Occasional Paper Series

No. 5 / June 2014

ESRB
European Systemic Risk Board
European System of Financial Supervision

Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options

Carsten Detken[*1]
Olaf Weeken[*2]
Lucia Alessi[*3]
Diana Bonfim[*4]
Miguel M. Boucinha[*4]
Christian Castro[*5]
Sebastian Frontczak[*6]
Gaston Giordana[*7]
Julia Giese[*8]
Nadya Jahn[*9]
Jan Kakes[*10]
Benjamin Klaus[*11]
Jan Hannes Lang[*3]
Natalia Puzanova[*12]
Peter Welz[*3]

*1 European Central Bank (Chair of the ESRB Expert Group)
*2 ESRB Secretariat (Secretary to the ESRB Expert Group)
*3 European Central Bank
*4 Banco de Portugal
*5 Banco de España
*6 ESRB Secretariat
*7 Banque centrale du Luxembourg
*8 Bank of England
*9 European Central Bank / Finance Center Muenster
*10 De Nederlandsche Bank
*11 Banque de France / European Central Bank
*12 Deutsche Bundesbank

This paper summarises the work of the ESRB Expert Group on establishing guidance for setting countercyclical buffer rates. The Expert Group has been working under the auspices of the ESRB Instrument Working Group and its members are listed in Annex G.

Contents

Executive summary ... 3

Introduction ... 5

Section 1: The data ... 9
- 1.1. The left-hand-side variable ... 9
- 1.2. The-right-hand-side variable ... 10

Section 2: The evaluation approach ... 13
- 2.1. Predictive models ... 13
- 2.2. Evaluation ... 14

Section 3: Evaluation results for the build-up phase ... 21
- 3.1. The credit-to-GDP gap ... 21
- 3.2. Other right-hand-side variables ... 31

Section 4: Evaluation results for the release phase ... 43
- 4.1. A prompt release ... 43
- 4.2. Gradual release ... 45
- 4.3. Use of judgement ... 46

Section 5: Operationalising the indicators ... 47
- 5.1. Thresholds and their mapping according to the BCBS ... 47
- 5.2. Alternative approaches for triggering and calibrating the CCB ... 48

Section 6: Conclusion ... 54

Annex A: Summary of data ... 56

Annex B: Estimation of the pAUROC and statistical inference ... 57

Annex C: Results for the static analysis ... 63

Annex D: Results for the dynamic analysis ... 67

Annex E: Credit-to-GDP ratios, gaps and buffer guides ... 76

Annex F: Multivariate logit model – Robustness tests ... 83

Annex G: List of members of the Expert Group ... 87

References ... 88

Executive summary

This paper presents the analysis underpinning the ESRB Recommendation on guidance on setting countercyclical buffer rates (ESRB 2014/1). The Recommendation is designed to help authorities tasked with setting the countercyclical capital buffer (CCB) to operationalise this new macro-prudential instrument. It follows on from the EU prudential rules for the banking system that came into effect on 1 January 2014.

The analysis underpinning the Recommendation was conducted by a dedicated ESRB Expert Group working under the auspices of the ESRB Instruments Working Group. The Expert Group was made up of representatives from over 30 ESRB member institutions (see Annex G). Its analysis focused on early warning models in order to identify indicators that signal the types of crises that the CCB is designed to mitigate. This includes identifying leading indicators and associated thresholds that signal that the CCB might need to be built up as well as indicators and associated thresholds that suggest that the CCB should be reduced or fully released. For one particular variable, the deviation of the ratio of credit to gross domestic product (GDP) from its long-term trend (credit-to-GDP gap), the analysis also focused on how to map specific levels of the credit-to-GDP gap into indicative settings of the CCB – a so-called benchmark buffer rate.

Consistent with the literature, this paper finds that, in univariate signalling, credit-to-GDP gaps (using bank, household and total credit) are the best single leading indicators for systemic banking crises associated with excessive credit growth. This finding is established here for the European Union *as a whole*. While the specification suggested by the Basel Committee on Banking Supervision (BCBS) performs well for the large majority of countries for which it can be analysed, the credit-to-GDP gap does not perform well in all cases. In particular, some specifications suggest implausibly persistent gaps for a number of countries. In addition, the results are based on little information from central and eastern European transition economies owing to a lack of data.

The main results for credit-to-GDP gaps are robust across a range of different specifications for the gap. The specification suggested by the BCBS is based on total credit to the domestic private non-financial sector and is among the best performing indicators. Few specifications of the credit-to-GDP gap perform better. Those that do are often based on narrower credit aggregates (e.g. bank credit and credit to households) that may be less robust to financial innovation than a broad measure based on total credit.

A number of other variables performed well in univariate signalling, and thus offer a good indication that the CCB may need to be built up. These variables include the residential property price-to-income ratio, residential and commercial property price gaps, the debt service-to-income ratio for households, real bank and household credit growth and the deviation of the (deflated) broad monetary aggregate M3 from its trend.

Multivariate analysis shows that when the credit-to-GDP gap is combined with other variables either in a multivariate signalling approach, a discrete choice model or a decision tree approach, the overall signalling performance improves. In addition to the above-mentioned variables, the overall debt service-to-income ratio, the current account-to-GDP ratio and real equity price growth are useful variables in a multivariate setting. In particular, this reduces the incidence of false alarms.

Market-based indicators have been found to be the best coincident or near-crisis indicators which can be used to signal that the CCB should be reduced or released. The LIBOR-OIS (overnight index swaps) spread, covered bond spreads and the ECB's composite indicator of systemic stress (CISS) perform particularly well. Moreover, these indicators not only perform well for the pooled sample, but also for most individual EU Member States. An important caveat is the short time series availability for many of these data series. This means that the results are largely driven by the recent global financial crisis and the possibility for generalising the results for the release phase is limited. Moreover, there may be circumstances other than stress, e.g. when cyclical systemic risks recede, where the buffer might be reduced, but which may not be captured by this analysis. Judgement may thus need to play an even greater role in the release phase than in the build-up phase.

The analysis in this paper suggests thresholds for each indicator and multivariate model. Policy-makers could view the breaching of such thresholds as a trigger for discussions on the implementation and release of the CCB. Of course, the results are based on an in-sample analysis and therefore the usual caveats for policy use apply.

This paper finds that mapping the credit-to-GDP gap into a benchmark buffer rate poses conceptual challenges. Since the way that the BCBS calibrates the benchmark buffer rate is ad hoc, the Expert Group investigated a number of alternative approaches. Some of the alternatives described in this paper may serve as cross-checks for calibration decisions, but would need to be further developed before becoming operational.

In addition to describing the analysis underpinning the ESRB Recommendation on the CCB, this paper contributes to the literature. First, to the best of our knowledge, this is the first comprehensive study covering all 28 Member Countries of the European Union to the extent possible with the data available. The paper establishes a reference dataset on systemic banking crises associated with domestic credit cycles, which will be useful for further research on CCBs. Second, this paper includes a conceptually more appealing measure of (broad) total credit than most of the existing empirical cross-country studies, which are often based on narrower aggregates. Third, a novel evaluation approach has been used, combining different criteria, i.e. AUROCs (and partial AUROCs), with a loss function approach for specific policy-makers' preferences, as well as taking into consideration the robustness of results across different countries. Fourth, while the analytical findings focus on the CCB, this paper provides a methodology for indicator selection, threshold identification and calibration options that can be adapted to help operationalise other macro-prudential instruments.

Introduction

Pro-cyclicality contributed to the origins of the global financial crisis and was an aggravating factor: during the economic upswing credit grew excessively as banks had easy access to debt funding, whereas credit supply contracted during the economic downturn as this funding dried up (see, for example, Borio et al., 2001; BIS, 2008; Brunnermeier, 2009; and Acharya and Merrouche, 2013). The large economic and social costs of the global financial crisis led to an international effort to address vulnerabilities in the global financial system. As part of this effort, the Basel Committee on Banking Supervision (BCBS) agreed a comprehensive package of reforms to strengthen global capital and liquidity rules for the banking sector – the so-called Basel III framework (BCBS, 2010b, 2011).

The countercyclical capital buffer (CCB) – a key macro-prudential instrument agreed under the Basel III framework – is designed to counter pro-cyclicality in the financial system (BCBS, 2010b). By strengthening the capital base during periods of excessive credit growth, the banking system can absorb losses during the downswing of the financial cycle without constraining the flow of credit to the economy. This makes it important that the CCB swings over the financial cycle – the CCB should be released during periods of stress or when systemic risks abate. In addition, during the upswing of the financial cycle, the build-up of the CCB may help dampen excessive credit growth.

In the European Union the CCB is implemented through the Capital Requirements Directive (CRD IV). The CRD IV (Directive 2013/36/EU), together with its accompanying Capital Requirements Regulation (CRR) (Regulation (EU) No 575/2013), came into force on 1 January 2014 and implements the Basel III framework in the European Union. The CRD IV outlines how authorities that are tasked with setting the CCB should do so. Reflecting the principle of "guided discretion", the designated authorities are obliged to calculate a buffer guide.[1] This buffer guide serves as a reference CCB rate. Authorities are also asked to regularly monitor a range of economic and financial variables that have in the past been associated with excessive credit growth and ensuing financial crises. This, together with qualitative assessments, should guide authorities in setting the CCB. The CCB rate should be higher than zero when credit growth is excessive and poses systemic risks.

Recognising the ESRB's role as the macro-prudential overseer of the EU financial system, the CRD IV gives it a role in guiding EU Member States in the use of the CCB (Art. 135(1), CRD IV). This guidance, which takes the form of an ESRB Recommendation, is designed to help the designated authorities to operationalise the CCB. It consists of four items:

- *Principles* to guide judgement as to the appropriate CCB rate.
- *General guidance* on the measurement and calculation of the credit-to-GDP gap and the calculation of buffer guides.
- *Guidance* on variables indicating the build-up of system-wide risks associated with excessive credit growth.
- *Guidance* on variables that indicate whether the buffer should be maintained, reduced or fully released.

[1] For more details on the rules vs discretion discussion in the context of macro-prudential policy, see Libertucci and Quagliarello (2010) and Agur and Sharma (2013).

To develop this guidance, a dedicated ESRB Expert Group consisting of representatives from over 30 ESRB member institutions was set up (see Annex G). This paper presents the analysis of this Expert Group, which has been working under the auspices of the ESRB Instruments Working Group. This analysis underpins the ESRB Recommendation (ESRB 2014/1) on guidance for setting countercyclical buffer rates.

The analysis presented in this paper is grounded in the vast empirical literature on Early Warning Models (EWMs). These models identify indicators that signal economic vulnerabilities sufficiently early to enable policy-makers to take appropriate action in order to avoid crises or mitigate their severity. EWMs were first developed in the 1990s following currency crises in emerging market economies (Kaminsky et al., 1998; Berg and Pattillo, 1999). Since then the literature has been extended to assess leading indicators for banking crises (Demirgüç-Kunt and Detragiache, 1998), "twin" banking and currency crises (Kaminsky and Reinhart, 1999) and asset price boom-bust cycles and the role of global liquidity factors (see Borio, 2008; Alessi and Detken, 2011; and Eickmeier, Gambacorta and Hofmann, 2013).

The results presented in this paper underscore the role of the credit-to-GDP gap (the deviation of the ratio of credit to GDP from its long-term trend) as a key leading indicator. Results in the literature tend to support this finding. On the one hand, there are a number of studies supporting the leading indicator properties of the credit-to-GDP-gap, for instance those based on its average performance for a group of countries (Borio and Lowe, 2002; Borio and Drehmann 2009; Drehmann et al., 2010, 2011; and Alessi and Detken, 2011) and those based on its performance in individual EEA countries (Denmark: Harmsen, 2010; Finland: VM, 2012; Germany: Deutsche Bundesbank, 2012; the Netherlands: De Nederlandsche Bank, 2010; Norway: Gerdrup et al., 2013; Portugal: Bonfim and Monteiro, 2013; Sweden: Juks and Melander, 2012; and the United Kingdom: Giese et al., 2014). In addition, Sveriges Riksbank (2011) suggests that the credit-to-GDP gap would have foreseen the build-up of imbalances in the Baltic countries. On the other hand, the credit-to-GDP gap performed less well as a predictor of the global financial crisis in Belgium (Keller, 2011) and Austria (Eidenberger et al., 2013). Moreover, its use as an indicator may be particularly challenging in central and eastern European (CEE) countries, where credit markets expanded rapidly in the 1990s following the transition from centrally planned economies (Geršl and Seidler, 2011).

Despite the good performance of the credit-to-GDP gap, investigating other indicators is important. As discussed by Giese et al. (2014), it is important to augment the credit-to-GDP gap with a range of complementary indicators. This is partly due to the fact that other indicators might provide additional information on the build-up of system-wide vulnerabilities. On the other hand, different kinds of variables may be better suited to providing information for the decision on whether to reduce the CCB (release phase) since the gap displays a poor signalling ability in this context. Some authors level a more conceptual criticism against the credit-to-GDP gap. For example, Repullo and Saurina (2011) point out that the credit-to-GDP gap tends to be negatively correlated with GDP growth. They advocate focusing instead on credit growth, which is positively correlated with GDP growth and is a good predictor of financial crises (Schularick and Taylor, 2012). Drehmann and Tsatsaronis (2014) argue that the relevant concept is the correlation of the credit-to-GDP gap with the financial cycle, not the business cycle. They show that the correlation with the former is indeed positive. Other authors point to more technical challenges in the calculation of the credit-to-GDP gap. These challenges are akin to those found in output gap estimations (Orphanides and van Norden, 2002) and are associated with the use of a filter devised by Hodrick and Prescott (1980) to estimate the trend component of the credit-to-GDP gap. In particular, the Hodrick-Prescott filter is sensitive to end points and may induce spurious results if applied to series which are integrated or driven by deterministic trends (Harvey and Jaeger, 1993; and Cogley and Nason, 1995). Edge and Meisenzahl (2011) discuss how such unreliable end-of-sample estimations and data revisions can affect the credit-to-GDP gap. Van Norden (2011) shows that ex post revisions do not, however, negatively affect the performance of the credit-to-GDP gap as an indicator for the build-up phase of the CCB.

This paper investigates a range of indicators for the build-up phase that have been found most promising in the literature. These include credit-related variables other than the credit-to-GDP gap such as credit growth (Repullo and Saurina, 2011 and Schularick and Taylor, 2012, among others); debt service ratios (Drehmann and Juselius, 2012); other macroeconomic and macro-financial variables, such as GDP growth, the consumer price index (CPI), the unemployment rate, the broad monetary aggregate M3, the real exchange rate, the current account balance, interest rates (Borgy et al., 2009; Borio and Drehmann, 2009; Barrell et al., 2010a; Drehmann et al., 2010, 2011; Babecky et al., 2011 and Kauko, 2012a) property variables, such as property prices, as well as price-to-income and price-to-rent ratios, and equity prices (Riiser, 2005; Mendoza and Terrones, 2008; Borio and Drehmann, 2009; Barrell et al., 2010a; Drehmann et al., 2010, 2011; and Claessens et al., 2011a, 2011b). Global liquidity indicators have also been found useful in the literature (see Borio, 2008; Alessi and Detken, 2011; and Eickmeier, Gambacorta and Hofmann, 2013).

In addition to the univariate analysis, this paper explores three main approaches that help extract signals from the EWMs that contain a combination of indicators. These multivariate approaches are the multiple indicator signalling approach, multivariate discrete choice models and decision trees. For instance, CGFS (2012, p. 47f) provides two examples of the multiple indicator signalling approach: the “weighted signalling” and the “multidimensional grid search” approaches. The “weighted signalling” approach takes a linear combination of the original set of indicators and applies the same signal extraction techniques to the resulting composite indicator as in the univariate case. By contrast, the “multidimensional grid search” calculates separate thresholds for each of the original indicators and computes a signal when all or a minimum number of indicators breach their respective thresholds. Relevant empirical evidence for such models is provided by Drehmann and Juselius (2014), who find that combining the credit-to-GDP gap with the debt service ratio improves its signalling performance. As for other multivariate models, Frankel and Rose (1996) and Licchetta (2011) use a methodology based on multivariate probit models in the context of predicting currency crises. Demirgüç-Kunt and Detragiache (1998) also apply a multivariate probit model approach for predicting banking crises. Behn et al. (2013a) find a logit model, in which they combine credit-related variables with equity, house price and banking sector variables, more useful for predicting banking crises. By using decision trees, Alessi and Detken (2014) show that the bank credit aggregate contains useful information, and that credit-to-GDP gaps, ratios of credit to GDP and rates of credit growth should be considered in a unified framework.

The literature on indicators for the release phase is scarce. Drehmann et al. (2010, 2011) suggest that real credit growth, bank losses and market indicators may all be useful in signalling the need to release the buffer. But their analysis does not distinguish between the two possible release scenarios which are relevant in the context of the CCB: a gradual release in the case of threats receding, and a prompt release during periods of financial stress. In the first case the indicators used for the build-up phase should be informative, although flow-based measures may be more helpful in identifying turning points in the financial cycle than stock-based ones. Recent research has provided important insights into the characterisation of the financial cycle: see Aikman et al. (2011), Claessens et al. (2011a, 2011b), Busch (2012), Drehmann et al. (2012) and Schularick and Taylor (2012). In addition, efforts have been made to identify the different phases of the cycles – i.e. expansion and contraction – as well as capturing the effects of asymmetric shocks on the credit supply (see, for example, Ceron and Suarez, 2006 and Anguren Martin, 2011). But the financial cycle is still less well understood than the business cycle, and further empirical work is needed to identify its turning points in real time. The prompt release of the buffer should be guided by near-coincident indicators that are readily available and can reflect stress in the financial sector. As an example, Juks and Melander (2012) illustrate how, for Sweden, a composite financial stress index could have informed the decision on the release of the CCB as the global financial crisis approached the country. Taking these considerations into account, this paper investigates the signalling performance of market based indicators (the ECB's CISS indicator described in Holló et al., 2012; the LIBOR-OIS spread[2]; bank and sovereign credit default swap (CDS) premia; covered bonds spreads, etc.) and an aggregate bank balance sheet indicator (the non-performing loans ratio). Not analysed in this paper but potentially informative for the release phase are model-based measures of systemic risk as proposed by Segoviano and Goodhart (2009), Schwaab et al. (2011), Adrian and Brunnermeier (2011), Acharya et al. (2012) and Puzanova and Düllmann (2013), among others.

This paper contributes to the literature in a number of ways. First, to the best of our knowledge, this is the first comprehensive study covering all 28 Member States of the European Union. The paper establishes a reference dataset on systemic banking crises associated with domestic credit cycles, which will be useful for further research on CCBs. Second, most of the empirical cross-country studies cited above only test the credit-to-GDP gap based on a narrow definition of bank credit to the domestic non-financial sectors. By contrast, this paper includes a conceptually more appealing measure of broad credit based on the new credit data compiled by the BIS (Bank for International Settlements, 2013), which have been additionally enhanced and extended by members of the Expert Group to include information on all 28 EU Member States. Third, a novel evaluation approach has been used, combining different criteria, i.e. the Area Under the Receiver Operating Characteristics Curve (AUROC and partial AUROC) with a loss function approach for specific policy-makers' preferences, as well as taking into account the robustness of results across different countries. Fourth, while the analytical findings presented focus on the CCB, this paper provides a methodology for indicator selection, threshold identification and calibration options that can be adapted to help operationalise other macro-prudential instruments.

The remainder of this paper is organised as follows. Section 1 describes the data. Section 2 introduces the evaluation approach. Section 3 discusses the evaluation results for the build-up phase. Section 4 discusses the evaluation results for the release phase. Section 5 considers how to operationalise the indicators through a variety of approaches in order to trigger and calibrate the CCB. Section 6 concludes.

[2] See Federal Reserve Bank of St. Louis (2009) for a short description of the LIBOR-OIS spread as a market-based solvency risk metric.

Section 1: The data

The empirical analysis requires two types of variables: A left-hand-side variable that identifies the type of financial crises the CCB is meant to mitigate, and right-hand-side variables that serve as leading indicators for crises. This section describes the underlying data. All data were adjusted for seasonal patterns, where applicable. Other data transformations and the evaluation results are described in Section 3.

1.1. The left-hand-side variable

One of the challenges in early warning models is to ensure consistency in the definition of a banking crisis in order for it to be used as dependent variable (Davis and Karim, 2008a). Given the different dynamics in crisis episodes, it is difficult to find a single quantitative variable for defining the periods of stress and a degree of subjectivity is usually necessary. In order to ensure consistency in the definition of a banking crisis for the sample, while taking into account the specific characteristics of crisis experiences in different EU Member States and the objectives and operation of the CCB, a two-step approach was followed.

First, the ESCB Heads of Research (HoR) Group's banking crises database, compiled by the Macro-prudential Research Network (MaRs), formed the basis and starting point for defining the binary crisis variable.[3] This database defines a banking crisis as one that demonstrates significant signs of financial distress in the banking system as evidenced by bank runs in relevant institutions or losses in the banking system (non-performing loans above 20% or bank closures amounting to at least 20% of banking system assets), or significant public intervention in response to losses in the banking system or taken to prevent the realisation of such losses.[4]

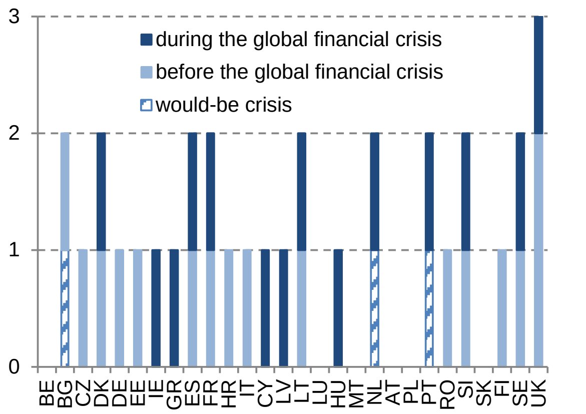
Second, in order to align the database with the objectives and operation of the CCB, members of the Expert Group were asked to amend the database. This was done as follows: (1) crises that were not systemic banking crises were excluded (2) systemic banking crises that were not associated with a domestic credit/financial cycle were excluded, and (3) periods where domestic developments related to the credit/financial cycle could well have caused a systemic banking crisis had it not been for policy action or an external event that dampened the financial cycle – henceforth "would-be" crises – were added. The variable takes a value of 1 if a period meets one of these criteria and a value of 0 otherwise.

Data were collected from the first quarter of 1970 to the fourth quarter of 2012 for all 28 EU Member States. However, six countries (Austria, Belgium, Luxemburg, Malta, Poland and Slovakia) did not experience any crisis consistent with the above criteria over the sample period. Of the remaining 22 countries 12 experienced one crisis, nine experienced two crises and one experienced three crises (Chart 1). These 33 crisis episodes were, however, clustered around the global financial crisis (Chart 2), with only 19 crisis episodes outside the global financial crisis (more detailed information about the duration of each crisis episode for all countries is provided in Table A1 in Annex A). The limited

[3] See the ECB's website (http://www.ecb.europa.eu) for more information on the Macro-prudential Research Network.
[4] The literature on early warning models usually relies on a number of studies providing lists of crisis events, for example Demirgüç-Kunt and Detragiache (2005) and Caprio and Klingebiel (2003).

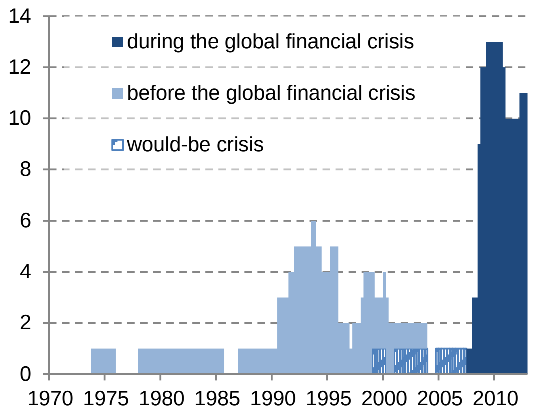
number of crisis observations in particular regions or countries limited the scope to conduct stringent performance evaluations, such as out-of-sample predictions, on the predicting models explored (see Berg et al., 2005). A further constraint was that a few crises occurred before the first observations for the indicator variables had been recorded. This reduced the number of crisis episodes that could be considered in the analysis from 33 to 25.

Chart 1: Number of actual and would-be crisis per country



Notes: BE=Belgium, BG=Bulgaria, CZ=Czech Republic, DK=Denmark, DE=Germany, EE=Estonia, IE=Ireland, GR=Greece, ES=Spain, FR=France, HR=Croatia, IT=Italy, CY=Cyprus, , LV=Latvia, LT=Lithuania, LU=Luxemburg, HU=Hungary, MT=Malta, NL=Netherlands, AT=Austria, PL=Poland, PT=Portugal, RO=Romania, SI=Slovenia, SK=Slovakia, FI=Finland, SE=Sweden, UK=United Kingdom.

Chart 2: Number of countries in actual or would-be crisis



1.2. The-right-hand-side variable

1.2.1. The credit and GDP data

The credit-to-GDP gap was constructed from quarterly nominal credit and nominal GDP data. Table A2 in Annex A (the sixth column in the section "credit variables") summarises the years for which data for each country were available. It shows that for 11 EU Member States both nominal GDP and credit data were available from the first half of the 1970s, with data for another three EU members available from the first half of the 1980s. However, for 14 EU members – mainly transition economies – both of these data series were available only from the mid-1990s or later. Moreover, few of the transition economies had developed mortgage markets before the first half of the 2000s, which further limited the information content of these data.

1.2.1.1. Nominal GDP data

Where available, nominal GDP data were seasonally adjusted and working day adjusted data from Eurostat. Members of the Expert Group were asked to extend the data where possible to cover the 1970s using national data sources or – where Eurostat data were not available – to provide alternative quarterly data from national data sources.

1.2.1.2. Nominal credit data

The starting point for the credit data was the BIS data for total credit to the domestic non-financial private sector (Dembiermont et al., 2013). Where BIS data were not available and/or where time series were short, members of the Expert Group were asked to provide additional, consistent data from national data sources and to liaise with the BIS where appropriate.

Although the CCB applies only to banks, the measure of credit aims to capture not just domestic bank credit, but credit from all sources. This includes credit from abroad and debt securities issued by non-financial corporations. The rationale for this broad concept of credit, which is consistent with the BCBS's approach (2010b), is twofold. First, it recognises that banks can suffer the consequences of a period of excess credit growth even if they were not the driving forces behind the growth. For example, a domestic corporation that defaults on borrowing from a foreign non-bank is also likely to default on its borrowing from domestic banks. Second, it recognises that limiting the definition of credit to bank credit may provide incentives for regulatory arbitrage and may drive the growth of the shadow banking system. However, alternative measures of credit may also provide useful signals to the designated authorities and were included in the analysis (Table 1).

1.2.2. Other right hand side variables

The Expert Group collected data on a number of other indicators. Most of these data were obtained from public databases (e.g. Eurostat, BIS, Bloomberg). Where appropriate, data series were extended and gaps were filled by members of the Expert Group, in some cases by incorporating data that are not publically available.

The data can be categorised into five subsets: real-economy variables, other credit-related variables, market-based variables, property variables and variables on bank balance sheets. These variables are listed in Table 1 below. As with the total credit and nominal GDP data described above, data availability differs from country to country, generally with shorter data availability in the transition economies. Data on bank balance sheets poses particular challenges. The data are summarised in Table A2 in Annex A.

Table 1 – Summary of data used

Category	Variable
Real-economy variables	- nominal GDP - real GDP* - consumer price index - unemployment rate - nominal M3 - real effective exchange rate - current account balance
Other credit-related variables	- nominal total credit to non-financial sector* - nominal total credit to non-financial corporations - nominal total credit to households - nominal bank credit to non-financial sector - alternative measure of total nominal credit – available for Belgium and Sweden - ratio of nominal public debt to nominal GDP - debt service ratio – all agents - debt service ratio – non-financial corporations - debt service ratio – households
Market-based variables	- nominal three-month money market rate - nominal long-term interest rate - nominal equity prices* - LIBOR-OIS spread** - average bank CDS premia** - sovereign CDS premia** - Merrill Lynch covered bond spread** - ECB's CISS indicator**
Property variables	- real residential property prices - nominal residential property prices - ratio of nominal residential property prices to nominal income - ratio of nominal property prices to nominal rent - nominal commercial property prices*
Bank balance sheet variables	- ratio of non-performing loans to total gross loans** - leverage ratio

*) These variables are used for both the build-up and the release phase.
**) These variables are used for the release phase only.

Section 2: The evaluation approach

There are three main EWM approaches used in the literature to extract signals from indicators in order to predict crises. The first – henceforth the "signalling approach" – uses the variable in question as an indicator without any further transformations (e.g. Drehmann et al., 2011). The second approach transforms the variable into crisis probabilities using a logit or probit model – henceforth the "discrete choice approach" (e.g. Demirgüç-Kunt and Detragiache, 1998). The third approach – henceforth "decision trees" (e.g. Alessi and Detken, 2014) – is based on numerical algorithms that allocate a set of variables with larger discriminatory power in a "decision tree" format and calculate optimal decision thresholds at each node of the tree. Decision trees remain largely unexplored in the literature on EWMs. The signalling properties of different indicators/models can be evaluated for each of the three approaches. The remainder of this subsection describes this in more detail.

2.1. Predictive models

2.1.1. Signalling approach

The signalling approach is one of the most common approaches for the statistical evaluation of EWMs. Following the approach of Kaminsky and Reinhart (1999), a number of papers have used this method for an ex post examination of how well various indicator variables signal approaching crises (e.g. Borio and Lowe, 2002; Borio and Drehmann, 2009; Drehmann et al., 2010, 2011; Alessi and Detken, 2011; and CGFS, 2012). The signal is extracted directly from the indicator variable when it breaches a pre-determined or optimised threshold. The signalling approaches used in the economics literature are mostly univariate approaches, i.e. they use one indicator variable at a time. In the more computationally intensive multivariate signalling approach several indicator variables are combined in order to derive a signal which is potentially more robust. For example, CGFS (2012) outlines a bivariate signalling approach that combines the information from the credit-to-GDP gap with the price-to-rent gap.

2.1.2. Discrete choice

Another workhorse for EWMs is the discrete choice approach. It became standard following the seminal papers by Frankel and Rose (1996), Hardy and Pararbasioglu (1998) and Demirgüç-Kunt and Detragiache (1999). Recent contributions in this field include Davis and Karim (2008b), Barrell et al. (2010b), Lund-Jensen (2012), Lo Duca and Peltonen (2013) and Behn et al. (2013a). In a discrete choice model a binary classification set-up first maps various explanatory variables into the probability of a systemic banking crisis, i.e. either a probit or a logit mapping function transforms the variables into a continuous indicator variable between 0 and 1. This indicates the crises probability. If the probability exceeds a specified threshold, a signal is issued. A discrete choice model can include one or several indicator variables at a time. While in the case of the multivariate signalling approach a joint condition needs to be fulfilled for a crisis to be signalled (e.g. all indicator variables breaching a specific threshold), in a multivariate discrete choice model each variable included reflects the marginal contribution of that variable. All variables then jointly determine a continuous crisis probability which, when exceeding a specific (optimised) threshold, signals a crisis.

