REM WORKING PAPER SERIES

Untapped Tax Revenue and Sovereign Credit Ratings

António Afonso, José Alves, Kakoum Eudes Somda

REM Working Paper 0439-2026

November 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.

fct Fundação para a Ciência e a Tecnologia
U LISBOA | UNIVERSIDADE DE LISBOA
ISEG Lisbon School of Economics & Management Universidade de Lisboa
REM RESEARCH IN ECONOMICS AND MATHEMATICS

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/

Untapped Tax Revenue and Sovereign Credit Ratings

António Afonso[1,*] José Alves[1, †] Kakoum Eudes Somda[2, ‡]

October 2026

Abstract

Do sovereign credit ratings reflect how much of their fiscal potential governments leave untapped? We estimate tax potential with a stochastic frontier and measure the tax gap as the share of this potential that remains uncollected, which averages 35% in our sample, 30% in advanced economies and 38% in emerging and developing economies. Using panel local projections for 133 countries over 1990–2021, we find that a one-standard-deviation increase in the tax gap, about 4.1 percentage points of potential, is followed by a gradual decline in the sovereign rating, which peaks after about four years at 0.13 to 0.20 notches, depending on the estimator. Such an increase raises the net probability of a downgrade by about 4 percentage points within three to four years. Ratings follow the tax gap with a lag and return to their initial level only once the gap has closed, about a decade after the shock. The decline is not driven by falling revenue. A fall in revenue has only a short-lived effect, whereas a widening of the gap at a given level of revenue, which signals a tax base that expands without being taxed, has twice the baseline effect. Untapped fiscal potential is thus read as a sign of potential fiscal weakness rather than as fiscal headroom. The response is present in both advanced and developing economies, peaking earlier in the latter. It is stronger in countries that rely on taxation rather than natural resource revenue, outside IMF programs, and when uncertainty is high.

Keywords: Sovereign credit ratings; Tax gap; Revenue mobilization; Fiscal capacity; Local projections

JEL Classification: E62, G24, H20, H63, O23

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.

[1] 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.
[2] ERUDITE, University of Paris-Est Créteil (UPEC), France.
[*]Email: aafonso@iseg.ulisboa.pt
[†]Email: jalves@iseg.ulisboa.pt
[‡]Email: kakoum-eudes.somda@u-pec.fr

1 Introduction

Domestic revenue mobilization lies at the heart of governments' ability to finance public goods and services and preserve the sustainability of public finances. This challenge is particularly important in emerging and developing economies, where financing needs remain high, fiscal space is often limited, and access to external financing is costly (Besley & Persson, 2014; Prichard, 2016). These pressures have intensified in recent years. Global public debt reached nearly 94% of GDP in 2025 and could exceed 100% by 2029 (IMF, 2026). In developing countries, net interest payments on public debt reached $921 billion in 2024, a 10% increase in one year, and 61 countries now devote at least 10% of their government revenue to interest payments (UNCTAD, 2025). At the same time, official development assistance fell by 23.1% in real terms in 2025 (OECD, 2026). Governments' ability to mobilize their own revenue is therefore becoming central, and sovereign ratings shape the cost at which they can finance themselves (Afonso, Alves, Gogas, & Papadimitriou, 2026).

Yet the observed level of revenue reveals only part of a country's fiscal position. What also matters is the share of fiscal potential that remains uncollected. This distinction is particularly important because two countries collecting the same share of GDP in taxes may be in very different positions. One may be mobilizing nearly all of a relatively limited fiscal potential, while the other may remain far below a much larger potential. Part of the revenue-mobilization literature captures this difference through tax effort, defined as the share of fiscal potential that is effectively mobilized (Fenochietto & Pessino, 2013; Mawejje & Sebudde, 2019). The resulting tax gap, in turn, represents the share of estimated fiscal potential that remains uncollected (Afonso, Alves, & Somda, 2026). It therefore places at the center of the analysis the amount of revenue that a government could potentially mobilize, given the structural characteristics of its economy, but does not collect. An increase in the tax gap thus reflects a lower utilization of available fiscal potential. In our sample, the tax gap averages about 35% of estimated fiscal potential, 30% in advanced economies and 38% in emerging and developing economies.

The existence of untapped fiscal potential gives rise to two competing interpretations. On the one hand, a large tax gap may represent a form of fiscal headroom: additional revenue remains potentially mobilizable if financing needs increase, providing the government with room for adjustment and potentially pushing its fiscal limit further away (Bi, 2012). On the other hand, it may reveal fiscal weakness: despite the existence of a mobilizable tax base, the government is unable or unwilling to convert it into actual revenue. These two interpretations have opposite implications for sovereign risk, particularly in an international environment characterized by tighter financing conditions.

This paper therefore asks the following question: when a country's tax gap widens, does its sovereign rating improve because the government retains additional revenue-raising capacity, or does it deteriorate because the larger gap reveals a weakness in revenue mobilization? We further examine whether the information contained in the tax gap is distinct from that already captured by observed revenue.

To address this question, we first estimate fiscal revenue potential using a stochastic frontier approach (Aigner et al., 1977; Battese & Coelli, 1988; Fenochietto & Pessino, 2013), based on tax revenue excluding social security contributions for a panel of 133 countries over 1990–2021. We then combine these estimates with long-term foreign-currency sovereign ratings from the World Bank's Fiscal Space database (Kose et al., 2022). This database harmonizes sovereign credit ratings from the three major rating agencies (Moody's, Standard & Poor's, and Fitch) on a common scale. We then examine the dynamic evolution of sovereign ratings in the years following an increase in the tax gap, conditional on the main macroeconomic and fiscal fundamentals, using panel local

projections (Jordà, 2005; Jordà & Taylor, 2025).

Four main findings emerge. First, a one-standard-deviation increase in the tax gap, about 4.1 percentage points of potential, is followed by a gradual deterioration in the sovereign rating. The rating does not react in the year of the shock and declines by 0.13 to 0.20 notches after four years, the lower bound corresponding to our baseline estimates and the upper bound to an estimator that corrects for the small-sample bias of fixed-effects local projections. Even the lower bound is comparable to the effect associated with a public debt level about 12 percentage points of GDP higher. This deterioration does not merely reflect marginal adjustments of an average index. For a given country, the net probability of experiencing a downgrade rather than an upgrade of its composite rating increases by about 4 percentage points within three to four years of the shock.

Second, the tax gap contains information that observed revenue does not reveal. A fall in tax revenue is followed by only a modest and short-lived decline in the rating. By contrast, when the change in revenue is held constant, a widening of the tax gap can only reflect fiscal potential that grows without being collected, as the economy broadens its tax base through trade openness, urbanization or structural change while revenue fails to follow. In this case, the rating declines by about 0.25 notches, twice the baseline effect. This result discriminates directly between the two interpretations of untapped potential. If it were perceived as fiscal headroom, the effect should be positive, and if agencies only paid attention to observed revenue, it should be zero. What rating agencies appear to penalize is therefore not low revenue as such, but the failure to convert an expanding tax base into public resources.

Third, ratings follow the tax gap with a lag, both when it widens and when it closes. The rating does not react in the year of the shock, and most of the deterioration materializes between the second and the fourth year, when it reaches its peak. The tax gap itself is persistent but not permanent, as it takes about eight years to be fully reabsorbed. The rating returns to its initial level only about two years later, once the gap has closed. This pattern is consistent with agencies waiting for a deterioration in revenue mobilization to be confirmed before downgrading, and for its correction to be confirmed before restoring the rating. Fourth, the response is present across country groups, but its timing and strength vary. In emerging and developing economies, ratings respond faster, with a decline that peaks two years after the shock, whereas in advanced economies the decline builds up more slowly and is larger and more persistent. The relationship is also more pronounced in countries that rely on taxation rather than natural resource revenue, outside IMF programs, and when uncertainty is high. Taken together, these findings are more consistent with interpreting untapped fiscal potential as revealed weakness than as fiscal headroom.

This paper makes three contributions. First, it extends the literature on the determinants of sovereign ratings (Cantor & Packer, 1996; Afonso et al., 2011) by distinguishing two dimensions that conventional fiscal indicators tend to conflate: the revenue actually collected and the share of fiscal potential that remains uncollected. We show that the tax gap contains information that is distinct from observed revenue. We also document the timing of this relationship: ratings adjust gradually, several years after the tax gap widens, thereby extending the literature on the timing of sovereign rating adjustments (El-Shagi & von Schweinitz, 2022).

Second, the paper connects the literature on revenue mobilization, which is central to development economics, with the literature on sovereign risk. Existing work has focused primarily on explaining why some countries leave a larger share of their fiscal potential uncollected than others (Gupta, 2007; Fenochietto & Pessino, 2013; Besley & Persson, 2014; Prichard, 2016; Mawejje & Sebudde, 2019). To our knowledge, no study has related untapped fiscal potential to the assessment of sovereign risk. We study one consequence of this gap: its informational content for sovereign risk assessment. This is particularly relevant for emerging and developing economies, where domestic

revenue mobilization and access to external financing constitute closely interconnected constraints. Our results also show that resource-rich countries receive no rating signal for their tax effort, which weakens the market incentives to diversify their revenue base.

Third, we contribute to the literature on fiscal space and fiscal limits (Bi, 2012; Aizenman et al., 2013; Kose et al., 2022). In principle, untapped fiscal potential may constitute a reserve that can be mobilized when needed. Our results suggest that sovereign ratings do not primarily reflect it in this way: an increase in the tax gap is followed by a deterioration in the rating, and the response is strongest precisely when fiscal potential grows without being collected. Untapped fiscal potential therefore appears to be more consistent with a signal of fiscal weakness than with a readily available fiscal reserve.

The remainder of the paper is organized as follows. Section 2 presents the conceptual framework and the literature on fiscal capacity, the tax gap, and sovereign creditworthiness. Section 3 describes the data and the construction of the tax gap. Section 4 presents the empirical approach. Section 5 reports the main results and their dynamics. Section 6 presents robustness checks, Section 7 examines how the relationship varies across countries and circumstances, and Section 8 concludes.

2 Background and Hypotheses

2.1 Fiscal Capacity and the Tax Gap

Fiscal capacity refers to the ability of a state to mobilize resources from its economic base on a sustained basis (Besley & Persson, 2009, 2011; Dincecco & Prado, 2012; Besley & Persson, 2014; Dincecco & Katz, 2016; Albers et al., 2023). The tax revenue-to-GDP ratio, commonly used as a proxy for this capacity, captures it only partially. Two countries with the same revenue ratio may in fact be in very different positions: the first may be mobilizing a large share of a relatively limited potential, while the second may be collecting only a small fraction of a much larger one. Part of the literature seeks precisely to capture this distinction through tax effort, that is, the share of fiscal potential that is effectively mobilized (Cyan et al., 2013; Fenochietto & Pessino, 2013; Mawejje & Sebudde, 2019; McNabb et al., 2021; Benitez et al., 2023). This potential can be estimated using stochastic frontier models, which distinguish systematic shortfalls from the frontier from random fluctuations affecting observed revenue (Aigner et al., 1977; Battese & Coelli, 1988). This distinction is particularly relevant in a macroeconomic setting, where tax revenue may be affected by transitory factors or measurement errors that do not necessarily reflect a persistent underutilization of fiscal potential.

In this analysis, we focus on a complementary measure, the tax gap, defined as the share of estimated fiscal potential that remains uncollected (Afonso, Alves, & Somda, 2026). An increase in the tax gap therefore reflects a lower utilization of available fiscal potential, whereas a decrease indicates that actual revenue is moving closer to the estimated frontier. This formulation is particularly well suited to our question, which concerns how the underutilization of fiscal potential is reflected in assessments of sovereign creditworthiness. It should be noted, however, that the tax gap does not reflect solely the efficiency of the tax administration. It may also result from tax policy choices regarding rates, exemptions, or the structure of the tax system, as well as from differences in the state's ability to enforce tax legislation. It is therefore best understood as a summary measure of the underutilization of fiscal potential rather than as a pure indicator of administrative performance. This distinction is nonetheless particularly useful for assessing fiscal soundness, as it makes it possible to examine not only how much a state collects, but also the

extent to which it mobilizes the resources that the characteristics of its economy appear to make available.

2.2 Sovereign Ratings and Fiscal Fundamentals

The literature identifies several factors that shape sovereign ratings and, through them, the conditions under which governments borrow on financial markets (Cantor & Packer, 1996; Afonso et al., 2011; Thornton & Vasilakis, 2018; Balima, 2020; Afonso, Alves, Gogas, & Papadimitriou, 2026). Since Cantor and Packer (1996), numerous studies have shown that sovereign ratings are closely associated with per capita income, growth, inflation, institutional quality and, of course, the state of public finances. In this vein, Afonso et al. (2011) show that changes in GDP per capita, growth, government debt and the fiscal balance explain short-run rating adjustments, whereas government effectiveness and default history matter more in the long run. El-Shagi and von Schweinitz (2022) also emphasize the importance of timing: they show that rating adjustments do not always coincide immediately with changes in fundamentals, as ratings may adjust gradually. This motivates studying the trajectory of ratings beyond their contemporaneous relationship with fiscal variables.

More recently, Blanchard et al. (2026) compare ratings with what a simple sustainability model based on debt and projected primary balances would predict, and show that ratings place much more weight on the debt stock than on expected primary balances. Ratings also respond little to the interest rate–growth differential, while country-specific effects play an important role: for comparable debt and balances, ratings can differ substantially across countries. In other words, ratings appear to focus more on the inherited state of public finances than on the government's capacity to restore them.

This finding directly raises the question of revenue. Future primary balances depend to a large extent on the revenue the government will be able to raise, and therefore on its capacity to mobilize its remaining fiscal potential. If ratings mainly reflect inherited debt, they may overlook this capacity, since it does not appear in observed revenue. Yet the importance of country effects suggests that ratings also incorporate persistent characteristics of states that cannot be reduced to their debt and balances; the way a government exploits its fiscal potential may be one of them. The literature, however, does not allow us to settle this question: it captures revenue through the amounts actually collected or the resulting balances, and therefore does not distinguish a government that collects little because its potential is limited from one that collects little despite having a substantially larger potential. An essential question thus remains open: do sovereign assessments account only for the amount of revenue actually mobilized, or do they also reflect available fiscal capacity and the extent to which the government manages to exploit it? It is this question that leads us to consider the tax gap as a distinct dimension of fiscal soundness.

2.3 The Tax Gap: Fiscal Headroom or Revealed Weakness?

The relationship between the tax gap and sovereign risk is, however, not clear-cut. Untapped fiscal potential can be interpreted in two opposite ways.

The first interpretation is that of fiscal headroom. In models of fiscal limits, a government's distance from its limit depends not only on its debt and spending, but also on the additional revenue it can raise (Bi, 2012). Measures of fiscal space rest on the same idea of a reserve that can be mobilized beyond current revenue (Aizenman et al., 2013; Kose et al., 2022). Other things equal, a country far from its revenue frontier would thus have greater room for adjustment than a country that already mobilizes nearly all of its potential.

