Economics Bulletin

Volume 43, Issue 2

Leading indicators of sovereign defaults in middle- and low-income countries: the role of foreign exchange reserve ratios in times of pandemic.

Sofiane El Ouardi
Université Paris-Est Créteil (UPEC)

Abstract

The context of the global pandemic has brought back to the foreground a renewed challenge of designing effective early warning systems for sovereign debt crises. This paper aims to empirically assess the predictive power of several foreign exchange reserve ratios in 66 middle- and low-income countries during 1973-2017. The main estimation results demonstrate that the reserves to total external debt and the reserves-to-GDP ratios stand out compared with the other predictors and yield good predictive power according to an array of performance criteria. The previous outcome is robust, even at more distant forecast horizons. I eventually show that the reserves to total external debt ratio also displays a fine predictive power from an out-of-sample perspective (i.e., in predicting defaults that occurred in the wake of the COVID-19 crisis). The previous outcome highlights that foreign exchange reserve buffer accumulation is an efficient macroprudential policy instrument that may enable to loosen constraints related to the Mundell trilemma, therefore preventing debt crises by reducing output volatility.

Citation: Sofiane El Ouardi, (2023) "Leading indicators of sovereign defaults in middle- and low-income countries: the role of foreign exchange reserve ratios in times of pandemic.", *Economics Bulletin*, Volume 43, Issue 2, pages 793-812
Contact: Sofiane El Ouardi - sofiane.el-ouardi@u-pec.fr.
Submitted: June 15, 2022.   **Published:** June 30, 2023.

1 Introduction

In response to the currency mismatches that led to large capital outflows in the 1990s, emerging market economies (EMEs) turned to the local currency bond market to overcome the "original sin" (Eichengreen & Hausmann (1999)). Nonetheless, this process exposed EME bond markets that relied on foreign portfolio investors who evaluate risk exposure in terms of dollars (Hong Kong Monetary Authority (2020)). Consequently, EMEs became more vulnerable to global financial shocks that accelerated capital flights during periods of financial turmoil (Hong Kong Monetary Authority (2019)). As a result of the COVID-19 pandemic, EMEs that relied on foreign investors to hold their domestic currency bonds suffered larger increases in their local currency bond spreads. According to Hofmann et al. (2021), multiple EMEs have undertaken inflation-targeting policy frameworks that employ macroprudential instruments such as foreign exchange reserve accumulation over the past decade. This strategy aims to mitigate the risks associated with large fluctuations in capital flows and exchange rate depreciation. Indeed, Rodrik & Velasco (1999) (1999) emphasized that countries that abide by the Greenspan-Guidotti (G-G) rule (i.e., holding reserves that equal at least 100% of their short-term debt) reduce the probability of experiencing capital outflows by 10 percentage points on average.

In the context of a global pandemic, Hofmann et al. (2021) highlight that sovereign spreads tend to increase as a result of domestic currency depreciation against the US dollar, as the authors state that this process has accelerated since 2013, reaching its peak in early 2020. This phenomenon is mirrored in the recent evolution of credit ratings: while 15% of the advanced economies have experienced rating downgrades since the start of 2020, emerging and developing countries recorded demotions that reached approximately 40 percent (Reinhart (2021)). In addition, many low- and middle-income countries have recorded significant capital outflows and sharply weakening currencies (in some cases, currency crashes such as in Turkey) during 2020. Thus, the specter of looming sovereign default resurfaces since a growing number of low-income countries, which are eligible for the Debt Service Suspension Initiative (DSSI), are in debt distress or at high risk (World Bank (2020)), while other emerging economies have recently restructured (Argentina, Belize, Ecuador) or remain in default (Lebanon, Suriname, Venezuela).[1]

This paper contributes to the literature on debt crises by demonstrating that consensual foreign exchange reserves metrics perform well at predicting sovereign defaults in middle- and low-income countries. To the best of my knowledge, there are no papers in the literature that implement a horse race among these reserve metrics for the forecasting of debt crises. In addition, the reserves to total external debt ratio, which produces the best overall performance among the horse races, also displays good predictive power from an out-of-sample perspective and is able to detect 4 out of the 5 current sovereign defaults that occurred in the wake of the COVID-19 pandemic. The individual assessments of foreign exchange reserves ratios demonstrate that the reserves-to-GDP ratio has a strong predictive power in the lower-middle-income group of countries, while the reserves to total external debt ratio seems more suitable for the upper-middle group. The previous outcome is robust at more distant forecast horizons. The main policy implication of those results is that foreign exchange reserves have a strong predictive power in the context of sovereign defaults in middle- and low-income countries, suggesting that reserve buffer accumulation should be employed as a

[1] Out of the 73 low-income countries in the DSSI, 34 are classified as in external debt distress or at high risk as of December 2021.

macroprudential policy instrument to prevent debt crises. The remainder of the paper is organized as follows: section 2 details the process of sample construction while describing both the model and the methodology employed for the main estimation. Section 3 displays the results of the horse race among reserve ratios and explores the individual performances of these consensual reserve metrics, focusing on their predictive power at more distant forecast horizons and in an out-of-sample context.

2 Data and methodology

Regarding sovereign defaults dating, I rely on the database proposed by Laeven & Valencia (2020), which is a compilation of information collected from various sources: Beim & Calomiris (2001), World Bank (2002), Sturzenegger & Zettelmeyer (2007) and Cruces & Trebesch (2013). The gathered data include the year of sovereign default as well as the restructuring date if the latter took place. Under this definition, the authors manage to capture 79 episodes during the 1970-2017 period.

2.1 Sample construction

My sample consists of 57 middle-income and 9 low-income countries covering the 1973-2017 period. The starting point is the Laeven & Valencia (2020) updated database that covers 112 middle- and low-income countries and provides annual data related to banking, currency and sovereign debt crises. Due to data unavailability regarding some of the main explanatory variables, 46 countries are dropped in the process of sample construction. Given that some observations are missing for a few economies at specific periods, the starting date for each country varies within the sample resulting in an unbalanced panel for a maximum of 1,889 observations for the full sample and 1,649 for middle-income countries, which corresponds to a 29-year period per country on average. The list of countries along with details on each crisis starting and ending date are displayed in Tables A4 and A5.

2.2 The role of foreign exchange reserves

Over the last 40 years, numerous emerging market economies have suffered multiple financial crises. The common feature of those events was the sudden stop in capital flows, which resulted in large and permanent output losses (Nakamura et al. (2013)). According to Arslan & Cantú (2019), EMEs addressed this issue by accumulating foreign exchange reserves as a form of self-insurance (the so-called precautionary motive). In fact, the empirical evidence since the global financial crisis and the taper tantrum episode demonstrates that reserves boost EME resilience, as countries that held more reserves suffered smaller currency depreciation compared with the others. Previous empirical findings suggest that large reserve buffer accumulation may mitigate some of the constraints implied by the Mundell trilemma.[2] Indeed, Aizenman et al. (2010) illustrate that while both Latin American and Asian EMEs liberalized their financial markets and maintained exchange rate stability since the 1990s, Asian EMEs stand out by displaying greater monetary independence. According to Aizenman et al. (2010), these two groups of economies are mostly differentiated from each other by their respective levels of international reserves holding. Therefore, EME

[2]The impossibility of simultaneously maintaining a fixed exchange rate regime, allowing free flows of capital and having autonomous monetary policy.

