The signaling approach to early warning:

Application for systemic banking crises

Mathias Dahle Bryde-Erichsen

Master of Philosophy in Economics

Department of Economics

University of Oslo

January 2016

Abstract

With a growing focus on macroprudential policy in the aftermath of the financial crisis of 2007/2008, there is a need for early warning systems. The object of the thesis is to present a toolbox for signaling systemic banking crises that can be applied to policy. To this end I evaluate the existing methodology, identify the best performing early warning indicators as well as their optimal threshold values.

The noise-to-signal ratio has been a workhorse of the signaling approach since the seminal papers of Kaminsky et al. (1997) and Kaminsky and Reinhart (1999), yet I will show that this may not be an appropriate tool for finding optimal thresholds. I will instead evaluate the signaling performance of indicators based on measures that either takes explicit account of the preferences of the policy maker or incorporate the full range of possible threshold values. The thesis also shows that country specific threshold values given as the percentile of the distribution seems to be best suited for Norwegian data

In line with most of the existing literature, the private credit to GDP gap is found to be the best performing single indicator, closely followed by private credit exuberance. Both indicators also produce stable threshold values for probable ranges of the policy makers relative preference between correctly and falsely signaling crises.

With the use of two indicators for signaling, more than one signaling scheme can be used to define the signal. The standard approach in the literature has been to require both indicators to breach their respective threshold values for a signal to be issued. I will in this thesis present an alternative scheme that will be shown to significantly increase the signaling performance in a bivariate analysis, compared with the standard scheme. The best performing pair of indicators is found to be private credit exuberance and the global house price to income gap.

Acknowledgments

This thesis has only been possible through my time in Norges Bank. In that regard I would like to thank everyone in the macroprudential unit for creating an inspiring and engaging work environment for me as a student. I would especially like to thank Karsten Gerdrup for introducing me to the signaling approach, and for the help along the way, as well as Frank Hansen for help with data. I would also like to thank Jan-Hannes Lang and Peter Welz from the European Central Bank for help with codes for bivariate thresholds.

Contents

1 Introduction

Following the financial crises that erupted in 2007/2008, a macroprudential approach to financial regulation has emerged, with focus on the systemic risk of the financial system (Borchgrevink et al., 2014). To help in this vein, there is a steadily growing literature on early warning systems (EWS), where the object is to predict upcoming crises. One of the EWSs that is applied is the signaling approach to early warning, where the values of one or more indicator variables are translated into a binary signal for upcoming crises. Systemic banking crises are generally understood to follow imbalances in the financial system, and in that sense the signaling approach is intuitive. Given a pre-defined threshold value of the indicator variable, a crisis is signaled whenever the indicator breaches its threshold, i.e. when the indicator takes a large enough value. In this thesis I will present the signaling approach as an early warning system for systemic banking crises and look more closely at how it can be applied to policy.

The signaling approach for indicator evaluation was first used as an early warning system for banking crises by Kaminsky and Reinhart (1999), where they investigate currency and banking crises and the link between the two in so called twin crises. Kaminsky et al. (1997) used the same concept to evaluate indicators for currency crises exclusively. These papers seek to find the best indicators to signal upcoming crises, and evaluate the indicators by how many crises they are able to signal and by their noise-to-signal ratio. The noise-to-signal ratio, defined as the ratio of falsely signaled crises to correctly signaled crises, has been a workhorse of the signaling literature since the start. The threshold values are found as those that minimize the said noise-to-signal ratio for each indicator. Kaminsky and Reinhart (1999) test a multitude of financial sector, external sector and real sector variables, among them domestic credit as a percentage of GDP. In their work they don't find this to be among the best indicators of banking crises, but this indicator has come to dominate the literature on early warning systems for systemic banking crises, and will be one of the indicators of this thesis.

Borio and Lowe (2002) investigate the role of asset prices, along with credit, in the build up to a crisis. They follow Kaminsky and Reinhart (1999) by applying a signaling approach to early warning and the minimization of the noise-to-signal ratio to find threshold values, but they expand on the latter paper in multiple ways. While Kaminsky and Reinhart (1999) mainly focused on the twelve month growth of the indicators, Borio and Lowe (2002) measure the data as deviations from a trend, calculated by a one sided HP filter. In addition, they explore the signaling ability of the indicators for different horizons, namely one, two and three years prior to a crisis. Lastly, the paper introduces a bivariate analysis of the indicators, where a signal is issued if two indicators simultaneously breach their respective threshold values. Among their indicators they find that the so called credit gap performs best in the univarate analyses. When combining indicators in a bivariate setting they find the noise-to-signal ratio to be reduced, but at the cost of fewer crises detected.

Borio and Drehmann (2009) continue to expand on the methodology related to the signaling approach. Amongst other contributions, they introduce two new methods for finding the optimal threshold values. The first is the minimization of a loss function, based on the method of Demirgüç-kunt and Detragiache (1999), where the type I and type II error rates are weighted by a preference parameter, θ. The second is the minimization of the noise-to-signal ratio, but conditional on that at least a given proportion of the crises are signaled. Borio and Drehmann (2009) then test the indicators' performance out of sample in signaling the financial crisis of 2007/2008, with the conclusion that:

> *"The out-of-sample performance is not an unqualified success"*

In their paper, Drehmann and Juselius (2014) evaluate early warning indicators by comparing the indicators in a new way. While previous papers have found optimal threshold values through minimization of the noise-to-signal ratio or a loss function, Drehmann and Juselius (2014) acknowledge the difficulty in assessing the costs of a crisis or of implementing countermeasures, or that of quantifying a policy makers preferences between the two. In addition, previous work has not been able to compare indicators in a clearcut quantitative way. The authors therefore introduce a new measure, called the AUROC.

As the threshold values of an indicator are varied, the corresponding number of correctly and falsely signaled crises also varies. The new measure evaluates the indicators based on their performance in signaling crises for all threshold values. The paper then introduces three criteria to evaluate the indicators based on this measure. These are the the timing of an indicator, the stability of an indicator, and lastly the ranking among indicators. Using these criteria they find that private credit to GDP, measured as deviation from trend, calculated by a one-sided HP filter, has the best signaling performance for long horizons, while the debt service ratio dominates in the short run.

Detken et al. (2014) seeks to operationalize the countercyclical capital buffer. They do this by investigating different approaches to early warning, among them the signaling approach. For the most part they present and use the evaluation tools already presented, but an innovation is the partial standardized AUROC, which is a modification of the measure described in the previous paragraph where some conservative assumptions about the policy makers preferences are made to enhance the performance of the measure.

The work presented so far is far from exhaustive when it comes to the literature concerning the signaling approach, but it illustrates some of the aspects of evaluating indicators within the signaling framework. The contribution of this thesis will be to give a more holistic and thorough description of the signaling approach and its challenges. I will also introduce a signaling scheme that increases the signaling performance in bivariate analyses. Using the described framework, I will find the indicators that perform best in-sample along with their optimal threshold values, with emphasis on the optimal thresholds for Norway. To perform the calculations of the thesis I have used the program MATLAB, and developed a class for signaling analyses called IndicatorEval[1].

In section 2 I will present the variables and data to be used throughout the thesis. Firstly, the crises will be defined and their start and end dates given. Secondly, the variables to be evaluated as indicators and the data used for this purpose will be presented. Section 3 presents the methodological framework, which will be given in four parts. As many of

[1] The code is available upon request

the indicators will be expressed in terms of their deviation from a trend, the HP filter used to generate the gap variables will be presented first. In the second part I present the general idea of the signaling approach and its workings. Part three gives a more thorough look at the evaluation techniques that will be applied to the indicators. In the last part the different ways of defining a signal criteria for one and two indicators are presented. Section 4 looks more closely at the assumptions that are made and parameters that are chosen, specifically the choice of the policy makers preferences, the setting of the threshold values and the signal horizon. After this, in section 5, the results of the analyses done using the methods of the earlier sections are presented. In addition, the best performing indicators and indicator combinations are found, along with their optimal thresholds. Section 6 concludes.

2 Data

The dataset that will be used in this thesis comprises the following 20 countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, Korea, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, USA. This is a fairly homogenous group of countries, with most of them being European. Keeping the dataset to a group of advanced economies is supported by Drehmann and Tsatsaronis (2014). They evaluate the credit to GDP gap for two group of countries, one comprising advanced economies and the other of emerging market economies. Their findings show that the credit to GDP gap performs differently for the two groups, with the indicator performing best as an early warning signal for the advanced economies.

The data is gathered for the period from the first quarter of 1970 to the last quarter of 2014. The data is not complete for all countries or all variables, and the series capetures varying amounts of crises with the most comprehensive catching 33 crises and the least 23. The details of the data will be laid out in this section, with the crisis definition and crises observations first, followed by a description of the indicator variables and data.

2.1 Crises

The crises relevant to the thesis are systemic banking crises. As stated by Davis and Karim (2008):

> *"Even if systemic crises unambiguously occur, identifying their starting and ending dates is hazardous and the same episode may have a different duration in different studies. Where runs do not occur and banking system data are either unavailable or unreliable, locating the exact time when the system became insolvent is impossible."*

This leads to a variety of different definitions of a crisis in the literature, but also to the reuse of previously defined crises. The crises dating in this thesis is based on Anundsen et al. (2015). There, the dates are drawn from multiple sources, which will be presented next.

Some dates are provided by Reinhart and Rogoff (2008, 2009a,b), but these papers again base their crises dating on multiple sources, among them Kaminsky and Reinhart (1999), with the following definition. The beginning of a banking crisis is marked by one of two events. The first is bank runs that lead to the closure, merging or takeover by the public sector of one or more financial institutions. The second is if there are no runs, the closure, merging, takeover, or large-scale government assistance of an important financial institution (or group of institutions) that marks the string of similar outcomes.

Crises dates are also based on Laeven and Valencia (2008, 2010, 2012). In the latter, the definition of a banking crisis is as follows. A banking crisis is defined as systemic if the two following conditions are met. Firstly, there must be significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system and/or bank liquidations). Secondly, there must be significant banking policy intervention measures in response to significant losses in the banking system. For policy interventions to be considered significant three out of the following six measures must have been used: extensive liquidity support (5 percent of deposits and liabilities to nonresidents), bank restructuring gross cost (at least 3 percent of GDP), significant bank nationalization, significant guarantees put in place, significant asset purchases (at least 5 percent of GDP) or deposit freezes and/or bank holidays. The start of a crisis is defined as the first year in which both criteria are met.

Based on, among others, the papers already presented for crisis dating, Babecky et al. (2014) compose a binary occurrence index for banking crises, which takes the value 1 if at least one of its sources claims that a crises occurs. In addition to this the authors conduct a survey among country experts, mostly from the national central banks, for all countries in the sample. This adds two features to the database. Firstly, the country specific issues are best known by the country experts which can amend the original findings. The second feature is that crises have in the past been dated mostly on an annual basis. Babecky et al. (2014) date the crisis quarterly and this is made more precise with the help of country experts.

As can be seen, there are multiple ways of defining a crisis, and although one tries to implement quantitative criteria in the definitions, discretion will always have a place in the dating. The exact crises dates for the different countries of the dataset are given in table 1, while figure 1 shows how these crises are distributed over time. As expected, the dataset shows a lot of systemic banking crises during the financial crises of 2007/2008, with 14 registered crises. There is also see a cluster of crises during the late 1980s early 1990s.

Table 1: Crises dates. Based on Anundsen et al. (2015)

Australia	1989Q4-1992Q4		
Austria	2008Q3-2013Q4		
Belgium	2008Q3-2013Q4		
Canada	1983Q1-1985Q4		
Switzerland	1991Q1-1994Q4	2008Q3-2012Q4	
Germany	1977Q1-1979Q4	2001Q1-2003Q4	
Denmark	1987Q1-1993Q4	2008Q3-2012Q4	
Spain	1978Q1-1985Q3	2008Q3-2013Q4	
Finaland	1991Q3-1995Q4		
France	1993Q3-1995Q4	2008Q3-2012Q4	
UK	1973Q4-1975Q4	1990Q3-1994Q2	2007Q3-2012Q4
Greece	2008Q3-2013Q4		
Italy	1994Q1-1995Q4	2008Q3-2012Q4	
Japan	1992Q1-2001Q4		
Korea	1997Q3-1998Q4		
Netherlands	2002Q1-2003Q4	2008Q3-2012Q4	
Norway	1988Q2-1993Q3	2008Q3-2009Q3	
Portugal	1999Q1-2000Q1	2008Q3-2013Q4	
Sweden	1990Q3-1993Q4	2008Q3-2010Q4	
USA	1988Q1-1990Q4	2007Q3-2013Q4	

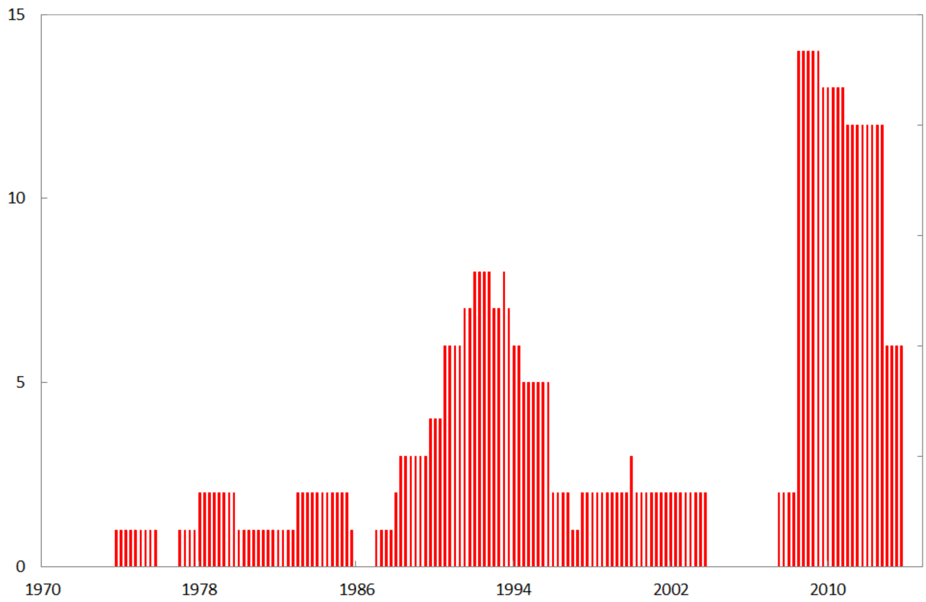


Figure 1: *The distribution of systemic banking crises over time in the sample. 1970Q1 - 2014Q4*

2.2 Variables and data

To identify financial imbalances, Norges Bank focuses particularly on four key indicators (Norges Bank, 2013), which are the ratio of total credit to GDP[2], the ratio of house prices to household disposable income, commercial property prices and the wholesale funding ratio of Norwegian credit institutions. Since the object of this thesis is to apply the signaling approach as an early warning system, and especially for Norway, it would be preferable to be able to include all of these variables in the dataset. Unfortunately, because of the lack of available data, it is not possible to include commercial property prices. On the other hand, there is available data on the other three indicators, so these will enter in the analyses. The ratio of total credit to GDP will from now on be referred to as private credit to GDP, and the ratio of house prices to disposable income as house price to income. Data on the decomposition of private credit to GDP, namely household credit to GDP and non-financial enterprise credit to GDP, is also included. Along with

[2]Total credit is here given by credit to households and non-financial enterprises, which in the Norwegian case comprises C2 households and C3 enterprises. Both credit and GDP are measured for mainland Norway

the mentioned variables five other variables will be included, four of which are calculated by Anundsen et al. (2015). The first variable is another banking variable, namely the equity ratio. The next two are measures of exuberance, or bubbles, in house prices and private credit. The last two variables are measures of global private credit to GDP and global house price to income. In table 2 some key statistics of the data for the indicators are presented. In column (1) we see the number of countries that enter in the dataset for each variable. (2) gives the total number of observations. (3) shows the number of crises that are covered by each variable. Lastly, (4) and (5) gives the timing of the first and last observation for each variable. The indicator variables are presented graphically in figures 2 and 3. In the former we can see the variables time series for Norway, where the shaded areas indicate systemic banking crises. Figure 3 on the other hand shows the behavior of the variables in the periods around the outbreaks of a crisis for the whole sample, more specifically from 20 quarters prior to, to 20 quarters following the outbreak. The solid line is the mean of the sample, while the dotted lines gives one standard deviation. The data sample for the indicators is the same as in Anundsen et al. (2015)

Private credit to GDP is the most widely applied indicator for early warning of banking crises. In the data sample the credit data is gathered from the Bank for International Settlements (BIS) and comprises credit to non-financial enterprises (both privately and publicly owned) and household credit, which is composed of credit to both households and non-profit institutions serving household. As previously stated the two components of private credit to GDP are also used as individual indicators. The data for the GDP is nominal GDP, gathered from the Organisation for Economic Co-operation and Development (OECD). The three indicator variables are all represented by their deviation from a trend, where the trend is calculated by a one sided HP filter, and the exact method will be presented in section 3.1. The time series for Norway and for the periods around crises can be seen in figures 2a, 2b, 2c and figures 3a, 3b, 3c respectively. From table 2 it is clear that private credit to GDP holds the most comprehensive data of the sample, starting in the 1970 Q1, ending in 2014 Q4 and covering all crises with a total of 3494 observations.

