The European Commission's expenditure benchmark

Charles Wyplosz[1]

The Graduate Institute, Geneva and CEPR

1 INTRODUCTION

In its proposal for the next revision of the Growth and Stability Pact (SGP), the European Commission takes two steps. The first is to shift the focus away from year-by-year evolution of national public finances to a longer-term view that is aligned with the concept of sustainability. Previously, the Pact was implemented on an annual basis, which resulted in often procyclical recommendations that were usually staunchly resisted by national governments for that reason. The second step is to move away from assessing the stance of fiscal policy from the budget balance and to attribute instead a central role to the expenditure benchmark, an in-house measure. Combining these two steps, the Commission proposes to conduct a medium to long-term analysis of debt sustainability from which it derives a multi-year path for acceptable level of public expenditures. Deviations from this path are to be used for surveillance and, if need be, to impose fines.

Until now, the SGP relied on two benchmarks that were easy to understand: the 3% of GDP limit for budget balances and the 60% of GDP limit on public debts. These benchmarks, however, were arbitrary and therefore too controversial to impose on recalcitrant member countries. As a result, even though the Pact may have succeeded in limiting fiscal indiscipline, it failed to establish fiscal discipline. This failure has long been recognised (e.g. Eichengreen and Wyplosz 1998, Canzoneri and Diba 1999, De Haan et al. 2004). It has prompted two reforms that were still be seen as ineffective (e.g. Feldstein 2005, Betsma and Debrun 2005, Wyplosz 2013), a view that was even eventually accepted by the Commission (Pench et al. 2018). In this Policy Insight, I argue that the first step, long-awaited and likely to help deliver debt discipline among euro area member countries, stands to be undermined by the second step – the shift from the budget balance to the expenditure benchmark.

Because the newly proposed steps are highly technical, the details may be overlooked by policymakers and observers and accepted without proper analysis. This is one reason why, in general, simple rules are better than complex ones. Yet, the history of the SGP is that it started with simple rules that could not be made adequately

---

operational, which prompted changes that became increasingly complex to the point of not being operational either (Larch and Turrini 2010, Pench et al. 2018). The new proposal by the Commission is a welcome attempt to find a balance between simplicity and complexity, but the focus on the expenditure benchmark deserves scrutiny even though it may seem unimportant.

In short, the expenditure benchmark (to be defined precisely below) indicates the rate of increase of public spending, adjusted for various factors including new revenue measures, that is compatible with debt sustainability. This concept was first introduced as part of the 2011 reform of the SGP.[2] It was presented as a measure complementary to the cyclically adjusted budget balance (CAB) that had been introduced in the previous reform of 2005. The initial Pact had relied on the headline budget balance. Due to its sensitivity to business cycles, using the headline budget balance could lead to procyclical fiscal policies as the automatic stabilisers would be prevented from operating as needed.[3] And, indeed, it emerged that fiscal policies were often procyclical. In principle, cyclically adjusted budget balances avoid this pitfall. However, the cyclical correction is imprecise and euro area member countries did not waste time in using this weakness to reject unpleasant recommendations by the Commission. In response to the disenchantment with the cyclically adjusted budget balance, the 'Six Pack' reform of 2011 introduced the expenditure benchmark as part of the preventive arm of the SGP:

> "The new expenditure benchmark is therefore not a completely additional requirement, but a means of making the surveillance mechanism of the preventive am of the pact more transparent and, by extension, more effective. By explicitly judging the conduct of fiscal policy on concrete existing figures rather than on estimates of underlying positions, it becomes easier for Member States' plans and outcomes to be judged against the requirements set out by the preventive arm, while the simple requirements of the expenditure benchmark provide more explicit guidance to Member States" (European Commission 2011).

The Commission now proposes to make the expenditure benchmark the only measure to be used: "A single operational indicator – net primary expenditure, i.e. the expenditure which is in a government's control – would serve as a basis for setting the fiscal adjustment path and carrying out annual fiscal surveillance." (European Commission 2022). In other words, the expenditure benchmark is presented as more reliable than the CAB. It is supposed to be simple, "more transparent and, by extension, more effective", based on "concrete existing figures rather than on estimates of underlying positions" and "in a government's control". My aim in this Policy Insight is to show that none of these justifications applies, that the measure lacks any logical basis, and that it stands to lead to undesirable policy actions.

The next section presents the formal definition of the expenditure benchmark and its relationship with the CAB. Section 3 shows that the signals from both measures are statistically similar. Section 4 digs deeper at the adjustments that drive the measures. Section 5 examines how the two measures differ relative to the stated intentions of the Commission. It argues that the expenditure benchmark may have been useful as

---

2   Carnot and de Castro (2015) and Marinhero (2021) provide details on the history of the expenditure benchmark and the related literature.

3   This would not necessarily be the case if the budget balance was sufficiently below the 3% limit in normal years, as was intended initially. As it turned out, several countries considered the 3% limit as the objective to be reached in good years.

part of year-to-year surveillance at the heart of existing SGP but is at odds with the new proposal that focuses on the medium to long run. The last section warns that the expenditure benchmark stands to undermine the proposed reform.

2 THE EXPENDITURE BENCHMARK

2.1 Definition
The derivation of the expenditure benchmark involves four steps, which are detailed in Section A1 of the Technical Appendix.

Step 1: The concept of adjusted expenditure
As explained in European Commission (2013), adjusted expenditure is defined as primary spending (net of debt service) *plus* one-off net tax revenues *less* cyclical unemployment spending *less* various EU-financed expenditures *less* public investment in excess of its previous trend. One-off net tax revenue is the difference between gross one-off tax revenues and one-off expenditures.

Adjusted expenditure is meant to capture public spending under direct government control, hence the exclusion of interest payments and of cyclical changes in unemployment benefits, which are legally mandated entitlements. Above-trend public investments are subtracted, which is surprising since these are government decisions. One explanation is it takes time to actually carry out investment projects, which often mature late in a legislature. This element, therefore, merely smooths out outlays, removing the period of under- and overspending. This can also be seen as a nod to the golden rule that considers that public investments eventually pay for themselves so they should not count as expenditure when they rise above trend.

Step 2: Net adjusted expenditure
Net adjusted expenditure is really an adjusted measure of the change in the primary budget balance. It is the difference between the change in adjusted expenditure and the change in discretionary revenue – a specific definition of the primary budget deficit. Discretionary revenue changes are those that are explicitly decided by the government to pay for any spending increase. The change in discretionary revenue differs from the change in actual revenue, which includes endogenous changes that occur without any government decision because of the evolution of GDP and tax bases. These endogenous changes give rise to revenue windfalls or shortfalls.

Step 3: Expenditure benchmark
At this stage, a rule is introduced. It prescribes that net expenditure cannot grow faster than potential GDP.[4] This limit determines an expenditure benchmark: the highest level of adjusted expenditure compatible with the rule.

