Supplementary Appendix A

Data, Sample Construction, and Fiscal-Vulnerability Measurement

Supplement to "Does Outcome Measurement Matter for Forecasting? Evidence from Fiscal Vulnerability in Developing Economies"

This Supplementary Appendix documents the data and measurement procedures underlying the main manuscript. It provides additional information on the Global Macro Database source data, sample construction, variable definitions and transformations, and the construction of the alternative fiscal-vulnerability measures used in the empirical analysis.

The appendix is organized as follows. Section A.1 documents the data source, sample construction, panel coverage, and treatment of forecast and missing observations. Section A.2 describes the variables, transformations, and information-timing conventions used to construct the analytical dataset. Section A.3 provides detailed implementation of the breadth, severity, and PCA-based vulnerability measures, including the development-sample thresholds, scaling procedures, and statistical-component construction. Section A.4 reports supplementary measurement diagnostics, including component incidence, temporal and income-group variation, and historical-consistency comparisons with independently recorded banking, sovereign-debt, and currency-crisis episodes.

The supplementary material is intended to make the construction of the analytical sample and vulnerability measures transparent and reproducible. The three vulnerability measures are alternative aggregations of the same underlying eight fiscal, macroeconomic, and external dimensions. Robustness across these measures therefore assesses sensitivity to alternative constructions within a common multidimensional measurement framework; it should not be interpreted as independent validation against realised fiscal distress.

A.1 Data source and sample construction

The analysis uses the June 2026 release of the Global Macro Database (GMD, version 2026_06). The source database contains 56,992 observations, 162 variables, and 239 countries or territories over its maximum historical coverage. The sample is restricted to economies classified in the release as low-, lower-middle-, or upper-middle-income. These income-group assignments are static release-level classifications and are used only as descriptive strata; they are not treated as time-varying historical income classifications.

The initial developing-economy window comprises 129 economies and 5,805 potential country-year observations over 1980–2024. Construction of the vulnerability measures requires contemporaneous observations for all eight underlying fiscal, macroeconomic, and external dimensions. Applying this complete-outcome requirement yields an analytical panel of 4,527 country-year observations from 127 economies over 1982–2024. The panel is intentionally unbalanced, with the number of observed economies varying across years according to source-data availability.

Table A1.1. Sample-period summary before complete-outcome restrictions

Sample window	Period	Potential observations	Economies
Developing-economy observed window	1980–2024	5,805	129
Development/model selection	1982–2005	3,096	129
Calibration	2006–2009	516	129
Internal evaluation	2010–2017	1,032	129
External evaluation	2018–2023	774	129
2024 sensitivity	2024	129	129

Notes: Counts refer to potential country-year observations before imposing complete contemporaneous availability of the eight variables used in vulnerability measurement. The analytical sample is smaller because outcome availability varies across countries and years.

The analytical sample is partitioned chronologically according to its role in model development and evaluation. The development period spans 1982–2005. Observations from 2006–2009 form the calibration sample for the internal forecasting exercise, while 2010–2017 constitutes the held-out internal evaluation period. For the external exercise, the locked forecasting architecture is re-estimated using information available through 2013, probabilities are recalibrated over 2014-2017, and performance is evaluated over 2018–2023. Observations for 2024 are reserved for a post-evaluation sensitivity exercise and are not included in the principal external evaluation.

Table A1.2. Analytical sample by temporal role

Analysis block	Period	Observations	Economies	Analytical role
Development/model selection	1982–2005	2,219	123	Recursive model selection
Calibration	2006–2009	493	124	Probability calibration
Internal evaluation	2010–2017	989	125	Held-out temporal evaluation
External evaluation	2018–2023	719	124	Principal external evaluation
2024 sensitivity	2024	107	107	Post-evaluation sensitivity

Notes: Analytical observations require complete contemporaneous data for all eight variables entering vulnerability measurement. Economy counts differ across periods because the panel is unbalanced. The 2024 observations are excluded from the principal external evaluation.

The data-processing protocol preserves the information structure of an annual prospective forecasting exercise. Observations identified by the GMD as forecasts are set to missing before annual changes, lagged variables, vulnerability thresholds, scaling parameters, or outcome measures are constructed. Published source-database forecasts are therefore not treated as realised historical observations. No interpolation, forward filling, or backward filling is applied.

Annual transformations and one-year lags are retained only when observations for the same economy occur in consecutive calendar years. For a generic variable x_{it}, an annual transformation is defined as

$$\Delta x_{it} = \begin{cases} g(x_{it}, x_{i,t-1}), & \text{if observations exist in both } t \text{ and } t-1, \\ \text{missing}, & \text{otherwise,} \end{cases}$$

where $g(\cdot)$ denotes the relevant growth-rate or change transformation. Likewise, $x_{i,t-1}$ is retained as a one-year lag only when the immediately preceding observation for economy i corresponds to calendar year $t-1$. Transformations are therefore never calculated across multi-year gaps.

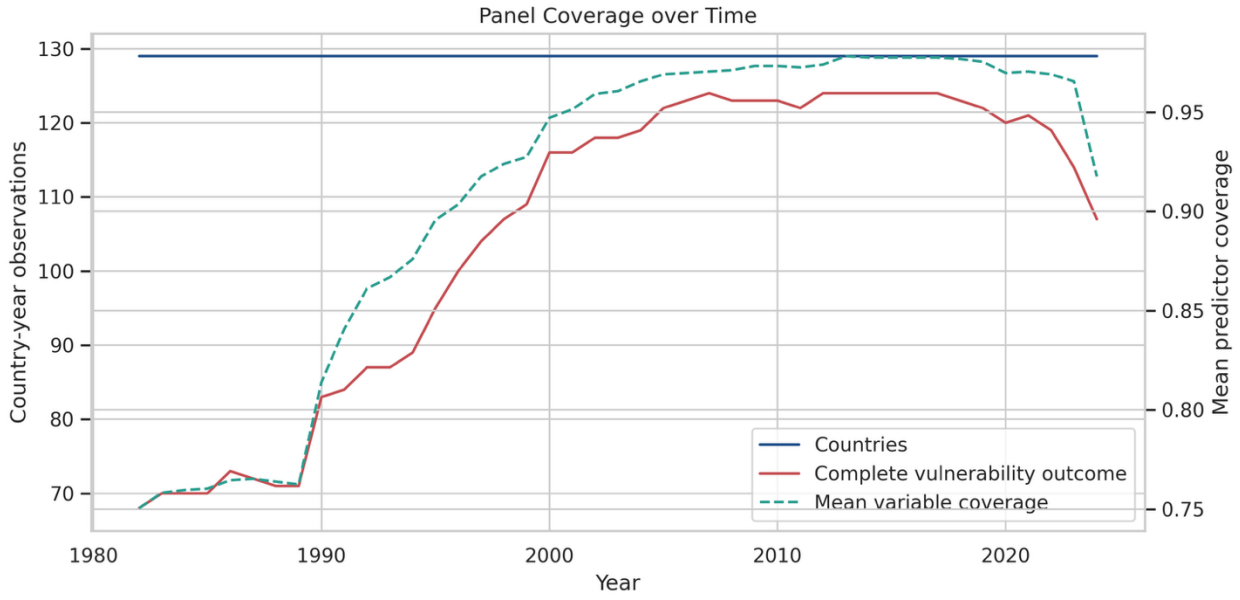


Figure A1.1. Panel coverage over time

Notes: The figure reports annual analytical-sample coverage after imposing complete contemporaneous availability of the eight vulnerability dimensions. Variation in coverage reflects the unbalanced structure of the source data. No interpolation or sample balancing is used.