The second interpretation leads to the opposite prediction. The existence of potential does not guarantee that the government can or will convert it into revenue, and debt repayment depends as much on the willingness as on the ability to pay (Eaton & Gersovitz, 1981). A gap between actual revenue and potential may reveal collection difficulties, weak administrative capacity, compliance problems, or policy choices that persistently limit resource mobilization. Untapped potential improves adjustment capacity only if it can be mobilized when constraints tighten. Otherwise, it undermines the credibility of future adjustment and instead signals higher deficits and debt accumulation.

Two competing hypotheses therefore emerge, and they differ not only in the sign of the relationship but also in its timing and in the contexts where it should be strongest.

Hypothesis 1 (fiscal headroom): A higher tax gap signals resources that can still be mobilized. Under this view, a government that collects well below its potential retains room to raise revenue if financing needs increase, which lowers the probability that it will have to default or to adjust abruptly. An increase in the gap should therefore not lead to a rating deterioration, and may even lead to an improvement. This effect should be strongest where markets have reason to believe that the untapped potential can actually be mobilized, for instance where institutions are strong or where an external anchor, such as an IMF program, makes future revenue measures more credible.

Hypothesis 2 (revealed weakness): A higher tax gap signals an inability or limited willingness to mobilize available resources. Under this view, the gap reveals weak administrative capacity, extensive exemptions or political resistance to taxation, and therefore lowers the credibility of future fiscal adjustment. An increase in the gap should be followed by a rating deterioration. Because agencies need time to distinguish a lasting shortfall from a transitory fluctuation, this deterioration should build up gradually and persist as long as the gap remains open. It should also be strongest where taxation is the main source of public finance, where corruption makes the gap more likely to reflect leakages and evasion, and where no external anchor offsets the loss of credibility.

The two hypotheses also make distinct predictions about the information content of the tax gap relative to observed revenue. A rise in the gap can result either from a fall in collected revenue or from a rise in fiscal potential that revenue fails to follow. If agencies only paid attention to observed revenue, the second case should leave ratings unaffected. Under fiscal headroom, it should improve them, since potential grows while debt-servicing capacity is unchanged. Under revealed weakness, by contrast, it should be penalized at least as strongly as the first, since it reveals a tax base that expands without being taxed. Comparing the response of ratings to the tax gap with and without holding revenue constant thus provides a direct test between the two interpretations, which we implement in Section 5.4. The sign of the relationship between the tax gap and sovereign ratings is therefore ambiguous a priori, but the two hypotheses can be distinguished empirically.

3 Data

3.1 Main Data Sources

The analysis covers a panel of 133 advanced and emerging and developing economies observed over 1990–2021. This broad coverage is essential for our question, since it is in emerging and developing economies that domestic revenue mobilization and access to external financing are most closely intertwined. The composition of the sample is determined by data availability.

Our dependent variable is the long-term foreign-currency sovereign rating, taken from the World Bank's Fiscal Space database (Kose et al., 2022), Spring 2026 version. For each country, the ratings assigned by the three main rating agencies (Moody's, Standard & Poor's and Fitch) are converted to a numerical scale ranging from 1 (C) to 21 (AAA) and then averaged across agencies (see Table 6 in the Appendix). This composite index has a twofold advantage: it covers a large number of developing countries and smooths out agency-specific differences. It is, however, highly persistent from one year to the next, as agencies favour the stability of their assessments, which calls for studying its response over several years.

Our variable of interest is the tax gap, that is, the share of estimated fiscal potential that remains uncollected. Its construction, detailed in Section 3.2, relies on the stochastic frontier developed in Afonso, Alves, and Somda (2026) and combines two types of information. Tax revenue excluding social contributions, expressed as a share of GDP, comes from the IMF World Revenue Longitudinal Database (Mansour et al., 2025), whose harmonization under the GFS framework ensures comparability across countries and over time; the UNU-WIDER Government Revenue Dataset (UNU-WIDER, 2025) serves as an alternative source in robustness tests. Fiscal potential, in turn, is conditioned on the structural characteristics that determine the size and taxability of the tax base, namely GDP growth, trade openness, the share of agriculture, natural resource rents and the urbanization rate, drawn from the World Development Indicators (Fenochietto & Pessino, 2013; Mawejje & Sebudde, 2019; McNabb et al., 2021). An augmented version of the frontier adds the level of development, population density and human capital from Penn World Table (Feenstra et al., 2015). In the empirical analysis, the annual change in the tax gap constitutes the shock under study.

To isolate the information specific to the tax gap, we account for the usual determinants of sovereign ratings (Cantor & Packer, 1996; Afonso et al., 2011). Macroeconomic fundamentals are captured by GDP per capita, inflation and GDP growth, from the World Development Indicators. The fiscal position is captured by public debt and the fiscal deficit, from the IMF Fiscal Monitor, as well as by the level of tax revenue. Finally, the institutional framework is taken into account through the presence of a fiscal rule, based on the IMF Fiscal Rules Dataset, and of an IMF program. The latter indicator, equal to one in years in which a country benefits from a Fund financing arrangement, combines the data of Vreeland (2007) for the earlier years with the IMF MONA database from 2002 onwards; it accounts for the fact that such programs often involve revenue mobilization measures and are closely monitored by rating agencies. The heterogeneity analyses further draw on indicators of countries' environment, including country-specific geopolitical risk (Caldara & Iacoviello, 2022), the World Uncertainty Index (Ahir et al., 2022) and the ICRG corruption score. Table 9 in the Appendix summarizes the definitions and sources of the main variables.

Table 1 reports descriptive statistics for the estimation sample, in which nearly two-thirds of the observations relate to emerging and developing economies (96 of the 133 countries). They first reveal a considerable rating gap between the two groups. The average rating of advanced economies (18.2) corresponds to the AA category, whereas that of emerging and developing economies (10.2) lies around BB, that is, in the speculative grade: below the BBB− threshold, debt is considered to carry a high risk of default, which restricts its holding by many institutional investors and raises borrowing costs. Ratings nevertheless change only slowly: their average annual change is close to zero and its standard deviation does not exceed two-thirds of a notch, an inertia that supports the choice of tracing their response over several years.

These rating differences go hand in hand with marked differences in revenue mobilization. Emerging and developing economies collect on average 15.7% of GDP, against 24.9% in advanced

economies, but their lower revenue does not merely reflect a narrower tax base: they also leave a larger share of their potential uncollected, 38% against 30%, for an average of 35% over the whole sample. Their tax gap is also more dispersed (standard deviation of 0.18 against 0.12) and twice as volatile from one year to the next (standard deviation of the annual change of 0.04 against 0.02). The variation in the tax gap on which identification relies thus comes to a large extent from these economies, which gives our question particular relevance for them.

Finally, fiscal fundamentals alone cannot account for the differences in ratings. Emerging and developing economies are on average less indebted (46% of GDP against 63%) and run comparable deficits, of around 2% of GDP, even though they are rated considerably lower. They differ instead by lower income, higher inflation, less widespread fiscal rules (41% of observations against 83%) and more frequent recourse to IMF programs (27% against 6%). This suggests that, for a given level of debt, rating agencies take into account other dimensions of fiscal soundness, among which the capacity to mobilize revenue appears as a natural candidate, which the analysis that follows allows us to examine.

Table 1: Descriptive statistics

	All countries		Advanced economies		Emerging & developing	
	Mean	SD	Mean	SD	Mean	SD
Sovereign ratings						
Rating index (1–21)	13.14	5.22	18.15	3.37	10.18	3.60
Δ Rating index	-0.02	0.64	-0.01	0.66	-0.02	0.62
Revenue mobilization						
Tax gap	0.35	0.17	0.30	0.12	0.38	0.18
Δ Tax gap	0.00	0.04	0.00	0.02	0.00	0.04
Tax revenue (% of GDP)	19.10	8.27	24.87	7.06	15.69	6.94
Controls						
Log GDP per capita	9.16	1.31	10.39	0.56	8.42	1.06
GDP growth (%)	3.38	4.24	2.53	3.53	3.89	4.53
Inflation, log(1+π)	0.04	0.05	0.02	0.02	0.05	0.06
Public debt (% of GDP)	52.50	33.56	62.92	38.81	46.34	28.29
Fiscal deficit (% of GDP)	2.01	5.58	1.94	4.44	2.05	6.15
Fiscal rule (0/1)	0.57	0.50	0.83	0.37	0.41	0.49
IMF program (0/1)	0.19	0.39	0.06	0.23	0.27	0.44
Observations	2,558		951		1,607	
Countries	133		37		96	

Note: Estimation sample of the baseline specification. Ratings are the average of Moody's, S&P and Fitch foreign-currency long-term ratings on a 1–21 scale (Kose et al., 2022). The tax gap is the share of estimated revenue potential that remains uncollected (Section 3.2). Inflation is measured as $\log(1+\pi)$, where π is the annual inflation rate; a value of 0.04 corresponds to about 4%.

3.2 Measuring the Tax Gap

Our measure of the tax gap relies on the stochastic frontier developed in Afonso, Alves, and Somda (2026), which makes it possible to assess the extent to which countries mobilize their estimated fiscal potential. The approach belongs to the stochastic frontier literature initiated by Aigner et al.

(1977) and applied to tax capacity and tax effort by Fenochietto and Pessino (2013). Rather than inferring fiscal capacity directly from observed revenue, it estimates a common revenue benchmark that accounts for the structural characteristics of each economy. We recall its main elements here.

For country i in year t, tax revenue is written as

$$\ln T_{it} = \alpha + x'_{it}\beta + \lambda_t + \varepsilon_{it}, \tag{1}$$

where T_{it} denotes tax revenue excluding social contributions as a percentage of GDP, x_{it} a vector of structural determinants of the tax base and λ_t year fixed effects. The composite error term is decomposed as

$$\varepsilon_{it} = v_{it} - u_{it}, \qquad v_{it} \sim \mathcal{N}(0, \sigma_v^2), \qquad u_{it} \sim \mathcal{N}^+(0, \sigma_u^2), \tag{2}$$

where v_{it} and u_{it} are independent of each other and of x_{it}. The two components have distinct interpretations. The symmetric term v_{it} captures measurement error and transitory shocks that may temporarily push revenue above or below the benchmark, such as an unusually severe recession or a revision of national accounts. The one-sided term $u_{it} \geq 0$, by contrast, captures the systematic shortfall of observed revenue relative to this benchmark. It is this difference in shape that allows the revenue shortfall to be identified. If revenue merely fluctuated around the frontier, residuals would be symmetric around zero, whereas a persistent tendency to remain below the frontier makes their distribution asymmetric. The frontier thus distinguishes temporary fluctuations from the lasting component associated with a revenue shortfall.

In line with the literature on tax capacity (Fenochietto & Pessino, 2013; Mawejje & Sebudde, 2019; McNabb et al., 2021), the vector x_{it} includes GDP growth, trade openness, the share of agriculture in GDP, natural resource rents and the urbanization rate. These variables reflect the size and taxability of the tax base. More urbanized and more open economies generally offer broader and more easily observable tax bases, whereas large agricultural sectors are harder to tax, as activity there is more dispersed and more informal. Natural resource rents reflect the particular revenue structure of resource-dependent economies, where the state may rely less on conventional taxation (Bornhorst et al., 2009; Besley & Persson, 2011; Crivelli & Gupta, 2014; Masi et al., 2024). GDP growth neutralizes the effect of the business cycle on revenue collection, and year fixed effects absorb shocks common to all countries. It is worth emphasizing that the frontier includes neither public debt, nor the fiscal balance, nor any rating-related variable, so that the tax gap is mechanically related neither to the fiscal fundamentals in our empirical specification nor to the ratings themselves.

Since u_{it} is not directly observed, it must be recovered from the composite residual. Jondrow et al. (1982) show that its distribution, conditional on the estimated residual ε_{it}, can be inferred. Building on this result, Battese and Coelli (1988) derive the conditional expectation of technical efficiency, which we use as our measure of tax effort,

$$TE_{it} = \mathbb{E}\left[\exp(-u_{it}) \mid \varepsilon_{it}\right], \qquad 0 < TE_{it} \leq 1. \tag{3}$$

The tax gap is its complement,

$$Gap_{it} = 1 - TE_{it}, \qquad 0 \leq Gap_{it} < 1. \tag{4}$$

A tax gap of 0.30 thus means that the country mobilizes about 70% of its estimated fiscal potential, given its structural characteristics.

The tax gap is a relative measure, which does not rank countries by the amount of revenue they collect. A country with a low revenue-to-GDP ratio may therefore lie close to its frontier, while

another country collecting more remains further from its own potential. As emphasized earlier, the tax gap does not measure the performance of the tax administration alone either, since it also reflects tax policy choices regarding rates, exemptions and the structure of taxation. We therefore interpret it as a measure of the underutilization of estimated fiscal potential.

Our baseline measure comes from a frontier common to all countries, with half-normal inefficiency, estimated by maximum likelihood over 1990–2021. This choice provides a comparable benchmark across countries while allowing the tax gap to vary over time within each of them, which is essential since our analysis relates annual changes in the gap to subsequent changes in ratings. A common frontier may admittedly attribute part of persistent cross-country differences to the gap, but these differences are absorbed by the country fixed effects of our specification, which exploits only within-country variation. Finally, we verify that our results do not depend on this choice by recomputing the tax gap from alternative frontier specifications, which change the distributional assumption on inefficiency, the treatment of cross-country heterogeneity and the set of structural determinants.

3.3 Stylized Facts

Before presenting the empirical strategy, we examine how sovereign ratings and the tax gap have evolved over the sample period, and how their dynamics relate to each other. Figure 1 traces the evolution of the average rating and the average tax gap between 1990 and 2021, for all countries as well as separately for advanced economies and for emerging and developing economies.

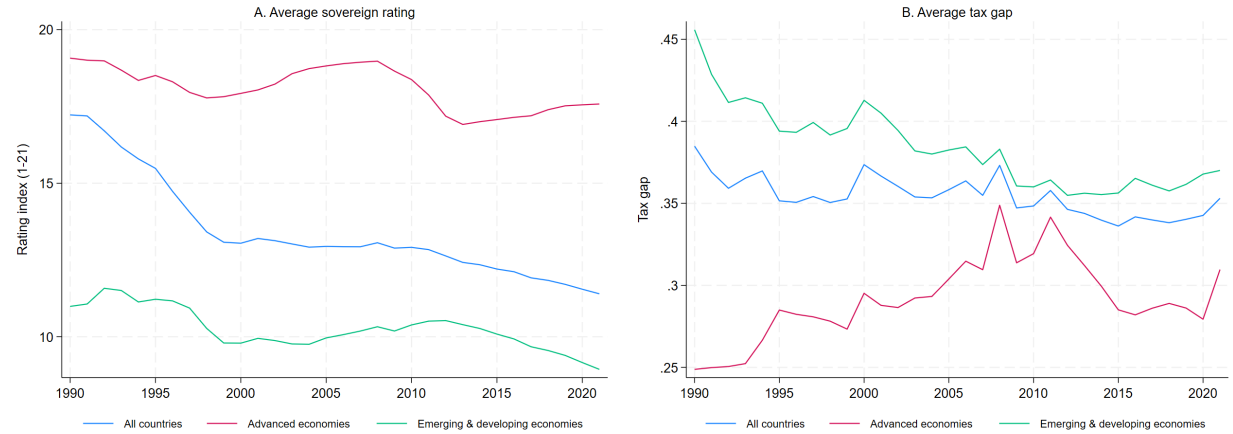
The first lesson from this figure concerns the size and persistence of the gap between the two groups of countries. Advanced economies are rated on average between 17 and 19 on the 1–21 scale, that is, in the AA category, whereas emerging and developing economies mostly lie between 9 and 11, around BB, in the speculative grade. This difference of seven to eight notches does not narrow noticeably over time. It goes hand in hand with an equally lasting difference in revenue mobilization, as emerging and developing economies leave a larger share of their fiscal potential uncollected every year, in the order of 36 to 41% over the past two decades, against 25 to 35% for advanced economies. The lowest-rated countries are thus also, on average, those that make the least use of their fiscal potential, which raises the question of the extent to which rating agencies take this dimension into account.

