ZSUZSANNA HOSSZÚ

GERGELY LAKOS

EARLY WARNING PERFORMANCE OF UNIVARIATE CREDIT-TO-GDP GAPS

MNB OCCASIONAL PAPERS | 142

2021

SEPTEMBER

EARLY WARNING PERFORMANCE OF UNIVARIATE CREDIT-TO-GDP GAPS

MNB OCCASIONAL PAPERS | 142

2021

SEPTEMBER

The views expressed are those of the authors and do not necessarily reflect the official view of the central bank of Hungary (Magyar Nemzeti Bank).

MNB Occasional Papers 142

Early warning performance of univariate credit-to-GDP gaps

(Egyváltozós hitel/GDP-rések korai előrejelző képessége)

Written by Zsuzsanna Hosszú, Gergely Lakos

Budapest, September 2021

Published by the Magyar Nemzeti Bank

Publisher in charge: Eszter Hergár

H-1013 Budapest, Krisztina körút 55.

www.mnb.hu

ISSN 1585-5678 (online)

Contents

Abstract

We use European and simulated Hungarian data to search for the univariate one-sided credit-to-GDP gap that predicts systemic banking crises most accurately. The credit-to-GDP gaps under review are optimized along four dimensions: (1) definition of outstanding credit, (2) forecasting method for extending credit-to-GDP time series, (3) filtering method and (4) maximum cycle length. Based on European data, we demonstrate that credit-to-GDP gaps calculated with narrow definition of outstanding credit and up to 1-year forecasts of credit-to-GDP outperform other specifications significantly and robustly. Regarding the other two dimensions, the Hodrick–Prescott filter with long cycles (popular in regulatory practice), the Christiano–Fitzgerald filter with medium-term cycles and the wavelet filter with short cycles prove to be the best. All three should be applied to credit-to-GDP time series calculated with narrow credit, and with no credit-to-GDP forecast, except the wavelet filter with short-term forecast. Credit-to-GDP gaps with most informative early warning signals exhibit the highest degree of comovement with the financial cycle, but not the lowest level of endpoint uncertainty. Analysis of Hungarian credit-to-GDP time series extended by ARIMA simulations reinforces the early warning quality of the Hodrick–Prescott credit gap and the wavelet credit gap to a lesser extent.

JEL-codes: C20, C52, E32, G28

Keywords: financial cycle, crises, early warning, univariate filtering methods

1 Introduction

In order to maintain an effective countercyclical capital buffer (CCyB) requirement, it is crucial to foresee financial stress early, several years in advance. Using the additional capital buffer built up in good times makes the banking system more resilient during financial stress, so banks need to resort less to curbing their lending activity to stabilize their capital position. The timing and size of the CCyB requirement can only be accurate if the extent of the risk of systemic banking crises expected within a few years is known exactly.

A widely used early warning indicator for systemic banking crises is the so-called credit-to-GDP gap[1], which is the deviation of the private non-financial sector credit-to-GDP ratio from its long-run trend.[2] By definition, credit gaps capture the part of the credit cycle that is independent from the business cycle, they characterize the earnings potential of outstanding loans, and they can also be regarded as macroeconomic indicators of loan repayment capacity. All interpretations attest that positive credit gaps suggest excessive lending, which increases the likelihood and enhances the severity of systemic banking crises. In line with the guidance of the Basel Committee on Banking Supervision (BCBS, 2010), the ESRB 2014/1 methodological recommendation assigns a central role to credit-to-GDP gaps in the build-up of the countercyclical capital buffer. Based on the recommendation, the macroprudential authorities of all ESRB member states calculate the so-called Basel gap, or standardized credit-to-GDP gap, and the corresponding benchmark buffer rate. Optionally, the additional credit-to-GDP gap can also be determined, the methodology of which can be tailored to the features of the country concerned. Moreover, the recommendation proposes other indicators in six risk categories to be taken into consideration during the build-up phase of the CCyB.

Our study aims to identify the univariate credit-to-GDP gap with the best early warning signals. In empirical studies and regulatory practice, credit gaps calculated with various methodologies are examined and used, without any consensus on which of these provides the best early warning signals. The paper seeks to contribute to a more accurate ranking of credit gaps with a widescale and in-depth assessment of them, which is unprecedented to the best of our knowledge.

Based on empirical studies examining pooled data of several countries, the Basel gap is generally not the most accurate predictor of financial crises, although it is still one of the best. Earlier papers argued that the Basel gap was the most reliable early warning indicator for systemic banking crises,[3] which is disputed by recent studies.[4] An overall conclusion is that the indicators outperforming the Basel gap are often some transformations of credit-to-GDP, sometimes even gap indicators.

Accordingly, the benchmark buffer rate based on the standardized credit gap is not followed closely in regulatory decisions on building up the CCyB. Instead, the mainly discretionary decisions take into account a much broader information base.[5] According to the survey of Arbatli-Saxegaard and Muneer (2020) and MNB (2020), more than half of the 30 ESRB member states have created an additional credit-to-GDP gap deviating from the standardized credit gap in the definition of the

[1] For the sake of simplicity, credit-to-GDP gaps are often referred to as credit gaps here.

[2] Gap indicators are usually considered the deviations of the credit-to-GDP gap from its estimated long-run trend, expressed in percentage points. Sometimes the ratio of this deviation and the long-run trend is referred to as a gap indicator, which is expressed as a percentage. The latter, used for example by Detken et al. (2014) and Tölö et al. (2018), is not discussed here.

[3] See: Drehmann et al. (2010), Drehmann et al. (2011), Drehmann and Juselius (2014), Bonfim and Monteiro (2013), Babecký et al. (2014), and Detken et al. (2014). These studies typically looked at two or three dozen, mostly advanced countries and four or five decades of quarterly time series to compare the Basel gap with several other indicators. In the case of Detken et al. (2014), this examination also covered numerous credit gaps with different specifications.

[4] According to Hamilton (2018), using a one-sided HP filter mostly yields unreliable gap values. This is because the filter is usually not used for time series where it could appropriately perform a trend-cycle decomposition. Drehmann and Yetman (2018) acknowledge this but defend the use of credit-to-GDP gaps calculated with an HP filter with the empirical argument that these gaps are good predictors of systemic banking crises. Hamilton and Leff (2020) raise technical objections to this argument. For further relevant studies, see Barrell et al. (2020), Tölö et al. (2018), Beltran et al. (2020), Kauko and Tölö (2020), and Lang et al. (2019). Tölö et al. (2018) offer the most comprehensive assessment, with around 50 early warning indicators with almost 400 specifications examined on data from 28 EU member states, starting in 1970.

[5] This is found in the most comprehensive and recent assessment by Arbatli-Saxegaard and Muneer (2020), the same conclusion is drawn by ESRB (2018) and Babić and Fahr (2019). For an earlier international review of the use of the CCyB, see Pekanov and Dierick (2016) as well as BCBS (2017).

credit-to-GDP ratio and in the method used for trend-cycle decomposition as well (Table 1). This variety demonstrates simultaneously that the ranking of credit gaps according to their early warning performance is ambiguous, and that national macroprudential authorities strive to find specifications best aligned with local circumstances.

Table 1
Calculation methods for the standardized and additional credit-to-GDP gaps in ESRB member countries, 2020 H1

	Stock of credit	GDP	Credit-to-GDP forecast	Trend-cycle decomposition method	Cycle length
Standardized	• Broad: Loans granted to non-financial corporations and households	• 4-quarter rolling GDP	• The credit-to-GDP forecast should not be taken into account when calculating the credit-to-GDP gap	• One-sided Hodrick–Prescott-filter	• 30 years (λ = 400 000)
Additional	• Narrower (typically loans granted by domestic credit institutions to non-financial corporations and households): **BE, CY, CZ, DE, EE, FR, HR, IE, IT, LU, LV, SK** • Narrower and exchange rate adjusted (foreign currency loans at a fixed exchange rate): **HU**	• 4-quarter rolling potential GDP: **SK** • Seasonally adjusted quarterly GDP: **HR** • 4-quarter rolling private sector gross added value: **CZ** • 4-quarter rolling GNI: **IE**	• Assumes the average of the last 4 quarters for the next 20 quarters: **NO** • Assumes a weighted average of the last 4 quarters for the next 20 quarters: **LT** • ARIMA(p,1, 0) forcast for the next 28 quarters: **PT**	• The cyclical component is the difference between the current value of the credit-to-GDP and the minimum value of the last 8 quarters: **CZ** • Approximate two-sided HP filtering: The current period gap calculated with one-sided HP filter is adjusted based on past differences in standardized one-sided and two-sided gaps: **IT**	• 8 years (λ = 1 600): **RO** • 10,5 years: **PL** • 15 years (λ = 25 000): **ES**

Note: ESRB member countries: EU member countries, Iceland, Liechtenstein, and Norway. Liechtenstein has substantial data constraints, so it is not included in the table.
Source: MNB (2020) Box 2.

Our paper seeks to rank as many relevant univariate and one-sided credit-to-GDP gaps as possible.[6] Based on the options in the literature and regulatory practice, we compare 48 credit gap specifications varying in four dimensions. We analyze credit-to-GDP time series with different definitions of the outstanding credit and with different forecasts that extend the credit-to-GDP time series. Trend-cycle decomposition is conducted by three filters (Hodrick–Prescott (HP), Christiano–Fitzgerald (CF) and wavelet) with three different cycle lengths (approx. 32 years, approx. 25 years, approx. 19 years).[7]

We use three methods to rank the 48 credit gaps. First, we assess the early warning performance of systemic banking crises on an unbalanced quarterly panel data from 18 European countries with a method introduced by Kaminsky and Reinhart (1999) for extracting early warning signals of economic crises, which is by now common in the literature. Most of the credit-to-GDP time series comes from the BIS credit database,[8] with the longest one starting in 1960 Q4. The crisis data are taken from the ESRB database (Lo Duca et al., 2017), which contains all the crises in the analyzed 18 countries in a monthly breakdown starting in 1970. Early warning performance is assessed primarily with AUROC values on various prediction horizons.[9]

[6] A univariate credit gap is derived solely from the credit-to-GDP time series, while a one-sided version only uses the credit-to-GDP values observed until the date of the gap value in question. (By contrast, a two-sided credit gap is calculated using the entire credit-to-GDP time series.) Multivariate credit gaps using explanatory variables exhibit usually better early warning properties, on account of their larger information base (see for example Castro et al., 2016; Galán and Mencía, 2018; Lang and Welz, 2018). Nevertheless, due to their simplicity, univariate credit gaps should not be left out from among the other individual indicators that are used for deciding on building up the CCyB. We narrow down the examination of credit gaps to one-sided versions because in practice macroprudential policy mostly needs the latest gap values, which can only be calculated in a one-sided manner since future credit-to-GDP values are not observable.
[7] These cycle lengths are often referred to as short, medium-term and long.
[8] We use credit-to-GDP data provided by national macroprudential authorities for only one country (Lithuania).
[9] For the definition of the AUROC value, see Appendix D.

We find that it is easier to answer the question 'What to filter?' than the question 'How to filter?'. First, the credit gaps calculated with narrow credit, where the lenders are only the domestic banks, usually fared better than the same credit gap specifications using broad credit. Second, the credit gaps calculated with credit-to-GDP time series that are extended by longer-term (12-quarter) 'perfect' forecasts (identical to the future credit-to-GDP values that will be realized later) are outperformed by the credit gaps using a shorter forecast horizon or none at all but otherwise having the same specifications. These results are robust to using RU (relative usefulness) performance measure based on the idea of Alessi and Detken (2011).[10] A breakdown of the sample into three country groups shows that these findings mostly hold true for Nordic countries, and less so for Mediterranean countries and core EU member countries. By dividing the sample into the last two decades containing the global financial crisis and the period before, the results are achieved in both subsamples in a weak sense because the narrow and broad credit gaps performed similarly in the first subsample. The 12- and 20-quarter ARIMA forecasts that are used instead of the 'perfect' credit-to-GDP forecasts are also unable to make long-term forecasts competitive. However, the out-of-sample evaluation based on the methodology in Lang et al. (2019) does not confirm the first two main findings. Early warning signals between 2000 and 2016 using experiences from earlier periods are often more accurate when credit gaps with broad credit provide them. Furthermore, none of the credit-to-GDP forecast alternatives dominate. Since the out-of-sample evaluation can be applied to a relatively short and special period, its relevance is limited.

No credit gap provides clearly better early warning signals than the so-called additional credit-to-GDP gap used by several European macroprudential authorities (narrow credit, no credit-to-GDP forecast, HP filter with a lambda of 400,000). Two other quite different specifications also belong to the best set of credit gaps with an AUROC value of around 0.75, which is relatively high based on the literature. Both of them are calculated with narrow credit. One of them uses no credit-to-GDP forecast and is derived by a CF filter with medium-term cycle, while the other uses a 1-year credit-to-GDP forecast and is produced by a wavelet filter with short-term cycle. The three best credit gaps generate different Type I and Type II errors, and their performance is varying over different prediction horizons, therefore in practice it might be worth applying all three together when assessing cyclical systemic risks. During the above-mentioned robustness checks, at least one of the three best credit gaps (typically different ones) retains its status with the only exception of the out-of-sample evaluation. During the robustness checks, no fourth credit gap provides consistently as informative early warning signals as the top three gaps.

Our second method quantifies how credit gap series can reflect the evolution of the financial cycle. The calibration and communication of the CCyB is facilitated if credit gaps not only provide an accurate binary signal of impending crises but are also able to capture the current position of the financial cycle as best as possible. Here, we follow two approaches. First, we estimate the average endpoint uncertainty of the credit gaps. It is assumed that regardless of the specification, the two-sided credit gaps would characterize the current position of the credit cycle better than their one-sided versions, therefore lower endpoint uncertainty can be seen as a beneficial feature. Among the top three early warning credit gaps, the CF filter credit gap comes close to the performance of the credit gaps revising the least over time, while the HP filter credit gap is revised to a moderate extent, and the wavelet filter credit gap is revised markedly. Second, we estimate the average correlation of credit gaps with aggregated indicators of the financial cycle created by national macroprudential authorities. In half of the 18 countries under review, a total of 11 indicators are found with a relatively long time series. All three credit gaps are among those with the highest correlation, with values around 0.7.

In our third analysis, we measure credit gaps' capacity to signal excessive credit growth in simulated future Hungarian data. The differences in earlier results across country groups may be attributable to inherent national economic specificities. Consequently, different credit gaps may be optimal for different countries.[11] Since we are primarily interested in finding the best credit gap to be applied in Hungary, this examination intends to consider mainly Hungarian economic specificities. The current Hungarian additional credit-to-GDP gap uses narrow definition of credit with exchange rate adjusted values. We extend this credit-to-GDP time series by forecast data, and further by simulated data until 2037 under four scenarios, with

---

[10] RU values are calculated assuming balanced preferences over Type I and Type II error rates of the early warning signals. (The value of the preference parameter is 0.5.) For the definition of the RU value, see Appendix D.

[11] This aspect is also noted in the international regulatory guidance and recommendations on the CCyB framework, e.g. BCBS (2010), Drehmann and Tsatsaronis (2014), ESRB (2014), Detken et al. (2014). Moreover, Buncic and Melecky (2014) also found that the structural estimates of equilibrium stock of credit depend significantly on country-specific factors.

200 simulated time series in each scenario. We incorporate periods with high cyclical systemic risk into the simulations, and with the same signaling method used before we evaluate how accurately credit gaps can indicate these periods. Out of the best three early warning credit gaps based on European data, the HP filter credit gap is also among the top credit gaps based on Hungarian data, under all scenarios. The wavelet filter credit gap makes it to the top only under the baseline scenario, which is, however, considered the most relevant. The CF filter credit gap fares only marginally worse.

Four aspects of our approach are novel compared to the existing literature. First, we are unaware of any other study that would have optimized the credit-to-GDP gap along the above four dimensions *together.* There is one that considered three characteristics simultaneously (Detken et al., 2014), albeit without a detailed analysis, and there are multiple papers on optimizing up to two attributes. The results confirm our approach, as any two of the top three credit gaps are markedly different along three dimensions. Second, previous examinations did not include a careful assessment of the early warning performance of the credit gaps that are calculated using ‘perfect’ or ARIMA model based credit-to-GDP forecasts. Third, the early warning performance of univariate credit gaps produced with a discrete wavelet filter has also not been examined in detail. Finally, we are not aware of any paper that would have assessed indicators’ early warning properties with respect to economic crises using simulated data.[12]

Our findings about the better performance of narrow credit are consistent with the results in recent studies. Drehmann (2013) provided a partial comparison of the credit-to-GDP gaps calculated with broad and narrow credit when BIS expanded its credit database with long-run series of broad credit to the private non-financial sector (see Dembiermont et al., 2013). The credit gaps calculated with broad credit predict a larger proportion of crises, with fewer false positives, than those calculated with narrow credit. Coudert and Idier (2018), aimed at developing an early warning system specifically for France, confirm this. The study uses time series starting in 1985 from 10 advanced EU member states overweighting French data. Tölö et al. (2018) and Lang et al. (2019) examine larger samples and draw the opposite conclusion. Detken et al. (2014), whose paper provides the analysis underpinning ESRB Recommendation 2014/1, also use many indicators and a larger dataset, and without offering a detailed description of the relevant results, give a summary that credit gaps calculated with narrow credit generally outperform those calculated with broad credit, albeit not significantly.

How much credit-to-GDP forecasts improve the predictive power of credit gaps has not been discussed in detail before. Our negative result corresponds to the findings in most of the existing studies. One of the main ways to reduce endpoint uncertainty is to conduct gap estimation on a time series extended with a forecast. In the case of the HP filter, a meticulous discussion can be found in Mise et al. (2005). However, lower endpoint uncertainty does not necessarily mean better early warning performance. This hypothesis has only been tested partially before. Drehmann et al. (2011) and Drehmann and Juselius (2014) find that the two-sided version of the Basel gap does not improve its predictive performance. Detken et al. (2014) examine linear and moving average forecasts for credit-to-GDP with different definitions of credit and varying cycle lengths. No detailed assessment is given, but it turns out that none of the specifications are significantly better than the Basel gap. Gerdrup et al. (2013), Valinskytė and Rupeika (2015), and Banco de Portugal (2015) use much smaller datasets and find weak arguments for that the Basel gap has a better early warning performance when it is calculated with credit-to-GDP forecasts. Martínez and Oda (2018) employ data from Chile in a period spanning from 1970 and containing three periods of financial stress, to examine credit gaps that are computed with narrow credit, HP and CF filters, different cycle lengths, using the full sample and a rolling window of 10 years. They mention the possibility of calculating a 3-year “perfect” forecast for credit-to-GDP (identical to the future credit-to-GDP values that will be realized later), but they do not adopt this in the end.

Some earlier studies have shown examples of univariate credit gaps calculated with methods other than the HP filter that have a competitive early warning performance. Even our thorough analysis presented here is unable to find a method that could clearly surpass the HP filter. Hamilton (2018) claims that the residual from linear projection outperforms the HP filter, however, this is disputed by Drehmann and Yetman (2018). The gap measure of Beltran et al. (2020) derived

[12] Boissay et al. (2016) use data simulated with a calibrated DSGE model to check whether certain variables predict systemic banking crises consistently with the model. Nevertheless, the number of variables under review is low, and the forecast horizon is short (one year). Hodrick (2020) checks on simulated time series consistent with the mean and standard deviation of the rate of growth of the US real GDP to what extent the gap indicators derived with the examined filtering methods match those in the simulated data series. Schüler (2018) illustrates the fact that HP filter credit gaps have a too large amplitude with an example in which an AR model fitted to the changes in the US credit-to-GDP ratio is used to simulate a credit-to-GDP time series and a credit gap time series.

with a Bayesian structural time series model is on a par with the HP filter. Barrell et al. (2020) make specific assumptions on the functional form of the cycle in the state space representation of the HP filter and find that using an AR(2) process or a stochastic cycle results in credit gaps that are better predictors of future crises than the Basel gap. Galán (2019) compares the crisis prediction power of credit gaps calculated with one-sided HP filters with different lambdas (different cycle lengths) on a Spanish credit-to-GDP time series starting in 1965 and containing three periods of financial stress. As a robustness check, the same is done with Butterworth and CF filters with various cycle lengths. None of these credit gaps perform better than the Basel gap. According to the above-mentioned Martínez and Oda (2018), the HP and CF filters produce credit gaps with similar prediction power on Chilean data.

There is no consensus in the literature about the cycle length that gives the best early warning credit gap. Our large-scale comparison resolves this by claiming that credit gaps with long, medium-term and short cycle length can all perform well, if combined properly with different other characteristics. Among the studies mentioned so far, Drehmann et al. (2010), Drehmann et al. (2011), Bonfim and Monteiro (2013), Detken et al. (2014), Valinskytė and Rupeika (2015) and Lang et al. (2019) support that a cycle length of about 30 years is the best. Geršl and Jašová (2018) also find this on the data from 36 emerging countries over a period starting in 1987, in the case of credit gaps computed with narrow credit and an HP filter. The following papers argue that the optimal cycle length is below 30 years. Galati et al. (2016) use the credit-to-GDP time series of the US and the five largest euro area members starting from 1970 and estimate cycle lengths of 8–25 years with a univariate unobserved components time series model. Kauko and Tölö (2020) examine long, annual time series of 15 advanced countries, going back until 1870 in the case of the longest one, finding that systemic banking crises are better predicted by a version of the Basel gap with a much shorter cycle length. The previously mentioned Galán (2019) analyzes Spanish credit-to-GDP time series and observes that out of the credit gaps calculated with one-sided HP filters, the one with a cycle length of about 15 years exhibits the best early warning property.

The rest of the paper is organized as follows. Section 2 gives an overview about the data used, then Section 3 describes the 48 credit gaps under review, with a detailed justification of the cycle lengths. Section 4 contains our findings about early warning performance on European data, and their robustness checks. Section 5 characterizes the degree of comovement of credit gaps with the financial cycle based on the endpoint uncertainty of credit gaps and the correlation of credit gaps with aggregated indicators of financial cycle. Section 6 discusses the performance of signaling the excessive credit growth in simulated Hungarian data. Section 7 concludes.

2 Data

Our study focuses on identifying the best univariate credit-to-GDP gap for Hungary, therefore we do not consider non-European countries (which are markedly different from Hungary). Three types of data are used from 18 European countries. First, credit-to-GDP time series are necessary to compute and compare credit-to-GDP gaps and to generate the simulated Hungarian credit-to-GDP time series as well. Second, periods of systemic banking crisis are required to measure the prediction power of credit gaps. Third, we employ aggregated indicators describing the position of the financial cycle to assess the comovement of credit gaps with the financial cycle. (See Appendix A for the details of the three datasets.)

We use credit-to-GDP time series mainly from the BIS private non-financial sector credit database.[13] These are the longest quarterly credit-to-GDP time series available, which are useful both for computing credit gaps and for testing their early warning performance. This database contains time series calculated with broad and narrow definition of the outstanding amount of credit. However, it only covers 19 ESRB member countries[14], from which we disregard two countries that we consider as outliers. Additionally, we have collected quarterly credit-to-GDP time series published by national macroprudential authorities to have as long time series from as many ESRB member countries as possible. In the end, the data from Lithuania are only used, because they contain sufficiently long time series of both narrow and broad definitions of credit stock.[15]

The unbalanced panel data span the period from 1960 Q4 to 2019 Q1. There are no observations from the 1970s for only three countries, because the Czech, Polish and Lithuanian time series begin in the 1990s. The examined 18 countries are classified into four country groups: Nordic countries, Central and Eastern European countries, Mediterranean countries, core EU countries. While credit-to-GDP evolves quite similarly within groups, it exhibits considerable differences across groups (Figure 1). Many of our examinations are conducted separately for each country group as well.

For assessing early warning performance, we make use of the crisis database of the ESRB published in 2017 (Lo Duca et al., 2017). This is currently considered the most comprehensive and accurate compilation of European financial crisis periods measured in months. The monthly observations are used on a quarterly basis from 1970–2016.[16] We focus only on the systemic financial crisis periods with the banking sector in stress that are not related to the post-socialist transition. This selection results in 26 crises out of which 15 are related to the global financial crisis started in 2007 (Figure 2). Since most credit-to-GDP time series extend well back into the 1960s, all selected crises can be used for evaluating early warning signals of credit-to-GDP gaps. Crises affected the 18 countries under review at a similar extent: Czech Republic and Poland did not experience a crisis, while there were two episodes in six countries and three episodes in two countries (see Appendix A).

