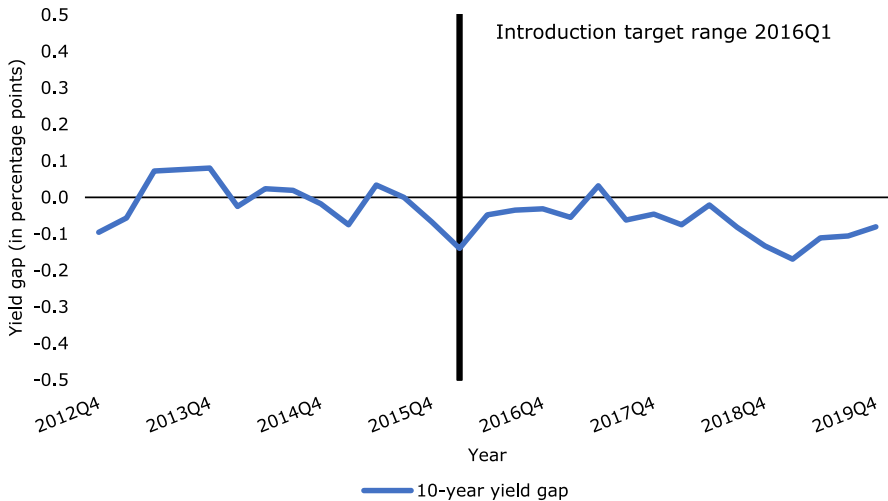


Fig. 2 No treatment effect on yield after 2016Q1. Source: Own calculations of synthetic control model for the yield

more aligned, but still worse than the predictor balance for SC or CR. The average yield spread is now 5 basis points higher than the yield spread for the Netherlands, and the S&P rating, government balance, and government debt ratio resemble the Dutch data better than the full DID. However, it still underperforms compared to SC and CR. This shows that the DID control group is not so similar to the Netherlands.

Concluding, both the SC and CR result in a good match for the yield and the yield spread. This means that the constructed synthetic control group is very similar to the Netherlands, so that it can be used for an analysis of the treatment effect. When DID is used, the fit is considerably worse, signaling that the DID control group is very different from the Netherlands, which may influence the results and their reliability.

Treatment effect.

Since SC and CR prove to generate a valid control group for the Netherlands when analyzing the yield and yield spread, the treatment effect can be estimated. Because the synthetic control group closely matches the Dutch yield (spread) in the pre-treatment period, and no similar policy changes have occurred to the control group countries, any divergence in the post-treatment period can be considered the treatment effect of the introduction of the target range. An additional condition is that there was no other shock to funding policy in the control countries. We have been unable to identify a similar shock, indicating that this does not affect the results. Additionally, most macroeconomic control variables developed in a similar way, as mentioned in Sect. 3.2.2.

The results suggest that the introduction of the target range did not have a significant effect on the Dutch 10-year yield. In the pre-treatment period, the gap generally hovers around zero in the pre-treatment period, with a bandwidth of 0.1% points (Fig. 2). The deviations from zero are both upwards and downwards, implying that on average the synthetic yield matches the Dutch yield well. At the moment of

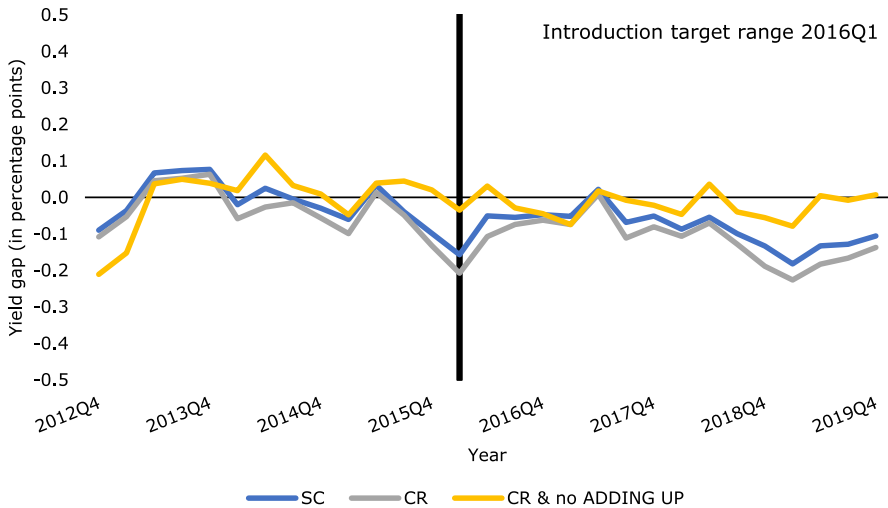


Fig. 3 No treatment effect on yield spread found with constrained regression. Source: Own calculations of synthetic control, constrained regression, and constrained regression with relaxing the ADDING UP constraint for the yield spread

treatment (2016Q1), the Dutch yield is 0.14% points below the synthetic yield. In the post-treatment period, the gap still hovers around 0, but is negative in most periods, indicating the Dutch yield is lower than the synthetic yield. This shows that the Dutch 10-year yield was not significantly increased by the introduction of the target range. Using CR to construct the synthetic control group for the yield also does not result in any significant treatment effect (in the Appendix).

The lack of significant pseudo-treatment effects could reflect the actual null distribution. However, the fluctuations in the pseudo-treated units persist post-treatment. As such, the lack of significance could as well be the result of a lack of precision rather than a lack of pseudo-effects. As a result, the null effect in the post period remains to some extent uncertain.