The second lesson concerns the joint dynamics of the two variables. In advanced economies, the tax gap widened almost continuously throughout the 2000s, from about 29% at the beginning of the decade to nearly 35% in 2008, and reached a new peak in 2011, at the height of the euro area sovereign debt crisis. The ratings of these countries, which had remained stable until 2008, subsequently declined markedly, losing about two notches between 2009 and 2013. Conversely, the narrowing of the tax gap from 2012 onwards preceded a gradual recovery of ratings from 2014. The sequence is thus the same in both directions, with changes in the tax gap preceding changes in ratings by a few years. In emerging and developing economies, the tax gap declined during the 2000s before stabilizing at around 36%, while ratings remained broadly stable and then weakened over the last decade, a development that most likely also reflects other factors, such as rising indebtedness and commodity price shocks.

Figure 2 examines the link between the two variables more directly, starting from rating decisions themselves. For each year, we distinguish countries whose composite rating falls by at least one notch, countries whose rating rises by at least one notch and countries whose rating remains unchanged, and we measure the change in their tax gap over the three years preceding the decision. This approach makes it possible to check whether rating decisions follow particular developments

in revenue mobilization.

The results are clear-cut. Downgrades are preceded by a widening of the tax gap of about 2.1 percentage points over the three previous years, whereas upgrades are preceded by a narrowing of about 2.6 percentage points. When the rating remains unchanged, which is the case for the vast majority of observations, the tax gap remains virtually stable. The confidence intervals associated with downgrades and upgrades exclude zero, indicating that these differences are not due to chance. Changes in the tax gap thus tend to precede rating decisions, in both directions. These regularities nevertheless remain purely descriptive. They account neither for the other determinants of ratings nor for shocks common to all countries, and may partly reflect other factors, such as crises or changes in macroeconomic conditions. Isolating the information specific to the tax gap, once these factors are taken into account, is precisely the purpose of the empirical strategy presented in the next section.



Source: Authors' computations.

Figure 1: Average sovereign rating and tax gap

Notes: Annual averages over the estimation sample. In the early 1990s, few emerging and developing economies were rated, so that averages partly reflect changes in sample composition as additional countries obtained a rating. Source: Authors' computations based on Kose et al. (2022) and the tax gap estimates described in Section 3.2.

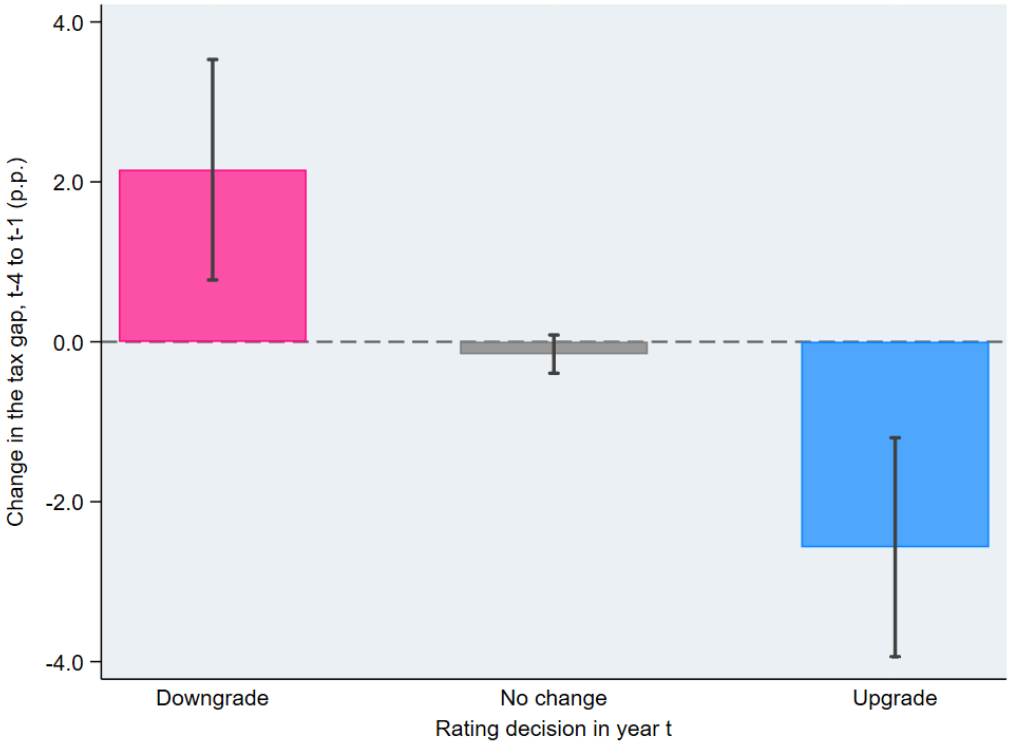


Figure 2: Change in the tax gap before rating decisions

Notes: Average change in the tax gap between $t-4$ and $t-1$ (percentage points), by rating decision in year t. A downgrade (upgrade) is a decline (increase) of at least one notch in the composite rating. Vertical bars indicate 95% confidence intervals. Source: Authors' computations.

4 Empirical Approach

We estimate the dynamic response of sovereign ratings to changes in the tax gap using panel local projections (Jordà, 2005; Jordà & Taylor, 2025). For each horizon $h = 0, \ldots, 6$, we estimate:

$$R_{i,t+h} - R_{i,t-1} = \alpha_i + \delta_t + \beta_h \, \Delta Gap_{it} + \gamma_h R_{i,t-1} + \Gamma_h X_{it} + \varepsilon_{i,t+h}, \tag{5}$$

where R_{it} is the rating index of country i in year t, ΔGap_{it} is the annual change in the tax gap, α_i and δ_t are country and year fixed effects, and X_{it} is a set of control variables. The dependent variable measures the cumulative change in the rating between the year preceding the shock and horizon h, so that β_h reads directly as the number of notches gained or lost h years after an increase in the tax gap. Country fixed effects absorb permanent characteristics that shape both fiscal capacity and creditworthiness, such as institutional quality or default history. Year fixed effects capture shocks common to all countries, such as commodity cycles, the global financial crisis or changes in agencies' methodologies. Standard errors are clustered by country, which accounts for the serial correlation in residuals induced by overlapping horizons. Responses are scaled to a one-standard-deviation increase in the change in the tax gap, about 4.1 percentage points of potential.

The choice of controls follows the literature on the determinants of sovereign ratings (Cantor & Packer, 1996; Afonso, 2003; Afonso et al., 2011). The vector X_{it} thus includes GDP per capita,

inflation, public debt, the fiscal deficit, the presence of a fiscal rule and participation in an IMF program, all measured at $t-1$ so that they are predetermined with respect to the shock. Two controls specific to our question complete the set. The lagged level of the tax gap separates the effect of a change in the gap from that of its initial level, and neutralizes any convergence of countries towards their potential. Lagged tax revenue ensures that β_h does not reflect the level of taxation already achieved, but the share of potential that remains untapped. Finally, GDP growth enters the specification contemporaneously. Since our index is an annual average of daily ratings, decisions taken during year t incorporate the cyclical information available within that year, whether high-frequency data or the forecasts on which agencies rely. Controlling for contemporaneous growth also rules out the possibility that a rise in the tax gap merely reflects a slowdown that weighs simultaneously on revenue and on the rating.

This approach suits our question for three reasons. First, ratings adjust gradually and with a lag (El-Shagi & von Schweinitz, 2022). A deterioration in revenue mobilization may thus show up in ratings only after several years, as its fiscal consequences materialize, so the relevant object is the path of ratings rather than their contemporaneous association with fiscal variables. Second, local projections estimate the response at each horizon directly, without imposing the dynamic restrictions of a VAR. They are therefore more robust to misspecification, easily accommodate an unbalanced panel and allow interactions to study heterogeneity in the responses (Jordà, 2005; Ramey & Zubairy, 2018). In population, local projections and VARs identify the same impulse responses (Plagborg-Møller & Wolf, 2021). In finite samples, local projections have lower bias at the cost of higher variance (Li et al., 2024), a cost that the size of our panel allows us to bear. Including the lagged rating further makes inference more robust to the strong persistence of ratings (Montiel Olea & Plagborg-Møller, 2021). Third, estimating the same specification for $h < 0$ provides a direct placebo test for pre-existing trends. If increases in the tax gap followed downgrades already under way, the pre-shock coefficients would be significant.

Local projections are subject to small-sample bias (Herbst & Johannsen, 2024), which the joint inclusion of country fixed effects and the lagged rating amplifies in panels. This is the Nickell bias (Nickell, 1981), which worsens with the horizon as the number of usable observations per country declines (Mei et al., 2026). We therefore also report estimates corrected with the split-panel jackknife proposed by Mei et al. (2026). Finally, we check that our results depend neither on the estimator nor on specification choices, using alternative measures of the tax gap, additional controls and different subsamples. These exercises, reported in Section 6, confirm the baseline effect.

5 Results

5.1 Baseline results

Do rating agencies penalize a government that leaves a growing share of its fiscal potential untapped? Our results show that they do, but with a delay. An increase in the tax gap does not trigger an immediate reaction. It translates into a downgrade that builds up gradually over the following years and persists for several years. Table 2 presents this result by enriching the specification step by step, from column (1), which only includes the lagged tax gap and country fixed effects, to column (4), our baseline specification, which also includes year fixed effects, macroeconomic and fiscal controls and lagged tax revenue. Figure 3 plots the dynamics estimated in this baseline specification, on which most of our interpretation rests.

In the year of the shock, the rating does not move. In column (4), the coefficient at $h = 0$ is close to zero and insignificant, as in all other columns. This initial silence is consistent with the way agencies form their judgment. Revenue data are published with a lag, rating reviews follow an annual calendar, and a one-off fall in collection may first be read as a cyclical accident (El-Shagi & von Schweinitz, 2022). The penalty comes once the deterioration is confirmed. One year after the shock, the rating has lost 0.07 notches, an effect significant at the 1% level. Two years after, the loss reaches 0.12 notches, and it peaks at 0.13 notches at $h = 4$. At this stage, agencies are no longer reacting to a temporary gap but revising their assessment of the state's capacity to mobilize its resources, and hence to service its debt. This revision is lasting, since at horizons $h = 5$ and $h = 6$ the rating remains about 0.10 notches below its initial level and the effect remains significant.

To assess the size of this effect, recall that our index is the average of the ratings of the three agencies. A one-notch downgrade by a single agency therefore lowers it by one third of a notch. Because ratings change infrequently, the estimated coefficient measures an expected effect, which combines many years without revision and a few discrete downgrades. At the peak, a one-standard-deviation increase in the tax gap, about 4.1 percentage points of potential, lowers the rating by 0.13 notches, which corresponds to about 40% of a one-notch downgrade by one of the three agencies. An increase of two standard deviations is thus equivalent, in expectation, to nearly a full downgrade by one agency. In the baseline specification, this effect is comparable to that associated with a public debt level about 12 percentage points of GDP higher, or with a fiscal deficit about 2.6 percentage points of GDP higher, two of the indicators agencies monitor most closely. It matters particularly near the investment-grade threshold, which separates BBB− (12 on our scale) from BB+ (11). Crossing this threshold forces many institutional investors to sell the bonds concerned and sharply raises borrowing costs, so that a fraction of a notch can have substantial consequences. Yet 15% of the observations in our sample lie between BB+ (11) and BBB (13), on either side of this threshold, and more than eight out of ten of them concern emerging and developing economies. For these countries, whose market access depends closely on their rating, revenue mobilization becomes a financing issue as much as a fiscal one.

Reading the table from left to right shows that this signal is specific to the tax gap. Moving from column (1) to column (2) leaves the coefficients virtually unchanged, which indicates that the effect does not reflect shocks common to all countries. Nor does the effect reduce to the fundamentals that agencies already follow closely. In column (3), once growth, inflation, debt, the deficit, the fiscal rule and IMF programs are accounted for, the downgrade remains of the same order of magnitude and significant at all horizons from $h = 1$ onwards, with a peak of 0.13 notches at $h = 4$. Above all, adding lagged tax revenue in column (4) barely changes the estimates. Agencies therefore do not react only to the observed level of taxation, but to the gap between what the state collects and what the structure of its economy would allow it to collect. Two countries with the same revenue ratio may thus be assessed differently depending on the share of their potential they leave untapped. These results also rest on a stable sample, since all 133 countries are present in the four columns and the introduction of controls removes only 49 observations out of more than 2,600.

Figure 3, which plots the estimates of column (4), also rules out the possibility that the rise in the tax gap merely accompanies a deterioration already under way. Before the shock, the rating shows no downward trend. The coefficients at negative horizons are close to zero, and the only ones that depart from zero, two and three years before the shock, are slightly positive. The break only occurs after the shock, when the path of the rating turns clearly downwards. This pattern confirms that agencies react to the deterioration in revenue mobilization itself, and not to earlier

difficulties that it would merely reflect.

These results raise several questions, which the following subsections address. An average effect of a fraction of a notch may reflect more or less frequent downgrade decisions, which we examine by studying agencies' decisions directly. The persistence of the effect then invites us to extend the horizon of analysis. Finally, the fact that the effect survives once revenue is controlled for suggests that the tax gap contains information that observed revenue does not reveal, an avenue we explore in the last subsection.