A.2 Variable definitions, transformations, and information timing

The empirical dataset distinguishes between the contemporaneous variables used to construct fiscal-vulnerability outcomes and the lagged variables available to the forecasting model. The vulnerability measures are constructed from eight fiscal, macroeconomic, and external-sector dimensions observed in year t, while the forecasting information set contains eleven variables observed in year $t-1$. Table A2.1 summarizes the economic concepts, analytical variables, source mappings, units, forecasting roles, and adversity directions used in the measurement stage.

Let

$$\mathbf{Z}_{it} = (D_{it}, B_{it}, R_{it}, E_{it}, G_{it}, \pi_{it}, CA_{it}, FX_{it})'$$

denote the eight-dimensional contemporaneous vector used in vulnerability measurement, where D is government debt, B is net lending or borrowing, R is government revenue, E is government expenditure, G is real GDP growth, π is consumer-price inflation, CA is the current-account balance, and FX is nominal exchange-rate depreciation. As shown in Table A2.1, the adverse direction differs across components: higher debt, expenditure, inflation, and exchange-rate

depreciation are treated as adverse in the upper tail, whereas weaker fiscal balances, revenue, real GDP growth, and current-account balances are treated as adverse in the lower tail.

The one-year-ahead forecasting vector is

$$\mathbf{X}_{i,t-1} = (D, B, R, E, G, \pi, CA, FX, \Delta REER, EX, IM)'_{i,t-1},$$

where $\Delta REER$ denotes the annual percentage change in the real effective exchange rate and EX and IM denote exports and imports as percentages of GDP. The final three variables expand the forecasting information set but do not enter the construction of the vulnerability outcomes. Thus, all variables used to predict vulnerability in year t are dated $t-1$, whereas the outcome itself is constructed from contemporaneous year-t conditions.

Baseline fiscal variables use the consolidated or harmonised government series supplied in the GMD. Government debt, net lending or borrowing, revenue, and expenditure are expressed as percentages of GDP. The analytical variable govdef_GDP represents net lending or borrowing rather than the absolute size of a fiscal deficit; lower and more negative values therefore correspond to larger deficits. General-government and central-government counterparts are retained separately for robustness analysis.

Real GDP growth, nominal exchange-rate depreciation, and the change in the real effective exchange rate are derived from the corresponding level series. For any positive level variable x_{it}, the annual percentage change is calculated as

$$g(x_{it}) = 100\left(\frac{x_{it}}{x_{i,t-1}} - 1\right),$$

provided that observations exist for both t and $t-1$ and that the observations correspond to consecutive calendar years. Applying this transformation gives

$$G_{it} = 100\left(\frac{rGDP_{it}}{rGDP_{i,t-1}} - 1\right),$$

$$FX_{it} = 100\left(\frac{USDfx_{it}}{USDfx_{i,t-1}} - 1\right),$$

and

$$\Delta REER_{it} = 100\left(\frac{REER_{it}}{REER_{i,t-1}} - 1\right).$$

The nominal exchange rate is measured as units of local currency per US dollar. Positive values of FX_{it} therefore indicate nominal depreciation, while negative values indicate appreciation. The real effective exchange-rate change is used only as an additional forecasting variable and is not one of the eight dimensions entering vulnerability measurement.

Table A2.1. Variables used in vulnerability measurement and forecasting

Economic concept	Analytical variable	GMD source or construction	Unit	Vulnerability dimension	Forecast predictor	Adverse direction in vulnerability construction
Government debt	govdebt_GDP	GMD	% of GDP	Yes	$t-1$	Higher
Net lending/borrowing	govdef_GDP	GMD	% of GDP	Yes	$t-1$	Lower
Government revenue	govrev_GDP	GMD	% of GDP	Yes	$t-1$	Lower
Government expenditure	govexp_GDP	GMD	% of GDP	Yes	$t-1$	Higher
Real GDP growth	rgdp_growth	Derived from rGDP	Annual % change	Yes	$t-1$	Lower
Consumer-price inflation	infl	GMD	Annual %	Yes	$t-1$	Higher
Current-account balance	CA_GDP	GMD	% of GDP	Yes	$t-1$	Lower
Nominal exchange-rate depreciation	fx_depreciation	Derived from USDfx	Annual % change	Yes	$t-1$	Higher
Real effective exchange-rate change	reer_change	Derived from REER	Annual % change	No	$t-1$	—
Exports	exports_GDP	GMD	% of GDP	No	$t-1$	—
Imports	imports_GDP	GMD	% of GDP	No	$t-1$	—

Notes: The eight variables marked as vulnerability dimensions constitute the common information set used to construct the breadth, severity, and PCA-based outcomes. The final three variables enter only the forecasting information set. Baseline fiscal variables use the consolidated or harmonised GMD government concepts. Alternative government-sector definitions are examined separately as robustness checks.

One-year lags are constructed by country after the annual transformations have been generated. For predictor x_{it},

$$L1(x_{it}) = x_{i,t-1}$$

only when the immediately preceding record corresponds to calendar year $t-1$. If an economy has a gap in its annual series, the lag is set to missing rather than carried across the gap. This restriction applies both to variables observed directly in the GMD and to variables derived from annual changes.

Complete contemporaneous observation of all eight vulnerability dimensions is required to construct an outcome for year t. Predictor completeness is not imposed as an additional sample-selection criterion. Country-years with a valid vulnerability outcome can therefore remain in the analytical panel when one or more lagged predictors are unavailable. Predictor missingness is handled subsequently within the forecasting pipeline using preprocessing parameters estimated only from the corresponding historical training sample.

Table A2.2 reports the extent of missingness in the contemporaneous model variables and their one-year lags before the complete-outcome restriction is imposed. Missingness is concentrated in the fiscal variables and the real effective exchange-rate series, whereas real GDP growth, trade variables, and nominal exchange-rate depreciation have substantially higher coverage. The same pattern is shown graphically in Figure A2.1, which makes clear that lag construction adds some missingness because a valid one-year lag requires both an observed predecessor and calendar-year continuity.

Table A2.2. Model-variable missingness in the developing-economy window

Variable	Missing observations	Missing share
L1_govdef_GDP	873	0.1574
reer_change	861	0.1552
L1_govrev_GDP	829	0.1495
govdef_GDP	825	0.1487
L1_reer_change	823	0.1484
L1_govexp_GDP	817	0.1473
govrev_GDP	783	0.1412
govexp_GDP	773	0.1394
L1_govdebt_GDP	766	0.1381
govdebt_GDP	724	0.1305
L1_CA_GDP	577	0.1040
CA_GDP	564	0.1017
L1_infl	440	0.0793
infl	421	0.0759
L1_fx_depreciation	204	0.0368
fx_depreciation	188	0.0339

Variable	Missing observations	Missing share
L1_imports_GDP	185	0.0334
L1_exports_GDP	185	0.0334
imports_GDP	166	0.0299
exports_GDP	166	0.0299
L1_rgdp_growth	75	0.0135
rgdp_growth	62	0.0112

Notes: Missing shares are calculated over potential developing-economy country-year observations for 1982–2024 before imposing complete contemporaneous outcome availability. Lagged variables additionally reflect the requirement that the preceding observation be available in the immediately preceding calendar year.

