Working Paper Series

No 102 / September 2019

Exposition, climax, denouement

Life-cycle evaluation of the recent financial crisis in the EU by linking the ESRB financial crisis database to the European Commission's Macroeconomic Imbalance Procedure Scoreboard

By
Szilard Erhart

ESRB
European Systemic Risk Board
European System of Financial Supervision

Abstract

The paper investigates the life-cycle of the 2008-2009 financial crisis by linking the Macroeconomic Imbalance Procedure (MIP) Scoreboard of the European Commission to the crisis database of the European Systemic Risk Board (ESRB). The novelty of the analysis is that early warning capacity of MIP indicators is empirically tested in case of various crisis events case by case (i) Currency/Balance-of-Payment/Capital flow events, (ii) Sovereign crisis events, (iii) Banking crisis events and (iv) Significant asset price corrections in EU Member States. Furthermore, we contribute to the literature by studying the predicting power of the MIP Scoreboard in the identification of the overheating in the economy in advance of crises (preventive arm of the MIP). We found that the predictive power of the MIP Scoreboard may be twice as high to capture sovereign and Currency/Balance-of-Payment/Capital flow type of crisis events than its power to capture a banking crisis or serious asset price corrections. We confirm the results of earlier empirical studies that some MIP indicators perform relatively well (current account and net international position) in all specifications. A simple composite indicator based on the threshold breaches of MIP Scoreboard Indicators, performed in most cases as good as the best individual indicator, and hence could be considered as an input to a simple, rule based and accountable decision making.

JEL classification: C40, G01, E44, E61, G28

Keywords: early warning system, macroeconomic imbalance procedure, signal approach, financial crisis, boom and bust, systemic risk

1 Introduction

The paper reconsiders the role of the Macroeconomic Imbalance Procedure (MIP) Scoreboard and its 11 headline indicators (see further details in Section 2.1). The European Commission introduced the MIP in 2011 as part of the "Six-Pack" legislation, so as to react to the Global Financial Crisis and to treat macroeconomic imbalances in the European Union (EU). The MIP Scoreboard aims at identifying imbalances both ex-ante, during the early phase in the crisis life cycle (so as to trigger pre-emptive actions) and also at monitoring imbalances ex-post (so as to trigger corrective actions).

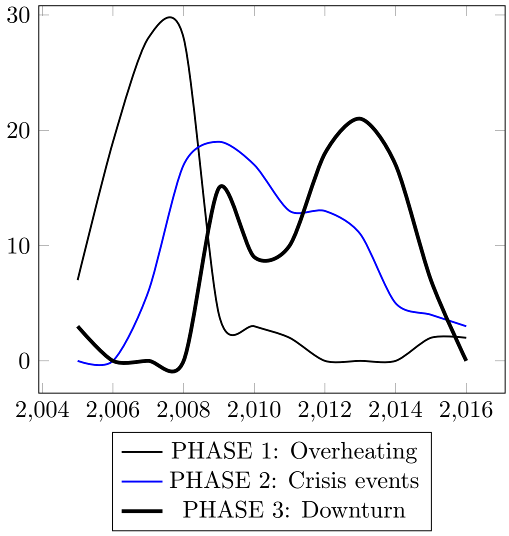
In this paper, the predictive power of the MIP Scoreboard Indicators are tested for all three phases of crisis life-cycle. Previous empirical studies on the MIP Scoreboard investigated only whether its indicators could have helped predict the 2nd and 3rd PHASE of crisis life-cycles, e.g. certain types of crisis events and economic downturns following crisis events (Figure 1) in EU Member States. This paper also tests whether MIP indicators could help to signal the 1st PHASE of crisis events, the overheating of the economy (preventive arm of the MIP).

Furthermore, we investigate systemically how the definition of a crisis event influences the predictive power of the MIP Scoreboard by linking the MIP Scoreboard to the European Systemic Risk Board's crisis database. We found that the MIP could have been able to signal two types of crisis events with the highest probability: (i) Sovereign crisis events and (ii) Currency/Balance-of-Payment/Capital flow crises in EU Members States, but not banking crises. This result can also help to reconcile the varying findings of earlier studies that defined crisis events differently. Csortos and Szalai (2014) used the cyclical GDP gap similarly to Domonkos et al. (2017), Boysen et al. (2015) analysed the union of financial crisis events and Knedlik (2014) the debt crisis events.

Our empirical findings confirmed the hypothesis of Erhart et al. (2018) who argued that the real house price growth indicator of the MIP and its current one-sided threshold in the MIP Scoreboard is able to identify the overheating of the economy well, but not economic downturns.

We combine the signals approach of empirical studies and the composite indicator methodology to evaluate the performance of the MIP Scoreboard. Such method has been proposed by the empirical study by Christensen and Li (2014) to capture the fragility of the economy prior to a financial stress event most efficiently. Lo Duca et al. (2017) also argued that the newly established financial crisis database confirms that multivariate methods can improve upon univariate signalling models.

