IMF Working Paper

Bunching at 3 Percent: The Maastricht Fiscal Criterion and Government Deficits

by Francesca Caselli and Philippe Wingender

INTERNATIONAL MONETARY FUND

IMF Working Paper

Fiscal Affairs Department and Research Department

Bunching at 3 Percent: The Maastricht Fiscal Criterion and Government Deficits[*]

Prepared by Francesca Caselli and Philippe Wingender

Authorized for distribution by Catherine Pattillo and Xavier Debrun

August 2018

Abstract

This paper estimates the effects of the Maastricht treaty's fiscal criterion on EU countries' general government deficits. We combine treatment effects methods with bunching estimation, and find that the 3 percent deficit rule acts as a "magnet", increasing the number of observations around the threshold, while reducing the occurrence of both large government deficits and surpluses. After the rule is adopted, the distribution of government deficits among EU countries displays 20 percent excess mass around the deficit ceiling compared to a counterfactual distribution in which countries have the same observable characteristics but without the fiscal rule. Most of the bunching response comes from a reduction in the number of high deficit observations. We also find that the average treatment effect on fiscal deficits is positive and statistically significant. Finally, we derive country-specific impacts under a rank invariance assumption and find that all EU countries have seen their fiscal position improve on average as a result of the deficit rule.

Contents

I. INTRODUCTION

The global financial crisis and the buildup of large fiscal imbalances followed by painful adjustment and fiscal distress in several countries brought the European fiscal framework under scrutiny once again. For instance, in 2008, just before the global financial crisis, only 11 of the 27 European Union (EU) countries were in compliance with the fiscal criterion, which stipulates a ceiling of 3 percent on general government deficits and a 60 percent limit on debt-to-GDP ratios (European Commission 1992).[2] The sovereign debt crisis that followed has prompted renewed debate in academic and policy circles on the impact of the European fiscal rules on fiscal outcomes.

This paper contributes to the debate on fiscal rules by studying the effect of the 3 percent fiscal criterion introduced by the Maastricht treaty on the headline general government deficits of EU countries. We extend previous studies by looking at the impact of the rule on the entire distribution of general government deficits. Looking beyond the average effect is important since the rule could affect low and high-deficit countries differently, for instance because the incentives for deviating from a numerical fiscal limit can be non-linear, i.e. governments might incur a cost for breaching the rule, while the benefits of compliance could be minimal (e.g. Debrun and Kumar 2009). Indeed, political economy considerations may lead governments with large budgetary surpluses to relax their fiscal efforts in the presence of widespread non-compliance with the rule by other countries in the union (Alesina 2000). In such cases, lobbying pressures by various interest groups over the apportionment of the surplus may lead to a less stringent fiscal position. While the political process might interact with fiscal rules to enforce a specific fiscal policy, markets can also provide further incentives. For example, sovereign spreads may vary based on whether countries are in compliance with the rule or not (Diaz Kalan, Popescu and Reynaud 2018). The flexibility of our estimation approach means that we can identify the impact of fiscal rules on countries with high deficits, for example testing whether they are more likely to change their fiscal policies than other countries that are closer to the numerical rule, or above it. Documenting such heterogeneous effects is crucial, especially on high-deficit countries given their importance to policy makers.

The estimation of the causal effect of fiscal rules on fiscal balances is challenging. First, the adoption of the deficit ceiling can be correlated with other characteristics that could affect the conduct of fiscal policy. For instance, countries with social preferences for fiscal prudence might run smaller deficits than countries with a history of fiscal profligacy, even in the absence of the rule. Second, "rulers" and "non-rulers" also have other characteristics that could impact subsequent conduct of fiscal policies in complex ways. Therefore, a simple comparison between two groups of countries with and without the fiscal ceiling would yield a biased estimate of the rule itself. Our empirical approach to estimate causality proceeds in three steps. First, to correct for confounding factors that could affect both the likelihood of adopting a fiscal rule and subsequent conduct of fiscal policy, we implement the inverse probability weighting (IPW) procedure proposed by Hirano, Imbens, and Ridder 2003. This allows us to build a suitable counterfactual group under the assumption that adoption of the fiscal rule is as-good-as random conditional on a broad set of covariates. Second, combining

---

nonparametric and bunching estimation, we recover the causal impact of the deficit rule on the entire distribution of fiscal outcomes, documenting how the introduction of the rule has affected its shape. Third, upon further assumptions to match observations in the treatment and counterfactual groups, we recover country-specific results for the counterfactual path of government balances.

We find that the introduction of the 3 percent rule had a statistically significant impact on the distribution of government balances. While the presence of the 3 percent deficit ceiling has a positive effect on countries with very high deficits, it had a negative impact on countries with high balances, although this effect is small and not statistically significant. We also find that, after the introduction of the rule, government balances tend to converge towards a specific interval (bunching region) ranging from roughly -4.5 to 3.5 percent of GDP. We document that about 20 percent of the sample bunches in this range around the 3 percent deficit ceiling in the treated group compared to the counterfactual group. From this pool of bunchers, more than 80 percent come from the bottom of the distribution below -3 percent, while the rest comes from the top of the government balance distribution. Our empirical results imply that the fiscal criterion had an effect on government deficits even when it was not complied with, since deficits among EU countries above the 3 percent ceiling would have been even larger absent the fiscal rule.

Finally, when zooming in on country-specific results, we find that all EU countries have seen their fiscal position improve, on average, as a result of the deficit rule. Across these countries, government balances improved by 1.5 percent of GDP on average in the years in which the rule has been adopted. The paper also provides a number of checks to assess the robustness of our results across alternative model specifications, sample composition, and different weighting schemes.

The novelty of our approach consists in studying the effect of the fiscal rule on the entire distribution of government deficits, and also showing how this can be combined with bunching estimation. To our knowledge, this is the first study to combine these two approaches.[3] Bunching estimators have first been used to analyze earning responses to non-linear labor taxes (Burtless and Moffitt 1984; Friedberg 2000). In a seminal contribution, Saez 2010 studied the effects on income distributions of kink points in individuals' budget sets created by changes in marginal tax rates, showing how bunching can be used to estimate elasticities.[4] Most applications of this approach have also been in the context of behavioral responses to non-linear income tax schedules (Chetty and others 2011; Chetty and others 2013; Bastani and Selin 2014, Kleven and Waseem 2013; Kleven and others 2014). However, with the increasing availability of large administrative datasets, the approach has become more popular in other areas of empirical research as well. Recent studies in other areas include Manoli and Weber 2015 in the context of pensions, Harasztosi and Lindner 2015 who study the effect of minimum wages, and DeFusco and Paciorek 2016 who look at

---

[3] Shimizutani 2013 and Shimizutani and Oshio 2013 use a similar weighting approach to estimate the effects of Japan's social security earnings test. They rely on the decomposition method proposed by DiNardo, Fortin and Lemieux, 1996 to build counterfactual earnings distributions. However, the papers do not quantify bunching behavior.

[4] Kleven 2016 provides a recent and comprehensive overview of this growing literature.

the interest rate elasticity of mortgage demand. This paper is, to our knowledge, the first study to apply bunching estimation to macroeconomic data.

The approach in this paper complements the conventional bunching estimation approach. The conventional approach is to estimate smooth counterfactual distributions that would have been observed in the absence of kinks or notches induced by a discrete change in marginal tax rates for instance. Empirically, this is done by fitting polynomials to observed earnings distributions, excluding data in a range around some threshold. The advantage of our approach is that it provides a full characterization of the shape of the counterfactual distribution without a need to arbitrarily choose the bounds of the excluded range. It can also be more readily applied to macroeconomic data, which often have much fewer observations than administrative datasets, and where the visual determination of bunching behavior is harder to make. Finally, the robustness of the counterfactual density can be readily assessed through the use of standard treatment effects methods, such as covariate balancing.[5]

From a methodological point of view, this work is also related to the growing literature on heterogeneous and distributional effects (see for example Chamberlain 1994; Stock 1989; Heckman and Vytlacil 2007; Koenker 2017 and Bedoya and others 2017 for a recent review). Studies using quantile or distribution regressions have analyzed the factors behind changes in income inequality, the effects of training programs on earnings, education and student performance, the impact of welfare reform on household labor supply, as well as various topics in empirical finance, among others (Koenker and Hallock 2001).[6]

Finally, this paper also contributes to the empirical literature on the effectiveness of fiscal rules (Eyraud and others 2018 for a recent survey). Debrun and others 2008 find a positive and quantitatively relevant impact of national fiscal rules on the government balance in European countries.[7] Von Hagen 2006 finds that the Maastricht Treaty had a significant impact on political budget cycles in the EU, but was not successful in constraining fiscal policy in the large member states.[8] Other empirical studies have shown that countries with fiscal rules tend to have better fiscal outcomes, including by running smaller fiscal deficits (Bergman, Hutchison and Hougaard Jensen 2016; Debrun and others 2008; Tapsoba 2012).

---

[5] This allows us to check how similar the means and other moments of the covariate distributions are after reweighting the samples.

[6] For instance, in their seminal study of wage inequality among US workers, DiNardo, Fortin, and Lemieux 1996 decompose the changes in the distribution of wages over time into several components due to changes in labor market institutions, individual and macroeconomic factors. Notably, they document the impact of minimum wage regulations and unionization on the shape of the wage distribution and provide precise estimates of the specific income ranges affected by these changes. Estimation of the entire distributions also allows them to assess changes in terms of traditional inequality indicators such as the Gini coefficient and Theil index. Firpo 2007 and Donald and Hsu 2014 develop similar weighting methods to estimate quantile treatment effects and counterfactual cumulative distribution functions. Chernozhukov, Fernández-Val and Melly 2013 provide a general framework for counterfactual analysis using regression methods.

[7] For a review of the transition from the Maastricht Treaty to the EMU see Bini-Smaghi, Padoa-Schioppa and Papadia 1994 and Buiter, Corsetti and Roubini 1993 for a discussion of the convergence criteria. Eichengreen and Wyplosz 1998 review the reasons for the introduction of the Stability and Growth Pact in 1997.

[8] In the US context, Von Hagen 1991 finds that fiscal restraints do not have a significant impact on average fiscal performance, except for the choice of debt instruments. However, the results point to a significant effect on the distribution of per capita debt and debt-income rations. This evidence suggests that the effect of fiscal rules might go beyond a simple average effect, and therefore an analysis of the entire fiscal outcome distribution might be required.

Overall the evidence is mixed, which this is largely due to the challenges posed by endogenous rule adoption for estimating a causal relation (Poterba 1996; Alesina and Perotti, 1999; Grembi, Nannicini and Troiano 2016, Heinemann, Moessinger and Yeter 2017). Our paper contributes to this literature by developing a new identification strategy and obtaining estimates that go beyond the average effect.

While a fiscal rule might reduce the deficit bias, it might also sacrifice the benefits of fiscal policy stabilization. We deliberately focus on the disciplining effect of fiscal policy, and we do not attempt to quantifying the effect of rule on the level of procyclicality of fiscal policies.[9] On this different, but related aspect, Gali and Perotti 2003 find that discretionary budget deficits have become more counter-cyclical in the European Monetary Union (EMU). Manasse 2007 shows that deficit rules generate procyclical policies in intermediate state of the cycle, and countercyclical ones in extreme cases (booms or crises).

The paper proceeds as follow. Section II describes the data and their properties. Section III introduces the treatment effects framework and presents the first stage estimation. Section IV presents the distributional effects of the fiscal rule and the associated bunching estimates. Section V presents robustness exercises. Section VI shows country specific results under rank invariance assumption and Section VII concludes.

