

REM WORKING PAPER SERIES

**Stock-Flow Adjustments and Fiscal Sustainability in the
European Union: A Measurement Reassessment**

José Alves, Laura Mendes, Frederico Silva Leal

REM Working Paper 0423-2026

July 2026

REM – Research in Economics and Mathematics

Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

Any opinions expressed are those of the authors and not those of REM. Short, up to two paragraphs can be cited provided that full credit is given to the authors.





REM – Research in Economics and Mathematics

Rua Miguel Lupi, 20
1249-078 LISBOA
Portugal

Telephone: +351 - 213 925 912

E-mail: rem@iseg.ulisboa.pt

<https://rem.rc.iseg.ulisboa.pt/>



<https://twitter.com/ResearchRem>

<https://www.linkedin.com/company/researchrem/>

<https://www.facebook.com/researchrem/>

Stock-Flow Adjustments and Fiscal Sustainability in the European Union: A Measurement Reassessment*

José Alves[†]

Laura Mendes[‡]

Frederico Silva Leal[§]

July 2026

Abstract

Public debt rises through two channels: the recorded deficit and a residual, the stock-flow adjustment (SFA), that reconciles the deficit with the actual change in debt. A growing literature interprets recurrent SFAs as hidden fiscal risk, yet this conclusion may reflect how the SFA is constructed rather than genuine fiscal dynamics. We show that commonly used measures mechanically embed the fiscal balance, generating strong but spurious correlations even when no true relationship exists. Using Monte Carlo simulations and a newly reconstructed Eurostat-consistent SFA series for the EU-27 (1995–2024), we show that these correlations largely disappear once the accounting identity is respected. The corrected SFA series displays much smaller values, no economically meaningful predictive power for future primary balances, and weakens previous instrumental-variable evidence. Our findings suggest that much of the evidence linking SFAs to fiscal risk reflects measurement error rather than underlying fiscal behaviour, and point to the need for accounting-consistent measures in both research and fiscal surveillance.

Keywords: stock-flow adjustments; fiscal sustainability; measurement; below-the-line operations; fiscal surveillance; European Union.

JEL: H63; H68; H83; C18; E62.

* This work was supported by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., in the framework of the unit UID/06522/2025 - <https://doi.org/10.54499/UID/06522/2025>. The opinions expressed herein are those of the authors and do not necessarily reflect those of the authors' employers. Any remaining errors are the authors' sole responsibility.

[†] ISEG Research, ISEG - Lisbon School of Economics and Management, Universidade de Lisboa, Lisbon, Portugal; UECE - Research Unit on Complexity and Economics; CESifo (Center for Economic Studies and Ifo Institute). Corresponding author: jalves@iseg.ulisboa.pt.

[‡] ISEG - Lisbon School of Economics and Management, University of Lisbon. Email: l60854@aln.iseg.ulisboa.pt

[§] ISCAL - Lisbon Accounting and Business School, Polytechnic University of Lisbon, Portugal; ISEG Research, UECE - Research Unit on Complexity and Economics. Email: fgleal@iscal.ipl.pt

1. Introduction

When a government's debt rises by more than its deficit can explain, the gap carries a bland official name, the stock-flow adjustment, and an increasingly uneasy reputation. It is where bank rescues, equity injections, debt assumptions, and the reclassification of off-budget vehicles come to rest: operations that raise debt without ever crossing the deficit line that fiscal surveillance watches most closely. Because they are large and recurrent, a growing literature has come to read persistent stock-flow adjustments (SFAs) as a form of hidden fiscal risk (von Hagen and Wolff, 2006; Milesi-Ferretti, 2004; Buti et al., 2007; Jaramillo et al., 2016), namely below-the-line debt that deficit-centred monitoring misses today and pays for tomorrow.

That reputation sits awkwardly beside how the European Union actually treats the SFA. Under the bloc's reformed fiscal rules, the projected stock-flow adjustment feeds directly into the net-expenditure path each member state is now legally bound to follow. Yet, the European Commission assumes it away, setting it to zero beyond the first projection years for nearly every country. The same residual is thus, on one reading, the next sovereign-debt surprise waiting to happen. On the other, it is a rounding term safely projected at zero. Both cannot be true, and the disagreement is not academic. It shapes the fiscal adjustment that twenty-seven governments are required to deliver.

We argue that, for the European Union in aggregate, the alarmist reading is an artefact of construction. The stock-flow adjustment is not a measurement in its own right. It is a residual, pinned down by the accounting identity that ties the change in debt to the deficit, and what one concludes about it depends sharply on how it is rebuilt. Built as the identity requires, netting the overall, interest-inclusive balance, it is modest and tells us little about future fiscal performance. Built in any way that loads the deficit back in, most realistically by netting the primary balance (the textbook decomposition stripped of its interest-growth term), but equally by the wrong-sign version we use as the cleanest illustration, it quietly absorbs the deficit, and the well-known "hidden risk" facts then follow almost mechanically. A residual that contains the deficit correlates strongly and negatively with the primary balance, because it is, in part, the primary balance. It appears to explode in crises, because crises are when deficits are largest, and it lends apparent strength to any instrument that tracks the cycle. The mistake is easy to make precisely because the correct object exists. Eurostat publishes the stock-flow adjustment with the right sign, so a user assumes any reconstruction inherits it, and the primary-balance decomposition is a principled-looking starting point that needs only one dropped term to go wrong. The

resulting stylized facts are not news about government balance sheets. They are news about how a residual is assembled.

We make the case in three steps, then take it to the data. First, we show that the conventional construction manufactures the negative SFA-balance correlation mechanically (Proposition 1). In a simulation where the true adjustment is, by construction, unrelated to the fiscal balance, the mis-built measure still returns a correlation near minus one. Second, we rebuild the EU-27 stock-flow series for 1995–2024 strictly to the canonical identity and validate it, cell by cell, against the figures Eurostat itself publishes (Section 6). Read correctly, the familiar facts dissolve: the average SFA is modest and, if anything, debt-reducing. The notorious peripheral “crisis spike” of over fourteen percent of GDP shrinks to essentially zero, and the strong negative SFA-balance correlation that anchors the creative-accounting reading falls from -0.77 to -0.13 . Third, we ask the question the descriptive literature mostly sets aside, whether the residual actually predicts later fiscal performance, and find that, correctly measured, it does not. Not at any horizon, no adverse signal in any subgroup, not even among the high-debt countries where the danger is supposed to be greatest (Section 7). Once the deficit is netted out, the instruments that seemed to identify a causal effect lose all their power. The apparent identification was an artefact, and for the genuine residual the causal question is left open rather than closed, for a telling reason: the interventions that made those instruments strong, Ireland’s bank rescues and the Greek programme, were booked above the line, in the deficit, not below it.

The lesson is not that the public balance sheet is harmless; single operations move debt by tens of percentage points and warrant close scrutiny. The implication is narrower, but more useful. In EU aggregate data the below-the-line margin, measured as the identity requires, is neither large nor a leading indicator of fiscal weakness, and the evidence offered to the contrary turns out to be an accounting reflection of the deficit. Conclusions about a residual are only as trustworthy as the residual’s construction, a caution that reaches well beyond the SFA to the many fiscal aggregates assembled the same way.

The remainder of the paper is organised as follows. Section 2 situates the contribution within the literatures on creative accounting, SFA measurement, accounting-identity econometrics, and fiscal sustainability. Section 3 describes how the SFA is defined in EU debt accounting and how it is embedded in the reformed surveillance framework. Section 4 states the measurement problem and proves the mechanical-artefact result. Section 5 sets out the empirical strategy. Section 6 describes the data and presents the

corrected stylized facts, contrasting the conventional measure with the correct one. Section 7 asks whether correctly measured SFAs predict subsequent fiscal performance and reports the robustness checks. Section 8 draws the implications for surveillance, and Section 9 concludes.

2. Related literature

Public debt rises through two channels: the recorded deficit and a residual, the stock-flow adjustment, that reconciles the deficit with the actual change in the debt stock. That this residual is large and systematic, rather than a rounding term, is by now well established. von Hagen and Wolff (2006), studying the EU, show that headline deficits explain only part of the change in debt and that the residual bunches just below the Stability and Growth Pact's three-percent ceiling, as though governments shift adjustment toward less-monitored margins as a deadline binds. Milesi-Ferretti (2004) supplies the mechanism: when fiscal rules constrain the visible deficit more tightly than the balance sheet, governments substitute below-the-line operations for on-budget adjustment, so that "creative accounting" becomes a rational response to rules rather than measurement noise. Buti et al. (2007) embed this in a broader account of how rules reshape accounting behaviour, while Koen and van den Noord (2005) catalogue the specific devices, including securitisations, sale-and-leaseback operations, pension-fund transfers, and deferred payments, deployed across Europe.

A largely political-economy literature has since tested and refined the substitution hypothesis. Reischmann (2016), measuring creative accounting precisely as the debt-deficit residual, finds that OECD governments inflate stock-flow adjustments around elections to flatter the reported balance. Alt et al. (2014) show that the Stability and Growth Pact raised gimmickry mainly in low-transparency member states, so that transparency blunts the incentive.

Two features of this body of work frame what follows. First, its object is the concealment of current deficits, a contemporaneous, behavioural phenomenon, not a claim that today's residual forecasts tomorrow's fiscal position. Second, its empirical signature is a strong negative association between the residual and the fiscal balance, and it is precisely this regularity that we show to be, under the constructions in common use, an accounting artefact.

A parallel literature asks how the residual should be measured and what it contains. The debt-accumulation identity decomposes the change in the debt ratio into an interest-

growth (“snowball”) term, the primary balance, and the stock-flow adjustment (Escolano, 2010; D’Erasmus et al., 2016). Equivalently, and this is the construction we insist upon, the SFA is the change in gross debt net of the overall, interest-inclusive balance. Weber (2012) relates cross-country SFAs to fiscal transparency. Seiferling (2013), exploiting the government’s integrated balance sheet, finds that correctly measured residuals are substantially smaller than commonly assumed and uncorrelated with transparency, a null our EU evidence reinforces. On magnitude and determinants, Campos et al. (2006) document large “unexplained” debt changes in emerging markets, Afonso and Jalles (2020) show by dynamic panel methods that SFAs contribute to debt accumulation almost one-for-one and that fiscal frameworks dampen them, and a related strand links the residual to debt management and interest costs (Afonso and Alves, 2019; Casalin et al., 2023). This work establishes that SFAs are quantitatively important and institutionally patterned; but it is overwhelmingly about accounting and determinants, not about whether the residual predicts later fiscal outcomes.