Table 1: Descriptive statistics

Upper-middle (1)

	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
Observations	906	906	906	906	906
Mean	18.458	107.702	70.301	5.715	3.969
Median	12.344	32.846	43.979	4.227	3.156
Standard deviation	19.695	352.065	168.111	5.162	4.165
Kurtosis	10.478	58.670	164.403	12.197	59.571
Minimum	0.196	0.389	0.000	0.073	0.191
5% percentile	2.571	8.221	1.063	1.233	0.823
95% percentile	62.339	332.973	172.166	16.586	9.422
Maximum	124.011	3840.105	2780.651	36.782	55.924
Data source	WDI	WDI	WDI	WDI	WDI

Lower-middle & low (2)

	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
Observations	983	983	983	983	983
Mean	11.113	36.299	487.097	3.881	8.401
Median	9.522	22.526	30.323	3.332	3.342
Standard deviation	8.244	42.184	2926.692	2.679	28.297
Kurtosis	4.481	13.055	31.132	6.380	139.967
Minimum	0.008	0.049	0.000	0.028	0.553
5% percentile	0.012	1.378	0.649	0.444	0.988
95% percentile	26.781	115.953	1129.028	9.185	17.676
Maximum	43.736	335.128	59755.560	19.209	505.726
Data source	WDI	WDI	WDI	WDI	WDI

Variables definitions: Res/GDP = International reserves over GDP ratio; $Res/Debt$ = International reserves to total external debt ratio; $StDebt/Res$ = Short-term debt in % of reserves; Res/Imp = International reserves in months of imports; $M2/Res$ = Broad money to reserves ratio.
Data source: World Development Indicators (WDI)

efforts to loosen the trilemma in the short run can involve an increase in international reserves holding. Jeanne & Rancière (2006) propose a cost-benefit model with the purpose of measuring the optimal level of reserves. The authors employ a sample of 34 middle-income countries spanning from 1975 to 2003 and determine that the optimal ratio is approximately 10.1% of the GDP, which is close to the empirical observations for the same time interval (9.4%) and corresponds to full coverage of the short-term debt according to the G-G rule.[3] Nevertheless, the study emphasizes that this ratio has tended to increase in recent years. The previous statement is also true in our sample, as Table 1 depicts means of 18.5% and 11.1% for the reserves-to-GDP ratio in upper-middle and lower-middle-income countries, respectively, over the 1973-2017 period. Note that all the reserve ratios are extracted from the World Development Indicators (WDI, World Bank).

2.3 The model

The construction of my EWS relies on the so-called "window crisis approach". Thus, the binary dependent variable SDC_{it} (Sovereign Debt Crisis) is set to one for the full duration of the crisis (as detailed in Table A4) and zero during tranquil periods. Nevertheless, I control in each specification for the post-crisis entry bias (Bussiere & Fratzscher 2006) by keeping

[3] Greenspan (1999) suggests that reserves should exceed official and officially-guaranteed short-term debt.

only the starting year of each crisis episode. The main idea behind this procedure is that successfully detecting crisis onsets seems more relevant from a policy-maker's perspective. The logit model employed to estimate the probability of default can be written as follows:

$$Pr(SDC_{it} = 1) = F(X_{it-1}\beta) = \frac{e^{X_{it-1}\beta}}{1 + e^{X_{it-1}\beta}} \tag{1}$$

where F is the cumulative logistic distribution, X_{it-1} is the vector of 1 period lagged independent variables, SDC_{it} designates the binary crisis variable and β is the vector of coefficients. Following Manasse et al. (2003), I allow for country-specific variances using the Huber-White robust variance estimator. Two models are considered following the World Bank's income classification: the one that encompasses upper-middle-income countries and a second one that contains lower-middle- and low-income countries.[4]

2.4 Horse-race methodology

To evaluate the forecasting performances of different reserves metrics, I rely on two approaches: the first one requires selecting the optimal cutoff probability (i.e., setting a threshold such that the model issues a signal of an imminent crisis if that threshold is exceeded), while the second one consists of employing several performance criteria.

2.4.1 Optimal cutoff probability

In this approach, assessing the predictive power of reserve ratios consists of comparing the actual dependent variable SDC_{it} to the issued signals. Thus, the following contingency matrix can be constructed:

$$\begin{array}{c} \\ \textit{Signal} \\ \textit{No Signal} \end{array} \begin{array}{c} \begin{array}{cc} \textit{Crisis} & \textit{Tranquil} \end{array} \\ \begin{pmatrix} TP & FA \\ MC & TN \end{pmatrix} \end{array}$$

TP denotes the true positives (i.e., correctly called episodes), TN the true negatives, FA refers to false alarms and MC to missed crises.

Nevertheless, selecting a cutoff probability can be challenging since setting an elevated threshold yields a higher rate of missed crises (Type I errors), while a low cutoff point triggers too many false alarms (Type II errors). Fuertes & Kalotychou (2007) state that focusing on missed crises is more relevant for policy-makers than focusing on false alarms since the cost of unforeseen sovereign defaults is substantially higher than that of undertaking precautionary measures. Nevertheless, Savona & Vezzoli (2015) pointed out that trivializing type II errors may lead to adverse effects on the international reputation, as a high rate of false alarms tends to issue a negative signal regarding domestic market stability. However, it is important to mention that a triggered alarm can occur as a result of a successful preemptive policy adopted by the authorities to avoid a crisis. Therefore, type II errors are not inevitably miscalculations but could also be the indication of an early intervention. The empirical literature on EWS acknowledges that selecting the cutoff point requires either minimizing the joint error measure (signal-to-noise ratio) or maximizing Youden's J-statistic. In their study, Savona & Vezzoli (2015) suggest that Youden's J-statistic is more suitable than the nose-to-ratio signal, as the authors state that the J-statistic is robust to extreme type I and

[4] Two additional models that encompass the full sample and middle-income countries are considered as a robustness check (Table A1).

type II errors. In contrast, minimizing the signal-to-noise ratio yields extreme thresholds in which false alarms are close to zero but the defaults are barely detected (Mulder et al. (2002)). Following Savona & Vezzoli (2015), I decide to implement Youden's J-statistic with the purpose of evaluating the performance of the reserve metrics. The latter can be written as follows:

$$J = \frac{TP}{TP + MC} + \frac{TN}{TN + FA} - 1 \tag{2}$$

where the left term on the right-hand side of the equation denotes the true positive rate (sensitivity) and the right term designates the true negative rate (specificity). The optimal cutoff probability point can be obtained by maximizing the J-statistic:

$$J = arg\ max[sensitivity + specifity - 1] \tag{3}$$

To assess the predictive power of reserve ratios, I consider three aspects: the percentage of correctly called crises and the rate of properly forwarded crisis entries while keeping attention on minimizing the rate of false alarms.