Regarding the role of house prices and financial stability Borio and Lowe (2002) state that:

> *"...asset prices stood out in historical accounts of financial instability ... In these accounts it is property prices in particular that have been highlighted..."*

In the aftermath of the financial crisis of 2007/2008 this role can not be said to have been diminished. As a measure of house prices I follow Norges Bank (2013) and Anundsen et al. (2015) in using house price to income. The data for house prices and disposable income are gathered from the International House Price Database at the Federal Reserve Bank of Dallas. For countries not covered by this database, the data is supplemented with similar measures collected for the OECD. As with the credit indicators, house price to income is given as the deviation from trend, calculated by a one sided HP filter, and will be referred to as the house price to income gap. Figure 2d and 3d shows the time series for Norway and around crises for the full sample. The house price to income series is slightly shorter than those of the credit series, having the first observation in 1975 Q1, but the only crisis not covered by the data is that of the UK starting in 1973 Q4.

The non-core (wholesale) funding for banks is defined as total assets less customer deposits and bank equity. Dividing the non-core funding with the banks' total assets gives the wholesale funding ratio. The representation of the indicator will be as deviation from trend, again using a one sided HP filter. The new variable, the equity ratio, is defined as the end-of-year amount of capital and reserves in the banking sector as a share of total assets. This will not be given as a gap variable. The data for both variables are obtained from the OECD Banking Statistics, which provides annual data on the different components in banks' assets and liabilities for most of the countries included in the sample of this thesis[3]. As the data of the OECD Banking Statistics is annual, linear interpolation methods are used to convert the data to quarterly series. Figures 2e and 3e shows the time series for Norway and around crises for the full sample for the wholesale funding ratio, while 2f and 3f are for the equity ratio. The database was discontinued in 2009. This, combined with the four missing countries, explain the numbers seen in table 2. The wholesale funding ratio gap and the equity ratio have the lowest amount of observations

[3]Exceptions are the countries Austria, Greece, Portugal and the UK

in the sample, with 1692 in total, and fewest crises covered with 23.

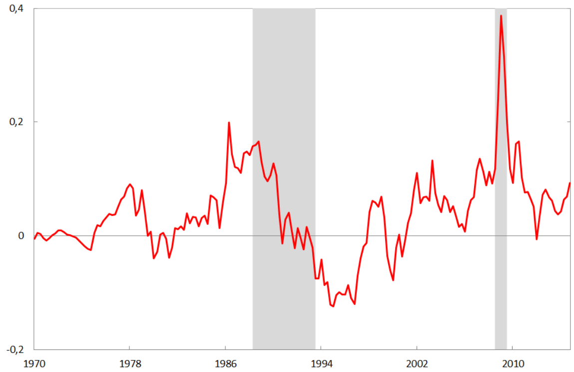
Table 2: Key statistics of the data for the indicator variables. Number of countries included, number of observations, number of crises covered, time of first observation and time of last observation.

Indicator	(1) Countries	(2) Observations	(3) Crises	(4) First	(5) Last
Private credit/GDP	20	3494	33	1970Q1	2014Q4
Household credit/GDP	20	2840	30	1970Q1	2014Q4
NFE credit/GDP	20	2816	29	1970Q1	2014Q4
House price/Income	20	2888	32	1975Q1	2014Q2
Wholesale/assets	16	1692	23	1979Q4	2009Q4
Equity ratio	16	1692	23	1979Q4	2009Q4
Credit exuberance	20	2774	30	1978Q4	2014Q4
HP exuberance	20	2152	27	1983Q1	2013Q4
Global credit	17	3054	28	1970Q1	2014Q4
Global HP/Income	17	2585	27	1975Q1	2014Q2

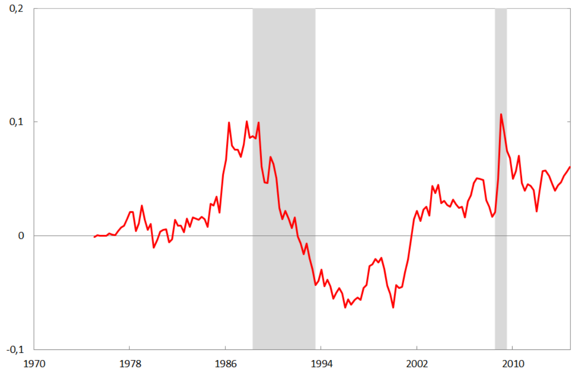
Private credit to GDP gap, household credit to GDP gap, non-financial enterprise credit to GDP gap, house price to income gap, wholesale funding ratio gap, equity ratio, private credit exuberance, house price to income exuberance, global credit to GDP gap, global house price to income gap

Periods of exuberance are characterized by extreme imbalances. The details of the calculations can be found in the online appendix of Anundsen et al. (2015). They have constructed country-specific exuberance measures for house prices and private credit based on econometric tests for a transition to a regime with explosive behavior.

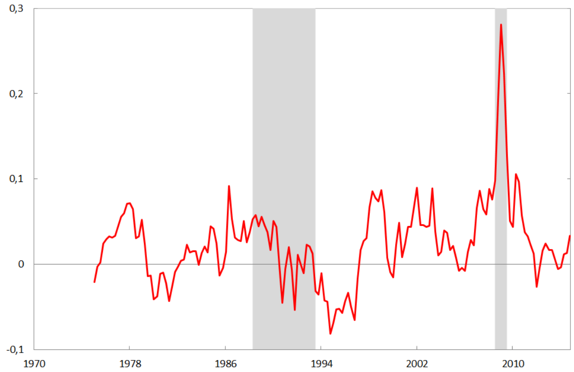
The measures of the global house price to income gap and private credit to GDP gap are included to capture possible contagion between countries through the financial system. The global variables are compiled using time-varying trade weights. The calculations are done by Anundsen et al. (2015), and further details can be found in their online appendix.



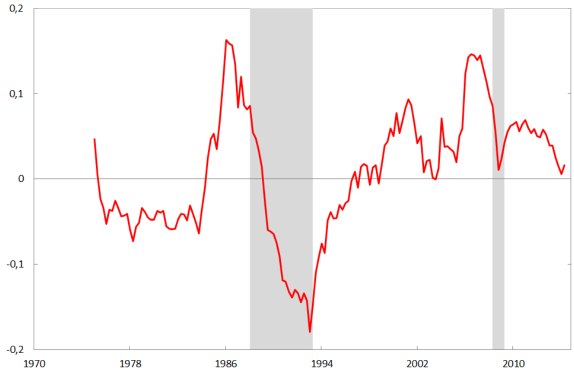
(a) Private credit to GDP gap



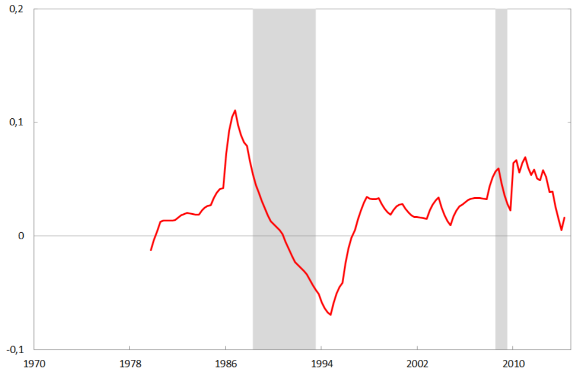
(b) Household credit to GDP gap



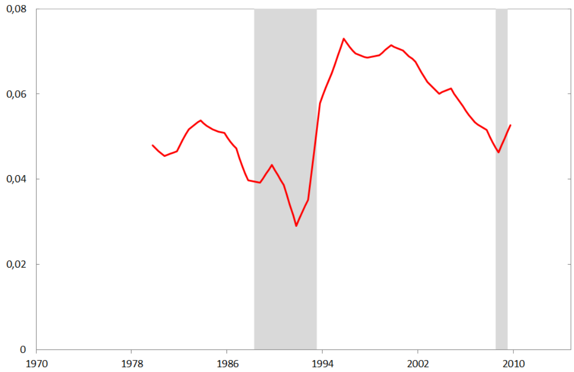
(c) Non-financial enterprise credit to GDP gap



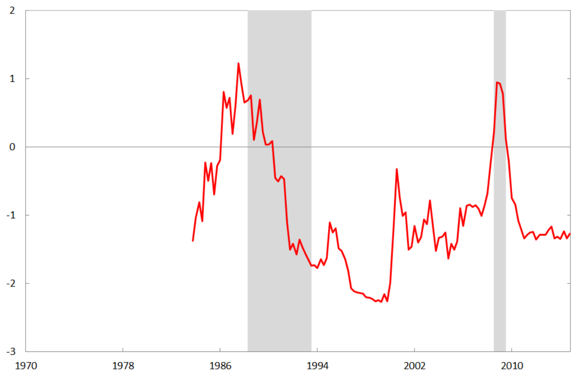
(d) House price to income gap



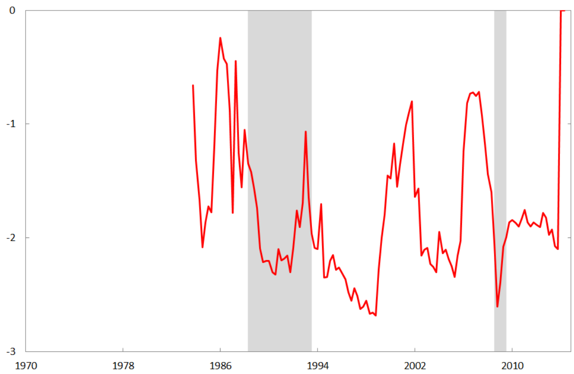
(e) Wholesale funding ratio gap



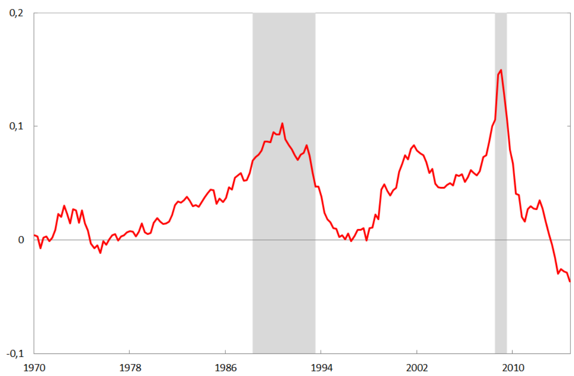
(f) Equity ratio



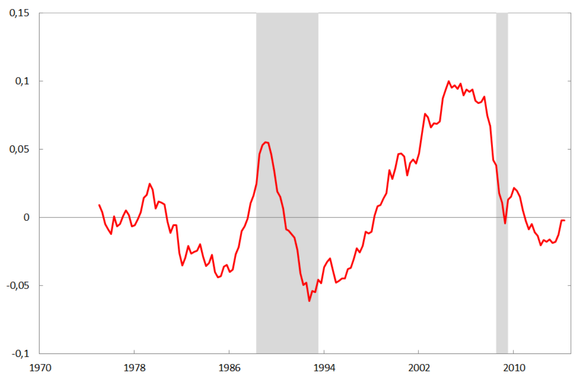
(g) Private credit exuberance



(h) House price exuberance

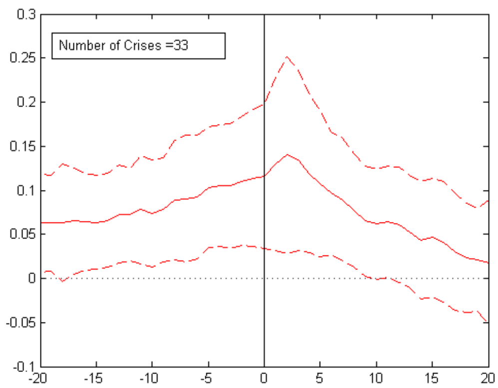


(i) Global credit to GDP gap

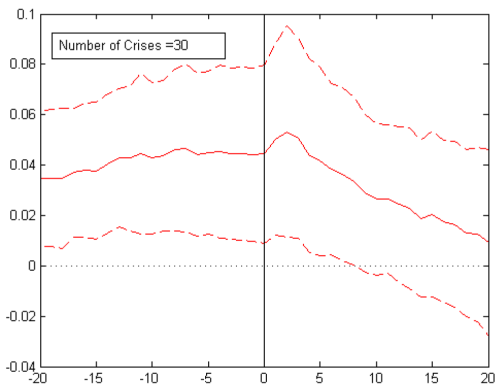


(j) Global house price to income gap

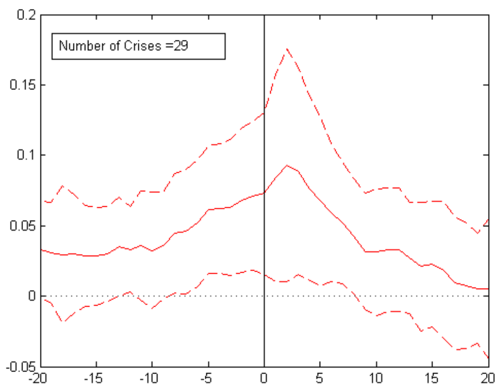
Figure 2: *Time series of all indicator variables for Norway. Gaps are given as deviations from trend calculated by a one sided HP filter with a rolling average forecast*



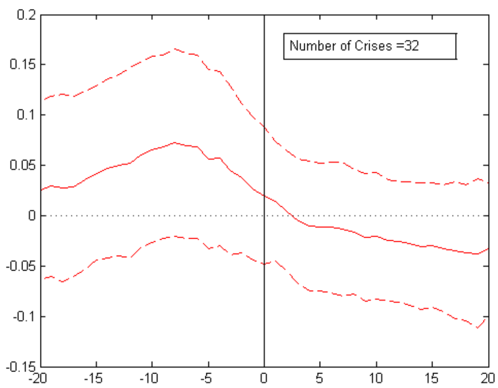
(a) Private credit to GDP gap



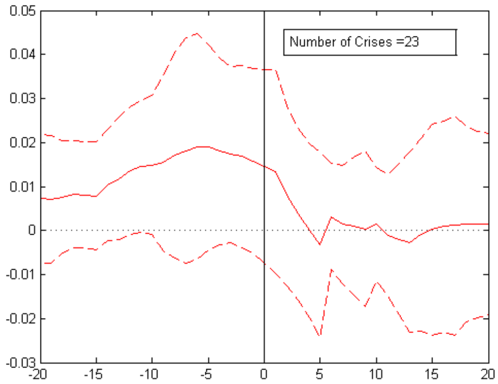
(b) Household credit to GDP gap



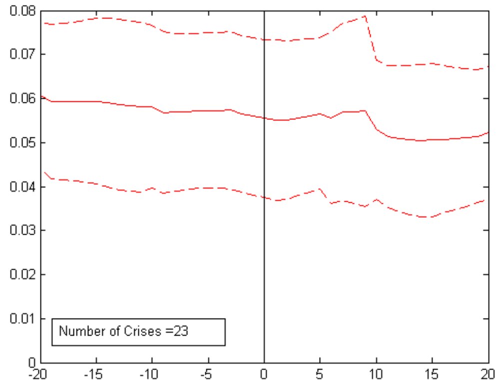
(c) Non-financial enterprise credit to GDP gap