That predictive, “hidden-risk” turn is more recent, and is the proximate target of this paper. Jaramillo et al. (2016) show that large public-debt spikes are driven less by primary deficits than by stock-flow adjustments, framing the SFA as a forecasting “blind side”. Subsequent work extends the SFA-debt-spike-hazard link to broader samples. The concern has since migrated into the operational core of European fiscal governance. Under the reformed framework – a turn toward “second-generation” rules built around medium-term debt projections (Eyraud et al., 2018) – each member state’s binding net-expenditure path is derived from a debt-sustainability analysis in which the projected SFA appears directly, and in which Bachleitner and Prammer (2024) identify it as a leading source of forecast error. The literature thus leaves an unresolved tension: surveillance assumes the SFA away beyond the short term, even as a growing body of work urges that recurrent SFAs be treated as a risk indicator.

Our methodological contribution draws on a literature, so far unconnected to public finance, on the hazards of regressing a variable on a component of its own accounting identity. The same logic animates the long debate over investment and cash-flow sensitivities (Fazzari et al., 1988), where regressor and outcome share balance-sheet components, and the general statistical critique of ratio correlations (Kronmal, 1993). Sánchez-Vidal (2025) formalises this as the accounting semi-identity problem and shows by simulation that it can generate most of the explained variation in widely used corporate-finance regressions. To our knowledge no one has brought this lens to the

stock-flow adjustment. Yet the case is exact, because the SFA is, by definition, the change in debt minus the deficit, so that regressing a balance measure on it recovers the algebra of the identity unless the SFA is constructed correctly. We advance on the two nearest precedents in specific ways. Sánchez-Vidal (2025) documents the semi-identity problem by simulation for corporate-finance ratios. We give a closed-form mechanical term for a specific, policy-relevant fiscal aggregate, together with three testable corollaries: horizon decay, first-stage contamination, and the construction-dependence of the descriptive facts. Seiferling (2013) finds correctly measured residuals smaller using a fuller government balance sheet, but our point is sharper: the apparent largeness is a property of the regressor's construction, present even when the data source is held fixed, and we pair it with a powered predictive null.

Finally, because “hidden fiscal risk” is ultimately a claim about sustainability, the paper connects to the sustainability literature and to the identification tools its causal version employs. That tradition asks whether primary balances respond to debt (Bohn, 1998; Bohn, 2007; Mendoza and Ostry, 2008) and how the response weakens at high debt, a phenomenon known as “fiscal fatigue”, thereby bounding fiscal space (Ghosh et al., 2013; D’Erasmus et al., 2016). The causal SFA claim is typically operationalised through local projections (Jordà, 2005), instrumented with shift-share (Bartik) designs (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022; Adão et al., 2019) and assessed with weak-instrument-robust inference (Montiel Olea and Pflueger, 2013; Lee et al., 2022). We adopt exactly this toolkit and show that the apparent first-stage strength of such instruments for the SFA is itself a by-product of the deficit contamination, dissolving once the SFA is measured correctly.

Three gaps emerge from this survey, and the paper is organised to fill them. First, although the negative SFA-balance association is the empirical bedrock of both the creative-accounting and the hidden-risk literatures, no study has asked whether that association is mechanical (an artefact of how the residual is constructed) rather than behavioural. We show that it is largely the former, and we are, to our knowledge, the first to bring the accounting-semi-identity critique to bear on fiscal aggregates. Second, the literature lacks a validated, reproducible EU stock-flow series built strictly to the canonical identity and benchmarked to official data, so that the descriptive facts motivating the hidden-risk reading, namely large average SFAs and crisis-era spikes, have never been re-examined under correct measurement. Third, the predictive claim itself, that recurrent SFAs foreshadow fiscal deterioration, has not been tested for the EU

on a correctly measured, adequately powered panel with credible identification. This paper fills all three: it makes the mechanical artefact precise and proves it; it supplies and validates the corrected EU series and overturns the contaminated stylized facts; and it shows that, correctly measured, EU stock-flow adjustments neither average large nor predict subsequent fiscal performance, while the instruments that appeared to identify a causal effect have no power for the genuine object. The lesson is not that the public balance sheet is unimportant, but that conclusions about it are unusually sensitive to construction and must be disciplined by the officially measured series.

3. Institutional background: the SFA in EU debt accounting and surveillance

To see why the construction of the stock-flow adjustment carries such weight, it helps to know exactly what the residual reconciles. The European Union monitors not one fiscal number but two: the general-government deficit, measured above the line on an accrual basis, and the consolidated gross “Maastricht” debt, a stock measured at nominal (face) value and covering only currency and deposits, debt securities, and loans (Council Regulation (EC) No. 479/2009, referencing ESA 2010). An accounting identity ties the two together, and the SFA is whatever makes that identity hold: the part of a year’s change in debt that the deficit does not explain. Eurostat treats the reconciliation as a routine quality check. If the deficit and the change in debt do not square, the figures are wrong, and member states report it twice a year, before 1 April and 1 October, in Table 3 of the Excessive Deficit Procedure notification.

Inside that residual is a strikingly heterogeneous list, codified in Eurostat’s Manual on Government Deficit and Debt and published, item by item, in a dedicated semi-annual note. It runs to some seventeen components in three families. The first is the government’s net acquisition of financial assets, including currency and deposits, debt securities, loans, and, most consequentially, equity and investment-fund shares. When a state borrows to recapitalise a bank, the cash it raises adds to debt while the stake it buys is an asset, so the operation lands here, lifting debt without ever crossing the deficit line. The second family is a set of accounting adjustments. Here sit liabilities that move debt but stay outside the deficit; valuation effects that arise precisely because the debt is carried at face value while transactions occur at market prices (a bond issued below par, the wedge between interest accrued and interest paid, a foreign-currency bond repriced by the exchange rate), and the item that has repeatedly moved European debt ratios by double digits: the reclassification of an entity into the government sector. The third family is the

residual of the residual: the statistical discrepancy between the deficit and its financing. The economic content of this list is exactly the “below-the-line” margin the hidden-risk literature has in mind. What that literature tends to overlook is that, by the same identity, the residual also contains the deficit it is measured against. That is the whole of our argument.

The episodes that earned the SFA its reputation are real and large. When Greece’s statistics were restated in 2009–2010, the reported 2009 deficit climbed from 3.7% of GDP in the April 2009 notification to 15.4% by November 2010, and Eurostat lodged formal reservations on the data. This was measurement, not policy, rewriting a country’s fiscal position. Germany’s debt jumped from 74.4 to 83.2% of GDP between 2009 and 2010, almost nine points with no matching deficit, when the wind-down (National Asset Management Agency) vehicles of Hypo Real Estate and WestLB (holding roughly EUR 240 billion of assets) were reclassified into the government sector (Eurostat News Release 153/2011). Ireland’s bank rescues and its NAMA vehicle worked through the same channel. More recently, Italy’s “Superbonus” tax credit has generated large debt-increasing adjustments through a timing wedge between when credits are granted and when they are cashed. Finland’s funded public pensions, which save in financial assets, throw off a structural debt-increasing SFA. On our correctly measured series, this averages about 0.9% of GDP a year (Section 7.1). These are not curiosities. They are the reason a careful observer would want the residual measured precisely.

That precision now has legal teeth, because the SFA has been written into binding rules. The Stability and Growth Pact was overhauled in April 2024 (Regulations (EU) 2024/1263 and 2024/1264), and the new regime replaces a patchwork of targets with a single operational indicator: a multi-year ceiling on nationally-financed net primary expenditure that each member state must respect over a four-year plan, extendable to seven. The ceiling is not set politically. It is reverse-engineered from the European Commission’s debt-sustainability analysis, which must show debt on a “plausibly downward path” by the end of the plan and declining with at least seventy-percent probability for five years after, subject to numerical safeguards (debt falling by at least one point of GDP a year while it exceeds 90%, half a point between 60 and 90). The stock-flow adjustment enters this machine through the debt-accumulation arithmetic itself, and here is the convention that prompts this paper. In the Commission’s own words, “stock-flow adjustments are in line with the Commission forecast up to T+2 and set to zero afterwards,” with named exceptions only for Finland, Greece, and Luxembourg

(European Commission, 2024). For every other member state, the projected residual is exactly zero beyond a two-year horizon, whether in the central scenario, in the three deterministic stress tests, or in the stochastic simulation of ten thousand shocks alike. Because a positive SFA would raise the projected debt path and demand a tighter expenditure ceiling, zeroing it is not a neutral simplification but a thumb on the scale. It is applied uniformly to a quantity that the historical record shows is anything but uniform, and one that ranks among the principal drivers of the Commission’s own debt-forecast errors (Bachleitner and Prammer, 2024).

The bodies set up to scrutinise the framework have noticed but cannot act. The European Fiscal Board, re-established under the 2024 regulation, has flagged the zero-SFA assumption and pressed the Commission to project the residual transparently for all member states rather than three. Yet it sits only as an observer in the expert group that fixes the methodology, free to criticise a convention it has no vote to change. National independent fiscal institutions have a thinner role still: their scrutiny of the plans’ macroeconomic assumptions becomes mandatory only in 2032. The contrast with international practice is sharp. The IMF’s Sovereign Risk and Debt Sustainability Framework does the opposite of assuming the residual away: it treats the SFA as a debt driver to be actively policed, colour-coding a country’s past SFA forecast errors against its peers, flagging implausibly favourable residuals, and, by default, excluding optimistic stock-flow contributions from its risk fan-chart unless they can be justified. Nor is that vigilance idle: across 172 countries, a one-point rise in the SFA-to-GDP ratio has been estimated to raise the hazard of a debt spike by roughly 17 to 26 percent (Andrian et al., 2025; Jaramillo et al., 2016). The stock-flow adjustment thus reaches the centre of EU surveillance bearing two contradictory institutional verdicts: assumed away by the rule that binds, hunted down by the framework that does not. This leaves an unresolved measurement question that the remainder of this paper takes up.

4. Theoretical framework

A theory of when the stock-flow adjustment is informative must begin with what the SFA actually is, and the answer is unusually literal: it is whatever the debt-accumulation identity says it is. Let b_t be the gross general-government debt-to-GDP ratio and g_t the overall fiscal balance (net lending, a surplus positive), both as shares of GDP. Debt accumulates according to

$$b_t - b_{t-1} = -g_t + sfa_t, \quad \Rightarrow \quad sfa_t \equiv (b_t - b_{t-1}) + g_t(1)$$

The overall balance itself splits into the primary balance p_t and the interest the government pays, $g_t = p_t - i_t$ with $i_t \geq 0$. Carrying the interest term one step further, into an interest–growth “snowball” component, recovers the textbook debt-dynamics decomposition (Escolano, 2010). Either way, the stock-flow adjustment in the identity is the change in debt net of the overall balance – equivalently, the change in debt minus the deficit – and it is the construction Eurostat applies in the Excessive Deficit Procedure. Two features of the identity drive everything that follows. First, the SFA is not an independent measurement but a residual, fixed by the identity once debt and the balance are known. Second, it is defined in terms of the balance, so that a regression of the SFA on any balance measure is, in part, estimating a relationship the identity has already imposed.

