Bank of Finland

BoF Economics Review

3 • 2024

Constructing a composite indicator to assess cyclical systemic risks: An early warning approach

Heidi Koponen, Economist, Financial Stability and Statistics Department

Abstract

The main contribution of this paper is the construction of a cyclical systemic risk indicator from early warning indicators of banking crises (EWIs) used in Finland. Previous research has shown that combining EWIs can enhance their early warning properties. This study evaluates the indicator's performance through AUROC and noise-to-signal ratios and finds that the early warning performance of the composite indicator is good (AUROC of 0.76), with a low noise-to-signal ratio (0.2). The indicator warns of an approaching crisis well beforehand and the pre-crisis level of the indicator seems to correlate with the severity of the crisis. The study also examines the impact and relative importance of individual EWIs within the composite indicator by analysing the performance of the composite indicator when individual EWIs are excluded. Results suggest that including an external balance indicator is crucial, while excluding the credit-to-GDP gap (also called Basel gap) has minimal effect on the indicator's performance. The limited usefulness of the Basel gap can be attributed to its redundancy, as it shares substantial similarities with other indicators, resulting in minimal influence on the composite indicator's performance.

Keywords: financial cycles, systemic risk, banking crises, early warning systems

JEL codes: G01, G17, E44, E47

1 Introduction

The prediction of banking crises relies on historical information and different kinds of empirical and mathematical models. While estimations can offer valuable information, accurately forecasting future developments is challenging due to the unpredictable nature of crises. Nevertheless, these estimations can still be helpful in preparing for and mitigating the effects of shocks and crises.

Central banks have developed early warning systems (EWS) to estimate the accumulation of financial imbalances and the probability of a banking crisis. Such systems consist of selected early warning indicators and a model used to analyze their development. These systems offer guidance for regulators in adjusting macroprudential policy, which is why identifying useful early warning indicators is important. In this study, I construct a cyclical systemic risk indicator for Finland, building upon the work of Lang et al. (2019), but tailoring the indicator to reflect the conditions present in Finland.

1.1 Early warning indicators

Early warning indicators are used to assess the state of the financial cycle and to aid in potential adjustments in the level of the countercyclical capital buffer (CCyB). The European Systemic Risk Board (ESRB) advises using the credit-to-GDP gap (also known as the Basel gap) as the primary risk indicator, along with other supplementary indicators (European Systemic Risk Board (2014b)). According to the recommendation, 1-2 indicators should be included in each of the following risk categories: credit developments, potential overvaluation of property prices, private-sector debt burden, external imbalances, potential mispricing of risk, and strength of bank balance sheets. In Finland, the Financial Supervisory Authority determines appropriate indicators for each category. There are currently 13 indicators in use. The indicators are monitored at a quarterly level so that changes in the risk level are noticed early enough and there is time to react to them. Indicators are chosen based on existing research (e.g. Tölö et al. (2018) and Detken et al. (2014)), and the regulation of the Ministry of Finance (2014) (1029/2014) and aim at monitoring financial market developments as comprehensively as possible.. In addition, technical considerations, such as data availability, update cycle of data, and the number of post-release revisions of data have to be taken into account in the choosing of risk indicators.

Indicators are chosen based on their ability to predict crises at a sufficiently early stage with sufficient reliability. A threshold commonly used in literature is the ability to predict

2/3 of future crises (Borio and Drehmann (2009)). The forecast horizon is approximately 1-5 years ahead to provide policymakers enough time to take necessary actions in case the probability of a future crisis increases. Drehmann and Juselius note that in addition to good predictive qualities, EWIs should also fulfill a larger set of requirements. They need to alarm policymakers early enough, provide stable information, and be easy to interpret. Making policy decisions takes time, and there is a lag related to implementing policy tools.

The predictive power can be improved by combining EWIs (ie. Tölö et al. (2018) and Aldasoro et al. (2018)). Combination indicators combine the properties of individual indicators and allow the monitoring of several trends with one indicator. Borio and Lowe (2002) claim that it is not the single symptoms that cause a problem, but the combination of them. When, for example, excessive credit growth is combined with asset price growth and abnormal capital accumulation, systemic risks increase. Therefore it makes sense to combine EWIs into combination indicators. Aldasoro et al. (2018) study combination indicators of property prices and credit growth and find that them to significantly improve the early warning properties compared to single indicators. This is intuitive since the simultaneous excessive growth of credit and property prices is often associated with the overheating of the financial cycle.

The key question related to the use of EWIs is how to predict future developments of the financial cycle from current data. The starting position is therefore forward-looking. However, indicators are chosen based on historical data and performance. Aldasoro et al. (2018) point out that for example, structural changes in macro-prudential policy and policy actions could reduce the predictive power of EWIs. Borio and Drehmann (2009) for example suggest that property prices peaked later than expected (6 years after equity price peak instead of 2) because of differences in monetary policy actions. In the 1980s crisis, monetary policy was used to fight inflation, causing property prices to decrease, whereas, during the financial crisis of 2007-2008, inflation remained relatively low. Also, due to the infrequency of financial crises and the limited instances within individual countries, identifying the most effective EWIs for a specific country presents a significant challenge. Analysis is therefore usually done with cross-country data that includes several crises. However, it should be considered that countries differ in terms of policy-making and overall fundamentals, and one indicator does not necessarily work well for all countries. The fundamental differences between developing and developed countries, for example, affect the performance of EWIs which is why it should be carefully considered, whether grouping these countries in one study is desirable.

In addition to including cross-country data of similar countries, considering a long enough period can help mitigate errors and enable the detection of consistent findings. Aldasoro et al. (2018) find that similar variables have been shown to have consistent predictive power, with data dating back to at least the 1870s. However, it may still be difficult to estimate beforehand, what is sustainable growth (for example in house prices or credit) and what is a sign of the market overheating. In many cases, the trend can be seen afterward, but during the upturn, the trend might not raise any worries (eg. the financial crisis).

1.2 Evaluating the predictive power of EWIs

Risk indicators are chosen based on their ability to issue correct signals and avoid issuing erroneous signals. The indicator can either issue a signal or not, and the signal (or no signal) can be either true or not. A false positive signal is a type I error and a false negative is a type II error. When an indicator is chosen, policymakers have to choose between the rate of type I and type II errors. A sensitive indicator that issues an alarm often would give several false alarms, whereas an unresponsive indicator would miss several crises. If a policy-maker is more concerned about missing crises, they would choose lower signaling thresholds, thus decreasing the number of missed crisis but increasing the noise-to-signal ratio. The noise-to-signal ratio depicts how many signals are correct from all signals given. A higher noise-to-signal ratio means that a higher amount of all signals are false, i.e. there is more noise. A perfect indicator predicts all crises without any errors, whereas the information value of a poor indicator resembles a coin toss. The usefulness of an indicator is commonly assessed with a receiving operating characteristic (ROC) curve. The ROC curve shows the tradeoff between true positives and false positives. The area under the ROC curve (AUROC) summarizes the signaling quality of a binary indicator. An AUROC of 0.5 indicates complete randomness in predictive power and a 1 is a perfect fit. The AUROCs of chosen indicators range typically from 0.6 to 0.85. (Aldasoro et al. (2018))

The AUROC is a useful tool for measuring the predictive ability of an indicator but does not provide any information on critical thresholds that would indicate a rise in the probability of a crisis (Aldasoro et al. (2018)). Instead, policymakers try to minimize a loss function related to false alarms and missed crises to determine appropriate threshold values. Borio and Drehmann (2009) name three alternative ways to approach the minimization problem:

1. $min_{\theta}[L_1] = min_{\theta}[\alpha T_2 + (1-\alpha)T_1]$

2. $min_\theta[L_2] = min_\theta[noise - to - signal - ratio] = min_\theta[\frac{T_1}{1-T_2}]$

3. $min_\theta[L_3] = min_\theta[\frac{T_1}{1-T_2}|(1 - T_2) \geq X]$

where θ is the threshold and X is a minimum threshold for predicted crises. The second approach i.e. minimizing the noise-to-signal ratio is popular in research. However, simply minimizing the loss function does not necessarily provide eligible results due to a higher number of missed crises. The third approach, i.e. minimizing the noise-to-signal ratio with a minimum threshold for predicted crises helps balance the requirements between the prediction rate and noise-to-signal ratio. Borio and Drehmann (2009) test the indicators with a minimum rate of 66% and 75% of predicted crises. Their preferred choice for an indicator minimizes the noise-to-signal ratio while predicting at least 2/3 of all crises. Policymakers can adjust the thresholds to their preferences, depending for example on how concerned they are about missing a crisis. The cost of missing a crisis can be costlier than a false signal, which is why policymakers may be incentivized to pay more attention to type II errors (false negatives).

