International Oversight of Fiscal Discipline: Online Appendix

Franchino, Fabio.

European Journal of Political Research.

Table of Contents

1. Data Collection and Measurement

Countries and periods of observation

The time period is 1994-2019 for Belgium, Denmark, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, and the United Kingdom, while it starts in 1995 for Austria, Finland, and Sweden. Eight Central and Eastern European countries (Czechia, Estonia, Latvia, Lithuania, Slovakia, Slovenia, Hungary, and Poland) are included since they joined the EU in 2004, and Bulgaria and Romania from 2007 onwards. Malta, Cyprus, and Croatia, which joined in 2013, are excluded because of missing information about budgetary rules.

Table A1: Descriptive Statistics for All European Union Countries (1994-2019)

Variables in Fortunato and Loftis (2018)	mean	s.d.	min	max
Deficit (% of GDP) [a b]	2.273	3.565	-6.853	32.06
EDP [c]	0.372	0.444	0	1
EMU			0	1
Expected duration (days)	801.4	444.9	-191.6	1,819
Parties in government [d]	2.456	1.212	0	7
Budgetary constraint index [e]	0.584	0.229	0	1
Effective number of parties [d]	4.080	1.502	1.997	9.051
Government ideology [d f]	-2.542	13.20	-47.87	35.41
Caretaker time (% of year) [d]	0.0556	0.146	0	1
Unemployment rate [g]	8.556	4.135	1.900	27.50
Trade openness [h]	107.9	62.02	36.16	416.4
GDP per capita (thousands) [h]	35.86	20.20	6.476	112.0
Dependency ratio [h]	49.65	4.316	38.46	61.27
Variables in the expanded models				
Temporal effects				
Debt (% of GDP) [a b]	60.54	34.02	3.800	180.8
Spatial effects				
Average deficit$_{\sim i}$ [a b]	2.263	1.965	-0.901	6.600
Augmented tax-smoothing				
Real GDP growth [g]	2.433	3.351	-14.84	25.16
Expected inability to repay [b i j]	-0.401	3.591	-12.42	25.10

Terms of trade [k]	0.972	0.0303	0.837	1.012
Augmented common pool resource problem				
Semipresidentialism [l]			0	1
Effective number of electoral districts [m]	96.37	181.0	1	659
Number of regions [n]	3.430	6.905	0	21
Agrarian-ethic (cabinet) party share [d o]	0.0369	0.104	0	0.818
Fractionalized governments and delay				
Government range [f]	15.20	14.56	0	76.81
Alternative budgetary rules				
Delegation index [e]			0	1
Contracts index [e]			0	1
Augmented partisan effects				
Replacement risk [d f]	0.944	0.801	0.0857	5.754
Fiscal Illusion				
Total tax share [a b]	0.866	0.0387	0.750	0.999
Control				
Financial assistance [j]	0.0555	0.219	0	1

Note: N = 516, number of countries = 25. One-year lag statistics, except for the contemporaneous *Deficit*, *EMU,* and *Average deficit*$_{-l}$ and the two-year lagged *Debt*; s.d. standard deviation.

Sources: [a] AMECO database (European Commission); [b] OECD; [c] EU Council of Ministers; [d] Parlgov database (Döring and Manow 2021); [e] Hallerberg, Strauch and von Hagen (2009), Hallerberg and Yläoutinen (2008); [f] Comparative party manifesto project (Volkens et al. 2020); [g] Eurostat; [h] World Bank; [i] European Commission; [j] European Central Bank; [k] International Monetary Fund; [l] Elgie (2011), Roper (2002); [m] Golder (2005), Birch (2001), *Electoral Studies* election reports, *European Journal of Political Research* political data; [n] Hooghe, Marks, and Schakel (2008); [o] Franzese (2002).

Deficit

I use the measure of deficit employed in the excessive deficit procedure: the difference between the total expenditure and the total revenue of the general government, as defined in the European system of national and regional accounts, as a percentage of the GDP. Data come from the AMECO database and the OECD for the 1993-4 missing information.

Economic variables

Trade openness is the sum of exports and imports of goods and services measured as a share of gross domestic product. The *GDP per capita* is at constant 2010 US dollars. The *Dependency*

ratio is the ratio of people younger than 15 or older than 64 (dependents) to those of age 15-64 (the working-age population).

Expected inability to pay

The expected inability to pay and service the debt is measured as the difference between the expected real interest and real growth rates. For the expected real growth rate, I use the forecast of the real GDP growth in the autumn of year $t-1$ for year t. The expected real interest is the difference between the annual average of the long-term government bond yields and the forecast of the GDP deflator in the autumn of year $t-1$ for year t. Forecasts are available from the OECD and the European Commission, while bond yields come from the OECD and the European Central Bank.

Expected duration

The measure of expected government duration developed by Fortunato and Loftis (2018) presents an improvement over the previous practice of using indicator variables for proximity to the next elections that assumes their timing to be fixed and known *ex-ante* (Franzese 2002: 78, 179-81; Hallerberg, Strauch, and Hagen 2009: 81-3; Harrinvirta and Mattila 2001; Mink and De Haan 2006; von Hagen 2010). In my case, this is only applicable to the presidential system of Cyprus. Fortunato and Loftis (2018) use Chiba, Martin, and Stevenson's (2015) model of government duration to estimate the number of days a given cabinet expects to remain in office. They then employ a nonparametric bootstrap to produce 1,000 predicted survival times for each cabinet. Three steps are required to replicate this procedure: a) the creation of a dataset for the government survival model, b) the choice of the government survival model, and c) the measurement of expected duration.

Dataset: First, I have added to Chiba, Martin, and Stevenson's (2015) dataset of 432 cabinets of western European parliamentary democracies from the late 1940s to the late 2000s, those that have been formed up to 2019. I have then further included the following countries and time periods: Bulgaria (1991-2019), Croatia (2000-2019), Czech Republic (1992-2019), Estonia (1992-2019), Hungary (1990-2018), Latvia (1993-2019), Lithuania (1992-2019), Malta (1996-2003), Poland (1991-2015), Romania (1990-2019), Slovakia (1990-2019) and Slovenia (1990-2018). Cyprus is omitted because government survival does not depend on a legislative majority to exist. The model is, therefore, not applicable. My expanded dataset has 673 governments.

Model: Next, Chiba, Martin, and Stevenson (2015) employ a joint model of government formation and duration that accounts for biases arising from non-random selection of governments at the formation stage. However, they find no evidence of a selection problem when the type of cabinet termination is parliamentary dissolution, that is, when cabinet duration is estimated in view of the risk of early election rather than the risk of replacement.[1] Results from their naive and joint Weibull models are statistically indistinguishable. I, therefore, re-estimate their survival model only, thus significantly expediting the measurement process. The six covariates of Chiba, Martin, and Stevenson's (2015) model of government survival are minority status, ideological divisions in the coalition, returnability, effective number of legislative parties, party system polarization, and time remaining to the constitutionally mandated deadline for legislative elections. Data come from Döring and Manow (2019), Seki and Williams (2014), Volkens et al. (2019), and Wratil (2018).

[1] Cabinets can be replaced by other means, but this variable is meant to measure government's responsiveness to elections only.

Measurement: Finally, to estimate expected duration, I have employed the same nonparametric bootstrap of Fortunato and Loftis (2018: 954) to produce 1,000 survival times for each cabinet and derived mean estimates of duration from these distributions. As I explain in the main text, *Expected duration* is the number of days a cabinet expects to remain in office at the time the budget for a given year has been submitted to parliament. The measure turns negative if the date a cabinet is expected to fall, due to parliamentary dissolution, is passed at the time of presenting the budget. If this value exceeds the constitutionally mandated deadline for legislative elections, it is trimmed back to such maximum duration. This is also the value for the Cypriot governments, for the Maltese and Polish governments whose duration cannot be estimated because of missing information, and for caretaker governments that are omitted from the survival model. Although the duration of these governments (including those in office during the post-election government formation period) is plausibly short, constraints on their actions suggest that I should expect them to be less responsive to electoral pressures. Croatia, Cyprus, and Malta are eventually dropped from the analysis because of missing data on budgetary rules.