The introduction of the target range also did not have a significant effect on the yield spread, using SC, CR, and CR no ADDING UP. The gap between the Dutch yield spread and the synthetic yield spread fluctuated around 0 between 2012 and 2016 (Fig. 3). There were small deviations in both directions, showing that the synthetic yield spread matches the Dutch yield spread well on average. At the moment of treatment (2016Q1), the gap is less than 0.2% points. Interestingly, CR no ADDING up resembles the Dutch yield spread best at the moment of treatment. In the post-treatment period, the gap is slightly negative or close to zero. Thus, using these three models, there is no evidence that the introduction of the target range raised the Dutch yield spread.

The results from the DID model differ substantially from SC and CR (Table 7).¹⁰ The DID results suggest that introduction of the target range for the Netherlands

¹⁰ Germany is excluded from the DID model as it has a zero yield spread by default. This is in line with SC and CR.

Table 7 Introduction of target range significantly raises the Dutch yield spread in DID-models

	(1)	(2)	(3)	(4)
Yield spread vs. Germany				
Constant	1.315*** (0.140)	2.243*** (0.805)	-1.286 (2.233)	0.287 (2.355)
NL	-1.024*** (0.143)	-1.024*** (0.184)	0.026 (0.067)	0.013 (0.138)
Post2016Q1	-0.390** (0.161)	-1.611** (0.822)	-0.177* (0.096)	-1.453*** (0.485)
NL * Post2016Q1	0.291* (0.164)	0.291 (0.205)	0.548*** (0.141)	0.454*** (0.159)
S&P rating			-0.239*** (0.023)	-0.255*** (0.023)
Ln(IP)			0.835 (0.517)	0.779 (0.516)
Government balance			-0.089** (0.036)	-0.030 (0.031)
Government debt ratio			-0.001 (0.002)	-0.001 (0.002)
Time FE	No	Yes	No	Yes
Observations	232	232	232	232
Adj. R-squared	0.089	0.078	0.723	0.799

*, **, and *** refer to significant at the 10%, 5%, and 1% level, respectively. Robust standard errors between brackets. NL is a dummy variable for the Netherlands, and Post2016Q1 is a dummy variable for the post-treatment period. IP refers to industrial production

Table 8 No significant effect in DID when periphery countries are excluded

	(1)	(2)	(3)	(4)
Constant	0.435*** (0.028)	0.582*** (0.110)	-1.506** (0.723)	0.714 (0.696)
NL	-0.143*** (0.039)	-0.143*** (0.031)	-0.059** (0.028)	-0.034 (0.024)
Post2016Q1	-0.089*** (0.031)	-0.286** (0.111)	-0.038 (0.030)	-0.334*** (0.081)
NL * Post2016Q1	-0.010 (0.042)	-0.010 (0.035)	0.144** (0.064)	0.177*** (0.051)
S&P rating			0.012 (0.032)	-0.138*** (0.036)
Ln(IP)			0.296* (0.171)	0.248 (0.157)
Government balance			-0.047*** (0.015)	-0.022* (0.013)
Government debt ratio			0.004*** (0.001)	0.000 (0.001)
Time FE	No	Yes	No	Yes
Observations	145	145	145	145
Adj. R-squared	0.207	0.402	0.487	0.793

*, **, and *** refer to significant at the 10%, 5%, and 1% level, respectively. Robust standard errors between brackets. NL is a dummy variable for the Netherlands, and Post2016Q1 is a dummy variable for the post-treatment period. IP refers to industrial production. The control group includes Austria, Belgium, Finland, and France

Table 9 Treatment effect with DID differs from SC and CR

Outcome variable	Method	Effect size (in percentage points)
Yield	SC	-0.073
Yield	CR	-0.023
Yield spread	SC	-0.086
Yield spread	CR	-0.120
Yield spread	CR & no ADDING UP	-0.021
Yield spread	DID (all countries)	0.454*
Yield spread	DID (AU, BE, FI, FR)	0.177*

SC: synthetic control. CR: constrained regression. DID: difference-in-differences. AU: Austria, BE: Belgium, FI: Finland, FR: France. For DID: *: significant at the 5% level. For SC and CR, the treatment effect is calculated as the average gap between the treatment unit and the synthetic control unit in the post-treatment period

led to a 0.45% points increase in the yield spread (last column).¹¹ The coefficient on the variable Post2016Q1 shows that all countries experienced a decreasing trend in the yield spread on average. However, this decrease was larger for the control group countries than for the Netherlands. Starting from an already low yield spread of 0.29% points in the pre-treatment period, the post-treatment average yield spread declined by around 0.10% points.

The discrepancy between SC and CR, and the DID results can be explained by the constraint CONSTANT WEIGHTS, which applies to the DID model, but not to SC and CR. In fact, it is the main contribution of SC to relax this constraint (D&I). Each country in the control group (Austria, Belgium, Finland, France, Italy, Portugal, Spain) is given the same weight. However, SC and CR show that Italy, Portugal, and Spain are so different from the Netherlands that they never obtain a large weight. Instead, the other countries are closer matches. This is disregarded in DID, leading to different results.