Step 4: Fiscal effort
The 'fiscal effort' measures the deviations of net expenditure from the benchmark. A negative fiscal effort signals that net adjusted expenditure is excessive so that the government is required to correct its intended budget, as explained in European Commission (2019).

---

4   For countries with high debts – those which miss their Medium-Term Objective (MTO) as defined by the preventive arm of the existing SGP – the allowed growth of net adjusted expenditure is reduced by a margin designed to bring the debt-to-GDP ratio down at a given speed.

The terms 'adjusted expenditure' and 'fiscal effort' hide the fact that net adjusted expenditure is really the primary budget balance (more precisely the deficit), partly cyclically adjusted in a different way than the traditional CAB, partly driven by the rule that sets the benchmark. This vocabulary amounts to a potentially confusing misnomer.

2.2 Evaluation of justifications

Beyond vocabulary, the concepts of net expenditure growth and fiscal effort are convoluted and much less intuitive than the CAB. Why, then, this new concept? A number of justifications have been advanced.

First, and maybe foremost, comes dissatisfaction with the CAB, which relies on estimates of the potential output gap – the difference between actual and potential GDP – to capture cyclical fluctuations. A minor issue is that it takes time to measure reasonably precisely such an all-encompassing concept as GDP. It usually takes more than one year to move from flash estimates of GDP to a stabilised measure, and revisions occur over time. More importantly, potential GDP is not observable. To evaluate potential GDP, the Commission uses two methods – a production function and a statistical procedure to determine its trend – each with well-known strengths and weaknesses,[5] which can be challenged endlessly. Then, moving from the output gap to the budget requires estimating the impact of cyclical fluctuations on the budget balance. This is done by using estimates of the elasticities meant to capture the effect of cyclical fluctuations on public spending and revenue. These elasticities are variable, changing with the detailed sources of cyclical fluctuations, which results in budgetary windfalls or shortfalls. Carnot and de Castro (2015) find that the budgetary windfalls or shortfalls are the main source of errors in estimates of the CAB and are biased towards more favourable estimates. This is an important criticism of the CAB measure, but it must be compared with how the expenditure benchmark treats cyclical fluctuations, which is considered next.

Second, like the CAB, the expenditure benchmark includes a cyclical adjustment but in a limited way. The adjusted expenditure takes into account cyclical changes in unemployment benefit outlays. The implicit argument must be that unemployment benefits, which may be the largest cyclical component of public spending, can be estimated more precisely than overall public spending fluctuations. This remains to be seen but, even so, it ignores that government revenues are generally more cyclical than spending. The CAB explicitly attempts to capture this effect while the expenditure benchmark ignores it altogether. Anyway, estimates of the unemployment gap are subject to the same criticism as the potential gap since, like potential GDP, the natural unemployment rate is not observable either. The exclusive focus on cyclical unemployment spending is arguably a weakness of the expenditure benchmark.

Third, in spite of the imprecision that mars estimates of potential GDP, the expenditure benchmark does not escape the issue. The benchmark explicitly involves the potential growth rate to determine the benchmark and therefore the assessment of how much the government can spend. Instead of the usual estimates of potential GDP, the procedure uses the ten-year average real GDP growth rate, calculated each year from year t-5 to year t+4. This requires forecasting growth over the next four years, which is likely to be at least as imprecise as the standard estimates of potential GDP in any

5   Orphanides (2001) famously documented how monetary policy in the US went astray because estimates of potential GDP had been off the mark.

given year. Anyway, if it turns out that this measure is better, it is possible to use it to make the cyclical adjustments to compute the CAB. This is just a measurement issue, which does not justify developing an alternative concept.[6]

Fourth, the benchmark allows for public expenditure to increase in proportion to potential growth from the level observed in the previous year. As noted in Marinhero (2021) and Larch et al. (2022), this can be misleading if the previous year's expenditure happens to be too high or too low for debt sustainability. This can happen if public spending was increased during a boom year, or if it was slashed because of pressure on the public debt. This base effect perpetuates a one-year outcome unrelated to the notion of debt sustainability. The CAB explicitly corrects for this situation, while the expenditure benchmark ignores it.

Fifth, the expenditure benchmark involves changes in discretionary revenues, which are provided by member countries as part of the surveillance process. This is a key element of the procedure since it directly determines the expenditure benchmark. However, the measurement of discretionary revenue changes – equivalently of windfalls or shortfalls – is bound to be imprecise and therefore open to challenge. Were the Commission to attempt to do so, it would have to second guess the estimates provided by governments, which is delicate and likely to result into an equally uncertain measurement. This, in turn, may provide an incentive for governments to resort to tax expenditures, leaving the burden of the proof with the Commission. Tax expenditures complicate tax systems and may undermine their main allocative or redistributive objectives.

Sixth, the expenditure benchmark explicitly tracks net revenue one-offs. This is a valid concern, borne out of a long experience in surveillance. Temporary spending cuts or tax increases improve the budget balance one year but do not affect debt sustainability over the longer run. The argument can be turned around, though. Temporary policy actions have a negligible effect on debt sustainability so they should not be of any concern.

Finally, as the one-off discussion illustrates, the expenditure benchmark was designed to manage previous versions of the SGP. It deals mostly with procedural weaknesses encountered when trying to enforce the preventive arm on a year-by-year basis. Errors in cyclical adjustments, imprecise budgetary figures backed by limited information, and one-offs are a key concern for annual surveillance, but they mostly wash out in the longer-term horizon that the Commission now proposes to structure the Stability and Growth Pact. This new framework largely invalidates the justifications in favour of abandoning the CAB for the expenditure benchmark.

3 THE DIFFERENCE BETWEEN THE EXPENDITURE BENCHMARK AND THE CYCLICALLY ADJUSTED BUDGET BALANCE

3.1 Formal differences
One way to evaluate the advantage of using the expenditure benchmark over the CAB is to look at the formal and empirical differences between these two measures. As already noted, the net adjusted expenditure growth rate is closely related to the change in the cyclically adjusted primary budget deficit which, following Blanchard

---

6   A related justification is that the estimation errors are smaller for potential growth than for the potential level. It is true that the cyclically adjusted budget balance is based on the level as it involves the output gap. However, it is the change in the change in the cyclically adjusted budget balance that is used to assess the fiscal stance.

(1980), is the most frequently used measure of discretionary fiscal policy. Note that the CAB under consideration here concerns the primary budget, after deduction of debt service, as is common practice.[7]

The fiscal effort is directly comparable to the change in the CAB, since it is the difference between the deviation of the change in net adjusted expenditure (which really is an adjusted measure of the primary budget balance, as previously noted) from its benchmark. Appendix A2 derives the formal difference between the change in the CAB (ΔCAB) and the fiscal effort (FE). This difference can be summarised as follows:

ΔCAB − FE = cyclical adjustment for CAB
– cyclical adjustment for fiscal effort (unemployment benefits)
– temporary effects (one-offs, windfalls/shortfalls, investment, EU receipts)
– benchmark

Importantly, the (change in the) actual budget balance does not appear in the difference. It cancels out since it is included in both measures – the change in CAB and the fiscal effort – which are both just two different ways of adjusting the primary budget balance. Clearly, the expenditure benchmark adjustments are more expansive than the mechanical CAB adjustment.

