Received: 6 December 2023 | Accepted: 19 December 2023

DOI: 10.1002/for.3058

RESEARCH ARTICLE

WILEY

"Combine to compete: Improving fiscal forecast accuracy over time"

Laura Carabotta[1] | **Peter Claeys**[2]

[1]Ministry of Economy and Finance of Italy, General Inspectorate of the Budgetary Policy, Public Finance Office, Rome, Italy

[2]Department of Economics, Universidad Pontificia Comillas, Madrid, Spain

Correspondence

Peter Claeys, Department of Economics, Universidad Pontificia Comillas, calle Alberto Aguilera, 23, E-28015 Madrid, Spain.

Email: pgaclaeys@comillas.edu

Abstract

Budget forecasts have become increasingly important as a tool of fiscal management to influence expectations of bond markets and the public at large. Difficulties in projecting macroeconomic variables in volatile economic times—together with political bias—thwart the accuracy of budget forecasts. Pooling information from many different forecasters can still lead to substantial gains in predictive accuracy when taking into account time variation. We combine the forecasts of both private and public agencies for Italy over the period 1993–2022, and test absolute and relative forecasting performance over time. Although forecast combinations do not necessarily result in less biased or more efficient forecasts, tracking better performing forecasters and combining their budget predictions produces significantly better predictions.

KEYWORDS

consensus forecasts, fiscal forecasting, forecast accuracy, forecast bias, forecast combination, public deficit

1 | INTRODUCTION

Budget forecasts are increasingly becoming a tool of fiscal management. Budget deficits that were rather contained in all industrialized economies before 2007 quickly gave way to deep budget deficits because of stimulatory tax cuts and spending hikes, financial bailouts, and the dragging on of the economic crisis. The Pandemic and geopolitical conflict have further fueled public debt, at a time rising interest rates are complicating refinancing. In Europe, budgetary forecasts now play a key role in the preparation of economic measures under the European Semester and in the monitoring of excessive deficits under the different reforms of the Stability and Growth Pact. Budget forecasts have always been a crucial part of a democratically controlled policy process. They are becoming a key input of informed budget drafting and decision-making, and a tool to manage expectations of fiscal responsibility in financial markets and the public at large.

Evidence tells us that budget forecasts have been a rather poor guide to correctly assessing the fiscal outlook. Budget projections often paint a too rosy picture of reality and are consistently biased towards too low deficits, especially when confronted with comparable predictions made by international institutions (Artis & Marcellino, 2001; Leal et al., 2008). Projections of fiscal adjustments are usually pushed forward over time and

We thank Barbara Rossi, Domenico Giannone, and participants at the 14th IWH-CIREQ Macroeconometrics Workshop in Halle, the III Time Series Conference in Zaragoza, the Centro Europeo di Ricerche in Rome, the 34th International Symposium on Forecasting in Rotterdam and AQR-IREA for useful comments and suggestions. The usual disclaimer applies.

revised when the decision nears (Beetsma & Giuliodori, 2010). A large literature argues that this bias in prediction performance is the consequence of setting politically motivated targets rather than realistic economic projections (Jonung & Larch, 2006). Nevertheless, even the forecasting performance of private institutions or public agencies is not stellar, which casts doubt on the forecastability of fiscal variables (Favero & Marcellino, 2005; Jalles et al., 2015). This might be the result of a lack of attention by private agencies as budget forecasting has not been a priority, or simply the inability to predict economic variables given structural changes over time.

Research into better forecasting practices of budget variables has not come to more conclusive findings. The bottom line of most applied work is that results depend on the choice of the forecasting procedure, the consistency of macroeconomic with fiscal forecasts, the forecast horizon, and the level of disaggregation of fiscal forecasts. Efforts to improve data availability on fiscal accounts over the last decade have paid off as attempts to incorporate more detailed information (Onorante et al., 2010; Pedregal & Pérez, 2010) and to apply more advanced econometric techniques (Asimakopoulos et al., 2013) have led to marginal improvements in forecasting performance.

Yet, forecasting the budget deficit is still considered to be more of an art than a science. Fiscal forecasts may require more judgment and expertise than econometric or modeling techniques (Leal et al., 2008). If progress depends on better inside knowledge of the dark box of the budget process, then the ultimate consequence is that there may be as many forecasts as there are forecasters. The fortunate implication, we argue in this paper, is that we can exploit the information contained in many individual budget forecasts to project a combined forecast. We pool the judgment and expertise of many forecasters. It is an established finding in the forecasting literature that combining improves upon the forecast of any single model (Hendry & Clements, 2004). We apply a variety of simple as well as more advanced combination techniques, which account for past forecasting performance, to compute a combined forecast.

In addition, we take the time variation in those forecasts into account by testing absolute and relative forecasting performance of all forecasts using recently developed tests to check forecasting accuracy over time (Giacomini & Rossi, 2010; Rossi & Sekhposyan, 2016). This is a necessity given the changing environment fiscal policymakers face, and we study, as an exemplary case, Italy. Fiscal policy in Italy has undergone many dramatic changes since the early 1990s, in particular during the run up to EMU entry with the large consolidation between 1997 and 1999, and this consolidation was reversed first during the 2008 Global Financial Crisis ending in the Sovereign Crisis, and then the Pandemic pushing the public debt ratio to 150% of GDP. We look at a dataset of 13 expert forecasts from both private agencies (from Consensus Forecasts) and projections by public institutions (Organization for Economic Co-operation and Development [OECD], European Commission [EC], International Monetary Fund (IMF), and the Ministry of Economy and Finance) for Italy over the period 1993 to 2022.

Our main finding is that different combinations of budget forecasts result oftentimes in more accurate forecasts. Although forecast combinations do not necessarily result in less biased or more efficient forecasts, tracking better performing forecasters and combining their budget predictions produces significantly better predictions. Standard static tests of forecasting accuracy show that some of the pooled forecasts indeed outperform each single forecaster, yet such an insight is not robust to structural changes. In fact, although not all individual forecasters produce forecasts efficiently, and even the better performing ones can produce mistakes in some specific time periods, tracking the bias and efficiency over time from the latter can lead to significant improvements when attributing more weight to better performing forecasters, especially over more recent years.

This is not a surprising result in light of prior evidence in the forecasting literature in general, yet the outcome is a novel one for the fiscal forecasting literature, which so far had little evidence on how to improve over a naïve random walk (Artis & Marcellino, 2001) or over a variety of forecasting time series models (Favero & Marcellino, 2005).

The paper is structured as follows. We first review in Section 2 the fiscal forecasts for Italy and discuss several techniques for combining forecasts from the different forecasters in the dataset. In Section 3, we first discuss standard static tests for absolute and relative forecasting performance, whereas in Section 4, we extend those tests to consider the evolution of forecasting performance over time. Section 5 concludes.

2 | COMBINING BUDGET FORECASTS

2.1 | Private and public budget forecasts

Budget forecasting may look like an exclusive task of ministries and international institutions, yet many other expert forecasters, like commercial or investment banks, industry, semi-governmental agencies, and university departments, have produced budget forecasts too. In recent years, some datasets have become available that include deficit forecasts from a larger set of expert forecasters over a continuous period of time. One of those datasets is collected by Consensus Economics Forecasts, Inc. (CEF). This company conducts a survey in several OECD countries among professional economists working for commercial or investment banks, industry, government-based agencies, and university departments. Most of the surveyed experts are at domestic institutions that provide forecasts for a single country only; a few work for international financial institutions or research institutes that provide forecasts for several countries simultaneously.

The CEF survey has gradually expanded its scope and coverage, and provides us with a large panel of private forecasters. The monthly survey on Italy covers 42 forecasters from January 1993 to December 2022. However, despite the gradual expansion of the dataset, fiscal forecasts have not always received the same attention by forecasters over time. Some forecasters stopped producing projections for the budget balance over time, whereas others that were initially included left the sample owing to closure, mergers, or other reasons. Moreover, new forecasters joined the CEF survey only at a later stage. Our sample is therefore a subset of the entire group of 42 expert forecasters. We do not consider those forecasters that have participated just a few times in the survey. In particular, any forecaster participating fewer than 36 consecutive months in the CEF survey and/or not producing budget forecasts over this period is excluded. This reduces the panel to a selection of nine forecasters among Italian banks and research institutes. To preserve confidentiality of the dataset, we call these forecasters A to I.

The survey enquires respondents every first week of each month about current and year-ahead forecasts for a number of macroeconomic variables, and these forecasts are published early in the second week of the same month.[1] We compute budget forecasts for both the current year ($F_{t,k}$) and the year ahead ($F_{t+1,k}$) over the sample period from January 1993 to December 2022. The forecasts require some transformation before they can be used in the empirical analysis. CEF asks respondents for a forecast of general government (overall) budget balance in nominal terms.[2] In order to transform this forecast into one of the net lending as a ratio to GDP—so positive numbers represent deficits—we divide the forecast of the nominal balance for year $t+1$ in a certain month m by the GDP forecast for the same year. As the CEF only provides forecasts of GDP *growth rates*, we compute the year-ahead nominal GDP forecast by applying the CEF growth rate to the latest available estimate for the same year GDP. The latter is taken from IMF World Economic Outlook (WEO; see Appendix B for more details).

In addition to the private forecasts, we also consider public budget forecasts for the current year and the year ahead. These forecasts come from four institutions: the OECD, the IMF Forecast, the European Commission, and the Italian Ministry of Economy and Finance (MEF). The international institutions do not produce forecasts at a monthly frequency. Generally speaking, they produce projections twice a year (in Spring and Autumn) at different moments. The OECD publishes its forecasts twice a year in June and December in the Economic Outlook; IMF forecasts are published in the WEO in May and October; and so are the forecasts by the European Commission. The publication of forecasts by the Italian MEF is part of the "Economic and Financial Planning Document (DPEF)" from 1992 to 1997, and the "Forecast and Planning Report (RPP)" from 1998 to 2022 that are used by the Italian government when submitting the budget to Parliament. These forecasts are produced in June, July, and October.

Table 1 shows how we match the timing of the four public forecasters with the nice CEF private forecasts. We can match 4 months where there is a correspondence between the nine forecasters (May, June, October, and December). However, for the purpose of combining and evaluating forecasts provided, the December forecast is too close to the end of the year to lead to divergent

TABLE 1 Timing of release of budget forecasts.

Month	Current year forecast	Year ahead forecast
May	EC	EC
	IMF	IMF
	Private forecasters (CEF, A to I)	Private forecasters (CEF, A to I)
June	OECD MEF	OECD MEF
	Private forecasters (CEF, A to I)	Private forecasters (CEF, A to I)
October	MEF	MEF
	EC	EC
	IMF	IMF
	Private forecasters (CEF, A to I)	Private forecasters (CEF, A to I)
	OECD	OECD
December	Private forecasters (CEF, A to I)	Private forecasters (CEF, A to I)

Notes: MEF projections are published in July during 1992–1995, June in 1996–1997, and October during 1998–2022. CEF, Consensus Economics Forecasts, Inc.; IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance.

forecasts of the budget deficit for that year. Hence, we use the Spring forecasts published in May and June in the rest of the paper.

In particular, in our database, we include the information from the months of May or June for public institutions (EC, OECD, IMF, and MEF) and May for the forecasters from the CEF database. In a few cases, some of the private forecasts were missing, and in that case, we used the forecasts from April that year. We add to the 13 forecasts also a simple random walk forecast, which is just the realized net lending ratio of last year.

Figure 1a shows a graph of the different current-year forecasts over time and compares them to the realized net lending ratio for that period. This series comes from the OECD Economic Outlook.[3] Figure 1b does the same for the budget forecasts 1 year ahead provided by the same respondents. In both panels of Figure 1, the forecasts broadly move in the same direction, but there is definitely more dispersion in the year-ahead forecasts than in the current-year forecasts.[4] Although the range of forecasts differs by no more than 1% of GDP in the latter, the range increases to 3% on average for the former. There are also considerable changes over time. Up to 2001, all forecasters agree on a quite fast consolidation, and this is inspired by the Maastricht criteria. Afterwards, the forecast tends to become less accurate. The exception is the

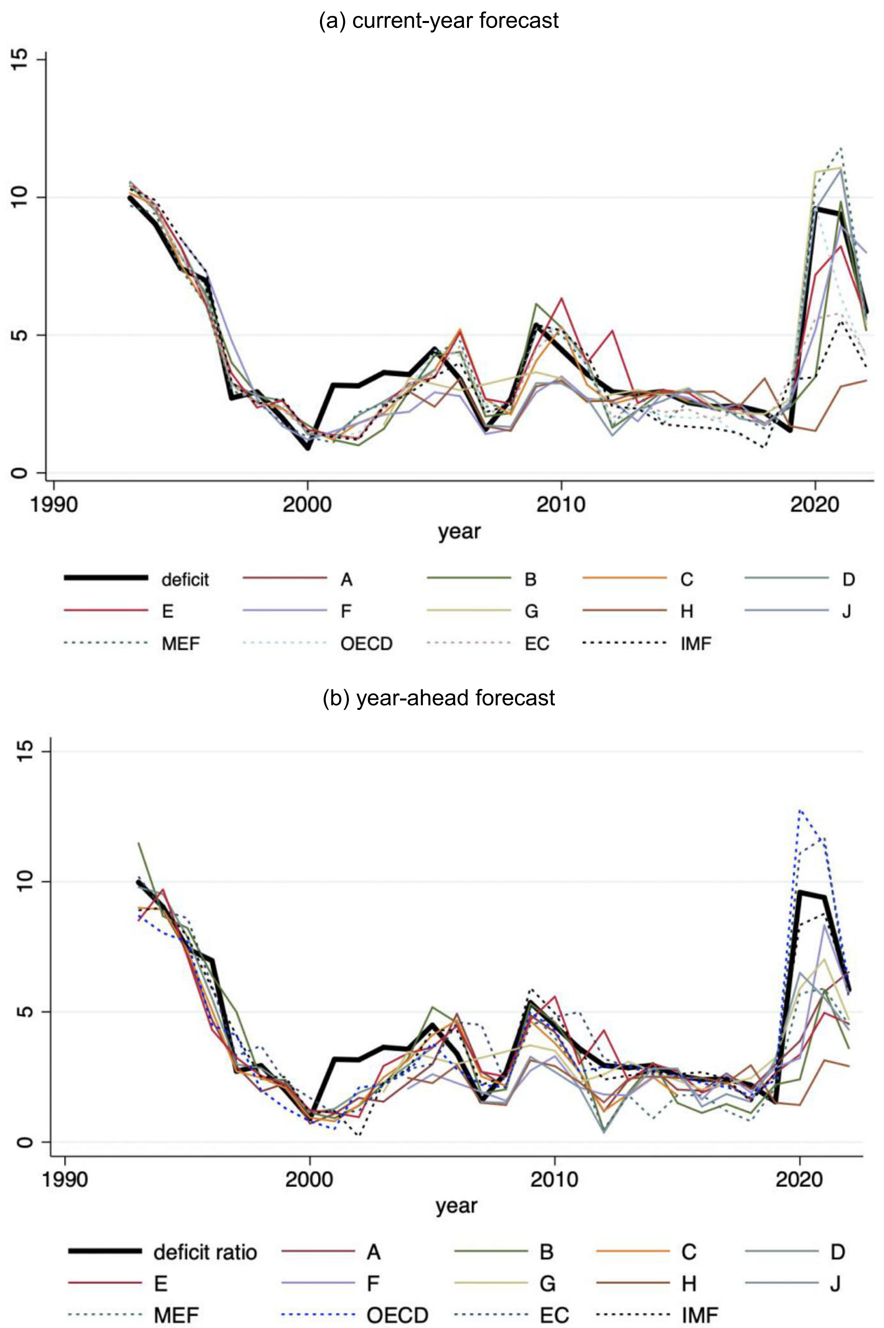

FIGURE 1 Realized and forecast net lending ratio, sample 1993–2022. (a) Current-year forecast. (b) Year-ahead forecast. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance.

rapid rise of deficits in the Global Financial Crisis starting in 2008: all forecasters agree the deficit would become much larger. The consolidation in the following year is not as easily foreseen. A similar phenomenon is seen in the Pandemic: starting in March 2020, some forecasters quickly adjusted in the next Spring forecast their projections whereas others predicted a much more gradual approach. Given the speed of the fiscal programs and the uncertainty surrounding the evolution of the Pandemic, forecasters had strongly different opinions.