Parties in government, Effective number of parties, Caretaker time

Data on these three variables come from Döring and Manow (2021). Cabinet parties without parliamentary seats are excluded in the computation of *Parties in government*, while non-affiliated parliamentarians and single-seat parties are excluded in the computation of *Effective number of parties*. The cabinets of Monti in Italy in 2012, and of Cioloș in Romania in 2016 have been in office for a full calendar year and they had no partisan ministers. *Parties in government*, therefore, takes the value of zero.

Budgetary constraint index

Data to construct the index of Martin and Vanberg (2013) come from Hallerberg, Strauch and von Hagen (2009) and Hallerberg and Yläoutinen (2008). The index sums up the values, ranging from zero to four, of five procedural rules, and divides the result by the maximum possible sum (20). Note that the attributes *global vote on the total budget* and *amendments offsetting* are trichotomous, rather than dichotomous, in Hallerberg and Yläoutinen (2008). Data from Croatia, Cyprus, and Malta are missing.

Semipresidentialism

According to Elgie (2011: 29), twelve member states are semipresidential in the period of my inquiry (aside from the pure presidential system of Cyprus, which is excluded because of other missing data). Austria has a president-parliamentary system, while Bulgaria, Croatia, Finland, France, Ireland, Lithuania, Poland, Portugal, Romania, Slovakia, and Slovenia have premier-presidential systems. Czechia could also be considered semipresidential from 2013 with the introduction of the popular election of the president. The five countries scoring six or above in Roper's (2002: 260) measure of presidential powers are Finland, France, Poland, Portugal, and Romania. Austria, Lithuania, and Slovenia score below this threshold. In Bulgaria, Croatia, Ireland, Slovakia, and Czechia, presidents have also limited powers, so *Semipresidentialism* takes the value of zero (Croatia is anyway excluded because of other missing data), as do all parliamentary systems.

Number of regions

The *Number of regions* setting the rate of at least one major tax and, hence, scoring three or above in Hooghe, Marks, and Schakel's (2008) regional fiscal autonomy measure is the following: In Belgium, the Flemish Community, Walloon Region, and the Bruxelles-Capital

Region since 1995; sixteen Danish counties (including three municipalities with county status, but excluding Greenland and the Faroe Islands) until 2006, then five regions since 2007; Åland in Finland; fifteen ordinary regions, four special statute regions and two provinces in Italy since 1997; Azores and Madeira in Portugal since 1999; Basque County and Navarre in Spain, and the remaining fifteen autonomous communities since 1997 (excluding Ceuta and Melilla); twenty-one Swedish regional councils; Scotland in the United Kingdom since 1999. Note that the score of German *Länder* and, until 1997, of most Spanish regions is below this threshold (Marks, Hooghe, and Schakel 2008).

Effective number of electoral districts

As I say in the text, I rely on several sources, including Golder (2005), Birch (2001), the election reports in *Electoral Studies*, and the political data of the *European Journal of Political Research*, to measure the *Effective number of electoral districts*. When there is more than one electoral tier, the effective number is the seat-weighted sum of the number of districts at each tier. It is measured as follows. Let L be the size of the legislative assembly, d_i and s_i respectively the number of districts and seats at tier i for n tiers. The effective number of districts is $\sum_{i=1}^{n} d_i \frac{s_i}{L}$. For instance, in the 2017 German elections, 299 seats of the lower chamber were assigned to single-member districts and the remaining 299 seats to 16 higher-level regional districts (*Länder*). The effective number is $(299 * 299/598) + (16 * 299/598)$. In bicameral systems, we use data only from the lower chamber.

Government ideology

As I say in the text, government ideology is the seat-weighted mean of the left-right scores of cabinet parties from the comparative manifestos project (Volkens et al. 2020). For the few missing data, I have used the scores from either the previous elections, the originating party (in

case of fission), or, after being re-scaled, from Döring and Manow (2019). For the cabinets of Monti in Italy in 2012 and of Cioloș in Romania in 2016, I use the positions of the parties in the parliamentary support coalition.

Replacement risk

As I say in the main text, the measure of *Replacement risk* follows Franzese's (2002: 148-149) operationalization. First, I measure the expected probability of losing office in a year as the inverse of the actual government duration and, then, I multiply this value by the standard deviation of the government ideology across nine years, centred on the present year. In the case of more than one government a year, the hazard rate is weighted by the part of the year a government has been in office (caretaker cabinets and post-election government formation periods are disregarded). For presidential and semipresidential systems, the hazard rate is the mean of the hazard rates of the president and the cabinet. Data on actual government duration comes from Döring and Manow (2021) while, as explained above, ideology is the seat-weighted mean of the left-right scores of cabinet parties from the comparative manifestos project (Volkens et al. 2020).

Financial assistance

Information on financial assistance in the European Union is available from the European Commission at https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/financial-assistance-eu_en. In the fall of 2008, the Hungarian and the Latvian governments applied to the EU and other international organizations for financial assistance because they were experiencing difficulties in refinancing their debts. A request from Romania followed in the spring of 2009. Greece was the first eurozone country to demand assistance in

May 2010, followed by Ireland later in the year, and Portugal, Spain, and Cyprus in the next three years.

2. Replication of Fortunato and Loftis (2018)

Table A2 reports the results from adding fiscal oversight (*EDP*) and eurozone membership (*EMU*) to the models of Table 2 in Fortunato and Loftis (2018). The comparison allows us to establish if I unearth similar patterns of association.

Table A2: Extension of Fortunato and Loftis' (2018) Model of Fiscal Discipline

Lagged variables	Western Europe	European Union	Western Europe	European Union
Excessive deficit procedure	-0.3164	-0.2765	0.2163	0.1074
	(0.4885)	(0.3584)	(0.4113)	(0.3241)
Eurozone (EMU) [a]	-0.3982	-0.3358	-0.1967	0.1788
	(0.5962)	(0.4050)	(0.4709)	(0.2519)
EDP × EMU [a]	-0.2651	-0.0175	-0.5822	-0.1946
	(0.5305)	(0.3830)	(0.4171)	(0.3313)
Expected duration	-0.0007**	-0.0004*	-0.0006**	-0.0003[†]
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Parties in government	-0.1101	0.0153	0.0079	0.0354
	(0.2690)	(0.2180)	(0.2123)	(0.2205)
Budgetary constraint index	-1.0832	-0.7557	0.9099	0.7400
	(1.3872)	(1.2173)	(0.8487)	(0.8174)
Parties in government × BCI	0.3878	0.3516	0.2356	0.1956
	(0.4536)	(0.3721)	(0.3287)	(0.3251)
Effective number of parties	-0.0478	-0.0672	-0.1790[†]	-0.1418
	(0.1592)	(0.1086)	(0.0982)	(0.1230)
Government ideology	-0.0133	-0.0134	-0.0088	-0.0100
	(0.0128)	(0.0101)	(0.0063)	(0.0061)
Caretaker time	1.1537	1.2510*	1.0776	1.0829*
	(0.7594)	(0.5308)	(0.7167)	(0.4845)
Deficit	0.5429**	0.5461**	0.5916**	0.6436**
	(0.0700)	(0.0562)	(0.0697)	(0.0514)
Unemployment rate	-0.6959**	-0.4717**	-0.7364**	-0.5130**
	(0.1937)	(0.1369)	(0.1946)	(0.1361)
Trade openness	-0.0001	0.0058	0.0057	0.0146
	(0.0248)	(0.0172)	(0.0273)	(0.0187)
GDP per capita	0.3173**	0.3986**	0.2551**	0.3168**
	(0.0953)	(0.1162)	(0.0932)	(0.1126)
Dependency ratio	-0.9784*	-0.5608*	-0.7319*	0.0134

	(0.3677)	(0.2494)	(0.3671)	(0.1446)
Concurrent variables				
Unemployment rate	0.7936*	0.5664**	0.7346**	0.5293**
	(0.2729)	(0.1774)	(0.2328)	(0.1608)
Trade openness	-0.0033	-0.0186	0.0035	-0.0179
	(0.0318)	(0.0233)	(0.0314)	(0.0197)
GDP per capita	-0.3079*	-0.3588**	-0.3182**	-0.3270**
	(0.1127)	(0.1196)	(0.0944)	(0.1102)
Dependency ratio	0.8750*	0.4474†	0.6633†	-0.0906
	(0.3621)	(0.2438)	(0.3535)	(0.1422)
Intercept	6.8280*	6.6574**	6.7231**	5.3597**
	(2.6352)	(1.9010)	(2.3415)	(1.3226)
N	372	516	372	516
Countries	15	25	15	25
R^2	0.641	0.618	0.747	0.606
Effects	Fixed	Fixed	Random	Random

Notes: Dependent variable: deficit/GDP. Autoregressive distributed lag models. Robust standard errors in parentheses adjusted for country clusters. ** p<0.01, * p<0.05, † p<0.1.
[a] Since the indicator *EMU* measures also expected membership, it enters the regression in its concurrent realization. Average marginal effects computed for one year of oversight under the excessive deficit procedure (about 2.3 standard deviations).