One reason for the development of the expenditure benchmark has been to stamp out one-offs that may be seen as 'creative accounting', and to capture windfalls and shortfalls that arise when cyclical fluctuations affect the budget over and above what could be expected given experience. These 'surprises' may matter a lot for annual surveillance, but they do not affect debt sustainability since, by definition, they average out over the longer run. CAB estimates typically ignore the windfalls or shortfalls.

Several papers (e.g. Darvas 2013, Fatas 2019, Larch and Turrini 2009, 2010), explain the shortcomings of the CAB. Others (e.g. Irish Advisory Council 2015, Marinhero 2021) do the same for the expenditure benchmark. Because the 'true' CAB is not observable, no measure is ever going to be precisely correct. Much of the same applies to the fiscal effort, which involves unobservables such as the natural rate of employment and, arguably, the one-offs.

An important question is which measure is more reliable. Benalal et al. (2021) and Carnot and de Castro (2015) compare the two methods. The first paper tracks the empirical impact of the underlying assumptions and concludes in favour of the expenditure benchmark, on the basis of criteria like predictability or sources of procyclical recommendations. The second paper, which is mostly a presentation of the expenditure benchmark, evaluates the importance of the various adjustment terms, with emphasis on the role of windfalls and shortfalls.

The difference also reflects the fact that the adjustments incorporated in the computation of the fiscal effort include policy choices. The distinction between public spending and public investment has long been used to portray public investment as 'good' because it enhances future growth (Blanchard and Giavazzi 2004). However, the distinction is fuzzy. Public infrastructures, education and research are arguably productive, along with justice and health policies, but the definition could include public services that also contribute to economic growth. Should defence and climate

---

7   The European Commission uses a difference measure of the CAB, called the structural balance, which nets out the one-offs (and gets closer to the fiscal effort measure). Here I retain the usual definition of the CAB.

change policies not be considered as investments? Dubbing public investment as 'good' not only stands to generate endless debates, it is also likely to result in creative accounting.

Furthermore, as noted above, the target for adjusted expenditure growth, the benchmark, is completely arbitrary as it prescribes keeping it constant as a share of GDP. Some countries may have good reasons to expand this ratio, others to reduce it. At any rate, the size of the public sector is a deeply political choice to be made democratically at the national level, not to be frozen by a European-level rule. A merit of the CAB is that it is devoid from such policy considerations.

3.2 Empirical evaluation

This section revisits Carnot and de Castro (2015) with updated data that extend the sample period from 2004–2013 to 2004–2022. Rather than focusing on the roles played by specific adjustment methods and their components, it looks at the evolution of each of the two measures – the change in the CAB and the fiscal effort. Even though the adjustment methods have different impacts across time and countries, this section shows that the CBA and the benchmark expenditure measures do not differ in any statistically significant way.

For 2004–2020, the fiscal data come from Carnot and de Castro (2015), updated and complemented. The update does not just add more years, it also includes a revision of past estimates with new and hopefully more accurate measures. The data therefore do not reflect the measures of CAB and of the fiscal effort that have been effectively used in real time. This does not affect the purpose of the exercise, which is meant to compare the properties of the two concepts, although data revisions matter for policy purposes since the availability of data may affect their real-time usefulness. This issue is discussed in several of the papers mentioned above and should be kept in mind. For 2021 and 2022, the data have been provided by the Commission, which kindly provided its 'transparency files' (which are not publicly available, another misnomer). The CAB data come from the AMECO database, which is also used by the Commission to feed into its computations of the fiscal effort. The CAB measure used here is based on estimates of potential GDP based on the statistical properties of its trend, which makes it more methodologically comparable to the expenditure benchmark than the other model-based measure. Both measures are expressed as percent of GDP.

Figure 1 shows the evolution of the fiscal effort and of the CAB for all euro area countries. For most countries, the two measures move very closely to each other. Where differences appear, they generally concern crisis years – the euro area crisis over 2010–12 and the Covid pandemic over 2020–2021. Are these differences meaningful? The figure also displays the two-standard deviations band around the CAB curve. Table 1 reports, country by country, the percentage of observations of the fiscal effort that lie outside the band. For most countries (nine out of 19), this never happens. It occurs five times in six countries, and three times in three countries. The case of Cyprus is special, with both measures being highly volatile. Once again, as far as annual surveillance is concerned, getting the numbers right is especially important in crisis years when procyclical reactions should be avoided, but the new proposed emphasis on longer horizons makes the issue moot. The suspension of the SGP during and after the Covid pandemic is a good illustration. It was perfectly right to face this exceptional shock with larger deficits and thus ignore any notion of fiscal effort but, afterward, it is essential to implement corrections to restore debt sustainability over the long run. The focus should not be on yearly decisions but on the timing and duration of the corrections when deviations emerge, for which the CAB is perfectly adapted.