2.2 | Forecast combinations

A vast literature shows that the combination of various forecasts results in improved prediction performance (Clemen, 1989; Hendry & Clements, 2004; Timmermann, 2006).[5] The reason for the improved performance is that single forecasts are the product of a specific forecasting model, which depends on specific econometric techniques and personal judgment each with some idiosyncratic error. Pooling many forecasts averages out these errors. Also, the empirical models used in forecasting are based on the assumption of stable relationships, but political events, crises, technological progress, etc. upset economic relations continuously. Combination levels out this instability (Pesaran & Timmermann, 2005). Further, combining reduces the risk of forecast bias when there are many macroeconomic variables that are endogenous over the economic cycle. If forecasts are used as a proxied input for forecasting other variables, these proxies introduce a systematic measurement bias and reduce forecast accuracy. Finally, each forecasting model assumes a loss function by the forecaster. With changes in volatility of the economic variables used in the model, combination of forecasts can produce more precise forecasts. The aim of combination is to make forecasting practices robust to the different types of uncertainty.

A combined forecast $Y^*_{i,t+h}$ of n different forecasts of a variable Y at horizon h is of the general form:

$$Y^*_{t+h} = \alpha_t + \sum_{i=1}^{n} \beta_{i,t} Y_{i,t+h} Y^*_{t+h} = \alpha_t + \sum_{i=1}^{n} \beta_{i,t} Y_{i,t+h} \quad (1)$$

A considerable amount of research has been undertaken to determine how to choose the coefficients, α_t and $\beta_{nt}\beta_{it}$. Evidence suggests that the simple approach of averaging the individual predictions works well (Clemen, 1989; Clemen & Winkler, 1986; Lupoletti & Webb, 1986). In this case, β_{nt} is equal to $1/n$ on all individual predictions. Alternatively, the geometric mean and harmonic mean and the median can be used as a summary. The simple average has often been found to be a quite robust forecast for a set of macro-economic variables, suggesting that forecasters are on average right (Clemen, 1989).[6]

More complex methods may further improve performance by attributing different weights $\beta_{i,t}$ to each forecaster (Stock & Watson, 2004). The typical way is to give more weight to better performing forecasters, for example, by attributing to each expert forecaster a weight that is inversely proportional to the predictor's Mean Square Forecast Error (MSFE). The proposal is to attribute more weight to those forecasters that were on average doing better in the past. Past performance might not be a good guide to future performance, nonetheless: resilience of a forecasting model to structural breaks distinguishes good from average forecasts. Stock and Watson (2004) discount past MSFE over a horizon h to attach greater weight to the recent predictive ability of each individual predictor. The weights in (1) are then given by the forecaster's MSFE compared to the overall MSFE, where past MSFE is discounted h periods back in time with a factor δ:

$$\beta_{n,t+h} = m_{n,t}^{-1} / \sum_{n=1}^{N} m_{\mathrm{n,t}}^{-1} \text{where } m_{n,t+h} = \sum\nolimits_{s=t_0}^{t-h} \delta^{t-h-s} \left(Y^h_{s+h} - \widehat{Y}^h_{n,s+h} \right)^2 \quad (2)$$

Stock and Watson (2004) propose to cut off information from all past performance after some relevant period of time, using time-varying weights. This corresponds to setting δ to 1, and reducing h to a short horizon, and gives the "recently best" forecast. As forecasters update their models quickly after bad performance, one should exclude outdated versions of forecasting models. Recent performance is therefore more relevant for forecast evaluation than average historical performance.

An alternative approach to computing weights is to estimate those weights from a simple regression of the forecast on the different forecasts, as in Equation (3):

$$Y_t = \alpha_t + \sum_{i=1}^{n} \beta_i \widehat{Y}_i + Y_t \quad (3)$$

This is nothing else than an extended version of the regression used for testing unbiasedness and weak efficiency of the forecast. The regression approach relaxes the assumption in (2) of unbiased and uncorrelated errors as the constant is not bound to be zero, and the weights do not have to sum to 1 (Lupoletti & Webb, 1986). Another approach is to apply to the sample of forecasts a principal components model and extract those factors that drive most of the forecasts.

A priori, there is no reason to treat private or public forecasters differently. Public (international) institutions have been found to produce less biased and more efficient forecasts (Artis & Marcellino, 2001). One of the benefits of the CEF forecasts is that unlike other surveys, individual forecasts in the CEF should not suffer a bias owing to the release of strategic forecasts, as often happens for official forecast released by governmental agencies (Ottaviani & Sorensen, 2006). CEF data are public, which prevents a participant from reproducing others' forecasts and also limits the possibility of herding (Trueman, 1994). Analysts are bound in their survey answers by their recommendations to their clients, and discrepancies between the survey and their private recommendation would be hard to justify (Keane & Runkle, 1990). In addition, and unlike other surveys, professional economists who participate in the CEF poll not only take a stance on the direction of the expected change of a macroeconomic variable but also forecast the level of the macroeconomic variable. Evidence shows that CEF forecasts are less biased and more accurate than other surveys in general, and this might also hold for budget forecast (Jalles et al., 2015).[7]

From the 13 expert forecasts, we thus compute an additional 11 different combined forecasts. These include five simple combination models that average the different budget forecasts. These include the simple average, the geometric average, the harmonic average, the median, and the trimmed mean (in which 20% of the top or bottom forecasts have been eliminated).

We also compute two regression weights-based combination models. The first one is based on the regression of the realized net lending ratio on all nine expert forecasts (weighted forecast combination).[8] The second one adds to the set of forecasters also a random walk forecast. The reason to include a simple AR(1) process is to make the forecast more robust to structural breaks in the series. Pesaran and Timmermann (2005) show that including models with different degrees of adaptability to breaks outperform the forecasts from alternative pooled forecasts. The random walk picks up any of such changes in the following period already, whereas the other forecasts may still deviate because of their dependence on past patterns.

We next construct four forecast combinations that select only the best performing forecasters over recent periods. We apply the weights from (1) that are inversely proportional to the predictor's MSFE relative to the realized net lending ratio. For the first three of these combinations, we discount past performance using a value of δ of 0.90, 0.95, and 0.99. Alternatively, we cut off the time horizon after 4 years and look only at the recently best performing forecasters (R_{best}).[9]

Figure 2 displays the realized net lending ratios together with the different forecast combinations; Figure 2a shows the current year forecast, and Figure 2b the year-ahead forecast. All combined forecasts track closely the net lending ratio over the first part of the sample (up to 2001, or between the Financial Crisis and the Pandemic). Afterwards, there is a tendency to deviate from the balance for a couple of years. In 2001, this was often considered to be the consequence of the election of a new government that undid most of the fiscal efforts pre-EMU; in 2020, as the Pandemic struck Italy early on, the economic downturn was very strong. Figure 2 shows that most expert forecasts fail in the same direction at the moment of an unexpected break, either undershooting or overshooting at the same time.[10] A comparison of Figure 2 to the original forecasts shows that combinations are less variable than the single forecasts. Figure 2a suggests that forecast combinations are performing equally well in tracking the realized net lending. In Figure 2b, the weighted forecast combination as well as the R_{best} combination are closest to the actual data after 2001 or 2020 when all the agencies tended to make large forecasting mistakes.

3 | STANDARD TESTS FOR PREDICTIVE ACCURACY OF BUDGET FORECASTS

3.1 | Testing absolute forecast performance

The forecast error is given in (4) by the difference between Y_t (the actual value of the net lending to GDP ratio in year t) and $F_{t,k}$—the forecast for the current year made by forecaster k ($F_{t+1,k}$ if it is the next year forecast)—and $e_{t,k}$ ($e_{t+1,k}$) is the corresponding forecast error:

$$Y_t - F_{t,k} = e_{t,k} \tag{4}$$

We first want to check whether fiscal forecasts are performing well on absolute standards, so we look into the bias, the efficiency, and the information rigidity inherent in forecasts. The first concern with fiscal forecasts is that they are unbiased, that is, whether the mean forecast error is significantly different from zero. This can be easily tested on the following expression for the forecast error:

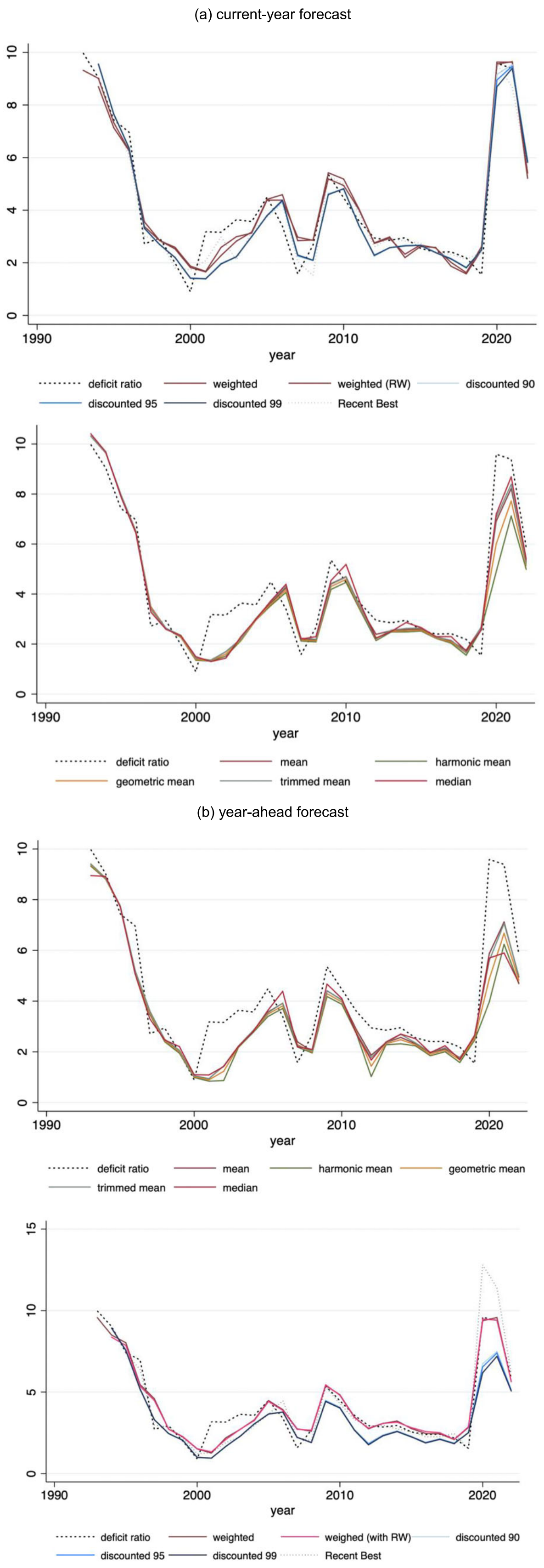

FIGURE 2 Realized net lending ratio and the combined forecasts. (a) Current-year forecast. (b) Year-ahead forecast.

$$Y_t - F_{t,k} = \alpha + \epsilon_{t,k} \tag{5}$$

Forecasts are unbiased if we cannot reject the null hypothesis that $\alpha = 0$ in Equation (5). A positive coefficient would indicate that forecasters are on average biased toward too optimistic forecasts. Column (a) of Table 2 shows that this is actually not the case for private and public forecasters, with the exception of forecaster H. Among the forecast combinations, the weighted regressions and the discounted forecast, as well as the recent best are all unbiased, but this is not the case for the different mean combinations. Over the entire panel of private and public forecasters, nevertheless, the same panel test (with fixed effects for each forecaster, and assuming robust standard errors), shows that forecasters are biased and tend to underpredict on average the budget deficit by about 0.54% of GDP. For the year-ahead forecasts, the average error is larger—around 0.68% of GDP—and there is a significant bias in the budget forecast, and few forecasters are showing a positive bias.

Holden and Peel (1990) show that the test above provides a sufficient and necessary condition for bias. The efficiency or rationality of the forecasts furthermore depends on the available information a forecaster has, and whether the forecaster uses this information efficiently. A standard test for efficiency regresses actual observations Y_t on a constant plus the forecast $F_{t,k}$. Forecasts are efficient or rational if we cannot reject the null hypothesis that $\alpha = 0$ and $\beta = 1$ in Equation (6)

$$Y_t = \alpha + \beta F_{t,k} + \epsilon_{t,k} \tag{6}$$

Table 2 presents our results for each forecaster and the different groups of combinations. In the tests on α and β, we find that the first condition is met, but not the second one. On the entire panel, nevertheless, the hypothesis that $\beta = 1$ is just marginally rejected, and the coefficient falls 20% short on average. For the year ahead forecasts, most forecasters do surprisingly well as the forecast matches the realized net lending, and the constant is in many cases significantly different from zero. This also shows up in the panel result: while the forecast is on average different from zero, it matches the forecast closely (at a coefficient 0.96) yet it is significantly different from zero, as indicated by the F-statistics and the associated p-values.

Forecasts can be inefficient because of two reasons: on the one hand, the forecast and the forecast error should not co-move, and on the other hand, forecast errors should not be serially correlated. Jalles et al. (2015)

TABLE 2 Tests for bias, efficiency, and information rigidity.