Because the SFA is a residual, constructing it requires a choice of which balance to net, and applied work does not always make the choice the identity dictates. Two departures recur. One enters the balance with the wrong sign; the other nets the primary balance in place of the overall balance:

$$(i) \text{ wrong sign: } \widetilde{sfa}_t = (b_t - b_{t-1}) - g_t \quad (2)$$

$$(ii) \text{ primary balance: } \widehat{sfa}_t = (b_t - b_{t-1}) - p_t \quad (3)$$

The second is especially telling. Netting the primary balance is what the textbook decomposition does after also subtracting the interest–growth term, so dropping that term leaves the government’s entire interest bill lodged inside the measured “SFA”. Substituting the identity shows what the two constructions really are:

$$\widetilde{sfa}_t = sfa_t - 2g_t, \quad \widehat{sfa}_t = sfa_t - g_t - p_t \quad (4)$$

Each equals the true adjustment minus a multiple of the balance. They do not mismeasure the SFA by adding noise. They add back the very deficit the SFA was built to remove.

Throughout we treat the series (p_t, g_t, sfa_t, i_t) as covariance-stationary with finite second moments, so that the population moments in what follows are well defined; the panel analogue requires the cross-country second moments to exist and is otherwise unchanged.

The remainder of this section develops the argument in two steps. Proposition 1 establishes the paper's central result: once the fiscal balance is embedded in the construction of the stock-flow adjustment, the resulting measure acquires a mechanical covariance with fiscal outcomes unrelated to behaviour. The three corollaries then trace the implications of this result for contemporaneous regressions, dynamic local projections, and instrumental-variable estimation. Proposition 2 subsequently shows that the same logic extends beyond stock-flow adjustments to residuals defined by accounting identities more generally.

Proposition 1. Decompose each construction against the identity $b_t - b_{t-1} = -g_t + sfa_t$. The correct construction reproduces the true adjustment, $sfa_t \equiv (b_t - b_{t-1}) + g_t$, so its covariance with the primary balance is purely behavioural: $cov(p_t, sfa_t)$ carries no mechanical component. The two contaminated constructions instead satisfy $\widetilde{sfa}_t = sfa_t - 2g_t$ and $\widehat{sfa}_t = sfa_t - g_t - p_t$, so that $cov(p_t, \widetilde{sfa}_t) = cov(p_t, sfa_t) - 2cov(p_t, g_t)$, $cov(p_t, \widehat{sfa}_t) = cov(p_t, sfa_t) - cov(p_t, g_t) - var(p_t)$.

Because $g_t = p_t - i_t$ with the interest bill i_t comparatively smooth, $cov(p_t, g_t) \approx var(p_t) > 0$. Both contaminated measures therefore carry a mechanical component of order $-2var(p_t)$ in their covariance with the primary balance, independent of the sign or magnitude of the true $cov(p_t, sfa_t)$.

Proof. From the identity $b_t - b_{t-1} = -g_t + sfa_t$, the correct construction satisfies $(b_t - b_{t-1}) + g_t = sfa_t$. Substituting into the two contaminated definitions, $\widetilde{sfa}_t = (b_t - b_{t-1}) - g_t = sfa_t - 2g_t$ and, using $p_t = g_t + i_t$, $\widehat{sfa}_t = (b_t - b_{t-1}) - p_t = sfa_t - g_t - p_t$. Taking the covariance with p_t and applying bilinearity gives the two displayed expressions. Finally, $g_t = p_t - i_t$ implies $cov(p_t, g_t) = var(p_t) - cov(p_t, i_t)$; when the interest bill is smooth relative to the primary balance, $cov(p_t, i_t) \approx 0$, so $cov(p_t, g_t) \approx var(p_t) > 0$ and the mechanical term dominates.

The proposition makes the contamination quantitative. The overall and primary balances move almost one for one, because the interest bill i_t is small and slow relative

to the swings in p_t ; hence $\text{cov}(p_t, g_t) \approx \text{var}(p_t)$, and each contaminated measure carries a covariance with the primary balance of order $-2\text{var}(p_t)$. The sign and size of that term owe nothing to government behaviour. They would be present if the true SFA were white noise, and, as the next two results show, they survive into every statistic the hidden-risk reading relies on.

Corollary 1 (Spurious regression slope). The population slope from regressing the contemporaneous primary balance on the wrong-sign SFA is $\text{plim } \hat{\beta} = \frac{\text{cov}(p_t, \widetilde{sfa}_t)}{\text{var}(\widetilde{sfa}_t)} = \frac{\text{cov}(p_t, sfa_t) - 2\text{cov}(p_t, g_t)}{\text{var}(\widetilde{sfa}_t)}$, which is negative whenever $\text{cov}(p_t, sfa_t) < 2\text{cov}(p_t, g_t)$ that is, for any true relationship short of an implausibly strong positive one. Under the correct construction the same regression returns $\text{cov}(p_t, sfa_t)/\text{var}(sfa_t)$, the behavioural slope.

Corollary 2 (Horizon decay). The same bias governs the predictive regression that motivates the hidden-risk reading. Regressing the primary balance h periods ahead on the contemporaneous wrong-sign SFA gives $\text{cov}(p_{t+h}, \widetilde{sfa}_t) = \text{cov}(p_{t+h}, sfa_t) - 2\text{cov}(p_{t+h}, g_t)$. Because the primary balance is persistent but mean-reverting, the cross-horizon covariance $\text{cov}(p_{t+h}, g_t)$ shrinks toward zero as h grows – geometrically, $\approx \rho^h \text{var}(p_t)$, when the balance follows an AR(1) with coefficient ρ , and at the rate of its own autocovariance function more generally. The spurious predictive coefficient therefore fades with the horizon even when the true SFA has no predictive content whatever. Corollary 2 matches the conventional profile in Table 5 closely: read on the contaminated SFA, the impact coefficient of -0.35 decays monotonically to -0.04 by the fifth year - the geometric fingerprint of a vanishing balance autocovariance, not of a genuine but fading economic effect.

Corollary 3 (Contaminated instruments). Let Z_t be any instrument proposed for the SFA (written Z_{it} in the panel of Section 5). Its first-stage coefficient on the wrong-sign measure is $\frac{\text{cov}(Z_t, \widetilde{sfa}_t)}{\text{var}(Z_t)} = \frac{\text{cov}(Z_t, sfa_t)}{\text{var}(Z_t)} - 2 \frac{\text{cov}(Z_t, g_t)}{\text{var}(Z_t)}$. An instrument that is correlated with the overall balance therefore acquires first-stage strength from the second term whether or not it predicts the genuine SFA.

Corollary 3 is the hinge between the framework and the data. Instruments built from crisis-era financial operations (including bank recapitalisations, programme entry, asset-class exposure) are mechanically correlated with the balance, because such operations cluster with large deficits. They will look strong against \widetilde{sfa}_t and weak against sfa_t . The corollary thus predicts, before any estimation, the pattern Section 7 documents: the same three instruments deliver first-stage F statistics of roughly six to twenty-three against the contaminated SFA and collapse to F below 1.5 against the correct one.

None of this is special to public finance. It is the local instance of the accounting-semi-identity hazard surveyed in Section 2: when a regressor is built from the same accounting aggregates as the outcome, ordinary least squares recover the algebra of the identity before it recovers any behaviour (Sánchez-Vidal, 2025). The SFA is a textbook semi-identity regressor, since it is the change in debt minus the deficit, and Corollaries 1–3 are simply what that logic implies for it.

Proposition 2 (General semi-identity bias). Let an accounting identity define a residual $r^* = y - \lambda x$ from an aggregate y , a balance-type component x , and a known weight λ . An analyst who rebuilds the residual with the wrong weight, $r = y - \tilde{\lambda}x$ with $\tilde{\lambda} \neq \lambda$, has $r = r^* - (\tilde{\lambda} - \lambda)x$, so the population slope of any regress and u on r is $\text{plim}\hat{\beta} = \frac{\text{cov}(u, r^*) - (\tilde{\lambda} - \lambda)\text{cov}(u, x)}{\text{var}(r)}$. The first term is behavioural. The second is a mechanical artefact of the construction error $\tilde{\lambda} - \lambda$, nonzero whenever the regress and u covaries with the netted component x , even if the genuine residual r^* is unrelated to u .

Proposition 1 is the case $y = \Delta b$, $x = g$, $\lambda = -1$ (so $r^* = sfa$), $\tilde{\lambda} = +1$, and $u = p$, for which $(\tilde{\lambda} - \lambda)\text{cov}(u, x) = 2\text{cov}(p, g) \approx 2\text{var}(p)$. The same template covers any fiscal residual that is a linear function of the balance it is later regressed against, so the discipline we urge generalises: build the residual exactly to the identity, and validate it against the official series before reading any correlation as behaviour.

This is not a thought experiment. Consider a second identity-defined residual on the same EU panel: the debt-stabilising primary gap $PG_t = p_t - p_t^*$, where $p_t^* = [(i_t - \gamma_t)/(1 + \gamma_t)] b_{t-1}$ is the primary surplus that would hold the debt ratio constant at the effective interest rate i_t and nominal growth γ_t , a standard sustainability indicator. The debt-accumulation identity makes it a semi-identity regressor in exactly the sense of Proposition 2: $\Delta b_t = -PG_t + sfa_t$, an identity that closes to machine precision on our data, with the implied residual correlating with the Eurostat SFA at 0.98. A researcher

who runs the natural “does fiscal effort reduce debt?” regression of the change in the debt ratio on the gap recovers a slope of -1.06 (s.e. 0.10): each point of gap appears to cut debt almost one for one. That number is the budget constraint, not a behavioural response (it is mechanically near -1 because Δb is $-PG + sfa$) and it would survive even if the gap carried no information whatever. The same trap catches the contemporaneous association between the primary balance and the gap, a mechanical +0.92 because the gap contains the balance. A second residual, the same lesson: identity-defined regressors recover the identity before they reveal behaviour, and the only protection is to build and validate them as the identity dictates.

Two conclusions for the empirical work follow at once. The correct construction is clean: because $(b_t - b_{t-1}) + g_t$ equals the true adjustment, its relationship with the balance is whatever the data make it, with no mechanical term to subtract. And a genuinely behavioural relationship between the SFA and later fiscal performance can be recovered only from variation in the SFA that is orthogonal to the contemporaneous balance (variation an instrument is meant to supply). Corollary 3 warns that supplying it is harder than it looks, because the obvious instruments are themselves balance-correlated; whether any instrument clears that bar for the European Union is the question Section 7 settles.