The impact of both government duration, which is the key variable of interest of Fortunato and Loftis (2018), and supranational oversight attenuates in the random-effects specification, especially if I include all EU countries. To investigate these differences, I have run the within-between variant of Mundlak's (1978) formulation proposed by Bell and Jones (2015: 143) which allows estimating separately within- and between-effects.[2]

[2] A Hausman test indicates than within- and between-effects are not significantly correlated. This does not mean, however, that they are orthogonal and that the random-effects estimator is unbiased. Because of insufficient statistical power, this test may not be able to distinguish a small correlation from zero correlation. The result of the Hausman test provides, therefore, "neither a sufficient nor a necessary condition" for choosing between the two models (Clark and Linzer 2015: 406).

The covariates of random-effects models are made up of a 'within' component that represents the difference across periods of the same countries, and a higher-level 'between'-country component. Pooled models assume the effects of these components to be equal. If they are not, as is likely the case, a random-effects estimator could be plagued by a kind of omitted variable bias as it cannot distinguish between the two processes (Bell and Jones 2015: 137). This heterogeneity can be modelled with Mundlak's (1978) formulation. The within-between variant of this formulation proposed by Bell and Jones (2015: 143) is straightforward. The fixed part can be represented as follows

$$y_{ij} = \beta_0 + \beta_1\left(x_{ij} - \overline{x_j}\right) + \beta_2\overline{x_j}$$

where y_{ij} is the dependent variable (the deficit-GDP ratio in year i and country j), β_0 is the intercept term, x_{ij} is a series of covariates that are measured at the country-year level, and $\overline{x_j}$ are the country means of these covariates. In this formulation, within-effects, estimated by the coefficient β_1, are clearly separated from between-effects, estimated by β_2. Additionally, country-mean centering eliminates by design the collinearity between the within-component x_{ij} and the between-component $\overline{x_j}$, leading to more stable and precise estimates. Finally, if there is collinearity between the multiple $\overline{x_j}$s, these covariates can be removed without the risk of engendering bias in the estimation of the effects of within-level variables x_{ij} (Bell and Jones 2015: 142).

Tables A3 and A4 report the results from running the within-between variant of Mundlak's (1978) formulation.

Table A3: Mundlak (1978) Formulation of Pooled Model of Table A2: Western Europe

Lagged variables	Within-effects coefficients	Between-effects coefficients
Excessive deficit procedure (EDP)	-0.2889	-0.4225
	(0.4693)	(1.1809)
Eurozone (EMU) [a]	-0.1782	
	(1.1400)	
EDP × EMU [a]	-0.2742	-0.8167
	(0.5209)	(2.1191)
Expected duration	-0.0007**	-0.0006*
	(0.0002)	(0.0003)
Parties in government	0.1449	0.6552*
	(0.1307)	(0.2565)
Budgetary constraint index (BCI)	-0.2386	6.9663
	(0.4614)	(5.6388)
Parties in government × BCI	-0.0311	[b]
	(0.7011)	
Effective number of parties	-0.0914	-0.3066*
	(0.1412)	(0.1224)
Government ideology	-0.0125	-0.0271[†]
	(0.0129)	(0.0144)
Caretaker time	1.1799	-3.4514
	(0.7985)	(5.3738)
Deficit	0.5475**	0.1773
	(0.0660)	(0.6342)
Unemployment rate	-0.6905**	-3.2230
	(0.1937)	(3.2614)
Trade openness	-0.0004	0.5494
	(0.0257)	(0.4756)
GDP per capita	0.3250**	-0.1536
	(0.0975)	(0.1221)
Dependency ratio	-0.9551**	-0.4580
	(0.3565)	(0.3095)
Concurrent variables		
Unemployment rate	0.7799**	3.0613
	(0.2614)	(3.1745)
Trade openness	-0.0028	-0.5267
	(0.0329)	(0.4580)
GDP per capita	-0.3171**	[b]
	(0.1077)	
Dependency ratio	0.8637*	[b]
	(0.3403)	
Intercept		28.9508
		(19.6412)
N	372	
Countries	15	

R^2	0.758

Notes: Dependent variable: deficit/GDP. Western European EU countries, 1994-2019. Autoregressive distributed lag models with random effects. Robust standard errors in parentheses adjusted for country clusters. ** p<0.01, * p<0.05, † p<0.1. [a] Since the indicator *EMU* measures also expected membership, it enters the regression in its concurrent realization. For the same reason, the country-mean of *EMU* is not included. [b] Variable dropped because of collinearity.

Table A4: Mundlak (1978) Formulation of Pooled Model of Table A2: European Union

Lagged variables	Within-effects coefficients	Between-effects coefficients
Excessive deficit procedure (EDP)	-0.2934	0.6920
	(0.3680)	(0.8167)
Eurozone (EMU) [a]	0.2435	
	(0.4001)	
EDP × EMU [a]	0.0699	-1.0577
	(0.4175)	(0.8190)
Expected duration	-0.0004*	-0.0017**
	(0.0002)	(0.0004)
Parties in government	0.2410	-0.1551†
	(0.1424)	(0.1934)
Budgetary constraint index (BCI)	-0.0112	-0.7484
	(0.5023)	(0.6361)
Parties in government × BCI	-0.0891	0.6981**
	(0.7202)	(0.2326)
Effective number of parties	-0.0775	-0.0545
	(0.1106)	(0.0715)
Government ideology	-0.0131	-0.0027
	(0.0104)	(0.0058)
Caretaker time	1.2651*	1.0006*
	(0.5472)	(0.5067)
Deficit	0.5506**	1.0327**
	(0.0543)	(0.0883)
Unemployment rate	-0.4728**	0.0404
	(0.1384)	(0.3932)
Trade openness	0.0052	0.1904**
	(0.0178)	(0.0594)
GDP per capita	0.4023**	-0.2245
	(0.1182)	(0.1482)
Dependency ratio	-0.5195*	0.1223†
	(0.2339)	(0.1981)
Concurrent variables		
Unemployment rate	0.5628**	-0.0337
	(0.1773)	(0.3972)
Trade openness	-0.0188	-0.1885**

	(0.0237)	(0.0587)
GDP per capita	-0.3665**	0.2311
	(0.1186)	(0.1429)
Dependency ratio	0.4099	-0.1448
	(0.2235)	(0.2122)
Intercept		1.8470
		(1.3320)
N	516	
Countries	25	
R^2	0.734	

Notes: Dependent variable: deficit/GDP. EU countries, 1994-2019. Autoregressive distributed lag models with random effects. Robust standard errors in parentheses adjusted for country clusters. ** p<0.01, * p<0.05, † p<0.1. [a] Since the indicator *EMU* measures also expected membership, it enters the regression in its concurrent realization. For the same reason, the country-mean of *EMU* is not included. [b] Variable dropped because of collinearity.