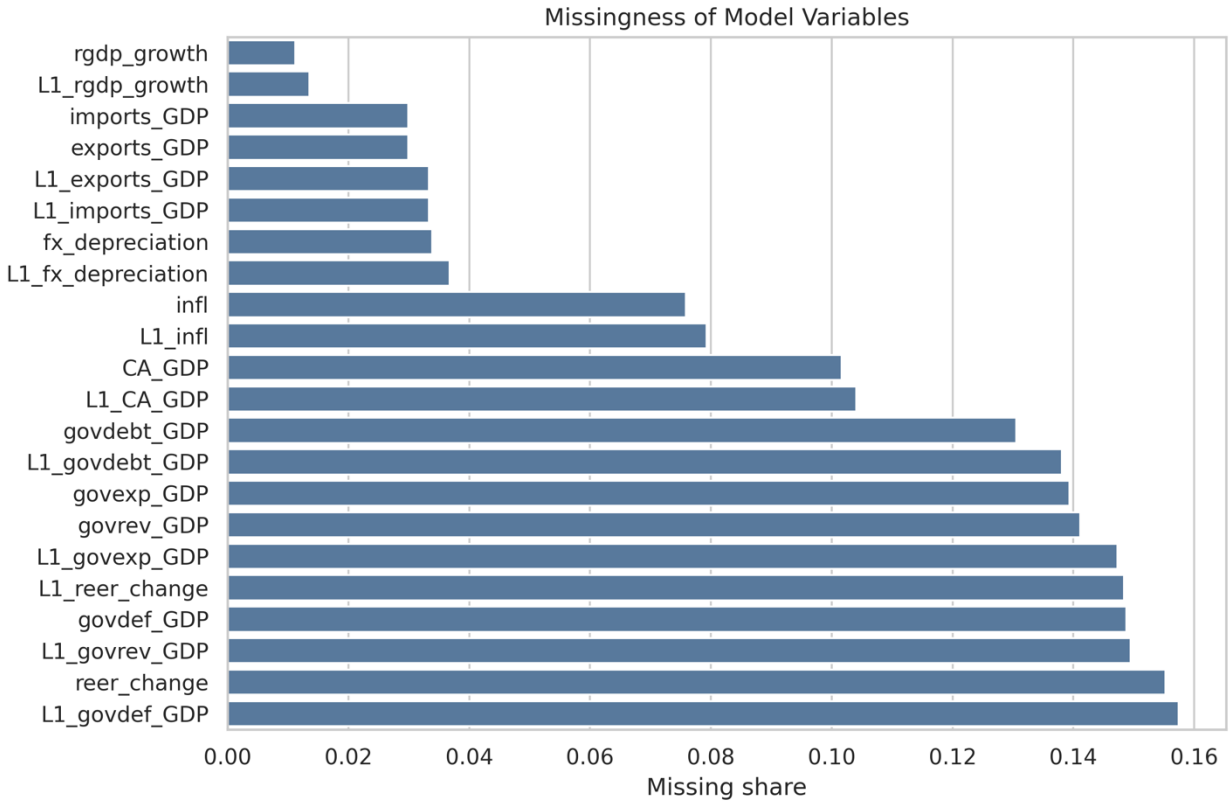


Figure A2.1. Model-variable missingness

A.3 Construction of the alternative fiscal-vulnerability measures

Fiscal vulnerability is measured using three alternative constructions applied to the same eight-dimensional information set: a breadth measure, a threshold-distance severity measure, and a PCA-based statistical component. Holding the underlying dimensions fixed allows differences in subsequent forecasting results to be attributed to the aggregation rule rather than to changes in the economic information entering the outcome.

Let Z_{kit} denote vulnerability dimension $k \in \{1, \dots, 8\}$ for economy i in year t, and let $\mathcal{D}_0$ denote the 1982-2005 development sample. The eight dimensions are government debt, net lending or borrowing, government revenue, government expenditure, real GDP growth, consumer-price inflation, the current-account balance, and nominal exchange-rate depreciation.

Define the adversity-orientation parameter

$$d_k = \begin{cases} +1, & \text{if higher values of dimension } k \text{ are adverse,} \\ -1, & \text{if lower values of dimension } k \text{ are adverse.} \end{cases}$$

Accordingly, $d_k = +1$ for government debt, government expenditure, inflation, and nominal exchange-rate depreciation, while $d_k = -1$ for net lending or borrowing, government revenue, real GDP growth, and the current-account balance.

For the baseline measurement system, all thresholds and other distributional parameters are estimated from $\mathcal{D}_0$ and subsequently held fixed. During recursive model selection, the same construction procedure is repeated using only the observations contained in each historical training fold, so that validation observations do not contribute to their own outcome definition.

A.3.1 Breadth of simultaneous vulnerability

The breadth measure records the number of dimensions simultaneously located in an adverse tail of the development-sample distribution. For component k, the adverse threshold is

$$c_k = \begin{cases} Q_{0.75}(Z_k \mid \mathcal{D}_0), & d_k = +1, \\ Q_{0.25}(Z_k \mid \mathcal{D}_0), & d_k = -1, \end{cases}$$

where $Q_p(\cdot)$ denotes the empirical p-th quantile.

The component-level adverse-state indicator is

$$R_{kit} = \mathbf{1}\{d_k(Z_{kit} - c_k) \geq 0\}.$$

The breadth score is then

$$B_{it} = \sum_{k=1}^{8} R_{kit}\,, \qquad B_{it} \in \{0, \ldots, 8\}.$$

The baseline binary breadth outcome is

$$V_{it}^{B} = \mathbf{1}\{B_{it} \geq 3\}.$$

Thus, a country-year is classified as breadth-vulnerable when at least three of the eight dimensions simultaneously cross their adverse development-sample thresholds. Alternative cut-offs of $B_{it} \geq 2$ and $B_{it} \geq 4$ are retained as measurement-sensitivity specifications.

Table A3.1 reports the numerical thresholds estimated from the 1982-2005 development sample. The thresholds are distributional reference points rather than normative fiscal limits, sustainability thresholds, or causal tipping points.

Table A3.1. Development-sample thresholds for the breadth measure

Vulnerability dimension	Analytical variable	Adverse direction	Development quantile	Threshold
High government debt	govdebt_GDP	Higher	0.75	80.31
Large fiscal deficit	govdef_GDP	Lower	0.25	-5.629
Low government revenue	govrev_GDP	Lower	0.25	13.91
High government expenditure	govexp_GDP	Higher	0.75	29.64
Weak real GDP growth	rgdp_growth	Lower	0.25	1.345
High inflation	infl	Higher	0.75	15.08
Current-account pressure	CA_GDP	Lower	0.25	-7.432
Exchange-rate pressure	fx_depreciation	Higher	0.75	15.01

Notes: Thresholds are estimated from complete-component observations in the 1982–2005 development sample and are fixed for subsequent descriptive and evaluation periods. Fiscal variables and the current account are expressed as percentages of GDP; growth, inflation, and exchange-rate depreciation are percentages.

The distribution of B_{it} is reported in Table A3.2 and Figure A3.1. Most country-years activate relatively few adverse dimensions, although a substantial upper tail exhibits simultaneous stress across three or more indicators. The baseline threshold $B_{it} \geq 3$ therefore identifies a subset of observations characterized by multidimensional rather than isolated adverse conditions.