For measuring the degree of comovement between a credit gap and the financial cycle, we have collected financial cycle indicators from publications of national macroprudential authorities. Although the exact information content and methodology of these indicators vary markedly, all of them aggregate several variables related to the financial cycle capturing a major part of the relevant determinants. We could find at least one such indicator for the half of the 18 countries, with 11 indicators in total (two for each of Denmark and Norway).

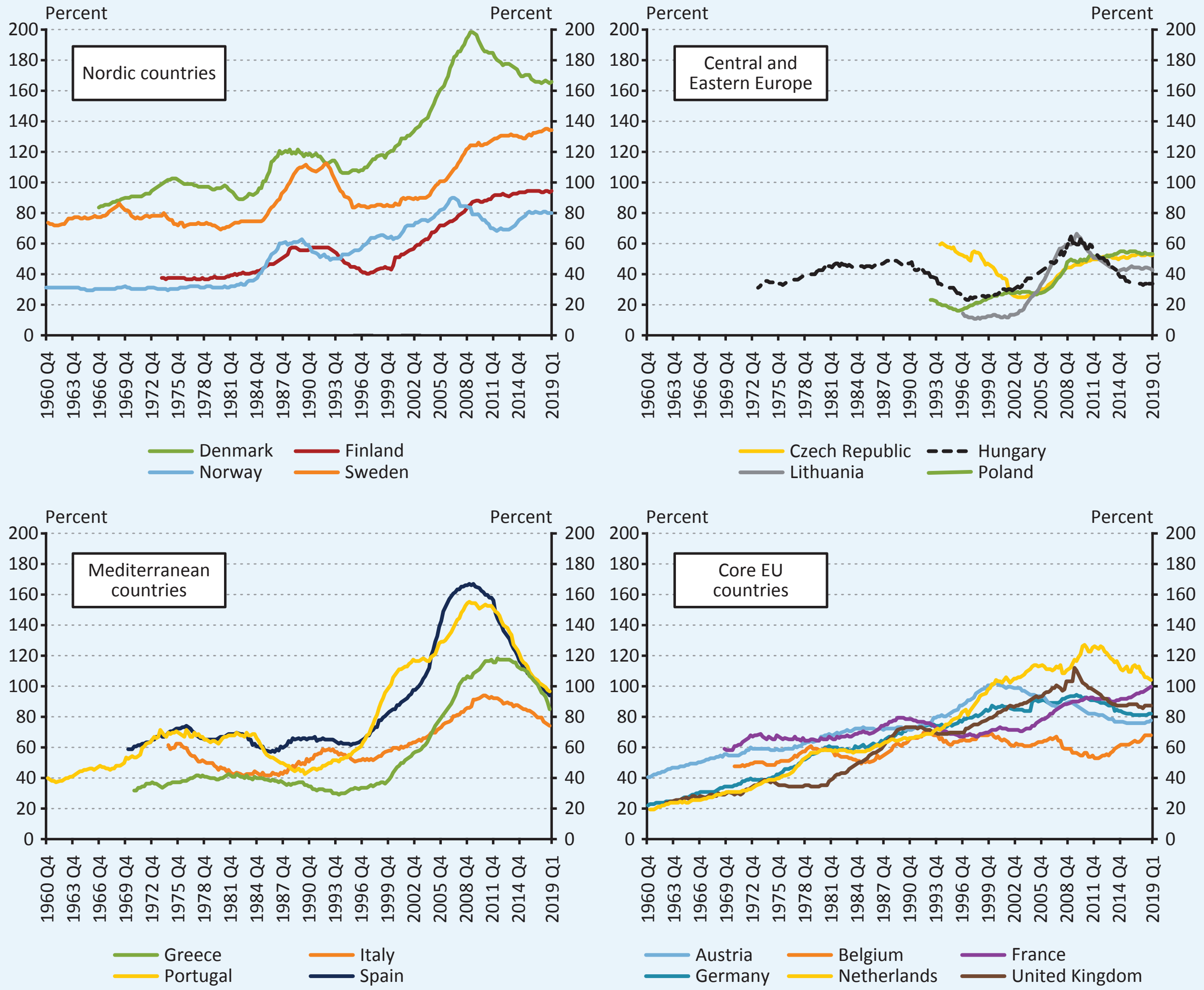
[13] Source: https://www.bis.org/statistics/totcredit.htm.

[14] ESRB member countries include all EU member countries (which for the present purposes covers the UK too), as well as Iceland, Liechtenstein, and Norway.

[15] Source: https://www.lb.lt/uploads/documents/files/EN/our-functions/financial-stability/CCyB_data_2019Q2.xlsx.

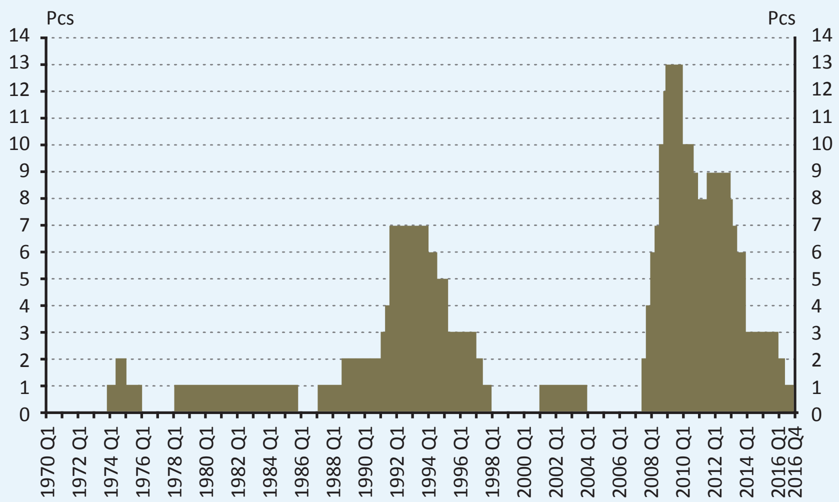
[16] Crisis periods include all quarters in which the original database registered a crisis for at least one month.

Figure 1
Credit-to-GDP series calculated with narrow definition of credit by country groups



Source: BIS, Lietuvos Bankas.

Figure 2
Number of countries in systemic banking crisis



Source: ESRB financial crisis database, Lo Duca et al. (2017).

3 Credit-to-GDP gap specifications

This section presents the exact specifications of the 48 credit gaps differing along four dimensions. Two characteristics are related to the credit-to-GDP time series, and the other two are linked to the trend-cycle decomposition procedures. Table 2 summarizes the options considered in each dimension.

Table 2
Specifications of the credit-to-GDP gaps

Which credit-to-GDP series do we use for gap calculations?	Definition of the stock of credit	2 options	Credit-to-GDP series calculated with broad and narrow definitions of the outstanding amount of credit
	Credit-to-GDP forecast	3 options	Actually available credit-to-GDP series and their extended versions with forecasts for 1 and 3 years
How do we calculate gaps?	Filtering method	3 options	Hodrick–Prescott filter, Christiano–Fitzgerald filter, Wavelet filter
	Cycle length	3 options	HP: up to 19, 25, and 32 years CF: from 2 up to 18, 24 and 30 years W: from 2 up to 16 and 32 years

3.1 CREDIT DEFINITIONS

We include two possible definitions for the outstanding amount of credit. The broader one contains all types of loans and debt securities taken out by the private non-financial sector, the narrower one contains only the loans and debt securities from the domestic banking system.[17] Both can be useful in the build-up phase of the CCyB. Credit gaps calculated with the narrow definition focus on the portion of excessive lending that can be most directly affected by the CCyB. Credit gaps calculated with the broad definition capture all forms of excessive lending including the part that eventually gets diverted outside the regulated banking sector as a possible response to the use of CCyB. ESRB member countries are divided on the credit definition they use for calculating their benchmark credit-to-GDP gap (Table 1).

3.2 CREDIT-TO-GDP FORECASTS

We focus on one-sided credit gaps, so we determine a credit gap value for a specific quarter based on the credit-to-GDP observations up to this quarter.[18] The use of credit-to-GDP forecasts aims to improve the accuracy of the one-sided credit gap estimates, or in other words to reduce the endpoint uncertainty of the one-sided approach. If the gap calculation can use credit-to-GDP values not only from the past and the present but also from the future, the cyclical part of the credit-to-GDP ratio, which changes faster, can be better distinguished from the trend, which changes slower.

We could not compile actual forecasts, so we rely on an approximation in the form of 'perfect forecasts', i.e. actual data from later on. We extend 'actually observable' credit-to-GDP time series by 4 or 12 quarters of perfect forecasts. Applying longer extensions would be unrealistic since macroeconomic variables can only be projected in practice with a great degree of uncertainty at these forecast horizons. Our approach approximates credit gaps computed with the most accurate forecasts of credit-to-GDP time series that can be determined in practice (such as central banks' regular macroeconomic forecasts).

[17] See Appendix A for the exact definitions.
[18] By contrast, two-sided credit gaps are calculated using the whole time series.

3.3 FILTERING METHODS

Economic time series are usually decomposed into four components: trend, cycle, seasonality, and noise. Seasonality is not addressed here, as it is negligible in the case of credit-to-GDP time series. This is because the credit stock in the numerator typically has a low degree of seasonality, and the denominator always includes 4-quarter rolling GDP.

In this paper, we use three filtering methods for calculating credit gaps: HP filter, CF filter and wavelet filter.[19] They differ significantly in their logic and methodology, therefore we can cover a wide range of possible univariate filtering techniques. The HP filter is the most common method of estimating trends for macroeconomic time series because it is simple and straightforward. The guidance of the Basel Committee on Banking Supervision (BCBS, 2010) and the recommendation of the European Systemic Risk Board (ESRB, 2014) advocate the HP filter (with a lambda of 400,000, assuming approximately 30-year long cycles) for calculating the main credit gap supporting regulatory decisions on the build-up of the CCyB. The main arguments against the HP filter point to its large endpoint uncertainty and its inappropriate applications. (Edge and Meisenzahl, 2011; Hamilton, 2018).

The second filtering method is the CF filter, which is a bandpass or frequency filter. Unlike with the HP filter, specific lower and upper thresholds control explicitly the range of cycle lengths that the CF filter 'passes' into the cyclical component of the time series decomposition. Thereby, noise filtering is also possible with CF filter: Specifically, we consider cycles shorter than two years as noise. In contrast, the lambda parameter of the HP filter determines only implicitly and only the maximum cycle length belonging to the cyclical component. Consequently, the CF filter generates smoother credit gaps than the HP filter.

Among the numerous specifications for the wavelet filter, we chose the MODWT (maximal overlap discrete wavelet transform) procedure, which is the most common version used for economic time series. An important feature of this discrete wavelet filter compared to other variants is that the length of the sample is not bound to the powers of 2 as with other methods. The wavelet filter has the advantage over the HP and CF filters that it can better manage structural breaks in the time series. This is because the wavelet filter divides the available time series into several parts, and the importance of the cycles with different lengths is determined separately for all parts. In addition, this is performed efficiently, so longer cycles are identified from a partition with longer time series parts (see Figure 11 in Appendix B). Therefore, the wavelet filter can also handle a structural change that occurs only at certain frequencies. However, short time series entail a greater disadvantage for the wavelet filter than for the other two, because the sample has to be at least as long as the maximum length of the cycle sought to be filtered with wavelet filter. By contrast, the HP and CF filters allow the frequency parameters to be chosen independently from the sample length. Similar to the CF filter, the wavelet filter also uses two parameters for setting the range of cycle lengths, but the choice is more limited because both the upper and lower thresholds have to be a power of 2. In this paper, the lower threshold is always two years (8 quarters).

Table 3 summarizes the main properties of the three methods. Assumptions about the symmetry have not been mentioned yet. This refers to whether the specification of a filtering method requires credit-to-GDP observations from dates following the date of the credit gap value to be estimated. This problem does not arise with the HP filter, and an asymmetric solution is chosen with the CF filter (no additional observations are required). However, the wavelet filter always operates symmetrically, so the credit-to-GDP samples have to be extended artificially in a special way.

[19] For more details on the three methods see Appendix B.

Table 3
Main properties of the filtering methods

Method	Parameters	Noise filtering	Structural breaks	Assumption about symmetry	Minimum required length of credit-to-GDP series
Hodrick–Prescott filter	Lambda: smoothness vs. goodness of fit	not possible	disregards it	no	no
Christiano–Fitzgerald filter	F_l, F_u: shortest and longest cycle length	possible	disregards it	asymmetric	no
Wavelet filter	N: maximum cycle length is 2^N	possible	takes into account	symmetric	2^{N-1}

There exist several other univariate methods that are not considered in this paper. These can be classified basically into two groups: simple deterministic procedures (e.g., moving average, linear, quadratic, or cubic trend) and methods based on time series econometric models (e.g., Beveridge–Nelson filter). The former group is left out because they are inflexible methods that cannot easily adapt to different cycle lengths, while our approach regards cycle length as an important dimension to account for. Furthermore, many of them can be approximated by a special case of the HP filter. The Beveridge–Nelson filter is not included because it is suitable for filtering cycles that have much higher frequencies than credit cycles (see Canova, 1998), and it cannot adequately manage permanent (15-30 year long) deviations. Other special methods (e.g. residual from linear projection, special latent variable models) are disregarded because we consider their properties not materially different from the filters examined in the paper or from any of the above-mentioned procedures left out for a reason.

3.4 CYCLE LENGTHS

We provide preliminary results about the typical cycle lengths in the credit-to-GDP time series in our database, which helps in determining the relevant cycle length for the credit gaps to be calculated. We apply two procedures for this purpose: a discrete wavelet filter, and a periodogram. The wavelet filter can be used to quantify which cycle explains what variance in the cyclical component of a time series. For each country, the credit-to-GDP time series calculated with broad and narrow definitions of credit are decomposed into trends and cycles of 2–4, 4–8, 8–16 and 16–32 years. Then we compute the shares of the total variance of the cyclical component explained by the different cycles.

The periodogram is based on the Fourier transform, and it shows the relative importance of the cycles with various frequencies in the time series. The greater the value assigned by the periodogram to a given cycle length, the greater the role it plays in the time series. The lowest frequency that can be examined with the periodogram corresponds to the entire length of the time series. In our case, this falls between 45 and 50 years, depending on the definition of credit. The half, third, quarter and other fractions of the sample length represent the different cycle lengths. One advantage of the periodogram over the wavelet filter is that it offers a more detailed decomposition, but a significant drawback is that it can only be applied to stationary time series. Therefore, the time series are first decomposed into trends and cyclical parts with a CF filter (assuming a maximum cycle length of 30 years[20]), and then the periodograms of the derived cyclical parts are computed.[21] Detrending with the CF filter is performed in two ways, using both a one-sided and a two-sided approach.

[20] Because of the detrending method, the cycles of over 30 years have negligible values in the periodogram.
[21] The CF filter is chosen for detrending, because it is also based on a Fourier transform like the periodogram.

We obtain similar results with the two methods (see Appendix C): Credit cycles are clearly longer than business cycles, and their average duration is usually 15–30 years. The exact values are clustered in the middle and upper part of this range and vary depending on the country, the definition of credit and the method chosen. We find that shorter cycles have greater significance mainly in the core EU countries (Austria, Belgium, Netherlands, and UK), and to a lesser extent in two Nordic countries (Denmark and Norway) and in Italy. Our findings are consistent with the literature establishing that credit cycles are longer than business cycles without a consensus on the exact value of the typical cycle length.[22] It may vary between 10 and 30 years depending on the country group, the analyzed period, and the estimation method.

In this paper, we examine credit gaps with relevant cycle lengths. HP filter is coupled not only with a lambda of 400,000, in line with international recommendations, but also with a lambda of 160,000 and 50,000 corresponding to cycles with maximum length of 32, 25 and 19 years, respectively. CF filter is used with three upper thresholds 120, 96 and 72, which means the maximum allowed cycle lengths are very similar to the ones with the HP filter (30, 24, and 18 years). The lower threshold is always 8, so noise is defined as cycles with length of less than two years. Given the less flexible nature of the wavelet filter, we can consider only two different maximum cycle lengths: 16 and 32 years, whereas noise is defined identically with the CF filter.

[22] See for example Drehmann et al (2012), Borio (2014), Aikman et al (2015), Galati et al (2016), Jordà et al (2016), Hiebert et al (2018).

4 Early warning performance on European data

4.1 UNIVARIATE SIGNALING APPROACH

We follow a common approach to assessing early warning signals of systemic banking crises. Kaminsky and Reinhart (1999) introduced the signaling approach for predicting macroeconomic crises, which was followed by several other similar applications.[23] The idea of the method is simple. A credit gap issues a warning signal if its value exceeds an appropriately chosen threshold. The method compares these signals to the signals of an ideal early warning indicator that issues warning signals only before crises. The pattern of misalignment can be used to quantify the prediction power of the credit gap.

Using pooled data, we expected from the ideal indicator to issue the following warning signals in different countries and quarters. We do not evaluate credit gap signals in some periods: the first 32 and the last 12 quarters of each credit-to-GDP time series,[24] 4 quarters preceding crises, and the whole crisis periods.[25] In the baseline case, warning signals are expected in the period between the 5th and 16th quarter prior to a crisis. Nearer to a crisis, credit gaps can more easily warn of an impending crisis, but macroprudential policy has less chance to deploy effective countermeasures. In order to manage this trade-off flexibly, we also consider the prediction horizons of 12–5 quarters and 20–5 quarters prior to the crises.[26] Credit gaps are expected to issue no warning signal in all other periods.[27]

We rank credit gaps with two evaluation statistics: AUROC (area under the receiver operating characteristic curve) and RU (relative usefulness) (see Appendix D for details). AUROC assesses credit gaps' performance across all possible thresholds. The AUROC value of an ideal early warning indicator is 1, while a completely uninformative indicator has an AUROC of 0.5. RU is defined to a specifically chosen threshold that is usually determined by introducing regulatory preferences over Type I and Type II errors of the early warning signals with different thresholds. The highest possible value of RU is 1, completely uninformative indicators produce a RU of 0. We consider the AUROC as the more instructive performance measure because it does not depend on a somewhat arbitrary regulatory preference, and it generally characterizes the strength of the trade-off between the two types of errors. Therefore, we employ the RU criterion in robustness checks.

[23] For example, Kaminsky et al. (1998), Borio and Lowe (2002), Borio and Drehmann (2009), Alessi and Detken (2011), Drehmann and Juselius (2014), Babecký et al. (2014), Detken et al. (2014), Alessi et al. (2015), Coudert and Idier (2018), Lee et al. (2020), Tölö et al. (2018), Lang et al. (2019).

[24] There is only one crisis out of the 26 that cannot be predicted because of this constraint: the Spanish crisis starting in 1978.

[25] We have the following reasons. At the beginning of the credit-to-GDP time series, either credit gaps are barely different from zero because of the one-sided calculation approach, or they cannot be calculated at all. Towards the end of the credit gap time series, we cannot define ideal warning signals unless we know where and when crises would occur after 2019. We have an additional reason for leaving out the last three years from evaluation: We want to use the same data to compare all credit gaps, including those calculated with the longest, 3-year credit-to-GDP forecasts. Credit gap signals just before a crisis cannot effectively modify macroprudential interventions, so we do not want signals from this period to influence the assessment of the predictive power. During systemic banking crisis periods, credit gap values fluctuate dramatically due to the volatility of GDP, and usually there is no need for predicting the next crisis because it typically occurs much later.

[26] This is consistent with the common practice in the literature, see, for example, Detken et al. (2014), Tölö et al. (2018), Lang et al. (2019).

[27] We implicitly assume here that no country experienced a systemic banking crisis between 2017 Q1 and 2019 Q1 (in fact, even longer than this in the case of prediction horizon larger than 12–5 quarters) because the employed ESRB crisis database has observations up to the end of 2016. At the end of 2016, it classifies only Greece as being in a systemic banking crisis period.

4.2 RESULTS

Table 4 contains the main results about the early warning performance of the different credit gaps. It shows that it is easier to answer the 'What to filter?' question than the 'How to filter?' question. The most distinct finding is that credit-to-GDP gaps perform the best when they are calculated with narrow definition of credit, in other words the credit provided by the domestic banking system to the private non-financial sector. All credit gap specifications give at least as good (and often significantly better) predictions with narrow credit than with broad credit. The finding that narrow credit outperforms broad credit reinforces recent results obtained with European data (Detken et al., 2014, Tölö et al., 2018, Lang et al., 2019).

The second main finding is that using longer-term credit-to-GDP forecasts in credit gap calculations tends to be counterproductive for early warning purposes. Credit gaps with a 3-year forecast perform worse than credit gaps with the other two options. Compared to the respective credit gaps with no forecast, using a 1-year forecast undermines the predictive power of the credit gap in the case of the HP and CF filters, while it improves in the case of the wavelet filter. This result is somewhat surprising, because using credit-to-GDP forecasts approximates credit gaps to their two-sided versions, which usually exhibit greater cyclical positions than the one-sided versions,[28] providing the possibility of issuing clear warning signals of impending crises.

Table 4
AUROC values of credit gaps in the baseline analysis

Using all crisis periods from all countries			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
We expect "Signal" in 20-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.74	0.74	0.71	0.73	0.74	0.73	0.71	0.73
		1-year forecast	0.74	0.73	0.70	0.66	0.67	0.66	0.72	0.73
		3-year forecast	0.64	0.61	0.55	0.54	0.55	0.54	0.68	0.57
	Broad definition of credit stock	0-year forecast	0.74	0.73	0.69	0.73	0.74	0.71	0.71	0.72
		1-year forecast	0.73	0.72	0.68	0.65	0.65	0.62	0.72	0.72
		3-year forecast	0.62	0.58	0.52	0.51	0.50	0.53	0.67	0.54
We expect "Signal" in 16-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.74	0.74	0.71	0.73	0.74	0.73	0.73	0.74
		1-year forecast	0.74	0.73	0.70	0.67	0.67	0.65	0.74	0.74
		3-year forecast	0.67	0.64	0.58	0.59	0.59	0.57	0.71	0.60
	Broad definition of credit stock	0-year forecast	0.74	0.72	0.68	0.73	0.73	0.70	0.72	0.72
		1-year forecast	0.73	0.71	0.67	0.65	0.65	0.60	0.73	0.72
		3-year forecast	0.62	0.59	0.53	0.53	0.52	0.52	0.68	0.54
We expect "Signal" in 12-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.75	0.75	0.73	0.74	0.75	0.73	0.74	0.75
		1-year forecast	0.75	0.74	0.72	0.70	0.70	0.68	0.75	0.76
		3-year forecast	0.72	0.70	0.67	0.67	0.68	0.66	0.74	0.68
	Broad definition of credit stock	0-year forecast	0.74	0.72	0.68	0.73	0.72	0.69	0.72	0.71
		1-year forecast	0.73	0.70	0.66	0.65	0.64	0.59	0.73	0.72
		3-year forecast	0.63	0.60	0.55	0.56	0.54	0.51	0.69	0.57

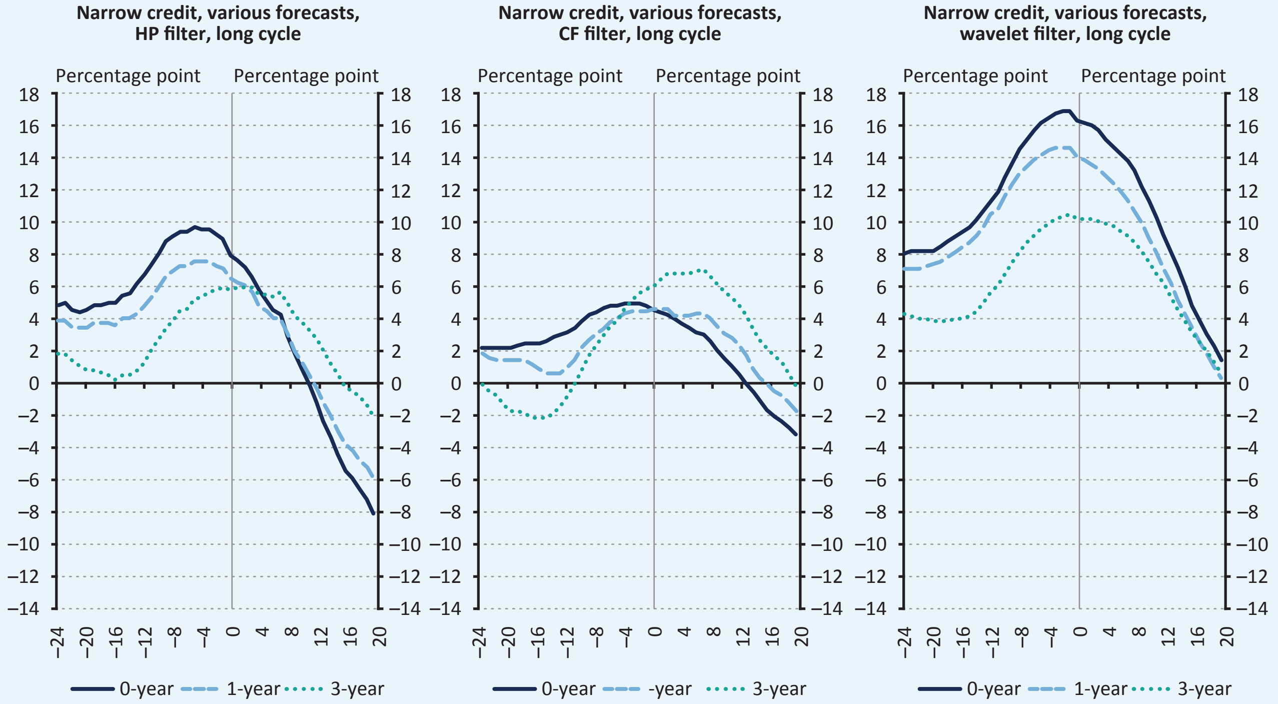
Note: Red cells show the highest AUROC values of various prediction horizons. Orange cells contain values that do not differ from the highest values at the 10 percent significance level. Yellow and green cells indicate additional AUROC values that do not differ from the highest AUROC value at the 5, and 1 percent significance levels, respectively. Tests of differences in AUROC values are based on Delong et al. (1988). The test statistics verify the identical nature of the ROC curves, rather than the AUROC values directly, and they also take into account sample size, so credit gaps with the same AUROC value may differ from the highest AUROC value at marginally different significance levels.