II. DATA AND DESCRIPTIVE STATISTICS

We estimate the impact of the Maastricht treaty's 3 percent fiscal deficit ceiling on the nominal fiscal balance in a panel of EU members and candidates from 1970 to 2017. The advantage of our sample is that the selected countries present some degree of homogeneity and detailed data on fiscal, macroeconomic and institutional variables are available over a long period of time both before and after the fiscal rule (FR) adoption for most countries. We use a treatment effects framework to estimate the causal impact of the FR, where our treatment variable is the adoption of the 3 percent deficit ceiling and the outcome variable is the nominal budget balance as a percent of GDP. We focus on the 3 percent headline deficit ceiling, first introduced with the Maastricht Treaty in 1992, since this is the operational threshold.[10] We control for a number of covariates that have been used in previous studies to model the probability of adoption of the rule in our first stage estimation. These include: lags of the government balance and government debt, inflation, GDP per capita, GDP per capita growth, an indicator variable for federal states, government fragmentation, and trade with EU-11 countries.[11] Appendix Table A2 provides a description of our estimation sample, data sources and construction of the variables. We explain the rationale for the inclusion of these variables in Section III below.

---

[9] Computing structural budget balances is difficult and subject to significant measurement error. Specifically, the structural balance is prone to ex-post revisions resulting from the measurement bias of potential GDP.

[10] As Manasse 2007 underlines, even when debt limits are put in place, they are not operational in setting annual fiscal policy.

[11] These are Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal and Spain.

Adoption of the 3 percent fiscal deficit rule among European countries followed 3 waves, in line with the enlargements of the EU: twelve countries adopted the rule in 1992 following the signing of the Maastricht Treaty.[12] Austria, Finland and Sweden subsequently adopted the rule in 1995 with the fifth EU enlargement, and Eastern European countries (plus Malta and Cyprus) introduced the rule between 2004 and 2007. Croatia was the last country in our sample to enter the European Union in 2013. Among the countries that entered the EU, Bulgaria, Croatia, the Czech Republic, Denmark, Hungary, Poland, Romania, Sweden and the United Kingdom did not adopt the Euro.[13] Countries in our sample that never adopted the fiscal criterion are: Albania, Iceland, Macedonia, Montenegro, Serbia, Switzerland and Turkey.

Table 1 presents the descriptive statistics for our variables separately for observations with and without the fiscal rule. Observations with the 3 percent rule (rulers) unsurprisingly have lower deficits on average compared to observations where the deficit rule is not in place (non-rulers). The means are -2.7 and -3 percent of GDP for country/years with the 3 percent deficit rule and country/years without, respectively. The table also shows that rulers and non-rulers are different in terms of their macroeconomic and institutional characteristics, supporting the idea that a simple comparison of the deficits between the two groups would give biased estimates of the impact of the fiscal rule. To have a full comparison of the two deficit distributions we also plot the kernel density estimates for the two groups separately in Figure 1. The figure shows that the two distributions differ beyond their means. The shapes of the density functions are also different. In particular, we observe that the distribution is somewhat more concentrated between -3 and 3 percent of GDP for observations with the fiscal rule, with thinner tails compared to observations without the rule. As already mentioned, the naïve comparison between these two distributions cannot be interpreted as the causal effect of the 3 percent deficit rule however, because the two groups have different characteristics which might also influence the outcome of interest.

---

[12] These countries are: Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and the UK. The Delors Report, issued in 1989, contained a plan for the realization of the European Monetary Union in three stages: 1) complete trade and capital markets integration 2) increase the coordination of monetary policy and strengthen economic convergence 3) fixing of the exchange rate and introduction of the Euro. The Report also referred to the need for "binding rules" for the functioning of stage 3, but not as a condition for entering that stage (Bini Smaghi and Padoa Schioppa 1994). The introduction of such rules was advocated on the grounds that "uncoordinated and divergent national budgetary policies would undermine monetary stability and generate imbalances in the real and financial sectors of the Community". The Maastricht Treaty also regulates inflation rates, exchange rates and interest rates. The Maastricht treaty constituted the basis for the subsequent Stability and Growth Pact (1995) that is the collection of secondary legislation that implements the Treaty on budgetary surveillance (European Commission 2013).

[13] Since the paper focuses on the FR imposed by the European Union fiscal framework, we note the process of adopting the 3 percent deficit ceiling coincides with EU accession for a number of countries. It is therefore not possible to identify separately these two effects in such cases without further assumptions. We think however unconfoundness is a reasonable assumption given our choice of first stage covariates. Specifically, it is unlikely that EU membership would have a material impact on fiscal balances beyond the deficit ceiling itself conditional on the predictors included in the first stage regression.

III. ESTIMATION UNDER SELECTION ON OBSERVABLES: INVERSE PROBABILITY WEIGHTING

In this section, we present the treatment effects framework by which we can recover a causal estimate of the impact of the fiscal rule on fiscal deficits. To do so, we introduce the potential outcome notation that will be used throughout the paper.[14] For a given value of the fiscal rule indicator $FR \in \{0, 1\}$ and a given outcome variable Y_{ct}, we define the potential outcomes $Y_{ct}(FR_{ct})$.

$$Y_{ct}(FR_{ct}) = \begin{cases} Y_{ct}(0) & if \ FR_{ct} = 0 \\ Y_{ct}(1) & if \ FR_{ct} = 1 \end{cases} \tag{1}$$

Since we cannot observe a country in both states simultaneously (with and without FR), we therefore have a *missing data problem* (Rubin 1974). This is the fundamental problem of causal inference: for any treatment allocated to an individual unit, the potential outcome associated with any alternative treatment cannot be observed. The solution to this *missing data problem* is to build a counterfactual that can answer the following question: what would have been the level of the outcome had the treatment not been assigned? In our case, the causal effect of the fiscal rule would be recovered by building a counterfactual value of the deficit had the country not adopted the rule and comparing it with the value of the deficit under adoption of the fiscal rule.

One particular moment of interest is the average effect of the rule on the rulers (the so called average treatment effect on the treated or $ATET$). This is given by:[15]

$$ATET = E\left[Y_{ct}(1) - Y_{ct}(0)|FR_{ct} = 1\right] \tag{2}$$
$$= E[Y_{ct}(1)|FR_{ct} = 1] - E[Yct(0)|FR_{ct} = 1]$$

This expression underlines the need to develop a counterfactual. The second term is in fact the average level of budget balance for rule adopters had they not adopted the fiscal rule.

If the adoption of the 3 percent fiscal rule was perfectly randomized, we could estimate the $ATET$ by simply comparing the average deficits in countries with and without the fiscal rule since treatment assignment would be independent of potential outcomes. However, the adoption of the fiscal rule is not exogenous and therefore a comparison of means or other moments of the distributions would lead to biased estimates due to omitted variables. As shown in Table 1, rulers and non-rulers differ along many important dimensions that can also

---

[14] For a comprehensive review of this topic, see Wooldridge 2010, and Imbens and Rubin 2015.

[15] An alternative quantity of interest also widely used in the empirical literature is the average treatment effect (ATE): $\beta = E\left[Y_{ct}(1)\right] - E[Y_{ct}(0)]$. We focus on ATET, since the required assumptions are weaker than the ones required for ATE. As noted by Wooldridge (2010), to estimate the ATE under selection on observables, the unobservables in the first stage model must be conditionally independent of the unobservables in both potential outcomes (treatment and control). On the contrary, to estimate the ATET, we only require the unobservables in the treatment model to be conditionally independent of the control-level potential outcome.

potentially impact their government budget balances beyond the impact of the rule itself. In other words, there is self-selection into rule adoption.

As is standard in observational studies (in contrast to randomized experiments), two assumptions are required to estimate causal effects, and in particular the $ATET$:

1.   Selection on observables: $Y_{ct}(FR_{ct}) \perp FR_{ct}| X_{ct}$. This means that potential outcomes, $Y_{ct}(0)$, $Y_{ct}(1)$, are jointly independent of treatment status, FR_{ct}, given observed covariates, X_{ct}.[16]

2.   Overlap: $0 < P(FR = 1|X_{ct}) < 1$. This means, for any value of the covariates, X_{ct}, countries have a positive probability of adopting a fiscal rule.

To overcome the selection bias into rule adoption, we use the inverse probability weighting scheme proposed by Hirano, Imbens, and Ridder 2003.[17] This strategy relies on a two-step procedure: first, we model the probability of FR adoption with a logit model that accounts for relevant characteristics that predict adoption. Second, we estimate the causal effect of FR adoption on the distribution of government balances, conditioning on the estimated propensity to adopt the FR. To compute the ATET and other moments of the treatment effect, we use the following weights to correct for the missing data on the potential outcomes:

$$\widehat{w}_{ct} = FR_{ct} - (1 - FR_{ct})\frac{\hat{P}_{ct}}{1 - \hat{P}_{ct}} \tag{3}$$

where $\hat{P}_{ct}$ are the propensity scores (or predicted probabilities) obtained from a logit model, and FR_{ct} is a dummy equal to 1 for observations where countries have the rule in place. By construction, outcomes of countries in the group of no-rulers that have been estimated to be more likely to adopt the FR (based on their observable characteristics) receive a greater weight $\widehat{w}_{ct}$. This produces a control group that is more comparable to the group of rulers, and therefore provides a suitable counterfactual for estimating the effect of having a rule on the fiscal balance.

It is important to note that we estimate the probability of "adopting" a rule, not "having" a rule.[18] Using adoption of the rule as left-hand-side variable and not presence in our first stage regression avoids the mechanical endogeneity given the panel structure of the data and our use of lagged government balance to explain adoption. This approach is supported by

---

[16] This assumption is alternatively described as uncounfoundness or ignobility of treatment.

[17] Recent macroeconomic studies that use this methodology include Acemoglu and others 2017, Angrist and Kuersteiner 2011, Angrist, Jordà, and Kuersteiner 2013, Serrato and Wingender 2016, Jordà and Taylor 2016.

[18] In practice, this means coding the left-hand side variable in the first-stage logit model as one only in the year of FR adoption and setting its value to missing for all the subsequent years when the rule is present.

Wooldridge 2005, who argues that controls that are themselves affected by treatment generally violates the key ignorability assumption.

We estimate the probability of adopting the FR using the following logit model:

$$\mathbb{P}(FR_{ct} = 1 | FR_{ct-1} = 0) = \Lambda(\alpha + \beta X_{ct}) \qquad (4)$$

The extensive literature on the determinants of fiscal rule adoption guides the choice of covariates used in the first stage estimation (Debrun and Kumar 2007, Combes and others 2014, Elbadawi, Schmidt-Hebbel and Soto 2015, and Badinger and Reuter 2017). Our identification strategy relies on the following predictors of fiscal rule adoption:

- Past fiscal behavior, captured by 3 lags of the deficit and 1 lag of the overall debt-to-GDP ratio have a direct impact on the probability of adopting a rule. We expect that fiscally conservative governments adopt rules to signal to agents the nature of their unobserved preferences. At the same time, high-deficit countries could introduce a rule to impose some discipline on their public finances. Therefore, the signs on these covariates are not clear *a priori*. Moreover, the difference in deficits displayed in Table 1 between rulers and non-rulers could be driven by systematic differences due to the business cycle. Including past fiscal behavior therefore controls for possible past variation in fiscal performance driven by country specific crises and other factors. Structural features that could be associated with systematic differences in fiscal policy are also captured by the age dependency ratio. Countries with high population dependency ratio will have a higher demand for social spending, making it more difficult to introduce fiscal discipline.
- Macro variables are also important to ensure identification: richer countries tend to have better institutions and human capital and thus will be more inclined to set up the institutional framework needed for fiscal rules. Countries with low levels of inflation (and in some cases with inflation targeting central banks) are more likely to have the necessary institutional framework to implement reforms to their fiscal institutions. For instance, inflation targeting adoption sometimes coincided with the introduction of fiscal reforms, including the establishment of fiscal rules (e.g. in Brazil, Norway, New Zealand or Sweden—see Combes and others 2014). Adding past inflation in the adoption regression also controls for the fact that nominal balances in the 1970s and 1980s were affected by high inflation rates. More generally, controlling for a broad set of macro and fiscal variables addresses the concerns that deficits in European countries might have been reduced, not because of the introduction of the rule, but because of other macro developments that impacted our sample during the period considered.[19]