2.4.2 Performance criteria

Regarding the performance criteria, I first consider the area under the receiver operating characteristic (AUROC) curve. The receiver operating characteristic (ROC) curve plots the correctly called episodes rate against the false alarms rate at various threshold settings. High values of the AUROC curve indicate that the binary classifier performs well at predicting zeros as zeros and ones as ones. AUROC curve values range from 0.5 (no discrimination) to 1 (perfect distinction). Nevertheless, there are several criticisms of this method. One is that the AUROC curve attributes the same importance to tranquil periods compared with crisis periods. However, one might argue that tranquil periods are less relevant than crisis observations for policy-makers. Therefore, the AUROC curve might overestimate the global performance of a classifier based on good predictions of true negative outcomes (i.e., classifying zeros as zeros). This limitation might be particularly significant in the case of forecasting sovereign default episodes since positive outcomes are sparse in the dataset. Furthermore, I employ the area under precision-recall (AUPR) curve as an alternative measure to address the class imbalance issue (i.e., datasets in which the number of negatives significantly outweighs the number of positives), as suggested by Saito & Rehmsmeier (2015). The precision-recall (PR) curve plots the ratio of correctly called crises to total episodes against the true positive rate at various threshold settings. Consequently, the AUPR curve eliminates the impact of true negative outcomes in the process of assessing the performance of each classifier with imbalanced data. Nevertheless, interpreting AUPR curve values might be challenging in contrast with the AUROC curve. Thus, the primary idea behind this implementation is not interpreting each value separately but rather establishing a hierarchy among predictors by comparing multiple criteria. In addition, I use two additional criteria to assess the accuracy of the predicted probabilities: the Brier (1950) score and the Tjur (2009) R^2. The Brier (1950) score measures the mean squared error of the predictions. Accordingly, a low Brier score indicates that the binary classifier performs well. Specifically, a Brier score approaching 0 is considered the best possible value (i.e., total accuracy). The Tjur (2009) R^2 , also called Tjur's coefficient of discrimination, is defined as the difference between the mean predicted probability of both positive and negative outcomes. Consequently, a high Tjur $R2$ indicates that the binary classifier performs well. More precisely, a coefficient near

1 suggests that there is a clear separation between the predicted values for zeros and ones. Finally, the predicted probabilities are obtained from a binary logit model estimated by maximum likelihood. Therefore, I also display the maximum likelihood and pseudo-$R2$ as a complement to compare the performance of the reserve ratios.

3 Results

This section is organized as follows: I first present the results of the horse race among the consensual exchange reserve ratios that are typically employed by central banks to determine foreign exchange reserve adequacy. Afterward, I more thoroughly explore the predictive power of these reserve ratios by comparing their respective predictive power at more distant forecast horizons. Finally, an out-of-sample forecast is considered to assess the model's ability to detect recent sovereign defaults that occurred as a result of the COVID-19 pandemic.

3.1 Reserve ratios as debt crises predictors: a horse-race

Following the methodology previously detailed, I implement a horse race between the main reserve ratios employed by central banks. Given the absence of a regulatory framework with which to assess reserve requirements for precautionary motives, central banks typically follow an array of measures:

• Reserves over imports: assesses the reserves coverage in terms of monthly imports. The benchmark is usually set to 3 months sustainability.

• Short-term debt to reserves: measures the need for repayment related to a country's short-term external liabilities in foreign currency with a remaining maturity of one year or less. The G-G rule suggests that this ratio should be equal to 1 (100% cover). The model designed by Jeanne & Rancière (2006) determines that the G-G rule is also a good approximation of the optimal amount of international reserve requirements.

• Broad money to reserves: evaluates the potential impact of a loss of confidence in the domestic currency. This ratio is well-suited for countries with large banking sectors and very open capital accounts (IMF (2015)).

In addition to those indicators, I include the ratio of reserves-to-GDP as well as the reserves to total external debt. Since I expect the race to be tight among the reserve ratios, I account for potential dissimilarities in financial development between countries that might affect the ranking of the ratios. Thus, in Table 2, Model (1) designates countries that belong to the upper-middle-income group of the World Bank's classification, while Model (2) encompasses both lower-middle- and low-income economies of the same classification. The upper part of Table 2 shows that the ratio of reserves to total external debt dominates the race according to 4 out of the 6 implemented criteria with an AUROC value approaching 0.88. I reach similar conclusions when the full sample (3) and middle-income countries (4) are considered, as shown in Table A1. The previous outcome implies that the reserves to total external debt is a relevant early warning indicator of sovereign default in financially developed middle-income countries. The remaining predictors in the upper-middle group display a fairly good performance with AUROC values near 0.8. Furthermore, the short-term debt to reserves ratio exhibits a fine performance in the upper-middle-income group (according to AUPR criteria) suggesting that short-term maturity coverage, which is a key indicator in the determination of optimal reserve requirements, is also relevant for debt crisis

predictions in financially developed middle-income countries.[5] A different scheme emerges in the lower part of Table 2, as I observe a slight dominance of the reserves-to-GDP ratio according to AUROC, AUPR, Tjur's $R2$ and Brier Score. In addition, the M2 to reserves ratio appears to perform poorly compared with the 4 other ratios, which corroborates the idea of financial development features that affect the predictors' ranking. Indeed, the M2 to reserves ratio is often deemed to be a relevant indicator in countries with open capital accounts and financially developed markets (IMF (2015)).
In terms of optimal cutoff probability, the upper part of Table 2 also demonstrates that the reserves to total external debt ratio stands out relative to other ratios as this predictor is able to individually detect 21 out of the 23 crises onsets in the upper-middle group of countries while the short-term debt to reserves ratio emits the lowest false alarm rate yet is still able to correctly call 74% of entries. In contrast, the reserves-to-GDP ratio significantly outperforms other predictors in the lower-middle group, with more than 94% of entries detected, although issuing slightly more false alarms than the reserves in months of imports, which dominates for this specific criterion (only 19.3% of Type II errors).