[28] See, for example, the discussion of the periodograms produced with one-sided and two-sided CF filters in Appendix C.

Although the 3-year forecast horizon can take into account further increase in credit-to-GDP ratios during the expansion phase of the credit cycle, it is too short to cover the contraction during the banking crisis. Furthermore, credit-to-GDP ratios typically do not shrink at the beginning of banking crises in the face of a sudden recession (falling GDP). Accordingly, the 'perfect' 3-year credit-to-GDP forecast increases the current trend (and reduces the gap) compared to the version without the forecast. Close to a crisis, however, a 'perfect' 3-year forecast can contain stagnation or even decline in credit-to-GDP ratios, which decreases current trend values and increases gap values. Accordingly, using a 3-year forecast results in lower credit gaps that rise more sharply before crises (Figure 3). Table 4 also demonstrates that these credit gaps tend to issue more informative early warning signals with shorter prediction horizons. These findings are consistent with other papers concluding that two-sided credit gaps have a poorer early warning performance than their one-sided versions: e.g. Drehmann et al. (2011) and Drehmann and Juselius (2014).

The third main conclusion from Table 4 is that none of the filtering methods and cycle lengths are dominated. Instead, special combinations of different alternatives provide the best credit gaps. The credit gap calculated with narrow credit, no credit-to-GDP forecast, an HP filter, and the longest cycle length (lambda: 400,000) reaches the highest AUROC value out of the 48 gap indicators if we employ the baseline prediction horizon of 16–5 quarters. The AUROC value of 0.74 makes it a reliable early warning indicator. The best credit gap does not stand out from the others with this performance: 6, 14, and 15 other credit gaps have an AUROC value that is not different from the highest one at the 10, 5, and 1 percent significance levels, respectively.[29]

Figure 3
Development of average credit gaps calculated with various credit-to-GDP forecasts around systemic banking crises



Note: The horizontal axis shows quarters, the 0th point in time is the quarter when systemic banking crisis begins. The charts show averages of credit gaps with a given specification across countries and banking crises.

[29] The test is based on Delong et al. (1988).

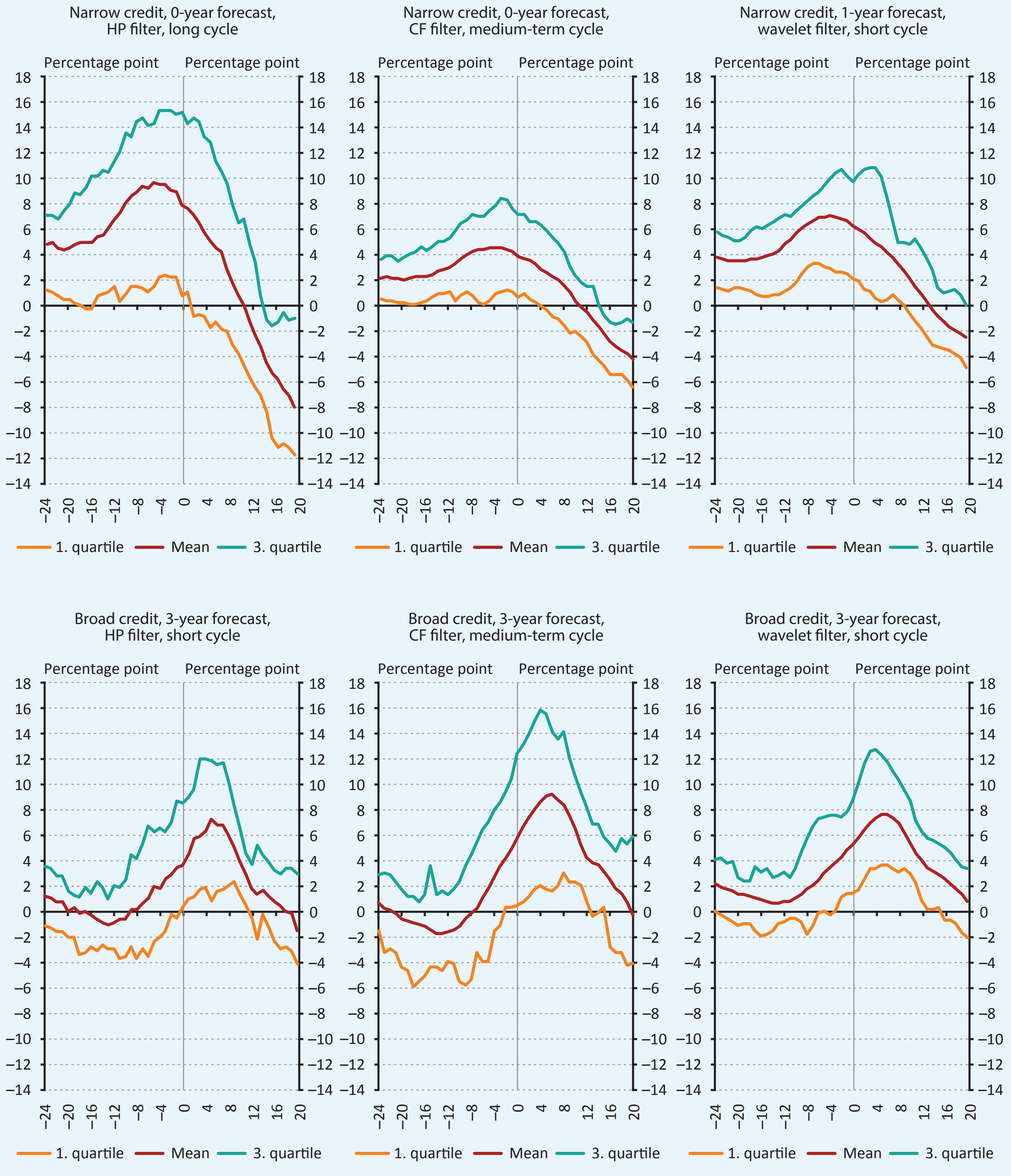
In order to identify the best early warning credit gap, we introduce an additional selection criterion. The best credit gap should attain the highest AUROC values at the 10 percent significance level with all three prediction horizons of 12–5, 16–5, and 20–5 quarters. The following three credit gaps can pass this test:

- narrow credit, 0-year forecast, HP filter, long cycle,
- narrow credit, 0-year forecast, CF filter, medium-term cycle,
- narrow credit, 1-year forecast, wavelet filter, short cycle.

We will refer to these specifications as the best HP filter, CF filter and wavelet filter credit gaps. The top row of Figure 4 shows that in the years preceding the systemic banking crises (especially in the first three years), all three credit gaps have materially higher values than earlier or during the crisis period. This demonstrates that these credit gaps can issue informative warning signals that are slightly more accurate with shorter prediction horizons, which can also be seen in Table 4. By comparison, the bottom row of Figure 4 includes credit gaps producing barely informative signals (AUROC values of 0.50–0.57). These credit gaps are typically near zero in the fourth and third year before a crisis, and they only rise considerably in the first year.

According to Figure 5, the top three credit gaps perform well for all the forecasting horizons that are between 6 and 16 quarters long. The highest AUROC value always belongs to one of these credit gaps with all prediction horizons. There are also meaningful differences among the three gaps. If the selection criterion also included the prediction horizon of 10–5 quarters, the CF filter credit gap could not be among the top ones. Still, this is the best early warning credit gap for the longest prediction horizons. The other two perform more evenly. The HP filter credit gap is particularly reliable for longer prediction horizons (15–20 quarters), while the wavelet filter credit gap is particularly reliable for shorter ones (10–15 quarters).

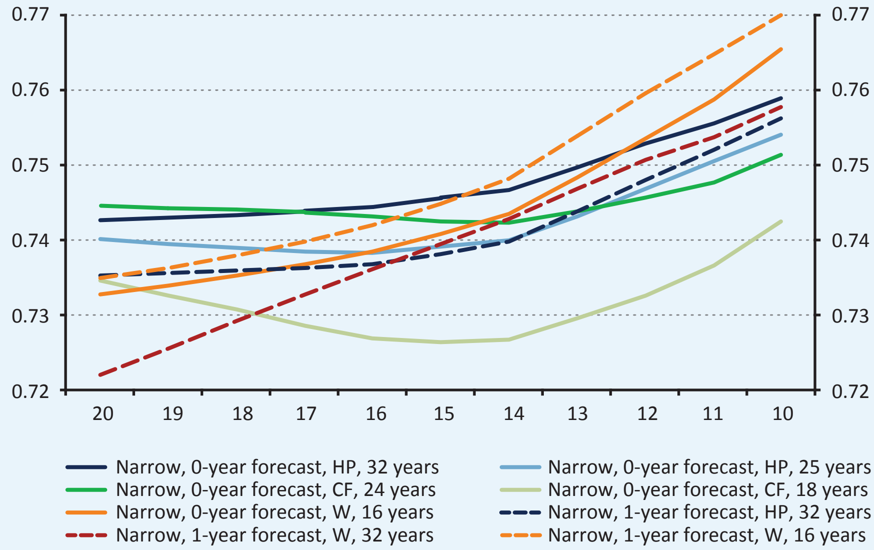
Figure 4
Development of the three best and the three worst early warning credit gaps around systemic banking crises



Note: The top row shows the three best early warning credit gaps, while the bottom row includes three from among the worst ones. Horizontal axes show quarters, 0th point in time is the quarter when systemic banking crisis begins. The charts characterize the distributions of credit gaps over countries and banking crises.

The results confirm our approach, namely the optimization of credit-to-GDP gaps along four characteristics at the same time. The top three early warning credit gaps are calculated with two different credit-to-GDP forecasts, all three filtering methods, and all three cycle lengths. This means that the top three credit gaps differ markedly from each other along three dimensions.

Figure 5
AUROC values of the best early warning credit gaps with various prediction horizons



Note: The horizontal axis includes the number of quarters by which the beginning of a given prediction horizon precedes the next crisis. The figure contains credit gaps with AUROC values that do not differ from the respective highest AUROC values with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters at the 1 percent significance level. Solid lines refer to credit gaps calculated from credit-to-GDP series without a forecast, while dashed lines refer to those with a forecast. The three colors differentiate the three filters. Darker shades denote longer cycles.

4.3 ROBUSTNESS CHECKS

The baseline analysis uses pooled data from the entire unbalanced panel database of European countries and employs AUROC to evaluate the early warning signals of the 48 credit gaps with respect to systemic banking crises. It turns out that the 48 credit gaps and the examined alternatives of the four characteristics of credit gaps can be ranked only roughly. We conduct a number of robustness checks to corroborate the results of the baseline analysis. This means the following: different evaluation statistic, out-of-sample prediction exercise, subsamples by country groups and time, alternative method for credit-to-GDP forecast.

4.3.1 RU

The baseline analysis employs AUROC because it quantifies the strength of the trade-off between Type I and Type II errors without assuming specific preferences over them. RU criterion is also popular in the literature, and it can be useful indeed, when policy makers' preferences are more or less known. It is commonly assumed that these preferences are balanced or assign a greater importance to not recognizing an impending crisis because of the large economic and social impact of systemic banking crises. Accordingly, we also assume balanced preferences over Type I and Type II error rates, which means a preference parameter of 0.5 in the loss function describing the preferences. Since systemic banking crises occur rarely, any value significantly higher than 0.5 would make preferences extremely sensitive to the error of not recognizing crises.[30] A great advantage of the RU criterion is that it enables the evaluation of out-of-sample prediction performance, which is important because it is more in line with how credit gaps are actually used in practice. Therefore, we conduct both in-sample and out-of-sample early warning exercises evaluated by the RU.

Table 5 presents the result of the in-sample evaluation with RU applying the same database as in the baseline analysis. No statistical test has been found to determine the significance of the differences between the RU values, so a rule of thumb is used: We consider RU values over 0.4 not to differ significantly from the highest RU values, which allows for an approximately 10 percent divergence.[31] In line with baseline analysis, we find that narrow credit outperforms broad credit and 3-year credit-to-GDP forecast worsens the prediction power. Among the top three credit gaps, the CF filter credit gap is clearly the best one according to the RU values. It achieves the highest RU values with the two longer prediction horizons, and it is a close runner-up with the shortest. The HP filter credit gap is also among the top credit gaps. This group is defined similarly to the baseline analysis and contains five credit gaps denoted with bold RU values in Table 5. The wavelet filter credit gap performs well at shorter prediction horizons, but it predicts crises a bit less accurately than the top credit gaps with the longest prediction horizon. Their relatively high RU values confirm that all three credit gaps are reliable early warning indicators, which has also been found with AUROC values.

Table 5
RU values of credit gaps with balanced preferences

Using all crisis periods from all countries			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
We expect "Signal" in 20-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.41**	0.39	0.36	**0.41**	**0.45**	**0.41**	0.36	0.38
		1-year forecast	**0.41**	0.38	0.33	0.27	0.28	0.26	0.37	0.39
	Broad definition of credit stock	0-year forecast	0.36	0.36	0.31	0.35	0.35	0.34	0.30	0.33
		1-year forecast	0.35	0.33	0.27	0.23	0.25	0.22	0.31	0.34
We expect "Signal" in 16-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.42**	0.39	0.35	**0.42**	**0.46**	**0.41**	0.38	0.40
		1-year forecast	**0.41**	0.37	0.32	0.30	0.30	0.27	0.40	0.40
		3-year forecast	0.29	0.23	0.16	0.18	0.20	0.15	0.34	0.16
	Broad definition of credit stock	0-year forecast	0.38	0.36	0.31	0.36	0.36	0.34	0.33	0.32
		1-year forecast	0.36	0.34	0.26	0.23	0.24	0.21	0.34	0.32
		3-year forecast	0.19	0.16	0.08	0.10	0.07	0.08	0.28	0.10
We expect "Signal" in 12-5 quarters before crises										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.43**	0.41	0.36	**0.44**	**0.44**	**0.42**	0.40	0.41
		1-year forecast	**0.43**	0.40	0.35	0.35	0.38	0.34	0.42	0.42
	Broad definition of credit stock	0-year forecast	0.38	0.36	0.30	0.36	0.35	0.34	0.33	0.32
		1-year forecast	0.36	0.33	0.25	0.23	0.22	0.19	0.34	0.33

Note: RU values are calculated assuming balanced preferences over Type I and Type II error rates. (The value of the preference parameter θ is 0.5. For more details, see Appendix D.) Red cells show the highest RU values of various prediction horizons. Grey cells include values higher than 0.40. Credit gaps belonging to cells with bold numbers have RU values of over 0.40 for each of the three prediction horizons. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

[30] For more details, see Appendix D.

[31] For the credit gaps computed with a 3-year credit-to-GDP forecast, RU values are only presented with the prediction horizon of 16–5 quarters, as they clearly show a poor forecasting performance even there.

Our out-of-sample early warning exercise is based on the method of Lang et al. (2019). The binary signals of credit gaps are calculated and assessed in a similar way as during the in-sample analysis with prediction horizons of 12–5 and 16–5 quarters. The only difference is that we use the sample only up to 1996 Q4 for threshold optimization assuming balanced preferences.[32] In turn, we compute the signal values in 2000 Q1 (or 2001 Q1 in the case of the longer prediction horizon) with the credit gap values in 2000 (2001) Q1 and the optimal threshold values. This is repeated for the signal values in 2000 (2001) Q2 by expanding the sample used for determining the optimal threshold with one quarter. We continue until all the signals up to 2016 Q1 are produced. The advantage of out-of-sample predictions compared to in-sample predictions is that they simulate the actual early warning signals in practice more accurately, but its drawback is that the evaluation can occur only for a short and special period, namely when the 2007–2008 global financial crisis has to be predicted.[33]

The out-of-sample analysis reaches different conclusions compared to the in-sample analysis (Table 6). Several credit gaps with near-zero (or even negative) RU values prove to be barely informative. The best credit gaps have similar RU values (around 0.4) as the ones in the in-sample analysis, but they are more clearly distinct from worse credit gaps, and they perform differently with different prediction horizons. The best credit gap with a three-year prediction horizon (broad credit, 1-year forecast, CF filter, medium-term cycle) and the best credit gap with a four-year prediction horizon (broad credit, 3-year forecast, HP filter, medium-term cycle) are the only ones that have an RU value of at least 0.3 on both prediction horizons.

Table 6
Out-of-sample RU values of credit gaps with three- and four-year predictions and balanced preferences

Using all crisis periods from all countries belonging to the sample for threshold optimization			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
Three-year out-of-sample prediction										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	-0.01	-0.06	-0.22	0.07	0.14	0.08	0.10	0.17
		1-year forecast	0.00	-0.11	-0.17	0.07	0.18	0.12	-0.08	0.17
		3-year forecast	-0.07	-0.13	0.01	-0.05	0.23	0.01	-0.07	-0.03
	Broad definition of credit stock	0-year forecast	-0.18	-0.15	-0.05	0.14	0.28	0.23	0.17	0.13
		1-year forecast	-0.05	0.00	0.15	0.29	0.38	0.23	0.21	0.28
		3-year forecast	0.25	0.32	0.26	0.28	0.21	0.13	0.09	0.31
Four-year out-of-sample prediction										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	-0.01	-0.04	0.04	0.13	0.26	0.11	0.00	0.12
		1-year forecast	-0.01	0.01	0.12	0.17	0.25	0.21	-0.11	0.15
		3-year forecast	0.18	0.39	0.38	0.20	0.26	0.24	0.01	0.03
	Broad definition of credit stock	0-year forecast	0.11	0.00	0.16	0.21	0.33	0.17	0.10	0.32
		1-year forecast	0.30	0.19	0.30	0.23	0.32	0.21	0.00	0.29
		3-year forecast	0.35	0.47	0.28	-0.14	0.16	0.10	0.01	0.29

Note: RU values are calculated assuming balanced preferences over Type I and Type II error rates. (The value of the preference parameter θ is 0.5. For more details, see Appendix D.) Red cells show the highest RU values of various prediction horizons. Grey cells include values higher than 0.30. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

[32] This also means that the out-of-sample analysis does not include the Czech, Polish and Lithuanian observations that start in the 1990s.

[33] The sample period cannot be divided in any markedly different manner, because either the subsample used for estimation or the one used for evaluation would be too short.

Credit gaps calculated with broad credit provide better predictions more often than their versions with narrow credit. Using credit-to-GDP forecasts pays off in several credit gap specifications. The top three credit gaps of the baseline analysis perform average at best in the out-of-sample analysis, including the HP credit gap being nothing short of useless. These findings cast doubt on the robustness of the in-sample results obtained with AUROC and RU.

The general lesson from robustness checks with RU is that in-sample analysis confirmed the predictive power of the HP and CF filter credit gaps out of the top three credit gaps from baseline analysis, while the out-of-sample exercise disputes the performance of all three, especially the HP filter credit gap. In our view, the out-of-sample analysis refers to a fairly specific time period, which makes the results based on the full sample more informative. This is partly the reason why the out-of-sample analysis is included among robustness checks rather than in the baseline analysis.[34]

4.3.2 Country groups

The baseline analysis uses pooled data of 18 countries. These countries can be classified into four more or less homogenous groups according to their overall economic conditions, including especially the evolution of their credit-to-GDP time series: core EU countries, Nordic countries, Mediterranean countries, Central and Eastern European countries. Additionally, we have seen in Section 3.4 that shorter cycles in the credit-to-GDP time series are more dominant in the core EU countries and in some Nordic countries. Consequently, the question arises whether the baseline analysis would yield different results for each country group.

Similar findings across country groups strengthen the baseline results, but different ones do not necessarily undermine them. We cannot determine whether a country specific credit gap history is due to country specific economic conditions or rather to an idiosyncratic coincidence that can easily occur in other countries in the future as well. If the former is more likely, all countries should focus on the experiences of its own country group. Unfortunately from a Hungarian perspective, this robustness check cannot be conducted reliably for the Central and Eastern European countries, due to the short Czech, Polish and Lithuanian time series.

We find a significant heterogeneity in the results across country groups (Table 7). Credit gaps are much worse predictors of systemic banking crises in core EU countries than in Nordic countries and especially in Mediterranean countries. The highest AUROC values are 0.63, 0.83 and 0.88, respectively. The first is only moderately higher than 0.5 belonging to completely uninformative indicators, therefore the ranking of the 48 credit gaps is even less clear in core EU countries than in the baseline analysis. AUROC values of more than half of the credit gaps are not different from the highest value at the 1 percent significance level, and even at the 10 percent level almost a quarter of credit gaps are like that. By contrast, Nordic and Mediterranean countries have a more distinct and relatively small group of top credit gaps with much higher AUROC values. A potential reason behind these differences is that credit-to-GDP time series vary less over time in core EU countries than in the other two country groups.

The finding in the baseline analysis that narrow credit yields credit gaps with at least as good predictive power than broad credit is mostly due to the observations from Nordic countries and partly from core EU countries. The two definitions of credit usually fare similarly in Mediterranean countries. Our second main finding in the baseline analysis that the 3-year credit-to-GDP forecast is useless is true for Nordic countries (with the exception of a few broad credit gaps), and less sharply for Mediterranean countries, and even less sharply and with a few exceptions in core EU countries.

None of the top three credit gaps from the baseline analysis reach the highest AUROC value in any country group, but the HP filter and CF filter credit gaps are among the best few credit gaps in two country groups each. This level of performance is achieved only by credit gaps with close specifications (HP filter with medium-term cycle, and CF filter with long cycle), and no other credit gap outperforms them. The wavelet filter credit gap is not among the best specifications in any country

[34] Lang et al. (2019) assign a two-thirds weight in their final evaluation to the results from the in-sample early warning exercise, and one-third to the out-of-sample results, similarly to the approach used here.

group. An interesting finding is that in Nordic countries, shorter cycles perform generally better than longer ones, unlike in Mediterranean countries, where the opposite is true.