---

[19] To further investigate the validity of our identification strategy, we also consider several dimensions of non-linearity in terms of the fiscal and macro variables. More specifically, we include squared and cubic terms in the deficit, debt and growth variables. We also separately include interaction effects for positive balances, balances above -3 percent and debt levels above 60 percent. This exercise is a further test to check, for instance, that our results are not affected by country

- Our set of controls also include institutional and political variables. For instance, we control for the degree of political stability since countries where the government is stable are more likely to prefer fiscal discipline and therefore to adopt rules. We also add a dummy variable for federal countries, since federal governments have weaker fiscal sovereignty and might have an interest in introducing rules to strengthen their bargaining position with states or provinces.
- Finally, to control for the European Union accession and the fact that many countries in our sample also adopted the euro, we follow Alesina, Barro and Tenreyro 2002, and we augment the baseline model with a variable that captures the intensity of the trade relationship with EU-11member countries. We also include a dummy equal to one for countries that joined the euro.[20]

Table 2 presents the marginal effects of these variables on the propensity to adopt the FR obtained via estimation of the first stage logit model. The variables that appear to have the strongest predictive power for the adoption of the FR are the age dependency ratio, the log of GDP per capita, lags of the balance and trade intensity with the EU-11. In terms of the fiscal variables, our results indicate that the adoption of FR appears to be motivated in part by the desire to address the increased spending demands from high age dependency ratios. Past fiscal deficits are also related to the adoption of FR, with the second and third lags showing up as particularly significant. Regarding levels of income, there is a strong indication that higher GDP per capita is associated with higher probability of adoption, in line with prior results in the literature. We note also that the negative coefficient on the trade variable is somewhat surprising given prior evidence that higher trade integration is associated with currency unions. However, this result seems to be due to the strong collinearity with other covariates. Indeed, the estimated coefficient in the univariate case (unreported) is statistically significant and positive. Finally, both the pseudo R-squared and the area under the receiver operating characteristic curve (AUROC) indicates that model IPW4 achieves the highest predictive power. We therefore proceed with this specification as the baseline for the distributional analysis that follows.

While we cannot directly test the selection on observables assumption to assess the quality of our counterfactual, we can check whether our weighting scheme improves the balancing of the covariates. To do so, we first compare the raw data series with the reweighted ones, using estimated weights calculated from Equation (3). Table 3 reports the differences in means between treatment and control in the raw and reweighted data together with the ratio of variances for the estimated model IPW4.[21] The comparison of the raw and reweighted series

---

specific crises, that disproportionally increased subsequent debt and deficits levels. Results are qualitatively unaffected by the inclusion of these non-linear terms.

[20] In a previous version of this paper, we also include output price co-movements across EU countries, and a measure that captures the extent to which governments undergo regular checks and balances. Results do not change.

[21] Results are qualitatively similar for the other models.

reveals improved covariate balancing, given that the standardized differences between the means of the reweighted series are in all cases closer to zero. To formally test whether the means are different from zero, we also implement the covariate balancing test developed by Imai and Ratkovic 2014.[22] Consistent with the statistics reported at the bottom of the previous Table 2, we cannot reject the null hypothesis that covariates in the reweighted sample are mean-balanced for all the four models. This provides further supporting evidence that our counterfactual is appropriately estimated and modelled. For the variance ratios, we note that we cannot always improve on the raw comparisons, as the variance ratios are in some cases further away from one for the reweighted data. While this could be due to a relatively small sample, we check the robustness of our approach by implementing an alternative balancing procedure based on entropy weighting (Hainmuller 2012). In Section V, we show that our results are robust when using alternative estimators.

The overlap assumption is satisfied when there is a sufficiently high probability of seeing observations in both the control and the treatment groups, at each combination of the covariate values. We visually check the degree of overlap by plotting the estimated densities of the probability of getting each treatment level. Figure 2 reports the densities of the estimated propensity scores for rulers and non-rulers. Busso, DiNardo, and McCrary 2014 argue that the overlap assumption is violated when the estimated densities have too much mass around either 0 or 1. For all four of the first stage models, neither samples present excessive probability mass near 0 or 1, with the densities also displaying significant overlap between the two groups.

Finally, in Figure 3 below we show the estimated inverse probability weights for individual observations in the counterfactual group.[23] The weights are estimated from Equation (4) using the four model specifications reported in the Table 2. Country-year labels are also included for a few observations with particularly large weights, notably Iceland in 2016, Slovakia in 2002 and Luxembourg in 1989. As discussed above, these observations were deemed to be have a high propensity to adopt the fiscal rule, despite not having the rule in place at the time. While these individual observations receive a larger weight than most other observations, the figure also shows that none of the control group observations are outliers that could singlehandedly drive subsequent distribution estimates.

IV. EFFECTS ON THE DISTRIBUTION OF DEFICITS AND BUNCHING ESTIMATION

To obtain the full distributional effects of the FR on fiscal deficits, we compare the density estimates of the treated observations with FR with the counterfactual (reweighted) distribution. Let $f(y|FR)$ denote the probability density functions (PDFs) of potential

---

[22] The test is implemented via Stata's *tebalance overid* command.

[23] Observations in the treatment group receive a weight of one to recover the treatment on the treated. These are omitted in the chart.

outcomes given FR status. We use a kernel estimator to recover the densities for both groups as follows:

$$\hat{f}(y|FR) = \frac{1}{NT}\frac{1}{h}\sum_{c=1}^{N}\sum_{t=1}^{T}\hat{w}_{ct}K\left(\frac{Y_{ct}(FR_{ct}) - y}{h}\right) \tag{5}$$

for country c in year t, with $\hat{w}_{ct}$ defined as in (3), bandwidth h and Epanechnikov kernel function K. We use a bandwidth of 0.75 in all baseline results presented below.[24] Firpo 2007 and Donald and Hsu 2014 use a similar weighting approach to estimate potential outcomes for quantiles and cumulative distribution functions. We choose to focus on PDFs since they allow us to further analyze bunching behavior. Figure 4 plots the treated and counterfactual government balance distributions estimated following Equation (5) for each group separately. The weights used to correct for selection bias are those obtained from the IPW4 model in Table 2. Robustness checks using the other first stage models are also presented in Figure 5. All counterfactual kernel density estimates display very similar shapes confirming that our results are not driven by specific modelling assumptions in the first stage regression.

Returning to Figure 4, we show in addition to the kernel estimates, three relevant moments for both distributions. First we compare the means and note that the average effect on government balances is positive—the middle blue solid line is located to the right of the middle dashed red line—and estimated at 1.5 percent of GDP. However, we also find that the effects at the 5th and 95th percentiles are of opposite signs.[25] The estimated effects on both quantiles indicate that the presence of the 3 percent deficit ceiling had a positive effect on countries with very high deficits, improving balances by 3.1 percent of GDP, and a negative effect on countries with very lower deficits (high balances), reducing balances by 1.2 percent of GDP. These effects therefore describe a "magnet" effect of the FR on government balances. As discussed earlier, while the average treatment effect is well defined in this setting, it nevertheless fails to describe the full extent of the impact of the fiscal rule on countries' deficits. It overstates the impact on observations with a high fiscal balance, but more importantly, it underestimates the impact on countries with high deficits, arguably those countries which are of most concern to policy makers.

We also find that the introduction of the 3 percent deficit rule changed the shape of the counterfactual distribution. Figure 4 shows important differences with Figure 1. The shape of the counterfactual distribution is different from the raw data distribution, most notably at the top end of the distribution with excess density around 5 percent of GDP compared to the treatment and no-ruler distributions. This is caused by the fact that in the counterfactual

---

[24] The bandwidth was chosen using Stata's optimal bandwidth selection for the treatment group. The baseline results are robust to a wide range of bandwidths.

[25] Appendix Table A3 provides the $ATET$ estimates for the four different IPW models as well as simple difference-in-differences regressions. The effects at different quantiles of the distribution can similarly be recovered using the quantile treatment effect approach proposed by Firpo 2007.

distribution, many observations with very high government balances were also assigned a high estimated probability of treatment and therefore a large weight $\widehat{w}_{ct}$. The density around the mean of the counterfactual group is also depressed in Figure 4. This leads to larger excess mass around the threshold when comparing treated and counterfactual.

Another interesting pattern emerges from Figures 4 and 5. We notice that the treated and counterfactual densities cross each other in two points of the distributions, namely where the budget balance is equal to -4.3 and 3.3, respectively in the case of the IPW4 model. This means that the introduction of the deficit rule generated a concentration of observations in the range bounded by the two densities' intersections.

To analyze this magnet effect, we adopt a bunching estimation approach. We define bunching B as the excess mass in the treatment distribution compared to the counterfactual distribution around the threshold. This can be expressed as:

$$B = \int_{y-}^{y+} f(y|FR = 1) - f(y|FR = 0)dy,$$

where $f(y|FR)$ are the treated and counterfactual densities, and y^+ and y^- are the bounds of the bunching range where these two densities cross each other around the threshold.

Empirically, we estimate total bunching B, as well as excess mass in the top and bottom ranges, using the following equation:

$$\hat{\alpha}^k, \hat{B}^k = \arg\min_{\alpha^k, B^k} \sum_{c,t} \widehat{w}_{ct} \cdot [1(Y_{ct} \in R^k) - \alpha^k - B^k FR_{ct}]^2, \quad k \in \{l, b, u\} \quad (6)$$

where R^k denotes the three bunching ranges with partitioning determined by the intersections of the treatment and counterfactual densities $(\hat{y}^-, \hat{y}^+)$, the dummy FR_{ct} defined as above, and the weights $\widehat{w}_{ct}$ estimated from the selection model in Equation (4). We show estimates for the three areas' $\hat{B}^k$ and their corresponding standard errors for the four models in Table 4.[26] From our preferred specification in column (4), we find that almost 20 percent of the sample is attracted towards the bunching area located between -4.3 and 3.3 percent of GDP. This estimate is also statistically significant at the 5 percent level. From this pool of bunchers, 86 percent come from the bottom of the distribution below -4.3, while 14 percent comes from the top above 3.3 percent of GDP. Table 4 also confirms that this pattern is robust across the different first stage specifications. These results suggest that the fiscal rule worked mainly by reducing large deficits, and particularly that the rule was successful in constraining fiscal policies in countries that do not comply with the threshold, showing that rules can therefore have an effect on deficits even when they are not complied with. At the

---

[26] Because the weights $\widehat{w}_{ct}$ in Equation (6) are estimated in a first stage, we estimate both steps simultaneously using GMM to obtain the correct standard errors for the bunching estimator.

same time, the 3 percent ceiling also had an impact on good performers, by causing them to reduce their fiscal effort and lower their high budget balances.

V. ROBUSTNESS CHECKS

We test the robustness of our results along several dimensions. We first adopt alternative weighting schemes to ensure the results are not driven by the particular weights imposed by the logit functional form assumption in the first stage. Also, as already mentioned, we notice that, unlike for mean differences, the IPW do not always improve covariate balancing in terms of variance ratios. We therefore implement the entropy weighting approach proposed by Hainmuller 2012.[27] The advantage of this method with respect to inverse probability weighting or propensity score matching approaches, is that covariate balancing is directly built into the weighting function that is used to adjust the control group. Instead of checking for covariate balancing after the first stage, we can *a priori* choose the level of balancing, up to the first, second, and potentially higher moments of the covariate distributions in the treatment and reweighted counterfactual groups. This will always lead to an improvement in terms of balancing with respect to propensity score weighting methods so that checking covariate balancing is not necessary to assess the model's identifying assumptions. Based on Table 3, we match the first as well as the first and second moments of the covariate distributions. Figure 6 shows the kernel density estimates obtained with this alternative first stage procedure. Both series adjusted to the first and second moments of the covariate distribution show similar shapes to the one obtained using the IPW scheme with logit first stage.[28]

In addition to entropy weighting, we estimate a local logit model in the first stage to obtain a non-parametric estimate of the propensity score.[29] This relaxes the parametric assumption based on the logit first stage model. The fully nonparametric estimation reported in Figure 6 once again provides very similar results to the one obtained in the baseline estimation. This is reassuring since it provides evidence that the baseline results are largely unaffected either by the level of covariate balancing, or by the functional form assumed in the first stage. The bunching results with these three alternative weighting schemes—using covariates from the first stage model IPW4—are also shown in Table 5.[30] They confirm the same pattern presented in the baseline results. Note that unlike results in Table 4, we use the bootstrap to conduct inference for these robustness checks. Specifically, we implement the pairs cluster bootstrap-t approach proposed by Cameron, Gelbach and Miller 2008. This also has the added benefit of correcting for the uncertainty associated with the bunching bounds, which are themselves estimates. It is reassuring to note that the statistical significance of the

---

[27] For applications of this weighting scheme, see Egger and Tarlea 2017, and Marcus 2013.

