WP/20/274

IMF Working Paper

Individual Treatment Effects of Budget Balance Rules

by Francesca Caselli, Daniel Stoehlker and Philippe Wingender

INTERNATIONAL MONETARY FUND

 WP/20/274

IMF Working Paper

Research Department

Individual Treatment Effects of Budget Balance Rules

Prepared by Francesca Caselli, Daniel Stoehlker and Philippe Wingender[1]

Authorized for distribution by Malhar Nabar

December 2020

Abstract

This paper investigates the heterogenous effects of budget balance rules on fiscal policy in a large sample of countries. To derive country-specific treatment effects of fiscal rules and conduct inference, we use a Synthetic Difference-in-Differences Method. Our results indicate that countries with a budget balance rule improve their fiscal balance on average by around 3 percent after its introduction. However, our results also illustrate the importance of going beyond the average treatment effect, as it masks significant heterogeneity in the country-specific impact of the rule. We find that countries that would have had large deficits in the absence of the fiscal rule exhibit positive treatment effects, thus reducing their budget deficits. On the other hand, countries with budget surpluses respond to fiscal rules by reducing their budget surplus and moving closer to the numerical target of the rule. Our results also suggest that rules' design matters: a small overall number of fiscal rules, and the presence of a monitoring process outside the government, especially at the supra-national level, improve significantly the effectiveness of the rules.

JEL Classification Numbers: E61, E62, H62, C33

Keywords: fiscal rules, fiscal policy, heterogeneous treatment effects, synthetic control, difference-in-differences
Author's E-Mail Address: fcaselli@imf.org, stoehlker@ifo.de, pwingender@imf.org.

[1] We wish to thank Tohid Atashbar, John Bluedorn, Raphael Espinoza, and Mehdi Raissi for helpful comments and suggestions. We are also grateful to participants of the IMF Research Department seminar for helpful discussions and suggestions. All errors remain our own.

I. INTRODUCTION

Fiscal rules have increasingly been used to strengthen fiscal discipline and constrain budget deficits: between 1985 and 2015, a total of 89 countries introduced a budget balance rule.[2] Despite their rapid diffusion, the empirical evidence on the effectiveness of fiscal rules remains mixed, mainly because of the challenges of establishing a causal link to fiscal outcomes.[3] Earlier empirical results suggest that countries that introduce a rule tend to have lower budget deficits (Poterba 1997, Debrun and others 2008, Badinger and Reuter 2017), however, studies that correctly account for rules' endogeneity tend to produce insignificant results (Heineman and others 2018, Caselli and Reynaud 2020). While most studies focus on the average effect of fiscal rule adoption, some recent papers have investigated the heterogeneous impacts of fiscal rules on budget balances. Reuter (2015) and Lledó and Reuter (2018) estimate how policymakers react to past deviations from rule thresholds. They find that rules-based frameworks prevent the fiscal variables from permanently deviating from the thresholds imposed by the rule. Caselli and Wingender (2018), by constructing counterfactual distributions of government balances, uncover a "magnet" effect of the 3 percent fiscal ceiling in the European Union: the adoption of the rule reduces the occurrence of both large government deficits and surpluses. This implies not only that the rule had an effect on prudent countries, but also that the rule had an effect on deficits even when it was not complied with.[4]

However, the effect of a policy on the distribution of outcomes is not the same as the distribution of the effects (Heckman, Smith and Clements 1997, Abbring and Heckman 2007, Bedoya et al. 2017). While subtle, the difference between the impact on the distribution and the distribution of the impacts is important. Estimating the distribution of treatment effects allows one to uncover further layers of heterogeneity by computing, for instance, the variance of treatment effects or what proportion of countries were 'helped' or 'hurt' by the rule. Also, it can shed light on how individual countries responded to the introduction of the rule. However, estimating individual treatment effects is challenging because it requires estimating a counterfactual for every unit. In this paper we overcome this challenge by adopting a Synthetic Difference in Differences (SDD) approach to estimate individual treatment effects of budget balance rules in a large sample of advanced, emerging economies and low-income countries. The SDD approach recently developed by Arkhangelsky and others (2019) is a generalization of the widely used Synthetic Control Method (SCM) (Abadie and Gardeazabal 2003). Importantly, it allows for inference which is lacking in most SCM studies.

The use of SDD allows us to obtain country-specific counterfactuals and hence to go one step further than conventional treatment effects approach and analyze the key determinants of rules' effectiveness. While previous studies have mostly relied on anecdotal evidence (for examples see e.g. Primo, 2007, and IMF, 2018), we exploit the cross-country variation in the timing and design of budget balance rules and in the magnitude (and sign) of the treatment effects to quantify the impact of particular features of the rule in a systematic way.

Our findings point to a sizable heterogeneity in the size of the treatment effect. Figure 1 reports the main result of our analysis: it plots the individual effect of introducing a budget balance rule on countries' government balance. Countries that would have had large deficits in the absence of the fiscal rule exhibit positive and statistically significant treatment effects, thus reducing their budget deficits. On the other hand, countries with

---

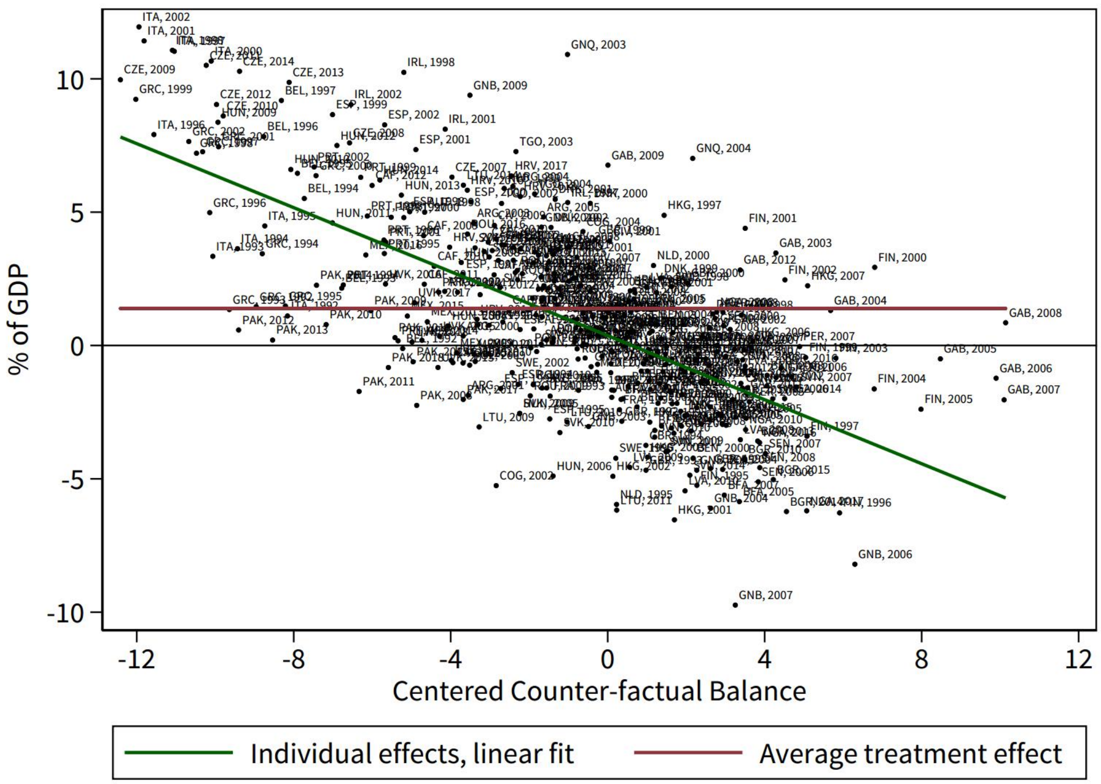
[2] IMF Fiscal Rules Database, accessible at: https://www.imf.org/external/datamapper/fiscalrules/map/map.htm.

[3] For example, countries with fiscal rules could have a stronger preference for fiscal prudence and discipline, even in the absence of the rule. Moreover, the adoption of such a rule could reflect the overall fiscal situation or can be part of a general fiscal overhaul. In both cases, countries with already smaller levels of public deficits would also be more likely to adopt a fiscal rule, thus rendering the causal interpretation of the impact of the fiscal rule on deficits invalid.

[4] Under a rank invariance assumption, Caselli and Wingender (2018) also derive country-specific effects and find that all countries have seen their fiscal position improve on average because of the deficit rule.

budget surpluses also respond to fiscal rules by reducing their budget surplus and moving closer to the numerical target of the rule. This is in line with previous findings that the fiscal rule has exerted a ‘magnet effect’ on budget balances by pulling deficits and surpluses closer to the center of the distribution (Caselli and Wingender 2018). Overall, the average treatment effect is significant, positive and equal to around 3 percent of GDP after 10 years. It is also important to note that the average effect (illustrated by the horizontal line in the figure) fails to capture the full heterogeneity in the individual treatment effects. While it largely underestimates the effect for countries with large deficits, it vastly overstates (and predicts the wrong sign of) the treatment effect for countries with small deficits and surpluses as indicated by the difference between the average and linear fit.