2.1.3. Trees

A third approach, which Alessi and Detken (2014) propose as a predictive model for the CCB, is decision tree learning (see Manasse, Savona, Vezzoli, 2013). A decision tree, and in particular a binary classification tree, is a partitioning algorithm that recursively identifies the indicator variables and the respective thresholds that best split the sample into two relevant classes: tranquil and pre-crises periods. For early warning purposes, the classification tree can be followed according to the current values of the relevant indicator variables, to check whether the model foresees a crisis. The selection of the set of indicator variables on which the tree is “grown” can affect the results. To overcome this issue, more advanced decision tree techniques such as the so-called “random forest” can be used. In a random forest a multitude of trees are bootstrapped and aggregated in order to robustly select the most relevant indicator variables to feature in the decision tree.

2.1.4. Discussion of the three approaches

There are few studies to guide the choice between these three approaches. An advantage of the discrete choice approach over the signalling approach is that a logit or probit model gives an immediate understanding of whether a variable is statistically significant in relation to crisis observations, even before a threshold for the crisis probabilities is set (see Barrell et al., 2010b). However, the results may not be robust in small samples. This makes out-of-sample performance difficult to test (see Berg et al., 2005). Moreover, the discrete choice approach allows different variables to be included in the same model and hence they can be evaluated jointly (see Frankel and Rose, 1996 and Licchetta, 2011 on the use of multivariate probit models for the prediction of currency crises; and Demirgüç-Kunt and Detragiache, 1998 for the prediction of banking crises). Alessi and Detken (2011) highlight that while the signalling approach assumes extreme non-linearity between the indicator and the binary crisis variable, it allows a large number of determinants to be tested without the risk of being misled by potentially wrong inference resulting from many regressions. Davis and Karim (2008a) suggest that the signalling approach is statistically superior for country-specific EWMs but that the discrete choice approach is more efficient if a global perspective is sought, as in Lo Duca and Peltonen (2013) or Behn et al. (2013a), for instance. The output of the discrete choice approach in the form of crisis probabilities may also be easier to interpret and provides for a straightforward time and cross-section comparison.

Classification trees retain some of the advantages of both the signalling and the discrete choice approaches: they are both easy to explain and use and able to provide an EWM where the relevant indicators are considered conditionally on whether other indicators breach certain thresholds. However, further research is needed in order to assess this method’s adequacy and robustness for guiding policy decisions.

2.2. Evaluation

Under each of the approaches an indicator either stays below a threshold and issues no signal or it breaches a threshold and issues a signal. The different outcomes can then be classified within a so-called “confusion matrix”: when a signal is issued and a crisis occurs within a predefined horizon, it is classified as correct (A) or, if no crisis occurs, it is classified as incorrect (B). When no signal occurs and a crisis occurs within the predefined horizon, it is classified as incorrect (C) or, if no crisis occurs, it is classified as correct (D). This is summarised in Figure 1.

Figure 1: Confusion matrix and definitions

	crisis	no crisis
signal	A	B
no signal	C	D

Signal ratio / true positive rate $= \frac{A}{A+C}$

Noise ratio / false positive rate / Type II error rate $= \frac{B}{B+D}$

Type I error rate (1 - true positive rate) $= \frac{C}{A+C}$

Policy makers' loss function: $L = \theta \frac{C}{A+C} + (1-\theta) \frac{B}{B+D}$

2.2.1. The noise-to-signal ratio

The noise-to-signal ratio is a common measure in economics literature which is used to compare the signalling qualities of different models for a given threshold. It is defined as the noise ratio (the fraction of false positives – i.e. type-II errors – over all non-crisis episodes) divided by the signal ratio (the fraction of correctly predicted crises over all crisis episodes). When two models are compared a lower noise-to-signal ratio indicates better signalling ability.

A problem with the noise-to-signal ratio as an evaluation criterion is that it relies on a specific threshold and often reaches its minimum value at both very low noise and signal ratios.[5] Such a constellation will usually be associated with a high threshold – the higher the threshold, the fewer signals will be issued. A high threshold would suggest that policy-makers are extremely averse to false alarms, but put little weight on missing financial crises. This is unlikely to reflect policy-makers' true preferences. To address this Borio and Drehmann (2009) suggest minimising the noise-to-signal ratio subject to at least two-thirds of the crises being correctly predicted.[6] In addition, following the seminal work of Demirgüç-Kunt and Detragiache (1999), several authors have derived the optimal threshold by minimising a loss function. A loss function explicitly takes into account policy-makers' preferences (measured by θ in Figure 1 above) with regard to type-I errors and type-II errors (see, for instance, Alessi and Detken, 2011).

2.2.2. The area under the receiver operating characteristic (AUROC) curve

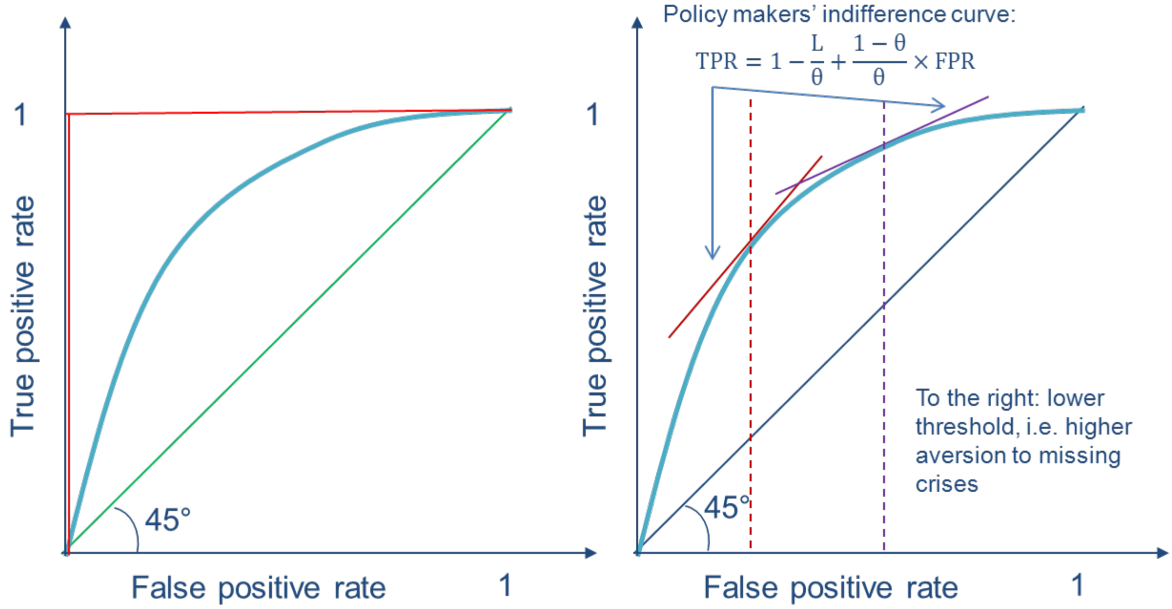
More recently the AUROC has been used as an evaluation criterion in the economics literature, for instance Berge and Jordà (2011), Jordà and Taylor (2011), Candelon et al. (2012), Jordà (2012), Drehmann and Juselius (2014), Betz et al. (2013) and Behn et al. (2013a). The advantage of using an AUROC is that it takes all possible threshold values into account. The Receiver Operating Characteristic (ROC) curve plots the noise ratio (false positive rate) against the signal ratio (true positive rate) for every possible threshold value. High thresholds are close to the origin (as few signals will be issued under a high threshold, few crises are correctly identified and few incorrectly

[5] To see this, note that the noise to signal ratio would be minimised for the true and false positive rates on the tangential point between a line from the origin and the ROC curve in Figure 2.

[6] Davis and Karim (2008a) discuss some policy implications from the selection of the optimisation procedure for the thresholds.

signalled) whereas low thresholds are close to the (1;1) point (as many signals will be issued under a low threshold, many crises will be correctly identified, but many false signals will also be issued). The left panel in Figure 2 illustrates the different shapes of the ROC.

Figure 2: The ROC and AUROC



The AUROC is then computed as the area under the ROC curve and provides a summary measure which ranges from 0 to 1. An AUROC value of 0.5 indicates uninformative indicators (the green line in the left panel of Figure 2). The AUROC is larger than 0.5 if an indicator is informative and tends to be higher ahead of crises than during normal times (e.g. the blue curve in the left panel of Figure 2). A value of 1 indicates a perfect indicator (the red line in the left panel of Figure 2).

2.2.3. Policy-makers' preferences and the partial standardised AUROC (psAUROC)

Since no indicator is perfect, there is always a trade-off between missed crises (type-I errors) and false alarms (type-II errors). Using a higher threshold to signal crises will result in more type-I errors and fewer type-II errors, whereas a lower threshold will result in the reverse. By taking all possible threshold values into account the AUROC is robust to different preferences.

If policy-makers' preferences were known with certainty, one could choose the threshold at which the policy-makers' indifference curve is tangential to the ROC. The right panel of Figure 2 illustrates this for two different policy preferences; the red indifference curve shows the case of a policy-maker that places a relatively low weight on missing crises, whereas the purple indifference curve further to the right shows the case of a policy-maker that places a relatively high weight on missing crises. While policy-makers' preferences between type-I and type-II errors cannot be known with certainty, it may be possible to specify a range for their likely preferences. For example, Demirgüç-Kunt and Detragiache (1999), Borio and Lowe (2002) and Borio and Drehmann (2009) argue that policy-makers may be more concerned about missing crises, since the costs of crises are high relative to the costs of taking preventive action in the case of a false alarm. Alessi and Detken (2011) argue that before the global financial crisis central bankers were generally less averse to missing a crisis than to receiving a false alarm and that after the crisis these preferences have become more balanced. As pointed out by Betz et al. (2013), the policy-makers' preference parameter may also depend on whether the model prediction is made public or not. If the model prediction is not made public, a higher preference parameter for detecting crises is likely. An early warning signal would trigger an internal in-depth review of the situation and would not lead to credibility losses if no crisis materialised afterwards.

This Expert Group assumed that, following the global financial crisis, policy-makers are at least as concerned about missing crises as about wrongly signalling crises. In other words, the preference parameter theta (θ), which is associated with a specific indifference curve in Figure 2 lies in the interval [0.5; 1]. So instead of considering only the full AUROC (e.g. Drehmann and Juselius, 2014), this paper also presents a partial standardised AUROC (psAUROC) that cuts off the area associated with a preference parameter of $\theta<0.5$. A further extension not pursued in this paper would be to further constrain the AUROC by cutting off an area associated with implausibly high values of theta. While the psAUROC has been used extensively in the area of medical statistics to assess the performance of a classifier only in specific regions of the ROC curve (e.g. McClish, 1989 and Jiang et al., 1996), it is a new approach in the literature evaluating EWMs. Annex B describes the psAUROC in more detail. The results reported in this paper show that the psAUROC can reveal useful additional information as long as the partial area does not become too restricted.

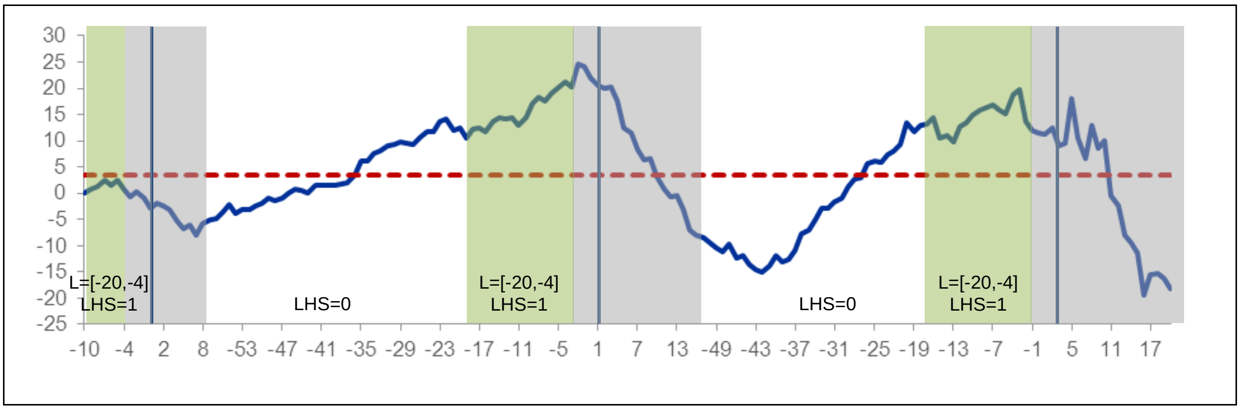
2.2.4. The evaluation procedures

As in Drehmann and Juselius (2014), indicators are evaluated using a "static" procedure and a "dynamic" evaluation procedure. The dynamic evaluation procedure has the advantage of highlighting the stability of the signalling performance of an indicator at different horizons. Drehmann and Juselius (2014) emphasise that this is important from a policy perspective. The resulting AUROCs or psAUROCs can also be averaged over the periods considered most relevant for the policy process in order to derive the most suited models (denoted av(ps)AUROC). However, the dynamic evaluation procedure can only be conducted for the pooled data as there are not sufficient data points for a country-by-country analysis. The static evaluation procedure has the advantage of being able to be conducted both for the pooled dataset and on a country-by-country basis (in the latter case the pooled optimal threshold is used). It can thus provide information about the number of countries in the European Union for which an indicator provides a significant signal. The remainder of this section explains in more detail the static and dynamic evaluation procedures and the associated decision criteria used to discriminate between indicators.

2.2.4.1. Static evaluation procedure

Chart 3 illustrates the static evaluation procedure. The thick blue line is the early warning indicator. The vertical blue lines denote the start date of a crisis and the grey shaded areas to the right of the vertical blue lines indicate the crisis duration. All crises quarters and three quarters prior to each crisis (the grey shaded areas) are excluded from the analysis. The green shaded areas represent the evaluation period. Given that banks should typically be given one year to build up the CCB, the Expert Group decided to use an evaluation period that lasted from five years to one year prior to a crisis (i.e. from 20 quarters to four quarters: L=[-20,-4]). During this period, the left-hand-side (LHS) variable is set to 1 or to 0. Signals are then classified into the different elements of the confusion matrix, depending on the value of the LHS variable and whether the indicator variable is above or below the specified threshold (indicated by the red dashed line).

Chart 3: Illustration of the design of the "static" procedure



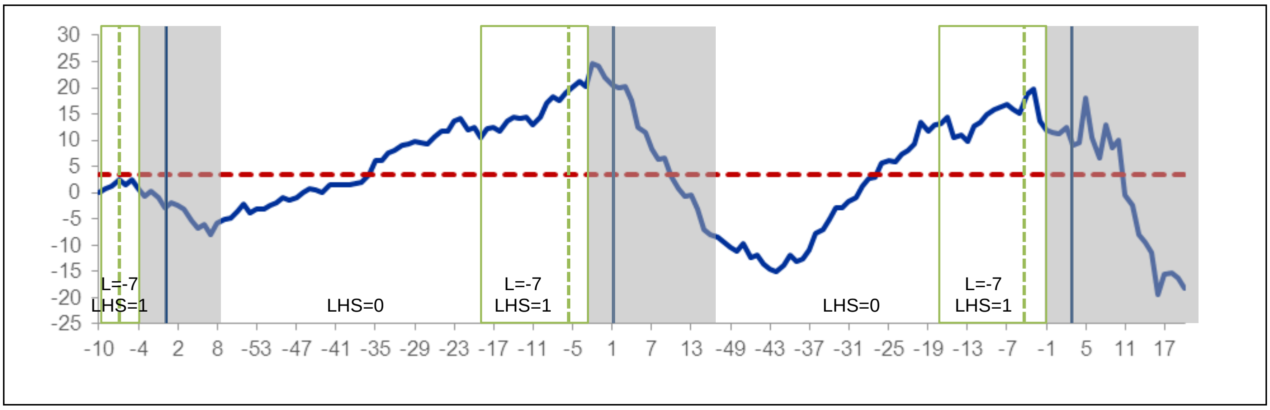
Notes: Data are for the United Kingdom's credit-to-GDP gap as defined by the BCBS (2010b).

To assess the overall performance of an early warning indicator in predicting a crisis, AUROCs and psAUROCs are estimated over the entire evaluation period L=[-20,-4]. Each quarter within the evaluation period during which a signal was issued is counted as a correctly identified pre-crisis quarter, whereas each quarter during which no signal was issued is counted as a missed pre-crisis quarter. Based on this, the signal ratio (true positive rate) is calculated as the number of correctly identified pre-crisis quarters as a share of the total number of quarters during the entire evaluation period. Similarly, the number of false alarms is determined by counting all signals issued prior to the five-year period ahead of a crisis and all signals issued after the end of a crisis. Based on this, the noise ratio (false positive rate) is calculated as the number of false alarms as a share of the total number of quarters outside the evaluation and crisis periods.

2.2.4.2. Dynamic evaluation procedure

Chart 4 illustrates the dynamic evaluation procedure. This procedure is the same as the static procedure described above with the difference that, when classifying observations into the confusion matrix based on a specific cut-off point (indicated by the red dashed line), this is done separately for each single quarter within the evaluation period [-20,-4] (the "evaluation horizon"). For example, for the evaluation horizon L=-7 (indicated by the dashed green line) the LHS variable is set to 1 only for observations seven quarters prior to a crisis, while it is set to 0 for all other observations. All other horizons within the green outline (except for L=-7 in the given example) are excluded from the analysis.

Chart 4: Illustration of the design of the “dynamic” procedure



Notes: Data are for the United Kingdom’s credit-to-GDP gap as defined by the BCBS (2010b).

To assess the appropriate timing of an indicator (ps)AUROCs are computed for a range of different evaluation horizons. This is done using the approach employed (for AUROCs) in Drehmann and Juselius (2014)[7]. The dynamic approach differs from the static procedure in terms of how the signal ratio is computed; it is calculated for each single evaluation horizon h within the evaluation horizon (i.e.: h=20,19,18,...,6,5,4). For example, at horizon L=-16 (i.e. 4 years ahead of crises), a signal is only recorded as a true positive if it is issued 16 quarters ahead of the beginning of a crisis. The noise ratio, on the other hand, is determined as in the static evaluation procedure.

Dynamic evaluation has implications for statistical inference. In particular, the significance of the results tends to be lower (i.e. the standard deviation tends to be higher) compared with the static evaluation approach. This reflects the fact that for the dynamic evaluation approach an AUROC and the associated standard deviation are calculated for each of the 17 lags separately. As explained earlier, the crisis dummy takes a value of 1 only for the horizon assessed. Therefore the number of crisis periods which enter the equation for the standard error (see Appendix B) is 17 times lower than the number of observations for the static evaluation approach. This implies that the standard deviation calculated for the dynamic evaluation approach would be about four times larger ($\sqrt{17}$=4.12) than that for the static evaluation approach. The width of the two standard deviation error bands in the charts shown in Appendix D is thus approximately eight times larger than the standard deviations shown in the tables in Appendix C.

[7] Special thanks to M. Drehmann for his support in comparing the Expert Group’s code with his code for the dynamic analysis.

2.2.5. Criteria for indicator selection

To discriminate between indicators, the Expert Group placed most weight on the static evaluation procedure, which allows for a country-by-country evaluation of indicators. This reflects the importance the Expert Group assigned to selecting indicators that provide meaningful signals to a large number of EU Member States. Reflecting the large number of indicators evaluated, a first set of criteria was chosen in order to eliminate indicators that did not perform well. The remaining indicators were then considered more closely, which included an assessment of their stability in the dynamic evaluation procedure. This approach, which is described in more detail below, is necessarily judgemental.

Indicators were eliminated if they failed any of the following criteria.

Overall performance: Any indicator with an AUROC that was smaller than 0.6 has not been recommended. This reflects the judgement that – given the fact that the toss of a fair coin would have a value of 0.5 – for an indicator to be considered as performing well, it would need to be somewhat better than a coin toss.

True positive and false positive rates: Any indicator or model with a true positive rate of less than 0.5 and/or with a false positive rate of more than 0.5 for balanced preferences (θ=0.5) has not been recommended. Indicators that did not provide a signal four to 20 quarters prior to a crisis more than half of the times and/or provided a signal when no crisis followed in the four to 20 quarters thereafter more than half of the times have been deemed inappropriate for policy purposes.

Performance across member states: Any indicator where the AUROC was insignificant for more than one third of relevant EU Member States has also not been recommended. The cut-off level of one third reflects the aim of identifying indicators that prove useful across the European Union. Indicators eliminated by this threshold may nevertheless provide useful signals in individual countries.

Usefulness: Any indicator with a usefulness measure of less than 0.1 for balanced policy preferences has not been recommended. A model is considered useful if the loss arising from the policy-makers' loss function is smaller than the loss obtained when disregarding the model. Alessi and Detken (2011) show that a policy-maker can always realise a loss of min (θ, 1-θ), i.e. 0.5, based on the balanced preferences assumed as a baseline in this paper. A cut-off of 0.1 implies that using the model improves usefulness by 20%

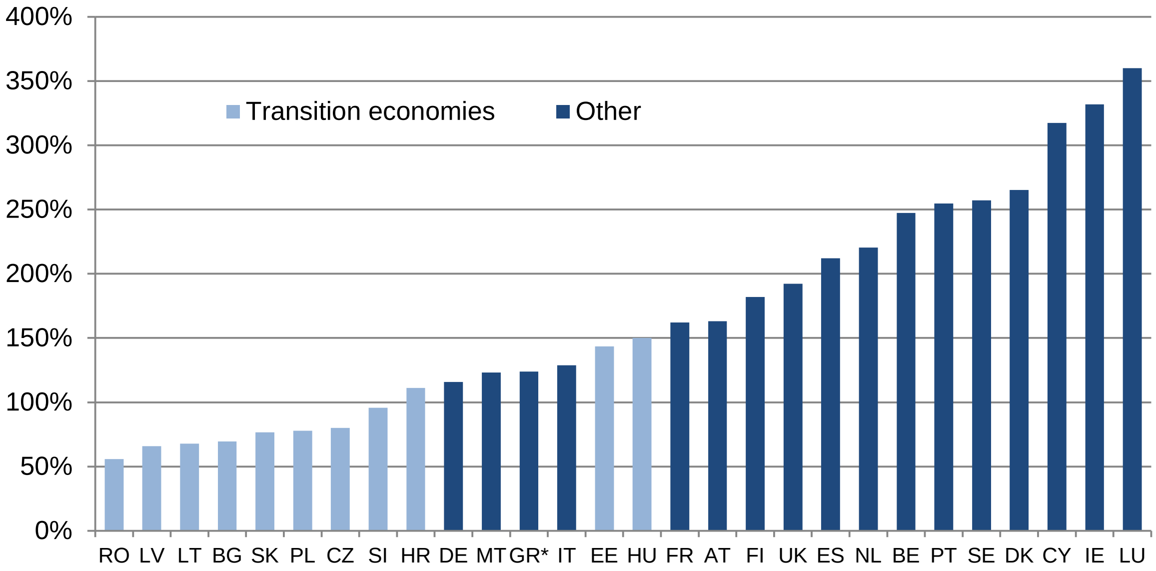
Stability: Indicators have not been recommended if – based on the dynamic evaluation procedure – they exhibited two successive quarters with an AUROC not significantly above 0.5 within the prediction horizon of 20 to 4 quarters ahead of a crisis. When an indicator breaches a threshold and signals a crisis, policy-makers are unable to tell whether this signal indicates, for example, a crisis in 20 or in four quarters. Without this information, indicators that display good signalling properties at many horizons, but perform poorly over a prolonged window within the evaluation horizon, are limited in their usefulness in a univariate analysis. This is not to say that such indicators cannot add useful information in a multivariate analysis.

Section 3: Evaluation results for the build-up phase

3.1. The credit-to-GDP gap

This subsection focuses on the credit-to-GDP gap. It builds on the BCBS's guidance for calculating credit-to-GDP gaps, with two important extensions. First, whereas the BCBS approach is based on the 27 BCBS member countries – including seven EU Member States – this paper in principle covers all 28 EU Member States, as long as data are available. This is an important difference, as the European Union includes several transition economies which, given their distinct economic history, tend to have a relatively short back-run of data (Tables A2) and relatively low credit-to-GDP ratios (Chart 5). Second, the paper analyses a broad range of alternative specifications of the credit-to-GDP gap. This is to provide a robustness check and to explore whether – for the European Union as a whole – alternative calculations perform better than the measure advocated by the BCBS.

Chart 5: Credit-to-GDP ratios, Q4 2012



* observation for GR is for Q1 2011

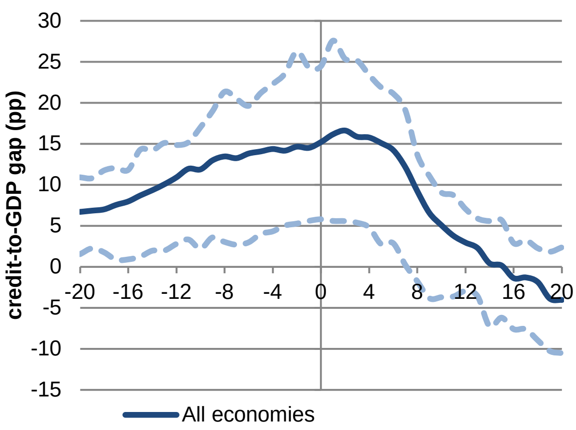
3.1.1. BCBS benchmark gap

In the BCBS's guidance, the credit-to-GDP gap (BCBS benchmark gap) is calculated as follows:

- The ratio of nominal broad credit to nominal GDP is calculated for each quarter, where GDP is annualised by taking the sum of the four most recent quarterly observations.
- The long-term trend is calculated with a one-sided (or recursive) Hodrick-Prescott (HP) filter, where the smoothing parameter lambda (λ) is set at 400,000.
- The credit-to-GDP gap is the difference between the credit-to-GDP ratio and its long-term trend, resulting in a gap in percentage points (pp).

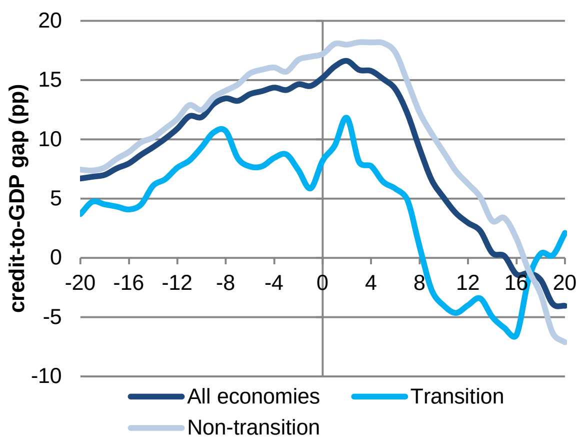
Visual inspection suggests that the BCBS benchmark gap is likely to perform well in signalling crises for the European Union as a whole. This is illustrated in Chart 6, which shows that the credit-to-GDP gap tends to be relatively high (i.e. above the 2 pp threshold suggested by the BCBS), even 20 quarters ahead of financial crises. Its wide distribution, however, also indicates that the performance may differ significantly across countries. For example, Chart 7 shows a markedly different level and pattern of average gaps for transition and non-transition economies, with the former peaking about two years prior to crises.

Chart 6: Average gap and ranges for EU-28



Notes: The solid line represents the average credit-to-GDP gap (percentage points) from 20 quarters prior to a crisis to 20 quarters after the start of a crisis. Averages are based on all crisis episodes in the set of countries considered. The dashed lines represent the 25th and 75th percentiles.

Chart 7: Average gaps by country grouping



Note: See notes for Chart 6

Nevertheless, the BCBS benchmark is unlikely to be suitable for each individual EU Member State. This is illustrated in Annex E, which shows persistent differences between the ratio of credit to GDP and its long-term trend for a number of countries. Technically, the usage of a high smoothing parameter as advocated by the BCBS (λ=400,000) results in a persistent positive credit-to-GDP gap for countries experiencing a period of falling and subsequently rising credit-to-GDP ratios. For instance, using this measure, Ireland would have had a continuously positive credit-to-GDP gap for more than 20 years. Economically, this illustrates the limits of statistical tools in differentiating between cyclical developments (e.g. a credit boom) and structural developments (e.g. ongoing financial deepening). Therefore, a holistic approach is likely to be needed, which should consider alternative statistical specifications, as well as the application of judgement. Future work should also try to derive normative benchmarks that have a stronger structural link with the economy in question, since pure statistical approaches will always be sample dependent.

3.1.2. Alternative specifications

Three main alternatives for calculating credit-to-GDP ratios and related credit-to-GDP gaps are explored in this paper. These are: (1) alternative ways for estimating the trend – different smoothing parameters, shorter time series and adding forecasts; (2) an alternative definition of the credit-to-GDP gap (i.e. the ratio instead of the absolute difference); and (3) alternative credit aggregates.

A number of other alternatives were considered but not pursued in this paper. First, other filter techniques besides the HP filter (e.g. the band-pass filter or moving averages) could have been used.[8] However, the HP filter has a number of practical advantages: it is widely used, included in many statistical packages and is therefore easy to implement by the authorities tasked with operating the CCB. Second, two-sided filters could have been used. While using all the information currently available to run a two-sided filter might provide "better" historical estimates, such an approach does not work for the evaluation of the signalling performance of indicators with respect to past crises, as no information about the future would have been available prior to a crisis. Finally, the impact of real-time "vintage" data for the calculation of each observation of the credit-to-GDP gap was not pursued. While Edge and Meisenzahl (2011) show with US data that the gap is prone to a considerable ex post revision when new data become available, van Norden (2011) and Drehmann et al. (2011) show that this does not have consequences for the performance of the credit-to-GDP gap as an indicator. Moreover, such vintage data are not available for many countries over a sufficiently long period.

3.1.2.1. Alternative trend estimates

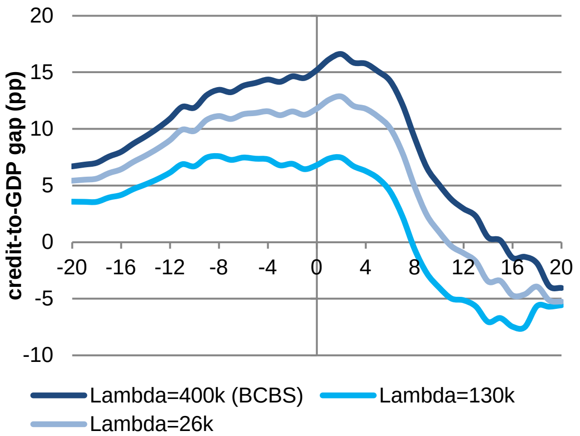
Statistical trend estimates are subject to a high degree of uncertainty. This is particularly the case for the initial years when only a few observations were available, especially if the series is far away from its "equilibrium value" at the start of the sample. There may thus be a case for evaluating the performance of the credit-to-GDP gap only from a later stage in the sample, when the trend has "settled down". Such an approach has to be weighed against not being able to use all the information contained in a longer time series. Unless otherwise stated, evaluations shown in this paper are conducted using the calculation of the credit-to-GDP gap from Q1 1975 onwards (or, for countries where data is not available from the start of the 1970s, from 5 years after the first data is available). Another source of uncertainty is the possibility of structural breaks, which can be partly addressed by choosing a more flexible trend with a lower smoothing parameter lambda.