Table A5 reports the marginal effects of both the within- and the between-components of my variables of interest. Reassuringly, the within-effects have a similar size to those reported in the fixed-effects model in Table A2 and they remain significant, at least for Western Europe. The between-effects are signed in the expected direction. Their size, relative to the within-effects, does not vary across the two datasets, but they fail to reach the nominal level of statistical significance. The same applies to the remaining political variables in all of the four models (in the full sample, caretaking seems to engender a deficit bias, perhaps as expected). The results appear to be similar in Fortunato and Loftis (2018), although these scholars pay less attention to the other variables. In other words, both the random-effects model and the Mundlak formulation strongly suggest that there is room for improving the random-effects specifications. Therefore, the models in Table 1 greatly extend the number of explanatory factors.

Table A5: Marginal Effects on Deficit of Oversight and Duration for Eurozone Countries

		Western European countries		
Effects	Random	Fixed		Mundlak formulation
Expected duration	0.4248**	0.4905**	within	0.4900**
	(0.1543)	(0.1367)		(0.1413)
			between	0.1188*
				(0.0483)
EDP	-0.3660	-0.5815**	within	-0.5631**
	(0.2245)	(0.1821)		(0.1994)
			between	-0.4089
				(1.049)
		European Union countries		
Expected duration	0.2508†	0.3088**	within	0.3050*
	(0.1420)	(0.1339)		(0.1368)
			between	0.3113**
				(0.0734)
EDP	-0.0872	-0.2940	within	-0.2235
	(0.2437)	(0.2182)		(0.2410)
			between	-0.1207
				(0.1990)

Notes: ** p<0.01, * p<0.05, † p<0.1. Average marginal effects on deficit/GDP, computed for a two-year or, for between-effects, six-month shortening of expected duration, and for a full year or, for between-effects, four months of oversight under the excessive deficit procedure (about 1.6 standard deviation differences). Estimates are based on the random and fixed-effects models of Table A2 and the Mundlak formulation models of Tables A3 and A4.

3. Determinants of Oversight

Model 1 of Table A6 reports the results from regressing the periods of oversight (the proportion of a given year a country has been under the excessive deficit procedure) on the lagged and average levels of deficit to account for temporal and spatial effects of deficits as well as on the lagged levels of debt. Model 2 adds economic and political factors to test whether right-leaning governments may choose to be subject to oversight to avoid the blame for fiscal consolidation. There is no evidence that government expected duration or ideology, across any level of replacement risk, is significantly related to the time under scrutiny. In further analysis, I also

find no evidence of a significant relationship between oversight periods and pro-anti EU government positions (-0.0072, *p*-value > 0.074) or national public support for the EU (0.0016, *p*-value > 0.092).

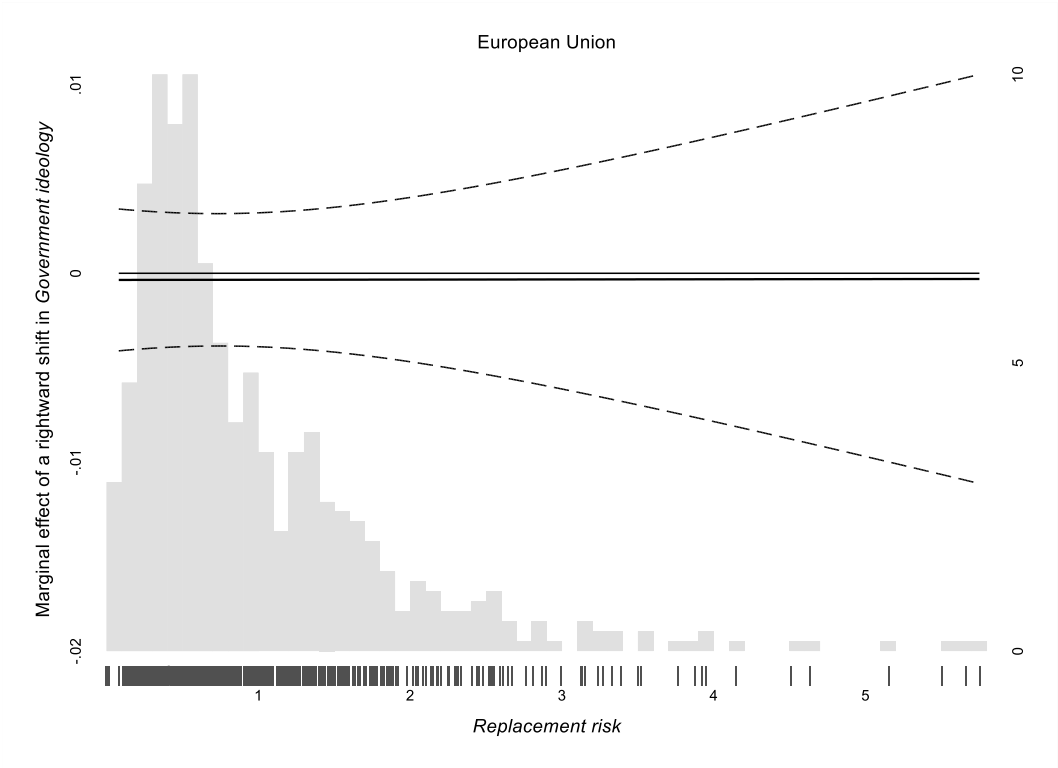
Table A6: Determinants of International Oversight

	Spatial and temporal effects	Economic and political conditions
Lagged variables	(1)	(2)
Deficit	0.0280**	0.0258**
	(0.0088)	(0.0068)
Deficit L2 [a]	0.0304**	0.0308**
	(0.0063)	(0.0054)
Average deficit$_{\sim i}$	0.0612**	0.0728**
	(0.0119)	(0.0120)
Debt	0.0039	0.0005
	(0.0036)	(0.0039)
Debt L2 [a]	-0.0022	0.0014
	(0.0035)	(0.0038)
Government ideology		-0.0003
		(0.0020)
Expected duration		-0.0000
		(0.0000)
Government ideology × Expected duration		-0.0000
		(0.0000)
Replacement risk		-0.0134
		(0.0225)
Government ideology × Replacement risk		0.0000
		(0.0010)
Unemployment rate		0.0007
		(0.0180)
Real GDP growth rate		0.0032
		(0.0073)
Expected inability to pay		-0.0055
		(0.0055)
Trade openness		0.0101
		(0.0115)
Terms of trade (ToT)		0.9023
		(2.1162)
Trade openness × ToT		-0.0109
		(0.0122)
GDP per capita		-0.0116
		(0.0153)
Dependency ratio		0.0372**

		(0.0054)
Concurrent variables		
EMU	-0.0528	-0.0936*
	(0.0367)	(0.0466)
ln(GDP)	0.0118	0.0164
	(0.0130)	(0.0129)
Unemployment rate		0.0074
		(0.0178)
Real GDP growth rate		-0.0040
		(0.0089)
Expected inability to pay		0.0108
		(0.0087)
Trade openness		0.0087
		(0.0094)
Terms of trade (ToT)		2.9358
		(2.0707)
Trade openness × ToT		-0.0083
		(0.0108)
GDP per capita		0.0112
		(0.0152)
Dependency ratio		-0.0461**
		(0.0059)
Intercept	-0.0892	-3.4610**
	(0.0825)	(1.3130)
N	580	580
Countries	28	28
R^2	0.527	0.617

Notes: Dependent variable: Excessive Deficit Procedure (EDP). Models with random effects. Robust standard errors in parentheses adjusted for country clusters. ** p<0.01, * p<0.05, † p<0.1. [a] These variables enter the regression at t-2.

Figure A1: Marginal Effects of a Rightward Shift in Ideology across Replacement Risk



Note: Histograms underlaid. Expected duration is set at its mean.

4. Treatment Effects: Extension to the Pre-Maastricht Period

Dataset: I have extended the dataset to the period before the Treaty of Maastricht and to the countries included in Fortunato and Loft (2018): Austria (1972-1994), Belgium (1972-1993), Denmark (1973-1993), Finland (1971-1994), France (1981-1993), Germany (1971-1993), Greece (1981-1993), Ireland (1971-1993), Italy (1971-1993), the Netherlands (1971-1993), Portugal (1982-1993), Spain (1980-1993) and Sweden (1971-1994). Luxembourg is excluded because it is missing in Franzese (2002). Deficit data of the United Kingdom are missing in Fortunato and Loft (2018), so I used the data from the UK Office of Budget Responsibility for 1971-1993.