2 Forming a composite indicator

2.1 Data

I use a panel dataset by Koponen (2023) to perform the empirical part of this study. To analyze the predictive power of early warning indicators, two types of data are needed, time series data on the indicators and a crisis dataset that indicates the starting point and the duration of the crisis.

2.1.1 Early warning indicators

The indicator set includes those indicators that have been selected to the Finnish early warning system. The indicators are defined by the FIN-FSA and are based on the Ministry of Finance's regulation on the countercyclical capital buffer of credit institutions and investment companies (1029/2014, Ministry of Finance (2014)). The regulation is set following EU regulations and ESRB recommendations (European Systemic Risk Board (2014b)). The recommendation includes six categories that can be used in addition to the primary risk indicator (the Basel gap). Due to data availability reasons, the Finnish framework includes an alternative primary risk indicator with a narrow definition of credit in addition to the primary risk indicator, which is calculated with a wider definition of private sector credit. Table 1 shows the indicators that are used in Finland for evaluating systemic risk in addition to the Basel gap (Bank of Finland (2022)).

Category	Indicators
Credit developments	Annual change in private sector credit relative to 3-year moving average of GDP Annual change in household credit relative to 3-year moving average of GDP
Private-sector debt burden	Estimated private sector debt service-to-income ratio Estimated household debt service-to-income ratio
Potential overvaluation of property prices	House prices relative to consumer prices, whole country House prices relative to consumer prices, Greater Helsinki area
External imbalances	Current account-to-GDP ratio Composite indicator: Current account-to-GDP ratio + primary risk indicator
Potential mispricing of risk	Average margin on new business loans to private sector Average margin on new housing loans
Strength of bank balance sheets	Domestic MFI's total assets-to-GDP ratio, quarter-on-quarter change

Table 1: EWIs in Finland

Data on EWIs was collected for the Bank of Finland project that updated EWIs used in the CCyB decision making. A new set of core EWIs were selected and taken into policy use in September 2022 (Bank of Finland (2022)). The new composite indicator of these EWIs is constructed to help the use and interpretation of the individual core indicators. The indicators are listed in table 1 by risk category. The data reaches from 1970 to 2017 and includes the variables Basel gap, Basel gap (narrow definition of credit), annual change in private sector credit relative to 3-year moving average of GDP, estimated private sector debt service-to-income ratio, house prices relative to consumer prices (whole country), current account-to-GDP ratio, the composite indicator combining the current account-to-GDP ratio and the primary risk indicator, interest rates of new business loans and interest rates of new housing loans.

From now on, I will call the set of EWIs included in the Finnish early warning framework as sub-indicators, for they are used as the base for calculating the composite indicator. Table 2 summarizes the sub-indicators that are used to perform this analysis.

The indicator dataset by Koponen (2023) includes a selection of EU countries and Great Britain[1] that contain a sufficient amount of observations for the analysis to be performed. Using data from a single country would not be desirable, since crises are rare events and there have not been enough banking crises in any singular EU country in recent history. In terms of country-specific fundamentals, EU countries are a suitable selection for this study, as they they share many fundamental similarities with Finland. The countries belong to the common economic area and most of them are euro countries.

Each sub-indicator must have over 1500 observations to ensure the robustness of the results. If there are fewer observations, the estimation cannot be done, because the gaps in time series are too large to run an estimation. Most of the categories include two indicators, but due to data availability reasons, I include one indicator per category except for external imbalances and mispricing of risk (see table 2). There are not enough observations on changes in household credit and debt service ratio, which is why these categories only include private sector credit growth and private sector debt service ratio. The same applies to the property prices category. House price developments are only included for the whole country. Data on house prices in capital areas are not available for every country in the dataset, and would not be as descriptive, as the importance and size of capital areas vary depending on the country. In the category of external imbalances, both indicators, the current account-to-GDP and the combination indicator, are included. I also include the Basel gap calculated with two alternative definitions of credit: one according to the ESRB standard and the other with a more narrow definition of credit. The indicators of potential mispricing of risk are the average margins for new business loans to the private sector and new housing loans to households. However, data on lending margins is only available from 2003 onwards, which is not sufficient for this analysis. That is why I proxy lending margins with interest rates of new business loans to the private sector and interest rates of new housing loans.

Due to data availability reasons, I omit the bank balance sheet indicator from the analysis. Lang et al. (2019) omit balance sheet indicators due to the same reason.

[1]AT BE CY EE FI FR DE DK GB GR IE IT LV LT LU MT NL PT SK SI ES SE

Variable	Obs	Mean	Std.Dev.	Min	Max
Basel gap	2,961	-0.481	10.87	-81.34	49.07
Basel gap, narrow credit	3,472	0.184	14.80	-62.86	86.45
Private sector credit-to-GDP, annual growth	3,213	0.0541	0.0667	-0.348	0.498
House prices relative to consumer prices, 3-year avg, growth	2,704	2.110	7.418	-18.05	61.96
Debt service-to-income ratio, 2-y avg. ear change	2,921	0.0977	1.550	-12.99	15.51
Current account-to-GDP	2,502	-0.819	4.758	-22.97	13.53
Combination indicator	2,045	1.506	6.061	0.275	170.9
Interest rate, new business loans, 2-year avg. growth	2,118	-0.311	0.943	-6.444	4.315
Interest rate, housing loans, 2-year avg. growth	2,189	-0.307	0.839	-6.444	4.315

Table 2: Summary of sub-indicators

2.1.2 Crisis data

The crisis data is based on the ECB/European Systemic Risk Board (2022)) and reaches from 1970 to 2016. The dataset classifies the nature and origin of the crisis so that it can be determined whether the crisis was systemic or not and whether it originated from the domestic system or a foreign one. In this exercise, a financial crisis is defined as a systemic event that is not purely of a foreign origin. The dataset includes a crisis dummy, which takes the value 1, during the crisis period. In addition, the dataset includes three benchmark variables that flag the pre-crisis periods. The benchmark variables receive a value of 1 in the quarters preceding the crisis and 0 in normal times. The value is empty from the end of a determined pre-crisis horizon until the end of the crisis. The pre-crisis horizons included are 12 to 5, 16 to 5, and 20 to 5 quarters. The main variable of interest in my exercise is the pre-crisis dummy which denotes the period from 12 to 5 quarters before a crisis.

In addition to ECB/ESRB dataset, I use the crisis dataset by Laeven and Valencia (2018) to conduct a robustness check. The dataset includes systemic banking crises between 1970-2017. Crisis duration is limited to 5 years in the Laeven and Valencia (2018) dataset, starting from the beginning of the crisis period. For this exercise, truncating the duration of the crisis is less significant, as the main interest is in pre-crisis events. A systemic event is defined as a crisis if it fills the following requirements: there are significant signs of financial distress in the banking system, such as bank runs or bank liquidations, and policy interventions are made as a response to the financial distress. Policy interventions include deposit freezes, liquidity support, bank restructuring, nationalization, asset purchases, and significant guarantees. Three of these have to be used, for the definition of policy intervention definition to hold.

2.2 Methods

I will be following in the footsteps of Lang et al. (2019) in forming the composite indicator. Lang et al. (2019) form the composite indicator, also called the domestic systemic risk indicator (d-SRI), by determining optimal weights for each sub-indicator. In this study, I use the indicator set of Finland as sub-indicators. The purpose of this exercise is to extend the early warning system to include a composite indicator formed from chosen EWIs. The composite indicator helps in analyzing the relevant importance of individual indicators and is easier to interpret compared to a dozen individual indicators.