Figure 1: Individual Treatment Effects by Level of Counterfactual Government Balance



Note: The figure depicts the treatment effect for all countries with fiscal rules for the 10-year post-treatment period obtained from the baseline SDD estimation. The counterfactual balance on the horizontal axis is centered around the numerical value of the fiscal rule in each country.

Finally, when we link the individual treatment effects with the rules’ other characteristics, our results suggest that the design of fiscal rules is indeed critical for their effectiveness (see IMF 2018 and Caselli and Reynaud 2020). Specifically, we find that maintaining a small overall number of fiscal rules is key. We also see that a monitoring process outside the government, especially at the supra-national level, improves the effect of the budget balance rule significantly while the existence of an escape clause has a diminishing effect.

The roadmap of the paper is as follows: sections 2 and 3 introduce the SDD, the specification of our baseline model and the data used. Section 4 presents the baseline individual treatment effect results and the average impact of the fiscal rule. Section 5 discusses the quality of our model fit and its robustness to alternative specifications. In section 6 we analyze the role of specific features of rules' design and their effectiveness. Section 7 concludes.

II. SYNTHETIC DIFFERENCE IN DIFFERENCES AND MODEL SPECIFICATION

This section introduces the notation used throughout the paper and briefly reviews the SDD estimation. The SDD is a two-step data-driven procedure to build counterfactual outcomes for observations that are subject to a treatment, e.g. the introduction of a fiscal rule. It is particularly well suited when units of observation are at an aggregate level such as countries for which it can be difficult to find a suitable counterfactual that resembles the treated country.[5] In a first step, synthetic control weights are estimated using the usual SCM algorithm. These weights are then used in a second step to recover the individual treatment effect in a standard difference in differences estimation.

A. Estimating synthetic control weights

In a set of countries $i = 1, \dots, N$ over T periods, country 1 is the only country that receives the treatment in period $s < T$. For a given value of the fiscal rule indicator $FR_i \in \{0,1\}$ for whether the country is ever treated and values of the government budget balance $Y_{i,t}$, we define potential outcomes $Y_{i,t}(FR_i)$ as follows:

$$Y_{i,t}(FR_i) = \begin{cases} Y_{i,t}(0) & \text{if } FR_i = 0 \\ Y_{i,t}(1) & \text{if } FR_i = 1 \end{cases}, \qquad \forall\, t \geq s$$

The size of the treatment effect on the outcome of interest would be straightforward to infer by comparing actual outcomes and their counterfactuals, i.e. $\tau_{1,t} = Y_{1,t}(1) - Y_{1,t}(0)\ \forall\, t \geq s$, where $Y_{1,t}(0)$ denotes the counterfactual for country 1 had it not adopted the fiscal rule. However, we cannot observe a country in both states simultaneously: while $Y_{i,t}(1)$ is observable for those countries that have adopted a fiscal rule, $Y_{1,t}(0)$ is not directly observable in the data.

The SCM allows to build a counterfactual for each treated observation, i.e. the outcome of interest in the absence of the treatment. Concretely, it finds the weighted average of all potential comparison units which is 'closest' to the treated unit in terms of pre-treatment outcomes.

To do this, we use the following nested minimization problem. We wish to match in the pre-treatment period both:

1) The weighted average of K predictors in the control group X_0 to the weighted average of those predictors for the treated unit X_1; and
2) The time path of the (weighted) average outcome among control units Y_0 to the outcome for the treated unit Y_1.

[5] As argued in Abadie and others (2010), with aggregate data and in an observational context it is more promising to use a weighted combination of several untreated units instead of a single untreated unit as a counterfactual. Doudchenko and Imbens (2016) note that the SCM can be viewed as a generalization of the standard difference-in-differences approach, with the weights on the control units chosen to better match the pre-treatment trend of the unit that is exposed to the treatment.

To match the treated country to the weighted average of the control units along the two dimensions, we need two sets of weights. A first set of country-level weights $W = (\omega_2, \dots, \omega_N)$ is used to build weighted averages of the K predictors and the outcome variable among control units. A second set of predictor-level weights denoted by the diagonal matrix V is used to linearly combine the K predictors. Under the usual identifying assumptions, any difference between the treated unit and this synthetic control is due to the treatment itself, since both units have the same (weighted) average values for predictors and outcome in the pre-treatment period.

Formally, objective 1) is expressed as

$$\widehat{W}(V) = \arg\min_{W \in W} (X_1 - X_0 W)' V (X_1 - X_0 W) \tag{1}$$

with country-level weights $W = (\omega_2, \dots, \omega_N)$, $\omega_i \geq 0$ and $\sum_{i=2}^{N} \omega_i = 1$ and taken as given the predictor-level weight matrix V, a diagonal positive semidefinite matrix of dimension $K \times K$ with trace equal to one.

For objective 2), we wish to find a linear combination of the K predictors that yields, through Eq. (1), the set of country-level weights $\widehat{W}(V)$ that in turn produces the smallest distance between the synthetic control and the treated unit. This can be expressed as

$$\widehat{V} = \arg\min_{V \in V} \left(Y_1 - Y_0 \widehat{W}(V)\right)' \left(Y_1 - Y_0 \widehat{W}(V)\right) \tag{2}$$

noting that $\widehat{W}(V)$, the solution to Eq. (1), implicitly depends on the choice of matrix V.

The data-driven approach just described is highly flexible and allows for a very broad definition of X_0. To see why, consider for example that the set of predictors X_0 could be defined as the pre-treatment country-level average of k predictors. Alternatively, X_0 could also be defined as the set of annual observations of the same k predictors, so that the total number of predictors used in Eq. (1) would instead be $K = s \times k$. Given some choice of X_0, Eq. (2) would then select the matrix V that combines all predictors K to best match pre-treatment outcomes across the two groups. Note the matrix V does not have an economic interpretation but serves only to achieve the best pre-treatment match possible between treated unit and synthetic control.

Without economic theory to guide the choice of the weighting matrix V, there is a risk researchers will use specification searching to cherry-pick the set of K predictors that yields the desired outcome for the subsequent treatment effect estimation. This ‘p-hacking’ would affect the size of tests for statistical significance of treatment effects as shown by Ferman, Pinto and Possebom, 2020. Consequently, the authors recommend using all the pre-treatment outcome lags as predictors. This also simplifies the algorithm above to

$$\widehat{W} = \arg\min_{W \in W} (Y_1 - Y_0 W)' (Y_1 - Y_0 W) \tag{3}$$

where we minimize the distance between outcomes for treated and synthetic control by searching directly for the country-level weights W.

After estimating country-level weights $\widehat{W}$ that produce the synthetic control with the closest pre-treatment match, we can construct post-treatment counterfactual outcomes for the treated unit using a weighted average of the control units

$$\widehat{Y}_{1,t}(0) = \sum_{i=2}^{N} \widehat{\omega}_i Y_{i,t}, \quad \forall\, t \geq s.$$

We estimate the treatment effect at each point in time using

$$\hat{\tau}_{1,t} = Y_{1,t} - \sum_{i=2}^{N} \widehat{\omega}_i \, Y_{i,t}, \quad \forall\, t \geq s. \tag{4}$$

If the number of pre-intervention periods in the data is sufficiently large, matching on pre-intervention outcomes can allow us to control for the heterogenous responses to multiple unobserved factors. The intuition here is that only units that are alike along both observables and unobservables would follow a similar trajectory pre-treatment (Cunningham 2019).

As a baseline, we choose the predictor-level weight matrix V such that pre-treatment outcomes are matched on the lags of the government balance with no additional controls (see Doudchenko and Imbens 2016; Ferman, Pinto and Possebom 2020).[6] In a set of robustness checks, we use instead other covariates in the nested SCM algorithm outlined above. We allow the pre-treatment period for calibrating the donor weights to be as long as data availability allows but not longer than 10 years.[7] We also allow for 10 years after the introduction of the fiscal rule in our baseline scenario.[8] Thus, the sample window for every treated country can be at most 21 years in total. The pool of donor countries is restricted to those that have complete budget balance data over the same time horizon as the treated country and did not introduce a fiscal rule themselves within this time horizon. Moreover, as our baseline specification, we only consider donor countries that are in the same or in an adjacent income class group, as classified by the World Bank. For example, high-income countries are matched to other high-income countries or those from the upper middle-income group. Similarly, low-income countries are only paired with other low-income countries and those in the lower middle-income country group.[9]

To account for country-specific and time invariant factors, we de-mean the government balance data by subtracting the pre-treatment average by country. If selection into treatment is only correlated with time-invariant common factors, e.g. a general taste for fiscal prudence, then the de-trended synthetic control estimator is unbiased. Finally, the use of SCM is particularly appealing because of its transparency in constructing counterfactual observations. Compared to standard regression methods, the chosen weights make explicit what each control unit contributes to the counterfactual (see Cunningham 2019 for a discussion).