	(a)		(b)				(c)				(d)	
	α		α		β		β		ρ		θ	
Current year	**Coefficient**	***p*-value**	**Coefficient**	**Se**	**Coefficient**	**Se**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**
Private												
A	−0.04	0.80	0.93	0.07	0.31	0.32	0.07	0.34	0.00	0.99	0.13	0.52
B	−0.29	0.28	0.84	0.10	0.91	0.45	0.16	0.11	−0.03	0.86	−0.10	0.64
C	−0.17	0.40	0.87	0.08	0.67	0.36	0.13	0.12	0.17	0.48	0.23	0.36
D	−0.17	0.50	0.86	0.07	0.79	0.37	0.14	0.07	0.33	0.28	0.61	0.06
E	−0.02	0.92	0.93	0.08	0.33	0.39	0.07	0.37	0.12	0.54	0.10	0.64
F	−0.41	0.13	0.84	0.12	0.96	0.47	0.16	0.18	0.08	0.73	0.43	0.05
G	−0.14	0.55	0.79	0.07	0.97	0.34	0.21	0.01	0.42	0.04	0.05	0.90
H	−1.20	0.04	0.21	0.86	3.31	2.36	0.79	0.37	0.49	0.04	−0.28	0.27
I	−0.34	0.16	0.84	0.08	0.91	0.34	0.16	0.06	0.41	0.12	0.08	0.83
Public												
MEF	−0.08	0.62	0.84	0.05	0.72	0.24	0.16	0.00	0.21	0.27	0.14	0.54
OECD	−0.31	0.10	0.90	0.07	0.71	0.32	0.10	0.14	0.35	0.07	0.10	0.63
EC	−0.45	0.06	0.96	0.10	0.60	0.45	0.04	0.69	0.24	0.21	0.32	0.10
IMF	−0.55	0.07	0.83	0.11	1.16	0.50	0.17	0.16	0.31	0.11	0.13	0.51
Combinations												
Mean	−0.30	0.06	0.98	0.06	0.39	0.30	0.02	0.73	0.13	0.49	0.35	0.09
Harmonic mean	−0.50	0.02	0.97	0.09	0.61	0.39	0.03	0.73	0.18	0.34	0.36	0.07
Geometric mean	−0.39	0.03	0.98	0.08	0.47	0.34	0.02	0.79	0.15	0.42	0.37	0.07
Median	−0.24	0.13	0.96	0.06	0.40	0.29	0.04	0.52	0.12	0.52	0.32	0.12
Trimmed mean	−0.29	0.07	0.97	0.06	0.39	0.29	0.03	0.69	0.11	0.56	0.34	0.10
Weighted forecast regression	0.00	1.00	1.00	0.05	0.00	0.24	0.00	1.00	0.19	0.33	020	0.37
Weighted forecast regression (RW)	0.00	1.00	1.00	0.05	0.00	0.25	0.00	1.00	0.10	0.62	0.17	0.43
Discounted MSFE ($\delta = 0.90$)	−0.17	0.17	0.94	0.05	0.41	0.23	0.06	0.23	0.16	0.41	0.21	0.33
Discounted MSFE ($\delta = 0.95$)	−0.18	0.15	0.95	0.05	0.39	0.23	0.05	0.30	0.14	0.48	0.22	0.30
Discounted MSFE ($\delta = 0.99$)	−0.20	0.13	0.95	0.05	0.37	0.24	0.05	0.40	0.11	0.58	0.23	0.27
R_{best}	−0.13	0.27	1.01	0.06	0.09	0.,23	−0.01	0.86	−0.10	0.63	0.01	0.98

TABLE 2 (Continued)

	(a)		(b)				(c)				(d)	
	α		α		β		β		ρ		θ	
Current year	**Coefficient**	***p*-value**	**Coefficient**	**Se**	**Coefficient**	**Se**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**
Panel												
	−0.54	0.00	0.93	0.02	0.80	0.06	0.07	0.00	0.21	0.00	0.19	0.00

	(a)		(b)				(c)				(d)	
	α		α		β		β		ρ		θ	
Year ahead	**Coefficient**	***p*-value**	**Coefficient**	**Se**	**Coefficient**	**Se**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**
Private												
A	−0.77	0.07	0.98	0.18	0.83	0.62	0.46	0.26	0.28	0.17	0.28	0.17
B	−0.50	0.30	0.76	0.12	1.52	0.51	0.55	0.15	0.45	0.01	0.45	0.01
C	−0.24	0.47	0.92	0.08	0.85	0.34	0.37	0.12	0.28	0.24	0.28	0.24
D	0.04	0.94	0.89	0.07	0.92	0.38	0.35	0.12	0.51	0.09	0.51	0.09
E	−0.44	0.28	0.98	0.15	0.69	0.63	0.38	0.19	0.39	0.04	0.39	0.04
F	−1.40	0.03	1.00	0.25	1.15	0.82	0.50	0.38	0.10	0.75	0.10	0.75
G	−0.68	0.18	1.52	0.18	−1.00	0.64	−0.12	0.37	0.11	0.74	0.11	0.74
H	−1.59	0.02	0.87	0.90	1.85	2.17	2.72	0.87	0.38	0.13	0.38	0.13
I	−1.46	0.01	1.35	0.20	0.24	0.63	−0.08	0.31	0.07	0.82	0.07	0.82
Public												
MEF	−0.90	0.02	1.02	0.10	0.98	0.37	0.29	0.15	0.35	0.07	0.35	0.07
OECD	−0.15	0.72	0.79	0.06	1.08	0.30	0.45	0.11	0.08	0.73	0.08	0.73
EC	0.27	0.55	0.80	0.06	0.73	0.33	0.49	0.12	0.20	0.37	0.20	0.37
IMF	−0.10	0.79	0.96	0.07	0.39	0.32	0.40	0.14	0.24	0.24	0.24	0.24
Combinations												
Mean	−0.49	0.19	1.06	0.08	0.44	0.35	0.32	0.15	0.30	0.13	0.30	0.13
Harmonic mean	−0.74	0.06	1.01	0.11	0.87	0.44	0.36	0.16	0.37	0.05	0.37	0.05
Geometric mean	−0.61	0.11	1.04	0.10	0.62	0.40	0.33	0.16	0.34	0.08	0.34	0.08
Median	−0.53	0.15	1.08	0.10	0.39	0.40	0.29	0.16	0.32	0.10	0.32	0.10
Trimmed mean	0.15	0.68	1.00	0.06	0.00	0.28	0.34	0.13	0.16	0.44	0.16	0.44

(Continues)

TABLE 2 (Continued)

	(a)		(b)				(c)				(d)	
	α		α		β		β		ρ		θ	
Year ahead	**Coefficient**	***p*-value**	**Coefficient**	**Se**	**Coefficient**	**Se**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**	**Coefficient**	***p*-value**
Weighted forecast regression	0.12	0.74	1.00	0.06	0.00	0.30	0.38	0.15	0.14	0.51	0.14	0.51
Weighted forecast regression (RW)	−0.50	0.17	1.05	0.09	0.46	0.36	0.32	0.15	0.31	0.11	0.31	0.11
Discounted MSFE ($\delta = 0.90$)	−0.49	0.19	1.08	0.08	0.34	0.31	0.36	0.17	0.21	0.30	0.21	0.30
Discounted MSFE ($\delta = 0.95$)	−0.51	0.17	1.08	0.08	0.36	0.32	0.36	0.17	0.22	0.27	0.22	0.27
Discounted MSFE ($\delta = 0.99$)	−0.54	0.15	1.09	0.09	0.37	0.34	0.36	0.17	0.24	0.23	0.24	0.23
R_{best}	0.05	0.93	0.75	0.05	0.95	0.22	0.60	0.13	−0.04	0.85	−0.04	0.85
Panel	−0.68	0.00	0.96	0.00	0.93	0.00	0.39	0.00	0.26	0.00	0.15	0.00

Notes: For the panel results, forecaster fixed effects are included in each regression but not reported for reasons of parsimony. Heteroskedastic-consistent robust standard errors are used for the *p*-value. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

call this the beta and rho tests for inefficiency in a joint test of Equation (7) respectively,

$$e_{t,k} = \alpha + \beta F_{t,k} + \epsilon_{t,k}$$
$$e_{t,k} = \phi + \rho e_{t-1,k} + \upsilon_{t,k} \tag{7}$$

For the current-year forecast, Table 2 (column c) shows that forecasts and their errors do not significantly co-move, except in the case of forecaster H and the Ministry of Economy and Finance. Nevertheless, for none of the forecasters is there evidence of forecast errors being persistent, implying mistakes do not carry over. Again, the small sample for each forecaster may bias the result as the panel test shows that forecasts are on average related to the errors, and that these are also persistent, although in both cases the co-movement is low.[11] For the year-ahead forecasts, the results are very similar, and the degree of co-movement and serial correlation is similar. These results confirm the outcomes of other papers (Jalles et al., 2015; Leal et al., 2008) that forecasters do not update forecasts with all information available. This pattern is confirmed if we look at of forecast revisions.

While rational forecasters should constantly update their projections, there can be reasons for not doing so. Mankiw and Reis (2002) argued that sticky information renders constant updating too costly, and forecasters are not willing to acquire and digest information to make new projections constantly. As described in Section 2, this seems to be borne out on our sample, where most private forecasters have paid just intermittently attention to the budget deficit. By contrast, Sims (2003) argues that information is not costly, but that forecasters receive very diverse signals about the state of the economy they need to interpret. Forecasters are therefore bound to make errors, and information rigidities imply that the forecast error will be correlated with forecast revisions. Dovern et al. (2015) show that both classes of models of information rigidities described earlier imply that regressions of revisions of forecasts on the past revision should yield a positive coefficient; in contrast, the full information rational expectations model would predict that the coefficient is zero. Results from a regressions of the budget forecasts revisions on earlier revisions are displayed for the different (groups of) forecasters in column d of Table 2. As before, for none of the forecasters do we find evidence that the revisions are serially correlated, yet for most the coefficient is positive (with the exception of forecasters B and H). As in the previous set of results, the small sample may result in insignificant findings, as the panel estimates for both the current-year and year-ahead forecasts show that revisions display correlation (about 0.15 and 0.19, respectively).

As the size of the budget revision may be too demanding a task to forecast, one would like the forecaster to at least make a correction in the same direction as the budget changes, in order to capture the turning points in the series. A simple way to test this is to use a Wilcoxon rank test for the equality of revision of the forecast versus the realized budget series. One way is to run a z-test on the equality of the forecast and realized budget series, or alternatively, run a two-sided test that the median of the difference between the realized net lending and the forecast revision move in the same direction. Table 3 shows that for all forecasters—public and private—or any combination of those, we cannot reject the null that the series are different, and that the direction of revision of the budget forecast is similar to the change in the budget, which indicates forecasters do not anticipate the changes. The same occurs for the year-ahead forecasts.

The overall conclusion is that combining budget forecasts does not improve absolute performance, as the average bias carries over from the private and public forecasters, and inefficient use of information is not improved upon by combining them, that is, combining

TABLE 3 Wilcoxon rank test on forecast revisions.

	Current year		Year ahead	
Private	**(a)**	**(b)**	**(c)**	**(d)**
A	0.90	0.70	0.77	0.99
B	0.94	0.99	0.75	0.71
C	0.68	0.50	0.97	0.99
D	0.94	0.77	0.94	0.99
E	0.94	0.71	0.72	0.71
F	0.44	0.54	0.39	0.61
G	0.20	0.18	0.10	0.18
H	0.59	0.48	0.56	0.81
I	0.95	1.00	0.99	0.99
Public				
MEF	0.77	0.99	0.69	0.46
OECD	0.99	0.46	0.57	0.46
EC	0.82	0.85	0.61	0.99
IMF	0.66	0.46	0.67	0.71
Combinations				
Mean	0.69	0.99	0.34	0.26
Harmonic mean	0.96	0.85	0.51	0.34
Geometric mean	0.74	0.99	0.39	0.26
Trimmed mean	0.77	0.99	0.36	0.26
Median	0.82	0.99	0.55	0.71
Principal component	0.74	0.99	0.37	0.26
Principal component (RW)	0.66	0.99	0.38	0.46
Weighted forecast regression	0.91	0.71	0.61	0.71
Weighted forecast regression (RW)	0.93	0.85	0.45	0.57
Discounted MSFE ($\delta = 0.90$)	0.99	0.34	0.45	0.34
Discounted MSFE ($\delta = 0.95$)	0.99	0.34	0.47	0.34
Discounted MSFE ($\delta = 0.99$)	0.98	0.34	0.47	0.34
R_{best}	0.94	0.70	0.89	0.70

Note: Columns (a) and (c) show the p-value of a z-test for the equality of revision of the forecast versus the realized series, whereas columns (b) and (d) show the p-value of a two-sided test that the median of the difference between the realized net lending and the forecast revision move in the same direction. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

does not solve the problems of information rigidities. We will analyze in Section 4 if evaluating the stability of absolute performance over time can provide more insight.

3.2 | Testing relative forecast performance

Even if forecasts are biased and inefficient, we can compare their relative performance to choose a combination of forecasts that can provide insight in future budget developments. We first analyze with some standard test whether a linear combination of forecasts outperforms any individual expert forecast—private or public. We apply a standard test and compute the RMSE (root mean squared error) and the Theil test.[12] These descriptive statistics are often used to rank the performance of different forecasting models but do not test whether difference in performance is statistically significant. In a second exercise, we therefore apply the Diebold–Mariano–West (DMW) test of predictive accuracy. The DMW test (1995) supposes that a forecaster has an identical loss function g() so that two different forecasts A and B lead to similar losses because of errors. Let $g(e_{t,A})$ and $g(e_{t,B})$ denote the loss from a forecast error evolving from prediction models A and B, then the DMW statistic is

$$DMW = \frac{\frac{1}{T}\sum_{1}^{t}\{g(e_{t+h,A}) - g(e_{t+h,B})\}}{\widehat{\sigma}_{g(e_{t+h,A})-g(e_{t+h,B})}}, \tag{8}$$

where $\widehat{\sigma}$ is a consistent estimate of the standard deviation of the difference of losses. The null hypothesis is that $g(e_{t+h,A}) = g(e_{t+h,B})$, and DM is simply distributed as N (0,1) under the null.

We compare each of the 24 single or combined forecasts in panels (a)–(c) of Table 4. We see that nearly all forecasters do better than a simple RW model would suggest, both for the current year and the year-ahead forecast. The only exception is forecaster H, whose Theil statistic is larger than 1. The result that forecasts of budget deficits are doing better than a random walk stands a bit in contrast with other results in the field that find the random walk to perform at least as well as public forecasts (Artis & Marcellino, 2001) or simple time-series models (Favero & Marcellino, 2005).