[28] We do not include inflation in the list of covariates when estimating entropy weights for the first and second moments as the algorithm fails to converge.

[29] Frolich and Melly 2010 provides an overview of this estimator.

[30] We omit the inference results for the entropy-based estimates since the bootstrap algorithm fails to converge almost 60 percent of the time across bootstrap iterations.

bunching estimator using the logit model in the first column is largely unchanged when the bootstrap is used.

Finally, we assess the robustness of our inference by adopting a different clustering scheme. Our baseline specification uses standard error clustered at the country level. However, there might be unobserved correlation in outcomes across countries, especially for countries that adopted (or intended to adopt) the fiscal rule at the same time, such as Austria, Finland and Sweden that joined the EU in 1995. To address this potential bias in our standard errors, we use clustering at the adoption wave level. We group countries in our sample in 9 such waves depending on the year in which they adopted the Maastricht treaty, joined the EU or became candidates for accession (see Table A1). Table 6 shows the results when clustering by adoption waves of adoption similar bunching similar results in terms of magnitude. Inference is one again conducted via the pairs cluster bootstrap-t, with p-values reported. As discussed in Cameron, Gelbach and Miller 2008, this bootstrapping approach is particularly well suited for inference with few clusters.

VI. COUNTRY-SPECIFIC RESULTS

The results from the previous sections provide evidence that the 3 percent deficit rule had an impact on the whole distribution of government deficits across European countries. However, the impact of the treatment on the distribution is distinct, and generally not equal, to the distribution of treatment effects (Bitler, Gelbach and Hoynes 2005). The reason is that the treatment effect on the distribution only requires information on the marginal distributions $\hat{f}(y|FR)$, which are always identified under assumptions a) and b), while to estimate the distribution of impacts, information on the joint distribution of outcomes are required (Heckman, Smith, and Clements 1997). Further assumptions are therefore needed to recover the effect of the rule on individual countries, since this requires observing jointly both treatment and counterfactual outcomes for every single observation. Given the *missing data problem* discussed above, a common approach in the literature has been to assume rank invariance. This assumption states that the ordering in the treatment group remains the same in the counterfactual group without the deficit rule, which means that the position of one country in the distribution of deficits is unaffected by the introduction of the rule. For example, Italy in 1996 was at the 10th percentile of the treatment group with a deficit of 6.75 percent of GDP. Under the rank invariance assumption, Italy's unobserved rank in the counterfactual group would therefore remain at the 10th percentile with a deficit of 7.8 percent of GDP. The assumption of rank invariance is quite strong, but does not seem unrealistic in the context of fiscal behavior, which displays some persistence (e.g., due to operational and political difficulties to reduce the deficit drastically in a short period of time, role of social preferences, etc.).

Figure 7 presents the average effect of the FR by country over the sample period under rank invariance assumption. All 28 countries that adopted the 3 percent deficit rule in our sample have seen an improvement of their average annual deficit across the sample period. This

means that all countries would have recorded a higher deficit on average without the rule in place. Across the sample, government balances improved by 1.5 percent of GDP on average, ranging from slightly over 3 percent of GDP improvement for Greece, to less than 0.25 percent of GDP for Luxembourg.[31] Finally, in Figures 8 and 9 we assess how the effect on the FR varies according to the observed value of the government balance and debt level. Consistent with evidence from the densities in Figures 4 through 6, in Figure 8 the treatment effect is largest at the bottom of the government budget balance distribution. For observations with deficit exceeding 10 percent of GDP, the FR rule improves fiscal outcomes by more than 4 percent of GDP. The effect is also negative at the top, when government balances exceed 2 percent of GDP. Another interesting pattern is the treatment effect on countries that are just above the -3 percent limit. For these observations, even though countries are in compliance with the fiscal criterion, the rule still leads them to tighten government balances, by around 1 percent of GDP. In Figure 9, we find that the impact of the rule was generally larger for countries with higher levels of debt, reaching around 3 percent of GDP in annual deficit adjustment. This suggests that the fiscal rule facilitated more ambitious fiscal adjustments in countries with higher debt.

VII. CONCLUSION

This paper shows that the effect of fiscal rules on fiscal performance goes beyond an impact on the average outcome and affects the whole distribution of fiscal balances in complex ways. Looking at the 3 percent deficit ceiling in EU countries, we find that the average treatment effect is statistically significant, positive and small. This average impact also masks important variation across the distribution. Notably, we observe large positive effects at the bottom of the distribution on the countries with the largest deficits. This evidence therefore suggests that the FR is effective in constraining fiscal policies even for countries that do not comply with the deficit threshold. We also find that the stronger fiscal performers with very large budget surpluses tend to reduce the government balance and converge towards the 3 percent ceiling compared to a counterfactual outcome without the FR in place. Taken together, these results show that the 3 percent deficit rule acts as a "magnet", increasing the number of observations around the threshold while reducing the occurrence of both large government deficits and surpluses. Consequently, policy makers should carefully consider the costs and benefits of introducing numerical targets since the impacts might be quite complex and interact in unexpected ways with countries' current fiscal position.

Another contribution of this paper is to demonstrate how treatment effects methods can be used to estimate bunching behavior in response to threshold-based policies. To our knowledge this is the first application combining the two methods and we show how this can yield new results. In particular, we provide clear evidence of bunching using macroeconomic data, something that had not been documented in the literature before.

---

[31] Counterfactual deficit paths for every country that adopted the FR are given in Figure A2.

While we carefully assess the robustness of our results along numerous dimensions, we do not investigate the channels by which FR affect fiscal policy. For instance, the fiscal rule could introduce a reference point (Kleven 2016) for voters such that the probability of reelection for a government could vary discontinuously around the threshold. Similarly, financial markets could respond differently to fiscal outcomes that are close, but on different sides of the deficit ceiling. Exploring these channels would be important to further understand how fiscal rules impact fiscal policies and how these responses should inform the design of fiscal rules going forward.

VIII. REFERENCES

Acemoglu, D., S. Naidu, P. Restrepo, and J. Robinson, 2016. "Democracy Does Cause Growth", Journal of Political Economy, *forthcoming*.

Alesina, A., 2000. "The Political Economy of the Budget Surplus in the United States", *Journal of Economic Perspectives,* Vol. 14, Issue. 3, pp. 3–19.

Alesina, A., R. J. Barro and S. Tenreyro, 2002. "Optimal Currency Areas," *NBER Macroeconomics Annual*, 17, 301–345.

Alesina, A., and R., Perotti, 1999. "Budget Deficits and Budget Institutions," Fiscal Institutions and Fiscal Performance, edited by J. Poterba and J. Von Hagen, 13-36. Chicago, IL: University of Chicago Press and NBER, 13–36.

Angrist, J.D., and G. M. Kuersteiner, 2011. "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests with a Multinomial Propensity Score" *Review of Economics and Statistics*. Vol. 93. Issue 3, pp. 725–747.

Angrist, J.D., Ò. Jordà, and G. M. Kuersteiner, 2013. *"Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited,"* NBER Working Papers 19355, National Bureau of Economic Research, Inc.

Badinger, H. and W. Reuter, 2017. "The Case for Fiscal Rules", *Economic Modelling*, 60, pp. 334–343.

Bastani S., and H. Selin, 2014. "Bunching and non-bunching at kink points of the Swedish tax schedule", *Journal of Public Economic*, Vol. 109, pp. 36–49.

Bedoya, G., Bittarello, L., J., Davis and N. Mittag, 2017. "Distributional Impact Analysis. Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect", World Bank Policy Research Working Paper 8139.

Bergman, U. M., M. M. Hutchison, and S.E. Hougaard Jensen, 2016. "Promoting sustainable public finances in the European Union: the role of fiscal rules and government efficiency," *European Journal of Political Economy.* Volume 44, pp. 1–19.

Bini Smaghi, L., Padoa-Schioppa, T., and F. Papadia, 1994, "The Transition to EMU in the Maastricht Treaty", Essays in International Finance, No. 194, Princeton University.

Buiter W., Corsetti G., and N. Roubini, 1993, "Excessive Deficits: Sense and Nonsense in the Treaty of Maastricht", *Economic Policy*, Vol. 8, No. 16, pp. 57–100.

Burtless, G., and R. A. Moffitt. 1984. "The Effect of Social Security Benefits on the Labor Supply of the Aged," In *Retirement and Economic Behavior*, ed. Henry J. Aaron and Gary Burtless, 135–71. Washington, DC: Brookings Institution.

Busso, M., J. DiNardo, and J. McCrary, 2014. "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators", *Review of Economics and Statistics*. Vol. 96. Issue 5.pp. 885–897.

Cameron, A.C., J. B. Gelbach, and D.L. Miller, 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90(3): 414–427.

Chamberlain, G., 1994. "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics*, Sixth World Congress, Vol. 1, ed. by C. A. Sims. Cambridge: Cambridge University Press.

Chernozhukov, V., I. Fernández-Val, and B. Melly, 2013. "Inference on Counterfactual Distributions," *Econometrica*, 81(6), 2205–2268.

Chetty R., Friedman J. N., Olsen T, and L. Pistaferri, 2011. "Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: evidence from Danish tax records", *Quarterly Journal of Economics*, Vol. 126, pp. 749–804.

Combes J. L., X. Debrun, A. Minea and R. Tapsoba, 2017. "Inflation Targeting, Fiscal Rules, and the Policy Mix: Cross-Effects and Interactions", forthcoming in *Economic Journal.*

DeFusco A., and A. Paciorek, 2016. "The interest rate elasticity of mortgage demand: evidence from bunching at the conforming loan limit", *American Economic Journal: Economic Policy*, Vol. 9, pp. 210–240.

Debrun, X., and M.S. Kumar. 2007. "Fiscal Rules, Fiscal Councils, and All That: Commitment Devices, Signaling Tools or Smokescreens?" Available at SSRN: https://ssrn.com/abstract=2004371

Debrun, X., and M.S. Kumar. 2009. "The Discipline-Enhancing Role of Fiscal Institutions: Theory and Empirical Evidence," in *Policy Instruments for Sound Fiscal Policies*, pp. 51–88. Palgrave Macmillan, London, 2009.

Debrun, X., L. Moulin, A. Turrini, J. Ayuso-i-Casals and M.S. Kumar, 2008. "Tied to the mast? National fiscal rules in the European Union," *Economic Policy*, 23(54), 297–362.

Diaz Kalan, F., Popescu, A. and J. Reynaud, 2018. "Thou Shalt Not Breach: The Impact on Sovereign Spreads of Noncomplying with the EU Fiscal Rules," IMF WP 18/87.

DiNardo, J., N. Fortin, and T. Lemieux, 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach," *Econometrica*, 64, 1001–1044.

Donald, S. G., and Y.C. Hsu, 2014. "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," *Journal of Econometrics*, 178, 383–397.

Egger, P., and F. Tarlea, 2017. "Comparing Apples to Apples: Estimating Consistent Partial Effects of Preferential Economic Integration Agreements," CEPR Discussion Papers 11894.

Eichengreen, B. and C. Wyplosz, 1998. "Stability Pact. More than a minor nuisance?", *Economic Policy*, Vol. 13, No. 26, pp. 65–113.