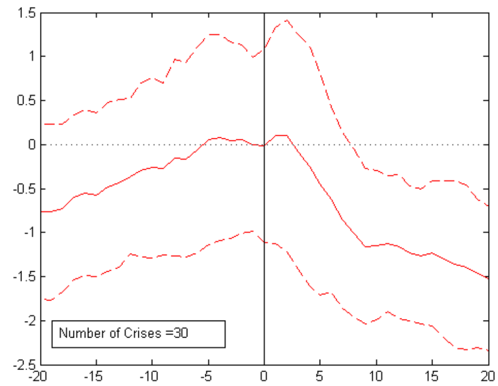
(d) House price to income gap



(e) Wholesale funding ratio gap

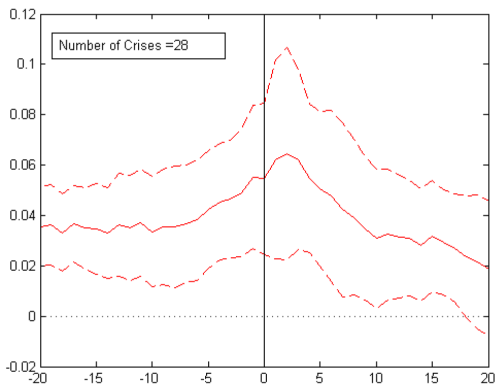


(f) Equity ratio

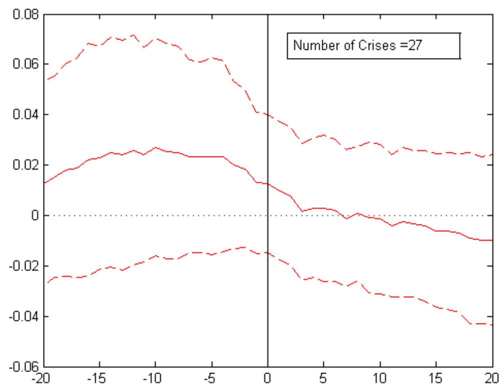


(g) Private credit exuberance

(h) House price exuberence



(i) Global credit to GDP gap



(j) Global house price to income gap

Figure 3: *Value of the indicator variables from 20 quarters prior to 20 quarters following the outbreak of a crisis. Solid lines are the mean of the sample. Dotted lines give one standard deviation*

3 Methodology

3.1 Data transformation: The HP filter

Most of the potential indicator variables in the sample are expressed by their deviation from a calculated trend, also referred to as a gap. This is to capture the cyclical component of the variable as the indicator. The method used for the calculations will be referred to as the Norges Bank method, and is a one-sided HP filter with a simple forecast, which is described in Gerdrup et al. (2013). The HP filter is named after the authors and presented in Hodrick and Prescott (1997). The HP filter is a method for calculating a trend from a time series, which will then make it possible to calculate the cyclical component as the deviation from trend. The filter is calculated by finding the trend series (μ_t) that minimizes the sum as given by (1):

$$\min_{\{\mu_t\}_{t=0}^{T}} (\sum_{t=0}^{T}(y_t - \mu_t)^2 + \lambda \sum_{t=1}^{T-1}((\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}))^2) \tag{1}$$

The parameter λ is also called the smoothing parameter. As λ increases, more weight will be put on the deviations in the trend from previous periods. This means that the higher λ is, the smoother will the trend be, as the sum is minimized by allowing for larger deviations between the trend and the observed variable.

The Basel Committee give guidance to national supervisory authorities about setting a so called buffer guide for the countercyclical capital buffer, and in on Banking Supervision (2010) they present the methodology to be used for this purpose. As with the signaling approach the buffer guide is based on the deviation of an indicator from its longterm trend. The trend is in this instance calculated using a one sided HP filter, where each trend observation is the end point of a two sided calculation. The smoothing parameter (λ) that is used is 400 000 to capture the long-term trend in the behavior of the credit to GDP ratio. In comparison, for business cycle analyses λ is often set to 1600 (Norges Bank, 2013). on Banking Supervision (2010) point to the fact that other methods could be used to calculate the trend, like a rolling average or linear trend, but that the HP filter has the advantage that it tends to place a higher weight on recent observations, thereby dealing more effectively with structural breaks.

To reduce the endpoint uncertainty Gerdrup et al. (2013) expand on this method by introducing a simple forecast to the time series when calculating the HP filter each period. Each period the time series is extended by H periods which is the forecast horizon. The minimization problem from (1) will now be formulated as:

$$\min_{\{\mu_t\}_{t=0}^{T+H}} \Big(\sum_{t=0}^{T+H} (y_t - \mu_t)^2 + \lambda \sum_{t=1}^{T-1+H} ((\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}))^2\Big) \tag{2}$$

The trend series that is calculated using the method of (2) will then compose of all T-period trend estimations, i.e. the original end-point had the forecast not been done. Gerdrup et al. (2013) present three different forecast schemes:

$$\text{Rolling average forecast:} \quad y_{t+h} = \frac{1}{4} \sum_{s=t-3}^{t} (y_s) \tag{3}$$

$$\text{Linear forecast:} \quad y_{t+h} = \alpha_{1:t} + \beta_{1:t} * (t + h) \tag{4}$$

$$\text{Rolling linear forecast:} \quad y_{t+h} = \alpha_{t-20:t} + \beta_{t-20:t} * (t + h) \tag{5}$$

These three forecast schemes are compared, along with a one sided HP filter without forecast, using the AUROC[4] from an early warning evaluation using the signaling approach. In the comparison they evaluate the timing of the indicators in predicting a crisis, the consistency of the signal, and how well the indicator signals a crisis, measured by the AUROC, in line with the three criteria of Drehmann and Juselius (2014). They find that the signaling quality of the indicator is best when the rolling average forecast is used. The analysis is done for the four key indicators of Norges Bank, on Norwegian data, with two crisis, admittedly a small samle to draw inference on. The method used in this thesis will be the Norges Bank method of using a one sided HP filter with a rolling average forecast to generate the gap. The forecast horizon is 20 quarters.

[4]The AUROC will be presented in greater detail in section 3.3.3

3.2 The signaling approach

As described in the introduction, the signaling approach to early warning seeks to transform an indicator or set of indicators into a binary signal that will signal an upcoming crisis prior to its outbreak. A good indicator will signal prior to most crises, while refraining from signaling when a crisis is not approaching. To be able to investigate whether an indicator is "good" or "bad" Kaminsky and Reinhart (1999) propose four judgments that must be made. Firstly, a well-defined notion of what classifies as a crisis is needed. Secondly, a list of variables that are potential leading indicators must be determined. Thirdly, a criteria that determines whether an indicator is signaling or not, and lastly a way to decide whether a signal is true or false.

The crises that will be used in this thesis and the potential leading indicators have already been presented. In the following the final two judgments will be set.

3.2.1 Signal horizon

As stated, the indicator, or set of indicators, is to signal prior to a crisis. The signal horizon is a predetermined time period prior to the crisis in which an indicator is expected to anticipate the crisis. If the indicator signals within the signal horizon it is called a true signal, while it is called a false signal if it signals outside of the horizon. In this thesis the signal horizon used will vary as part of the analysis, but unless otherwise specified the default will be the time period from 12 to 5 quarters prior to a crisis. There are three main reasons for dropping the last four quarters before the start of the crisis. The job of the indicator is predicting crisis in order to be able to implement measures to avoid them, and at the onset of a crisis this will be to late. Secondly, the behavior of the indicators may change in, or close to, a crisis, as indicated by figure 3. Lastly, the exact timing of the start of the crisis can be hard to determine (Davis and Karim (2008)), and with the previous point in mind this may skew the results.

3.2.2 The signal and categorization of periods

To make the two final judgment the signal is defined, along with a method of categorizing the periods, based on Drehmann and Juselius (2014). There they categorize the economy to be in three possible states each period, a normal state, a boom ("good times") (B) or a crisis ("bad times") which always follows the boom. Whether the economy is in a normal state (B=0) or a boom (B=1) is not directly observable in a given period, but an indicator (S), carrying imperfect information, is observed instead. A policy maker wants to evaluate this indicator to be able to say with some certainty whether or not the economy is in a pre-crisis boom state and measures must be implemented. The policy maker sets a threshold value (ϕ) for the indicator and defines the signal through a mapping from the continuous indicator to the binary signal, by the function $f : \mathbb{R} \rightarrow \{0, 1\}$:

$$f(s;\phi) = \begin{cases} 1 & \text{if } s \geq \phi \\ 0 & \text{if } s < \phi \end{cases} \tag{6}$$

The mapping states that whenever the indicator takes a value larger than or equal to the threshold value, this signals that the economy is in a boom and that a crisis is upcoming.

Table 3: Confusion matrix for categorizing indicator periods into true positives, false positives, false negatives or true negatives

	Boom	No boom
Signal	TP	FP
No signal	FN	TN

We are now in a position to categorize the observations of the indicator in each period. Every observation falls in one of the four categories given in the confusion matrix of table 3. In the matrix the boom periods are taken as the signal horizon. TP is the number of periods, in this case quarters, in which the indicator signals an upcoming crisis during a boom, $TP(\phi) = \sum_{i=1}^{n}(f(s_i;\phi) * b_i)$, meaning that the indicator gives a true signal, or a true positive. FP is the number of periods where the indicator signals a crisis outside of the signal horizon, $FP(\phi) = \sum_{i=1}^{n}(f(s_i;\phi) * (1 - b_i))$, which is a false signal, also known as a false positive or type II error. FN is the number of periods when a crisis is upcoming, but no signal is issued by the indicator, $FN(\phi) = \sum_{i=1}^{n}((1 - f(s_i; \phi)) * b_i)$. This is known as a false negative or type I error. Lastly, TN is the number of periods where no signals were issued and no crisis were upcoming, $TN(\phi) = \sum_{i=1}^{n}((1 - f(s_i; \phi)) * (1 - b_i))$, known as true negatives.

These quantities are the foundation of the signaling approach. The true positive rate (TPR) is the ratio of true positive periods to the total number of signal horizon periods. Using the confusion matrix of table 3 the ratio is defined as:

$$True\ positive\ rate = \frac{TP}{TP + FN} \tag{7}$$

The true positive rate is directly linked to the false negative rate (FNR). This is also known as the type I error rate and is the ratio of false negative periods to the total number of signal horizon periods:

$$False\ negative\ rate = \frac{FN}{TP + FN} = 1 - TPR \tag{8}$$

The false positive rate (FPR), or type II error rate, is the number of periods where a crisis is falsely signaled relative to the total number of periods outside of the signal horizons, given by equation (9). Just like the true positive and false negative rates sum to one, so does the false positive and true negative rates. The true negative rate is given by equation (10):

$$False\ positive\ rate = \frac{FP}{FP + TN} \tag{9}$$

$$True\ negative\ rate = \frac{TN}{FP + TN} \tag{10}$$

With the aforementioned rates, it is important to have an understanding of what we are looking for in a good indicator. The perfect indicator will have true positive and true negative rates of one, meaning type I and type II error rates of zero, but finding an indicator with these attributes will be almost impossible. With a low threshold value for the indicator it will signal prior to more of the actual crises, but at the same time it will issue more false signals. With a higher threshold the noise of the false positives will be reduced, but in doing so the probability of not signaling an upcoming crisis will increase. This trade-off is visualized in figure 4, showing the type I and type II error rates for different threshold values, using the private credit to GDP gap as an indicator. As can be seen from the graph, there are no threshold values that give both zero type I and type II error rates, and so there will be a trade-off between the two.

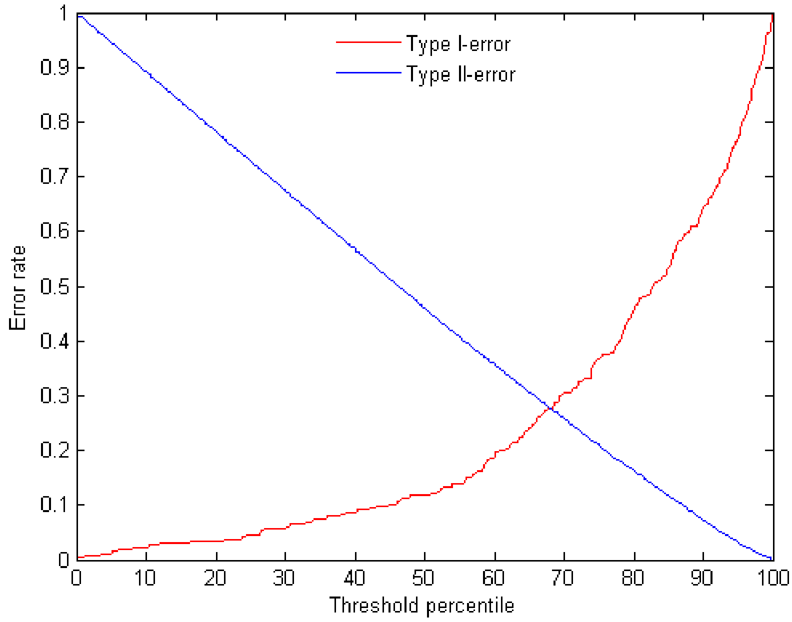


Figure 4: *Type I and type II error rates for different threshold values. Individual thresholds using the percentile method. Private credit to GDP gap. Signal horizon from 12 to 5 quarters prior to crisis*

3.2.3 Defining the thresholds

A more subtle choice that must be made when utilizing the signaling approach as an early warning system is how to set the threshold values. A threshold is the value above which an indicator is said to signal, but how we define this threshold can vary along two dimensions.

The first dimension is how to define the grid of threshold values between the lowest and highest. Two different methods are used for this purpose, the linear method and the percentile method. As the name implies, the linear method gives a linear grid of potential threshold values. By locating the lowest and highest value of the indicator, a grid is made with equal spacing between each threshold. A threshold can then be characterized by its percentage of the difference between the lowest and highest value. Alternatively, the percentile method gives a grid of thresholds comprising the corresponding percentiles of the indicator. E.g. with a grid of 11 points, the second entry gives the 10th percentile of the indicator, while the fifth gives the 40th percentile.

The second dimension is whether to calculate the grid of possible thresholds based on the whole sample, so called common threshold, or have individual grids for each country in the sample, individual thresholds. When applying the percentile method using individual thresholds, the same percentile is used for all countries, but the actual threshold value corresponding to that percentile will generally differ among all countries. The same goes for applying the linear method with individual thresholds, where the same percentage between the lowest and highest indicator value is used for all countries, but with correspondingly different threshold values .

3.2.4 Performing the calculations

A thorough walk-through of the calculations will be to extensive, but a brief summary of the basic concept will be given here. The observations of the indicator variable/variables, are given by a vector/vectors where each country's observations are stacked to give a vector of observations. With n observations per country and m countries, this will be an $(m * n)x1$ vector. Likewise, the crises will be represented through a binary vector, which takes the value 1 in all signal horizon periods and the value 0 in all other. Since this is an early warning system, the periods from the end of the signal horizon to the end of the crisis are not of interest to the evaluation. The observations for these periods are therefore removed from both the indicator and crises vectors. Introducing the signal criteria, each observation can now be categorized based on the confusion matrix of table 3. This is done iteratively for all the threshold values of the threshold grid, or tuples of threshold values given multiple indicators. With g being the number of grid-points for the thresholds and v being the number of indicators[5], the true positive, false positive, false negative and true negative rates corresponding to each combination of threshold values can now be stored in arrays of size g^v. These can then be used by the evaluation techniques presented next.