Remark 1 (Not classical measurement error). The bias in the decomposition above is not the attenuation familiar from errors-in-variables. Classical measurement error adds noise orthogonal to the truth and biases slopes toward zero. Here, the “error” is the balance itself (a term systematically related to the outcome) so the bias is structured rather than attenuating, and it can reverse the sign of the estimated relationship. A larger sample does not cure it, and a more precisely measured debt series does not either. Thus, the problem is the definition, not the data.

Remark 2. The artefact’s magnitude is easiest to see by simulation. Drawing a true SFA that is, by construction, uncorrelated with the primary balance, the correct measure returns a correlation of zero while the wrong-sign and primary-balance measures return -0.91 and -0.94. The result is invariant to the truth. The true correlation can be set anywhere from -0.6 to +0.6 without moving the contaminated measures off -0.9.¹

¹ For reasons of parsimony, the Monte Carlo illustration of Proposition 1’s results is available upon request.

5. Empirical strategy

The framework of Section 4 dictates the empirical agenda. By Proposition 1 and its corollaries, the descriptive moments and regression slopes that anchor the hidden-risk reading are largely predictable from the construction of the SFA alone, so the empirical task is twofold: measure the residual correctly and find variation in it that is orthogonal to the contemporaneous balance. We attack the first with a measurement experiment and the second with instrumental variables, and we use the contrast between the two SFA constructions as a running diagnostic. The object of interest throughout is the sequence of dynamic responses $\{\beta_h\}$ of the primary balance to a stock-flow shock; the hidden-risk hypothesis is that they are negative and economically meaningful at medium horizons, so that recurrent below-the-line accumulation foreshadows fiscal deterioration.

The measurement experiment. The first and simplest design is the paper’s identifying idea in miniature (Section 6). Holding the EU-27 sample, the control set, and every line of downstream code fixed, we vary only the balance term, computing $\widetilde{sfa}_{it} = \Delta b_{it} - g_{it}$ and $sfa_{it} = \Delta b_{it} + g_{it}$ from the same Eurostat debt and balance series. Any statistic that differs between the two, whether a mean, a correlation, a regression slope, or a first-stage F , is attributable to the balance term and to nothing else. The experiment invokes no exclusion restriction and no functional-form assumption. It is the cleanest possible isolation of the contamination the theory describes, and it carries the bulk of the descriptive evidence.

Dynamic specification. To trace the response of fiscal performance to the SFA we use local projections (Jordà, 2005). For each horizon $h = 0, 1, \dots, 5$ we estimate

$$p_{i,t+h} = \beta_h sfa_{it} + \gamma_h' \mathbf{x}_{i,t-1} + \varepsilon_{it}^h \quad (5)$$

where $p_{i,t+h}$ is the primary balance h years ahead, sfa_{it} is the residual under whichever construction is in play, and $\mathbf{x}_{i,t-1}$ stacks lagged controls – the debt ratio, the output gap, real GDP growth, and inflation, each standardised. The sample begins in 1997, because the instrument shares introduced below are built from a 1994–1996 pre-rule window. We deliberately keep the local projection pooled, without country fixed effects: the shift-share instrument is populated for only a handful of member states, so country effects would

absorb almost all of its identifying cross-sectional variation. Standard errors are clustered by country, and we also report heteroskedasticity-robust (HC1) errors, which are if anything smaller. Estimated by ordinary least squares, the sequence $\{\beta_h\}$ from the local projection is exactly the object of Corollaries 1 – 2. On the conventional SFA it must trace a mechanical negative gradient that decays with the horizon. On the correct SFA it isolates whatever genuine relationship remains – an association, at this stage, rather than a causal effect, a distinction we take up next.

Identification. Least squares on the local projection cannot deliver a causal β_h . Because sfa_{it} is the residual of the contemporaneous balance it is correlated with the error by construction (Corollary 1), and, behaviourally, the same banking distress or rule-driven pressure that produces below-the-line operations also moves later primary balances directly. A causal effect can be recovered only from an instrument Z_{it} that shifts the SFA without otherwise predicting the outcome. Two of the required conditions are standard: relevance ($\text{cov}(Z_{it}, sfa_{it}) \neq 0$), and exclusion ($\text{cov}(Z_{it}, \varepsilon_{it}^h) = 0$). However, the theory adds a third, specific to this setting and easy to overlook.

Assumption 1 (Orthogonality to the balance). The instrument is uncorrelated with the contemporaneous overall balance: $\text{cov}(Z_{it}, g_{it}) = 0$.

Assumption 1 is not a validity requirement – a valid instrument for the true SFA needs only relevance and exclusion – but a diagnostic condition. By Corollary 3, it is precisely the condition under which the first stages on the contaminated and the correct SFA coincide; when it fails, the contaminated first stage is inflated by $-2\text{cov}(Z_{it}, g_{it})/\text{var}(Z_{it})$, so the instrument looks strong while tracking only the deficit. The gap between the two first stages is therefore a direct measure of how much an instrument’s apparent strength is borrowed from the balance, and we read every first-stage comparison in that light.

Instruments. We bring three sources of below-the-line variation to the local projection. The first is a Bartik shift-share (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022; Adão et al., 2019). Country i ’s exposure to asset class k is its pre-rule composition of net asset-class acquisitions, ω_{ik} , averaged over 1994–1996; the shift is the leave-one-out EU innovation in that class, $\Delta \bar{a}_{kt}^{-i} = (N - 1)^{-1} \sum_{j \neq i} \Delta a_{jkt}$; and the instrument is their inner product, $Z_{it} = \sum_k \omega_{ik} \Delta \bar{a}_{kt}^{-i}$, the leave-one-out construction preventing a country’s own balance-sheet moves from entering its own instrument.

Identification follows the shares-exogeneity logic of Goldsmith-Pinkham et al. (2020): conditional on the controls, a country’s pre-rule portfolio is taken to predict its later primary balance only through its exposure to common SFA shocks. This is a strong assumption. A country’s pre-rule asset-class portfolio plausibly correlates with persistent fiscal institutions that also shape its later balances, and we do not lean on it for a causal estimate. The instrument’s role here is the contrast of Corollary 3: it is informative for the contaminated SFA and uninformative for the correct one, a diagnostic of deficit-contamination rather than a source of a structural parameter, which clears a far lower bar than full shares-exogeneity (Borusyak and Hull, 2025). We are candid that these shares are populated for only six member states (Belgium, Denmark, Spain, Hungary, Italy, and Sweden), so the design identifies a response across a narrow, transparent subset of the panel rather than an EU-wide elasticity, and we enter the four asset-class components separately to permit over-identification testing.

The second instrument is the year-on-year change in the Eurostat-verified debt impact of bank recapitalisations, from the April-2026 supplementary table on government interventions. It is relevant because a recapitalisation is itself a below-the-line operation, but its exclusion restriction is the weakest of the three, since rescues cluster with the deficits that crises generate. The third is the debt impact of ESA-2010 sector reclassifications: because the timing and incidence of a reclassification are decided by Eurostat under statistical rules rather than by national fiscal authorities, it has the strongest claim to Assumption 1, even though its first stage is the thinnest.

Estimation and inference. Each instrument enters a two-stage least-squares version of the local projection; with the four asset-class components we additionally report limited-information maximum likelihood, which is less biased under weak identification, together with the Hansen J over-identification test. Because identification rests on so few clusters, we do not lean on the “ $F > 10$ ” rule of thumb or on asymptotic cluster-robust intervals for the structural coefficient: we report the Montiel Olea-Pflueger effective F (Montiel Olea and Pflueger, 2013) as the appropriate gauge of instrument strength under heteroskedasticity, and we form weak-instrument-robust confidence sets by Anderson-Rubin inversion (Lee et al., 2022), which has correct size at any first-stage strength. The affirmative evidence that the correct SFA carries no large predictive content does not rest on the instrument at all – it is the powered equivalence bound of Section 7, which states “no predictive relationship” as a quantitative upper bound on the effect we could have

missed rather than a bare failure to reject. The conventional and correct SFAs pass through identical code at every step; the comparison of their first stages and coefficients is the paper’s central piece of evidence, and Assumption 1 is what tells us how to read it.

6. Data and the measurement experiment

The analysis uses an annual panel of the EU-27 member states over 1995–2024. The two series that define the SFA, namely the consolidated gross (Maastricht) debt ratio and the overall balance (net lending/borrowing, B.9), are taken from Eurostat’s Excessive Deficit Procedure statistics (`gov_10dd_edpt1`), so that both are measured on the same vintage and definition. The correct stock-flow adjustment is constructed cell-by-cell from the identity $sfa_{it} = (b_{it} - b_{i,t-1}) + g_{it}$, where b_{it} is the gross debt ratio and g_{it} the overall balance. By construction, this equals Eurostat’s published stock-flow adjustment. We use the reconstructed series rather than the published one so that the same code generates the conventional and the correct measures, and the contrast that follows reflects only the balance term. The primary balance, output gap, real GDP growth, and inflation (used as the outcome and as lagged controls) are from the IMF World Economic Outlook (October 2024 vintage).

Two checks validate the reconstruction. First, the panel’s overall balance matches Eurostat’s (B.9) at a correlation of 0.96, confirming that the series entering the identity is the overall, not the primary, balance. Second, an SFA reconstructed independently from the macro-fiscal panel’s debt and balance matches the Eurostat-native series at a correlation of 0.97 in the pool and 0.98 at the country-level median, with 24 of 27 members above 0.90, whereas the same panel built with the wrong sign matches it at only 0.54, a first indication that the balance term, not the data source, drives the discrepancy. Descriptively, the correctly measured SFA averages -2.4% of GDP with a standard deviation of 7.0, somewhat less volatile than the conventional construction, and opposite to it in average sign (mean +2.8%, standard deviation 9.9). A single observation (Bulgaria in 1997), a currency-board nominal-GDP collapse rather than a below-the-line operation, accounts for the extreme minimum. It inflates the dispersion but leaves the correlations and the predictive null unchanged (Section 7.1).

We now run the measurement experiment of Section 5, tracing its consequences for every descriptive fact the hidden-risk reading relies on.

Table 1 reports the descriptive half of the result. The conventional measure reproduces the familiar “hidden risk” moments: a mean SFA of +2.8% of GDP and a pooled

correlation with the primary balance of -0.77. The correct measure overturns both: the mean is -2.4% and the correlation is -0.13. The correct measure is the one Eurostat actually publishes, as the validation above confirms, while the wrong-sign construction departs from it. The wrong sign is only the cleanest illustration. Consider the primary-balance construction $\widehat{sfa} = \Delta b - p$ – the textbook decomposition stripped of its interest-growth term, and the variant an applied user is most likely to reach for. On the same panel it reproduces the same alarmist facts, if anything more strongly (mean +1.0% of GDP, correlation with the primary balance -0.82, impact local-projection gradient -0.38). It is loading the deficit, not the wrong sign specifically, that manufactures the hidden-risk reading. The same contamination that drives these descriptive moments also manufactures the instrument strength on which the literature’s causal reading depends. But because instrument relevance is an inferential object, we defer it to Section 7 (Table 6), where every first stage that looks usable on the conventional measure collapses on the correct one.