Three alternative trend specifications are explored:

1. *Different smoothing parameters*. A smoothing parameter of λ=400,000 hinges on the assumption that the financial cycle tends to be four times as long as the business cycle (Drehmann et al., 2010) and implies a relatively inflexible trend. The review of the financial cycle by Borio (2012) suggests that the business cycle involves frequencies from one to eight years, whereas the average length of the financial cycle in his sample of seven industrialised countries since the 1960s has been around 16 years. Therefore, we use smoothing parameters that reflect different cycle lengths. Specifically, estimates for λ=26,000 and λ=130,000 which imply that the financial cycle is, respectively, two and three times as long as the business cycle were also tested. Chart 8 plots the resulting credit-to-GDP gaps against the BCBS benchmark gap. The trend in the credit-to-GDP ratio was also determined by separately calculating the one-sided trends of the numerator (credit) with a smoothing parameter of λ = 400,000 and the denominator (GDP) with a smoothing parameter of λ = 1600 (a value commonly used for quarterly GDP data). Chart 9 plots the resulting credit-to-GDP gaps against the BCBS benchmark gap.

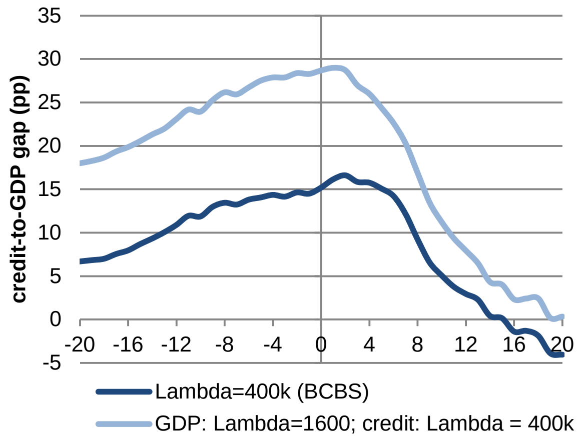
[8] Harvey and Jaeger (1993) and Cogley and Nason (1995) discuss some shortcomings of HP filters.

Chart 8: Average gaps for different *lambda*



Note: See notes for Chart 6

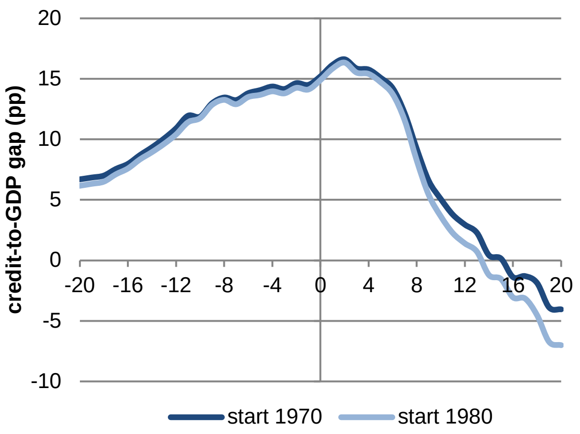
Chart 9: Average gaps for different *trends*



Note: See notes for Chart 6

2. *Shorter time series*. A way to deal with structural breaks is to estimate the trend for a shorter sample period. Given the fact that the transition economies already have short time series, the focus is on removing periods related to financial liberalisation in the non-transition economies, by conducting sensitivity analysis using only data from Q1 1980 onwards. Reflecting the shorter time series, the credit-to-GDP gap has been evaluated from Q1 1980, despite the aforementioned concerns about the stability of trend at the start of the period. Chart 10 shows that the average credit-to-GDP gaps for sample periods starting in Q4 1970 and Q1 1980 are very similar.

Chart 10: Average gap for different *periods*



Notes: See notes for Chart 6

3. *Using forecasts*. Another alternative is to augment credit and GDP series with model forecasts and use these in the calculations of the credit-to-GDP gap for each observation. This may help to stabilise the trend and reduce the end-point problem and has been found to improve the signalling performance in the case of Norway (Gerdrup, Kvinlog, and Schaanning, 2013).

3.1.2.2. An alternative definition of the gap

The gap could be computed as the ratio between credit/GDP and its trend, rather than the absolute difference between the two. The BCBS defines the credit-to-GDP gap as the *difference* between the actual credit-to-GDP ratio and its trend. This leads to a gap measured in *percentage points*. Alternatively, the gap could be based on the *ratio* between the actual series and the trend, leading to a gap measured as a *percentage*.

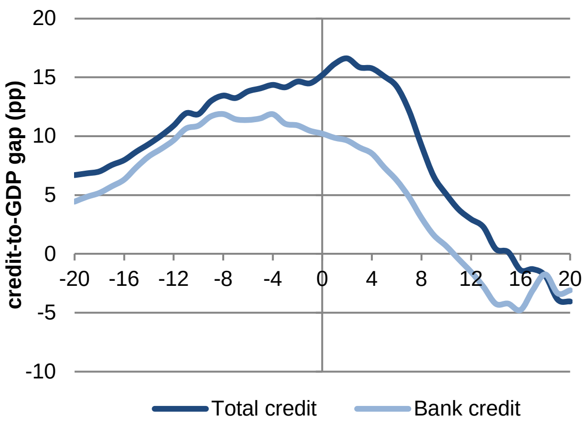
This difference may be important when thresholds for activating the CCB are selected. Combining a percentage point gap with thresholds that are fixed at specific absolute levels implies that the benchmark buffer rates are not independent of the level of the credit-to-GDP ratio. Specifically, countries with a lower credit-to-GDP ratio can experience a faster increase in credit as a percentage of credit outstanding before the CCB would be activated. In contrast, combining a gap measured as a percentage with thresholds that are fixed at specific absolute levels means that the benchmark buffer rates are independent of the level of the ratio of credit to GDP.

Economic arguments can be made for both approaches. The resilience of the banking system to a given increase in credit may be independent of the level of credit relative to GDP. But in a system where bank credit plays a smaller role, the real economy may be more resilient to banking sector strains, and fast credit growth in such systems may be less of an overall concern than fast credit growth in systems that rely more heavily on banks. To better understand the empirical relevance of these differences, both variants are considered.

3.1.2.3. Alternative credit aggregates

Instead of a broad credit aggregate as proposed by the BCBS, narrower aggregates may be considered. The advantage of a broad aggregate is that it captures all credit and related risks that are building up in the economy and, hence, is not very sensitive to disintermediation of the banking sector. Moreover, as this is the measure advocated by the BCBS, it is consistent internationally. On the other hand, for some countries the availability of data on credit granted by banks is better than that of broader credit aggregates that capture credit granted by the non-bank sector. Moreover, broad credit aggregates can include elements – such as intra-company loans – that are not considered relevant for the purposes of the CCB. In addition to bank credit, sectoral aggregates – borrowing by households and non-financial corporations – are considered as well. Although the CCB has not been designed to address sectoral risk, such aggregates may help the designated authorities to better understand developments. Chart 11 shows the differences between average credit-to-GDP gaps based on broad credit and those based on bank credit.

Chart 11: Average gaps for different *credit definitions*



Notes: See notes for Chart 6

3.1.3. Results for the credit-to-GDP gap

The static evaluation procedure described in Section 2.2.4.1 shows that signalling qualities are stable across a range of gap specifications. Table 2 summarises the results. Overall, the credit-to-GDP gaps perform well, with AUROCs of around 0.8 (psAUROCs around 0.9). Moreover, the credit-to-GDP gaps generally perform well according to the other performance criteria (true/false positive rates, performance across member states, usefulness and stability).

Evaluation of the alternative specifications reveals some commonality. First, results for credit-to-GDP gaps based on low smoothing parameters (not shown) generally tend to be worse than those based on high smoothing parameters. Second, results for credit-to-GDP gaps based on shorter time series (not reported) tend to be worse than those based on the full sample. Third, for the European Union as a whole, neither the use of forecasts nor the calculation of the credit-to-GDP gaps as a ratio between credit/GDP and its trend significantly improve the signalling performance relative to the BCBS benchmark gap. Finally, specifications based on bank credit or on total credit to the household sector tend to perform better than the BCBS benchmark gap, but not in statistically significant terms.

Table 2 Results of the static evaluation procedure

Credit gaps	1	2	3	4	5	6	7	8	9	10	11
Indicator	**AUROC**	**sd(AUROC)**	**ps(AUROC)**	**Optimal threshold**	**TPR**	**FPR**	**Usefulness**	**Crisis quarters within evaluation**	**Number of evaluated quarters**	**Percentage of MS for which indicator has no significant AUROC**	**Countries for which AUROC has been calculated**
Bank credit-to-GDP gap, rw60	0.81	0.012	0.95	1.93	0.90	0.42	0.24	267	1003	0.20	5
Credit to households-to-GDP gap, nrw	0.80	0.012	0.86	2.78	0.67	0.20	0.23	345	1207	0.08	13
Credit to households-to-GDP gap, rw60	0.79	0.012	0.90	27.40	0.71	0.27	0.22	229	631	0.00	5
Credit-to-GDP ratio minus ratio of trends	0.79	0.012	0.86	12.67	0.72	0.24	0.24	431	2164	0.11	9
Total credit-to-GDP gap, moving average forecast**	0.79	0.012	0.91	4.65	0.83	0.38	0.23	369	1699	0.14	7
Credit to households-to-GDP gap, rw60	0.79	0.012	0.98	0.07	0.95	0.47	0.24	213	607	0.20	5
Total credit-to-GDP gap, linear forecast**	0.79	0.012	0.90	3.95	0.80	0.33	0.24	369	1699	0.14	7
Total credit-to-GDP gap, nrw (Basel gap)	**0.79**	**0.012**	**0.90**	**2.70**	**0.86**	**0.39**	**0.24**	**362**	**1652**	**0.07**	**15**
Differenced relative banking credit*	0.79	0.012	0.90	0.10	0.79	0.34	0.23	408	2176	0.00	17
Total credit gap multiplied by credit-to-GDP level	0.77	0.012	0.86	382.4	0.67	0.23	0.22	431	2164	0.11	19
Bank credit-to-GDP relative gap, rw60	0.77	0.012	0.93	3.97	0.81	0.36	0.22	267	1003	0.20	5
Bank credit-to-GDP gap, nrw	0.77	0.012	0.88	2.49	0.79	0.36	0.21	377	1761	0.00	14
Credit to households-to-GDP relative gap, nrw	0.76	0.012	0.87	11.78	0.78	0.37	0.21	362	1297	0.07	14
Total credit-to-GDP gap, rw60	0.76	0.012	0.88	3.02	0.81	0.41	0.20	280	983	0.00	5
Credit to households-to-GDP relative gap, nrw	0.75	0.012	0.87	4.83	0.77	0.34	0.22	342	1204	0.08	13
Total credit-to-GDP relative gap, nrw	0.74	0.013	0.91	2.54	0.84	0.41	0.21	359	1649	0.07	15
Bank credit-to-GDP relative gap, nrw	0.74	0.013	0.94	8.27	0.89	0.50	0.20	394	2120	0.07	15
Total credit relative gap, nrw	0.74	0.013	0.98	7.15	0.95	0.61	0.17	382	1976	0.13	16
Credit to households-to-GDP relative gap, rw60	0.74	0.013	0.99	-0.59	0.97	0.50	0.24	213	607	0.20	5
Total credit-to-GDP relative gap, r60	0.74	0.013	0.88	3.34	0.76	0.37	0.19	280	983	0.00	5
Differenced relative total credit*	0.74	0.013	0.85	0.16	0.67	0.31	0.18	425	2087	0.06	17
Bank credit relative gap, rw60	0.73	0.013	0.98	5.23	0.90	0.54	0.18	304	1252	0.00	6
Total credit gap multiplied by log of credit-to-GDP level	0.73	0.013	0.81	141.1	0.72	0.33	0.20	383	1400	0.06	17
Bank credit-to-GDP relative gap, nrw	0.73	0.013	0.89	3.65	0.78	0.39	0.20	377	1761	0.00	14
Total credit relative gap, rw60	0.70	0.013	0.81	11.05	0.64	0.35	0.15	317	1230	0.00	6
Credit to NFC relative gap, nrw	0.69	0.013	0.97	1.70	0.94	0.63	0.15	362	1274	0.21	14

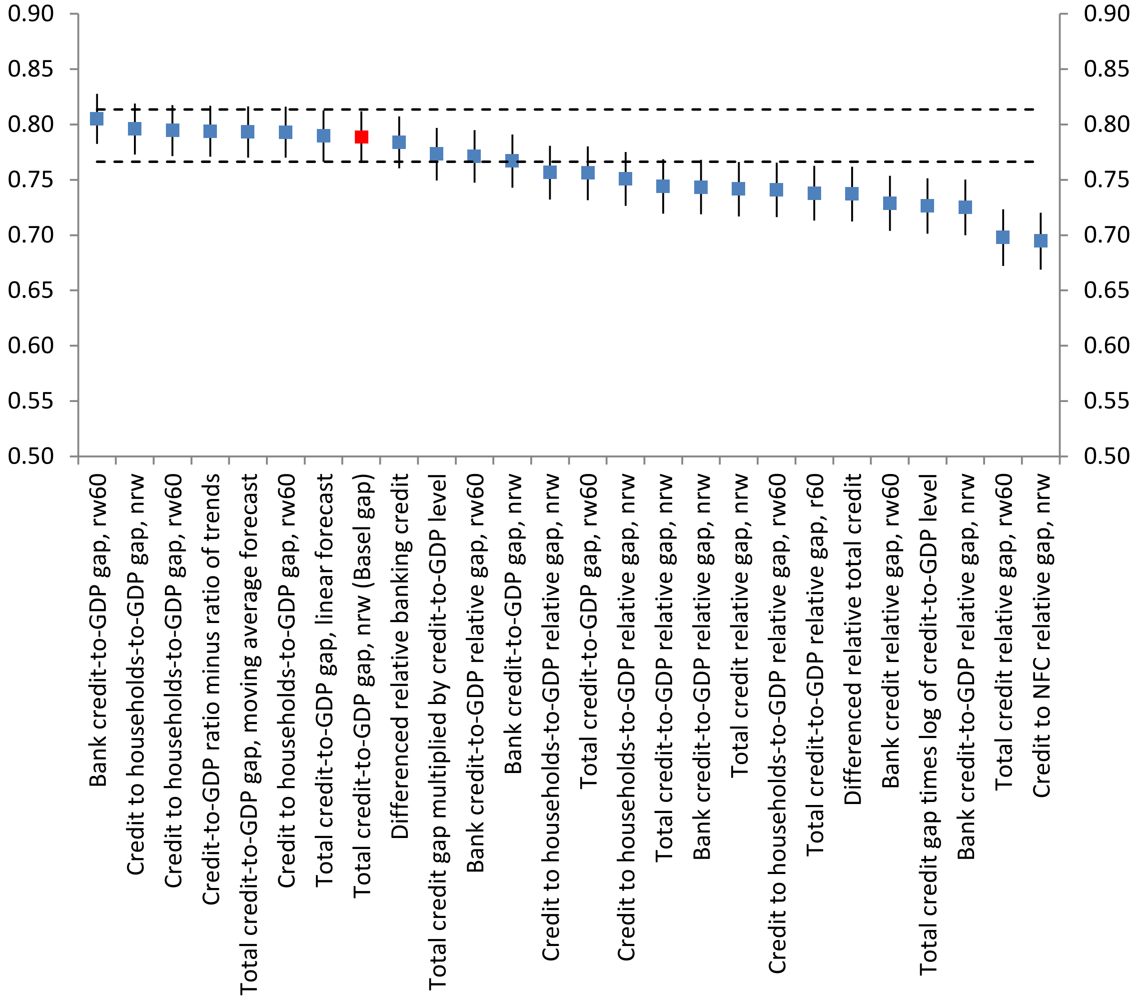
Notes: The first and second columns show the AUROC and its standard deviation, respectively. An entry is marked red if at least one of the performance indicators mentioned in Section 2.2.5 is violated. The third column shows the partial standardised AUROC (psAUROC); see Annex B for further details. The fourth column shows the optimal threshold (i.e. the one which minimises the relevant loss function for balanced preferences between type-I and type-II errors. The fifth and sixth columns show, respectively, the true positive rate and the false positive rate for balanced preferences. The seventh column shows a usefulness measure based on Alessi and Detken (2011). The model is considered useful, if its loss is smaller than the loss obtained when disregarding it. The usefulness (U) is defined as U=min (θ;1-θ)-L, where L is the loss (see Figure 1). The eighth, ninth and eleventh columns provide information about the sample size. The tenth column shows the percentage of Member States for which the AUROC at the optimal pooled threshold is insignificant using only the crisis data for the country in question. The row in bold font is the credit-to-GDP gap according to the BCBS (2010b). For further details on the indicators, see the notes for Tables C1 to C3 in Annex C.

* $\frac{5 \cdot L_t}{\sum_{i=0}^{4} GDP_{t-i}} - \frac{5 \cdot L_{t-4}}{\sum_{i=4}^{7} GDP_{t-i}}$, where L_t stands for credit in period t (Karlo Kauko, 2012b);

** Variables described in Section 3.1.2.1 (alternative trend specifications, third point in the numbered list, p. 24)

The relative performance of the BCBS benchmark is also illustrated by Chart 12. This graph shows AUROCs as well as associated confidence intervals. None of the gap measures significantly outperforms the BCBS benchmark in statistical terms.

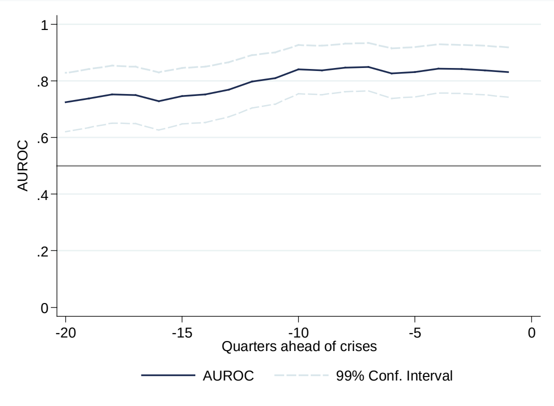
Chart 12 **AUROCs and two standard deviation intervals – static approach**



Notes: The blue squares display AUROCs for different calculations of the credit-to-GDP gap. The bars represent two standard deviations above and below this measure, which can be interpreted as a 95% confidence interval. The red square is the AUROC according to the BCBS's calculation of the credit-to-GDP gap. For further details on the indicators see the notes for Tables C1 to C3 in Annex C.

The dynamic evaluation procedure leads to similar conclusions. Chart 13 presents the AUROC of the BCBS benchmark credit-to-GDP gap based on pooled EU data for each evaluation horizon up to 20 quarters ahead of a crisis. The measure is significantly higher than 0.5 over the entire period considered and gradually increases to about 0.8. It thus gives a consistent warning signal long before the start of financial crises. This pattern is similar to the one found by Drehmann and Juselius (2014), who calculate dynamic AUROCs for the BCBS benchmark gap based on data from EU countries, as well as data from non-EU countries.

Chart 13: BCBS benchmark gap based on pooled EU data – dynamic approach



In addition to the credit-to-GDP gaps displaying good performance for the European Union as a whole, most specifications are informative for the majority of individual countries as well. When issuing its recommendation, the ESRB is mandated to "… take into account the differences between Member States…" (Art 135(2) CRD IV). However, a statistical evaluation of indicators at the individual country level is complicated by practical problems such as insufficient observations to calculate (ps)AUROCs. Some countries did not report any crises for the sample period, which makes any evaluation based on statistical methods impossible. Insofar as the performance of indicators for individual EU Member States can be assessed, the tenth column in Table 2 shows that most of the best performing indicators are statistically significant for all the Member States for which the indicators could be evaluated. Table 3 offers further details on the performance of the BCBS's measure of the credit-to-GDP gap for all EU Member States. The calculation of (ps)AUROCs is based on the pooled optimal threshold of 2.7, obtained with a preference parameter of 0.5. It shows that the gap performs well (i.e. it has an AUROC above 0.6, delivers a true positive rate higher than 0.5 and a false positive rate lower than 0.5 and a usefulness measure of no less than 0.10) for 11 out of 16 Member States for which it could be evaluated. For two of these 11 countries (Spain and Portugal) however, the gap has been found to be particularly persistent (defined as exhibiting at least one positive gap for at least ten consecutive years).

Table 3: Performance of the BCBS's measure by EU Member State

	No crisis	Short time series	Persistent gap[1]	AUROC	psAUROC	Results based on pooled sample optimal threshold				Number of evaluated quarters	Crisis quarters within evaluation
						Optimal threshold	TPR	FPR	Usefulness		
pooled sample				0.79	0.90	2.7	0.86	0.39	0.24	1652	362
BE	x		x					0.84		86	0
BG	x[2]	x					1.00	1.00		4	2
CZ	x[2]	x						0.55		22	0
DK				0.72	0.72		0.56	0.18	0.19	94	34
DE				0.77	0.78		0.47	0.04	0.21	107	17
EE	x[2]	x								0	0
IE			x	0.82	1.00		1.00	0.68	0.16	85	17
GR		x					1.00	0.00		3	3
ES			x	0.90	0.90		0.85	0.47	0.19	94	20
FR				0.89	1.00		1.00	0.14	0.43	112	34
HR	x[2]	x						0.60		10	0
IT			x	0.95	1.00		1.00	0.57	0.21	115	17
CY				0.96	0.97		1.00	0.32	0.34	48	17
LV		x		1.00	1.00		1.00	0.22	0.39	26	17
LT		x		1.00	1.00		1.00	0.56	0.22	26	17
LU	x		x							0	0
HU		x		0.96	0.96		0.88	0.00	0.44	25	17
MT	x		x					0.78		118	0
NL				0.69	0.97		0.45	0.20	0.12	114	31
AT	x							0.26		126	0
PL	x	x						0.04		26	0
PT			x	0.75	1.00		1.00	0.34	0.33	90	34
RO	x[2]	x						1.00		6	0
SI		x								0	0
SK	x	x								0	0
FI				0.86	1.00		0.82	0.20	0.31	105	17
SE				0.89	1.00		0.94	0.26	0.34	108	34
UK				0.91	1.00		1.00	0.29	0.35	102	34

Notes: (1) Persistence is defined as a gap that lasts for at least ten successive years. Excludes the countries reported in the second column. (2) An actual or would-be crisis was reported, but no credit-to-GDP gap can be calculated for the period before the crisis.
The cells are highlighted in red if TPR<0.5 or FPR>0.5. The country name is highlighted in grey if the AUROC for the given country cannot be calculated. The FPR for these countries have been calculated, but are in a number of cases based on few observations

3.2. Other right-hand-side variables

3.2.1. Results for the univariate case

The Expert Group analysed a number of other indicators that could complement the credit-to-GDP gap. These indicators fall into five broad categories: other credit-related variables, property variables, variables related to the real economy, market-based variables and variables referring to bank balance sheets. This analysis brings together the static and dynamic approaches described above. The results broadly confirm common findings in the literature. In particular, credit-related variables and some property-related variables perform well in predicting crises (detailed results of the static and dynamic approaches are shown in Tables C1 to C3 and D1 to D3 in Annex C and Annex D, respectively).

Other credit-related variables: As financial imbalances are often driven by an unsustainable economic expansion reflected in the rapid growth of credit and asset prices, credit-related variables are key predictors of banking crises (Borio and Lowe, 2002). These show a picture consistent with the year-on-year growth rate, displaying AUROCs of around 0.7 (and psAUROCs approaching 0.9) and thus perform well. Since quarter-on-quarter growth rates tend to be more volatile and generally perform worse than year-on-year growth rates, these results are not reported. The dynamic analysis shows that credit growth variables tend to provide stable results, although their signalling performance declines close to the start of financial crises (see Figure D2 in Annex D).

Debt service ratio: When households and firms are highly indebted, large income shortfalls are likely to trigger bankruptcies. As a consequence, the repayment of loans becomes more challenging leading to an increase in banks' losses. This is one of the variables that performs particularly well in Drehmann and Juselius (2012, 2014). For the European Union this indicator performs less well, with only the household measure passing the evaluation criteria set out above. This might reflect particularities of the two datasets; relative to the data for the countries covered in Drehmann and Juselius (2014) the dataset for the European Union tends to span a shorter time period. In addition, the composition of the two datasets considerably differs: 11 out of 26 countries considered in Drehmann and Juselius (2014) are non-EU countries. The fact that the debt service ratio for non-financial firms is insignificant may be related to the fact that it only accounts for bank loans. Corporations' access to other sources of finance results in funding diversification, but it also exposes them to other vulnerabilities. It may also be an insignificant measure because business sector debt service ratios, at least in some countries, tend to be more closely linked to the business cycle (Drehmann and Juselius, 2012) than to the financial cycle. The dynamic analysis shows that debt service ratios are only significant towards the end of the forecast horizon, i.e. from about 6 quarters before the crisis (see Figure D5 in Annex D).

Property variables: Falling house prices will prompt banks to require more collateral from borrowers, which increases their repayment burden. Hence, substantial house price increases can be considered as indicators of future house price corrections and thus predictors of banking crises if the borrowers are highly indebted and/or deteriorating macroeconomic conditions adversely affect their ability to repay their loans. These exhibit AUROCs of around 0.6-0.7 and thus perform well. In particular, measures that signal a potential overvaluation of house prices (e.g. the house price-to-income ratio, the house price-to-rent ratio and a house price gap) display AUROCs of around 0.7. A number of these variables, however, fail the evaluation criterion of a true positive rate of at least 0.5.

As is the case in Drehmann and Juselius (2014), the dynamic analysis shows that the signalling performance of property variables tends to be less stable than that of other variables. In particular, the signalling performance of property variables tends to drop close to the start of crises and – in some cases – the AUROC even reverses sign, revealing that at some point closer to the start of the crisis it is negative house price developments which signal the forthcoming crisis. These results are not surprising, as it is well-known that the rapid unwinding of imbalances in property markets often precedes banking crises (see Behn et al., 2013).

Macroeconomic variables: Macroeconomic conditions affect the ability of households and firms to repay their loans and hence banks' asset quality and profitability. For example, the money supply might provide additional information because changes in net external assets and marketable instruments, which might affect asset prices, are reflected in money but not in credit (Adalid and Detken, 2007). While excessive credit growth often precedes banking crises, foreign debt (the flow of which is proxied by the current account deficit) could be most problematic as foreign lenders may have a disadvantage in identifying risks compared with an informed domestic creditor (Kauko, 2012a). Therefore, foreign sources of financing might also be more volatile and more susceptible to herding behaviour. Several macroeconomic variables (e.g. GDP growth, CPI growth, unemployment and the ratio of government debt to GDP) generally perform poorly. With AUROCs below 0.6 most fail the evaluation criterion. While psAUROCs are sometimes high, these cases are often misleading due to high false positive rates as also shown by the low usefulness measures. This is not necessarily surprising, given the higher frequency of the economic cycle relative to the financial cycle. In contrast, the inverted ratio of the current account to GDP performs relatively well (AUROC of 0.62) suggesting that a current account deficit can help signal crises. Similarly, the gap between real M3 and its long-term trend shows an AUROC of close to 0.7 and is significant for the large majority of EU Member States. The dynamic analysis shows that those variables that perform poorly in the static analysis also tend to display a less stable signalling performance over time.

Market-based variables: Financial market indicators can complement quantity-based measures, as they reflect market participants' expectations about the future state of the economy and may hence be more forward looking, even though their lead time is expected to be limited to a few months (Dufrénot et al., 2012 and Arsov et al., 2013). Most market-based variables fail the evaluation criterion of an AUROC of at least 0.6. Although psAUROCs tend to be high, these results often have high false positive rates and a low usefulness measure for balanced preferences. A couple of exceptions are the inverted measures of short-term and long-term interest rates, which have AUROCs approaching 0.7. While low nominal interest rate levels seem to contain some information relevant for crisis prediction, the thresholds derived from the current analysis are likely to be very backward looking on account of a trend for the downward adjustment of equilibrium interest rates. Again, the dynamic analysis shows that the variables that perform poorly in the static analysis also tend to display less stable signalling performance over time.

Balance sheet variables: Given that early warning models aim to predict banking crises, developments in bank balance sheets related to asset quality, liquidity, solvency and profitability are likely to signal upcoming problems. Owing to the short length of time series, few balance sheet variables were tested. One variable considered was the leverage ratio. This ratio has been computed from banks' published accounts, using data on equity and reserves and total assets (see Bush et al., forthcoming, for details). It shows an AUROC of below 0.5, suggesting that on its own it is an uninformative leading indicator of banking crises. Bush et al. (forthcoming) and Behn et al. (2013), however, find that a leverage ratio does help to predict banking crises in a multivariate cross-country panel data setting.

3.2.2. Results for the multivariate case

This subsection summarises the results of the three multivariate approaches explored by the Expert Group. They are the discrete choice (logit) model, the bivariate signalling approach and the decision tree approach described in Section 2.1.

3.2.2.1. Logit

The analysis presented is based on a subset of variables that was selected based on economic reasoning and data availability. This choice reflects a common challenge with multivariate models: while using all possible combinations of early warning indicators for which data is available becomes unfeasible even for a relatively small number of variables, a stepwise variable selection does not evaluate all variable combinations and hence might not identify a particular variable combination with a high crisis prediction power. To mitigate this issue, a subset of variables was selected based on economic reasoning and data availability. For this selection all possible combinations were evaluated.[9] The chosen subset of variables listed in Table 4 contains 15 indicators broadly covering most of the data categories outlined in Section 1.

Table 4: Indicator variables used in the multivariate logit models

Variable name	Category	Obs.	Countries	Stationarity
Total credit-to-GDP gap, no rolling window, (Basel gap)	Credit	1870	25	No
Bank credit-to-GDP gap, no rolling window	Credit	1966	26	No
Year-on-year growth rate of real total credit	Credit	2690	28	Yes
Year-on-year growth rate of real bank credit	Credit	2809	27	Yes
Year-on-year growth rate real commercial property prices	Housing	745	20	Yes
Year-on-year growth rate of real residential property prices	Housing	1897	21	Yes
Annual absolute change in house price-to-income ratio	Housing	1739	21	Yes
Annual absolute change in house price-to-rent ratio	Housing	1726	18	Yes
Debt service ratio	Housing	2410	27	No
Year-on-year growth rate of equity prices	Market	2376	27	Yes
Real three-month money market interest rate	Market	2517	27	Yes
Year-on-year growth rate of real GDP	Real/macro	2359	28	Yes
Current account to GDP ratio	Real/macro	2152	28	Yes
Public debt to GDP ratio	Real/macro	1550	28	No
Year-on-year growth rate of real M3	Real/macro	2179	27	Yes

Note: This table shows the indicator variables used in the multivariate logit regressions. The third and fourth columns report the number of quarterly observations and countries available, respectively, for a given variable. The fifth column reports whether a Fisher-type test (Choi, 2001) rejected the null hypothesis of a unit root in all panels at the 10% significance level. For further details on the indicators see the notes for Tables C1 to C3 in Annex C.

