Room to Respond*

Justin Eloriaga
International Monetary Fund

Abstract

When a financial crisis strikes, governments with room to act are believed to escape with smaller recessions, and one number dominates the assessment of that room, the ratio of public debt to GDP. Across 140 systemic banking crises since 1976, we find the ratio matters less, and differently, than the advice assumes. It adds nothing to crisis prediction once credit-boom symptoms are accounted for, and its association with severity is narrow, concentrated in advanced economies at long horizons and above 90 percent of GDP, while in emerging market and developing economies the average penalty at those horizons is demonstrably small, though tail outcomes are estimated less precisely. What debt does predict, strongly, is the response. Advanced economies entering a crisis with one standard deviation more debt tightened their budgets by more than four percentage points of GDP relative to low-debt peers. We therefore construct two indices, the transparent Policy Buffers Index and the outcome-based Effective Space Index, which carry information the ratio alone does not.

JEL classification. E32, E52, E62, G01, H63.
Keywords. Policy space, fiscal buffers, financial crises, local projections, heterogeneous treatment effects, crisis recovery.

*Economist, IMF Shanghai Center. Email jeloriaga@imf.org. The author thanks Qing Peng and Johannes Weigand for helpful comments. The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management.

1 Introduction

When a financial crisis strikes, does it matter how much room the government has to fight back? A prominent body of evidence says yes. Romer and Romer (2019) find that among 30 OECD economies since 1980, output falls by less than one percent after a financial crisis when the government enters it with low debt, but by roughly seven percent when it enters with high debt because, they argue, constrained governments cut spending into the downturn while unconstrained ones support their economies. Jordà et al. (2016) and Romer and Romer (2018) reach similar conclusions for public debt and for the room to cut interest rates. On this evidence rests one of the most common prescriptions in macroeconomics, the call to rebuild buffers in good times.

Our question is the one that comes before the prescription. What actually measures the room to respond? In practice the answer is a handful of balance-sheet ratios, chiefly public debt relative to GDP, and we refer to these observable ratios as measured buffers. Measured buffers turn out to predict far less than the advice assumes. Raw comparisons show at most a loose and unstable link between pre-crisis debt and crisis depth in any country group, and the deepest crises contradict the ratio directly. The waves of 1997 and 2008 struck broadly, most of emerging Asia in the first and most of the advanced world in the second, yet among the hardest hit were the governments whose debt looked enviably low, Ireland, Iceland, and Spain in 2008 and Korea in 1997, plausibly because the same credit booms that preceded those crises were temporarily inflating tax revenues and GDP and so making the public finances look healthier than they were, a windfall mechanism documented for exactly these episodes (Morris and Schuknecht, 2007).

The lesson we draw is that a low debt ratio is not the same thing as room to respond. What a government can actually do in a crisis, commonly referred to as effective policy space, depends on more than the ratios. It depends on whether the buffer is genuine or an artifact of a boom, on whether markets will keep lending, and on the currency in which the debt is owed. Ireland's buffer in 2008 was, on this reading, an illusion of its boom. The property boom filled the treasury with windfall revenues and inflated the GDP denominator, so a debt ratio near 25 percent of GDP concealed a fiscal position that depended on the boom itself, and when the boom collapsed the revenues went with it and the cost of rescuing the banks arrived on top. The United States carries very high debt and, on the logic of Brunnermeier et al. (2022), retains room because it borrows in the currency it prints. We turn this distinction into measurement. We use observed crisis outcomes to learn what actually predicts the room to respond.

We use the Global Macro Database (Müller et al., 2025), which provides harmonized

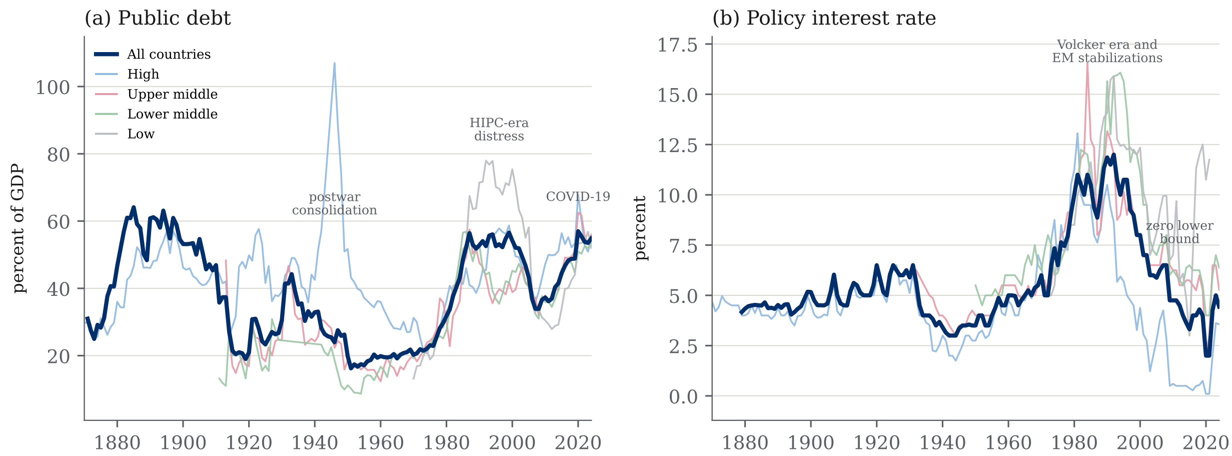


Figure 1: Policy buffers over 150 years.

Notes: Cross-country median (dark line) and income-group medians (lighter lines) of government debt in percent of GDP (panel a) and the policy rate in percent (panel b), 1870 to 2024, over the countries reporting each year. Source: Global Macro Database.

fiscal, monetary, external, and crisis data for a far broader set of economies and years than the samples behind the advice, reaching back to the nineteenth century. The long-run crisis flags supply the historical backdrop, and the newly updated Laeven and Valencia (2026) chronology supplies the estimation events. Banking crises are the laboratory throughout, and we test the same design on currency and sovereign debt crises as well. That is enough events, across enough regimes, to ask not just whether space matters on average but which space matters where.

Before any estimation, the long sweep of this record already tells much of the story. Four of its features shape everything that follows.

Policy buffers are historically depleted. Figure 1 plots the cross-country median of the two canonical buffer measures, with income-group medians layered in lighter colors to show how differently the groups have traveled.

The debt panel is a history of buffers built and spent. The median ratio falls from its interwar peaks to a trough of 17 percent of GDP in 1955, as postwar growth, inflation, and financial repression melted the war debts away. It then climbs almost without interruption. The low-income median peaks near 80 percent in the early 1990s, at the height of the distress that led to relief under the Heavily Indebted Poor Countries (HIPC) Initiative, before debt forgiveness and the 2000s boom bring it back down. The global financial crisis and the pandemic then push the overall median to roughly 55 percent, its highest sustained level in the sample outside the world wars. The rate panel tells a parallel story. The median policy rate holds near five percent through the gold standard era, peaks at 12 percent in 1992 at the end of the global disinflation, and then falls for two decades to a floor of two percent

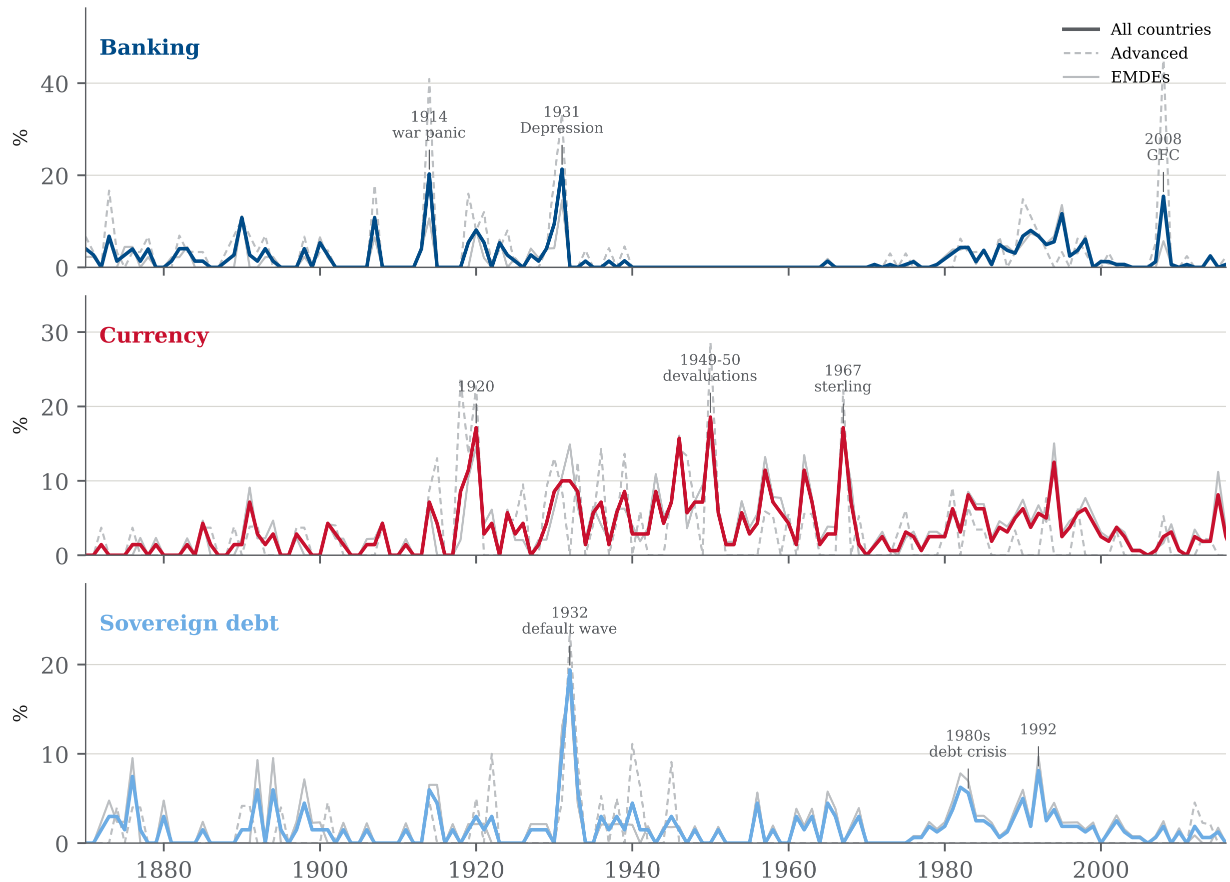


Figure 2: Share of countries in crisis, 1870 to 2016.

Notes: Percent of GMD countries with data in each year that are in each crisis type, for all countries (colored) and by income group (light lines). Source: Global Macro Database.

in the 2010s before settling near 4.4 percent in 2024. Today's configuration pairs debt near historic highs with rates near their long-run average, which leaves less measured room than at almost any point in the sample, and the spread across income groups is a first hint that no single ratio summarizes the room to respond.

Systemic crises keep coming. The GMD marks, for each country and year, whether the country was experiencing a systemic banking, currency, or sovereign debt crisis, drawing on the standard chronologies, including Laeven and Valencia (2020) and the historical dating surveyed in Reinhart and Rogoff (2009). Figure 2 shows the share of countries in each type of crisis in every year since 1870.

Banking crises arrive in global waves roughly once a generation. The share of countries affected peaks at 20 percent in 1914, 21 percent in 1931, and 15 percent in 2008, and the light lines show that these waves strike advanced economies and emerging market and developing economies (EMDEs) alike. The celebrated quiet of the Bretton Woods era turns out to be

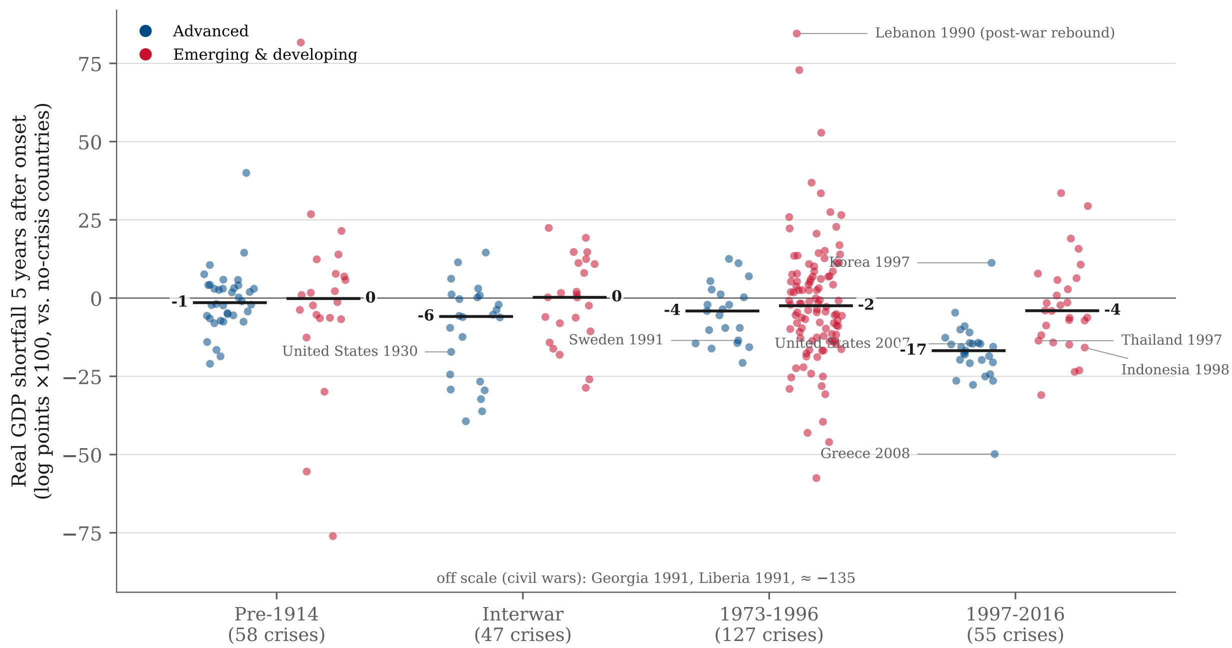


Figure 3: The heterogeneity to be explained.

Notes: Real GDP five years after each of 289 banking crises, world-war onsets excluded, relative to countries with no crisis in the same window, in log points, by era and income group. Black bars mark group medians. Source: Global Macro Database.

quiet only for banks. Currency crises spike at 19 percent of countries in 1950, reflecting the wave of parity realignments that followed the 1949 devaluation of sterling, and again in 1967 when sterling fell once more. Sovereign debt crises peak at 19 percent in 1932 and return in the EMDE debt crisis of the 1980s and 1990s, and the light lines show they have been almost exclusively an EMDE phenomenon since 1945. Repressed finance appears to have suppressed one kind of crisis and rerouted the pressure into others. The prudent baseline is that another wave will come, and the policy question is who will be able to respond when it does.

Aftermaths vary enormously, and the variance is what buffer policy has to explain. We date each banking crisis by the first year the chronologies record it, exclude onsets during the two world wars, and ask a simple question about each of the 289 resulting events. How did the economy perform over the following five years, compared with the countries that had no banking crisis at the same time? The answer, one dot per crisis, is Figure 3.

Two things stand out in Figure 3. The first is the sheer spread. Within every era, some economies end the five years well ahead of their peers while others fall 20 or more log points behind, with Korea's swift rebound after 1997 and Greece's lost decade after 2008 marking the extremes of the modern advanced group. The second is the drift of the black

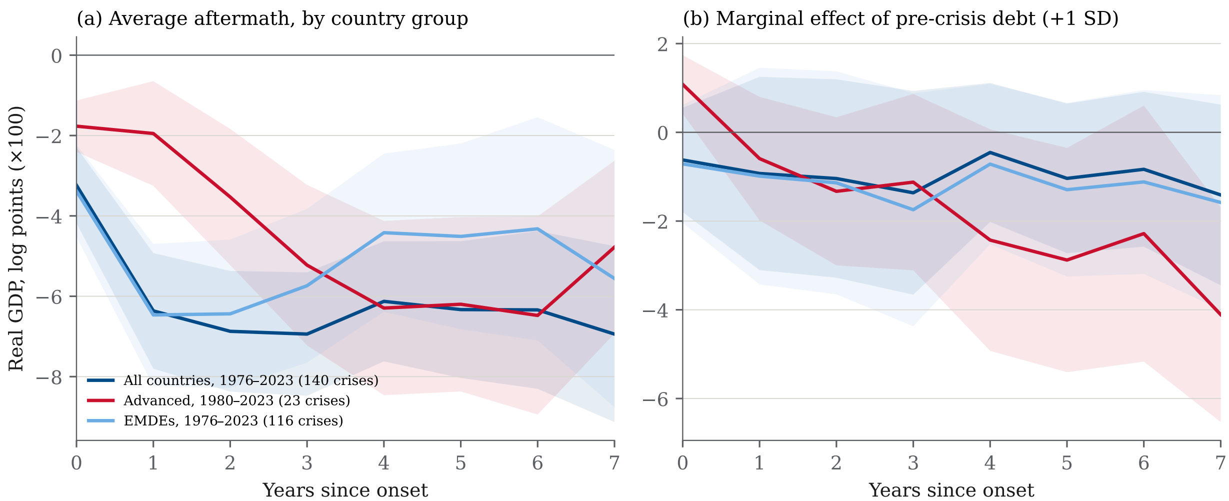


Figure 4: Crisis aftermaths and pre-crisis debt.

Notes: Average post-crisis path of real GDP (panel a) and the marginal effect of pre-crisis debt (panel b), by sample, in log points, with one-standard-error bands clustered by country. Sources: Global Macro Database; Laeven and Valencia (2026).

median bars. The typical advanced-economy crisis has grown far more costly over time, from roughly one log point before 1914 to 17 in the era since 1997, while the typical EMDE crisis has worsened more gently. Crises have become more expensive over time, alongside financial deepening (Bordo et al., 2001), and most expensive of all for the richest countries most recently.[1] Averages conceal nearly everything about crises. The case for buffers stands or falls on whether something observable before a crisis predicts where a country lands in this distribution.

Measured buffers predict less than the advice assumes. We estimate the local projections of Section 3 on the adopted chronology, on the pooled panel and separately by country group. Figure 4 reports the average aftermath in panel (a) and the marginal effect of pre-crisis debt in panel (b).