The measurement experiment, descriptive moments: same data and code, only the balance term changes.

Table 1. Descriptive Results from the Measurement Experiment

	Conventional sfa (deficit-contaminated)	Correct sfa (Eurostat-native)
Mean SFA (% GDP)	+2.83	-2.37
Std. dev. (% GDP)	9.89	6.95
Corr. with primary balance	-0.77	-0.13

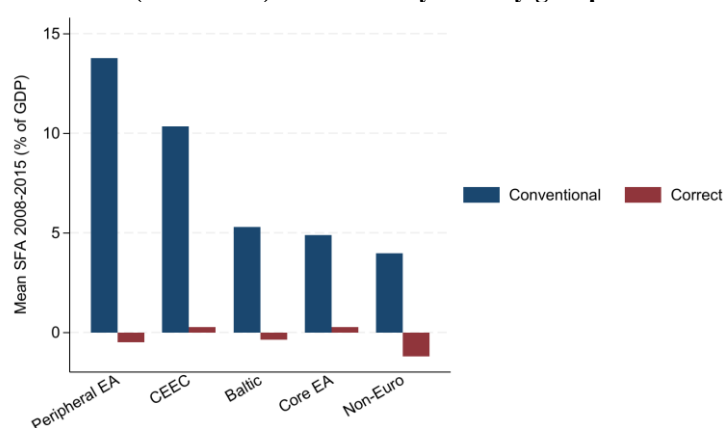
Notes: EU-27, 1995–2024. The conventional measure is $\widehat{sfa} = \Delta b - g$ (wrong sign); the correct measure is $sfa = \Delta b + g$, identical in construction to Eurostat’s published SFA. The three rows are descriptive moments on the common Eurostat sample (N=777), differing only in the balance term; the reconstruction’s cross-source validation against Eurostat’s published series is reported in Section 6, and the inferential half of the experiment (instrument relevance) in Table 6.

The most telling exhibit is the cross-group, crisis-window comparison in Table 2 and Figure 1. Read on the conventional measure, the peripheral euro area shows a 2008–2015 mean SFA of 14.5% of GDP and the central-European members 10.4% – the headline “crisis explosion” of below-the-line debt. Read on the correct measure, those same cells are -0.5% and +0.3%: essentially zero. What looked like a wall of hidden crisis-era debt accumulation was the crisis-era deficit, double-counted by the wrong-sign construction.

Table 2. Mean SFA (% of GDP) by country group: conventional vs. correct

Group	Full sample 1995–2024		Crisis window 2008–2015	
	Conventional	Correct	Conventional	Correct
Core euro area	1.85	-1.25	4.85	0.28
Peripheral euro area	4.49	-3.18	14.55	-0.50
Central/Eastern (CEEC)	5.16	-2.38	10.36	0.26
Non-euro EU	1.22	-3.29	4.23	-1.20
Baltic	2.65	-0.94	5.60	-0.37
EU-27	2.83	-2.37	7.42	-0.49

Notes: Group means of the conventional ($\Delta b - g$) and correct ($\Delta b + g$, Eurostat-native) SFA. The conventional crisis-window means reproduce the stylized facts reported in the “hidden risk” literature; the correct means are near zero.

Figure 1. Crisis-window (2008–2015) mean SFA by country group: conventional vs. correct.

Notes: The conventional measure shows the headline “crisis explosion” (peripheral euro area +14.5% of GDP) and the correct measure is near zero everywhere.

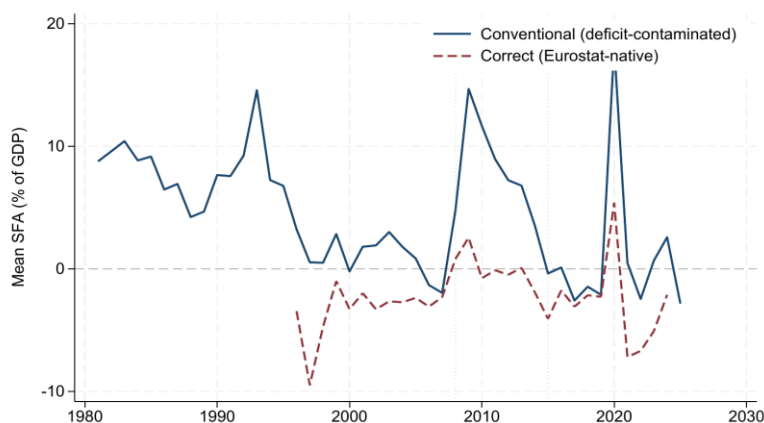
On the correct measure, the EU-27 SFA averages -2.4% of GDP over 1995–2024, with a standard deviation of 7.0 (Table 3). The negative average reflects net debt-reducing stock-flow operations over the sample, including privatisation receipts, drawdowns of financial assets, and valuation effects, rather than systematic below-the-line debt accumulation. By period, the Eurostat-native mean is -3.4% before 2008, -0.5% over the 2008–2015 crisis window, and about -3.0% thereafter. The correct SFA is therefore modestly higher (less negative) during the crisis than in adjacent periods, but it comes nowhere near the +14.5% “explosion” that the conventional measure shows. The crisis concentration is almost entirely an artefact of the deficit, which crises enlarge and which the wrong-sign construction absorbs.

Table 3. Descriptive statistics, EU-27, 1995–2024

Variable	Mean	Std. dev.	Min	Max	N
Correct SFA ($\Delta b + g$, % GDP)	-2.37	6.95	-146.30	18.30	777
Conventional SFA ($\Delta b - g$, % GDP)	2.83	9.89	-149.30	56.50	777
Primary balance (% GDP)	-0.67	3.15	-29.92	7.62	951
Overall balance (% GDP)	-2.92	3.59	-32.11	7.14	981
Gross debt (% GDP)	59.70	34.66	3.77	213.15	938
GDP growth (%)	2.41	3.66	-14.69	24.62	1120
Inflation (%)	9.64	60.64	-1.68	1518.53	1141

Notes: Eurostat EDP for gross debt and the overall balance; IMF WEO for the primary balance and macro controls. The correct-SFA minimum (-146.3) is Bulgaria in 1997 – a currency-board/nominal-GDP collapse (the debt ratio fell from 245% to 97% of GDP), a denominator effect rather than a below-the-line operation; excluding it shifts the mean from -2.4 to -2.2 and the standard deviation from 7.0 to 4.6, leaving the correlations and the predictive null unchanged (the conventional measure’s standard deviation likewise falls, from 9.9 to 8.3). Inflation reflects early-1990s transition hyperinflation and enters the regressions standardised. Both stock-flow rows are computed from the same Eurostat debt and balance series on the common sample (N=777), so they differ only in the balance term. The within-country primary-balance standard deviation, used for the equivalence bound of Section 7, is 2.9, against the pooled 3.15 reported above.

Figure 2 traces the EU-27 mean SFA over time on both measures. The conventional series rises sharply during 2008–2014. The correct series does not move with the crisis at all. The contamination is therefore concentrated exactly where the “hidden risk” reading is most rhetorically powerful (the crisis years) because that is when deficits, which the wrong-sign measure absorbs, were largest.

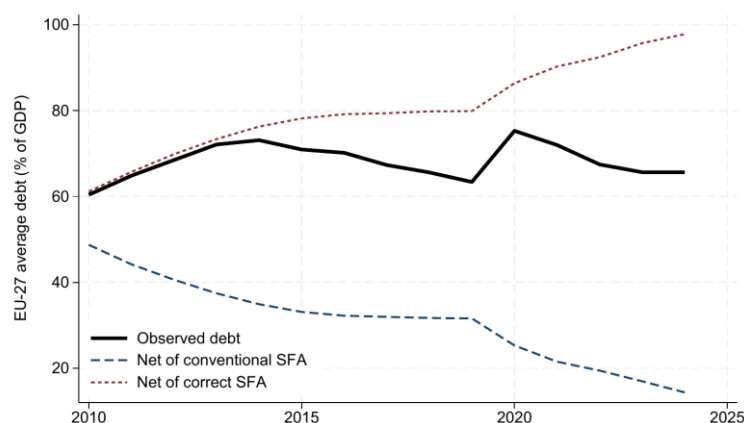
Figure 2. EU-27 mean stock-flow adjustment, 1995–2024: conventional vs. correct

Notes: conventional ($\Delta b - g$) vs. correct ($\Delta b + g$, Eurostat-native). Dotted lines mark the 2008–2015 crisis window. The conventional series spikes during the crisis; the correct series barely moves.

The same point can be put as a debt-accounting decomposition. Figure 3 subtracts the accumulated SFA from observed debt from 2010 onward and averages across the EU-27. Netting the conventional SFA mechanically erases roughly forty percentage points of debt by 2024, which is the basis for the “debt is much lower once you remove hidden adjustments” claim. Netting the correct SFA does the opposite: it raises the path above

observed debt, because correctly measured EU SFAs are mildly debt-reducing on net. The flagship counterfactual is an artefact of the sign error.

Figure 3. Mechanical debt path, EU-27 average: observed debt vs. debt net of accumulated SFA from 2010



Notes: Netting the conventional SFA spuriously removes ~40 points of debt; netting the correct SFA raises the path.

The composition of SFAs (net acquisitions of currency and deposits, debt securities, loans, and equity) is reported in Table 4 as descriptive background that the hidden-risk reading attributes to it. The substantive point is that the aggregate below-the-line margin in the EU is modest and does not display the size or the crisis concentration. Currency and deposits are the largest positive component in every group except the core euro area, where equity and investment-fund shares dominate. The equity component is positive only for the core euro area and negative for every other group, most markedly for the central-European members, reflecting different balance-sheet strategies. These compositional differences are real and unaffected by the measurement issue. Only the aggregate SFA, and its relationship to the primary balance, is distorted by the balance-term error.

Table 4. Net acquisition of financial assets by instrument and country group, 1997–2024 (% of GDP)

<i>Group</i>	<i>Currency & deposits</i>	<i>Debt securities</i>	<i>Loans</i>	<i>Equity & inv. funds</i>
Core euro area	0.25	0.24	0.17	0.40
Peripheral euro area	0.43	0.12	0.08	-0.03
Central/Eastern (CEEC)	0.67	0.02	0.24	-0.54
Non-euro EU	0.48	0.05	0.25	-0.20
Baltic	0.51	0.20	0.14	-0.15

Notes: Group means of net acquisitions of financial assets by instrument, 1997–2024 (% of GDP). Currency and deposits are the largest positive component in every group except the core euro area, where equity and investment-fund shares dominate; the equity component is positive only for the core euro area and negative elsewhere, most markedly for the central-European members.