CEPR

Figure 1   Comparison of the CAB and fiscal effort, 2004-2023 (% of GDP)
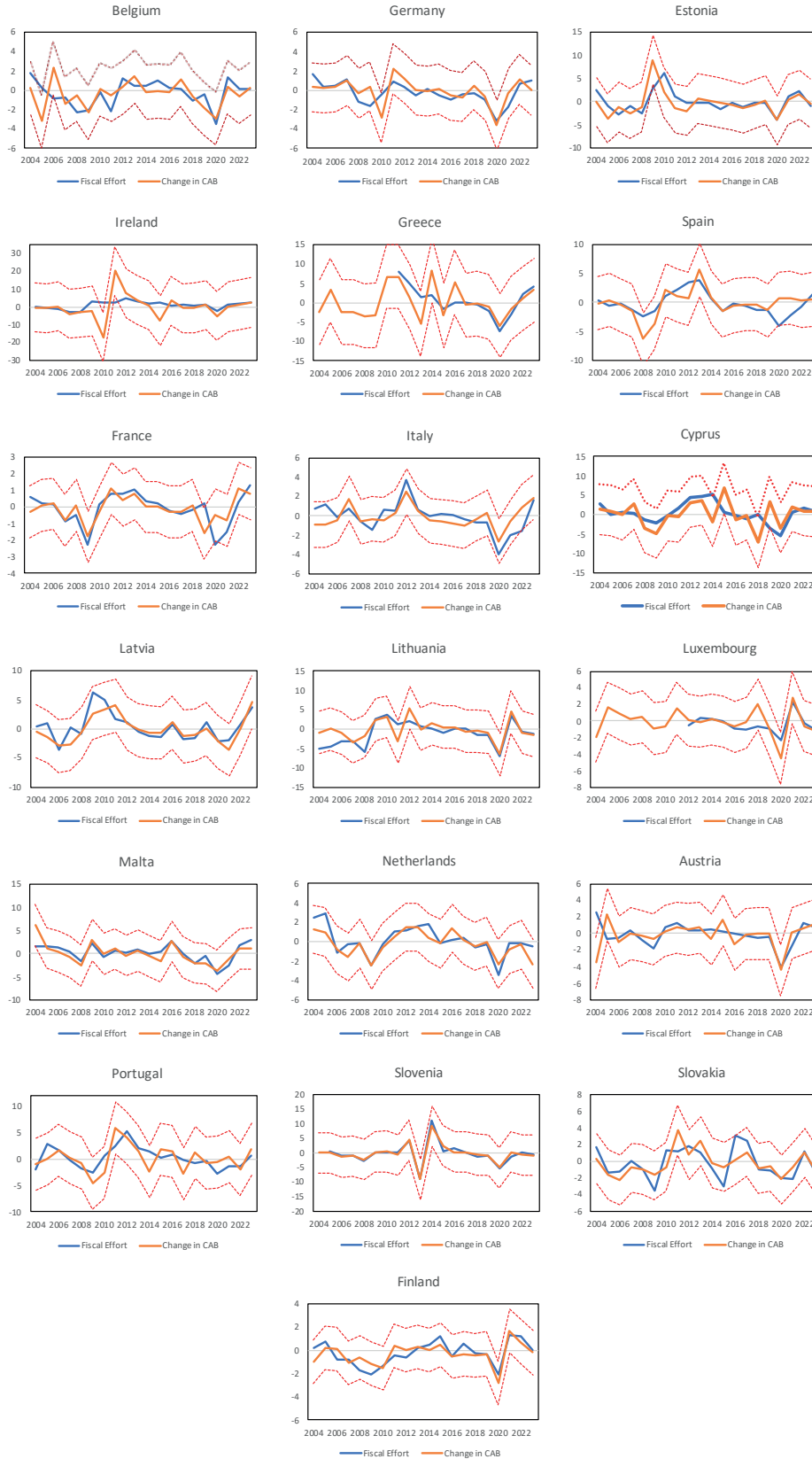
Sources: European Commission and AMECO

Table 1   Percent of years when the expenditure benchmark measures lie outside the two standard-deviations band of the CAB

Country	Value	Country	Value
Belgium	10	Lithiania	0
Germany	0	Luxembourg	0
Estonia	5	Malta	5
Ireland	10	Netherlands	0
Greece	0	Austria	5
Spain	5	Portugal	0
France	10	Slovenia	0
Italy	5	Slovakia	5
Cyprus	15	Finland	0
Latvia	0		

Sources: See Figure 1.

These results confirm previous evidence that differences between the two measures are not statistically significant. [8] When significant differences emerge, it is impossible to determine which one is better. In fact, there is no better measure since each one reflects a specific methodology to determine what a government is doing, which is unobservable. Nor is it possible to decide which measure leads to better policy recommendations because we cannot observe the counterfactual.[9]

4 IMPLICATIONS OF THE SHIFT TO MULTIYEAR EVALUATION OF SUSTAINABILITY

In its latest proposal of SGP reform, the Commission envisages to shift from year-by-year assessments to a medium-term outlook that focuses on the debt path. This change is highly welcome as it gets closer to the concept of debt sustainability, which is ignored by the 3% and 60% annual deficit and debt ceilings that drove previous versions of the Pact, even if their importance had been declining. A long-term horizon allows for short-term flexibility in using fiscal policy to deal with cyclical fluctuations and other shocks while retaining the longer-term debt sustainability constraint.

To recall, public debt is deemed sustainable when it is matched by current and future primary budget balance surpluses. Since the government (or the state) is expected to exist forever, the relevant future is infinite. In practical terms, this formal definition must be approximated if only because it is impossible to look at an infinite horizon. The Commission proposes to look at a horizon of ten years and to focus on the path of primary surpluses over the next four years that put the debt on a safe path, defined as moving toward 60% in ten years. This is a form of debt sustainability analysis (DSA). While DSA is now routinely performed, it is highly arbitrary, as explained in Wyplosz (2011). It requires making assumptions about the evolution of the interest and growth

8   "Overall, the size of forecast errors appears broadly similar regardless of whether the fiscal effort is based on the expenditure benchmark or the structural balance" (European Commission 2020: 11).
9   A counterfactual analysis is presented in Andrle et al. (2015). It relies on an IMF model and combines the indicator (CAB and benchmark expenditure) with policy rules, so it is not clear what is being compared. At any rate, the CAB and benchmark rules deliver virtually identical results regarding the variability of output and a significantly larger variability regarding the evolution of the debt to GDP ratio.

rates until the horizon is reached. In addition, there is an infinity of budget balance paths that deliver a sustainable debt.[10] Keeping these limitations in mind, it remains the only logical procedure.

Once we focus on debt sustainability adequately defined and look reasonably far into the future, temporary annual fluctuations become irrelevant. What matters is the average evolution of the actual primary balance over time.[11] The cyclically adjusted balance fits well this definition. It is the primary budget balance that enters the definition of sustainability, not the fiscal effort. On average over time, the CAB, if correctly estimated, should be equal to the actual primary budget balance. The CAB is therefore the logical measure to identify a sustainable path to be used as a benchmark to evaluate the actual balances. Since it is another form of adjustment of the actual balance, the expenditure benchmark could also be used, but it goes beyond cyclical considerations and involves some assumptions, which are difficult to formulate and justify.

The DSA procedure is subject to significant uncertainty and therefore highly imprecise.[12] Using imprecise estimates of the cyclically adjusted budget balance, either the CAB or the expenditure benchmark, does not help of course. A natural question is how each measure's uncertainty is combined with the uncertainty of DSA. This has led Benalal et al. (2022) and Carnot and de Castro (2015) to study the statistical properties of each measure. A different piece of evidence is provided in Figure 2, which displays the adjustments used for each measure, and are summarised above. The standard cyclical adjustment of the CAB is transparently provided in the AMECO database. The database updates its evaluation as more precise data become available, so this is not real-time information. There is not corresponding information for the fiscal effort adjustment, but it can be inferred from available data.[13]

Figure 2 displays the adjustments over 2004–2022 for each country. These measures are all expressed as changes in the ratios to GDP. In most cases, both adjustment methods yield similar results: for 79% of all observations the signs are the same, the proportion being highest in the case of Finland (95%) and lowest for Lithuania (58%).

An important difference concerns the variability of these adjustments. The right-hand chart in Figure 3 displays the country averages of the annual standard deviations presented in Figure 2. It shows that the adjustments to the expenditure benchmark are significantly more volatile than the adjustments to the primary budget balance. This confirms the role, noted by Carnot and de Castro (2015) and others, of the size of the windfalls and shortfalls, which are ignored in the CAB adjustment.