A robustness check is to only exclude Italy, Portugal, and Spain from the sample, as they substantially differ from the Netherlands. The control countries are now Austria, Belgium, Finland, and France. Using this smaller, but more similar control group, the DID model still shows a significantly positive treatment effect on the yield spread (Table 8). However, it is 2.5 times smaller than in the baseline DID model. The model including the macroeconomic variables and time fixed effects is the preferred specification. Thus, even when the control group consists of more similar countries, there is a significantly positive treatment effect, whereas SC and CR showed no treatment effect. The key finding is that DID does not perform well in this case, even with a control group of similar countries.

¹¹ Excluding the log of industrial production and the government balance as control variables, as these variables have a zero v-weight in the synthetic control method, also leads to a significant and positive treatment effect.

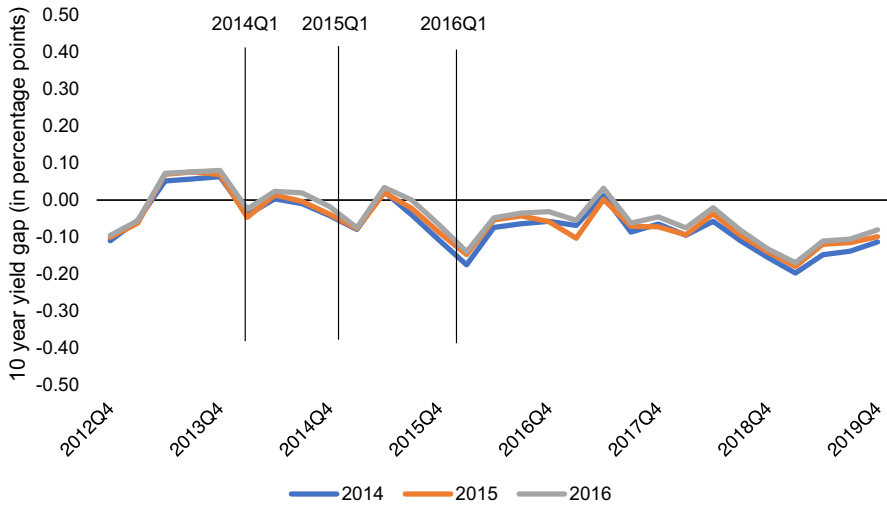


Fig. 4 SC results for yield are robust to time placebo, no treatment effect. Source: Own calculations of synthetic control model

Concluding, DID leads to different results from SC and CR (Table 9). While SC and CR (with or without ADDING UP) do not find a sizeable effect on the yield (spread), DID shows a positive and significant treatment effect of around 0.45% points. Moreover, when the control group only includes countries similar to the Netherlands, the effect size decreases but remains significant. Even with a similar control group, the DID delivers different results than SC and CR. As explained above, the CONSTANT WEIGHTS constraint causes the discrepancy in results. When the number of control units are small, DID leads to erroneous results if the control units are not similar to the treatment unit. In that case, SC or CR is the preferred method.

Placebo test.

SC and CR do not have inference tests; thus it is hard to say something about the statistical significance of the results. For inference, D&I estimate pseudo-treatment effects that are estimated for the control units. As such, the distribution of the pseudo-estimates represents the null distribution of the effect of no treatment, and is compared to the actual treatment effect (Hollingsworth & Wing, 2020).

There are some methods to test the robustness of the results, namely placebo tests (Abadie et al., 2014). Placebo tests can be applied in two different dimensions. One dimension is the time placebo test, where treatment is assigned to a different time period, before the actual treatment took place. The treatment unit does not change with the time placebo test. The second dimension is across control units, where treatment is assigned to different control units, but the time period remains unchanged. In both cases, the treatment effect if present, should disappear compared to the baseline model. If there is a treatment effect observed before actual treatment took place, it is hard to argue that the observed divergence is truly due to the treatment, as the divergence appeared before treatment took place. Similarly, if a treatment effect is

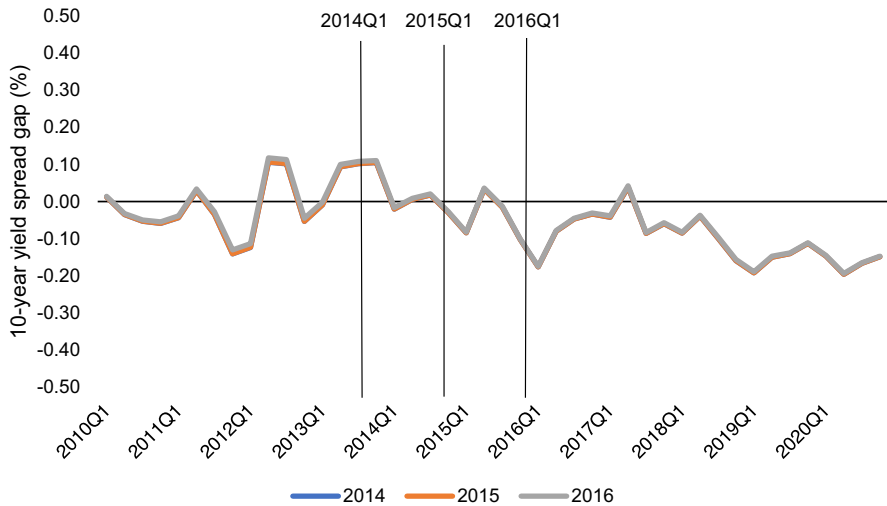


Fig. 5 CR results for yield spread are robust to time placebo, no treatment effect. Source: Own calculations of constrained regression model

observed when treatment is assigned to another unit, then the results are not very reliable. At the moment of treatment, a control unit shows a similar divergence as the treatment unit. This would indicate that treatment is either not limited to the treatment unit, or that there is an underlying trend influencing the results.