[9] Bayesian model averaging (BMA) is another way to address the issue and has been used in the context of early warning models by Babecky et al. (2011). For a comprehensive introduction to BMA see Hoeting et al. (1999).

The analysis has largely been based on an unbalanced panel. It shows that not all variables are available for all countries and over the same time span. In particular housing variables are available only for the more recent years. Against this background, the analysis could have been carried out for different subsets of countries with more homogeneous data coverage. Since the number of crisis episodes is rather limited, the Expert Group instead opted to use as many countries as possible in the analysis. Limited data availability for some variables means that the use of a balanced panel in the estimation would have reduced the sample in terms of variables, countries and time. This would have focused the analysis on the more recent crisis episodes, in particular on the global financial crisis and could have created a sample selection bias. To mitigate such bias, an unbalanced panel has been used for the initial analysis. In order to assess whether the best multivariate models have a better crisis prediction power than the best univariate model, the different models have subsequently been compared using a common sample. The persistency of some of the explanatory variables used in the logit model reduces the robustness of the results. The last column in Table 4 reports whether a Fisher-type test (Choi, 2001) rejected the null hypothesis of a unit root in all panels at the 10% significance level. Four out of 15 variables exhibit strong persistency. In particular, credit-to-GDP gap variables show a higher persistency than credit growth variables. According to Park and Phillips (2000), high persistency can be problematic for binary choice regression-based models and can affect inference. For robustness, multivariate models could be estimated using growth rates instead of credit-to-GDP gaps for credit-related variables and appropriate transformations of other variables with a high persistency.

The estimations of all possible combinations of the 15 indicator variables listed in Table 4 are carried out using a pooled logit regression. All variables, except the annual equity price growth and three-month money market rates, enter the models with a one quarter lag to account for publication lags. In addition, all left-hand-side observations that were not classified as a pre-crisis period after the last quarter of 2007 are dropped from the analysis, because it is not possible to determine the true value of these pre-crisis observations until at least five years have passed. It is important to note that, depending on data availability, each of the approximately 32,700 estimated models could be based on a different sample. This limits the comparability of the results. To assess how often each variable has been included in a model, only those variables with a significant coefficient are considered. In addition to the share of models in which a variable was included, Table 5 also reports the average AUROC of all models in which the respective variable was included, the average number of explanatory variables and the average coefficient estimate across these models.

Table 5: Model inclusion of indicator variables – unbalanced panel

Variable	Included in % of models	Average AUROC	Average of RHS variables	Average coefficient estimate
Year-on-year growth rate of equity prices	99.7%	0.84	5.9	0.02
Bank credit-to-GDP gap, nrw	97.5%	0.86	5.6	0.13
Annual absolute change in house price-to-income ratio	92.4%	0.85	6.1	0.15
Year-on-year growth rate of real GDP	91.2%	0.85	6.2	* -0.31
Debt service ratio	87.2%	0.85	6.0	5.97
Public debt to GDP ratio	86.1%	0.85	5.8	* -0.02
Current account-to-GDP ratio	83.4%	0.85	6.1	-0.12
Year-on-year growth rate real commercial property prices	80.6%	0.84	6.0	0.05
Year-on-year growth rate of real bank credit	74.2%	0.84	6.0	0.11
Total credit-to-GDP gap, nrw, (Basel gap)	64.8%	0.85	5.9	0.03
Year-on-year growth rate of real total credit	57.5%	0.84	6.2	* -0.06
Year-on-year growth rate of real residential property prices	53.6%	0.84	6.5	* -0.08
Year-on-year growth rate of real M3	48.3%	0.83	5.8	0.06
Real three-month money market interest rate	40.6%	0.84	6.4	* 0.06
Annual absolute change of house price-to-rent ratio	30.0%	0.84	5.9	0.03

Note: The second column reports the share of models in which a given variable was included. Only coefficients significant at the 5% level were considered based on HAC standard errors. For robustness, the Bonferroni adjustment was used to account for a potential alpha-inflation. The results, which are available upon request, are qualitatively similar. Due to the more conservative significance level, the average model with the Bonferroni adjustment contains only two variables. The third and fourth columns report the average AUROC and the average number of right-hand-side variables, respectively, of all models in which a given variable was included. The fifth column reports the average coefficient estimate of all models in which a given variable was included. Average coefficients marked with an asterisk have a sign which differs from expectations. For further details on the abbreviations and terms used see the notes for Tables C1 to C3 in Annex C.

The bank credit-to-GDP gap, equity price growth, the debt service ratio, the change in the house price-to-income ratio and the current account-to-GDP ratio are among those variables most often included in the estimated models. Even though equity price growth and the debt service ratio – when evaluated in the univariate case – rank lower in terms of their crisis prediction power, they appear to improve the prediction accuracy when included in a model together with other variables, in particular credit-related variables. The results also reveal that the estimated coefficients of some variables have signs which go against economic intuition or prior findings. One potential reason could be that the pattern of a variable may change during the five-year evaluation window. House prices, for example, while exhibiting a strong increase up to two to three years prior to the onset of a crisis started to decrease afterwards. A shorter sample size and more complex interaction effects among different variables may also result in unexpected coefficient signs.

In order to assess the extent to which multivariate models can improve the crisis prediction power compared with the best univariate model, robustness tests have been carried out using balanced panels. The results shown in Annex F are qualitatively similar to the previous ones and hence suggest that multivariate models can improve crisis prediction power compared with the best univariate model based on the bank credit-to-GDP gap. However, since some variables cover a shorter time span the results depend on the specific sample being used.

3.2.2.2. Multivariate signalling

The work on multivariate signalling follows the approach by Borio and Lowe (2002), Alessi and Detken (2011) and Drehmann and Juselius (2014). It is also consistent with the approach described in a paper by the Swiss National Bank (2013). In a bivariate signalling approach a crisis signal is only issued when both indicators that are included in the respective analysis exceed their individual optimal threshold at the same time.[10] The optimal signalling threshold for each of the indicators is determined jointly based on a two-dimensional grid search, where the highest usefulness for a given preference parameter of the policy-maker's loss function determines the level of the thresholds.

For the benchmark results, a balanced preference between missing crises and issuing false alarms was assumed. The analysis was performed on the pooled dataset described earlier, assuming a common signalling threshold for all countries.

Within this general set-up, 37 of the best performing univariate signalling indicators from different variable categories were selected. This set of variables included various credit-based indicators, house price-related variables, debt service ratios, macroeconomic variables and interest rates. In total, almost 500 different bivariate signalling models were estimated by combining the best 17 univariate indicators with each other and with the remaining variables of the pre-selected variable list.

Overview of the main results

Table 6 summarises the results of the bivariate analysis for the top 20 models according to the AUROCs. For comparison, the results for the best univariate indicator of each indicator pair are displayed as well.[11] The main message from the results is that the bivariate signalling approach could in principle improve on univariate analysis when the AUROC or usefulness serves as the evaluation criterion. However, the results need to be interpreted with care, given that data availability differs considerably across countries and variables and therefore also across the models.[12] The best bivariate model in terms of the AUROC attained a value of 0.89 and combines the household credit-to-GDP gap and the annual growth rate of commercial property prices; whereas the best univariate indicator is the bank credit-to-GDP gap with an AUROC of 0.81. In terms of usefulness, the corresponding numbers for the best bivariate and univariate signalling model are 0.63 and 0.48 respectively. The results suggest that the best bivariate signalling model is closer to a perfect model of crisis prediction for the policy-maker by 15% compared with the best univariate model, although no formal significance test was conducted. In terms of the shares of correctly signalled crises and absent false alarms, the best bivariate model achieved values of 83% and 19% respectively.

[10] In principle a multivariate analysis with more than two indicators could be conducted. However, this is computationally quite demanding as it requires higher dimensional grid searches and the testing of a large number of indicator combinations. It therefore remains something to consider in future work.

[11] For the bivariate signalling approach the grid search was conducted with a percentile grid of about 2500 grid points in total. For consistency, the univariate signalling analysis was repeated with a percentile grid as well. This could lead to some differences in the thresholds as compared with the linear grid.

[12] See the discussion below for further details.

Table 6: Results for the top 20 bivariate signalling models

			AUROC		Thresholds				True postive rate (TPR)		False positive rate (FPR)		Loss		Relative usefulness		Best uni-variate
					Indicator 1		Indicator 2										
Indicator 1	Indicator 2	Corr-elation	bivariate	best univ.	univariate	bivariate	bivariate	univariate	bivariate	best univ.	bivariate	best univ.	bivariate	best univ.	bivariate	best univ.	indicator
Total credit to HH/GDP gap, nrw	Comm. real estate price inflation (yoy)	0.15	0.89	0.80	2.77	3.58	-0.78	-0.84	0.83	0.67	0.19	0.20	0.18	0.27	0.63	0.47	1
Res. property price/income	Comm. real estate price inflation (yoy)	0.02	0.89	0.73	95.51	96.23	-0.84	-0.84	0.87	0.79	0.25	0.40	0.19	0.31	0.62	0.39	1
Total credit to HH/GDP gap, nrw	Real commercial property price gap, nrw	0.19	0.89	0.80	2.77	3.58	-8.25	8.83	0.82	0.67	0.20	0.20	0.19	0.27	0.62	0.47	1
Bank credit/GDP gap, r60	Debt service/income	0.07	0.88	0.81	1.93	1.93	0.14	0.18	0.80	0.90	0.20	0.41	0.20	0.26	0.60	0.48	1
Res. property price/income	Real M3 gap, nrw	0.04	0.86	0.73	95.51	95.27	1.42	2.21	0.82	0.79	0.24	0.40	0.21	0.31	0.58	0.39	1
Bank credit/GDP gap, r60	Total credit to HH/GDP	0.15	0.86	0.81	1.93	2.46	17.47	55.56	0.86	0.90	0.30	0.41	0.22	0.26	0.55	0.48	1
Total credit to HH/GDP gap, nrw	Res. property price/income gap, nrw	0.40	0.86	0.80	2.77	1.83	-9.90	13.33	0.84	0.67	0.31	0.20	0.24	0.27	0.53	0.47	1
Res. property price/income	Gov debt/GDP	-0.22	0.86	0.73	95.51	96.27	6.91	6.94	0.91	0.79	0.34	0.40	0.21	0.31	0.57	0.39	1
Total credit to HH/GDP gap, nrw	Res. property price/income	0.47	0.85	0.80	2.77	0.45	95.27	95.51	0.78	0.67	0.20	0.20	0.21	0.27	0.58	0.47	1
Total credit to HH/GDP gap, nrw	Res. property price/rent	0.45	0.85	0.80	2.77	1.77	83.48	92.25	0.78	0.67	0.21	0.20	0.21	0.27	0.57	0.47	1
Bank credit/GDP gap, r60	Res. property price/income	0.38	0.85	0.81	1.93	0.05	95.27	95.51	0.82	0.90	0.26	0.41	0.22	0.26	0.56	0.48	1
Bank credit/GDP gap, r60	Total credit to HH/GDP gap, nrw	0.75	0.85	0.81	1.93	1.85	-1.30	2.77	0.88	0.90	0.33	0.41	0.22	0.26	0.56	0.48	1
Total credit to HH/GDP gap, nrw	Debt service/income	0.08	0.85	0.80	2.77	1.94	0.01	0.18	0.80	0.67	0.28	0.20	0.24	0.27	0.52	0.47	1
Bank credit/GDP gap, r60	Debt service/income HH	0.18	0.85	0.81	1.93	-5.60	0.12	0.11	0.77	0.90	0.15	0.41	0.19	0.26	0.62	0.48	1
Res. property price/income	Real commercial property price gap, nrw	-0.05	0.85	0.73	95.51	103.85	-1.42	8.83	0.69	0.79	0.20	0.40	0.26	0.31	0.49	0.39	1
Total credit to HH/GDP gap, nrw	Res. property price/rent gap, nrw	0.48	0.85	0.80	2.77	1.77	-7.10	13.81	0.78	0.67	0.25	0.20	0.23	0.27	0.53	0.47	1
Bank credit/GDP gap, nrw	Gov debt/GDP	-0.26	0.84	0.77	2.50	2.76	6.91	6.94	0.82	0.79	0.28	0.36	0.23	0.29	0.54	0.43	1
Bank credit/GDP gap, r60	Bank credit/GDP	0.21	0.84	0.81	1.93	1.93	76.15	72.48	0.67	0.90	0.14	0.41	0.23	0.26	0.53	0.48	1
Debt service/income HH	Res. property price/rent	0.36	0.84	0.71	0.11	0.12	80.27	92.25	0.69	0.79	0.12	0.45	0.22	0.33	0.56	0.34	2
Total credit to HH/GDP gap, nrw	Real M3 gap, nrw	0.45	0.84	0.80	2.77	1.77	-5.72	2.21	0.81	0.67	0.30	0.20	0.25	0.27	0.51	0.47	1

Notes: Results for balanced preferences θ = 0.5. The results for the univariate analysis in this table are slightly different from those in Section 3.2.1 in some cases because the bivariate analysis was conducted with a grid based on percentiles. In order to ensure consistency between the two approaches the univariate analysis was repeated with a percentile grid with 15000 grid points. For computational reasons the grid size for the bivariate models is limited to 751 times 751 points. The ROC-curve for the bivariate case is computed by taking the maximum true positive rate for each false positive rate. For further details on the indicators see the notes for Tables C1 to C3 in Annex C.

Overall, the results suggest that the bivariate signalling approach yields a lower rate of type-II errors, and in some cases even a lower rate of type-I and type-II errors at the same time. In particular, among the top 20 bivariate models, 15 have a lower rate of false alarms than the best univariate indicator of each bivariate indicator pair. When it comes to the share of correctly signalled crises, 12 of the best bivariate models attain a higher value than the best univariate indicator of each bivariate indicator pair. For seven of the bivariate models, both the type-I and type-II error rates are lower than for the best univariate indicator of each corresponding bivariate indicator pair.

The top 20 bivariate models combine indicators related to various credit aggregates, residential and commercial property prices, and debt service ratios. These results are in line with the findings in Borio and Drehman (2009) and Drehmann and Juselius (2014), for instance. In terms of specific indicator combinations that provide good in-sample signalling properties the best five pairs are:

1. Total household credit-to-GDP gap; Growth rate of commercial property prices.
2. Residential property prices to income; Growth rate of commercial property prices.
3. Total household credit-to-GDP gap; Real commercial property price gap.
4. Bank credit-to-GDP gap, r60[13]; Debt service-to-income ratio
5. Residential property prices to income; Real M3 gap

Generally, the best models feature indicators that are almost uncorrelated, underlining that the greatest gains in a bivariate signalling approach should be expected from combinations of indicators that contain complementary information about vulnerabilities. This additional information could, for instance, be derived from a different measurement of the same underlying economic concept or from indicators that capture a different economic concept.

It is important to note that the optimal (in-sample) signalling thresholds can change considerably and even change sign in some cases between the bivariate and univariate signalling models. Intuitively, one would expect that the optimal thresholds in a bivariate signalling model should be lower than in a univariate model because for given thresholds it should be less likely that both indicators exceed their thresholds at the same time compared with only one of the indicators. The results displayed in Table 6 confirm this intuition; for the best 20 bivariate models at least one indicator features a lower threshold compared with the respective univariate case. More specifically, 16 of the top 20 bivariate models have a lower threshold for the second indicator, while 10 have a lower threshold for the first indicator.

It is also crucial to note that in some bivariate models the combination of optimal thresholds appears counter-intuitive. For example, the thresholds in the second best bivariate model imply that a crisis signal is issued whenever the ratio of residential property prices to income is higher than 96.23% of its historical country-specific mean and at the same time annual growth of commercial property prices is higher than -0.84%. In other words, a crisis signal could be issued by this model in cases where the ratio of residential property prices to income is below the historical average and commercial property prices are declining. There are other examples in the table, where at least one

[13] For an explanation of this abbreviation see the notes for Tables C1 to C3 in Annex C.

of the indicator thresholds appears counter-intuitive, e.g. when a credit-to-GDP gap has a negative signalling threshold. These results highlight that the bivariate signalling approach, while in principle useful, should not be applied mechanically and that judgement, as well as a robustness analysis, should always be employed in order to develop models that are useful for policy purposes. Notably, for some indicators, the chosen signalling window of 20 to 4 quarters before the crisis might not be the most appropriate choice. This becomes relevant, in particular, if an indicator normally features a turning point during the above window.

Finally, a potential caveat that needs to be kept in mind when interpreting the results is that data availability differs considerably between the variables employed in the analysis. This implies that the in-sample performance measures reported in Table 6 are not necessarily comparable across models, because some models are estimated on different samples of crisis occurrences. For example, commercial property prices are mostly only available from the end of the 1990s, whereas many other data series are available from the 1970s. Hence, any bivariate model containing commercial property prices is based on a shorter common sample. However, different sample sizes across indicators are also used in the univariate signalling approach.

Robustness with respect to policy preferences and sample size

As with the univariate case, the ways in which the performance of the models and the optimal thresholds change when different policy preferences are assumed were explored. Table 7 shows how the results are affected when policy-makers have a higher preference for correctly predicting crisis, namely when they assign a weight of 60% or 70% to type-I errors in their loss function. In line with what intuition would suggest and the results obtained from the univariate analysis, true positive rates and false positive rates increase for all of the top 20 models when the preference parameter is increased from 0.5 to 0.7. In addition, lower thresholds are estimated in many cases. Specifically, for seven of the models both thresholds are lower when there is a weight of 70% for type-I errors and in all models at least one of the thresholds is lower in this scenario. The loss and usefulness measures are not directly comparable across Tables 6 and 7, because the relevant benchmark losses without a model are different on account of the different policy preferences.[14] The results also show that the magnitude and in some cases also the sign of the indicator thresholds are sensitive to the policy preferences. This highlights the fact that optimal signalling thresholds always need to be considered within the specific policy context.

[14] The benchmark loss without a model is equal to the lower value of either θ, the preference parameter, or 1-θ.

Table 7: Bivariate signalling results for different policy preferences

Indicator 1	Indicator 2	AUROC	Thresholds						TPR			FPR			Loss			Relative usefulness		
			Indicator 1			Indicator 2														
			0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7
Total credit to HH/GDP gap, nrw	Comm. real estate price inflation (yoy)	0.89	3.58	3.28	2.85	-0.78	-2.37	-14.15	0.83	0.86	0.92	0.19	0.23	0.34	0.18	0.18	0.16	0.63	0.56	0.47
Res. property price/income	Comm. real estate price inflation (yoy)	0.89	96.23	96.23	95.53	-0.84	-11.57	-11.57	0.87	0.95	0.98	0.25	0.37	0.41	0.19	0.18	0.14	0.62	0.56	0.54
Total credit to HH/GDP gap, nrw	Real commercial property price gap, nrw	0.89	3.58	2.85	1.07	-8.25	-8.22	-7.97	0.82	0.86	0.93	0.20	0.25	0.38	0.19	0.19	0.17	0.62	0.53	0.45
Bank credit/GDP gap, r60	Debt service/income	0.88	1.93	1.93	1.93	0.14	0.11	0.11	0.80	0.89	0.89	0.20	0.31	0.31	0.20	0.19	0.17	0.60	0.53	0.44
Res. property price/income	Real M3 gap, nrw	0.86	95.27	95.27	92.40	1.42	-0.05	-4.52	0.82	0.85	0.95	0.24	0.28	0.45	0.21	0.20	0.17	0.58	0.50	0.43
Bank credit/GDP gap, r60	Total credit to HH/GDP	0.86	2.46	1.93	1.93	17.47	17.47	17.47	0.86	0.90	0.90	0.30	0.35	0.35	0.22	0.20	0.18	0.55	0.50	0.41
Total credit to HH/GDP gap, nrw	Res. property price/income gap, nrw	0.86	1.83	0.30	0.30	-9.90	-13.25	-13.25	0.84	0.95	0.95	0.31	0.44	0.44	0.24	0.20	0.16	0.53	0.49	0.45
Res. property price/income	Gov debt/GDP	0.86	96.27	95.55	95.55	6.91	6.91	6.91	0.91	0.93	0.93	0.34	0.36	0.36	0.21	0.19	0.16	0.57	0.53	0.47
Total credit to HH/GDP gap, nrw	Res. property price/income	0.85	0.45	0.30	0.25	95.27	93.43	85.53	0.78	0.83	0.88	0.20	0.26	0.36	0.21	0.21	0.20	0.58	0.49	0.35
Total credit to HH/GDP gap, nrw	Res. property price/rent	0.85	1.77	0.30	0.25	83.48	84.18	83.48	0.78	0.88	0.89	0.21	0.33	0.35	0.21	0.20	0.18	0.57	0.50	0.41
Bank credit/GDP gap, r60	Res. property price/income	0.85	0.05	0.10	0.10	95.27	93.96	76.65	0.82	0.84	0.93	0.26	0.29	0.45	0.22	0.21	0.18	0.56	0.48	0.39
Bank credit/GDP gap, r60	Total credit to HH/GDP gap, nrw	0.85	1.85	1.85	-5.99	-1.30	-1.30	-0.03	0.88	0.88	0.96	0.33	0.33	0.46	0.22	0.20	0.17	0.56	0.50	0.44
Total credit to HH/GDP gap, nrw	Debt service/income	0.85	1.94	1.32	0.30	0.01	0.01	0.04	0.80	0.85	0.94	0.28	0.35	0.49	0.24	0.23	0.19	0.52	0.43	0.37
Bank credit/GDP gap, r60	Debt service/income HH	0.85	-5.60	-5.72	-3.20	0.12	0.11	0.08	0.77	0.78	0.95	0.15	0.16	0.47	0.19	0.20	0.18	0.62	0.51	0.41
Res. property price/income	Real commercial property price gap, nrw	0.85	103.85	95.85	95.85	-1.42	-17.61	-17.61	0.69	0.97	0.97	0.20	0.52	0.52	0.26	0.22	0.17	0.49	0.44	0.42
Total credit to HH/GDP gap, nrw	Res. property price/rent gap, nrw	0.85	1.77	0.47	-0.23	-7.10	-7.01	-7.01	0.78	0.88	0.93	0.25	0.36	0.45	0.23	0.21	0.19	0.53	0.47	0.38
Bank credit/GDP gap, nrw	Gov debt/GDP	0.84	2.76	1.98	-0.84	6.91	6.91	6.91	0.82	0.85	0.98	0.28	0.32	0.55	0.23	0.22	0.18	0.54	0.45	0.41
Bank credit/GDP gap, r60	Bank credit/GDP	0.84	1.93	1.93	1.93	76.15	49.24	49.24	0.67	0.90	0.90	0.14	0.39	0.39	0.23	0.22	0.19	0.53	0.45	0.36
Debt service/income HH	Res. property price/rent	0.84	0.12	0.11	0.08	80.27	80.27	84.42	0.69	0.69	0.87	0.12	0.13	0.42	0.22	0.24	0.22	0.56	0.41	0.28
Total credit to HH/GDP gap, nrw	Real M3 gap, nrw	0.84	1.77	0.30	0.25	-5.72	-4.45	-5.18	0.81	0.93	0.94	0.30	0.48	0.50	0.25	0.23	0.19	0.51	0.42	0.37

Notes: The columns with the headers “0.5”, “0.6” and “0.7” show the results for the respective policy preference parameter θ. For further details on the indicators see the notes for Tables C1 to C3 in Annex C. TPR = true positive rate, FPR = false positive rate.

As mentioned above, the comparison between bivariate and univariate results in Table 10 does not account for potentially different sample sizes between the two approaches. In a robustness exercise the univariate signalling models were therefore re-estimated with the same sample size as in the associated bivariate case. In general, the findings of this exercise confirm the results found with heterogeneous sample sizes. First, in most cases type-II errors and optimal signalling thresholds are lower for the bivariate models. Second, the AUROC and usefulness performance measures are generally higher for the bivariate models. However, with the adjusted sample sizes the differences to the univariate models are smaller in most cases and are potentially not statistically significant as the shorter sample sizes lead to higher univariate performance measures.[15] One explanation for this could be that the smaller samples give more weight to the global financial crisis. Therefore, indicators associated with credit and residential house price developments display an improved univariate signalling performance owing to the strong dynamism in these market segments in a number of EU Member States during the last decade.[16]

3.2.2.3. Decision trees

Alessi and Detken (2014) show that a decision model based on a random forest with 1,000 trees tends to misclassify periods in less than 10% of cases and yields an AUROC of above 0.9. The associated classification tree, assuming balanced preferences between type-1 and type-2 errors, yields a true positive rate of 87% and a false positive rate of 9%. The usefulness measure is 39%. The classification tree in Figure 3 gives an example of how regulators could use this tool to inform CCB decisions in real time. Figure 3 is derived by using the 15 most robust indicators from the random forest exercise. The tree shows that different indicators become relevant depending on whether bank credit exceeds 93% of GDP.

If bank credit exceeds 93% of GDP, a crisis is likely to occur within the next 4 to 20 quarters if the following three conditions are met: i) the bank credit-to-GDP gap is above -0.8, ii) the debt service ratio is above 0.14, and iii) household credit is above 56% of GDP. This branch is indicated by the red ellipse.

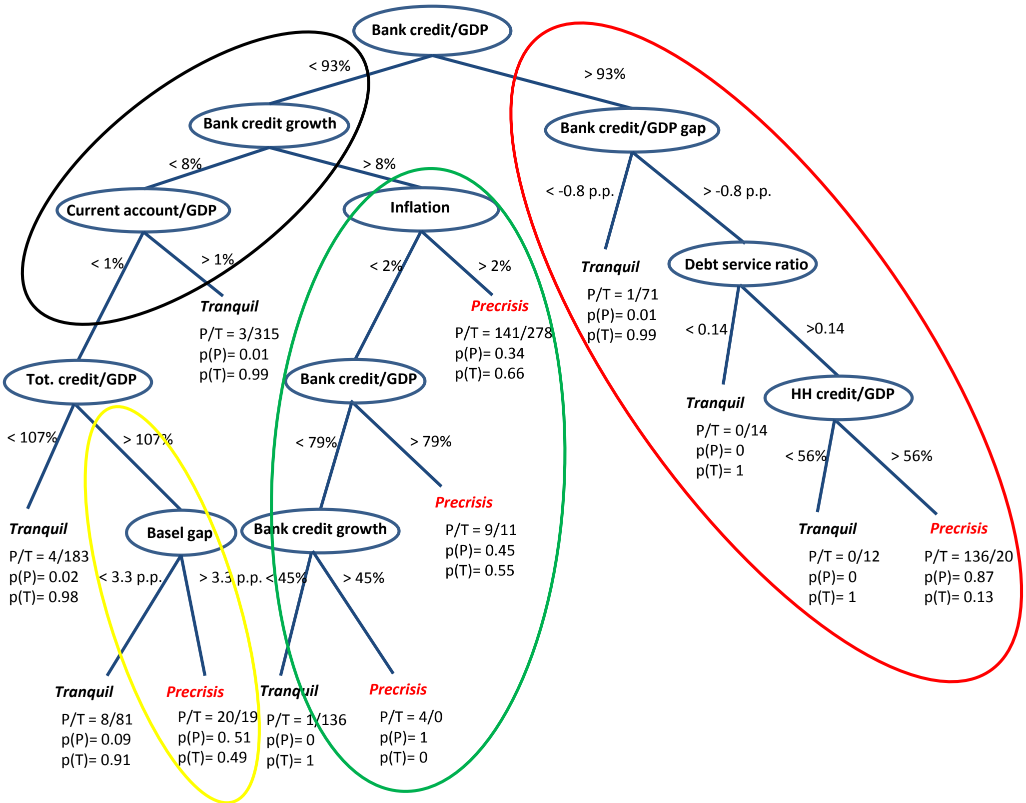
If bank credit does not exceed 93% of GDP, bank credit growth becomes the relevant indicator. If bank credit growth breaches the threshold of 8% year-on-year, the probability of a crisis happening is between 34% and 100% if one of the following three conditions is verified: i) inflation is above 2%, or ii) bank credit is above 79% of GDP, or iii) bank credit growth is above 45% year-on-year. This branch is indicated by the green ellipse. If bank credit growth does not breach the threshold of 8% and the country is running a current account surplus of more than 1% of GDP, the probability of crisis is virtually zero (black ellipse). If the country is running a current account deficit or only a very small surplus, it becomes important to consider whether it is structurally vulnerable due to the degree of credit intermediation by the banking sector. In this respect, the identified threshold for total credit is 107% of GDP. If this is breached and at the same time there are signs of credit overheating (i.e. the

[15] The statistical significance of the differences in AUROCs between univariate and bivariate signalling models was not formally tested.

[16] For example, the ratio of residential property prices to nominal income improves its AUROC from 0.72 to 0.84 and the AUROC measure for the household credit-to-GDP gap increases from 0.78 to 0.88 when the sample period is shortened to the availability of data for commercial property prices.

BCBS benchmark gap is above 3.3 percentage points), the probability of a crisis materialising is slightly above 50%. This branch is indicated by the yellow ellipse.

Figure 3: An example of a binary classification tree



Notes: In each terminal node (leaf) of the tree, P/T indicates the number of pre-crisis quarters over the number of non-pre-crisis (tranquil) quarters ending up in that particular leaf, considering the historical data on which the tree has been grown. For forecasting purposes, p(P) is the derived probability that a crisis will materialise if the relevant indicators breach their respective thresholds, while p(T) is the probability that no crisis will materialise over the relevant prediction horizon.

Section 4: Evaluation results for the release phase

This section presents the findings for cases when the countercyclical capital buffer could be reduced or fully released. Following the BCBS's guidance, Jahn (2014) considers two scenarios: a prompt release in times of financial stress and a release when system-wide financial stability risks recede. The analysis is based on the dataset described in Section 1 (see Table 1 and Table A2 in Annex A) for all 28 EU Member States. However, data limitations, particularly concerning market-based indicators, considerably reduce the country coverage and the lengths of the time series that can be used in the empirical analysis.

4.1. A prompt release

This subsection considers the case of a prompt release of the countercyclical capital buffer in times of financial stress. Three types of right-hand-side variable that could potentially provide useful signals guiding the CCB release are considered: (1) high-frequency market-based indicators, as they can reflect rapidly weakening financing conditions or act as near-coincident indicators of financial stress; (2) indicators that are found to be relevant at a short-term prediction horizon for the build-up phase of the buffer, as they could also be informative for the buffer release phase. This includes flow-based measures such as growth in credit and asset prices; and (3) indicators reflecting banking sector losses or banks' asset quality, such as banks' non-performing loans.

The signalling approach used in the analysis corresponds to the static procedure described in Section 2.2.4.1. Reflecting the fact that the release phase requires a near-coincident rather than a leading indicator of financial stress, the evaluation is based on a short-term prediction horizon of [-2, 0] quarters rather than the longer evaluation horizon used for the build-up phase. All signals issued over the indicated period are evaluated against the occurrence of a banking crisis to assess the signalling abilities of the candidate indicators. Market-based indicators, such as CDS premia or interest rates, as well as banking sector losses, tend to increase ahead of a banking crisis and are evaluated against an increase ahead of a banking crisis. In contrast, real-economy variables, credit-related variables and equity prices tend to fall ahead of a banking crisis and are evaluated against a decrease ahead of a banking crisis.