7. Do correctly measured SFAs predict fiscal performance?

The motivating claim is that positive SFAs foreshadow weaker future primary balances. On the correct measure this association is absent: the pooled correlation between SFAs and primary balances is -0.13, and local-projection regressions of future primary balances on contemporaneous SFAs yield small and unstable coefficients.

Quantitatively, the correctly measured SFA has no economically meaningful predictive content for the primary balance. In our baseline local projection, conditioning on lagged debt and the contemporaneous output gap, growth, and inflation, the coefficient on the SFA is +0.14 on impact and lies between -0.04 and +0.06 at horizons one to five (Table 5, Panel B). The conventional measure, by contrast, reproduces the familiar adverse gradient (-0.35 on impact decaying to -0.04, Panel A). We read nothing into the sign of the impact coefficient: it is small and not robustly signed. Entering the cyclical controls with a one-period lag turns it mildly negative (-0.05 to -0.09), while country and two-way fixed effects leave the baseline essentially unchanged (+0.110 and +0.12; Panel C). What is robust is the magnitude: across every specification we examined, $|\hat{\beta}_h| \leq 0.14$ at all horizons, well below the contaminated gradient (which reaches -0.35).

An equivalence test makes this precise. Fix the smallest predictive effect we would treat as economically meaningful as one under which a one-standard-deviation stock-flow innovation (about seven points of GDP) moves the future primary balance by a full within-country standard deviation -2.9 points, so $|\beta_h| = 0.4$. A two-one-sided-tests (TOST) procedure rejects an effect that large at every horizon ($p < 0.001$), and the realized 90% confidence intervals are far tighter: within ± 0.22 (a width set by the not-robustly-signed impact coefficient) and within ± 0.13 from the first year onward. The medium-horizon cells of Panel C make the same point from the other direction. The only correct-SFA coefficient even marginally distinguishable from zero is the country-fixed-effects estimate at $h=2$ (-0.062). It implies that a one-standard-deviation shock moves the primary balance two years out by about 0.4 points, roughly a seventh of its within-country standard deviation, with the bad sign but no economic content, and gone by $h=4$. These bounds qualify the debt-spike result of Jaramillo et al. (2016), whose positive within-episode SFA association is estimated on a selected pooled sample of advanced and developing debt spikes rather than a general lagged EU panel, when it is extrapolated to the contemporary EU. And they accord with the integrated-balance-sheet evidence of

Seiferling (2013) that correctly measured residuals are smaller and less consequential than partial measures imply.

Table 5. Local projection of the primary balance at $t+h$ on the SFA

Horizon h	0	1	2	3	4	5
<i>Panel A: conventional SFA ($\widetilde{sfa} = \Delta b - g$)</i>						
$\hat{\beta}_h$	-0.348*** (0.043)	-0.245*** (0.036)	-0.190*** (0.034)	-0.122*** (0.034)	-0.061* (0.033)	-0.038 (0.028)
<i>Panel B: correct SFA ($sfa = \Delta b + g$, Eurostat-native)</i>						
$\hat{\beta}_h$	0.137*** (0.048)	0.059 (0.037)	-0.042 (0.037)	-0.027 (0.036)	-0.008 (0.050)	0.059 (0.041)
MDE _{0.8}	0.136	0.104	0.103	0.100	0.140	0.115
<i>Panel C: correct SFA, country fixed effects</i>						
$\hat{\beta}_h$	0.110** (0.043)	0.036 (0.039)	-0.062** (0.029)	-0.044 (0.028)	-0.031 (0.043)	0.026 (0.033)

Notes: OLS local projections of the primary balance at $t + h$ on the contemporaneous SFA, conditioning on lagged debt and the contemporaneous output gap, GDP growth, and inflation, EU-27, post-1997 (N between 447 and 522). Country-clustered standard errors in parentheses. MDE_{0.8} is the minimum effect detectable with 80% power at the 5% level; the 90% confidence intervals lie within ± 0.22 at every horizon. The conventional SFA reproduces the adverse “hidden-risk” gradient; the correct-SFA coefficients are economically negligible at every horizon and not robustly signed – two-way fixed effects give an impact coefficient of 0.124 (0.049) and entering the cyclical controls with a one-period lag gives -0.046 (0.041). No instrumental-variables estimate is reported for the correct SFA because its first stage collapses (Table 6) and its Anderson-Rubin confidence set is unbounded. *, **, and *** represent statistical significance at levels of 10%, 5% and 1%, respectively.

The causal reading the prior literature gives the SFA also fails to survive correction, but the right conclusion is sharper than “the effect is zero”. We instrument the SFA with the three sources of below-the-line variation of Section 5: the Bartik shift-share, bank-recapitalisation debt impacts, and ESA-2010 reclassification debt impacts. All three are strong against the conventional measure (first-stage F of roughly six, seven, and twenty-three) and collapse against the correct one (F below 1.5; Table 6), exactly the pattern Corollary 3 predicts when an instrument’s apparent strength is borrowed from the deficit. Two things follow, and they are different. First, the apparent causal identification in the hidden-risk reading was spurious: the instruments tracked the contaminated SFA only because they tracked the deficit it contains. On the conventional measure the over-identified shift-share returns a limited-information maximum-likelihood estimate of -0.63 with a Hansen J that does not reject ($p=0.20$), a clean but meaningless “causal” number. Second, for the genuine SFA no available instrument is informative. The Anderson-Rubin confidence set for the impact effect is unbounded, so the causal question is open, not answered in the negative. The affirmative evidence that the correct SFA carries no large predictive content is therefore the powered equivalence bound of Table 5, which uses no instrument. Here the instruments serve only as a diagnostic of contamination, not as a source of a causal estimate.

Table 6. The measurement experiment, instrument relevance: first stages collapse on the correctly measured SFA

	Conventional \widehat{sfa} (deficit-contaminated)	Correct sfa (Eurostat-native)
Shift-share (Bartik), first-stage effective F	6.0	0.3
Bank-recapitalisation, first-stage F	7.1	1.2
ESA-2010 reclassification, first-stage F	23.2	1.4
Shift-share LP-IV, impact $\hat{\beta}$ (LIML)	-0.63	(uninformative)

Notes: The inferential half of the measurement experiment of Table 1: same sample, same code, only the balance term in the SFA changes. First-stage effective F statistics (Montiel Olea and Pflueger, 2013) for the three instruments of Section 5 – a Bartik shift-share from pre-1997 asset-class composition, Eurostat-verified bank-recapitalisation debt impacts, and ESA-2010 reclassification debt impacts – together with the impact local-projection instrumental-variables (LP-IV) elasticity for the shift-share instrument. On the conventional (deficit-contaminated) measure all three instruments appear usable – the over-identified four-component shift-share gives a limited-information maximum-likelihood estimate of -0.63 with a Hansen J that does not reject ($p=0.20$); on the correct measure every first stage collapses to F below 1.5, the pattern Corollary 3 predicts for an instrument whose apparent strength is borrowed from the deficit, and the impact effect is not interpretable – its Anderson–Rubin confidence set is unbounded.

The reason is instructive. The instruments derive their apparent strength from the deficit component that the wrong-sign construction injects, because financial interventions coincide with deficits. But the large EU interventions themselves were predominantly recorded *above* the line, in the deficit, not in the SFA. Ireland in 2010 is the clean illustration: a +27.5 pp. reclassification of bank-support vehicles coincided with a net-lending figure of -32% of GDP, so the correctly measured SFA that year is only -7.7%. The events that make the instruments powerful were booked in the balance. Once the balance is netted correctly, the genuine SFA residual is small and idiosyncratic, and none of the available instruments predicts it. The causal design cannot be implemented on the correctly measured object with the data at hand.

7.1. Robustness

The null is stable along the dimensions a referee would check. The by-period pattern leaves no window with a large positive crisis mean: the correct EU-27 SFA averages -3.4% before 2008 and -0.5% across the 2008–2015 crisis window and stays around -3% in the years after. Nor does the single extreme observation flag in Table 3 drive anything. That observation of Bulgaria in 1997, where the currency-board stabilisation collapsed nominal GDP and so the debt ratio fell from 245% to 97%, a denominator effect rather than a below-the-line operation, inflates the dispersion but not the result. Dropping it moves the sample mean from -2.4 to -2.2 and the standard deviation from 7.0 to 4.6, while the SFA-balance correlations (-0.13 for the correct measure, and -0.77 for the conventional) and the local projections of Table 5 are unchanged. Winsorising at the first and ninety-ninth percentiles gives the same picture. The conclusion is invariant to

aggregation: the contamination $\widetilde{sfa} - sfa = -2g$ operates within each country-year through the balance term, so it is present identically whether country-years are weighted equally or by economic size. Finally, because the correct series is validated cell-by-cell against Eurostat's published SFA (median country-level correlation 0.98), the facts do not depend on our reconstruction.

The null also holds across subgroups (Table 8). Splitting the sample at the median debt ratio, the correctly measured SFA has no adverse medium-horizon relationship with the primary balance in either high-debt or low-debt countries; in particular the high-debt subsample, where the "hidden risk" reading should bite hardest, gives an impact coefficient of 0.083 and an $h=2$ coefficient of -0.031 (s.e. 0.033), both economically negligible. The low-debt subsample is estimated less precisely, and its larger impact point estimate (+0.220, s.e. 0.093) is wrong-signed for the hidden-risk story in any case. We therefore read the subgroup evidence as showing no adverse signal in any group, with precision that varies across the smaller cells, rather than a single uniform bound. The same picture holds splitting by membership of the peripheral euro area versus the rest of the EU. The contaminated narrative's strongest claim is that below-the-line accumulation is especially dangerous in already-stressed, high-debt economies, and that is precisely where the corrected data are most clearly null.

If there is a genuine below-the-line signal in the EU, it is concentrated, and the surveillance framework already knows where: the three member states it exempts from the zero-SFA assumption are the ones with the largest average correctly measured SFA, each driven by a documented institutional mechanism (Table 7). Finland and Luxembourg accumulate public pension-fund assets and so run debt-increasing SFAs averaging +0.9% and +2.1% of GDP. Greece, through deferred official-sector interest, runs a debt-reducing one averaging -4.2%. These cases are structural in their average level rather than their year-to-year autocorrelation, which is low across the EU (mean country-level first-order autocorrelation 0.16). Excluding the three leaves the predictive null intact (impact coefficient -0.13 under country fixed effects). The constructive reading is therefore not that SFAs are uninformative, but that the informative cases are few, identifiable, and already flagged by the framework.