Model: The dependent variable (deficit as a percentage of GDP) is from Fortunato and Loft (2018) (i.e. OECD estimates) for the pre-Maastricht period and the AMECO database for the post-Maastricht period. For the country-years the two datasets overlap, the correlation is very high (0.937, *p*-value > 0.000) but they are not identical. The data of the covariates for the pre-Maastricht period come from Fortunato and Loft (2018) and Franzese (2002), with small differences. Data on *GDP per capita* comes from the World Bank for the entire period. I use the general government gross debt as GDP share. Portugal is set as an effective semipresidential system. To compute the pre-Maastricht spatial effects, I use the countries in Fortunato and Loft (2018). The model is specified like the full model of the last column of Table 1 without the interaction between the variables *EMU* and *Excessive deficit procedure*.

Table A7: Auxiliary Models for the Difference-in-Differences Estimates of Treatment Effects

Lagged variables	Western Europe	Europe
Excessive deficit procedure	-0.5471*	-0.5508*
	(0.2534)	(0.2341)
Eurozone (EMU) [a]	-0.0074	-0.0330
	(0.3243)	(0.2816)
Expected duration	-0.0005**	-0.0004*
	(0.0001)	(0.0001)
Parties in government	0.4263	0.5525*
	(0.2409)	(0.2380)
Budgetary constraint index	0.4403	0.5975
	(0.5086)	(0.5444)
Parties in government × BCI	-0.3761	-0.3628
	(0.2301)	(0.2349)
Parties in government × Debt	-0.0007	-0.0017
	(0.0018)	(0.0017)
Government range	0.0085	0.0103
	(0.0197)	(0.0151)
Government range × Debt [b]	-0.0002	-0.0002
	(0.0001)	(0.0001)
Effective number of parties	-0.2026	-0.2044
	(0.1167)	(0.1116)
Semipresidentialism	[c]	[c]

Ethnic-agrarian party share	0.6507	0.7997
	(0.8013)	(0.6240)
Number of regions	0.0029*	0.0028*
	(0.0013)	(0.0012)
Eff. num. of electoral districts	-0.0509**	-0.0558**
	(0.0109)	(0.0095)
Government ideology	-0.0343	-0.0309
	(0.0190)	(0.0156)
Replacement risk (RR)	-0.1004	-0.0342
	(0.1846)	(0.1542)
Government ideology × RR	0.0130	0.0117
	(0.0131)	(0.0106)
Total tax share	0.0976	1.1428
	(5.7745)	(5.1090)
Caretaker time	0.4425	0.5801
	(0.6842)	(0.5375)
Financial assistance	-2.0606	-1.8628*
	(1.0826)	(0.7644)
Deficit	0.5803**	0.5637**
	(0.0641)	(0.0584)
Deficit L2 [b]	0.0650	0.0734
	(0.0471)	(0.0447)
Debt [b]	-0.0143	-0.0124*
	(0.0068)	(0.0051)
Unemployment rate	-0.1958	-0.1890
	(0.1027)	(0.1017)
Real GDP growth rate	-0.0874	-0.0258
	(0.0492)	(0.0464)
Expected inability to pay	-0.0308	-0.0115
	(0.0812)	(0.0508)
Trade openness	-0.0866	-0.0095
	(0.0731)	(0.0571)
Terms of trade (ToT)	1.8169	6.3968
	(5.6271)	(4.9277)
Trade openness × ToT	0.0929	0.0162
	(0.0710)	(0.0556)
GDP per capita	-0.0235	0.1475
	(0.1479)	(0.1247)
Dependency ratio	0.0569	0.0557*
	(0.0274)	(0.0242)
Concurrent variables		
Average deficit$_{\sim i}$	0.1634**	0.1860**
	(0.0539)	(0.0453)
Unemployment rate	0.2770*	0.2787*
	(0.1039)	(0.1105)
Real GDP growth rate	-0.1862*	-0.1101*
	(0.0731)	(0.0485)

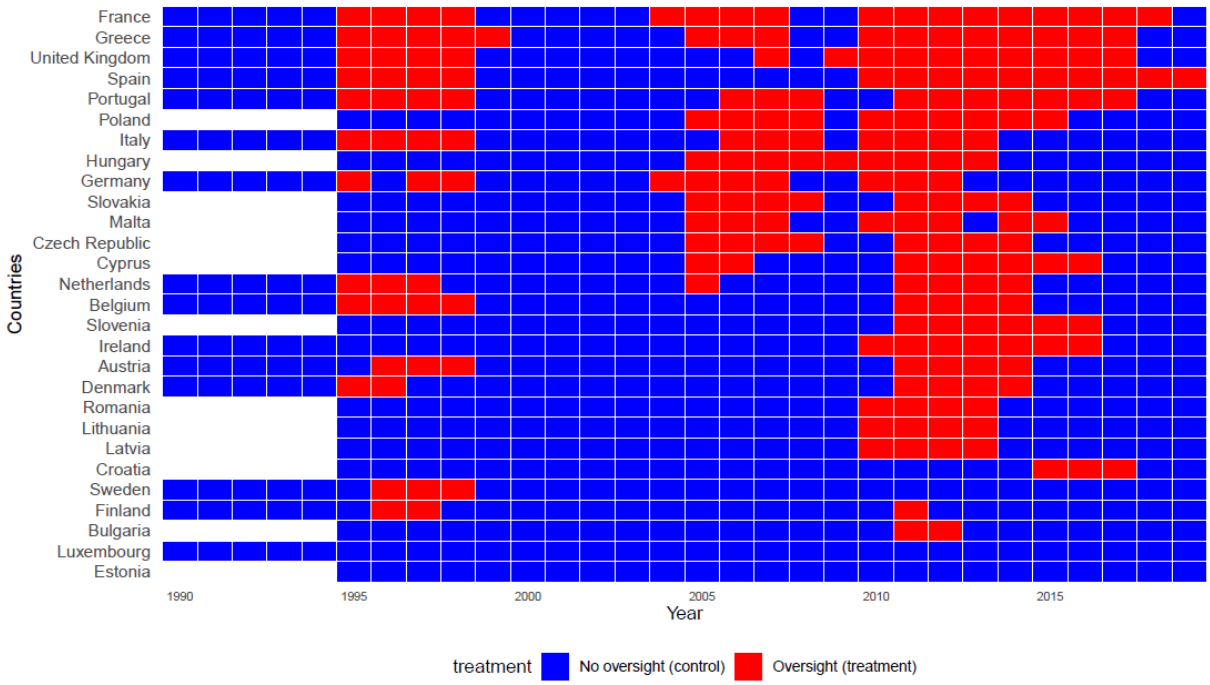
Expected inability to pay	0.0632	0.0300
	(0.0785)	(0.0633)
Trade openness	0.0659	-0.0088
	(0.0474)	(0.0496)
Terms of trade (ToT)	-0.4836	-4.8401
	(4.7082)	(4.8470)
Trade openness × ToT	-0.0714	-0.0008
	(0.0431)	(0.0464)
GDP per capita	0.0155	-0.1480
	(0.1418)	(0.1189)
Dependency ratio	-0.0471	-0.0451
	(0.0339)	(0.0278)
Intercept	0.4723	-1.4442
	(4.8073)	(3.9649)
N	658	804
Countries	15	25
R^2	0.746	0.724

Notes: Dependent variable: deficit/GDP. Country fixed effects. Robust standard errors in parentheses adjusted for country clusters. ** p<0.01, * p<0.05. [a] Since the indicator *EMU* measures also expected membership, it enters the regression in its concurrent realization. [b] These variables enter the regression at t-2. [c] Omitted variable because of collinearity.

5. Treatment Effects: Exact Matching on Treatment History

Figure A2 illustrates the variation of oversight across countries and time periods.