We also rank the credit gaps in each country group with the RU instead of the AUROC, assuming balanced preferences. The results are very similar to those presented here. See Appendix E for more details.

Table 7
AUROC values of credit gaps by country groups

Using all crisis periods, we expect "Signal" in 16-5 quarters before crises.			**Gap calculation**							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
Core EU countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.62**	**0.62**	0.60	**0.62**	**0.63**	0.61	0.59	0.59
		1-year forecast	**0.62**	**0.61**	0.59	0.59	0.59	0.57	0.59	0.59
		3-year forecast	0.60	0.59	0.58	0.61	0.62	0.59	0.59	0.57
	Broad definition of credit stock	0-year forecast	0.61	0.59	0.55	**0.63**	0.61	0.58	0.57	0.56
		1-year forecast	0.61	0.59	0.55	0.60	0.59	0.57	0.58	0.56
		3-year forecast	0.59	0.58	0.55	0.55	0.55	0.54	0.56	0.54
Nordic countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.80	**0.83**	**0.83**	0.77	**0.81**	**0.83**	0.74	0.79
		1-year forecast	0.79	0.80	0.80	0.64	0.70	0.69	0.76	0.80
		3-year forecast	0.65	0.61	0.54	0.51	0.55	0.55	0.70	0.55
	Broad definition of credit stock	0-year forecast	0.76	0.77	0.77	0.73	0.77	0.77	0.72	0.73
		1-year forecast	0.74	0.74	0.72	0.57	0.60	0.58	0.73	0.74
		3-year forecast	0.50	0.55	0.61	0.67	0.65	0.66	0.64	0.62
Mediterranean countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.87**	0.82	0.72	**0.88**	0.84	0.77	0.83	0.83
		1-year forecast	0.87	0.81	0.71	0.85	0.76	0.68	0.83	0.83
		3-year forecast	0.85	0.77	0.65	0.74	0.63	0.58	0.83	0.71
	Broad definition of credit stock	0-year forecast	0.87	0.83	0.72	0.87	0.86	0.75	0.84	0.84
		1-year forecast	0.87	0.83	0.73	0.86	0.81	0.70	0.84	0.83
		3-year forecast	0.85	0.81	0.72	0.81	0.73	0.61	0.83	0.75

Note: Red cells show the highest AUROC values of various country groups. Orange cells contain values that do not differ from the highest values at the 10 percent significance level. Yellow and green cells indicate additional AUROC values that do not differ from the highest AUROC value at at the 5, and 1 percent significance levels, respectively. Credit gaps belonging to cells with bold numbers have AUROC values calculated with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters that do not differ significantly from the respective highest values at the 10 percent significance level. Tests of differences in AUROC values are based on Delong et al. (1988). The test statistics verify the identical nature of the ROC curves, rather than the AUROC values directly, and they also take into account sample size, so credit gaps with the same AUROC value may differ from the highest AUROC value at marginally different significance levels. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

4.3.3 Time periods

More than half of the crisis periods (15 out of 26) are related to the global financial crisis. In this subsection, we explore how the baseline results depend on the observations around the global financial crisis. We derive the credit gaps just like in the baseline analysis, but we evaluate them with the AUROC separately on two subsamples spanning from 1960 Q4 to 1997 Q4 and from 1998 Q1 to 2019 Q1. This division localizes the crises related to the global financial crisis together with their prediction periods into the second subsample. Similar findings across subsamples strengthen baseline results, but different findings undermine them only if one of the subsamples definitely holds a more relevant experience for the future. This condition is not necessarily met in the case of the global financial crisis, as it can be regarded as a unique event from some aspects.

According to Table 8, our first two main findings hold true for the subsamples as well, with some minor restrictions (Table 8). First, credit gaps with narrow credit perform significantly better than their broad credit versions in the first subsample, and they perform similarly in the second. Second, credit gaps calculated with a 3-year credit-to-GDP forecast prove to have a clearly worse early warning accuracy in both subsamples. The results pertaining to the trend-cycle decomposition procedures differ more in the two subsamples. The most informative early warning signals in the subsample lasting until 1997 are issued by credit gaps calculated with narrow credit, no credit-to-GDP forecast, HP or CF filters, and various cycle lengths. In contrast, credit gaps calculated with wavelet filter and long cycles predict the global financial crisis most accurately. The performance of these credit gaps with HP and CF filter differs more between the two subsamples than that of the mentioned credit gaps with wavelet filter. This also means that the global financial crisis can be less precisely predicted (highest AUROC: 0.71) than the earlier ones (highest AUROC: 0.81). The more even performance of the wavelet filter is probably due to the fact that it is more robust to structural changes in the credit-to-GDP time series than the other two filters.

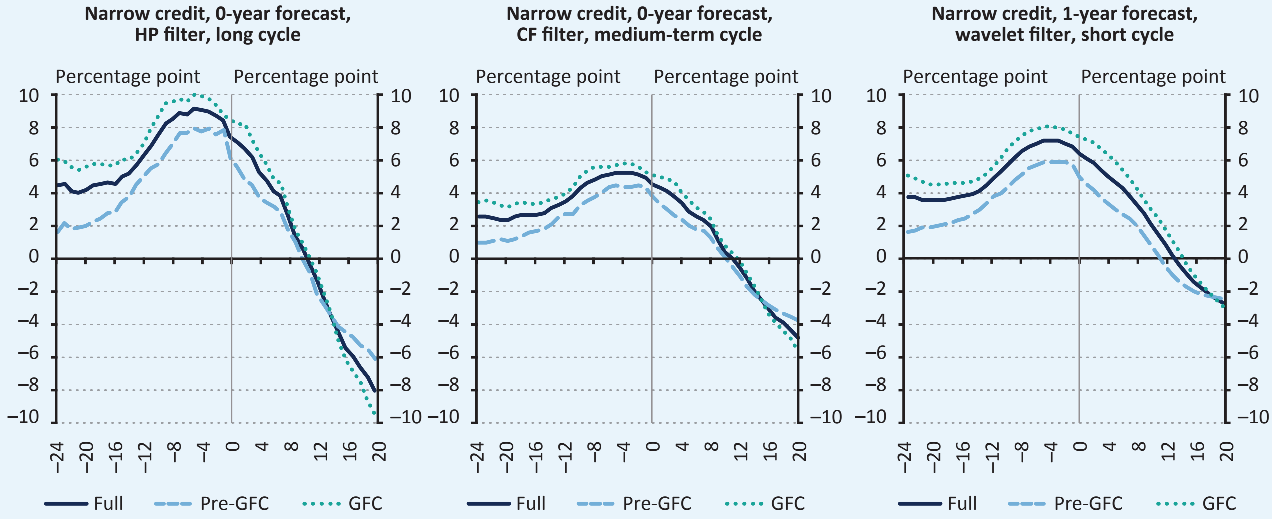
Table 8
AUROC values of credit gaps by different periods

Using data from all countries, we expect "Signal" in 16-5 quarters before crises.			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
1960-1997										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.81**	**0.80**	**0.80**	0.79	0.79	0.79	0.69	0.75
		1-year forecast	0.79	0.79	0.77	0.68	0.67	0.65	0.71	0.76
		3-year forecast	0.68	0.65	0.60	0.59	0.60	0.57	0.66	0.59
	Broad definition of credit stock	0-year forecast	0.77	0.77	0.77	0.75	0.75	0.75	0.64	0.69
		1-year forecast	0.75	0.75	0.73	0.62	0.62	0.58	0.65	0.70
		3-year forecast	0.59	0.55	0.49	0.53	0.53	0.57	0.59	0.53
1998-2019										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.66	0.65	0.63	0.65	0.66	0.65	**0.70**	0.68
		1-year forecast	0.64	0.62	0.58	0.60	0.59	0.57	**0.71**	0.67
		3-year forecast	0.61	0.58	0.55	0.56	0.56	0.56	0.69	0.57
	Broad definition of credit stock	0-year forecast	0.65	0.62	0.58	0.64	0.65	0.62	0.70	0.67
		1-year forecast	0.65	0.64	0.62	0.60	0.61	0.60	**0.71**	0.69
		3-year forecast	0.59	0.55	0.49	0.53	0.53	0.57	0.59	0.53

Note: Red cells show the highest AUROC values of various periods. Orange cells contain values that do not differ from the highest values at the 10 percent significance level. Yellow cells indicate additional AUROC values that do not differ from the highest AUROC value at the 5 percent significance level. Credit gaps belonging to cells with bold numbers have AUROC values calculated with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters that do not differ significantly from the respective highest values at the 10 percent significance level. Tests of differences in AUROC values are based on Delong et al. (1988). The test statistics verify the identical nature of the ROC curves, rather than the AUROC values directly, and they also take into account sample size, so credit gaps with the same AUROC value may differ from the highest AUROC value at marginally different significance levels. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

The three best credit gaps in the baseline analysis are less good at predicting the global financial crisis than the earlier crises. This is reflected in Figure 6 demonstrating that even if the values of these credit gaps are strikingly high before the global financial crisis, they are less distinct from the values of earlier quarters compared to the case of the crises in the first subsample. Only the HP filter credit gap is among the best credit gaps, and only in the first subsample. However, all three credit gaps prove to be the best in both subsamples among the other credit gaps with HP, CF and wavelet filter, respectively. There is only one exception, the best credit gap in the second subsample, which is a credit gap with wavelet filter, but it differs from the wavelet filter credit gap.

Figure 6
Development of the three best early warning credit gaps around different systemic banking crises



Note: Horizontal axes show quarters, 0th point in time is the quarter when systemic banking crisis begins. The charts show averages of credit gaps with a given specification over countries and banking crises within the given subsample. The subsample 'Pre-GFC' contains observations from 1960 Q4 until 1997 Q4, while the subsample 'GFC' (global financial crisis) includes those between 1998 Q1 and 2016 Q1.

4.3.4 ARIMA forecasts

There are at least two reasons for conducting a robustness check using alternative credit-to-GDP forecasts. First, with a 'perfect' forecast we disregard forecast errors. Second, in the few European countries that employs credit-to-GDP forecasts for calculating credit gaps, simple methods are implemented rather than the best available expert estimates.[35] Following this practice, for all countries and all quarters except for the first 8 years[36] of every credit-to-GDP time series we estimate the best fitting ARIMA model for the credit-to-GDP time series lasting until the given quarters, using the Bayesian information criterion. The estimated model is employed to extend the credit-to-GDP time series with 1-year, 3-year and 5-year forecasts. Credit gap values for the given quarters are calculated with these extended time series. Unlike in the baseline analysis, 5-year projections are also prepared in line with the prevailing European practice.

Table 9 displays baseline results from Table 4 supplemented by credit gap specifications using ARIMA forecasts. With these new credit gaps, it also remains true with a minimal number of exceptions that a longer-term forecast decreases the AUROC value of the respective credit gap. However, in the case of the ARIMA forecasts, this deterioration in AUROC values is more limited. As a consequence, the practice followed by a few macroprudential authorities leads to suboptimal credit gaps from an early warning perspective based on pooled data from European countries. Credit gaps calculated with ARIMA forecasts usually entail higher AUROC values than their respective versions with 'perfect' forecasts. This allows only one new credit gap, the 1-year ARIMA forecast variant of the best HP filter credit gap, to match the performance of the three best credit gaps in the baseline analysis.

[35] In Norway and Lithuania, some simple average of the credit-to-GDP (with broad credit) observations from the last 4 quarters is projected 20 quarters ahead. Norway uses an arithmetic mean (Gerdrup et al., 2013), whereas Lithuania uses a weighted average, with weights of 0.4, 0.3, 0.2 and 0.1 employed to the observations, starting from the most recent one (Valinskytė and Rupeika, 2015). In Portugal, the credit-to-GDP time series calculated with broad credit is extended by 28 quarters, by using an ARIMA(p,1,0) forecast. The value of optimal lag order *p* is determined recursively until 2015 Q1, and it is set to 3 quarters from 2015 Q2, which is the optimal value when the time series lasting until 2015 Q1 is used (Banco de Portugal, 2015).

[36] This was warranted by the fact that the early warning signals are not evaluated anyway in the first 8 years of the credit-to-GDP time series.

Table 9
AUROC values of credit gaps calculated with various credit-to-GDP forecasts

Using all crisis periods from all countries, we expect "Signal" in 16-5 quarters before crises.			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
Narrow definition of credit stock	Forecast: with real data	0-year forecast	**0.74**	0.74	0.71	0.73	**0.74**	0.73	0.73	0.74
		1-year forecast	0.74	0.73	0.70	0.67	0.67	0.65	0.74	**0.74**
		3-year forecast	0.67	0.64	0.58	0.59	0.59	0.57	0.71	0.60
	Forecast: with ARIMA model	1-year forecast	**0.74**	0.74	0.71	0.72	0.73	0.71	0.72	0.73
		3-year forecast	0.73	0.72	0.71	0.69	0.70	0.68	0.72	0.70
		5-year forecast	0.69	0.69	0.69	0.68	0.69	0.67	0.70	0.66
Broad definition of credit stock	Forecast: with real data	0-year forecast	0.74	0.72	0.68	0.73	0.73	0.70	0.72	0.72
		1-year forecast	0.73	0.71	0.67	0.65	0.65	0.60	0.73	0.72
		3-year forecast	0.62	0.59	0.53	0.53	0.52	0.52	0.68	0.54
	Forecast: with ARIMA model	1-year forecast	0.73	0.71	0.68	0.71	0.71	0.69	0.70	0.70
		3-year forecast	0.72	0.71	0.70	0.69	0.70	0.67	0.71	0.68
		5-year forecast	0.70	0.70	0.70	0.69	0.69	0.67	0.68	0.66

Note: The red cell shows the highest AUROC value. Orange cells contain values that do not differ from the highest value at the 10 percent significance level. Yellow and green cells indicate additional AUROC values that do not differ from the highest AUROC value at at the 5, and 1 percent significance levels, respectively. Credit gaps belonging to cells with bold numbers have AUROC values calculated with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters that do not differ significantly from the respective highest values at the 10 percent significance level. Tests of differences in AUROC values are based on Delong et al. (1988). The test statistics verify the identical nature of the ROC curves, rather than the AUROC values directly, and they also take into account sample size, so credit gaps with the same AUROC value may differ from the highest AUROC value at marginally different significance levels. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

4.4 NOTES ON PRACTICAL APPLICATIONS

Each of the three best credit gaps identified above issues significantly more informative early warning signals of systemic banking crises than the Basel gap in the baseline analysis and in most robustness checks.[37] While the AUROC value of the Basel gap is 0.74 in all three prediction horizons, only slightly lower than the values produced by the three best credit gaps (Table 4), there are considerable differences regarding RU values (Table 5). It is able to match the performance of the three best credit gaps in core EU countries and Mediterranean countries (although there are no large differences between the various credit gaps in core EU countries to begin with) (Table 7). When assessed on the subsample of 1998–2019, the Basel gap is only marginally worse than them (Table 8).[38]

The additional credit-to-GDP gap[39] used by several European macroprudential authorities (including the Hungarian one) (Table 1) cannot be clearly outperformed by any other credit gaps. Yet, two others are identified that matched its early warning performance. In this subsection, we argue that all these three credit gaps should be included among the indicators monitoring the build-up of cyclical systemic risks, on account of their different characteristics.

According to the AUROC evaluation statistic, HP filter and CF filter credit gaps perform better with longer prediction horizons, while the wavelet filter credit gap is better with shorter horizons (Figure 5). According to the RU evaluation statistic with balanced preferences over Type I and Type II error rates, the relative rank of the wavelet filter credit gap is worse in all prediction horizons (Table 10). The RU values of the HP filter credit gap and especially the CF filter credit gap surpass the RU values of the wavelet filter credit gap even in shorter prediction horizons. The CF filter credit gap not only has the highest RU values always but also the lowest Type I error rates (missed crisis signals) almost always. The

[37] The Basel credit gap is a one-sided HP filter gap using broad credit, without a credit-to-GDP forecast and with a lambda of 400,000.

[38] These results are consistent with the current stance of the literature, namely that the Basel gap is not the most accurate early warning indicator of financial crises, although it is still one of the best. See the introduction for the references.

[39] Credit gap calculated with narrow credit, no credit-to-GDP forecast and one-sided HP filter with a lambda of 400,000.

values around 30 percent mean that the CF filter credit gap does not signal an impending crisis in around three out of ten quarters in the prediction period. In accordance with this, in the 20th, 16th, and 12th quarters prior to the systemic banking crises, it signals the coming crisis 61, 65, and 74 percent of the time, respectively (crisis signal ratio). These ratios are also higher than the values of the other two credit gaps in all three quarters. It is somewhat surprising in the case of the HP filter credit gap that Type I error rate does not decline and crisis signal ratio does not rise monotonically as the crisis nears, unlike in the case of the other two credit gaps. The HP filter credit gap steadily produces the lowest Type II error rates (false positive crisis signals), at about 25 percent, although it always has the highest Type I error rate, too. The former means that the HP filter credit gap (falsely) signals an impending crisis in 25 percent of the quarters outside the prediction periods (and crisis periods). The wavelet filter credit gap always has the highest Type II error rate, while it proves to be average along the rest of the performance measures. Overall, the three credit gaps predict systemic banking crises with the same degree of accuracy, albeit somewhat differently from each other.

Table 10
In-sample early warning properties of the best three early warning credit gaps

Credit-to-GDP gap specification				Prediction horizon of 20-5 quarters					Prediction horizon of 16-5 quarters					Prediction horizon of 12-5 quarters				
				AUROC	in-sample RU	Type I error rate	Type II error rate	Crisis predic-tion	AUROC	in-sample RU	Type I error rate	Type II error rate	Crisis predic-tion	AUROC	in-sample RU	Type I error rate	Type II error rate	Crisis predic-tion
Narrow credit	0-year forecast	HP filter	30-year cycle	0.74	0.41	35%	25%	61%	0.74	0.42	36%	23%	52%	0.75	0.43	32%	25%	61%
Narrow credit	0-year forecast	CF filter	24-year cycle	0.74	0.45	31%	26%	61%	0.74	0.46	28%	27%	65%	0.75	0.44	28%	29%	74%
Narrow credit	1-year forecast	Wavelet filter	16-year cycle	0.73	0.39	34%	27%	57%	0.74	0.40	31%	29%	57%	0.76	0.42	27%	30%	70%

Note: Values other than AUROCs are calculated assuming balanced preferences over Type I and Type II error rates. (The value of the preference parameter θ is 0.5.) Type I and Type II error rates are determined at the optimal threshold value. Crisis predictions indicate shares of 'Signal' values in the 20th, 16th and 12th quarters prior to the onset of banking crises.

The simultaneous monitoring of the somewhat different top three credit gaps raises the question whether it is worth aggregating the three (or even more) credit gaps into a single indicator, and if so, which method is suitable for that. The detailed analysis of this issue falls outside the scope of this study. However, the first part can be confirmed, because even two simple aggregate indicators can issue at least as good early warning signals as the three individual credit gaps. One of them is the arithmetic mean of the credit gaps, and the other is the arithmetic mean of the normalized versions of the credit gaps, where normalization means dividing with the standard deviations (Table 11). Despite its simplicity, the arithmetic mean is one of the best aggregating methods from the perspective of forecasting performance. The arithmetic mean of normalized credit gaps can perform well because the simple arithmetic mean is influenced more by a credit gap with a larger standard deviation, but when normalized values are used, all three indicators contribute equally. This method produces a slightly better early warning indicator than simple averaging according to Table 11.

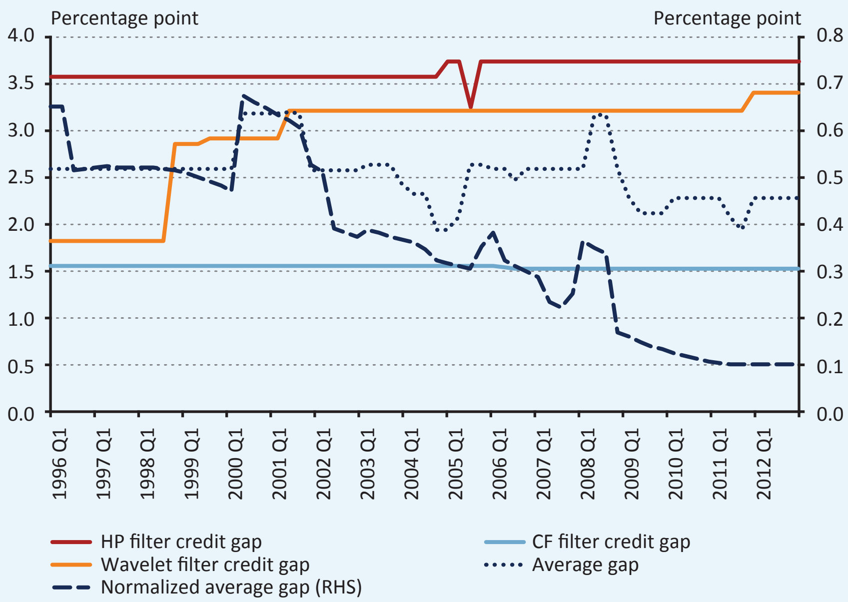
Table 11
Early warning performance of the best three credit gaps and their averages

Credit-to-GDP gap specification				AUROC			In-sample RU			Out-of-simple RU	
				pred. hor. 20-5 q.	pred. hor. 16-5 q.	pred. hor. 12-5 q.	pred. hor. 20-5 q.	pred. hor. 16-5 q.	pred. hor. 12-5 q.	pred. hor. 16-5 q.	pred. hor. 12-5 q.
Narrow credit	0-year forecast	HP filter	30-year cycle	0.74	0.74	0.75	0.41	0.42	0.43	-0.01	-0.01
Narrow credit	0-year forecast	CF filter	24-year cycle	0.74	0.74	0.75	0.45	0.46	0.44	0.26	0.14
Narrow credit	1-year forecast	Wavelet filter	16-year cycle	0.73	0.74	0.76	0.39	0.40	0.42	0.15	0.17
Average of the three credit gaps				0.75	0.75	0.76	0.43	0.43	0.44	0.06	0.02
Average of the three normalized credit gaps				0.75	0.75	0.76	0.44	0.45	0.45	0.15	0.26

Note: Red cells show the highest values for various prediction horizons.

In practice, monitoring of credit gaps is more effective if the corresponding optimal threshold values are stable over time. Figure 7 shows that the threshold values optimized with balanced preferences are not sensitive to the length of the data sample in the case of the HP filter and the CF filter versions of the top three credit gaps. The threshold value for the wavelet filter credit gap rises monotonically, which is not surprising, as credit gaps typically soared higher before the global financial crisis than prior to earlier crises (see Figure 6.). The threshold values and their stable order over time are also consistent with the typical pre-crisis credit gap values seen in Figure 6. Yet the thresholds of the two aggregated indicators are less stable over time.

Figure 7
Robustness of the optimal thresholds belonging to the best credit gaps



Note: The horizontal axis shows the last quarter of the sample of credit gap signals used for optimizing the threshold values. Optimization is performed with a prediction horizon of 16–5 quarters and balanced preferences (a preference parameter of 0.5). The three individual credit gaps are the best three early warning credit gaps: (narrow credit, 0-year forecast, HP filter, long cycle), (narrow credit, 0-year forecast, CF filter, medium-term cycle), (narrow credit, 1-year forecast, wavelet filter, short cycle). The other two indicators are the arithmetic average of these three and the arithmetic average of the normalized versions of these three (normalization means dividing by their own standard deviation).

5 Comovement with the financial cycle

From a macroprudential policy perspective, the most important requirement of one-sided credit gaps is that they should be as reliable in predicting systemic banking crises as possible. In addition, the calibration and communication of macroprudential interventions is also facilitated if credit gaps are able to capture the current position of the credit cycle. The two requirements are not the same, as early warning simply demands that credit gaps take values higher than an appropriate threshold before crises, and otherwise stay under it. This is a necessary but not a sufficient condition of accurately characterizing the credit cycle. We quantify the degree of comovement of the 48 credit gaps with the credit cycle and the financial cycle in general in two ways. First, we estimate the endpoint uncertainty of the individual credit gaps. Second, we calculate correlations between credit gaps and aggregated indicators of the financial cycle published by national macroprudential authorities.