Panel (a) shows that a systemic banking crisis is expensive everywhere. Output falls six to seven log points below the no-crisis benchmark within two years and does not return within seven, with advanced crises deepening more slowly and EMDE crises striking harder on impact before a partial rebound. Panel (b) shows the pattern that motivates the paper. The point estimates for the debt interaction are small and sit well inside their one-standard-error bands at almost every horizon and in every group, so the debt ratio barely separates mild crises from disasters, and what structure exists appears late and mostly in the advanced group. Panel (b) should be read as suggestive, and the formal estimates with controls follow

[1] These patterns are not artifacts of the five-year window. Appendix B verifies them at three and seven years.

in Section 3. The weakness is exactly what the boom story predicts, since the low-debt busts of Ireland, Iceland, Spain, and crisis-era Korea sever the link between measured debt and true vulnerability. Any measure of policy space that leans on the debt ratio alone risks saying nothing about the outcomes it is supposed to insure against.

We make four contributions. First, the debt ratio predicts less, and predicts differently, than the advice assumes. On the newly updated Laeven and Valencia (2026) chronology, pre-crisis debt adds nothing to crisis prediction once boom symptoms, the rapid money growth, investment surges, and external deficits that mark a credit boom, are accounted for, and its association with severity is imprecise in most linear projections we run, under fixed effects, growth controls, a historically accurate income classification, and clean-control designs in Section 3. The signal that survives is narrow and specific. It is concentrated in advanced economies, sharpest above the 90 percent debt threshold, and consistent there with the narrative evidence.[2] In EMDEs the intervals rule out penalties of that size. A doubly robust cross-sectional design finds a larger advanced-economy estimate, which the appendix attributes to its different weighting of the 2008 cohort.

Second, booms hide the danger, and correcting for them moves the answer in the ratio's favor. Because credit booms can make debt ratios look artificially low exactly when crises approach, we compare crises that followed booms of similar size. Once that comparison is made, the advanced-economy estimates deepen and grow with the horizon. In this sense the second finding rehabilitates part of the first. The ratio is not meaningless. It is contaminated, and the signal that emerges once the boom is held fixed is the debt penalty of the narrative literature, only with wider bands. What fails is the raw ratio read on its own.

Third, the ratio predicts behavior even when it fails to measure resilience. Governments that entered a crisis with more debt consolidated sharply relative to low-debt peers in the aftermath, by more than four percentage points of GDP per standard deviation of debt among advanced economies. However, we cannot tell whether governments respond to the ratio itself or to the market pressure and program conditionality that travel with it.

Fourth, we turn the evidence into usable measures. The Policy Buffers Index ranks countries by observable buffers in a form anyone can audit, the Effective Space Index ranks them by resilience learned from crisis outcomes themselves, and the same estimates yield a first answer to what a buffer is worth in expected output.

The argument proceeds as follows. Section 2 describes the data, the adopted crisis chronology, and the historical income classification. Section 3 presents the empirical analysis, from whether debt predicts severity to what a buffer is worth. Section 4 estimates effective

[2] The 90 percent line traces to Reinhart and Rogoff (2010) and the critique of Herndon et al. (2014). The narrative evidence for advanced economies is Romer and Romer (2019).

policy space from crisis outcomes, and Section 5 builds the two indices of policy space. Section 6 concludes.

Related literature. A first strand documents the aftermath of financial crises. Cerra and Saxena (2008) establish that output losses after crises are large and persistent, Reinhart and Rogoff (2009) document the regularities across eight centuries, Romer and Romer (2017) measure aftermaths in advanced economies with a narrative distress series, Baron et al. (2021) re-date crises using bank equity returns, and Müller et al. (2025) show in the data we use that losses can persist for decades, and Claessens et al. (2012) document how financial-cycle conditions shape recession depth. This literature characterizes the average aftermath. Our focus is the variance around it, which is where the policy content lies. On the samples where the space advice originated, our estimates agree with it. The advanced-economy gradient we find is consistent in sign and broad magnitude with Romer and Romer (2019), only with the wider bands that a design-based specification carries.[3] What is new here is everything outside that sample, the EMDE null with intervals tight enough to exclude penalties of that size, the boom conditioning that moves both margins, and the behavioral margin.

A second strand connects crisis outcomes to the policy environment and asks what fiscal capacity is. Ostry et al. (2010) introduce the fiscal-space concept the modern debate uses. Romer and Romer (2018) and Romer and Romer (2019) show that monetary and fiscal space shape aftermaths in OECD economies, and Jordà et al. (2016) show that high public debt slows recoveries in 17 advanced economies. On measurement, Ghosh et al. (2013) define fiscal space from the distance to a debt limit implied by fiscal fatigue, Kose et al. (2022) assemble the leading cross-country database of fiscal space indicators from 1990 onward, Aizenman and Jinjarak (2010) propose a de facto fiscal-space measure that is a direct ancestor of our buffers index, and Brunnermeier et al. (2022) provide a theory of why a safe-asset issuer retains room at debt levels that would sink other sovereigns, the exorbitant privilege Gourinchas and Rey (2007) document, which is the reserve-currency wedge our index has to confront. That one ratio means different things in different countries is an old lesson. Reinhart et al. (2003) document debt intolerance at low ratios in serial defaulters, Eichengreen and Hausmann (1999) trace it to the currency composition of liabilities, and Arslanalp and Tsuda (2014) measure the investor base that composition implies. Reserves are the buffer this literature emphasizes for exactly those economies (Obstfeld et al., 2010; Aizenman and Lee, 2007), and the IMF's own multi-indicator framework (International Monetary

[3] Section 3.1 re-runs their comparison directly on our data and chronology. Splitting the advanced crises at their median entry debt, low-debt entrants suffer almost no five-year loss while high-debt entrants lose about five and a half log points, a contrast that matches theirs in direction and rough magnitude but is not significant on 23 events (Appendix Table A3).

Fund, 2018) reflects the same conclusion, which is why our buffers index carries reserve and market-access components alongside the fiscal ones. The measures closest to ours are the indicator dashboards of Kose et al. (2022) and International Monetary Fund (2018) and the de facto space measure of Aizenman and Jinjarak (2010), all of which aggregate what is observable. Our buffers index extends that logic to a within-year percentile spanning fiscal, monetary, external, reserve, and market-access margins on a much longer panel, and the Effective Space Index is, to our knowledge, the first policy-space measure whose weights are disciplined by realized crisis outcomes rather than chosen in advance. Relative to this strand, we show that the workhorse single-ratio metric predicts crisis outcomes poorly outside the samples where it was developed, and we propose and estimate measures that do better.

A third strand studies the booms that precede crises. Schularick and Taylor (2012) show that credit expansions are the most robust predictor of financial crises, Jordà et al. (2013) show that excess credit deepens the recessions that follow, which is the severity conditioning our boom controls implement with public-debt interactions added, and Mian et al. (2017) document the household-debt version worldwide. The claim that booms flatter fiscal positions has its own lineage. Morris and Schuknecht (2007) document the revenue windfalls that asset-price booms hand to budgets, and Borio et al. (2017) and Borio et al. (2016) argue that finance-neutral output gaps overturn assessments of cyclically adjusted positions precisely in boom years. Our contribution is to carry that logic into the crisis-severity regression, showing that controlling for boom exposure moves the estimated debt gradient exactly as the flattery mechanism predicts, which offers one explanation for the unstable gradients we document. On methods we draw on Jordà (2005) for local projections, Dube et al. (2023) and Jordà and Taylor (2016) for design-based corrections, Wager and Athey (2018) and Athey et al. (2019) for heterogeneous treatment effects, and Adrian et al. (2019) for the distributional lens.

2 Data

The core dataset is the Global Macro Database as of June 2026. We use real GDP, government debt, revenue, expenditure, and balances, the central bank policy rate with market rates as fallbacks, consumer price inflation, the dollar exchange rate and the real effective exchange rate, the current account, and the systemic banking, currency, and sovereign debt crisis indicators.

Crisis events. All estimation uses a single chronology, the updated database of Laeven and Valencia (2026), which records 164 systemic banking crises in 120 countries over 1970 to 2025, of which 14 are borderline episodes that we exclude. Table A5 in the appendix traces every step from these episodes to each estimation sample. We also test the design separately on other chronologies. The GMD's own multi-source flags reproduce the analysis over a broader event set in Appendix B, and the currency and sovereign chronologies receive the same treatment in Section 3.3. Appendix B details the construction and verification of the episode list and its relation to the GMD's own multi-source crisis flags. Those flags serve the descriptive figures of the introduction and two clearly labeled auxiliary exercises where longer coverage matters, the currency and sovereign extension of Section 3.3 and the buffers-index validation of Section 5. In the historical figures, onsets during the two world wars, 1914 to 1919 and 1939 to 1947, are excluded. Appendix A reports a full event-level audit reconciling the two lists.

Historical income classification. The World Bank's own analytical history (OGHIST) classifies countries from 1987 onward. For earlier years we backfill with a transparent rule. We compute each country's real GDP per capita relative to the United States in the same year, calibrate the three class boundaries on the pooled 1987 to 1990 overlap so that the rule best reproduces the official classes, and apply the calibrated cutoffs, at 4, 14, and 45 percent of United States income per capita, to every year back to 1870. The results match contemporary assessments in the cases we checked. Argentina classifies as high income in 1910, Russia as lower middle income in 1994, and Chile as upper middle income in 1981, while Korea counts as high income at its 1997 crisis by the World Bank's own contemporary assessment. Countries the rule cannot reach fall back to the current classification recorded in the GMD.

Exchange-rate regimes. We merge the Ilzetzki et al. (2019) annual coarse classification for 1946 to 2016, carrying the 2016 category forward, and define a peg indicator for the two most rigid categories. The regime data measure monetary space under pegs in the mechanism tests and enter the state vector of Section 4. Appendix D cross-validates both indices against the Kose et al. (2022) fiscal space database.

3 Empirical analysis

This section presents the empirical designs and their findings together. Throughout, events are the 140 crises of Laeven and Valencia (2026) that meet the estimation data requirements, outcomes come from the GMD, and standard errors are clustered by country, with estimation details in Appendix B.

3.1 Does pre-crisis debt predict crisis severity?

Let $y_{i,t}$ denote 100 times the log of real GDP in country i and year t, and let $D_{i,t}$ equal one in the first year of a systemic banking crisis. The average aftermath is estimated from the panel local projection of Jordà (2005),

$$y_{i,t+h} - y_{i,t-1} = \alpha_i^h + \gamma_t^h + \beta^h D_{i,t} + \sum_{k=1}^{2} \theta_k^h \, \Delta y_{i,t-k} + \sum_{k=1}^{2} \phi_k^h \, D_{i,t-k} + \varepsilon_{i,t}^h, \qquad h = 0, 1, \ldots, 7, \tag{1}$$

where α_i^h and γ_t^h are country and year fixed effects, so the crisis path is identified by comparing crisis countries with countries that had no crisis in the same year. We want to know how this path varies with the state a country was in when the crisis hit. With $S_{i,t-1}$ the standardized debt ratio at the end of the previous year, the interacted design of Romer and Romer (2019) is

$$\begin{aligned} y_{i,t+h} - y_{i,t-1} &= \alpha_i^h + \gamma_t^h + \beta^h D_{i,t} + \delta^h \left(D_{i,t} \cdot S_{i,t-1}\right) + \kappa^h S_{i,t-1} \\ &\quad + \Phi^h \, B_{i,t-1} + \Psi^h \left(D_{i,t} \cdot B_{i,t-1}\right) + \sum_{k=1}^{2} \theta_k^h \, \Delta y_{i,t-k} + \sum_{k=1}^{2} \phi_k^h \, D_{i,t-k} + \varepsilon_{i,t}^h. \end{aligned} \tag{2}$$

The boom-exposure vector $B_{i,t-1}$ collects the footprints a credit boom leaves in the data before the bust, rapid growth of money relative to GDP over five years, a surge in the investment share, and a widening external deficit, with exact definitions and summary statistics in Appendix D.[4] Because the symptoms enter both directly and in interaction with the onset, δ^h compares crises that followed booms of similar size, so a low ratio is neither credited for safety that belongs to a small boom nor blamed for damage that belongs to a large one. Throughout, we treat the boom symptoms as confounders to hold fixed rather than mediators to explain away, so conditional estimates are direct associations given the boom and unconditional ones include whatever debt shares with it. Two further designs guard against known failure modes, a clean-control variant in the spirit of Dube et al. (2023) that drops post-crisis years from the control pool and the doubly robust estimator of Jordà and Taylor (2016) that reweights events and controls to balance on the pre-crisis state, with alternative boom measures in Table A6 and the interaction under both designs in Table A9.

Figure 5 shows the answer. Once boom controls enter, the interaction is negative at short horizons in every sample and at every horizon among advanced economies, where it deepens from −1.6 log points per standard deviation at three years to −3.6 at five, while the pooled

[4] The credit gap of Drehmann and Juselius (2014) is the benchmark early-warning boom measure. Our proxies substitute for it where long credit series are missing, and Appendix B verifies the results on a credit-based vector where coverage allows.

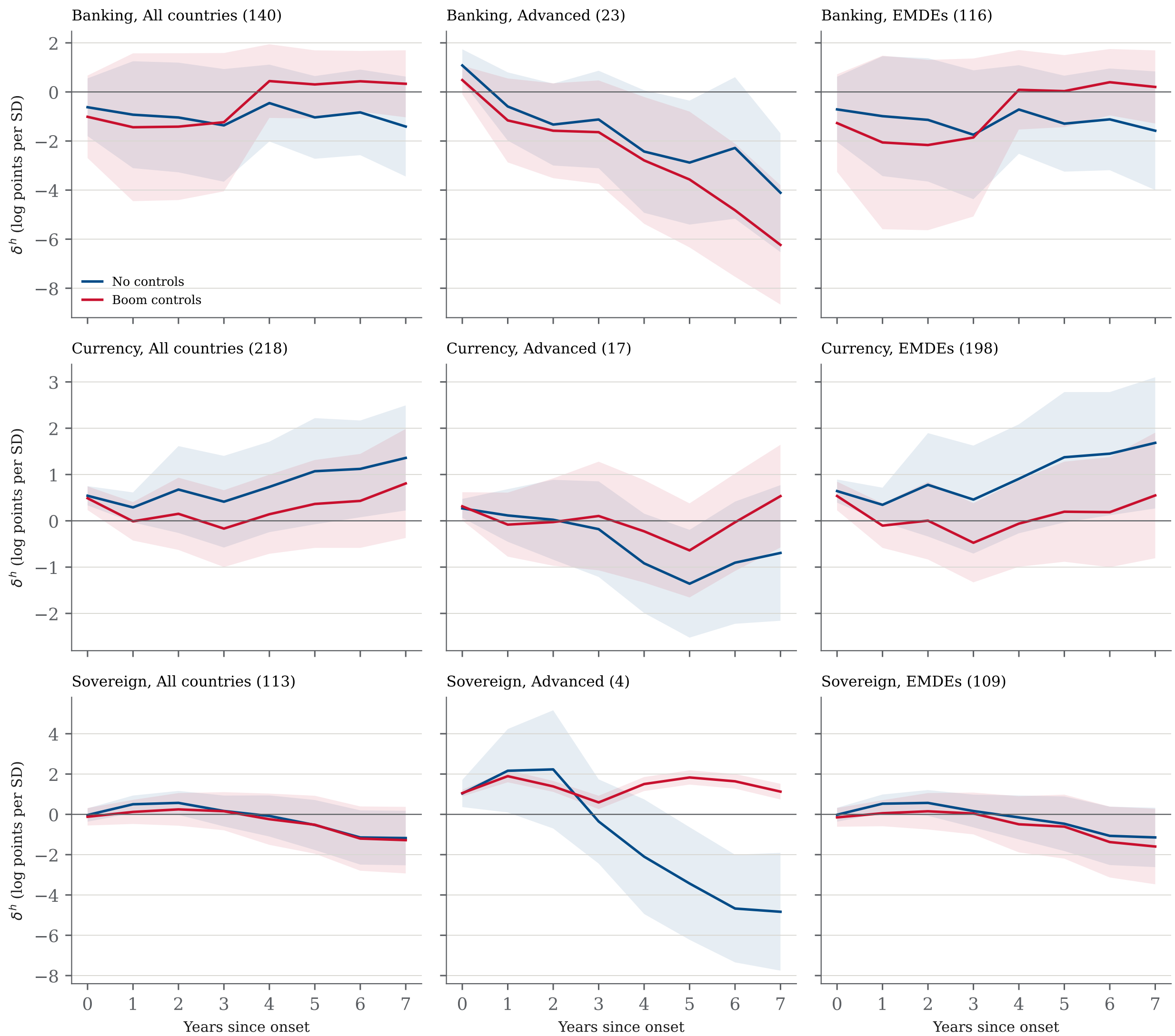


Figure 5: The debt interaction across crisis types.

Notes: The coefficient δ^h from equation (2) without controls and with boom controls. Rows are crisis types and columns are samples, with event counts in parentheses. Banking onsets follow the adopted chronology through 2023. Currency and sovereign onsets follow the GMD flags, which are verified through 2016. Bands are one standard error, clustered by country. The advanced sovereign panel rests on four events and is shown for completeness. The numeric counterparts, including the clean-control and doubly robust designs, are Tables A2, A9, and 2.

and EMDE estimates drift back toward zero at longer horizons. The clean-control design pushes the advanced estimates further, to −5.4 and −6.2, and the doubly robust estimate confirms that the average crisis is expensive, −9.7 log points of output over five years with a standard error of 5.1, as Appendix B reports. The consistency of sign and shape across all three designs is the substantive message, and it is exactly what the credit-boom logic predicts. The lower rows of the figure carry a preview of Section 3.3. For currency and sovereign onsets the interaction sits at or above zero in nearly every sample, so the penalty the banking row displays is specific to the crisis type the advice is implicitly about. The wide bands are themselves informative. At five and seven years the pooled and EMDE estimates are precise enough to rule out a penalty of the size the narrative aftermath literature implies, while the advanced-economy estimates remain consistent with one, and only the advanced seven-year cell approaches conventional significance under the most conservative inference we use.[5]

The closest antecedent makes the point sharply. Romer and Romer (2019) compare post-crisis paths across advanced-economy crises entered with low and high debt and read a large gap between them. Appendix Table A3 re-runs their headline comparison on our data and chronology, splitting advanced-economy crises at the median entry debt ratio of 43 percent of GDP, 12 low-debt and 11 high-debt events. The pattern reproduces. Low-debt entrants suffer almost no loss, −1.6 log points at five years and above trend by seven, while high-debt entrants lose 4.4 to 5.6 log points from three years on, matching the published contrast in direction and rough magnitude. The difference between the halves, however, is 2.8 to 5.1 log points from three years on, with standard errors of similar size, so it is not significant at any horizon. What changes on the broader data is not the shape of the comparison but its precision, and that is the motivation for everything that follows. The advice rests on a pattern that is real in point estimates, concentrated in advanced economies, and far less certain than its influence suggests.