Table A3.2. Distribution of the breadth score

Breadth score B_{it}	Observations	Share
0	904	0.1997
1	1,364	0.3013
2	998	0.2205

Breadth score B_{it}	Observations	Share
3	656	0.1449
4	369	0.0815
5	158	0.0349
6	57	0.0126
7	21	0.0046

Notes: Shares are calculated over the 4,527 analytical country-year observations with complete data for all eight vulnerability dimensions. No observation in the analytical sample has all eight dimensions simultaneously flagged.

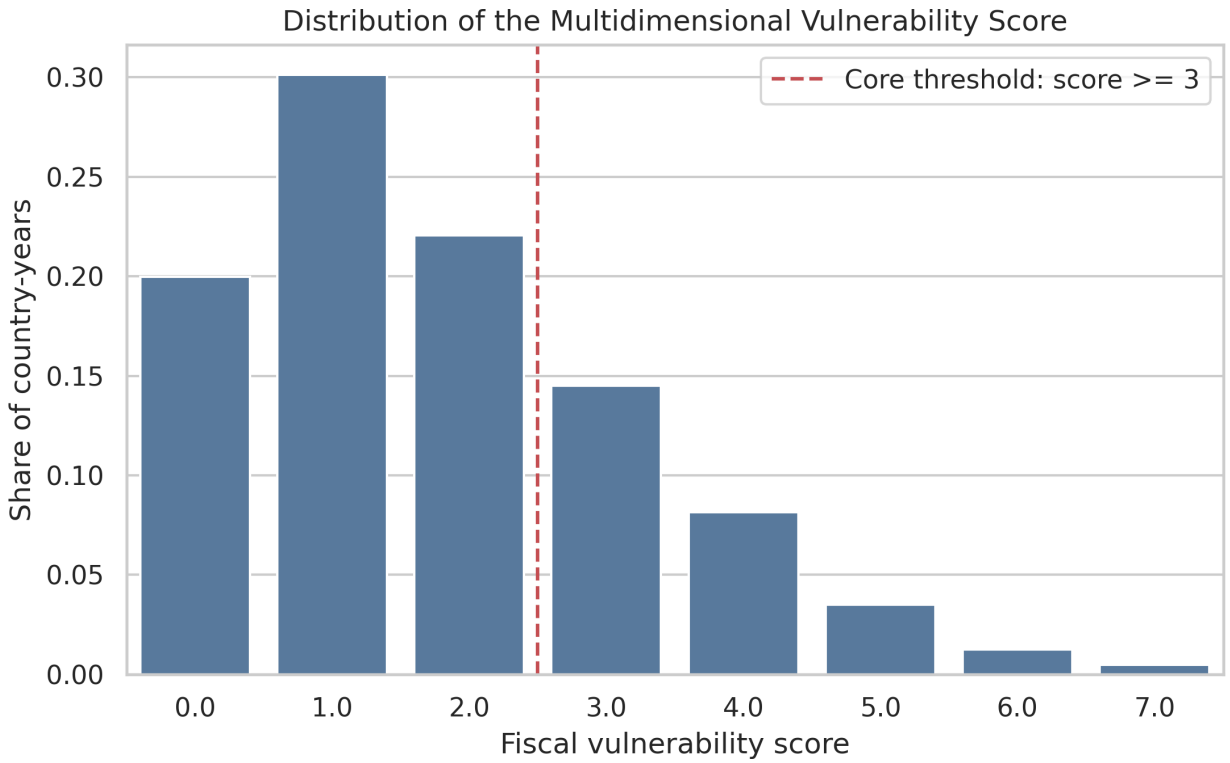


Figure A3.1. Vulnerability-score distribution

Component incidence is heterogeneous across the eight dimensions. As reported in Table A3.3 and illustrated in Figure A3.2, high expenditure and current-account pressure are the most frequently activated components, while high inflation and exchange-rate pressure occur less frequently. The unequal ex post incidence is expected because the thresholds are estimated from an earlier development period and then applied to a longer panel whose marginal distributions evolve over time.

Table A3.3. Component-level adverse-state prevalence

Vulnerability component	Share flagged	Flagged observations
High expenditure	0.3086	1,397
Current-account pressure	0.2878	1,303
Weak growth	0.2322	1,051
Large deficit	0.2275	1,030
Low revenue	0.1970	892
High debt	0.1873	848
Exchange-rate pressure	0.1754	794
High inflation	0.1696	768

Notes: Component indicators use the fixed development-sample thresholds reported in Table A3.1 and are evaluated over the full analytical panel.

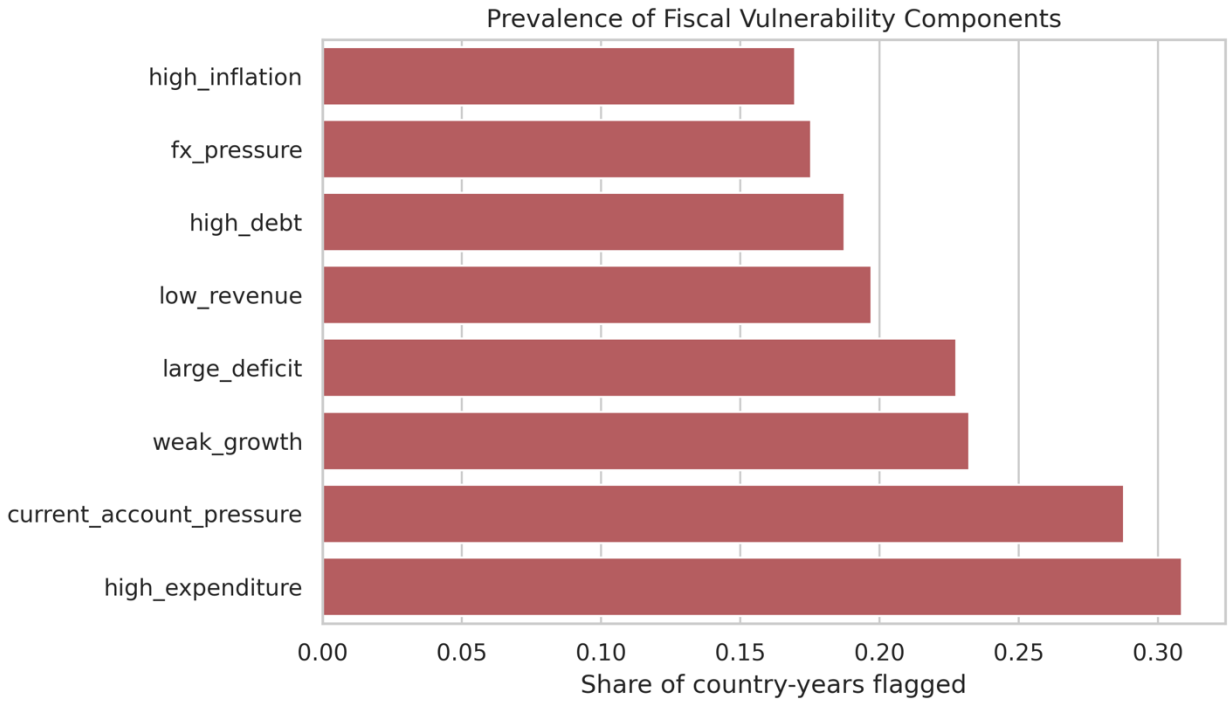


Figure A3.2. Component-level risk-flag prevalence

A.3.2 Threshold-distance severity

The breadth measure treats all threshold exceedances equally once a component becomes active. The severity construction instead incorporates the distance by which each observation moves beyond its adverse threshold. For dimension k, define the positive adverse distance

$$H_{kit} = \left[\frac{d_k(Z_{kit} - c_k)}{\sigma_{k,0}}\right]_+,$$

where

$$[x]_+ = \max(x, 0)$$

and $\sigma_{k,0}$ is the standard deviation of dimension \(k\) estimated from the development sample.