Market-based indicators display the highest AUROCs. Table 8 shows that, with an AUROC of 0.85, the LIBOR-OIS spread performs best. This partly reflects the fact that the spread had already peaked in the third quarter of 2007 and remained at an elevated level until the second quarter of 2009, while in most EU Member States the banking crises started in the third quarter of 2008. Other market-based indicators such as covered bond spreads and the ECB's CISS indicator show AUROCs in excess of 0.8. Average bank CDS premia and equity price growth display AUROCs above 0.65, but these fail the criterion of sufficient performance across Member States.

All other variables fail at least one of the evaluation criteria. Table 8 shows that real-economy variables, GDP growth and commercial property price growth provide some useful signals for the pooled sample, with AUROCs of 0.68 and 0.75, respectively. But these variables perform poorly for individual Member States. This is partly driven by the fact that on a country-by-country basis the short time series availability (especially for financial market indicators) considerably reduces the number of crisis observations. Credit-related variables and the indicator for bank asset quality were not found to provide useful signals. In particular, non-performing loans (NPL) over total gross loans start to increase two quarters after the onset of a crisis, and although credit growth is among the best performing indicators for the build-up phase of the CCB, it displays little signalling quality for the release phase. This is likely to reflect the fact that the use of credit lines can result in credit growth declining only slowly ahead of banking crises.

The short time series availability for many financial market series is an important caveat that affects the robustness of the results. In particular, many of the findings are driven by the global financial crisis. This is particularly true for some of the best performing market-based indicators, namely the LIBOR-OIS spread and the ECB CISS indicator for which the data only starts in 1999 (see Table A2 and Table A3 in Annex A). This means that judgement may need to play an even greater role for the prompt release of the countercyclical capital buffer than for the build-up phase.

Table 8: Evaluation of variables for the release phase

Indicator	AUROC	SD(AUROC)	psAUROC	Optimal Threshold	True Positive Rate	False Positive Rate	Percentage of MS for which indicator has no sig AUROC	Coefficient of variation for national usefulness measure
Real-economy variables								
GDP growth	0.68	0.19	0.69	1.19	0.61	0.32	100	1.16
Comm. Price Growth	0.75	0.30	0.70	0.07	0.64	0.22	92	0.81
Other credit-related variables								
Credit growth	0.52	0.10	0.56	2.20	0.43	0.34	100	2.16
Bank balance sheet								
NPL/Total Gross Loans	0.44	0.04	0.44	1.90	0.06	0.03	73	8.56
Market based variables								
Equity price growth	0.69	0.18	0.67	-0.29	0.64	0.35	100	0.99
Libor-OIS	0.85	0.03	0.89	30.49	0.77	0.07	13	0.48
Av. Bank CDS premia	0.66	0.07	0.90	64.22	0.83	0.04	33	0.49
Sovereign CDS premia	0.32	0.04	.	15.17	0.87	0.36	60	0.80
ML covered bond spreads	0.82	0.07	0.88	47.18	0.67	0.13	0	0.70
ECB CISS indicator	0.83	0.04	0.87	0.30	0.77	0.08	13	0.48

Note: The first and second columns show the AUROC and its standard deviation. An entry is marked red if the AUROC is below 0.6. The third column shows the psAUROC. The forth column shows the optimal threshold signalling a crisis for balanced preferences between type-I and type-II errors: for real-economy and other credit related variables as well as equity price growth (quarter on quarter growth rates) in percent; bank balance sheet variable in percent; for remaining market based indicators in basis points; for CISS unit free and (0,1]. Preference parameter is at 0.5. The fifth and sixth columns show the true positive rate and the false positive rate for balanced preferences. Entries have been highlighted in red if the true positive rate is smaller than 0.5 or the false positive rate is greater than 0.5. The seventh column shows the percentage of member states for which the AUROC at the optimal pooled threshold is insignificant using only the crisis data for the country in question. Entries have been highlighted red if the percentage is larger than 25. Last column shows the cross-country coefficient of variation of a usefulness measure.

A publication lag of 1 quarter is included for the real-economy, other credit-related variables, equity price growth and the bank balance sheet variable. Bank non-performing loans (NPL) over total gross loans are available on an annual basis only and have been interpolated by spline function. Average bank CDS premia refer to a simple average of the CDS premia of the five largest banks in terms of market capitalisation by country.

4.2. Gradual release

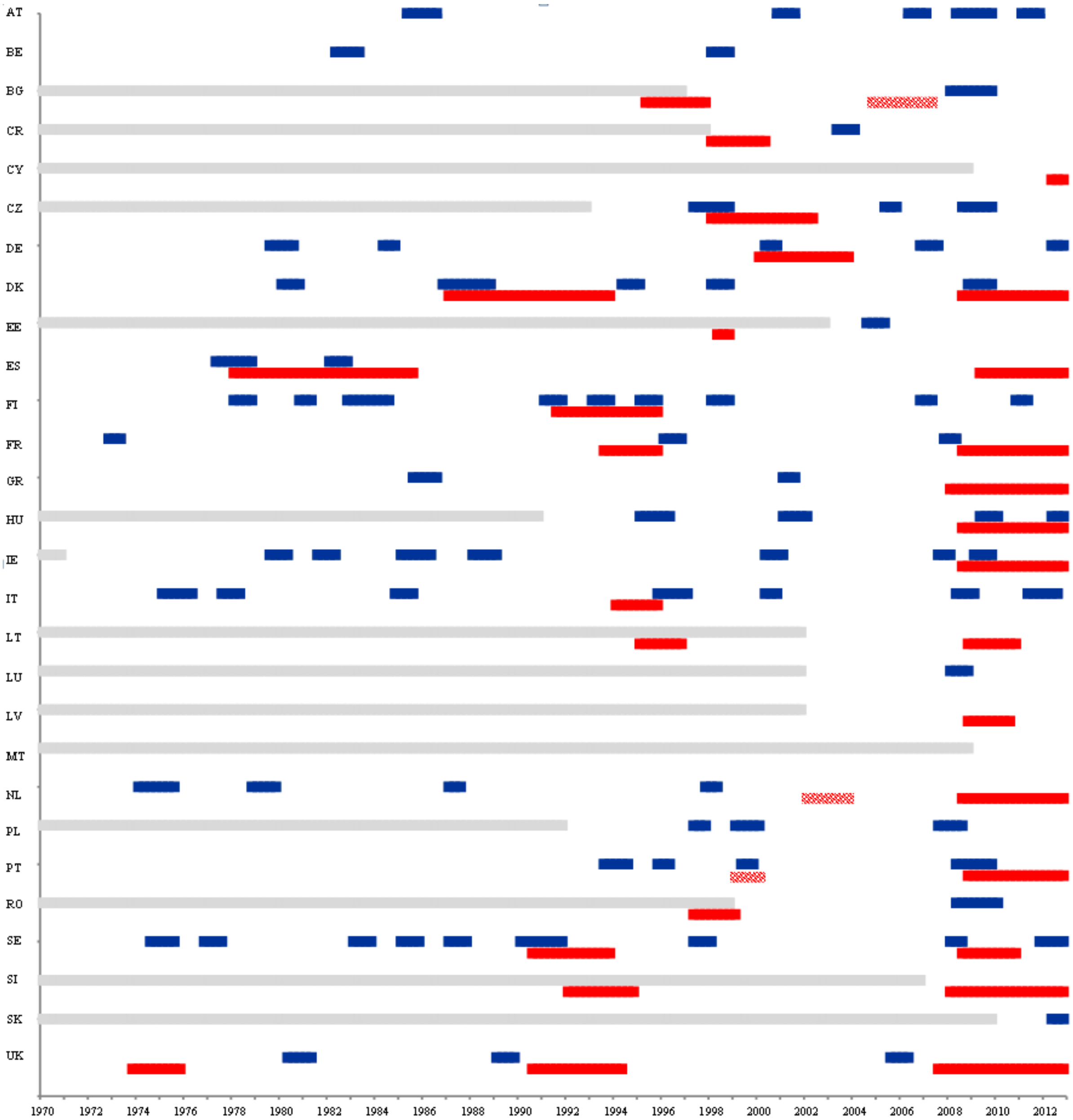
This subsection considers the case of a gradual release of the countercyclical capital buffer when risks to financial stability recede. To operationalise this, the first derivative of credit growth has been taken as a momentum indicator which signals that risks to financial stability from excessive credit growth are receding. This momentum indicator of credit growth is defined as at least three consecutive quarters of acceleration in credit growth that is followed by at least three consecutive quarters of deceleration of credit growth.[17]

Such a momentum indicator could be used to signal a gradual release of the CCB. This is displayed in Figure 4 for all EU Member States. The blue bars denote periods in which the momentum indicator would have signalled that financial stability risks from excessive credit growth were receding. Periods for which no data are available to compute the momentum indicator are shown in grey and periods of actual or would-be banking crises are shown in red. For example, if a CCB framework had been already in place, the momentum indicator would have pointed to a gradual release of any previously built up buffer in Portugal during the mid-1990s. Similarly, in Austria, Italy, Greece, Hungary and Ireland, credit growth slowed around the time of the burst of the so-called “dot-com bubble”.

The momentum indicator would also have given false signals. Figure 4 shows that there have been periods during which the momentum indicator coincided with, or was closely followed by, a materialisation of banking crises. Such periods (indicated in Figure 4 by a close overlap and/or sequence of blue and red bars for a given country) could have resulted in the objective of the buffer being defeated: the buffer might have been released – and the capital potentially used in dividend payouts – leaving no buffer available to absorb losses when they materialised in a subsequent banking crisis. As with the results for the prompt release, this suggests that judgement needs to play an important role in the gradual release of the CCB. For example, the use of the momentum indicator would have suggested a gradual release of the buffer at the beginning of 2008; however, this would have issued a false signal, given the beginning of the banking crisis shortly afterwards. In order to avoid the issuance of false signals, a cross-check with similarly important indicators for the build-up of the CCB, e.g. asset prices or current account deficits, could be recommended.

[17] Drehmann et al. (2012) find that the average duration of the credit cycle contraction phase, i.e. peak to trough, lasts a minimum of three quarters in EU Member States.

Figure 4: Credit contraction periods and banking crises



Note: The blue bars denote periods in which the momentum indicator would have signalled that financial stability risks from excessive credit growth were receding. Periods for which no data are available to compute the momentum indicator are shown in grey. Periods of actual banking crises are shown in red. Periods of would-be banking crises are shown in a patterned red.

4.3. Use of judgement

The distinction between (1) a gradual release in the case where threats to financial stability are receding and (2) a prompt release during periods of financial stress complicates the analysis of the release phase of the CCB. Moreover, data limitations, particularly concerning market-based indicators, considerably reduce the country coverage and the lengths of the time series that can be used in the empirical analysis. This places even greater weight on using judgement during the release phase than during the build-up phase.

Section 5: Operationalising the indicators

The evaluation of indicators described in the preceding sections provides guidance on thresholds beyond which the CCB might be raised. But it does not provide guidance on when the CCB should be lowered or at what level it should be set. BCBS (2010b) provides such guidance in the context of the credit-to-GDP gap: different levels of the credit-to-GDP gap are mapped into different benchmark buffer rates – the so-called "buffer guide". While this type of mapping helps operationalise indicators such as the credit-to-GDP gap, finding a suitable approach for the mapping poses a number of conceptual challenges. This section describes the BCBS's methodology and alternatives that could address these challenges. While the focus here is on the credit-to-GDP gap, the discussion applies to EWMs in general.

5.1. Thresholds and their mapping according to the BCBS

5.1.1. Selection of a lower and higher threshold according to the BCBS

The BCBS (2010b) sets two thresholds for the buffer guide. A lower threshold (L) is defined as the point beyond which the buffer guide would indicate that the CCB may need to be activated. A higher threshold (H) is defined as the point where no more capital would be required to fulfil the objective of the CCB, even if the gap continues to increase.

The BCBS suggests that these thresholds should be calibrated based on the following criteria.

- *"L should be low enough, so that banks are able to build up capital in a gradual fashion before a potential crisis. As banks are given one year to raise capital, this means that the indicator should breach its threshold at least 2-3 years prior to a crisis."*

- *"L should be high enough, so that no additional capital is required during normal times."*

- *"H should be low enough, so that the buffer would be at its maximum prior to major banking crisis…"*

Based on these criteria and the noise-to-signal ratio at different thresholds, the BCBS suggests a lower threshold of a two percentage point gap (L=2) and an upper threshold of a ten percentage point gap (H=10).

5.1.2. Mapping different levels of the gap into a buffer guide according to the BCBS

The BCBS (2010b) specifies that at or below the lower threshold (L=2), the CCB would be zero. However, it only provides an illustration for the appropriate CCB rate at the higher threshold and the mapping function between these thresholds:

"…we could assume for illustrative purposes that the maximum buffer add-on (VBmax) is 2.5% of risk weighted assets. … When the credit-to-GDP ratio is between 2 and 10 percentage points of its trend, the buffer add-on will vary linearly between 0% and 2.5%. This will imply, for example, a buffer of 1.25% when the credit-to-GDP gap is 6 (i.e. half way between 2 and 10)."

5.1.3. Conceptual challenges with the BCBS's approach

Finding a suitable approach for the mapping of different levels of the credit-to-GDP gap into different buffer rates poses a number of conceptual challenges. While the selection of the lower threshold L is fairly straightforward (see Section 5.2.1), the selection of a higher threshold H and mapping different levels of the credit-to-GDP gap into different benchmark buffer rates is less clear. Signals in early warning models, such as the ones investigated in this paper, are binary variables. If a threshold is breached a crisis is likely and the authorities should consider taking action. However, these models do not place any weight on the extent to which a threshold is breached. In other words, in the context of the early warning analysis conducted here, a credit-to-GDP gap of three percentage points cannot necessarily be interpreted as a stronger signal than credit-to-GDP gap of two percentage points. Not only do different levels of the gap in the signalling approach provide little information of the relative likelihood of a crisis and thus the probability of default (PD) of households and firms on their obligations, they also provide no information on the loss given default (LGD). In addition, the "optimal thresholds" in the signalling approach are likely to differ widely, depending on whether global or country-specific signal extraction models are used (Davis and Karim, 2008b).

5.2. Alternative approaches for triggering and calibrating the CCB

There are at least three conceptually different approaches to calibrating the CCB. These can be grouped into "probability-based", "loss-based" and "cost/benefit" approaches. Each approach poses conceptual challenges. And there are trade-offs in terms of how resource intense it is to derive them and how complex they are to explain and use. The approaches are summarised in Table 9 with the remainder of this subsection describing them in more detail without advocating one particular approach.

Table 9: Pros and cons of different approaches to calibrating benchmark buffer rates

	Method	Pros	Cons	
Probability-based approach	'Basel–revisited'	Simple, mechanistic approach to link indicator levels to benchmark buffer rates. Easy to communicate.	Ad hoc (especially the choice of the upper benchmark buffer rate). Only probability based.	Least costly
Probability-based approach	Containing crises Probabilities	Structured approach. Easy to communicate.	Only probability-based. Robust estimate of feedback assumptions difficult.	
Loss-based approach	Unexplained residual losses	Explicitly associates losses linked to credit cycle to indicators.	Based on typical losses that have occurred in past crises episodes.	
Loss-based approach	Stress tests	Focuses on resilience in a forward-looking way.	Requires high degree of transparency and good communication.	
	Cost-benefit	Most rigorous approach. Provides estimate of the maximum buffer rate.	Complex functional assumptions or estimations needed to link CCBs to costs and benefits.	Most costly

5.2.1. Selection of the lower threshold L

The lower threshold for triggering the activation of the CCB can be based on the policy-maker's assumed preferences. This requires two steps: (1) choosing the policy-maker's preference parameter θ, i.e. the relative weight between type-I and type-II errors, (2) calculating the threshold where the policy-maker's loss function is minimised over the desired horizon.

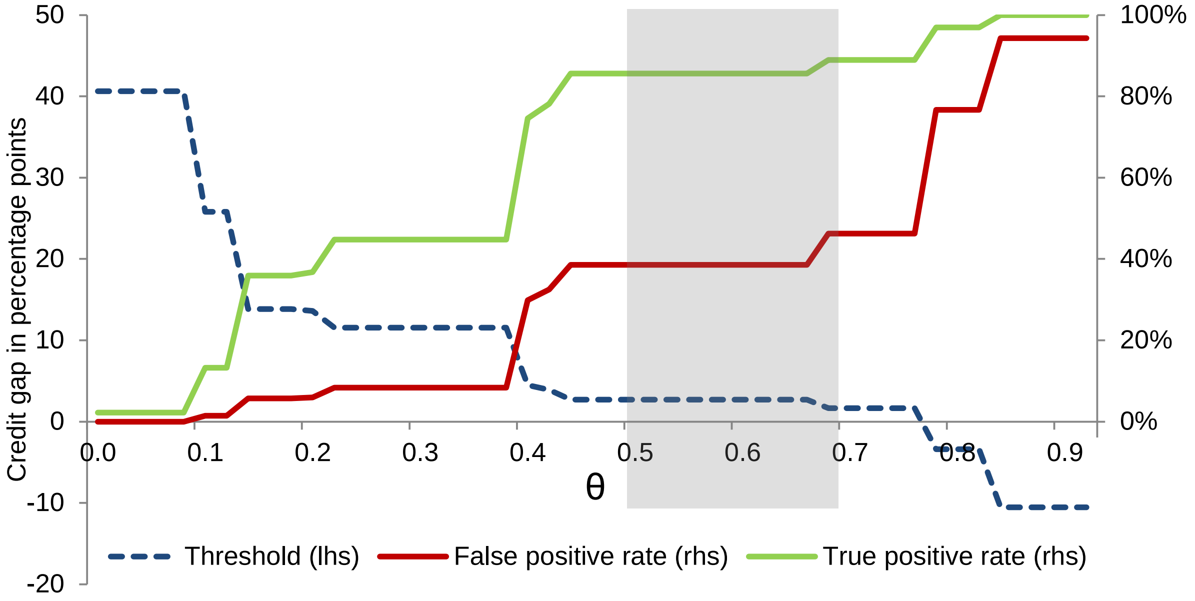
Selection of the preference parameter: This paper postulates that policy-makers would be at least as concerned about missing a crisis as they would be about false alarms, with the preference parameter theta thus in the interval $\theta \in [0.5;1]$. To derive optimal thresholds and in the absence of more accurate information on policy-makers' specific preferences, the thresholds computed in this paper were for θ 0.5, 0.6 and 0.7. The results for the optimal thresholds using these three preference parameters for all indicators tested can be found in Tables C1-C3.

Selecting the threshold that minimises the loss function: The threshold that minimises the loss function for different values of θ informs policy-makers about the range of optimal lower thresholds, L, the breaching of which would activate the CCB. A policy-maker's choice of threshold then depends on judgment regarding the appropriate value of θ. Results show that the thresholds derived for $\theta=0.7$ are often already very low, e.g. some thresholds for gaps even turn negative. The smaller the variation of the threshold across preference parameters, the more confidence might be placed on a recommendation to enact policies based on an indicator breaching its threshold. Also in this respect the credit gaps stand out, as their variation in thresholds is typically smaller than for many of the other indicators. Nevertheless, even the Basel gap's optimal thresholds vary between 3.2 and 1.3 percentage points across our range of selected preference parameters.

Figure 5 shows how the optimal threshold for the Basel credit gap depends on the preference parameter selected. The more the policy-maker is averse to missing crises (the higher θ), the lower the optimal threshold (black line). The corresponding true and false positive rates are also depicted. Figure 5 shows the relatively small variation of optimal thresholds in the selected policy preference range (shaded area).

Most importantly, despite the differences in country coverage between the BCBS's approach and this paper, the minimisation of the loss functions yield a range for optimal lower thresholds which encompasses the two percentage point gap suggested by the BCBS.

Figure 5: The Basel credit gap thresholds and policy preferences (θ)



5.2.2. CCB calibration once the lower threshold is breached

5.2.2.1. Probability based approaches

"Basel–revisited":

The BCBS suggests that the higher value H of the credit-to-GDP gap should be low enough for the CCB to be at its maximum prior to major crisis. Out of the 33 crisis episodes experienced by EU Member States since the 1970s, a threshold of H=10 as suggested in BCBS (2010b) would have been too high for the buffer to reach its maximum in the run-up to 12 of them.

These were mainly crisis episodes in the early part of the sample. Results for this period are likely to be less reliable, as the trend of the ratio of credit to GDP was estimated based on a small number of observations. For the remaining 15 episodes the credit-to-GDP gap has been higher than H=10 and in many cases substantially so (see Annex E).

Conceptually, the value chosen for H should also be high enough so that it is not reached outside financial crises and the corresponding evaluation period. Hence, in order to investigate this condition in the data, the period running from 5 years prior to a crisis until the end of the crisis is excluded. It is found that the credit-to-GDP gap exceeded the BCBS's threshold of H=10 numerous times. In some cases this reflects the credit-to-GDP gap breaching H=10 before the beginning of the evaluation period and remaining above this level for an extended period. In other cases, the gap breached this threshold in periods where a crisis did not occur. Overall, the credit-to-GDP gap exceeded H=10 outside crises or evaluation periods in 18 EU Member States.

All things considered, there is no strong case to set an upper threshold much in excess of that suggested by BCBS (2010). Taken in isolation, the analysis suggests that for the European Union as a whole the upper threshold could be set at a somewhat higher rate than H=10. Considering that the buffer guide is only a starting point which intends to provide a common basis for the setting of buffer rates, the advantages of using the same threshold, in terms of simplicity and harmonisation, would, however, seem to outweigh those of setting a different value.

The mapping of the higher and lower thresholds into benchmark buffer rates remains ad hoc. Although data for the European Union appears consistent with higher and lower thresholds of L=2 and H=10 as in BCBS (2010b), this approach still provides little guidance on how to map these thresholds into benchmark buffer rates. In particular, neither the BCBS nor the CRD IV set a maximum CCB rate, albeit reciprocity under Basel and the CRD IV is only mandatory until 2.5% of risk-weighted assets.

Containing crisis probabilities:

This approach requires the estimation of a discrete choice model (e.g. multivariate logit) including the credit gap and a leverage ratio or capital ratio among the regressors to determine the probability of crises. In this type of model the lower threshold L is derived in terms of a specific crisis probability (pL). Following this approach, the CCB would then at each quarter be set to the level which – given the estimated model – would keep the resulting crisis probability below pL. The main advantage of this approach is that it allows for the inclusion of multiple macroeconomic and macro-financial variables in the signalling process, which is likely to lead to a more informed decision on setting CCB rates.

Still, it is important to take into account that higher banking sector capitalisation is expected not only to strengthen banking sector resilience (thereby reducing the probability of a future banking crisis) but also to a certain degree to dampen the financial cycle and reduce financial imbalances. However, evidence on the impact of banks' capital adequacy on their risk-taking is mixed. For example, Blum (1998) finds that increasing capital requirements may increase banks risk-taking. In order to tackle this issue, a multivariate logistic regression model combined with a Global Vector Auto-Regression (GVAR) model (Pesaran, Schuerman and Weiner, 2004) can be used to guide policy-makers in the calibration of the CCBs. The main advantage of this approach is that it allows for an analysis of the possible effects of higher capital levels on financial vulnerabilities across countries, and thus also the CCB policy spillovers, while controlling for macro-financial feedback effects.

The simulation results from Behn et al. (2013b) using a GVAR approach suggest that the macro-financial feedback effects associated with higher banking sector capitalisation are non-negligible and important to take into account when calibrating the CCB rates. Their analysis shows that in order to have contained crisis probabilities below the optimal threshold before the global financial crisis, the CCB would have needed to be increased by more than 2.5 percentage points when the model excludes the macro-financial feedback effects. In contrast, when the model is augmented with feedback effects, substantial declines in predicted banking crisis probabilities could have been achieved by increasing the CCB by 1.25 percentage points ahead of the crisis.

5.2.2.2. Loss-based approaches

"Unexplained residual losses"[18]

This approach is based on the premise that banks' capital should be able to absorb unexpected losses. Under Basel III/CRD IV capital becomes multi-layered with ordinary capital being complemented by additional buffers, namely the capital conservation buffer, the CCB and, possibly, other surcharges related to systemic risk. Each of these layers is designed to serve a particular purpose. While ordinary capital should enable banks to withstand the impact of unexpected losses on capital in the case of severe but plausible stresses, the other two have more specific purposes. For example, the capital conservation buffer seeks to ensure that banks build up capital buffers outside periods of stress, which can be drawn down as losses are incurred so that breaches of minimum capital requirements are avoided.[19] And, as described earlier, the CCB can protect the banking sector from periods of excess aggregate credit growth.

Consistent with this, a loss-based approach could be calibrated such that it abstracts from losses that other layers of capital are meant to absorb and only accounts for loan losses that can be attributed to *excess* aggregate credit growth. In this case, the size of the CCB would only be linked to the "unexplained residual losses" after accounting for losses that have *typically* been associated with crises. This approach would be conceptually attractive in the sense that the calibration of the credit-to-GDP gap, for example, would not be linked to its early warning property, but to the type of losses the CCB is meant to absorb.

In the first stage the loan losses incurred after the turning point of GDP are empirically linked to both the contraction in GDP and credit growth during the boom phase. To this end (based on data from 2007-11) the accumulated loan impairments and provisions (percentage of total assets) over the two years following the peak in real GDP in each EU Member State are in a first step regressed on the corresponding post-peak two-year contraction in real GDP.[20] The estimation errors from this regression are interpreted as residual losses that cannot be explained by the decline in economic activity.

In a second stage, these residual losses are regressed on one of two variables[21] that detect excessive credit growth: the maximum credit-to-GDP gap or, alternatively, the compound annual growth rate of the credit-to-GDP ratio[22] during the individual upswing phase of each EU Member State. In order to operationalise the credit-to-GDP gap for the range of EU Member States, which differ with respect to financial development, each credit-to-GDP gap has been normalised by dividing it by the level of credit-to-GDP of the respective EU Member State.[23] To correct for the weakness of

[18] This approach has been developed by Torsten Wezel (ECB/IMF).

[19] BCBS (2011), "Basel III: A global regulatory framework for more resilient banks and banking systems".

[20] Since data on losses are available only at an annual frequency, the beginning of the loss accumulation does in some instances lag the turning point in GDP by a few quarters.

[21] Other variables were tested such as different variants of the credit gap or house and equity prices, but these were generally less significant than the two variables mentioned.

[22] This metric has a number of advantages; it takes into account the developments in credit and real activity during the entire upswing phase which differs strongly across EU countries. Also, the variable is not biased by individual starting levels in this ratio as opposed to the credit-to-GDP gap that requires normalisation to work well in the EU context.

[23] It appears intuitive that a 1 percentage point credit gap has a stronger impact in a country with low financial development (e.g. with a credit-to-GDP ratio below 50%) than in a mature economy (with a level well above 100%).

the credit-to-GDP gap in dealing with sudden starts and stops in credit, as experienced by some central and eastern European countries, a smaller smoothing parameter lambda has been used in the analysis (e.g. lambda = 1600, as opposed to lambda = 400,000, as advocated in BCBS, 2010b).

The outcome of such estimations is promising – the explanatory variables in each of the two regressions are highly significant at the 1%-level. The first regression explains more than 80% of the overall variation, while either of the two credit variables manages to explain 40% of the remaining variation in the second regression. For the mapping into the size of the CCB as a percentage of risk-weighted assets (RWA), the predicted values for each country from the second regression are divided by the ratio of RWA to total assets in each case and the resulting numbers are regressed again on either the maximum credit-to-GDP gap or the compound annual growth rate of the credit-to-GDP ratio, thereby obtaining the slope coefficient used in the mapping.

The resultant mapping of credit-to-GDP gaps into buffer rates is less conservative than the one suggested by the BCBS. This mapping suggests that up to a normalised credit-to-GDP gap of 4 percentage points the CCB would remain at zero. Thereafter, the buffer would grow by 0.25 percentage points for every 1.6 additional percentage points in the normalised credit-to-GDP gap, reaching 2.5% of RWA at a credit-to-GDP gap of 20 percentage points.

Top-down stress tests:

When the chosen indicator breaches L, the designated authorities could conduct stress tests to calibrate buffer settings. The top-down scenarios would be calibrated to depict a turning point in the credit cycle. A decision would have to be made as to whether the resulting capital shortfalls would be mapped 1:1 into a CCB or whether only shortfalls leading to the breach of some to be determined overall average capital ratio for a country's banking sector would be covered by the CCB. The issue is whether existing levels of capitalisation would be taken into account when deciding on the necessity and size of the CCB.

5.2.2.3. Cost-benefit approach

The most rigorous approach would be to acknowledge that the setting of CCBs entails costs and benefits and might also affect the probability of crises. The CCB should then be set at the level which at each point in time maximises the expected net benefit of setting or not setting a buffer. In order to calculate the expected net benefit the expected beneficial effects of CCBs in times of crises, as well as their costs in terms of expected lost output in times when the CCBs are activated would have to be estimated. In addition, the crisis probability would most likely not be independent from the level of CCBs, as the costs associated with higher capital requirements might dampen the upswing of the financial cycle and thus affect the credit-to-GDP gap over time. Depending on the functional forms estimated or calibrated, the final result may be that the optimal buffer rate would be a (most likely non-linear) function of the credit-to-GDP gap. The attractive feature would be that a higher gap would not lead to an ever-rising buffer rate, as at some point costs would dominate benefits. A maximum buffer rate would result endogenously from this approach.

Section 6: Conclusion

The analysis presented in this paper underpins the ESRB Recommendation on guidance for setting countercyclical buffer rates (ESRB 2014/1). It was conducted by a dedicated ESRB Expert Group working under the auspices of the ESRB Instruments Working Group.

Variables that perform well for the European Union as a whole in signalling that the CCB may need to be built up mirror those found in the literature. The credit-to-GDP gap (based on total, bank and household credit) is the best single leading indicator for systemic banking crises associated with excessive credit growth. This finding is here established for the European Union *as a whole*. The credit gap suggested by the BCBS performs well for the large majority of countries for which it can be analysed. But the credit-to-GDP gap does not perform well in all cases.

The main results for the credit-to-GDP are robust across a range of different specifications for the gap. The specification suggested by the Basel Committee on Banking Supervision (BCBS) is based on total credit to the domestic private non-financial sector and is among the best performing single indicators. Few specifications of the credit-to-GDP gap perform better. Those that do are often based on – from a forward-looking perspective – less robust, narrower credit aggregates (e.g. bank credit and credit to households).

Other variables that perform well as single indicators include the residential property price-to-income ratio, commercial and residential property price gaps, year-on-year (bank and household) real credit growth, the household debt service ratio and the real M3 gap.