4.1 Time placebo tests

First, instead of assigning treatment to 2016Q1, treatment is assigned to 2015Q1 and 2014Q1. We also shift the timing of control variables such that we use the time period for the outcome variable and the control variables. In 2015, and 2014, there was no introduction of a target range so one would not expect to find a significant effect on the yield spread. Figure 4 shows that the SC results for the yield spread are robust to the time placebo, as there is no treatment effect at earlier dates, neither as there is a treatment effect at 2016Q1. In fact, the lines are hardly distinguishable from each other. This also applies to the results for CR and the yield spread, as shown in Fig. 5. After 2014Q1, the gap between the Netherlands and the synthetic yield spread is slightly positive, but quickly decreases again, to continue hovering around zero, with deviations towards both directions. Similarly, after 2015Q1 there is no persistent deviation in the gap between the Dutch yield spread and the synthetic yield spread. The only persistent deviation takes place from 2017Q3 onwards, but it is unlikely that this occurred due to the introduction of the target range 18 months earlier. Additionally, we would expect a positive treatment gap if it were due to the target range. Concluding, all lines follow a similar pattern, showing that the CR results are robust to the time placebo.

Thus, assigning the treatment to a time period before the actual treatment took place does not result in any treatment effects. This suggests that the effects that

Table 10 DID results do not pass time placebo test

	(1)	(2)	(3)
Constant	0.287 (2.355)	-0.461 (2.353)	-0.175 (2.287)
NL	0.013 (0.138)	-0.183 (0.147)	-0.478*** (0.173)
Post2016Q1	-1.453*** (0.485)		
NL * Post2016Q1	0.454*** (0.159)		
Post2015Q1		-1.481*** (0.483)	
NL * Post2015Q1		0.678*** (0.174)	
Post2014Q1			-1.504*** (0.482)
NL * Post2014Q1			0.915*** (0.207)
S&P rating	-0.255*** (0.023)	-0.256*** (0.023)	-0.254*** (0.22)
Ln(IP)	0.779 (0.516)	0.944* (0.519)	0.884* (0.507)
Government balance	-0.030 (0.031)	-0.033 (0.030)	-0.033 (0.029)
Government debt ratio	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Time FE	Yes	Yes	Yes
Observations	232	232	232
Adj. R-squared	0.799	0.804	0.806

*, **, and *** refer to significant at the 10%, 5%, and 1% level, respectively. Robust standard errors between brackets. NL is a dummy variable for the Netherlands, and Post2016Q1/Post2015Q1/Post2014Q1 is a dummy variable for the (placebo) post-treatment period. IP refers to industrial production

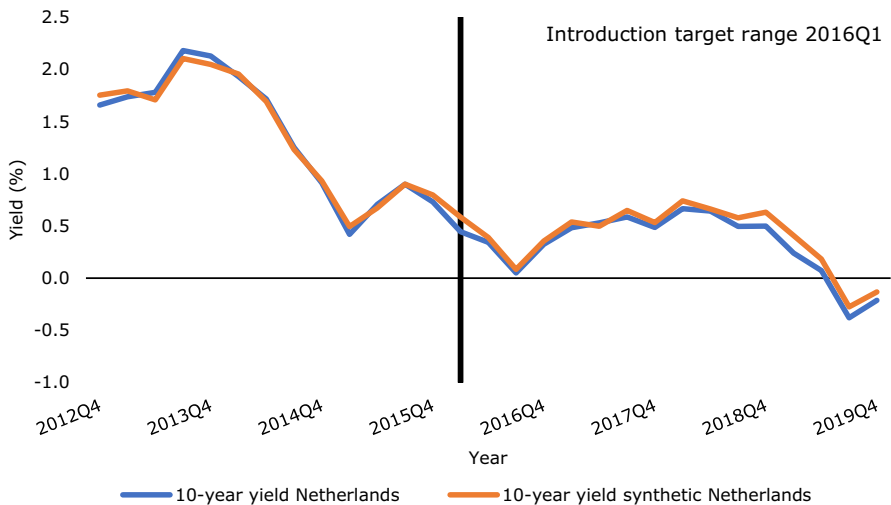


Fig. 6 Yield Netherlands closely matched by synthetic control group, no treatment effect. Source: Own calculations of synthetic control model

occur during the treatment period arise only because of the actual treatment (Fig. 6, 7, 8, 9, 10 and 11).

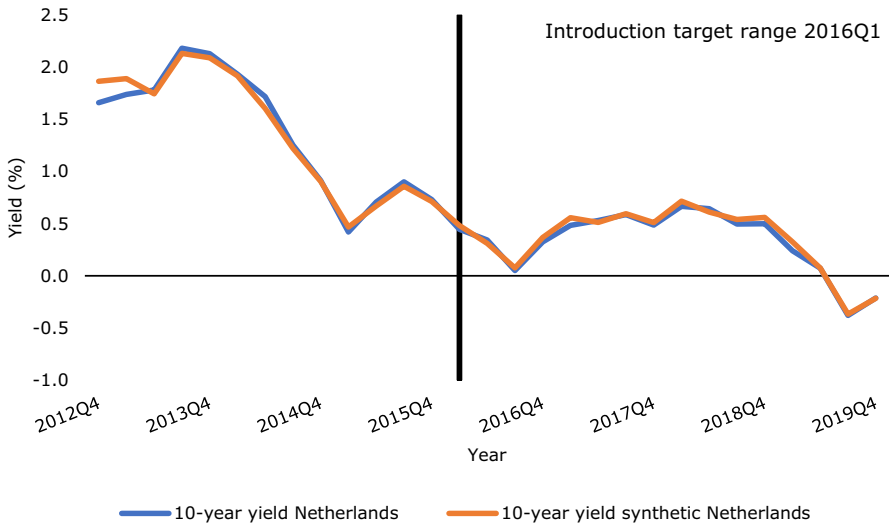


Fig. 7 Dutch yield matched well with synthetic control group under CR, no treatment effect. Source: Own calculations of constrained regression