5.1 ENDPOINT UNCERTAINTY

Endpoint uncertainty refers to the uncertainty of estimated values of the credit gap caused by the absence of the subsequent, currently unobservable credit-to-GDP values.[40] The credit-to-GDP values that become observable over time can continuously refine the gap value estimates, which thus converges to a certain limit. This value deduced from an appropriately large number of observations and with a two-sided filter is considered the most accurate credit gap estimate. The available credit-to-GDP time series are too short to examine how well these credit gap estimates describe the credit cycle. We work on the assumption that the further away the credit gaps calculated with the one-sided approach are from the 'most accurately' estimated values, the more misleading they are about the current position of the credit cycle. The larger the difference, the greater the chance that the one-sided credit gap has not only a false value but also the wrong sign. In this paper, those credit gaps are considered to reflect the credit cycle more accurately whose values are revised less over time due to the observations in subsequent quarters.

We examine only 16 of the 48 credit gaps for the sake of easier comparability. The characteristics deemed not useful for early warning purposes in Section 4 are disregarded, in particular broad credit and 3-year credit-to-GDP forecast. We use the data from as long a period as possible where observations for all the countries are available and the planned analysis can be performed. Therefore, the revisions in the credit gap values between 1983 Q1 and 2008 Q1 due to subsequent data are evaluated.[41] This is a period of 101 quarters (over 25 years), covering 8 complete crises out of the 11 preceding the global financial crisis. The period is long enough to contain complete credit cycles. However, we have to disregard the short time series of three Central and Eastern European countries (Czech Republic, Poland, Lithuania), and focus on the data from the remaining 15 countries.

The average revision due to new data is calculated as follows. Having computed the value of a one-sided credit gap in a particular country and quarter, we recalculate this value taking also into account the credit-to-GDP values of the 4 following quarters.[42] We take the absolute value of the difference of these two values, and then we divide the result by

[40] In the case of two-sided gaps, this mostly concerns the credit gap values near the end of the time series, hence the name.

[41] Three aspects have to be simultaneously considered when choosing this subsample. Credit-to-GDP observations going back far enough before the subsample are needed to make the credit gaps calculated with a two-sided filter accurate enough within the chosen subsample. The subsample itself should cover as long a period as possible because in that case the revisions to the one-sided credit gaps can be observed along at least one full credit cycle. Finally, at the end of the credit-to-GDP time series, a sufficient number of observations are needed to calculate revisions at the end of the subsample.

[42] In the case of credit gaps with a 1-year credit-to-GDP forecast, the additional data of the 4 quarters are taken into account as follows. The 1-year forecast is replaced with actual data (or itself, because 'perfect' forecast is used), and a 1-year forecast (containing the actual subsequent data, because 'perfect' forecast is used) is placed after it.

the average over the 101 quarters of the absolute values of the credit gaps calculated with the two-sided version of the method concerned, on the entire credit-to-GDP time series. So, the revision is measured relative to the average extent of the cyclical position in the given country. Finally, the derived ratios are averaged over the 101 quarters and the countries, to arrive at the measure of 1-year revision for the given credit gap specification. The values for the 2-, 3- and 10-year revisions are produced in a similar manner.

Table 12 and Figure 8 contain the results. Data of one additional year bring about nearly the same revisions in credit gaps with 1-year credit-to-GDP forecast compared to the same specifications without the forecast, but credit gaps with the forecast are clearly revised less when more data are added. This is probably because the 'limit' of both versions of credit gaps is the same two-sided credit gap, but the version with the credit-to-GDP forecast is further along in 'convergence'.[43] Our best approximation of the long-term revisions is derived from considering the most additional data. Based on these 10-year revisions, the credit gaps with 1-year credit-to-GDP forecast, CF filter, long- and medium-term cycles exhibit the lowest endpoint uncertainty. Among the top three early warning credit gaps, only the CF filter version can come close to this. The HP filter credit gap has a moderate level of endpoint uncertainty, while the wavelet filter credit gap has one of the highest, which is twice as large as the best value. The 10-year revisions are lower bounds of the full revisions, but even these values are significant, as they vary between two-thirds and one and a half times the extent of the respective typical two-sided cyclical positions.[44]

Typically, one-quarter or one-third of the 10-year revisions occur within a year, and two-thirds happen after the first three years. This suggests that the 10-year revisions are good approximations of the long-term full revisions. At any stage of additional data considered, credit gaps with CF filter tend to produce the smallest revisions compared to similar specifications with other filters. Likewise, longer cycles entail smaller revisions. An average 20–50 percent revision in credit gaps over a year highlights that in practice, the perception about the exact position of the credit cycle can change considerably even in the short run. This requires the related macroprudential measures to be carefully considered and communicated.

Table 12
Average change in credit gaps due to 1, 2, 3 and 10 years of new data (percent)

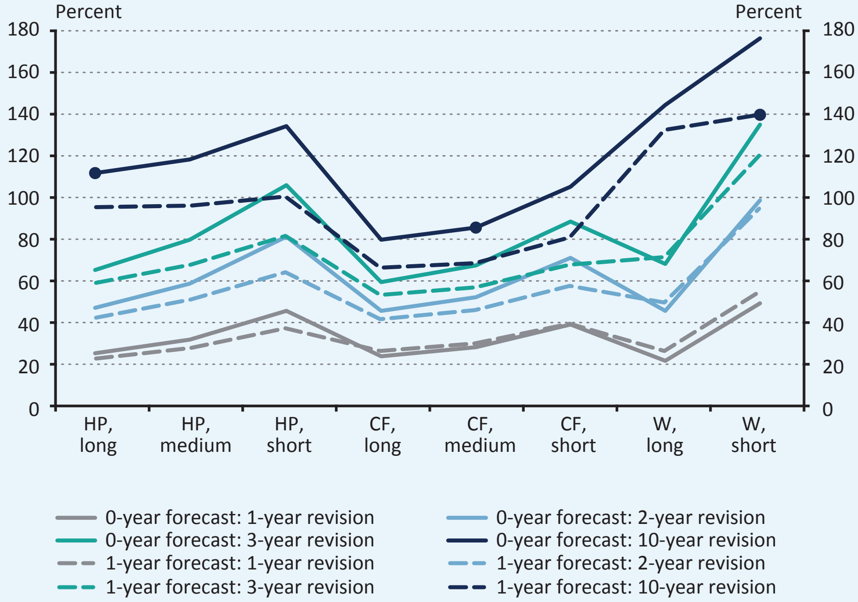
Using data between 1983 Q1 and 2008 Q1 from all countries			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
Narrow definition of credit stock	0-year credit-to-GDP forecast	1-year revision	25	32	46	24	28	39	21	49
		2-year revision	47	59	81	45	52	71	45	99
		3-year revision	65	80	106	59	67	88	68	136
		10-year revision	112	118	135	80	85	105	145	177
	1-year credit-to-GDP forecast	1-year revision	23	28	37	26	30	39	26	54
		2-year revision	42	51	64	42	46	58	50	95
		3-year revision	59	68	82	53	57	68	71	121
		10-year revision	96	96	101	66	69	81	133	140

Note: Red numbers denote values of the smallest average revisions with a given number of additional observations used. Yellow numbers denote the same, if only the three credit gaps are compared to each other that proved to have the best early warning properties using European data in Section 4. Cells with thick borders refer to the best three early warning credit gaps.

[43] Due to the same reason, the endpoint uncertainties of the credit gaps using a 3-year credit-to-GDP forecast are even lower than those of the credit gaps using 1-year forecast and otherwise having the same specifications.

[44] This is consistent with the results of Edge and Meisenzahl (2011).

Figure 8
Average change in credit gaps due to 1, 2, 3 and 10 years of new data (percent)



Note: Highlighted points denote 10-year revisions of the credit gaps that proved to have the best early warning properties using European data in Section 4.

The above findings regarding 10-year revisions also hold true, with small differences, for Nordic, Mediterranean and core EU countries separately. (See Appendix F for more details.)

5.2 CORRELATION BETWEEN CREDIT-TO-GDP GAPS AND FINANCIAL CYCLE INDICATORS

We calculate correlations between different credit gaps and aggregated indicators of the financial cycle for each country where macroprudential authorities publish such indicators, and then we average these values over countries. We could find 11 such indicators from 9 countries (for more details, see Appendix A, Table 19).[45] This method has a number of shortcomings. First, only half of the countries have a financial cycle indicator, and these are typically much shorter than the credit gap time series. Second, methodologies of the financial cycle indicators vary widely across countries. Finally, some financial cycle indicators use credit-to-GDP gaps as an input, which unjustifiably favors the credit-to-GDP gaps whose methodology is close to those used as input.

According to Table 13, all the top three early warning credit gaps are among the indicators with the highest correlations, and the wavelet credit gap has the highest value out of the 48 credit gaps.[46] Broad credit and longer credit-to-GDP forecasts perform poorly not only in early warning evaluated in Section 4 but also in the correlation with financial cycle indicators.

[45] The countries are taken into account with the same weight, irrespective of the number of indicators they have.

[46] The latter is noteworthy because even if the calculation of some financial cycle indicators uses a credit-to-GDP gap, none of these gaps are computed with a wavelet filter.

Table 13
Average correlations between different credit gaps and financial cycle indicators (percent)

Using all financial cycle indicators from all countries			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	69	70	67	64	67	65	54	69
		1-year forecast	67	68	66	49	54	53	68	74
		3-year forecast	57	56	52	40	44	44	30	32
	Broad definition of credit stock	0-year forecast	58	59	57	55	57	54	45	56
		1-year forecast	55	56	54	38	42	37	56	60
		3-year forecast	43	40	34	28	30	25	19	21

Note: The red cell shows the highest average correlation, while grey cells have values that are no more than 10 percent lower than that. Cells with thick borders refer to the best three early warning credit gaps based on the analysis with European data in Section 4.

6 Early warning performance on simulated Hungarian data

The examination of European data does not identify the single best credit-to-GDP gap, and different country groups seem to have different credit gap specifications with the best prediction power. In this section, we use only Hungarian observations to find a more precise identification of the best credit gap for Hungary. We extend the short credit-to-GDP time series with simulated future values under various realistic scenarios. The credit gaps are ranked according to the accuracy of their warning signals indicating excessive lending periods built into the simulations.

6.1 DATA: ARIMA SIMULATIONS

Analyses with European data in the previous sections show that the credit gaps with narrow credit are better early warning indicators and show the position of the financial cycle more accurately than the credit gaps with broad credit. Therefore, we produce the Hungarian simulated data based on credit-to-GDP observations with narrow credit. The actual data are extended first with the most reliable forecast as long as it is possible, and the simulated data of various credit-to-GDP projections are only added after that. We use exchange rate adjusted[47] credit stocks and nominal GDP seasonally adjusted by the Magyar Nemzeti Bank (MNB), because the credit-to-GDP calculated with these elements is a common specification in Hungary and it is the only one with a reliable forecast. The observations span from 1995 Q1 until 2019 Q3,[48] the forecasts[49] last from 2019 Q4 until 2022 Q2, and the simulated data are from 2022 Q3 until 2037 Q2. The Hungarian financial cycle was in the middle of the recovery phase in 2019, and cyclical financial systemic risks were still low.[50] It will probably take many years until a future excessive lending phase is reached, if at all. The simulation seeks to cover the entire expansion phase of lending, but further projections of the credit-to-GDP time series require handling too complex scenarios,[51] so this is not performed.

All the realistic future paths of credit-to-GDP are sought to be included among the simulated data series. First, a Monte Carlo simulation produces data series with ARIMA model parameters estimated on the actual and forecasted values of credit-to-GDP. Then, we apply adjustment on these data series with the assumed extent of financial deepening and excessive lending. 15 year-long data are generated, because in such a long time span, even late and protracted periods of excessive lending can be accommodated. The simulations are done along four scenarios (Table 14). To produce a simulated data series for a given scenario, first a random choice is made from the possible sets of parameters enabled by a scenario. The data series is generated through the Monte Carlo simulation, the financial deepening adjustment, and the excessive lending adjustment. We prepare 200 such simulated data series for all scenarios. All simulated time series contain exactly one future excessive lending period, although their characteristics can differ significantly from each other.

[47] FX denominated lending has been a major factor in corporate lending in the last two decades, and in household lending until the global financial crisis as well.

[48] The observations start much later than the BIS data series used in Section 4, where the first values are from 1970 Q4.

[49] The credit-to-GDP forecast uses data from the MNB's Financial Stability Report and Inflation Report from December 2019 (MNB, 2019a; MNB 2019b). This is when the MNB's last macroeconomic forecast was published before the coronavirus pandemic. The forecast and the simulated data describe potential future credit-to-GDP rates that the economy is likely to return to in the medium term, as the coronavirus pandemic abates. Since realistic, future periods of excessive lending in the medium and long run need to be taken into account, disregarding the present pandemic is considered acceptable.

[50] See, for example, the MNB's Macroprudential Report from 2020.

[51] For example, the starting date, severity and duration of the next financial crisis and the speed and duration of the subsequent recovery have to be accounted for.

The differences between scenarios are in (1) whether the shocks have the same or greater standard deviation than the estimated one in the ARIMA model, (2) whether financial deepening is assumed and (3) the extent of the excessive lending considered. The credit-to-GDP time series is an I(1) process at all usual significance levels. We estimate an ARIMA(1,1,1) model including a constant because choosing one AR and one MA term is optimal according to the Akaike and Bayesian information criteria. The estimated variance of the error term in the ARIMA model is lower than those for euro area countries, therefore it makes sense to consider different values during the simulations. Hungarian credit-to-GDP ratios are much lower than euro area figures, thus the Hungarian ratio is expected to rise, as part of economic convergence without overheating, so financial deepening is a justified assumption. A good early warning credit gap is expected to distinguish between overheating and convergence. We take into account financial deepening by adding extra 0.25 percentage points to simulated credit-to-GDP data in all quarters, so an annual convergence of 1 percentage point is assumed. The excessive lending periods can vary along three characteristics: starting date, length, and extent. The possible values of these parameters are determined based on the European data used in previous sections. Periods of excessive credit-to-GDP growth usually last for 4–8 years, and the credit-to-GDP ratio rises by an average of 0.5–2 percentage points in each quarter prior to systemic banking crises. So, parameter sets of the various scenarios are chosen from these two ranges. The period of excessive credit growth does not always start right at the beginning of the simulation, the simulated data series are adjusted after 0, 2 or 4 calm years. Time series include only the excessive lending period without any subsequent crisis episode.

Table 14
Parameters of the different simulation scenarios

Scenario	ARIMA-simulation		Adjustment: financial deepening	Adjustment: excessive credit growth		
	Period	Standard deviation of shocks used in simulation	Quarterly value	Starting date	Length	Quarterly value
		multiples of the standard deviation of ARIMA error term	*quarterly change in credit-to-GDP, percentage point*	*number of quarters after 2022 Q3*	*number of quarters*	*quarterly change in credit-to-GDP, percentage point*
Baseline	2022 Q3 - 2037 Q2	1	0,25	0 8 16	16 20 24	0.5 1.0 1.5
Baseline with higher uncertainties	2022 Q3 - 2037 Q2	2	0,25	0 8 16	16 20 24	0.5 1.0 1.5
Lower systemic risk	2022 Q3 - 2037 Q2	1	0	0 8 16	16 20 24	0.5 1.0
Higher systemic risk	2022 Q3 - 2037 Q2	1	0	0 8 16	20 24 28 32	0.5 1.0 1.5 2.0

The 'Baseline' scenario contains the most likely parameters regarding the long-term evolution of Hungarian credit-to-GDP ratios. This includes financial deepening, in other words a steadily increasing trend in credit-to-GDP ratios, without the rise in systemic risks.[52] It is plausible that the financial stability framework strengthened after the global financial crisis is able to prevent the protracted and significant increases in indebtedness that was only seen in the 2000s, and only in southern European countries according to the data used in Section 4. At the same time, the scenario does not preclude the possibility of Hungarian data running on a path seen before the global financial crisis.[53] The Baseline scenario also

[52] The euro area credit-to-GDP calculated with narrow credit was 89 percent at the end of 2018, 100 percent at the end of 2007 and 94 percent on average between 1999 and 2018, according to the BIS database also used in Section 4. We presume that it is not unrealistic that the credit-to-GDP ratio in Hungary would rise from 41 to 56 percent during the 15 years of the simulation period, solely on account of financial deepening.

[53] This does not mean that a repeat of the rise in *cyclical systemic risks* as seen before the crisis is deemed possible in the Baseline scenario. Prior to the crisis, systemic risks were considerably increased even by those excessive lending factors that are measured inaccurately by credit gaps. Examples included FX denominated household lending, loans with variable lending rate that could be adjusted at a nontransparent way by banks, and the excessive reliance on short-term external funds. Due to the strengthened macroprudential policy framework, these are not expected to arise again simultaneously in the Baseline scenario.

includes the possibility that, because of more efficient policy interventions, excessive lending will only arise marginally and for a short period of time. A version of the scenario is also examined ('Baseline with higher uncertainties') where the simulated data are derived from shocks with greater standard deviation that is typical in the ARIMA models estimated for the euro area countries. Figure 9 demonstrates that the simulated time series of the latter scenario contain more noise and cover a wider range of credit-to-GDP values. However, vast majority of the time series in the two scenarios do not reach the historic maximum seen in the global financial crisis by 2037.

In the 'Higher systemic risk' scenario, we assume no financial deepening, but allow the largest overheating seen in the European data. Accordingly, there are more credit-to-GDP time series that rise more sharply and quickly compared to the 'Baseline' scenario (Figure 10). The 'Lower systemic risk' scenario considers the cases where the activity of the financial intermediary system expands only moderately. This means not only a more limited lending boom, but also an absence of financial deepening. This scenario includes the credit-to-GDP time series that increase the least (Figure 10), and the excessive credit growth periods that were mostly typical in core EU countries.

Figure 9
Minimum, maximum, and median simulated time series of the Hungarian additional credit-to-GDP ratio, according to the 'Baseline' and 'Baseline with higher uncertainties' scenarios

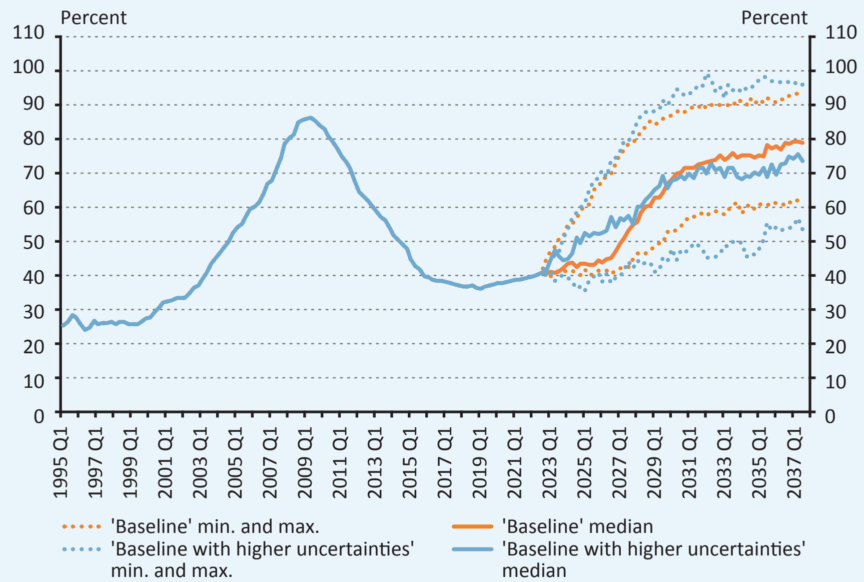
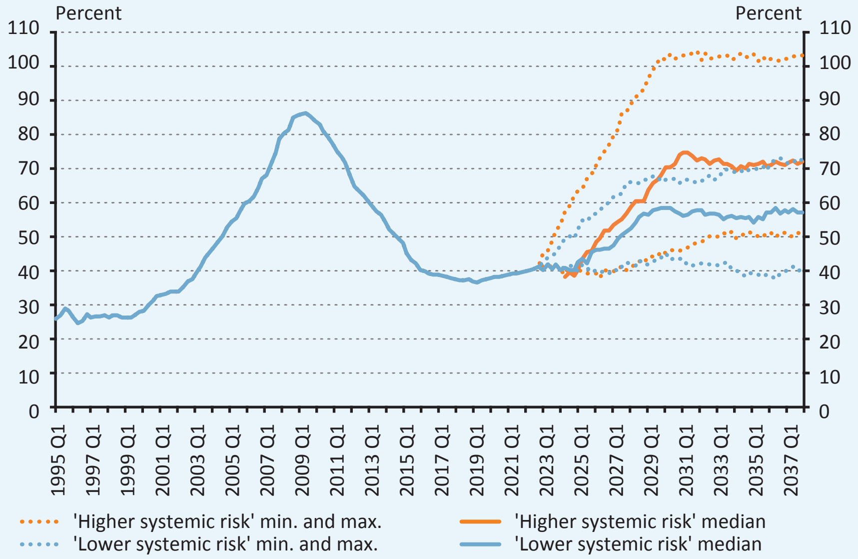


Figure 10
Minimum, maximum, and median simulated time series of the Hungarian additional credit-to-GDP ratio, according to the 'Lower systemic risk' and 'Higher systemic risk' scenarios



An advantage of the Hungarian simulated data over the European data is that they factor in more of the distinctive attributes of the Hungarian economy, which is important from the perspective of identifying the best credit gap for Hungary. Another argument for the simulations is that they produce larger sample size, which improves the accuracy and reliability of the results. Moreover, on account of the incorporated overheating periods, it is known when exactly credit gaps need to issue warning signals. A drawback of this approach is that it cannot handle potential structural breaks or greater changes, as the simulation of the (unadjusted) credit-to-GDP time series is driven exclusively by the parameters estimated on the past Hungarian credit-to-GDP observations. If none of the ARIMA specifications can provide a good approximation of the data generating process, the simulated data are less instructive. Finally, the simulations also required parameters of financial deepening and excessive lending, whose potential errors can undermine the reliability of the results.

6.2 SPECIFICATION OF THE EARLY WARNING SIGNALS

We evaluate only a third of the 48 credit gaps. Since narrow credit and shorter credit-to-GDP forecasts dominated broad credit and 3-year forecast respectively in Section 4, this section examines only credit gaps calculated with narrow credit, and no longer than 1-year credit-to-GDP forecasts. The 16 credit gaps under review are calculated for the full credit-to-GDP time series using actual data, forecasts, and simulated data. We rank credit gaps based on their early warning performance on forecasted and simulated data. Since only excessive lending periods and no crisis episodes are simulated, an ideal early warning indicator is expected to issue warning signals in and only in the quarters of excessive lending. This is a sensible requirement because the build-up of cyclical systemic risks in such periods significantly increases the probability of banking crises in the near future. We determine the signals from the beginning of the forecast data period until the end of the excessive lending period.[54] The signals are assessed separately for each scenario, on the pooled samples of the 200 time series belonging to one scenario. Like in Section 4, the AUROC and RU evaluation statistics are used, and balanced preferences (a preference parameter of 0.5) are assumed for the latter.[55]

[54] We disregard later periods because financial stress is especially likely to occur after major credit booms, so it is unrealistic to require 'No signal' in these quarters. Since no systemic banking crises are simulated, realistically no 'Signal' can be expected either.

[55] Unlike during the analysis with European data in Section 4, Type I and Type II error rates also have the same weight in the sense that the unconditional probabilities of the two errors here are more or less the same. While earlier 'Signal' was expected much less often, here the two signals are expected in a similar number of quarters.

6.3 RESULTS

Tables 15 and 16 summarize the early warning performance of the 16 credit gaps, under different simulation scenarios. As the simulation allows the examination of much more time series than the use of European data, there are much fewer credit gaps whose predicting power does not differ significantly from the best one. The simulated credit-to-GDP time series differ from each other much less than the time series of various countries, so higher AUROC and RU values are derived. The highest AUROC values are over 0.96 and the highest RU values are over 0.78, while there are considerable differences among other AUROC values and among other RU values.