[5]Assuming that each indicator has only one threshold

3.3 Evaluation techniques

This section will present the main evaluation techniques to be used for the indicator evaluation. As stated in the introduction, the noise-to-signal ratio has been a workhorse of the signaling literature for a long time, and as such it will be presented first, in section 3.3.1. An alternative method to find optimal threshold values for the indcators is the minimization of a loss function. This will be presented in section 3.3.2. The method of Drehmann and Juselius (2014), mentioned in the introduction, of comparing the indicators based on their performance for all possible threshold values is presented in section 3.3.3 with the modified version of Detken et al. (2014) presented in section 3.3.4.

3.3.1 Noise to signal

The noise-to-signal ratio is defined as the false positive rate divided by the true positive rate. A lower noise-to-signal ratio can therefore be the result of less noise in the form of false signals, or of more correctly signaled crises.

$$Noise - to - signal = \frac{\frac{FP}{FP+TN}}{\frac{TP}{TP+FN}} \tag{11}$$

When evaluating indicators in their early warning models Kaminsky et al. (1997) and Kaminsky and Reinhart (1999), amongst other methods, compare the indicators by the proportion of crises detected. The thresholds are set by minimizing the noise-to-signal ratio. Although a low noise-to-signal ratio is a desired trait, the method of minimization doesn't take into consideration the preferences of the policy maker. At the same time, the method will generally lead to unjustifiably high threshold values, a point illustrated by figure 5, showing the noise-to-signal ratios for four of the indicators in the sample as the thresholds increase. In other words, the cost of few false signals is few detected crises. This fact is also brought up by Borio and Drehmann (2009):

> *"...minimizing the noise-to-signal ratio generally results in an unacceptably low percentage of crises predicted."*

In line with this, when evaluating the private credit to GDP as an indicator of banking crises, Kaminsky and Reinhart (1999) found it to be far from the best indicator in the sample and only signaled prior to 50 percent of the crises. This most likely stems from the fact that the threshold that minimizes the noise-to-signal ratio in their paper is in the 95th percentile.

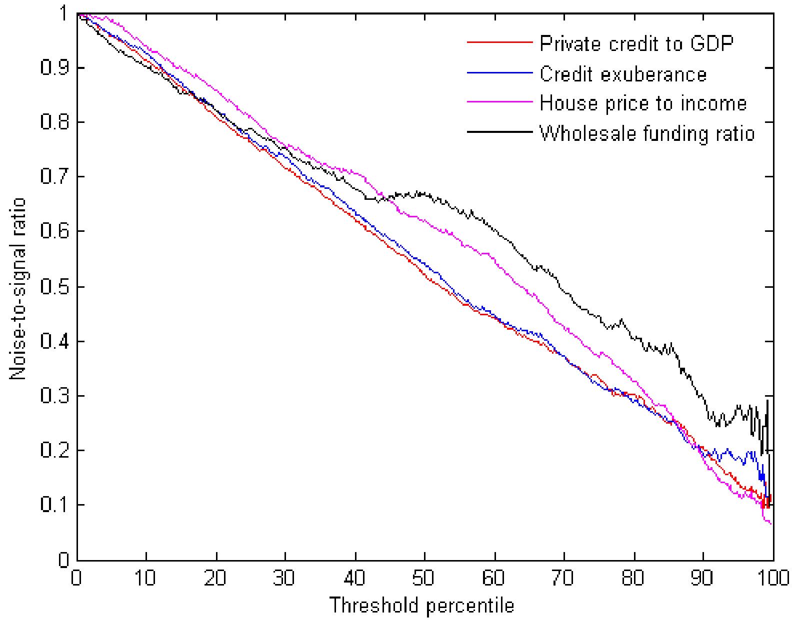


Figure 5: *Noise-to-signal ratios for different threshold values. Private credit to GDP gap, private credit exuberance, house price to income gap and wholesale funding ratio. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

To correct for this Borio and Drehmann (2009) propose implementing an ad hoc requirement that at least X percent of crises are detected. A point they neglect to discuss in this is that, although this in general will increase the number of crises detected it will do so up to the lower limit set in the condition. We again see this fact clearly from figure 5. If we for instance minimize the noise-to-signal ratio subject to at least 70 percent of crises being detected, the number of crisis that will be detected will generally be the closest possible to 70 percent from above. An exception might be if the data contains few observations and few crises, where one more crisis detected will reduce the ratio substantially. Nevertheless, this in practice means that the optimal threshold will be the one that gives the smallest amount of noise for an implicitly set true positive rate. It will still not take regard of the preferences of the policy maker. To sum up, although the noise-to-signal ratio has been the tool of choice for most applications of the signaling approach, it does not seem to perform well, at least not for the variables and data of this thesis. As such, it will not be used further in this thesis.

3.3.2 Loss function

An alternative to the noise-to-signal ratio as a way of establishing optimal thresholds is a loss function. The loss function suggested by Borio and Drehmann (2009) can be expressed as follows:

$$L = \theta * TypeI + (1-\theta) * TypeII, \qquad \theta \in [0,1], \tag{12}$$

where $TypeI$ is the type I error rate/false negative rate, $TypeII$ is the type II error rate/false positive rate, while θ is the preference parameter for the policy maker between failing to signal a crisis and falsely signaling one. The higher θ is, the more costly does the policy maker view missing crises relative to falsely signaling them. Since each threshold value corresponds to a specific pair of type I and type II error rates, the loss function is minimized with respect to the threshold value that generates the smallest possible loss.

Minimization of a loss function has previously been used by Demirgüç-kunt and Detragiache (1999) and Bussière and Fratzscher (2008)[6], but in both these cases they use a multivariate logit approach instead of the signaling approach. Figure 6 illustrates the relationship between the type I and type II error rates and the value of the loss function for different threshold values. The analysis is done for the private credit to GDP gap, using individual threshold values and the percentile method. Note that the loss function never takes a value larger than 0.5 in this case, which is the value of θ. This comes from the convex shapes of both error rates as functions of the threshold value. If they instead had been concave in the threshold value, a linear combination of the two would produce loss values that were larger than or equal to θ. If that were the case the policy maker would always be able to limit the loss to the smallest of θ and $(1-\theta)$, by setting the threshold value to the lowest possible value and acting as though the indicator is always

[6] Instead of type I and type II error ratios, they weigh the probability of missing a crisis and that of issuing a signal

signaling whenever $\theta < 0.5$ and the opposite when $\theta > 0.5$. This implies that:

$$max(L) = min(\theta, 1 - \theta)$$

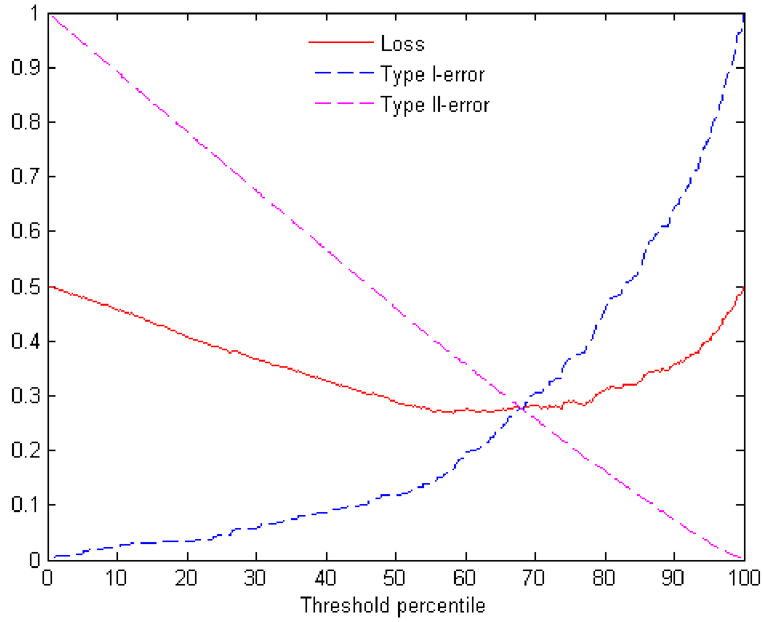


Figure 6: *Loss values and type I and type II error rates for different threshold values. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis.* $\theta = 0.5$

3.3.3 Area Under the Reciever Operating Characteristic curve (AUROC)

Finding the best performing indicator by the minimization of a loss function relies heavily on the preference parameter of the policy maker, and so the findings are highly sensitive to the choice of θ. It may be hard, if not impossible, to determine its true value and a more general approach may therefore be preferable for indicator evaluation. A possibility is to use the area under the receiver operating characteristic (ROC) curve, which is based on true and false positive rates given by equations (7) and (9) respectively. ROC analysis has its origin from the analysis of radar signal detection (van Erkel and Pattynama, 1998), where the name "receiver operating characteristic" stems from, but it also has a long history in machine learning and medical science (Fawcett, 2006). Corresponding to each threshold value is a pair of true and false positive rates. The ROC curve expresses the true positive rate as a function of the false positive rate. The ROC curve for the private credit to GDP gap can be seen in figure 7, with individual thresholds by the percentile method, and a signal horizon from 12 to 5 quarters prior to a crisis.

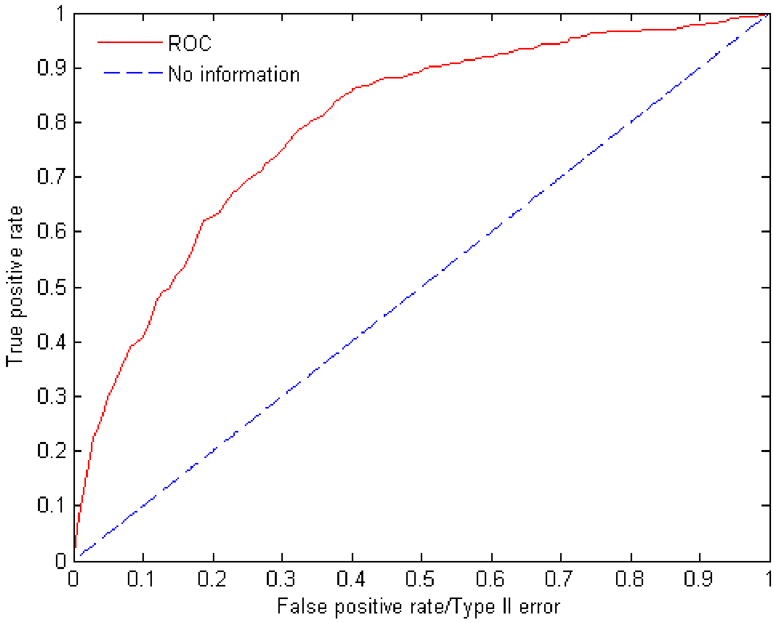


Figure 7: *ROC curve expressing the true positive rate as a function of the false positive rate. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

In the lower left corner of the graph, at the origin, both rates take the value zero. This is the case for all threshold values above the maximum value of the indicator variable, when no signals are issued. On the other end of the spectrum is the upper right corner where the threshold value is lower than the minimum of the indicator variable, and a signal is issued in every period. Along the 45 degree line connecting (0, 0) to (1, 1) the rates are equal and the indicator will signal randomly, meaning that there is no information in the indicator to help signal a crisis. The point (0, 1) is said to be perfect since this it has a true positive rate of one and a false positive rate of zero, thereby zero type I and type II error rates. Any point above and to the left of the diagonal indicate a signaling performance better than random. This also entails that any point below or to the right of the diagonal is worse than random, but by reversing the classification decisions, i.e. true positives become false negatives and false positives become true negatives, the same indicator will now perform better than random (Fawcett, 2006)

For any given loss the loss function can be rewritten as an indifference curve expressing the true positive rate as a function of the false positive rate:

$$L = \theta * TypeI + (1-\theta) * TypeII$$
$$L = \theta * (1-TPR) + (1-\theta) * FPR$$
$$TPR = \frac{\theta - L}{\theta} + \frac{1-\theta}{\theta} * FPR \tag{13}$$

From equation (13) it is clear that as θ increases, the slope of the indifference curve will be less steep, which generally moves the tangent point with the ROC curve to the right along the curve. The result is lower threshold values and more crises being signaled, both true and false. An example of the indifference curves is provided by figure 8.

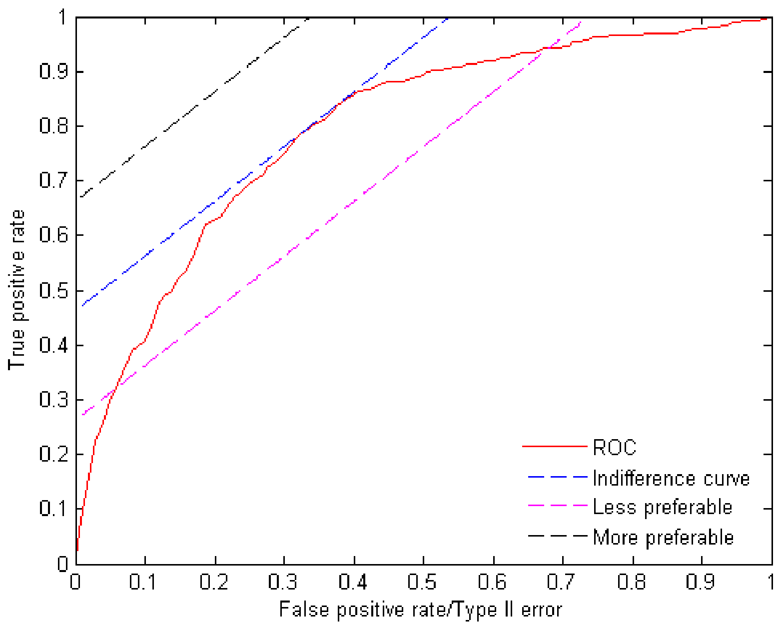


Figure 8: *ROC curve and indifference curves for the policy maker. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis.* $\theta = 0.5$

By reference to the previous point, evaluating an indicator by using the ROC curve in isolation, although it gives a graphic representation of the trade-off between true and false positive rates, doesn't give any more information than the minimization of the loss function with varying values for θ. On the other hand, based on the ROC curve it is possible to calculate the area under the ROC curve (AUROC). This area will take values between zero and one, where one represents a perfect indicator. An indicator that traces the diagonal line will signal randomly and have an AUROC of 0.5. To be able to establish threshold values for the indicators and compare them fully, one still needs a grasp of the policy makers preferences. Yet, by comparing indicators by the use of their AUROC it is possible to, at least generally, establish which indicators have the best signaling performance for a broad specter of threshold values. The higher the AUROC, the higher will the true positive rate generally be relative to the false positive rate, i.e. the more precise will the indicator be when signaling a crisis.

It is also possible to calculate standard errors the AUROC, and the method presented here will be based on Hanley and McNeil (1982). There the method is used for calculating the standard error of the AUROC related to analyses in radiology. The standard error of the AUROC is given by the formula:

$$SE(A) = \sqrt{\frac{A(1-A) + (n_a - 1)(Q_1 - A^2) + (n_n - 1)(Q_1 - A^2)}{n_a n_n}} \tag{14}$$

Here, A is the calculated AUROC, n_a is the number of signal horizon periods and n_n is the number of non-signal horizon periods. Q_1 and Q_2 are of a more complex nature. In this case, Q_1 equals the probability that the indicator in two randomly chosen signal horizon periods will have higher values than the indicator in a random non-signal horizon period. Q_2 equals the probability that the indicator in a randomly chosen signal horizon period will have a value higher than the indicator in two randomly chosen non-signal horison periods. The two probabilities can be found using the following formulas:

$$Q_1 = \frac{A}{2-A}$$
$$Q_2 = \frac{2A^2}{1+A}$$

Hanley and McNeil (1983) present a method for calculating the standard error for the difference between two AUROCs based on the same data sample. The method relies on the individual standard errors presented earlier, and is given by the formula:

$$SE(A_1 - A_2) = \sqrt{SE(A_1)^2 + SE(A_2)^2 - 2rSE(A_1)SE(A_2)} \tag{15}$$

The parameter r represents the correlation introduced by studying the AUROC for the same sample. Detken et al. (2014) choose to set this to zero to keep the analysis as conservative as possible.