Table 7. The genuine signal is concentrated: the structural cases the framework exempts

	Mean SFA (% GDP)	Institutional mechanism
Finland	+0.93	Public pension-fund asset accumulation
Luxembourg	+2.07	Public pension-fund asset accumulation
Greece	-4.23	Deferred official-sector (EFSF) interest

Notes: Country-level mean correctly measured SFA, 1996–2024, and the institutional mechanism behind it, for the three member states the European Commission exempts from the zero-SFA-beyond-T+2 assumption. These are the EU’s largest average SFAs in magnitude; they are structural in their average level – a recurring mechanism – not in autocorrelation, which is low throughout (the mean country-level first-order autocorrelation of the correct SFA across the EU-27 is 0.16). The predictive null of Table 5 is unchanged when the three are excluded (impact coefficient -0.13 under country fixed effects).

Table 8. The predictive null across subgroups (correct SFA)

	Impact ($h = 0$)	Medium ($h = 2$)
All countries	0.137*** (0.048)	-0.042 (0.037)
High debt (> median)	0.083 (0.054)	-0.031 (0.033)
Low debt (< median)	0.220** (0.093)	-0.029 (0.176)
Peripheral euro area	0.159* (0.089)	-0.048 (0.034)
Rest of the EU	0.156** (0.071)	0.058 (0.054)

Notes: OLS coefficient of the primary balance at $t+h$ on the correctly measured (Eurostat-native) SFA with lagged controls (debt, output gap, GDP growth, inflation), by subsample, country-clustered standard errors (in parenthesis). The high/low-debt split is at the median lagged debt ratio. No subsample shows an adverse medium-horizon relationship. *, **, and *** represent statistical significance at levels of 10%, 5% and 1%, respectively.

8. Implications for fiscal surveillance

The measurement result bears directly on how stock-flow adjustments are, and should be, used in fiscal surveillance, and it cuts in two directions that are worth keeping apart. The case for monitoring the SFA is untouched, and the case for treating it as a leading indicator of fiscal risk does not survive correct measurement.

The transparency case is, if anything, reinforced. Below-the-line operations have a long and well-documented history of being used to flatter headline deficits, including debt-decreasing one-offs, the timing of asset sales, the drawing of the public-sector perimeter (Milesi-Ferretti, 2004; von Hagen and Wolff, 2006; Koen and van den Noord, 2005; Reischmann, 2016; Alt et al., 2014). Nothing in our findings argues for watching the public balance sheet less closely. Debt-deficit reconciliation should be published, individual large operations (recapitalisations, sector reclassifications, guarantee calls) audited, and projected SFAs reported alongside projected deficits. Our claim is not that the SFA is uninformative. It is that the aggregate SFA, correctly measured, does not carry the predictive content the “hidden risk” reading attributes to it.

That distinction has become concrete under the European Union’s reformed fiscal framework (European Union, 2024). Regulation (EU) 2024/1263, in force since 30 April

2024 and operational from 2025, with the first medium-term fiscal-structural plans having been endorsed by the Council in January 2025 (European Central Bank, 2025), makes a single multi-year path for net primary expenditure the operational anchor of surveillance, over a standard four-year adjustment period extendable to seven (Articles 2, 5, 14). For a member state whose debt exceeds 60% of GDP or whose deficit exceeds 3%, the Commission derives a “reference trajectory” for that path, and the trajectory must place the projected debt ratio on a plausibly declining path by the end of the adjustment period (Articles 6 and 10), reinforced by a debt-sustainability safeguard and a deficit-resilience safeguard (Articles 7 and 8). The whole architecture therefore runs through the Commission’s debt-sustainability analysis (DSA), and the DSA must in turn take a stand on the SFA.

It does so in a way worth stating precisely, because it is exactly the choice our results speak to. In the deterministic debt projections, the SFA follows the Commission’s short-term forecast for two years and is then set to zero from the third year onward, “zero beyond the first two years”, for every member state but three (European Commission, 2024). The exceptions are revealing. Finland and Luxembourg are carved out because their public pension funds accumulate financial assets and so generate recurrent positive SFAs (accumulated assets of roughly 90% and 30% of GDP, respectively). Greece is carved out for the unrelated reason that deferred interest on its official-sector (European Financial Stability Facility, EFSF) loans generates recurrent negative SFAs: the accrued interest enlarges the deficit while the face-value debt rises by less, so debt grows by less than the balance implies and the residual is debt-reducing. The same three-country carve-out remains in force in the latest vintage (European Commission, 2025). The SFA is then silenced a second time in the stochastic layer. The fan-chart simulations that quantify the risk around the central path shock five variables, namely the primary balance, short- and long-term interest rates, nominal growth, and, for non-euro members, the exchange rate, but hold the SFA at its deterministic baseline. By construction, the SFA contributes nothing to the projected debt path beyond two years and nothing to the measured uncertainty around it.

This double silencing sits uneasily with the evidence on where debt projections actually go wrong. Bachleitner and Prammer (2024) find that the SFA is one of the two main drivers of EU debt-forecast errors, alongside misprojected GDP growth, and a driver the literature had largely neglected. The framework thus assumes away, and assigns no risk to, precisely the component that empirically moves the debt ratio away from forecast.

The magnitude of the assumption is not trivial. Were correctly measured SFAs to continue at their historical EU-27 average of -2.4% of GDP a year, zeroing them from the third projection year would understate the debt reduction over a four-year adjustment plan by about ten percentage points of GDP, and by roughly fourteen for a high-debt member state, whose correctly measured SFA averages -3.4%, a wedge of the same order as the consolidation the plans are built to deliver. The direction here is conservative (the assumption over-projects debt), but the point is general: zeroing a correctly measured residual of this size is a substantive modelling choice, not a neutral simplification. The natural inference would then be to build an SFA-based alarm into surveillance, and here our result earns its keep.

On the contaminated construction, the SFA looks like a large, adverse, persistent force, and an alarm calibrated to it would fire constantly and spuriously. On the correct construction, the aggregate EU SFA is modest, mildly debt-reducing on net, and does not predict deterioration, so an alarm calibrated to it would rarely fire and should not. The framework's instinct to discipline the SFA is sound. The hazard is disciplining, or being alarmed by, a mismeasured one.

The contrast with the International Monetary Fund's practice is instructive, because the Fund confronts the same residual and treats it as a signal of data quality rather than a forecast of distress. Its Sovereign Risk and Debt Sustainability Framework includes realism tools that flag a projected SFA which is large by a country's own historical experience (IMF, 2022); but the flag fires on the optimistic side, where a large assumed SFA would flatter the debt path, and it triggers a demand to explain or revise rather than a verdict of impending trouble. The lineage is older still: the 2013 market-access DSA guidance instructed staff to scrutinise the size of the debt-stock residual and named definitional and measurement mismatch, general-government debt set against a narrower balance, as a leading reason residuals appear at all (IMF, 2013). "Explain or revise," applied to a residual read as a diagnostic of measurement, is close to the right posture. Mechanical zeroing forgoes it.

The operational recommendation that follows is modest, and it builds on functions that already exist rather than inventing a mandate. National independent fiscal institutions already endorse or assess the official macroeconomic forecasts underpinning national budgets, and a substantial share of them conduct their own debt-sustainability analysis (Network of EU IFIs, 2021); the European Fiscal Board evaluates the framework's implementation and has publicly criticised the opacity of the Commission's DSA-based

reference trajectories (European Fiscal Board, 2025). The corrective these bodies are well placed to supply is measurement discipline: any SFA-based risk metric should be benchmarked against Eurostat's officially measured series, and projected SFA paths reported alongside projected deficits and debt, decomposed into their main balance-sheet components. Where a member state exhibits a genuinely persistent, correctly measured SFA (as Finland, Luxembourg, and Greece already do, and as the framework itself acknowledges by exception), that pattern can be flagged and stress-tested on its own terms. What the evidence here cautions against is the symmetric errors of inferring risk from a contaminated correlation and of assigning a mismeasured residual a blanket value of zero.

The broader methodological lesson generalises well beyond the EU. Because the SFA is a residual defined by an accounting identity, conclusions about it are unusually sensitive to how it is built: a wrong sign or the wrong balance does not merely add noise. It injects the deficit and reverses the central stylized fact. The integrated-balance-sheet evidence teaches the same lesson from the other side: once the government's accounts are measured completely and consistently, the stock-flow residual shrinks and loses its apparent link to fiscal opacity (Seiferling, 2013). The phenomenon was always smaller than the partial, mismeasured series implied. Applied work on stock-flow adjustments should therefore report the exact construction it uses and validate it against the official series before drawing inferences. That is a small price for avoiding a sign error that can flip a literature's central finding.

9. Conclusion

Stock-flow adjustments are a real and consequential feature of public-debt dynamics, but the recent reading of recurrent EU SFAs as "hidden fiscal risk" does not survive correct measurement. The reading rests on a residual that, as conventionally constructed, contains the very balance it is regressed against. Build the SFA from the right balance term, and the deficit no longer leaks into the measure: the evidence then reverses. The large average SFA, the crisis-era spike, and the negative SFA-sustainability gradient are not findings about fiscal behaviour. They are mechanical properties of an accounting semi-identity: subtract a deficit, correlate with it, and this is what you get. Measured the way Eurostat actually publishes it, the EU SFA is modest, shows no crisis spike precisely where the conventional narrative says it explodes, does not predict future primary balances, and is not identified by the instruments that appeared to give the causal claim

its bite. None of this argues for less transparency or scrutiny of the public balance sheet. In the EU aggregate, however, it is not the reservoir of concealed risk that a mismeasured residual suggests.

The case is made on three levels that reinforce one another. First, a mechanical result shows analytically why regressing a balance on a residual that embeds it yields a spuriously signed coefficient, with a horizon-decaying profile that mimics a genuine but fading economic effect (Proposition 1 and its corollaries). Second, a measurement experiment isolates the contamination empirically: holding the sample, the controls, and every line of code fixed and changing only the balance term, each familiar “hidden risk” fact flips. This is a design that invokes no exclusion restriction and leaves nothing else to which the reversal could be attributed. Finally, a quantified, fixed-effects-robust null bounds the medium-term predictive content of the correctly measured SFA well below any policy-relevant margin: we can rule out a one-standard-deviation stock-flow shock moving the future primary balance by more than about half of its own within-country standard deviation. Measurement, mechanism, and magnitude point the same way.

Three caveats bound the claim. The result concerns the EU-27 aggregate relationship: individual episodes (such as bank rescues or large reclassifications) do move debt materially and remain a transparency concern, and our null is silent on country-specific or tail risk. The corrected SFA is itself a residual and inherits the measurement limits of the underlying debt and balance series, including the occasional one-off valuation or reclassification spike. And because no instrument survives correction, we identify associations and bound effect sizes rather than estimate a structural parameter. What we can bound, however, is exactly what matters: no adverse effect of the magnitude the contaminated measure implies survives correct measurement.