On the current-year forecast, public forecasters generally outperform the private ones, with the exception perhaps of forecaster A or I. The performance of the combined forecasts is a bit mixed. Unsurprisingly, the simple combinations of private and public forecasts tend to do worse than the public forecasters. The weighted forecast combination improves over the public forecasts as it puts less weight on the private forecasts, and aggregates the information of the public ones. Adding a simple RW model to the forecast combination incorporates any structural breaks, such that this model is ranked first on all current-year forecasts by all criteria. The robustness to structural breaks also explains why the Recent Best is more accurate than the discounted combination of forecasts.

Forecasting performance for the budget 1-year ahead is not surprisingly worse than for the current year. All forecasters—or any combination of them—do better than the simple random walk, except for forecaster H or the principal components. Evidence is more mixed on the relative performance of private and public forecasters: the IMF or EC forecasts are ranked above those of the MEF, OECD, or any private forecaster. And they even perform better than any of the combinations of forecasts. Figure 2 already showed that year-ahead forecasts are deviating more from the realized net lending than for the current-year forecast. As a result, the combinations do not provide much gain over the single forecasts. The weighted forecast combinations are now performing better, even in comparison to the discounted forecasts and the Recent Best one. The reason for the relative underperformance of combinations is that nearly all private forecasters are now prone to make mistakes because of structural breaks, and the insufficiently incorporate new information. This can be seen from a weighted forecast combination that includes the random walk among the regressors.

The accuracy criteria show that the improvements in forecast performance from one forecast to another are often marginal, so they may not be significant. We therefore compare each pair of forecasters and test the relative predictive performance. Table 5 summarizes the results of the DMW test and shows *p*-values.

Panel (a) of Table 5 confirms some of the previous findings in Tables 2 and 4 for current-year budget forecasts. In contrast to most of the literature on forecasting deficits, we find that expert forecasts or pooled forecasts always outperform the naïve AR (1) model, and this is also a significant improvement. Between private or public forecasters, it is hard to detect any single one that stands out. Instead for the forecast combinations, the weighted forecast combinations, the discounted weighted forecasts, and the Recent Best combination consistently outperform any of the private or public forecasters. Yet, it is hard to distinguish any of these combinations to consistently outperform any of the other ones. For example, even if the

TABLE 4 Accuracy test of single and combination forecasts.

	Current year		Year ahead	
	RMSE	Theil U	RMSE	Theil U
Private				
A	0.82	0.29	1.67	0.77
B	1.42	0.78	1.86	0.91
C	0.92	0.81	1.03	0.94
D	0.83	0.79	0.92	0.82
E	1.08	0.48	1.72	0.80
F	1.36	0.59	1.94	0.83
G	0.95	0.26	1.38	0.50
H	2.57	1.01	2.66	1.02
I	0.96	0.24	1.77	0.49
Public				
MEF	0.86	0.43	1.54	0.71
OECD	1.02	0.41	1.17	0.63
EC	1.32	0.60	1.14	0.49
IMF	1.65	0.79	0.93	0.44
Combinations				
Mean	0.89	0.48	1.18	0.61
Harmonic mean	1.22	0.64	1.55	0.78
Geometric mean	1.03	0.55	1.36	0.69
Trimmed mean	0.88	0.47	1.21	0.63
Median	0.86	0.47	1.31	0.61
Weighted forecast regression	0.65	0.33	0.74	0.39
Weighted forecast regression (RW)	0.64	0.32	0.74	0.38
Discounted MSFE ($\delta = 0.90$)	0.67	0.37	1.04	0.55
Discounted MSFE ($\delta = 0.95$)	0.68	0.38	1.08	0.56
Discounted MSFE ($\delta = 0.99$)	0.69	0.38	1.13	0.59
R_{best}	0.60	0.27	0.95	0.51

Notes: IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error; RMSE, root mean square forecast error; Theil U, the Theil test.

weighted forecast combination improves considerably over the single forecasts and nearly all other combined forecasts, if we make this forecast robust to structural changes by adding a RW, then this combination does not improve in a significant way. This set of results shows that pooling may result in improvements in forecasting accuracy, and that those gains in accuracy are also statistically significant.

The results in panel (b) shows a rather similar picture for the year-ahead forecasts. The combination of forecasts is better than most expert forecasts, yet between those models, differences are not always significant. Only the Recent Best combination is on many occasions performing marginally better than other combinations.

4 | STABILITY IN FORECASTS

4.1 | Stable absolute forecasting performance

The tests for forecast rationality, efficiency, and information rigidity have so far assumed stationarity, but fiscal policy in Italy has undergone many dramatic changes since the early 1990s, in particular during the run up to EMU entry with the large consolidation between 1997 and 1999, the 2008 Global Financial Crisis ending in the Sovereign Crisis, and the Pandemic. Tests of forecaster performance are invalid in the presence of such instabilities.

TABLE 5 DMW test for relative forecasting performance.

Panel (a): current year forecast	B	C	D	E	F	G	H	I	MEF	OECD	EC
A	0.16	0.04	0.08	0.00	0.04	0.75	0.21	0.34	0.58	0.30	0.23
B		0.65	0.09	0.40	0.84	0.21	0.19	0.37	0.16	0.11	0.01
C			0.32	0.11	-	0.62	-	0.01	0.34	0.45	0.66
D				0.00	0.00	-	-	-	0.62	0.59	0.32
E					0.14	-	0.26	0.19	0.01	0.68	0.53
F						0.10	0.26	0.26	0.01	0.00	0.79
G							0.20	0.66	0.75	0.42	0.27
H								0.24	0.20	0.20	0.19
I									0.71	0.66	0.41
MEF										0.18	0.22
OECD											0.22
EC											
IMF											
Weighted forecast regression											
Weighted forecast regression (RW)											
Mean											
Harmonic mean											
Geometric mean											
Median											
Trimmed mean											
Discounted MSFE ($\delta = 0.90$)											
Discounted MSFE ($\delta = 0.95$)											
Discounted MSFE ($\delta = 0.99$)											
R_{best}											

TABLE 5 (Continued)

Panel (a): current year forecast	IMF	Weighted forecast regression	Weighted forecast regression (RW)	Mean	Harmonic mean	Geometric mean	Median	Trimmed mean	Discounted MSFE ($\delta = 0.90$)	Discounted MSFE (δ=0.95)	Discounted MSFE (δ=0.99)	R_{best}
A	0.23	0.19	0.20	0.51	0.26	0.31	0.65	0.56	0.26	0.27	0.30	0.02
B	0.34	0.13	0.13	0.13	0.03	0.11	0.16	0.14	0.14	0.14	0.14	0.11
C	0.96	0.05	0.07	0.12	0.28	0.19	0.29	0.13	0.02	0.02	0.02	0.00
D	0.04	0.00	0.10	0.39	0.34	0.36	0.34	0.45	0.94	0.91	0.89	0.00
E	0.37	0.00	0.00	0.09	0.68	0.79	0.02	0.04	0.00	0.00	0.00	0.00
F	0.49	0.02	0.03	0.00	0.21	0.00	0.01	0.01	0.03	0.02	0.02	0.03
G	0.24	0.00	0.01	0.97	0.32	0.49	0.40	0.83	0.01	0.01	0.01	0.00
H	0.18	0.20	0.20	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
I	0.30	0.08	0.08	0.84	0.47	0.72	0.58	0.77	0.03	0.04	0.05	0.03
MEF	0.22	0.09	0.08	0.56	0.25	0.28	0.93	0.68	0.11	0.12	0.12	0.02
OECD	0.22	0.12	0.12	0.02	0.24	-	0.17	0.07	0.15	0.15	0.15	0.09
EC	0.23	0.18	0.17	0.20	0.15	0.20	0.23	0.21	0.19	0.20	0.20	0.15
IMF		0.20	0.19	0.21	0.21	0.21	0.23	0.22	0.21	0.21	0.21	0.18
Weighted forecast regression			0.53	0.14	0.19	0.17	0.10	0.13	0.79	0.67	0.50	-
Weighted forecast regression (RW)				0.13	0.18	0.16	0.10	0.12	0.71	0.62	0.51	0.31
Mean					0.23	0.22	0.53	0.34	0.17	0.18	0.18	0.08
Harmonic mean						0.23	0.26	0.23	0.21	0.21	0.21	0.16
Geometric mean							0.29	0.23	0.19	0.19	0.20	0.12
Median												
Trimmed mean									0.16	0.16	0.17	0.06
Discounted MSFE ($\delta = 0.90$)										0.05	0.09	0.25
Discounted MSFE ($\delta = 0.95$)											0.13	0.18
Discounted MSFE ($\delta = 0.99$)												0.10
R_{best}												

(Continues)

TABLE 5 (Continued)

Panel (b): year ahead forecast	B	C	D	E	F	G	H	I	MEF	OECD	EC	IMF	Weighted forecast regression	Weighted forecast regression (RW)	Mean	Harmonic mean	Geometric mean	Median	Trimmed mean	Discounted MSFE ($\delta = 0.90$)	Discounted MSFE (δ=0.95)	Discounted MSFE (δ=0.99)	R_{best}
A	0.31	0.15	0.00	0.57	0.47	0.12	0.19	0.83	0.61	0.17	0.29	0.17	0.11	0.30	0.10	0.05	0.11	0.11	0.10	0.11	0.11	0.11	0.05
B		0.00	-	0.15	0.48	0.14	0.19	0.40	0.38	0.22	0.27	0.20	0.16	0.20	0.17	0.14	0.15	0.16	0.16	0.17	0.17	0.17	0.11
C			0.00	0.54	0.03			0.01	0.08	0.78	0.49	0.84	0.08	0.19	0.74	0.21	0.16	0.18	0.06				0.00
D				0.00					0.00	0.00	0.06	0.00	0.00	0.72		0.75							0.22
E					0.90	0.27	0.13	0.60	0.55	0.26	0.32	0.23	0.19	0.34	0.22	0.18	0.17	0.18	0.19	0.20	0.20	0.20	0.13
F						0.01	0.26	0.50	0.16	0.00	0.03	0.02	0.00	0.06	0.00	0.00	0.02	0.02	0.00	0.02	0.01	0.01	0.00
G							0.18	0.00	0.02	0.03	0.14	0.10	0.04	0.16	0.28	0.44	0.06	0.06	0.25	0.08	0.09	0.07	0.02
H								0.30	0.24	0.16	0.23	0.16	0.17	0.18	0.17	0.17	0.16	0.16	0.16	0.17	0.17	0.17	0.16
I									0.79	0.00	0.06	0.01	0.00	0.89	0.27	0.01	0.13	0.01	0.01	0.00	0.00	0.00	0.00
MEF										0.07	0.23	0.08	0.01	0.94	0.08	0.01	0.03	0.03	0.03	0.02	0.01	0.01	0.00
OECD											0.90	0.23	0.95	0.26	0.41	0.78	0.52	0.04	0.04	0.25	0.41	0.78	0.12
EC												0.23	0.88	0.34	0.50	0.78	0.60	0.06	0.06	0.51	0.65	0.90	0.23
IMF													0.35	0.20	0.25	0.33	0.31	0.29	0.34	0.54	0.47	0.41	0.96
Weighted forecast regression														0.12	0.14	0.24	0.26	0.14	0.14	0.14	0.15	0.15	0.00
Weighted forecast regression (RW)															0.11	0.11	0.07	0.13	0.13	0.13	0.13	0.12	0.05
Mean																0.12	0.03	0.14	0.14	0.14	0.14	0.13	0.03
Harmonic mean																	0.26	0.15	0.15	0.16	0.17	0.16	0.01
Geometric mean																		0.17	0.18	0.22	0.23	0.23	0.05
Median																				0.17	0.16	0.16	0.49
Trimmed mean																							
Discounted MSFE (δ =0.90)																						0.18	0.12

TABLE 5 (Continued)

Discounted MSFE (δ =0.95)																						0.21	0.01
Discounted MSFE ($\delta = 0.99$)																							0.00
R_{best}	0.31	0.15	0.00	0.57	0.47	0.12	0.19	0.83	0.61	0.17	0.29	0.17	0.11	0.30	0.10	0.05	0.11	0.11	0.10	0.11	0.11	0.11	0.05

Notes: Each cell shows the *p*-value for the null hypothesis of the DMW test that the forecast accuracy of both models is equal. DMW, Diebold–Mariano–West; IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

Rossi and Sekhposyan (2016) propose a general Fluctuation Rationality test. This test for relative forecasting performance is a Wald test for forecasting properties in a rolling window regression of the tests shown in Section 3.1. The null hypothesis of stability of the specific test is rejected if the supremum value of the Wald test statistic series is larger than the critical value.

We choose a rolling window of 10 years to evaluate forecaster performance and set significance at 5% (the corresponding critical value is 9.1724 in this case). Table 6 shows the maximum of the test statistic series, whereas Figures 3 and 4 show the history of the rolling test statistic, together with the full sample critical value for current year and year-ahead forecasts.

For current year forecasts, the Fluctuation Rationality test rejects the null of forecast rationality in all cases, with a few exceptions. Forecaster A is able to keep forecasts rational at all horizons. This might be the consequence of good performance: in Table 4, we showed that forecaster A produces a particularly well performing budget prediction overall. Figure 3 plots the sequence of rationality tests over time, and forecaster A remains below the critical value line. Among the other private forecasters, performance declines either after the Sovereign Crisis (forecasters B, E, H, and I), or at the start of the Pandemic (forecasters C or G). It is perhaps surprising to see that public forecasts are not found to be rational, yet the picture is quite different. The EC certainly stands out for producing a rational forecast, except at the start of the Pandemic, but even during the Sovereign Crisis, the EC did not lose track. Similarly, the Ministry of Economy and Finance kept a good record of budget deficit forecasts till the start of the Pandemic, whereas the OECD (during the Sovereign Crisis) and the IMF (over the full sample) have trouble producing rational forecasts.

It would be hard to anchor forecasts on any of these forecasters, yet in this case, an encouraging result is that several combinations of forecasts produce rational forecasts over the full sample. This is the case for a weighted forecast combination (including the random walk) and any of the discounted forecasts—including the Recent Best forecast. While some of these combinations still had problems in capturing the deviation of forecasts during the Sovereign Crisis, these combinations manage to stay rational throughout the entire sample period. For some of the simpler combination methods, there is evidence against forecast rationality at the start (Sovereign Crisis) or the end of the sample (Pandemic).

A combination of forecasts aggregates information and reduces uncertainty by eliminating judgment errors on structural changes. The outcome is still based on the

TABLE 6 Fluctuation test for relative forecasting performance.