The continuous severity index is

$$S_{it} = \frac{1}{8}\sum_{k=1}^{8} H_{kit}.$$

Consequently, observations that remain on the non-adverse side of a threshold contribute zero, whereas observations beyond the threshold contribute in proportion to their standardized distance from that threshold. Equal weighting across the eight standardized distances is retained so that no component weight is chosen on the basis of forecasting performance.

The binary severity outcome is defined by the upper quartile of the severity distribution in the development sample:

$$V_{it}^{S} = \mathbf{1}\{S_{it} \geq s_{0.75}\},$$

where

$$s_{0.75} = Q_{0.75}(S_{it} \mid \mathcal{D}_0).$$

The construction therefore differs from breadth in two respects. First, it preserves information about the magnitude of adverse threshold exceedances. Second, the final binary classification is based on the development-sample distribution of the resulting continuous severity score rather than on a fixed number of activated dimensions.

Table A3.4 reports the development-sample scale parameters used in the baseline severity measure. The large dispersion of inflation and nominal exchange-rate depreciation illustrates why the severity definition is additionally evaluated using robust scale estimators in the sensitivity analysis.

Table A3.4. Development-sample scale parameters for the severity measure

Vulnerability dimension	Development-sample standard deviation
Government debt/GDP	70.20
Net lending/borrowing/GDP	19.49
Government revenue/GDP	14.99
Government expenditure/GDP	24.82
Real GDP growth	5.69
Inflation	485.45
Current-account balance/GDP	11.79
Nominal exchange-rate depreciation	110,058.09

Notes: Scale parameters are estimated from the 1982–2005 development sample and are fixed when the baseline severity measure is applied subsequently. Because conventional standard deviations can be influenced by extreme observations, robustness analysis also considers severity measures based on robust dispersion estimates.

A.3.3 PCA-based statistical component

The third construction summarizes common variation across the same eight dimensions using principal component analysis. Before estimation, each variable is oriented so that the sign convention is comparable across indicators:

$$A_{kit} = d_k Z_{kit}.$$

The oriented variables are standardized using development-sample location and scale parameters,

$$\tilde{Z}_{kit} = \frac{A_{kit} - \mu^a_{k,0}}{\sigma^a_{k,0}},$$

where $\mu^a_{k,0}$ and $\sigma^a_{k,0}$ are respectively the mean and standard deviation of the adversity-oriented variable in the development sample.

Let

$$\tilde{\mathbf{Z}}_{it} = \left(\tilde{Z}_{1it}, \ldots, \tilde{Z}_{8it}\right)'.$$

Principal component analysis is estimated using $\tilde{\mathbf{Z}}_{it}$ in the development sample. If $\hat{\boldsymbol{\omega}}_1$ denotes the loading vector associated with the largest eigenvalue of the development-sample correlation matrix, the first component score is

$$P_{it} = \hat{\boldsymbol{\omega}}_1{}' \tilde{\mathbf{Z}}_{it}.$$

Because the sign of a principal component is algebraically arbitrary, the loading vector is sign-normalized so that P_{it} is positively correlated with the breadth score B_{it} in the development sample. The estimated loadings are then held fixed for subsequent observations.

The PCA-based binary outcome is

$$V_{it}^{P} = \mathbf{1}\{P_{it} \geq p_{0.75}\},$$

where

$$p_{0.75} = Q_{0.75}(P_{it} \mid \mathcal{D}_0).$$

The first principal component accounts for approximately 26.0% of the standardized variance in the eight-dimensional development-sample indicator set. Table A3.5 reports its loading structure.

Table A3.5. Development-sample loadings of the first principal component

Dimension	Original adverse direction	PC1 loading after adversity orientation
Government debt	Higher	0.124
Net lending/borrowing	Lower	0.566
Government revenue	Lower	-0.415
Government expenditure	Higher	0.684
Real GDP growth	Lower	-0.086
Inflation	Higher	0.017
Current-account balance	Lower	0.128
Nominal exchange-rate depreciation	Higher	-0.025

Notes: PCA is estimated from adversity-oriented and standardized variables using the 1982–2005 development sample. The sign of the first component is normalized to produce a positive development-sample correlation with the breadth score. PC1 explains approximately 26.0% of total standardized variation.

The mixed loading signs in Table A3.5 are substantively important. Even after the original indicators are oriented toward adversity, the first principal component is not a monotone equally signed combination of all eight dimensions. It therefore captures the dominant covariance pattern in the indicator set rather than an economic factor for which deterioration in every underlying variable necessarily increases the component score. For this reason, the PCA construction is interpreted throughout as a latent statistical component, not as a structural fiscal-vulnerability factor.

A.3.4 Relationship among the three constructions

The three outcome definitions deliberately preserve different aspects of the same underlying information. Breadth emphasizes the simultaneous occurrence of adverse states but discards the

magnitude of threshold exceedances. Severity retains information on the standardized distance beyond adverse thresholds while preserving equal treatment of the eight dimensions. PCA instead extracts the dominant linear covariance pattern and does not impose monotonicity across the individual dimensions.

Accordingly, the three measures are neither interchangeable nor independent validation targets. Their common eight-dimensional basis provides a controlled measurement-robustness exercise: differences in incidence, persistence, forecastability, or country rankings arise from alternative aggregation and scaling of the same underlying fiscal, macroeconomic, and external conditions. Subsequent forecasting comparisons therefore assess whether empirical conclusions survive these alternative representations of fiscal vulnerability rather than whether one constructed measure is externally validated by another.

A.4 Distributional diagnostics, cross-sectional variation, and historical consistency

This section reports supplementary diagnostics for the variables and vulnerability measures used in the analysis. The diagnostics are descriptive and are not used to select predictors, determine model specifications, or alter the vulnerability definitions. They provide evidence on the shape and dependence of the underlying macro-fiscal variables, variation in vulnerability across static income groups and time, and the correspondence between the constructed breadth measure and independently recorded crisis episodes.

Figure A4.1 summarizes the distributions of the eleven variables entering the forecasting information set. For visual presentation, each distribution is displayed within its empirical 1st–99th percentile range so that the central mass of the data remains visible in the presence of extreme observations. The distributions show substantial heterogeneity in scale and shape. Government debt, expenditure, exports, and imports exhibit pronounced right skewness, while real GDP growth and the current-account balance are more symmetrically concentrated around their central ranges. Inflation and nominal exchange-rate depreciation display particularly long upper tails. These features motivate the use of fold-specific preprocessing in the forecasting analysis but do not affect the construction of the baseline breadth indicators, which depends only on development-sample quantiles.