3.3.4 Partial standardized AUROC (psAUROC)

Detken et al. (2014) present a modification of the AUROC, called the standardized partial AUROC (psAUROC). Instead of taking into account all possible pairs of false and true positive rates, they only consider those that can be seen as relevant for evaluation. The calculation of the psAUROC can be said to be divided into three steps. As already described, a higher preference parameter of the policy maker (θ) will, in general, lead to a lower optimal threshold value given minimization of a loss function, which again leads to more false signals. If it is now assumed that θ will have a minimum value in the eyes of the policy maker, this means that there is a lowest possible false positive rate corresponding to the optimal threshold value given for the minimum value of θ. This can be called the minimum false positive rate. The implication of this is that the only relevant part of the ROC curve when evaluating indicators is the part to the right of the minimum false positive rate.

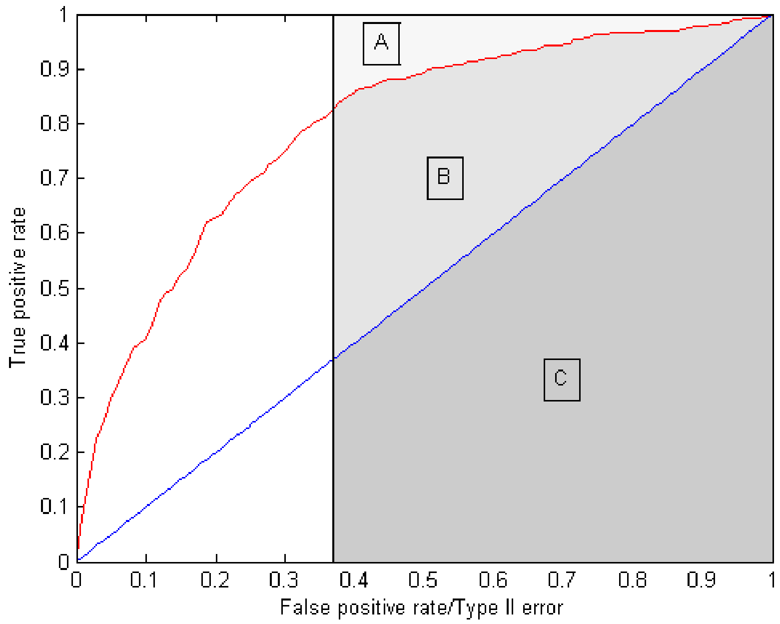


Figure 9: *Decomposition of the calculation of the psAUROC. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

Figure 9 shows the relevant parts of the graph for the calculation of the psAUROC. The first step is to find the minimum false positive rate. This thesis will follow Detken et al. (2014) in assuming that the lowest possible preference of the policy maker is $\theta = 0.5$. The area to the right of the minimum false positive rate comprises the areas A, B and C in figure 9, and will in the calculation be referred to as max, ($max = A + B + C$). The second step is to calculate the partial AUROC, which is the area under the ROC curve to the right of the minimum false positive rate, $pAUROC = B + C$. Lastly, the partial AUROC must be standardized so that a perfect indicator takes the value 1 and an uninformative indicator takes the value 0.5. First, area min is defined as the area under the diagonal curve to the right of the vertical line, $min = C$. The calculation of the partial standardized AUROC is given by the formula:

$$psAUROC = \frac{1}{2}\left[1 + \frac{pAUROC - min}{max - min}\right] \tag{16}$$

As can be seen from (16), if there is no information in the indicator, i.e. $pAUROC =$

min then the $psAUROC = 0.5$, while $psAUROC = 1$ with a perfect indicator where $pAUROC = max$.

3.4 Number of variables

When performing the analyses one or more indicators can be used to signal a crisis. Some of the signaling schemes that can be used will be presented next.

3.4.1 Univariate analysis

In the univariate analysis we only look at one variable in isolation as an indicator. This gives an easy to interpret signal, where the indicator signals whenever it takes a value above the threshold value. This is the most commonly used approach to signal evaluation, and is used in every paper on the subject.

3.4.2 Bivariate analysis

When doing a bivariate analysis two different variables are used as indicators. In this case there are different approaches available to generate a signal. In the easiest one, from now on referred to as the standard method, a threshold value is prescribed for each variable, and the indicators signal whenever both indicators breach their respective threshold. This is the commonly used method (see for instance Borio and Lowe (2002), Borio and Lowe (2004), Borio and Drehmann (2009) and Alessi and Detken (2011))

The third approach is an innovation in the literature. In this case one of the variables is recognized as a main indicator, while the other is a support indicator. The indicators will signal an upcoming crisis if the main indicator breaches its main threshold, or if it breaches a secondary threshold and the support indicator breaches its threshold value. This can be illustrated using the private credit to GDP gap as the main indicator and house price to income gap as a support. If there is a large private credit to GDP gap at the same time as a large house price to income gap, the indicators signal a crisis. But a signal will also be issued if the private credit to GDP gap breaches a higher threshold alone. The intuition is that although large deviations from trend for private credit to GDP alone gives a good indication of a crisis, the signal is even "stronger" when coupled with large deviations for house price to income from trend, requiring a lower threshold value for the private credit to GDP gap. By setting the main threshold of the main indicator to its maximum, the signal will work as under the standard method. Alternatively, by setting the secondary threshold of the main indicator to its minimum, the criteria will be as for a and/or criteria.

4 Sensitivities

In this section I will look more closely at how sensitive the analytical framework is to changes in the underlying assumptions or choices of methodology. I will first look at the preferences of the policy maker, before I move on to the definition of the threshold grid. Lastly, I will look at the signal horizon.

4.1 The policy makers preferences

To be able to use the signaling approach as an early warning system for systemic banking crises we need to be able to say when an indicator is signaling. To be able to do this a threshold value for the indicator must be defined, above which the signal is issued. By the loss function presented earlier, the optimal threshold value is defined as the one that minimizes the function. Figur 10 presents the type I and type II error rates for different threshold values when the indicator variable is the private credit to GDP gap.

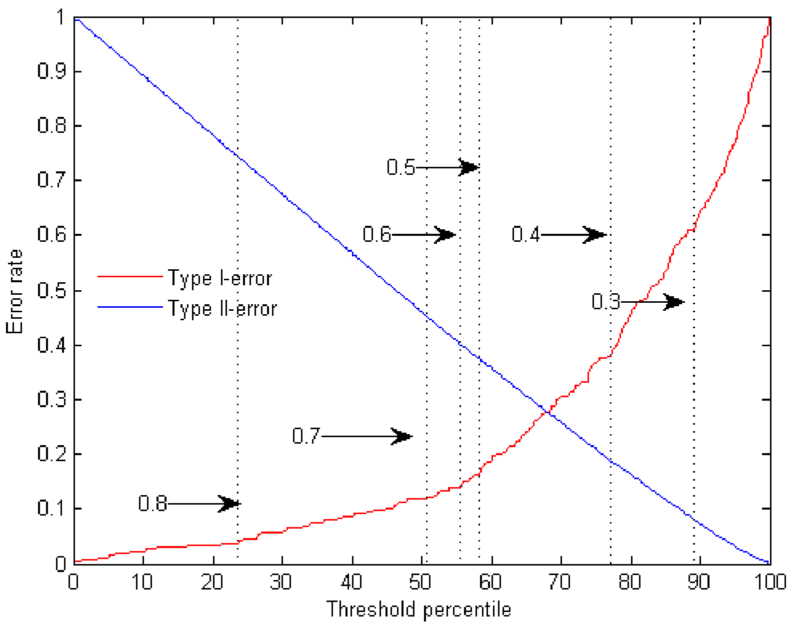


Figure 10: *Optimal threshold values expressed as verticle lines, along with the type I and type II error rates for the corresponding threshold values. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

The dotted vertical lines represent the optimal threshold values, expressed as the percentile value for each country's indicator value, when the loss function is minimized using different θ values, i.e. for different preferences of the policy maker. It is clear from this figure that without having an opinion about the preferences between type I and type II errors, setting the threshold value will be impossible. For the case of the private credit to GDP gap, the threshold values corresponding to $\theta \in [0.5, 0.7]$ are pretty close, with that of $\theta = 0.5$ being 58.2, $\theta = 0.5$ being 55.6, and that of $\theta = 0.7$ being 50.8.

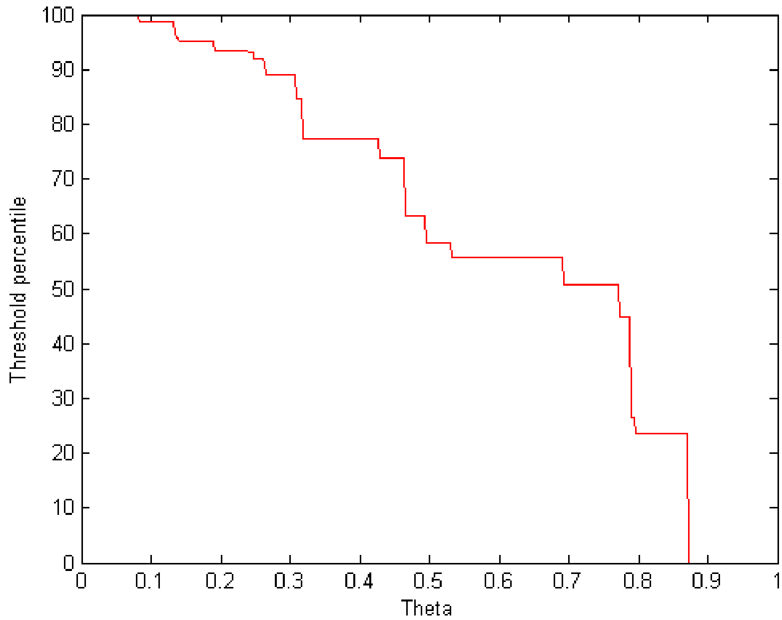


Figure 11: *Optimal threshold values expressed as percentiles for different θ values. Private credit to GDP gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

Figure 11 shows in greater detail how the optimal threshold values change with changes in the policy makers preferences for the case of the private credit to GDP gap. The potentially very high costs of a systemic banking crisis makes it probable that the policy maker is more inclined to allow for type II errors than type I errors, i.e. that $\theta > 0.5$. The question is then how averse the policy maker is to missing an upcoming crisis. With θ values from 0.5 to almost 0.8 the optimal threshold is quite stable, but at 0.8 it has dropped to 23.6.

4.2 How to set the thresholds

This section shows how the choice of methods for setting the threshold values may affect the results. The different methods were presented in section 3.2.3, and can be separated into linear or percentile grids and common or individual method. Figure 12 shows the ROC curves for the cases percentile and common method, percentile and individual method, and lastly linear and individual method, all for the private credit to GDP gap. As stated in the presentation of the ROC curve, the more we are up and to the left in the graph, the better. From the graph it is hard to determine the best performing indicator. As seen in the previous section, knowledge about the preferences of the policy maker are needed to be able to make a definitive decision about which of the methods for setting the threshold values is prefered. Among the three cases, the linear individual seems to perform the worst, with the lowest AUROC and the most unstable shape.

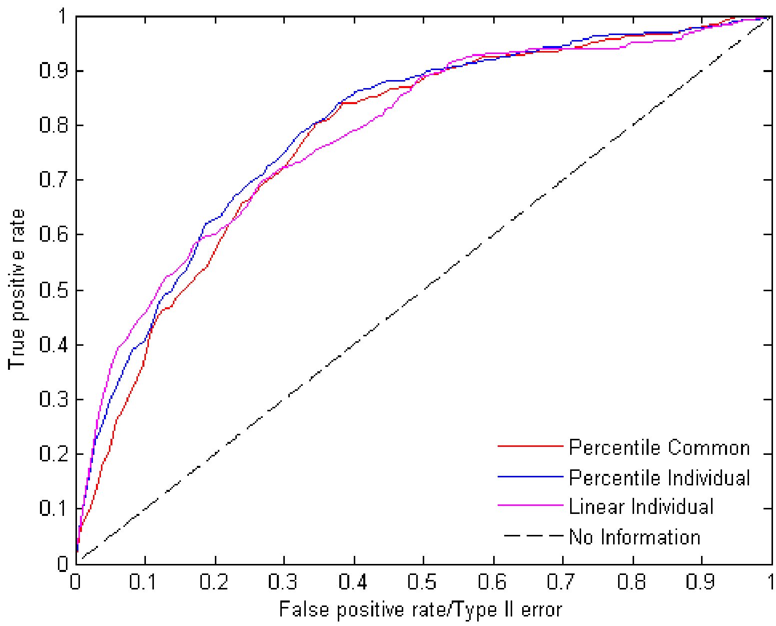
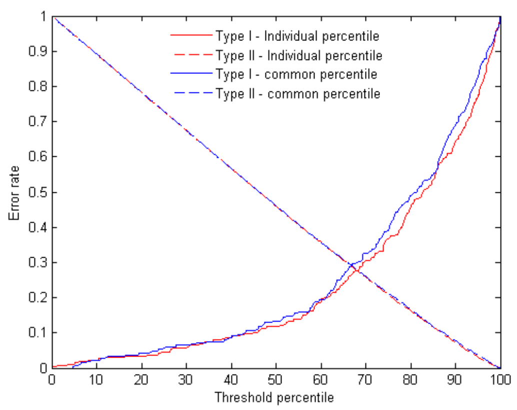


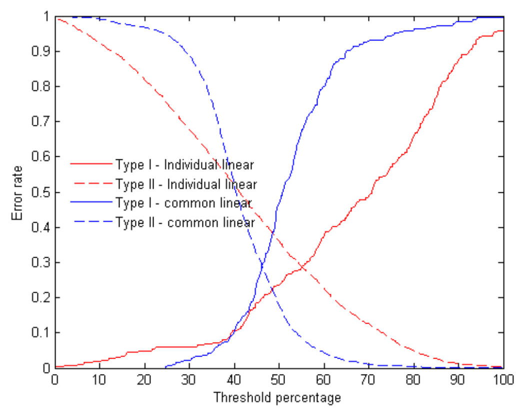
Figure 12: *ROC curves for different methods of setting the threshold values. Private credit to GDP. Signal horizon from 12 to 5 quarters prior to crisis*

In figure 13 the type I and type II error rates are given for all thresholds given for all four possible methods. This is done for the three key indicators of Norges Bank in the data set, with a signal horizon from 12 to 4 quarters prior to a crisis. Figure 13a, 13c and 13e shows the error rates for the case of the percentile method for both common and individual thresholds. The curves for the type II error rates can be considered to be identical for all three cases. For the type I error, although not clearcut, the individual thresholds seems to systematically have slightly lower error rate than the common method. Figure 13b, 13d and 13f shows the corresponding error rates for the linear method. Even though the difference in the minimized loss not necessarily is to large between the common and individual in these cases, the interval of acceptable thresholds seems to be too narrow for practical use in the case of a common threshold. Based on the figures 12 and 13 the main method applied for the rest of the thesis will be the percentile method with individual thresholds.