	Current year	Year ahead
Private		
A	7.82	14.99
B	20.12	38.77
C	14.12	18.68
D	-	-
E	26.48	11.77
F	23.40	392.58
G	87.03	63.61
H	33.10	65.78
J	82.58	34.56
Public		
MEF	52.98	46.82
OECD	27.32	129.84
EC	11.19	29.76
IMF	28.57	8.87
Combinations		
Mean	16.63	4.47
Harmonic mean	13.63	4.28
Geometric mean	16.41	26.14
Trimmed mean	10.35	18.14
Median	15.99	20.62
Weighted forecast regression	17.54	4.28
Weighted forecast regression (RW)	7.97	25.06
Discounted MSFE ($\delta = 0.90$)	4.52	47.98
Discounted MSFE ($\delta = 0.95$)	4.47	43.27
Discounted MSFE ($\delta = 0.99$)	5.01	36.05
R_{best}	8.58	196.44

Notes: The table reports the maximum value attained by the (rolling) test statistic over the sample of the Rossi and Sekhposyan (2016) Fluctuation Rationality test. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.
*A rejection of the null at 5% for a test of stable forecasting rationality (with a window of 10 years, and three lags). The critical value at 5% is 9.17240.

global performance of forecasters, however, and not on the change in performance over time of different competing forecasts. One of the reasons for the good performance of R_{best} or the weighted forecast combinations is that we select the best performing forecasters by their RMSE over the last couple of years. This time frame is sufficiently narrow to eliminate any forecasters that after structural changes do not update their forecasts. We observed in Figures 1 and 2 that forecasting performance on the Italian budget deficit changed over time. Up to 2001, most forecasters performed quite well, and projections were mostly aligned with actual budget outcomes. Afterwards, performance has diverged.

As would be expected, predicting the budget for the next year is much harder. Table 6 shows that only the IMF is a rational forecaster, and that only a simple mean combination or a weighted forecast combination model is significantly "rational." This is the outcome of all forecasters having difficulties of maintaining a good forecasting performance over the sample. In fact, the "success" of forecaster A—and to some extent forecaster E—is entirely because of a good prediction during the Pandemic. Although the international institutions tend to perform well over the entire sample, the EC or the OECD miss the Pandemic, but the IMF does not, making it the best performer for year-ahead budget forecasts throughout. By contrast, the Ministry of Economy and Finance is rejected over the entire sample to produce efficient forecasts.

The combinations of forecasts reflect these outcomes too. The mean and the Recent Best combinations cannot eliminate the effects of the bad forecasts during the Pandemic, so the mistakes by most forecasters in 2019 spill over to most combinations too. This is most visible in the behavior of the Recent Best forecast that suddenly spikes up. The only model to protect against this change is the weighted forecast combination (including a RW or not) as it attributes some weight to outlier forecasts.

These results thus indicate that in spite of the negative finding in Section 3.1 that forecast combinations do not result in unbiased and efficient forecasts, and that information rigidities stay present, looking at the absolute performance over time allows us to obtain combinations that are unbiased and efficient. Nevertheless, and perhaps unsurprisingly, the results also show that updating the information (by weighing more strongly recent "strong" forecasts) is going to improve performance. As in the literature, predicting well in recent periods will result in better forecasts over later periods, or seen in another perspective, fewer information rigidities could substantially improve budget forecasts, implying forecasters should have access to timely and correct budget data, next to paying attention to the budget process.

4.2 | A fluctuation test for relative forecasting performance

As even expert forecasters are unable to anticipate all economic and political changes, forecasting models have to be adaptive. Finding an indicator that predicts well in

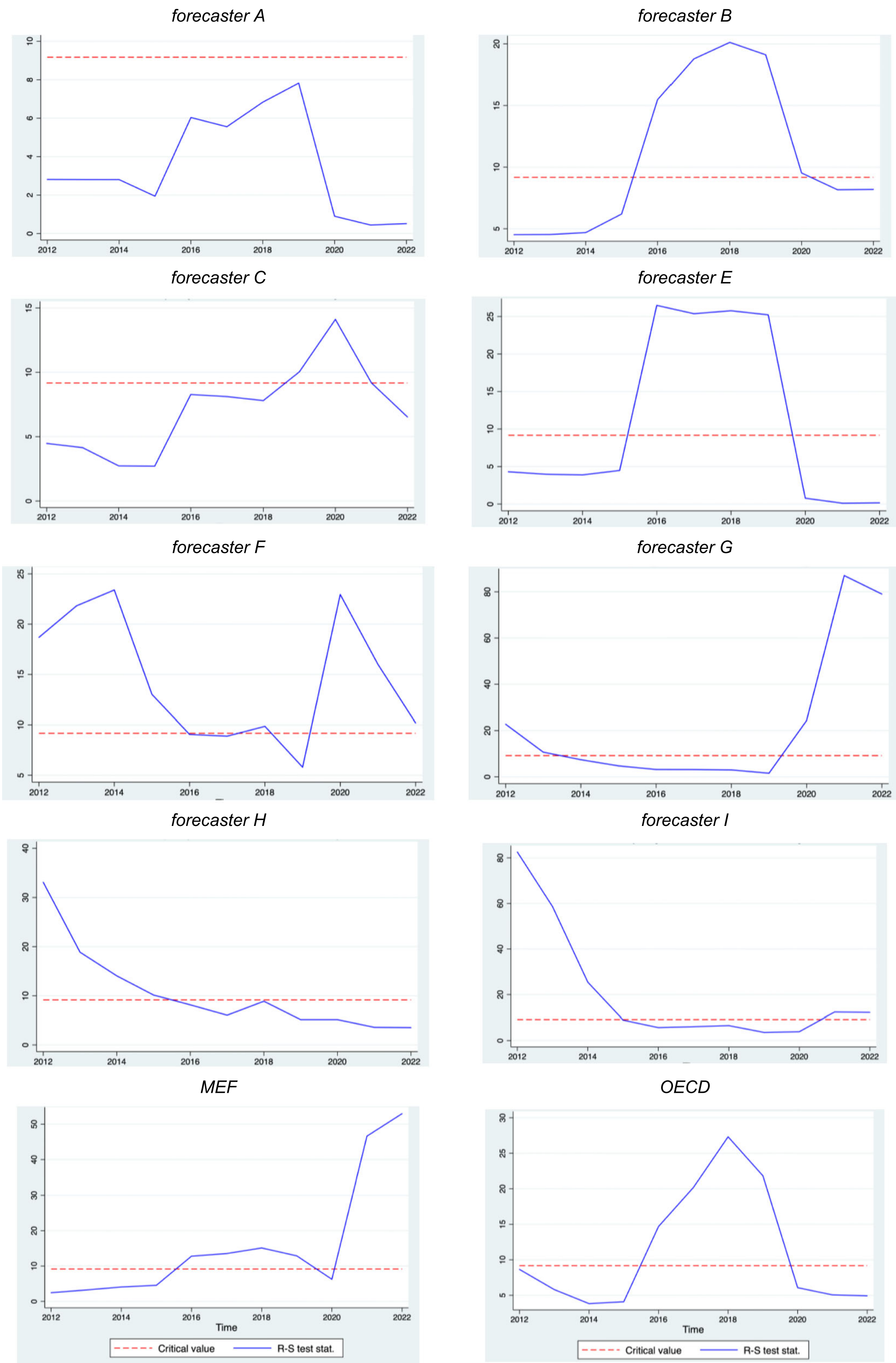

FIGURE 3 Rossi and Sekhposyan (2016) Fluctuation Rationality test (current year forecasts).

one period is no guarantee that it will predict well in later periods. This explains the success of simple time series models in forecasting fiscal variables (Favero & Marcellino, 2005). More generally, expert forecasters using the same model are unlikely to outguess other experts at all points in time. Rather, the best forecasting

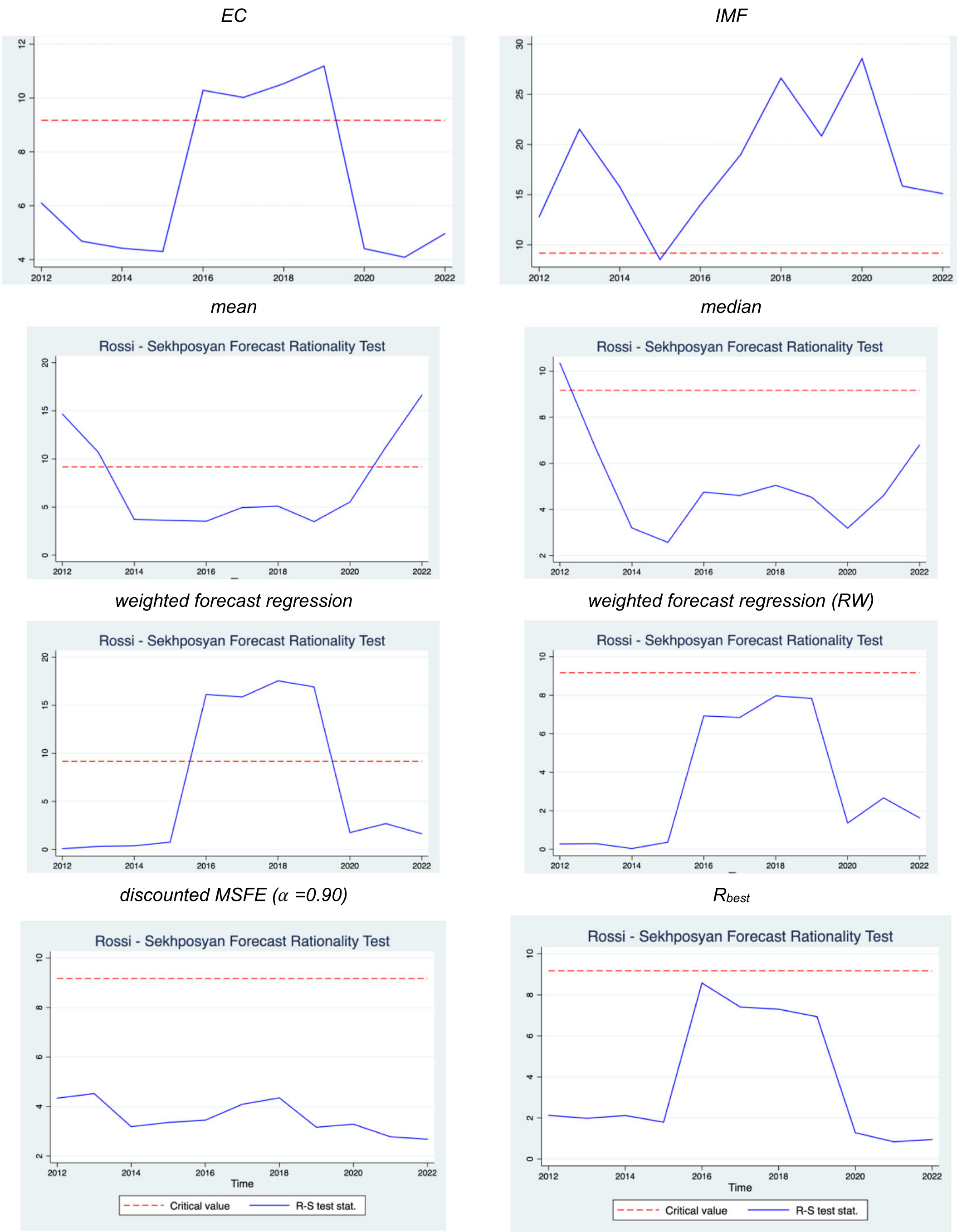

FIGURE 3 (Continued)

model may change over time in ways that can be difficult to track on the basis of past forecasting performance (Timmermann, 2006). The results in Section 3.2 already pointed out that models weighing more recent forecasts tend to perform better.

If the predictive accuracy of a model relative to a competitor forecaster appears very much connected to some specific period of time, we would like to test if predictive accuracy changes over time. Giacomini and Rossi (2010) develop two tests that examine the fluctuations in relative predictive performance of two forecasting methods X and Y. Each method produces a sequence of out-of-sample forecasts based on a rolling window of m observations used for constructing the forecasting model at each point in time. At each point in time, we can then compute the difference in the loss of the two models as

$$\left\{ \Delta L_t \left(\widehat{\beta}_{t-h,R}, \widehat{\theta}_{t-h,R} \right) \right\}_{t=R+h}^{T} \tag{9}$$

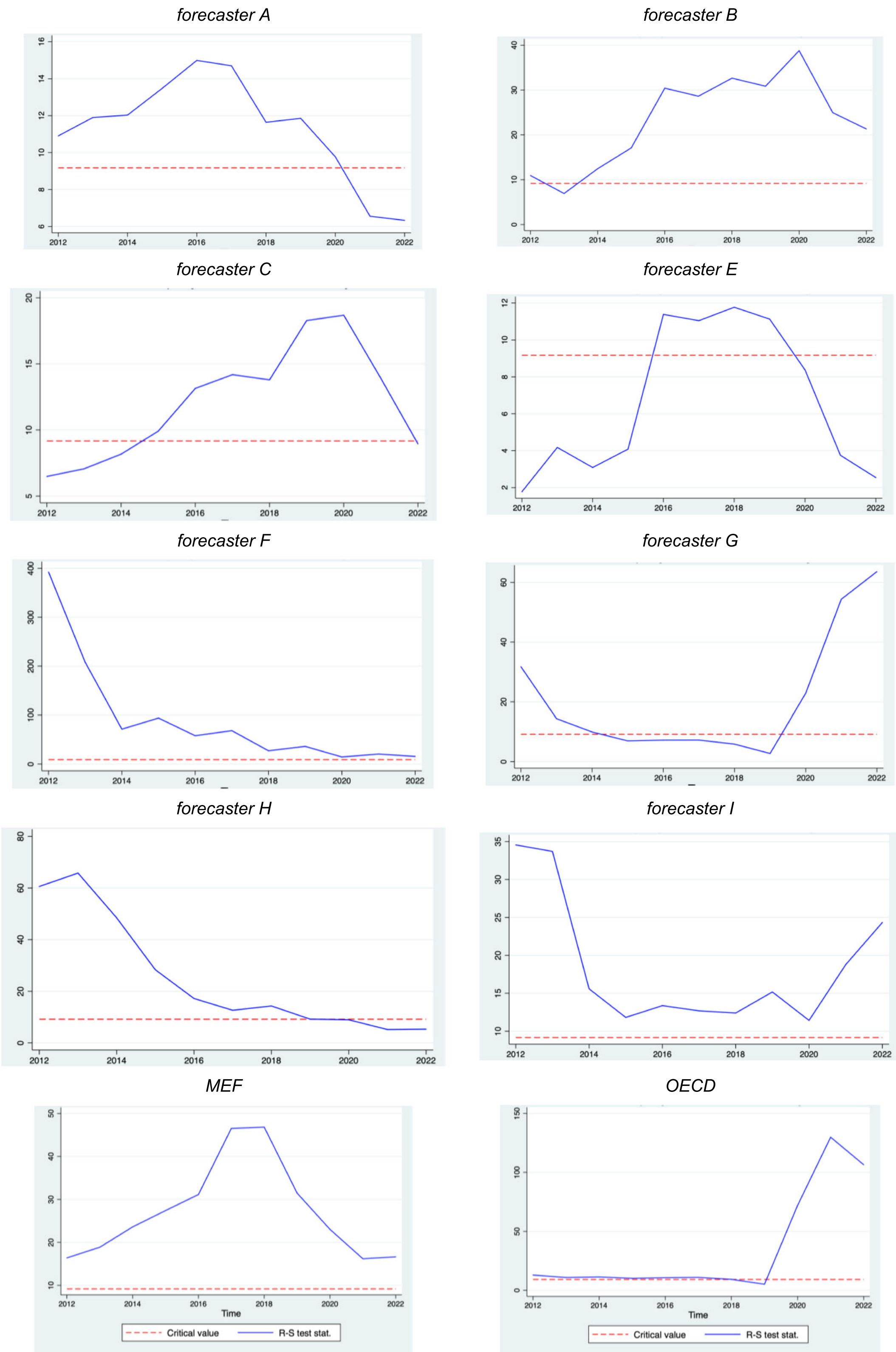

FIGURE 4 Rossi and Sekhposyan (2016) Fluctuation Rationality test (year-ahead forecasts).