3.2 Reserve ratios performance for different forecast horizons

In section 3.1, all the predictors are lagged by one period (i.e., 1 year prior to a crisis). An effective predictor should, however, start to issue a signal earlier than 1 year so that it provides policy-makers with some lead time to adopt preemptive policies. In addition, debt crisis predictors should provide a stable signal throughout multiple consecutive periods to reduce uncertainty regarding the risk of default. Therefore, I run a sensitivity test with a forecast horizon covering a 5-year window prior to a crisis. For each forecast horizon and debt crisis predictor, Figures 1 and 2 plot the AUROC curve to highlight the quality of the signals.[6] Figures 1 and 2 corroborate the main result obtained from section 3.1 concerning the difference between the upper-middle and the lower-middle groups. For all forecast horizons, Figure 2 highlights that the reserves to debt ratio provides a better signal to defaults in upper-middle-income countries, while the signal issued by the M2 to reserves ratio substantially drops at t-4 and t-5. In contrast, Figure 2 demonstrates that starting from t-1, the reserves-to-GDP ratio remains the best performing predictor for all forecast horizons in the lower-middle-income group of countries. With respect to the remaining ratios, the import coverage ratio yields a stable performance for all the forecast horizons, with values ranging from 0.63 in t-5 to 0.8 in t-1 for both groups. Conversely, the short-term debt over reserves ratio only produces a stable performance in the upper-middle group, while the same does not hold for the lower-middle cluster, as the issued signal dramatically drops prior to t-2 in Figure 2 (more details are displayed in Table A2).

3.3 Out-of-sample performance

Finally, I attempt to assess the ability of the ratios to detect the current sovereign defaults that occurred as a result of the COVID-19 crisis. According to Beers et al. (2021), multiple upper-middle-income countries defaulted in 2020 and remain in default in 2021 (Argentina, Belize, Ecuador, Lebanon, Suriname and Venezuela). Further, most of these observations are confirmed in the Standard and Poor's sovereign ratings.[7] Considering that all of these

[5]This pattern is confirmed by a time-varying Granger causality test displayed in Table A3.
[6]Similar conclusions are reached with the other criteria.
[7]Excluding Venezuela, for which rating is currently not available.

Table 2: Assessing the predictive power of ratios

Upper-middle (1)

Predictor :	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
β	-0.2111**	-0.1337***	0.0020**	-0.6444***	0.1367***
(σ_β)	(0.0874)	(0.0312)	(0.0010)	(0.2214)	(0.0417)
Observations	906	906	906	906	906
Num. countries	32	32	32	32	32
Crises episodes	23	23	23	23	23
Log likelihood	-91.6196	**-81.8984**	-98.8133	-92.8422	-92.9106
Pseudo$-R^2$	0.1453	**0.2360**	0.0782	0.1339	0.1333
$AUROC$ curve	0.8149	**0.8792**	0.8039	0.8023	0.7818
$AUPR$ curve	0.1672	0.1605	**0.2519**	0.1244	0.1685
Tjur R^2	0.0480	0.0860	0.0500	0.0440	**0.1070**
Brier score	0.0235	**0.0227**	0.0238	0.0237	0.0242
Optimal cutoff (%)	5	3	2	2	2
% Correctly called	74	**91.3**	74	**91.3**	87
Detected entries	17	**21**	17	**21**	20
% False alarms	17.4	25.3	**11.4**	38.7	41.3

Lower-middle & low (2)

Predictor :	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
β	-0.2548**	-0.1040***	0.0001	-0.7424***	0.0054***
(σ_β)	(0.1063)	(0.0221)	(0.0001)	(0.2064)	(0.0027)
Observations	983	983	983	983	983
Num. countries	34	34	34	34	34
Crises episodes	17	17	17	17	17
Log likelihood	-74.4187	**-73.8466**	-85.6244	-74.6260	-85.2179
Pseudo$-R^2$	0.1329	**0.1396**	0.0024	0.1305	0.0071
$AUROC$ curve	**0.8291**	0.8145	0.8082	0.8146	0.7227
$AUPR$ curve	**0.0637**	0.0501	0.0584	0.0618	0.0506
Tjur R^2	**0.0300**	0.0250	0.0000	0.0290	0.0010
Brier score	**0.0165**	0.0166	0.0169	0.0166	0.0166
Optimal cutoff (%)	2	3	2	2	2
% Correctly called	**94.1**	76.5	82.4	82.4	70.6
Detected entries	**16**	13	14	14	12
% False alarms	25	22.5	23.1	**19.3**	31

** and *** denote the 5% and 1% significance levels, respectively.
Variables definitions: Res/GDP = International reserves over GDP ratio; $Res/Debt$ = International reserves to total external debt ratio; $StDebt/Res$ = Short-term debt in % of reserves; Res/Imp = International reserves in months of imports; $M2/Res$ = Broad money to reserves ratio.

Figure 1: Reserve ratios for different forecast horizons (upper-middle group)