Three checks establish that this pattern is what it appears to be (Table A7). The advanced-economy penalty is not the euro area in disguise. Excluding every crisis in a euro member leaves a similar and better determined path, and euro-membership or hard-peg interactions leave the debt term intact. The EMDE null is not a twin-crisis artifact of the kind Kaminsky and Reinhart (1999) document, since controlling for concurrent currency and sovereign crises moves nothing.[6] And the sharpest version of the penalty is the threshold form the advice actually uses, the 90 percent line that Reinhart and Rogoff (2010) made

[5] Appendix B contains the formal inference, the bootstrap and interval detail, the pre-trend and placebo checks, and comparisons with the broader multi-chronology event set.

[6] The foreign-currency share of sovereign debt in the Kose et al. (2022) database covers too few crisis years to estimate a denomination interaction.

Table 1: Debt and the policy response.

Sample	$h = 1$	$h = 3$	$h = 5$
Cyclically adjusted balance			
All countries (102 crises)	-0.47 (0.58)	0.10 (0.68)	-0.52 (0.42)
Advanced (20 crises)	-0.24 (1.81)	4.43*** (1.08)	2.67** (1.27)
EMDEs (81 crises)	-0.06 (0.63)	0.28 (0.76)	-0.30 (0.50)
Policy rate			
All countries (87 crises)	-0.40 (0.98)	-0.64 (0.51)	-0.78 (0.55)
Advanced (20 crises)	-0.24 (0.44)	0.41 (0.61)	0.37 (0.65)
EMDEs (68 crises)	-0.16 (1.21)	-0.68 (0.63)	-0.35 (0.87)

Notes: The debt interaction with policy responses as outcomes, boom-control specification, 1976 to 2023, in percentage points. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels. Bootstrap inference is in Appendix B.

famous and Herndon et al. (2014) made notorious. Crises in advanced economies entered with debt above 90 percent of GDP cost 8 to 15 log points more than those entered below, significant at every horizon, while pooled threshold effects stay imprecise.[7] The failure we document is therefore not the absence of a debt penalty in advanced economies. It is that the penalty is invisible to the linear ratio in most samples, absent in EMDEs, and absent at the onset margin.

3.2 Does pre-crisis debt predict the policy response?

Policy space is supposed to matter because it is used. So we ask a second question. Does the debt ratio predict what governments actually do once a crisis arrives? Here the data give a clear answer. We re-estimate equation (2) with policy responses as outcomes. The fiscal outcome is the change in the fiscal balance from $t-1$ to $t+h$ minus one half times the change in log output, the cyclical-adjustment approximation of Romer and Romer (2019), and the monetary outcome is the change in the policy rate, both trimmed at their 1st and 99th percentiles to remove hyperinflation-era artifacts.

We find that one standard deviation of pre-crisis debt predicts a cyclically adjusted balance of an advanced economy 4.4 percentage points of GDP higher three years after a crisis (Table 1).[8] The response echoes the narrative consolidation record of Guajardo et al.

[7] Four of the 21 advanced events entered above 90 percent, Greece and Italy in 2008, Japan in 1997, and Israel in 1983. No single episode carries the estimate. Dropping the events one at a time leaves the seven-year threshold coefficient between -10 and -18 and the euro-exclusion coefficient between -5 and -9 (Appendix B). A four-event cell is nonetheless better read as describing those episodes than as a universal threshold, in line with the evidence against common debt thresholds (Eberhardt and Presbitero, 2015; Chudik et al., 2017).

[8] The estimate is robust where it needs to be (Table A10). It moves little with the cyclical-adjustment

(2014) and Alesina et al. (2015) and the crisis-era austerity accounting of House et al. (2020). The implied paths are stark. A government entering with debt one standard deviation below the mean eases by about seven and a half percentage points of GDP, while one entering a standard deviation above tightens slightly. The policy-rate response is positive at three and five years but never distinguishable from zero, and we detect no EMDE response, a null that sits naturally with the procyclicality literature, in which EMDE fiscal policy is constrained in bad times regardless of the inherited ratio (Kaminsky et al., 2004; Frankel et al., 2013). It appears that the debt ratio reliably predicts what advanced-economy governments do in a crisis while only weakly measuring what the economy can withstand. We cannot tell from these regressions whether policymakers respond to the ratio itself or to the market pressure and program conditionality that accompany high debt, and if consolidation into a downturn is contractionary, part of any debt penalty could operate through the response rather than through structural vulnerability. An illustrative calculation sizes that channel. Applying multipliers between 0.5 and 1.5, the range the fiscal literature treats as standard for downturns (Ramey, 2019; Blanchard and Leigh, 2013; Auerbach and Gorodnichenko, 2012), to a consolidation of 4.4 percentage points of GDP implies an output effect of 2.2 to 6.6 log points, between a third and all of the seven-year advanced-economy severity gradient of 6.2, itself only marginally significant. Consolidations of this kind are exactly the ones Fatás and Summers (2018) find leave permanent scars, and Bianchi et al. (2023) supply the theory in which sovereign risk makes austerity in a downturn the constrained optimum. The behavioral channel could therefore carry much of the measured debt penalty, though this is an accounting statement rather than a mediation estimate.

3.3 Do other crises behave the same way?

Table 2 extends the analysis to currency and sovereign debt crises, which the banking-focused Laeven and Valencia (2026) update does not cover, using the GMD flags and equation (2) without boom controls. Currency crises are sharp but transitory, with output about two log points below the benchmark in the first year and back near it by the fifth, and their debt interaction is positive at long horizons though weakly determined. Sovereign crises are deep and persistent, and we do not interpret their debt interaction, since high debt enters the

elasticity, survives dropping Ireland and Iceland, whose crisis-era deficits carry enormous one-off bank recapitalizations, and survives excluding the 2008 to 2012 wave entirely, and it disappears at pseudo-onsets assigned to random no-crisis advanced economies, so it is crisis-specific. It also survives an IMF or EU program-participation control, entered in levels and interacted with the onset, at 3.7 with a standard error of 0.8 at three years, so conditionality does not absorb it (Table A10). Appendix B carries the details, the bootstrap inference, the caveats on program conditionality and sample selection, and the level paths behind the implied responses.

Table 2: Average paths and debt interactions for the other crisis types.

Coefficient	$h = 1$	$h = 3$	$h = 5$	$h = 7$
Currency crises (347 events)				
Average path β^h	-2.33** (0.95)	-1.17 (1.17)	-0.48 (1.33)	-0.16 (1.49)
Debt interaction δ^h	0.56 (0.38)	1.06 (0.98)	1.67 (1.11)	2.12* (1.14)
Sovereign debt crises (162 events)				
Average path β^h	-2.91*** (0.79)	-1.31 (1.20)	-1.36 (1.44)	-2.36 (1.69)
Debt interaction δ^h	0.49 (0.43)	0.22 (0.89)	-0.22 (1.35)	-0.75 (1.58)

Notes: Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

definition of a sovereign event. Whatever severity signal the debt ratio carries is specific to banking crises rather than generic to macroeconomic distress.

3.4 What is a buffer worth?

A buffer can pay off through two margins, by making crises less likely and by making them less costly, the same insurance arithmetic that Jeanne and Rancière (2011) apply to international reserves. Both margins follow from the expected discounted crisis loss

$$V(s) \;=\; \lambda(s)\, L(s), \qquad L(s) \;=\; \sum_{h=0}^{H} \rho^h\, \mathbb{E}[\text{output loss at horizon } h \mid \text{crisis}],$$

where s is measured policy space, defined as the negative of the standardized debt ratio $S_{i,t-1}$, and $\lambda(s)$ is the annual crisis probability. We set $\rho = 0.97$ for the discount factor and $H = 7$, the longest horizon the projections estimate. Differentiating the product and flipping the sign so that a gain is positive,

$$-\frac{dV}{ds} \;=\; \underbrace{-\frac{\partial\lambda}{\partial s}\, L}_{\text{probability margin}} \;\; \underbrace{-\,\lambda\,\frac{\partial L}{\partial s}}_{\text{severity margin}} . \tag{3}$$

Every piece of equation (3) is estimated. The expected output loss of a crisis country at horizon h is $-(\beta^h + \delta^h S)$ from equation (2), so under $s = -S$ the severity derivative is $\partial L/\partial s = \sum_h \rho^h \delta^h$, taken from the boom-controlled pooled rows of Table A2. The level L is the discounted sum of the average-path coefficients β^h of equation (1). The probability pieces come from the logit

$$\Pr\left(D_{i,t} = 1 \mid \cdot\right) \;=\; \Lambda\!\left(a_{d(t)} + b\, S_{i,t-1} + c' B_{i,t-1}\right), \tag{4}$$

Table 3: The two margins of buffer value.

Component	Estimate
Annual crisis probability λ	1.8%
Discounted output loss per crisis L	42 log points
Effect of 10pp lower debt on λ, boom-conditional	0.000pp
Effect of 10pp lower debt on λ, raw cross-decade	-0.027pp
Severity gradient per 10pp lower debt, $-\sum_h \rho^h \delta^h$	0.70 (1.09) log points
Implied value, probability margin, boom-conditional	0.02 bp of GDP per year
Implied value, probability margin, raw upper bound	1.1 bp of GDP per year
Implied value, severity margin	1.2 (1.9) bp of GDP per year

Notes: All entries are per 10 percentage points of lower debt. The probability rows transform the logit margins of Table A11, which reports their standard errors, and the severity row rescales the pooled boom-controlled δ^h to ten points of debt using the pooled standard deviation of 51 points. Standard errors in parentheses where available.

with decade intercepts $a_{d(t)}$, estimated with and without each block of regressors in Table A11, so λ is the sample hazard and $\partial\lambda/\partial s$ is the average marginal effect of debt. Table 3 assembles the pieces for a ten percentage point reduction in the debt ratio, and each of its rows maps to one of these estimates.

The probability margin depends on what one conditions on, and Table A11 reports the onset regressions in full, with average marginal effects, clustered standard errors, and discrimination statistics by sample. Debt carries little onset signal of its own. Within decades, ten points more debt is associated with a 0.006 percentage point higher annual crisis frequency, with a standard error twice that size, and adding debt to the boom symptoms leaves the area under the ROC curve unchanged at 0.73 while the boom symptoms alone reach it, a comparison Figure A7 in the appendix draws curve by curve. A raw association does appear when decade-level crisis waves are left uncontrolled, in line with the early-warning evidence (Gourinchas and Obstfeld, 2012), worth about 0.03 points per ten points of debt and up to a basis point of GDP per year in the valuation, and Table 3 carries it as an upper bound. The probability margin of a debt buffer is therefore at most about one basis point per year and plausibly nothing, consistent with the boom symptoms carrying most of the ratio's onset signal.[9] The severity margin is worth about one basis point per year with a standard error of two. Even at the upper bound, the two margins together value a ten point debt buffer at roughly two basis points of GDP per year. The ingredients behind that number are individually indistinguishable from zero, and that is not a defect of the exercise but its finding. The valuation prices exactly what the data support, so near-zero margins translate

[9] The null is not an artifact of the boom proxies or the specification. With the boom vector built from WDI private credit to GDP, the debt margin is 0.026 percentage points with a standard error of 0.018, and a linear probability model with country and year effects puts it at -0.008 (0.037), per Appendix B.

into a near-zero value, with the reported standard errors carrying the uncertainty rather than concealing it. Whatever stronger case exists rests in the tails, where quantile projections in Appendix B put the point estimates in the worst tenth of outcomes though with wide bands, and on the behavioral channel of Section 3.2, since the most reliably estimated correlate of entering a crisis with high debt is the consolidation that follows. Put simply, buffers are worth the most when they can be used. The case is weakest for holding a particular number and strongest for having room that does not force tightening in a downturn.

4 Effective policy space

The projections of equation (2) do their job, but they answer one question at a time. Each coefficient summarizes how crisis cost varies with a single conditioning variable, linearly, holding the others fixed, and Section 3 shows that the answer differs by income group, by threshold, and by boom exposure. What we want next is the expected cost of a crisis as a flexible function of a country's entire pre-crisis situation. Formally, the object is the conditional average treatment effect of a crisis,

$$\tau(x) \;=\; \mathbb{E}\left[\, y_{i,t+5} - y_{i,t-1} \,|\, D_{i,t} = 1,\, X_{i,t-1} = x\right] - \mathbb{E}\left[\, y_{i,t+5} - y_{i,t-1} \,|\, D_{i,t} = 0,\, X_{i,t-1} = x\right], \quad (5)$$

which answers the question, for a country whose pre-crisis situation is described by x, of how much five-year output a banking crisis would be expected to cost. The vector x collects thirteen inputs, the debt ratio, the fiscal balance, the policy rate in nominal and real terms, the external balance, foreign reserves over GDP, the government bond spread over the United States as a market-access measure, the three boom symptoms, the exchange-rate regime, the income group, and inflation. The predicted cost of a crisis is $-\hat{\tau}(x)$, and effective policy space is the predicted effect itself, $\mathrm{ES}_{i,t} = \hat{\tau}(X_{i,t})$, higher when the predicted crisis is milder. The state is measured in year t, so $\mathrm{ES}_{i,t}$ prices a crisis arriving the following year.

We estimate $\tau(x)$ with the honest causal forest of Wager and Athey (2018) and Athey et al. (2019). The method proceeds in two stages. The first stage removes the predictable parts of both the outcome and the crisis indicator by fitting flexible models $\hat{m}(x) = \hat{\mathbb{E}}[Y \mid x]$ and $\hat{e}(x) = \hat{\Pr}(D = 1 \mid x)$ and taking residuals, the double machine learning construction of Chernozhukov et al. (2018), which turns equation (5) into the partially linear moment

$$Y_{i,t} - \hat{m}(X_{i,t-1}) \;=\; \tau(X_{i,t-1})\,(D_{i,t} - \hat{e}(X_{i,t-1})) + \varepsilon_{i,t}, \quad (6)$$

so that the remaining variation mimics a comparison between otherwise similar country-years with and without a crisis. The second stage grows many trees over the residualized data and solves, at each query point x, the locally weighted version of equation (6),

$$\hat{\tau}(x) \;=\; \frac{\sum_i \alpha_i(x)\;(Y_i - \hat{m}(X_i))\,(D_i - \hat{e}(X_i))}{\sum_i \alpha_i(x)\,(D_i - \hat{e}(X_i))^2}, \tag{7}$$

where the weights $\alpha_i(x)$ count how often observation i falls in the same leaf as x across trees. Each tree uses one half of its sample to choose splits and the other half to estimate within leaves, the honesty device that prevents the forest from certifying patterns it invented, and the averaged $\hat{\tau}(x)$ carries valid confidence intervals. These estimates carry a causal reading only if selection into crisis is captured by the observables in x, and unmeasured institutions or politics that both trigger crises and shape recoveries (Funke et al., 2016) would bias them, one more reason the text treats the fitted effects as predictive. Nuisance models are cross-fit in country-grouped folds so that no country's own outcomes inform its predictions, and hyperparameters are chosen by rolling-origin validation, in which the forest is fit on all events through year Y and scored on the events of the following five years, rolled forward from 1990 to 2015. A wider feature set loses this contest to the thirteen-input baseline, a reminder that with 139 events parsimony is protective. Full estimation detail is in Appendix B.

This estimator suits the problem better than the alternatives for three reasons. Linear interactions, as in Section 3, impose that one standard deviation of debt means the same thing everywhere, which is the hypothesis under examination rather than a fact to assume, while kernel and local methods that would relax it break down over a thirteen-dimensional state vector. Off-the-shelf machine learning applied directly to outcomes would confound the effect of a crisis with selection into crisis, since the same booms that predict crises predict their aftermaths, whereas the orthogonalized first stage mimics a comparison between observably similar country-years and thereby isolates the event itself. And unlike most flexible learners, the honest forest delivers confidence intervals for each prediction and an importance decomposition for each input, which is what an index needs.

Three results summarize where the effective-space estimates stand. First, the fitted effects carry signal, with the held-out validation below clearing conventional significance across countries and a calibration test in Appendix B consistent with sizable if imprecisely measured heterogeneity, and Table 4 reports how much each input contributes. Each input's share of importance is the frequency with which the forest splits on it, weighted by how early in the tree the split occurs, normalized to sum to one across inputs. Reserve and external positions dominate. The ratio of reserves to GDP carries the largest share at 18 percent, the external balance and its three-year average together carry about a quarter, and inflation and the debt ratio follow at about an eighth each with the fiscal balance under a tenth. The ordering is corroborated rather than idiosyncratic. External and reserve positions are the

Table 4: Forest importance shares by crisis type.

Input	Banking	Currency	Sovereign
External balance	12%	15%	14%
Reserves / GDP	18%	15%	8%
External balance, 3y avg	12%	14%	13%
Debt / GDP	12%	12%	35%
Investment change, 5y	8%	10%	6%
Inflation	12%	10%	6%
Fiscal balance	9%	9%	4%
Money growth, 5y	5%	6%	5%
Policy rate	6%	4%	3%
Real policy rate	3%	3%	3%
Pegged regime	0%	1%	0%
Advanced economy	1%	1%	1%
Bond spread over US	1%	0%	1%
Events	139	216	112

Notes: Share of split importance for each input, by the crisis type the forest is trained on. The banking column uses the adopted Laeven and Valencia (2026) chronology, 1970 to 2019. The currency and sovereign columns use the GMD flags over their verified window, 1970 to 2016, with the same thirteen inputs and settings.

leading predictors in the early-warning literature (Gourinchas and Obstfeld, 2012; Frankel and Saravelos, 2012; Catão and Milesi-Ferretti, 2014), and machine-learning crisis prediction reaches a similar ranking (Bluwstein et al., 2023; Fouliard et al., 2021). The ordering is also crisis-type specific in an instructive way. Fitting the same forest to the GMD currency and sovereign crisis flags reorders the inputs, with the debt ratio's share rising from about an eighth for banking crises to a third for sovereign crises, where debt is the crisis itself, and the external balances and the investment boom carrying more of the weight for currency crises (Table 4). Different crisis types warrant different buffers, which is itself an argument against a single all-purpose ratio. Figure 6 traces the three crisis-specific scores through the same four histories, and Figure A5 in the appendix maps them across countries in 2019, where the three scores are uncorrelated or negatively correlated with each other, banking against sovereign at -0.38. The banking ordering also matches the rest of our evidence, since the inputs that predict severity are plausibly the ones that bind regardless of the policy response, and it explains in one line why no debt threshold can summarize the room to respond.