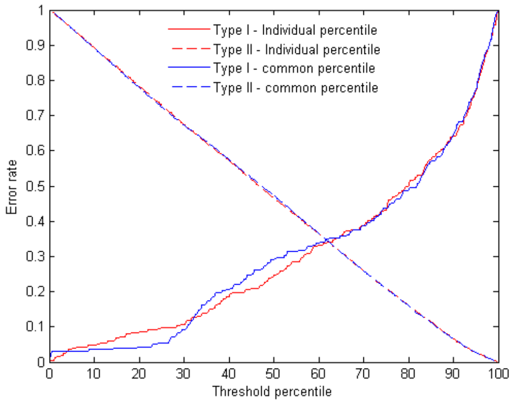
The choice is not crystal clear though. Borio and Lowe (2002) suggest that it is the absolute value of an indicator that is important, and that the top percentile of a variable will be a poor indicator for a country with moderate size of its observations. When comparing common and individual thresholds with the percentile method, Davis and Karim (2008) find the results to be ambiguous in the sense that the common method leads to higher type I error rates and the individual to higher type II error rates. In their analyses though, they use the minimization of the noise-to-signal ratio to find optimal thresholds.



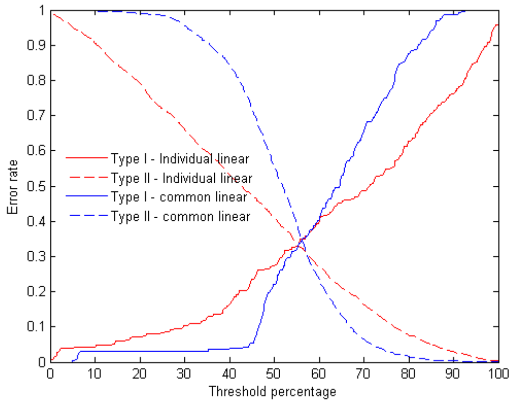
(a) Private credit to GDP gap. Percentile



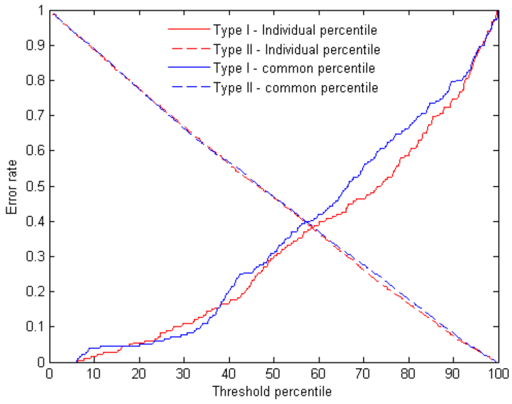
(b) Private credit to GDP gap. Linear



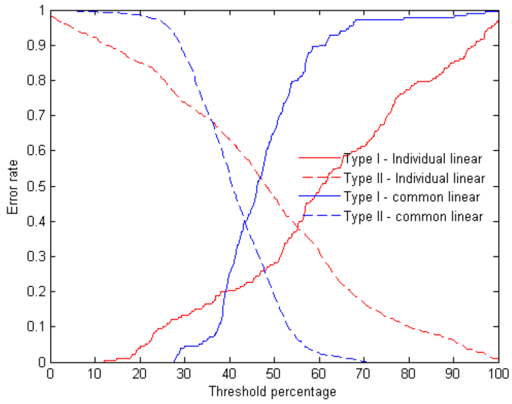
(c) House price to income gap. Percentile



(d) House price to income gap. Linear



(e) Wholesale funding ratio gap. Percentile



(f) Wholesale funding ratio gap. Linear

Figure 13: *Type I and type II error rates for different thresholds given by percentile and linear grids for the individual and common methods. Private credit to GDP gap. Signal horizon from 12 to 5 quarters prior to crisis*

4.3 Horizons

The signaling approach is an early warning tool used to signal a crisis before it hits, and how long before is given by the signal horizon. There seems to be a trade-off in that regard. On the one hand, the earlier an upcoming crisis is signaled, the more time is available to implement counter measures to try to negate it. On the other, when there is too much time between the signal and the outbreak of a crisis it may be hard to uphold the trust in the policy maker with regards to forecasting, thereby reducing the legitimacy of any policy action.

Figure 14 shows the AUROC for the private credit to GDP gap, using a signal horizon of only a single period. The solid line is the calculated AUROCs while the dotted lines give one standard deviation. The thresholds are individual and given by the percentile method. When performing the analysis for each period in this case, the periods between the single signal horizon period and the outbreak of the crisis are excluded. Figure 14 illustrates that the AUROC for the indicator generally increases as the signal approaches the crisis. The private credit to GDP gap has a fairly good predictive power also further away from the crisis. Although the AUROC is a little unstable, 12 quarters prior to the outbreak it is still almost 80.

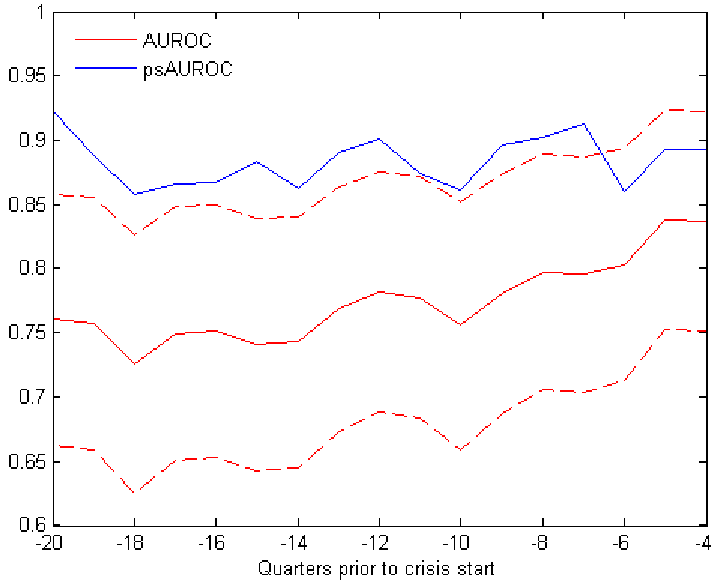


Figure 14: *AUROC and psAUROC values for different single signal horizon periods. Dotted lines give one standard error from the calculated AUROC. Private credit to GDP gap. Individual thresholds using the percentile method*

Although the AUROC falls as the signal horizon moves away from the crisis it is not clear that the signaling ability of the indicator worsens for these periods. This again depends on the preferences of the policy maker. Figure 15 gives the ROC curves for the analysis one, two, three and five years prior to the outbreak of a crisis. From figure 14 one can see that the AUROC two years prior is higher than that of three and five years prior to the crisis. It is clear from figure 15 that this comes from the better signaling ability at higher threshold values, i.e to the left of the graph. If the policy maker has a fairly high θ this implies a lower optimal threshold value and a higher false positive rate relative to the true positive rate. This again means that we are more to the right in the graph, and in this range the signaling ability will be better three and five years prior to the outbreak than 2 years. Indeed, figure 14 also show the psAUROC calculated with a minimum preference of the policy maker of $\theta = 0.5$. This curve shows no systematic tendency of being lower for longer signal horizons. One would expect that as the crisis approached, the indicators would signal more precisely. Although it will not elaborated on, a possible explanation for the lack of this may lie in that construction of the private credit to GDP gap, which is the indicator in question. If a crisis follows a prolonged period of instability, evident by a consistent increase in private credit to GDP this may be incorporated in the trend calculated by the Norges Bank method, thereby expressing the gap as smaller than it in reality is as one gets closer to the crisis.

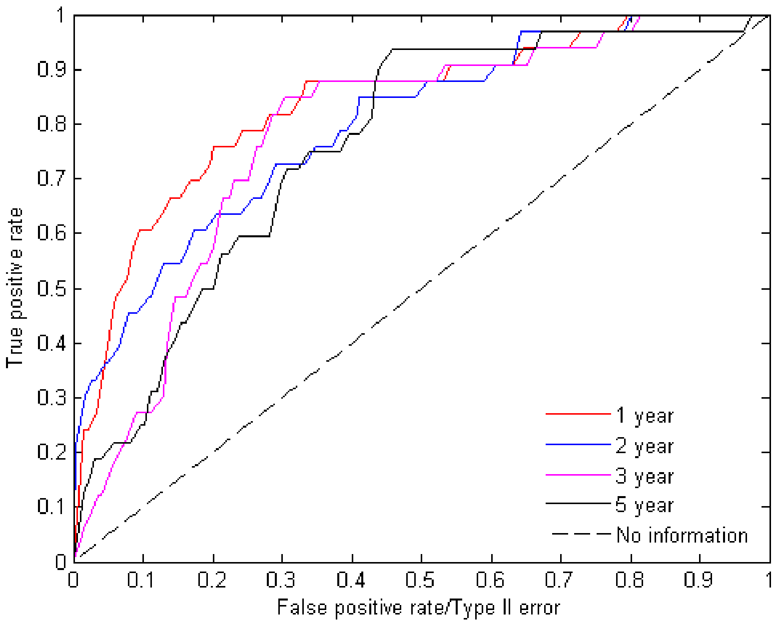
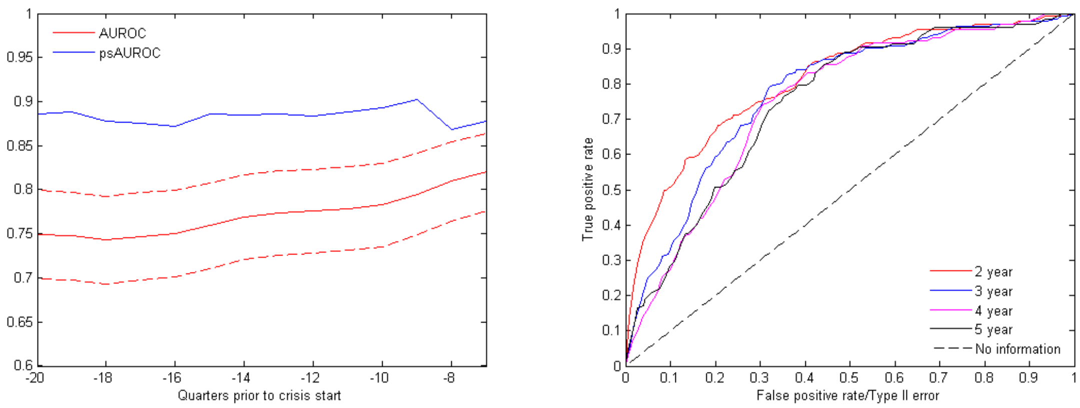


Figure 15: *ROC curves for different single signal horizon periods, given as years prior to crisis. Private credit to GDP gap. Individual thresholds using the percentile method*

It is worth mentioning again that when performing the analyses of this section the signal horizons consisted of only a single period prior to each crisis. This gives relatively few observations for signal horizon periods relative to the non signal horizon periods, with a total of 33 signal horizon periods for the private credit to GDP gap. As can be seen from figure 15 this creates fairly stepwise ROC curves. This also goes for the changes in the AUROC between periods in figure 14. If each signal horizon instead consist of four periods this is likely to change, as there are four times as many periods for an indicator to signal. The corresponding results for these signal horizons are illustrated in figures 16a and 16b. The x-axis of figure 16a gives the quarter prior to the crisis outbreak in which the signal horizon starts. As such, the far right observation is the AUROC calculated using the signal horizon from 7 to and including 4 quarters prior to the crisis. The fact that this curve is smoother should come as no surprise, as each adjacent observation share three out of four signal horizon periods. Likewise for figure 16b, the year stated is of the start of the signal horizon, e.g. 2 years is from 8 quarters to and including 5 quarters prior to the crisis. The ROC curves are, as expected smoother in this case. An interesting feature is that although the single signal horizon five years prior to the crisis outbreak seemed to dominate for some parts, this dominance is gone with the extended signal horizon. For the area of interest the different four period horizons seem to perform about the same, a fact that is substantiated by the psAUROCs of figure 16a.



(a) AUROC and psAUROC (b) ROC curves

Figure 16: *AUROC, psAUROC and ROC curves for different four quarter signal horizons. Dotted lines give one standard error from the calculated AUROC in 16a. Private credit to GDP. Individual thresholds using the percentile method*

5 Results

In this section I will evaluate the indicators to find the best performing indicator and pair of indicators, along with their optimal threshold values. The indicator evaluation will be based on the AUROC of each indicator or indicator pair, while the optimal thresholds are found by minimization of the loss function given by (12).

5.1 Univariate analyses

The results of the univariate analyses are presented in table 4 for all the potential leading indicators. The analyses are done using individual thresholds, defined by the percentile method, and with a signal horizon from 12 to 5 quarters prior to the a crises.

Table 4: Univariate analyses. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Indicator	AUROC	se(A)	psA	0.5	0.6	0.7	0.8	Crises
Private credit/GDP	0.7939	0.0169	0.8979	58.2	55.6	50.8	23.6	33
Credit exuberance	0.7769	0.0182	0.8822	61.6	57.0	54.6	25.8	30
Household credit/GDP	0.7628	0.0185	0.8396	68.0	50.4	50.4	0.6	30
NFE credit/GDP	0.7329	0.0194	0.8021	70.0	44.0	41.4	1.2	29
House price/Income	0.7269	0.0188	0.7696	73.4	46.6	30.2	0.0	32
Wholesale/Assets	0.7005	0.0227	0.8631	49.2	32.6	23.0	23.0	23
House price exub.	0.6851	0.0211	0.7478	59.8	59.8	16.0	0.0	27
Global HP/Income	0.6723	0.0211	0.6942	85.4	29.8	5.2	5.2	27
Global credit/GDP	0.6376	0.0207	0.6968	63.2	57.2	0.0	0.0	28
Equity ratio	0.5073	0.0232	0.5047	93.0	0.0	0.0	0.0	23

Columns (1) and (2) gives the AUROC and its calculated standard error, while column (3) gives the psAUROC. (4)-(7) give the optimal thresholds for $\theta \in \{0.5, 0.6, 0.7, 0.80\}$. Lastly, the number of crises covered by each indicator evaluation are given by column (8).

There are several points to notice. Firstly, note that of the 10 indicators, the private credit to GDP gap fares best in terms of both the AUROC and the psAUROC. It can also be noted that when ranking the performance of the indicators by their AUROC the two decompositions of private credit to GDP fare worse individually. With the exception of global credit to GDP, the credit variables clearly has the best signaling performance in terms of the AUROC. If the indicators instead are ranked by the psAUROC the ranking stays about the same, but the wholesale funding ratio gap moves up to third place, with a psAUROC not far below the private credit to GDP gap. The psAUROC is calculated based on the assumption that the minimum preference parameter of the policy maker, θ, is 0.5. Table 4 shows that the optimal threshold value for the wholesale funding ratio gap, given minimization of a loss function with $\theta = 0.5$, is by far the lowest. This will likely lead to a correspondingly higher false positive rate. The result of this can be seen in figure 17, which shows the ROC curves for the private credit to GDP gap and the wholesale funding ratio gap, together with the areas used to calculate their psAUROCs.

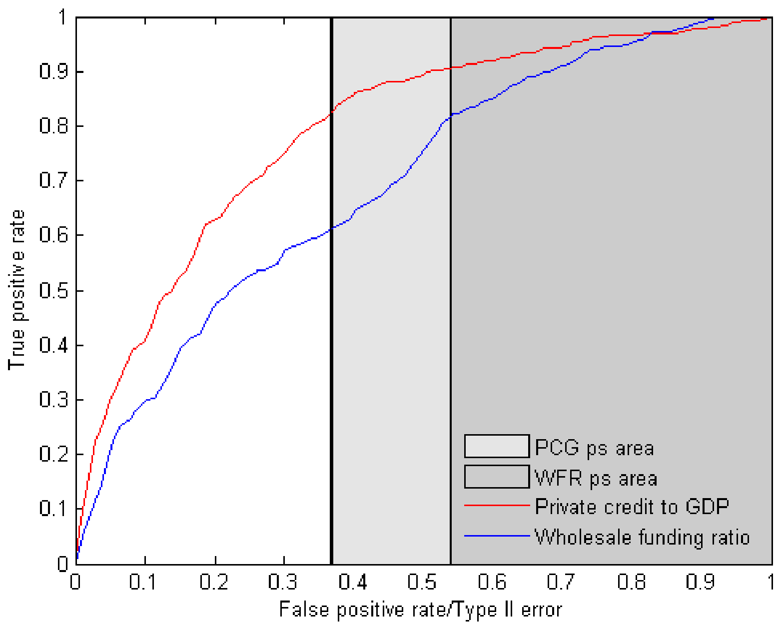


Figure 17: *The ROC curves and areas used for the calculation of the psAUROC. Private credit to GDP gap and wholesale funding ratio gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

It is clear that for almost the entire range of possible preferences the private credit to GDP gap dominates the wholesale funding ratio gap. Specifically, for $\theta = 0.5$, the true positive rate of the former is higher than that of the latter, and with a much lower false positive rate. The seemingly high performance of the wholesale funding ratio gap, when measured by the psAUROC, stems from the smaller area used for the standardization. This is weakness of the measure that must be considered when applying it to indicator evaluation.