Table 2: Tax gap and sovereign ratings: baseline results

	(1) Tax gap	(2) + Year FE	(3) + Macro-fiscal controls	(4) + Tax revenue (baseline)
$h = 0$	-0.007	-0.009	-0.013	-0.013
	(0.015)	(0.015)	(0.014)	(0.014)
$h = 1$	-0.083***	-0.077***	-0.072***	-0.071***
	(0.024)	(0.024)	(0.024)	(0.023)
$h = 2$	-0.119***	-0.125***	-0.122***	-0.119***
	(0.028)	(0.028)	(0.031)	(0.030)
$h = 3$	-0.126***	-0.138***	-0.126***	-0.122***
	(0.036)	(0.036)	(0.041)	(0.040)
$h = 4$	-0.152***	-0.159***	-0.133***	-0.128***
	(0.044)	(0.046)	(0.050)	(0.049)
$h = 5$	-0.129***	-0.140***	-0.109**	-0.103*
	(0.047)	(0.050)	(0.054)	(0.053)
$h = 6$	-0.095**	-0.113**	-0.092*	-0.100*
	(0.045)	(0.046)	(0.050)	(0.051)
Lagged tax gap	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	No	Yes	Yes	Yes
Macroeconomic and fiscal controls	No	No	Yes	Yes
Lagged tax revenue	No	No	No	Yes
Observations ($h = 0$)	2,607	2,607	2,558	2,558
Countries	133	133	133	133

Note: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t - 1$ and $t + h$ following a one-standard-deviation increase in the tax gap. Panel local projections including the lagged rating. Macroeconomic and fiscal controls are lagged log GDP per capita, inflation, public debt, fiscal deficit, fiscal rule and IMF program, and contemporaneous GDP growth. Observations and countries refer to the sample at $h = 0$. Standard errors clustered by country in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

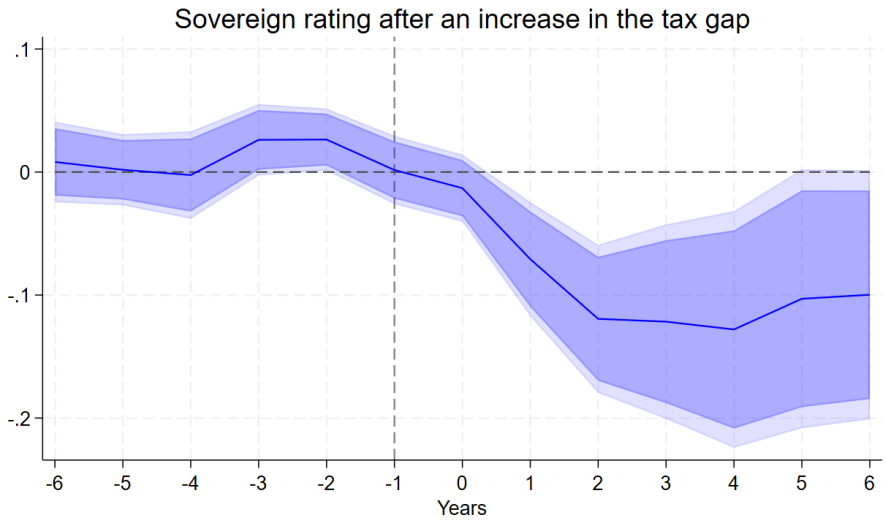


Figure 3: Sovereign rating after an increase in the tax gap

Notes: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap (about 4.1 percentage points of potential), estimated with the baseline specification (Table 2, column 4). The solid line shows the point estimates. Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country. Year $t-1$ is the reference period, and coefficients at negative horizons are placebo tests for pre-existing trends.

5.2 Rating decisions

One might object that our dependent variable is an average across the three agencies, smoothed over the year, so that a decline of 0.13 notches could merely be an artefact of this averaging, with no actual rating decision behind it. The question is not only statistical. For a government, it is downgrade decisions, publicly announced and widely reported, that alter market access and borrowing costs. To examine this point, we count for each country downgrades, defined as a decline of at least one notch in the composite rating index within the year, and upgrades, defined as an increase of the same size. We then estimate the same baseline specification on the cumulative difference between downgrades and upgrades between t and $t+h$, expressed per 100 countries. Expressing this balance as net upgrades instead would simply reverse its sign, since net upgrades are by construction the exact opposite of net downgrades; the corresponding response is reported in Online Appendix (Figure OA.1).

Figure 4 shows that the rise in the tax gap does translate into decisions. In the year of the shock, the balance of decisions is zero. It turns positive in the first year, though not yet significantly, and then rises markedly. Two years after a one-standard-deviation increase in the tax gap, there are about 3.3 additional net unfavourable decisions per 100 countries, and this balance peaks at around 4.3 in the third year. In other words, for a given country, the net probability of experiencing an unfavourable rather than a favourable rating decision increases by about 4 percentage points in the three to four years following the shock. This is a sizeable effect. Downgrades of at least one notch of the composite rating are rare events, which occur in 3.8% of country-years in our sample, against 2.6% for upgrades. A one-standard-deviation increase in the tax gap thus adds, within

three to four years, about as many net downgrades as are observed on average over an entire year. The balance then remains around 3.6 until the sixth year, which means that these decisions are not offset by subsequent upgrades. This timing mirrors the response of the rating index in the previous subsection. Agencies wait for the deterioration in revenue collection to be confirmed, then act, and do not reverse their decision. Estimating downgrades and upgrades separately shows that both margins contribute to this net effect in similar proportions: two years after the shock, downgrades increase by about 1.7 per 100 countries and upgrades fall by about 1.6 (Online Appendix Figures OA.2 and OA.3).

The decline in ratings therefore reflects actual agency actions rather than an artefact of averaging. This result also sheds light on the nature of the average effect reported in Table 2. Each decision counted here corresponds to a decline of at least one notch in the composite index, so the roughly four net unfavourable decisions per 100 countries account on their own for at least one third of the 0.13-notch loss observed at the peak. The remainder reflects smaller adjustments, typically a downgrade by a single agency, which moves the average by only one third of a notch. Agencies' sanction is thus not limited to marginal adjustments. A substantial part of it operates through clear-cut decisions, of the kind that weigh on financing conditions.

Figure 2, presented in Section 3.3, leads to the same conclusion from the opposite direction by looking backward from observed decisions. In the three years preceding a downgrade, the tax gap rises on average by about 2.1 percentage points, whereas it falls by about 2.6 points before an upgrade and remains stable when the rating does not change. Local projections look forward from changes in the tax gap, accounting for fixed effects, macroeconomic and fiscal fundamentals and the level of revenue. The two approaches agree on two points. They place the delay between the deterioration in collection and the decision at three to four years, and they show that the relationship works in both directions. An improvement in revenue mobilization precedes upgrades, just as its deterioration precedes sanctions. Agencies therefore do not merely punish setbacks. They also reward progress, which makes reducing the tax gap a lever from which governments can expect a rating benefit.

These decisions carry particular weight for emerging and developing economies. For a country close to the investment-grade threshold, a clear-cut downgrade can trigger forced sales by institutional investors and a sharp rise in borrowing costs. The persistence of net unfavourable decisions until the sixth year further suggests that the effect may extend beyond the horizon considered so far, which we examine in the next subsection.

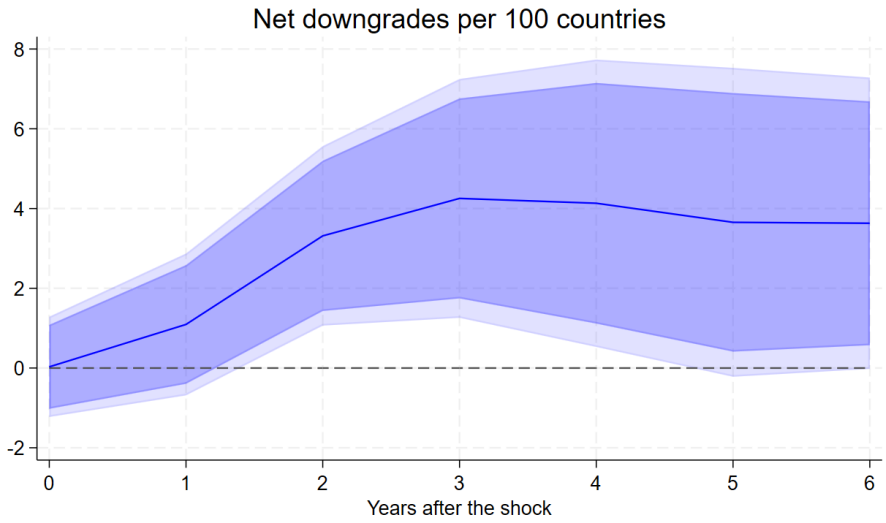


Figure 4: Net unfavourable rating decisions after an increase in the tax gap

Notes: Cumulative number of downgrades minus upgrades between t and $t+h$, per 100 countries, following a one-standard-deviation increase in the tax gap (about 4.1 percentage points of potential), estimated with the baseline specification (Table 2, column 4). A downgrade (upgrade) is a decline (increase) of at least one notch in the composite rating index within the year. The solid line shows the point estimates. Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country.

5.3 Long-run dynamics

The previous results show that the rating downgrade persists for at least six years after the shock. It remains to be seen whether this sanction is permanent or whether agencies eventually revise their judgment. The answer depends on what happens to the tax gap itself. A one-off increase in the gap may be quickly reabsorbed or, on the contrary, become entrenched, and agencies have no reason to treat these two situations in the same way. We therefore extend the horizon of the projections to twelve years, and estimate in parallel the response of the tax gap to its own shock. Figure 5 presents the two paths.

Panel (b) shows that an increase in the tax gap is persistent but not permanent. After a one-standard-deviation shock, about 4.1 percentage points of potential, the gap is still 2.2 points above its initial level two years later. About half of the shock has then been reabsorbed. The tax gap then continues to close gradually, still exceeding its initial level by about 1.3 points after four years and by 0.6 points after seven years. It only returns to this level after about eight years, before falling slightly below it in the following years. A deterioration in revenue mobilization is therefore not a one-year accident. It leaves a mark that takes nearly a decade to disappear.

The path of the rating, in panel (a), follows that of the tax gap with a lag. The downgrade builds up while the gap remains high and reaches its maximum, 0.13 notches, in the fourth year. It remains significant until the sixth year and stays around 0.08 to 0.10 notches until the ninth, although the widening of confidence intervals at the longest horizons makes these estimates less precise. The rating only returns to its initial level from the tenth year onwards, about two years after the tax gap has closed, and the coefficients become slightly positive beyond that, without being significant.

This parallel sheds light on the logic of the agencies. The sanction is not a definitive penalty

attached to a past episode. It lasts as long as the state's capacity to mobilize revenue remains below its potential, and it is only lifted once the improvement is confirmed. The lag of about two years between the closing of the tax gap and the recovery of the rating echoes the delay observed on the way down. Agencies wait for a deterioration to be confirmed before downgrading, and they likewise wait for a recovery to be confirmed before restoring the rating. The slight overshooting of the tax gap below its initial level from the ninth year onwards, which coincides with the recovery of the rating, is consistent with the symmetry highlighted in the previous subsection, where improvements in collection preceded upgrades.

For governments, these dynamics carry two lessons. The loss of credibility associated with a decline in revenue mobilization is reversible, and a sustained collection effort can restore the rating. But the road is long. A deterioration that occurs within a few years weighs on the rating for nearly a decade. For emerging and developing economies, which rely on markets to finance their development, this prolonged cost argues for preserving collection capacity rather than counting on later catch-up. The next subsection examines why agencies attach such importance to the tax gap, beyond what observed revenue reveals.

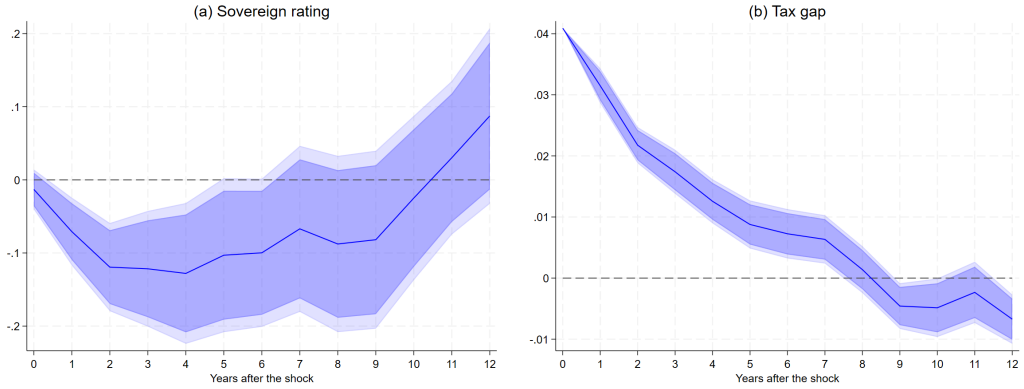


Figure 5: Long-run dynamics of the rating and the tax gap

Notes: Responses to a one-standard-deviation increase in the tax gap (about 4.1 percentage points of potential), estimated with the baseline specification over horizons $h = 0$ to 12. Panel (a): cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$. Panel (b): change in the tax gap relative to $t-1$; at $h = 0$, the response equals the shock by construction. The solid lines show the point estimates. Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country.

5.4 Beyond observed revenue

One might worry that the tax gap merely reflects falling revenue. Agencies would then be reacting to lower revenue rather than to untapped potential. The distinction matters. A fall in revenue may stem from the business cycle, a reform or a one-off shock. A rise in the tax gap means that the state collects less than the structure of its economy would allow it to collect. To separate these two signals, Table 3 and Figure 6 compare three adverse fiscal shocks, each of one standard deviation. Column (1) considers a fall in tax revenue. Column (2) reproduces the rise in the tax gap from the baseline specification. Column (3) measures the effect of a rise in the tax gap holding the change in revenue constant.

A fall in revenue, taken in isolation, triggers only a modest and short-lived sanction. In column (1), the rating declines by about 0.06 notches one and two years after the shock, and the effect then fades. From the third year onwards, the coefficients approach zero and are no longer significant. Agencies seem to treat a fall in revenue as a fluctuation that the economy will absorb, rather than as a sign of lasting weakness. The rise in the tax gap, in column (2), produces instead the gradual and persistent downgrade described in the previous subsections, with a peak of 0.13 notches at $h = 4$ and an effect still significant at $h = 6$. Beyond the second year, only the tax gap continues to weigh on the rating.

The decisive comparison is that of column (3), which holds the change in revenue constant. In this case, the tax gap can only rise because fiscal potential grows without being collected. The economy changes in ways that broaden its tax base, through greater trade openness, stronger urbanization or a declining weight of agriculture, but revenue does not follow. This case allows us to discriminate between three readings. If agencies only paid attention to revenue, the effect should be zero. If they perceived untapped potential as fiscal headroom, a reserve of revenue that could be mobilized when needed, it should be positive. Instead, the rating declines, and more sharply than in the baseline specification. The loss reaches 0.16 notches at $h = 2$, 0.23 notches at $h = 3$ and 0.25 notches at the peak, at $h = 4$, twice the baseline effect. It remains about 0.22 notches at $h = 6$, and all coefficients are significant at the 1% level from the second year onwards. At the peak, this loss represents in expectation three quarters of a one-notch downgrade by one of the three agencies.

Agencies therefore do not see untapped potential as a reassuring reserve, but as a sign of an inability to mobilize available resources. They judge not only how much revenue a state collects, but how much of its potential it leaves untapped. A country whose revenue stagnates while its economic base expands fails to turn this expansion into public resources. This points to weaknesses in tax administration, extensive exemptions or political resistance to taxation, all of which weigh on the capacity to service debt. The revenue-to-GDP ratio, often used as the main indicator of fiscal capacity, is thus misleading on its own. Two countries with the same revenue path may be assessed very differently depending on whether their potential grows or not.

For emerging and developing economies, this lesson is particularly important. Structural transformation rapidly broadens their tax base, and domestic revenue mobilization lies at the heart of strategies to finance development. Our results indicate that agencies watch whether this transformation translates into revenue. A broadening of the economic base that is not converted into public resources is not neutral for the rating. It is penalized more severely than a simple fall in revenue. The next section checks that these results hold under alternative measures of the tax gap, additional controls and different samples.