Figure 1: Life cycle of serious imbalances in EU Member States (Number of Member States in a situation of overheating, crisis, downturn)



*All 3 crisis phases are defined in Section 2 and 3

Empirical works have concluded so far that the predictive power of the MIP Scoreboard Indicators is low (Boysen et al. (2015); Knedlik (2014) and Csortos and Szalai (2014)). Furthermore, the difficulty of early warning in case of macroeconomic surveillance is confirmed by the fact that different studies identified different indicators useful as alarm bells (Table 1). Csortos and Szalai (2014) showed that only in the cases of the current account deficit and the unemployment rate were the prediction ratios better than the ratios of false alarms to alarms total. Boysen et al. (2015) found that house prices, private sector debt, and private sector credit flow are the best early warning indicators of crisis events. Knedlik (2014) showed that the usefulness is the highest for the current account, net international investment position and nominal unit labour costs. Domonkos et al. (2017) found that in the short run private sector debt is the best performing indicator among the headline indicators, complemented by current account balances in the long term.

Knedlik (2014) also remarked that some indicators perform differently as early warning indicators in different countries. For instance, ULC performs excellently in Central and Eastern European countries but poorly in the Eurozone. Public sector debt is a very important indicator for the Eurozone but not for the rest of the European Union; the unemployment rate and export market share are relevant early warning indicators in the Eurozone but not in Central and Eastern European countries.

Finally, different explanations could be given as to why establishing effective early warning systems is a difficult analytical and policy challenge. First, vulnerabilities are time and place dependent.

Second, early warning indicators may fail to trigger action and in many cases they did fail in practice in the past, because policy makers were resistant to act on vague warnings in good times. The MIP is definitely subject to these challenges (Erhart et al. (2018)).

Table 1: MIP empirical literature review : the best indicators to signal crisis events

Studies (in chronological order)	**Target variable**	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities
Csortos-Szalai (2013, 2014)	Downturn (cycl. GDP gap)	■									■	
Boysen-Hogrefe et al. (2015)	Crisis (UNION of syst., banking, sov. debt, currency)						■	■	■			
Domonkos et al.(2017)	Downturn (cycl. GDP gap)	■							■			
Knedlik (2014)	Debt crisis	■	■			■						

The rest of the paper is structured as follows. Section 2 describes our two data-sets (i) the MIP Scoreboard and (ii) the ESRB's Financial Crisis database. Section 3 presents the methodology. Section 4 discusses the results on the signaling power of the MIP Scoreboard indicators in terms of different phases of the recent financial crisis life-cycle. The conclusions are presented in Section 5.

2 Data

The analysis in this paper was based on two datasets: (1) the annual MIP Scoreboard dataset of the European Commission and (2) the ESRB financial crisis database.

2.1 The MIP Scoreboard Indicators

The MIP Scoreboard is a set of headline indicators aiming at the assessment of macroeconomic risks (Commission (2012)). It consists of 14 indicators currently. Initially, the MIP Scoreboard was built upon 11 indicators grouped into (I) five *'external imbalance indicators'*, and (II) six *'internal imbalance indicators'*, which were augmented by additional 3 employment indicators Commission (2015). However, these new headline indicators do not play a direct role in the identification of macro-financial risks and do not trigger by themselves steps in the MIP, hence this report does not cover them.

Headline MIP Indicators and their indicative upper and/or lower thresholds in parenthesis:

I. EXTERNAL IMBALANCE INDICATORS
- **Current account balance**: in % of GDP, 3 years average (upper: +6% and lower - 4%)
- **Net international investment position**: in % of GDP (-35%)
- **Real effective exchange rate**: 42 trade partners, 3 years % change (-/+5)% for euro-area countries and (- /+) 11% for non-euro-area countries)
- **Export market shares**: 5 years % change (- 6%)
- **Nominal unit labour cost index**: 3 years % change (+9% for euro-area countries and +12% for non-euro-area countries)

II. INTERNAL IMBALANCE INDICATORS
- **House price index, deflated**: 1 year % change (6%)
- **Private sector credit flow**: consolidated, in % of GDP (15%)
- **Private sector debt**: consolidated, in % of GDP (133%)
- **General government sector debt**: in % of GDP (60%)
- **Unemployment rate**: 3 years average (10%)
- **Total financial sector liabilities**: non-consolidated - 1 year % change (16.5%)

In the MIP dataset, there are 336 observations for each indicators for the period between 2005 and 2016 (except for real house prices for which we had 333 observations) summing to 3693 observations (11 Indicators X 28 Member States X 12 years, 3 missing values). The great advantage of the dataset is that it combines observations from the pre- and post-crisis (boom and bust) periods. Hence, the database and the statistical analysis is less subject to possible biases stemming from economic cyclicality.

2.2 The ESRB Financial Crisis database

The ESRB financial crisis database provides the latest chronological summary of crisis periods to support the calibration of models in macroprudential analysis. Lo Duca et al. (2017) discusses how the ESRB database identifies financial crises by combining a quantitative approach based on a financial stress index with expert judgement from national and European authorities. The crises database is updated on a regular basis. Alessi and Detken (2018) showed that there is a large overlap between the widely used two datasets (Laeven and Valencia (2018) and ESRB) and the results across the two are robust).