Different types of multivariate analysis show that when other variables are combined with the credit-to-GDP gap, the resulting models primarily further reduce false alarms and can improve on the signalling performance of the credit-to-GDP gap considered in isolation. The variables found to add most value in the multivariate analyses, in addition to the ones listed for the univariate case, are commercial property price growth rates, the combined household and non-financial corporation debt service ratio, year-on-year changes in equity prices, the current account-to-GDP ratio and various credit-to-GDP ratios. Overall these results show that when taking into account other developments beyond credit, signalling performance improves because signals derived from excessive credit developments can be explicitly made conditional on other imbalances occurring at the same time. In other words false alarms can be reduced in a multivariate setting because not all credit growth is bad and likely to trigger a banking crisis. The authorities involved in setting CCBs should consider this finding.

Other variables that did not perform well in the joint analysis for the European Union as a whole might of course nevertheless be relevant for signalling crises in some countries. While this could not be tested empirically due to the scarcity of crises in individual countries, it points to the need to exercise judgement when setting the CCB and to further explore possibilities of how to address country idiosyncrasies in analytical work of this type.

For the release phase of the buffer, judgement may have to play an even greater role, as empirical results are less robust. Market-based indicators were found to display the best performance of coincident or near-crisis indicators which can be used to signal that the CCB should be fully released.

The LIBOR-OIS spread, covered bond spreads and the ECB's CISS indicator performed particularly well. Moreover, these indicators did not only perform well for the pooled sample, but also for most individual EU Member States. An important caveat is the short time series availability for many financial market series implying that the results are to a large extent driven by the global financial crisis. This limits the possibility for generalising these results. Moreover, there may be circumstances other than stress, e.g. when cyclical systemic risks recede, where the buffer might be reduced gradually, but which may not be captured by this analysis.

This paper finds that mapping the credit-to-GDP gap into a benchmark buffer rate poses conceptual challenges. Since the way the BCBS calibrates the benchmark buffer rate is ad hoc, the Expert Group investigated a number of alternative approaches. Some of the alternatives described in this paper are promising but would need to be developed further to serve as a practical, less ad hoc calibration of the buffer guide. Therefore, this paper does not advocate any particular calibration approach.

Annex A: Summary of data

Table A1: Start and end of crises across EU members

Country	Start and end of crises					
	Start	End	Start	End	Start	End
BE	no crisis according to the definition					
BG	Q2 1995	Q4 1997	Q4 2004	Q2 2007		
CZ	Q1 1998	Q2 2002				
DK	Q1 1987	Q4 1993	Q3 2008	ongoing*		
DE	Q1 2000	Q4 2003				
EE	Q2 1998	Q4 1998				
IE	Q3 2008	ongoing*				
GR	Q1 2008	ongoing*				
ES	Q1 1978	Q3 1985	Q2 2009	Q2 2013		
FR	Q3 1993	Q4 1995	Q3 2008	ongoing*		
HR	Q1 1998	Q2 2000				
IT	Q1 1994	Q4 1995				
CY	Q2 2012	ongoing*				
LV	Q4 2008	Q3 2010				
LT	Q1 1995	Q4 1996	Q4 2008	Q4 2010		
LU	no crisis according to the definition					
HU	Q3 2008	ongoing*				
MT	no crisis according to the definition					
NL	Q1 2002	Q4 2003	Q3 2008	ongoing*		
AT	no crisis according to the definition					
PL	no crisis according to the definition					
PT	Q1 1999	Q1 2000	Q4 2008	ongoing*		
RO	Q2 1997	Q1 1999				
SI	Q1 1992	Q4 1994	Q1 2008	ongoing*		
SK	no crisis according to the definition					
FI	Q3 1991	Q4 1995				
SE	Q3 1990	Q4 1993	Q3 2008	Q4 2010		
UK	Q4 1973	Q4 1975	Q3 1990	Q2 1994	Q3 2007	ongoing*

Notes: Dates in red indicate periods where domestic developments related to the credit/financial cycle could well have caused a systemic banking crisis had it not been for policy action / an external event that dampened the credit cycle (referred to as “would-be” crisis).

* - ongoing refers to the cut-off date of Q4 2012

BE=Belgium, BG=Bulgaria, CZ=Czech Republic, DK=Denmark, DE=Germany, EE=Estonia, IE=Ireland, GR=Greece, ES=Spain, FR=France, HR=Croatia, IT=Italy, CY=Cyprus, , LV=Latvia, LT=Lithuania, LU=Luxemburg, HU=Hungary, MT=Malta, NL=Netherlands, AT=Austria, PL=Poland, PT=Portugal, RO=Romania, SI=Slovenia, SK=Slovakia, FI=Finland, SE=Sweden, UK=United Kingdom

ESRB
European Systemic Risk Board
European System of Financial Supervision

Table A2: Data availability by country and start date (sorted by country)

	real economy variables							credit variables										market-based variables								real estate					Bank BS	Crisis episodes	
	GDP nominal	GDP real*	CPI	Unemployment	M3 nominal	Real effective ex-change rate	Current account	Nominal total credit*	Nominal total credit to NFC	Nominal total credit to HHS	Nominal bank credit	Alternative measure of credit	Total credit-to-GDP ratio	Public debt to GDP	Debt service ratio	Debt service ratio NFC	Debt service ratio HHS	3M money market rate	Long-term interest rate	Equity prices*	Libor-OIS **	ECB CISS Indicator **	ML Covered Bonds Spreads **	Sovereign CDS Premia **	Average Bank CDS Premia **	Real residential property prices	Nominal residential property prices	Ratio of residential property prices to income	Ratio of residential property estate prices to ren	Nominal commercial property prices*	Non-Performing Loans / Total Gross Loans **	Start of the crisis episoce before the global financial crisis	During the global financial crisis
BE	1980	1980	1970	1970	1970	1975	1995	1970	1980	1980	1970	1980	1980	1980	1980	1980	1995	1970	1980	1985	1999	1999		2003	2001	1970	1970	1970	1976	1997	2000	no crisis	
BG	1997	1997	1997	1997	2004	1992	1994	1997	1997	1997			1997	2000	1997	1997	1997	1998	2003		1999	1999		2000							2000	1995 2004^	
CZ	1995	1996	1991	1993	2002	1990	1993	1993	1995	1995	1993		1995	2000	1995	1995	1995	1993	2000	1994	1999	1999		2006		2008	1999	2008	2008	1997	2000	1998	
DK	1970	1970	1970	1970	1991	1975	1988	1970	1970	1970	1970		1970	2000	2003	2003	2003	1970	1970	1970	1999	1999		2008		1970	1970	1981	1970	1997	2001	1987	2008
DE	1970	1970	1970	1970	1970	1972	1971	1970	1970	1970	1970		1970	2000	1970	1991	1991	1970	1980	1970	1999	1999	1996	2003	2001	1970	1970	1980	1970	1997	2000	2000	
EE	1995	1995	1995	2000	2008	1994	1992	2003	2003	2003	1997		2003	2000	1996	2003	2003	1996	1997	1996	1999	1999		2008			2003				2000	1998	
IE	1980	1980	1970	1990	1970	1975	1981	1971	2002	2002	1971		1980	1990	1980	2002	2002	1971	1988	1970	1999	1999		2009	2003	1970	1970	1977	1970	1994	2000		2008
GR	2000	2000	1970	1983	1980	1980	1980	1970	1994	1994	1970		2000	2000	2000	2003	2003	1980	1992	1985	1999	1999		2003	2008	1997	1997	1997	1997	1998	2000		2008
ES	1970	1995	1970	1976	1970	1980	1990	1970	1980	1980	1970		1970	2000	1970	1980	2000	1977	1980	1985	1999	1999	1999	2004	2001	1971	1971	1971	1971	1997	2000	1978	2009
FR	1970	1970	1970	1970	1970	1979	1994	1970	1977	1977	1970		1970	1995	1970	1977	1977	1970	1980	1970	1999	1999	1996	2003	2001	1970	1970	1978	1970	1997	2000	1993	2008
HR	1996	1996	1992	1998	1994	1992	1999	1999	1993	1993	1993		1999	2000	2011	2011	2011	2001	1999	2000							1997			2009		1998	
IT	1970	1970	1970	1970	1970	1980	1970	1970	1970	1970	1974		1970	1970	1970	1992	1992	1970	1970	1970	1999	1999	2005	2003	2001	1970	1970	1970	1970	1997	2000	1994	
CY	1993	1995	1980	2000	2005	1980	1999	1993	2004	2004	1993		1993	2000	1993	2004	2004	1999	1998	2004	1999	1999		2008		2005	2002		2009		2008		2012
LV	1995	1995	1996	1998	1998	1994	1992	1995	1997	1997	1995		1995	2000	1995	2004	2004	1997	2001	1998	1999	1999		2006		1999	1999	1999	1999		2000		2008
LT	1995	1995	1993	1998	2004	1994	1993	1993	1993	1993	1993		1995	2000	1995	2004	2004	1999	2001	1999	1999	1999		2008		1998	1998	1998	2005	2004	2000	1995	2008
LU	1980	1995	1970	1985	1970	1975	1995	2003	2005	2005	1980		2003	2000	1980				1985	1999	1999	1999				2007	2007	1995	2007	1997	2000	no crisis	
HU	1995	1995	1980	1992	2003	1979	1993	1989	1989	1989	1989		1995	2000	1995	1995	2003	1994	2001	1991	1999	1999		2002		1998	1998	1998		1997	2000		2008
MT	1972	1975	1970	2000	1972	1975	1972	1972	2003	2003	1972		1972	2000	1972	2000	2000	1995	1999	1995	1999	1999				2005	1980	2000			2004	no crisis	
NL	1970	1970	1970	1970	1970	1975	1970	1970	1982	1982	1970		1970	1970	1970	1982	1982	1970	1970	1970	1999	1999		2008		1970	1970	1970	1970	1997	2000	2002^	2008
AT	1970	1970	1970	1970	1970	1975	1982	1970	1995	1995	1970		1970	1970	1970	1995	1995	1970	1985	1970	1999	1999		2006	2003	2000	2000	2000	2000	1997	2000	no crisis	
PL	1995	1995	1990	1992	2004	1980	1999	1992	1995	1995	1992		1995	2000	2005	2005	2005	1995	2001	1991	1999	1999		2000		2002	2002	2002		1997	2001	no crisis	
PT	1977	1978	1970	1970	1970	1975	1993	1970	1979	1979	1970		1977	1980	1977	1995	1995	1970	1986	1988	1999	1999		2003	2002	1988	1988	1995	1991	2002	2000	1999^	2008
RO	2000	2000	1995	1997	2004	1990	1991	1999	1999	1999	1991		2000	1998	2007	2007	2007	1983	2005	1997	1999	1999		2002			2009				2001	1997	
SI	1995	1995	1989	1996	2004	1994	1994	2001	1993	1993	1993		2001	2000	2004	2004	2004	1998	2002	1994	1999	1999		2008		2007	2007	2007	2007	2008	2000	1992	2008
SK	1993	1993	1991	1993	2006	1990	1994	2004	2003	2003	2004		2004	2000	2006	2006	2006	1995	2001	1993	1999	1999		2001		2005	2005	2005	2005	2001	2000	no crisis	
FI	1970	1970	1970	1970	1970	1970	1970	1970	1970	1970	1974		1970	1970	1970	1975	1990	1980	1987	1970	1999	1999		2008		1970	1970	1975	1970	1997	2000	1991	
SE	1970	1993	1970	1970	2001	1975	1982	1970	1980	1980	1970	1970	1970	2000	1970	.	.	1983	1987	1970	1999	1999	2005	2006	2003	1970	1970	1970	1980	1997	2000	1990	2008
UK	1970	1970	1970	1970	1999	1975	1971	1970	1976	1970	1970		1970	2000	1970	1976	1976	1970	1984	1970	1999	1999	2003	2008	2002	1970	1970	1975	1970	1986	2000	1973 1990	2007

Notes: * - variable used for both the tightening and the release phase; ** - variable used for the release phase only; ^ - would-be crisis
Colours for variables correspond to the year for which the first observation is available, with green indicating the longest time series. Red colour in the starting date of the crisis indicates that observations for less than half of the variables start before the crisis episode.
BE=Belgium, BG=Bulgaria, CZ=Czech Republic, DK=Denmark, DE=Germany, EE=Estonia, IE=Ireland, GR=Greece, ES=Spain, FR=France, HR=Croatia, IT=Italy, CY=Cyprus, , LV=Latvia, LT=Lithuania, LU=Luxemburg, HU=Hungary, MT=Malta, NL=Netherlands, AT=Austria, PL=Poland, PT=Portugal, RO=Romania, SI=Slovenia, SK=Slovakia, FI=Finland, SE=Sweden, UK=United Kingdom

Table A3: Data availability by country and start date (sorted by data availability)

	real economy variables							credit variables										market-based variables								real estate					Bank BS	Crisis episodes	
	GDP nominal	GDP real*	CPI	Unemployment	M3 nominal	Real effective ex-change rate	Current account	Nominal total credit*	Nominal total credit to NFC	Nominal total credit to HHS	Nominal bank credit	Alternative measure of credit	Total credit-to-GDP ratio	Public debt to GDP	Debt service ratio	Debt service ratio NFC	Debt service ratio HHS	3M money market rate	Long-term interest rate	Equity prices*	Libor-OIS **	ECB CISS Indicator **	ML Covered Bonds Spreads **	Sovereign CDS Premia **	Average Bank CDS Premia **	Real residential property prices	Nominal residential property prices	Ratio of residential property prices to income	Ratio of residential property estate prices to rer	Nominal commercial property prices*	Non-Performing Loans / Total Gross Loans **	Start of the crisis episoce before the global financial crisis	During the global financial crisis
NL	1970	1970	1970	1970	1970	1975	1970	1970	1982	1982	1970		1970	1970	1970	1982	1982	1970	1970	1970	1999	1999		2008		1970	1970	1970	1970	1997	2000	2002^	2008
FI	1970	1970	1970	1970	1970	1970	1970	1970	1970	1970	1974		1970	1970	1970	1975	1990	1980	1987	1970	1999	1999		2008		1970	1970	1975	1970	1997	2000	1991	
IT	1970	1970	1970	1970	1970	1980	1970	1970	1970	1970	1974		1970	1970	1970	1992	1992	1970	1970	1970	1999	1999	2005	2003	2001	1970	1970	1970	1970	1997	2000	1994	
DE	1970	1970	1970	1970	1970	1972	1971	1970	1970	1970	1970		1970	2000	1970	1991	1991	1970	1980	1970	1999	1999	1996	2003	2001	1970	1970	1980	1970	1997	2000	2000	
FR	1970	1970	1970	1970	1970	1979	1994	1970	1977	1977	1970		1970	1995	1970	1977	1977	1970	1980	1970	1999	1999	1996	2003	2001	1970	1970	1978	1970	1997	2000	1993	2008
UK	1970	1970	1970	1970	1999	1975	1971	1970	1976	1970	1970		1970	2000	1970	1976	1976	1970	1984	1970	1999	1999	2003	2008	2002	1970	1970	1975	1970	1986	2000	1973 1990	2007
BE	1980	1980	1970	1970	1970	1975	1995	1970	1980	1980	1970	1980	1980	1980	1980	1980	1995	1970	1980	1985	1999	1999		2003	2001	1970	1970	1970	1976	1997	2000	no crisis	
DK	1970	1970	1970	1970	1991	1975	1988	1970	1970	1970	1970		1970	2000	2003	2003	2003	1970	1970	1970	1999	1999		2008		1970	1970	1981	1970	1997	2001	1987	2008
SE	1970	1993	1970	1970	2001	1975	1982	1970	1980	1980	1970	1970	1970	2000	1970	.	.	1983	1987	1970	1999	1999	2005	2006	2003	1970	1970	1970	1980	1997	2000	1990	2008
ES	1970	1995	1970	1976	1970	1980	1990	1970	1980	1980	1970		1970	2000	1970	1980	2000	1977	1980	1985	1999	1999	1999	2004	2001	1971	1971	1971	1971	1997	2000	1978	2009
IE	1980	1980	1970	1990	1970	1975	1981	1971	2002	2002	1971		1980	1990	1980	2002	2002	1971	1988	1970	1999	1999		2009	2003	1970	1970	1977	1970	1994	2000		2008
AT	1970	1970	1970	1970	1970	1975	1982	1970	1995	1995	1970		1970	1970	1970	1995	1995	1970	1985	1970	1999	1999		2006	2003	2000	2000	2000	2000	1997	2000	no crisis	
PT	1977	1978	1970	1970	1970	1975	1993	1970	1979	1979	1970		1977	1980	1977	1995	1995	1970	1986	1988	1999	1999		2003	2002	1988	1988	1995	1991	2002	2000	1999^	2008
MT	1972	1975	1970	2000	1972	1975	1972	1972	2003	2003	1972		1972	2000	1972	2000	2000	1995	1999	1995	1999	1999				2005	1980	2000			2004	no crisis	
GR	2000	2000	1970	1983	1980	1980	1980	1970	1994	1994	1970		2000	2000	2000	2003	2003	1980	1992	1985	1999	1999		2003	2008	1997	1997	1997	1997	1998	2000		2008
LU	1980	1995	1970	1985	1970	1975	1995	2003	2005	2005	1980		2003	2000	1980				1985	1999	1999	1999				2007	2007	1995	2007	1997	2000	no crisis	
HU	1995	1995	1980	1992	2003	1979	1993	1989	1989	1989	1989		1995	2000	1995	1995	2003	1994	2001	1991	1999	1999		2002		1998	1998	1998		1997	2000		2008
PL	1995	1995	1990	1992	2004	1980	1999	1992	1995	1995	1992		1995	2000	2005	2005	2005	1995	2001	1991	1999	1999		2000		2002	2002	2002		1997	2001	no crisis	
CZ	1995	1996	1991	1993	2002	1990	1993	1993	1995	1995	1993		1995	2000	1995	1995	1995	1993	2000	1994	1999	1999		2006		2008	1999	2008	2008	1997	2000	1998	
BG	1997	1997	1997	1997	2004	1992	1994	1997	1997	1997			1997	2000	1997	1997	1997	1998	2003		1999	1999		2000							2000	1995^ 2004	
LV	1995	1995	1996	1998	1998	1994	1992	1995	1997	1997	1995		1995	2000	1995	2004	2004	1997	2001	1998	1999	1999		2006		1999	1999	1999	1999		2000		2008
LT	1995	1995	1993	1998	2004	1994	1993	1993	1993	1993	1993		1995	2000	1995	2004	2004	1999	2001	1999	1999	1999		2008		1998	1998	1998	2005	2004	2000	1995	2008
CY	1993	1995	1980	2000	2005	1980	1999	1993	2004	2004	1993		1993	2000	1993	2004	2004	1999	1998	2004	1999	1999		2008		2005	2002		2009		2008		2012
RO	2000	2000	1995	1997	2004	1990	1991	1999	1999	1999	1991		2000	1998	2007	2007	2007	1983	2005	1997	1999	1999		2002			2009				2001	1997	
HR	1996	1996	1992	1998	1994	1992	1999	1999	1993	1993	1993		1999	2000	2011	2011	2011	2001	1999	2000							1997			2009		1998	
EE	1995	1995	1995	2000	2008	1994	1992	2003	2003	2003	1997		2003	2000	1996	2003	2003	1996	1997	1996	1999	1999		2008			2003				2000	1998	
SI	1995	1995	1989	1996	2004	1994	1994	2001	1993	1993	1993		2001	2000	2004	2004	2004	1998	2002	1994	1999	1999		2008		2007	2007	2007	2007	2008	2000	1992	2008
SK	1993	1993	1991	1993	2006	1990	1994	2004	2003	2003	2004		2004	2000	2006	2006	2006	1995	2001	1993	1999	1999		2001		2005	2005	2005	2005	2001	2000	no crisis	

Notes: * - variable used for both the tightening and the release phase; ** - variable used for the release phase only; ^ - would-be crisis
Colours for variables correspond to the year for which the first observation is available, with green indicating the longest time series. Red colour in the starting date of the crisis indicates that observations for less than half of the variables start before the crisis episode.
BE=Belgium, BG=Bulgaria, CZ=Czech Republic, DK=Denmark, DE=Germany, EE=Estonia, IE=Ireland, GR=Greece, ES=Spain, FR=France, HR=Croatia, IT=Italy, CY=Cyprus, , LV=Latvia, LT=Lithuania, LU=Luxemburg, HU=Hungary, MT=Malta, NL=Netherlands, AT=Austria, PL=Poland, PT=Portugal, RO=Romania, SI=Slovenia, SK=Slovakia, FI=Finland, SE=Sweden, UK=United Kingdom

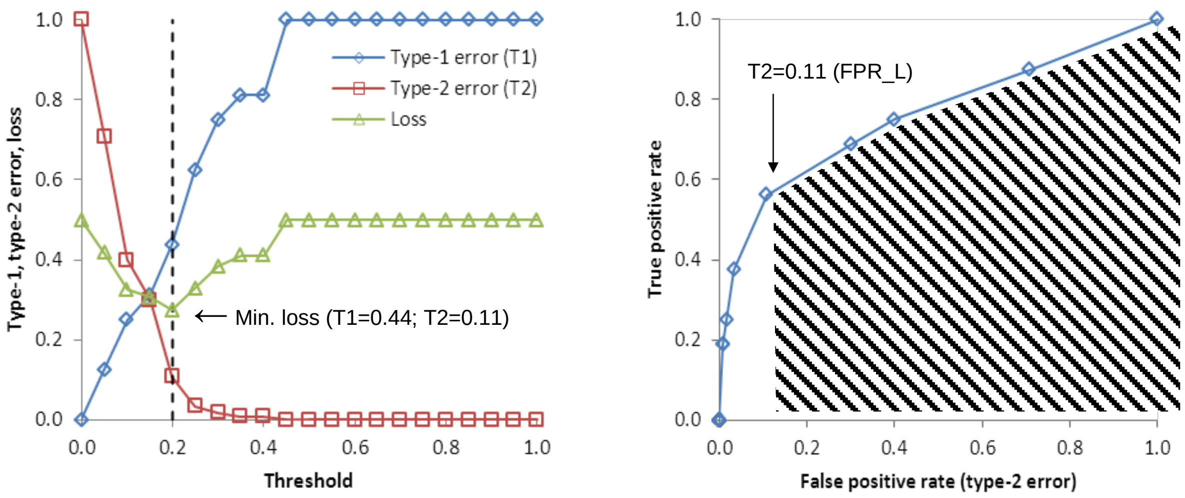
Annex B: Estimation of the pAUROC and statistical inference

Estimation of the pAUROC

The estimation of the pAUROC involves three broad steps: (1) the specification of the restricted range of false positive (or alternatively true positive) rates, (2) the computation of the partial area under the ROC curve, and (3) the standardisation of the pAUROC measure.

In the first step, the loss for every possible threshold based on a simple loss function is calculated – the policy preference, theta, relating to type-I errors is set to 0.5. The threshold at which the loss function is minimised then provides the lower bound for the false positive rate over which the partial area under the ROC curve is computed. This is illustrated in Figure B.1. The left panel depicts the type-I and type-II errors and calculates the loss function (for theta=0.5) for an indicator over all threshold values between 0 and 1. It shows that the loss function is minimised at a threshold of 0.2, when the true positive rate is 0.56 (i.e. one minus the type-I error rate of 0.44) and the false positive rate is 0.11 (i.e. the type-II error rate). For a higher theta, the green loss function would move up and the minimum moves to the left – e.g. for theta=1, the minimum of the loss function would be at the origin of the left panel. The right panel shows the corresponding ROC curve and the pAUROC (the hatched area).

Figure B.1: Restricting the range of false positive rates



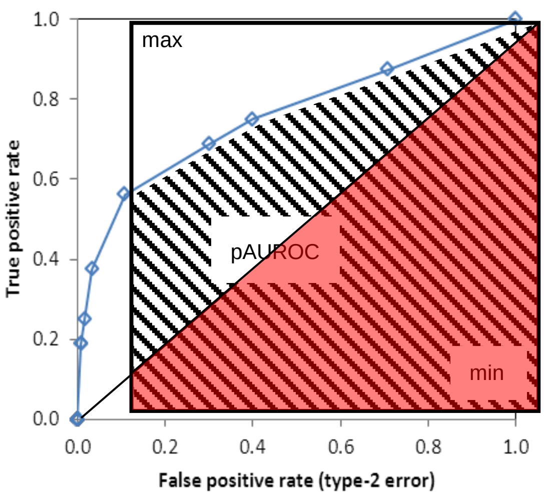
In the second step, the pAUROC is computed as the integral under the ROC curve for false positive rates (FPR) between the lower bound FPR_L and the upper bound FPR_U=1.

In the third step, the pAUROC is standardised so that the measure can be used to rank different indicators with respect to their discriminatory accuracy. In the case of the area under the full ROC curve, the AUROC measure ranges between 0 and 1 as it refers to the full range of false and true positive rates. Since all values of 0.5 or below indicate an uninformative model, an AUROC of 0.5 indicates the worst model, while a value of 1 indicates the best model. In the case of the pAUROC this is no longer true, as its minimum and maximum attainable values vary depending on the specific range of false positive rates over which the partial area is computed. Hence, to standardise the pAUROC, McClish (1989) proposes the use of the following transformation:

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

where min is the pAUROC under the diagonal ROC curve and max is the pAUROC under the perfect ROC curve. As a consequence, for a perfect ROC curve (when pAUROC=max) psAUROC equals 1 and for a non-discriminant ROC curve (when pAUROC=min) psAUROC equals 0.5. Figure B2.2 illustrates the procedure to standardise the pAUROC by showing the different parameters in the above equation. The maximum attainable area (over the restricted false positive rate range) is represented by the black rectangle, the pAUROC is the hatched area as in Figure B2.1, and the minimum attainable area is indicated by the red area.

Figure B2.2: Procedure to standardise the pAUROC measure



Statistical inference on AUROCS

The statistical inference on AUROC is intimately related with the method employed for the calculation of a summary statistic of this kind. Traditionally, in medical studies specific parametric distributions are assumed as data generating process of normal and abnormal tests (i.e. no crises or crisis events in this case). On this basis, the maximum-likelihood estimations of AUROC and the relevant parameters (i.e. mean and standard deviation) can be obtained and then confidence intervals (CIs) and statistical comparison tests can be built. This approach has made wide use of a binormal model assumption (e.g. Dorfman and Alf, 1969). In order to take into account the correlation between ROC curves of different indicators obtained from the same sample of individuals (i.e. countries), a bivariate binormal model assumption is advocated (DeLong et al. 1988, and references therein).

The AUROC can also be calculated non-parametrically by trapezoidal integration as is the case in the present work. The statistical inference can be then performed on the basis of fully non-parametric or semi-parametric approaches. While the former are almost exclusively based on bootstrap techniques, the latter make use of alternative methods for estimating AUROC standard errors (SEs).

Then, for binary classification problems, which tend to be well characterised by a binomial distribution, the Chebyshev's inequality allows CIs to be built for statistical inference once the SE has been estimated. In what follows some of these approaches are reviewed briefly.

Building on the relationship between the AUROC calculated using trapezoidal integration (henceforth empirical AUROC) and the Wilcoxon statistic, Hanley and McNeil (1982) propose a formula for the calculation of AUROC's standard errors. Their methodology takes account of situations where the numbers of abnormal and normal cases (i.e. crises and no crises) are not the same. They calculate the SE for an AUROC (*A*) as:

$$SE(A)=\sqrt{\frac{A(1-A)+(n_a-1)(Q_1-A^2)+(n_n-1)(Q_2-A^2)}{n_a n_b}}$$

where, n_a and n_b are the number of crisis and non-crisis periods, respectively. Q_1 and Q_2 are quantities that can be calculated non-parametrically from the sample (see Section 4 in Hanley and McNeil (1982) for a detailed explanation) or assuming particular distribution models for the behaviour of crisis and non-crisis events. In the latter approach Q_1 and Q_2 depend in a complex manner on the parameters of the indicators' distributions. Hanley and McNeil (1982) have shown that changing the assumed distribution of the indicator in crisis and non-crisis periods has little impact on the standard error. In addition, they conclude that a negative exponential distribution results in the most conservative estimation of the SE. Fortunately, such a model produces Q_1 and Q_2 which can be easily calculated:

$$Q_1=A/(2-A)$$
$$Q_2=2A^2/(1+A)$$

Following the approach of Hanley and McNeil (1982, 1983), the standard error of the difference between two AUROCs for the comparison of different areas (A_1 and A_2) that have been calculated from the same sample can be obtained by:

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

where r is a quantity that represents the correlation induced between the two AUROCs by the study of the same sample. In our calculations we set $r=0$ with the aim of keeping the analysis as conservative as possible.

DeLong et al. (1988) propose a fully non-parametric test for the comparison of ROC curves based on the comparison of AUROCs obtained from the same sample of individuals. As Hanley and McNeil (1982) explain, the relationship between the empirical AUROC and the Wilcoxon statistic can be exploited and they propose an estimator of the variance-covariance matrix using a non-parametric kernel estimator based on generalised U-statistics theory.

In line with the papers reviewed above, Cortes and Mohri (2004) provide CIs for the AUROC without imposing distributional assumptions on the behaviour of crisis and non-crisis events. They make extensive use of combinatorial analysis using simple parameters taken from the confusion matrix (e.g. error rate). While the formula proposed is rather complex, it should not pose any problem given the calculation power of modern computers.

The use of bootstrap techniques is common in modern statistics. Such techniques require a random resampling with a repositioning of the data in order to build *m* additional samples and to obtain robust estimated statistics (Efron and Tibshirani, 1994). The bootstrap permits both a fully non-parametric calculation of CIs using the percentiles of the bootstrapped distribution of the AUROCs and an empirical estimation of SEs based on the bootstrapped distribution which can then be used for building CIs. One particular advantage of CIs based on bootstrap percentiles is that their limits always lie within the AUROC domain (i.e. the interval [0, 1]). In order to obtain meaningful estimated parameters the resampling algorithm should take proper account of the structure of the reference sample (Pepe, 2009).

The comparison of ROC curves at specific false positive rates (or true positive rates) can be useful in some cases. As for the statistical inference on AUROCs, several parametric and non-parametric approaches can be implemented for building confidence intervals around ROC curve points. Macskassy and Provost (2004) survey several methods used in the medical literature and the literature on machine learning.