First, each sub-indicator is normalized so that the sub-indicators can be interpreted together. This is done by subtracting the median of the sub-indicator from its current value and dividing the result by the standard deviation of the pooled distribution of the sub-indicator. The pre-crisis horizon is then regressed on the normalized sub-indicators. Robust standard errors are clustered at the country level. Here, the pre-crisis horizon is 12-5 quarters before the event. After running the regression, the regression weights are summed up and standardized to one. If the weight of any single sub-indicator is under 5%, a minimum weight of 5% is set to the indicator and a constrained regression is run to retrieve the coefficients and weights.

Two versions are generated of each sub-indicator. One is used for in-sample evaluation and the other is used for out-of-sample evaluation. The in-sample sub-indicators are formed by calculating the normalization from the full data sample. The out-of-sample estimation on the other hand is done by generating recursive real-time variables (Sarlin (2013)). First, a starting point, in this case, 2000q1, is selected, and data up to 12 quarters pre-starting point is used to form the normalized sub-indicator. The same exercise is performed over and over, moving forward by a quarter each round until 2016q4. The recursive real-time variable contains therefore the normalized values based on data leading up to a given point in time.

The regressions are run with both in-sample and out-of-sample variables and the performance of these estimations is evaluated by retrieving AUROC values. AUROC values are retrieved by running a logit analysis with the pre-crisis horizon as the dependent variable and the sub-indicators as explanatory variables. AUROC, or the area under the curve, describes the performance of a chosen indicator. It is a measure of how well a given indicator can separate between classes 0 and 1. In this case, the pre-crisis horizon is the class of 1. The AUROC is retrieved by plotting the true positive rate against the false positive rate. It is the area under the ROC curve. (Fawcett (2006))

$$true\ positive\ rate = \frac{true\ positives}{true\ positives + false\ negatives} \tag{1}$$

$$false\ positive\ rate = \frac{false\ positives}{false\ positives + true\ negatives} \tag{2}$$

The relative usefulness is estimated in addition to the performance by minimizing a loss function defined in Alessi and Detken (2011). Other definitions of a loss function could be used as well. The loss function is used to determine signaling thresholds by minimizing the loss associated with missing crises and issuing false signals.

$$L = \theta * \frac{C}{A+C} + (1-\theta) * \frac{B}{B+D} \tag{3}$$

where A is the number of correct signals (true positives), B is the number of false positives, C is the number of false negatives and D is the number of true negatives. θ is a preference parameter that indicates the policy-maker's preference between missing crises and issuing false signals. A θ of 0.5 means both types of errors are weighted equally, $\theta < 0.5$ a preference for avoiding false alarms, and $\theta > 0.5$ a preference for avoiding missed crises.

3 Results

The in-sample results estimated with data reaching from 1970 to 2017 sets the most weight on the debt service ratio (29%), current account to GDP gap (20%), interest rates on new business loans (12%), and real estate prices relative to the consumer price index (12%). The out-of-sample estimates offer similar results but place more value on private sector credit growth, the combination indicator, and the interest rate on new business loans (see table 3). All the estimations are statistically significant with $p < 0.01$.

The composite indicator is formed by looking only at crises that are of domestic origin. This is because a policymaker can more reliably detect systemic build-up from inside the system. Crises due to external or sudden events do not necessarily show in the indicators.

The highest in-sample weights are assigned to the DSR, the current account-to-GDP gap, the interest rate on new business loans, and the real estate price development. The sub-indicators are all interconnected and tend to move similarly, even though they monitor different financial variables. When the private sector debt burden starts to increase significantly, it threatens financial stability. Impaired ability to pay loans decreases consumption and may result in bankruptcies and weakened firm profitability. This is

Variable	In-sample	Out-of-sample
Basel gap	5%	5%
Basel gap, narrow definition of credit	7%	5%
Private sector credit-to-GDP, annual growth	5%	23%
House prices relative to consumer prices, 3-year avg, growth	12%	5%
Debt service-to-income ratio, 2-y avg. ear change	29%	11%
Current account-to-GDP	20%	12%
Combination indicator	5%	11%
Interest rate, new business loans, 2-year avg. growth	12%	21%
Interest rate, housing loans, 2-year avg. growth	5%	5%

Table 3: Sub-indicator weights, in-sample and out-of-sample

reflected in banks' balance sheets, as their risks increase and credit losses rise. Finally, the negative cycle paralyzes banks' lending ability, leading to large disruptions in the financial markets. If lending is halted, investments and economic growth stop, and the risk of a banking crisis increases considerably.

In small open economies, such as Finland, increased current account deficits is a signal of an overheating financial cycle. When domestic demand is high, the demand for credit is also elevated. The amount of foreign credit is increased to meet domestic demand, which consequently increases the current account deficit. A large deficit means that the country is a net importer. Eventually, running a deficit will start to affect the purchasing power of households and firms. Foreign products will become more expensive which reduces the income available for other consumption. Reduced purchasing power negatively affects households' and firms' loan repayment capacity and increases the riskiness of banks.

Compared to the significance that is given to the Basel gap in the ESRB recommendation and country frameworks, the weight of the sub-indicator is surprisingly low. More weight is assigned to the Basel gap with a narrow definition of credit than to the standard Basel gap. This in itself is not surprising, as the standard Basel gap is more volatile compared to the one with a narrow definition of credit (see figure 1). When the sub-indicator on private sector credit growth is omitted, the weight assigned to the Basel gap (narrow definition of credit) increases to 10.5%. The sub-indicators on private sector credit and the Basel gap consist of similar data, which partly explains the results.

Whilst being a good standalone indicator, Lang et al. (2019) note that the Basel gap has limitations related to trend smoothing. This is a possible reason to why the composite indicator prefers the sub-indicator on private sector credit growth over the Basel gap indicators. A prolonged period of booming credit may bias the indicator downwards.

Also, the trend component may affect the interpretation of the indicator, if credit-to-GDP growth does not match the trend smoothing component.

The development of the Basel gap is illustrated in figure 1. The blue line demonstrates the standard Basel gap and the red line is the version with a narrower definition of credit, which includes bank credit to the non-financial private sector. The red lines indicate the thresholds for evaluating the need for the CCyB. The upper red line is crossed when the gap exceeds 10 and based on the European Systemic Risk Board (2014a) recommendation would mean activating the maximum CCyB of 2.5%. Both versions of the Basel gap have similar upward trends before the large events of the 1990s and 2008. In Finland, only the 1990s banking crisis was considered a systemic crisis. By comparing figures 1 and 2, the predictive ability of the d-SRI seems therefore better. The composite indicator reaches its highest values during the 1990s crisis, whereas the Basel indicator gives a stronger signal prior to the global financial crisis.