CRISIS TYPES IN THE ESRB DATABASE

1. **Currency/Balance-of-Payment/Capital flow**
2. **Sovereign**
3. **Banking**
4. **Significant asset price correction**

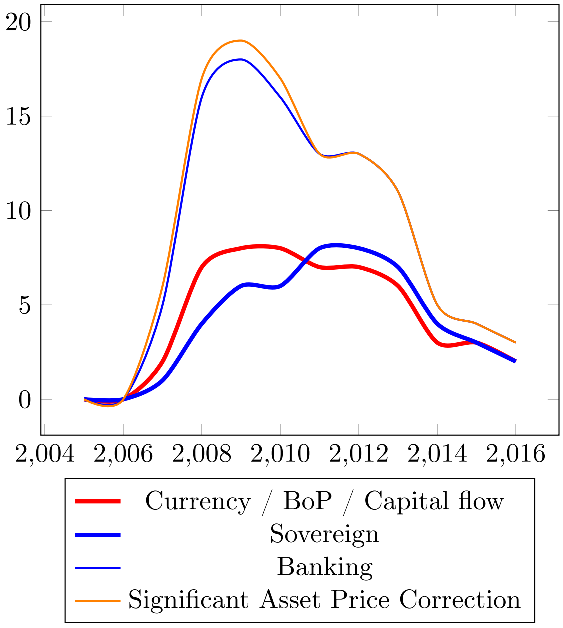
The ESRB financial crisis database is a set of binary variables for all crisis types (1 if crisis, and 0 otherwise) for the 28 EU Member States, in the period between 2005 and 2016 summing to 1344 observations (4 crisis types X 28 Member States X 12 years).

3 Methodology

3.1 Signals Approach

We followed the best practice of numerous empirical studies on the MIP and used the so-called *'Signals Approach'*, (Boysen et al. (2015), Knedlik (2014), Domonkos et al. (2017), Csortos and Szalai (2014)). The idea of the signals approach is that an indicator or a system of indicators can forecast a crisis, whenever it exceeds or falls below a certain threshold in a given forecast horizon (see Boysen et al. (2015)). Such identification of the best performing early warning system (EWS) indicators could also be advantageous, because it could help design an optimal scoreboard or com-

Figure 2: Number of EU Member State crisis events by ESRB crisis types



posite measure.

In the framework of the signals approach an indicator signals a crisis, if it exceeds or falls below a certain threshold. Overall, there are four cases (**A, B, C, D**) depending on whether a crisis signal or the absence of a crisis signal was correct or incorrect.

A the number years, in which the indicator **CORRECTLY** provides signal.

B the number of years, in which the indicator **INCORRECTLY** provides signal.

C the number of years, in which the indicator **INCORRECTLY** provides NO signal.

D the number of years, in which the indicator **CORRECTLY** provides NO signal.

Table 2: Probability space of the Signals Approach

		Crisis	
		YES	NO
Signal	YES	**A**	**B**
	NO	**C**	**D**

Signal performance evaluation

The less binding the thresholds, the more crises are correctly predicted (A) but at the same time the number of incorrect signals (B) and, thus, the number of *'false alarms'*, eg. Type II errors increases. In the same vein, a stricter threshold increases the number of cases, in which one correctly predicts that no crisis occurs (D) but it also increases the number of crises without any preceding signal (C) and, thus the number of Type I errors. The Commission aimed at striking a healthy balance when setting MIP thresholds at prudent levels, in order to avoid both excessive numbers of *'false alarms'*, and on the other hand also unacceptable delays of signals. (Commission (2012)).

$$\textbf{Correct crisis signals} = \frac{A}{(A+C)},$$

$$\textbf{Type I error} \;\; = \frac{C}{(A+C)},$$

the share of crisis events, when the early warning system failed to give signal,

$$\textbf{Type II error} \;\; = \frac{B}{(B+D)},$$

the share of events, when the early warning system incorrectly signalled crisis,

$$\textbf{Noise-to-Signal Ratio} = \frac{B/(B+D)}{A/(A+C))},$$

the share of incorrectly signalled crises events and correct crisis signals.

Empirical evaluations of early warning indicators are frequently based on a **usefulness function** *'U'* proposed by Alessi and Detken (2011) that is based on a loss function L. In the loss function certain weights θ and (1-θ) are assigned to Type I errors (C/(A + C)) and Type II errors (B/(B + D)). An early warning indicator is the more useful, **the lower the NSR** and **the higher the value of U** is and is generally considered useful if U > 0.

$$\textbf{Usefulness} = min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}.$$

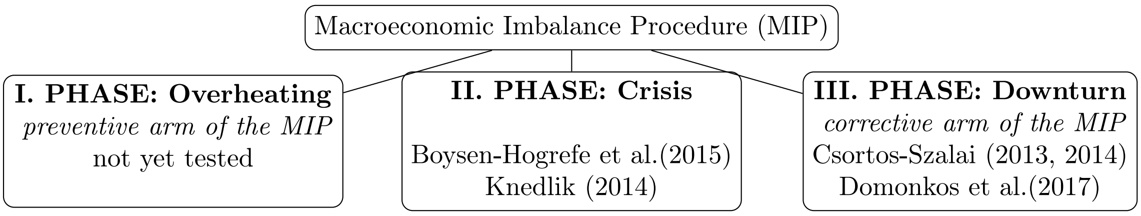
3.2 Differentiation of crisis events and of life-cycle phases

In this paper, the predictive power of the MIP Indicators are tested for all three phases of the crisis life-cycle (Figure 3). The 1st PHASE is called *'Overheating'*, which is often observed during the run-up to crisis events. The 2nd PHASE is called *'Crisis'* and the 3rd PHASE is the *'Downturn'* in the aftermath of the crisis. Previous empirical studies on the MIP Scoreboard investigated only whether its indicators could have forecast the 2nd and 3rd PHASE of crisis life-cycles, e.g. the crisis events and downturns in EU Member States. This paper also tests whether MIP indicators could help to signal the 1st PHASE of crisis events, the overheating of the economy (preventive arm of the MIP).