Figure A2: Treatment Variation Plot



Notes: The plot covers the years for which we have excessive deficit data from the AMECO database and the OECD for the 1990-4 missing information. For some countries, the data cover also the years preceding accession to the EU. A country is treated if it has been under oversight for an entire year, including the year when the Council has terminated the procedure.

The histograms in Figure A3 illustrate the number of control countries for each treated country (i.e., under oversight) that match exactly on treatment histories of one to three years. The pink line represents the number of unmatched treated countries. The choice of duration of the treatment history faces a bias-variance trade-off (Imai, Kim, and Wang 2022: 7). All treated countries are matched on a one-year treatment history but carryover effects are controlled for only very partially, whereas a three-year treatment history allows for better control but nine treated units are not matched. A choice of two years is a reasonable balance between controlling partially for carryover effects and having enough variance since only one treated country is not matched.

Figure A3: Frequency Distribution of the Number of Matched Control Units

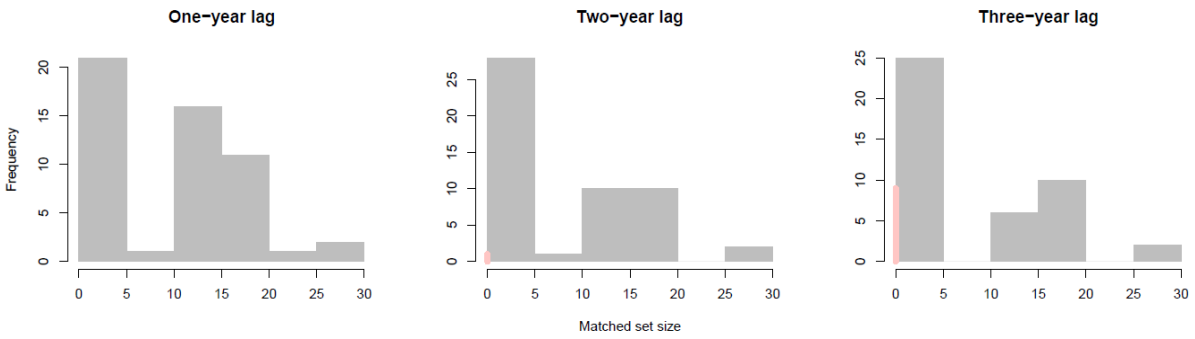
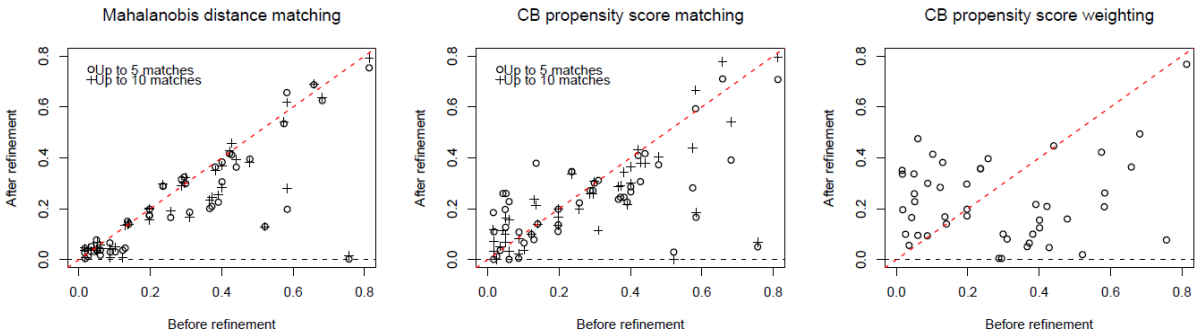


Figure A4 shows the performance in terms of improved covariance balance of three refinement methods that employ all the lagged covariates of the expanded model of Table 1 (including expected eurozone membership). Matching-based methods include up to five and up to ten matches. A dot below the 45-degree lines implies an improvement of the standardized mean balance since the mean difference after refinement is lower than the mean difference before refinement. Methods based on covariance balancing propensity score display both significant

improvement and worsening.[3] Overall, Mahalanobis distance matching produces the best improvement in covariance balance.

Figure A4: Improved Covariance Balance due to Refinement of Matched Sets

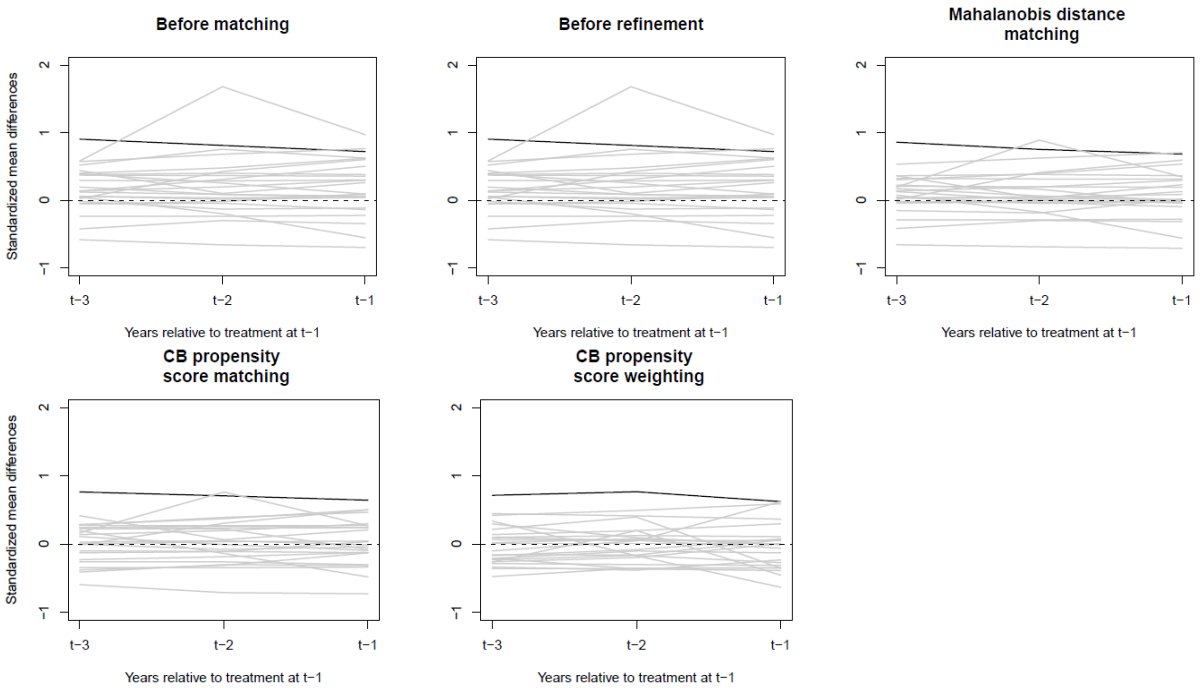


Notes: Absolute value of standardized mean difference before and after refinement. Diagonal covariance matrix employed. Two-year lag treatment history matching. CB = covariance balancing.

Figure A5 shows the improvement of covariance balance due to matching over the two years prior to the administration of the treatment at $t-1$. Refinement produces some improvement without appreciable differences across the different methods. Notably, the standardized mean differences for the lagged deficit, represented by the black lines, stay relatively constant over the pre-treatment period, lending support for the appropriateness of the parallel trend assumption.

[3] Other propensity score methods produce numerically 0 or 1 fitted probabilities.