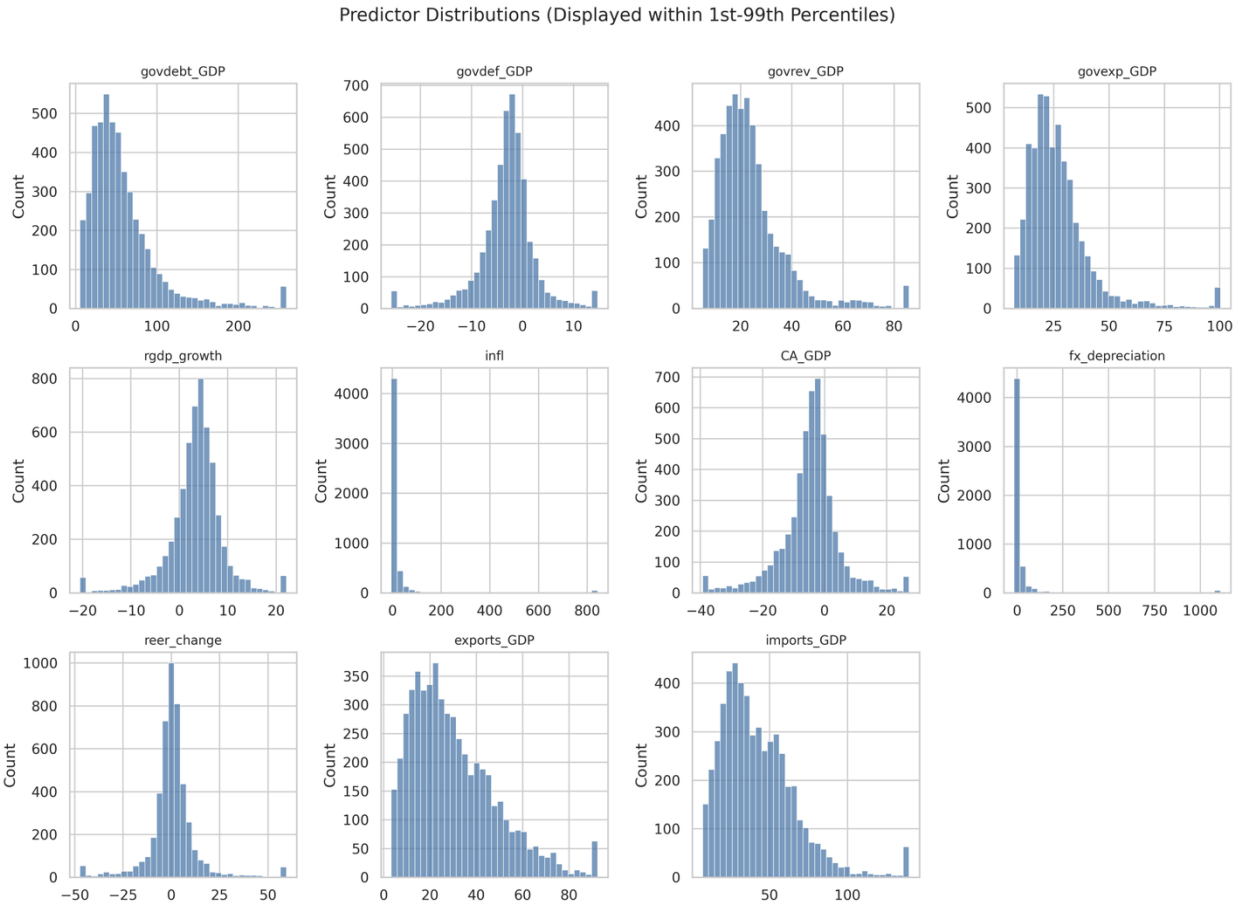


Figure A4.1. Predictor distributions

Notes: Histograms are displayed within the empirical 1st–99th percentile range for visualization. The display restriction does not redefine the analytical variables or vulnerability outcomes.

Dependence among the model variables is summarized by the Spearman rank-correlation matrix in Figure A4.2. The strongest positive associations occur among government revenue, government expenditure, exports, and imports, while nominal exchange-rate depreciation is negatively associated with the change in the real effective exchange rate. Most other pairwise correlations are moderate. The correlation matrix is used only as a descriptive diagnostic; predictors are not mechanically excluded on the basis of pairwise dependence.

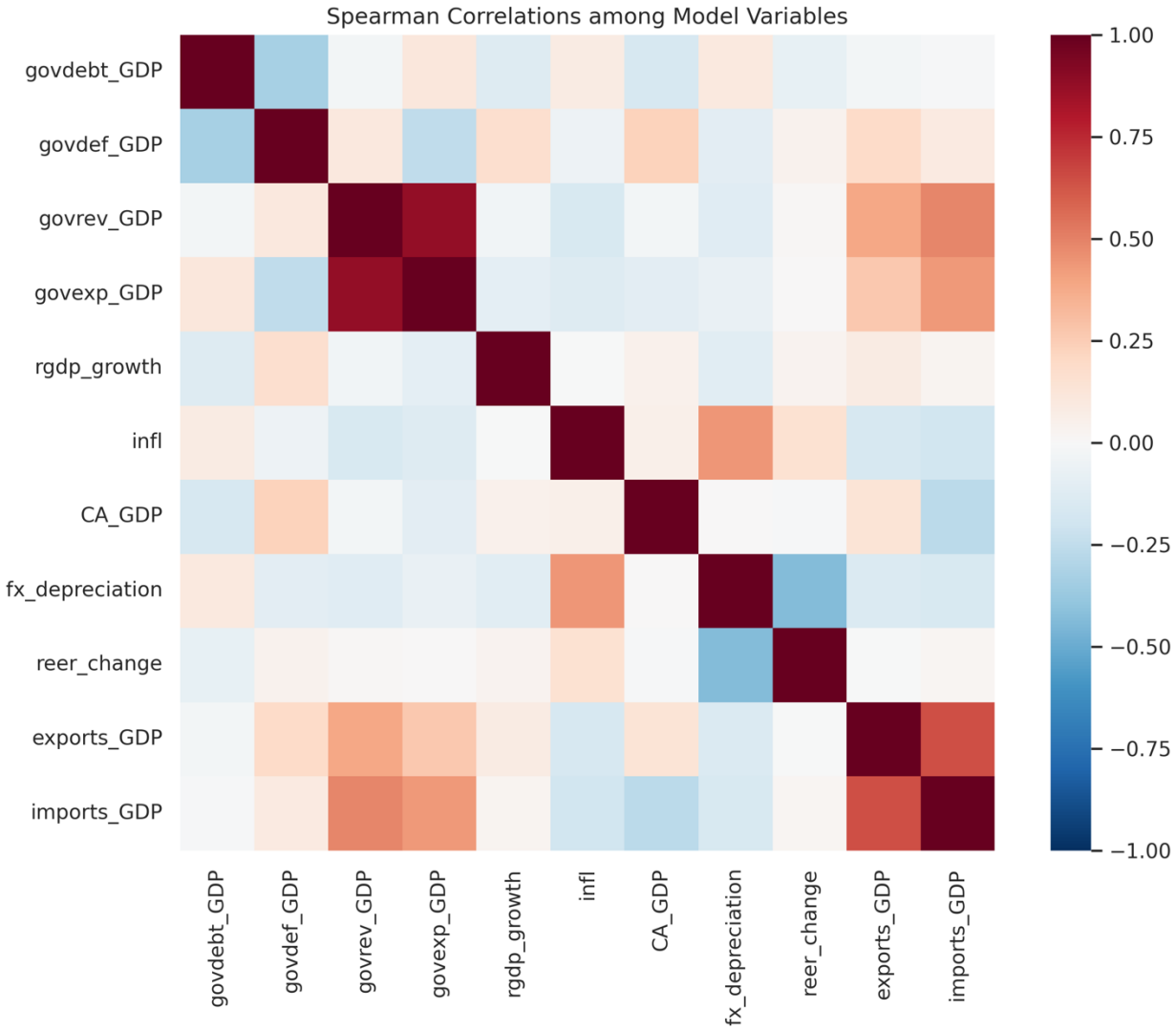


Figure A4.2. Spearman correlations among model variables

Notes: *Entries are pairwise Spearman rank correlations calculated from available observations. The matrix is descriptive and is not used as a variable-selection rule.*

The incidence of vulnerability differs across the static GMD income groups. Table A4.1 reports the three principal vulnerability definitions together with selected components of the breadth measure. Breadth- and severity-based vulnerability are most prevalent among low-income economies and decline across the lower-middle- and upper-middle-income groups. The PCA-based state displays a different cross-sectional ordering, with the highest incidence among upper-middle-income economies. This reversal illustrates why the PCA construction is treated as a statistical component rather than as an interchangeable representation of the breadth measure.