10   This is an important and hardly ever discussed aspect of DSA. In Wyplosz (2021), I argue that the choice of a particular path should be the responsibility of governments, not of technical staffs.
11   Strictly, this statement must concern the present value of present and future budget balances.
12   This has led to adopting another procedure, stochastic debt sustainability analysis where past observations are used to assess the probabilities of the assumptions. But the assumption that future uncertainty will be similar to past uncertainty is dubious at best.
13   The adjustments A^{CAB} and A^{FE} used to derive the change in CAB and to the fiscal effort are defined as follows: $\Delta CAB = \Delta \text{Balance} - A^{CAB}$ and $FE = \Delta \text{Balance} - A^{FE}$. It follows that $A^{FE} = \Delta CAB - FE + A^{CAB}$, and all the terms on the right are available.

Figure 2   Cyclical adjustments, 2004-2022 (% of GDP)
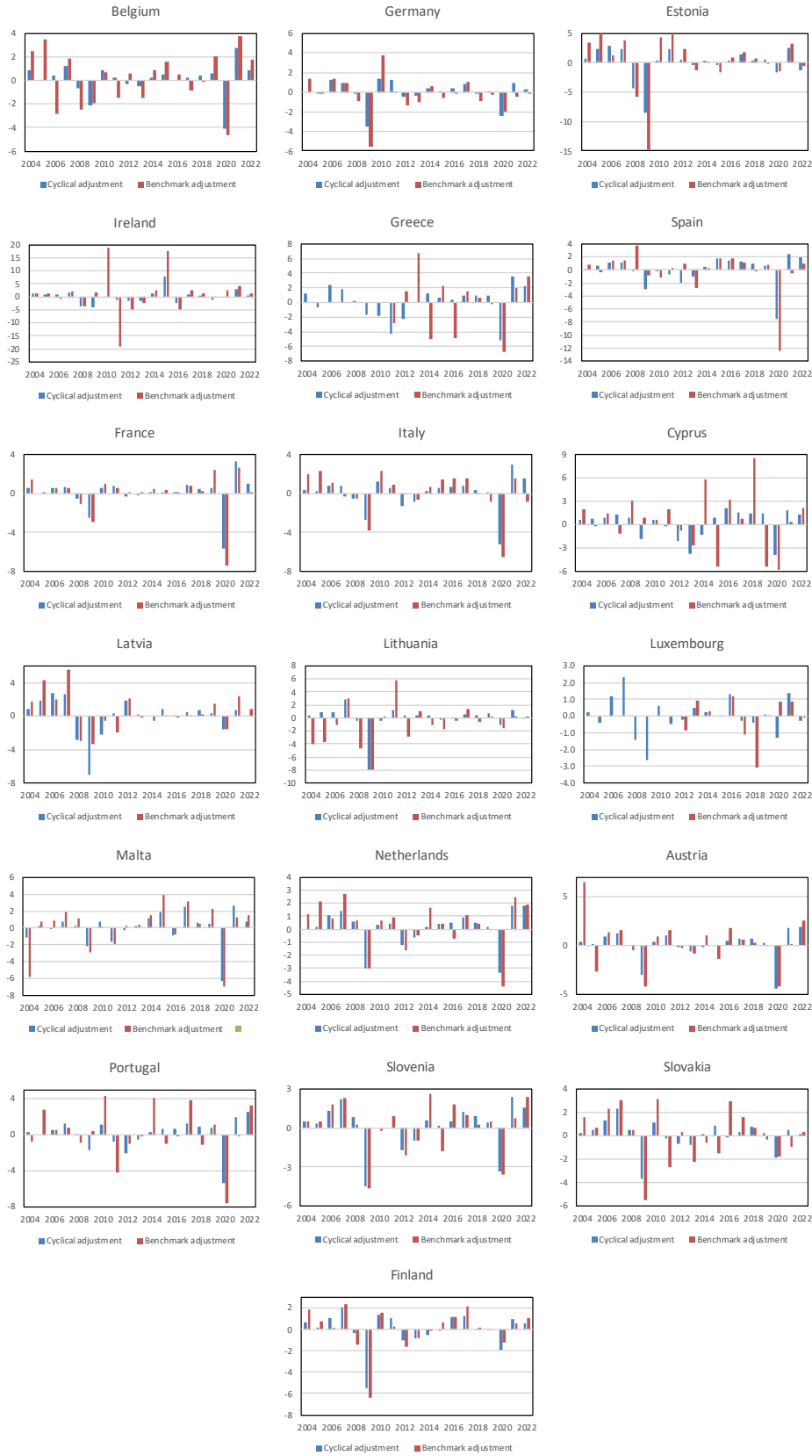
Sources: See Figure 1.

Figure 3   Cyclical adjustments: average and standard deviations, 2004–
2022 (% of GDP)
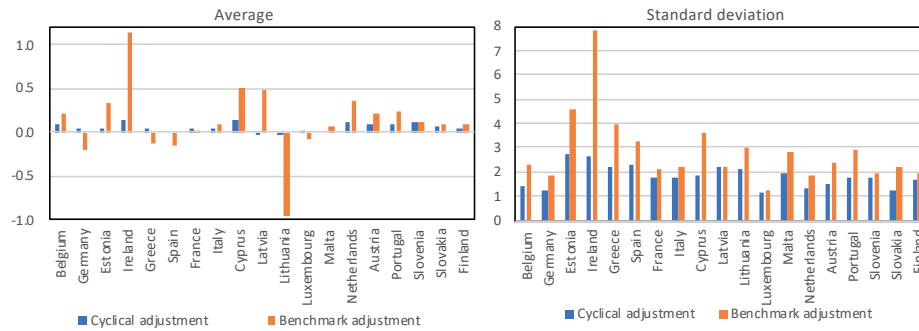
Sources: See Figure 1.

It is impossible to determine whether this variability is justified and, more generally, to assess the quality of the estimates. The CAB adjustments correspond to cyclical variations, which should average zero over the long run. Most terms of the expenditure benchmark adjustment should also average zero, except for the EU-financed expenditure and the expenditure increase benchmark. Ignoring this remark for the time being, the left-hand chart in Figure 3 displays these two adjustment averages over 2004–2022 for all member countries. It shows that the average adjustments to the primary budget balance are indeed close to zero, which is not the case of the average adjustments to the expenditure benchmark.

Taking now into account the non-cyclical parts of the benchmark adjustment, a back-of-the-envelope estimate of the average difference between the benchmark and EU funded spending is -0.1% of GDP.[14] Subtracting this very rough estimate does not alter the conclusion.

It could be that the large averages shown in Figure 3 are dominated by one or two exceptional years, but Figure 2 does not support this interpretation. Alternatively, some member countries may regularly provide information that is not trustworthy, in which case the proper response would be to conduct a proper audit rather than relying on corrections by the Commission. Whatever the explanation, this observation does not provide support for the assertion that the expenditure benchmark is more reliable than the CAB over the medium or long term. If anything, it suggests that the CAB is better adapted than the expenditure benchmark to the medium-term approach advocated by the Commission.