Table 3: Tax gap, tax revenue and sovereign ratings

	(1) Fall in tax revenue	(2) Rise in tax gap (baseline)	(3) Rise in tax gap at given revenue
$h = 0$	-0.019	-0.013	0.007
	(0.013)	(0.014)	(0.019)
$h = 1$	-0.057**	-0.071***	-0.059*
	(0.025)	(0.023)	(0.034)
$h = 2$	-0.055*	-0.119***	-0.164***
	(0.031)	(0.030)	(0.045)
$h = 3$	-0.022	-0.122***	-0.231***
	(0.036)	(0.040)	(0.058)
$h = 4$	-0.014	-0.128***	-0.254***
	(0.043)	(0.049)	(0.072)
$h = 5$	0.001	-0.103*	-0.229***
	(0.048)	(0.053)	(0.074)
$h = 6$	0.026	-0.100*	-0.219***
	(0.057)	(0.051)	(0.075)
Shock variable	Δ Tax revenue	Δ Tax gap	Δ Tax gap
Lagged tax gap	No	Yes	Yes
Lagged tax revenue	Yes	Yes	Yes
Contemporaneous Δ tax revenue	–	No	Yes
Macro-fiscal controls and fixed effects	Yes	Yes	Yes
Observations ($h = 0$)	2,647	2,558	2,558
Countries	133	133	133

Note: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$. Each column reports the response to a one-standard-deviation shock. Column (1): fall in log tax revenue. Column (2): rise in the tax gap (baseline, Table 2, column 4). Column (3): rise in the tax gap holding the contemporaneous change in tax revenue constant, so that the shock reflects a rise in revenue potential. All specifications include the lagged rating, country and year fixed effects, and the macroeconomic and fiscal controls of the baseline. Standard errors clustered by country in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

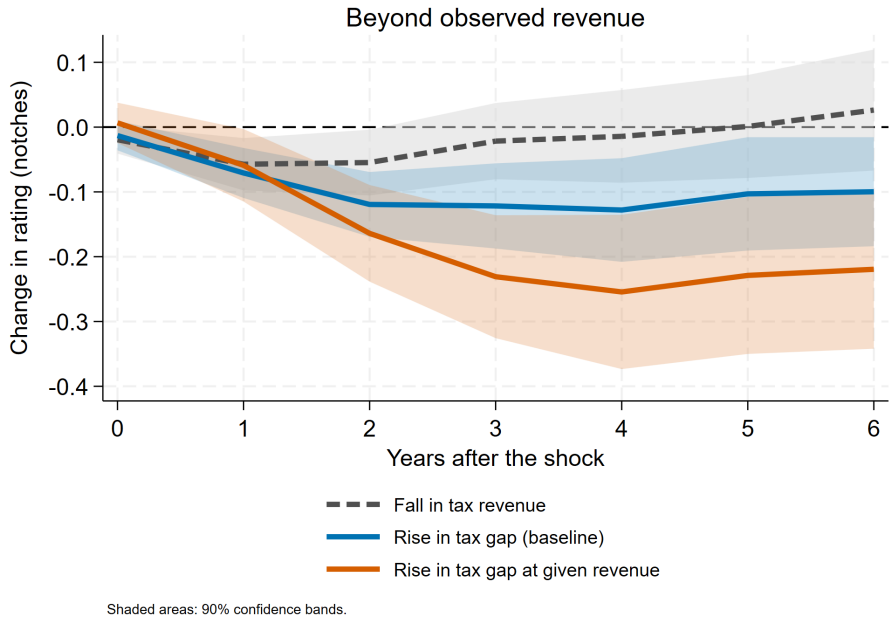


Figure 6: Tax gap and sovereign ratings beyond observed revenue

Notes: Cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$ following three one-standard-deviation shocks: a fall in log tax revenue (dashed line), a rise in the tax gap (baseline), and a rise in the tax gap holding the contemporaneous change in tax revenue constant. All responses are estimated with the controls and fixed effects of the baseline specification. Shaded areas are 90% confidence bands based on standard errors clustered by country. Estimates are reported in Table 3.

6 Robustness

This section checks that our results do not depend on particular methodological choices. Table 4 shows that the effect of the tax gap on ratings withstands the inclusion of additional controls, alternative measures of the tax gap, alternative samples and an estimator that corrects for the Nickell bias. Each column changes a single element of the baseline specification (Table 2, column 4) and keeps all others unchanged.

6.1 Additional controls

A first objection concerns possible omitted variables. Country fixed effects absorb the permanent characteristics of each country, but not factors that vary over time. If some of these factors affect both revenue mobilization and the sovereign rating, the estimated effect could reflect them rather than the tax gap. Three candidates deserve particular attention.

Institutional quality directly shapes the effectiveness of tax administration, and it is also one of the pillars of agencies' assessments. A deterioration in governance could therefore widen the tax gap and lower the rating without the former causing the latter. Corruption raises a related but distinct concern. Beyond administrative capacity, it affects the integrity of tax collection: bribes, negotiated exemptions and weak enforcement allow part of the tax base to escape taxation, while agencies treat corruption as a source of political and institutional risk in their own right. Natural

resource rents raise a similar issue. Countries that depend on them tend to rely less on taxation, and their ratings follow commodity price cycles.

Column (1) adds to the baseline specification government effectiveness, natural resource rents and the ICRG corruption score, all lagged by one period. The effect of the tax gap remains negative and significant from the first to the fourth year. The rating declines by 0.08 notches at $h = 1$ and by about 0.11 to 0.12 notches between $h = 2$ and $h = 4$, magnitudes close to those of the baseline specification, although the sample is smaller because of the coverage of the corruption score. The effect of the tax gap is thus not driven by institutional quality, corruption or dependence on natural resources.

6.2 Alternative measures of the tax gap

The tax gap is not observed. It is estimated from a stochastic frontier, which requires choosing a distribution for inefficiency, a source of revenue data and a set of determinants of fiscal potential. One might therefore worry that our results are an artefact of these choices rather than the reflection of an economic relationship. To address this concern, we reconstruct the tax gap in four different ways. Each alternative measure replaces the baseline measure, including in its lagged level, and each shock is expressed in standard deviations of the corresponding measure.

Column (2) assumes an exponential rather than a half-normal distribution of inefficiency. The results are virtually identical to those of the baseline specification, with a loss of 0.12 notches at $h = 4$ and an effect that remains significant until the sixth year. Column (3) estimates the frontier using UNU-WIDER revenue data, which harmonize national and international sources. The effect is somewhat larger, peaking at 0.15 notches in the fourth and fifth years. Column (4) excludes revenue from natural resources, whose evolution depends largely on commodity prices rather than on tax effort. This measure yields the strongest effect. The rating declines by 0.17 notches as early as the second year and by more than 0.20 notches between the fourth and fifth years, and all coefficients are significant at the 1% level from the first to the sixth year. Once the most volatile component of revenue is set aside, the signal that agencies draw from the tax gap becomes sharper, which confirms that they react to domestic collection effort rather than to commodity-driven revenue fluctuations. Finally, column (5) augments the frontier with three additional determinants of fiscal potential, namely the square of log GDP per capita, log population density and human capital. The estimates remain very close to the baseline, with a peak of 0.13 notches at $h = 4$. Whichever way the tax gap is constructed, the profile of the response thus remains the same.

6.3 Alternative samples

We next check that our results do not hinge on a few particular observations. Changes in the tax gap can be very large during exceptional episodes, such as a crisis or a break in revenue series, and a handful of extreme values could be enough to produce the average effect. Likewise, years of global crisis could bring about both a widespread rise in the tax gap and a wave of downgrades, without any causal link between the two.

Columns (6) and (7) reduce the weight of extreme shocks in two ways. Column (6) sets changes in the tax gap beyond the 1st and 99th percentiles to the value of these percentiles. Column (7) excludes observations below the 5th or above the 95th percentile. Far from weakening, the effect becomes stronger in both cases. It reaches 0.16 notches at $h = 4$ in column (6), and 0.22 to 0.23 notches between the fourth and fifth years in column (7). It is therefore ordinary changes in the

tax gap, and not a few exceptional episodes, that drive the result. Column (8) excludes 2008 and 2020, marked by the global financial crisis and the pandemic. The coefficients remain very close to the baseline, with a loss of 0.12 to 0.13 notches between the second and third years, which shows that the effect is not driven by these two global shocks. Column (9), finally, restricts the sample to the period after 2000, during which agencies' coverage of emerging and developing economies expanded markedly. The effect remains of a magnitude comparable to that of the baseline results.

6.4 SPJ correction

A final objection concerns the estimator itself. Mei et al. (2026) show that panel local projections estimated with fixed effects suffer from a Nickell bias, even when the lagged dependent variable is not among the regressors. This bias pulls the estimates towards zero, grows with the horizon and distorts inference, particularly when the number of countries is large relative to the time dimension, as is the case in our panel. To correct it, Mei et al. (2026) propose a split-panel jackknife estimator, which extends the approach of Dhaene and Jochmans (2015) to local projections. The model is estimated over the full period, then separately over its first and second halves. The corrected estimator equals twice the full-period estimate minus the average of the two half-period estimates. Since the bias is inversely proportional to the length of the period, this combination removes its first-order term. The resulting estimator is asymptotically unbiased and restores valid inference, while preserving the flexibility of local projections.

Column (10) of Table 4 and Figure 7 report the corrected estimates. The profile of the response is unchanged. The rating does not react in the year of the shock, declines by 0.08 notches at $h = 1$ and by 0.16 notches at $h = 2$, then reaches a trough of 0.20 notches in the fourth year, before stabilizing at around 0.16 to 0.17 notches. The magnitude of the effect, however, is markedly larger. The gap with the baseline specification is small in the short run, about 15% at $h = 1$, but it widens with the horizon, exceeding 50% at the peak and approaching 70% in the sixth year, in line with the prediction of an attenuation bias that grows with the horizon. Our baseline specification therefore provides a conservative estimate of the effect. Once the Nickell bias is corrected, a two-standard-deviation shock lowers the rating by about 0.4 notches at the peak, compared with a quarter of a notch in the baseline specification.

Table 4: Robustness checks

	Additional controls	Alternative measures of the tax gap				Alternative samples				SPJ correction
	(1) Gov. effectiveness, rents, corruption	(2) Exponential inefficiency	(3) UNU-WIDER revenue	(4) Non-resource tax gap	(5) Augmented frontier	(6) Winsorized shock (1%)	(7) Without 5% extreme shocks	(8) Without 2008, 2020	(9) Since 2000	(10) Split-panel jackknife
$h = 0$	-0.016	-0.015	-0.020	-0.021	-0.004	-0.011	-0.014	-0.015	-0.016	-0.024
	(0.013)	(0.014)	(0.018)	(0.018)	(0.013)	(0.016)	(0.025)	(0.014)	(0.014)	(0.016)
$h = 1$	-0.076***	-0.066***	-0.090***	-0.113***	-0.062***	-0.082***	-0.126***	-0.080***	-0.073***	-0.081***
	(0.024)	(0.024)	(0.031)	(0.032)	(0.022)	(0.026)	(0.038)	(0.022)	(0.025)	(0.030)
$h = 2$	-0.115***	-0.103***	-0.128***	-0.166***	-0.116***	-0.142***	-0.200***	-0.126***	-0.113***	-0.156***
	(0.033)	(0.030)	(0.037)	(0.036)	(0.029)	(0.032)	(0.048)	(0.031)	(0.034)	(0.044)
$h = 3$	-0.112***	-0.102***	-0.135***	-0.189***	-0.122***	-0.150***	-0.209***	-0.123***	-0.103**	-0.188***
	(0.042)	(0.039)	(0.046)	(0.046)	(0.039)	(0.042)	(0.056)	(0.040)	(0.043)	(0.065)
$h = 4$	-0.115**	-0.117**	-0.149***	-0.211***	-0.127***	-0.164***	-0.224***	-0.110**	-0.096**	-0.200**
	(0.050)	(0.049)	(0.057)	(0.059)	(0.046)	(0.052)	(0.067)	(0.045)	(0.049)	(0.082)
$h = 5$	-0.086*	-0.093*	-0.146***	-0.237***	-0.101**	-0.140**	-0.232***	-0.081*	-0.068	-0.160*
	(0.051)	(0.056)	(0.055)	(0.055)	(0.048)	(0.055)	(0.076)	(0.049)	(0.053)	(0.088)
$h = 6$	-0.051	-0.100**	-0.070	-0.182***	-0.108**	-0.121**	-0.187**	-0.082*	-0.061	-0.169*
	(0.055)	(0.051)	(0.056)	(0.056)	(0.043)	(0.053)	(0.080)	(0.050)	(0.048)	(0.093)
Observations ($h = 0$)	2,051	2,558	2,418	2,202	2,326	2,558	2,363	2,322	2,308	2,558
Countries	115	133	126	117	118	133	133	133	133	133

Notes: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap. All specifications include the lagged rating, country and year fixed effects and the controls of the baseline (Table 2, column 4), unless stated otherwise. Column (1) adds lagged government effectiveness (World Bank, Worldwide Governance Indicators), total natural resource rents in percent of GDP (World Bank, World Development Indicators) and the ICRG corruption score (PRS Group). Columns (2) to (5) replace the baseline tax gap, including its lagged level, with a measure based on an exponential inefficiency distribution, on UNU-WIDER revenue data, on non-resource tax revenue, and on a frontier augmented with additional determinants; each shock is scaled to one standard deviation of the corresponding measure. Column (6) winsorizes the change in the tax gap at the 1st and 99th percentiles, column (7) drops changes below the 5th or above the 95th percentile, column (8) excludes 2008 and 2020, and column (9) restricts the sample to 2000 onwards; this sample loses relatively few observations because rating coverage is limited in the 1990s (35 rated countries in 1990, against 90 in 2000). Column (10) applies the split-panel jackknife correction of Mei et al. (2026) to the baseline specification. Observations and countries refer to the sample at $h = 0$. Standard errors clustered by country in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

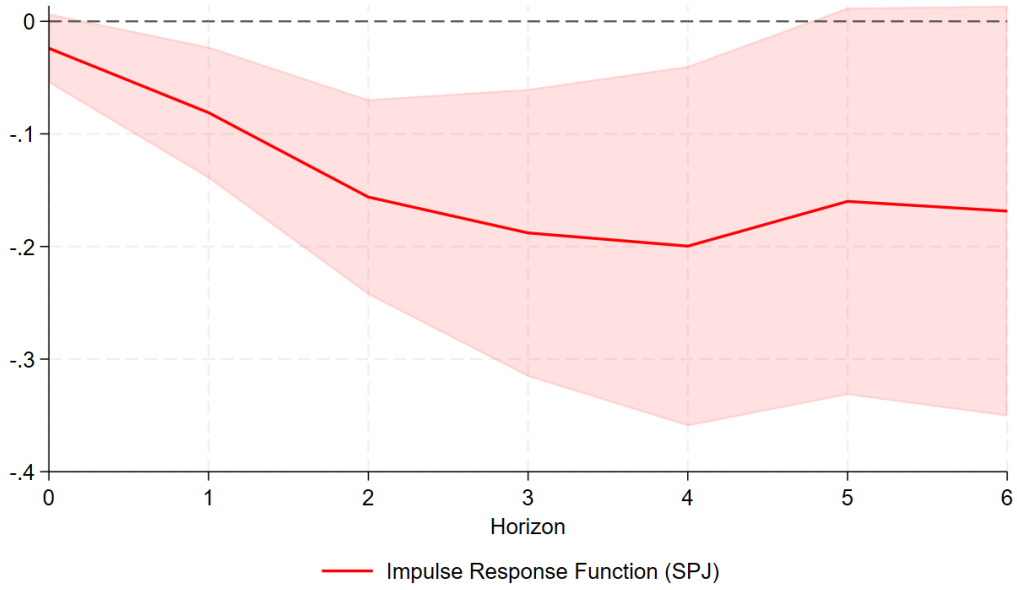


Figure 7: Sovereign rating after an increase in the tax gap: SPJ estimates

Notes: Cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated with the split-panel jackknife estimator of Mei et al. (2026), which corrects the Nickell bias of fixed-effects panel local projections. The specification includes the controls of the baseline (Table 2, column 4) and country and year fixed effects. The shaded area is the 95% confidence band based on standard errors clustered by country. Estimates are reported in Table 4, column (10).