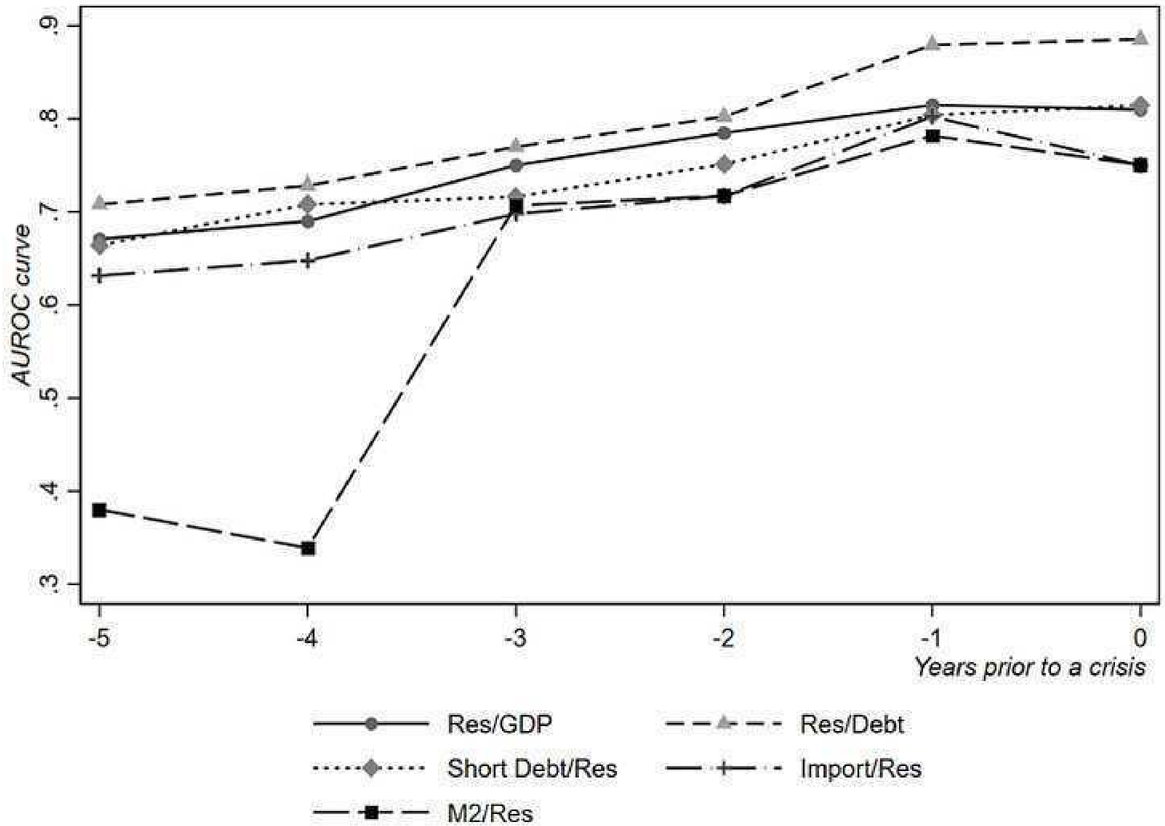
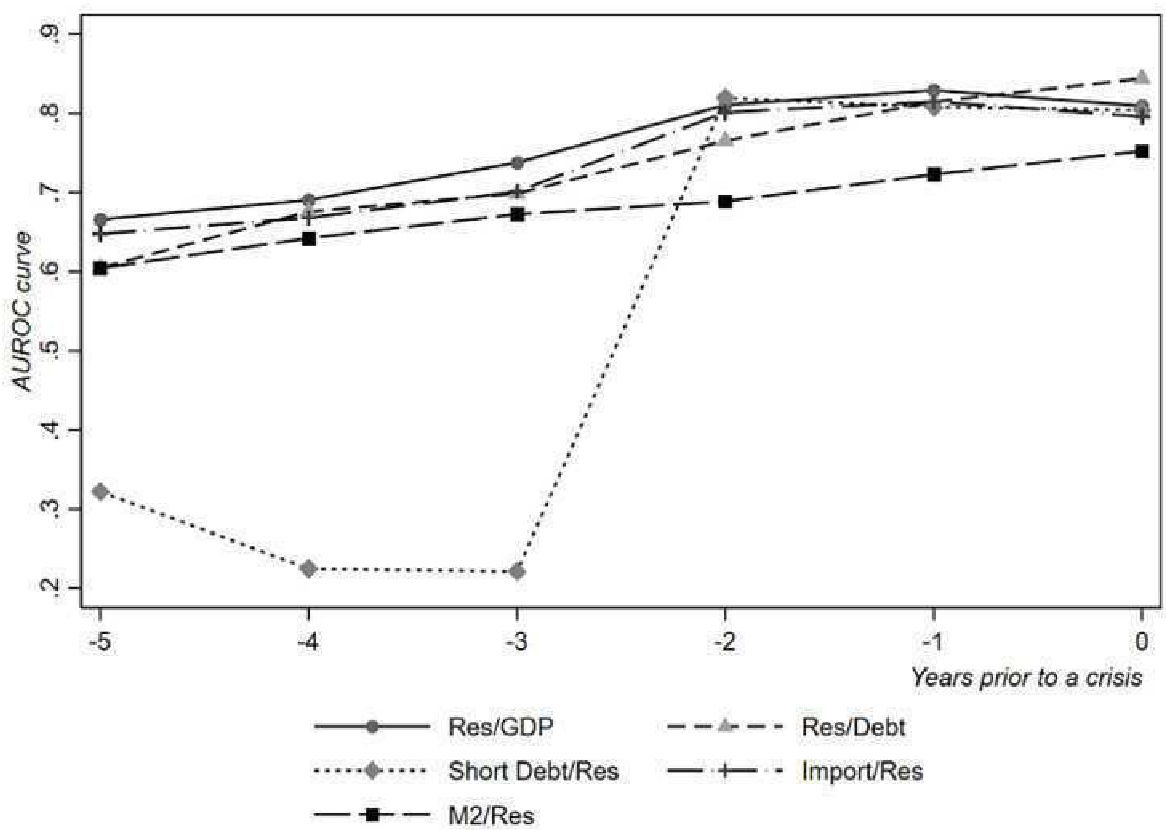


Figure 2: Reserve ratios for different forecast horizons (lower-middle & low group)



countries belong to the upper-middle-income group, I employ the reserves to debt ratio, as this indicator yielded the best overall performance. The estimation results are displayed in the upper part of Table 3 and relate to the 2018-2020 period (for the out-of-sample forecast), in which only 4 defaults are kept because of data unavailability regarding foreign exchange reserves in Suriname and Venezuela. Starting-off with the in-sample performance, the model is able to detect 21 of 23 crisis onsets although issuing 25.3% of false alarms in the process. Turning to the out-of-sample forecast performance, the model detects 75% of crisis onsets that occurred in 2020 (Argentina, Belize, Ecuador) and only misses the Lebanese default. In addition, the lower part of Table 3 depicts the out-of-sample performance using the reserves to external debt ratio in the lower-middle- and low-income group of countries in which only one episode is currently recorded for Zambia in 2020 (according to Beers et al. (2021) and Standard & Poor's). Thus, the model correctly predicts the Zambian default on the top of emitting a lower rate of false alarms (9.2%) compared with the upper-middle-income group (13.3%). Nevertheless, false alarms are not necessarily miscalculations. Indeed, as mentioned in section 2.4, false alarms could also be the sign of an early intervention or the signal of an important financial distress that does not inevitably morph into a sovereign default. From that perspective, I investigate the reserves to external debt ratio observations that exceed the optimal threshold during the tranquil period for the 2018-2020 period. Figure 4 shows that a significant proportion of false alarms come from countries such as Chad, Congo, Laos, Papua New Guinea and Sudan. Interestingly, the DSSI database from the World Bank (2020) classifies all of these countries as being at high risk of overall debt distress (as of December 2021).[8] Finally, Figure 3 reveals a similar feature regarding the upper-middle-income group since countries that are classified as "CCC" by the Standard & Poor's ratings (namely, Sri Lanka) also appear to be increasing the rate of false alarms.

4 Conclusion and policy recommendations

This paper aims to develop efficient tools in the process of calibrating an EWS for sovereign debt crises in middle- and low-income countries. The horse-race implementation demonstrates that each of 5 predictors yields a fair performance according to an array of 6 criteria from a policy-maker's perspective while displaying a few differences depending on income groups. The previous outcome is robust, even at more distant forecast horizons. I finally illustrate that the reserves to external debt ratio performs well at predicting the current defaults that occurred in the wake of the COVID-19 pandemic. The main policy implication of those results is that debt crisis episodes tend to occur when foreign exchange reserves are lower than the long-term benchmark, which corresponds to 100% short-term debt coverage in accordance with the G-G rule. Therefore, reserve buffer accumulation should be a strong macroprudential policy instrument for central banks, as this process enables economies to mitigate the harmful effect of debt crises by preventing large capital outflows and exchange rate depreciation. From an operational perspective, my findings suggest that the G-G rule should be employed in financially developed middle-income countries to determine reserve adequacy for debt crises prevention. Forthcoming research on sovereign defaults in middle- and low-income countries should focus on optimal reserve requirements that may enable to loosen constraints related to the trilemma, therefore preventing debt crises by reducing output volatility.

[8] Zambia is classified as being in overall risk of debt distress.