B. Weighted Difference in Differences

The second step of the SDD approach uses the estimated weights from the previous SCM results in a conventional difference in difference estimation. Specifically, Arkhangelsky and others (2019) propose to estimate the treatment effect in Eq. (4) via weighted least squares regression with time fixed effects of the following form:

---

[6] See also Kaul et al. (2015) for a discussion of covariate selection in SCM.

[7] Given our data on fiscal balances starts in 1980 and the first fiscal rules were introduced in the early 1990s, this is the longest pre-treatment horizon that the data would allow to use.

[8] This leaves enough time to let the effects of the fiscal rule materialize and become visible while being also short enough to minimize the risk that other major country-specific changes could potentially alter the true underlying country weights. It is not uncommon in the literature to use substantially longer post-treatment horizons. For example, Campos and others (2019) analyze the country-specific growth effects of European integration and use the obtained weights for extrapolating counterfactual GDP series for as much as 30 years in some cases, however very.

[9] We also report results without this restriction on the set of potential donor countries in the section on robustness checks.

$$\left(\hat{\mu}, \hat{\beta}, \hat{\tau}_{1,t}\right) = \arg\min_{\mu,\beta,\tau_{1,t}} \sum_{i=1}^{N} \sum_{j=-10}^{10} \left(Y_{i,t} - \mu - \beta_t - FR_i \tau_{1,t}\right)^2 \widehat{\omega}_i, \quad (5)$$

where β_t are year fixed effects and FR_i is an indicator variable that is equal to one in the case of the treated country. Importantly, the authors show that under standard assumptions, the estimation uncertainty of the SCM weights $\widehat{\omega}_i$ can be ignored. This simplifies inference considerably as the weights obtained in the first step can be used in estimating Eq. (5) without further adjustment. It is also straightforward to stack observations across treatment groups to recover moments of the treatment effects distribution via pooled regression.

III. Data

We investigate the country-specific impact of budget balance rules (BBRs) on the nominal government balance in a sample of 151 countries between 1980 and 2018. The sample combines data from the IMF World Economic Outlook database with historic fiscal balance series from Mauro and others (2015) and the IMF Government Finance Statistics.[10] Among the 151 countries in our sample, 53 countries have adopted a budget balance rule with varying numerical targets.[11] Appendix Table A.1 lists the rule adopters, year of adoption, the numerical target of the rule and its coverage (general or central government). Most countries have adopted fiscal rules restricting the deficit to levels not larger than 3 percent or 0 percent of GDP.

Table 1 presents the summary statistics for the fiscal balance for countries that have adopted a fiscal rule ("rulers") and those without fiscal rule ("non-rulers") between 1980 and 2018. Not surprisingly, deficits are larger on average for country-year observations without fiscal rules compared to the situation with fiscal rule in place. The table also shows that rulers and non-rulers are different with respect to other macroeconomic and institutional characteristics. This is particularly true for income per capita, inflation and interest rates as well as the old-age dependency ratio. A simple comparison of the average deficits of rulers and non-rulers is therefore likely to give a biased estimate of the causal effect of fiscal rules on deficits.

IV. Individual Treatment Effects

Figure 1 presents the main result of our analysis by plotting the individual effect of introducing a budget balance rule on countries' government balance. As mentioned in the introduction, we find that while the rule has improved the balance of high-deficit countries, it has reduced the balance among those with smaller deficits or surpluses. Thus, the introduction of the fiscal rule has moved individual observations towards the value of the target, in line with Caselli and Wingender (2018). This suggests that simply computing an average effect would overstate the effect on high-balance countries while it would understate the effect for high-deficit countries.

To formally look at the average treatment effect, we average all the individual treatment effects across all countries together with the corresponding 95 percent confidence interval obtained from Eq. (5). Figure 2 shows that the size of the treatment effect in the post-treatment period increases gradually in the first six years after the introduction of the rule before reaching a long-run effect of around 2 percent of GDP. This suggests that the introduction of the fiscal rule led to an economically and statistically significant reduction in the fiscal deficit for the average country.

---

[10] We restrict the set of countries to those that have at least one million inhabitants in 2018.

[11] We exclude six countries (Cameroon, Chad, Estonia, Indonesia, Mali and Singapore) due to limitations in data availability in the early 1990s. We also excluded India from our analysis as it had a budget balance rule in place only for one year.

Figure 3 plots the time paths of the 10th and the 90th percentiles of the actual and the counterfactual distribution before and after the introduction of the fiscal rule. The actual budget deficit of countries at the lower end of the budget balance distribution is considerably smaller in magnitude than in the case without fiscal rules in place. As already mentioned, these countries have reduced their deficits substantially.

V. QUALITY OF THE SYNTHETIC WEIGHTS S AND ROBUSTNESS EXERCISES

In this section we assess the quality of our baseline model and test alternative specifications. First, we want to assess the quality of the pre-treatment fit as it is the key aspect for the validity of the SC estimates. However, there is a lack of consensus on what constitutes a sufficiently good pre-treatment fit of the synthetic counterfactual when compared to the actual outcome. We follow Adhikari and others (2018) who propose to use the ratio of the Root Mean Squared Prediction Error (RMSPE) and a benchmark RMSE for each treated country. The RMSPE is defined as:

$$RMSPE = \sqrt{\frac{1}{10}\sum_{t=-9}^{0}\left(Y_{1,t} - \sum_{i=2}^{N}\widehat{\omega}_i Y_{i,t}\right)^2} \tag{6}$$

and the benchmark RMSPE as:

$$Benchmark\ RMSPE = \sqrt{\frac{1}{10}\sum_{t=-9}^{0}\left(Y_{1,t}\right)^2} \tag{7}$$

The index takes values greater or equal than 0, where smaller values indicating a better fit. Adhikari and others (2018) suggest that if the index exceeds the value of 1, the counterfactual fails to describe the actual path of the outcome variable of the treated country reasonably well and should be discarded. Figure 4 shows the distribution of the indices which are well below 1, suggesting a sufficiently high quality of the fit.[12]

Second, as already noted, another key feature of the SCM is to make the weights assigned to the treated unit explicit. Figure 5 displays the weights of donor countries for every treated country in our sample. Treated countries are ranked according to their GDP per capita in 2018 in descending order on the vertical axis, and the corresponding donor countries for each country on the horizontal axis, also ranked according to GDP per capita. Each row contains the corresponding weights for every treated country, with larger weights displayed in darker red. The figure confirms that treated units tend to be matched with countries at a similar level of income. In our sample, positive weights are assigned to more than 40 countries from the pool of donor countries on average, but it can be as high as 60 in some cases. This comes with the advantage that potential country-specific shocks in the group of donor countries or cross-country spillover effects to geographically neighboring countries are negligible if single country-specific weights are small.

Third, we implement a series of robustness checks to assess the stability of our baseline results to a different set of donor countries and the use of covariates instead of lagged outcomes in the SCM algorithm. We also assess whether the results are influenced by restricting the pool of potential donors to countries only in the same

[12] As an alternative assessment of the quality of the pre-treatment fit, we perform a formal F-test for whether all pre-treatment differences are jointly statistically different from zero based on Equation (5). In line with the assessment based on the index proposed by Adhikari and others (2018), the corresponding test statistic suggests that we cannot reject the hypothesis that all pre-treatment differences are jointly different from zero.

or adjacent income level. Figure 6 shows the distribution of the Adhikari and others (2018) index without donor pool restriction. This means that all countries, for which complete balance data is available, can be potentially used for the construction of the synthetic counterfactual independently of their level of economic development. Figure 6 shows that the shape of the distribution including those countries for which the pre-treatment fit is worst within the sample is only marginally changed as compared to the baseline scenario when conditioning on income levels (Figure 4). This suggests that conditioning on income classes is for most cases a non-binding constraint in terms of the quality of the pre-treatment fit.

Fourth, we test the robustness of our main results by using a set of relevant macro-economic predictors of the fiscal balance instead of lagged dependent variables. Specifically, we condition on the debt-to-GDP ratio, GDP per capita, the inflation rate (all from the IMF World Economic Outlook database), a commodity terms of trade index (Gruss and Kabhaj 2019) and whether there is currently an IMF program in place. Donor weights are estimated based on the pre-treatment means of all regressors. The results for the country-year-specific treatment effects, the time paths of the 10^{th} and the 90^{th} percentiles as well as the average treatment effect are displayed in Figures 7 to 9. They confirm the magnet effect at the individual country-level and the positive average treatment effect of the rule. However, we also find it is challenging to construct similar synthetic controls in terms of the pre-treatment trends based on covariates as often series are not completely available for the full set of countries.