There are two credit gap specifications whose AUROC values do not differ significantly from the highest value in any scenario: (no credit-to-GDP forecast, HP filter, long cycle) and (1-year credit-to-GDP forecast, wavelet filter, long cycle). A similarly good performance is seen under all scenarios by all the HP filter gaps that are calculated with medium-term or long cycles. The same is true for the following credit gaps calculated without a credit-to-GDP forecast: the credit gaps with CF filter and medium-term or long cycle, and the credit gap with wavelet filter and long cycle. These are eight indicators in total, which include only two of the three credit gaps that have the best early warning performance based on European data: the HP filter and the CF filter credit gap. However, the third one (the wavelet filter credit gap) fares the best in the 'Baseline' scenario here, which is considered the most relevant.

Table 15
AUROC values of credit gaps calculated with simulated future Hungarian credit-to-GDP data

Using 200 simulated future time series in each scenario, we expect "Signal" in qaurters with simulated excessive credit growth.			**Gap calculation**							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
'Baseline' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.98**	0.98	0.85	0.97	0.97	0.47	0.97	**0.98**
		1-year forecast	0.97	0.97	0.85	0.94	0.93	0.62	**0.98**	0.98
'Baseline with higher uncertainties' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.97**	0.96	0.83	0.96	0.96	0.48	0.96	0.95
		1-year forecast	0.96	0.95	0.81	0.94	0.90	0.62	**0.97**	0.95
'Lower systemic risk' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.98**	0.97	0.70	0.97	0.96	0.63	0.97	0.93
		1-year forecast	0.97	0.96	0.71	0.94	0.91	0.71	**0.98**	0.93
'Higher systemic risk' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.99**	0.99	0.78	0.98	0.97	0.53	0.98	0.96
		1-year forecast	0.99	0.98	0.79	0.96	0.92	0.68	**0.99**	0.96

Note: Red cells show the highest AUROC values for various scenarios. Green cells contain values that do not differ from the highest values at the 1 percent significance level. Credit gaps belonging to cells with bold numbers have AUROC values calculated under each of the different scenarios that do not differ significantly from the best ones at the 1 percent significance level. Cells with thick borders refer to the best three early warning credit gaps based on the analysis with European data in Section 4.

The results based on the RU are very similar. There are four credit gaps that are the best under all scenarios: the two that are also chosen by the AUROC and additionally the specifications (no credit-to-GDP forecast, HP filter, medium-term cycle), and (1-year credit-to-GDP forecast, HP filter, long cycle). These four are also among the eight good performers according to the AUROC. The other four do not fare as well based on the RU as based on the AUROC. Accordingly, only the HP filter credit gap out of the best three early warning credit gaps based on European data is among the top four here. However, the wavelet filter credit gap still fares the best in the 'Baseline' scenario, which is considered the most relevant.

Table 16
RU values of credit gaps calculated with simulated future Hungarian credit-to-GDP data

Using 200 simulated future time series in each scenario, we expect "Signal" in qaurters with simulated excessive credit growth.			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
'Baseline' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.83**	**0.85**	0.59	0.78	0.80	0.14	0.80	0.84
		1-year forecast	**0.82**	0.82	0.54	0.71	0.69	0.29	**0.82**	0.86
'Baseline with higher uncertainties' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.79**	**0.76**	0.57	0.78	0.75	0.13	0.78	0.75
		1-year forecast	**0.77**	0.73	0.52	0.70	0.64	0.29	0.79	0.76
'Lower systemic risk' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	**0.83**	**0.81**	0.36	0.78	0.78	0.32	0.81	0.71
		1-year forecast	**0.81**	0.77	0.32	0.71	0.63	0.38	0.83	0.73
'Higher systemic risk' scenario										
Credit-to-GDP	Narrow definition of credit stock	0-year forecast	0.89	**0.87**	0.53	0.84	0.81	0.25	0.86	0.81
		1-year forecast	**0.87**	0.84	0.48	0.78	0.66	0.36	**0.88**	0.82

Note: Red cells show the highest RU values for various scenarios. Green cells contain values that are no more than 5 percent lower than the highest values. Credit gaps belonging to cells with bold numbers have RU values calculated under each of the different scenarios that are no more than 5 percent lower than the highest values. Cells with thick borders refer to the best three early warning credit gaps based on the analysis with European data in Section 4.

All in all, the combined results from the analyses of European data and simulated Hungarian data supports that the HP filter credit gap, which is chosen as the additional credit-to-GDP gap by several European macroprudential authorities, can be considered as the best early warning credit gap. However, there are some other credit gaps with fairly different specifications that are not significantly worse. Two of these stand out: the CF filter credit gap and the wavelet filter credit gap.

7 Summary

The study used European data and simulated Hungarian data to determine the univariate one-sided credit-to-GDP gap specification that issues the most accurate early warning signals of systemic banking crises. The paper sought to include as many relevant credit gaps as possible, so credit gaps differing in four characteristics with a total of 48 specifications were compared. They were generated (1) from broad and narrow definition of credit, (2) with credit-to-GDP time series optionally extended with forecasts, (3) using three different filtering methods (HP, CF, wavelet filter), and (4) assuming three different cycle lengths. When looking at European data, we identified the best credit gap as the one that issues the most informative early warning signals on all prediction horizons under review. Three specifications meet this requirement, which shows that the question 'What to filter?' has a more straightforward answer, than the question 'How to filter?': (narrow credit, no forecast, HP filter, long cycle), (narrow credit, no forecast, CF filter, medium-term cycle), (narrow credit, 1-year forecast, wavelet filter, short cycle).

Then, we conducted the following robustness checks. Instead of AUROC, we used the RU evaluation statistics for both in-sample and out-of-sample prediction exercises. Instead of the full sample, we assessed the credit gaps on subsamples by country groups and different time periods. Finally, the paper tested the prediction power of the credit gaps calculated with credit-to-GDP time series extended with an alternative forecasting method (ARIMA). With the exception of the out-of-sample analysis, the robustness checks confirms that narrow credit and shorter credit-to-GDP forecasts dominate broad credit and 3-year forecasts respectively, and that special combinations of some filtering methods and cycle lengths produce credit gaps with accurate early warning performance. During the sensitivity analyses, no credit gap is found that performs at least as good as the originally identified best three credit gaps.

We examined two other properties of the credit gaps that are important for macroprudential policy, namely endpoint uncertainty and the correlation with the financial cycle. On European data, we find that credit gaps with more solid prediction power tend to correlate more strongly with financial cycles. However, the set of credit gaps with the lowest endpoint uncertainty is not identical to the set of the best early warning credit gaps. One of the main reasons is that the extension of credit-to-GDP time series with forecasts improves the former but undermines the latter feature.

Finally, we extended the Hungarian credit-to-GDP time series with ARIMA simulations under various future scenarios, and we evaluated how accurately credit gaps signal the excessive lending periods incorporated into the simulations. Out of the three credit gaps that prove to be the best on European data, this exercise confirms the performance of the HP filter credit gap and the wavelet filter gap to a lesser extent (only under the baseline scenario, which is considered the most relevant). The CF filter credit gap follows close behind.

This study presents the most comprehensive and detailed comparison of univariate one-sided credit gaps to date. The results partly underpin the practice widely used in the EU, that the main credit gap indicator is calculated with narrow credit, without a credit-to-GDP forecast, with a one-sided HP filter using a lambda of 400,000. The paper also highlights that this indicator should be supplemented with the above-mentioned CF filter credit gap and wavelet filter credit gap that have similarly good prediction power. These three can complement each other well, as they generate different Type I and Type II errors, and exhibit varying relative performance over different prediction horizons. Even their simple averages give indicators that have a slightly better early warning accuracy than any of the three. Furthermore, CF filter credit gap has much smaller endpoint uncertainty than the HP filter credit gap.

An interesting question that is beyond the scope of the present study is what the result of similar analyses performed on data from countries outside Europe would be. The simulation method for the Hungarian case could also be generalized for testing various early warning indicators of systemic banking crises other than credit gaps. Finally, the aggregation method producing the best early warning 'credit gap index' from the individual credit gaps should also be determined.

References

Aikman, D.–Haldane, A. G.–Nelson, B. D. [2015]: Curbing the credit cycle. *The Economic Journal*, Vol. 125. No. 585. 1072–1109. https://doi.org/10.1111/ecoj.12113.

Alessi, L.–Antunes, A.–Babecký, J.–Baltussen, S.–Behn, M.–Bonfim, D.–Bush, –Detken, C.–Frost, J.–Guimarães, R.–Havránek, T.–Joy, M.–Kauko, K.–Matějů, J.–Monteiro, N.–Neudorfer, B.–Peltonen, T.–Rodrigues, P. M. M.–Rusnák, M.–Schudel, W.–Sigmund, M.–Stremmel, H.–Šmídková, K.–van Tilburg, R.–Vašíček, B.–Žigraiová, D. [2015]: Comparing different early warning systems: Results from a horse race competition among members of the Macro-prudential Research Network. *MPRA Paper*, No. 62194.

Alessi, L.–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. No. 3. 520–533. https://doi.org/10.1016/j.ejpoleco.2011.01.003.

Antunes, A.–Bonfim, D.– Monteiro, N.–Rodrigues, P. M. M. [2018]: Forecasting banking crises with dynamic panel probit models. *International Journal of Forecasting*, Vol. 34. No. 2. 249–275. https://doi.org/10.1016/j.ijforecast.2017. 12.003.

Anundsen, A. K.–Gerdrup, K.–Hansen, F.–Kragh-Sørensen, K. [2016]: Bubbles and crises: The role of house prices and credit. *Journal of Applied Econometrics,* Vol. 31. No. 7. 1291–1311. https://doi.org/10.1002/jae.2503.

Arbatli-Saxegaard, E. C.– Muneer, M. A. [2020]: The countercyclical capital buffer: A cross-country overview of policy frameworks. *Norges Bank Staff Memo*, No. 6/2020.

Babecký, J.–Havránek, T.–Matějů, J.–Rusnák, M.–Šmídková, K.–Vašíček, B. [2014]: Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, Vol. 15. 1–17. https://doi.org/10.1016/j.jfs.2014.07.001.

Babić, D.–Fahr, S. [2019]: Shelter from the storm: recent countercyclical capital buffer (CCyB) decisions*. ECB Macroprudential Bulletin*, No. 7.

Banco de Portugal [2015]: Countercyclical capital buffer in Portugal: How will it work? December 2015.

Banco de Portugal [2019]: Financial stability report, June 2019, Box 3. 118–124.

Barrell, R.–Karim, D.–Macchiarelli, C. [2020]: Towards an understanding of credit cycles: Do all credit booms cause crises? *The European Journal of Finance*, Vol. 26. No. 10. 978–993. https://doi.org/10.1080/1351847X.2018.1521341.

BCBS [2010]: Guidance for national authorities operating the countercyclical capital buffer. Basel Committee on Banking Supervision, Bank for International Settlements.

BCBS [2017]: Implementation, Range of practices in implementing the countercyclical capital buffer policy. Basel Committee on Banking Supervision, Bank for International Settlements.

BIS [2020]: Long series on total credit and domestic bank credit to the private non-financial sector. Documentation on data, Bank for International Settlements, https://www.bis.org/statistics/totcredit/credpriv_doc.pdf.

Beltran, D. –Jahan-Parvar, M. R.–Paine, F. A. [2020]: Optimizing credit gaps for predicting financial crises: Modelling choices and tradeoffs. Manuscript.

Boissay, F.–Collard, F.–Smets, F. [2016]: Booms and Banking Crises. *Journal of Political Economy*, Vol. 124. No. 2. 489–538. https://doi.org/10.1086/685475.

Bonfim, D.–Monteiro, N. [2013]: The implementation of the countercyclical capital buffer: Rules versus discretion. Published in the Financial Stability Report (November), Banco de Portugal. 87–110.

Borio, C. [2014]: The financial cycle and macroeconomics: What have we learnt? *Journal of Banking and Finance*, Vol. 45. 182–198. https://doi.org/10.1016/j.jbankfin.2013.07.031.

Borio, C.–Drehmann, M. [2009]: Assessing the risk of banking crises – revisited. *BIS Quarterly Review*, March 2009. 29–46.

Borio, C.–Drehmann, M.–Dora, X. [2018]: The financial cycle and recession risk. *BIS Quarterly Review*, December 2018. 59–71.

Borio, C.–Lowe, P. [2002]: Assessing the risk of banking crises. *BIS Quarterly Review*, December 2002. 43–54.

Buncic, D.–Melecky, M. [2014]: Equilibrium credit: The reference point for macroprudential supervisors. *Journal of Banking and Finance*, Vol. 41. 135–154. https://doi.org/10.1016/j.jbankfin.2014.01.005.

Canova, F. [1998]: Detrending and business cycle facts. *Journal of Monetary Economics*, Vol. 41. No. 3. 475–512. https://doi.org/10.1016/S0304-3932(98)00006-3.

Castro, C.–Estrada, Á.–Martínez, J. [2016]: The countercyclical capital buffer in Spain: An analysis of key guiding indicators. *Banco de España Working Paper Series*, No. 1601.

Christiano, L. J.–Fitzgerald, T. J. [2003]: The band pass filter. *International Economic Review*, Vol. 44. No. 2. 435–465. https://doi.org/10.1111/1468-2354.t01-1-00076.

Coudert, V.–Idier, J. [2018]: Reducing model risk in early warning systems for banking crises in the euro area. *International Economics*, Vol. 156. 98–116. https://doi.org/10.1016/j.inteco.2018.01.002.

Crowley, M. P. [2005]: An intuitive guide to wavelets for economist. *Bank of Finland Research Discussion Paper,* No. 2005/1.

Davis, E. P.–Karim, D. [2008]: Comparing early warning systems for banking crises. *Journal of Financial Stability*, Vol. 4. No. 2. 89–120. https://doi.org/10.1016/j.jfs.2007.12.004.

DeLong, E. R.–DeLong, D. M.–Clarke-Pearson, D. L. [1988]: Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics,* Vol. 44. No. 3. 837–845. https://doi.org/10.2307/2531595.

Dembiermont, C.–Drehmann, M.–Muksakunratana, S. [2013]: How much does the private sector really borrow? A new database for total credit to the private nonfinancial sector. *BIS Quarterly Review*, March 2013. 65–81.

Demirgüç-Kunt, A.–Detragiache, E. [2000]: Monitoring Banking Sector Fragility: A Multivariate Logit Approach. *The World Bank Economic Review,* Vol. 14. No. 2. 287–307. https://doi.org/10.1093/wber/14.2.287.

Detken, C.–Weeken, –Alessi, L.–Bonfim, D.–Boucinha, M. M.–Castro, Ch.–Frontczak, S.–Giordana, G.–Giese, J.–Jahn, N.–Kakes, J.–Klaus, B.–Lang, J. H.–Puzanova, N.–Welz, P. [2014]: Operationalising the countercyclical capital buffer: indicator selection, threshold, identification and calibration. *ESRB Occasional Paper Series*, No. 5.

Deutsche Bundesbank [2018]: Financial stability review, 2018. 47–49.

Drehmann, M. [2013]: Total credit as an early warning indicator for systemic banking crises. *BIS Quarterly Review*, June 2013. 41–45.

Drehmann, M.–Borio, C.–Gambacorta, L.–Jiminez, G.–Trucharte, C. [2010]: Countercyclical capital buffers: Exploring options. *BIS Working Papers*, No. 317.

Drehmann, M.–Borio, C.–Tsatsaronis, K. [2011]: Anchoring countercyclical capital buffers: The role of credit aggregates. *International Journal of Central Banking*, Vol. 7. No. 4. 189–240.

Drehmann, M.–Borio, C.–Tsatsaronis, K. [2012]: Characterising the financial cycle: Don't lose sight of the medium term! *BIS Working Papers*, No. 380.

Drehmann, M.–Juselius, M. [2014]: Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting*, Vol. 30. No. 3. 759–780. https://doi.org/10.1016/j.ijforecast.2013.10.002.

Drehmann, M.–Tsatsaronis, K. [2014]: The credit-to-GDP gap and countercyclical capital buffers: Questions and answers. *BIS Quarterly Review*, March 2014, 55–73.

Drehmann, M.–Yetman, J. [2018]: Why you should use the Hodrick-Prescott filter — at least to assess credit gaps. *BIS Working Paper*s, No. 744.

Edge, R. M.–Meisenzahl, R. R. [2011]: The unreliability of credit-to-GDP ratio gaps in real time: Implications for countercyclical capital buffers. *International Journal of Central Banking*, Vol. 7. No. 4. 261–298.

ESRB [2014]: Recommendation of the European Systemic Risk Board of 18 June 2014 on guidance for setting countercyclical buffer rates. ESRB/2014/1.

ESRB [2018]: A Review of macroprudential policy in the EU in 2017, Special feature B: Use of the countercyclical capital buffer – a cross-country comparative analysis. European Systemic Risk Board, April 2018.

Frankel, J. A.–Rose, A. K. [1996]: Currency crashes in emerging markets: An empirical treatment. *Journal of International Economics*, Vol. 41. No. 3-4. 351–366. https://doi.org/10.1016/S0022-1996(96)01441-9.

Galán, J. E. [2019]: Measuring credit-to-GDP gaps, The Hodrick-Prescott filter revisited, *Banco de España Occasional Papers*, No. 1906.

Galán, J. E.–Mencía, J. [2018]: Empirical assessment of alternative structural methods for identifying cyclical systemic risk in Europe, *Banco de España Working Paper Series*, No. 1825.

Galati, G.–Hindrayanto, I.–Koopman, S. J.–Vlekke, M. [2016]: Measuring financial cycles in a model-based analysis: Empirical evidence for the United States and the euro area. *Economics Letters*, Vol. 145. 83–87. https://doi.org/10.1016/j.econlet.2016.05.034.

Galimberti, J. K.–Moura, M. L. [2014]: Improving the reliability of real-time Hodrick-Prescott filtering using survey forecasts. *ETH Zurich Working Paper*, No. 360.

Gallegati, M.–Gallegati, M. [2007]: Wavelet variance analysis of output in G-7 countries. *Studies in Nonlinear Dynamics and Econometrics*, Vol. 11. No. 3. Article 6. https://doi.org/10.2202/1558-3708.1435.

Gerdrup, K.–Kvinlog, A. B.–Schaaning, E. [2013]: Key indicators for a countercyclical capital buffer in Norway – trends and uncertainty. *Norges Bank Staff Memo*, No. 13/2013.

Geršl, A.–Jašová, M. [2018]: Credit-based early warning indicators of banking crises in emerging markets. *Economic Systems*, Vol. 42. No. 1. 18–31. https://doi.org/10.1016/j.ecosys.2017.05.004.

Giordani, P.–Spector, E.–Zhang, X. [2017]: A new early warning indicator of financial fragility in Sweden, *Sveriges Riksbank Economic Commentaries,* No. 1/2017.

Grinderslev, O. J.–Kramp, P. L.–Kronborg, A. F.–Pedersen, J. [2017]: Financial cycles: What are they and what do they look like in Denmark? *Danmarks Nationalbank, Working Papers,* No. 115.

Hamilton, J. D. [2018]: Why you should never use the Hodrick-Prescott filter, *Review of Economics and Statistics*, Vol. 100. No. 5. 831–843. https://doi.org/10.1162/rest_a_00706.

Hamilton, J. D.–Leff, D. [2020]: Measuring the Credit Gap. Mimeo.

Hiebert, P.–Jaccard, I.–Schüler, Y. [2018]: Contrasting financial and business cycles: Stylized facts and candidate explanations. *Journal of Financial Stability*, Vol. 38. 72–80. https://doi.org/10.1016/j.jfs.2018.06.002.

Hodrick, R. J. [2020]: An exploration of trend-cycle decomposition methodologies in simulated data. *NBER Working Paper Series*, No. 26750.

Hodrick, R. J.–Prescott, E. C. [1997]: Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, Vol. 29. No. 1. 1–16. https://doi.org/10.2307/2953682.

Hosszú, Zs.–Körmendi, Gy.–Mérő, B. [2016]: Egy- és többváltozós szűrők a hitelrés alakulásának meghatározására (Univariate and multivariate filters to measure the credit gap). Közgazdasági Szemle (Economic Review), Vol. 63. 233–259. http://dx.doi.org/10.18414/KSZ.2016.3.233.

Jordà, Ò.–Schularick, M.–Taylor, A. M. [2015]: Leveraged bubbles. *Journal of Monetary Economics*, Vol. 76. Supplement. S1–S20. https://doi.org/10.1016/j.jmoneco.2015.08.005.

Jordà, Ò.–Schularick, M.–Taylor, A. M. [2016]: Macrofinancial history and the new business cycle facts. *NBER Macroeconomics Annual*, Vol. 31. 213–263. https://doi.org/10.1086/690241.

Kaminsky, G. L.–Lizondo, S.–Reinhart, C. M. [1998]: Leading indicators of currency crises. *IMF Staff Papers*, Vol. 45. No. 1. 1–48. https://doi.org/10.2307/3867328.

Kaminsky, G. L.–Reinhart, C. M. [1999]: The twin crises: The causes of banking and balance-of-payments problems. American Economic Review, Vol. 89. No. 3. 473–500. https://doi.org/10.1257/aer.89.3.473.

Kauko, K.–Tölö, E. [2020]: On the long-run calibration of the credit-to-GDP gap as a banking crisis predictor. *Finnish Economic Papers*, Vol. 29. No. 2.

Lang, J. H.–Izzo, C.–Fahr, S.–Ruzicka, J. [2019]: Anticipating the bust: A new cyclical systemic risk indicator to assess the likelihood and severity of financial crises. *ECB Occasional Paper Series*, No. 219.

Lang, J. H.–Welz, P. [2018]: Semi-structural credit gap estimation. *ECB Working Paper Series*, No. 2194.

Lee, S. J.–Posenau, K. E.–Stebunovs, V. [2020]: The anatomy of financial vulnerabilities and banking crises. *Journal of Banking and Finance*, Vol. 112. Article 105334. https://doi.org/10.1016/j.jbankfin.2018.04.013.

Lietuvos bankas [2014]: Lithuanian economic review, December 2014.

Lietuvos bankas [2019]: Financial Stability Review, 2019.

Lo Duca, M.–Koban, A.–Basten, M.–Bengtsson, E.–Klaus, B.–Kusmierczyk, P.–Lang, J. H.–Detken, C.–Peltonen, T. [2017]: A new database for financial crises in European countries. *ESRB Occasional Paper Series,* No. 13.

Lo Duca, M.–Peltonen, T. A. [2013]: Assessing systemic risks and predicting systemic events. *Journal of Banking and Finance*, Vol. 37. No. 7. 2183–2195. https://doi.org/10.1016/j.jbankfin.2012.06.010.

Martínez, J. F.–Oda, D. [2018]: Characterization of the Chilean financial cycle, early warning indicators and implications for macro-prudential policies. *Working Papers of the Central Bank of Chile*, No. 823.

Mencía, J.–Saurina, J. [2016]: Macroprudential policy: objectives, instruments and indicators. *Banco de España, Occasional Papers*, No. 1601.

Mise, E.–Kim, T.-H.–Newbold, P. [2005]: On suboptimality of the Hodrick–Prescott filter at time series endpoints. *Journal of Macroeconomics*, Vol. 27. No. 1. 53–67. https://doi.org/10.1016/j.jmacro.2003.09.003.

MNB [2019a]: Inflation Report, December 2019.

MNB [2019b]: Financial Stability Report, December 2019.

MNB [2020]: Macroprudential Report, 2020.

Narodowy Bank Polski [2016]: Antycykliczny bufor kapitałowy i modele wczesnego ostrzegania przed kryzysami bankowymi: Zastosowanie dla Polski (Countercyclical capital buffer and early warning models for banking crises: Application for Poland).

Pekanov, A.–Dierick, F. [2016]: Implementation of the countercyclical capital buffer regime in the European Union. *ESRB Macro-prudential Commentaries*, No. 8.

Phillips, P. C. B.– Jin, S. [2021]: Business cycles, trend elimination, and the HP filter. *International Economic Review,* Vol. 62. No. 2. 469–520. https://doi.org/10.1111/iere.12494.

Plašil, M.–Seidler, J.–Hlaváč, P. [2016]: A new measure of the financial cycle: Application to the Czech republic. *Eastern European Economics*, Vol. 54. No. 4. 296–318. https://doi.org/10.1080/00128775.2016.1192950.