Second, the fitted model turns the weights into usable gradients. Table 5 reports the average predicted five-year crisis cost for paired pre-crisis conditions, with the within-pair gap in the last column, so the margins that separate mild crises from deep ones are visible at a glance. The external margin is by far the widest. Countries in the weakest external-

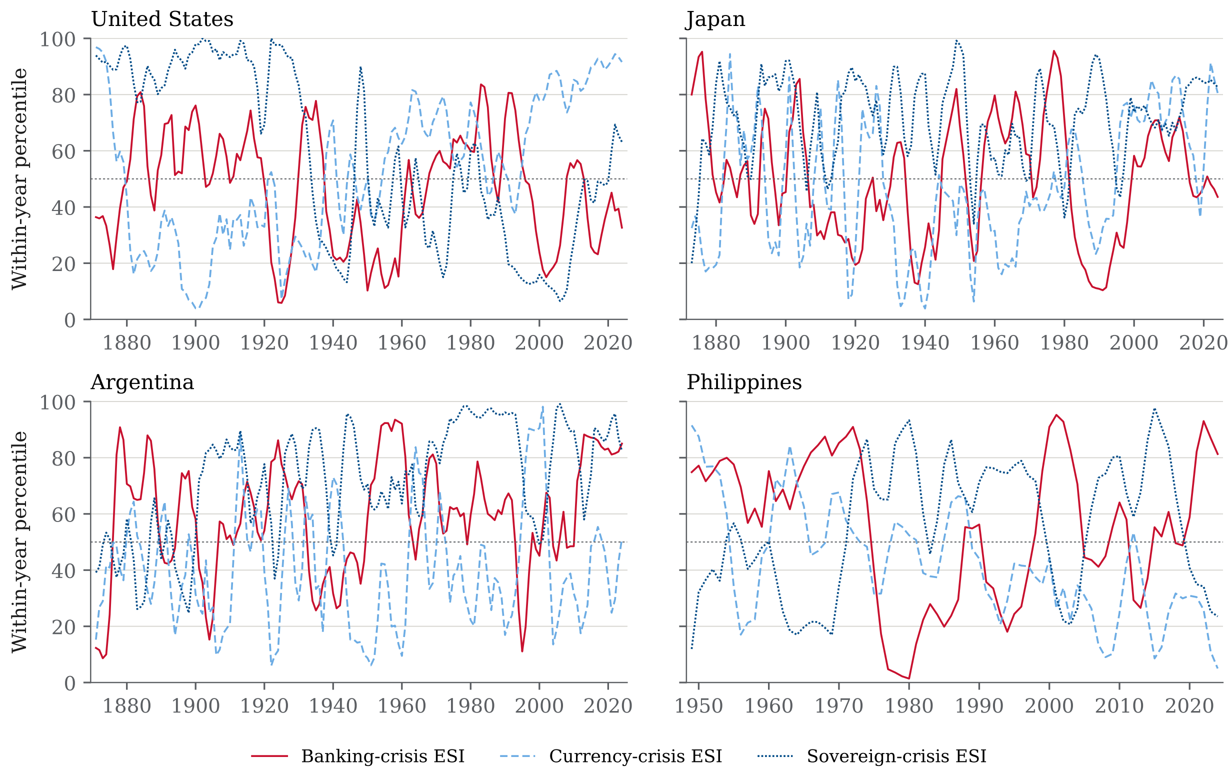


Figure 6: Effective space by crisis type, four histories.
Notes: Within-year percentiles of the effective-space scores from forests trained separately on the adopted banking chronology and on the GMD currency and sovereign flags, for the same four economies as Figure 8. Higher percentiles mean more effective space against that crisis type. All series are three-year centered moving averages of the annual percentiles. Source: author's calculations.

balance tercile face predicted losses of 6.3 log points against 3.5 in the strongest, a gap of 2.9 before rounding, and pegged regimes face about a point more than floats. The debt pair is computed at similar boom exposure, averaging within boom-symptom terciles, because the raw comparison is confounded by the fact that booming economies carry both low debt ratios and high predicted losses. Once that adjustment is made, the gap between the highest and lowest debt terciles is less than half a log point, the fitted model reproducing the regression finding that debt by itself separates little. These gradients, not any single ratio, are what a useful summary of policy space has to aggregate.

Third, the model carries positive out-of-sample information in both dimensions that matter, and reserves are part of the reason. Predicted crisis costs correlate with realized five-year shortfalls at a rank correlation of 0.20 for countries held out of the fit, significant under permutation inference, and 0.19 for events predicted strictly from earlier data across 94 rolling-origin predictions, just short of significance. Without reserves in the state vector the held-out country correlation drops to about 0.14. The cross-country correlation clears

Table 5: Predicted crisis costs by pre-crisis condition.

Condition	Predicted five-year loss		Gap
Income group	EMDEs: 4.8	Advanced: 4.8	-0.0
Exchange-rate regime	Floating: 4.2	Pegged: 5.1	+0.9
External balance tercile	Strongest: 3.5	Weakest: 6.3	+2.9
Debt tercile, at similar boom exposure	Bottom: 4.4	Top: 4.8	+0.4

Notes: Average predicted five-year cost for paired conditions, in log points of output, with the within-pair gap in the final column. The debt-tercile row averages within boom-exposure terciles to compare at similar boom intensity.

conventional significance, the cross-period one narrowly misses it, and the calibration test above is consistent with real but imprecisely measured heterogeneity, so we present the estimated measure as an informative companion whose country-level calls deserve skepticism rather than as a validated ranking. The noise has a size. The average 68 percent interval on a 2019 score spans 7.9 log points against a cross-country interquartile range of 3.6, so within-year ordering is dominated by estimation error, and year-to-year rank persistence is 0.72 against the buffers index's 0.83. To make it usable alongside the buffers index, we define the Effective Space Index as the within-year percentile rank of $\mathrm{ES}_{i,t}$, which Section 5 takes up. Scored on the 2019 cross section, it places the United States 122nd of 154 countries against the buffers index's 139th of 144, and between 70th and 122nd across the forest configurations we examined, reflecting inputs the balance sheet cannot see, though the estimated rankings remain specification-sensitive and we publish them as a companion to the buffers index rather than a replacement. Scoring any single cross section also leans on the fitted model at the edge of its training window, so published values inherit model error beyond what the validation measures.

5 Two indices of policy space

We propose two indices, named for what each one measures. The Policy Buffers Index (PBI) aggregates observable buffers with equal weights and no estimation, so that anyone can audit it. The Effective Space Index (ESI) is the effective-space measure of Section 4, expressed on the same percentile scale. Keeping them apart is deliberate. Measured buffers and effective space are different objects, and the gap between the two indices is where the message of the evidence lives. This section constructs the buffers index and compares the two.

Three features of the data shape every ranking that follows, and they belong up front. Coverage is uneven, since a country is ranked once it reports four of the six components, so economies with sparse statistics can rank on partial information, and the spread component in particular is missing for much of the sample. The indices reward what is measured. Large reported stocks of reserves and fiscal surpluses lift commodity exporters, while capacities no input measures, reserve-currency status, swap lines, deep local investor bases, and institutional strength, count for nothing, which mechanically favors some countries over others. And inputs arrive with error and reporting lags that are largest outside advanced economies, so neighboring ranks should not be read as meaningful distinctions.

The buffers index aggregates six components, the inverse debt ratio, the fiscal balance, the real policy rate, the current account balance, foreign reserves over GDP from the World Bank, and the inverse long-rate spread over the United States as a market-access measure, each signed so that a higher value means more room to respond. Because the components carry different units and heavy tails, we convert each to a within-year percentile rank before averaging, the standard rank-based normalization for composite indicators (Nardo et al., 2008), which is robust to outliers and makes scores comparable across eras by construction. Formally, let $z_{i,t}^k$ be the value of component k for country i in year t, let N_t^k be the number of countries reporting component k in year t, and let $\text{rank}(z_{i,t}^k)$ be the ascending rank of country i among those reporters, from 1 for the least room to N_t^k for the most. The component score and the index are

$$p_{i,t}^k = 100 \cdot \frac{\text{rank}\left(z_{i,t}^k\right) - 1}{N_t^k - 1}, \qquad \text{PBI}_{i,t} = \frac{1}{K_{i,t}} \sum_{k \in \mathcal{K}_{i,t}} p_{i,t}^k, \tag{8}$$

where $\mathcal{K}_{i,t}$ is the set of components available for country i in year t and $K_{i,t}$ is its size, with a minimum of fifteen reporters per component and four of the six components required. Equal weights are the neutral benchmark recommended when no agreed importance ordering exists (Nardo et al., 2008), and multi-indicator aggregation follows the fiscal-space practice of Kose et al. (2022) extended to monetary and external dimensions. The Effective Space Index changes both the recipe and the ingredients, replacing the equal-weighted average with a forest prediction estimated from crisis outcomes over a wider state vector.

Figure 7 maps the index at four dates, and two things stand out. The first is persistence. In every panel the dark band runs through fiscal-surplus commodity exporters. Kuwait, Oman, Venezuela, and Libya lead in 1980 much as the Gulf leads today, so measured room has always been concentrated where resource revenues sit. The second is the long fade of the advanced world. The United States slides from the 58th percentile in 1980 to the 46th in 1990, the 29th in 2010, and the 23rd in 2024, as debt, deficits, and a thin reserve stock come to dominate its score, and by the last panel much of the lightest shading sits in the largest advanced economies and in low-income economies with thin buffers on every dimension. Table A14 lists the leaders and laggards by population size and income group in 2019, the

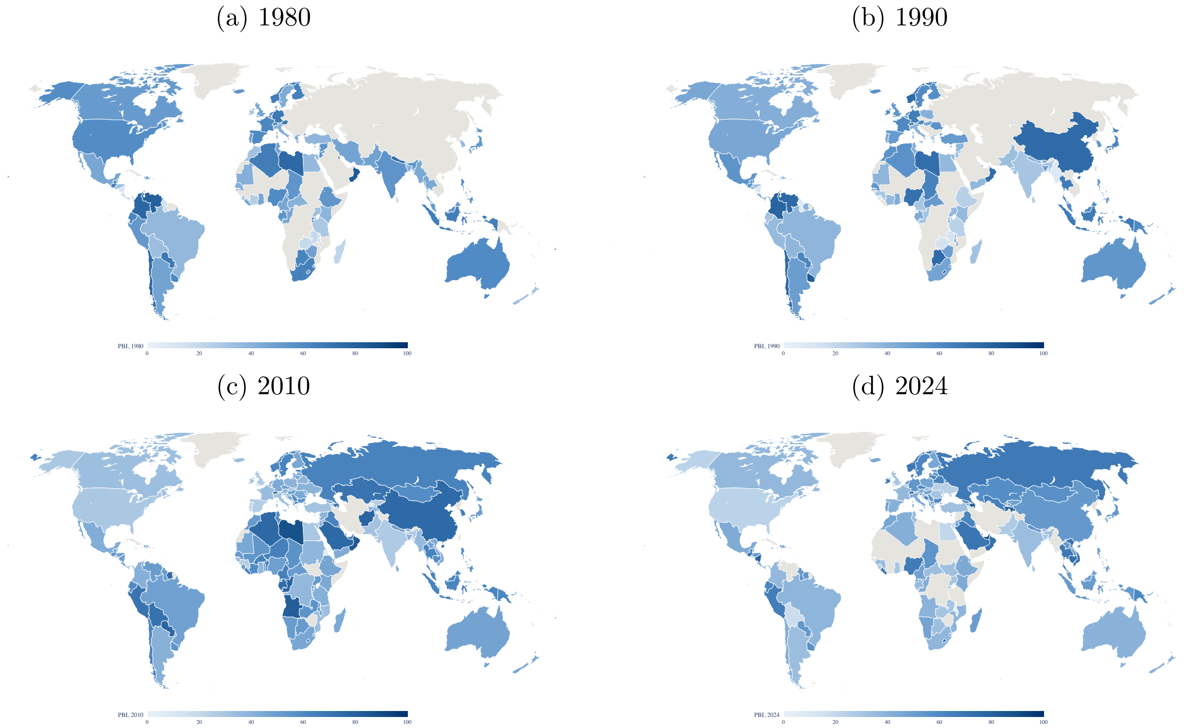


Figure 7: The Policy Buffers Index across four decades.

Notes: Within-year percentiles in 1980, 1990, 2010, and 2024. Darker shading indicates more room to respond, and gray indicates insufficient data. Source: author's calculations.

last pre-pandemic year, which serves as the benchmark cross section for every ranking exhibit. The top carries a coverage caution too. The United Arab Emirates, Iraq, and Cambodia are ranked on four reported components each, so their scores lean on the coverage rule as well as on genuine strength. The reserve and market-access components do visible work. Japan, nearly last on fiscal components alone, sits mid-table once its reserve stock and negative spread count, and Denmark and Switzerland join the commodity exporters near the top. Component-by-component detail for selected countries and maps of each component and of the Effective Space Index, all on the same 2019 cross section, are in Figures A3, A4, and A8 of Appendix C. Two failures of the buffers index are deliberate exhibits. Russia sits in the top tenth on low debt and a large reserve stock while the sanctions regime in place since 2014 already constrained exactly the market access an effective measure must capture, and the United States at 139th of 144 is mechanically correct yet hard to credit for the issuer of the world's reserve currency. The reason the United States sits near the bottom of both columns is mechanical rather than mysterious. It pairs a bottom-decile debt score and persistent fiscal and external deficits with a measured reserve stock near zero relative to GDP, because a country that issues the currency others hold as reserves has little need to hold reserves itself, so the inputs the indices reward are precisely the ones the United States alone can do without. The Effective Space Index narrows these gaps, lifting the United States to 122nd of 154 under the adopted configuration and near the middle of the distribution under the wider input sets, because it weights the reserve, external, and boom conditions that historically predicted crisis outcomes in our estimates. The distance that remains is consistent with the reserve-currency wedge of Brunnermeier et al. (2022), which no input measures, since the forest carries no denomination or reserve-currency indicator.

Table A13 in the appendix lists both indices side by side for every ranked country. Across the 143 countries both indices cover, the rank correlation between them is -0.01. We do not read that number as proof that buffers and resilience are unrelated, because a noisy ranking is uncorrelated with everything, and a split-half check shows the noise directly. Two forests fit on disjoint halves of the training countries produce 2019 rankings that correlate at 0.09 with each other, so the Effective Space Index carries no measurable country-level reliability at this sample size, and its information lives in the aggregate objects, the weights, the insight gradients, and the held-out validation, not in individual ranks. The disagreement between the indices is not an artifact of the benchmark year. Across years their within-year rank correlation averages 0.04 and never exceeds 0.27, the median within-country correlation over time is 0.02, and in 2019 only 26 percent of countries sit in the same quartile of both. Measured buffers and estimated resilience move almost independently at every cut, though the reliability numbers above counsel care in judging how much of that is concept and how much is noise. The largest divergences are still instructive, and we display rather than curate them. The buffers index puts Singapore, Kuwait, and Uzbekistan near the top where the Effective Space Index puts them near the bottom, and it puts Kenya and Yemen near the bottom where the estimated index puts them near the top. Figure A6 in the appendix dissects the extremes of the estimated ranking input by input on the eve of the global financial crisis, where Spain and Switzerland sit near the bottom of the 2007 ranking ahead of their 2008 distress and the widest calls rest on the most extreme inputs. The starkest caveat comes from scoring the model on the most recent data rather than the benchmark year. Scored on 2024 inputs, the Effective Space Index ranks wartime Ukraine first, and the anatomy of that call is the index's caveats in one row. Ukraine's 2024 inputs pair above-median reserves, swollen by official wartime financing, with a worse-than-median external balance, missing market-access data imputed at the sample median, and debt well above the median, a configuration outside the training support. The index has no conflict input and no way to read official financing as distress rather than strength, so a call no analyst would endorse sits at the top of the column, and wartime or program economies should be read off it with particular care. Both sets of failures argue for publishing the two indices together, with the buffers index as the auditable summary of what governments hold and the Effective

Space Index as a disciplined reading of what those holdings have historically bought.

External validation supports the ingredients and sharpens the limitation. Table A12 correlates the indices and their components with the Kose et al. (2022) fiscal space database. Our debt and balance components match theirs at rank correlations of 1.00 and 0.97, so the raw material agrees across sources. Against their balance-sheet composition indicators neither index correlates at more than 0.24 in absolute value, a reminder that no denomination input enters either index, so composition risk of the Eichengreen and Hausmann (1999) type is precisely what both indices still miss.