7 Heterogeneity

The average effect documented so far may conceal very different reactions across countries and circumstances. Agencies do not necessarily interpret a decline in revenue mobilization in the same way in an advanced economy and in a developing country, in a country whose revenue depends on commodities or in a country under an IMF program. This section therefore estimates the baseline specification separately on subsamples defined by economic structure, institutional constraints and the political environment. Table 5 reports the estimates, and Figures 8 to 10 plot the corresponding dynamics.

7.1 Development and economic structure

Agencies' reaction depends first on the level of development, although the tax gap is followed by a rating deterioration in both groups of countries. In emerging and developing economies, the response is faster. The rating declines by 0.05 notches after one year and by 0.08 notches after two years, when the effect peaks, before fading from the fourth year onwards (Table 5, column 2). In advanced economies, the decline builds up more gradually but is larger and more persistent, reaching 0.29 notches after two years and nearly 0.40 notches in the fourth year (column 1). This

contrast partly reflects the difference in the volatility of the tax gap between the two groups. The tax gap varies much less in advanced economies, so that a one-standard-deviation rise is a rare event there, which may be read as a break in a collection capacity previously regarded as solid. In emerging and developing economies, where collection fluctuates more, agencies react quickly to the deterioration but seem to regard it as partly temporary, so that the decline does not extend beyond a few years.

Dependence on natural resources produces an even sharper contrast. In resource-rich countries, the rating does not react to the tax gap, and no coefficient is significant (column 3). In other countries, the downgrade is strong and persistent. The rating declines by 0.12 notches in the first year, by 0.19 notches in the second, and by 0.25 notches between the fourth and fifth years, with all coefficients significant at the 1% level (column 4). This result echoes the one obtained with the non-resource tax gap in the robustness section. When government revenue relies on commodities, agencies assess solvency in light of export prices and volumes rather than tax effort. Where taxation is the main source of public finance, by contrast, the failure to exploit fiscal potential becomes a central signal of fiscal fragility. Figure 8 illustrates these contrasts. Panels (a) and (d) show a clear break after the shock, whereas the paths in panels (b) and (c) move back towards zero at longer horizons.

7.2 Institutional constraints

The institutional framework in which fiscal policy operates also shapes how agencies read the tax gap. Countries under an IMF program offer a first contrast. In these countries, the rating does not react to a rise in the tax gap, and the coefficients remain close to zero at all horizons (column 5). Outside programs, the downgrade is significant and persistent, from 0.07 notches in the first year to 0.13 notches in the fourth (column 6). An IMF program provides agencies with an external anchor. Revenue targets are set, monitored and subject to conditionality, and Fund financing reduces default risk in the short run. A decline in collection is then perceived as a deviation that the program is meant to correct, rather than as a sign that control over public finances is being lost.

The level of corruption works in the same direction. In countries where corruption is low, the effect of the tax gap remains limited and is no longer significant beyond the first year (column 7). In countries where it is high, the rating declines by 0.07 notches after one year and by 0.11 to 0.12 notches between the second and fourth years (column 8). Where corruption is widespread, a rise in the tax gap is likely to reflect evasion, exemptions granted to vested interests or leakages in tax administration, all weaknesses of state capacity that agencies penalize. Figure 9 highlights these asymmetries. The response is flat under a program and muted when corruption is low, whereas it is clearly negative in the other two cases.

7.3 Political environment and uncertainty

The political calendar also affects agencies' reaction. In election years, a rise in the tax gap does not lead to a significant downgrade before the sixth year (column 9). Outside election years, the sanction follows the profile of the baseline specification, with a loss of 0.12 notches in the second year and of 0.13 notches between the third and fourth years (column 10). Agencies seem to consider that a decline in collection during an election period stems from the political cycle, through tax relief or laxer enforcement, and is therefore likely to be corrected once the election is over. The

same decline outside an election period cannot be attributed to the political calendar, and it is read as a sign of a more structural weakness.

Finally, uncertainty amplifies the sanction. When the World Uncertainty Index of Ahir et al. (2022) is above its median, the rating declines by 0.16 notches in the second year and by 0.20 to 0.21 notches between the third and fourth years, an effect that remains significant until the fifth year (column 11). When uncertainty is low, the downgrade is about half as large, around 0.07 to 0.09 notches, although it remains significant at several horizons (column 12). In an uncertain environment, financing needs are more exposed to shocks and market access is more fragile. The state's capacity to mobilize its own resources then becomes an all the more decisive criterion of its creditworthiness. Figure 10 illustrates these two contrasts. The responses in panels (b) and (c) are markedly stronger than those in panels (a) and (d).

Overall, the sanction associated with the tax gap is strongest where taxation is the state's main source of finance and where no external anchor reassures agencies. It is also stronger when the rise in the gap cannot be attributed to a temporary factor, such as the electoral calendar, and when the environment makes revenue mobilization more necessary. Appendix A.3 reports further splits, which show that the sanction is also stronger when public debt is high and when country-specific geopolitical risk is high (Caldara & Iacoviello, 2022), and that it is concentrated in the year following an election rather than the year preceding it.

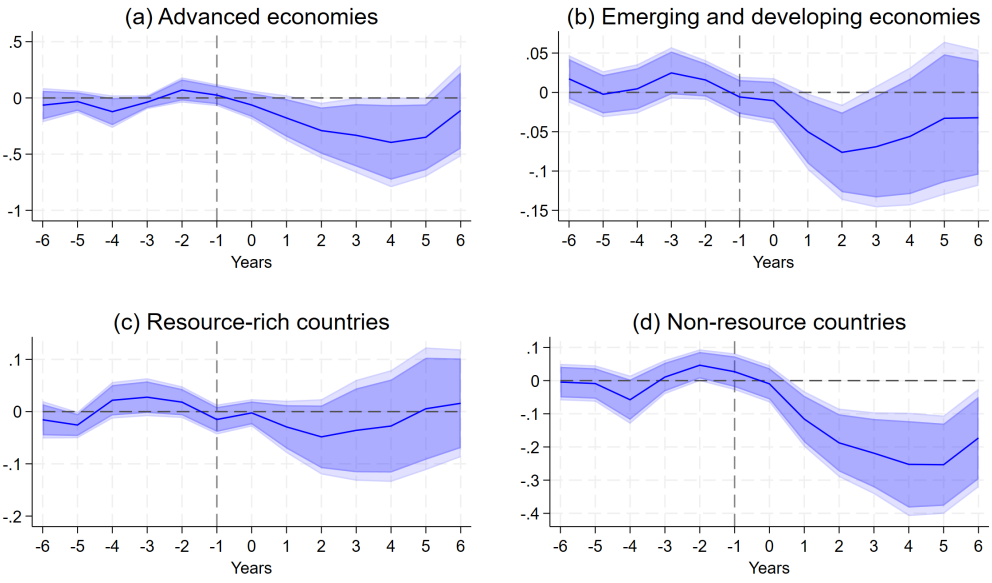


Figure 8: Heterogeneity: development and economic structure

Notes: Cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated separately in each subsample with the baseline specification (Table 2, column 4). Panels (a) and (b) compare advanced economies with emerging and developing economies, and panels (c) and (d) resource-rich countries with other countries. Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country. Horizons $h < 0$ are placebo tests. Estimates are reported in Table 5, columns (1) to (4).

Table 5: Heterogeneity

	Development and economic structure				Institutional constraints				Political environment and uncertainty			
	Development level		Natural resources		IMF program		Corruption		Elections		Uncertainty	
	(1) Advanced economies	(2) Emerging and developing	(3) Resource-rich	(4) Non-resource	(5) Under a program	(6) No program	(7) Low corruption	(8) High corruption	(9) Election year	(10) No election	(11) High uncertainty	(12) Low uncertainty
$h = 0$	-0.063	-0.011	-0.003	-0.010	0.065	-0.024**	-0.059*	0.002	-0.008	-0.007	0.013	-0.016
	(0.063)	(0.014)	(0.013)	(0.028)	(0.060)	(0.011)	(0.032)	(0.019)	(0.030)	(0.014)	(0.026)	(0.018)
$h = 1$	-0.178*	-0.050**	-0.029	-0.116***	-0.045	-0.073***	-0.080*	-0.070**	-0.040	-0.062***	-0.064*	-0.049*
	(0.099)	(0.024)	(0.025)	(0.042)	(0.077)	(0.022)	(0.044)	(0.028)	(0.049)	(0.023)	(0.035)	(0.029)
$h = 2$	-0.291**	-0.076**	-0.048	-0.188***	-0.067	-0.121***	-0.085	-0.111***	-0.059	-0.117***	-0.161***	-0.072**
	(0.122)	(0.030)	(0.036)	(0.052)	(0.070)	(0.032)	(0.060)	(0.035)	(0.065)	(0.027)	(0.045)	(0.030)
$h = 3$	-0.332**	-0.069*	-0.036	-0.219***	-0.021	-0.124***	-0.099	-0.108**	-0.045	-0.135***	-0.213***	-0.055
	(0.165)	(0.039)	(0.048)	(0.062)	(0.052)	(0.043)	(0.087)	(0.044)	(0.072)	(0.039)	(0.069)	(0.040)
$h = 4$	-0.396**	-0.056	-0.028	-0.253***	0.001	-0.128**	-0.082	-0.117**	-0.068	-0.133***	-0.198**	-0.091*
	(0.197)	(0.044)	(0.053)	(0.078)	(0.069)	(0.055)	(0.108)	(0.050)	(0.090)	(0.048)	(0.085)	(0.050)
$h = 5$	-0.350**	-0.033	0.005	-0.253***	-0.003	-0.099	-0.056	-0.084*	-0.137	-0.101*	-0.171**	-0.077
	(0.173)	(0.049)	(0.058)	(0.074)	(0.066)	(0.062)	(0.134)	(0.049)	(0.106)	(0.053)	(0.071)	(0.060)
$h = 6$	-0.112	-0.032	0.016	-0.173**	-0.014	-0.114*	-0.127	-0.067	-0.201*	-0.081	-0.102	-0.126**
	(0.201)	(0.043)	(0.051)	(0.075)	(0.081)	(0.063)	(0.116)	(0.048)	(0.111)	(0.062)	(0.065)	(0.061)
Observations ($h = 0$)	951	1,607	672	1,886	506	2,052	849	1,490	709	1,759	1,145	1,146
Countries	37	96	39	94	77	131	59	95	123	127	117	110

Notes: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated separately in each subsample with the baseline specification (Table 2, column 4). Columns (1) and (2) split the sample between advanced economies and emerging and developing economies, and columns (3) and (4) between resource-rich and other countries. Columns (5) and (6) split the sample according to participation in an IMF program in $t-1$, and columns (7) and (8) at the median of the lagged ICRG corruption score. Columns (9) and (10) separate election and non-election years, and columns (11) and (12) split the sample at the median of the lagged World Uncertainty Index (Ahir et al., 2022). Medians are computed on the estimation sample. Observations and countries refer to the sample at $h = 0$. Standard errors clustered by country in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

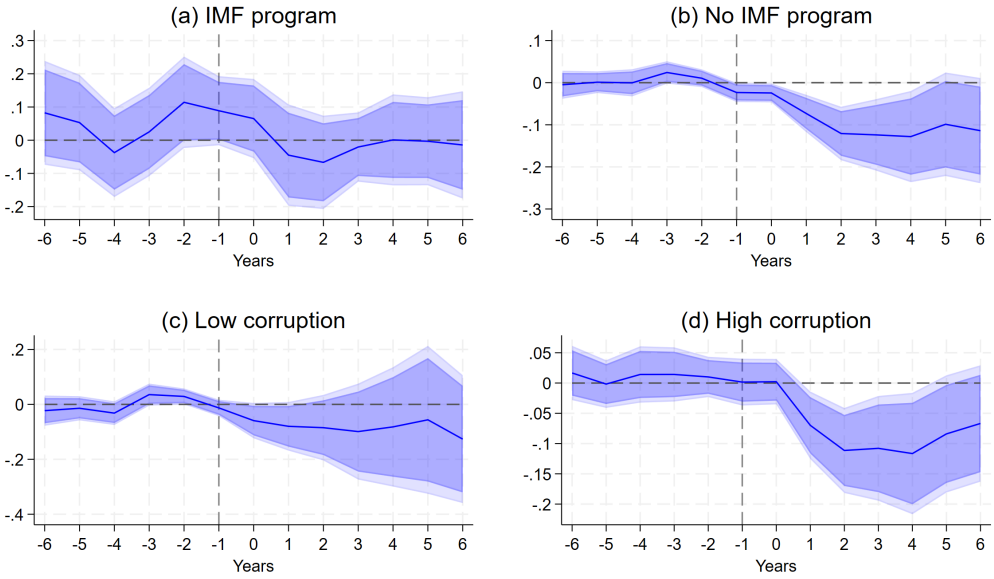


Figure 9: Heterogeneity: institutional constraints

Notes: Cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated separately in each subsample with the baseline specification (Table 2, column 4). Panels (a) and (b) compare countries under and without an IMF program in $t-1$, and panels (c) and (d) countries with low corruption (ICRG score above the median) and high corruption (ICRG score below the median). Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country. Horizons $h < 0$ are placebo tests. Estimates are reported in Table 5, columns (5) to (8).