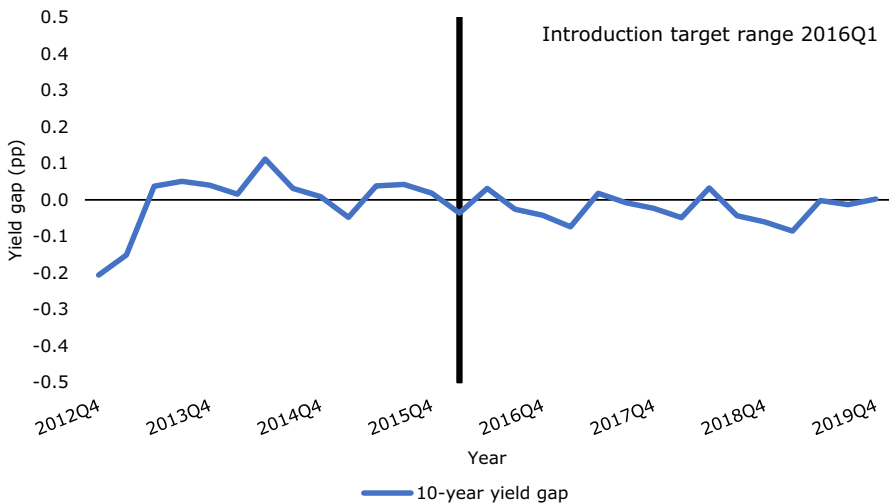


Fig. 8 Constrained regression for yield shows no significant treatment effect. Source: Own calculations of constrained regression model

The same time placebo tests are applied to DID. Again, one would expect insignificant results as for the previous placebo tests. However, the results in Table 10 show that the estimated treatment effect for the Netherlands is even larger when treatment is assigned to 2015Q1 (column 2) or 2014Q1 (column 3). And the results are also statistically significant at the 1% level. This shows that

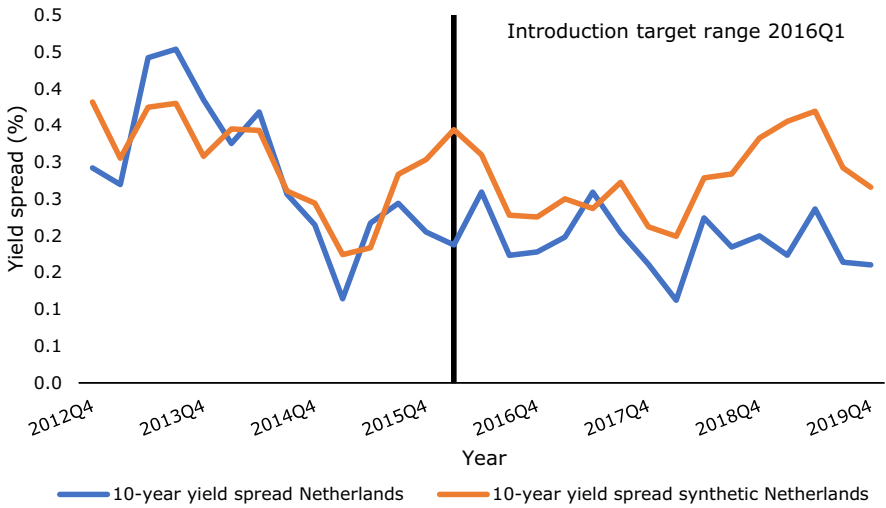


Fig. 9 Synthetic control group has a looser fit for the yield spread than for the yield. Source: Own calculations of synthetic control model

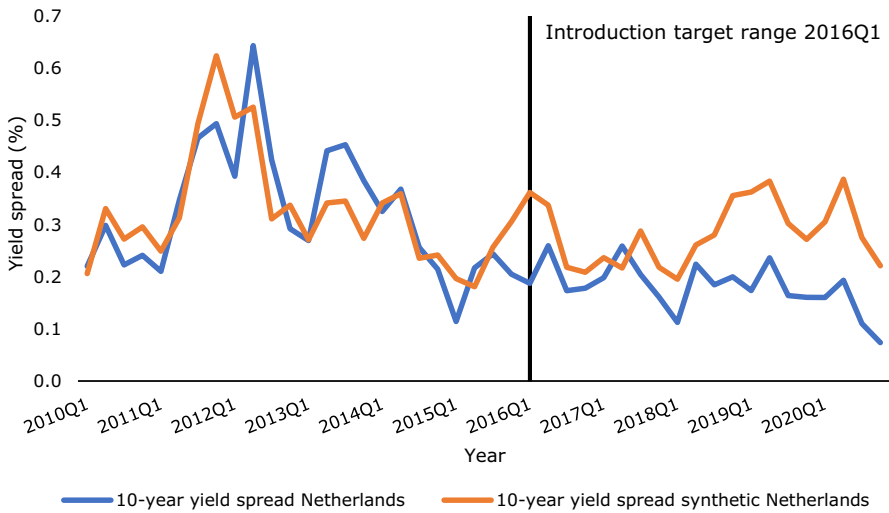


Fig. 10 Yield spread is matched relatively well with constrained regression, but no treatment effect. Source: Own calculations of constrained regression model

the DID results do not pass the time placebo tests. Even if the treatment period is assigned differently, DID shows a positive and significant treatment effect.