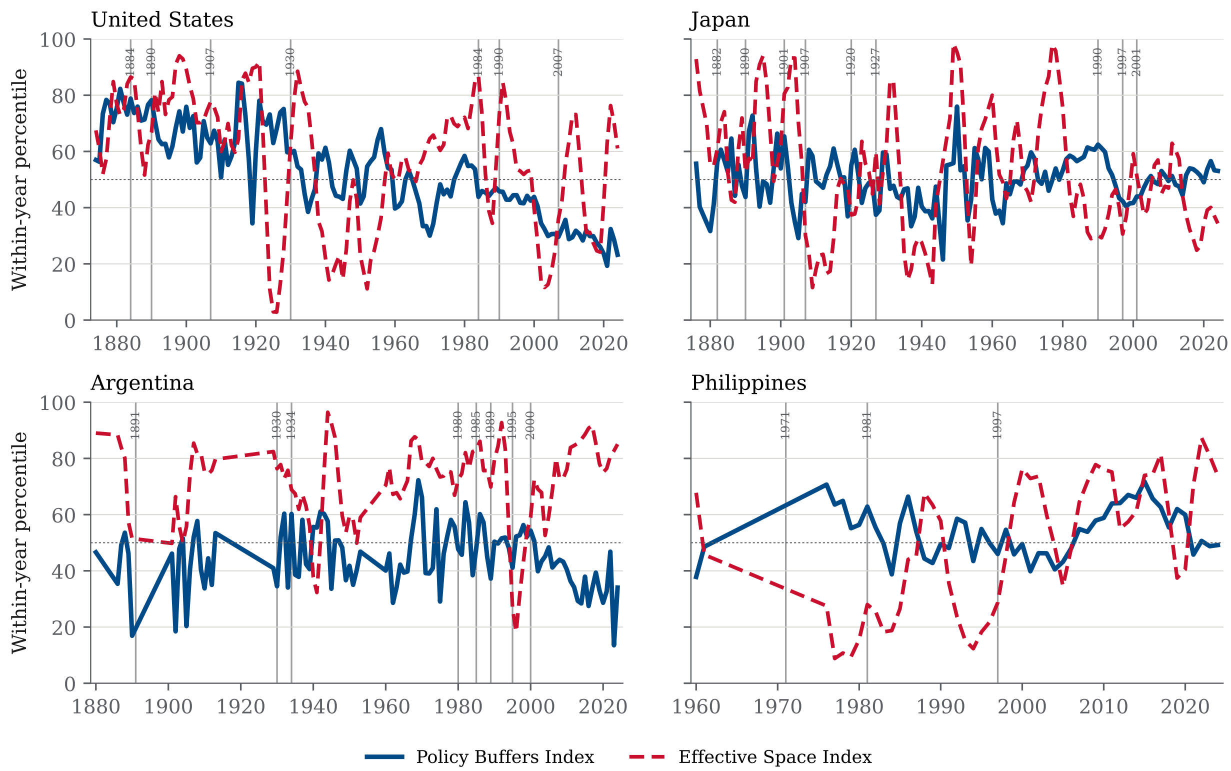


Figure 8: Two indices, four histories.

Notes: The Policy Buffers Index (solid) and the Effective Space Index percentile (dashed) for the United States, Japan, Argentina, and the Philippines, with vertical lines at GMD-flag banking-crisis onsets and the horizontal dashed line at the 50th percentile. Higher percentiles mean more room on both indices. The Effective Space Index series are three-year centered moving averages of the annual percentiles. Source: author's calculations.

The buffers index behaves sensibly in the four histories we plot, though an index collapsing around a crisis partly reflects the crisis itself depleting buffers. Figure 8 shows Japan's 1997 crisis arriving after a slide from the 62nd percentile in 1990 to the 42nd, the United States entering the 1930s and 2007 with above-median space and spending the post-2008 era depleting it, Argentina's crises clustering where its index collapses, and the Philippines' 1981 crisis arriving at a local trough before its post-2000 consolidation shows as a sustained climb. The dashed Effective Space Index lines make the distinction between the two objects visible country by country. The two series track each other only loosely within countries, and the gaps are as informative as the levels. The estimated index sits well above the buffers index for the United States in recent years, crediting the external and reserve inputs a balance sheet ratio cannot see, and below it for Japan, discounting measured buffers that history says bought little. Its Argentine reading is high in ways that mostly illustrate the reliability caveat of Section 4, and we read the overlay as a qualitative exhibit rather than a scorecard. Figure 9 asks the harder question of whether the composite predicts crisis resilience, re-estimating the interaction of Section 3.1 with the PBI in place of the debt ratio. The composite is right-signed though small in advanced economies, essentially flat in the pooled long-run panel, and wrong-signed though imprecise in EMDEs, where components mean different things, since a high real rate signals room to cut in an advanced economy and risk premia in an emerging one. Fixed weights do not appear to travel across development levels, which is exactly why the Effective Space Index lets crisis outcomes set the weights. A sharper version of the same test puts the indices where the debt ratio stood, re-estimating equation (2) on the adopted chronology with the lagged index in place of debt (Table A8). The buffers index passes where its most famous component fails. Its advanced-economy interaction is right-signed and significant, +6.0 with a standard error of 2.5 at five years, so a composite of measured buffers does predict resilience for advanced economies even though the debt ratio alone does not, while the pooled interaction stays flat, the fixed-weights limitation again. The out-of-fold Effective Space Index carries the right sign in the pooled sample and the wrong one among advanced economies, one more display of its negligible country-level reliability.[10]

6 Conclusion

When a financial crisis strikes, does it matter how much room the government has to fight back? Our answer comes in three parts, each with a policy edge.

First, the measure used to gauge that room predicts far less than the advice built on it assumes. Across 140 consistently defined systemic banking crises in half a century of data, the public debt ratio matters less, and differently, than the advice assumes. It adds nothing to crisis prediction once boom symptoms are accounted for. Its severity signal is real but narrow, concentrated in advanced economies at long horizons, outside the euro area, and above the 90 percent threshold, where crises entered with more debt cost significantly

[10] Appendix B diagnoses the wrong sign. It is not the 2008 cohort, whose events the out-of-fold scores if anything rank correctly as worse than others, and no single input drives it. It is what near-zero reliability looks like when interacted in a 21-event cell.

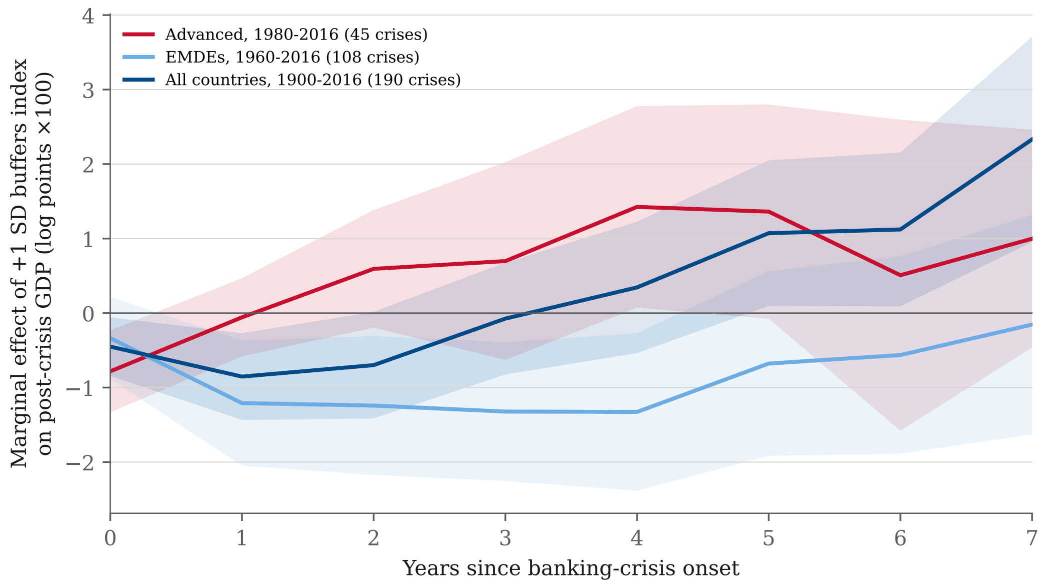


Figure 9: Does the buffers index predict resilience?

Notes: The interaction of the banking-crisis onset with the standardized lagged index, from equation (2) without boom controls, on the long-run GMD event set, in log points, with event counts in the legend. Bands are one standard error, clustered by country. Source: author's calculations.

more, and it is absent in EMDEs, where the five- and seven-year intervals rule out average penalties of the size the narrative literature finds, though deep-tail outcomes are estimated less precisely. The most plausible reason, on this evidence, is not that fiscal positions are irrelevant but that the ratio is contaminated exactly when it matters most, since the credit booms that precede the worst crises appear to inflate revenues and GDP and so make the most vulnerable balance sheets look the safest. The clearest victims of that illusion, Ireland, Iceland, and Spain in 2008, entered their crises with debt ratios their peers envied.

Second, what the debt ratio does predict, reliably, is behavior. Advanced economies that entered a crisis with one standard deviation more debt raised their cyclically adjusted balances by more than four percentage points of GDP relative to low-debt peers three years after onset, the most precise estimate in the paper. The ratio predicts the response far better than it predicts the resilience, and under standard multipliers that consolidation is large enough to account for between a third and all of the advanced-economy severity gradient, so part of any penalty high-debt countries suffer is plausibly administered rather than structural. A measure this consequential for behavior should measure what it claims to measure.

Third, measuring the room to respond directly is feasible. A transparent percentile index of buffers already organizes the cross-country landscape and exposes its own failures, ranking the reserve-currency United States near the bottom and sanction-bound Russia near the top, and the Effective Space Index, which learns from crisis outcomes, corrects in the right direction and weights reserve and external positions above the debt ratio. Its out-of-sample information is modest, consistently positive, significant across countries and marginal across periods, so we offer it as an illustration of outcome-disciplined measurement rather than a validated ranking. Both indices, the underlying event ledger, and all replication code are released with the paper. Debt-denomination inputs as their coverage grows, and era-specific training for the Effective Space Index, are natural extensions, but the central lesson does not wait on them. Buffers are worth rebuilding only if we can measure them, and the measure that international policy advice leans on is, on this evidence, the wrong one to lean on alone.

CRediT authorship contribution statement

Justin Eloriaga. Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The crisis event ledger, both indices, and all replication code are available from the author and will be posted publicly with the published article. The Global Macro Database and the Laeven and Valencia chronology are available from their original sources.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author used Claude Code in order to assist with drafting and editing the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article. All errors are the author's own.

References

Tobias Adrian, Nina Boyarchenko, and Domenico Giannone. Vulnerable growth. *American Economic Review*, 109(4):1263–1289, 2019.

Joshua Aizenman and Yothin Jinjarak. De facto fiscal space and fiscal stimulus: Definition and assessment. Working Paper 16539, National Bureau of Economic Research, 2010.

Joshua Aizenman and Jaewoo Lee. International reserves: Precautionary versus mercantilist views, theory and evidence. *Open Economies Review*, 18(2):191–214, 2007.

Alberto Alesina, Carlo Favero, and Francesco Giavazzi. The output effect of fiscal consolidation plans. *Journal of International Economics*, 96:S19–S42, 2015.

Serkan Arslanalp and Takahiro Tsuda. Tracking global demand for advanced economy sovereign debt. *IMF Economic Review*, 62(3):430–464, 2014.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *Annals of Statistics*, 47(2):1148–1178, 2019.

Alan J. Auerbach and Yuriy Gorodnichenko. Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy*, 4(2):1–27, 2012.

Matthew Baron, Emil Verner, and Wei Xiong. Banking crises without panics. *Quarterly Journal of Economics*, 136(1):51–113, 2021.

Javier Bianchi, Pablo Ottonello, and Ignacio Presno. Fiscal stimulus under sovereign risk. *Journal of Political Economy*, 131(9):2328–2369, 2023.

Olivier J. Blanchard and Daniel Leigh. Growth forecast errors and fiscal multipliers. *American Economic Review*, 103(3):117–120, 2013.

Kristina Bluwstein, Marcus Buckmann, Andreas Joseph, Sujit Kapadia, and Özgür Şimşek. Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Journal of International Economics*, 145:103773, 2023.

Michael Bordo, Barry Eichengreen, Daniela Klingebiel, and Maria Soledad Martinez-Peria. Is the crisis problem growing more severe? *Economic Policy*, 16(32):52–82, 2001.

Claudio Borio, Marco Lombardi, and Fabrizio Zampolli. Fiscal sustainability and the financial cycle. BIS Working Paper 552, Bank for International Settlements, 2016.

Claudio Borio, Piti Disyatat, and Mikael Juselius. Rethinking potential output: Embedding information about the financial cycle. *Oxford Economic Papers*, 69(3):655–677, 2017.

Markus K. Brunnermeier, Sebastian A. Merkel, and Yuliy Sannikov. Debt as safe asset. Working Paper 29626, National Bureau of Economic Research, 2022.

Ivan A. Canay. A simple approach to quantile regression for panel data. *Econometrics Journal*, 14(3):368–386, 2011.

Luis A.V. Catão and Gian Maria Milesi-Ferretti. External liabilities and crises. *Journal of International Economics*, 94(1):18–32, 2014.

Valerie Cerra and Sweta Chaman Saxena. Growth dynamics: The myth of economic recovery. *American Economic Review*, 98(1):439–457, 2008.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21(1):C1–C68, 2018.

Alexander Chudik, Kamiar Mohaddes, M. Hashem Pesaran, and Mehdi Raissi. Is there a debt-threshold effect on output growth? *Review of Economics and Statistics*, 99(1): 135–150, 2017.

Stijn Claessens, M. Ayhan Kose, and Marco E. Terrones. How do business and financial cycles interact? *Journal of International Economics*, 87(1):178–190, 2012.

Mathias Drehmann and Mikael Juselius. Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, 30(3):759–780, 2014.

Arindrajit Dube, Daniele Girardi, Òscar Jordà, and Alan M. Taylor. A local projections approach to difference-in-differences event studies. Working Paper 31184, National Bureau of Economic Research, 2023.

Markus Eberhardt and Andrea F. Presbitero. Public debt and growth: Heterogeneity and non-linearity. *Journal of International Economics*, 97(1):45–58, 2015.

Barry Eichengreen and Ricardo Hausmann. Exchange rates and financial fragility. Working Paper 7418, National Bureau of Economic Research, 1999.

Antonio Fatás and Lawrence H. Summers. The permanent effects of fiscal consolidations. *Journal of International Economics*, 112:238–250, 2018.

Jeremy Fouliard, Michael Howell, and Hélène Rey. Answering the queen: Machine learning and financial crises. Working Paper 28302, National Bureau of Economic Research, 2021.

Jeffrey Frankel and George Saravelos. Can leading indicators assess country vulnerability? evidence from the 2008–09 global financial crisis. *Journal of International Economics*, 87 (2):216–231, 2012.

Jeffrey A. Frankel, Carlos A. Végh, and Guillermo Vuletin. On graduation from fiscal procyclicality. *Journal of Development Economics*, 100(1):32–47, 2013.

Manuel Funke, Moritz Schularick, and Christoph Trebesch. Going to extremes: Politics after financial crises, 1870–2014. *European Economic Review*, 88:227–260, 2016.

Atish R. Ghosh, Jun I. Kim, Enrique G. Mendoza, Jonathan D. Ostry, and Mahvash S. Qureshi. Fiscal fatigue, fiscal space and debt sustainability in advanced economies. *Economic Journal*, 123(566):F4–F30, 2013.

Pierre-Olivier Gourinchas and Maurice Obstfeld. Stories of the twentieth century for the twenty-first. *American Economic Journal: Macroeconomics*, 4(1):226–265, 2012.

Pierre-Olivier Gourinchas and Hélène Rey. From world banker to world venture capitalist: US external adjustment and the exorbitant privilege. In *G7 Current Account Imbalances: Sustainability and Adjustment*, pages 11–66. University of Chicago Press, 2007.

Jaime Guajardo, Daniel Leigh, and Andrea Pescatori. Expansionary austerity? international evidence. *Journal of the European Economic Association*, 12(4):949–968, 2014.

Thomas Herndon, Michael Ash, and Robert Pollin. Does high public debt consistently stifle economic growth? a critique of reinhart and rogoff. *Cambridge Journal of Economics*, 38 (2):257–279, 2014.

Christopher L. House, Christian Proebsting, and Linda L. Tesar. Austerity in the aftermath of the great recession. *Journal of Monetary Economics*, 115:37–63, 2020.

Ethan Ilzetzki, Carmen M. Reinhart, and Kenneth S. Rogoff. Exchange arrangements entering the twenty-first century: Which anchor will hold? *Quarterly Journal of Economics*, 134(2):599–646, 2019.

International Monetary Fund. Assessing fiscal space: An update and stocktaking. Imf policy paper, International Monetary Fund, 2018.

Olivier Jeanne and Romain Rancière. The optimal level of international reserves for emerging market countries: A new formula and some applications. *Economic Journal*, 121(555): 905–930, 2011.

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

Òscar Jordà and Alan M. Taylor. The time for austerity: Estimating the average treatment effect of fiscal policy. *Economic Journal*, 126(590):219–255, 2016.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor. When credit bites back. *Journal of Money, Credit and Banking*, 45(s2):3–28, 2013.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor. Sovereigns versus banks: Credit, crises, and consequences. *Journal of the European Economic Association*, 14(1):45–79, 2016.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor. Macrofinancial history and the new business cycle facts. In *NBER Macroeconomics Annual 2016*, volume 31, pages 213–263. University of Chicago Press, 2017.

Graciela L. Kaminsky and Carmen M. Reinhart. The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review*, 89(3):473–500, 1999.

Graciela L. Kaminsky, Carmen M. Reinhart, and Carlos A. Végh. When it rains, it pours: Procyclical capital flows and macroeconomic policies. In *NBER Macroeconomics Annual 2004*, volume 19, pages 11–53. MIT Press, 2004.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

M. Ayhan Kose, Sergio Kurlat, Franziska Ohnsorge, and Naotaka Sugawara. A cross-country database of fiscal space. *Journal of International Money and Finance*, 128:102682, 2022.

Luc Laeven and Fabián Valencia. Systemic banking crises database II. *IMF Economic Review*, 68:307–361, 2020.

Luc Laeven and Fabián Valencia. Systemic banking crises database: 1970–2025. IMF Working Paper 2026/094, International Monetary Fund, 2026.

Atif Mian, Amir Sufi, and Emil Verner. Household debt and business cycles worldwide. *Quarterly Journal of Economics*, 132(4):1755–1817, 2017.

Richard Morris and Ludger Schuknecht. Structural balances and revenue windfalls: The role of asset prices revisited. ECB Working Paper 737, European Central Bank, 2007.

Karsten Müller, Chenzi Xu, Mohamed Lehbib, and Ziliang Chen. The global macro database: A new international macroeconomic dataset. Working Paper 33714, National Bureau of Economic Research, 2025.

Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffmann, and Enrico Giovannini. Handbook on constructing composite indicators: Methodology and user guide. Technical report, OECD Publishing, 2008.

Maurice Obstfeld, Jay C. Shambaugh, and Alan M. Taylor. Financial stability, the trilemma, and international reserves. *American Economic Journal: Macroeconomics*, 2(2):57–94, 2010.

Jonathan D. Ostry, Atish R. Ghosh, Jun I. Kim, and Mahvash S. Qureshi. Fiscal space. IMF Staff Position Note SPN/10/11, International Monetary Fund, 2010.