5.2 Bivariate analyses

First in this section I will evaluate indicators by using the standard signaling scheme, requiring both indicators to breach their respective thresholds. After that I will present the results using the alternative scheme introduced in this thesis, with one main indicator and one support indicator. The number of possible indicator pairs is to large to present as a whole, so a selection will be presented[7].

Based on the potential indicator variables all possible combinations of bivariate evaluations are run using the standard scheme. The thresholds are individual, found by the percentile method, and the signal horizon is from 12 to 5 quarters prior to a crisis. Table 5 holds the results for some of the evaluations done, including the best performing. In column (1) and (2), the AUROCs for the univariate analyses of the two indicators are reported, while (3) holds that of the bivariate. Column (4) holds the number of crises covered by each indicator pair and lastly column (5) reports the correlation of the two series for the time periods under consideration.

The univariate analyses of table 5 are done on the same sample as the bivariate they are reported with, i.e. for the intersection of the samples of both indicators, leading to differences in the reported AUROCs between table 4 and table 5. The consequences of this will vary, and two different cases will be illustrated. Firstly, the reported AUROC for the combination of the private credit to GDP gap and the house price to income gap

[7] All results are available upon request

is lower than that of private credit to GDP gap alone for the full sample (table 4). Since the univariate case is nested in the bivariate by setting the threshold value of the other variable to its minimum, a lower AUROC for the bivariate case is not possible when using the same sample for the univariate and bivariate analyses. The lower performance in the bivariate case stems from the reduced performance of the private credit to GDP gap in this sample as seen from column (1) of table 5. Secondly, the household credit to GDP gap and private credit exuberance now ranks higher on a solo level than the private credit to GDP gap for most combinations, given the limited samples of the bivariate analyses. A result of this is that the household credit to GDP gap is represented in both of the highest ranking indicator pairs. As can be seen from column (1) this clearly comes from the high performance of the indicator for the sample that is the intersection with that of the banking indicators. In this regard it must be mentioned that this is the smallest sample of the evaluation, covering only 21 of the 33 crises, and ending in the fourth quarter of 2009.

Table 5: Bivariate analysis. Individual thresholds with percentile method. Signal horizon is one to three years prior to crisis outbreak.

		(1)	(2)	(3)	(4)	(5)
Indicator 1	Indicator 2	Uni 1	Uni 2	Biv	Crises	Corr
Household credit/GDP	Wholesale/Assets	0.8146	0.6815	0.8305	21	0.226
Household credit/GDP	Equity ratio	0.8146	0.4663	0.8152	21	0.167
Credit exuberance	Global HP/Income	0.7966	0.6854	0.8099	25	0.119
Credit exuberance	Global credit/GDP	0.7984	0.6385	0.8096	25	0.124
Private credit/GDP	Credit exucberance	0.7830	0.7769	0.8091	30	0.546
Private credit/GDP	Global HP/Income	0.7873	0.6723	0.7961	27	0.078
Household credit/GDP	NFE credit/GDP	0.7744	0.7329	0.7829	29	0.448
House Price/Income	House Price exub.	0.7600	0.6851	0.7807	27	0.0.78
Private credit/GDP	House Price/Income	0.7606	0.7269	0.7741	32	0.445
Wholesale/Assets	House Price exub.	0.6735	0.6945	0.7512	22	-0.210

When comparing the AUROC for the bivariate analyses with those of the univariate, using the same sample, the bivariate analyses under the standard method doesn't seem to increase the performance of the indicators by much. The highest ranking pair, given by the AUROC, is household credit to GDP and the wholesale funding ratio gap, but as stated in the previous paragraph, this comes from the high performance of the household credit to GDP gap in this small sample. Decomposing private credit to GDP does not increase the signaling ability, as the bivariate signal of the household credit to GDP gap and non-financial enterprise credit to GDP gap has a lower AUROC than private credit to GDP alone, when evaluated on the full sample. Of the indicator pairs shown here, the only one that has a significant increase in the performance from the univariate to the bivariate case is the wholesale funding ratio and house price exuberance, but the pair still have the lowest bivariate AUROC of the ones reported here. The ROC curves for the two univariate and the bivariate cases are shown in figure 18.

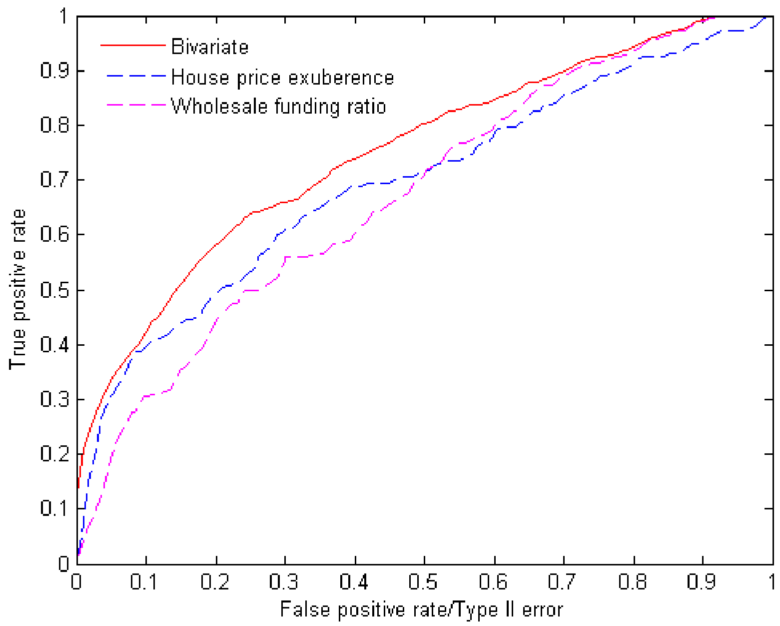


Figure 18: *The ROC curves for house price exuberance and the wholesale funding ratio gap. Both univariate and bivariate. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

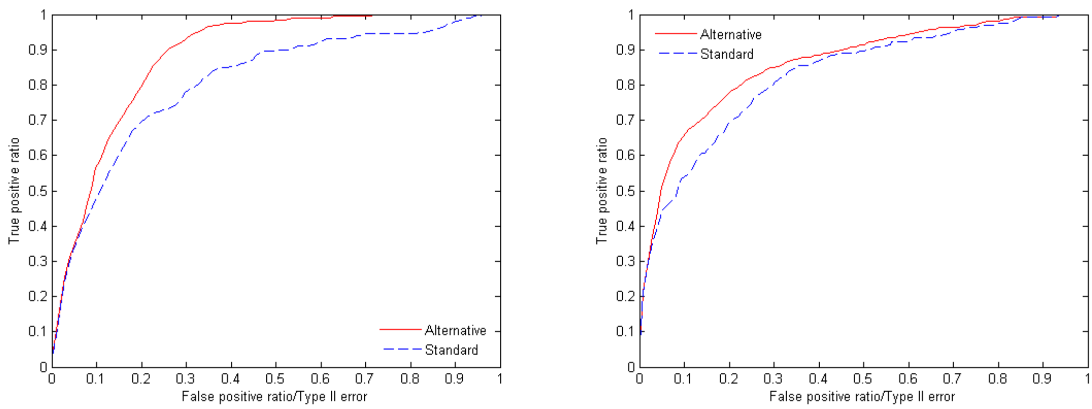
This thesis has presented a new signaling scheme for bivariate analysis. All possible combinations of the indicators have been tested, and the results are presented in table 6 for the top ten performing indicator pairs based on the AUROC, using the new alternative. The thresholds are individual, found by the percentile method, and the signal horizon is from 12 to 5 quarters prior to a crisis. Column (1) shows the AUROC when the standard criteria has been applied. Column (2) show the AUROC with the alternative criteria and column (3) shows the number of crises covered by each pair. Lastly, column (4) presents a conservative measure of the standard error of the difference between the AUROCs of the two methods.

Table 6: Bivariate analyses for the standard and alternative bivariate signaling scheme. Individual thresholds using percentile method. Signal horizon from 12 to 5 quarters prior to crisis

		(1)	(2)	(3)	(4)
Main indicator	Support indicator	Stand.	Alt.	Crises	Se(Diff)
Credit exuberance	Global HP/Income	0.8076	0.8820	25	0.0248
Household credit/GDP	Wholesale/Assets	0.8287	0.8615	21	0.0276
Private credit/GDP	Global HP/Income	0.7936	0.8483	27	0.0254
Credit exuberance	Household credit/GDP	0.7952	0.8471	28	0.248
NFE credit/GDP	Household credit/GDP	0.7802	0.8367	29	0.249
Private credit/GDP	Wholesale/Assets	0.8040	0.8364	23	0.0278
Credit exuberance	Global credit/GDP	0.8072	0.8349	25	0.0262
NFE credit/GDP	Global HP/Income	0.7379	0.8329	25	0.280
House price/Income	Wholesale/Assets	0.7832	0.8318	23	0.0285
Credit exuberance	House price exub.	0.7851	0.8301	27	0.0261

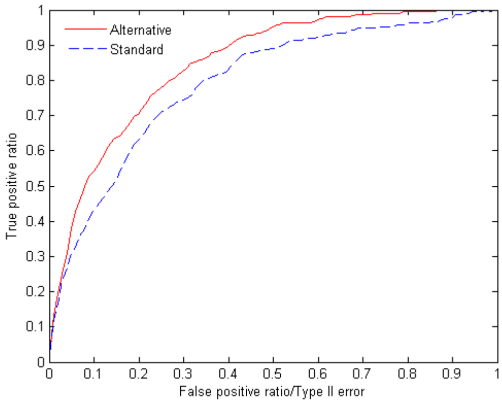
By comparing the bivariate AUROCs from table 5 and the AUROCs from table 6 column (1), where the standard schemes have been used, there are clearly differences. Given that they are based on the same samples and the same methodology one would expect them to be equal. The reason for the discrepancies lie in the limitations of the computations. As previously described, when evaluating the indicators they are tested over a grid of threshold values, where the standard number of grid points for each indicator is 501. For the univariate case this simply gives a one dimensional array of 501 grid points. For the standard bivariate case, where one tests over all combination of threshold pairs, this gives an array of size 501^2. Under the alternative scheme on the other hand, there are two different threshold values for the main indicator, so arrays of the size 501^3, or 125 751 501 entries, are needed to store the information. Along with a lot of other information, the MATLAB class IndicatorEval, used for the computations need to hold 11 arrays of this size, which simply takes up too much memory. Therefore, all calculations done for table 6 are done using 101 grid points for the threshold values.

From table 6 it is clear that using the alternative signaling scheme has a significant effect on the signaling performance of the indicator pairs when compared to the standard scheme. Figure 19 shows the ROC curves for the 10 indicator combinations, for both the standard and the alternative. From the ROC curves of the best performing indicator pair, private credit exuberance and the global house price to income gap, the higher performance of the alternative scheme is reinforced by the fact that the increase in the AUROC comes better signaling performance to the right in the graph, which is the region of most interest. Although the household credit to GDP gap ranks as number two with the wholesale funding ratio gap as a support indicator, there are some drawbacks with this pair, mainly the few observations and crises covered. On the other hand, coupled with the non-financial enterprise credit to GDP gap and private credit exuberance, the household credit to GDP gap holds up fairly well, with significant increases in the signaling performance from the standard approach.

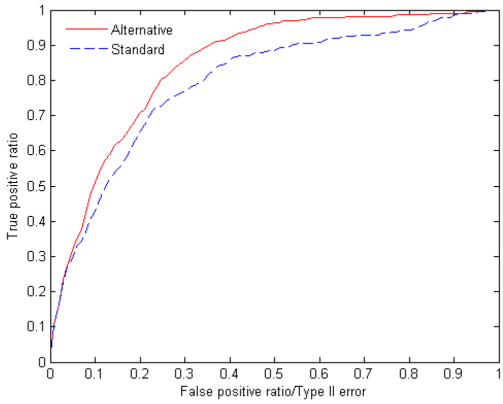


(a) Private credit exuberance and global house price to income gap

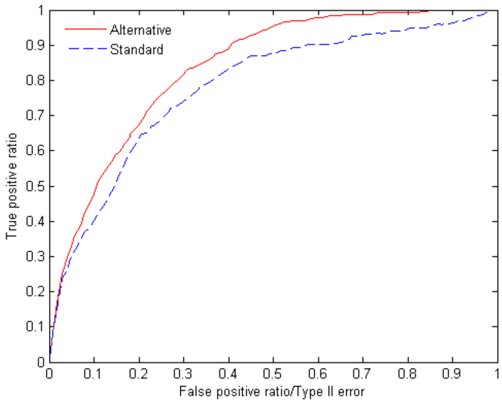
(b) Household credit to GDP gap and wholesale funding ratio gap



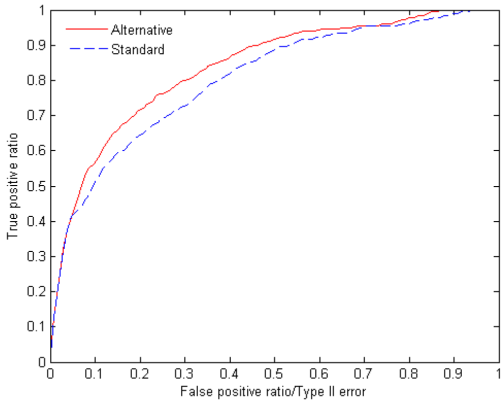
(c) Private credit to GDP gap and global house price to income gap



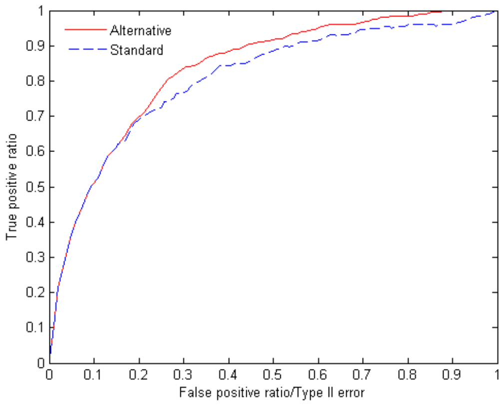
(d) Private credit exuberance and household credit to GDP gap



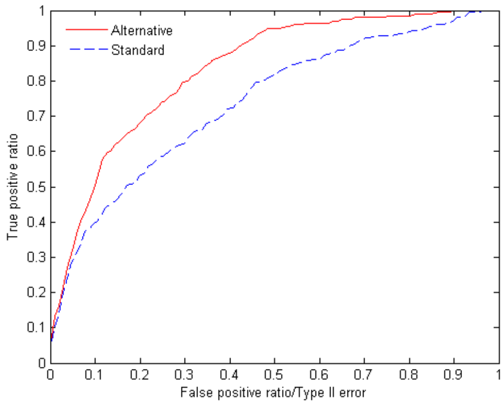
(e) Non-financial enterprise credit to GDP gap and household credit to GDP gap