Elbadawi, I., K. Schmidt-Hebbel and R. Soto, 2015. "Why do Countries have Fiscal Rules" in *Economic Policies in Emerging-Market Economies*, Festschrift in Honor of Vittorio Corbo, edited by Ricardo J. Caballero and Klaus Schmidt-Hebbel, Santiago, Chile.

European Commission, 2013. "Building a Strengthened Fiscal Framework in the European Union: A guide to the Stability and Growth Pact"*, Occasional Papers 150, May 2013.*

European Commission, 2017. *Vade Mecum on the Stability and Growth Pact,* Institutional Paper 052, https://ec.europa.eu/info/sites/info/files/ip052_en_0.pdf

Eyraud, E., X. Debrun, A. Hodge, V. Lledó, and C. Pattillo, 2018. "Second-Generation Fiscal Rules: Balancing Simplicity, Flexibility, and Enforceability," IMF Staff Discusssion Note 18/04.

Firpo, S. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects," Econometrica, 75, 259–276.

Friedberg, L. 2000. "The Labor Supply Effects of the Social Security Earnings Test," *Review of Economics and Statistics*, 82(1): 48–63.

Frolich, M. and B. Melly, 2010. "Estimation of quantile treatment effects with Stata," The Stata Journal, 10(3): 423–457.

Gali, J., and R. Perotti, 2003, Fiscal Policy and Monetary Integration in Europe, *Economic Polic*y, Vol. 18, No. 37, pp. 533–572.

Grembi, V., Nannicini, T., and U. Troiano, 2016. "Do Fiscal Rules Matter?", *American Economic Journal: Applied Economics,* Vol. 8, Issue 3, pp. 1–30.

Harasztosi, P., and A. Lindner, 2015. "Who pays for the minimum wage?" Working Paper, University College London.

Heckman, J. J., J. Smith, and N. Clements, 1997. "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

Heckman, J.J. and E. Vytlacil, 2007. "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics,* Vol. 6, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, Chapter 70.

Heinemann, F., M-D Moessinger, and M. Yeter, 2017. "Do Fiscal Rules Constrain Fiscal Policy? A Meta-Regression Analysis," *European Journal of Political Economy*.

Hainmueller, J. 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis, 20*(1), pp. 25–46.

Hirano, K., G. W. Imbens, and G. Ridder, 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica*, 71(4), 1161–89.

Imai, K., and M. Ratkovic, 2014. *"Covariate balancing propensity score"*, Journal of the Royal Statistical Society, Series B 76: 243–263.

Imbens, G. W., and J. M. Wooldridge, 2009. "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Imbens, G. W., and D. R. Rubin, 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences. An Introduction*. Cambridge University Press.

Jordà, Ò., and A. Taylor, 2016. "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy", The *Economic Journal*, Vol. 126, Issue. 590, pp. 219–255.

Kleven H. 2016. "Bunching", *Annual Review of Economics*, Vol. 8, pp. 435–464.

Kleven H, and M., Waseem, 2013. Using notches to uncover optimization frictions and structural elasticities: theory and evidence from Pakistan. Q. J. Econ. 128:669–723

Kleven H, C. Landais, E. Saez, and E. Schultz, 2014. "Migration and wage effects of taxing top earners: evidence from the foreigners' tax scheme in Denmark", *Quarterly Journal of Economics*, Vol. 129, pp. 333–78.

Koenker, R. 2017. "Quantile Regression: 40 Years On," *Annual Review of Economics*, 9:1, 155–176.

Koenker, R., and K.F. Hallock, 2001. "Quantile Regression," *The Journal of Economic Perspectives*, 15(4), 143–156.

Manasse, P. 2007. "Deficit Limits and Fiscal Rules for Dummies," IMF Staff Papers, vol. 54(3), pp. 455–473.

Manoli D., and A. Weber, 2015. "Nonparametric evidence on the effects of financial incentives on retirement decisions", Work. Pap., Univ. Mannheim.

Marcus, J. 2013. "The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany," *Journal of Health Economics*, Vol. 32, Issue 3, pp. 546–558,

Poterba, J.M. 1996. "Budget institutions and fiscal policy in the U.S. states," *American Economic Review* 86 (2), 395–400.

Saez, E. 2010. "Do Taxpayers Bunch at Kink Points?", *American Economic Journal: Economic Policy*, Vol. 2, Issue. 3, pp. 180–212.

Schaechter, A., Kinda, T., Budina, N. and A. Weber, 2012. "Fiscal Rules in Response to the Crisis—Toward the "Next-Generation" Rules: A New Dataset," IMF WP 12/187.

Serrato, J.C., and P. Wingender, 2016. "Estimating Local Fiscal Multipliers," NBER Working Paper 22425.

Shimizutani, S. 2013. "Social Security Earnings Test and the Labor Supply of the Elderly: New Evidence from Unique Survey Responses in Japan," *Japanese Economic Review*, 64(3): 399–413.

Shimizutani, S., and T. Oshio, 2013. "Revisiting the Labor Supply Effect of Social Security Earnings Test: New Evidence from its Elimination and Reinstatement in Japan," *Japan and the World Economy*, 28, 99–111.

Stock, J. H. 1989. "Nonparametric Policy Analysis, "*Journal of the American Statistical Association*, 84, 567–575. [2205,2207]

Tapsoba, R. 2012. "Do National Numerical Fiscal Rules really shape fiscal behaviors in developing countries? A treatment effect evaluation", *Economic Modelling*, 29, pp.1356–1369.

Von Hagen, J. 1991. "A note on the empirical effectiveness of formal restraints", *Journal of Public Economics,* Vol. 44, Issue. 2, pp. 199–210.

Von Hagen, J. 2006. "Fiscal Rules and Fiscal Performance in the European Union and Japan," *Monetary and Economic Studies*, Institute for Monetary and Economic Studies, Bank of Japan, vol. 24, pp. 25–60.

Wooldridge, J. M. 2005. "Violating Ignorability of Treatment by Controlling for too many factors," *Econometric Theory,* Volume 21, pp. 1026–1028.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

Table 1. Summary statistics

	Rulers					Non-rulers				
	Obs.	Mean	St.dev	Min	Max	Obs.	Mean	St.dev	Min	Max
Government balance	548	-2.7	3.6	-32.1	6.9	743	-3.0	4.0	-19.6	12.8
Government debt	548	61.5	33.0	3.7	180.8	784	41.9	26.4	1.0	200.0
Age dependency ratio	548	49.2	4.2	38.5	60.9	1,132	53.1	7.0	41.2	85.5
Log GDP per capita	548	10.4	0.3	9.5	11.5	932	9.6	0.6	7.9	11.0
GDP per capita growth	548	0.0	0.0	-0.1	0.2	897	0.0	0.1	-0.5	0.5
Inflation	548	2.3	2.1	-4.5	15.9	855	17.4	37.9	-1.3	250.0
Federation	548	0.5	0.5	0.0	1.0	1,132	0.3	0.5	0.0	1.0
Legislative fractionalization	548	71.5	10.2	50.1	89.5	903	59.7	25.8	0.0	90.6
Trade with EU-11	548	18.0	10.1	6.0	54.6	785	13.3	8.5	0.7	45.2
Euro dummy	548	0.7	0.5	0.0	1.0	1,132	0.4	0.5	0.0	1.0

Note: See Table A2 for data sources.

Table 2. Logit Marginal Effects

	IPW 1	IPW 2	IPW 3	IPW 4
Lag 1 balance	-0.121	-0.102	-0.049	0.014
	(0.266)	(0.245)	(0.280)	(0.257)
Lag 2 balance	-0.871**	-0.807**	-0.877**	-0.784**
	(0.366)	(0.342)	(0.372)	(0.334)
Lag 3 balance	0.614**	0.556*	0.720**	0.683**
	(0.313)	(0.290)	(0.320)	(0.286)
Lag 1 debt	0.016	0.041	0.034	0.073
	(0.036)	(0.040)	(0.042)	(0.046)
Age dependency ratio	-0.625***	-0.551***	-0.688***	-0.661***
	(0.201)	(0.185)	(0.227)	(0.183)
Log GDP per capita	3.818*	5.390***	3.555	5.770***
	(2.321)	(2.000)	(2.471)	(2.208)
GDP per capita growth	49.556**	44.255*	50.757**	44.650
	(23.514)	(25.420)	(25.069)	(27.712)
Inflation	-0.205	-0.233*	-0.223	-0.269*
	(0.140)	(0.134)	(0.144)	(0.142)
Federation		-3.730*		-5.054**
		(1.922)		(2.310)
Legislative fractionalization		0.027		0.019
		(0.059)		(0.069)
Euro dummy			2.434	3.365*
			(1.707)	(2.037)
Trade with EU-11			-0.088	-0.166*
			(0.059)	(0.088)
Observations	652	644	640	632
Pseudo R-squared	0.159	0.169	0.167	0.189
AUROC	0.817	0.818	0.828	0.846
Covariate balance (p-value)	0.966	0.877	0.915	0.881

Note: The table reports the marginal effects from the first stage logit model. The dependent variable is a dummy equal to one in the year that a country adopts a fiscal rule. We restrict the sample to observations without the FR in the previous year to estimate the probability of adoption. The area under the receiver operating characteristic curve (AUROC) is a measure of predictive power of the first stage model. A model with no predictive power has a value of 0.5, while perfect predictive power would yield a value of 1. Covariate balance reports the p-value of a chi-squared test for whether the estimated propensity score balances the reweighted means of the covariates. Standard errors are clustered at the country level in parentheses.
* p<10%, **p<5%, *** p<1%.

Table 3. Covariate Balancing

	Standardized difference		Variance ratio	
	Raw	Weighted	Raw	Weighted
Lag 1 balance	-0.274	-0.003	0.906	0.665
Lag 2 balance	-0.354	0.017	1.379	0.924
Lag 3 balance	-0.101	-0.012	1.193	0.927
Lag 1 debt	0.259	0.011	1.276	1.269
Age dependency ratio	-0.746	0.014	0.549	1.407
Log GDP per capita	0.520	0.001	0.407	0.530
GDP per capita growth	0.352	0.016	0.681	0.957
Inflation	-0.425	-0.003	0.017	0.464
Euro dummy	0.306	-0.021	0.951	1.013
Trade with EU-11	0.289	-0.015	0.926	0.719
Federation	-0.132	-0.005	0.974	0.997
Legislative fractionalization	0.225	0.018	0.903	0.993

Note: The table reports the standardized difference in covariate means between treated and control groups in the raw and reweighted samples using the regression specification from IPW4 in Table 2. The table also reports the ratio of variances of the first stage covariates before and after the reweighting procedure. Perfectly balanced covariates have a standardized difference of zero and variance ratios of one.

Table 4. General Government Balance: Bunching Estimates

|	IPW1	IPW2	IPW3	IPW4
Bunching area	0.193**	0.192**	0.195**	0.189**
	(0.078)	(0.081)	(0.079)	(0.080)
Top	-0.023	-0.024	-0.025	-0.025
	(0.027)	(0.029)	(0.023)	(0.024)
Bottom	-0.170**	-0.167**	-0.171**	-0.163*
	(0.079)	(0.082)	(0.081)	(0.084)
Bunching range				
Lower bound	-5.16	-4.63	-5.16	-4.32
Upper bound	3.41	3.20	3.41	3.30
Observations	1,172	1,164	1,160	1,152

Note: The table reports the coefficients from Equation (6) as well as the bounds of the estimated bunching area. The bounds, defined as the intersections of the treatment and counterfactual densities, are taken from Figure 5. Standard errors clustered at the country level in parentheses. * p<10%, **p<5%, *** p<1%.