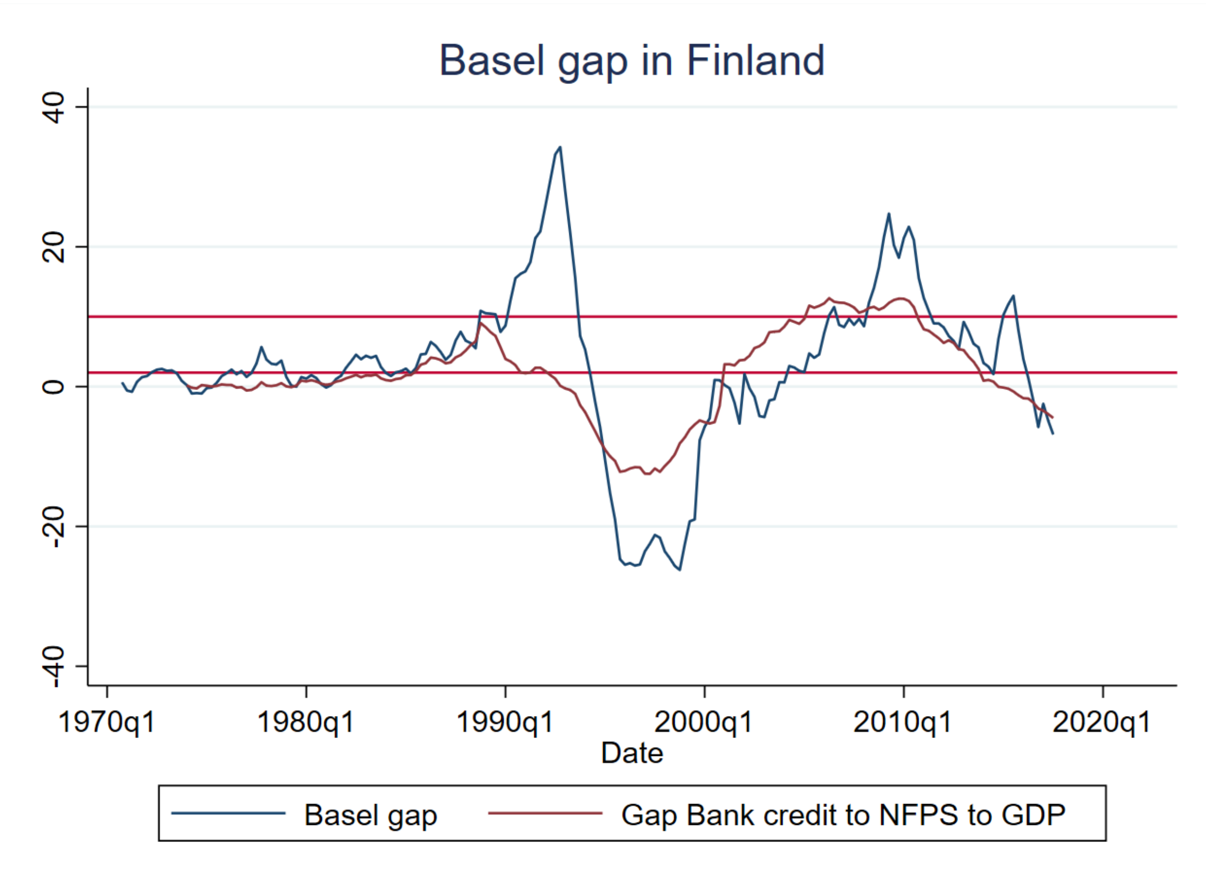

Figure 1: Development of the Basel gap

Figure 2 displays the different versions of the composite indicator (d-SRI). d-SRI benchmark illustrates the development of the indicator composed by the in-sample analysis. The d-SRI (1995) is the same indicator but with optimal weights based on 1995. Finally, the d-SRI (real-time) illustrates how the indicator based on the recursive real time variables performs in the review period. Out of these different versions, the d-SRI benchmark seems to have the best performance. It reaches the highest pre-crisis values and remains modest before the 2008 global financial crisis. Still, this conclusion should be considered with caution, since the time frame includes only one systemic crisis in Finland. The 2008 financial crisis did not trigger a banking crisis in Finland even though the financial markets were under significant stress. The signaling threshold of the d-SRI benchmark is at 0.593 (see table 4) and the indicator does not exceed the threshold prior to 2008. The signaling threshold is estimated with the Alessi and Detken (2011) loss function (formula 3), where $\theta = 0.5$.

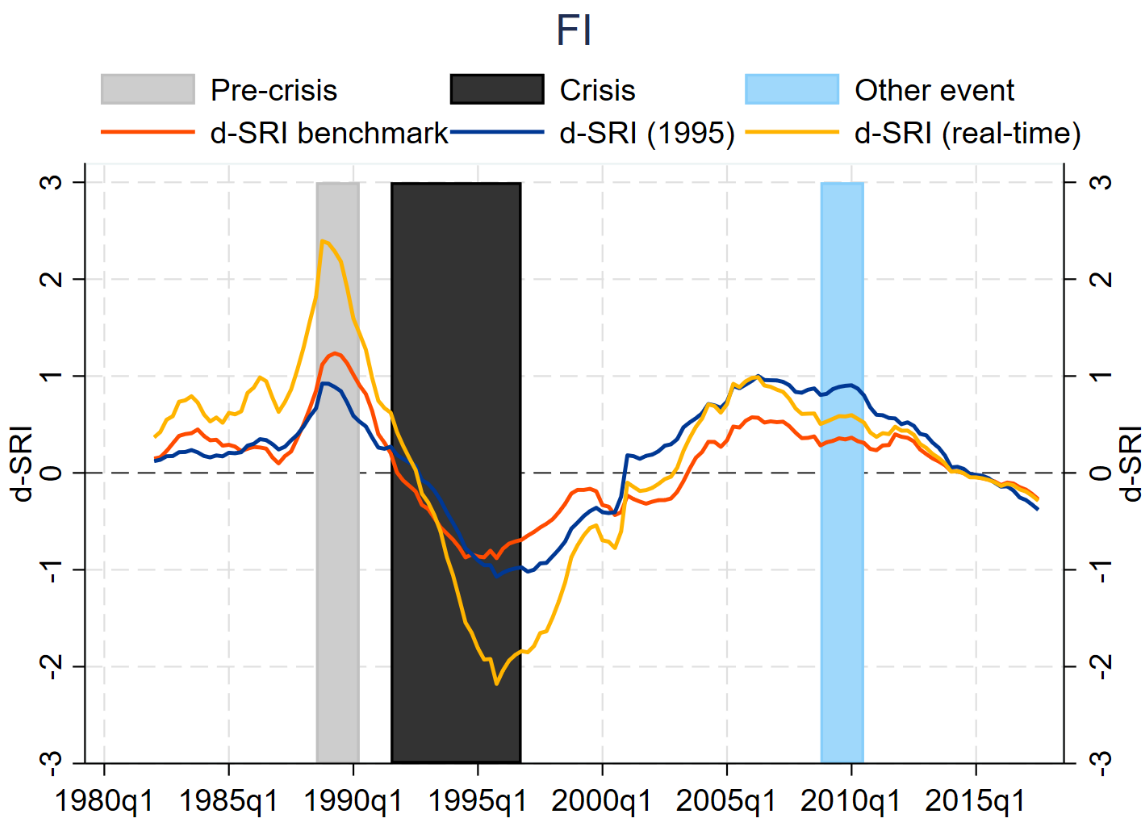

Figure 2: The performance of the composite indicator in Finland

3.1 Indicator performance

The d-SRI has an AUROC value of 0.756 and a noise-to-signal ratio of 0.218 (see table 4). Therefore, the AUROC value is slightly higher compared to the standalone AUROC of the Basel gap, which is currently used as the primary risk indicator. It can therefore be concluded that the d-SRI has a good predictive power.

Table 4 displays the predictive performance of the composite indicator in-sample and out-of-sample with a pre-crisis horizon of 12-5 quarters. Usefulness measures are calculated based on the loss function. First, policy-makers losses are calculated based on the Alessi and Detken (2011) loss function. Here, the preference parameter is set to $\theta = 0.5$. Opting for a middle-ground calibration acknowledging the trade-off between missed crises and false alarms, without taking a stance on preferences or economic costs. Loss no-model indicates the losses associated with not using an early warning model and is defined as $min(\theta, 1 - \theta)$.

Absolute usefulness = loss no-model - loss
Relative usefulness = Absolute usefulness / loss no-model

Table 4 shows that the d-SRI performs well both in-sample and out-of-sample. Noise-to-signal ratio is actually slightly smaller in the out-of-sample estimation. Usefulness measures are slightly better in the in-sample estimation.

The composite indicator receives good AUROC values. The signaling performance is similar to the best-performing individual indicators, whilst maintaining an acceptable noise-to-signal ratio. The composite indicator does not outperform the single indicators but is a good addition to the early warning framework used in Finland. The d-SRI also helps in estimating how to weigh increased indicator values, if the EWIs would issue contradictory signals.

Finally, the d-SRI fulfills the characteristics of a good indicator. As Drehmann and Juselius stated, a good EWI should alarm the policymakers early enough, provide stable information, and be easy to interpret. The d-SRI meets these requirements, as it is elevated up to 12 quarters before a systemic crisis. The noise-to-signal ratio is also fairly good at 0.218. A clear elevation and crossing of a predetermined threshold signal that the probability of a systemic crisis has increased, fulfilling the final requirement of easy interpretation.