Valerie A. Ramey. Ten years after the financial crisis: What have we learned from the renaissance in fiscal research? *Journal of Economic Perspectives*, 33(2):89–114, 2019.

Carmen M. Reinhart and Kenneth S. Rogoff. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press, Princeton, 2009.

Carmen M. Reinhart and Kenneth S. Rogoff. Growth in a time of debt. *American Economic Review*, 100(2):573–578, 2010.

Carmen M. Reinhart, Kenneth S. Rogoff, and Miguel A. Savastano. Debt intolerance. *Brookings Papers on Economic Activity*, 2003(1):1–74, 2003.

Christina D. Romer and David H. Romer. New evidence on the aftermath of financial crises in advanced countries. *American Economic Review*, 107(10):3072–3118, 2017.

Christina D. Romer and David H. Romer. Why some times are different: Macroeconomic policy and the aftermath of financial crises. *Economica*, 85(337):1–40, 2018.

Christina D. Romer and David H. Romer. Fiscal space and the aftermath of financial crises: How it matters and why. *Brookings Papers on Economic Activity*, Spring:239–313, 2019.

Moritz Schularick and Alan M. Taylor. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008. *American Economic Review*, 102(2):1029–1061, 2012.

Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

A Adjudication of audit disagreements

Table A1 reports the adjudication of the 27 GMD banking-crisis onsets with no Laeven and Valencia (2020) counterpart within two years, based on the BVX Crisis List of Baron et al. (2021), the dating surveyed in Reinhart and Rogoff (2009), and narrative sources for the initially unresolved events. Eighteen are borderline events that broader chronologies or narrative accounts record but Laeven and Valencia do not classify as systemic, including Peru in 1998, where Banco Latino was intervened and the bank count fell from 26 to 14, and Venezuela, where the 2009 takeover wave reached about 12 percent of deposits and the GMD flag at 2008 runs a year early. Five are dating divergences in which both sources record the same underlying crisis, including Japan, where the BVX list dates the crisis to the 1990 asset collapse while Laeven and Valencia date it to the 1997 bank failures. Three are continuations of earlier spells flagged as new onsets, and one, the Philippines in 1971, cannot be resolved from the statistical sources and is retained as the single unresolved case. The three Laeven and Valencia events without a GMD counterpart are the mirror image. Norway 1991 and Russia 1998 are the other side of dating divergences recorded in the table, and the Democratic Republic of the Congo 1994 is missing from the GMD flags and is a candidate addition. Because estimation uses the Laeven and Valencia list, none of these GMD-only events enters the baseline. The audit reconciles the flags with the 2020 vintage, which the adopted 2026 chronology nests, so the update's added episodes do not alter it. The adjudication instead governs the auxiliary exercises that use the GMD flags and identifies which disagreements are candidates for the robustness variants, re-dating category B events, excluding category C continuations, and excluding the pending category D event.

Table A1: Adjudication of GMD onsets absent from Laeven and Valencia (2020).

Country	Year	Category	Note
A. Borderline, in broader chronologies (18 events)			
Argentina	1985	A	Austral-era distress, on the BVX Crisis List
Brazil	1985	A	On the BVX Crisis List, borderline for LV
Switzerland	1990	A	Regional bank distress, on the BVX Crisis List
Denmark	1992	A	Nordic distress, in the Reinhart-Rogoff dating
United Kingdom	1973	A	Secondary banking crisis of 1973–75
United Kingdom	1991	A	Small banks crisis, on the BVX Crisis List
Indonesia	1990	A	On the BVX Crisis List, borderline for LV
Iceland	1985	A	In the Reinhart-Rogoff dating
Iceland	1993	A	In the Reinhart-Rogoff dating
Italy	1992	A	ERM-era distress, on the BVX Crisis List
Malaysia	1985	A	On the BVX Crisis List, borderline for LV
Norway	2008	A	GFC distress, borderline for LV
Peru	1998	A	Banco Latino under intervention Dec. 1998, Wiese bailout 1999, banks fell 26 to 14
Thailand	1979	A	Finance-company crisis, in the Reinhart-Rogoff dating
Turkey	1991	A	On the BVX Crisis List, borderline for LV
United States	1984	A	Continental Illinois and the S&L distress, on the BVX list
Venezuela	1981	A	Early-1980s distress, in the Reinhart-Rogoff dating
Venezuela	2008	A	2009–10 takeover wave, 12 percent of deposits, narrative dates 2009
B. Dating divergence with LV (5 events)			
Czech Republic	1991	B	LV dates the transition crisis 1996
Hungary	1995	B	LV dates the transition crisis 1991
Japan	1990	B	BVX dates the crisis 1990, LV dates it 1997
Norway	1987	B	BVX dates the crisis 1987, LV dates it 1991
Russian Federation	1995	B	GMD dates the 1995 interbank crisis, LV dates 1998
C. Continuation of an earlier spell (3 events)			
Italy	2016	C	MPS and Veneto resolutions in the tail of the 2008 spell
Japan	2001	C	Continuation of the 1997 crisis spell
Portugal	2014	C	Banco Espírito Santo resolution in the tail of the euro spell
D. Pending narrative verification (1 event)			
Philippines	1971	D	Post-devaluation distress, documented run is Continental Bank 1974

B Estimation details and diagnostics

Sample construction. Estimation rows on the adopted chronology run through 2023, with each horizon using the outcomes the June 2026 data vintage contains, so the latest onsets contribute only to the short-horizon cells and 138 of the 140 events carry realized five-year outcomes. Auxiliary exercises that use the GMD flags stop at 2016, the end of those flags' verified coverage. In the historical figures, onsets during 1914 to 1919 and 1939 to 1947 are excluded, and observations with an absolute annual change in log output above 30 log points are dropped, which removes wartime collapses, hyperinflation artifacts, and splice errors. The boom proxies use fallback chains, with M1 and M0 substituting for M2, fixed investment for total investment, and the trade balance for the current account, which extends coverage to 96 of the 116 EMDE crises in the estimation sample. Table A5 traces every step from the 164 recorded episodes to each estimation sample. Because the trimming of extreme outcome changes is done within each estimation sample, subsample event counts can sum to slightly more than the pooled count, as in the policy-rate rows of Table 1, and because the advanced subsample begins in 1980, its events plus the EMDE events fall one short of the pooled count, the difference being Spain 1977. Within the advanced sample the raw money proxy is well behaved, with a standard deviation of 28 percentage points, so the currency-splice artifacts noted in Table A4 never enter the headline advanced cells.

Attrition. Among the 140 events with baseline data, inclusion in the boom-restricted sample is unrelated to pre-crisis debt, with a coefficient of 0.02 and standard error 0.03 per standard deviation, and unrelated to realized five-year severity. Inclusion in the mechanism sample is likewise unrelated to debt but mildly related to severity, with a coefficient of 0.003 per log point and a p-value of 0.05, so events with milder outcomes are slightly overrepresented among those reporting budget data, a selection worth keeping in mind when reading Table 1.

Inference for the interacted projection. Standard errors are clustered by country throughout, and on the adopted chronology they exceed the interaction point estimate in every cell except the advanced-economy interactions at long horizons, which is why the text reports imprecision as the headline. A restricted wild cluster bootstrap with Rademacher weights and the null imposed on the interaction, run on the adopted chronology for the advanced boom-controlled specification, yields p-values of 0.62, 0.48, 0.24, and 0.05 at horizons one, three, five, and seven, so only the seven-year cell approaches conventional significance under the most conservative inference we use. On the broader multi-chronology event set the corresponding p-values lie between 0.34 and 0.44. The confidence intervals tell the other half of the story. At five and seven years the pooled and EMDE estimates carry standard errors near 1.4, so their confidence intervals exclude a penalty of 3.5 log points per standard deviation while the advanced intervals, with standard errors near 2.6, do not.

Pre-trends and placebo. Estimating the interacted projection at horizons -3 and -2, the last horizons not fixed at zero by the timing convention, shows no differential movement of output by pre-crisis debt before onset. The interaction coefficients never exceed 0.2 log points in absolute value across the three samples and none is statistically significant. Reassigning each onset to a randomly drawn no-crisis country in the same year, holding the number of events per year fixed, yields placebo interactions that are statistically indistinguishable from zero at every horizon, near zero on impact and positive rather than negative at longer horizons, so the advanced-economy gradient does not arise mechanically from the specification.

Scale and measurement. One standard deviation of the lagged debt ratio is 51 percentage points of GDP in the pooled estimation sample, 39 among advanced economies, and 55 in EMDEs, so per standard deviation magnitudes describe wide differences in positions. Attenuation from measurement error in the historical debt series works in the same direction as boom flattery, and we cannot separate the two.

Credit-based booms and onset specification. Building the boom vector from WDI private credit to GDP instead of money reproduces the severity pattern, with the advanced interaction at -6.0 with a standard error of 2.4 at seven years on 20 crises (Table A6), and reproduces the onset null. In the onset logit with credit-based booms and decade effects, ten points of debt carries a margin of 0.026 percentage points with a standard error of 0.018 and the area under the ROC curve is 0.77, and in a linear probability model with country and year effects and the baseline boom vector, the debt term is -0.008 with a standard error of 0.037.

Mechanism inference. Restricted wild cluster bootstrap p-values for the advanced cyclically adjusted balance cells of Table 1 are 0.96, 0.008, and 0.061 at one, three, and five years, and between 0.55 and 0.61 for the advanced policy-rate cells. The three-year cell is the largest in an eighteen-cell table, and family-wise adjustment would leave it marginal on its own, though the same-signed five-year cell at a bootstrap p-value of 0.06 makes the finding a horizon profile rather than one cell. The advanced onset path for the balance at three years is -3.2 with a standard error of 0.9, the anchor of the implied paths in the text, and the pseudo-onset placebo interaction is 0.5 with a standard error of 0.8 at three years. Two cautions apply to the headline cell. Fourteen of the twenty advanced events belong to the 2008 to 2012 wave, in which high-debt entrants were largely euro-area economies under program conditionality, so the interaction may partly measure institutional constraints that travel with debt rather than the ratio itself. And events with milder outcomes are slightly overrepresented among those reporting budget data, per the attrition tests above. The program-participation row of Table A10 enters an IMF or EU program indicator in levels and interacted with the onset, using the documented program windows for Korea, Iceland, Greece, Ireland, Portugal, Spain, Cyprus, Latvia, and Hungary, and the debt interaction survives it at 3.7 with a standard error of 0.8 at three years. Rows of Table A10 with fewer than ten events carry no significance stars because clustered inference is unreliable there.

Small-cell diagnostics. Four of the 21 advanced events entered with debt above 90 percent of GDP, Greece (105 percent) and Italy (103 percent) in 2008, Japan (98 percent) in 1997, and Israel (133 percent) in 1983. Dropping the advanced events one at a time leaves the threshold coefficient between −5.8 and −10.7 at three years, −6.7 and −15.1 at five, and −10.2 and −17.9 at seven, always negative and large. The eleven crises outside the euro area are Denmark, Iceland, the United Kingdom, and the United States in 2007 to 2008, Finland, Norway, and Sweden in 1991, Japan and Korea in 1997, Israel in 1983, and Kuwait in 1982, and the same leave-one-out exercise keeps that coefficient between −1.8 and −5.1 at three years and −5.2 and −8.6 at seven.

The doubly robust gradient. The average five-year crisis effect under the doubly robust design is −9.7 log points with a standard error of 5.1. Its debt gradient in Table A9 is −17.0 with a standard error of 3.0 in the advanced sample, several times the projection estimate, and the gap is not cross-country level variation, since adding country effects to the score regression moves it further, to −28.5 with a standard error of 7.0. The design differs from the projections in two ways, and each can be tested by swapping one ingredient. Replacing the score regression's decade effects with year effects moves its gradient from −17.0 to −11.0 with a standard error of 1.4, so time granularity explains about a third of the gap. Giving the projection decade effects instead of year effects leaves it near −3.2, so the remainder reflects the score regression's propensity weighting and cross-sectional identification rather than the time controls. We therefore carry the projection magnitudes, the design with the tightest time controls, and read the doubly robust row as corroborating the sign under a different weighting of the same cohort.

Chronology adoption. The episode list of Laeven and Valencia (2026) is parsed from the published working paper and nests every episode of the 2020 vintage, adding 13 new episodes through 2025. The 14 borderline episodes are excluded from the baseline and retained for robustness. Including them moves the baseline interaction by less than a tenth of a log point at every horizon, with the advanced seven-year estimate at −6.2 on 26 events (Table A7). Pooling events from several chronologies instead produces estimates that move with the era composition of the sample, because sources differ in how they define and date distress, which is why we commit to one consistently defined list for estimation.

Chronology comparison. The multi-chronology GMD flags yield 322 recorded onsets over 1870 to 2024, verified through 2016, against the 140 estimation events over 1976 to 2023, and estimates on the broader set are stronger, in particular an EMDE interaction near -2.3 log points at three years. The difference traces to the events only the broader chronologies record, borderline distress episodes and pre-1970 crises, so results on that set answer a different question, about distress broadly defined, than results on the single consistently defined systemic set. The audit of Appendix A reconciles the two event lists. The five-year shortfall measure behind Figure 3 is robust to the window, with rank correlations of 0.90 and 0.93 against three-year and seven-year variants.

Partial pooling across eras and groups. When the debt interaction is estimated separately for each era and group cell, the cells vary far more than sampling error explains, and the pattern has an identifiable source. Re-estimating the cells on the single LV chronology reproduces it, so chronology mixing is not the cause. The tell is which cells turn positive. Within-era interactions without boom controls are positive precisely in the credit-boom crisis waves, +3.5 with a standard error of 2.6 for advanced economies in 2008 to 2023 and a significant +5.5 with 2.3 for EMDEs in 1997 to 2007, while the early EMDE cell is near zero and the post-2008 EMDE cell is negative at -4.9 with 2.9. The positive cells are the episodes in which booming low-debt economies suffered the deepest busts, so the uncontrolled within-era coefficient is pushed positive exactly where boom flattery is strongest. The instability is not noise and not a data artifact. It is the boom confound varying in strength with the era composition of the sample, which is the severity mechanism showing up in the pooling diagnostics. Figure A1 reports the cells.

Effective-space diagnostics and the cross-era failure. The causal forest is fit with 1,000 honest trees, minimum leaf size 20, and gradient-boosting nuisance models cross-fit in five country-grouped folds, with missing features median-imputed alongside missingness indicators. For the final configuration, held-out countries' predicted crisis costs correlate with realized five-year shortfalls at a Spearman rank correlation of 0.195 across 138 events, with a permutation p-value of 0.02, and rolling-origin predictions made strictly from earlier data correlate at 0.192 across 94 events, with a permutation p-value of 0.07. Event counts differ across exercises by construction. The projections use 140 events over 1976 to 2023, the forest trains on the 139 events between 1970 and 2019 with an observable entry state vector, and 138 of those carry realized five-year shortfalls for validation. The rolling-origin design also selected the configuration among three candidates, which imparts a mild

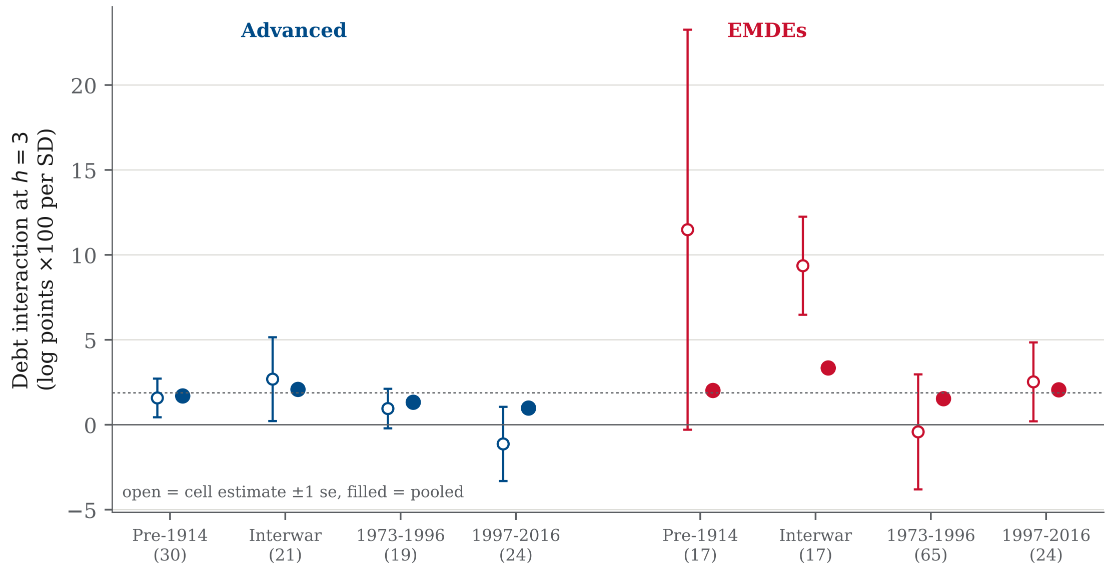


Figure A1: Partial pooling across eras and groups.

Notes: Cell-level debt interactions at the three-year horizon on the adopted Laeven and Valencia (2026) chronology without boom controls (open markers, one standard error) and empirical Bayes pooled counterparts (filled markers). Event counts in parentheses. Source: author's calculations.

optimism to the second number. A best-linear-predictor calibration test in the spirit of the generalized-random-forest literature, regressing cross-fit doubly robust scores on the demeaned out-of-fold predictions, gives a differential coefficient of 0.78 with a standard error of 0.62, consistent with sizable true heterogeneity but too imprecise to rule out none, which is why the text presents the Effective Space Index as a companion. The 2019 rank of the United States is 122nd under the chosen configuration and between 70th and 122nd across the three configurations examined. A diagnosis-stage forest with the same features, before reserves entered and with no configuration search, gives country-fold correlations of 0.13 to 0.14, and a forest fit through 1999 predicts the 38 later crises at only 0.04 to 0.09, weak but no longer wrong-signed. Running the identical design on the multi-chronology GMD flags over the same period gives country-fold correlations of 0.17 to 0.25 but cross-era correlations near −0.08. The comparison separates two problems. Wrong-signed cross-period transfer arises when event definitions are mixed across time, and the single chronology removes it, while the remaining weakness is genuine, since even consistently defined pre-2000 crises carry limited information about post-2000 severity. Era-specific training and richer market-access inputs are the natural refinements. The wrong-signed advanced-economy row of Table A8 has the same character. It is not the 2008 cohort, since the out-of-fold scores rank the wave's events as worse than other advanced events and excluding the wave makes the coefficient more negative on the seven events that remain, and no single input drives it, with the advanced scores' correlations with debt and the boom symptoms all below 0.2 in absolute value. A 21-event interaction of a score with near-zero split-half reliability can load significantly in either direction, which is why the text publishes the ranking as an illustration and rests the index's case on its aggregate objects.