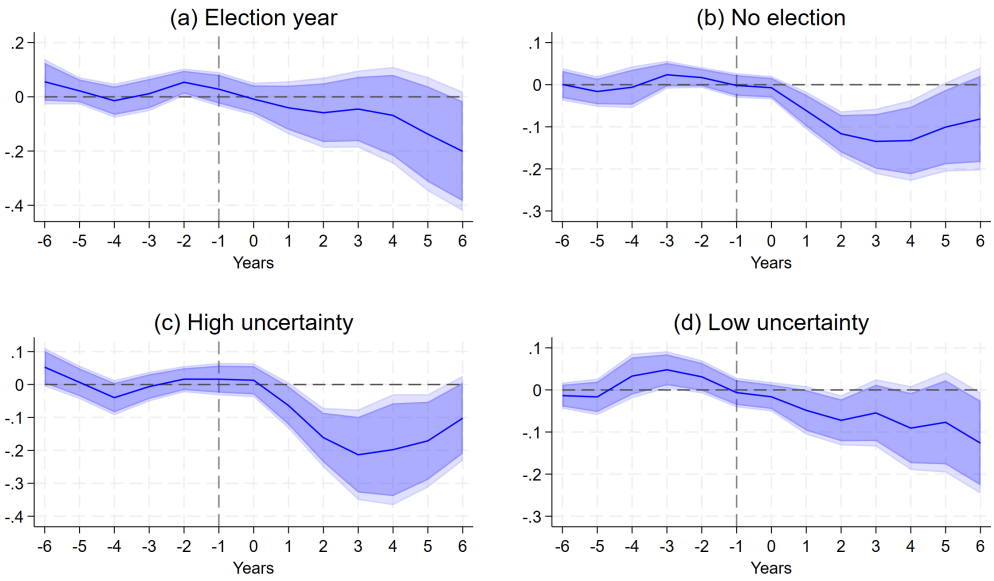


Figure 10: Heterogeneity: political environment and uncertainty

Notes: Cumulative change in the sovereign rating index (notches) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated separately in each subsample with the baseline specification (Table 2, column 4). Panels (a) and (b) compare election and non-election years, and panels (c) and (d) periods of high and low uncertainty, defined at the median of the lagged World Uncertainty Index (Ahir et al., 2022). Dark and light shaded areas are 90% and 95% confidence bands based on standard errors clustered by country. Horizons $h < 0$ are placebo tests. Estimates are reported in Table 5, columns (9) to (12).

8 Conclusion

This paper asks whether sovereign credit ratings reflect the share of fiscal potential that governments leave uncollected. Untapped potential can be read in two opposite ways, either as fiscal headroom that a government can mobilize when needed, or as a sign that it is unable or unwilling to convert its tax base into revenue. To distinguish between these interpretations, we estimate tax potential with a stochastic frontier, measure the tax gap as the share of this potential that remains uncollected, and trace the response of sovereign ratings to changes in the gap with panel local projections for 133 countries over 1990–2021.

The evidence points to revealed weakness. A one-standard-deviation increase in the tax gap is followed by a gradual decline in the sovereign rating of 0.13 to 0.20 notches after four years, an effect comparable to that associated with a public debt level about 12 percentage points of GDP higher. This decline translates into actual downgrades. For a given country, the net probability of a downgrade of its composite rating rises by about 4 percentage points within three to four years, more than the average annual frequency of such downgrades. Ratings follow the tax gap with a lag, both when it widens and when it closes. Most of the deterioration materializes between the second and the fourth year, and ratings return to their initial level only once the gap has been

reabsorbed, about a decade after the shock. Most importantly, the tax gap carries information that observed revenue does not reveal. A fall in revenue has only a short-lived effect, whereas a widening of the gap at a given level of revenue, which signals a tax base that expands without being taxed, has twice the baseline effect. What ratings appear to penalize is therefore not low revenue as such, but the failure to convert an expanding tax base into public resources. These results are robust to additional controls, alternative measures of the tax gap, alternative samples and a bias-corrected estimator.

The response is present across country groups, but its timing and strength vary with the context. In emerging and developing economies, ratings respond faster, while in advanced economies the decline builds up more slowly and is larger and more persistent. The response is also stronger where taxation is the main source of public finance and where no external anchor is in place, namely in countries that do not rely on natural resource revenue and outside IMF programs. It is stronger as well where corruption is high, when the rise in the gap cannot be attributed to the electoral calendar, and when uncertainty makes the mobilization of domestic resources more necessary. Ratings of resource-rich countries, by contrast, do not respond to the tax gap.

These findings carry implications for policy. For governments, they suggest that domestic revenue mobilization is not only a fiscal objective but also a determinant of the cost and availability of external financing. This matters particularly for emerging and developing economies, many of which are rated close to the investment-grade threshold, where a fraction of a notch can have large consequences for borrowing costs. The lag with which ratings recover also argues for preserving collection capacity rather than counting on later catch-up, since a deterioration that occurs within a few years weighs on the rating for nearly a decade. For resource-rich countries, the absence of any rating response to the tax gap means that markets provide little incentive to strengthen tax effort, which reinforces the case for domestic institutions that do. For rating agencies and analysts, the results suggest that the revenue-to-GDP ratio is an incomplete indicator of fiscal capacity, and that the distance between collected revenue and potential deserves explicit attention.

Several limitations open avenues for future research. The tax gap is estimated rather than observed, and although our results hold across several frontier specifications, it remains a summary measure that reflects tax policy choices as well as administrative capacity. Our identification relies on the timing of rating responses and on a rich set of controls rather than on exogenous variation in the tax gap, so that the estimates are best interpreted as the dynamic response of ratings to changes in revenue mobilization conditional on observed fundamentals. Future work could use agency-level ratings, outlooks and rating reports to identify more precisely how agencies process information on revenue mobilization, and examine whether bond spreads and other market indicators price the tax gap in the same way.

A Appendix

A.1 Ratings correspondence

Table 6: Correspondence between the rating index and agency ratings

Index	S&P / Fitch	Moody's	Index	S&P / Fitch	Moody's
21	AAA	Aaa	10	BB	Ba2
20	AA+	Aa1	9	BB−	Ba3
19	AA	Aa2	8	B+	B1
18	AA−	Aa3	7	B	B2
17	A+	A1	6	B−	B3
16	A	A2	5	CCC+	Caa1
15	A−	A3	4	CCC	Caa2
14	BBB+	Baa1	3	CCC−	Caa3
13	BBB	Baa2	2	CC	Ca
12	**BBB−**	**Baa3**	1	C / SD / D	C
11	BB+	Ba1			

Note: The table reports the standard 21-notch correspondence implied by this scale, consistent with the Fiscal Space database (Kose et al., 2022). The investment-grade threshold (12, BBB−/Baa3) is shown in bold.

A.2 Full estimates of the baseline specification

Table 7 reports all coefficients of the baseline specification for horizons $h = 0$ to $h = 6$. The first row reproduces the response of the rating to a one-standard-deviation increase in the tax gap, identical to that in column 4 of Table 2. The following rows give the coefficients of the control variables, expressed in their original units.

The control variables generally have the expected sign in light of the literature on the determinants of sovereign ratings (Cantor & Packer, 1996; Afonso, 2003; Afonso et al., 2011). The level of development is one of the most robust determinants. A 10% higher GDP per capita is associated with an improvement in the rating of about 0.05 notches in the year of the shock and about 0.2 notches four years later, an effect significant at the 1% level at all horizons. Growth works in the same direction, as one additional percentage point of growth raises the rating by 0.06 notches in the short run and by about 0.12 notches after two to three years. Fiscal variables, by contrast, weigh on the rating. Public debt 10 percentage points of GDP higher lowers the rating by 0.07 notches in the year of the shock and by 0.11 to 0.13 notches between the first and third years, an effect that then fades and is no longer significant in the sixth year. A fiscal deficit one percentage point of GDP higher lowers the rating by about 0.04 to 0.06 notches, an effect significant at the 1% level at all horizons.

Other variables have a more limited effect. Inflation is significant only at $h = 0$, at the 10% level. The presence of a fiscal rule has no significant effect, which suggests that, given countries' permanent characteristics, the adoption of a rule does not in itself change agencies' assessment. An IMF program is associated with a rating about 0.13 notches lower in the same year, which most likely reflects the financing difficulties that led to the program. The effect is no longer significant at subsequent horizons.

Two coefficients shed more direct light on the interpretation of our results. The lagged level of the tax gap has a negative and significant effect between the first and fourth years. A 10 percentage point higher initial gap lowers the rating by about 0.30 to 0.40 notches over this horizon. Agencies therefore take into account both the level of untapped potential and its evolution. Lagged tax revenue, by contrast, has no significant effect at any horizon. Once the tax gap is accounted for, the level of taxation provides no additional information to agencies, in line with the results of Section 5.4. Finally, the coefficient on the lagged rating is negative and increases in absolute value with the horizon, reaching about -1 in the sixth year. Deviations of the rating from its country-specific level thus tend to dissipate, and the initial rating has almost no predictive power six years later once fixed effects are accounted for.

The explanatory power of the specification declines with the horizon, as is usual in local projections. The R^2 falls from 0.89 in the year of the shock to 0.15 in the sixth year, as the lagged rating loses predictive power and shocks occurring after year t accumulate. The F-statistics nonetheless remain significant at all horizons. The consistency of these coefficients with the literature strengthens the credibility of the baseline specification and, consequently, that of the estimated effect of the tax gap.

Table 7: Baseline specification: full estimates

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
Tax gap shock (ΔGap_{it}, one s.d.)	-0.013	-0.071***	-0.119***	-0.122***	-0.128***	-0.103*	-0.100*
	(0.014)	(0.023)	(0.030)	(0.040)	(0.049)	(0.053)	(0.051)
Lagged rating (γ_h)	-0.142***	-0.324***	-0.501***	-0.653***	-0.788***	-0.913***	-1.031***
	(0.016)	(0.028)	(0.038)	(0.045)	(0.050)	(0.056)	(0.067)
Lagged tax gap	-0.650	-1.812**	-2.995**	-3.711**	-4.050*	-3.822	-3.109
	(0.428)	(0.844)	(1.219)	(1.689)	(2.123)	(2.615)	(3.093)
Lagged log tax revenue	0.012	-0.309	-0.710	-1.012	-1.288	-1.562	-1.631
	(0.122)	(0.265)	(0.454)	(0.681)	(0.896)	(1.186)	(1.315)
Lagged log GDP per capita	0.549***	1.232***	1.771***	1.990***	2.153***	2.440***	2.727***
	(0.184)	(0.370)	(0.517)	(0.611)	(0.706)	(0.777)	(0.863)
Lagged inflation, $\log(1+\pi)$	-0.727*	-0.043	0.830	1.203	0.751	0.219	-0.140
	(0.378)	(0.871)	(1.069)	(1.131)	(1.221)	(1.186)	(1.165)
Lagged public debt (% of GDP)	-0.0074***	-0.0113***	-0.0127***	-0.0113***	-0.0103**	-0.0089*	-0.0070
	(0.0016)	(0.0023)	(0.0031)	(0.0040)	(0.0046)	(0.0052)	(0.0057)
Lagged fiscal rule	-0.068	-0.073	-0.051	0.017	0.028	0.033	0.028
	(0.059)	(0.117)	(0.176)	(0.236)	(0.291)	(0.326)	(0.361)
Lagged IMF program	-0.134**	-0.133	-0.120	-0.209	-0.265	-0.282	-0.291
	(0.063)	(0.102)	(0.127)	(0.157)	(0.176)	(0.182)	(0.186)
GDP growth (%)	0.062***	0.109***	0.122***	0.117***	0.100***	0.092***	0.084***
	(0.011)	(0.017)	(0.022)	(0.023)	(0.023)	(0.022)	(0.021)
Lagged fiscal deficit (% of GDP)	-0.036***	-0.058***	-0.062***	-0.058***	-0.050***	-0.051***	-0.056***
	(0.006)	(0.011)	(0.014)	(0.016)	(0.017)	(0.019)	(0.019)
Constant	-2.534*	-4.923*	-6.046	-4.832	-3.422	-3.386	-4.199
	(1.455)	(2.866)	(4.067)	(4.816)	(5.617)	(6.197)	(6.772)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,558	2,427	2,295	2,166	2,041	1,918	1,797
R^2	0.885	0.707	0.527	0.375	0.256	0.183	0.147
F-statistic	834.81	155.16	73.49	26.92	12.85	5.96	4.72

Note: Full estimates of the baseline panel local projection (Table 2, column 4) at horizons $h = 0$ to 6. The dependent variable is the cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$. The first row reports the response to a one-standard-deviation increase in the tax gap, as in Table 2. The other rows report the coefficients of the control variables in their original units; the coefficient on the lagged rating corresponds to γ_h in equation (5). The R^2 refers to the equivalent regression of the rating at $t+h$ on the same regressors. Standard errors clustered by country in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

A.3 Additional checks

Table 8 gathers several complementary checks, which concern the measurement of the rating, country groupings, the fiscal position and the political environment. Column (1) addresses an objection about the nature of our dependent variable. The rating index is an average of the three agencies, computed over the year, and therefore takes non-integer values. One might worry that the estimated effect stems from this smoothing rather than from actual rating changes. When the index is rounded to the nearest notch, the results remain virtually identical to those of the baseline specification. The rating declines by 0.06 notches after one year and by 0.11 to 0.12 notches between the second and fourth years, then by 0.09 to 0.10 notches in the fifth and sixth years. This result complements the one in the subsection on rating decisions, and confirms that

the effect is not an artefact of averaging.

Columns (2) to (7) examine several country groupings. Members of the European Union and of the OECD react more strongly than other countries. The rating of EU countries declines by more than 0.50 notches in the fourth year, that of OECD countries by about 0.34 notches, and the effect remains significant until the sixth year for the latter. In other countries, the downgrade is more modest, around 0.08 notches, and significant between the first and third years. These results echo those obtained for advanced economies in the heterogeneity section, with which these groups largely overlap. Within the euro area, by contrast, no significant effect emerges, either in core or in periphery countries, following the classification of Afonso et al. (2014).[1] These estimates, however, rely on only five or six countries and should be read with caution. Membership of the monetary union, with its common support mechanisms, might also lead agencies to attach less weight to each state's revenue mobilization.

Columns (8) and (9) distinguish countries by their level of indebtedness. The sanction is stronger when debt is high. The rating then declines by 0.18 notches between the second and fourth years, compared with 0.08 to 0.09 notches when debt is low. This result is consistent with the logic of solvency. The heavier the debt, the more the state's capacity to mobilize revenue matters for servicing it, and the more a decline in collection worries agencies.

Columns (10) and (11) revisit the analysis of uncertainty using the country-specific geopolitical risk index of Caldara and Iacoviello (2022), which is available for a smaller number of countries. When geopolitical risk is high, the rating declines by 0.18 notches in the second year, by 0.23 notches in the third and by about 0.30 notches in the fourth. When it is low, no coefficient is significant. This contrast confirms, with a different measure, the result obtained with the World Uncertainty Index. The sanction associated with the tax gap is concentrated in environments where external shocks threaten financing conditions.

Columns (12) and (13), finally, clarify the role of the electoral calendar. When the rise in the tax gap occurs in the year before an election, the effect is limited and significant only in the second year. When it occurs in the year after an election, the sanction is instead strong and lasting. The rating declines by 0.17 notches in the second year, by 0.26 notches in the third and by 0.33 notches in the fourth, and the effect remains significant until the fifth year. This contrast extends the result of the heterogeneity section. Agencies seem to tolerate a decline in collection that they can attribute to the electoral context. A deterioration occurring after the election, however, can no longer be blamed on the campaign, and it is read as a deliberate choice or a lasting weakness of the new government.

[1] The euro area core comprises Austria, Belgium, Finland, France, Germany and the Netherlands, and the periphery comprises Greece, Ireland, Italy, Portugal and Spain.