	Including Basel gap indicators		Basel gap indicators excluded	
	Pre-crisis period, 12-5 quarters	Pre-crisis period, 12-5 quarters	Pre-crisis period, 12-5 quarters	Pre-crisis period, 12-5 quarters
	In-sample	Out-of-sample	In-sample	Out-of-sample
In-sample	2.060***		1.820***	
	(0.505)		(0.332)	
Out-of-sample		1.869***		1.810***
		(0.342)		(0.326)
Constant	-2.813***	-2.584***	-2.726***	-2.550***
	(0.431)	(0.424)	(0.377)	(0.340)
Observations	770	770	770	770
Pseudo R2	0.202	0.192	0.213	0.206
AUROC	0.756	0.770	0.768	0.773
Signalling Threshold	0.593	0.609	0.705	0.618
Relative Usefulness	0.401	0.392	0.419	0.407
Absolute Usefulness	0.200	0.196	0.210	0.203
Loss	0.300	0.304	0.290	0.297
Loss No-Model	0.500	0.500	0.500	0.500
Noise-to-Signal Ratio	0.218	0.174	0.140	0.143
False negatives	0.487	0.525	0.512	0.525
False positives	0.112	0.0826	0.0681	0.0681
True Positives	41	38	39	38
False Positives	77	57	47	47
True Negatives	613	633	643	643
False Negatives	39	42	41	42
Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1				
Policy-makers preference parameter $\theta = 0.5$				

Table 4: d-SRI performance in-sample and out-of-sample

3.2 The impact of excluding variables from the d-SRI

Excluding variables impacts both the optimal sub-indicator weights and the performance of the composite indicator. A potential challenge with this type of estimation is that even though a sub-indicator would perform well as a standalone indicator, it may lose its predictive power in a combination indicator. This is because the information captured by a specific indicator might overlap with that of another indicator. Conversely, this logic applies in reverse as well. Indicators that would not perform well as standalone indicators may contain an additional informational value in a composite indicator, enhancing the performance of the composite indicator.

I test whether the Basel gap adds value to the composite indicator by running the exercise without either version of these sub-indicators. The results are displayed in appendices in tables 7 and 8. The results suggest that including Basel gap sub-indicators is not necessary. The AUROC values of both versions are fairly similar with the sub-indicators (0.765) and without (0.768) the indicators and the noise-to-signal values are 0.218 and 0.140 respectively (see tables 7 and 8). When excluding the Basel gap indicators from the scenario, the weight of the private sector credit-to-GDP indicator increases significantly to

35%. The underlying time series that are used to construct the Basel gap indicators and the private sector credit development indicator are similar, which to some extend explains the results. Additionally, the weight assigned to the DSR decreases to 18%. Other than that, the sub-indicator weights remain fairly similar (See appendices, table 5)

Excluding the DSR from the analysis decreases the noise-to-signal ratio considerably, from 0.2 to 0.09. However, when the DSR is excluded, the relative weight of the combination indicator increases to 53% which is disproportionately much. Assigning that much weight to a single sub-indicator defeats the purpose of forming a composite indicator that would monitor the financial market as comprehensively as possible. Thus, I propose including the DSR as a component of the d-SRI. While it demonstrates strong predictive capabilities independently, it also appears to perform well as a part of the d-SRI.

In comparison, excluding the current account-to-GDP indicator raises the noise-to-signal ratio to 0.595. It appears that including an indicator of external balance is important for predicting crises. The possible reason behind this is that banking crises are often contagious. Therefore, significant concentrations in the current account can expose the country more severely to the contagion risk. Based on figure 3, it appears that excluding the current account from the GDP indicator weakens the predictive power of the d-SRI in Finland. The warning signal before the 1990s banking crisis would have been issued later and the indicator does not reach as high values, compared to the benchmark.

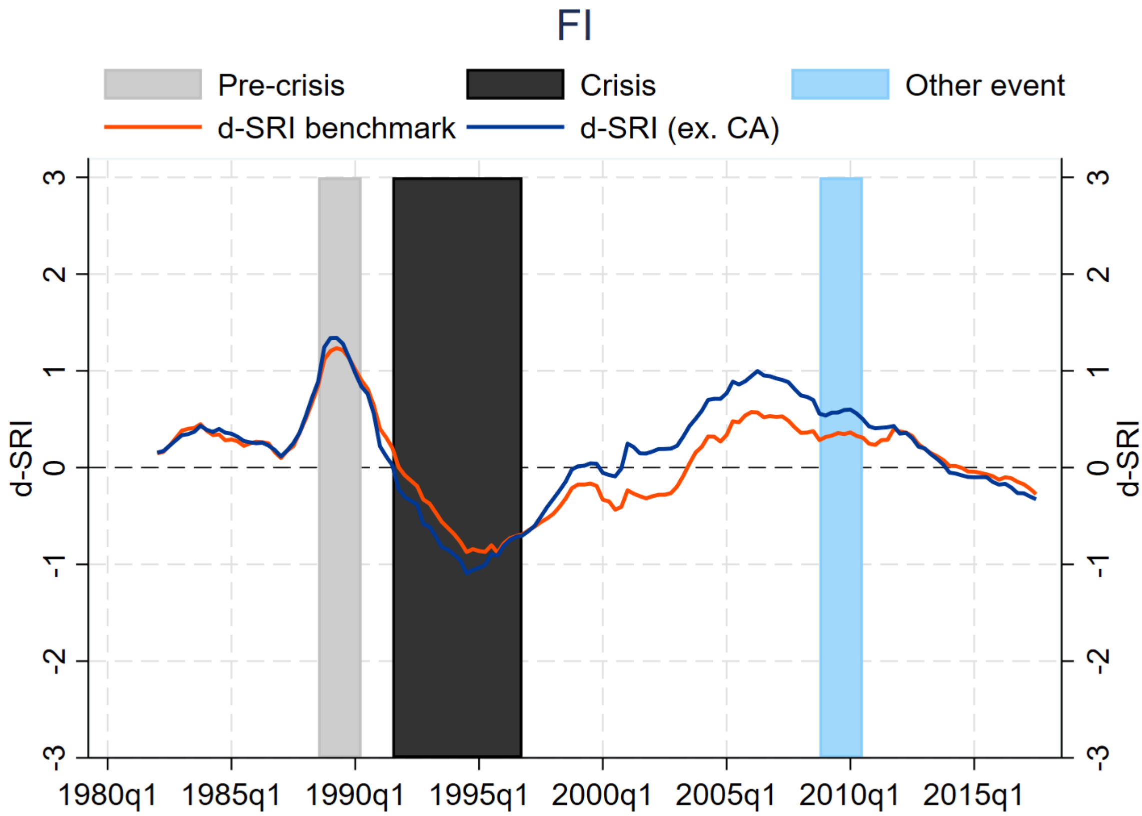

Figure 3: Composite indicator performance, excluding current account sub-indicator

In addition to good predictive power, the level of the composite indicator seems to negatively correlate with the severity of the crisis. Figure 4 illustrates this for a selection of countries. The regression includes those countries, to which the d-SRI could be generated and which have at least one pre-crisis period in which the d-SRI receives values. The countries that had the highest d-SRI values before the systemic crisis also had the largest drops in GDP. This can be explained by the accumulation of systemic risk. When financial imbalances are large, the crash associated is more likely costlier. That is why smoothing the financial cycle is important. When imbalances are not allowed to increase into large bubbles, the potential crisis and cost associated with it remain more moderate.