The definition of crisis events vary across studies. Boysen et al. (2015) used a dataset of three types of crisis (banking crises, currency crises, and debt crises) for the EU-28 countries and merged these sets of dummies to construct a new dummy variable (labeled as *'financial crisis dummy'*) equaling one if at least one of these three types of crises occurs. Knedlik (2014) defined its crisis variable based on long-term government bond spreads.

We followed the practice of most empirical studies and considered a forecast horizon between 1 to 3 years (Borio et al. (2009), Knedlik (2014). Downturn and overheating in EU Member States

Figure 3: Possible and tested channels of the *'Signals Approach'*



are defined on the basis of the cyclical GDP gap similarly to earlier studies Domonkos et al. (2017); Csortos and Szalai (2014) as follows. The real GDP of each country has been filtered by the HP (Hodrick–Prescott)-filter. To calculate the GDP gap, the difference between the basic time series and trend has been taken. We determine the threshold of the critical difference at minus 2 per cent for downturns, as proposed by the studies referred in this report. Macroeconomic overheating situations are identified the same way, if the gap exceeded plus 2 percent.

3.3 Composite Indicators

We combine the signals approach of empirical studies and the composite indicator methodology to provide a new method for evaluating the performance of the MIP Scoreboard. Such method has been proposed by the empirical studies of Babecky et al. (2014), Erhart et al. (2018), Borio et al. (2009), Christensen and Li (2014) to capture the fragility of the economy prior to a financial stress event most efficiently. Lo Duca et al. (2017) results on the newly established financial crisis database also confirm that multivariate methods can improve upon univariate signaling models. Babecky et al. (2014) note that a combination of several early warning indicators delivers a better-performing early warning model compared to a single early warning predictor.

We defined a composite measure as the sum of the red flags (or signals) based on official MIP indicator thresholds. The number of red flags has already been incorporated in the Commission analysis, without being the only or main determinant of the Alert Mechanism Report outcomes Commission (2016). The threshold for the total number of red flags has never been communicated yet by the European Commission. We set three as a starting value for the threshold of the composite measure for all countries in each year. (Reminder: on average four indicators were flagged in member states during the run-up and aftermath of the 2008-2009 crisis.)

4 Results

Our key result is that the usefulness of MIP Scoreboard Indicators' signals differ markedly for the different kind of crisis events. The MIP Scoreboard Indicators gave relatively good signals for the (i) Currency/BoP/Capital flow crises and (ii) Sovereign crisis events. For these two crisis types, usefulness of best MIP indicators is in the range between 0.2 and 0.3. However, in case of (iii) Banking crises and (iv) Significant asset price corrections the usefulness of MIP indicators is although mostly positive, but much smaller, around 0.1.

The current account was found as best forecasting indicator for all crisis events (its usefulness ranged between 0.1 and 0.3). Furthermore, the net international investment position was also a relatively good indicator for Currency/BoP/Capital flow and Sovereign crisis events (U = 0.2) compared to other MIP indicators. Frankel and Saravelos (2011) also found on the basis of a literature review of 83 papers that the current account balance is one of the most frequent statistically significant indicators in explaining crisis incidence.

Earlier studies aimed at the identification of the events in the second half of the crisis life-cycle, e.g. crisis events and ex-post downturns. Our analysis shows that some MIP indicators can signal the first phase of the crisis life-cycle, e.g. the economic overheating in EU Member States. In particular the real house price and total financial sector liabilities performed well in this setting. Real house prices were not flagged by earlier studies amongst best indicators, because these studies focused on the ex-post crisis manifestation indicators. The current threshold of 6 percent annual change can help to identify the build-up of imbalance, but not the corrections (see further discussion on the advantages of two-sided MIP thresholds in the paper of Erhart et al. (2018).)

The usefulness of MIP indicators in general, is higher for signalling overheating of the economy than that of downturns. This finding suggests that the preventive arm of the MIP could be stronger than its corrective arm, which is perhaps a preferred policy choice as *'prevention is better than cure'*.

Importantly, the simple composite of threshold breaches performed in most cases as good as the best MIP Scoreboard indicator in the baseline scenario (last columns in Table 4). Taking into account that decisions cannot be based on sole indicators, using a composite MIP measure could be an option to be considered by decision makers, who prefer simple, rule based decisions.

Table 3: Summary of the signals approach usefulness metric for individual indicators and the composite (forecast horizon = 3 Y, threshold for the number of red flags= 4, rounded to one decimal place)

Target variable	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)	Best MIP indicator
ESRB (Currency, BoP, Capital flow)	0.3	0.2	0.0	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.3	0.3
ESRB (Sovereign)	0.2	0.2	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.2	0.2
ESRB (Banking)	0.1	0.0	-0.1	0.0	0.0	0.0	0.1	0.1	0.1	-0.1	0.0	0.1	0.1
ESRB (Significant asset price correction)	0.1	0.0	-0.1	0.0	0.0	0.1	0.1	0.0	0.1	-0.1	0.1	0.1	0.1
Downturn (more severe than - 2pp.)	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1
Overheating (min + 2pp.)	0.1	0.0	0.1	-0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.2	0.1	0.2

Our results are in line with the conclusions of the majority of earlier empirical studies on the MIP. They confirm that the definition of the crisis could be possible explanation for the differences in the empirical results. Csortos and Szalai (2014) and Domonkos et al. (2017) used the cyclical GDP gap, Boysen et al. (2015) analysed the union of financial crisis events and Knedlik (2014) debt crisis events (Table 1.)