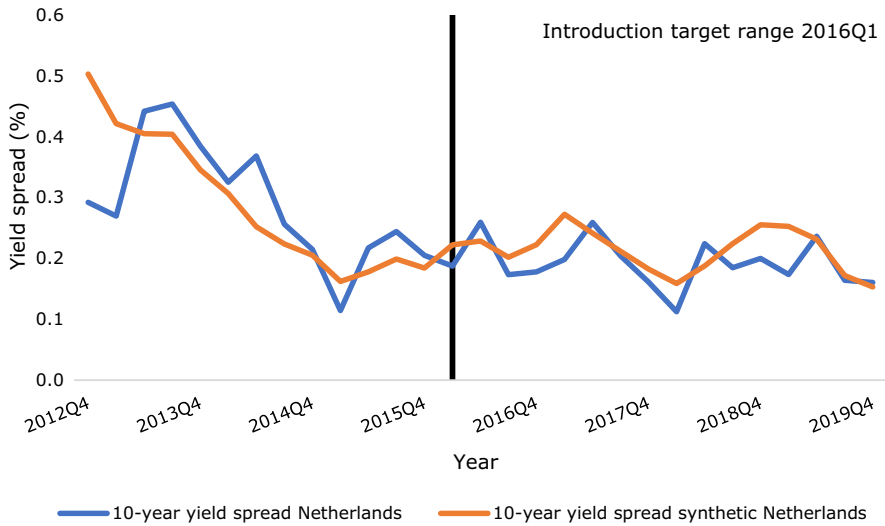


Fig. 11 Constrained regression and relaxing ADDING up leads to a close match of the Dutch yield spread. Source: Own calculations of constrained regression model, with relaxing ADDING UP constraint

Table 11 SC & CR results pass unit placebo test

Country	SC & yield		CR & yield spread	
	Post RMSPE/Pre RMSPE	Mean post-treatment gap (%)	Post RMSPE/Pre RMSPE	Mean post-treatment gap (%)
Netherlands	1.50	-0.07	1.75	-0.09
Austria	1.60	0.03	1.04	-0.03
Belgium	2.01	-0.12	1.07	-0.06
Finland	1.98	0.10	1.42	0.08
France	1.48	0.00	1.83	0.00
Germany	1.01	-0.27		
Italy	5.77	0.66	5.64	0.65
Portugal	0.90	0.95	0.90	0.95
Spain	2.56	-0.55	1.74	-0.28

SC: synthetic control. CR: constrained regression. Pre/Post RMPSE: root mean squared prediction error in pre/post-treatment period

Table 12 DID results do not fully pass unit placebo test

Country	Effect size (%)
Netherlands	0.454*** (0.159)
Austria	0.130 (0.155)
Belgium	0.060 (0.116)
Finland	0.127 (0.143)
France	0.110 (0.125)
Italy	0.118 (0.220)
Portugal	-0.740*** (0.366)
Spain	-0.018 (0.156)

*, **, and *** refer to significant at the 10%, 5%, and 1% level, respectively. Robust standard errors between brackets. Coefficients are the difference-in-differences estimates in models with macroeconomic control variables and time fixed effects

4.2 Unit placebo tests

The SC & CR results pass the unit placebo test (Table 11). The ratio between the post and pre RMSPE shows the size of the post-treatment gap relative to the pre-treatment gap. The pre-treatment gap is the inaccuracy of the matching procedure. A large post-treatment gap relative to the pre-treatment gap gives an indication of the size of the post-treatment gap. A large post-treatment gap itself does not give such information, as it can be the result of a generally weak fit. The RMSPE ratio is fifth highest for the Netherlands. Other countries have a higher RMSPE ratio, such as Italy, Spain, Belgium, and Finland. Germany and Portugal have a ratio close to or below 1, which means that the post-treatment gap is smaller than the pre-treatment gap. There is thus no treatment effect. Looking solely at this table, one might expect the treatment to have taken place in Italy, as Italy has an RMSPE ratio of 5.8.¹² As Italy receives a minimal weight in SC, it does not influence the results. Finland, which is generally similar to the Netherlands in terms of the yield (spread), also has a higher RMSPE ratio for the yield. The mean post-treatment yield gap is 0.10% points for Finland, compared to -0.07 for the Netherlands. In general, it seems that the SC method passes the unit placebo test.

The CR method for the yield spread also passes the unit placebo test, as Italy, which has a high RMPSE ratio, again has a small weight and does not influence the results. Next to Italy and France, the RMSPE ratio is highest for the Netherlands, though close to the Finnish ratio as well. With CR, the mean post-treatment gap for the Netherlands is -0.09% points. This shows that in the post-treatment period, the Dutch yield spread hardly differs from the synthetic yield spread on average. There are also no sizeable placebo treatment effects for other countries.