Table 8: Additional checks

	Alternative rating	Country groupings						Fiscal position		Political environment			
	Rounded	European Union		OECD		Euro area		Public debt		Geopolitical risk		Election timing	
	(1) Rounded rating	(2) EU members	(3) Non-EU countries	(4) OECD members	(5) Non-OECD countries	(6) Core countries	(7) Periphery countries	(8) High debt	(9) Low debt	(10) High risk	(11) Low risk	(12) Year before election	(13) Year after election
$h = 0$	-0.008	-0.012	-0.011	-0.049	-0.010	-0.014	0.147	-0.008	-0.016	-0.020	0.060	0.011	-0.009
	(0.017)	(0.047)	(0.015)	(0.052)	(0.014)	(0.026)	(0.112)	(0.033)	(0.012)	(0.032)	(0.047)	(0.040)	(0.044)
$h = 1$	-0.058**	-0.164*	-0.054**	-0.182**	-0.051**	0.005	-0.082	-0.115**	-0.051**	-0.114*	-0.014	-0.079	-0.134**
	(0.026)	(0.087)	(0.025)	(0.076)	(0.025)	(0.069)	(0.158)	(0.054)	(0.024)	(0.067)	(0.041)	(0.052)	(0.061)
$h = 2$	-0.114***	-0.312**	-0.088***	-0.259***	-0.083**	0.025	-0.006	-0.177**	-0.091***	-0.175*	-0.063	-0.132**	-0.173***
	(0.033)	(0.133)	(0.031)	(0.092)	(0.033)	(0.111)	(0.362)	(0.074)	(0.032)	(0.099)	(0.053)	(0.066)	(0.064)
$h = 3$	-0.119***	-0.401**	-0.082**	-0.289**	-0.077*	-0.030	0.272	-0.185**	-0.079*	-0.231**	-0.069	-0.097	-0.260***
	(0.042)	(0.183)	(0.039)	(0.114)	(0.042)	(0.134)	(0.595)	(0.080)	(0.041)	(0.116)	(0.085)	(0.084)	(0.095)
$h = 4$	-0.123**	-0.525***	-0.069	-0.344**	-0.067	0.033	-0.006	-0.182**	-0.084	-0.297*	-0.018	-0.096	-0.334***
	(0.049)	(0.203)	(0.045)	(0.137)	(0.048)	(0.128)	(0.214)	(0.080)	(0.051)	(0.179)	(0.112)	(0.095)	(0.114)
$h = 5$	-0.100*	-0.378*	-0.043	-0.363***	-0.040	0.056	-0.192	-0.142*	-0.057	-0.217	0.027	-0.048	-0.273**
	(0.055)	(0.204)	(0.050)	(0.126)	(0.051)	(0.140)	(0.363)	(0.079)	(0.056)	(0.199)	(0.098)	(0.091)	(0.109)
$h = 6$	-0.090*	-0.113	-0.034	-0.299***	-0.021	0.030	0.026	-0.067	-0.060	-0.155	0.091	-0.006	-0.127
	(0.053)	(0.247)	(0.045)	(0.111)	(0.044)	(0.126)	(0.531)	(0.097)	(0.041)	(0.166)	(0.102)	(0.102)	(0.134)
Observations ($h = 0$)	2,558	711	1,847	980	1,578	180	135	1,279	1,279	525	526	667	702
Countries	133	27	106	38	95	6	5	109	109	37	31	121	121

Notes: Cumulative change in the sovereign rating index (notches, 1–21 scale) between $t-1$ and $t+h$ following a one-standard-deviation increase in the tax gap, estimated with the baseline specification (Table 2, column 4). Column (1) replaces the rating index with its value rounded to the nearest notch. Columns (2) to (13) estimate the baseline specification separately in each subsample. Columns (2) and (3) split the sample between EU members and other countries, columns (4) and (5) between OECD members and other countries, and columns (6) and (7) between core and periphery euro area countries (Afonso et al., 2014). Columns (8) and (9) split the sample at the median of lagged public debt, and columns (10) and (11) at the median of the lagged country-specific geopolitical risk index (Caldara & Iacoviello, 2022). Columns (12) and (13) restrict the sample to shocks occurring in the year before and in the year after an election. Euro area estimates rely on few clusters and should be read with caution. Medians are computed on the estimation sample. Observations and countries refer to the sample at $h = 0$. Standard errors clustered by country in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Data sources

Table 9: Main variables, definitions and data sources

Variable	**Definition**	**Source**
Sovereign rating	Average of Moody's, S&P and Fitch long-term foreign-currency sovereign ratings, converted to a 1–21 scale (21 = AAA) and averaged over the year	Kose et al. (2022)
Tax gap	Share of estimated tax potential not collected (1− tax effort), from a stochastic tax frontier (Section 3.2)	Authors' estimation following Afonso, Alves, & Somda (2026)
Tax revenue	Tax revenue excluding social contributions (% of GDP)	IMF World Revenue Longitudinal Data
GDP per capita, GDP growth, inflation	Log of real GDP per capita; real GDP growth (%); consumer price inflation, $\log(1+\pi)$	World Development Indicators
Public debt, fiscal deficit	General government gross debt and overall fiscal deficit (% of GDP)	IMF Fiscal Monitor
Fiscal rule	Dummy equal to one if at least one national or supranational fiscal rule is in force	IMF Fiscal Rules Dataset
IMF program	Dummy equal to one in years under an IMF lending arrangement	Vreeland (2007); IMF MONA
Government effectiveness	Perceptions of the quality of public services and policy implementation	Worldwide Governance Indicators
Resource rents	Total natural resource rents (% of GDP)	World Development Indicators
Alternative tax revenue	Total and non-resource tax revenue (% of GDP), used to re-estimate the tax gap	UNU-WIDER GRD
Human capital, population density	Human capital index; log of population per km^2, used in the augmented frontier	Penn World Table (Feenstra et al., 2015); World Development Indicators
Country groups	Advanced vs emerging and developing economies; resource-rich countries	IMF classifications
Corruption	ICRG corruption score (higher = less corruption)	International Country Risk Guide
Election years	Dummy equal to one in years with a presidential or legislative election	Database of Political Institutions
Uncertainty	World Uncertainty Index	Ahir et al. (2022)
Geopolitical risk	Country-specific geopolitical risk index	Caldara & Iacoviello (2022)

A.5 List of countries

Table 10: Countries in the estimation sample

Albania	Denmark	Kyrgyz Republic	Rwanda
Angola	Dominican Republic	Latvia	San Marino
Argentina	Ecuador	Lebanon	Saudi Arabia
Armenia	Egypt	Lesotho	Senegal
Aruba	El Salvador	Lithuania	Serbia
Australia	Estonia	Luxembourg	Seychelles
Austria	Eswatini	Malaysia	Singapore
Azerbaijan	Ethiopia	Mali	Slovakia
Bahamas	Finland	Malta	Slovenia
Bahrain	France	Mauritius	South Africa
Bangladesh	Gabon	Mexico	Spain
Belarus	Gambia	Moldova	Sri Lanka
Belgium	Georgia	Mongolia	Sweden
Benin	Germany	Montenegro	Switzerland
Bolivia	Ghana	Morocco	Tajikistan
Bosnia and Herzegovina	Greece	Mozambique	Tanzania
Botswana	Guatemala	Namibia	Thailand
Brazil	Honduras	Netherlands	Togo
Bulgaria	Hong Kong SAR	New Zealand	Tunisia
Burkina Faso	Hungary	Nicaragua	Türkiye
Cabo Verde	Iceland	Niger	Turkmenistan
Cambodia	India	North Macedonia	Uganda
Cameroon	Indonesia	Norway	Ukraine
Canada	Iran	Oman	United Arab Emirates
Chile	Iraq	Pakistan	United Kingdom
China	Ireland	Panama	United States
Colombia	Israel	Paraguay	Uruguay
Congo, Dem. Rep.	Italy	Peru	Uzbekistan
Congo, Rep.	Japan	Philippines	Venezuela
Costa Rica	Jordan	Poland	Viet Nam
Côte d'Ivoire	Kazakhstan	Portugal	Zambia
Croatia	Kenya	Qatar	
Cyprus	Korea, Rep.	Romania	
Czechia	Kuwait	Russia	

Note: The estimation sample of the baseline specification covers 133 countries over 1990–2021.

References

Afonso, A. (2003). Understanding the determinants of sovereign debt ratings: Evidence for the two leading agencies. *Journal of Economics and Finance*, *27*(1), 56–74. https://doi.org/10.1007/BF02751590

Afonso, A., Alves, J., Gogas, P., & Papadimitriou, T. (2026). *Sovereign ratings and risk pricing, agency divergences in the European Union* (CESifo Working Paper No. 12832). CESifo. Munich.

Afonso, A., Alves, J., & Somda, K. E. (2026). *Fiscal policy and its watchers: Do fiscal councils help close the tax gap?* (CESifo Working Paper No. 13081). CESifo. Munich.

Afonso, A., Arghyrou, M. G., & Kontonikas, A. (2014). Pricing sovereign bond risk in the european monetary union area: An empirical investigation. *International Journal of Finance & Economics*, *19*(1), 49–56.

Afonso, A., Gomes, P., & Rother, P. (2011). Short- and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, *16*(1), 1–15.

Ahir, H., Bloom, N., & Furceri, D. (2022). *The world uncertainty index* (NBER Working Paper No. 29763). National Bureau of Economic Research.

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*(1), 21–37.

Aizenman, J., Hutchison, M., & Jinjarak, Y. (2013). What is the risk of european sovereign debt defaults? fiscal space, CDS spreads and market pricing of risk. *Journal of International Money and Finance*, *34*, 37–59.

Albers, T. N. H., Jerven, M., & Suesse, M. (2023). The fiscal state in Africa: Evidence from a century of growth. *International Organization*, *77*(1), 65–101.

Balima, H. W. (2020). Coups d'état and the cost of debt. *Journal of Comparative Economics*, *48*(3), 509–528.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*(3), 387–399.

Benitez, J. C., Mansour, M., Pecho, M., & Vellutini, C. (2023). *Building tax capacity in developing countries* (IMF Staff Discussion Note No. SDN/2023/006). International Monetary Fund.

Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review*, *99*(4), 1218–1244.

Besley, T., & Persson, T. (2011). *Pillars of prosperity: The political economics of development clusters*. Princeton University Press.

Besley, T., & Persson, T. (2014). Why do developing countries tax so little? *Journal of Economic Perspectives*, *28*(4), 99–120.

Bi, H. (2012). Sovereign default risk premia, fiscal limits, and fiscal policy. *European Economic Review*, *56*(3), 389–410.

Blanchard, O. J., Leigh, D., & Mishra, P. (2026). *Ratings, debt, and deficits: An exploration* (IMF Working Paper No. WP/26/195). International Monetary Fund.

Bornhorst, F., Gupta, S., & Thornton, J. (2009). Natural resource endowments and the domestic revenue effort. *European Journal of Political Economy*, *25*(4), 439–446.

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, *112*(4), 1194–1225.

Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Federal Reserve Bank of New York Economic Policy Review*, *2*(2), 37–53.

Crivelli, E., & Gupta, S. (2014). Resource blessing, revenue curse? Domestic revenue effort in resource-rich countries. *European Journal of Political Economy*, *35*, 88–101.

Cyan, M. R., Martinez-Vazquez, J., & Vulovic, V. (2013). *Measuring tax effort: Does the estimation approach matter and should effort be linked to expenditure goals?* (ICePP Working Paper No. 1308). International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Dhaene, G., & Jochmans, K. (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies*, *82*(3), 991–1030.

Dincecco, M., & Katz, G. (2016). State capacity and long-run economic performance. *Economic Journal*, *126*(590), 189–218.

Dincecco, M., & Prado, M. (2012). Warfare, fiscal capacity, and performance. *Journal of Economic Growth*, *17*(3), 171–203.

Eaton, J., & Gersovitz, M. (1981). Debt with potential repudiation: Theoretical and empirical analysis. *Review of Economic Studies*, *48*(2), 289–309.

El-Shagi, M., & von Schweinitz, G. (2022). Why they keep missing: An empirical investigation of sovereign bond ratings and their timing. *Scottish Journal of Political Economy*, *69*(2), 186–224.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, *105*(10), 3150–3182.

Fenochietto, R., & Pessino, C. (2013). *Understanding countries' tax effort* (IMF Working Paper No. 13/244). International Monetary Fund.

Gupta, A. S. (2007). *Determinants of tax revenue efforts in developing countries* (IMF Working Paper No. 07/184). International Monetary Fund.

Herbst, E. P., & Johannsen, B. K. (2024). Bias in local projections. *Journal of Econometrics*, *240*(1), 105655.

IMF. (2026). *Fiscal Monitor: Fiscal policy under pressure: High debt, rising risks*. International Monetary Fund.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, *19*(2-3), 233–238.

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), 161–182.

Jordà, Ò., & Taylor, A. M. (2025). Local projections. *Journal of Economic Literature*, *63*(1), 59–110.

Kose, M. A., Kurlat, S., Ohnsorge, F., & Sugawara, N. (2022). A cross-country database of fiscal space. *Journal of International Money and Finance*, *128*, 102682.

Li, D., Plagborg-Møller, M., & Wolf, C. K. (2024). Local projections vs. vars: Lessons from thousands of dgps. *Journal of Econometrics*, 105722.

Mansour, M., Verhoeven, M., Sawadogo, F., & Tan, B. C. S. (2025). *Methodology and overview of the IMF's World Revenue Longitudinal Database* (tech. rep.). International Monetary Fund. Washington, DC.

Masi, T., Savoia, A., & Sen, K. (2024). Is there a fiscal resource curse? Resource rents, fiscal capacity and political institutions in developing economies. *World Development*, *177*, 106532.

Mawejje, J., & Sebudde, R. K. (2019). Tax revenue potential and effort: Worldwide estimates using a new dataset. *Economic Analysis and Policy*, *63*, 119–129.

McNabb, K., Danquah, M., & Tagem, A. M. E. (2021). *Tax effort revisited: New estimates from the Government Revenue Dataset* (WIDER Working Paper No. 2021/170). UNU-WIDER. Helsinki.

Mei, Z., Sheng, L., & Shi, Z. (2026). Nickell bias in panel local projection: Financial crises are worse than you think. *Journal of International Economics*, 104210.

Montiel Olea, J. L., & Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, *89*(4), 1789–1823.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

OECD. (2026). *ODA projections for 2026 and the near-term: Implications for vulnerable countries and sectors* (OECD Policy Briefs No. 59). OECD Publishing. Paris.

Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses. *Econometrica*, *89*(2), 955–980.

Prichard, W. (2016). Reassessing tax and development research: A new dataset, new findings, and lessons for research. *World Development*, *80*, 48–60.

Ramey, V. A., & Zubairy, S. (2018). Government spending multipliers in good times and in bad: Evidence from US historical data. *Journal of Political Economy*, *126*(2), 850–901.

Thornton, J., & Vasilakis, C. (2018). Fiscal rules and government borrowing costs: International evidence. *Economic Inquiry*, *56*(1), 446–459.

UNCTAD. (2025). *A world of debt 2025: It is time for reform* (tech. rep.). United Nations Conference on Trade and Development. Geneva.

UNU-WIDER. (2025). UNU-WIDER Government Revenue Dataset [Version 2025]. https://doi.org/10.35188/UNU-WIDER/GRD-2025

Vreeland, J. R. (2007). *The International Monetary Fund: Politics of conditional lending*. Routledge.