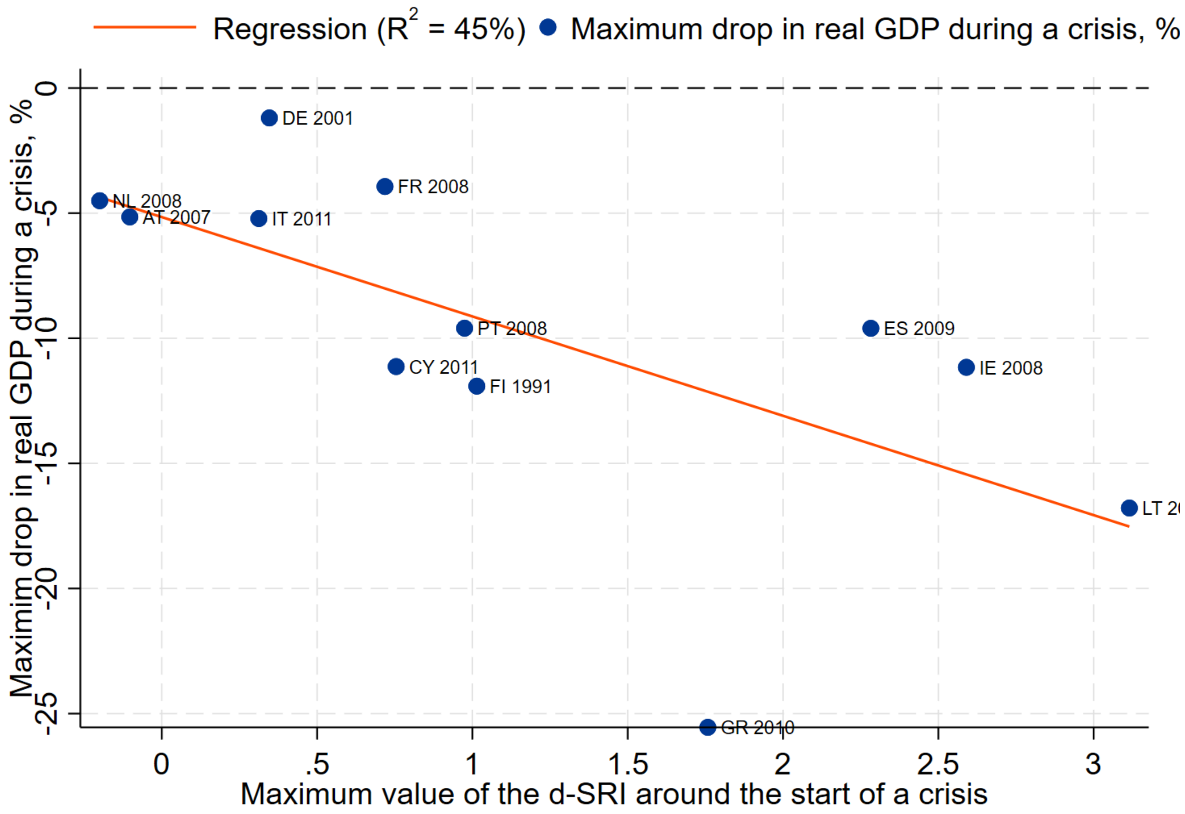

Figure 4: Drop in GDP compared to the d-SRI value

3.3 Robustness check

I run the estimation with an alternative systemic banking crisis dataset constructed by Laeven and Valencia (2018). The results vary depending on the crisis data. The in-sample weights based in this case assign 28% weight to private sector credit growth and 37 % to the average interest rate of new business loans. All other sub-indicators receive a 5% weight. However, when either of the two most weighted sub-indicators is excluded, the debt service-to-income ratio receives a higher weighting. It increases to 33% when private sector credit is excluded and 15% when the average interest rate of new business loans is excluded. These results are more in line with the weightings derived from running the estimation with the ECB crisis data. Private sector credit is often found to increase significantly before a systemic crisis, as happened for example prior to 2008. Excessive credit builds up systemic risk, which is why the sub-indicator on private sector credit growth may receive a high relative weight. A possible explanation for the high weight assigned to the sub-indicator on the interest rate of new business loans is related to the increased demand for credit. When credit is booming and demand for loans is high, interest rates tend to increase as well. A sudden increase in interest rates signals the policymaker that risks in the financial market are increasing.

Firstly, a significant difference between the crisis datasets is that Laeven and Valencia (2018) do not distinguish between foreign and domestic origin crises. As foreign crises may be first visible in different indicators (or in none of the indicators at all), such as the current account to GDP ratio, the crisis definition may cause alternative results. Secondly, since crises originating from foreign countries are more unpredictable, the noise-to-signal ratio is higher when (0.306) the estimation is run with Laeven and Valencia (2018) crisis data.

Another possible cause for differing estimates is that the Laeven and Valencia (2018) dataset includes fewer crises, most of which are in the 1990s or the global financial crisis. For example in the case of Italy, Laeven and Valencia (2018) define a systemic crisis beginning in 2008. The ECB crisis data defines this as an "other event", but places a systemic crisis starting in 2012. The similarity of crises and their origins in the Laeven and Valencia (2018) data may cause the estimation to place the largest weight on those singular sub-indicators that fit the data the best.

Identifying indicators that have the best predictive power in estimating one or two similar systemic events is not desirable, as the cause may differ the next time a crisis occurs. Here, the in-sample performance is good, with an AUROC of 0.838. However, the estimated noise-to-signal ratio (0.306) is higher compared to the d-SRI estimated with ECB data (0.218). As the ECB crisis data considers a wider set of crises systemic, the data includes more crises. Consequently the variation in predictive sub-indicators increases resulting in more dispersed weighting.

3.4 Considerations

The study is subject to several limitations. The most prevalent of these is data availability issues. Long enough time series are available only for a limited number of countries, and I had to omit some sub-indicators due to data availability issues. Data collection was standardized with the EU, which is why historical values may not be as reliable. This may affect the observed indicator performance and weights, especially when looking at historical values. Sample fit could be biased upwards or downwards, depending on how good the historical approximations are. However, this cannot be determined simply by looking at the data.

Furthermore, the study is conducted with a selection of EU countries, meaning that optimal country-level weights cannot be retrieved from the exercise. The exercise could not be performed in any single country anyways, because there is not enough data available for any single country that would contain enough systemic crises. Even if a country like this would exist, the estimation would not be reliable, because historical crises and country fundamentals (eg. legislation) differ significantly from the current day state. The best option is to try and select countries that are similar in terms of economic fundamentals and governance. That is why a selection of EU countries and Great Britain were selected for this empirical exercise.

However, studying only selected EU countries and GB introduces another limitation: the historical systemic crises are relatively similar. Similar origins of crises may put additional weight on those sub-indicators that prior to those crises experienced overheating. These sub-indicators would not necessarily be the most informative in the case of predicting future crises.

Additionally, the definition of a crisis may vary. I address this issue by performing the analysis with an alternative crisis dataset. However, the results of the robustness check provide little additional value. Due to the crisis definition of Laeven and Valencia (2018), their crisis dataset is more limited. This results in a higher noise-to-signal ratio and the estimation giving the most weight to only two sub-indicators.

In addition to data issues, the results are influenced by the calibration of the preference parameter. In this study, the preference parameter (denoted as θ) is set to $\theta = 0.5$ in order to take a neutral stance on policy preferences. However, with a different calibration the regulator can choose between a preference on missed crises and false alarms. A lower number of missed crises leads to more false alarms and a higher noise-to-signal ratio, and vice versa, as illustrated in figure 5. Particularly, setting the preference parameter above 0.55 incurs a higher number of false signals when aiming for fewer missed crises. As the purpose of this study is not to take a stand on optimal calibration and analyse the costs of this trade-off, results are derived using the mentioned $\theta = 0.5$.