¹² In their study on the German reunification, Abadie et al. (2014) report a ratio of 16 for West-Germany, and a ratio between 1 and 8 for the control units. Compared to that study, the ratios here are thus small.

The DID results do not fully pass the unit placebo test. Table 12 below shows the estimated treatment effects, using the DID model with macroeconomic control variables and time fixed effects. For most countries, there is no significant effect, which indicates passing the placebo test. However, Portugal does not pass the placebo test. This shows that the estimated DID treatment effect on the Netherlands is not due to the introduction of the target range, but due to a generally declining trend, which is not fully captured in DID.

5 Conclusion

This paper applies different policy evaluation methods to the DSTA case study. As such, the implications of our findings are twofold.

First, the results yield practical applications for policy making. It addresses the question whether or not the switch to a target range for capital market issuances has led to an increase in the yield spread on Dutch sovereign debt. Our results suggest that the switch to a target range did not lead to a higher risk premium. Specifically, we find that introducing a target range for capital market issuances by the DSTA did not increase the risk premium on Dutch government debt, which would have become visible by an increased yield spread. The DSTA can afford some flexibility in the funding policy without a significant risk of higher funding costs.

Second, our findings have methodological applications. By applying different methods to the DSTA case study, this paper adds to the literature by testing relative suitability of various methods within a funding policy context.

The synthetic control method and constrained regression show that introducing more flexibility in the funding policy framework by introducing a target range for capital market issuances does not lead to a higher cost of debt. The risk premium on Dutch government debt did not change after the introduction of the target range in 2016. A standard difference-in-differences method shows other results, namely that the change in the funding policy did raise the risk premium.

The difference in results derives from the fact that the synthetic control method, the constrained regression and difference-in-differences method choose control unit weights differently. The difference-in-differences method gives all control units an equal weight, whereas the synthetic control method (and constrained regression) assigns weights such that the control group closely matches the treatment group in the pre-treatment period. Control units that do not match the treatment group well, receive a small or zero weight under the synthetic control method. Under the difference-in-differences method, these units receive a weight equal to $1/n$, thus they weigh more heavily on the treatment effect.

The conclusion is that the synthetic control method yields more reliable results than the difference-in-differences method when treatment happened at one country at a specific point in time, the number of control units is small, and the range of time series data is sufficiently large. Thus, the synthetic control method, or variations on the synthetic control method, deserves a place in the Dutch policy evaluation toolbox.

For the synthetic control method and constrained regression, results are as well insignificant when treatment is excluded across time or across units in placebo tests. In order to test for robustness, a time placebo test estimates the effects for the period in which treatment did not yet take place. Second, a unit placebo test estimates the effects for units that are in fact not subjected to treatment. In contrast to the synthetic control method and constrained regression, the difference-in-difference method still yields significant results, even when actual treatment has been taken out of the equation.

Furthermore, the synthetic control method is a flexible method, as it can be adapted in several ways to suit the structure of the data. The constrained regression only includes the y-variable as predictor but yields similar results as the synthetic control method. When the treatment unit has the minimum or maximum value in the sample, the ADDING UP constraint can easily be relaxed, to ensure the synthetic control method or constrained regression fits the data well.

Further research could apply the synthetic control method to evaluate the effects of changes in debt funding policy for different case studies. This would shed light on the stability of the results presented in this paper, as they rest on a sole case study. The synthetic control method is also suitable for other policy evaluations in the area of financial markets and finance. For example, the Dutch government introduced a withholding tax for dividends and royalties in 2021. In the research preceding the implementation, Hers et al. (2018) suggested that this tax could be evaluated with a difference-in-differences model, but the synthetic control method is also applicable to this case study.

Appendix

See Tables 13 and 14.

Table 13 Macroeconomic control variables

Country	S&P credit rating (numerical)	Industrial production	Government debt ratio	Government balance
<i>Controls pre-treatment (2012Q4–2015Q4)</i>				
Netherlands	10.38	97.14	67.24	-2.80
Austria	10.00**	108.56***	83.28***	-2.06***
Belgium	9.00***	102.98***	108.45***	-3.20
Finland	10.61	95.62	58.37***	-2.58
France	9.31***	99.52	94.60***	-4.11***
Germany	11.00***	107.43***	76.52***	-0.38***
Italy	2.85***	91.49***	134.47***	-2.87
Portugal	0.15***	95.12	132.45***	-5.88***
Spain	2.61***	91.50***	97.23***	-7.03***
All except Netherlands	6.94***	99.03	98.17***	-3.42**
All except Netherlands, Italy, Portugal, Spain	9.98**	102.82***	84.25***	-2.31*
<i>Controls post-treatment (2016Q1–2019Q4)</i>				
Netherlands	11.00	95.70	56.11	0.74
Austria	10.00***	120.28***	77.94***	-0.53***
Belgium	9.00***	111.96***	104.04***	-1.50***
Finland	10.00***	102.20***	61.18***	-1.19***
France	9.00***	102.53***	98.84***	-3.05***
Germany	11.00	112.01***	64.80***	1.43**
Italy	2.56***	94.61	136.08***	-2.23***
Portugal	1.88***	95.70***	126.79***	-1.70***
Spain	4.63***	98.15**	98.86***	-3.39***
All except Netherlands	7.26***	105.40***	96.07***	-1.52***
All except Netherlands, Italy, Portugal, Spain	9.80***	109.80***	81.36***	-0.97***