Finally, we test the robustness of the baseline specification to alternative pre- and post-treatment windows. We restrict the length of the pre- and post-treatment to 5 years only. The result on the distribution of individual treatment effects is illustrated in Figure 10 showing a very similar pattern to Figure 1.

VI. Effectiveness of Fiscal Rules

The effective design of fiscal rules has been a longstanding question in the literature. A range of studies exploiting either cross-country or cross-state variation with respect to the 'strength' of fiscal rules document that stronger rules are generally better in restraining deficits (see e.g. Poterba, 1994, for the US and Badinger and Reuter, 2017, for the EU; Caselli and Reynaud, 2020). In many cases, these studies rely on a composite index of strength that aggregates several features, such as the coverage of the rule, the degree of independence of monitoring and enforcement bodies, existence of correction mechanisms and sanctions, etc. Based on a small set of case studies for countries that have adopted fiscal rules in recent years, IMF (2018) compares the actual outcome of public debt and expenditures with their counterfactual paths in absence of the fiscal rule over a 10-year time horizon as derived from a SCM estimation. It then relates improvements with respect to those indicators (or the lack thereof) to some specific features of the rule, based on anecdotical evidence.

The granular information on country-specific treatment effects together with information on the specific design of fiscal rules in these countries allows us to test in a systematic way what makes fiscal rules more or less effective at constraining short-term fiscal policy. Special attention is devoted to some of the hypotheses that have been put forward in IMF (2018): first, an increasing number and complexity of fiscal rules in place increases the potential of conflict between the rules and the probability of non-compliance. Second, the existence of monitoring and enforcement mechanisms can incentivize compliance with the rules, e.g. through raising public awareness.

The information on fiscal rule design is summarized in Table 2.[13] On average, countries in our sample have 3 rules in place, including rules at the national and the supra-national level. The most frequent combination is to have budget balance rules, many of them at the supra-national level, together with restrictions on public debt. We note that if a country has adopted a supra-national budget balance rule, it always includes a monitoring and an enforcement mechanism. For BBRs at the national level, this is only true in half of the cases. Differences exist with respect to the existence of escape clauses. It appears that this feature is included only in a minority of rules adopted.

Table 3 shows the estimation results for the number and types of rules in place. Specifically, we estimate the following regression equation:

$$\hat{\tau}_{i,t} = \alpha + \underbrace{\sum_{k=2}^{7} \beta_k \cdot \mathbb{1}_{k,i,t}}_{\text{\# of rules}} + \underbrace{\sum_{l \in \{\text{exp, rev,debt}\}}^{K} \gamma_l \cdot \mathbb{1}_{l,i,t}}_{\text{other rules in place}} + \underbrace{\sum_{m \in \{sup,nat+sup\}} \delta_m \cdot \mathbb{1}_{m,i,t}}_{\text{level of deficit rule}} + \varepsilon_{i,t}.$$

The indicator variables $\mathbb{1}_{k,i,t}$ capture different aspects of the design of fiscal rules. The first set of dummy variables refers to the total number of rules in place. In this case, the indicator variable is equal to 1 if country i has a total amount of k rules (of any sort) in place at time t. In general, there can be a total amount of 8 different rules (rules on the deficit, debt, revenue and expenditure – both at the national and the supra-national level). The largest number of rules that we observe in our sample is a total of 7. We are interested in the coefficient β_k related to each of the 6 potential outcomes (having only 1 rule in place is taken as the benchmark outcome). Column 1 shows that a smaller set of rules positively correlates with a positive effect of the rule on budget balances and the effect is statistically significant. For example, fiscal deficits in countries with two rules are lower by almost 4 percent of GDP. Column 2 adds to the discussion of how the existence of other, potentially conflicting rules, can have an impact on the effect of budget balance rules. We include an indicator variable in the regression that is equal to 1 if country i has, for instance, a debt rule in addition to a fiscal deficit rule in period t. The results do not suggest that the existence of other rules has an impact on the size of the treatment effect of budget balance rules, except for the case of debt rules which appears to have a positive effect. Similarly, also the fact that some countries have supra-national BBRs in addition to rules at the national level does seem to have a positive impact on the treatment effect size (column 3).

Next, we test the impact of rule design and institutional features, namely the existence of a monitoring mechanism, an enforcement procedure, and the possibility to trigger an escape clause. As before, we estimate the following regression equation:

$$\hat{\tau}_{i,t} = \alpha + \underbrace{\sum_{k \in \{nat,sup,nat+sup\}} \beta_k \cdot \mathbb{1}_{k,i,t}}_{\text{monitoring mechanism}} + \underbrace{\sum_{l \in \{nat,sup,nat+sup\}} \gamma_l \cdot \mathbb{1}_{l,i,t}}_{\text{enforcement mechanism}} + \delta_{esc} \cdot \mathbb{1}_{esc,i,t} + \varepsilon_{i,t}.$$

Table 4, Column 1 shows that the improvement of the fiscal balance is particularly strong for countries that have monitoring mechanisms at the supra-national level, as compared to countries that have either national monitoring mechanisms or none. The size of the treatment effect is large in magnitude and statistically highly significant. It would suggest that those countries with monitoring at the supra-national level would reduce their deficit by more than 2 percent of GDP vis-à-vis countries with no monitoring at all. Similarly, countries with

[13] We consider all the post-treatment years for the 46 countries with fiscal rules in place. For some countries, e.g. those that adopted the fiscal rule only after 2010 or that have removed the fiscal rule after a few years again, we have less than 10 years of post-treatment observations.

enforcement mechanisms at the supra-national level also improve their balances after introducing fiscal rules (column 2). The effect is smaller in magnitude, though. Lastly, the existence of an escape clause as part of the fiscal rule has a detrimental effect on the size of the treatment effect. The results indicate that allowing for the possibility to trigger an escape clause reduces the effect of the fiscal rule from around 2percent of GDP without escape clause by 2 percentage points to a zero-effect.

VII. CONCLUSION

This paper investigates the heterogenous effects of budget balance rules on fiscal performances in a large sample of countries. In order to derive country-specific treatment effects of the fiscal rule under weak and reasonable assumptions, we use a Synthetic Difference in Differences approach (Arkhangelsky and others 2019). This allows us to conduct inference on the treatment effects esimates. Our results indicate that countries with a fiscal rule improve their fiscal balance on average by around 3 percent of GDP in the long run after its introduction, thus documenting an overall positive and statistically significant average effect of the rule. However, our results also illustrate the importance of going beyond the average treatment effect, as it masks significant heterogeneity in the country-specific impact of the rule. Consistent with previous findings, we show that the improvement of the fiscal balance is largest for countries with large deficits in the counterfactual case where no rule is adopted, but we also find that countries with budget surpluses are affected. Despite not being binding for them, these countries see their fiscal balance converge to the numerical target of the fiscal rule after its introduction. Taken together, these two results confirm the hypothesis that budget balance rules exert a 'magnet effect' on fiscal balances by pulling deficits and surpluses closer to the center of the distribution. The results are robust to various modifications of the SDD specification, including changes to the length of the pre- and post-treatment period as well as the selection of covariates based on which the SCM weights are fitted.

Our results also add to the discussion of what concrete features determine budget balance rules' effectiveness in constraining fiscal deficits. Based on the country-specific treatment effects from the SDD analysis and significant cross-country variation with respect to the specific design of fiscal rules, our results indicate that features and interactions with the overall rule environment of the country under consideration are both critical determinants of its success. In particular, we demonstrate that a smaller overall number of rules and the existence of a credible and independent monitoring process for rule compliance are the most important contributors to the success of the budget balance rule. On the other hand, the existence of an escape clause appears to have a detrimental effect on the effectiveness of the rule. From a policy perspective, the results highlight the fact that fiscal rules, despite their overall potential to improve fiscal performances, are a complex matter and can trigger unexpected consequences, as in the case of diminishing fiscal surpluses.

References

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

Abbring, J. H. & Heckman, J. J. (2007). Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. Handbook of econometrics, 6:5145–5303.

Adhikari, B., Duval, R., Hu, B., & Loungani, P. (2018). Can reform waves turn the tide? Some case studies using the synthetic Control Method. Open Economies Review, 29(4), 879-910.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2019). Synthetic difference in differences (No. w25532). National Bureau of Economic Research.

Badinger, H., & Reuter, W. H. (2017). The case for fiscal rules. Economic Modelling, 60, 334-343.