The distribution of outcomes. Policymakers buy buffers as insurance against disasters, not against average outcomes, so the relevant question may be distributional. We estimate quantile analogues of the interacted projection at the three-year horizon, without the boom-exposure terms, by quantile regression (Koenker and Bassett, 1978), using the two-step approach of Canay (2011). The first step recovers the country effects from the mean regression,

$$\hat{\alpha}_i \;=\; \frac{1}{T_i}\sum_t \left(y_{i,t+3} - y_{i,t-1} - w_{i,t}'\hat{b}\right), \tag{9}$$

where $w_{i,t}$ stacks the onset, its interaction with debt, the debt level, and the lag controls of equation (2), and T_i is the number of observations for country i. The second step estimates, for each quantile q of the demeaned outcome $\tilde{y}_{i,t} = y_{i,t+3} - y_{i,t-1} - \hat{\alpha}_i$,

$$Q_q(\tilde{y}_{i,t}\,|\,\cdot) \;=\; \gamma_{d(t),q}+\beta_q D_{i,t}+\delta_q\,(D_{i,t}\cdot S_{i,t-1})+\kappa_q S_{i,t-1}+\sum_{k=1}^{2}\kappa_{k,q}\,\Delta y_{i,t-k}+\sum_{k=1}^{2}\phi_{k,q}\,D_{i,t-k}, \tag{10}$$

where $\gamma_{d(t),q}$ are decade effects indexed by the decade $d(t)$ of year t, and inference uses a country block bootstrap.

At the median and above, the debt interaction δ_q is zero in every sample. At the 10th percentile the point estimates turn negative, -3.6 log points per standard deviation in EMDEs and -0.6 in the pooled sample, as Figure A2 shows. The point estimates suggest that whatever role debt plays in severity operates in the worst tenth of outcomes, which connects the estimates to the growth-at-risk framework of Adrian et al. (2019), though the block-bootstrap bars are wide and we read the tail pattern as suggestive.

Quantile inference. A country block bootstrap with 200 replications widens the 10th-percentile standard errors to 5.8 in the EMDE sample against a point estimate of -3.6, so the tail concentration reported in the text is a pattern in point estimates rather than an established fact. Median and upper-quantile interactions are indistinguishable from zero under either inference method.

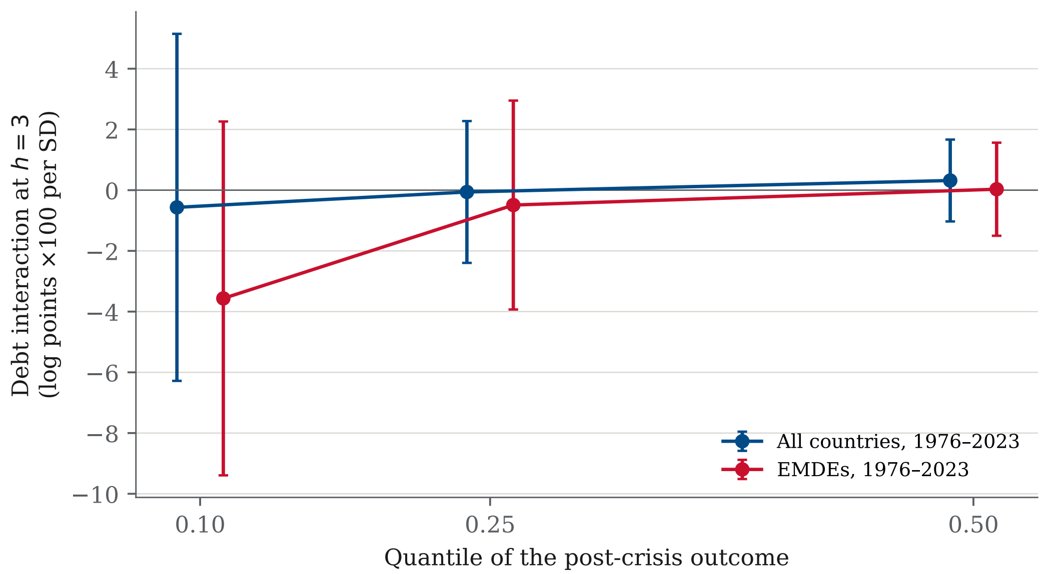


Figure A2: The debt interaction across the outcome distribution.

Notes: Quantile estimates δ_q at the three-year horizon on the adopted chronology, for the pooled and EMDE samples, in log points per standard deviation of debt. Bars are one country-block-bootstrap standard error from 200 replications. Sources: Global Macro Database; Laeven and Valencia (2026).

C Additional maps and figures

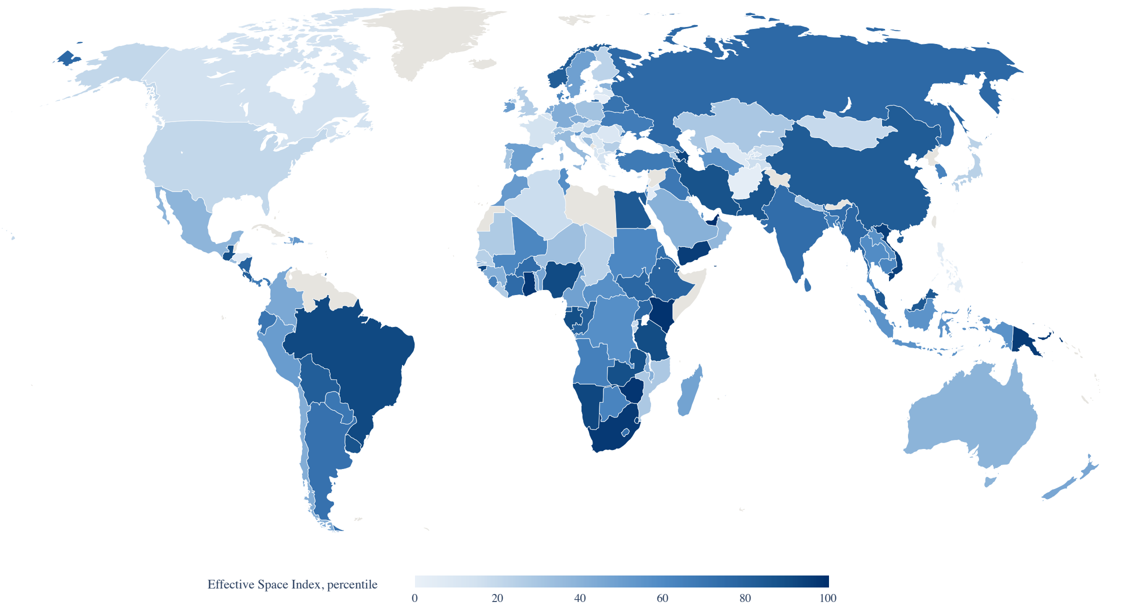


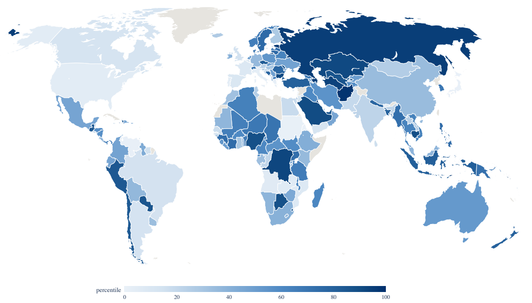
Figure A3: The Effective Space Index, 2019.

Notes: Within-year percentile of estimated effective space in 2019 across the 154 covered countries. Darker shading indicates more room to respond, and gray indicates insufficient data. Source: author's calculations.

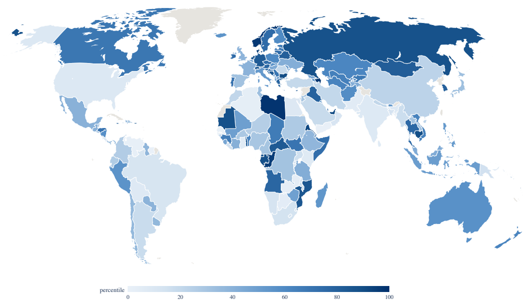
Inside the tails of the estimated ranking. Figure A6 reads the estimated index on the eve of the global financial crisis, showing the five highest- and five lowest-ranked countries in the 2007 cross section among the 125 of 154 ranked economies that report every displayed input, so no cell is imputed. The top five hold the configuration the training data associate with mild outcomes. Hong Kong SAR, Botswana, China, and the United Arab Emirates combine large reserve stocks, reserves reach 81 percent of GDP in Botswana, with external surpluses, fiscal surpluses, and debt ratios below the cross-country median. Syria in third place, with thin reserves and double-digit inflation, is the counterexample inside the same tail, a call resting on no single visible strength, and the kind of draw the split-half diagnostics price.

The bottom tail is the more instructive one. Nicaragua carries the one classic weak profile, an external deficit of 13 percent of GDP. Algeria, Oman, and Singapore are the standing caveat, since reserve stocks up to 93 percent of GDP and external surpluses above 10 percent of GDP sit far outside the training support and the forest reads exceptional strength as exceptional risk. Switzerland pairs large surpluses with a five-year money contraction of 29 percent of GDP, an extreme in the opposite direction. What makes this tail worth

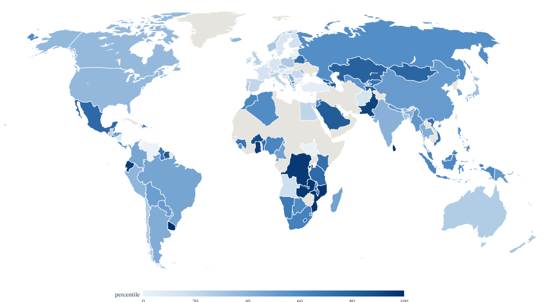
(a) Debt



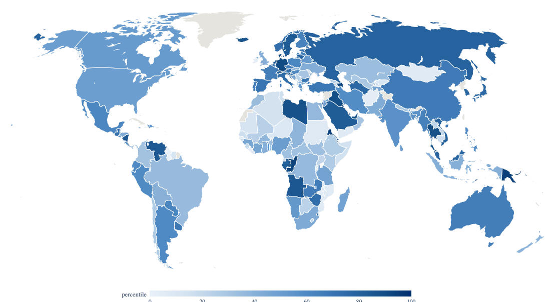
(b) Fiscal balance



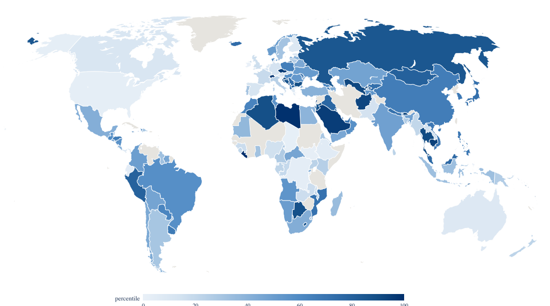
(c) Real policy rate



(d) Current account balance



(e) Reserves over GDP



(f) Inverse spread

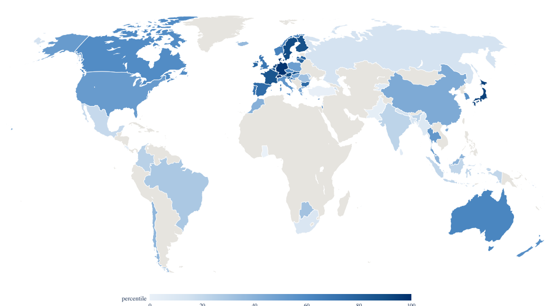


Figure A4: Components of the Policy Buffers Index, 2019.

Notes: Within-year percentiles, darker indicating more room. Source: author's calculations.

recording is the year that followed. Spain, at 140th, entered its systemic crisis in 2008, and Switzerland, at 145th, saw its largest bank recapitalized by the state that same year, an episode our chronology records as borderline. The forest is trained on the full event history, so this is an in-sample reading rather than a forecast, and the middle of the ranking carries little information, but the exhibit shows where the estimated index concentrates its signal, in the extremes of the reserve, external, and monetary margins.

(a) Currency-crisis forest (b) Sovereign-crisis forest

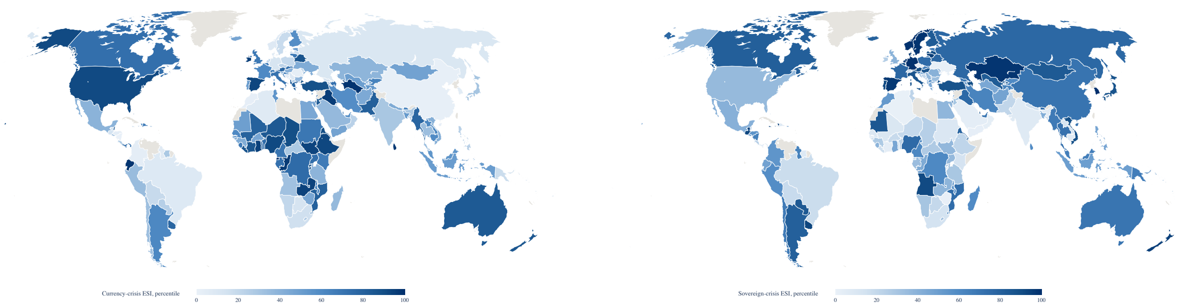


Figure A5: Effective space by crisis type, 2019.

Notes: Within-year percentile of the 2019 effective-space scores from the currency-crisis and sovereign-crisis forests. Darker shading indicates more room to respond, and gray indicates insufficient data. The 2019 rank correlations across crisis types are 0.00 between banking and currency, −0.38 between banking and sovereign, and −0.05 between currency and sovereign. Source: author's calculations.

	Reserves	External	Ext., 3y	Debt	Balance	Inflation	Money	Investment	ESI
Highest effective space									
Hong Kong SAR (1)	95	86	88	1	78	13	98	14	100
Syria (3)	38	67	68	64	38	84	1	36	99
Botswana (4)	96	95	90	6	93	89	87	7	98
China (5)	91	80	79	33	54	8	89	67	97
UAE (6)	44	93	90	5	92	80	74	15	97
Lowest effective space									
Algeria (144)	94	96	95	30	92	19	14	42	7
Switzerland (145)	54	85	89	68	59	4	5	38	6
Singapore (146)	97	97	96	90	70	3	25	12	5
Nicaragua (147)	49	7	5	69	61	79	32	53	5
Oman (150)	43	86	87	8	83	36	16	90	3

Figure A6: Inside the 2007 tails of the Effective Space Index.

Notes: Each cell is the country's percentile in the 2007 scoring cross section for the raw input, not a measure of room, so dark debt cells mean high debt. In the last column higher percentiles mean more effective space, and the overall rank among the 154 ranked countries is in parentheses, with rank 1 the most effective space. Countries missing any displayed input are excluded, so no cell is imputed. Reserves, external balances, debt, the fiscal balance, and inflation are lagged one year. Money and investment are the five-year boom-symptom changes. Source: author's calculations.

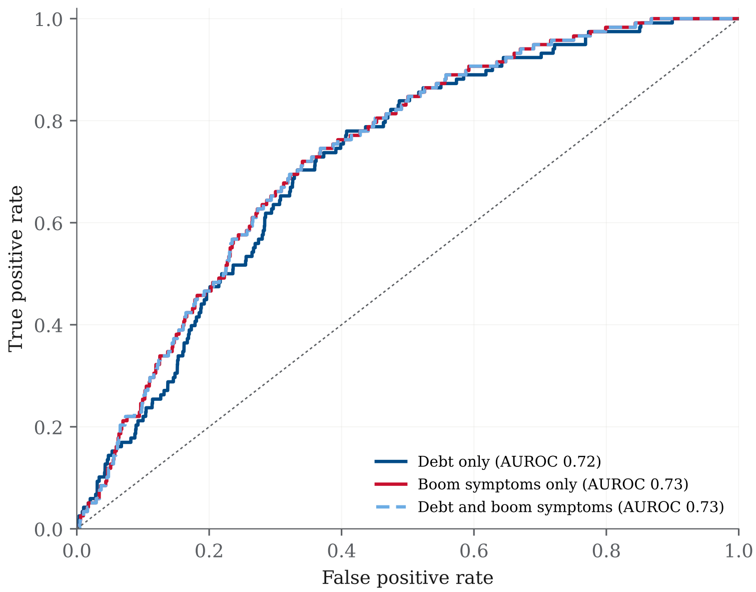


Figure A7: Receiver operating characteristic curves for the onset logits.

Notes: Pooled sample, 1976 to 2023, with decade dummies. The debt-and-booms curve lies on top of the booms-only curve, so debt adds no discrimination. Source: author's calculations.

	Debt	Balance	Real rate	External	Reserves	Spread	Index
Most room							
United Arab Emirates (1)	86	91		88	71		84
Azerbaijan (2)	95	97	84	92	44		82
Iraq (3)	60	80		89	77		77
Kuwait (4)	96	55	57	94	81		77
Hong Kong (5)	100	61	35	84	99		76
Switzerland (6)	70	83	23	81	98	100	76
Saudi Arabia (7)	91	19	82	83	95		74
Bosnia and Herzegovina (8)	79	84		45	87		74
Selected large economies							
China (72)	35	23	51	67	66	44	48
India (131)	22	8	40	57	49	23	33
Japan (61)	1	33	32	76	73	94	51
Germany (43)	36	83	12	91	14	98	56
Brazil (134)	15	12	47	35	58	29	33
Indonesia (70)	81	51	61	41	34	21	48
Philippines (27)	75	52	56	58	69		62
Least room							
Bahrain (137)	11	17		49	28		26
Lebanon (138)	3	3		3	97		26
United States (139)	5	9	37	50	3	52	26
Egypt (140)	19	9	22	37	39		25
Laos (141)	24	31		23	16		24
Venezuela (142)	2	2	0	84			22
Yemen (143)	12	2		7	45		17
Burundi (144)	33	11		9	6		15

Figure A8: Component detail for selected countries, 2019.