The moral travels well beyond stock-flow adjustments. Public finance is full of residuals defined by accounting identities, such as the stock-flow adjustment, the debt-stabilising primary gap, and the unexplained component of debt changes. Whenever such a residual is itself a linear function of the aggregate it is regressed on, OLS reproduces the identity before it reveals any behaviour. We generalise this result (Proposition 2) and show it twice on EU data: on the SFA, and on the debt-stabilising primary gap, where the natural test of whether fiscal effort reduces debt returns a spurious one-for-one slope that is merely the budget constraint. The discipline we urge is unglamorous and decisive: state the construction explicitly, validate it against the official series, and ask whether a reported correlation could have been produced by the algebra alone. For surveillance, the

same discipline counsels benchmarking any SFA-based risk metric against the officially measured series rather than zeroing, or sounding an alarm over, a residual no one has checked. It also points to where the genuine action lies. That is not in an aggregate the algebra inflates, but in the few countries, such as Finland, Luxembourg, and Greece, whose correctly measured SFAs really are persistent, and whose structural sources the framework already singles out by exception. Applied to the EU stock-flow adjustment, that discipline turns a striking “hidden risk” into a cautionary tale about measurement, and returns the below-the-line margin to its proper place: an object to be monitored transparently, not a hidden reservoir of fiscal danger.

References

1. Adão, R., Kolesár, M., & Morales, E. (2019). Shift-Share Designs: Theory and Inference. *Quarterly Journal of Economics*, 134(4), 1949–2010. <https://doi.org/10.1093/qje/qjz025>
2. Afonso, A., & Alves, J. (2019). Short and long-term interest rate risk: The sovereign balance-sheet nexus. *Finance Research Letters*, 31, 294-299. <https://doi.org/10.1016/j.frl.2018.11.014>
3. Afonso, A., & Jalles, J. T. (2020). Stock flow adjustments in sovereign debt dynamics: The role of fiscal frameworks. *International Review of Economics & Finance*, 65, 1-16. <https://doi.org/10.1016/j.iref.2019.09.007>
4. Alt, J. E., Lassen, D. D., & Wehner, J. (2014). It Isn't Just about Greece: Domestic Politics, Transparency and Fiscal Gimmickry in Europe. *British Journal of Political Science*, 44(4), 707-716. <https://doi.org/10.1017/S0007123414000064>
5. Andrian, L. G., Rodríguez, C. M., & Valencia, O. M. (2025). Surges in the shadows: Stock-flow adjustments and public debt spikes. *Emerging Markets Review*, 67, 101306. <https://doi.org/10.1016/j.ememar.2025.101306>
6. Bachleitner, A., & Prammer, D. (2024). Don't blame the government!? An assessment of debt forecast errors with a view to the EU Economic Governance Review. *European Journal of Political Economy*, 82, 102524. <https://doi.org/10.1016/j.ejpoleco.2024.102524>
7. Bohn, H. (1998). The Behavior of U.S. Public Debt and Deficits. *Quarterly Journal of Economics*, 113(3), 949-963. <https://doi.org/10.1162/003355398555793>

8. Bohn, H. (2007). Are stationarity and cointegration restrictions really necessary for the intertemporal budget constraint? *Journal of Monetary Economics*, 54(7), 1837-1847. <https://doi.org/10.1016/j.jmoneco.2006.12.012>
9. Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *Review of Economic Studies*, 89(1), 181-213. <https://doi.org/10.1093/restud/rdab030>
10. Borusyak, K., & Hull, P. (2025). A Practical Guide to Shift-Share Instruments. *Journal of Economic Perspectives*, 39(1), 181-204. <https://doi.org/10.1257/jep.20231370>
11. Buti, M., Martins, J. N., & Turrini, A. (2007). From Deficits to Debt and Back: Political Incentives under Numerical Fiscal Rules. *CESifo Economic Studies*, 53(1), 115-152. <https://doi.org/10.1093/cesifo/ifm003>
12. Campos, C. F. S., Jaimovich, D., & Panizza, U. (2006). The unexplained part of public debt. *Emerging Markets Review*, 7(3), 228-243. <https://doi.org/10.1016/j.ememar.2006.06.001>
13. Casalin, F., Cerniglia, F., & Dia, E. (2023). Stock-flow adjustments, public debt management and interest costs. *Economic Modelling*, 129, 106531. <https://doi.org/10.1016/j.econmod.2023.106531>
14. D'Erasmus, P., Mendoza, E. G., & Zhang, J. (2016). What is a sustainable public debt? In *Handbook of Macroeconomics* (Vol. 2, pp. 2493-2597). Elsevier. <https://doi.org/10.1016/bs.hesmac.2016.03.013>
15. Escolano, J. (2010). *A practical guide to public debt dynamics, fiscal sustainability, and cyclical adjustment of budgetary aggregates* (IMF Technical Notes and Manuals No. 2010/02). International Monetary Fund. <https://doi.org/10.5089/9781462396955.005>
16. European Central Bank. (2025). Medium-term fiscal-structural plans under the revised Stability and Growth Pact. *ECB Economic Bulletin*, 3/2025. Available at: https://www.ecb.europa.eu/press/economic-bulletin/articles/2025/html/ecb.ebart202503_01~1b1a292f83.en.html
17. European Commission. (2024). *Debt Sustainability Monitor 2023* (European Economy Institutional Paper No. 271). Directorate-General for Economic and Financial Affairs. Available at: https://economy-finance.ec.europa.eu/publications/debt-sustainability-monitor-2023_en
18. European Commission. (2025). *Debt Sustainability Monitor 2024* (European Economy Institutional Paper No. 306). Directorate-General for Economic and Financial

Affairs. Available at: https://economy-finance.ec.europa.eu/publications/debt-sustainability-monitor-2024_en

19. European Fiscal Board. (2025). *Annual report 2025*. European Commission. Available at: https://commission.europa.eu/publications/2025-annual-report-european-fiscal-board_en

20. European Union. (2024). Regulation (EU) 2024/1263 of the European Parliament and of the Council of 29 April 2024 on the effective coordination of economic policies and on multilateral budgetary surveillance, repealing Council Regulation (EC) No 1466/97. *Official Journal of the European Union*, L, 2024/1263.

21. Eyraud, L., Debrun, X., Hodge, A., Lledó, V. D., & Pattillo, C. (2018). *Second-generation fiscal rules: Balancing simplicity, flexibility, and enforceability* (IMF Staff Discussion Note No. 18/04). International Monetary Fund. <https://doi.org/10.5089/9781484350683.006>

22. Fazzari, S. M., Hubbard, R. G., & Petersen, B. C. (1988). Financing Constraints and Corporate Investment. *Brookings Papers on Economic Activity*, 1988(1), 141-206. <https://doi.org/10.2307/2534426>

23. Ghosh, A. R., Kim, J. I., Mendoza, E. G., Ostry, J. D., & Qureshi, M. S. (2013). Fiscal Fatigue, Fiscal Space and Debt Sustainability in Advanced Economies. *Economic Journal*, 123(566), F4-F30. <https://doi.org/10.1111/ecoj.12010>

24. Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8), 2586-2624. <https://doi.org/10.1257/aer.20181047>

25. International Monetary Fund. (2013). *Staff guidance note for public debt sustainability analysis in market-access countries* (IMF Policy Paper). International Monetary Fund. Available at: <https://www.imf.org/external/np/pp/eng/2013/050913.pdf>

26. International Monetary Fund. (2022). *Staff guidance note on the sovereign risk and debt sustainability framework for market access countries* (IMF Policy Paper No. 2022/039). International Monetary Fund. Available at: <https://www.imf.org/en/publications/policy-papers/issues/2022/08/08/staff-guidance-note-on-the-sovereign-risk-and-debt-sustainability-framework-for-market-521884>

27. Jaramillo, L., Mulas-Granados, C., & Kimani, E. (2016). *The blind side of public debt spikes* (IMF Working Paper No. 16/202). International Monetary Fund. <https://doi.org/10.5089/9781475545043.001>

28. Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1), 161-182.
<https://doi.org/10.1257/0002828053828518>
29. Koen, V., & van den Noord, P. (2005). *Fiscal gimmickry in Europe: One-off measures and creative accounting* (OECD Economics Department Working Paper No. 417). OECD Publishing. Available at:
https://www.oecd.org/content/dam/oecd/en/publications/reports/2005/02/fiscal-gimmickry-in-europe_g17a1685/237714513517.pdf
30. Kronmal, R. A. (1993). Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *Journal of the Royal Statistical Society: Series A*, 156(3), 379-392.
<https://doi.org/10.2307/2983064>
31. Lee, D., McCrary, J., Moreira, M., & Porter, J. (2022). Valid t-Ratio Inference for IV. *American Economic Review*, 112(10), 3260-3290.
<https://doi.org/10.1257/aer.20211063>
32. Mendoza, E. G., & Ostry, J. D. (2008). International evidence on fiscal solvency: Is fiscal policy “responsible”? *Journal of Monetary Economics*, 55(6), 1081-1093.
<https://doi.org/10.1016/j.jmoneco.2008.06.003>
33. Milesi-Ferretti, G. M. (2004). Good, bad or ugly? On the effects of fiscal rules with creative accounting. *Journal of Public Economics*, 88(1-2), 377-394.
[https://doi.org/10.1016/S0047-2727\(02\)00076-2](https://doi.org/10.1016/S0047-2727(02)00076-2)
34. Montiel Olea, J. L., & Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, 31(3), 358-369.
<https://doi.org/10.1080/00401706.2013.806694>
35. Network of EU Independent Fiscal Institutions. (2021). *The role of independent fiscal institutions in assessing the sustainability of high public debt in the post-Covid era*. Network of EU IFIs. Available at: <https://www.euifis.eu/publications/14>
36. Reischmann, M. (2016). Creative accounting and electoral motives: Evidence from OECD countries. *Journal of Comparative Economics*, 44(2), 243-257.
<https://doi.org/10.1016/j.jce.2015.07.001>
37. Sánchez-Vidal, J. (2025). *An empirical assessment of the accounting semi-identity problem’s pervasiveness and severity* (arXiv:2511.08591). arXiv. Available at: <https://arxiv.org/pdf/2511.08591>

38. Seiferling, M. (2013). *Stock-Flow Adjustments, Government's Integrated Balance Sheet and Fiscal Transparency* (IMF Working Paper No. 13/63). International Monetary Fund. <https://doi.org/10.5089/9781475537642.001>

39. von Hagen, J., & Wolff, G. (2006). What do deficits tell us about debt? Empirical evidence on creative accounting with fiscal rules in the EU. *Journal of Banking & Finance*, 30(12), 3259-3279. <https://doi.org/10.1016/j.jbankfin.2006.05.011>

40. Weber, A. (2012). *Stock-Flow Adjustments and Fiscal Transparency: A Cross-Country Comparison* (IMF Working Paper No. 12/39). International Monetary Fund. <https://doi.org/10.5089/9781463933821.001>