Table 3: Out-of-sample forecast (optimal cutoff)

Upper-middle

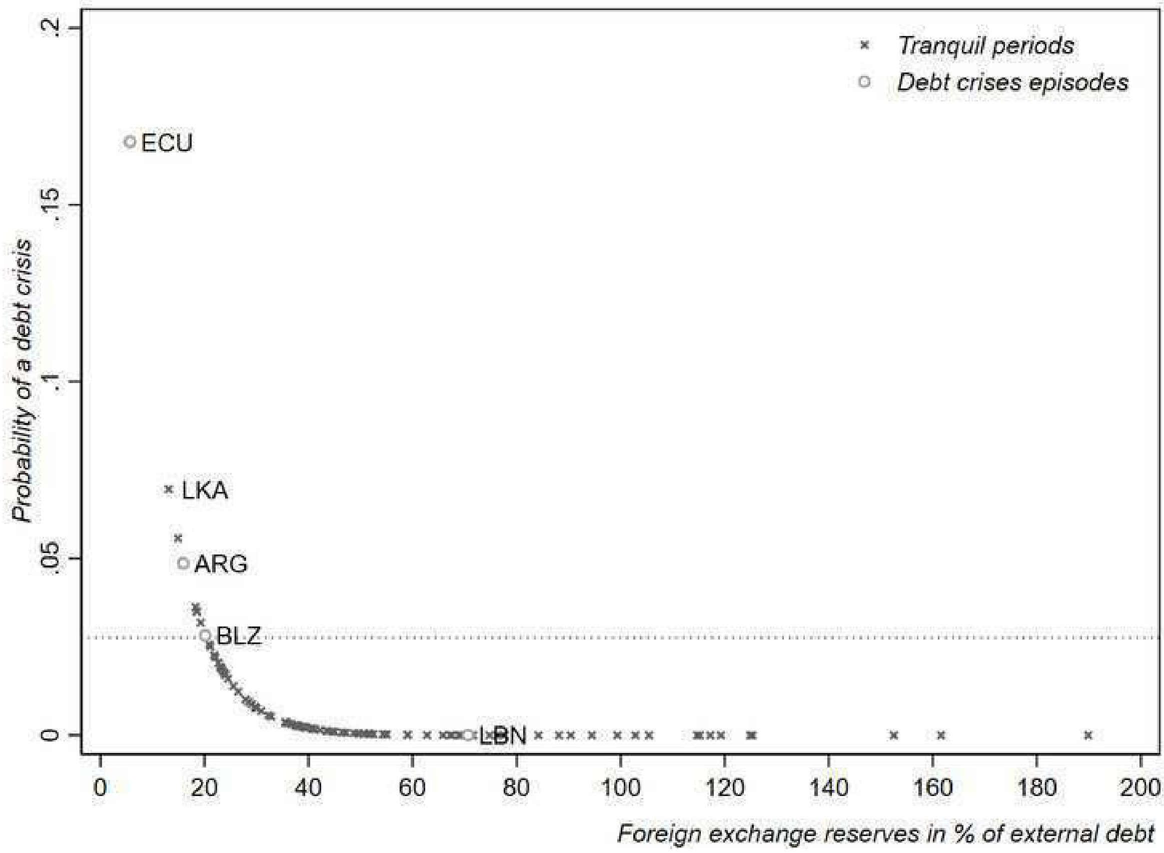
Predictor : $Res/Debt_{t-1}$	***In-sample***	***Out-of-sample***
Observations	906	124
Num. countries	32	44
Crises episodes	23	4
Optimal cutoff (%)	2.76	2.76
% Correctly called	91.3	75
Detected entries	21	3
% False alarms	25.3	13.3

Lower-middle & low

Predictor : $Res/Debt_{t-1}$	***In-sample***	***Out-of-sample***
Observations	983	165
Num. countries	34	57
Crises episodes	17	1
Optimal cutoff (%)	2.73	2.73
% Correctly called	76.5	100
Detected entries	13	1
% False alarms	22.5	9.2

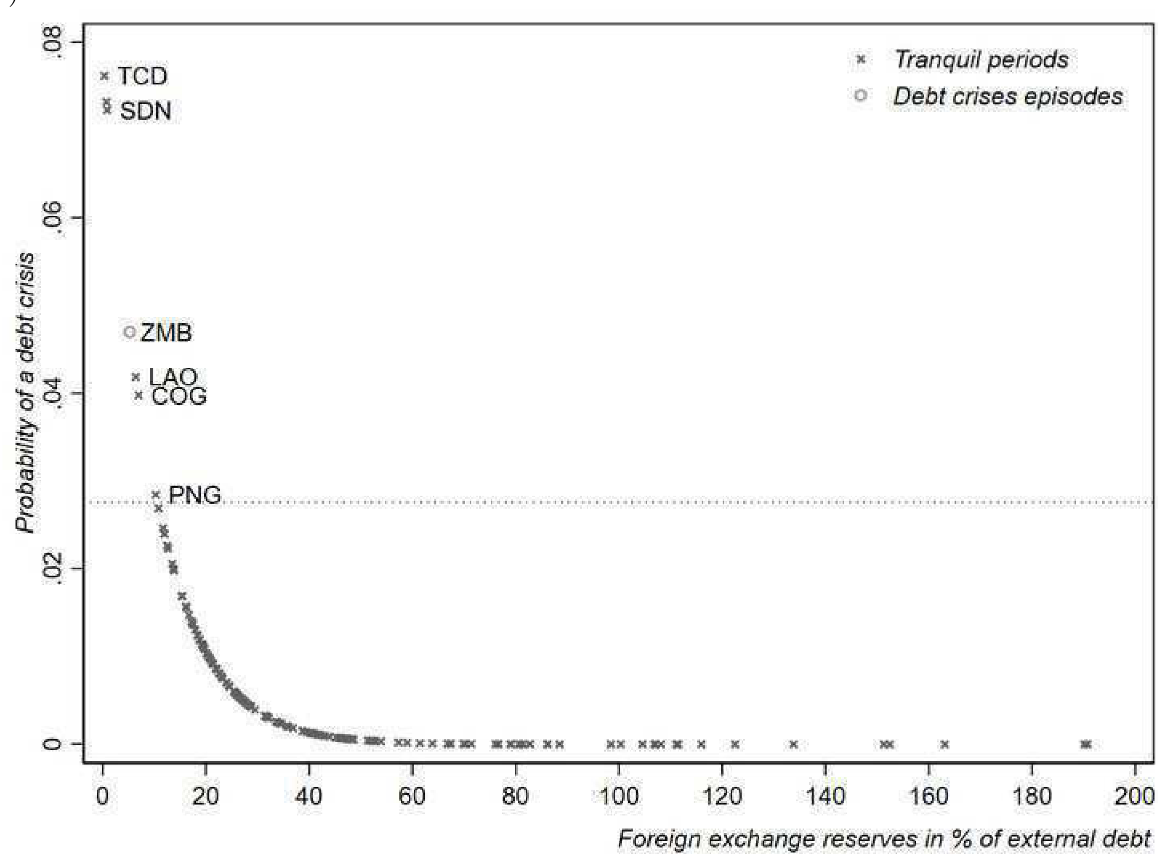
In-sample covers the 1973-2017 span while *Out-of-sample* refers to the 2018-2020 period.