The upshot is that, qualitatively, the adjustments mostly go in the same direction. Quantitatively, however, the fact that the average of the expenditure benchmark adjustment is often not close to zero is disquieting. At the very least, it contradicts claims that the expenditure benchmark is more precise than the CAB. In addition, since Figure 1 shows that the difference between the two measures of the fiscal stance is practically irrelevant, the choice between them must factor in other criteria.

5 BROADER OBJECTIVES

So far, the discussion has focused on the formal characteristics of each measure of the fiscal stance. What other criteria can be used to compare the two methods? The Commission has set the following objectives for the SGP reform:

14 The average of EU funding for all countries and years is about 1% of GDP, with significant difference across countries. The expenditure increase allowance is the product of potential growth, say 2%, and of the expenditure to GDP ratio is on average 45% of GDP, so the overall average is 0.9% of GDP.

"The orientations seek to ensure that the framework is simpler, more transparent and effective, with greater national ownership and better enforcement."

"Building an economic governance framework fit for the challenges ahead." (press release, 9 November 2022).

The aim is further explained as follows:

"An ambitious simplification of the Stability and Growth Pact is proposed. To improve effectiveness and transparency, a single operational indicator anchored on debt sustainability would serve to set the fiscal trajectory. We propose to use net primary expenditure as the single indicator, i.e. expenditure net of discretionary revenue measures and excluding interest expenditure as well as cyclical unemployment expenditure, which would ensure a higher degree of macroeconomic stabilisation. This simplification would increase transparency, including among policymakers, and would allow to do away with a number of fiscal surveillance tools that are overly complex, rely on unobservable indicators, or have not stood the test of time. While annual fiscal surveillance at the EU level would be conducted solely on the basis of the expenditure path, Member States could translate net primary expenditure into alternative indicators for national budgetary purposes (e.g. a structural balance)." (Buti et al. 2022)

Thus efficiency, transparency and simplicity are identified as key goals of the reform proposal and offered as reasons to replace the CAB with the expenditure benchmark. These criteria are also put forward by several contributions, including Carnot and de Castro (2015), European Commission (2020), Eyraud et al. (2018), Gaspar and Amaglobeli (2019) and Thygesen et al. (2020).

5.1 Efficiency

Efficiency means that the preferred measure is more accurate and therefore less open to debate than the CAB. The results presented above do not provide any evidence to that effect, on the contrary. It is not even clear how this criterion can be evaluated. The Commission has a long experience of using the CAB as its key gauge, no doubt a source of frustration (Larch and Turrini 2009), but not with the expenditure benchmark, which was only meant to serve as a check on the CAB. In addition, the CAB estimates potential GDP with two methods (a model and past trend) while the expenditure benchmark uses a different method (centred average). If the latter proves to be superior for efficiency, it can be used to estimate the CAB.

5.2 Transparency

The transparency criterion clearly favours the CAB. In order to avoid the cyclical adjustment method used to compute the CAB, which relies on estimates of potential GDP, the expenditure benchmark combines several approaches. The cyclical adjustment of unemployment benefit spending requires estimating the natural rate of unemployment using the NAWRU (non-accelerating wage rate of unemployment), which suffers from similar difficulties as estimating potential GDP. The expenditure benchmark allows public spending to increase at the same rate as potential GDP, which is estimated as the GDP growth rate over a ten-year horizon that starts five years before and ends four years after the current year. In addition, the NAWRU is calculated using the same potential estimates as those used for the CAB while simultaneously using a different measure to set the benchmark. This inconsistency is detrimental to the transparency objective.

Moreover, the Commission's own estimates of potential GDP can be evaluated in comparison with other estimates routinely produced by several national and international institutions, which enhances transparency. The expenditure

benchmark's estimates of average GDP require forecasting GDP growth four years ahead, which is a highly imprecise and untransparent procedure, with no available comparative estimates.

The expenditure benchmark also relies on one-off net revenues, which are difficult to detect *ex ante*. Similarly, one component of the adjustment variable is EU-financed expenditure. In principle, these two elements are computed using the national reporting provided annually by member states. The Commission checks these reports and interprets them. So far at least, the procedure and its outcome have been quite opaque (Marinhero 2021)[15] and occasionally misleading, as explained in Irish Fiscal Advisory Council (2015).

5.3 Simplicity

The Commission has long agreed that, from one reform to the next one, the SGP has become exceedingly complex. According to Pench et al. (2018), this is partly the result of the "sedimentation of reforms" that have added over the years new procedures without always eliminating older ones that were becoming redundant. They also argue that complexity is unavoidable when allowing for flexibility in the face of unforeseen events. This may be true, and the new proposal goes some way toward cleaning up the sediments and adopting the long-run view that is inherently well-designed to deal with unexpected events. But these improvements are likely to be offset by the shift to the expenditure benchmark.

It is enough to look at the (simplified) definition of the benchmark in Section 2.1 to see that it is anything but simple. It relies on nonstandard concepts (fiscal effort, adjusted net expenditure) that combine cyclical adjustments, specific accounting (the one offs) and normative rules (the golden rule, the allowed net expenditure growth). If sedimentation is a source of complexity, it should be avoided in the new proposal. Some elements, like the 3% and 60% rules or the combination of a preventive and a corrective arm, may be difficult to formally remove because they are enshrined in formal agreements, including the Treaty on the Functioning of the European Union (TFEU). Yet, they can be quietly deemphasised, as has been the case of the 60% rule. It is surprising, therefore, that the Commission wishes to contribute to more sedimentation with the expenditure benchmark, which has just been an internal measuring instrument, ill-adapted to the new multi-year proposal.

This sediment is already a source of complexity (and opacity) at the proposal stage of the envisaged reform. Indeed, in its proposal, the Commission resorts to a truncated explanation of what exactly is the expenditure benchmark. In contrast, the concept of a CAB is simple to grasp, if only because it is well established. Like potential GDP, measures of CABs have been computed by many institutions for many decades. Furthermore, the existence of readily available and widely disseminated measures of CAB is a source of both transparency and efficiency. It is easy to compare the Commission's estimates with the alternatives, which is good for transparency and for improvement.

---

15 To quote Marinhero (2021: 423-4): "In practice, Euro-area Member States are compelled to include in the Draft Budget Plan (…) a table stating the total discretionary measures and the amounts to be excluded from the expenditure benchmark. (…) The Commission makes a judgment about it, which is not public. (…) There is a complete lack of transparency regarding how the final figures are obtained since there isn't any known published source for the data. (…) The EB excludes the expenditure financed by the EU funds since such expenditure is neutral for the budget balance. It is, however, hard to obtain accurate information in real-time on such flows of financing. Given the operational complexity of structural funds, it is difficult to forecast the transfer of funds from the EU budget. Regarding public investment, the expenditure benchmark considers a moving average of four years for the nationally financed investment instead of the actual investment in year (…) When a large variation in investment occurs, this treatment is a source of inconsistency with the other rules applying to the budget balance."

6 CONCLUSIONS

This Policy Insight shows that the two measures – the CAB and the expenditure benchmark – provide similar results leading to similar policy conclusions. The question is: why should anyone object to the Commission's proposal to ditch the CAB and adopt the expenditure benchmark? I have argued that there are many reasons to prefer the CAB. Rather than summarising these arguments, this concluding section focuses on the political economy aspects of this choice.