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

Caselli, F. & Reynaud, J. (2020). Do Fiscal Rules Cause Better Fiscal Balances? A New Instrumental Variable Strategy. European Journal of Political Economy (63), 2020

Caselli, F. and P. Wingender (2018). Bunching at 3 Percent: The Maastricht Fiscal Criterion and Government Deficits. IMF Working Paper 18/182.

Debrun, X., Moulin, L., Turrini, A., Ayuso-i-Casals, J., & Kumar, M. S. (2008). Tied to the mast? National fiscal rules in the European Union. Economic Policy, 23(54), 298-362.

Doudchenko, N., & Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. National Bureau of Economic Research WP 22791.

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

Gruss, B., & Kebhaj, S. (2019). Commodity Terms of Trade: A New Database. International Monetary Fund WP 19/21.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. The Review of Economic Studies, 64(4):487–535.

Heinemann, F., M. D. Moessinger, and M. Yeter (2018). Do Fiscal Rules Constrain Fiscal Policy? A Meta-Regression Analysis. European Journal of Political Economy 51, 69–92

International Monetary Fund (2018). Second-generation fiscal rules: Balancing simplicity, flexibility, and enforceability. Staff Discussion Note No. 18/04.

Kaul, A., Klößner, S., Pfeifer, G., & Schieler, M. (2015). Synthetic control methods: Never use all pre-intervention outcomes together with covariates.

Lledó, V., Yoon, S., Fang, X., Mbaye, S., & Kim, Y. (2017). Fiscal rules at a glance. IMF Background Note.

Mauro, P., Romeu, R., Binder, A., & Zaman, A. (2015). A modern history of fiscal prudence and profligacy. Journal of Monetary Economics, 76, 55-70.

Primo, D. M. (2007). Rules and Restraint: Government Spending and the Design of Institutions. University of Chicago Press.

Poterba, J. M. (1997). Do Budget Rules Work? Fiscal Policy: Lessons from Economic Research.

Poterba, J. M. (1994). State responses to fiscal crises: The effects of budgetary institutions and politics. Journal of Political Economy 102 (4), 799-821.

Reuter, W. H. (2015). National numerical fiscal rules: Not complied with, but still effective?. European Journal of Political Economy, 39, 67-81.

VIII. TABLES

Table 1: Summary Statistics on Main Economic Variables for Countries with and without Fiscal Rule

	Rulers					Non-Rulers				
	Obs.	Mean	St. Dev.	Min.	Max.	Obs.	Mean	St. Dev.	Min.	Max.
Government Balance (percent of GDP)	808	-2.17	4.00	-15.14	18.55	3,817	-2.60	4.66	-19.8	19.39
Population Size (in million inhabitants)	815	24.72	35.36	0.67	200.96	4,516	34.67	118.75	0.24	1395.38
GDP per capita (in 1,000 USD, PPP adj.)	815	24.67	15.32	0.57	70.03	4,451	13.41	18.06	0.28	172.99
Government Debt (percent of GDP)	811	61.01	36.83	0.05	233.72	4,167	60.79	60.02	0	2092.92
Old Age Dependency Ratio (percent of working age population)	809	18.73	9.36	4.07	35.59	4,943	9.8	5.87	0.80	46.17
Inflation Rate (annual percentage change)	815	2.62	2.84	-3.98	25.87	4,422	260.78	13991	-72.73	929790
Trade openness (sum of imp. & exp., percent of GDP)	812	91.79	58.05	20.72	442.62	4,149	72.77	37.72	0.02	376.22
Short-Term Interest Rate (percent)	691	3.68	3.11	-0.69	19.92	2,641	34.2	827.97	-0.78	41280
Federal State (time-independent dummy)	815	0.19	0.40	0	1	5,035	0.15	0.35	0	1
Participation in IMF Program (time-dependent dummy)	815	0.26	0.44	0	1	5,035	0.34	0.47	0	1
Commodity Terms of Trade Index	802	100.06	6.28	48.56	111.81	4,455	98.03	12.60	31.92	130.96

Source: IMF World Economic Outlook Database (April 2019), IMF International Financial Statistics (IFS), World Bank World Development Indicators Database, Monitoring of Fund Arrangements (MONA Database), and Gruss and Kabhaj, 2019.

Table 2: Summary Statistics on Rule Designs

Outcome Variable	Obs.	Mean.	St. Dev.
Number of national rules	420	0.89	1.03
Number of supra-national rules	420	2.09	1.14
Number of rules, national and supra-national	420	2.97	1.30
Dummy, expenditure rule	420	0.32	0.47
Dummy, revenue rule in place	420	0.21	0.41
Dummy, debt rule in place	420	0.87	0.33
Dummy, BBR at the national level only	420	0.18	0.38
Dummy, BBR at the supra-national level only	420	0.64	0.48
Dummy, BBR at both national and supra-national level	420	0.18	0.39
If BBR at supra-national level exists: Dummy, monitoring outside gov't	346	1.00	0.00
If BBR at national level exists: Dummy, monitoring outside gov't	166	0.42	0.50
Either BBR at national or supra-national: Dummy, monitoring outside gov't	420	0.90	0.30
If BBR at supra-national level exists: Dummy, enforcement mech.	345	1.00	0.00
If BBR at national level exists: Dummy, enforcement mech.	149	0.40	0.49
Either BBR at national or supra-national: Dummy, enforcement mech.	420	0.92	0.27
If BBR at supra-national level exists: Dummy, escape clause	345	0.44	0.50
If BBR at national level exists: Dummy, escape clause	148	0.30	0.46
Either BBR at national or supra-national: Dummy, escape clause	420	0.47	0.50

Source: IMF Fiscal Rules Database, 1985-2015.

Table 3: The Effect of Fiscal Rule Complexity on the Size of the Treatment Effect

	(1)	(2)	(3)
Dummy: 2 numerical rules	4.17*** (0.81)		
Dummy: 3 numerical rules	1.54** (0.65)		
Dummy: 4 numerical rules	2.65*** (0.75)		
Dummy: 5 numerical rules	2.56** (1.19)		
Dummy: 6 numerical rules	1.21 (1.12)		
Dummy: 7 numerical rules	1.13 (1.83)		
Dummy: Expenditure rule		-0.16 (0.62)	
Dummy: Revenue rule		-0.95 (0.97)	
Dummy: Debt rule		2.33*** (0.81)	
Dummy: BBR at the supranational level			2.32** (0.80)
Dummy: BBR at the national × supranational level			0.93 (0.61)
Constant	-1.10*** (0.19)	-0.30 (0.61)	-0.17 (0.41)
Observations	420	420	420

Note: Regression results are based on OLS estimations. The dependent variable is the treatment effect of the fiscal rule as obtained from the baseline SDD analysis. Standard errors are clustered at the country-level and are displayed in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

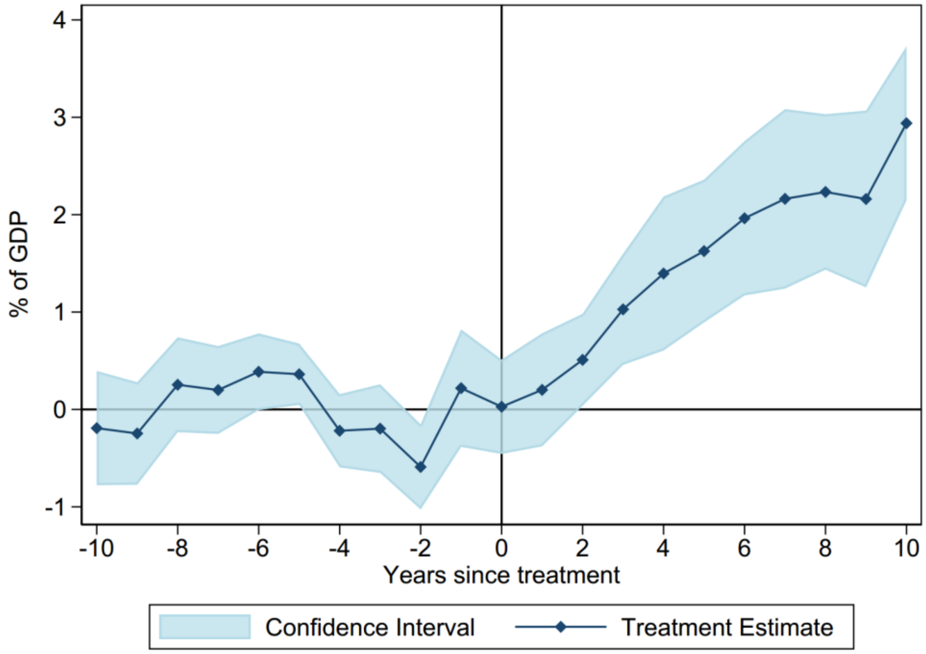
Table 4: The Effect of Fiscal Rules' Design on the Size of the Treatment Effect