Own calculations, based on data from S&P for credit ratings, IMF for industrial production, and European Commission the government debt ratio and government balance. S&P credit ratings are transformed into numerical values, where 11 equals AAA, 10 equals AA, 9 equals A, etc. *, **, ***, ****; mean control variable differs from Dutch mean at 10, 5, and 1% level, respectively

Table 14 Cross-country comparison yields

Country	Pre-treatment	Post-treatment	Change post-treatment (%)
	2012Q4–2015Q4	2016Q1–2019Q4	2016Q1–2019Q4 versus 2012Q4–2015Q4
	10-year yield (in percentage points)		
Netherlands	1.39	0.33	–76%
Austria	1.45	0.43	–70%
Belgium	1.70	0.55*	–68%
Finland	1.37	0.41	–70%
France	1.61	0.41*	–75%
Germany	1.10	0.14*	–87%
Italy	3.11***	2.04***	–34%
Portugal	4.45***	2.21***	–50%
Spain	3.20***	1.26***	–61%
All except Netherlands	2.25***	0.95***	–58%
All except Netherlands, Italy, Portugal, Spain	1.44	0.41	–72%

Own calculations, based on OECD data. *, **, ***: mean yield (2012Q4–2015Q4) differs from Dutch mean at 10, 5, and 1% level, respectively

Acknowledgements The authors thank Bas ter Weel, Michiel Bijlsma, the editor Marno Verbeek and two anonymous referees for their useful comments on earlier versions of this paper.

Funding This research was funded by SEO Amsterdam Economics.

Data Availability Data available from: Eurostat, IMF, OECD, and S&P.

Code Availability Available upon request.

Declarations

Conflict of interest No conflicts of interest.

References

- Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425.
- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Abadie, A., Diamond, A., & Hainmueller, J. (2014). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495–510.
- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1), 112–132.
- Afonso, A., Arghyrou, M., & Kontonikas, A. (2015). *The determinants of sovereign bond yield spreads in the EMU*. ECB Working Paper Series No. 1781, 1–39.
- Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3–32.

- Barone, G., & Moncetti, S. (2014). Natural disasters, growth and institutions: A tale of two earthquakes. *Journal of Urban Economics*, 84(1), 52–66.
- Campos, N. F., Coricelli, F., & Moretti, L. (2014). Economic growth and political integration: Estimating the benefits from membership in the European Union using the synthetic counterfactuals method. *IZA Discussion Papers, No., 8162*, 1–37.
- Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Industrial and Labor Relations Review*, 43(2), 245–257.
- Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772–793.
- Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic natural disasters and economic growth. *The Review of Economics and Statistics*, 95(5), 1549–1561.
- Coffman, M., & Noy, I. (2011). Hurricane Iniki: Measuring the long-term economic impact of a natural disaster using synthetic control. *Environment and Development Economics*, 17(2), 187–205.
- Doudchenko, N., & Imbens, G. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *Working Paper*, pp. 1–38.
- DSTA. (2015). *Outlook 2016*. Dutch State Treasury Agency.
- Eren, O., & Ozbekli, S. (2017). What do Right-to-Work laws do? Evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, 35(1), 173–194.
- Falkenhall, B., Mansson, J., & Tano, S. (2020). Impact of VAT reform on Swedish restaurants: A synthetic control group approach. *Scandinavian Journal of Economics*, 122(2), 824–850.
- Gautier, P. A., Siegmann, A., & Van Vuuren, A. (2009). Terrorism and attitudes towards minorities: The effect of the Theo van Gogh murder on house prices in Amsterdam. *Journal of Urban Economics*, 65(2), 113–126.
- Heersink, B., Peterson, B. D., & Jenkins, J. (2017). Disasters and elections: Estimating the net effect of damage and relief in historical perspective. *Political Analysis*, 25(2), 260–268.
- Hers, J. F. P., Witteman, J. P., & Rougoor, W. (2018). *Balance sheets, income and expenditure of special financial institutions (SFIs)*. SEO-report 2018–86. SEO Amsterdam Economics.
- Hers, J. F. P., Beetsma, R. M. W. J., Witteman, J. P., & Verheuel, N. H. (2019). *Assessment of DSTA's 2016–2019 risk framework and funding policy*. SEO-report 2019–20. SEO Amsterdam Economics.
- Hollingsworth, A., & Wing, C. (2020). Tactics for design and inference in synthetic control studies: An applied example using high-dimensional data. Available at SSRN 3592088.
- Opatrny, M. (2020). The impact of the Brexit vote on UK financial markets: a synthetic control method approach. *Empirica*.
- Peri, G., & Yasenov, V. (2019). The labor market effects of a refugee wave: Synthetic control method meets the Mariel boatlift. *Journal of Human Resources*, 54(2), 267–309.
- Powell, D. (2017). Synthetic control estimation beyond case studies: Does the minimum wage reduce employment? *RAND Working Paper 1142*, (pp. 1–44).
- Ratcliffe, A., Scholder, V. H. K., & S. (2013). The London bombings and racial prejudice: Evidence from the housing and labor market. *Economic Inquiry*, 53(1), 276–293.
- Saunders, J., Lundberg, R., Braga, A. A., Ridgeway, G., & Miles, J. (2015). A synthetic control approach to evaluating place-based crime interventions. *Journal of Quantitative Criminology*, 31(3), 413–434.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.