First, both measures of the fiscal stance are imperfect, and the quest for perfection is hopeless. As such, the longer they are used, the more the limitations are bound to emerge. The Commission developed the expenditure benchmark after it became disenchanted with the CAB that it once championed. The same disenchantment will affect the expenditure benchmark once it becomes the only gauge of the fiscal policy stance. However, I have argued that the imperfections of the expenditure benchmark are deeper than those of the CAB in many dimensions: logical characteristics, actual performance, simplicity, efficiency, and transparency. What may seem like a good idea today will become an even more serious weakness for discussing fiscal discipline with member countries.

Second, the Commission fails to recognise that the expenditure benchmark has been designed to carry out the failed year-by-year approach that it now proposes to replace with a medium to long-term framework. Going from debt sustainability analysis to feasible paths of primary budget balances is straightforward, both logically and technically. The link with the expenditure benchmark is marred by the simple fact that it is... a benchmark that arbitrarily sets the path of public expenditure growth over time. It is always a bad idea to mix up technical analysis (debt sustainability and cyclical adjustment) with normative targets. For most citizens, and for many governments, this will only reinforce the impression that fiscal discipline is a highly technical black box.

Third, the expenditure benchmark establishes the notion that the size of government, as a share of GDP, must be maintained at last year's level. This can be changed if revenue is modified accordingly, which is what the CAB implies as well. The risk is that, like the 3% and 60% benchmarks, it will focus public attention on a misleading single norm and that it will become obsolete as time passes. Some countries may need to increase the ratio of expenditure to GDP because it is too low, other countries should reduce an already high ratio and, anyway, the government size is a political decision that individual countries have to make while keeping the debt sustainable.

The key reason behind the design of the complex expenditure benchmark is well described by Pench et al. (2018): "The costs of enforcing the rules argue for taking into account economic circumstances so as to avoid undesirable economic outcomes, a source of 'design complexity'. (…) In order to capture all kinds of possible situations while maintaining a sufficient degree of predictability, the tendency has been to put in place incremental, detailed ex ante specifications in an elusive quest for a 'complete contract'. (…) This approach disregarded the fact that no system could eliminate or template the necessary role of economic and political judgement in unusual circumstances or in borderline cases. Moreover, the lack of simplicity generated by this over-specification turns against its proclaimed objective of predictability."

This lucid observation seems to have been lost. There is a serious risk that the new, much improved version of the SGP will be undermined by a construction that is both deeply technical and confusing. In addition to their logical flaws, previous versions of

the SGP have failed for the same reason. Hopefully, the complexity of the expenditure benchmark will not discourage serious analyses of what may look like a minor issue better left to the technicians.

TECHNICAL APPENDIX

A1. The expenditure benchmark
The derivation of the expenditure benchmark can be described in four steps.

First comes the concept of adjusted expenditure $\tilde{G}$, which can be written as:

$$\tilde{G} = G + NT^o - (u - \bar{u}) - X, \tag{1}$$

where G represents public expenditures net of debt service, NT^o one-off net revenues, $u - \bar{u}$ is the unemployment gap and X stands for the other expenditure items mentioned above (deviation from average of public investment spending and EU-financed expenditure).

Second, net adjusted expenditure consists in subtracting from adjusted expenditure $\tilde{G}$ the discretionary revenue changes $\Delta\tilde{T}$ that are explicitly decided by the government to pay for any spending increase. The change in discretionary revenue $\Delta\tilde{T}$ differs from the change in revenue ΔT because the latter can change endogenously – i.e. without any government decision – as tax bases or the effect of tax rates on revenue. This is a windfall revenue – a shortfall when negative – which is denoted as Ω:

$$\Delta T = \Delta\tilde{T} + \Omega \tag{2}$$

The growth rate of net adjusted expenditure in year t is computed as:

$$\Delta N\tilde{G}_t = \tilde{G}_t - G_{t-1} - \Delta\tilde{T}. \tag{3}$$

Third, a benchmark is introduced. It formalises a rule according to which net expenditure growth should not grow faster than the rate of growth of potential GDP. The expenditure benchmark is the adjusted expenditure G^* that just satisfies the rule:

$$\Delta NG_t^* \stackrel{\text{def}}{=} G_t^* - G_{t-1} - \Delta\tilde{T} = \bar{y}G_{t-1}, \tag{4}$$

where $\bar{y}$ is the potential GDP growth rate. This rule is not expected to always be followed; this is why it is a benchmark.

Finally, the 'fiscal effort' looks at deviations from (4). It is defined as the difference between the benchmark increase $\Delta NG_t^* = \bar{y}G_{t-1}$ and the planned change of net adjusted expenditure $\Delta N\tilde{G}$:

$$FE_t = \Delta NG_t^* - \Delta N\tilde{G}_t = \bar{y}G_{t-1} - \Delta N\tilde{G}_t. \tag{5}$$

A negative fiscal effort signals that net adjusted expenditure growth exceeds the benchmark, so that the government is required to correct its intended budget as explained in European Commission (2019). Using (1) and (2), and dropping the time subscript for the current year t, the fiscal effort can be rewritten as:

$$FE = \Delta T - \Delta G - \Omega - NT^o + \theta(u - \bar{u}) + X + \bar{y}G_{t-1} \tag{6}$$

This expression reveals that the fiscal effort is the change in the primary budget balance, $\Delta(T - G)$, corrected for the various variables used in the adjusted expenditure measure (1) and for revenue windfalls/shortfalls Ω, as well as the 'allowed' spending growth $\bar{y}G_{t-1}$.

A2 Difference between the CAB and the fiscal effort

As noted above, in addition to imprecise estimations of the output gap, the CAB must estimate the impact of cyclical fluctuations on the budget. This is usually done by using estimates of the relevant elasticities.[16] The CAB can be written as: [17]

$$CAB = (T - \varepsilon^T \psi T) - (G - \varepsilon^G \psi G)$$

where ε^T and ε^G are the elasticities of revenues T and primary expenditure G relative to changes in the output gap ψ.

The CAB estimates typically ignore the windfalls or shortfalls, assuming that the elasticities ε^T and ε^G are constant. Under this assumption, the change in the CAB is:

$$\Delta CAB = (1 - \varepsilon^T \psi)\Delta T - (1 - \varepsilon^G \psi)\Delta G - (\varepsilon^T T - \varepsilon^G G)\Delta \psi \tag{7}$$

Combining (6) and (7), the difference between the two measures of the fiscal policy stance is:

$$\Delta CAB - FE = \underbrace{\psi(\varepsilon^G \Delta G - \varepsilon^T \Delta T) + (\varepsilon^T T - \varepsilon^G G)\Delta \psi}_{\text{Cyclical adjustment}} + \underbrace{NT^o + \Omega - \theta(u - u^-) + X - y^- G_{t-1}}_{\text{Benchmark adjustment}}, \tag{8}$$

The first terms in (8) capture the cyclical adjustment used to derive the CAB. The next four terms capture the adjustments to the budget balance that lead to the adjusted net expenditure $\tilde{G}$: the one-offs, the windfalls or shortfalls, the cyclical adjustment of unemployment benefits and the terms captured by X, EU-financed expenditure and public investment deviations from trend. The last term corresponds to the normative prescription for adjusted expenditure to grow at the same rate as potential GDP.