Knedlik (2014) investigated debt crisis events and found similarly to us that the usefulness is the highest for the current account, net international investment position and nominal unit labour costs (Figure 4).

Earlier empirical studies focused on specific crisis types, when measuring the predicting power of the MIP Scoreboard indicators and there have not been any attempt to analyse the importance of crisis types systemically to the best of our knowledge. Boysen et al. (2015) focused on the union of financial crisis events and found that real house prices, the private sector credit flow and the private sector debt are the best indicators (Figure 5). Our results are similar with the exception of private sector debt, and that we found the current account a relatively good indicator.

If the MIP was used to signal any kind of crisis events the usefulness of indicators would be in most cases close to or below 0.1. If the MIP Scoreboard was used to signal (i) Currency/BoP/-Capital flow crises and (ii) Sovereign crisis events the signalling power would improve and reach the level of 0.2-0.3.

Figure 4: Usefulness of MIP Indicators to signal SOVEREIGN crisis events - Comparing our results to Knedlik (2014)

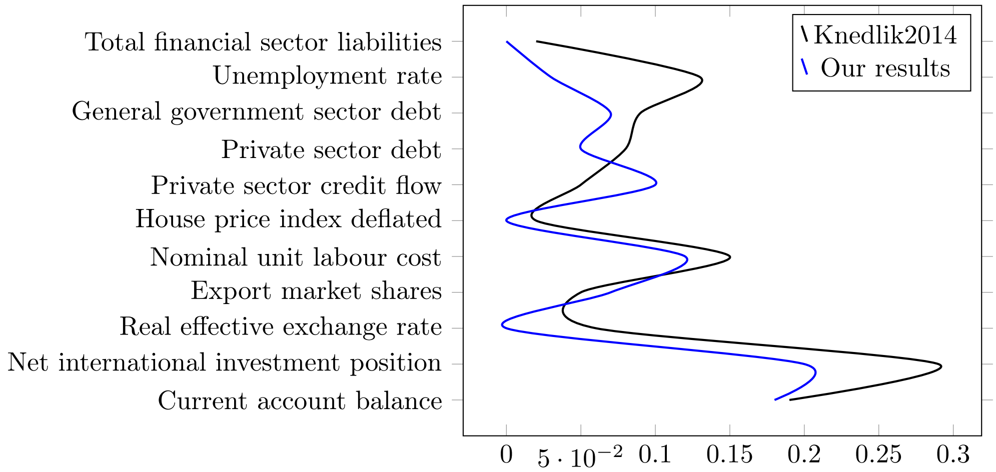
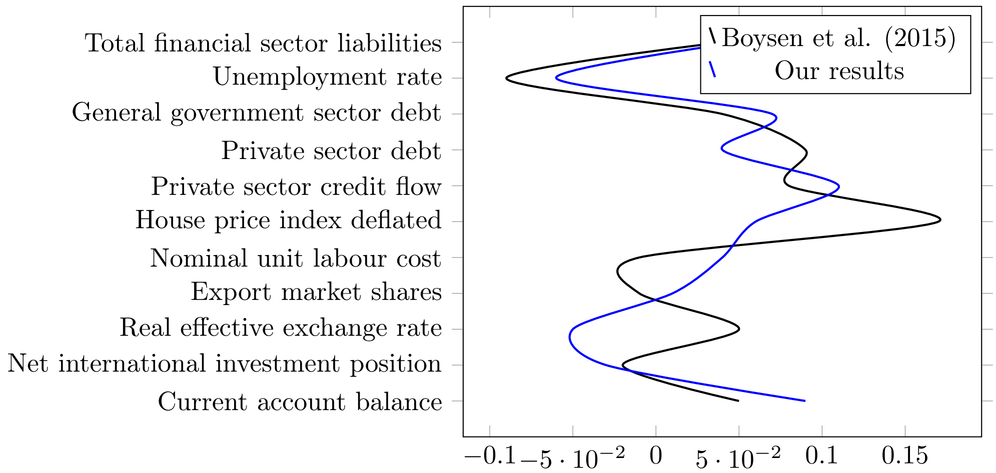


Figure 5: Usefulness of MIP Indicators to signal UNION OF CRISIS events - Comparing our results to Boysen et al. (2015)



Csortos and Szalai (2014) also showed that the current account is the best indicator to forecast downturns (it was the only MIP indicator, for which prediction ratios were better than the ratios of false alarms to alarms total). According to our results, the current account, net international investment position and nominal unit labour cost performed equally well as signalling indicators (usefulness = 0.1).

Our key result that the MIP can presumably better capture macroeconomic problems than financial imbalances confirms the findings of Detken et al. (2018) that simple credit and asset price indicators have better early warning properties for domestic financial crises in euro area countries than macroeconomic indicators, such as the current account balance.

Most of our results are robust to the length of the forecast horizon (Table 4). The MIP could help to recognize easier (i) Currency/BoP/Capital flow crises and (ii) Sovereign crisis events. However, the shorter the forecast horizon the better the predictive power of some indicators to identify the overheating of the economy (current account, private sector credit flow). Obviously, due to the life-cycle and order of crisis events, macroeconomic imbalances start usually as an economic overheating, followed by crisis and later by downturn (See Figure 1). The export market share was a relatively good indicator of (i) Currency/BoP/Capital flow crises and (ii) sovereign crisis events if the forecast horizon was set at 1 year.