	(1)	(2)	(3)
Dummy: Monitoring outside gov't at the national level	1.03 (0.79)		
Dummy: Monitoring outside gov't at the supranational level	2.48*** (0.69)		
Dummy: Monitoring outside gov't at the national × supranational level	2.46*** (0.58)		
Dummy: Formal enforcement at the national level		-0.13 (0.75)	
Dummy: Formal enforcement at the supranational level		1.93*** (0.69)	
Dummy: Formal enforcement at the national × supranational level		1.95*** (0.47)	
Dummy: Existence of escape clause			-2.29** (0.86)
Constant	-0.64** (0.30)	-0.09 (0.35)	2.56*** (0.67)
Observations	420	420	420

Note: Regression results are based on OLS estimations. The dependent variable is the treatment effect of the fiscal rule as obtained from the baseline SDD analysis. Standard errors are clustered at the country-level and are displayed in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

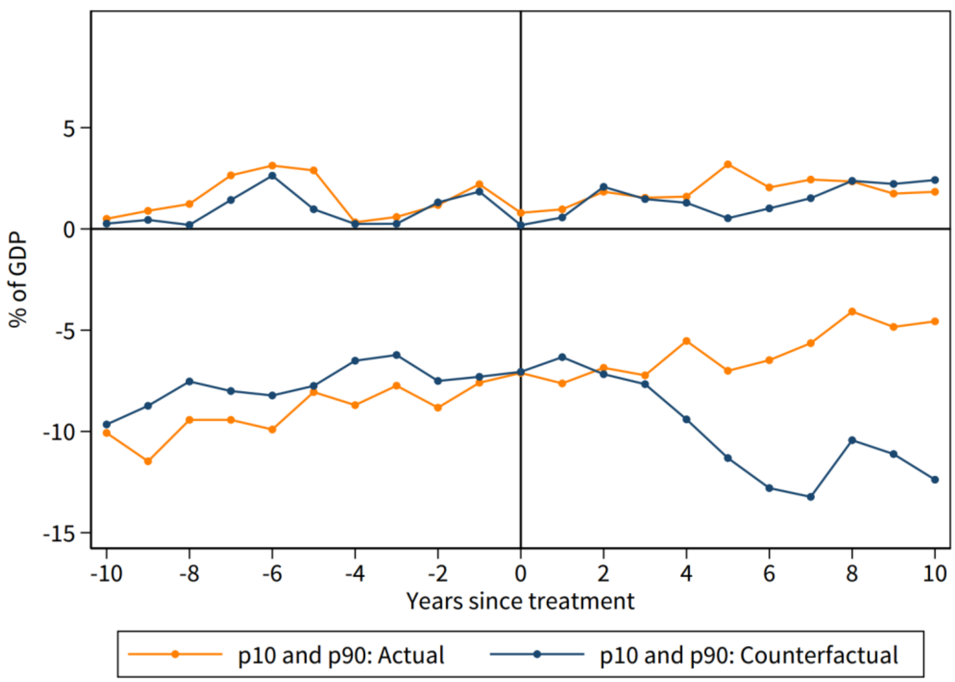
IX. FIGURES

Figure 2: Average Treatment Effect of the Fiscal Rule on Governments' Balances



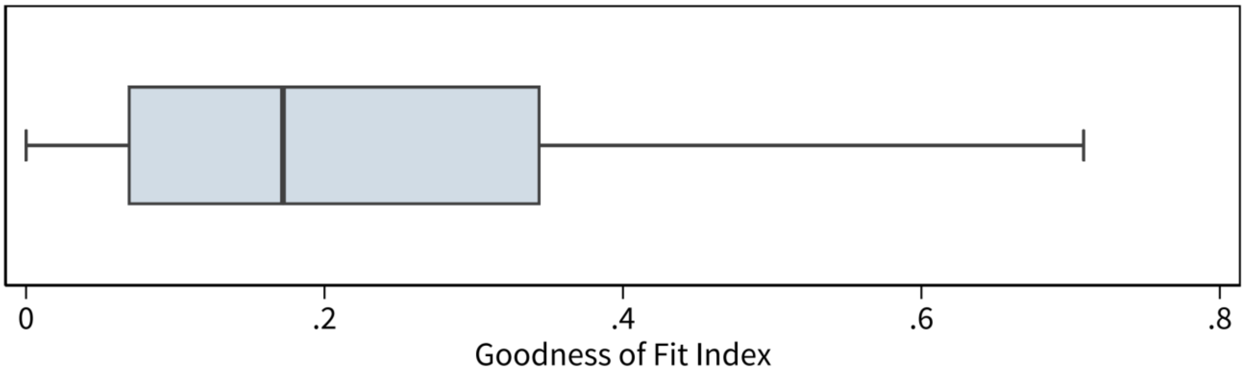
Note: The figure reports the average treatment effect obtained from a linear combination of country-specific individual treatment effects from the baseline SDD specification and Equation 5. The fiscal rule is introduced in year 0.

Figure 3: Time Paths of the 10th and the 90th Percentiles of the Treated and Counterfactual Balances



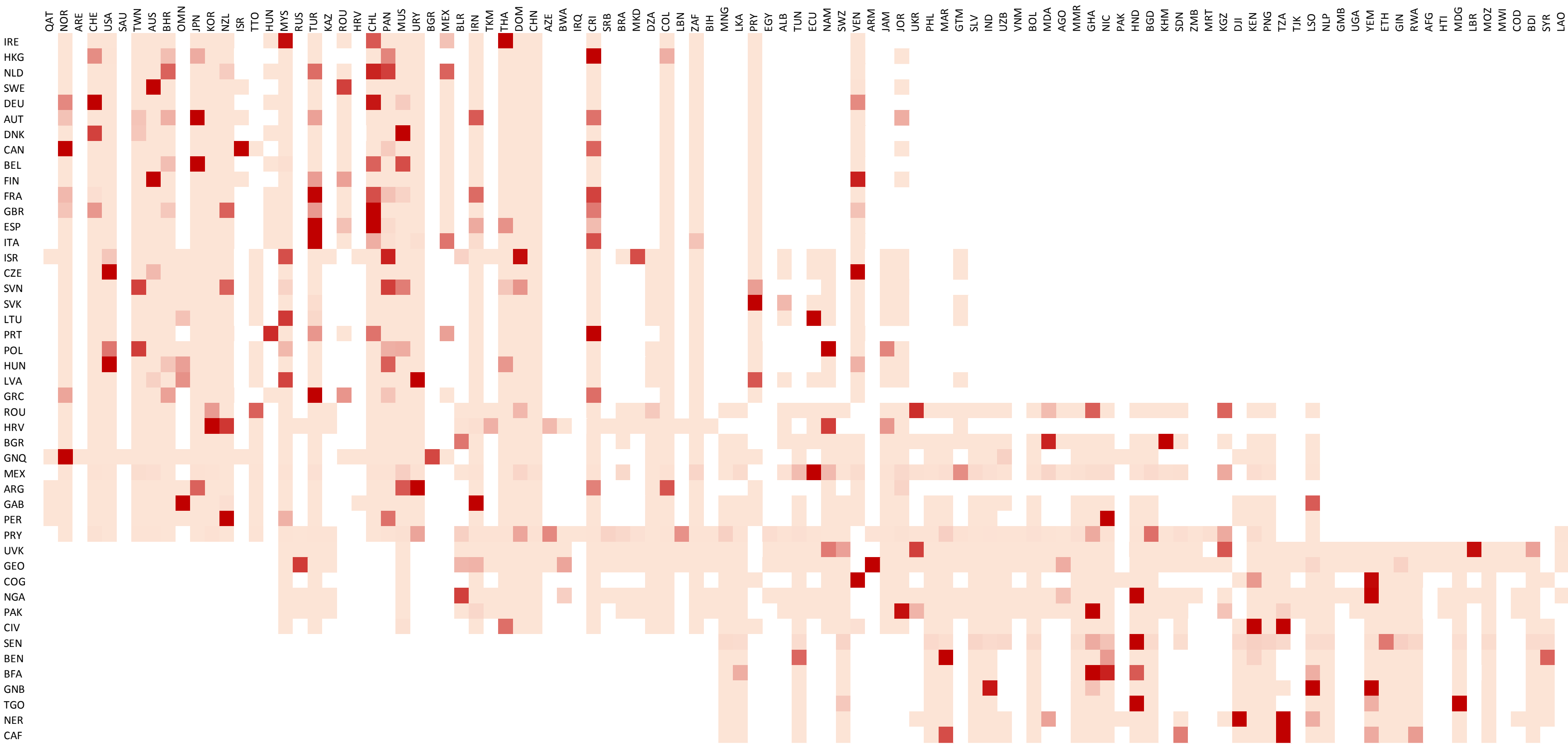
Note: The figure reports the 10th and the 90th percentiles of the actual budget balance series and the counterfactual balance series obtained from the baseline SDD specification. The fiscal rule is introduced in year 0.