REFERENCES

Andrle, M, J Bluedorn, L Eyraud, T Kinda, P Koeva Brooks, G Schwartz and A Weber (2015), "Reforming Fiscal Governance in the European Union", IMF Discussion Note SDN 15/09.

Benalal, N, M Freier, W Melyn, S Van Parys and L Reiss (2022) "Towards a Single Operational Indicator in The SGP – A Look at the Properties of the Current Expenditure Benchmark", SUERF Policy Note No. 267

Betsma, R and X Debrun (2005) "The New Stability and Growth Pact: A First Assessment", *European Economic Review* 51(3): 453-477.

Blanchard, O J (1990) "Suggestions for a New Set of Fiscal Indicators", OECD Working Paper No. 79.

Blanchard, O J and F Giavazzi (2004) "Improving the SGP through a Proper Accounting of Public Investment", CEPR Discussion Paper 4420.

Buti, M, J W Friis and R Torre (2022) "How to make the EU fiscal framework fit for the challenges of this decade", VoxEU.org, 10 November.

16 Some institutions use many detailed elasticities, each one for an expenditure or tax component.

17 The usual measures for the expenditure benchmark and the budget balances are computed as a ratio to GDP. For simplification, and without affecting the reasoning, the presentation considers nominal measures.

Canzoneri, M B and B T Diba (1999) "The Stability Pact: A Delicate Balance or an Albatross?", *Empirica* 26: 241-258.

Carnot, N and F de Castro (2015) "The Discretionary Fiscal Effort: an Assessment of Fiscal Policy and its Output Effect", *Hacienda Pública Española/Review of Public Economics* 215: 63-94.

Darvas, Z (2013) "Mind the Gap! And the Way Structural Budget Balances are Calculated", blogpost, Bruegel.

De Haan, J, H Berger and D-J Jansen (2004) "Why Has the Stability and Growth Pact Failed?", *International Finance* 7(3): 235-260.

Eichengreen, B and C Wyplosz (1998) "The Stability Pact: Minor Nuisance, Major Diversion?", *Economic Policy* 26: 67-113.

European Commission (2011), "Public Finance in EMU 2011", *European Economy* 3: 72.

European Commission (2013) "Vade Mecum on the Stability and Growth Pact", European Economy Occasional Papers 151.

European Commission (2019) "Vade Mecum on the Stability and Growth Pact - 2019 Edition", European Economy Institutional Paper 101,

European Commission (2020) "Report on Public Finances in EMU, 2019", European Economy Institutional Paper 133.

European Commission (2022) "Building an economic governance framework fit for the challenges ahead", press release, 9 November.

Eyraud, L, X Debrun, A Hodge, A ViVctor Lledó and C Pattillo (2018), "Second-generation fiscal rules: Balancing simplicity, flexibility and enforceability", IMF Staff Discussion Note 18/04.

Fatás, A (2019) "Fiscal Policy, Potential Output and the Shifting Goalposts", *IMF Economic Review* 67: 684–702.

Feldstein, M (2005) "The Euro and the Stability Pact", *Journal of Policy Modeling* 27(5): 421-426.

Gaspar, V and D Amaglobeli (2019) "Fiscal Rules", SUERF Policy Note 60.

Irish Fiscal Advisory Council (2015) "The EU Expenditure Benchmark: Operational Issues for Ireland in 2016", Analytical Note 7.

Larch, M and A Turrini (2009), "The Cyclically-Adjusted Budget Balance in EU Fiscal Policy Making: A Love at First Sight Turned into a Mature Relationship", European Economy Economic Papers No. 374.

Larch, M and A Turrini (2010), "The Cyclically Adjusted Budget Balance in EU Fiscal Policymaking", *Economic Trends* 45: 48-60.

Larch, M, O Claeys and W van der Wielen (2022) "The Scarring Effects of Major Economic Downturns: The Role of Fiscal Policy and Government Investment", EIB Working Paper 2022/14.

Marinhero, C F (2021) "The Expenditure Benchmark: Complex and Unsuitable for Independent Fiscal Institutions", *Comparative Economic Studies* 63: 411-431.

Orphanides, A (2001) "Monetary Policy Rules Based on Real-Time Data," *American Economic Review* 91: 964-85.

Pench, L, S Derose, G Mourre and N Carnot (2018) "EU fiscal rules: Root causes of its complexity", VoxEU.org, 14 September.

Thygesen, N, S Santacroce, M Szczurek, M Larch, E Orseau, M Gabrijelcic, X Debrun, M Busse, M Bordignon and R Beetsma (2020) "Reforming the EU fiscal framework: Now is the time", VoxEU.org, 26 October.

Wyplosz, C (2011), "Debt Sustainability Assessment: Mission Impossible", *Review of Economics and Institutions* 2(4): 1-37.

Wyplosz, C (2013) "Europe's Quest for Fiscal Discipline", European Economy Economic Papers 498.

Wyplosz, C (2021) "Toward a top-down fiscal rule", unpublished manuscript.

ABOUT THE AUTHOR

CHARLES WYPLOSZ is Emeritus Professor of International Economics at the Graduate Institute in Geneva where he has been Director of the International Centre for Money and Banking Studies (ICMB). He also was Policy Director at CEPR. Previously, he has served as Associate Dean for Research and Development at INSEAD and Director of the PhD programme in Economics at the Ecole des Hautes Etudes en Science Sociales in Paris. His main research areas include financial crises, European monetary integration, monetary policy, fiscal policy discipline and regional integration in various parts of the world. He is the co-author, jointly with Michael Burda, of a popular textbook on macroeconomics and, jointly with Richard Baldwin, of the leading textbook on European economic integration. He was a founding Managing Editor of the review *Economic Policy*. He is a founding contributor to the websites VoxEU and Telos. He has served on the Advisory Scientific Committee of the European Systemic Risk Board, on the Panel of Experts of the European Parliament's Economic and Monetary Affairs Committee, on the 'Conseil d'Analyse Economique', which reports to the Prime Minister of France, and on the French Finance Minister's 'Commission des Comptes de la Nation'. He has been a consultant to several international organizations such as the European Commission, the IMF, the World Bank, the United Nations, the Asian Development Bank, and the Inter-American Development Bank. He has advised the Presidents of France and of the European Commission as well as the governments of the Russian Federation and of Cyprus. Charles Wyplosz holds degrees in Engineering and Statistics from Paris and a PhD in Economics from Harvard University. He has been awarded the Legion d'Honneur by the President of France.