Table 5. Bunching Estimates with alternative Weighting Schemes

	Logit	Entropy 1st moment	Entropy 1-2nd moments	Local logit
Bunching area	0.189**	0.173*	0.167	0.169**
	{0.018}	{0.051}		{0.043}
Top	-0.025	-0.024	-0.001	-0.024
	{0.311}	{0.296}		{0.322}
Bottom	-0.163**	-0.149*	-0.166	-0.145*
	{0.027}	{0.057}		{0.058}
Bunching range				
Lower bound	-4.32	-4.22	-2.55	-4.11
Upper bound	3.30	3.30	4.24	3.41
Observations	1,152	1,152	1,152	1,152

Note: The table reports the coefficients from Equation (6) as well as the bounds of the estimated bunching area. The bounds, defined as the intersections of the treatment density and the counterfactual densities, are taken from Figure 6. p-values from a Wald test for whether coefficients equal zero are reported in curly brackets. The p-values are obtained from 1,000 pairs cluster bootstrap-t replications at the country level. We do not report p-values for the third column because the bootstrap algorithm fails to converge in more than 60 percent of the bootstrap iterations. See text for details. * p<10%, **p<5%, *** p<1%.

Table 6. Bunching Estimates with Clustering by Wave

|	IPW 1	IPW 2	IPW 3	IPW 4
Bunching area	0.193**	0.192**	0.195**	0.189**
	{0.022}	{0.046}	{0.019}	{0.028}
Top	-0.023	-0.024	-0.025	-0.025
	{0.300}	{0.302}	{0.195}	{0.423}
Bottom	-0.170*	-0.167	-0.171*	-0.163*
	{0.061}	{0.120}	{0.073}	{0.078}
Bunching range				
Lower bound	-5.16	-4.63	-5.16	-4.32
Upper bound	3.41	3.20	3.41	3.30
Observations	1,171	1,163	1,160	1,152

Note: The table reports the coefficients from Equation (6) as well as the bounds of the estimated bunching area. p-values from a Wald test for whether coefficients equal zero are reported in curly brackets. The p-values are obtained from 1,000 pairs cluster bootstrap-t replications at the adoption wave level. See text for details. * p<10%, **p<5%, *** p<1%.

**Figure 1. Government Balance Distributions**
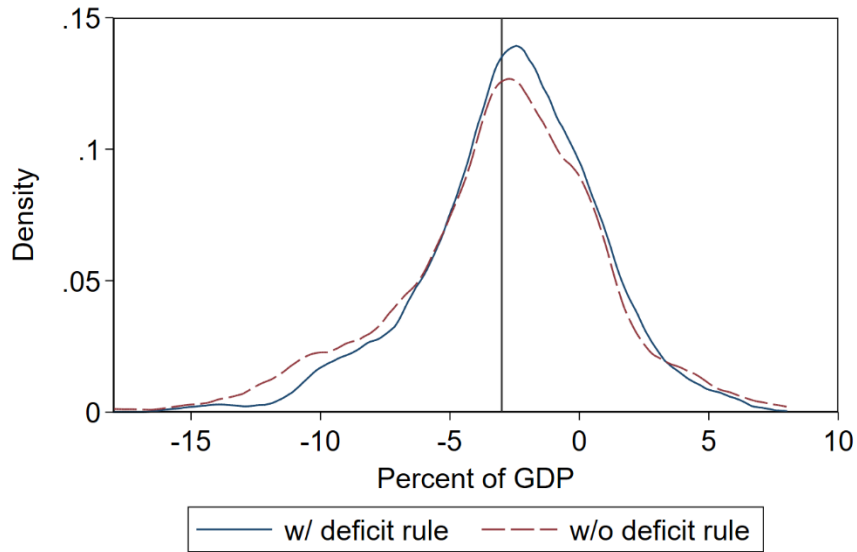
Note: The figure plots the kernel density estimates of general government balances. Raw data without with and without the fiscal rule. Epanechnikov kernel with bandwidth of 0.75 is used for both densities. The vertical line shows the FR limit at -3 percent of GDP.

Figure 2. Densities of the Estimated Propensity Scores
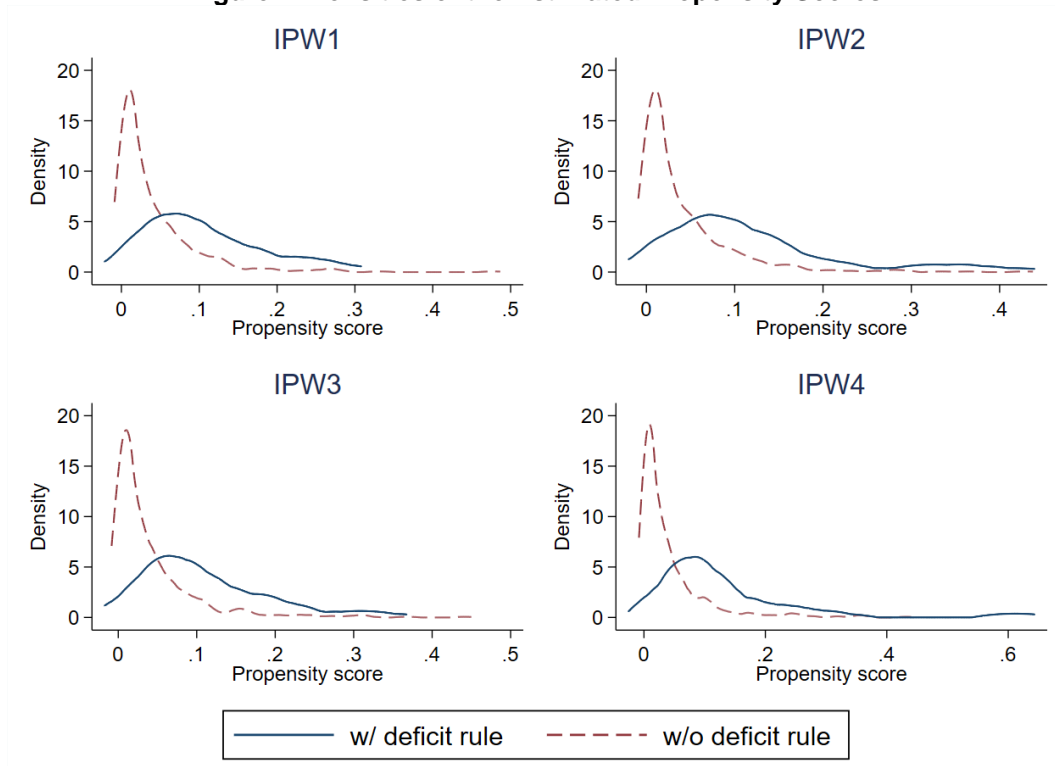
Note: The figure plots the densities of the estimated propensity to adopt the FR according to the four first stage logit models. The blue solid line plots the density for adopters, while the red dashed line plots the density for observations in the control group. Epanechnikov kernels are used with Stata's default bandwidth values.

**Figure 3. Inverse Probability Weights for the Counterfactual Group**
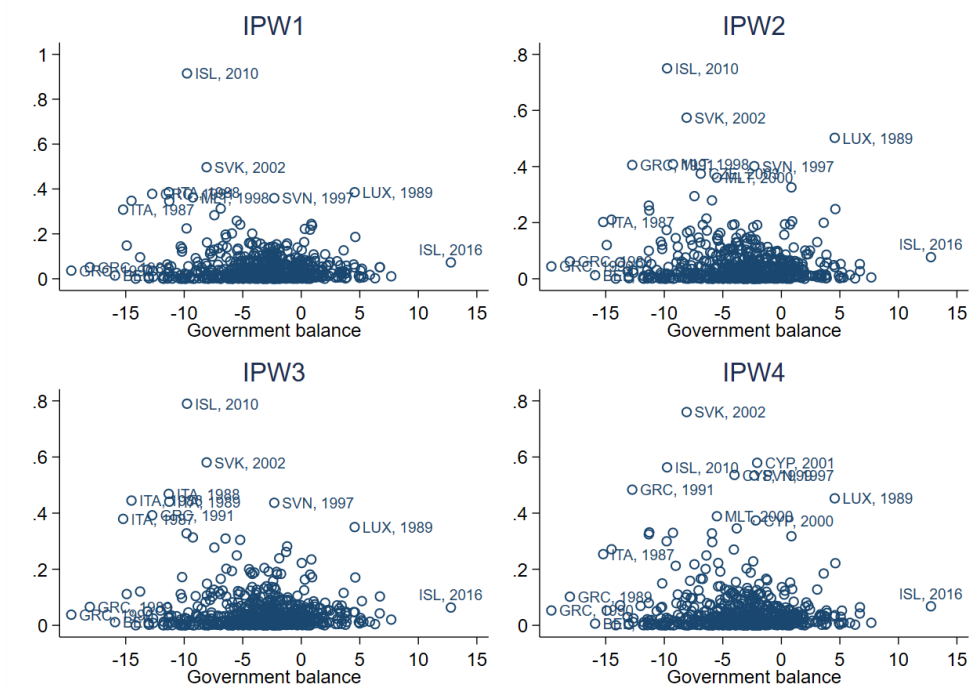
Note: The chart reports individual country-year inverse probability weights for observations in the control group calculated from Equation (3) and estimated using each of the four first stage logit models in Table 2. X-axes denote government balances in percent of GDP.

**Figure 4. Treated and Counterfactual Government Balance Distributions**
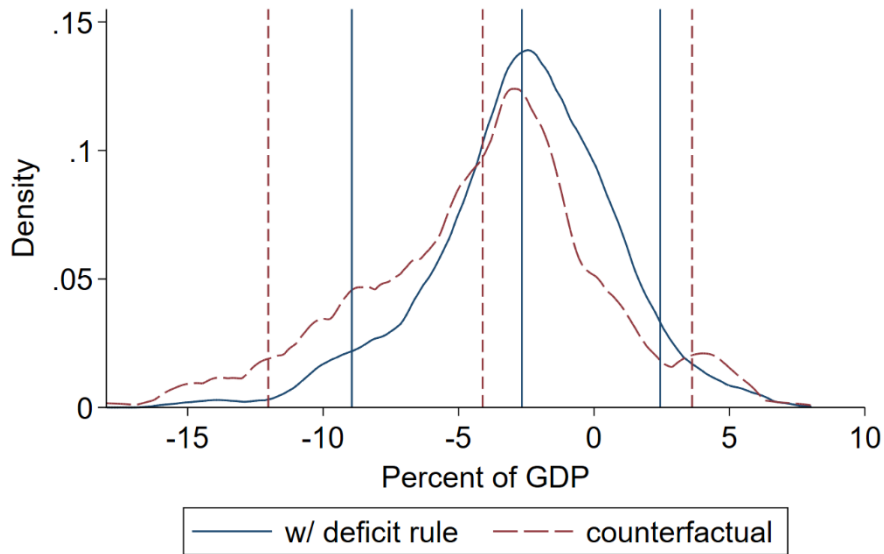
Note: The figure plots estimated kernel densities for the treated group with FR and the counterfactual group using weights estimated from Equation (5) and first stage model IPW4. The vertical lines indicate (from left to right) the 5th percentiles, means and 95th percentiles of the respective distributions. The bandwidth for the Epanechnikov kernel is 0.75. See text for details.