Table A4.1. Vulnerability incidence by static income group

Income group	Economies	Observations	Breadth vulnerable (%)	Severity vulnerable (%)	PCA vulnerable (%)	Mean breadth score	High debt (%)	Large deficit (%)	Low revenue (%)
Low income	24	819	38.2	28.9	18.4	2.20	32.5	26.0	35.3
Lower-middle income	50	1,846	26.7	24.0	26.5	1.79	20.7	26.5	24.2
Upper-middle income	53	1,862	24.4	20.8	37.4	1.60	10.7	17.6	8.4
Total	127	4,527	27.9	23.6	29.5	1.79	18.7	22.8	19.7

Notes: Income groups are static classifications recorded in the June 2026 GMD release. Breadth vulnerability is defined by \(B_{it}\geq3\). Severity and PCA classifications use their respective development-sample upper-quartile cut-offs. Income-group differences are descriptive and are not interpreted as causal effects of development status.

Figure A4.3 provides the corresponding graphical comparison for the baseline breadth measure. The gradient in breadth vulnerability partly reflects differences in the incidence of high debt, low revenue, and other component-level conditions across the three groups. Because the panel is unbalanced and the income labels are static rather than historical, these comparisons should be interpreted as descriptive sample characteristics.

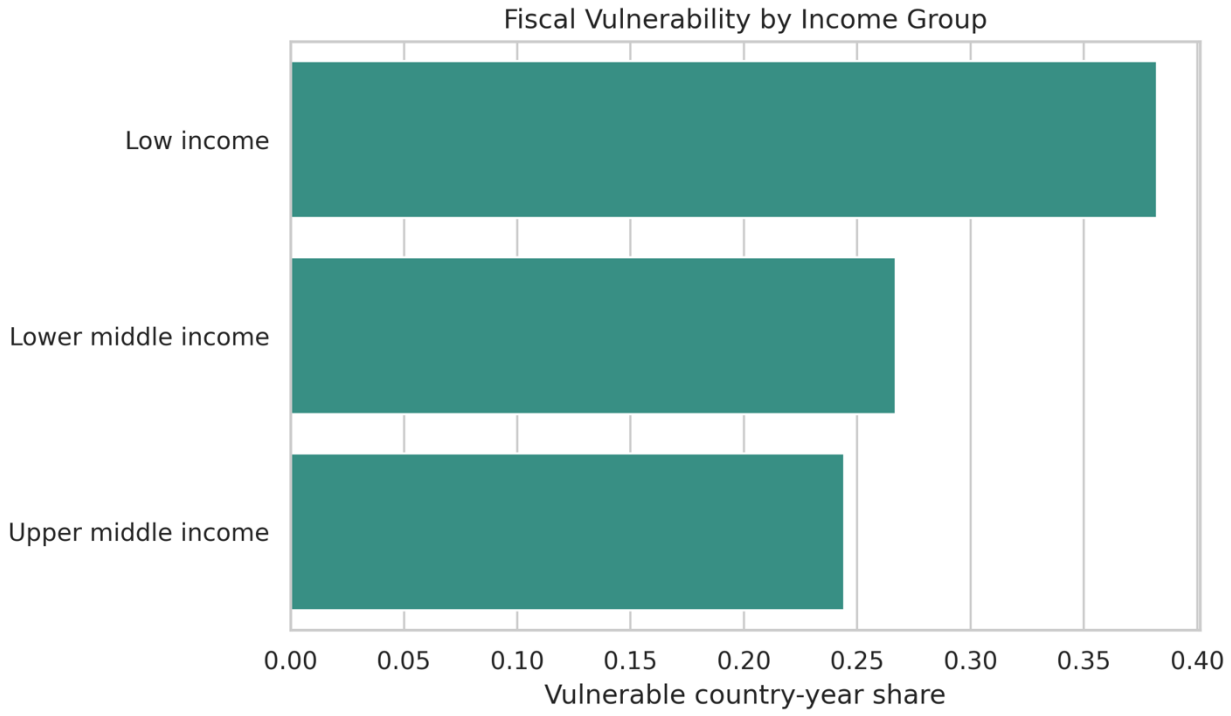


Figure A4.3. Baseline breadth vulnerability by static income group

Annual breadth prevalence also varies substantially over the sample period. Table A4.2 reports the vulnerable share, mean breadth score, and number of available country-years for each year. The breadth measure is particularly elevated in the early part of the sample and during major periods of macroeconomic disruption. Because annual country coverage changes over time, movements in prevalence should be interpreted jointly with the number of observations reported in the final column.

Table A4.2. Annual prevalence and coverage of baseline breadth vulnerability

Year	Vulnerable share (%)	Mean breadth score	Observations
1982	61.8	2.956	68
1983	52.9	2.857	70
1984	42.9	2.429	70
1985	42.9	2.457	70
1986	42.5	2.301	73
1987	47.2	2.486	72
1988	43.7	2.296	71
1989	36.6	2.183	71
1990	39.8	2.313	83
1991	39.3	2.286	84
1992	37.9	2.264	87
1993	40.2	2.184	87
1994	44.9	2.416	89
1995	38.9	2.011	95
1996	28.0	1.820	100
1997	19.2	1.635	104
1998	26.2	1.953	107
1999	28.4	1.844	109
2000	29.3	1.845	116
2001	26.7	1.741	116
2002	27.1	1.695	118
2003	24.6	1.458	118
2004	17.7	1.277	119
2005	18.0	1.254	122
2006	13.0	1.065	123
2007	12.1	1.089	124
2008	14.6	1.520	123
2009	27.6	1.902	123
2010	14.6	1.252	123
2011	18.0	1.369	122
2012	20.2	1.484	124
2013	14.5	1.331	124
2014	17.7	1.403	124

Year	Vulnerable share (%)	Mean breadth score	Observations
2015	33.1	1.976	124
2016	27.4	1.734	124
2017	19.4	1.395	124
2018	15.5	1.423	123
2019	18.0	1.500	122
2020	48.3	2.642	120
2021	25.6	1.777	121
2022	32.8	1.849	119
2023	24.6	1.702	114
2024	27.1	1.636	107

Notes: Annual shares use the baseline breadth definition \(B_{it}\geq3\). Observation counts vary because the panel is unbalanced. The 2024 observations are included for descriptive completeness but remain outside the principal external forecasting evaluation.