Figure 3: Reserves to external debt ratio and sovereign default probability (upper-middle)



.... Optimal cutoff.

Figure 4: Reserves to external debt ratio and sovereign default probability (lower-middle & low)



References

Aizenman, J., Chinn, M. D., & Ito, H. 2010. The emerging global financial architecture: Tracing and evaluating new patterns of the trilemma configuration. *Journal of international Money and Finance*, *29*(4), 615–641.

Arslan, Y., & Cantú, C. 2019. The size of foreign exchange reserves. *BIS Paper*(104a).

Beers, D., Jones, E., Quiviger, Z., & Walsh, J. 2021. *BoC–BoE Sovereign Default Database: What's new in 2021?* (Staff Analytical Notes No. 2021-15). Bank of Canada.

Beim, D., & Calomiris, C. 2001. *Emerging Financial Markets. Appendix to Chapter 1. New York: McGraw-Hill*.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly weather review*, *78*(1), 1–3.

Bussiere, M., & Fratzscher, M. 2006. Towards a new early warning system of financial crises. *Journal of International Money and Finance*, *25*(6), 953–973.

Cruces, J. J., & Trebesch, C. 2013. Sovereign defaults: The price of haircuts. *American Economic Journal: Macroeconomics*, *5*(3), 85–117.

Eichengreen, B., & Hausmann, R. 1999. *Exchange rates and financial fragility.* National bureau of economic research Cambridge, Mass., USA.

Fuertes, A.-M., & Kalotychou, E. 2007. Optimal design of early warning systems for sovereign debt crises. *International Journal of Forecasting*, *23*(1), 85–100.

Greenspan, A. 1999. Currency reserves and debt. *Remarks Before the World Bank Conference on Recent Trends in Reserves Management, Washington, D.C., April 29, 1999.*, *20*.

Hofmann, B., Shim, I., & Shin, H. S. 2021. Emerging market economy exchange rates and local currency bond markets amid the covid-19 pandemic. *Available at SSRN 3761875*.

Hong Kong Monetary Authority. 2019. The double-edged sword of foreign participation in local currency government bond markets. *Research Memorandum*, 14/2019.

Hong Kong Monetary Authority. 2020. Foreign participation in the local currency bond markets of emerging market economies: good or bad for market resilience and financial stability? In (Vol. 113, p. 133-140). *Bank for International Settlements*.

IMF. 2015. Assessing reserve adequacy—specific proposals. *IMF policy paper*.

Jeanne, O. D., & Rancière, R. 2006. The optimal level of international reserves for emerging market countries: Formulas and applications. *IMF Working Papers*, *2006*(229).

Laeven, L., & Valencia, F. 2020. Systemic banking crises database ii. *IMF Economic Review*, 1–55.

Manasse, P., Roubini, N., & Schimmelpfennig, A. 2003. Predicting sovereign debt crises. *IMF Working Papers*, *2003*(221).

Mulder, C., Perrelli, R., & Rocha, M. 2002. *The role of corporate, legal and macroeconomic balance sheet indicators in crisis detection and prevention* (Vol. 2). International Monetary Fund.

Nakamura, E., Steinsson, J., Barro, R., & Ursúa, J. 2013. Crises and recoveries in an empirical model of consumption disasters. *American Economic Journal: Macroeconomics*, *5*(3), 35–74.

Reinhart, C. M. 2021. From health crisis to financial distress. *World Bank, Washington, DC*.

Rodrik, D., & Velasco, A. 1999. *Short-term capital flows* (Tech. Rep.). National Bureau of Economic Research, Inc.

Saito, T., & Rehmsmeier, M. 2015. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3), 1-21.

Savona, R., & Vezzoli, M. 2015. Fitting and forecasting sovereign defaults using multiple risk signals. *Oxford Bulletin of Economics and Statistics*, *77*(1), 66–92.

Sturzenegger, F., & Zettelmeyer, J. 2007. Debt defaults and lessons from a decade of crises. *MIT Press Books*, *1*.

Tjur, T. 2009. Coefficients of determination in logistic regression models—a new proposal: The coefficient of discrimination. *The American Statistician*, *63*(4), 366–372.

Toda, H. Y., & Yamamoto, T. 1995. Statistical inference in vector autoregressions with possibly integrated processes. *Journal of econometrics*, *66*(1-2), 225–250.

World Bank. 2002. *Global development finance. appendix on commercial debt restructuring. Washington, DC.* World Bank.

World Bank. 2020. *Global covid-19 debt service suspension initiative.* World Bank.

Table A1: Assessing the predictive power of ratios (performance criteria)

Full sample (3)

Predictor :	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
β	-0.2074***	-0.0971***	0.0001	-0.6118***	0.0068**
(σ_β)	(0.0644)	(0.0173)	(0.0001)	(0.1507)	(0.0032)
Observations	1889	1889	1889	1889	1889
Num. countries	66	66	66	66	66
Crises episodes	40	40	40	40	40
Log likelihood	-169.7063	**-163.0820**	-193.4248	-171.2995	-192.2605
Pseudo$-R^2$	0.1242	**0.1584**	0.0018	0.1160	0.0078
$AUROC$ curve	0.8085	**0.8314**	0.8003	0.7927	0.7485
$AUPR$ curve	0.0845	0.0708	**0.0902**	0.0765	0.0724
Tjur R^2	0.0330	**0.0360**	0.0000	0.0300	0.0010
Brier score	0.0200	0.0200	0.0207	0.0200	0.0206

Middle-income countries (4)

Predictor :	Res/GDP_{t-1}	$Res/Debt_{t-1}$	$StDebt/Res_{t-1}$	Res/Imp_{t-1}	$M2/Res_{t-1}$
β	-0.1846***	-0.0940***	0.0001	-0.5523***	0.0056**
(σ_β)	(0.0607)	(0.0182)	(0.0001)	(0.1498)	(0.0027)
Observations	1649	1649	1649	1649	1649
Num. countries	57	57	57	57	57
Crises episodes	35	35	35	35	35
Log likelihood	-149.9809	**-142.4468**	-169.2728	-151.8690	-168.6604
Pseudo$-R^2$	0.1150	**0.1594**	0.0011	0.1038	0.0048
$AUROC$ curve	0.7980	**0.8335**	0.8161	0.7796	0.7376
$AUPR$ curve	0.0772	0.0700	**0.0904**	0.0691	0.0648
Tjur R^2	0.0290	**0.0360**	0.0000	0.0260	0.0010
Brier score	0.0201	**0.0200**	0.0207	0.0201	0.0207

** and *** denote the 5% and 1% significance levels, respectively.
Variables definitions: Res/GDP = International reserves over GDP ratio; $Res/Debt$ = International reserves to total external debt ratio; $StDebt/Res$ = Short-term debt in % of reserves; Res/Imp = International reserves in months of imports; $M2/Res$ = Broad money to reserves ratio.