Notes: Percentile ranks of each component and the composite in 2019, the last pre-pandemic year, with overall ranks in parentheses, for the eight highest- and lowest-ranked countries and eight large economies. Higher percentiles mean more room on every component, and rank 1 is the most room. Source: author's calculations.

D Additional tables

Table A2: The debt interaction by sample and control set.

Controls	$h = 1$	$h = 3$	$h = 5$	$h = 7$
All countries, 1976–2023 (140 crises, 118 in restricted sample)				
None, full sample	-0.93 (2.18)	-1.36 (2.30)	-1.04 (1.69)	-1.41 (2.04)
None, restricted sample	-1.02 (2.44)	-1.05 (2.36)	-0.20 (1.28)	-0.36 (1.35)
Boom controls	-1.44 (3.01)	-1.23 (2.82)	0.31 (1.39)	0.33 (1.37)
Advanced, 1980–2023 (23 crises, 21 in restricted sample)				
None, full sample	-0.59 (1.39)	-1.12 (1.99)	-2.88 (2.53)	-4.11* (2.42)
None, restricted sample	-0.92 (1.64)	-1.13 (2.37)	-3.15 (2.93)	-5.00* (2.58)
Boom controls	-1.16 (1.71)	-1.64 (2.11)	-3.57 (2.77)	-6.23** (2.43)
EMDEs, 1976–2023 (116 crises, 96 in restricted sample)				
None, full sample	-0.99 (2.44)	-1.75 (2.63)	-1.29 (1.96)	-1.58 (2.42)
None, restricted sample	-1.06 (2.71)	-1.00 (2.57)	0.22 (1.39)	0.28 (1.54)
Boom controls	-2.06 (3.54)	-1.86 (3.22)	0.03 (1.47)	0.20 (1.49)

Notes: Numeric counterpart of the no-control and boom-control paths of Figure 5, with the clean-control design in Table A9. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels using those clustered standard errors. Restricted wild cluster bootstrap p-values for the advanced boom-controlled row are 0.62, 0.48, 0.24, and 0.05 at horizons 1, 3, 5, and 7.

Table A3: The Romer and Romer comparison on our data.

Advanced economies	$h = 1$	$h = 3$	$h = 5$	$h = 7$
Low-debt entrants	0.01 (1.99)	-1.88 (2.42)	-1.55 (3.06)	0.69 (2.30)
High-debt entrants	-1.52 (1.59)	-4.64 (3.30)	-5.57 (3.84)	-4.37 (3.99)
Difference	-1.52 (2.29)	-2.76 (3.80)	-4.02 (4.49)	-5.06 (4.05)

Notes: Average post-crisis output paths from equation (1) with separate onset indicators for crises in advanced economies entered below and above the median entry debt ratio of 43 percent of GDP, 12 and 11 events, in log points. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A4: Boom-exposure controls, definitions and summary statistics, 1970 to 2023.

Variable	Definition	Mean	SD	Crises covered
Money growth, 5y	5y change in money over GDP, pp	4.4	15.2	119 of 150
Investment change, 5y	5y change in investment share, pp	0.1	7.3	144 of 150
External balance, 3y	Average external balance, $t-1$ to $t-3$, pp	-4.5	13.1	147 of 150

Notes: Means and standard deviations are computed after winsorizing at the 2nd and 98th percentiles because the raw money series contains currency-splice artifacts in a few economies. Estimation standardizes the raw proxies within each sample, and Table A6 shows the results are unchanged when the proxies are winsorized instead. Coverage counts are out of the 150 non-borderline episodes of Table A5. Source: Global Macro Database.

Table A5: Sample accounting.

Stage	Events	Dropped at this stage
Episodes recorded, 1970–2025	164	–
Excluding borderline episodes	150	14 borderline
Onset by 2023, country in the GMD panel	150	0 outside the panel or after 2023
Output, lagged growth, and lagged debt available	140	10 missing debt or output data
Boom proxies available (restricted sample)	118	22 missing money, investment, or external data
Fiscal balance path available	105	13 missing budget data
After trimming extreme balance changes (mechanism)	102	3 trimmed at 1/99 percent
Policy rate path available	94	24 missing policy rates
After trimming extreme rate changes	87	7 trimmed at 1/99 percent

Notes: Events at the reference horizon of each specification, tracing the recorded episodes to the estimation samples.

Table A6: The debt interaction under alternative boom measures.

Boom measure	$h = 1$	$h = 3$	$h = 5$	$h = 7$
Baseline boom controls (21 crises)	-1.16 (1.71)	-1.64 (2.11)	-3.57 (2.77)	-6.23** (2.43)
Winsorized at 2/98 (21 crises)	-1.05 (1.67)	-0.90 (2.06)	-3.15 (2.66)	-5.78** (2.38)
Money depth in logs (21 crises)	-1.34 (1.65)	-1.45 (2.14)	-3.40 (2.69)	-5.91** (2.43)
JST loans to GDP (14 crises)	0.09 (2.68)	-1.88 (3.50)	-5.43 (3.64)	-6.43* (3.31)
WDI credit to GDP (20 crises)	-2.14 (1.71)	-2.57 (2.03)	-3.86 (2.78)	-6.02** (2.38)

Notes: Advanced economies, 1980 to 2023, in log points per standard deviation of debt. The log money-depth row is winsorized at 2/98, and the JST row uses the 18-economy loans-to-GDP series of Jordà et al. (2017). Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A7: Identification robustness for the debt interaction.

Specification	$h = 1$	$h = 3$	$h = 5$	$h = 7$
Advanced economies				
Baseline (21 crises)	-1.16 (1.71)	-1.64 (2.11)	-3.57 (2.77)	-6.23** (2.43)
Excluding euro area (11 crises)	-3.27*** (1.13)	-3.46** (1.58)	-4.84** (2.00)	-6.76*** (1.70)
Euro interaction (21 crises)	-1.37 (1.69)	-1.60 (2.10)	-2.85 (2.67)	-5.93** (2.31)
Peg interaction (21 crises)	-1.07 (1.57)	-2.20 (2.31)	-4.85 (2.95)	-7.11*** (2.75)
Including borderline episodes (26 crises)	-1.08 (1.73)	-1.61 (2.13)	-3.44 (2.77)	-6.17** (2.40)
EMDEs, and pooled where noted				
Baseline (96 crises)	-2.06 (3.54)	-1.86 (3.22)	0.03 (1.47)	0.20 (1.49)
Twin controls (96 crises)	-2.18 (3.65)	-1.95 (3.30)	0.02 (1.50)	0.20 (1.50)
Twin controls, pooled (118 crises)	-1.50 (3.06)	-1.25 (2.85)	0.32 (1.41)	0.36 (1.39)
Including borderline, pooled (128 crises)	-1.32 (2.94)	-1.27 (2.80)	0.23 (1.36)	0.43 (1.35)
Nonlinear debt terms				
Debt over 90 percent, pooled (119 crises)	–	-9.69 (8.49)	-5.35 (4.93)	-5.63 (5.01)
Top vs bottom tercile, pooled (119 crises)	–	-2.09 (5.57)	2.31 (4.04)	3.34 (4.26)
Debt over 90 percent, advanced (21 crises)	–	-8.26** (3.77)	-11.59** (5.44)	-15.06*** (5.08)
Top vs bottom tercile, advanced (21 crises)	–	-1.05 (3.89)	-4.12 (5.11)	-10.30** (4.76)

Notes: Boom-control specification, in log points per standard deviation of debt. Euro-area exclusion drops each euro member's crisis window. Interaction rows add the named indicator to the control set in levels and interacted with the onset. Borderline rows add the 14 borderline episodes to the event set. Twin-crisis controls are the GMD currency and sovereign flags. Nonlinear rows replace the standardized debt term with the named indicator, in log points, with $h = 1$ omitted. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A8: The indices in place of the debt ratio.

Sample	$h = 1$	$h = 3$	$h = 5$	$h = 7$
Buffers index in place of debt				
All countries (106 crises)	-1.85** (0.85)	-1.51 (1.29)	-0.90 (1.83)	0.27 (2.07)
Advanced (21 crises)	0.67 (1.19)	2.98 (1.86)	6.00** (2.48)	6.13** (2.53)
Effective Space Index, out of fold, in place of debt				
All countries (117 crises)	0.64 (0.92)	0.72 (1.10)	1.15 (1.43)	0.32 (1.49)
Advanced (21 crises)	-1.08 (1.57)	-3.46 (2.14)	-5.02* (2.82)	-6.98*** (2.55)

Notes: The interaction of the onset with the standardized lagged index, from equation (2) with boom controls on the adopted chronology, in log points per standard deviation of the index. A positive coefficient means more measured room predicts milder crises. The Effective Space Index rows use out-of-fold predictions, so no event's own outcome informs its score. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A9: The debt interaction under design-based estimators.

Estimator	$h = 3$	$h = 5$	$h = 7$
All countries			
Fixed-effects projection, boom controls	-1.23 (2.82)	0.31 (1.39)	0.33 (1.37)
Clean-control projection	-0.76 (1.41)	-1.60 (1.80)	-2.34 (2.22)
Doubly robust score regression	–	-1.14 (1.82)	–
Advanced economies			
Fixed-effects projection, boom controls	-1.64 (2.11)	-3.57 (2.77)	-6.23** (2.43)
Clean-control projection	-5.43 (4.16)	-6.24 (3.91)	-6.79 (5.01)
Doubly robust score regression	–	-16.97*** (3.00)	–

Notes: Boom-control sample. The doubly robust row regresses the event-level robust score on standardized debt at the five-year horizon. It is identified without country or year fixed effects, so its scale is not comparable to the projections, as Appendix B discusses. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A10: Robustness of the advanced-economy consolidation response.

Variant	$h = 1$	$h = 3$	$h = 5$
Cyclical-adjustment elasticity 0.3 (20 crises)	-0.54 (1.76)	3.92*** (0.80)	1.86** (0.78)
Cyclical-adjustment elasticity 0.5 (20 crises)	-0.24 (1.81)	4.43*** (1.08)	2.67** (1.27)
Cyclical-adjustment elasticity 0.7 (20 crises)	0.05 (1.88)	4.82*** (1.52)	2.47 (1.61)
Excluding the 2008–2012 wave (6 crises)	-1.62 (0.73)	4.90 (0.75)	3.14 (0.75)
Excluding Ireland and Iceland (18 crises)	-0.08 (1.89)	4.75*** (1.01)	3.27*** (1.15)
Pseudo-onset placebo (17 crises)	-0.01 (0.80)	0.53 (0.84)	0.40 (1.05)
Program-participation interaction (20 crises)	-0.56 (1.81)	3.65*** (0.83)	1.92 (1.18)

Notes: Cyclically adjusted balance, boom-control specification, 1980 to 2023, in percentage points. Country-clustered standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels. Rows with fewer than ten events carry no stars because clustered inference is unreliable there, and the placebo row, which reassigns each onset to a random no-crisis advanced economy in the same year, carries none by design.

Table A11: Onset regressions.

Specification	AME of +10pp debt	AUROC	Events	Obs.
All countries, 1976–2023				
Debt only	0.006 (0.012)	0.72	118	6717
Boom symptoms only	–	0.73	118	6717
Debt and boom symptoms	0.000 (0.014)	0.73	118	6717
Advanced, 1980–2023				
Debt only	0.030 (0.064)	0.76	21	1449
Boom symptoms only	–	0.79	21	1449
Debt and boom symptoms	0.022 (0.069)	0.79	21	1449
EMDEs, 1976–2023				
Debt only	0.005 (0.013)	0.75	96	5026
Boom symptoms only	–	0.75	96	5026
Debt and boom symptoms	0.003 (0.013)	0.75	96	5026

Notes: Logits of the banking-crisis onset on the standardized lagged debt ratio and the boom symptoms with decade dummies, on the common sample with all regressors available, 1976 to 2023. The AME column reports the average marginal effect of ten percentage points of debt on the annual crisis probability, in percentage points, with delta-method standard errors clustered by country. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table A12: Cross-validation against the fiscal space database.

Pair, 2019 cross section	Countries	Rank correlation
Debt component vs. Kose et al. government debt	154	+1.00
Balance component vs. Kose et al. fiscal balance	157	+0.97
Buffers index vs. foreign-currency debt share	45	+0.24
Buffers index vs. nonresident debt share	43	-0.15
Effective space vs. foreign-currency debt share	46	+0.15
Effective space vs. nonresident debt share	43	-0.20

Notes: Comparison with Kose et al. (2022), 2019 cross section. Spearman rank correlations, signed so that agreement on the concept is positive for the component rows. The rows are descriptive and carry no inference.

Table A13: The full 2019 rankings of both indices.

Country	PBI	ESI	Country	PBI	ESI	Country	PBI	ESI
UAE	1	1	Central African R.	49	67	Slovakia	97	129
Azerbaijan	2	11	Burkina Faso	50	40	Panama	98	51
Iraq	3	49	Mozambique	51	110	Mauritius	99	145
Kuwait	4	136	Eritrea	52	42	Hungary	100	98
Hong Kong SAR	5	2	Latvia	53	137	Rwanda	101	123
Switzerland	6	111	Poland	54	105	Cameroon	102	80
Saudi Arabia	7	92	Afghanistan	55	148	Canada	103	131
Bosnia and Herz.	8	102	Nicaragua	56	33	Myanmar	104	37
Cambodia	9	59	Congo DR	57	66	Sierra Leone	105	54
Denmark	10	56	Papua N. Guinea	58	7	Albania	106	135
Bulgaria	11	115	Serbia	59	138	France	107	132
Djibouti	12	23	Israel	60	62	Romania	108	139
Russia	13	38	Japan	61	117	Italy	109	99
Vietnam	14	9	Estonia	62	93	West Bank and Gaza	110	101
Timor-Leste	15	149	Moldova	63	121	Belgium	111	134
Thailand	16	64	Nigeria	64	15	Zambia	112	18
Singapore	17	153	Botswana	65	58	Malawi	113	90
Czech Republic	18	87	Madagascar	66	81	Cyprus	114	125
Korea	19	57	New Zealand	67	84	Colombia	115	85
Uzbekistan	20	142	Lesotho	68	43	Bangladesh	116	53
Qatar	21	71	Jordan	69	147	Spain	117	78
Peru	22	76	Indonesia	70	69	Bolivia	118	29
Norway	23	28	Angola	71	55	Ireland	119	77
Trinidad and Tob.	24	24	China	72	27	Oman	120	97
Belarus	25	45	El Salvador	73	100	South Africa	121	6
Kazakhstan	26	109	Mexico	74	95	Türkiye	122	50
Philippines	27	146	Austria	75	130	Greece	123	133
Uruguay	28	21	Armenia	76	108	Costa Rica	124	30
Sweden	29	79	Kosovo	77	152	United Kingdom	125	114
Paraguay	30	48	Slovenia	78	72	Ghana	126	5
Benin	31	82	Kyrgyz Republic	79	126	Uganda	127	35
Croatia	32	140	The Gambia	80	63	Kenya	128	3
Tanzania	33	16	Ecuador	81	47	Sri Lanka	129	46
Honduras	34	144	Algeria	82	127	South Sudan	130	36
Congo, Rep.	35	32	Ukraine	83	52	India	131	41
Tajikistan	36	128	Eq. Guinea	84	31	Argentina	132	44
Lithuania	37	150	Portugal	85	107	Tunisia	133	65
Gabon	38	17	Australia	86	94	Brazil	134	14
Georgia	39	104	Dominican Rep.	87	73	Pakistan	135	22
Malaysia	40	25	Nepal	88	106	Ethiopia	136	34
North Macedonia	41	141	Namibia	89	13	Bahrain	137	75
Guatemala	42	19	Chad	90	118	Lebanon	138	154
Germany	43	88	Mauritania	91	112	United States	139	122
Netherlands	44	86	Finland	92	119	Egypt	140	26
Eswatini	45	10	Chile	93	89	Lao P.D.R.	141	68
Guinea	46	91	Liberia	94	113	Venezuela	142	–
Haiti	47	83	Jamaica	95	151	Yemen	143	8
Mongolia	48	124	Morocco	96	74	Burundi	144	120

Notes: Sorted by the buffers index. Columns report ordinal ranks, and on both indices lower ranks mean more room, so rank 1 is the most room to respond. ESI ranks are among the 154 countries the Effective Space Index covers, so they can run above 144, and a dash means the country is not covered. The Spearman rank correlation between the two across the 143 countries both cover is −0.01, a number to read against the index's split-half reliability reported in Section 5. The ESI column is illustrative, per the diagnosis in Appendix B.

Table A14: Policy Buffers Index rankings, 2019.

	Most room to respond			Least room to respond	
Rank	Country	PBI	Rank	Country	PBI
Panel A. All countries above one million population (144 ranked)					
1	UAE	84	135	Pakistan	30
2	Azerbaijan	82	136	Ethiopia	28
3	Iraq	77	137	Bahrain	26
4	Kuwait	77	138	Lebanon	26
5	Hong Kong SAR	76	139	United States	26
6	Switzerland	76	140	Egypt	25
7	Saudi Arabia	74	141	Lao P.D.R.	24
8	Bosnia and Herz.	74	142	Venezuela	22
9	Cambodia	73	143	Yemen	17
10	Denmark	72	144	Burundi	15
Panel B. Economies above thirty million population (45 ranked)					
1	Iraq	77	38	India	33
2	Saudi Arabia	74	39	Argentina	33
3	Russia	70	40	Brazil	33
4	Vietnam	70	41	Pakistan	30
5	Thailand	69	42	Ethiopia	28
6	Korea	66	43	United States	26
7	Uzbekistan	66	44	Egypt	25
8	Peru	64	45	Yemen	17
Panel C. Leaders within income groups					
High income			*Upper middle income*		
1	UAE	84	1	Azerbaijan	82
2	Kuwait	77	2	Iraq	77
3	Hong Kong SAR	76	3	Bosnia and Herz.	74
4	Switzerland	76	4	Bulgaria	70
Lower middle income			*Low income*		
1	Cambodia	73	1	Tajikistan	57
2	Djibouti	70	2	Guinea	55
3	Vietnam	70	3	Haiti	55
4	Timor-Leste	70	4	Central African R.	54

Notes: Within-year percentile scores in 2019. Higher scores and lower ranks mean more room to respond, so rank 1 is the country with the most room. Countries reporting fewer than four of the six components are not ranked. Source: author's calculations.