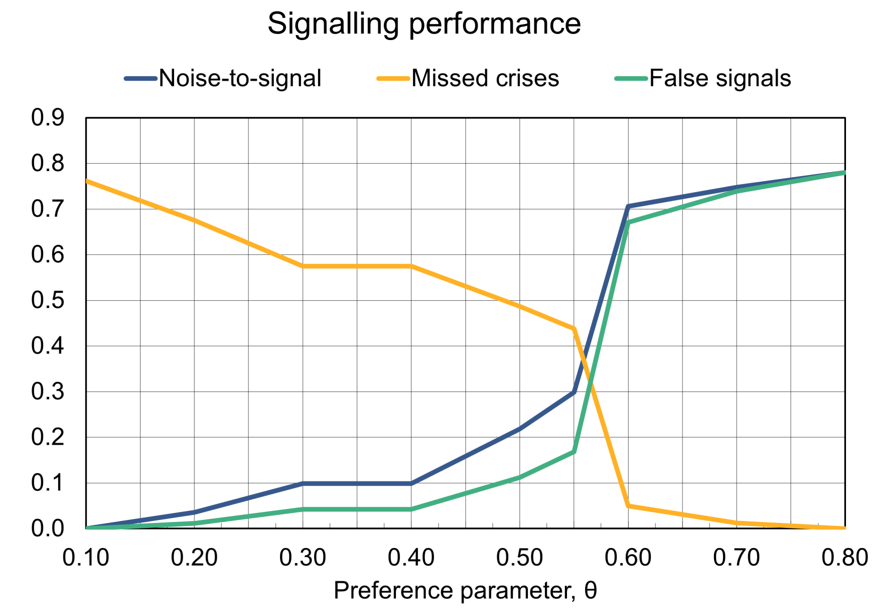

Figure 5: Trade-off between missed crises and false signals

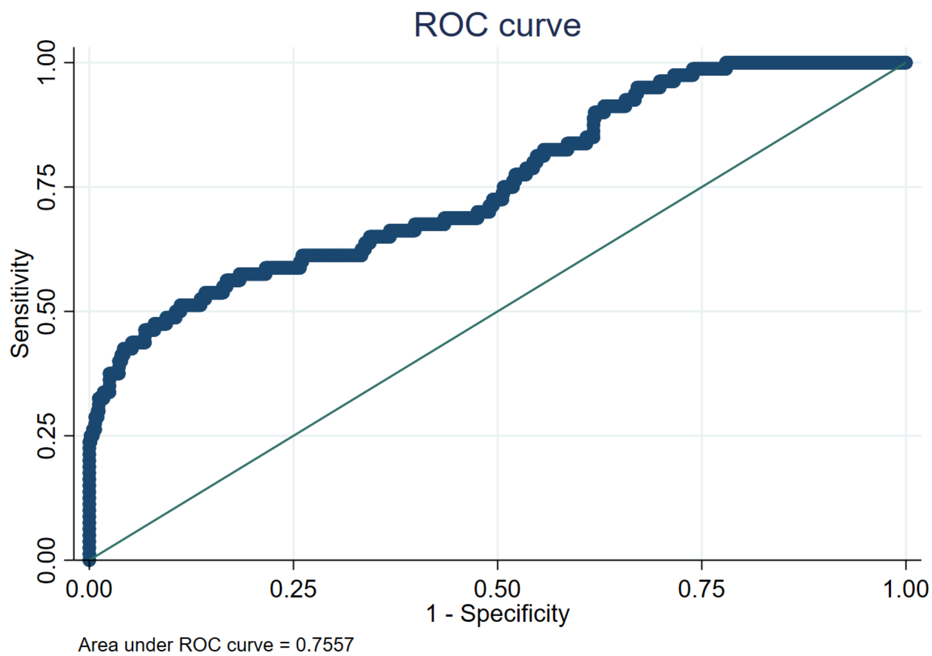

Figure 6: The ROC curve

4 Conclusion

The most important finding of this study is that the composite indicator (d-SRI) constructed from the EWIs chosen by the Bank of Finland fulfills the requirements of a good early warning indicator. It has a good predictive ability (AUROC value of 0.756) and a low noise-to-signal ratio (0.218). It is also easy to interpret and consistent in the data sample. When the d-SRI reaches high values, the probability of a banking crisis in the future increases. Low values indicate that the financial cycle remains subdued.

I perform a robustness check with an alternative crisis dataset by Laeven and Valencia (2018), but due to a stricter definition of a systemic crisis, the results are not as reliable. There are not enough systemic events from alternative causes to make convincing conclusions on the sub-indicator weights and performance of the d-SRI.

I also test how the d-SRI performs, when various sub-indicators are omitted from the estimation. Surprisingly, excluding the Basel gap sub-indicators doesn't significantly impact the performance but notably increases the relative weight of the sub-indicator on private sector credit development. This departure from expectations, given the Basel gap's perceived importance according to European Systemic Risk Board (2014a), is likely due to the inclusion of a sub-indicator focusing on private sector credit developments. The time series from which the Basel gap indicators are calculated from are highly similar with those used to calculate the private sector credit development.

Excluding the sub-indicator on current account-to-GDP development increases the noise-to-signal ratio significantly up to 0.6. This can be explained by contagion effects. Banking crises are often contagious in foreign countries. This is also the case for Finland because the most important systemic risks of the banking sector are related to large exposures in the Nordic housing markets. Excluding the sub-indicator on DSR, which received the highest relative weighting (29%) decreases the noise-to-signal ratio from 0.2 to 0.09. In this case, the combination indicator receives a weight of 53% which is disproportionately much and therefore undesirable. The composite indicator would not allow for a comprehensive evaluation of the financial market if a single indicator would receive such a high value.

The value the d-SRI reached in the pre-crisis period is correlated with the maximum GDP drop associated with the crisis. Based on the data, it seems that a higher d-SRI value preceded a larger drop. This is only an observation and does consider whether other factors could affect both variables. Therefore this observed correlation should be regarded only as a preliminary finding.

Finally, I address the limitations related to this study. The most prevalent of these are data availability issues. I had to omit the sub-indicator on bank balance sheets because there was no consistent cross-country data available. I also had to proxy lending margins with interest rates. Data collection was harmonized with the EU, which is why historical values before 2003 are not as reliable. These data issues may potentially bias my estimations upwards or downwards and affect the weightings of sub-indicators and observed performance of the d-SRI.

References

Iñaki Aldasoro, Claudio Borio, and Mathias Drehmann. Early warning indicators of banking crises: Expanding the family. *IS Quarterly Review*, 2018. URL https://ssrn.com/abstract=3139160.

Lucia Alessi and Carsten Detken. Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3):520–53, 2011.

Bank of Finland. Suhdannesidonnaisten haavoittuvuuksien uudistettu riskimittaristo. *Makrovakausraportti*, 1, 2022.

Claudio Borio and Matthias Drehmann. Assessing the risk of banking crises – revisited. Technical report, 2009.

Claudio Borio and Philip Lowe. Asset prices, financial and monetary stability: exploring the nexus. *BIS Working Papers*, 114, 2002.

Carsten Detken, Diana Bonfim Olaf Weeken, Lucia Alessi, Miguel M. Boucinha, Christian Castro, Sebastian Frontczak, Gaston Giordana, Julia Giese, Nadya Jahn, Jan Kakes, Benjamin Klaus, Jan Hannes Lang, Natalia Puzanova, and Peter Welz. Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options. *ESRB, Occasional Paper Series*, 5, 2014.

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

European Systemic Risk Board. Recommendation of the european systemic risk board on guidance for setting countercyclical buffer rates. *Official Journal of the European Union*, C 293/01, 2014a.

European Systemic Risk Board. Recommendation of the european systemic risk board on guidance for setting countercyclical buffer rates. *Official Journal of the European Union*, C 293/1, 2014b.

European Systemic Risk Board. European financial crises database. 2022. URL https://www.esrb.europa.eu/pub/financial-crises/html/index.en.html.

Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27 (8):861–874, 2006.

Heidi Koponen. Suhdannesidonnaisten riskimittareiden yhdistelmäindikaattori. *Makrovakausraportti, Suomen Pankki*, 1, 2023.

Luc Laeven and Fabian Valencia. Systemic banking crises revisited. *IMF working paper*, 18/206, 2018.

Jan Hannes Lang, Cosimo Izzo, Stephan Fahr, and Josef Ruzicka. Anticipating the bust: a new cyclical systemic risk indicator to assess the likelihood and severity of financial crises. *ECB Occasional Paper Series*, 219, 2019.

Ministry of Finance. Valtiovarainministeriön asetus luottolaitoksen ja sijoituspalveluyrityksen muuttuvasta lisäpääomavaatimuksesta. 1029, 2014.

Peter Sarlin. On policymakers' loss functions and the evaluation of early warning systems. *Economics Letters*, 119:1–7, 2013.

Eero Tölö, Helinä Lakkonen, and Simo Kalatie. Evaluating indicators for use in setting the countercyclical capital buffer. *International Journal of Central Banking*, 14, 2018.