Annex C: Results for the static analysis

Table C1 – Results for credit-to-GDP gaps and other credit variables

Credit-to-GDPgaps	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
Bank credit-to-GDP gap, rw60	0.81	0.012	0.95	1.93	1.93	1.93	0.90	0.90	0.90	0.42	0.42	0.42	0.24	0.17	0.10	267	1003	0.20	5
Credit to households-to-GDP gap, nrw	0.80	0.012	0.86	2.78	1.68	0.30	0.67	0.77	0.89	0.20	0.32	0.50	0.23	0.14	0.07	345	1207	0.08	13
Credit to households-to-GDP gap, rw60	0.79	0.012	0.90	27.40	2.70	2.70	0.71	1.00	1.00	0.27	0.58	0.58	0.22	0.17	0.13	229	631	0.00	5
Credit-to-GDP ratio minus ratio of trends	0.79	0.012	0.86	12.67	9.11	8.05	0.72	0.82	0.84	0.24	0.38	0.42	0.24	0.14	0.07	431	2164	0.11	9
Total credit-to-GDP gap, moving average forecast**	0.79	0.012	0.91	4.65	3.61	2.69	0.83	0.89	0.92	0.38	0.45	0.52	0.23	0.15	0.09	369	1699	0.14	7
Credit to households-to-GDP gap, rw60	0.79	0.012	0.98	0.07	-0.42	-0.42	0.95	0.97	0.97	0.47	0.51	0.51	0.24	0.18	0.13	213	607	0.20	5
Total credit-to-GDP gap, linear forecast**	0.79	0.012	0.90	3.95	2.32	1.59	0.80	0.87	0.90	0.33	0.43	0.47	0.24	0.15	0.09	369	1699	0.14	7
Total credit-to-GDP gap, nrw (Basel gap)	**0.79**	**0.012**	**0.90**	**2.70**	**2.70**	**1.64**	**0.86**	**0.86**	**0.89**	**0.39**	**0.39**	**0.46**	**0.24**	**0.16**	**0.08**	**362**	**1652**	**0.07**	**15**
Differenced relative banking credit*	0.79	0.012	0.90	0.10	0.10	0.04	0.79	0.79	0.89	0.34	0.34	0.52	0.23	0.14	0.07	408	2176	0.00	17
Total credit gap multiplied by credit-to-GDP level	0.77	0.012	0.86	382.37	91.03	-13.02	0.67	0.81	0.90	0.23	0.42	0.62	0.22	0.12	0.05	431	2164	0.11	19
Bank credit-to-GDP relative gap, rw60	0.77	0.012	0.93	3.97	1.47	0.01	0.81	0.90	0.94	0.36	0.48	0.53	0.22	0.15	0.10	267	1003	0.20	5
Bank credit-to-GDP gap, nrw	0.77	0.012	0.88	2.49	2.23	-0.67	0.79	0.80	0.93	0.36	0.38	0.63	0.21	0.13	0.07	377	1761	0.00	14
Credit to households-to-GDP relative gap, nrw	0.76	0.012	0.87	11.78	6.18	6.18	0.78	0.89	0.89	0.37	0.53	0.53	0.21	0.12	0.07	362	1297	0.07	14
Total credit-to-GDP gap, rw60	0.76	0.012	0.88	3.02	3.02	0.89	0.81	0.81	0.87	0.41	0.41	0.51	0.20	0.12	0.06	280	983	0.00	5
Credit to households-to-GDP relative gap, nrw	0.75	0.012	0.87	4.83	4.83	-0.07	0.77	0.77	0.90	0.34	0.34	0.57	0.22	0.13	0.06	342	1204	0.08	13
Total credit-to-GDP relative gap, nrw	0.74	0.013	0.91	2.54	2.54	1.34	0.84	0.84	0.88	0.41	0.41	0.49	0.21	0.14	0.07	359	1649	0.07	15
Bank credit-to-GDP relative gap, nrw	0.74	0.013	0.94	8.27	8.27	7.07	0.89	0.89	0.91	0.50	0.50	0.53	0.20	0.14	0.08	394	2120	0.07	15
Total credit relative gap, nrw	0.74	0.013	0.98	7.15	7.15	6.07	0.95	0.95	0.97	0.61	0.61	0.66	0.17	0.12	0.08	382	1976	0.13	16
Credit to households-to-GDP relative gap, rw60	0.74	0.013	0.99	-0.59	-0.59	-1.11	0.97	0.97	0.98	0.50	0.50	0.51	0.24	0.18	0.13	213	607	0.20	5
Total credit-to-GDP relative gap, r60	0.74	0.013	0.88	3.34	0.60	-4.07	0.76	0.87	1.00	0.37	0.52	0.74	0.19	0.11	0.08	280	983	0.00	5
Differenced relative total credit*	0.74	0.013	0.85	0.16	0.05	0.04	0.67	0.90	0.91	0.31	0.59	0.61	0.18	0.10	0.05	425	2087	0.06	17
Bank credit relative gap, rw60	0.73	0.013	0.98	5.23	3.00	3.00	0.90	0.97	0.97	0.54	0.62	0.62	0.18	0.13	0.09	304	1252	0.00	6
Total credit gap multiplied by log of credit-to-GDP level	0.73	0.013	0.81	141.14	141.14	36.11	0.72	0.72	0.89	0.33	0.33	0.64	0.20	0.10	0.03	383	1400	0.06	17
Bank credit-to-GDP relative gap, nrw	0.73	0.013	0.89	3.65	3.29	-1.50	0.78	0.80	0.95	0.39	0.41	0.65	0.20	0.12	0.07	377	1761	0.00	14
Total credit relative gap, rw60	0.70	0.013	0.81	11.05	2.47	-0.42	0.64	0.93	1.00	0.35	0.70	0.80	0.15	0.08	0.06	317	1230	0.00	6
Credit to NFC relative gap, nrw	0.69	0.013	0.97	1.70	1.47	0.20	0.94	0.95	0.97	0.63	0.64	0.69	0.15	0.11	0.07	362	1274	0.21	14

Credit growth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
y-o-y growth rate (%) of real bank credit	0.70	0.013	0.86	5.54	4.73	-0.25	0.77	0.81	0.97	0.46	0.52	0.80	0.15	0.08	0.04	432	2539	0.11	9
y-o-y growth rate (%) of real total credit	0.67	0.013	0.83	6.42	1.84	0.98	0.69	0.94	0.97	0.43	0.74	0.78	0.13	0.07	0.04	441	2434	0.44	9
y-o-y growth rate (%) of real credit to NFC	0.65	0.013	0.77	5.34	1.81	-2.16	0.67	0.85	0.97	0.43	0.67	0.87	0.12	0.04	0.02	413	1698	0.44	9
y-o-y growth rate (%) of real household credit	0.65	0.013	0.81	7.72	4.02	4.02	0.71	0.90	0.90	0.46	0.67	0.67	0.12	0.07	0.03	420	1710	0.22	9

* $\frac{5\cdot L_t}{\sum_{i=0}^{4} GDP_{t-i}} - \frac{5\cdot L_{t-4}}{\sum_{i=4}^{7} GDP_{t-i}}$, where L_t stands for credit in period t (Karlo Kauko, 2012b);

** Variables described in section 3.1.2.1 (alternative trend specifications, third point on the numbered list, p. 24)

Notes for this Table and naming conventions for variables are described at the end of Annex C.

Table C2 – Results for financial market variables and macro variables

Market variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
Inverted 3M money market interest rate	0.65	0.013	0.71	-4.89	-14.09	-14.27	0.57	0.97	0.97	0.32	0.83	0.84	0.12	0.05	0.03	439	2311	0.30	10
Inverted long-term interest rate	0.64	0.013	0.67	-4.94	-4.94	-14.87	0.56	0.56	0.99	0.23	0.23	0.94	0.17	0.05	0.01	382	1749	0.25	8
y-o-y growth rate (%) of real equity prices	0.60	0.013	0.79	3.89	0.48	-79.03	0.76	0.80	1.00	0.51	0.56	1.00	0.12	0.05	0.00	406	2166	0.25	8
y-o-y growth rate (%) of equity prices	0.58	0.013	0.82	5.88	5.88	-75.64	0.79	0.79	1.00	0.57	0.57	1.00	0.11	0.05	0.00	406	2166	0.38	8
Annual absolute change (pp) in 3M money market interest rate	0.58	0.013	0.87	-0.40	-1.74	-2.64	0.70	0.91	0.96	0.51	0.76	0.84	0.09	0.04	0.02	415	2196	0.40	10
Inverted real long-term interest rate	0.58	0.013	0.60	-2.16	-13.15	-13.15	0.42	1.00	1.00	0.25	1.00	1.00	0.09	0.00	0.00	382	1737	0.63	8
Annual absolute change (pp) in real 3M money market interest rate	0.55	0.013	0.96	-0.67	-1.76	-5.57	0.77	0.92	1.00	0.62	0.78	0.96	0.08	0.04	0.01	411	2143	0.20	10
Annual absolute change (pp) in long term-interest rate	0.55	0.013	0.65	-0.33	-1.40	-8.79	0.61	0.91	1.00	0.48	0.82	1.00	0.06	0.01	0.00	369	1652	0.29	7
Inverted real 3M money market interest rate	0.55	0.013	0.63	-1.63	-10.26	-10.26	0.54	0.99	0.99	0.40	0.98	0.98	0.07	0.00	0.00	433	2252	0.50	10
y-o-y growth rate (%) of effective exchange rate	0.54	0.013	0.74	-1.43	-5.67	-90.46	0.78	0.95	1.00	0.69	0.91	1.00	0.05	0.01	0.00	468	2587	0.36	11
Annual absolute change (pp) in long-term real interest rate	0.49	0.013	0.60	-1.29	-8.15	-8.15	0.80	1.00	1.00	0.77	1.00	1.00	0.02	0.00	0.00	369	1640	0.57	7

Macro variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
Real M3 (notional stock) relative gap, nrw	0.68	0.013	0.87	2.22	2.10	2.10	0.84	0.84	0.84	0.46	0.47	0.47	0.19	0.12	0.05	222	1403	0.00	5
y-o-y growth rate of real M3 (notional stock)	0.65	0.013	0.68	8.53	0.79	0.70	0.39	0.93	0.94	0.16	0.79	0.80	0.11	0.04	0.02	334	1788	0.29	7
Real M3 (notional stock) relative gap, rw60	0.63	0.013	0.79	3.00	1.67	-6.45	0.72	0.80	1.00	0.44	0.54	0.95	0.14	0.06	0.02	201	974	0.20	5
Current account surplus (minus means deficit)	0.62	0.013	0.71	0.30	-0.97	-11.13	0.70	0.79	1.00	0.50	0.62	0.97	0.10	0.03	0.01	395	1935	0.35	20
Nominal GDP relative gap, nrw	0.61	0.013	0.71	4.60	-1.61	-1.61	0.62	0.99	0.99	0.41	0.91	0.91	0.11	0.03	0.02	388	1913	0.43	7
Public debt-to-GDP relative gap, r60	0.59	0.013	0.61	1.01	0.50	-36.36	0.51	0.52	0.98	0.23	0.24	1.00	0.14	0.01	-0.01	95	417	1.00	1
Public debt-to-GDP gap, nrw	0.58	0.013	0.71	-5.42	-6.41	-19.57	0.78	0.82	1.00	0.61	0.66	0.98	0.09	0.03	0.01	116	664	0.50	2
Nominal GDP relative gap, rw60	0.58	0.013	0.69	2.22	1.57	-3.43	0.68	0.75	0.99	0.46	0.56	0.96	0.11	0.03	0.00	288	1077	0.20	5
Real GDP relative gap, nrw	0.56	0.013	0.78	-0.75	-1.33	-4.20	0.86	0.91	1.00	0.73	0.79	0.96	0.07	0.03	0.01	365	1687	0.38	8
Annual absolute change (pp) in inflation rate	0.55	0.013	1.00	1.00	0.73	0.73	0.93	0.96	0.96	0.86	0.90	0.90	0.04	0.02	0.00	469	2829	0.45	11
y-o-y growth rate (%) of real GDP	0.51	0.013	0.69	2.08	-0.22	-1.21	0.78	0.98	0.99	0.71	0.94	0.96	0.04	0.01	0.00	400	2140	0.60	10
Unemployment rate	0.50	0.013	0.91	3.21	2.71	1.89	0.94	0.96	1.00	0.83	0.86	0.92	0.06	0.04	0.02	449	2360	0.44	9
Public debt to GDP relative gap, nrw	0.48	0.013	0.48	0.75	-31.29	-31.29	0.50	0.99	0.99	0.38	1.00	1.00	0.06	-0.01	-0.01	116	664	0.50	2
y-o-y growth rate (%) of nominal GDP	0.46	0.013	0.90	3.52	-4.34	-4.34	0.91	1.00	1.00	0.89	1.00	1.00	0.01	0.00	0.00	438	2370	0.55	11
Real GDP relative gap, rw60	0.41	0.012	0.98	-4.17	-4.17	-4.17	1.00	1.00	1.00	0.97	0.97	0.97	0.01	0.01	0.01	229	835	0.75	4

Notes for this Table and naming conventions for variables are described at the end of Annex C.

Table C3 – Results for debt service ratio, balance sheet variables and property variables

Debt service ratios	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
Debt service ratio, households	0.65	0.013	0.70	0.11	0.11	0.01	0.59	0.59	1.00	0.19	0.20	0.97	0.20	0.07	0.01	314	886	0.11	9
Debt service ratio	0.61	0.013	0.64	0.18	0.02	0.02	0.44	1.00	1.00	0.20	0.99	0.99	0.12	0.00	0.00	421	2177	0.18	11
Debt service ratio, NFC	0.52	0.013	0.51	0.32	0.03	0.03	0.43	1.00	1.00	0.22	0.97	0.97	0.10	0.01	0.01	329	1093	0.11	9

Balance sheet variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	crisis quarters within evaluation	number of evaluated quarters	Percent of MS for which indicator has no significant AUROC	countries for which AUROC has been calculated
Leverage ratio	0.46	0.016	0.85	17.03	13.45	13.45	0.92	1.00	1.00	0.85	0.95	0.95	0.03	0.02	0.01	286	1086	0.64	11

Property variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
				Range of optimal thresholds			Range of TPR			Range of FPR			Usefulness			Stability			
Indicator	AUROC	sd(AUROC)	ps(AUROC)	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	0.5	0.6	0.7	Crisis quarters within evaluation	Number of evaluated quarters	Percentage of MS for which indicator has no significant AUROC	Countries for which AUROC has been calculated
House price-to-income ratio gap, nrw	0.73	0.013	0.76	13.36	2.45	-2.39	0.50	0.79	0.91	0.13	0.50	0.69	0.19	0.08	0.03	314	1186	0.17	6
House price-to-income ratio (index)	0.73	0.013	0.85	95.47	95.10	93.51	0.79	0.80	0.83	0.40	0.41	0.46	0.19	0.11	0.04	366	1622	0.27	15
House price-to-rent ratio gap, nrw	0.72	0.013	0.76	13.57	-6.83	-8.27	0.50	0.95	0.97	0.17	0.71	0.73	0.16	0.09	0.06	306	1242	0.33	6
House price-to-rent ratio (index)	0.71	0.013	0.86	92.07	84.99	83.57	0.80	0.89	0.90	0.45	0.57	0.59	0.17	0.11	0.05	365	1592	0.15	13
House price-to-income ratio relative gap, nrw	0.71	0.013	0.75	14.21	-2.60	-2.60	0.47	0.92	0.92	0.16	0.69	0.69	0.16	0.08	0.04	314	1186	0.17	6
House price-to-income gap, rw60	0.70	0.013	0.78	8.78	8.01	-4.10	0.69	0.71	0.94	0.37	0.40	0.75	0.16	0.07	0.03	187	707	0.20	5
Residential property prices gap, rw60	0.70	0.013	0.74	5.88	1.22	-3.98	0.56	0.78	0.98	0.25	0.51	0.84	0.16	0.07	0.03	257	878	0.00	5
House price-to-rent ratio relative gap, nrw	0.70	0.013	0.73	21.11	-7.95	-7.95	0.37	0.97	0.97	0.09	0.72	0.72	0.14	0.10	0.06	306	1242	0.50	6
Real residential property prices gap, nrw	0.69	0.013	0.70	12.58	3.13	-19.57	0.41	0.65	1.00	0.07	0.37	1.00	0.17	0.04	0.00	348	1349	0.17	6
Residential property prices relative gap, rw60	0.69	0.013	0.77	12.41	-4.38	-4.38	0.54	0.96	0.96	0.28	0.72	0.72	0.13	0.09	0.06	257	878	0.00	5
Real residential property prices relative gap, nrw	0.69	0.013	0.71	16.25	-7.75	-9.15	0.45	0.93	0.95	0.13	0.76	0.79	0.16	0.05	0.02	348	1349	0.33	6
Real residential property prices gap, rw60	0.68	0.013	0.72	9.86	-3.56	-7.96	0.45	0.91	1.00	0.18	0.73	0.89	0.14	0.05	0.03	257	878	0.20	5
Residential property prices gap, nrw	0.68	0.013	0.68	6.27	1.95	-15.52	0.50	0.68	1.00	0.14	0.38	1.00	0.18	0.06	0.00	363	1398	0.17	6
House price-to-rent ratio gap, nrw	0.66	0.013	0.69	17.64	-5.12	-8.70	0.40	0.93	0.96	0.16	0.72	0.78	0.12	0.07	0.04	231	818	0.40	5
Residential property prices relative gap, nrw	0.66	0.013	0.68	15.64	-9.13	-9.13	0.46	0.98	0.98	0.14	0.87	0.87	0.16	0.04	0.02	363	1398	0.17	6
y-o-y growth rate (%) of commercial property prices	0.66	0.013	0.82	4.57	0.08	0.08	0.75	0.95	0.95	0.51	0.73	0.73	0.12	0.08	0.05	199	666	0.33	6
House price-to-income ratio relative gap, rw60	0.66	0.013	0.82	3.71	-3.16	-6.22	0.78	0.94	0.97	0.54	0.72	0.78	0.12	0.08	0.04	187	707	0.20	5
y-o-y growth of real commercial property prices	0.65	0.013	0.92	-0.85	-0.85	-2.12	0.91	0.91	0.94	0.65	0.65	0.72	0.13	0.08	0.04	199	666	0.17	6
Real residential property prices relative gap, rw60	0.65	0.013	0.81	5.34	-5.91	-7.30	0.73	0.96	0.98	0.51	0.74	0.77	0.11	0.08	0.05	257	878	0.40	5
Annual absolute change (pp) in house price-to-income ratio	0.64	0.013	0.67	6.41	-4.15	-5.74	0.42	0.93	0.96	0.19	0.79	0.86	0.11	0.04	0.02	361	1541	0.29	7
Commercial real estate prices gap, nrw	0.63	0.013	0.67	8.49	7.56	-12.35	0.52	0.55	1.00	0.23	0.27	0.98	0.15	0.03	0.01	146	362	0.00	4
y-o-y growth of real residential property prices	0.63	0.013	0.66	7.70	-2.92	-8.29	0.46	0.92	0.98	0.20	0.80	0.94	0.13	0.03	0.01	395	1687	0.50	8
Commercial property prices gap, nrw	0.63	0.013	0.67	9.38	8.97	-12.18	0.54	0.55	1.00	0.23	0.24	0.97	0.16	0.03	0.01	146	362	0.00	4
House price-to-rent ratio relative gap, rw60	0.62	0.013	0.98	-6.69	-6.69	-7.18	0.97	0.97	0.98	0.73	0.73	0.73	0.12	0.09	0.06	231	818	0.40	5
Commercial property prices relative gap, nrw	0.62	0.013	0.66	7.90	-8.96	-10.90	0.52	0.97	1.00	0.26	0.90	0.96	0.13	0.02	0.01	146	362	0.00	4

Notes for this Table and naming conventions for variables are described at the end of Annex C.

Notes for Table C1 to Table C3

The first and second columns show the AUROC and its standard deviation, respectively. An entry is marked red if the AUROC is below 0.6. The third column shows the psAUROC. The next four clusters, each containing three columns show the optimal threshold for signalling a crisis (4-6), true positive rate (7-9), false positive rate (10-12) and usefulness measure (13-15) (Alessi and Detken, 2011). Numbers in column headings indicate the preferences of the policy-maker. The last cluster of columns (16-19) provides information about the sample and robustness of results. Column 18 shows the percentage of Member States for which the AUROC at the optimal pooled threshold is insignificant using only the crisis data for the country in question. Entries have been highlighted red if – assuming the policy-maker's preferences are balanced – the true positive rate is smaller than 0.5, the false positive rate is greater than 0.5, the usefulness measure is less than 0.1 or the percentage of Member States for which the AUROC is not significant is greater than 33%. Indicators are highlighted in red if they exhibit one column indicated in red.

In accordance with the explanation in Section 3.1.1. a **gap** is calculated as the difference between the actual value of the variable and its long-term trend. The gap is always expressed in units of the underlying variable (percentage points for ratios and index points for index variables).

The **relative gap** is used in line with the explanation in Section 3.1.2.2. It is calculated as a percentage deviation of the actual value from its trend ((actual value/trend-1)*100), rather than the absolute difference between the two.

Unless specified otherwise all trends have been calculated using **lambda** 400,000.

The abbreviation **nrw** indicates that no rolling window has been used, i.e. at each point in time all observations available up to the given data have been used to calculate the trend.

The abbreviation **rw60** indicates that a rolling window of 60 quarters has been used, i.e. at each point in time observations from the past 60 quarters have been used to calculate the trend

The term **inverted** means that the raw variable has been multiplied by -1. The reported results apply to the transformed variable.

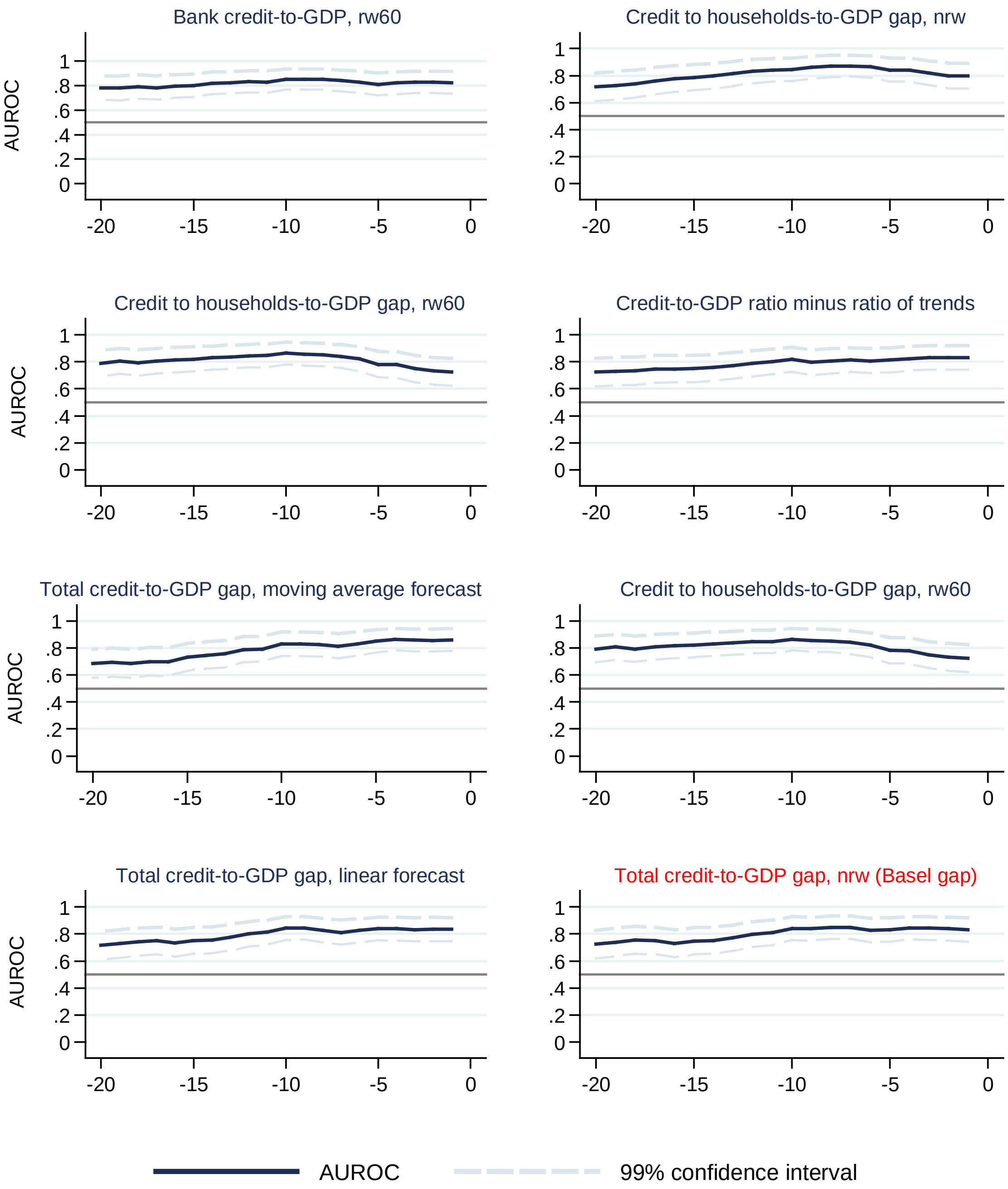
The terms **deflated** or **real** mean that the raw variable has been divided by the CPI index.

The **year-on-year growth rate** for a variable X has been calculated as $(X_t/X_{t-4} - 1)*100$.

The **annual absolute change** for a variable X has been calculated as $(X_t - X_{t-4})$.

Annex D: Results for the dynamic analysis

Table D1 – Results for credit-to-GDP gaps



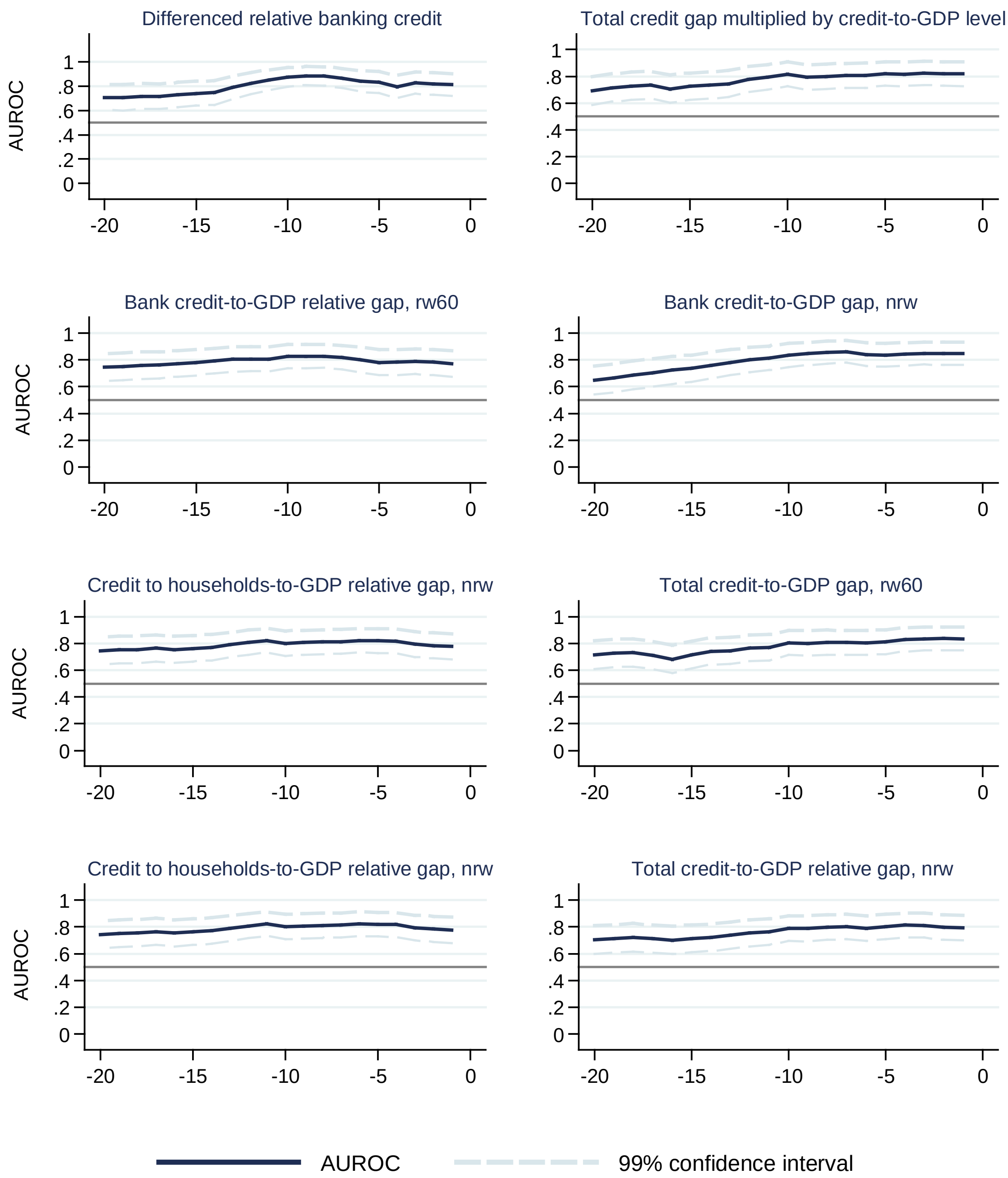

Differenced relative banking credit
Total credit gap multiplied by credit-to-GDP level
Bank credit-to-GDP relative gap, rw60
Bank credit-to-GDP gap, nrw
Credit to households-to-GDP relative gap, nrw
Total credit-to-GDP gap, rw60
Credit to households-to-GDP relative gap, nrw
Total credit-to-GDP relative gap, nrw
AUROC
1
.8
.6
.4
.2
0
-20
-15
-10
-5
0
AUROC
99% confidence interval