Figure A5: Improved Covariance Balance due to Matching over the Pre-treatment Period

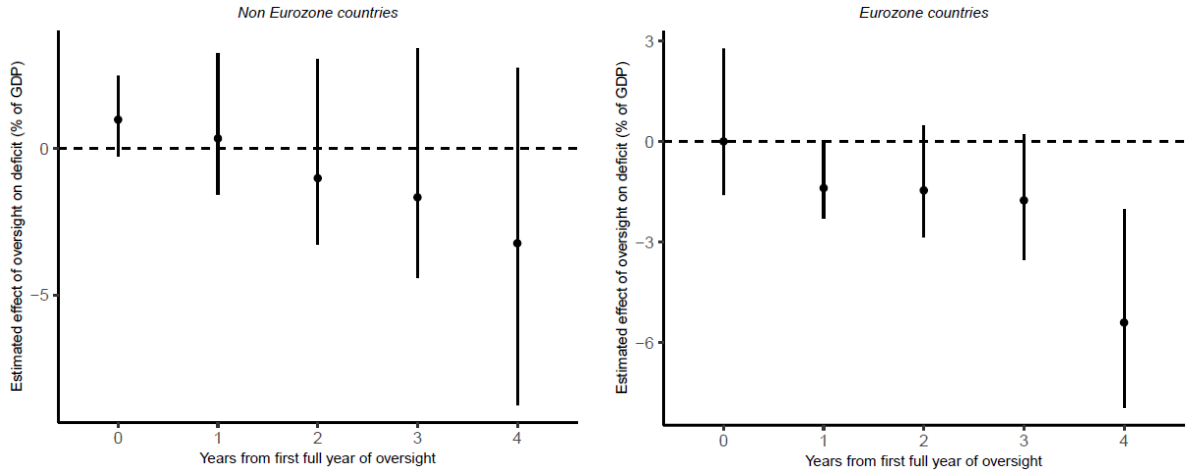


Notes: Standardized mean difference over the pre-treatment period of two years. The black lines plot the covariate balance of the lagged outcome variable, the grey lines that of the time-varying covariates. Diagonal covariance matrix employed. Two-year lag treatment history matching. CB = covariance balancing. Up to five matches are used in the matching methods.

Matching Estimates without Treatment Reversal

If I forbid treatment reversal, that is, I match only treated countries that have been continuously under oversight in the lead window, the procedure leads to deficits in the fourth and fifth year that are on average 1.75% and 5.40% of GDP lower those of untreated countries (see Figure A6). However, balance diagnostics requires dropping some covariates to assess the improvement across refinement methods.

Figure A6: Matching Estimates without Treatment Reversal

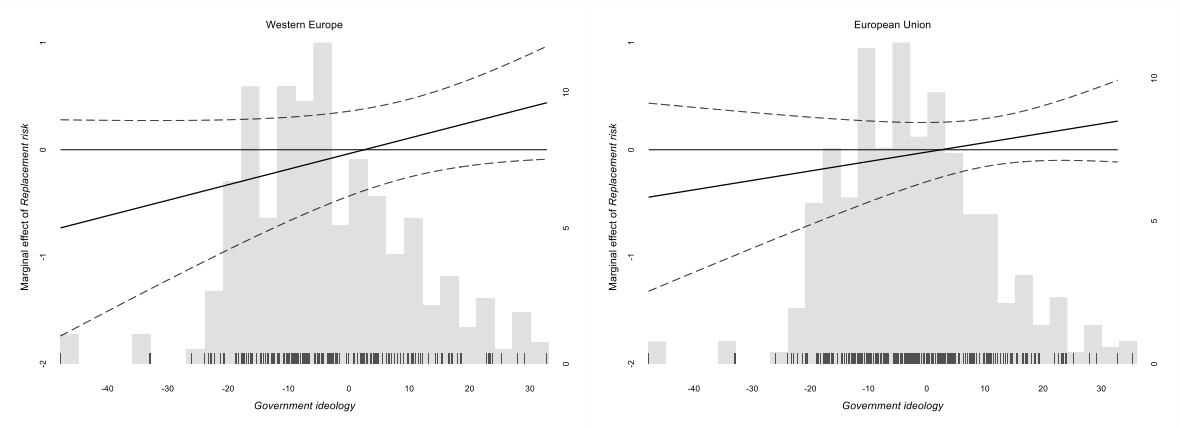


Note: The estimates are based on the exact matching of the treatment histories during the two years preceding the first full year of oversight. Matched sets are refined with Mahalanobis distance matching up to five matches, using the lagged covariates of the expanded model of Table 1. Treatment reversal is forbidden. Vertical bars are the 95% confidence intervals from a bootstrapping procedure of 500 iterations.

6. More Results from Table 1

Replacement Risk and Government Ideology

Figure A7: Marginal Effect of Replacement Risk across Government Ideology



Note: Histograms underlaid.

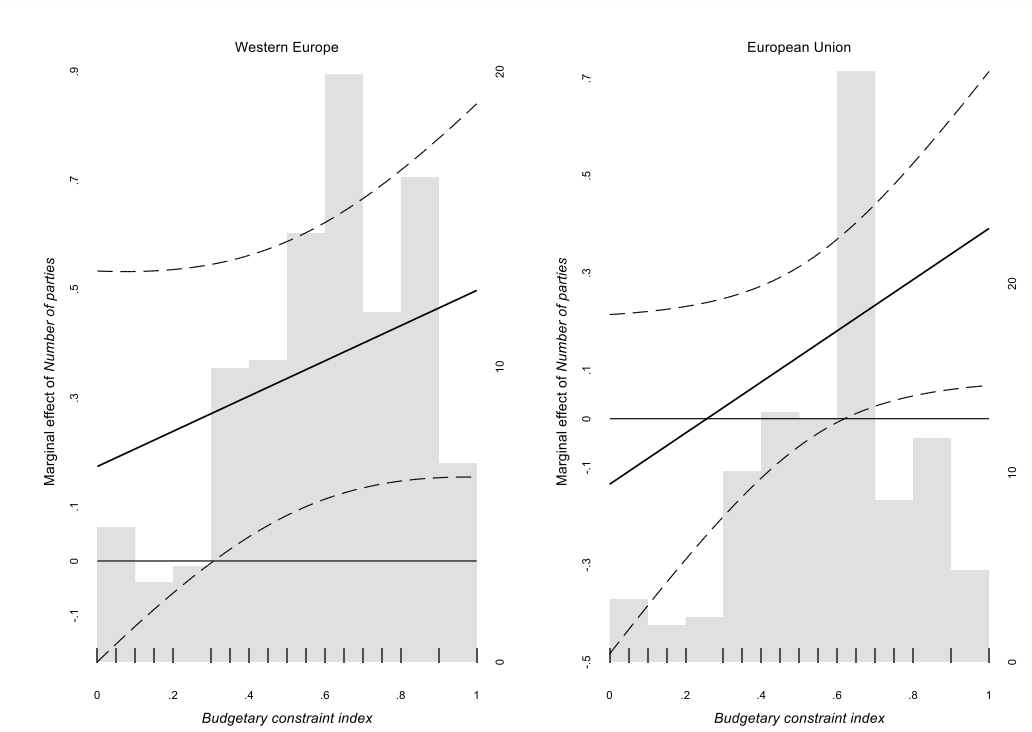
Government Fractionalization and Budgetary Rules

Contrary to expectations, the historical record indicates that, if the indebtedness is high, government fractionalization is associated with *higher* deficits when the rules are *more* constraining (see Figure A8).[4] In Western Europe, the curve in the left panel of Figure A8 shifts upward as the debt-to-GDP ratio increases. The marginal effect, therefore, becomes positive and significant at even lower values of debt ratios and budgetary constraints. In the whole European Union, the curve in the right panel of Figure A8 shifts downward as the debt-to-GDP

[4] Since fractionalization interacts with both the budgetary constraint index and the debt, the analysis of the marginal effect of fractionalization across budgetary contraints is carried out for given levels of debt. Similarly, in the main text, the analysis of the marginal effect of fractionalization across levels of debt is carried out for given values of the budget constraint index.

ratio increases and the confidence intervals expand as the debt-to-GDP ratio decreases. In both circumstances, the marginal effect is below the standard level of significance, so it is positive and significant only at mid-high values of debt ratios and high values of budgetary constraints. These results suggest that fractionalized governments that are confronting high public debt and find it hard to fiscally consolidate may be inclined to adopt more constraining rules. Indeed, Hallerberg, Strauch, and von Hagen (2009: 123-134) offer some evidence on how cabinet heterogeneity affects the types of fiscal rules adopted.

Figure A8: Marginal Effect of Executive Fractionalization across Budgetary Rules



Note: Derived from the models of Table 1. The lagged debt-to-GDP ratio is set at a standard deviation above its mean. Histogram underlaid.