Figure 5. Government Balance Distributions with alternative First Stage Models
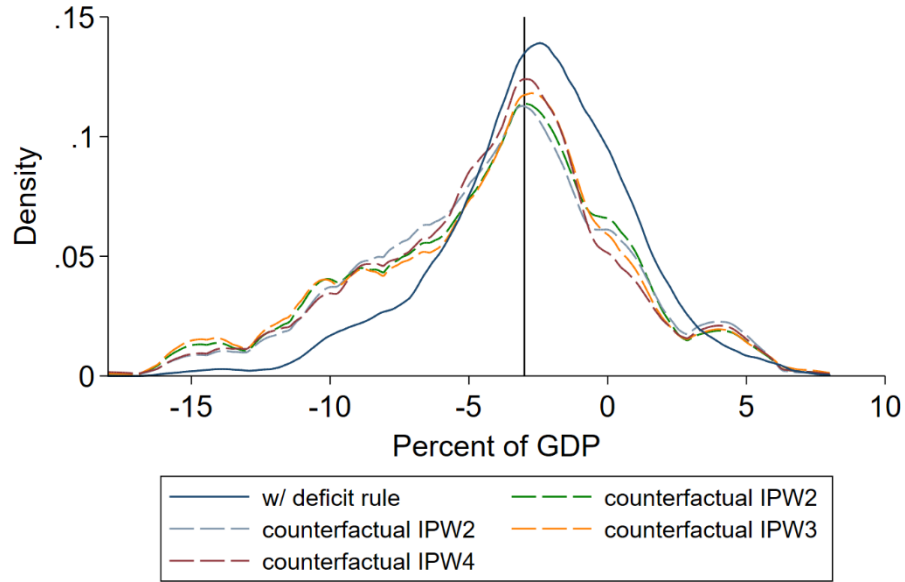

Note: The figure plots kernel densities for the treated group with FR and the counterfactual group using estimates from Equation (5) and the four IPW specifications for the first stage estimate. The bandwidth for the Epanechnikov kernel is 0.75. The vertical line shows the FR limit at -3 percent of GDP. See text for details.

**Figure 6. Government Balance Distributions with alternative Weighting Schemes**
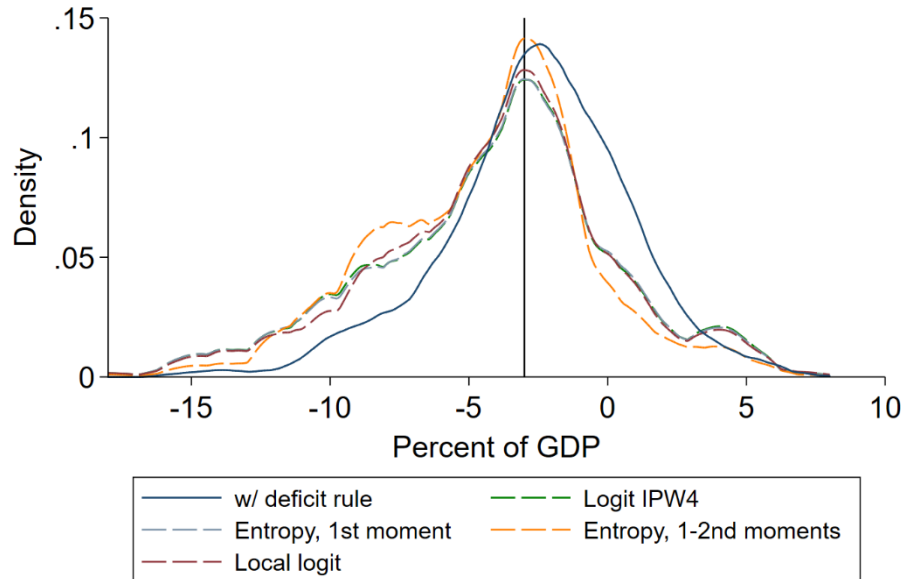
Note: The chart presents the kernel density for the counterfactual distribution obtained with different weighting procedures, using covariate specification IPW4. The bandwidth for the Epanechnikov kernel is 0.75. The vertical line shows the FR limit at -3 percent of GDP. See text for details.

**Figure 7. Average Impact of the FR by Country Based on Rank Invariance**
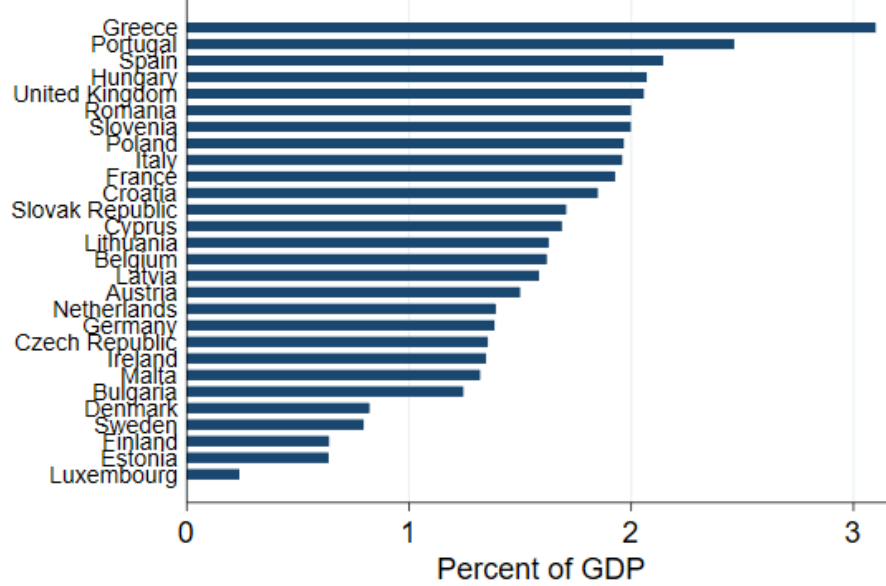
Note: The chart presents the average treatment effect by country obtained by averaging across year the country-specific results based on rank invariance. We recover individualized treatment effects by imposing rank invariance to match observations in the treatment and counterfactual groups.

Figure 8. Individualized Treatment Effects by Level of Deficit
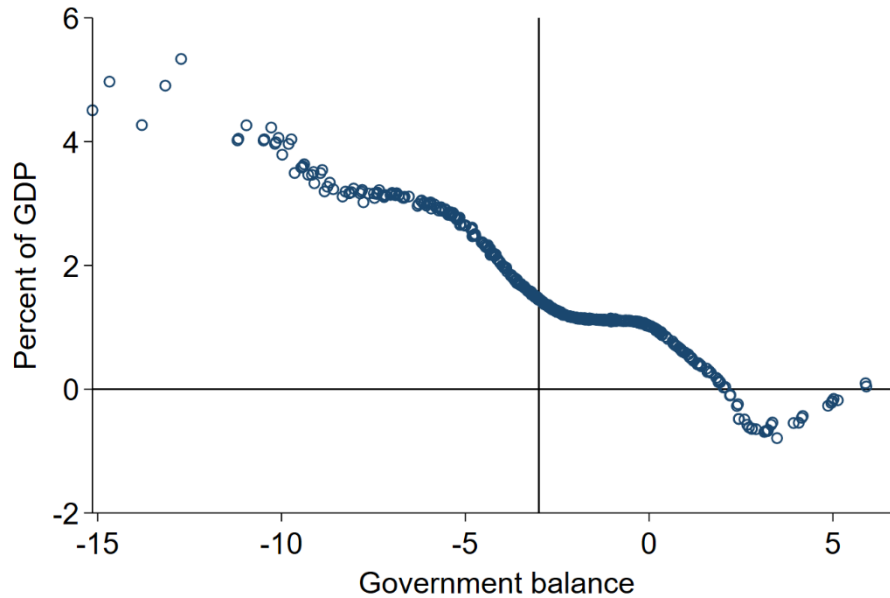
Note: The chart shows the individualized treatment effects by level of general government balance in percent of GDP. Counterfactual outcomes are derived based on the rank invariance assumption to match observations in the treatment and counterfactual groups. The vertical line shows the FR limit at -3 percent of GDP. See text for details.

Figure 9. Individualized Treatment Effects by Level of Debt
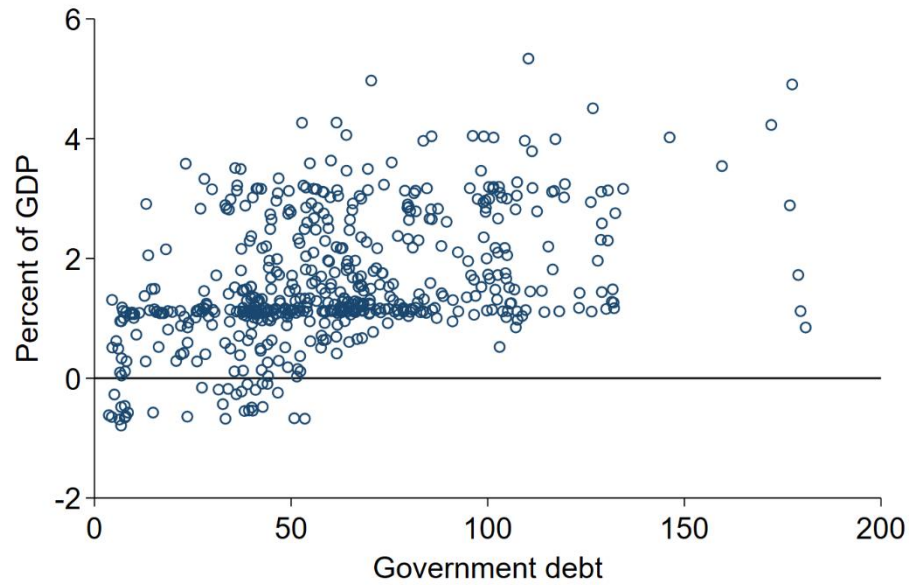
Note: The chart shows the individualized treatment effects by level of general government debt in percent of GDP. Counterfactual outcomes are derived based on the rank invariance assumption to match observations in the treatment and counterfactual groups. See text for details.

IX. APPENDIX

Table A1. Countries and 3 percent Rule's Adoption Dates

Rulers		Non-rulers	
Country	Year	Country	Year
Belgium	1992	Turkey	1987
Denmark	1992	Switzerland	1992*
France	1992	Macedonia	2004
Germany	1992	Montenegro	2008
Greece	1992	Albania	2009
Ireland	1992	Iceland	2009
Italy	1992	Serbia	2009
Luxembourg	1992		
Netherlands	1992		
Portugal	1992		
Spain	1992		
United Kingdom	1992		
Austria	1995		
Finland	1995		
Sweden	1995		
Cyprus	2004		
Czech Republic	2004		
Estonia	2004		
Hungary	2004		
Latvia	2004		
Lithuania	2004		
Malta	2004		
Poland	2004		
Slovakia	2004		
Slovenia	2004		
Bulgaria	2007		
Romania	2007		
Croatia	2013		

Note: Year for rulers indicates year of adoption of the Maastricht treaty or accession to the EU. Year for non-rulers indicates when the application for accession to the EU was formally issued. For Switzerland, we use the year when its application for EU membership was suspended following a referendum.
Source: Schaechter and others, 2012; IMF Fiscal Affairs Department 2017.

Table A2. Determinants of Fiscal Rules Adoption

Fiscal variables	Description	Sources
Deficit	Three lags of the budget balance	AMECO, World Economic Outlook and Government Finance Statistics
Government debt	One lag of general gov. consolidated gross debt	AMECO, World Economic Outlook
Age dependency ratio	Ratio of dependents-to the working-age population (%)	Eurostat
Macro variables		
GDP pc	Real GDP per capita in 2011 USD	AMECO, Penn World Tables, World Economic Outlook
GDP pc growth	Real GDP per capita growth	AMECO, Penn World Tables, World Economic Outlook
Inflation	Log change in CPI	AMECO, World Economic Outlook
Institutions		
Checks and balance	Political Constraint Index	Polity
Gov. fractionalization	Probability that two random legislators are from different parties	World Bank DPI
Federal state	Dummy variable for whether a country is a federal state	Elbadawi, Schmidt-Hebbel and Soto 2015
Trade with EU-11	Imports and exports with EU-11 countries as a share of GDP	IMF DOTS

Note: Individual government deficit series are spliced across data sources for Austria, Croatia, Ireland, Malta, Netherlands, Portugal, Spain and Sweden. Inflation and government debt-to-GDP are winsorized at 250 percent and 200 percent of GDP. This affects 13 and 2 observations, respectively.

Table A3. Average Treatment Effect Results

	DiD1	DiD2	IPW1	IPW2	IPW3	IPW4
ATET	1.028	1.774***	1.451**	1.332**	1.549**	1.449**
	(0.613)	(0.623)	(0.687)	(0.639)	(0.696)	(0.637)
Observations	1,291	1,291	1,172	1,164	1,160	1,152

Note: The first two columns show the average effect of the FR using difference-in-differences estimations without reweighting: the first column includes year fixed effects, the second column adds country fixed effects. Columns labeled IPW1 through IPW4 show the average treatment effect on the treated from Equation (2) using inverse probability weights derived from Table 2. Standard errors clustered at the country level in parentheses. * p<10%, **p<5%, *** p<1%.