Figure 4: Quality of the Pre-Treatment Fit in the Baseline SC Specification



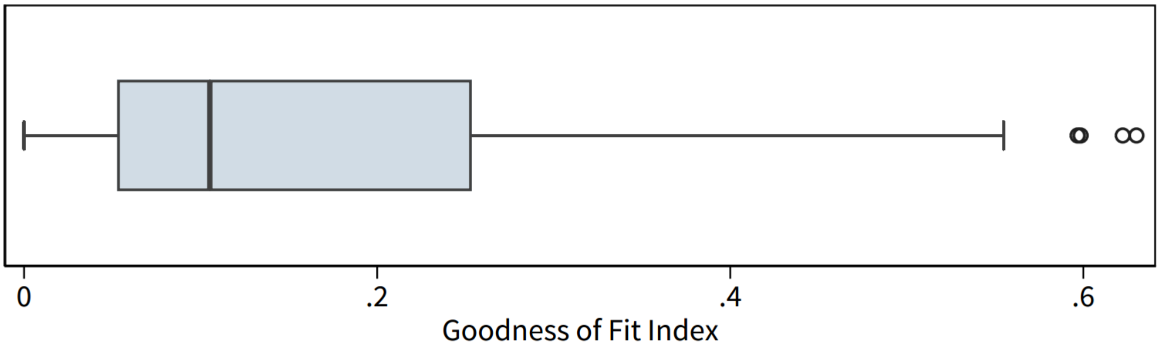
Note: The figure shows the distribution of the Adhikari and others (2018) indices based on Equation (6) and (7). The index takes values greater or equal than 0, with smaller values indicating a better fit.

Figure 5: Matrix of Donor Weights in the Baseline SDD Specification



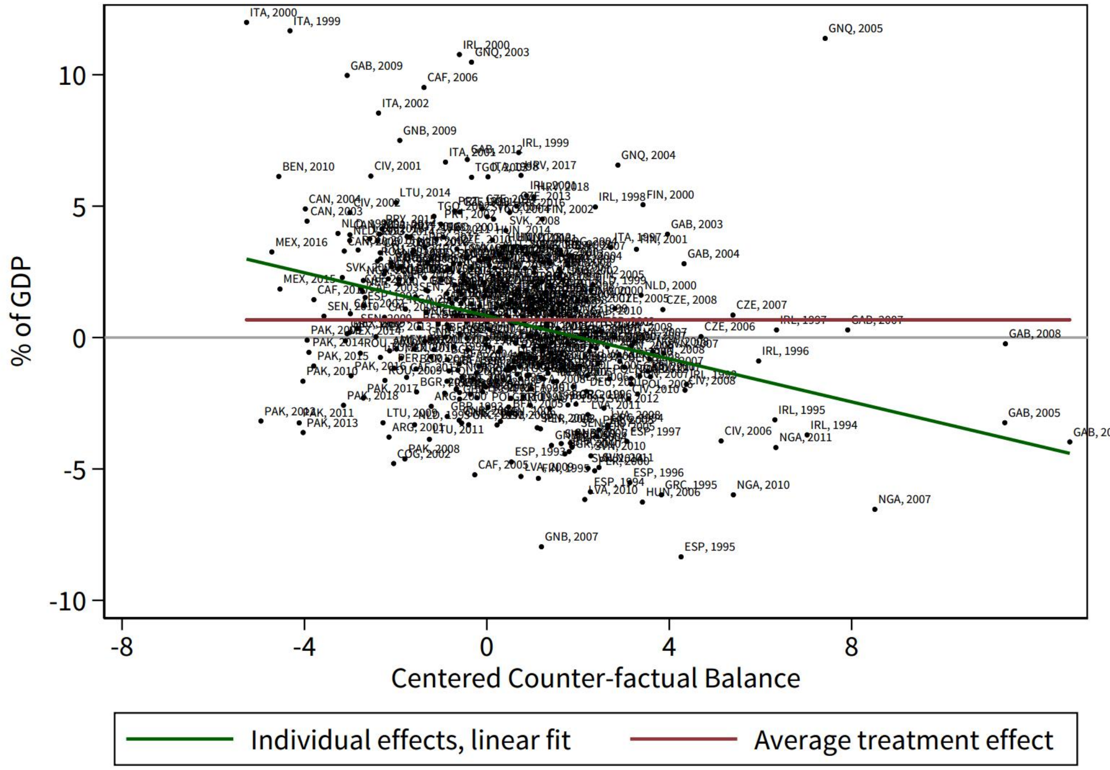
Note: The figure reports the weights from the baseline SDD estimation. Treated countries are ranked according to their GDP per capita in 2018 in descending order on the vertical axis, and the corresponding donor countries for each country on the horizontal axis, also ranked according to GDP per capita. Each row contains the corresponding weights for every treated country, with larger weights displayed in darker red.

Figure 6: Quality of the Pre-Treatment Fit in the Robustness SC Specification without Income Group Conditioning



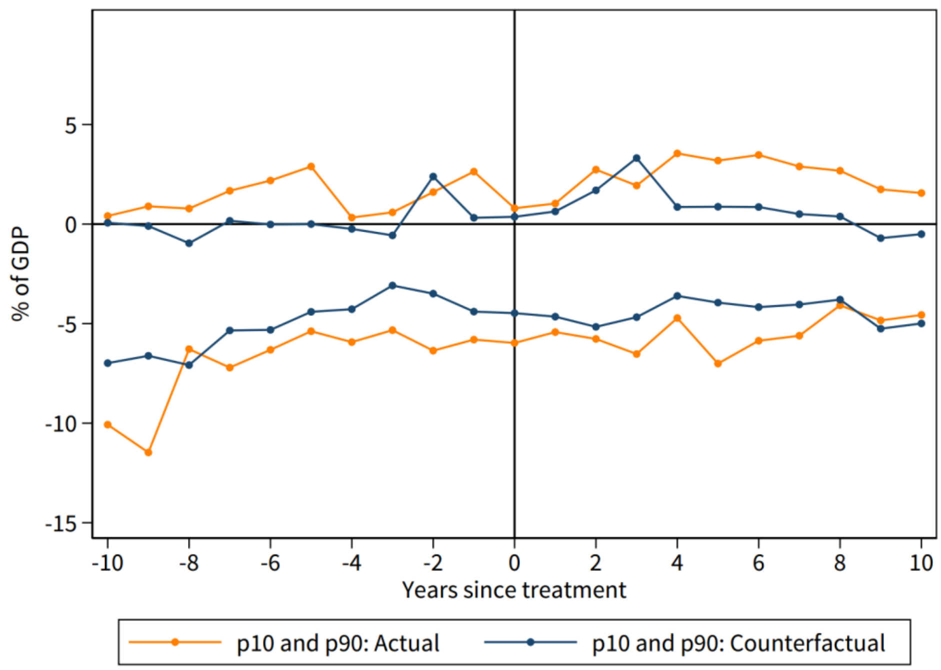
Note: Note: The figure shows the distribution of the Adhikari and others (2018) indices based on Equation (6) and (7) based on a robustness exercise where the set of donor countries is not restricted to countries of the same or adjacent income groups. The index takes values greater or equal than 0, with smaller values indicating a better fit.