Pönkä, H. [2017]: The role of credit in predicting US recessions. *Journal of Forecasting,* Vol. 36. No. 5. 469–482. https://doi.org/10.1002/for.2448.

Ravn, M. –Uhlig, H. [2002]: On adjusting the Hodrick-Prescott filter for the frequency of observations. *The Review of Economics and Statistics,* Vol. 84. No. 2. 371–380. https://doi.org/10.1162/003465302317411604.

Sarlin, P. [2013]: On policymakers' loss functions and the evaluation of early warning systems. *Economics Letters*, Vol. 119. No. 1. 1–7. https://doi.org/10.1016/j.econlet.2012.12.030.

Schleicher, C. [2002]: An introduction to wavelets for economists. *Bank of Canada Working Paper*, No. 2002/3.

Schüler, Y. S. [2018]: Detrending and financial cycle facts across G7 countries: Mind a spurious medium term! *ECR Working Paper Series*, No. 2138.

Sondermann, D.–Zorell, N. [2019]: A macroeconomic vulnerability model for the euro area. *ECB Working Paper Series,* No. 2306.

Tölö, E.–Laakkonen, H.–Kalatie, S. [2018]: Evaluating indicators for use in setting the countercyclical capital buffer. *International Journal of Central Banking*, Vol. 14. No. 2. 51–112.

Uliha, G. [2016]: Az olajár és a makrogazdaság kapcsolatának elemzése folytonos wavelet transzformáció segítségével (Analysis of the relationship between oil price and macroeconomy with continuous wavelet transforms). *Statisztikai Szemle (Hungarian Statistical Review)*, Vol. 94. No. 5. 505–534. https://doi.org/10.20311/stat2016.05.hu0505.

Valinskytė, N.–Rupeika, G. [2015]: Leading indicators for the countercyclical capital buffer in Lithuania. *Bank of Lithuania Occasional Paper Series,* No. 4/2015.

Wolf, E.–Mokinski, F.– Schüler, Y. [2020]: On adjusting the one-sided Hodrick-Prescott filter. *Deutsche Bundesbank Discussion Paper,* No. 11/2020.

Appendix A: Data

A.1 CREDIT-TO-GDP

The primary data source of the credit-to-GDP time series is the BIS credit-to-GDP database, containing indicators from 19 ESRB member countries (Table 17).[56] Broad credit includes all the loans and debt securities taken out by private non-financial sector (together with the loans extended by the private non-financial sector).[57] The BIS typically takes the data from financial accounts statistics of the countries, typically published by central banks. Narrow credit includes only the loans and debt securities taken out by private non-financial sector from 'domestic deposit-taking corporations except the central bank'. The exact group of institutions differs from country to country and period to period, but domestic banks are always included, and mostly only those. Narrow credit data come from the monetary statistics of the countries, usually also compiled by central banks. The denominator of the credit-to-GDP ratio is always the 4-quarter moving sum of nominal GDP, which is estimated by the BIS, if necessary, by the linear interpolation of annual data.

The secondary data source of the credit-to-GDP time series is the websites of national macroprudential authorities in the ESRB member countries. In accordance with Recommendation 2014/1 of the ESRB, all ESRB member countires need to calculate the so-called standardized credit-to-GDP gap[58] to support the decisions on building up the CCyB. Furthermore, pursuant to the Recommendation, a so-called additional credit-to-GDP gap may also be determined optionally, which is calculated differently from this. The standardized indicator needs to use broad credit, while more than a third of the members calculated an additional credit-to-GDP gap with narrow credit.[59]

We have found credit-to-GDP time series from 13 ESRB member countries, out of which 8 countries are also included in the BIS database. The two data sources are merged as follows. The study sought to assess the credit-to-GDP gaps calculated with broad and narrow credit on the data from the same group of countries. Therefore, the data from the national macroprudential authorities of Norway, Portugal, Romania, and the UK are disregarded, as these countries only published credit-to-GDP data with broad credit. In the case of the remaining countries, there are considerable differences between the credit-to-GDP data of the BIS and national authorities. This may be because while the BIS seeks to compile its database with a common methodology, the ESRB Recommendation 2014/1 does not give an exact definition of broad and narrow credit. Furthermore, in all the five countries concerned, the BIS time series are longer than those of national authorities (sometimes considerably), so in each case BIS data are used.

Two further selection criteria are applied to the data from the remaining 19 plus 4 countries. First, Ireland and Luxembourg are left out of the analysis because their economic structure and thus their credit-to-GDP time series are deemed too special as well. Both countries rely heavily on cross-border financial services, and the rise in the broad credit-to-GDP ratio was extremely fast before the global financial crisis. According to BIS data, it soared from around 120–130 percent at the turn of the millennium to over 300 percent in a decade, coming close to 400 percent in the case of Luxembourg.

[56] Source: https://www.bis.org/statistics/totcredit.htm. In this database, there are also data for additional European countries: Russia, Switzerland and Turkey.
[57] For details on the data sources and the method of the compilation of the BIS database, see: BIS (2020) and Dembiermont et al. (2013).
[58] The method follows the guidance by the BCBS (2010), so the main elements are as follows: broad credit and a one-sided HP filter with a lambda of 400,000.
[59] See the ESRB records about the decisions by the national authorities on the CCyB: https://www.esrb.europa.eu/national_policy/ccb/html/index.en.html.

Table 17
Length of credit-to-GDP time series by country group and data source

Country group	Country	Credit-to-GDP data from BIS		Credit-to-GDP data from national macroprudential authorities	
		Starting dates of 'broad' credit-to-GDP time series	Starting dates of 'narrow' credit-to-GDP time series	Starting dates of 'broad' credit-to-GDP time series	Starting dates of 'narrow' credit-to-GDP time series
Core EU countries	Austria	1960 Q4	id.		
	Belgium	1970 Q4	id.		
	France	1969 Q4	id.		
	Germany	1960 Q4	id.	1968 Q4	id.
	Netherlands	1961 Q1	id.		
	United Kingdom	1963 Q1	id.	1963 Q1	
Nordic countries	Denmark	1966 Q4	id.	1970 Q1	1981 Q1
	Finland	1970 Q4	1974 Q1		
	Norway	1960 Q4	id.	1983 Q1	
	Sweden	1961 Q1	id.		
Mediterranean countries	Greece	1970 Q4	id.		
	Italy	1960 Q4	1974 Q4	1980 Q1	id.
	Portugal	1960 Q4	id.	1977 Q4	
	Spain	1970 Q1	id.		
Central and Eastern Europe	Czech Republic	1993 Q1	id.	1995 Q4	id.
	Hungary	1970 Q4	id.		
	Lithuania			1995 Q4	id.
	Poland	1992 Q1	id.		
'Outlier' countries	Ireland	1971 Q2	id.	2005 Q1	id.
	Luxembourg	1999 Q1	2003 Q1		
Countries with short time series	Estonia			2003 Q4	1996 Q4
	Latvia			1995 Q4	1996 Q4
	Romania			1991 Q1	
	Slovakia			2003 Q4	id.

Note: In the paper, we use the time series pertaining to the shaded cells. Time series starting in the same decade have the same shading.
Source: BIS, websites of national macroprudential authorities.

The other selection criterion was the length of the time series. We need long time series for two reasons. Firstly, one-sided credit gaps start to take reliable values after about a decade of credit-to-GDP observations. Secondly and less importantly, short time series are, obviously, from the recent past, therefore they lead to an overrepresentation of recent experiences. Both time series from a country are either discarded or retained. If one of the time series of a country proves to be too short, neither of them are taken into account, irrespective of the length of the other. The countries that are left out because of this criterion are: Estonia, Latvia and Slovakia. (The time series of the 'outlier' Luxembourg would also have been too short.)

The cells of Table 17 with a blue shading show the starting dates of the credit-to-GDP time series from 18 countries that we use in the study. The time series end in 2019 Q1, with one exception, the time series with broad credit of Lithuania, which ends in 2018 Q4. Only the Lithuanian time series are from a national macroprudential authority.[60] The starting dates of the credit-to-GDP time series with broad and narrow credit differ in two countries. The longer one is not aligned with the shorter one, because the same crisis episodes can be used for assessing early warning performance both with and without a shortening.

A.2 CRISIS PERIODS

The ESRB crisis database contains the financial crisis episodes of all EU member countries and Norway between 1970 and 2016. This includes all the 18 countries with credit-to-GDP time series selected in the previous subsection. A distinctive feature of the database is that it characterizes crisis episodes from many aspects. This allows to identify systemic financial crisis episodes that are unrelated to the post-socialist transition, and during which systemic risks also materialized in the banking sector. We need these episodes because we want to evaluate early warning signals of impending systemic banking crises.

Table 18
Systemic banking crises by country group

Country group	Country	Global financial crisis (15 periods)	Pre-GFC crises (11 periods)	
Core EU countries	Austria	2007 Q4 - 2016 Q2		
	Belgium	2007 Q4 - 2012 Q4		
	France	2008 Q2 - 2009 Q4	1991 Q2 - 1995 Q1	
	Germany	2007 Q3 - 2013 Q2	1974 Q2 - 1974 Q4	2001 Q1 - 2003 Q4
	Netherlands	2008 Q1 - 2013 Q1		
	United Kingdom	2007 Q3 - 2010 Q1	1973 Q4 - 1975 Q4	1991 Q3 - 1994 Q2
Nordic countries	Denmark	2008 Q1 - 2013 Q4	1987 Q1 - 1995 Q1	
	Finland		1991 Q3 - 1996 Q4	
	Norway	2008 Q3 - 2009 Q4	1988 Q3 - 1993 Q4	
	Sweden	2008 Q3 - 2010 Q4	1991 Q1 - 1997 Q2	
Mediterranean countries	Greece	2010 Q2 -		
	Italy	2011 Q3 - 2013 Q4	1991 Q3 - 1997 Q4	
	Portugal	2008 Q4 - 2015 Q4		
	Spain	2009 Q1 - 2013 Q4	1978 Q1 - 1985 Q3	
Central and Eastern Europe	Czech Republic			
	Hungary	2008 Q3 - 2010 Q3		
	Lithuania	2008 Q4 - 2009 Q4		
	Poland			

Note: Crisis period refers to the period between the 'Start date' and the 'End of crisis management date' in the ESRB database, rounded to quarters. All the quarters are included in which there was a crisis as defined before for at least a month. The crisis periods in the blue cells are categorized by the ESRB crisis database as 'imported', the other crises either erupted domestically or both are true about them (Lo Duca et al., 2017).

Source: ESRB.

[60] The data are compiled by the Lithuanian central bank: https://www.lb.lt/uploads/documents/files/EN/our-functions/financial-stability/CCyB_data_2019Q2.xlsx. Broad credit data are sourced from the financial accounts and comprise the loans to households and non-financial corporations as well as debt securities issued by non-financial corporations. Narrow credit data are sourced from the balance sheets of monetary financial institutions and comprise the adjusted amount of loans to households and non-financial corporations from domestic monetary financial institutions. The adjustment seeks to eliminate certain structural breaks, the methodology is described in Appendix 2 of Lietuvos bankas (2014).

We retain all the resulting 26 crisis periods, even the six, including the only Hungarian one, that are considered 'imported' in the database, i.e. triggered by cross-border contagion effects (Table 18). The other two potential categories are domestic crisis and mixed-origin crisis. As the Hungarian example shows, the immediate reason for the eruption of the crisis may be an outside shock, but the domestic systemic financial risks that have built up endogenously may considerably exacerbate the calamity. So, this classification of the database does not accurately characterize all the potential reasons of a crisis. Moreover, according to the database, in all five cases other than in Norway, earlier 'excessive credit growth' played a role, which is exactly what credit gaps can measure. Consequently, we do not consider it justified to leave out the six crisis episodes from our analysis.

As described in Section 4.1, we evaluate credit gap values only from 8 years after the start of the credit-to-GDP time series. In the case of the credit-to-GDP time series beginning in the 1960s, there are some quarters at the end of the 1960s that are not covered by the crisis database. We assume that no systemic banking crisis occurred in these quarters.

A.3 AGGREGATED INDICATORS OF FINANCIAL CYCLE

We find 11 aggregated indicators that characterize the position of the financial cycle. The indicators can be classified into three main groups (Table 19). Most of them, six in total, are financial cycle indices. These indices are usually not too complex functions of individual indicators, often taking the form of weighted sums, attempting to use all the main individual indicators determining the financial cycle. Therefore, they can be readily decomposed, but the index value cannot be directly interpreted, it can characterize the current position of the financial cycle only in comparison with its historical distribution. The next biggest group comprises three estimated probabilities of a financial crisis in the near future. These indicators can be interpreted in themselves, but their decomposition is less straightforward. The third group includes two multivariate gap indicators. The methodology used here conducts trend-cycle decompositions with several indicators relevant to the financial cycle.

Table 19
Main characteristics of the aggregated indicators measuring financial cycle position

Country group	Country	Starting date	Ending date	Indicator type	Details about the methodology
Core EU countries	Germany	1983 Q1	2018 Q2	financial cycle index	Deutsche Bundesbank (2018)
		https://www.bundesbank.de/en/publications/reports/financial-stability-reviews/charts/early-warning-indicator-and-spillover-indicator-for-germany-768294			
	Spain	1999 Q1	2019 Q1	financial cycle index	Mencía and Saurina (2016)
		https://www.bde.es/f/webbde/INF/MenuHorizontal/Publicaciones/Boletines%20y%20revistas/InformedeEstabilidadFinanciera/IEF_Autumn2019.pdf			
Nordic countries	Denmark	1971 Q1	2018 Q4	2 multivariate gaps	Grinderslev et al. (2017)
		https://systemicriskcouncil.dk/Media/8/1/DSRR31%20-%20CCB%20Danmark%20(English).xlsx			
	Norway	1983 Q1	2019 Q1	2 financial crisis probabilities	Anundsen et al. (2016)
		https://www.norges-bank.no/en/topics/financial-stability/macroprudential-supervision/Countercyclical-capital-buffer/framework-countercyclical-capital-buffer/			
	Sweden	1980 Q4	2018 Q4	financial cycle index	Giordani et al. (2017)
		https://www.riksbank.se/globalassets/media/rapporter/fsr/engelska/2019/190522/data-financial-stability-report-2019_1.xlsx			
Mediter-ranean countries	Portugal	1991 Q1	2018 Q4	financial cycle index	Banco de Portugal (2019)
		https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/ref_06_2019_en.pdf			
Central and Eastern Europe	Czech Republic	2004 Q1	2018 Q4	financial cycle index	Plašil et al. (2016)
		https://www.cnb.cz/export/sites/cnb/en/financial-stability/.galleries/macroprudential_policy/countercyclical_capital_buffer/ccb_web_cnb_en.xlsx			
	Lithuania	2001 Q1	2018 Q4	financial cycle index	Lietuvos Bankas (2019)
		https://www.lb.lt/en/media/force_download/?url=/uploads/publications/docs/22310_a0027c03a416ff96725ed404b9cc9bcc.pdf			
	Poland	2005 Q1	2018 Q4	financial crisis probability	Narodowy Bank Polski (2016)
		https://www.nbp.pl/systemfinansowy/rsf062019.xlsx			

Note: The time series that begin in the same decade have the same shading.
Source: Websites of national macroprudential authorities.

Appendix B: Filtering methods

B.1 HODRICK–PRESCOTT FILTER

The trend is the result of the following minimization problem, where r_t denotes the original time series, and $\bar{r}_t$ denotes its trend value in period t (Hodrick and Prescott, 1997):

$$\min_{\bar{r}_1,\ldots,\bar{r}_T} \sum_{t=1}^{T} (r_t - \bar{r}_t)^2 + \lambda \sum_{t=2}^{T-1} ((\bar{r}_{t+1} - \bar{r}_t) - (\bar{r}_t - \bar{r}_{t-1}))^2$$

The exact solution can be found, for example, in Phillips and Jin (2021), which also gives a detailed description of its properties. The HP filter is popular in the case of macroeconomic time series due to its simple and intuitive objective function. The trend values should stay 'close enough' to the original time series (first sum), while they should evolve 'smoothly enough' (second sum). The parameter λ controls the relative importance of the two goals. Higher values result in smoother trends and thus longer cycles. For example, Ravn and Uhlig (2002) provide details on the relationship between parameter λ and estimated cycle length in the case of two-sided HP filters. We are not aware of any similar description of this relationship in the case of one-sided HP filters. Wolf et al. (2020) suggest that one-sided versions should use a lower λ for the same cycle length than two-sided ones.

Despite its popularity, several papers criticize the HP filter, mainly due to its large endpoint uncertainty and inconsistent ad hoc application, see, for example, Hamilton (2018), or Edge and Meisenzahl (2011) for credit gaps on an international sample, or Hosszú et al. (2016) for credit gaps on Hungarian data. However, different analyses make different aspects of the HP filter more important. In the case of output gap estimation, reducing endpoint uncertainty is a primary concern, because policymakers are particularly interested in the current position of the business cycle. By contrast, in the case of credit-to-GDP gap estimation, prediction of financial crises is more important than assessing the exact current position of the financial cycle, therefore the endpoint uncertainty in itself does not have a primary relevance.

The endpoint uncertainty of the HP filter can be significantly reduced if the time series data are extended with forecasts. This is especially true if professional forecasts are used rather than model-based forecasts (Galimberti and Moura, 2014).

B.2 CHRISTIANO–FITZGERALD FILTER

Frequency filters decompose a time series into sine and cosine functions of different frequencies. They use Fourier transform for this, which is the following in the case of a time series $x(t)$ (ω is the so-called angular frequency, which is derived by multiplying the frequency with 2π):

$$F_x(\omega) = \int_{-\infty}^{\infty} x(t)[\cos(\omega t) - i * sin\,(\omega t)]dt$$

Frequency filters require a lower and an upper cut-off value that controls the range of frequencies that belongs to the cyclical component in the decomposition of the time series. Frequencies exceeding the upper threshold are noise, while frequencies below the lower threshold are included in the trend. Among frequency filters, there are symmetric and asymmetric versions. These differ in how they handle the beginning and the end of the time series. Symmetric approaches truncate the first and last few observations and apply the same method to decompose the remaining observations into a trend and a cyclical component. Asymmetric methods, in contrast, use the full time series, but treat the observations at the beginning and the end slightly differently. Since we need a trend-cycle decomposition for the last observation, we chose an asymmetric approach, which is one of the most popular method, the asymmetric CF filter (for the exact formula, see Christiano and Fitzgerald, 2003). This choice does not have a significant impact on the results of the study, as the various frequency filters generate very similar trend-gap decompositions.

B.3 WAVELET FILTER

The continuous wavelet transform (which is comparable to the Fourier transform) of the time series $x(t)$ at time τ with a scale parameter s:

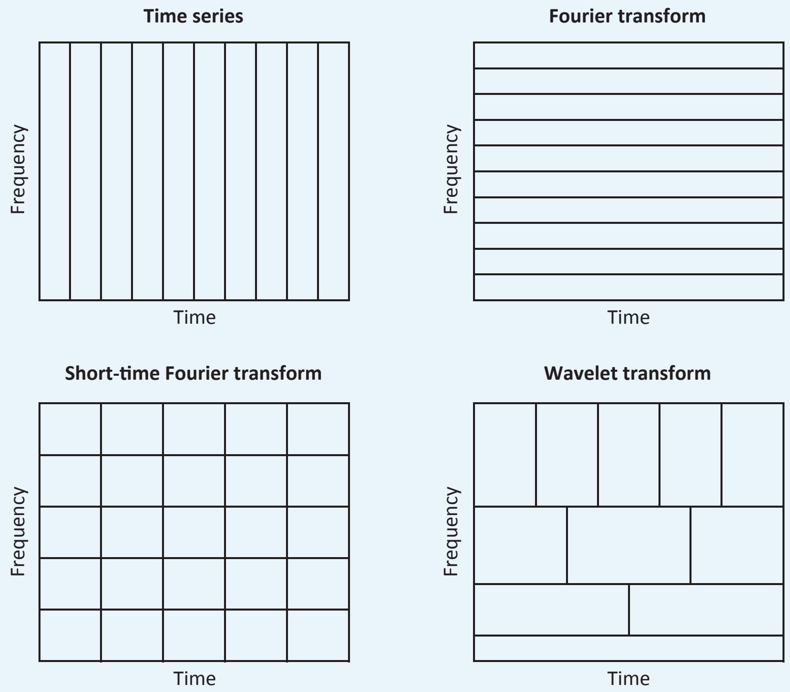
$$W_x(\tau, s) = \int_{-\infty}^{\infty} x(t)\psi^*_{\tau,s}(t)dt,$$

where $\psi^*_{\tau,s}(t)$ denotes the complex conjugate of the wavelet,

$$\psi^*_{\tau,s}(t) = s^{-0.5}\psi\left(\frac{t-\tau}{s}\right).$$

If the scale parameter s is low, the sample is broken down into many small parts and the high frequencies are analyzed, while the low frequencies are explored when the parameter is high. See Schleicher (2002) and Crowley (2005) for a more detailed description of wavelets for economists.

Figure 11
Schematic comparison of wavelet and frequency filters



Source: Uliha (2016).

Wavelet filters have various specifications depending on the exact definition of the function $\psi()$. Furthermore, they can be classified into two main groups: There are continuous and discrete decompositions. In the case of discrete decomposition, the procedure clusters cycle lengths according to the powers of 2 (2–4, 4–8, 8–16 etc. quarters). We employ the MODWT (maximal overlap discrete wavelet transform) specification, which is the most often used wavelet transformation for examining economic time series. (For the exact formulas and a detailed description of this method, see, for example, Gallegati and Gallegati, 2007.)

Simpler forms of discrete wavelet filters require a time series length equal to a power of 2. The MODWT has no such strict constraint, but the end of the time series has to be extended artificially. From the various methods available, we chose the 'reflection' approach, which 'reflects' the time series in the last data point, repeating the observations in the reverse order. Further assumptions have to be made regarding the specific function of the wavelets, we use the Daubechies function typically employed in the literature.

Appendix C: Cycle lengths

First, using the wavelet filter, we decompose the credit-to-GDP time series into a trend, noise and four cyclical components with different cycle lengths. Table 20 shows the shares of the individual cyclical components from the sum of the absolute values of cyclical components. In the case of credit-to-GDP calculated with broad credit, the cycles of 16–32 years dominate in all countries, amounting to around 60–70 percent of the full cycle. The second greatest share (15–30 percent) is represented by cycles of 8–16 years. The proportion of the cycles of 2–8 years, corresponding to business cycles, is the lowest, at about 5–20 percent. The calculations with narrow credit produce very similar results. Only two countries exhibit significant differences: Belgium and Norway. Here, the cycles of 8–16 years dominate, accounting for 42 percent of the full cycle, while the longest cycles are the second most significant.

Table 20
Decomposition of credit gaps calculated with a wavelet filter into cycles of different lengths by countries (percent)

	Credit-to-GDP with broad definition of credit stock				Credit-to-GDP with narrow definition of credit stock			
	16-32 years	8-16 years	4-8 years	2-4 years	16-32 years	8-16 years	4-8 years	2-4 years
Core EU countries								
Austria	60.7	20.6	6.7	11.9	59.9	22.6	7.2	10.3
Belgium	62.2	14.9	8.6	14.3	32.9	42.1	13.7	11.3
France	69.7	18.5	5.7	6.1	62.3	24.9	7.4	5.3
Germany	70.3	18.4	4.4	6.9	74.1	18.7	3.6	3.6
Netherlands	57.8	22.9	9.6	9.6	60.2	22.4	8.2	9.2
United Kingdom	65.7	23.3	5.5	5.6	60.1	26.3	6.3	7.4
Nordic countries								
Denmark	65.9	24.9	4.7	4.5	62.9	28.3	5.7	3.2
Finland	57.7	24.5	8.9	9.0	73.7	20.4	2.9	3.0
Norway	55.2	22.9	12.3	9.6	39.6	41.9	12.0	6.6
Sweden	60.8	23.3	8.4	7.5	66.9	24.3	5.2	3.6
Mediterranean countries								
Greece	78.2	16.6	2.7	2.5	75.7	18.8	3.2	2.2
Italy	70.0	21.9	3.9	4.2	62.1	28.1	5.1	4.6
Portugal	68.5	22.3	5.3	4.0	73.0	21.5	3.7	1.8
Spain	74.2	20.9	3.4	1.5	70.0	25.0	3.8	1.2
Central and Eastern Europe								
Czech Republic	67.2	20.0	4.9	8.0	69.5	22.7	4.2	3.5
Hungary	68.9	22.8	3.4	4.9	70.5	22.2	3.1	4.2
Lithuania	59.0	29.8	7.2	3.9	64.4	28.3	5.1	2.2
Poland	67.8	16.3	8.6	7.3	67.0	19.5	7.8	5.6

Note: Yellow cells show the highest values for a given definition of the credit stock in each country.