5 Conclusions

Our key finding is that each unhappy crisis is 'unhappy in its own way'. Although it is very difficult to detect economic imbalances in advance, the MIP Scoreboard indicators could effectively signal (i) BoP/Currency/Capital flow and (ii) Sovereign type of crisis events in our sample period between 2005 and 2016. They were less effective, however, as signalling means in case of (iii) Banking crises and (iv) Significant asset price corrections.

We give further evidence that the current account is perhaps the most important indicator in the MIP Scoreboard. Similarly to earlier studies, we find it as the best forecasting indicator for all crisis events (it's usefulness ranged between 0.1 and 0.3). Furthermore, the net international investment position is also a relatively good indicator for Currency/BoP/Capital flow and Sovereign crisis events.

Previous empirical works tested the predictive power of indicators on the basis of events in the second half of the crisis life-cycle, e.g. crisis events and ex-post downturns. Our analysis extended the focus to the run-up to crisis events. We showed that some MIP indicators can signal the first phase of the crisis life-cycle, e.g. the overheating in EU Member States. In particular, the real house prices, private sector credit flow and total financial sector liabilities. Earlier papers could not have identified real house prices as good indicators, probably because these papers focused on the ex-post crisis manifestation indicators. The current threshold of 6 percent annual change in real house prices can help to identify the build-up of imbalances, but not the corrections (see further discussion on the advantages of two-sided MIP thresholds in the paper of Erhart et al. (2018).) Apparently, the relatively large Type I errors of real house prices (70-80% in all specifications except for overheating) confirm the challenge to identify macroeconomic imbalances.

We found that the *'usefulness'* metric of MIP indicators is usually higher for signalling overheating of the economy than that of downturns. This finding suggests that the preventive arm of the MIP could be stronger than its corrective arm in line with the second principle of the MIP Scoreboard *'the scoreboard (indicators and thresholds) are chosen as to provide a reliable signalling device for potentially harmful imbalances and competitiveness loss at an early stage of their emergence'* (Commission (2012)).

Due to the heterogeneity of crises, standardisation of their detection is obviously difficult. Our simple composite of MIP indicator threshold breaches performed in most cases as good as the best MIP Scoreboard indicator. Therefore, further aggregation of scoreboard indicators could be considered so as to gain a more simple, rule based and perhaps more accountable decision making in the European Semester.

References

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

Alessi, L. and Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability*, Vol. 35.

Babecky, A., Havranek, T., Mateju, J., Rusnak, M., Smidkova, K., and Vasicek, B. (2014). Banking, debt, and currency crises in developed countries. *Journal of Financial Stability*, Vol. 15.

Borio, C., Koban, A., and Drehmann, M. (2009). Assessing the risk of banking crises – revisited. *BIS Quarterly Review.*

Boysen, A., Jansen, N., Poedt, M., and Schwarzmueller, T. (2015). An empirical evaluation of macroeconomic surveillance in the European Union. *Kiel Working Papers, Kiel Institute for the World Economy.*

Christensen, I. and Li, F. (2014). Predicting financial stress events: a signal extraction approach. *Journal of Financial Stability*, Vol. 14:54–65.

Commission, E. (2012). Scoreboard for the surveillance of macroeconomic imbalances. *Occasional Papers 92*, doi: 10.2765/20196(92).

Commission, E. (2015). Adding employment indicators to the scoreboard of the macroeconomic imbalance procedure to better capture employment and social developments, Ares (2015) 5426195. *na.*

Commission, E. (2016). The macroeconomic imbalance procedure, rationale, process, application: A compendium. *Institutional Paper*, doi:10.2765/958247(39).

Csortos, O. and Szalai, Z. (2014). Early warning indicators: financial and macroeconomic imbalances in central and eastern european countries. *MNB Working Papers*, Vol. 1(2/2014).

Detken, Fahr, S., and Lang, J. H. (2018). Predicting the likelihood and severity of financial crises over the medium term with a cyclical systemic risk indicator. *Financial Stability Review.*

Domonkos, T., Ostrihon, F., Sikulova, I., and Siranova, M. (2017). Analysing the relevance of the MIP scoreboard's indicators. *National Institute Economic Review*, Vol. 239(No 238).

Erhart, S., Becker, W., and Saisana, M. (2018). The Macroeconomic Imbalance Procedure - From Thresholds to Decisions. *JRC Technical Reports.*

Frankel, J. and Saravelos, G. (2011). Can Leading Indicators Assess Country Vulnerability? Evidence from the 2008-09 Global Financial Crisis. *Faculty Research Working Paper Series.*

Knedlik, T. (2014). The impact of preferences on early warning systems - The Case of the European Commission's Scoreboard. *European Journal of Political Economy*, Vol. 34:157–166.

Laeven, L. and Valencia, F. (2018). Systemic Banking Crises Revisited? *IMF Working Paper No. 18/206.*

Lo Duca, M., Koban, A., Barsten, M., Bengtsson, E., Klaus, B., Kusmierczyk, P., Lang, J. H., Detken, C., and Peltonen, T. (2017). A new database for financial crises in European countries. *European Central Bank Occasional Paper Series No 194* .