Table A4. Bunching Results for Different Subperiods

	IPW 1	IPW 2	IPW 3	IPW 4
Panel A: Excluding Global Financial Crisis				
Bunching area	0.261***	0.268***	0.270***	0.264***
	(0.076)	(0.077)	(0.082)	(0.092)
Top	-0.061	-0.079	-0.060*	-0.052*
	(0.042)	(0.050)	(0.037)	(0.030)
Bottom	-0.200**	-0.190**	-0.210**	-0.212**
	(0.080)	(0.078)	(0.083)	(0.094)
Bunching range				
Lower bound	-6.20	-6.51	-6.31	-3.28
Upper bound	3.41	3.51	3.61	4.35
Observations	724	719	713	708
Panel B: Excluding 1970s and 1980s				
Bunching area	0.148	0.169*	0.124	0.151*
	(0.101)	(0.098)	(0.084)	(0.080)
Top	-0.004	-0.004	-0.004	-0.003
	(0.004)	(0.004)	(0.004)	(0.003)
Bottom	-0.144	-0.165*	-0.120	-0.148*
	(0.101)	(0.098)	(0.084)	(0.079)
Bunching range				
Lower bound	-4.53	-4.43	-2.55	-2.55
Upper bound	7.27	7.27	8.00	7.48
Observations	878	875	877	874

Note: The table reports the coefficients from Equation (6) as well as the bounds of the estimated bunching area. In Panel A, the estimation sample excludes the Global Financial Crisis from 2008 to 2012. In Panel B, the sample excludes years from 1970 to 1989. Standard errors clustered at the country level in parentheses. * p<10%, **p<5%, *** p<1%.

Figure A1. Vertical Difference between Treatment and Counterfactual Densities
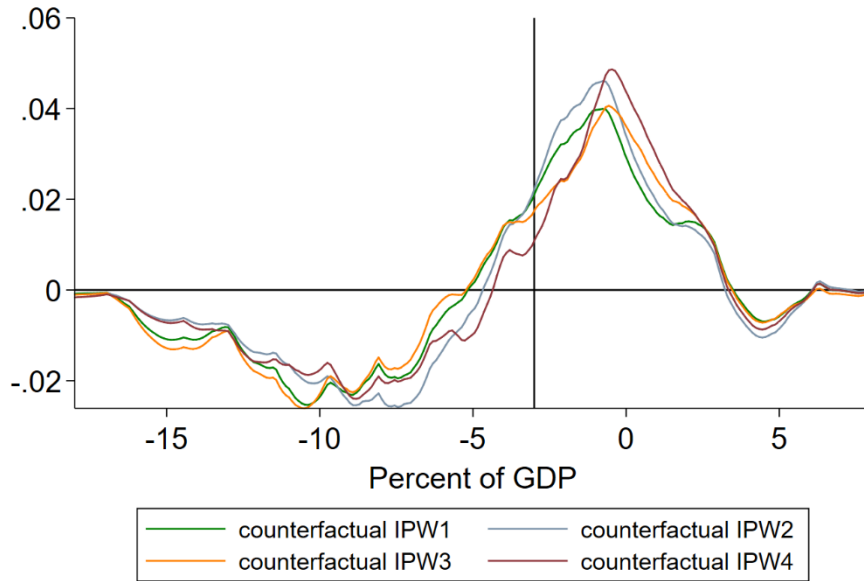
Note: The figure plots vertical difference between the kernel densities for the treated group with FR and the counterfactual group using estimates from Equation (5). The probability weights are calculated according to first stage models IPW1-IPW4 from Table 2. The bandwidth for the Epanechnikov kernel is 0.75. The vertical line shows the FR limit at -3 percent of GDP.

Figure A2. Counterfactual Time Paths for General Government Balance

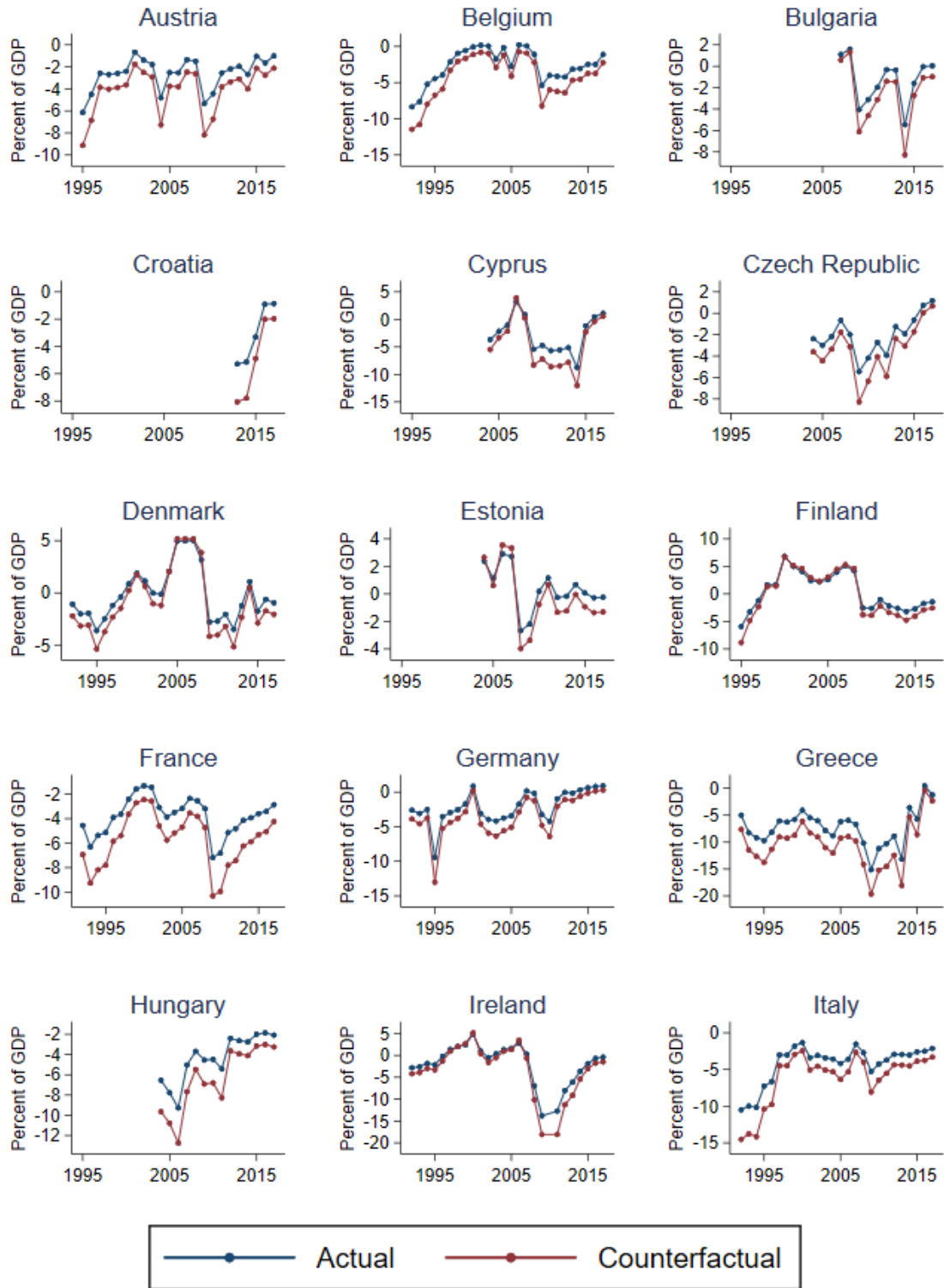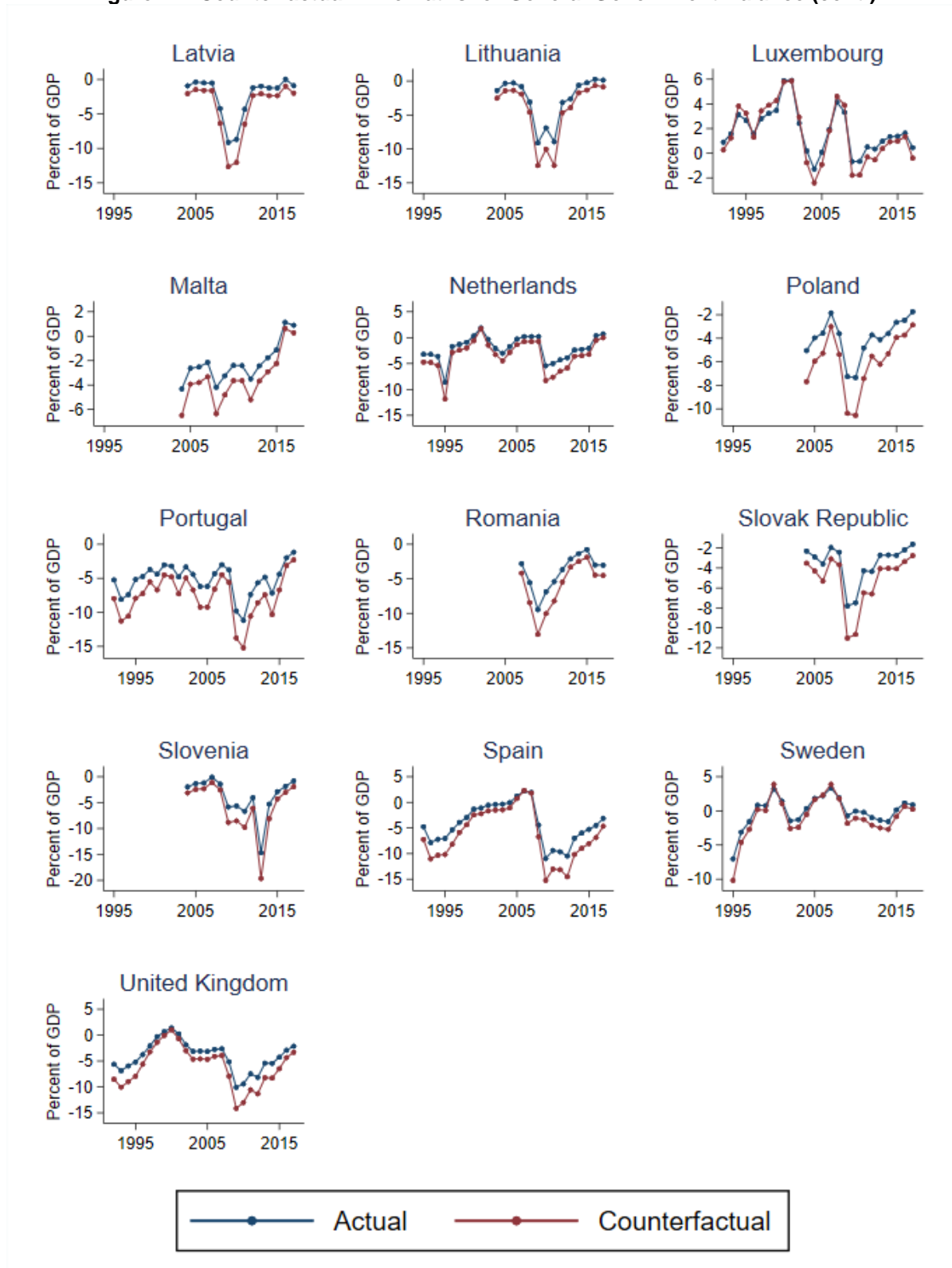
Actual — Counterfactual

Figure A2. Counterfactual Time Paths for General Government Balance (cont.)

Note: The figure plots actual and counterfactual government balances assuming rank invariance. Counterfactual time series obtained using first stage model IPW4 from Table 2.