5 Appendices

	CBNF_GDP	RREPN_development	DSR_2y	combination_indic	CA_GDP	corp_interest	hh_interest
Coefficients: nis	0.08	0.01	0.04	0.01	0.04	0.03	0.01
Weights: nis	0.35	0.05	0.18	0.05	0.17	0.15	0.05
Coefficients: nis2	0.06	0.01	0.05	0.01	0.03	0.04	0.01
Weights: nis2	0.26	0.05	0.22	0.06	0.15	0.21	0.05

Table 5: Coefficients and weights of sub-indicators when Basel gap sub-indicators are omitted

	GAP_CBNFA	GAP	CBNF_GDP	RREPN	DSR	comb_indic	CA_GDP	corp_interest	hh_interest
Coefficients: in-sample	0.015	0.011	0.011	0.027	0.064	0.011	0.045	0.027	0.011
Weights: in-sample	0.067	0.050	0.050	0.121	0.288	0.050	0.205	0.120	0.050
Coefficients: out-of-sample	0.011	0.011	0.052	0.011	0.029	0.025	0.026	0.047	0.011
Weights: out-of-sample	0.050	0.050	0.230	0.050	0.130	0.111	0.118	0.210	0.050
Coefficients: nis; excl. CBNF_GDP	0.036	0.011	0.000	0.035	0.011	0.011	0.052	0.050	0.011
Weights: nis; excl. CBNF_GDP	0.168	0.050	0.000	0.162	0.050	0.050	0.239	0.232	0.050
Coefficients: nis; excl. RREPN	0.020	0.013	0.013	0.000	0.077	0.048	0.044	0.026	0.013
Weights: nis; excl. RREPN	0.080	0.050	0.050	0.000	0.304	0.191	0.173	0.102	0.050
Coefficients: nis; excl. DSR_2y	0.025	0.023	0.023	0.032	0.000	0.241	0.049	0.042	0.023
Weights: nis; excl. DSR_2y	0.055	0.050	0.050	0.069	0.000	0.527	0.108	0.091	0.050
Coefficients: nis; excl. comb_indic	0.010	0.010	0.068	0.010	0.024	0.000	0.030	0.037	0.010
Weights: nis; excl. comb_indic	0.050	0.050	0.342	0.050	0.120	0.000	0.151	0.186	0.050
Coefficients: nis; excl. CA_GDP	0.032	0.015	0.015	0.025	0.076	0.110	0.000	0.015	0.015
Weights: nis; excl. CA_GDP	0.105	0.050	0.050	0.081	0.251	0.363	0.000	0.050	0.050
Coefficients: nis; excl. corp_interest	0.027	0.027	0.027	0.028	0.027	0.322	0.038	0.000	0.038
Weights: nis; excl. corp_interest	0.050	0.050	0.050	0.053	0.050	0.604	0.072	0.000	0.071
Coefficients: nis; excl. hh_interest	0.017	0.017	0.017	0.026	0.055	0.130	0.043	0.040	0.000
Weights: nis; excl. hh_interest	0.050	0.050	0.050	0.075	0.159	0.376	0.125	0.115	0.000

Table 6: Coefficients and weights of sub-indicators

VARIABLES	In-sample	Out-of-sample	excl. Basel gap	excl. Basel gap (narrow)	excl. credit growth
Coeff.	2.060***	1.869***	2.958***	4.496***	2.103***
Constant	-2.813***	-2.584***	-2.763***	-2.788***	-2.823***
Observations	770	770	770	770	770
Pseudo R2	0.202	0.192	0.197	0.207	0.195
AUROC	0.756	0.770	0.745	0.758	0.762
Signalling Threshold	0.593	0.609	0.439	0.280	0.526
Relative Usefulness	0.401	0.392	0.400	0.426	0.422
Absolute Usefulness	0.200	0.196	0.200	0.213	0.211
Loss	0.300	0.304	0.300	0.287	0.289
Loss No-Model	0.500	0.500	0.500	0.500	0.500
Noise-2-Signal Ratio	0.218	0.174	0.159	0.170	0.250
False negatives	0.487	0.525	0.525	0.487	0.438
False positives	0.112	0.0826	0.0754	0.0870	0.141
True Positives	41	38	38	41	45
False Positives	77	57	52	60	97
True Negatives	613	633	638	630	593
False Negatives	39	42	42	39	35

Table 7: Performance, excluded variables

VARIABLES	**excl. House price growth**	**excl. DSR**	**excl. Comb. indic.**	**excl. CA-to-GDP**	**excl. Corp. int. rate**
Coeff.	2.303***	4.199***	1.717***	2.880***	4.634***
Constant	-2.766***	-2.773***	-2.754***	-2.885***	-2.780***
Observations	770	770	770	770	770
Pseudo R2	0.193	0.195	0.208	0.191	0.185
AUROC	0.757	0.752	0.772	0.765	0.742
Signalling Threshold	0.544	0.385	0.808	-0.0256	0.289
Relative Usefulness	0.395	0.432	0.410	0.380	0.414
Absolute Usefulness	0.197	0.216	0.205	0.190	0.207
Loss	0.303	0.284	0.295	0.310	0.293
Loss No-Model	0.500	0.500	0.500	0.500	0.500
Noise-2-Signal Ratio	0.190	0.0915	0.137	0.595	0.152
False negatives	0.512	0.525	0.525	0.0625	0.512
False positives	0.0928	0.0435	0.0652	0.558	0.0739
True Positives	39	38	38	75	39
False Positives	64	30	45	385	51
True Negatives	626	660	645	305	639
False Negatives	41	42	42	5	41

Table 8: Performance, excluded variables

BoF Economics Review

2021	No 1	Kärkkäinen, Samu; Nyholm, Juho: Economic effects of a debt-to-income constraint in Finland : Evidence from Aino 3.0 model
	No 2	Nyholm, Juho; Voutilainen, Ville: Quantiles of growth : household debt and growth vulnerabilities in Finland
	No 3	Juselius, Mikael; Tarashev, Nikola: Could corporate credit losses turn out higher than expected?
	No 4	Nelimarkka, Jaakko; Laine, Olli-Matti: The effects of the ECB's pandemic-related monetary policy measures
	No 5	Oinonen, Sami; Vilmi, Lauri: Analysing euro area inflation outlook with the Phillips curve
	No 6	Pönkä, Harri; Sariola, Mikko: Output gaps and cyclical indicators : Finnish evidence Analysing euro area inflation outlook with the Phillips curve
	No 7	Hellqvist, Matti; Korpinen, Kasperi: Instant payments as a new normal : Case study of liquidity impacts for the Finnish market
	No 8	Markkula, Tuomas; Takalo, Tuomas: Competition and regulation in the Finnish ATM industry
	No 9	Laine, Tatu; Korpinen, Kasperi: Measuring counterparty risk in FMIs
	No 10	Kokkinen, Arto; Obstbaum, Meri; Mäki-Fränti, Petri: Bank of Finland's long-run forecast framework with human capital
2022	No 1	Norring, Anni: Taming the tides of capital – Review of capital controls and macroprudential policy in emerging economies
	No 2	Gulan, Adam; Jokivuolle, Esa; Verona, Fabio: Optimal bank capital requirements: What do the macroeconomic models say?
	No 3	Oinonen, Sami; Virén, Matti: Has there been a change in household saving behavior in the low inflation and interest rate environment?
	No 4	Nyholm, Juho; Silvo, Aino: A model for predicting Finnish household loan stocks
	No 5	Oinonen, Sami; Virén, Matti: Why is Finland lagging behind in export growth?
	No 6	Mäki-Fränti, Petri: The effects of age and cohort on household saving
2023	No 1	Obstbaum, Meri; Oinonen, Sami; Pönkä, Harri; Vanhala, Juuso; Vilmi, Lauri: Transmission of recent shocks in a labour-DSGE model with wage rigidity
	No 2	Kärkkäinen, Samu; Silvo, Aino: Household debt, liquidity constraints and the interest rate elasticity of private consumption
	No 3	Nippala Veera, Sinivuori Taina: Forecasting private investment in Finland using Q-theory and frequency decomposition
	No 4	Hokkanen, Topi: Externalities and market failures of cryptocurrencies
2024	No 1	Kortalainen, Mika: How effective quantitative tightening can be with a higher-for-longer pledge?
	No 2	Norring, Anni: Geoeconomic fragmentation, globalization, and multilateralism
	No 3	Koponen, Heidi: Constructing a composite indicator to assess cyclical systemic risks: An early warning approach

https://www.suomenpankki.fi/en/media-and-publications/publications/studies-and-reports/bof-economics-review/