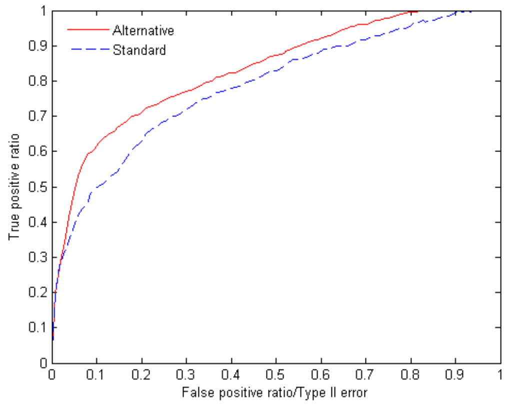
(f) Private credit to GDP gap and wholesale funding ratio gap



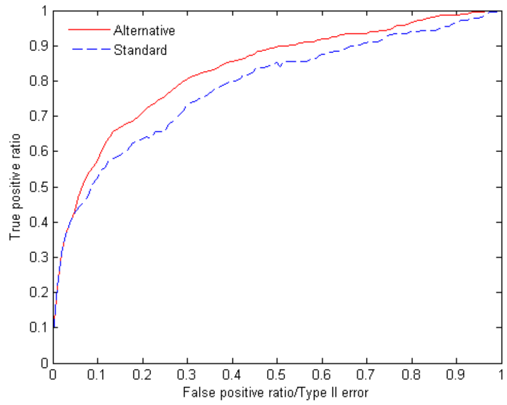
(g) Private credit exuberance and global credit to GDP gap



(h) Non-financial enterprise credit to GDP gap and global house price to income gap



(i) House price to income gap and wholesale funding ratio gap

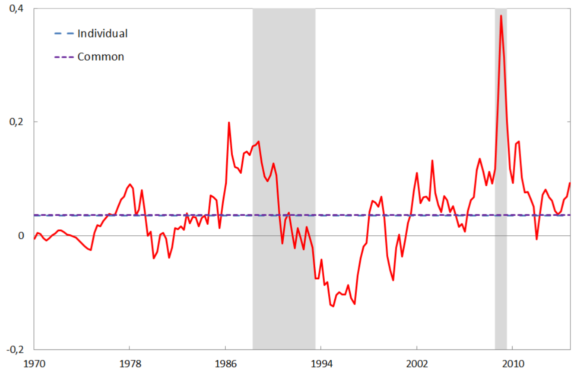


(j) Private credit exuberance and house price exuberance

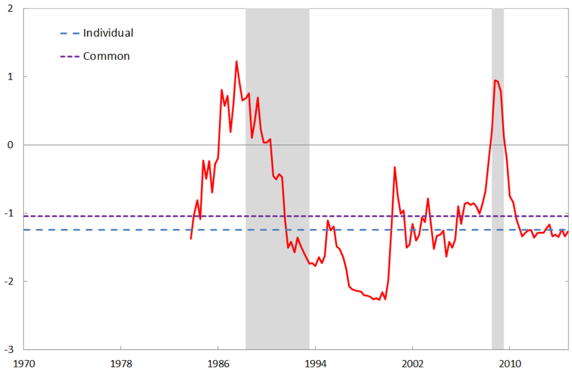
Figure 19: *ROC curves for bivariate analyses using both the standard and alternative signal criteria, for the top 10 performing bivariate pairs. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

5.3 Optimal threshold

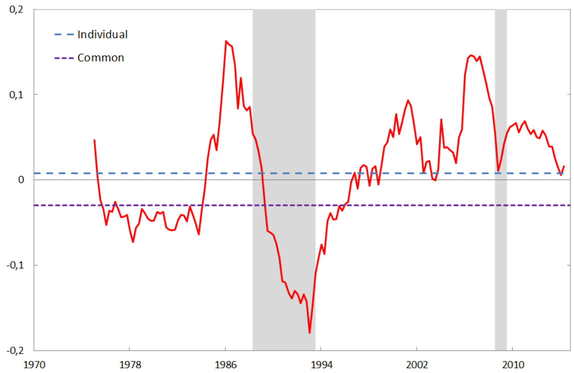
For the single indicator table 4 showed that the best performing indicators are the private credit to GDP gap and the private credit exuberance. It is assumed that the policy maker prefers to falsely signal crises rather than missing them, so $\theta > 0.5$. Both of the best performing indicators seem to have fairly stable optimal threshold values for $\theta \in [0.5, 0.7]$, so the preference parameter will be set to $\theta = 0.6$ for the minimization of the loss. Section 3.2.3 investigated the choice of methods for defining the threshold grid, but a definitive answer wasn't given for whether the individual or the common thresholds was preferable. In figure 20 the time series for the Norwegian key indicators, along with private credit exuberance, are shown together with the optimal thresholds from the minimization of the loss function with $\theta = 0.6$ for both common and individual thresholds.



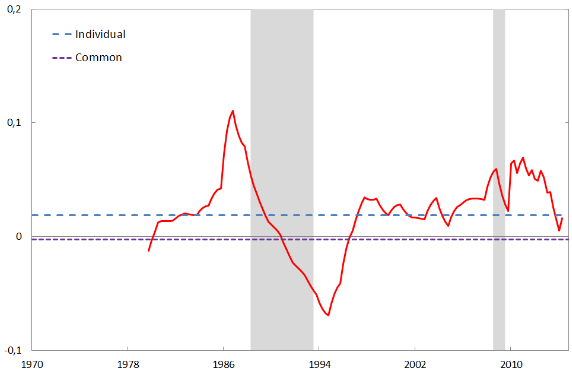
(a) Private credit to GDP gap



(b) Private credit exuberance



(c) House price to income gap



(d) Wholesale funding ratio gap

Figure 20: *Optimal threshold values for Norway for both individual and common thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis*

The first observation is that the choice has no relevance for the private credit to GDP gap under the current specification. For private credit exuberance both methods signal prior to the banking crisis and the financial crisis, but the individual method signals earlier for the latter crisis, but at the same time the lower threshold leads to more false signals. For this indicator the common method may seem to fit the data slightly better. For the remaining two indicators, the house price to income gap and the wholesale funding ratio gap, the individual method clearly fits better with the Norwegian data. Based on this, the optimal thresholds for the single indicators will be given by individual thresholds using the percentile method and a signal horizon from 12 to 5 quarters prior to a crisis. The optimal thresholds for Norway can be seen in table 7

Table 7: Optimal threshold percentiles and values for Norway for the private credit to GDP gap, private credit exuberance, house price to income gap and the wholesale funding ratio gap. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis. $\theta = 0.6$

	Private credit/GDP	Credit exub.	HP/Income	Wholesale/Assets
Percentile	55.6	57.0	46.6	32.6
Value	0.0356	1,2479	0.0078	0.0187

For the bivariate case, the optimal thresholds are only reported for the new alternative signaling scheme, since the standard method is nested in this. Table 8 displays the optimal thresholds for four of the best performing indicator pairs, based on individual thresholds using the percentile method and a signal horizon from 12 to 5 quarters prior to a crisis. The rows identified as "Main" and "Support" shows the thresholds, given as percentiles. The story laid out when presenting the alternative scheme in section 3.4.2 seems to hold for the pair house price to income gap and wholesale funding ratio gap, but for the rest the and/or criteria dominates. Although the optimal threshold for private credit exuberance was stable over a range of θ values in the univariate setting it is even more so now.

A word of caution must still be made. Although the financial crisis of 2007/2008 increased the number of systemic banking crises, the sample is still relatively small to conclude with full certainty how best to signal a systemic banking crisis. Using the alternative criteria allows for more flexibility in the signaling process, which may be a good thing seen as not all crises have the same root cause. Yet, with the increased flexibility the risk of over fitting will also increase.

Table 8: Optimal threshold values expressed as percentiles, true positive rates and false positive rates given by the minimization of the loss function for different preferences of the policy maker, for four pairs of indicators. Individual thresholds using the percentile method. Signal horizon from 12 to 5 quarters prior to crisis

		(1) $\theta = 0.5$	(2) $\theta = 0.6$	(3) $\theta = 0.7$	(4) $\theta = 0.8$
Credit exuberance	Main	74	74	71	69
	Support	2	0	0	0
Global HP/Income	Support	90	84	82	82
	True positive rate	0.905	0.940	0.965	0.970
	False positive rate	0.264	0.309	0.349	0.364
Private credit/GDP	Main	71	71	50	50
	Support	0	0	0	0
Global HP/Income	Support	85	85	85	82
	True positive rate	0.849	0.849	0.958	0.962
	False positive rate	0.317	0.317	0.508	0.521
NFE credit/GDP	Main	88	69	66	66
	Support	2	2	2	2
Household credit/GDP	Support	69	68	55	55
	True positive rate	0.828	0.914	0.966	0.966
	False positive rate	0.306	0.421	0.524	0.524
House price/Income	Main	89	81	89	89
	Support	54	44	2	0
Wholesale/Assets	Support	66	75	29	18
	True positive rate	0.828	0.914	0.966	0.966
	False positive rate	0.306	0.421	0.524	0.524

Credit exuberance = Private credit exuberance, Global HP/Income = Global house price to income gap, Private credit/GDP = Private credit to GDP gap, NFE credit/GDP = Non-financial enterprice credit to GDP gap, Household credit/GDP = Household credit to GDP gap, House price/Income = House price to income gap, Wholesale/Assets = Wholesale funding ratio gap

6 Conclusion

In this thesis I have presented the signaling approach as an early warning system of systemic banking crises. This thesis has illustrated that the noise-to-signal ratio is not an attractive tool for indicator evaluation. Since the actual preferences of the policy maker is hard to quantify, the AUROC has been the main evaluation tool for indicator selection in this thesis, both for univariate and bivariate analyses. When evaluating different signal horizons, it was shown that although the AUROC is higher in one instance than in the next, this does not unequivocally imply that the signaling performance is better in the former case. When defining the threshold values it has been shown that the optimal method is not obvious, but that for Norwegian data, individual thresholds using the percentile method is preferable.

When using a single indicator variable, the different measures of credit dominates when it comes to signaling performance. The private credit to GDP gap ranks as the best, in line with previous work, while the private credit exuberance, a measure calculated by Anundsen et al. (2015), takes second. When finding optimal threshold values by the minimization of a loss function, the two indicators have relatively stable threshold values for a range of preference parameters that seem plausible for the policy maker, thereby increasing their usefulness as early warning indicators. The signaling performance of the private credit to GDP gap has also been shown to be relatively stable over a range of signal horizons starting from 20 to 4 quarters prior to a crisis. The potential loss of credibility for the policy maker in signaling a crisis too early still supports the notion of choosing a signal horizon from 12 to 5 quarters prior to a crisis.

This thesis has introduced a new signaling scheme that has been shown to increase the signaling performance of bivariate signals. The standard criteria, as well as an and/or criteria, can be seen to be nested in the alternative signaling scheme, and for the best performing indicator pairs it is the and/or criteria that gives the optimal thresholds, although there are exceptions. The highest ranking pair is by a clear margin private credit exuberance and the global house price to income gap, with the former being the main indicator. The optimal thresholds can be seen to be higher, but also much more stable than under the univariate analyses.

The two biggest obstacles for the application of the signaling approach to policy seems to be the availability of suitable data and the lack of a measure of the preferences of the policy maker. The former has two components. Firstly, the lack of long and stable time series for the indicators hampers the search for good indicators. For private credit to GDP, relatively good series are available, but for instance for the banking variables, the number of observations are half of that of private credit to GDP. Secondly, the number of systemic banking crises defined in the data is relatively small. With the introduction of new macroprudential policies, the number of systemic banking crises hopefully stays low, but it may at the same time complicate the use of the signaling approach. Countercyclical policies may cloud the indicators, while choices must be made for when a would-be crises should be characterized as a crisis in the data.

References

L. Alessi and C. Detken. Quasi Real Time Early Warning Indicators for Costly Asset Price Boom/Bust Cycles: A Role for Global Liquidity. *Eurpean Journal of Political Economy*, 27(3):520–533, 2011.

A. K. Anundsen, K. Gerdrup, F Hansen, and Kragh-Sørensen. Bubbles and crises: The role of house prices and credit. *Journal of Applied Econometrics*, Forthcoming, 2015.

J. Babecky, T. Havranek, J Mateju, M Rusnak, K Smidkova, and B Vasicek. Banking, debt and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*, 15:1–17, 2014.

H. Borchgrevink, S. Ellingsrud, and F. Hansen. Macroprudential Regulation - What, Why and How? Staff memo 13, Norges Bank, 2014.

C. Borio and M. Drehmann. Assessing the Risk of Banking Crisis - Revisited. BIS Quarterly Review March, Bank for International Settlements, 2009.

C. Borio and P. Lowe. Asset Prices, Financial and Monetary Stability: Exloring the Nexus. Working Paper 114, Bank for International Settlements, 2002.

C. Borio and P. Lowe. Securing sustainable price stability: should credit come back from the wilderness? Working Paper 157, Bank for International Settlements, 2004.

M. Bussière and M. Fratzscher. Low Probability, High Impact: Policy Making and Extreme Events. *Journal of Policy Modeling*, 30(1):111–121, 2008.

E. P. Davis and D. Karim. Comparing Early Warning Systems for Banking Crises. *Journal of Financial Stability*, 4(2):89–120, 2008.

A. Demirgüç-kunt and E. Detragiache. Monitoring Banking Sector Fragility: A Multivariate Logit Approach with Application to the 1996-97 Banking Crisis. Working Paper 2085, World Bank, 1999.

C. Detken et al. Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options. Occsional Paper 5, European Systemic Risk Board, 2014.

M. Drehmann and M. Juselius. Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of forecasting*, 30(3):759–780, 2014.

M. Drehmann and K. Tsatsaronis. The Credit-to-GDP Gap and Countercyclical Capital Buffers: Questions and Answers. BIS Quarterly Review March, Bank for International Settlements, 2014.

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

K. Gerdrup, A. B. Kvinlog, and E. Schaanning. Key indicators for a countercyclical capital buffer in norway - trends and uncertainty. Staff Memo 13, Norges Bank, 2013.

J. A. Hanley and B. J. McNeil. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology*, 143(1):29–36, 1982.

J. A. Hanley and B. J. McNeil. The method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3):839–843, 1983.

R. J. Hodrick and E. C. Prescott. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997.

G. Kaminsky and C. M. Reinhart. The Twin Crisis: The Causes of Banking and Balance-of-Payment Problems. *The American Economic Review*, 89(3):473–500, 1999.

G. Kaminsky, S. Lizondo, and C. M. Reinhart. Leading Indicators of Currency Crisis. Working Paper 79, International Monetary Fund, 1997.

L. Laeven and F. Valencia. Systemic Banking Crises: A New Database. Working Paper 224, International Monetary Fund, 2008.

L. Laeven and F. Valencia. Resolution of Banking Crises: The Good, the Bad and the Ugly. Working Paper 146, International Monetary Fund, 2010.

L. Laeven and F. Valencia. Systemic Banking Crises Database: An Update. Working Paper 163, International Monetary Fund, 2012.

Norges Bank. Criteria for an appropriate countercyclical capital buffer. Norges Bank Papers 1, Norges Bank, 2013.

Basel Committee on Banking Supervision. Guidance for national authorities operating the countercyclical capital buffer. Technical report, Bank for International Settlements, 2010.

C. M. Reinhart and K. S. Rogoff. Is the 2007 U.S. sub-prime financial crisis so different? An international comparison. *American Economic Review*, 98(2):339–344, 2008.

C. M. Reinhart and K. S. Rogoff. The aftermath of financial crises. *American Economic Review*, 99(2):466–472, 2009a.

C. M. Reinhart and K. S. Rogoff. *This Time Is Different: Eight Centuries of Financial Folly.* Princton University Press, Princton, NJ, 2009b.

A. R. van Erkel and P. M. Th. Pattynama. Receiver operating characteristic (ROC) analysis: Basic principles and application to radiology. *European Journal of Radiology*, 27:88–94, 1998.