that depend on the realizations of the variable and on the in-sample estimates for each model re-estimated at each time $t = R + h,\ldots T$ over a window of size R. The local relative loss for the two models is defined over centered rolling windows of size m as

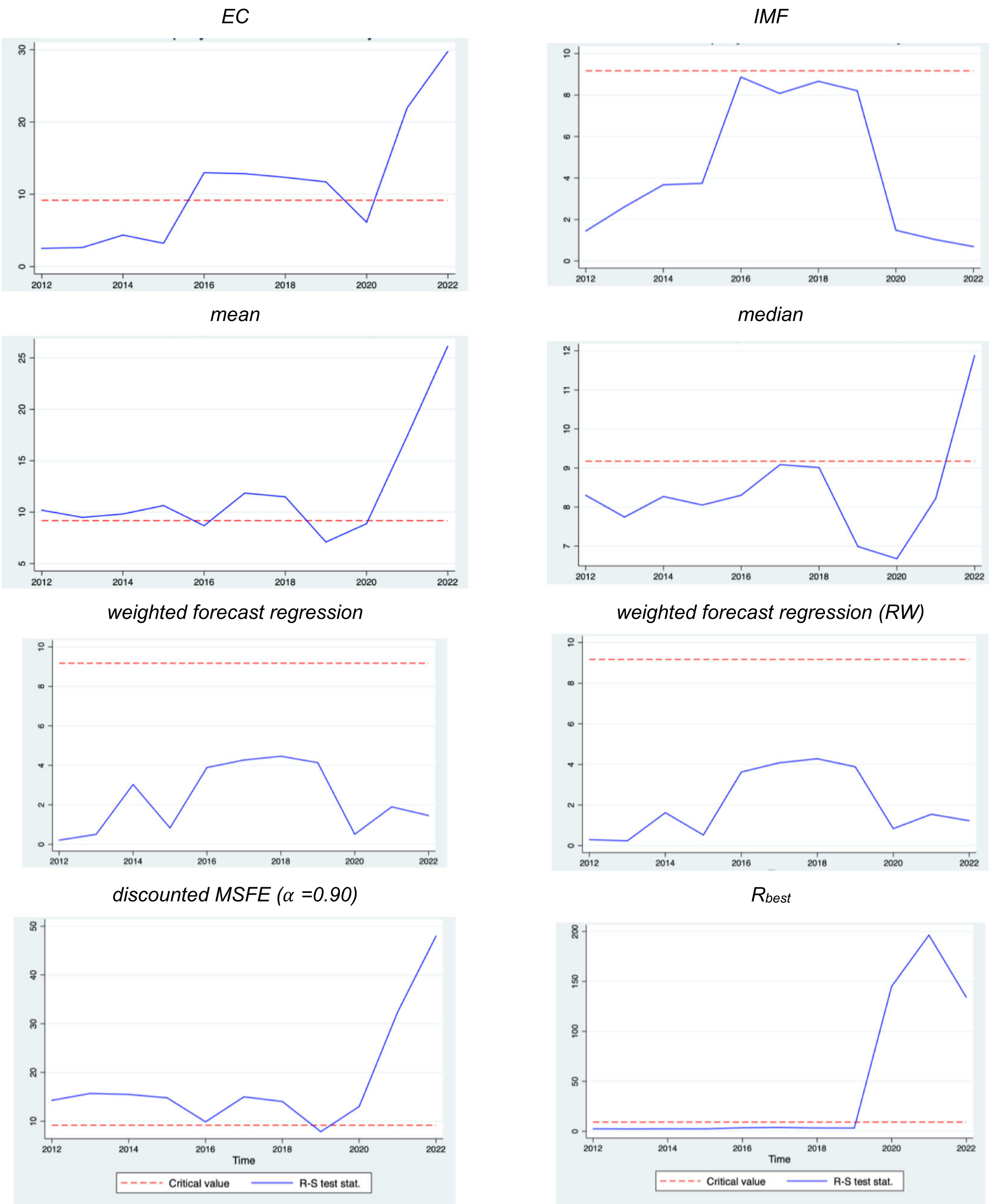

FIGURE 4 (Continued)

$$\frac{1}{m}\sum_{j=t-m/2}^{t+\frac{m}{2}-1}\Delta L_j\left(\widehat{\beta}_{j-h,R},\widehat{\theta}_{j-h,R}\right) \tag{10}$$

The sequence produced by repeated application of Equation (6) allows to evaluate at each point in time the relative performance of both models. The fluctuation test statistic by Giacomini and Rossi (2010) is then defined as

$$F_{t,m}=\widehat{\theta}^{2}m^{1/2}\sum_{j=t-m/2}^{t+\frac{m}{2}-1}\Delta L_j\left(\widehat{\beta}_{j-h,R},\widehat{\theta}_{j-h,R}\right) \tag{11}$$

where the null hypothesis is that

$$H_0=E\Delta L_t\left(\widehat{\beta}_{t-h,R},\widehat{\theta}_{t-h,R}\right)=0 \tag{12}$$

against a two-sided alternative that relative forecasting performance is not similar. As with a structural break test, if the difference in performance exceeds a certain threshold in some time period, the null is rejected.[13] Giacomini and Rossi (2010) derive the critical values for testing the null hypothesis that the local relative MSFE equals zero at each point in time.

TABLE 7 Test for equal predictive ability for forecast (combinations) using the Giacomini and Rossi's (2010) fluctuation test.

	MEF	OECD	EC	IMF	Mean	Harmonic mean	Geometric mean	Trimmed mean	Median	Weighted forecast regression	Weighted forecast regression (RW)	Discounted MSFE ($\delta = 0.90$)	Discounted MSFE (δ=0.95)	Discounted MSFE ($\delta = 0.99$)	R_{best}
	a. Current year forecast														
A	4.36	3.02	5.62	2.91	3.20	2.24	2.71	2.64	3.53	3.61	2.85	3.20	3.18	3.19	3.26
B	2.77	4.42	3.06	3.42	3.46	3.87	3.68	3.60	3.00	2.88	3.56	3.58	3.54	3.50	3.45
C	10.81	7.34	5.45	10.72	4.25	3.65	3.09	13.55	8.81	7.19	3.51	5.08	5.06	5.15	7.00
D	1.41	1.31	4.76	4.49	4.76	0.45	4.86	5.45	5.52	5.58	4.67				
E	5.04	5.80	4.21	3.75	4.72	7.98	5.44	7.20	4.76	4.49	4.86	5.45	5.52	5.58	4.67
F	4.54	4.76	4.03	4.30	4.86	4.75	4.83	4.41	4.11	4.30	4.64	4.53	4.52	4.50	4.72
MEF	-	-	-	-	2.44	1.93	2.08	2.11	2.16	2.13	2.22	3.10	2.50	2.41	3.98
OECD	-	-	-	-	5.67	4.50	5.10	4.18	6.69	6.27	5.36	5.48	5.48	4.84	6.13
EC	-	-	-	-	5.38	2.48	3.95	5.65	3.34	2.29	4.95	6.00	5.27	4.63	0.00
IMF	-	-	-	-	4.05	3.79	4.49	4.79	2.80	3.33	4.34	4.00	4.02	4.04	0.03
R_{best}	3.98	6.13	4.19	3.06	3.00	2.95	2.85	4.09	4.52	4.01	2.86	4.91	0.33	3.21	-
									b. Year ahead forecast						
A	4.37	4.79	3.05	4.96	3.15	2.88	2.41	3.67	3.49	5.00	5.05	3.63	3.48	3.30	5.91
B	2.60	3.93	3.82	3.85	3.25	3.28	3.43	3.31	3.17	6.60	6.86	3.59	3.54	3.50	4.97
C	4.68	5.64	4.21	5.21	4.51	2.31	6.43	5.36	5.34	5.22	6.31	4.66	4.40	4.21	5.64
E	3.47	6.25	3.03	7.87	2.74	3.62	2.38	3.68	2.62	6.89	7.31	3.30	3.30	3.31	4.16
MEF	-	6.01	4.32	4.74	4.90	4.05	3.94	4.30	5.03	5.35	5.49	5.57	6.17	6.13	6.39
OECD	6.01	-	4.43	1.81	5.35	3.71	4.25	4.51	4.25	5.21	5.49	5.23	5.30	5.34	3.90
EC	4.32	4.43	-	3.25	3.40	3.18	2.74	3.60	3.27	6.12	6.10	3.82	3.75	3.69	8.56
IMF	4.74	1.81	3.25	-	6.16	3.85	4.31	5.31	4.54	3.55	3.40	4.55	4.73	4.84	3.79
R_{best}	6.39	3.90	8.56	3.79	5.35	5.44	5.39	4.25	5.08	2.44	2.71	4.49	4.59	4.66	-

Notes: The table reports the maximum absolute value of the (rolling) test statistic of the Giacomini and Rossi's (2010) Fluctuation Test over the sample. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

*A rejection of the null at 5% for a two-sided test is of equal predictive ability (with a window of 10 years).

FIGURE 5 Fluctuation test for relative forecasting performance of current year budget forecasts over time by Giacomini and Rossi (2010).

We compute the MSFE differences over a rolling window of 10 years and test the null hypothesis that the MSFE is equal to zero for each private forecaster relative to the public forecasters, any of the combination forecasts, and a naive AR(1) process. If the relative MSFE exceeds the critical value in some part of the sample, we reject the null hypothesis and we conclude that there are periods during the sample that one model outperforms the other. A visual check of the relative MSFE will show the periods in which differences in performance are significant.

Table 7 reports the critical values for each of the set of forecasts we compare, and a rejection of the null hypothesis implies that the models' forecasting performance is significantly better for the first model (the row of Table 7) than for the second one (the column of Table 7). The critical value is in all cases 2.89 at 5% significance level. Our sample is limited as a few forecasters did not produce budget forecasts over a sufficiently long period of time to compute time-varying performance, as was the case for forecasters G, H, and I (in the current year) and in addition forecasters D and F (for the year-

FIGURE 5 (Continued)

ahead forecast). We limit the analysis to a comparison of each public/private forecast to the different forecast combinations.

Panel (a) of Table 7 reveals a few striking findings in comparison to the full sample tests in Section 3.2. First, if we first compare the private forecasters to the public ones, the relative performance of the latter is generally better. There are just a few exceptions when forecaster A or D perform better than the MEF, and forecaster D than the OECD.

Second, we observe a similar behavior for the comparison against the combination of forecasts. Private forecasters now generally perform better than any of the combinations. The only exception is forecaster A who performs worse as compared to any of the mean forecasts or the weighed forecast combination (including a random walk), or forecaster D against the harmonic mean.

Third, if we do the same exercise for public forecasters, then the MEF stands apart from the other forecasters: it consistently underperforms against any combination, except the Recent Best one. The EC and IMF consistently perform better than any of the combinations, except the Recent Best one, and occasionally against a mean, median, or weighted forecast

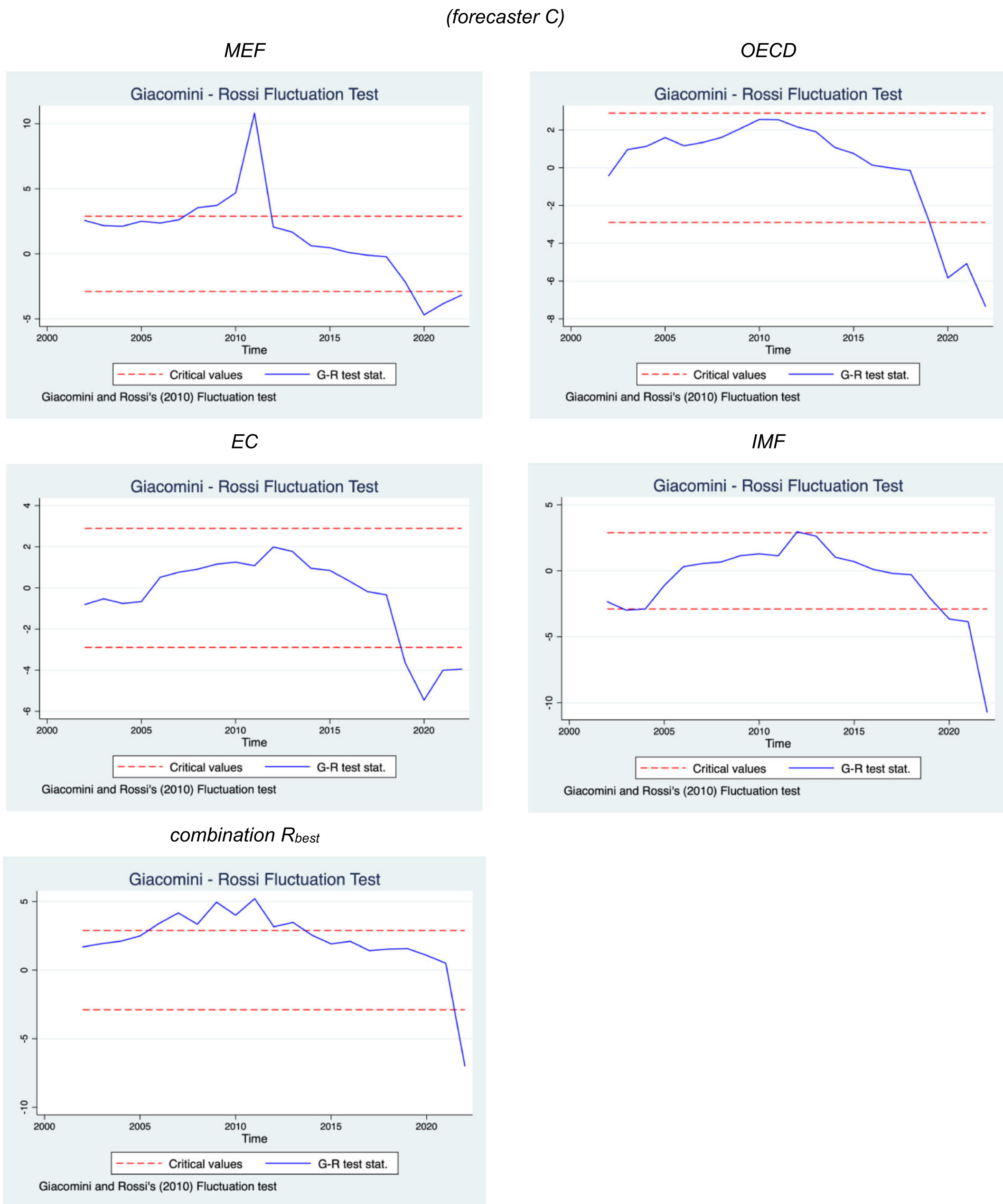

FIGURE 5 (Continued)

combination. Finally, given that the Recent Best combination was found to be outperforming most forecasts or their combinations, we also look in its relative performance and find that while it outperforms any of the private or public forecasts taken separately, it does not do so in a consistent way with all the combinations of forecasts.