Table A2: Reserve ratios for different forecast horizons (AUROC values)

Upper-middle

Predictor :	Res/GDP_{t-1}	Res/GDP_{t-2}	Res/GDP_{t-3}	Res/GDP_{t-4}	Res/GDP_{t-5}
$AUROC$ curve	**0.8149**	0.7849	0.7501	0.6899	0.6707
Predictor :	$Res/Debt_{t-1}$	$Res/Debt_{t-2}$	$Res/Debt_{t-3}$	$Res/Debt_{t-4}$	$Res/Debt_{t-5}$
$AUROC$ curve	**0.8792**	0.8024	0.7698	0.7280	0.7080
Predictor :	$StDebt/Res_{t-1}$	$StDebt/Res_{t-2}$	$StDebt/Res_{t-3}$	$StDebt/Res_{t-4}$	$StDebt/Res_{t-5}$
$AUROC$ curve	**0.8039**	0.7512	0.7162	0.7080	0.6641
Predictor :	Res/Imp_{t-1}	Res/Imp_{t-2}	Res/Imp_{t-3}	Res/Imp_{t-4}	Res/Imp_{t-5}
$AUROC$ curve	**0.8023**	0.7170	0.6978	0.6478	0.6315
Predictor :	$M2/Res_{t-1}$	$M2/Res_{t-2}$	$M2/Res_{t-3}$	$M2/Res_{t-4}$	$M2/Res_{t-5}$
$AUROC$ curve	**0.7818**	0.7170	0.7069	0.3389	0.3801

Lower-middle & low

Predictor :	Res/GDP_{t-1}	Res/GDP_{t-2}	Res/GDP_{t-3}	Res/GDP_{t-4}	Res/GDP_{t-5}
$AUROC$ curve	**0.8291**	0.8106	0.7377	0.6904	0.6658
Predictor :	$Res/Debt_{t-1}$	$Res/Debt_{t-2}$	$Res/Debt_{t-3}$	$Res/Debt_{t-4}$	$Res/Debt_{t-5}$
$AUROC$ curve	**0.8145**	0.7653	0.6986	0.6758	0.6047
Predictor :	$StDebt/Res_{t-1}$	$StDebt/Res_{t-2}$	$StDebt/Res_{t-3}$	$StDebt/Res_{t-4}$	$StDebt/Res_{t-5}$
$AUROC$ curve	0.8082	**0.8193**	0.2211	0.2244	0.3223
Predictor :	Res/Imp_{t-1}	Res/Imp_{t-2}	Res/Imp_{t-3}	Res/Imp_{t-4}	Res/Imp_{t-5}
$AUROC$ curve	**0.8146**	0.8013	0.7010	0.6680	0.6477
Predictor :	$M2/Res_{t-1}$	$M2/Res_{t-2}$	$M2/Res_{t-3}$	$M2/Res_{t-4}$	$M2/Res_{t-5}$
$AUROC$ curve	**0.7227**	0.6888	0.6725	0.6422	0.6044

Variables definitions: Res/GDP = International reserves over GDP ratio; $Res/Debt$ = International reserves to total external debt ratio; $StDebt/Res$ = Short-term debt in % of reserves; Res/Imp = International reserves in months of imports; $M2/Res$ = Broad money to reserves ratio.

Table A3: Granger causality from ratios to SDC_{it}

Upper-middle

Predictor :	Observations	F-Statistic	$p-$value
Res/GDP_{t-1}	874	7.57320	0.0060
$Res/Debt_{t-1}$	874	18.1392	0.0000
$StDebt/Res_{t-1}$	874	7.62984	0.0059
Res/Imp_{t-1}	874	6.01928	0.0143
$M2/Res_{t-1}$	874	10.2099	0.0014

Lower-middle & low

Predictor :	Observations	F-Statistic	$p-$value
Res/GDP_{t-1}	949	11.6548	0.0007
$Res/Debt_{t-1}$	949	11.2669	0.0008
$StDebt/Res_{t-1}$	949	0.10605	0.7448
Res/Imp_{t-1}	949	10.5617	0.0012
$M2/Res_{t-1}$	949	2.20339	0.1380

Note: H_0: The ratio does not Granger-cause SDC_{it}.

Note: number of selected lags for Granger causality test = 1.

Note: the test is performed following the Toda & Yamamoto (1995) approach.

Variables definitions: Res/GDP = International reserves over GDP ratio; $Res/Debt$ = International reserves to total external debt ratio; $StDebt/Res$ = Short-term debt in % of reserves; Res/Imp = International reserves in months of imports; $M2/Res$ = Broad money to reserves ratio.

Table A4: Sovereign defaults episodes by country (full sample)

Country	Crises episodes	Country	Crises episodes
Argentina	1982-1993	Mexico	1982-1990
	2001-2005		
	2014-2016		
Belize	2007-2007	Morocco	1983-1986
	2012-2013		
	2017-2017		
Bolivia	1980-1992	Nigeria	1983-1992
Brazil	1983-1994	Peru	1978-1996
Cameroon	1989-1992	Philippines	1983-1992
Congo Rep.	1986-1992	Sierra Leone	1977-1995
Costa Rica	1981-1990	Sudan	1979-1985
Dominican Rep.	1982-1994	Turkey	1978-1982
	2003-2005		
Ecuador	1982-1995	Uganda	1981-1993
	1999-2000		
	2008-2009		
Egypt	1984-1992	Ukraine	1998-1999
			2015-2015
Gabon	1986-1994	Venezuela	1982-1990
	2002-2002		
Gambia	1986-1988		
Guyana	1982-1986		
Honduras	1981-1992		
Indonesia	1999-2002		
Jamaica	1978-1990		
	2010-2013		
Jordan	1989-1993		
Madagascar	1981-1992		
Malawi	1982-1988		

Table A5: Middle- and Low-income countries in the sample (full sample)

Middle				Low	
Albania	Algeria	Indonesia	Jamaica	Burundi	Central African Rep.
Angola	Argentina	Jordan	Kazakhstan	Gambia	Haiti
Azerbaijan	Armenia	Kenya	Kyrgyzstan	Madagascar	Malawi
Bangladesh	Belize	Lebanon	Macedonia	Nepal	Sierra Leone
Bolivia	Botswana	Mexico	Mongolia	Uganda	
Brazil	Bulgaria	Morocco	Nigeria		
Cambodia	Cameroon	Pakistan	Papua Guinea		
China	Colombia	Paraguay	Peru		
Comoros	Congo Rep.	Philippines	Russia		
Costa Rica	Dominican Rep.	Sri Lanka	Sudan		
Ecuador	Egypt	Swaziland	Thailand		
El Salvador	Fiji	Tunisia	Turkey		
Gabon	Georgia	Ukraine	Venezuela		
Guatemala	Guyana	Vietnam			
Honduras	India				