Figure A4.4 plots the same annual breadth series. The figure highlights the secular and episodic variation in multidimensional vulnerability and also illustrates why a fixed development-sample measurement rule does not imply a fixed event prevalence in later periods.

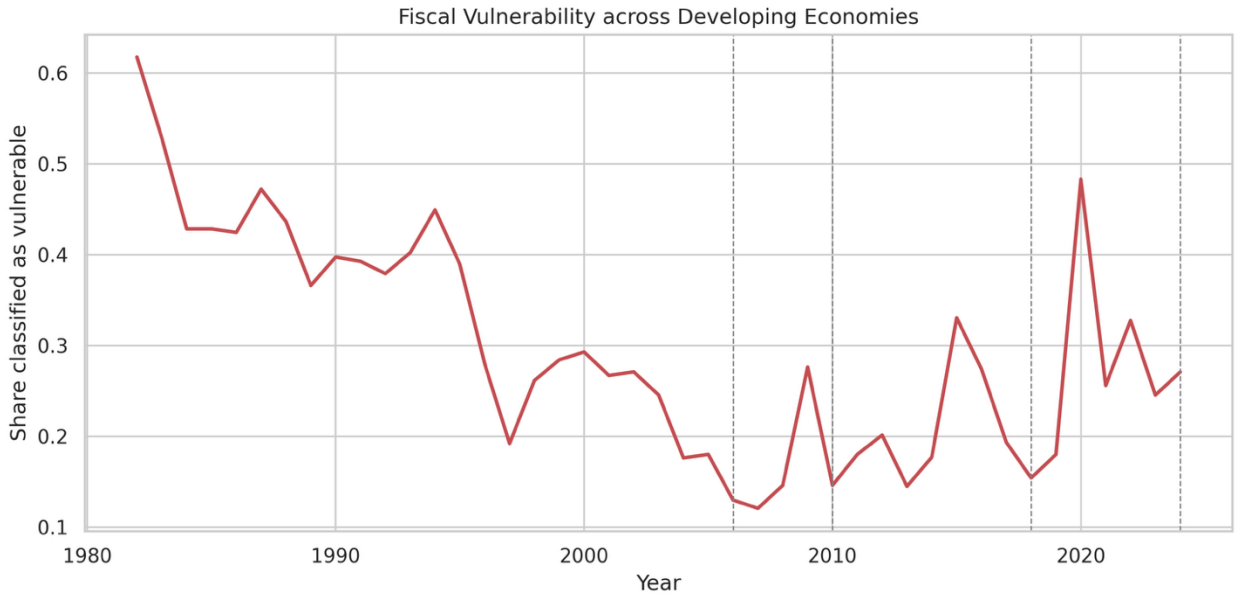


Figure A4.4. Annual prevalence of baseline breadth vulnerability

Notes: *The figure reports the annual share of complete-component country-years satisfying* $B_{it} \geq 3$. *Annual prevalence is affected by both macroeconomic conditions and changing country coverage in the unbalanced panel.*

A.4.1 Historical crisis indicators and matching procedure

The historical-consistency exercise uses the three crisis indicators distributed with the June 2026 GMD release: BankingCrisis, SovDebtCrisis, and CurrencyCrisis. These indicators are not constructed from the eight vulnerability dimensions and do not enter either outcome construction or forecasting. They are retained solely to examine whether the baseline breadth measure tends to be elevated during independently recorded episodes of banking, sovereign-debt, and currency stress.

The GMD crisis series are themselves harmonized indicators assembled from established crisis chronologies. Banking crises combine information from Baron, Verner, and Xiong (2019), Laeven and Valencia (2020), Jordà, Schularick, and Taylor (2017), and Reinhart and Rogoff (2019), with that source ordering used by the GMD when overlapping information is available. Sovereign-debt and currency-crisis indicators prioritize Laeven and Valencia (2020) and use Reinhart and Rogoff (2019) where the preferred source is unavailable. For sovereign debt, the GMD combines sovereign-debt crisis and restructuring indicators from Laeven and Valencia and domestic and external sovereign-default indicators from Reinhart and Rogoff. The GMD harmonization additionally suppresses repeated crisis starts occurring within the preceding three years so that adjacent observations do not mechanically generate multiple new crisis episodes.

Table A4.3. Historical crisis indicators used for the consistency diagnostic

Crisis type	GMD variable	Principal underlying chronologies	GMD harmonization
Banking crisis	BankingCrisis	Baron, Verner and Xiong (2019); Laeven and Valencia (2020); Jordà, Schularick and Taylor (2017); Reinhart and Rogoff (2019)	Composite indicator with source priority; repeated crisis starts within three years suppressed
Sovereign-debt crisis	SovDebtCrisis	Laeven and Valencia (2020); Reinhart and Rogoff (2019)	Laeven–Valencia prioritized; domestic/external default and restructuring information combined; repeated starts within three years suppressed
Currency crisis	CurrencyCrisis	Laeven and Valencia (2020); Reinhart and Rogoff (2019)	Laeven–Valencia prioritized; repeated crisis starts within three years suppressed

Notes: The analysis uses the crisis indicators as distributed in GMD version 2026_06 and does not modify the underlying crisis dates. The indicators are matched directly to the analytical panel by GMD ISO3 country code and calendar year. No fuzzy country matching, interpolation, or imputation of missing crisis indicators is applied.

Because the crisis variables are contained in the same GMD country-year file, matching is performed directly by ISO3 code and calendar year. For each crisis type, observations with a missing crisis indicator are excluded from that type-specific comparison. This produces different

available sample sizes across banking, sovereign-debt, and currency crises. The exercise is contemporaneous: vulnerability status in a country-year is compared with the recorded crisis indicator in the same country-year. It does not treat vulnerability as a crisis prediction and does not impose lead or lag windows around crisis dates.

Table A4.4 reports the resulting historical associations. Breadth vulnerability is observed in 56.4% of banking-crisis observations, compared with 28.2% outside banking crises; in 46.7% of sovereign-debt-crisis observations, compared with 28.4% outside such episodes; and in 78.2% of currency-crisis observations, compared with 26.5% outside currency crises. The corresponding differences are 28.2, 18.3, and 51.7 percentage points.

Table A4.4. Historical consistency of baseline breadth vulnerability with recorded crises

Crisis type	Available country-years	Crisis observations	Economies with crisis	Vulnerable during crisis (%)	Vulnerable outside crisis (%)	Difference (percentage points)
Banking crisis	3,302	78	60	56.4	28.2	28.2
Sovereign-debt crisis	3,309	75	41	46.7	28.4	18.3
Currency crisis	3,325	147	85	78.2	26.5	51.7

Notes: Available country-years are analytical observations for which the corresponding GMD crisis indicator is observed. Differences equal the vulnerable share during recorded crisis observations minus the vulnerable share outside crisis observations. Crisis indicators are excluded from vulnerability construction and forecasting and are used solely as a historical-consistency diagnostic.

The uniformly higher breadth-vulnerability incidence during recorded crisis observations supports the historical consistency of the constructed measure, particularly for currency crises. The correspondence is intentionally incomplete. Fiscal vulnerability can emerge before a recorded crisis, remain elevated after an episode, or occur without culminating in a formally recorded crisis. Conversely, a recorded crisis can arise through mechanisms not fully represented by the eight dimensions used here. These comparisons therefore do not constitute independent validation of the vulnerability measure and do not imply equivalence between constructed fiscal vulnerability and realized banking, sovereign-debt, or currency crises.