Figure 5 allows us to look in the period in which relative performance changes. It plots the sequence of the two-sided test statistics over time and both critical value lines. We compare each forecaster against several other ones. For forecaster A, we can see that the good performance of this forecaster is entirely because of the period before the Global Financial Crisis, and its performance is significantly better only in this period, but then gradually fades away. A similar change happens for forecaster B or F. Instead, forecaster C performs well during the Global Financial Crisis, but as the crisis unfolds, all public forecasters are actually doing significantly better during the Sovereign Crisis and the Pandemic. Forecaster E instead keeps up a good performance throughout the sample.

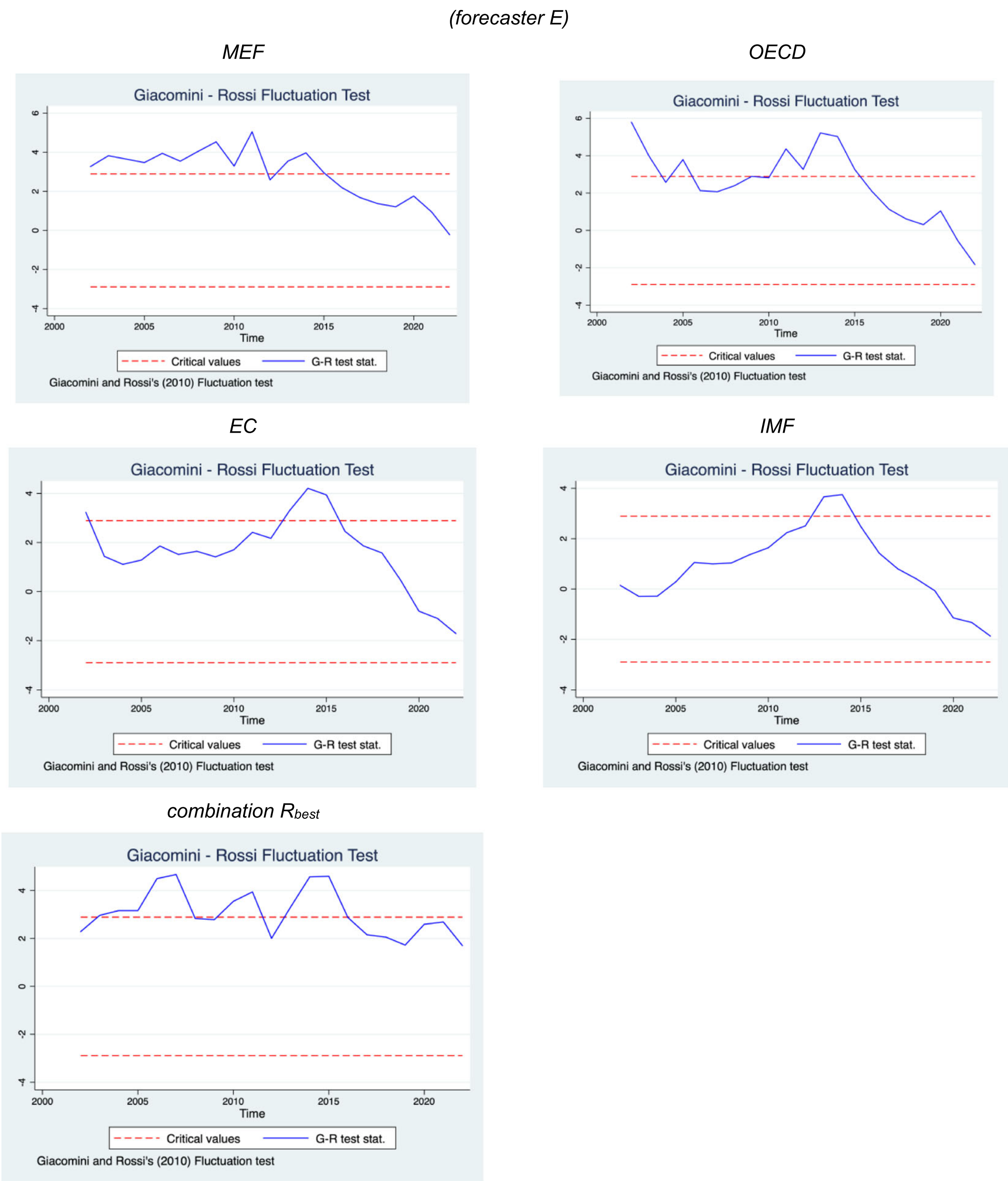

FIGURE 5 (Continued)

It is interesting to look now also at the change in relative performance of the public forecasters: the MEF is never significantly doing better or worse than the other institutions, and the international organizations can deliver significantly better, but this only occurs in specific time periods.

Finally, if we look at the Recent Best combination forecast, we see in Figure 5 that its performance is borderline significant and stays relatively constant over time, as would be expected given that this combination selects the best performing forecasters over the last 4 years. It does not outperform as it averages over those forecasters, rather than selecting one winner. The gain of using this combination can be seen at the start of the Pandemic: the strongest evidence against the null appears to be around 2020; this is when the empirical evidence in favor of the Recent Best combination is the strongest.

Panel (b) of Table 7 shows us the rolling relative MSFE for the year-ahead forecasts. Given the static results discussed in Section 3.2, the much less significant outcomes for the year-ahead forecast of the net lending ratio should not come as a surprise. In fact, we can hardly detect any significant improvement in performance for any of the forecasters or its combinations at 5%.[14] The graphs of the rolling test statistics never showed significant results, so we do not report those in the text. It

FIGURE 5 (Continued)

underlines the difficulty of providing good budget forecasts for a horizon over 18 months.

These empirical conclusions are very different from those obtained in Table 5, in which the DMWP test ignores the time variation in relative forecasting performance. The Fluctuation Test allows us to confirm the findings in the literature that budget forecasts suffer mostly from large changes in fiscal policy, because of major economic events, yet although not all forecasters produce interesting forecasts, a combination that pays attention to recent performance can improve the outcomes, at least within the year.

5 | CONCLUSIONS

Despite the growing importance of fiscal projections in the short-term to inform policymakers, control fiscal monitoring, and manage expectations, practitioners seem to require a lot of judgment in making better fiscal projections. We show that exploiting the information from many different forecasters can still lead to substantial gains in predictive accuracy. Applying different combination techniques to the current year and year-ahead forecasts of the Italian budget deficit over the period 1993 to 2022 from both private forecasters (from Consensus Economics) or public forecasters (EC, OECD, IMF, and the MEF), we find that forecast combinations do not necessarily result in less biased or more efficient forecasts, and they might still suffer from information rigidities. Nevertheless, forecast combinations that take into account the better performing forecasters, especially over recent periods, and using models robust to recent changes, perform better.

This can be particularly seen from examining absolute and relative forecasting performance of budget deficits over time. Using time-varying rationality tests, we find that some forecasters—in particular from international institutions—manage to keep their forecasts efficient and unbiased in spite of economic headwinds. While budget forecasts suffer mostly from large changes in fiscal policy, because of major economic events, and while not all forecasters produce reliable forecasts, a combination of budget forecasts that pays attention to recent performance can significantly improve outcomes.

ACKNOWLEDGMENTS

We thank Barbara Rossi, Domenico Giannone and participants at the 14th IWH-CIREQ Macroeconometrics Workshop, the III Time Series Conference, the Centro Europeo di Ricerche, the 34th International Symposium on Forecasting, the 23rd ICMAIF conference, and AQR-IREA for useful comments and suggestions. The usual disclaimer applies.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Peter Claeys https://orcid.org/0000-0003-2263-4250

ENDNOTES

[1] Further information on how the survey is conducted is available via the Internet: www.consensuseconomics.com.

[2] For Italy, specialists forecast the general budget balance for the calendar (end of the) year.

[3] Note that revisions to the net lending ratio are subject to adjustment for a couple of years after its first publication.

[4] In fact, correlations between current-year forecasts never drop below 0.80, and for year-ahead forecasts below 0.50, with the exception of forecaster H who follows a rather different path. Correlation matrices are available in Appendix A.

[5] Many authors have approached the pooling of forecasts. Zamowitz (1967) noted that the published averages of inflation and GNP growth forecasts was better than the individual ones. Bates and Granger (1969) discovered that the simple average outperform the forecasts taken individually. The idea was also to use the relative combination of variances and covariances to construct a weighted average of the forecasts that minimizes the mean square error of the combined forecasts. Likewise, Nelson (1972) and Cooper and Nelson (1975) showed that the combination of forecasts with ARIMA estimates produces a smaller error compared to the models alone. The suggested reasons for the better performance of ARIMA models in their paper are the incapacity of econometric models to arrange structural changes in the economy. Granger and Newbold (1973) also start from the similar point in terms of forecast evaluation. Makridakis (1989) studied a large variety of time series forecasting methods which were applied to 1.001 different economic time series. The forecast performance was measured using various error summary measures. Two different combining schemes were studied: both of these combinations performed well relative to the individual techniques, with the simple average having the better performance of the two. Clemen (1989) provided a very deep review of the methods used in combining and confirming these results. Clemens and Winker (1989) give root to a combination in their philosophical approach.

[6] Note that estimating the combination weights might induce uncertainty, especially when the sample size is small relative to the number of forecasts (Elliott, 2004).

[7] Batchelor (2001) shows that CEF forecasts are less biased and more accurate in terms of mean absolute error and root mean square error than OECD and IMF forecasts. Dovern and Weisser (2011) also find that the participants in the CEF poll provide rational and unbiased inflation and growth forecasts for the G7 countries.

[8] For the details of the regressions to compute weighted forecast combinations, see Appendix B.

[9] We also considered several other models for producing combinations, yet the outcomes in all cases were very similar to the ones obtained with a weighted forecast combination, the reason being that these alternatives attribute more weight to the better performing forecasters. We used a cluster combination following Aiolfi and Timmermann (2006), yet this method gave identical results as the trimmed mean. Instead, a principal components models on the set of 13 expert forecasts (and an augmented one including a random walk) failed picking up the level of net lending, even if it managed to track the forecasts in its changes. We therefore discarded including these forecast combinations in the further analysis.

[10] In fact, the correlation between forecast (combinations) never falls below 0.60 (see Appendix A).

[11] A Wooldridge test for autocorrelation in the panel of private, public and the combination of forecasts shows that forecast errors are serially correlated.

[12] The Theil test we apply compares the forecast error of the model versus the one of a naïve RW forecast; hence, a value of 1 is the threshold for forecasting accuracy against the naïve RW benchmark.

[13] The fluctuation test is in fact the Andrews–Ploberger version of the DM test. As all instability tests, the fluctuation test may suffer from low power as it does not specify an alternative hypothesis, so we do not know in which direction forecasters may diverge.

[14] With the exception of forecaster A or the OECD doing better than a single other combination forecast or the IMF, respectively.

REFERENCES

Artis, M., & Marcellino, M. (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *Ecopermetrics Journal*, *4*, 1–23. https://doi.org/10.1111/1368-423X.00051

Asimakopoulos, S., Paredes, J., & Warmedinger, T. (2013). Forecasting fiscal time series using mixed frequency data (No. 1550). ECB Working Paper.

Batchelor, R. (2001). How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. *Applied Economics*, *33*, 225–235.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*, 451–468.

Beetsma, R., & Giuliodori, M. (2010). Fiscal adjustment to cyclical developments in the OECD: An empirical analysis based on real-time data. *Oxford Economic Papers*, *62*(3), 419–441.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Clemen, R. T., & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business and Economic Statistics*, *4*, 39–46.

Cooper, J. P., & Nelson, C. R. (1975). The ex-ante prediction performance of the St. Louis and FRB-MIT-PENN econometric models, and some results on composite predictions. *Journal of Money, Credit and Banking*, *7*, 1. https://doi.org/10.2307/1991250

Dovern, J., Fritsche, U., Loungani, P., & Tamirisa, N. (2015). Information rigidities: Comparing average and individual forecasts for a large international panel. *International Journal of Forecasting*, *31*(1), 144–154. https://doi.org/10.1016/j.ijforecast.2014.06.002

Dovern, J., & Weisser, J. (2011). Accuracy, unbiasedness, and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7. *International Journal of Forecasting*, *27*, 452–465.

Favero, C. A., & Marcellino, M. (2005). Modelling and forecasting fiscal variables for the Euro area. *Oxford Bulletin of Economics and Statistics*, *67*, 755–783.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*, 595–620.

Granger, C. W. J., & Newbold, P. (1973). Some comments on the evaluation of forecasts. *Applied Economics*, *5*, 35–47.

Hendry, D. F., & Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, *7*(1), 1–31.

Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. *The Manchester School of Economic & Social Studies*, *58*(2), 120–127.

Jalles, J. T., Karibzhanov, I., & Loungani, P. (2015). Cross-country evidence on the quality of private sector fiscal forecasts. *Journal of Macroeconomics*, *45*, 186–201.

Jonung, L., & Larch, M. (2006). Fiscal policy in the EU: Are official output forecasts biased? *Economic Policy*, *21*(47), 491–534.