Second, for each country, we compute the periodogram of the credit gap calculated with broad credit, CF filter, and maximum cycle length of 30 years. The Czech Republic, Poland and Lithuania are left out because their credit-to-GDP time series are much shorter than in other countries, so they would have considerably reduced the largest cycle length to be examined with a periodogram. Table 21 shows the results where the maximum length of a cycle is the largest possible limited by the sample size, while the other cycle lengths are half, a third, a quarter etc. of this. This maximum length is 50 years because the shortest credit-to-GDP time series with broad credit starts in 1970 Q4. The greater the value assigned by the periodogram to a given cycle length, the more dominant this cycle length is in the total gap.

This method also shows that the cycles significantly longer than business cycles, specifically, 17–25-year-long cycles dominate, and 25-year-long cycles have the largest values in the periodograms. The exceptions to the latter are Belgium, Denmark, Italy, the Netherlands, and the UK, where the values of the 17-year cycles are the greatest. Other interesting examples include Austria and Belgium, where the relatively shorter 10-year cycles are also important.

Table 21
Periodogram of credit gaps calculated with broad definition of credit stock and two-sided CF filter by countries

	Cycle length in years														
	50,00	25,00	16,67	12,50	10,00	8,33	7,14	6,25	5,56	5,00	4,55	4,17	3,85	3,57	3,33
Core EU countries															
Austria	13	1215	55	73	919	13	57	9	15	43	11	26	28	0	16
Belgium	169	27	1476	199	1472	730	561	174	264	137	28	243	597	160	183
France	35	1839	1062	53	494	238	67	32	87	15	0	35	19	14	5
Germany	14	3012	1298	43	92	95	43	2	11	40	18	56	22	10	38
Netherlands	276	1012	4072	1582	1407	958	42	260	318	348	192	80	96	3	62
United Kingdom	89	3272	11209	131	1113	416	7	15	215	15	16	4	71	67	89
Nordic countries															
Denmark	378	13837	14793	1508	1647	105	239	268	287	83	103	43	123	167	74
Finland	37	13610	7476	1541	2081	1164	1005	183	660	30	339	118	61	10	35
Norway	139	11816	3370	3468	3050	106	398	1324	1642	329	391	289	101	120	184
Sweden	49	11549	9961	1025	5246	500	372	218	86	718	25	184	475	41	16
Mediterranean countries															
Greece	465	8809	764	620	361	44	90	168	7	68	40	32	75	13	5
Italy	239	2384	4867	561	729	61	120	15	19	6	3	27	40	2	23
Portugal	905	24687	602	4539	6534	659	850	391	49	840	5	124	687	247	76
Spain	535	23887	10001	2387	1572	340	401	1281	15	206	103	148	139	49	145
Central and Eastern Europe															
Hungary	143	19797	3130	2463	1837	209	6	93	123	259	88	7	62	101	23

Note: The darkest green cells include the highest values from the periodogram in each country. The lighter the cell, the smaller the value in it compared to the maximum.

A similar analysis with narrow credit shows an even more homogenous picture (Table 22). As a result of the same calculations, the 22.5-year-long cycles are the most important in all countries, except for Norway. (The longest cycle that can be examined on credit-to-GDP time series with narrow credit is 45 years.) In the case of Austria, Belgium and the Netherlands, there are relatively significant shorter cycles (9–15 years) as well, but these are all still longer than typical business cycles.

Table 22
Periodogram of credit gaps calculated with narrow definition of credit stock and two-sided CF filter by countries

	Cycle length in years														
	45,00	22,50	15,00	11,25	9,00	7,50	6,43	5,63	5,00	4,50	4,09	3,75	3,46	3,21	3,00
Core EU countries															
Austria	34	573	432	275	133	6	46	1	22	27	40	3	10	8	6
Belgium	18	1026	633	246	729	127	44	9	12	13	0	23	29	3	8
France	7	1478	427	269	254	0	9	18	1	4	6	3	1	0	0
Germany	38	2525	562	80	90	8	9	9	35	3	17	4	7	15	9
Netherlands	260	1444	495	994	15	166	194	103	35	22	15	12	10	16	32
United Kingdom	32	3462	1261	633	404	29	55	8	33	31	7	9	44	15	19
Nordic countries															
Denmark	186	14478	5128	3562	303	112	296	447	8	11	5	19	25	58	17
Finland	41	4815	1353	221	80	0	19	25	2	29	8	8	27	1	9
Norway	36	1543	3073	907	1089	44	75	29	20	41	92	7	4	1	10
Sweden	26	11796	2938	1210	607	27	15	9	115	27	41	57	32	25	13
Mediterranean countries															
Greece	1535	7873	1224	355	28	67	155	26	24	18	8	87	34	2	51
Italy	141	2528	1183	315	388	94	22	52	73	90	23	4	17	11	4
Portugal	1331	11772	2136	2086	2499	196	341	11	45	16	26	43	101	102	2
Spain	1327	21581	10418	2398	1010	641	594	6	35	138	47	13	23	21	18
Central and Eastern Europe															
Hungary	234	8422	973	687	103	65	28	29	35	17	4	5	11	5	11

Note: The darkest green cells include the highest values from the periodogram in each country. The lighter the cell, the smaller the value in it compared to the maximum.

The two-sided CF filter identifies credit gaps more accurately, but this study is interested in typical cycle lengths for one-sided credit gaps. Therefore, we also calculate the periodograms of one-sided credit gaps (Table 23 and Table 24).

Table 23
Periodogram of credit gaps calculated with broad definition of credit stock and one-sided CF filter by countries

	Cycle length in years														
	48,00	24,00	16,00	12,00	9,60	8,00	6,86	6,00	5,33	4,80	4,36	4,00	3,69	3,43	3,20
Core EU countries															
Austria	0	168	12	44	158	2	4	0	1	13	2	8	1	2	4
Belgium	21	87	210	86	255	216	50	13	60	17	25	76	103	23	22
France	0	377	182	25	106	28	3	4	6	3	1	9	1	2	1
Germany	22	564	355	10	17	19	13	1	2	6	5	22	2	2	6
Netherlands	124	53	732	574	89	196	28	54	123	9	9	16	15	3	19
United Kingdom	34	1359	1758	15	329	51	7	10	22	6	9	18	11	5	8
Nordic countries															
Denmark	27	3182	2517	497	257	15	105	6	16	23	3	8	25	21	51
Finland	62	3661	1586	329	420	249	104	67	120	49	59	13	12	10	3
Norway	110	2991	841	688	600	103	45	151	364	56	112	30	73	63	9
Sweden	6	3978	1791	492	1080	41	56	61	113	61	11	82	55	1	24
Mediterranean countries															
Greece	52	1351	95	110	66	3	10	27	2	23	15	3	15	4	4
Italy	20	1077	687	69	152	11	26	1	2	2	1	9	2	1	0
Portugal	43	4256	508	371	1480	10	70	84	48	92	13	58	18	16	26
Spain	41	3669	1345	302	379	38	22	148	5	16	12	6	3	4	2
Central and Eastern Europe															
Hungary	26	3811	540	604	262	46	1	11	31	37	11	1	22	8	8

Note: The darkest green cells include the highest values from the periodogram in each country. The lighter the cell, the smaller the value in it compared to the maximum.

Table 24
Periodogram of credit gaps calculated with narrow definition of credit stock and one-sided CF filter by countries

	Cycle length in years														
	45,00	22,50	15,00	11,25	9,00	7,50	6,43	5,63	5,00	4,50	4,09	3,75	3,46	3,21	3,00
Core EU countries															
Austria	2	53	50	35	46	3	13	0	5	5	9	1	1	1	3
Belgium	8	170	103	47	143	33	6	2	3	2	0	6	6	0	1
France	1	399	122	77	51	0	3	3	0	1	1	1	0	0	0
Germany	24	505	215	4	24	9	4	3	4	3	9	0	1	2	4
Netherlands	10	276	72	160	2	46	50	29	6	2	4	1	3	2	5
United Kingdom	0	821	312	116	103	6	15	2	3	3	2	1	6	1	4
Nordic countries															
Denmark	20	2980	1259	737	101	39	41	64	2	2	1	2	3	9	1
Finland	29	1133	362	41	31	1	8	6	2	10	1	3	7	0	2
Norway	2	473	759	190	279	17	13	4	6	9	18	2	1	0	1
Sweden	80	3392	833	330	127	20	15	8	17	14	4	19	10	4	3
Mediterranean countries															
Greece	139	1136	185	85	12	14	36	15	8	9	4	19	6	0	9
Italy	6	554	178	30	61	5	3	3	5	8	1	2	1	1	1
Portugal	152	1774	270	279	517	14	18	11	13	4	1	3	10	12	2
Spain	97	3336	1897	462	332	43	61	16	3	10	7	2	1	2	2
Central and Eastern Europe															
Hungary	25	1827	288	200	39	10	5	10	8	2	0	2	1	1	1

Note: The darkest green cells include the highest values from the periodogram in each country. The lighter the cell, the smaller the value in it compared to the maximum.

The results about the relative importance of various cycle lengths with the one-sided and two-sided approaches are almost identical across countries and across credit definitions. However, the one-sided approach produces smaller coefficients in all countries than the two-sided approach. This means that while the same cycle lengths are typical in both credit gaps, the one-sided approach assigns a smaller share of the changes in credit-to-GDP to the cyclical part and a higher share to the trend than the two-sided approach.

Appendix D: Univariate signaling approach

The early warning accuracy of an indicator can be assessed not only with the signaling approach but also with discrete choice models. These usually mean the application of logit or probit models, where the dependent variable is typically 1 before the crisis, and otherwise 0. Compared to the signaling approach, the advantage of these is that the statistical significance of the indicator's prediction performance can be directly evaluated. Additionally, they also provide the estimated probability of a crisis in the near future. However, a serious drawback is that no crisis probabilities can be observed, so it is difficult to say how accurate the signal of the indicator is in the given quarter. We consider this a considerable limitation from the practical point of view of macroprudential policy, therefore in this paper, we follow the signaling approach, which is not affected by this problem.

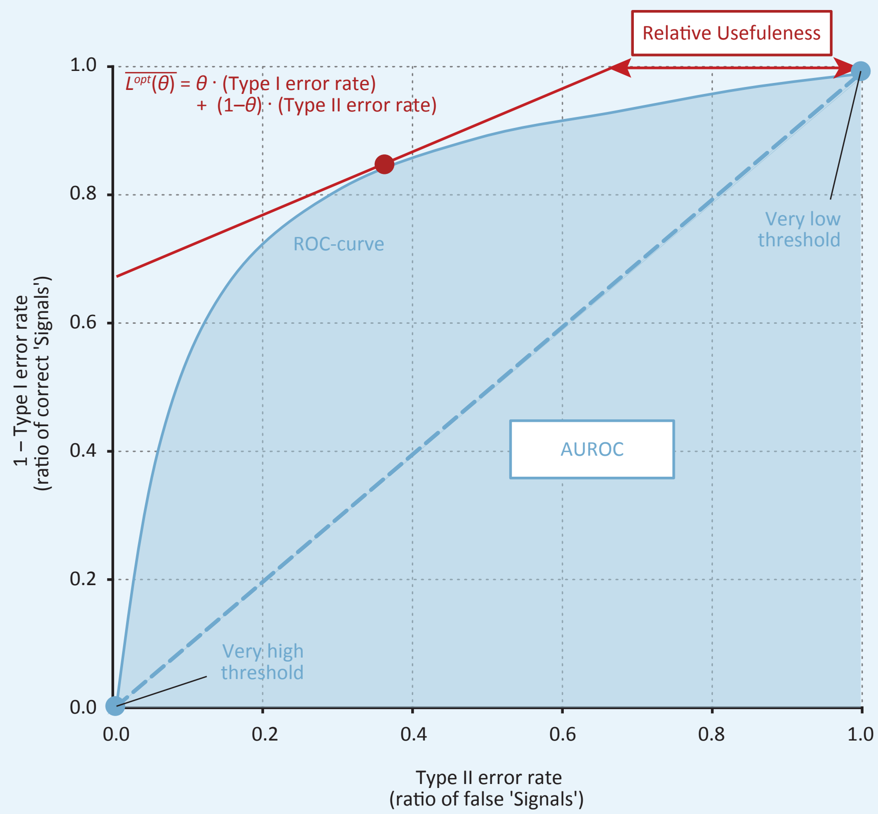
The crisis probabilities estimated by discrete choice models can also be treated as an indicator to which the signaling approach can be applied. In this case, estimated probabilities issue a warning signal when they exceed an appropriate threshold value. We reject this composite method because the estimated crisis probabilities are positive monotonic transformations of the credit gaps, so the composite method produces exactly the same set of time series of binary signals as the use of the signaling approach directly to the credit gaps. However, the probit and logit models can be readily employed for crisis prediction with multiple indicators, which is actually their typical application, see, for example, Frankel and Rose (1996), Demirgüç-Kunt and Detragiache (2000), Davis and Karim (2008), Lo Duca and Peltonen (2013), Jordà et al. (2015), Pönkä (2017), Antunes et al. (2018), Borio et al. (2018), Sondermann and Zorell (2019).

Table 25
Number of correct and false signals

		State	
		Crisis	No crisis
Signal	Signal	A	B
	No signal	C	D

In Section 4, where we use the signaling approach, we elaborate on what signals an ideal early warning indicator should issue. So, the number of country-quarter observations with correct and false signals can be computed for every credit gap time series and threshold value (Table 25). The number of correct 'Signal' values is denoted by A, while the number of correct 'No signal' values is denoted by D. The credit gap commits a Type I error of missing the impending crisis in C number of cases, and a Type II error of falsely signaling a crisis in B number of cases. From these, the ratio of false 'Signals' (B⁄(B+D)) and the ratio of correct 'Signals' (A⁄(A+C)) can be defined for all potential thresholds. These points plot out a curve connecting point (0,0) and point (1,1), which is referred to as a ROC ('receiver operating characteristic') curve (Figure 12). There are hardly any 'Signal' values with sufficiently high thresholds, so both ratios are close to zero. There are almost only 'Signal' values with sufficiently low thresholds, so both ratios are close to one. The more accurate early warning signals the credit gap can issue, the closer the two ratios can be to the point (0,1) with moderately high thresholds. So, the area under the ROC curve, the AUROC ('area under the receiver operating characteristic curve') value is large when the examined credit gap is a good predictor. The AUROC value of credit gaps producing a perfect prediction is 1, while the credit gaps that are completely uninformative have an AUROC of 0.5. In Section 4, credit gaps are ranked mostly based on the AUROC, which evaluate credit gap performance taking into account all potential thresholds.

Figure 12
Relationship between Type I errors, Type II errors, AUROC criterion, and RU criterion



Note: The figure applies to the values of preference parameter θ between 0.5 and 1.

We also use another criterion, the RU ('relative usefulness') value. A specific threshold has to be chosen to define this, which is usually done by assuming some preferences over the Type I and Type II errors. These preferences can be described for example with the following loss function, first used by Demirgüç-Kunt and Detragiache (2000) for predicting crises:

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

The parameter θ that can be chosen between 0 and 1 characterizes the relative sensitivity to Type I error rate compared to Type II error rate. Systemic banking crises cause large social losses, so macroprudential authorities are particularly concerned about missing an impending crisis. Therefore, the parameter value is usually at least 0.5,[61] which we follow in this paper.[62] The optimal threshold minimizes the loss *L(θ)* (the minimum value is $L^{opt}(\theta)$). The relationship between the Type I error rate (*C/*(*A+C*)) and the ratio of correct 'Signals' (*A/*(*A+C*)) can be written as follows: *C/*(*A+C*)=1-*A/*(*A+C*). Therefore, the point where the ROC curve just touches the indifference curve related to the loss $L^{opt}(\theta)$ determines the optimal threshold (Figure 12). The RU value based on the idea by Alessi and Detken (2011) is a normalized version of the least possible loss:

$$RU(\theta) = \frac{\min\{\theta; 1-\theta\} - L^{opt}(\theta)}{\min\{\theta; 1-\theta\}}$$

[61] See for example Babecký et al (2014), Lo Duca and Peltonen (2013), Detken et al (2014), Tölö et al (2018), Lang et al (2019).
[62] While comparing the RU values derived with different parameters *θ*, it should keep in mind that the value *RU(θ)* declines monotonically in the parameter *θ* if *θ* is between 0.5 and 1.

The maximum possible value of RU is 1, reached when both types of error rates are 0. Completely uninformative credit gaps produce a RU of 0.

Sarlin (2013) proposes slightly different preferences and RU. In this version, the loss function does not depend solely on the Type I and Type II error rates within the corresponding states of the world (whether there is an impending crisis or not), but also on the unconditional probabilities of the states of the world. The concrete specification is as follows:

$$L'(\mu) = \mu P \frac{C}{A+C} + (1-\mu)(1-P)\frac{B}{B+D}$$

Here, μ is the preference parameter and P is the ratio of the number of country-quarters falling into the prediction period to the total number of country-quarters. Accordingly, $P=(A+C)/(A+B+C+D)$, so

$$L'(\mu) = \mu \frac{C}{A+B+C+D} + (1-\mu)\frac{B}{A+B+C+D}$$

According to this loss function, policymakers are interested in the unconditional error rates of the two types of errors rather than the error rates conditional on the state. Systemic banking crises occur rarely, so the balanced preferences for conditional error rates (θ=0,5) are analogous to the preferences over unconditional error rates that assign much more weight to missing a crisis (μ>0,5). Consequently, the parameters θ that are significantly greater than 0.5 correspond to preferences over unconditional error rates that are very sensitive to missed crises.

Appendix E: In-sample early warning performance with RU criterion by country group

Table 26 presents the RU values calculated with balanced preferences (a preference parameter of 0.5) in the different country groups. The set of best credit gaps comprises credit gaps that have RU values calculated with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters that are no more than 12 percent lower than the respective highest RU values.

The results exhibit a heterogeneity across country groups, which is very similar to the heterogeneity obtained with AUROC (Section 4.3.2). Here, credit gaps are also much worse at predicting systemic banking crises in core EU countries than in Nordic and Mediterranean countries. Like with AUROC, narrow credit outperforms broad credit mostly in Nordic countries, just like in core EU countries, albeit with smaller differences, and no clear-cut ranking is seen in Mediterranean countries. Out of the top three credit gaps from the baseline analysis, the HP filter and CF filter credit gaps are included among the best few credit gaps in two country groups each (although not in the same country groups as when assessed with the AUROC). Apart from these two, only the long-cycle version of the CF filter credit gap can achieve the best credit gap status in at least two country groups. In the analysis with the AUROC, the medium-term cycle version of the HP filter credit gap attains this as well. Like with AUROC, the wavelet filter credit gap is not included among the best credit gaps in any country group.

Table 26
RU values of credit gaps with balanced preferences by country groups

Using all crisis periods, we expect "Signal" in 16-5 quarters before crises.			**Gap calculation**							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
core EU countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	0.25	0.25	0.21	0.24	**0.29**	0.25	0.25	0.20
		1-year forecast	0.23	0.25	0.19	0.18	0.23	0.20	0.25	0.21
	Broad definition of credit stock	0-year forecast	0.22	0.16	0.11	0.24	0.17	0.18	0.14	0.17
		1-year forecast	0.21	0.16	0.12	0.19	0.18	0.19	0.15	0.16
Nordic countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.60**	**0.60**	0.54	**0.67**	**0.65**	**0.61**	0.51	0.52
		1-year forecast	0.60	0.57	0.47	0.28	0.40	0.22	0.52	0.53
	Broad definition of credit stock	0-year forecast	0.47	0.48	0.39	0.56	0.51	0.42	0.51	0.44
		1-year forecast	0.42	0.41	0.30	0.19	0.16	0.15	0.51	0.40
Mediterranean countries										
Credit-to-GDP specification	Narrow definition of credit stock	0-year forecast	**0.71**	0.54	0.43	**0.72**	0.59	0.48	0.63	0.58
		1-year forecast	**0.69**	0.51	0.40	0.67	0.44	0.39	0.63	0.55
	Broad definition of credit stock	0-year forecast	**0.64**	0.59	0.46	**0.67**	0.65	0.47	**0.67**	0.58
		1-year forecast	**0.64**	0.58	0.49	0.66	0.57	0.45	**0.68**	0.56

Note: RU values are calculated assuming balanced preferences over Type I and Type II error rates. (The value of the preference parameter θ is 0.5. For more details, see Appendix D.) Red cells show the highest RU values of various country groups. Grey cells include RU values that are lower than the respective highest values by no more than 12 percent. Credit gaps belonging to cells with bold numbers have RU values calculated with each of the prediction horizons of 20–5, 16–5, and 12–5 quarters that are no more than 12 percent lower than the respective highest values. Cells with thick borders refer to the best three early warning credit gaps according to the baseline analysis.

Appendix F: Endpoint uncertainty by country group

Table 27
Average change in credit gaps due to 1, 2, 3 and 10 years of new data by country groups (percent)

Using data between 1983 Q1 and 2008 Q1			Gap calculation							
			Hodrick–Prescott			Christiano–Fitzgerald			Wavelet	
			32 years	25 years	19 years	30 years	24 years	18 years	32 years	16 years
core EU countries										
Narrow definition of credit stock	0-year credit-to-GDP forecast	1-year revision	23	31	45	26	29	35	29	54
		2-year revision	43	57	79	46	51	60	59	105
		3-year revision	60	76	101	57	61	70	86	140
		10-year revision	101	109	124	72	73	83	176	184
	1-year credit-to-GDP forecast	1-year revision	21	27	36	28	30	35	32	55
		2-year revision	39	48	61	42	44	48	60	94
		3-year revision	53	64	77	51	52	54	85	119
		10-year revision	88	89	93	61	60	66	155	139
Nordic countries										
Narrow definition of credit stock	0-year credit-to-GDP forecast	1-year revision	28	36	51	25	27	39	19	55
		2-year revision	53	66	91	49	52	73	41	109
		3-year revision	74	91	121	66	68	93	63	149
		10-year revision	116	128	154	87	88	116	132	184
	1-year credit-to-GDP forecast	1-year revision	26	32	42	28	29	40	25	62
		2-year revision	49	58	74	46	46	61	48	107
		3-year revision	67	78	95	59	58	72	70	135
		10-year revision	100	105	118	73	72	89	124	153
Mediterranean countries										
Narrow definition of credit stock	0-year credit-to-GDP forecast	1-year revision	23	29	41	21	28	42	14	40
		2-year revision	45	54	73	42	56	83	33	85
		3-year revision	64	74	96	58	75	107	52	121
		10-year revision	120	119	128	86	101	121	128	167
	1-year credit-to-GDP forecast	1-year revision	22	25	33	23	30	43	20	49
		2-year revision	41	47	58	39	50	67	40	86
		3-year revision	57	63	75	52	64	82	59	112
		10-year revision	101	94	93	69	79	91	122	135

Note: Red numbers denote values of the smallest average revisions for a given country group with a given number of additional observations used. Yellow numbers denote the same, if only the three credit gaps are compared to each other that proved to have the best early warning properties using European data in Section 4. Cells with thick borders refer to the best three early warning credit gaps.

MNB OCCASIONAL PAPERS 142

EARLY WARNING PERFORMANCE OF UNIVARIATE CREDIT-TO-GDP GAPS

September 2021

Print: Prospektus Kft.

H-8200 Veszprém, Tartu u. 6.

mnb.hu

H-1013 BUDAPEST, KRISZTINA KÖRÚT 55.