Figure 7: Individual Treatment Effects by Level of Counterfactual Government Balance from Estimation based on Covariates



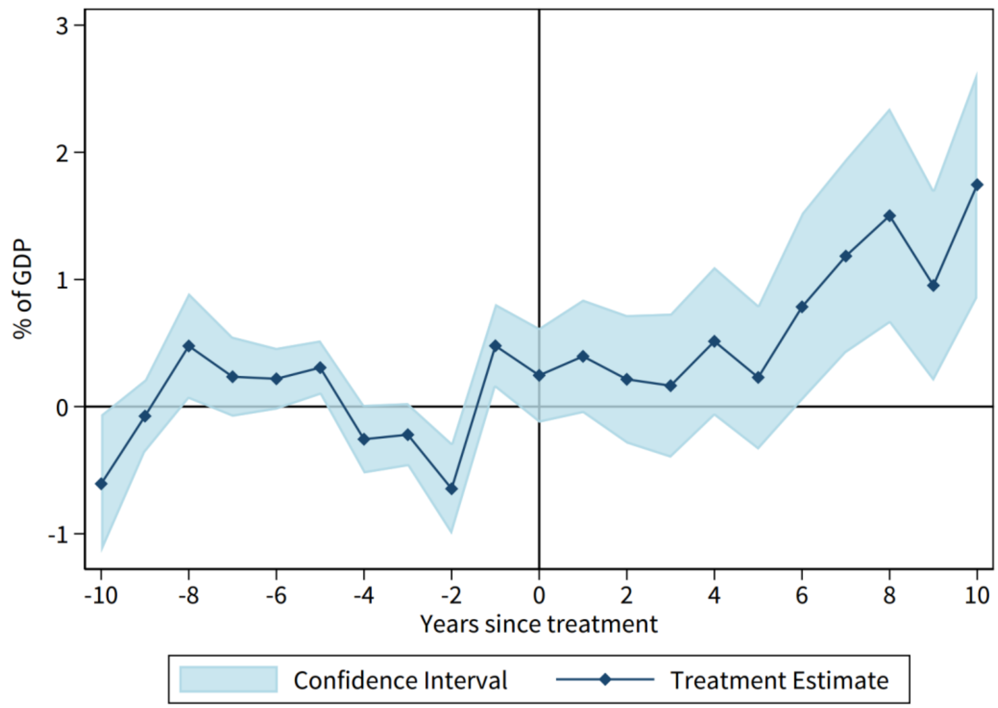
Notes: The figure reports the individual treatment effects obtained from a SDD specification that includes the covariates instead of lagged outcome. The counterfactual balance on the horizontal axis is centered around the numerical value of the fiscal rule in each country.

Figure 8: Time Paths of the 10th and the 90th Percentiles of the Treated and Counterfactual Balances - Estimation based on Covariates



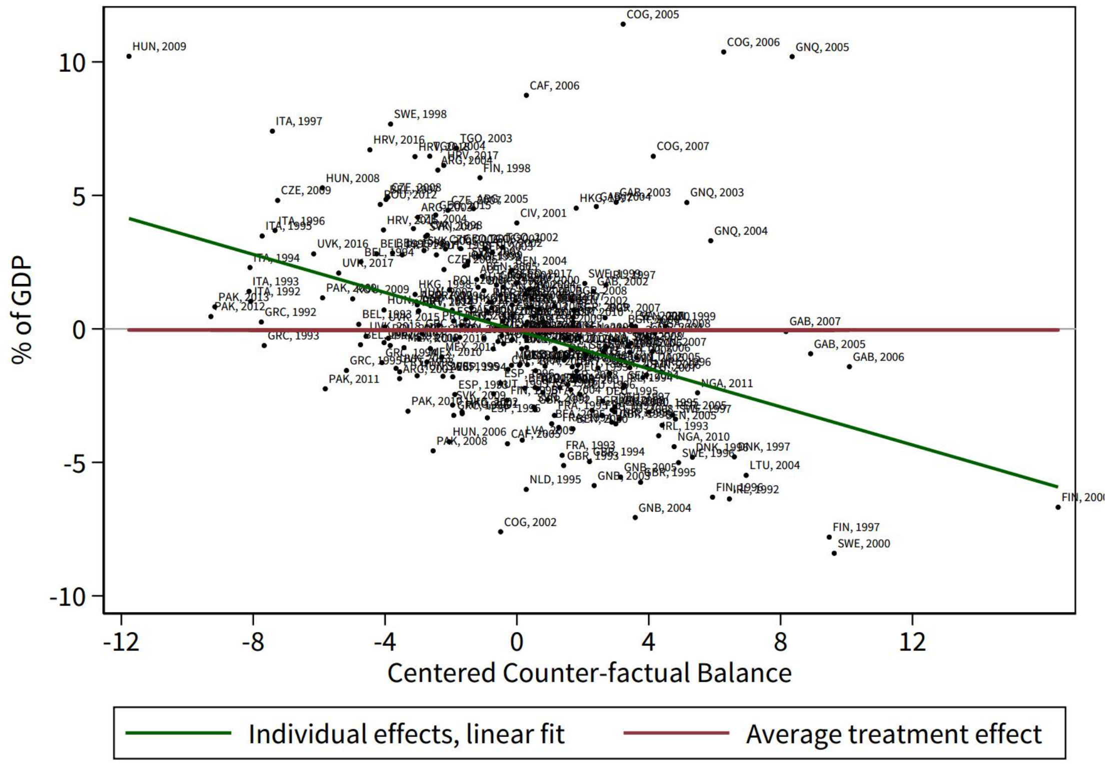
Notes: The figure reports the 10th and the 90th percentiles of the actual budget balance series and the counterfactual balance series obtained from the SDD specification based on covariates. The fiscal rule is introduced in year 0.

Figure 9: Average Treatment Effects of Fiscal Rule Introduction - Estimation based on Covariates



Note: The figure reports the average treatment effect as obtained from a linear combination of country-specific individual treatment effects from the SDD specification based on covariates and Equation 5. The fiscal rule is introduced in year 0.

Figure 10: Individual Treatment Effects by Level of Counterfactual Government Balance with Shortened Pre- & Post-Treatment Horizon



Note: The figure reports the individual treatment effects by level of the counterfactual budget balance obtained from a SDD specification with only 5 years of pre- and post-treatment windows.

Appendix

a) Tables

Table A.1: Countries included in the Sample and their National Fiscal Rules

Country	Year of adoption	Numerical Target (percent of GDP)	Coverage of FR	Included in baseline sample
Antigua and Barbuda	1998	-3	GG	no, < 1 million population in 2018
Argentina	2000	0	GG	Yes
Austria	1995	-3	GG	Yes
Belgium	1992	-3	GG	Yes
Benin	2000	-3	CG	Yes
Bulgaria	2006	-2	GG	Yes
Burkina Faso	2000	-3	CG	Yes
Cabo Verde	1998	-3	CG	no, < 1 million population in 2018
Cameroon	2002	0	CG	no, FR introduced in 2002 but too short pre-treatment period
Canada	1998	0	CG	Yes
Central African Republic	2002	0	CG	Yes
Chad	2002	0	CG	no, FR introduced in 2002 but too short pre-treatment period
Congo, Republic of	2002	0	CG	Yes
Croatia	2013	-3	GG	Yes
Cyprus	2004	-3	GG	no, < 1 million population in 2018
Czech Republic	2004	-3	GG	Yes
Côte d'Ivoire	2000	-3	CG	Yes
Denmark	1992	-3	GG	Yes
Dominica	1998	-3	GG	no, < 1 million population in 2018
Equatorial Guinea	2002	0	CG	Yes
Estonia	1993	-3	GG	no, FR already in place since 1993 but lack of balance data
Finland	1995	-3	GG	Yes
France	1992	-3	GG	Yes
Gabon	2002	0	CG	Yes
Georgia	2014	-3	CG	Yes
Germany	1992	-3	GG	Yes
Greece	1992	-3	GG	Yes
Grenada	1998	-3	GG	no, < 1 million population in 2018
Guinea-Bissau	2000	-3	CG	Yes
Hong Kong SAR	1997	0	GG	Yes
Hungary	2004	-3	GG	Yes
India	2008	-3	GG	no, FR was only for 1 year in place
Indonesia	1980	-3	GG	no, FR already in place since 1967 but lack of data
Ireland	1992	-3	GG	Yes
Israel	2007	varying	CG	Yes
Italy	1992	-3	GG	Yes
Kosovo	2013	2	GG	Yes
Latvia	2004	-3	GG	Yes
Lithuania	2004	-3	GG	Yes
Luxembourg	1992	-3	GG	no, < 1 million population in 2018
Mali	2000	-3	CG	no, lack of fiscal balance data
Malta	2004	-3	GG	no, < 1 million population in 2018
Mexico	2006	0	CG	Yes
Montenegro, Rep. of	2014	-3	CG	no, < 1 million population in 2018
Netherlands	1992	-3	GG	Yes
Niger	2000	-3	CG	Yes
Nigeria	2007	-3	CG	Yes
Pakistan	2008	0	CG	Yes
Paraguay	2015	-1.5	CG	Yes
Peru	2000	-2 and -1.5 (varying targets)	CG	Yes
Poland	2004	-3	GG	Yes
Portugal	1992	-3	GG	Yes
Romania	2007	-3	GG	Yes
Senegal	2000	-3	CG	Yes
Singapore	1980	0	CG	no, FR already in place since 1962 but lack of data

Slovak Republic	2004	-3	GG	Yes
Slovenia	2004	-3	GG	Yes
Spain	1992	-3	GG	Yes
St. Kitts and Nevis	1998	-3	GG	no, < 1 million population in 2018
St. Lucia	1998	-3	GG	no, < 1 million population in 2018
St. Vincent and the Grenadines	1998	-3	GG	no, < 1 million population in 2018
Sweden	1995	1	GG	Yes
Togo	2000	-3	CG	Yes
United Kingdom	1992	-3	GG	Yes

Source: IMF Fiscal Rules Dataset, 2016.