Keane, M. P., & Runkle, D. E. (1990). Testing the rationality of price forecasts: New evidence from panel data. *American Economic Review*, *80*, 714–735.

Leal, T., Pérez, J. J., Tujula, M., & Vidal, J. P. (2008). Fiscal forecasting: Lessons from the literature and challenges. *Fiscal Studies*, *29*(3), 347–386.

Lupoletti, V. M., & Webb, R. H. (1986). Defining and improving the accuracy of macroeconomic forecasts: Contributions from a VAR model. *Journal of Business*, *5*(9), 263–285.

Makridakis, S. (1989). Why combining works? *International Journal of Forecasting*, *5*(4), 601–603. https://doi.org/10.1016/0169-2070(89)90017-4

Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve. *The Quarterly Journal of Economics*, *117*(4), 1295–1328.

Nelson, C. R. (1972). The prediction performance of the F.R.B.-M.I.T.-PENN model of the U.S. economy. *American Economic Review*, *62*, 902–917.

Onorante, L., Pedregal, D. J., Pérez, J. J., & Signorini, S. (2010). The usefulness of infra-annual government cash budgetary data for fiscal forecasting in the euro area. *Journal of Policy Modeling*, *32*(1), 98–119.

Ottaviani, M., & Sorensen, P. (2006). The strategy of professional forecasting. *Journal of Financial Economics*, *81*, 441–466.

Pedregal, D. J., & Pérez, J. J. (2010). Should quarterly government finance statistics be used for fiscal surveillance in Europe? *International Journal of Forecasting*, *26*(4), 794–807.

Pesaran, M. H., & Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, *129*(1–2), 183–217.

Poplawski-Ribeiro, M. M., & Rülke, J. C. (2011). Fiscal expectations under the stability and growth pact: Evidence from survey data. International Monetary Fund.

Rossi, B., & Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to Federal Reserve and survey forecasts. *Journal of Applied Econometrics*, *31*(3), 507–532.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690.

Stock, J., & Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*, 405–430.

Timmermann, A. (2006). Forecast Combinations. In G. Elliot, C. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. I, pp. 136–196). North-Holland. https://doi.org/10.1016/S1574-0706(05)01004-9

Trueman, B. (1994). Analyst forecasts and herding behaviour. *Review of Financial Studies*, *7*(1), 97–124.

Zamowitz, V. (1967). *An appraisal of short-term economic forecasts*. National Bureau of Economic.

AUTHOR BIOGRAPHIES

Laura Carabotta is a PhD economist and senior advisor at Ministry of Economy and Finance in Italy (Rome) on monitoring, assessing, and forecasting public finance. She is also a researcher in economics particularly related to forecasting methods and econometric analysis on public finance. She obtained her PhD at the University of Barcelona (Spain).

Peter Claeys is an Associate Professor at the Department of Economics of the Universidad Pontificia Comillas. He is also a Visiting Professor at the Department of Economics of the College of Europe and the Universiteit Gent. He obtained his PhD at the European University Institute, and after a Marie Curie Fellowship at AQR, he became an Assistant Professor at the Universitat de Barcelona. His work focuses on the econometric modeling of various aspects of fiscal policy, including the macroeconomic effects of government spending, fiscal consolidation, spillover of budget decisions, fiscal federalism, and more generally the interaction between monetary and fiscal policies. He has worked on various projects on fiscal policy for the European Commission, the ECB, the OECD, the European Parliament, and the Spanish and Swedish governments. He teaches in the Erasmus Mundus Master on Globalization and European Integration, and he is a Research Fellow at the United Nations University.

How to cite this article: Carabotta, L., & Claeys, P. (2024). "Combine to compete: Improving fiscal forecast accuracy over time". *Journal of Forecasting*, 1–35. https://doi.org/10.1002/for.3058

APPENDIX A: Correlation matrix of forecasts

Year ahead / Current year	A	B	C	D	E	F	G	H	J	MEF	OECD
A	-	0.83	0.96	0.99	0.83	0.83	0.73	0.54	0.75	0.86	0.70
B	0.82	-	0.96	0.96	0.89	0.60	0.59	0.57	0.61	0.91	0.63
C	0.96	0.97	-	0.99	0.94	0.63	0.68	0.74	0.70	0.96	0.94
D	0.97	0.99	0.99	-	0.98	0.99	0.99	0.99	0.99	0.98	0.98
E	0.90	0.91	0.97	0.99	-	0.59	0.55	0.39	0.43	0.87	0.67
F	0.87	0.85	0.85	0.95	0.84	-	0.83	0.43	0.72	0.68	0.71
G	0.90	0.66	0.54	0.99	0.81	0.81	-	0.11	0.89	0.85	0.93
H	0.07	0.38	0.71	−0.99	0.18	0.24	−0.25	-	0.19	0.28	0.08
J	0.93	0.74	0.72	0.99	0.82	0.86	0.98	0.03	-	0.84	0.90
MEF	0.95	0.87	0.98	0.99	0.91	0.80	0.95	0.03	0.97	-	0.84
OECD	0.91	0.86	0.98	0.99	0.93	0.73	0.84	−0.13	0.83	0.93	-
EC	0.89	0.93	0.99	1.00	0.95	0.81	0.72	0.12	0.79	0.87	0.95
IMF	0.76	0.93	0.97	0.99	0.91	0.75	0.49	0.18	0.55	0.79	0.90
Mean	0.96	0.95	0.99	1.00	0.97	0.91	0.92	0.14	0.95	0.94	0.96
Harmonic mean	0.92	0.96	0.99	1.00	0.96	0.92	0.83	0.25	0.89	0.89	0.93
Geometric mean	0.95	0.96	0.99	1.00	0.97	0.92	0.89	0.18	0.93	0.92	0.95
Trimmed mean	0.96	0.95	0.99	1.00	0.97	0.90	0.92	0.12	0.96	0.95	0.96
Median	0.96	0.95	0.99	1.00	0.97	0.90	0.91	0.14	0.94	0.95	0.96
Weighted forecast regression	0.95	0.88	0.98	1.00	0.94	0.80	0.94	−0.01	0.94	0.99	0.96
Weighted forecast regression (RW)	0.95	0.85	0.97	0.99	0.92	0.81	0.94	−0.02	0.95	0.99	0.95
Discounted MSFE ($\delta = 0.90$)	0.98	0.88	0.99	1.00	0.94	0.87	0.96	0.06	0.97	0.98	0.95
Discounted MSFE ($\delta = 0.95$)	0.98	0.89	0.99	1.00	0.94	0.87	0.95	0.07	0.97	0.98	0.95
Discounted MSFE ($\delta = 0.99$)	0.98	0.89	0.99	1.00	0.95	0.88	0.95	0.08	0.97	0.98	0.95
R_{best}	0.96	0.81	0.93	0.97	0.90	0.82	0.94	0.09	0.94	0.97	0.93

Note: IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

Year ahead / Current year	EC	IMF	Mean	Harmonic mean	Geometric mean	Trimmed mean	Median	Weighted forecast regression	Weighted forecast regression (RW)	Discounted MSFE ($\delta = 0.90$)	Discounted MSFE ($\delta = 0.95$)	Discounted MSFE ($\delta = 0.99$)	R_{best}
Current year									regression (RW)				
A	0.74	0.86	0.92	0.94	0.93	0.93	0.81	0.80	0.92	0.89	0.90	0.91	0.71
B	0.72	0.83	0.91	0.95	0.94	0.92	0.81	0.75	0.92	0.85	0.86	0.87	0.56
C	0.94	0.96	0.99	0.99	0.99	0.99	0.97	0.96	0.99	0.99	0.99	0.99	0.94
D	0.99	0.97	1.00	0.99	0.99	1.00	0.98	0.98	0.99	0.99	0.99	0.99	0.99
E	0.75	0.85	0.91	0.93	0.92	0.93	0.81	0.76	0.92	0.86	0.87	0.88	0.61
F	0.74	0.77	0.84	0.87	0.87	0.77	0.76	0.75	0.86	0.82	0.82	0.83	0.70
G	0.91	0.92	0.94	0.86	0.91	0.88	0.95	0.95	0.94	0.96	0.95	0.95	0.92
H	0.09	0.28	0.35	0.51	0.43	0.38	0.20	0.19	0.38	0.32	0.32	0.35	0.16
J	0.84	0.89	0.90	0.83	0.88	0.92	0.91	0.91	0.90	0.92	0.92	0.91	0.88
MEF	0.87	0.94	0.97	0.96	0.97	0.97	0.94	0.92	0.97	0.96	0.96	0.96	0.81
OECD	0.95	0.92	0.86	0.77	0.82	0.83	0.97	0.97	0.85	0.90	0.89	0.87	0.98
EC	-	0.94	0.91	0.84	0.88	0.88	0.96	0.96	0.90	0.92	0.91	0.90	0.96
IMF	0.97	-	0.97	0.93	0.95	0.95	0.98	0.98	0.96	0.98	0.97	0.97	0.93
Mean	0.97	0.93	-	0.99	0.99	0.99	0.96	0.95	1.00	1.00	1.00	1.00	0.88
Harmonic mean	0.98	0.96	0.99	-	0.99	0.99	0.90	0.88	0.99	0.96	0.97	0.97	0.77
Geometric mean	0.98	0.95	1.00	0.99	-	0.99	0.93	0.92	1.00	0.98	0.99	0.99	0.83
Trimmed mean	0.97	0.92	1.00	0.99	0.99	-	0.93	0.92	0.99	0.98	0.98	0.99	0.84
Median	0.97	0.92	1.00	0.98	0.99	0.99	-	1.00	0.95	0.97	0.97	0.96	0.96
Weighted forecast regression	0.90	0.84	0.96	0.91	0.94	0.97	0.97	-	0.94	0.97	0.96	0.95	0.97
Weighted forecast regression (RW)	0.88	0.79	0.95	0.90	0.93	0.96	0.99	0.96	-	0.99	0.99	0.99	0.87

Year ahead / Current year	EC	IMF	Mean	Harmonic mean	Geometric mean	Trimmed mean	Median	Weighted forecast regression	Weighted forecast regression (RW)	Discounted MSFE ($\delta = 0.90$)	Discounted MSFE ($\delta = 0.95$)	Discounted MSFE ($\delta = 0.99$)	R_{best}
Discounted MSFE ($\delta = 0.90$)	0.91	0.82	0.98	0.93	0.97	0.99	0.99	0.99	0.98	-	0.99	0.99	
Discounted MSFE ($\delta = 0.95$)	0.91	0.82	0.98	0.94	0.97	0.99	0.99	0.99	0.99	0.99	-	0.99	0.91
Discounted MSFE ($\delta = 0.99$)	0.92	0.83	0.99	0.94	0.97	0.99	0.99	0.98	0.99	0.99	0.99	-	0.90
R_{best}	0.87	0.73	0.95	0.89	0.93	0.95	0.98	0.97	0.95	0.98	0.98	0.98	-

Note: IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.

APPENDIX B: Computation of weighted forecast combinations

The weighted forecast combination is based on an OLS pooled regression of the realized net lending on the various forecasts over the full sample. The weights are nothing else than the coefficients, and the combined forecast the fit of the model.

We compute two forecast combinations, based on a regression including all nine forecasters, or all nine forecasters and a random walk. Table B1 reports the regression results for the current year and year-ahead forecasts. We use all available forecasts from 1993 to 2022 but drop missing values.

APPENDIX C: Calculation of the forecasted budget balance (as a ratio of GDP)

The CEF provides forecasts for the total deficit only in nominal values (local currency). Hence, we follow Poplawski-Ribeiro and Rülke (2011) to construct a forecast measure of deficit ratio to GDP (percentage of GDP). For that, we cannot simply scale the nominal value deficit forecast by the GDP forecast, because the CEF surveys for growth rates only, and not for the GDP in nominal value.

We construct a measure of the expected nominal year-ahead GDP forecast of forecaster i at month m and year t as follows. In the first step, we take a real-time measure of real GDP in levels for a particular year t. We use the real-time forecast of the same-year real GDP (in levels) coming from the most recent IMF WEO vintage available at any particular month m of year t. The IMF WEOs are published either in April or October; hence, from May to October, we use the April issue, and the October issue in the other months.

The second step is to compute the year-ahead GDP forecast in nominal value. We multiply the real-time (WEO) measure of same-year real GDP (in levels), $E_{WEO,t}[y_t]$, by the year-ahead market (Consensus) forecasts for GDP *growth*, $E_{i,t,m}[\Delta y_{t+1}]$, and inflation, $E_{i,t,m}[\pi_{t+1}]$, for each forecaster i at a particular month m of year t. The expected year-ahead nominal GDP value for each country is then

$$E_{i,t,m}[y_{t+1}] = E_{WEO,t}[y_t] \times \left(1 + E_{i,t,m}[\Delta y_{t+1}] + E_{i,t,m}[\pi_{t+1}]\right). \tag{C.1}$$

The year-ahead expected budget balance for each country is then

$$E_{i,t,m}[b_{t+1}] = \frac{E_{i,t,m}\left[b_{t+1}^{\text{nom}}\right]}{E_{i,t,m}[y_{t+1}]}, \tag{C.2}$$

where $E_{i,t,m}\left[b_{t+1}^{\text{nom}}\right]$ is the (CEF) forecast of the nominal budget balance by forecaster i in month m of year t for 1 year-ahead $t+1$.

TABLE B1 Weights for weighted forecast combinations.

	(a)		(b)	
	Forecast current year (1)	**Forecast year ahead (1)**	**Forecast current year (2)**	**Forecast year ahead (2)**
C	0.59 (1.72)	0.78 (1.64)	0.62 (1.63)	0.86 (1.67)
B	−0.61 (−1.44)	0.42 (0.98)	−0.62 (−1.40)	0.51 (1.07)
E	−0.54 (1.68)	0.15 (0.34)	−0.53 (−1.58)	0.17 (0.39)
MEF	0.81 (1.6)	1.22 (1.00)	0.80 (1.53)	1.15 (0.91)
OECD	0.56 (0.6)	−0.87 (−0.71)	0.51 (0.53)	−0.75 (−0.59)
EC	0.04 (0.05)	0.02 (0.06)	0.10 (0.12)	0.08 (0.25)
IMF	0.64 (1.24)	−0.04 (−0.07)	0.66 (1.21)	−0.14 (−0.25)
RW	-	-	−0.039 (−0.24)	−0.13 (−0.24)
R^2	0.94	0.87	0.94	0.88
F	36.42	15.54	28.9	12.61

Note: The numbers in parenthesis are the t-statistics. IMF, International Monetary Fund; OECD, Organization for Economic Co-operation and Development; MEF, Ministry of Economy and Finance; MSFE, Mean Square Forecast Error.