A Appendix - Signal approach results for different forecast horizons

Table 4: Summary of the signals approach usefulness metric for individual indicators and the composite (forecast horizon = 2 Y, threshold for the number of red flags= 4, rounded to one decimal place)

Target variable	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)	Best MIP indicator
ESRB (Currency, BoP, Capital flow)	0.2	0.2	0.0	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.3	0.2
ESRB (Sovereign)	0.2	0.2	0.0	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.2	0.2
ESRB (Banking)	0.1	0.0	-0.1	0.1	0.0	0.0	0.1	0.0	0.1	-0.1	0.0	0.1	0.1
ESRB (Significant asset price correction)	0.1	0.0	-0.1	0.1	0.0	0.0	0.1	0.0	0.1	-0.1	0.0	0.1	0.1
Downturn (more severe than - 2pp.)	0.0	0.1	0.0	0.0	0.1	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.1
Overheating (min + 2pp.)	0.1	0.0	0.0	-0.1	0.0	0.2	0.1	0.0	0.0	0.0	0.2	0.1	0.2

Table 5: Summary of the signals approach usefulness metric for individual indicators and the composite (forecast horizon = 1 Y, threshold for the number of red flags= 4, rounded to one decimal place)

Target variable	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)	Best MIP indicator
ESRB (Currency, BoP, Capital flow)	0.2	0.2	0.1	0.1	0.1	-0.1	0.1	0.0	0.1	0.1	0.0	0.3	0.2
ESRB (Sovereign)	0.1	0.2	0.1	0.2	0.0	-0.1	0.0	0.0	0.1	0.1	-0.1	0.2	0.2
ESRB (Banking)	0.1	0.0	0.0	0.2	0.0	-0.1	0.0	0.0	0.1	0.0	-0.1	0.1	0.2
ESRB (Significant asset price correction)	0.1	0.0	0.0	0.1	0.0	-0.1	0.1	0.0	0.1	0.0	-0.1	0.1	0.1
Downturn (more severe than - 2pp.)	0.0	0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.1	-0.1	0.0	0.1
Overheating (min + 2pp.)	0.2	0.0	0.0	-0.1	0.0	0.2	0.2	0.0	-0.1	-0.1	0.2	0.1	0.2

Table 6: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **CURRENCY/BOP/Capital flow**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness = $\min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	46	45	12	27	28	14	22	25	29	16	18	47
B	75	89	41	87	51	42	30	84	77	49	33	78
C	5	6	39	24	23	37	29	26	22	35	33	4
D	126	112	160	114	150	159	171	117	124	152	168	123
Correct forecasts	68%	62%	68%	56%	71%	69%	77%	56%	61%	67%	74%	67%
% of correct crisis signals	90%	88%	24%	53%	55%	27%	43%	49%	57%	31%	35%	92%
type I error	10%	12%	76%	47%	45%	73%	57%	51%	43%	69%	65%	8%
type II error	37%	44%	20%	43%	25%	21%	15%	42%	38%	24%	16%	39%
NSR	0.41	0.50	0.87	0.82	0.46	0.76	0.35	0.85	0.67	0.78	0.47	0.42
Usefulness	0.26	0.22	0.02	0.05	0.15	0.03	0.14	0.04	0.09	0.03	0.09	0.27
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Table 7: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **SOVEREIGN**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness $= \min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	37	41	10	27	24	11	18	25	26	15	10	38
B	84	93	43	87	55	45	34	84	80	50	41	87
C	11	7	38	21	24	37	30	23	22	33	38	10
D	120	111	161	117	149	159	170	120	124	154	163	117
Correct forecasts	62%	60%	68%	57%	69%	67%	75%	58%	60%	67%	69%	62%
% of correct crisis signals	77%	85%	21%	56%	50%	23%	38%	52%	54%	31%	21%	79%
type I error	23%	15%	79%	44%	50%	77%	63%	48%	46%	69%	79%	21%
type II error	41%	46%	21%	43%	27%	22%	17%	41%	39%	25%	20%	43%
NSR	0.53	0.53	1.01	0.76	0.54	0.96	0.44	0.79	0.72	0.78	0.96	0.54
Usefulness	0.18	0.20	0.00	0.07	0.12	0.00	0.10	0.05	0.07	0.03	0.00	0.18
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Table 8: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **BANKING**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness = $\min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	57	49	12	47	34	28	33	49	51	19	24	57
B	64	85	41	67	45	28	19	60	55	46	27	68
C	42	50	87	52	65	71	66	50	48	80	75	42
D	89	68	112	86	108	125	134	93	98	107	126	85
Correct forecasts	58%	46%	49%	53%	56%	61%	66%	56%	59%	50%	60%	56%
% of correct crisis signals	58%	49%	12%	47%	34%	28%	33%	49%	52%	19%	24%	58%
type I error	42%	51%	88%	53%	66%	72%	67%	51%	48%	81%	76%	42%
type II error	42%	56%	27%	44%	29%	18%	12%	39%	36%	30%	18%	44%
NSR	0.73	1.12	2.21	0.92	0.86	0.65	0.37	0.79	0.70	1.57	0.73	0.77
Usefulness	0.08	-0.03	-0.07	0.02	0.02	0.05	0.10	0.05	0.08	-0.05	0.03	0.07
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Table 9: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **SIGNIFICANT ASSET PRICE CORRECTION**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness = $\min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	60	51	15	47	37	30	35	49	51	19	27	60
B	61	83	38	67	42	26	17	60	55	46	24	65
C	42	51	87	55	65	72	67	53	51	83	75	42
D	89	67	112	83	108	124	133	90	95	104	126	85
Correct forecasts	59%	47%	50%	52%	58%	61%	67%	55%	58%	49%	61%	58%
% of correct crisis signals	59%	50%	15%	46%	36%	29%	34%	48%	50%	19%	26%	59%
type I error	41%	50%	85%	54%	64%	71%	66%	52%	50%	81%	74%	41%
type II error	41%	55%	25%	45%	28%	17%	11%	40%	37%	31%	16%	43%
NSR	0.69	1.11	1.72	0.97	0.77	0.59	0.33	0.83	0.73	1.65	0.60	0.74
Usefulness	0.09	-0.03	-0.05	0.01	0.04	0.06	0.11	0.04	0.07	-0.06	0.05	0.08
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Table 10: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **DOWNTURN (NEGATIVE GDP GAP (-2%))**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness = $\min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	57	62	23	43	46	17	18	46	36	22	15	57
B	64	72	30	71	33	39	34	63	70	43	36	68
C	40	35	74	54	51	80	79	51	61	75	82	40
D	91	83	125	84	122	116	121	92	85	112	119	87
Correct forecasts	59%	58%	59%	50%	67%	53%	55%	55%	48%	53%	53%	57%
% of correct crisis signals	59%	64%	24%	44%	47%	18%	19%	47%	37%	23%	15%	59%
type I error	41%	36%	76%	56%	53%	82%	81%	53%	63%	77%	85%	41%
type II error	41%	46%	19%	46%	21%	25%	22%	41%	45%	28%	23%	44%
NSR	0.70	0.73	0.82	1.03	0.45	1.44	1.18	0.86	1.22	1.22	1.50	0.75
Usefulness	0.09	0.09	0.02	-0.01	0.13	-0.04	-0.02	0.03	-0.04	-0.03	-0.04	0.07
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Table 11: Detailed statistics of the signals approach metrics for individual indicators and the composite (crisis type: **OVERHEATING (POSITIVE GDP GAP (+2%))**, forecast horizon = 3 Y, threshold for the number of red flags= 4)