Effective Semipresidentialism and Electoral Districting

Semipresidentialism in Western Europe appears to exert the expected downward influence on deficits. Effective semipresidential systems (Finland, France, and Portugal) run on average

deficits that are 1.25% of GDP lower than figuratively semipresidential or parliamentary systems, but the effect does not extend to the full sample where I include Poland and Romania. Countries with the effective number of electoral districts a standard deviation above the mean, approximately the difference between the mixed electoral systems of Italy (when it had one) and Germany, are associated with deficits that are 0.61% of GDP higher. This impact is however attenuated and fails to reach standard significance in the full sample. Anyway, given their slowly changing nature, effective semipresidentialism and electoral districting seem to be macroinstitutional features that account for some cross-country differences, especially in Western Europe. Lastly, the effect of federalism is opposite to the expectations, as in Franzese (2002: 173-4).

Fiscal Illusion and Caretaking Time

At least in Western Europe, greater simplicity in the tax structure appears to discourage opportunistic behaviour that exploits voters' fiscal illusion. A standard deviation increase in the share of government taxes over revenues, about 3.5% - a bit less than the average difference between Spain and Portugal -, is associated with a reduction in the deficit by 0.49% of GDP. Lastly, periods of caretaking and government formation have a small but significant impact on deficits. If they carry on two months longer than the mean of twenty days, deficits increase by 0.20% of GDP in the full sample.

7. Independent Fiscal Institutions

From the onset of the financial crisis in December 2007, independent fiscal institutions have started to proliferate across Europe (Davoodi et al. 2022: 11). Their establishment eventually became compulsory for Eurozone countries with the adoption of Regulation 473/2013. Their influence is undertheorized and no empirical study has shown that they directly affect budget deficits.[5] Nevertheless, to address a concern of a reviewer, I have added to the models in Tables 1 and 2 an indicator variable (*Independent Fiscal Institution*) that takes the value of 1 if such a body is operative during budget drafting. The data comes from the OECD Independent Fiscal Institutions Database[6] and Georgescu and Căpraru (2020). I use the earliest year in which any of such institutions began operating in a country. Table A8 reports the coefficients of this variable. The presence of these bodies is not associated with lower deficits and our main results hold.

[5] There is some evidence that they may induce greater adherence to numerical budget balance and expenditure rules, but only if the two types of rule are pooled in the same model (Beetsma et al. 2019).

[6] Accessible at https://www.oecd.org/gov/budgeting/OECD-Independent-Fiscal-Institutions-Database.xlsx .

Table A8: Inclusion of Independent Fiscal Institutions to the Models in Tables 1 and 2

	Table 1		Table 2	
	WE	EU	WE	Europe
Independent Fiscal Institution	0.1912	0.2319	0.7658†	0.4435
	(0.2460)	(0.1863)	(0.3652)	(0.3021)
Marginal effects of EDP	-0.479*	-0.5498*		
	(0.2419)	(0.2403)		
Average EDP treatment effects on the treated			-0.4950*	-0.4909*
			(0.2299)	(0.2311)

Notes: WE: Western Europe; EU: European Union. Robust standard errors in parentheses adjusted for country clusters. * p<0.05, † p<0.1.

8. References

Beetsma, Roel, Xavier Debrun, Xiangming Fang, Young Kim, Victor Lledó, Samba Mbaye, and Xiaoxiao Zhang. 2019. "Independent Fiscal Councils: Recent Trends and Performance." *Fiscal Frameworks in Europe* 57 (March): 53–69.

Bell, Andrew, and Kelvyn Jones. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3 (1): 133–53.

Birch, Sarah. 2001. "Electoral Systems and Party Systems in Europe East and West." *Perspectives on European Politics and Society* 2 (3): 355–77.

Chiba, Daina, Lanny W. Martin, and Randolph T. Stevenson. 2015. "A Copula Approach to the Problem of Selection Bias in Models of Government Survival." *Political Analysis* 23 (1): 42–58.

Clark, Tom S., and Drew A. Linzer. 2015. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3 (2): 399–408.

Davoodi, R. Hamid, Paul Elger, Alexandra Fotiou, Daniel Garcia-Macia, Xuehui Han, Andresa Lagerborg, Waikei R. Lam, and Paulo A. Medas. 2022. "Fiscal Rules and Fiscal Councils: Recent Trends and Performance during the COVID-19 Pandemic." *IMF Working Papers* 22 (011): 41.

Döring, Holger, and Philip Manow. 2019. "Parliaments and Governments Database (ParlGov): Information on Parties, Elections and Cabinets in Modern Democracies." Development Version. Technical report.

———. 2021. "Parliaments and Governments Database (ParlGov): Information on Parties, Elections and Cabinets in Modern Democracies." Development Version. Technical report.

Elgie, Robert. 2011. *Semi-Presidentialism: Sub-Types and Democratic Performance*. Oxford: Oxford University Press.

Fortunato, David, and Matt W. Loftis. 2018. "Cabinet Durability and Fiscal Discipline." *American Political Science Review* 112 (4): 939–53.

Franzese, Robert J. Jr. 2002. *Macroeconomic Policies in Developed Democracies*. Cambridge: Cambridge University Press.

Georgescu, George, and Bogdan Căpraru. 2020. “Fiscal Councils in European Union. A Short Retrospective Review and Current Challenges in Terms of Functionality and Effectiveness.” Working Paper 1. Bucarest: Romanian Fiscal Council.

Golder, Matt. 2005. “Democratic Electoral Systems around the World, 1946-2000.” *Electoral Studies* 24 (1): 103–21.

Hagen, Jürgen von. 2010. “Sticking to Fiscal Plans: The Role of Institutions.” *Public Choice* 144 (3/4): 487–503.

Hallerberg, Mark, Rolf Rainer Strauch, and Jürgen von Hagen. 2009. *Fiscal Governance in Europe*. Cambridge: Cambridge University Press.

Hallerberg, Mark, and Sami Yläoutinen. 2008. “Fiscal Governance in Central and Eastern Europe Before and After European Union Accession: What Role Europeanisation?” Working paper no. 37. Berlin: Hertie School of Governance.

Harrinvirta, Markku, and Mikko Mattila. 2001. “The Hard Business of Balancing Budgets: A Study of Public Finances in Seventeen OECD Countries.” *British Journal of Political Science* 31 (3): 497–521.

Imai, Kosuke, In Song Kim, and Erik H. Wang. 2022. “Matching Methods for Causal Inference with Time-Series Cross-Sectional Data.” *American Journal of Political Science*.

Marks, Gary, Liesbet Hooghe, and Arjan H Schakel. 2008. “Patterns of Regional Authority.” *Regional & Federal Studies* 18 (2): 167–81.

Martin, Lanny W., and Georg Vanberg. 2013. “Multiparty Government, Fiscal Institutions, and Public Spending.” *The Journal of Politics* 75 (4): 953–67.

Mink, Mark, and Jakob De Haan. 2006. “Are There Political Budget Cycles in the Euro Area?” *European Union Politics* 7 (2): 191–211.

Mundlak, Yair. 1978. “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46 (1): 69–85.

Roper, Steven D. 2002. “Are All Semipresidential Regimes the Same? A Comparison of Premier-Presidential Regimes.” *Comparative Politics* 34 (3): 253–72.

Seki, Katsunori, and Laron K. Williams. 2014. “Updating the Party Government Data Set.” *Electoral Studies* 34 (June): 270–79.

Volkens, Andrea, Werner Krause, Pola Lehmann, Theres Matthieß, Sven Regel, and Bernhard Weßels. 2019. “The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR).” Technical report Version 2019b. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Volkens, Andrea, Werner Krause, Pola Lehmann, Theres Matthieß, Sven Regel, Bernhard Weßels, Tobias Burst, Nicolas Merz, and Lisa Zehnter. 2020. “The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR).” Technical report Version 2020a. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Wratil, Christopher. 2018. “Modes of Government Responsiveness in the European Union: Evidence from Council Negotiation Positions.” *European Union Politics* 19 (1): 52–74.