A - Correct signal (crisis signalled)
B - Wrong signal (NO crisis , signal)
C - Wrong signal (crisis, NO signal)
D - Correct signal (NO crisis, NO signal)
Correct forecasts (A+D)/(A+B+C+D)
Share of correct crisis signals A/(A+C)
type I error C/(A+C)
type II error B/(B+D)
NSR - Noise to signal ratio B/(B+D)/A/(A+C)
Usefulness = $\min(\theta; 1-\theta) - L = min(\theta; 1-\theta) - \theta \cdot \frac{C}{(A+C)} - (1-\theta) \cdot \frac{B}{(B+D)}$

	Current account balance	Net international investment position	Real effective exchange rate	Export market shares	Nominal unit labour cost	House price index, deflated	Private sector credit flow	Private sector debt	General government sector debt	Unemployment rate	Total financial sector liabilities	COMPOSITE (No of red flags >3)
A	25	22	13	9	13	20	15	17	17	13	25	29
B	96	112	40	105	66	36	37	92	89	52	26	96
C	16	19	28	32	28	21	26	24	24	28	16	12
D	115	99	171	106	145	175	174	119	122	159	185	115
Correct forecasts	56%	48%	73%	46%	63%	77%	75%	54%	55%	68%	83%	57%
% of correct crisis signals	61%	54%	32%	22%	32%	49%	37%	41%	41%	32%	61%	71%
type I error	39%	46%	68%	78%	68%	51%	63%	59%	59%	68%	39%	29%
type II error	45%	53%	19%	50%	31%	17%	18%	44%	42%	25%	12%	45%
NSR	0.75	0.99	0.60	2.27	0.99	0.35	0.48	1.05	1.02	0.78	0.20	0.64
Usefulness	0.08	0.00	0.06	-0.14	0.00	0.16	0.10	-0.01	0.00	0.04	0.24	0.13
Total number of events	252	252	252	252	252	252	252	252	252	252	252	252

Imprint and acknowledgements

The author gratefully acknowledges the help, comments and suggestions received from Zsuzsa Turóczy (ex Trafo), Anna Naszódi (Joint Research Centre, European Comission), Michaela Saisana (Joint Research Centre, European Comission), Daniel Vértesy (Joint Research Centre, European Comission), Lucia Alessi (Joint Research Centre, European Comission), Sven Langedijk (Joint Research Centre, European Comission), Claudia Ghisetti (Joint Research Centre, European Comission), from Jean-Charles Bricongne (DG ECFIN, European Commission), from Carsten Detken (European Systemic Risk Board), Tuomas Peltonen (European Systemic Risk Board). The views expressed in this paper are those of the author and do not necessarily reflect the official views of the European Commission.

Szilard Erhart
European Commission, Brussels, Belgium; email: **szilard.erhart@ec.europa.eu**

Postal address	60640 Frankfurt am Main, Germany
Telephone	+49 69 1344 0
Website	**www.esrb.europa.eu**

Note:
The views expressed in ESRB Working Papers are those of the authors and do not necessarily reflect the official stance of the ESRB, its member institutions, or the institutions to which the authors are affiliated.

ISSN	2467-0677 (pdf)
ISBN	978-92-9472-116-7 (pdf)
DOI	10.2849/681425 (pdf)
EU catalogue No	DT-AD-19-016-EN-N (pdf)