Table D2 – Results for credit growth variables

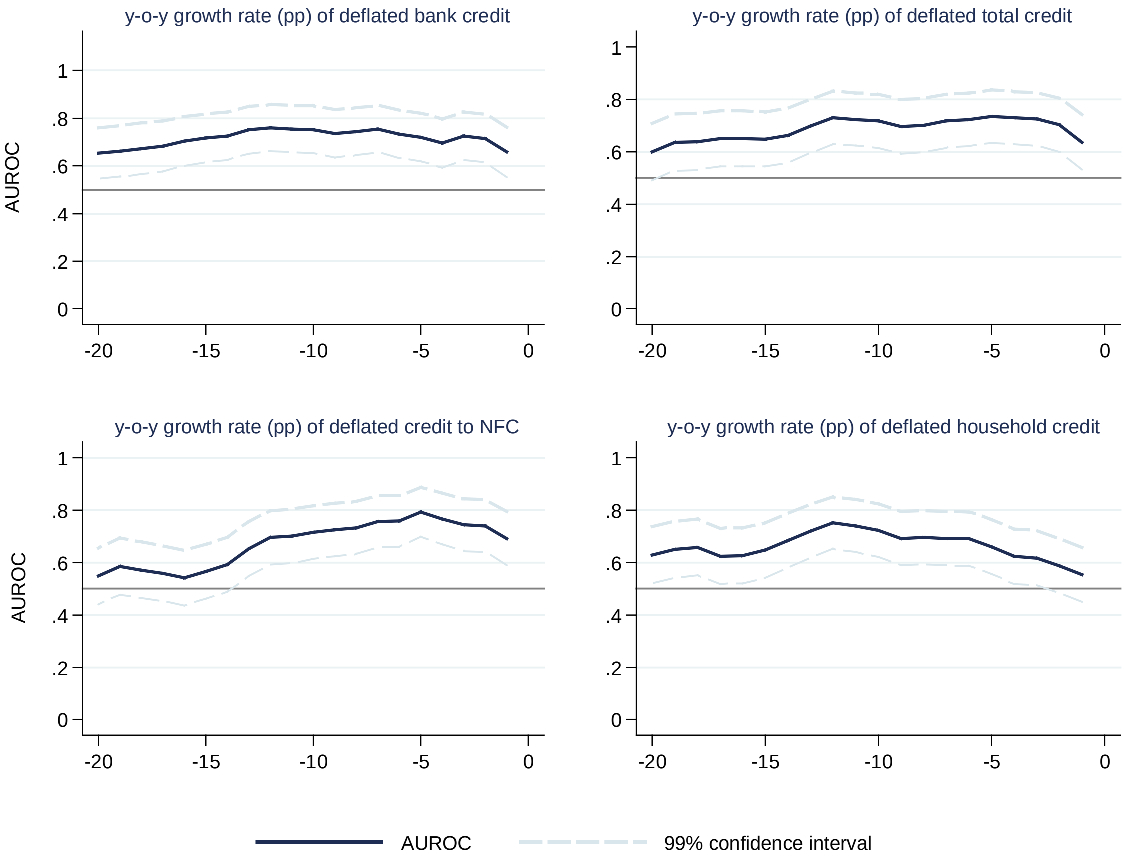


Table D3 – Results for market variables

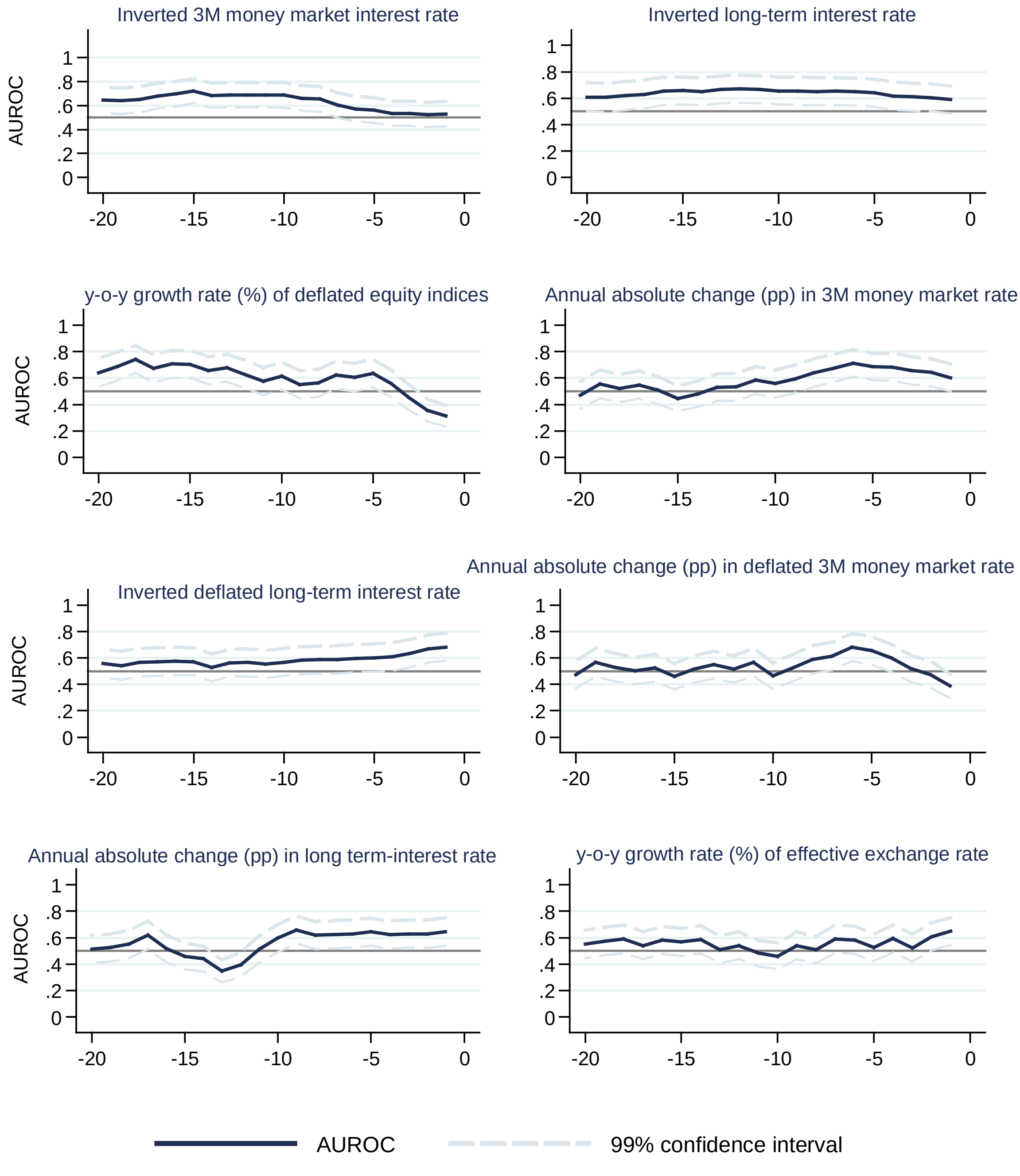
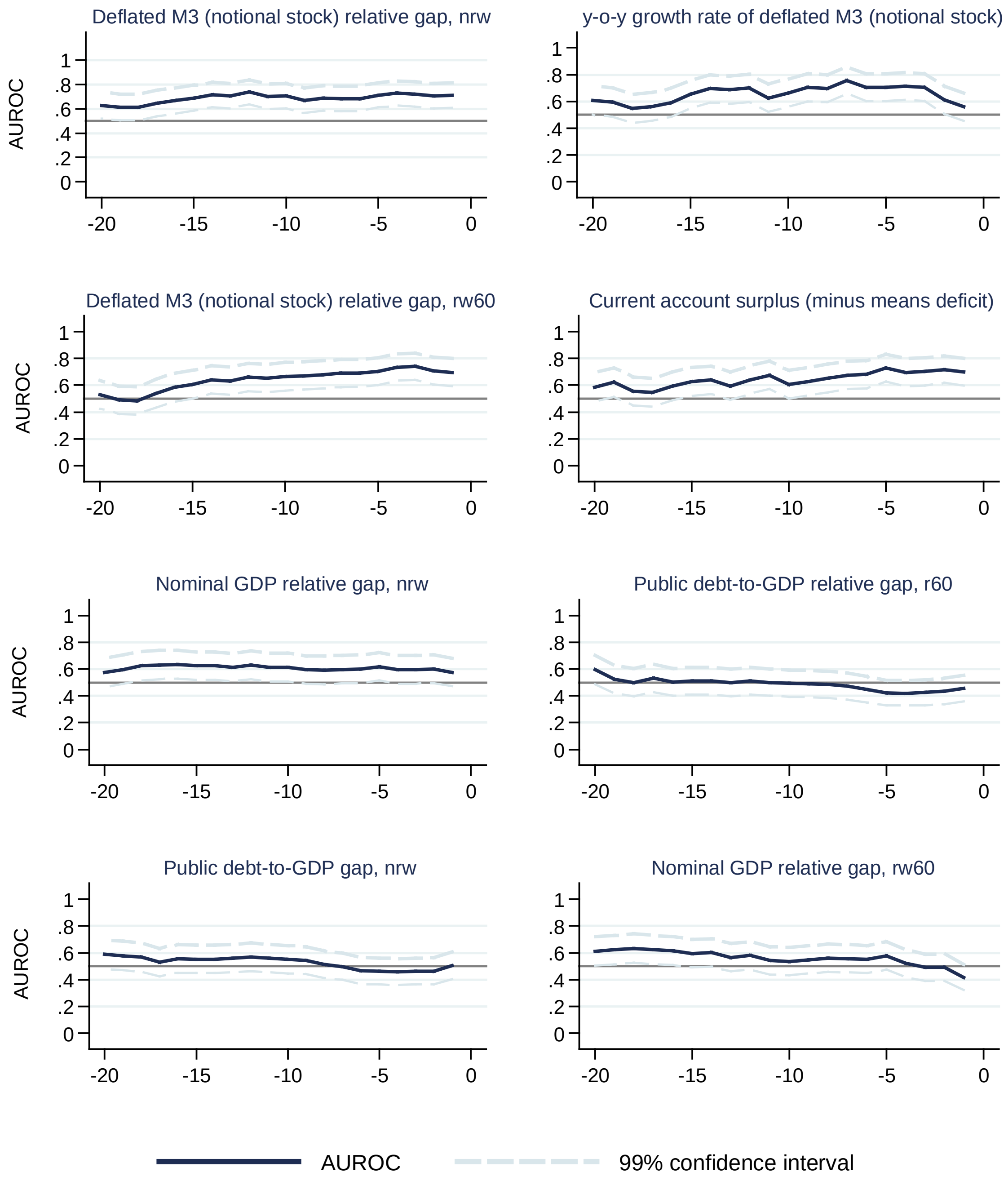


Table D4 – Results for macro variables



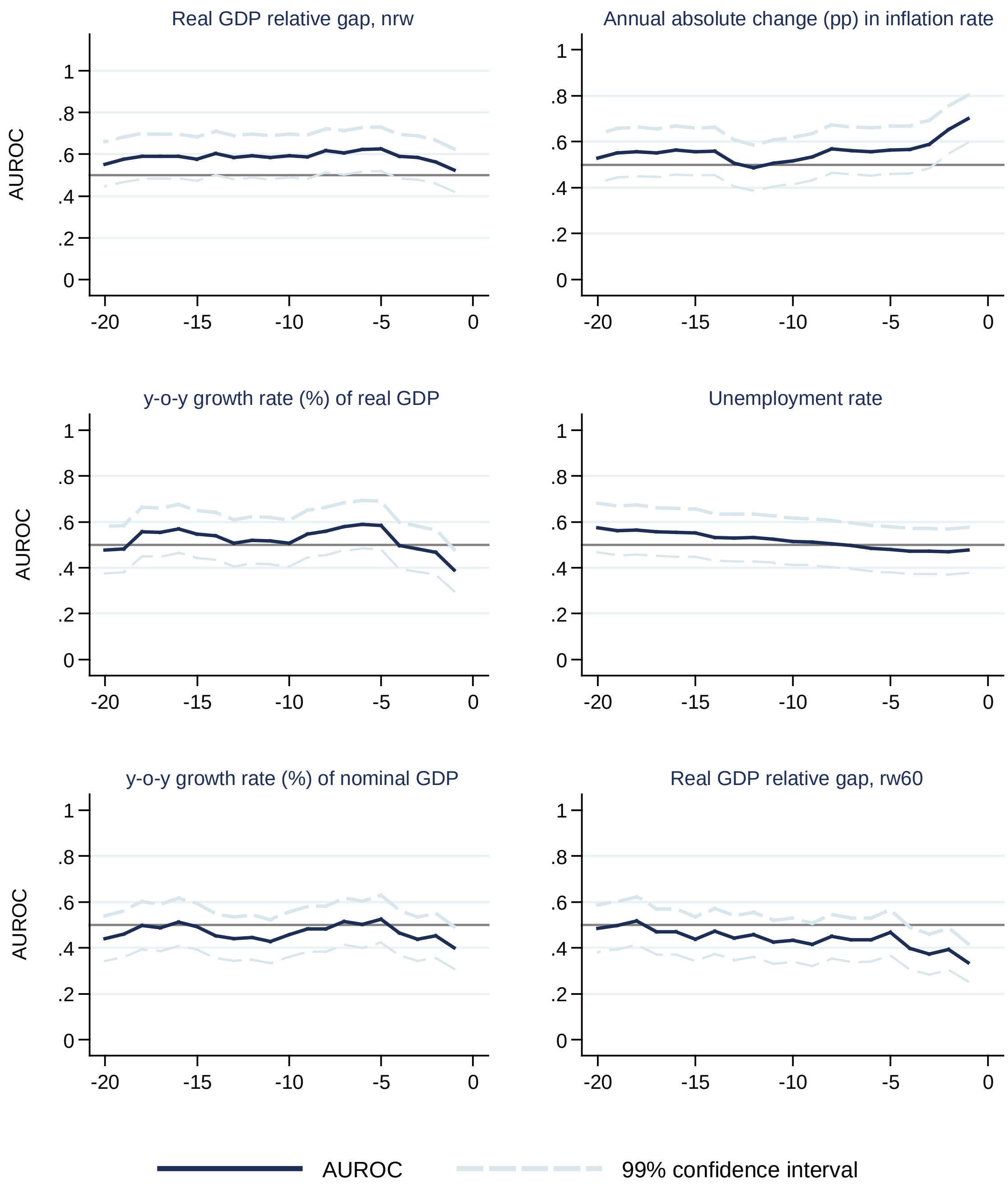

Real GDP relative gap, nrw
Annual absolute change (pp) in inflation rate
y-o-y growth rate (%) of real GDP
Unemployment rate
y-o-y growth rate (%) of nominal GDP
Real GDP relative gap, rw60
AUROC
AUROC
99% confidence interval

Table D5 – Results for debt service ratios

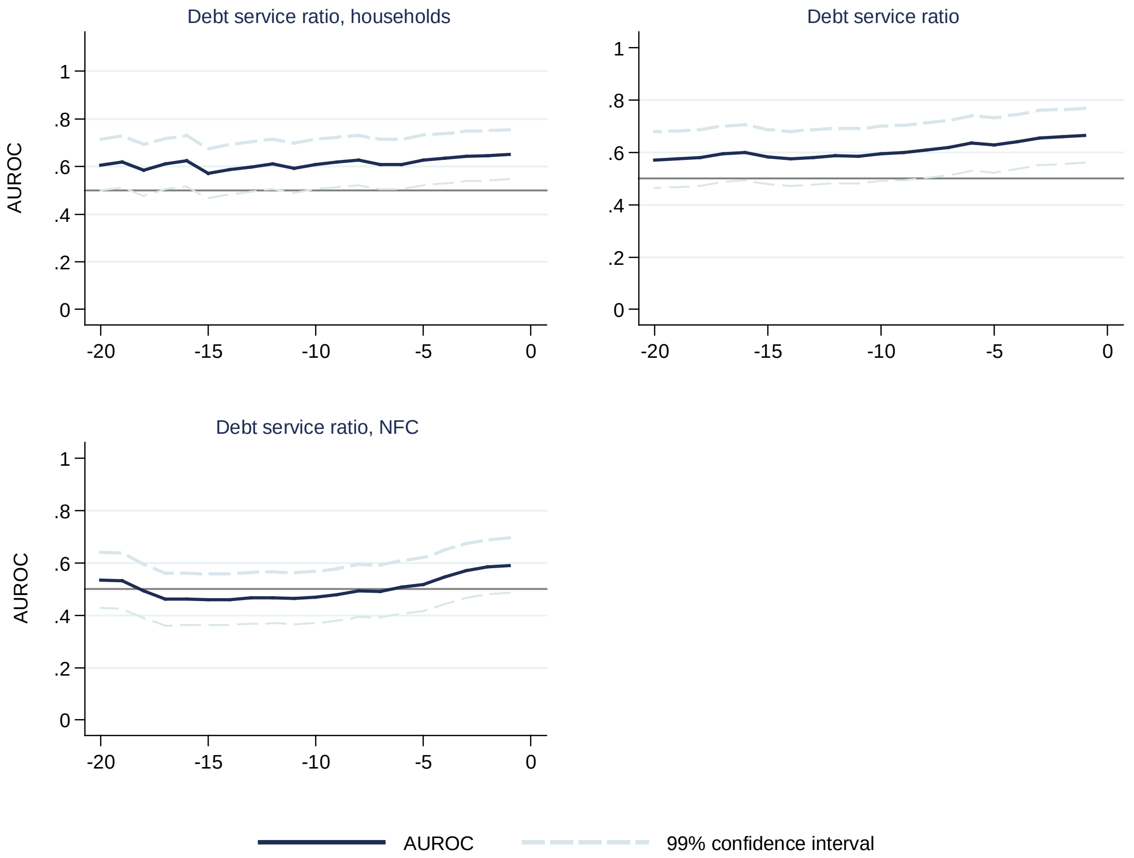
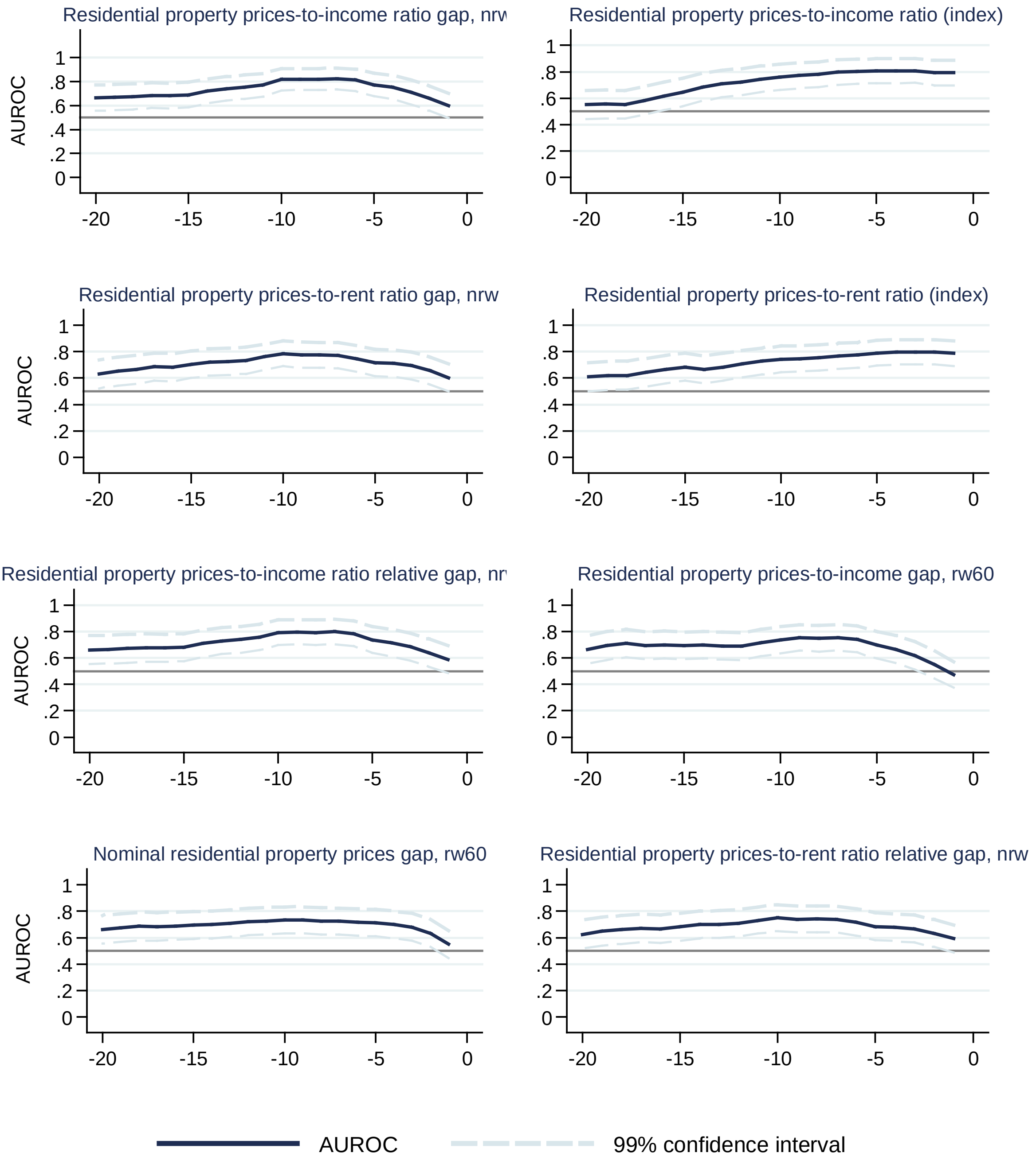
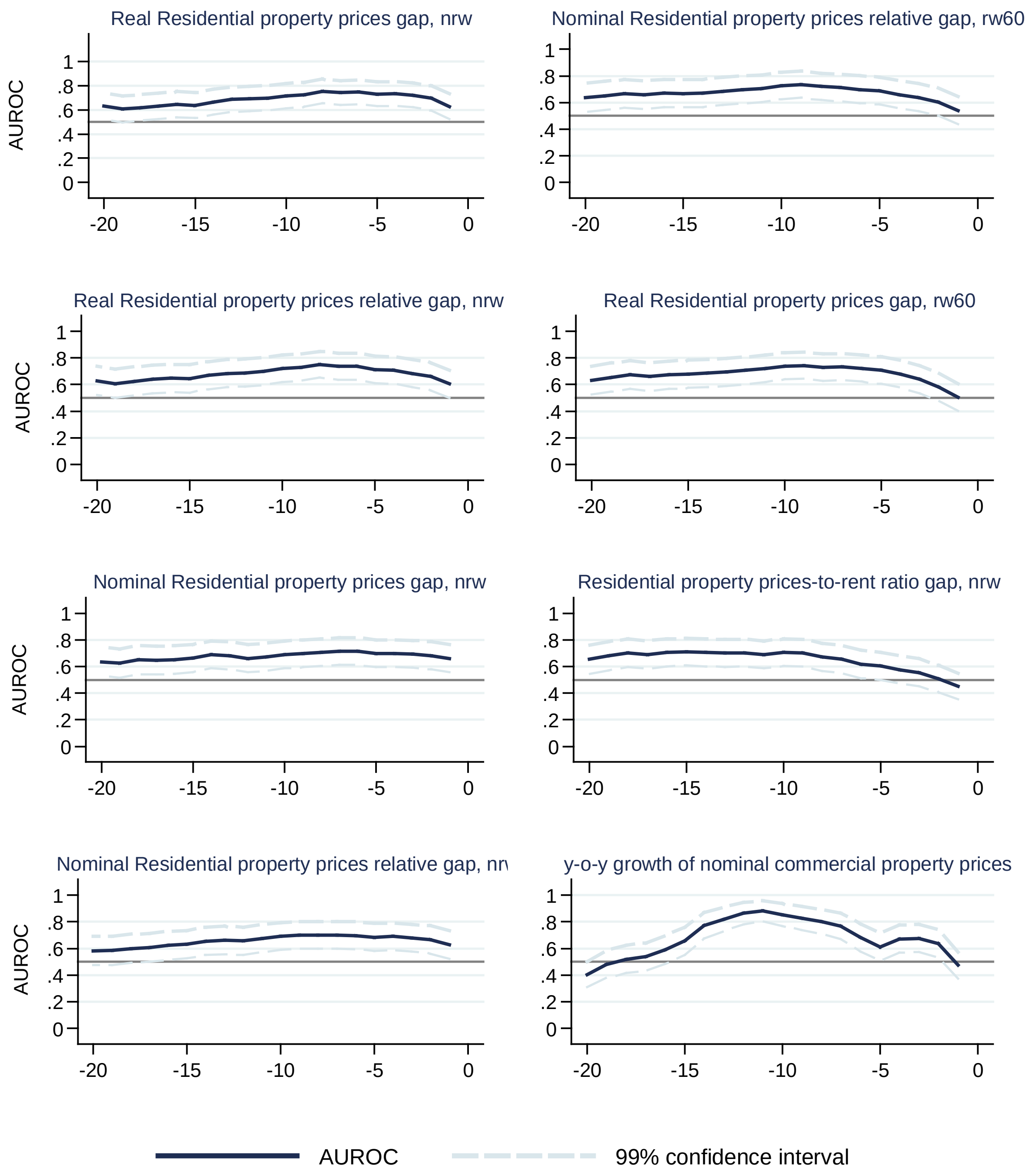
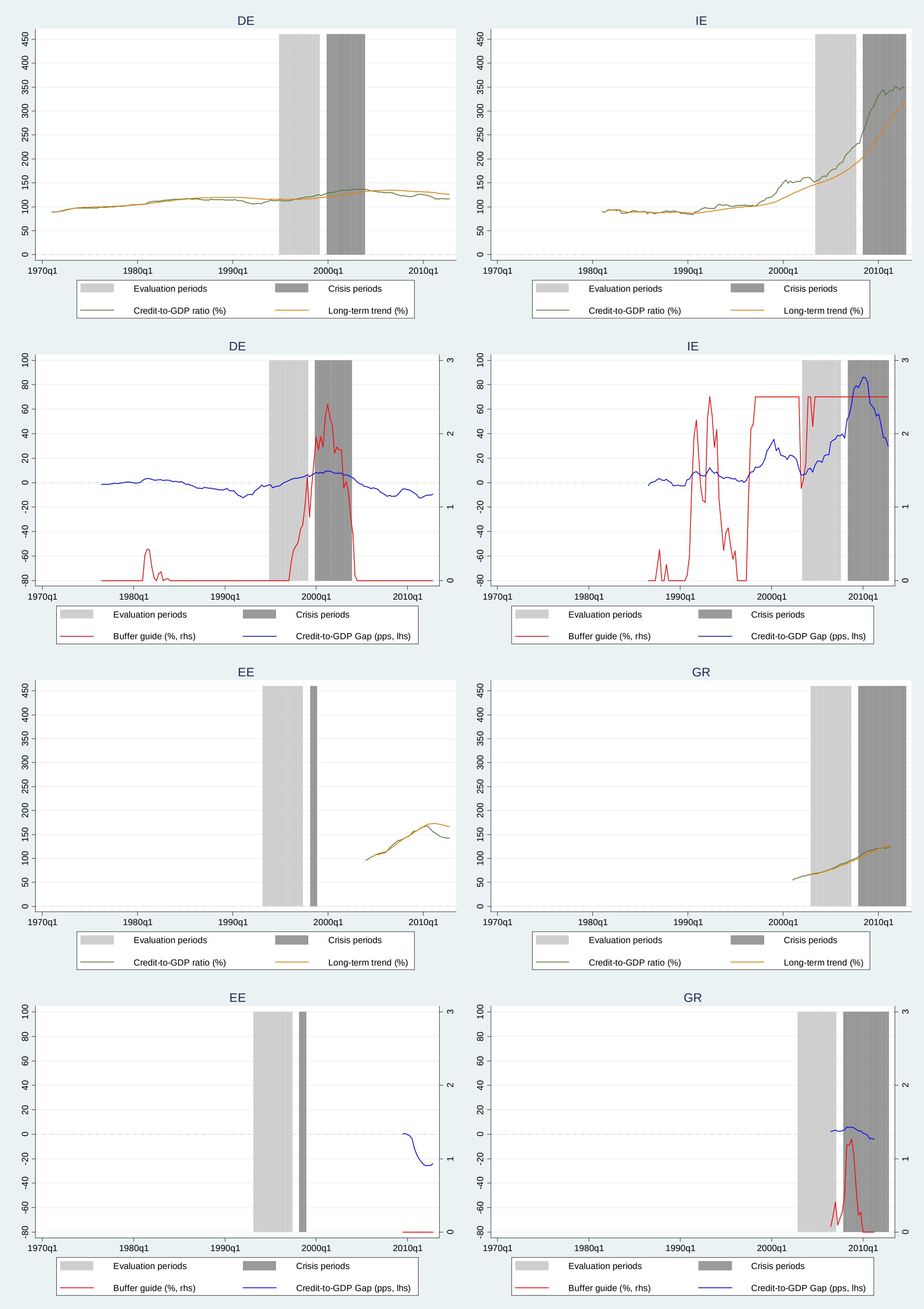


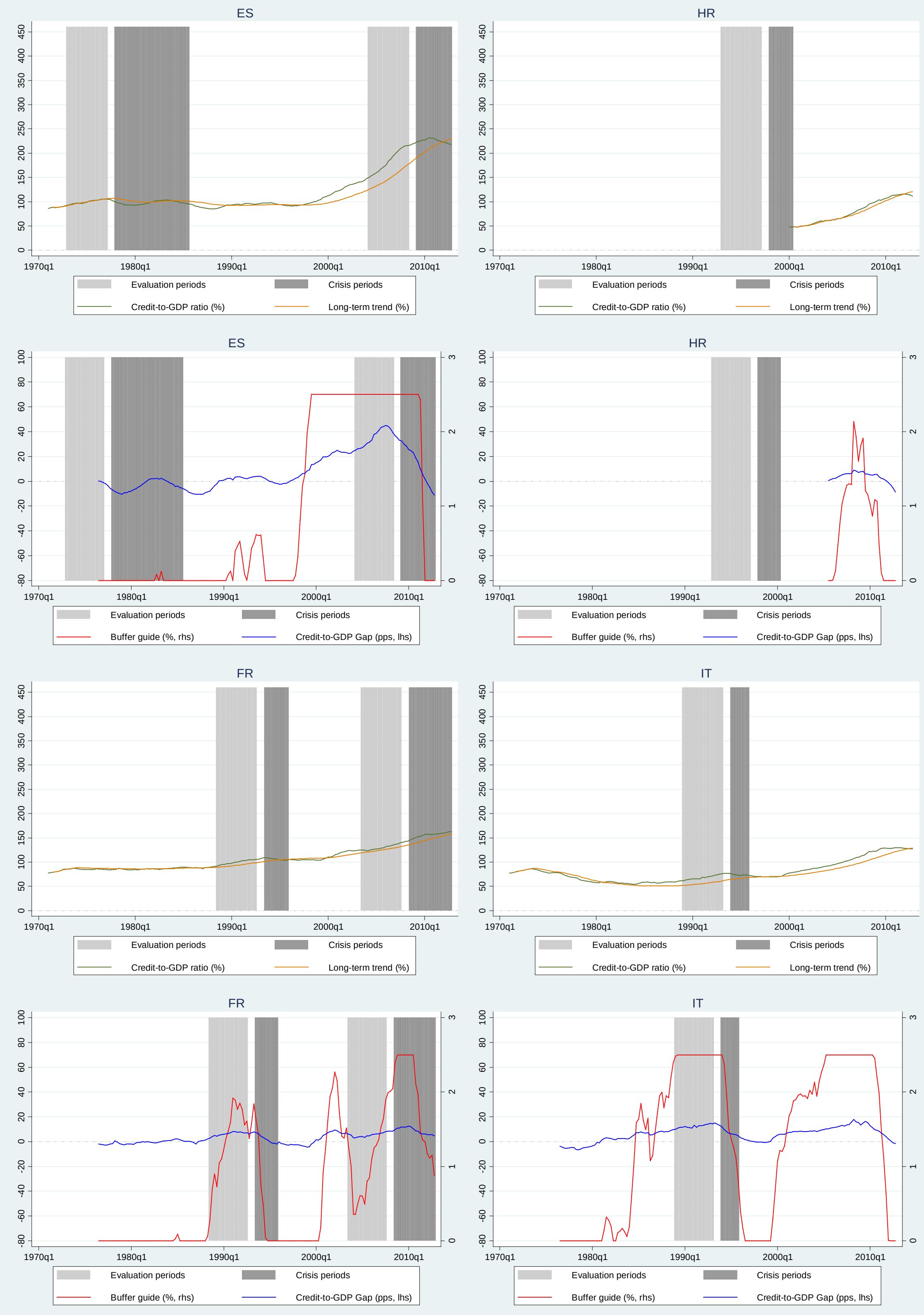
Table D6 – Results for property variables



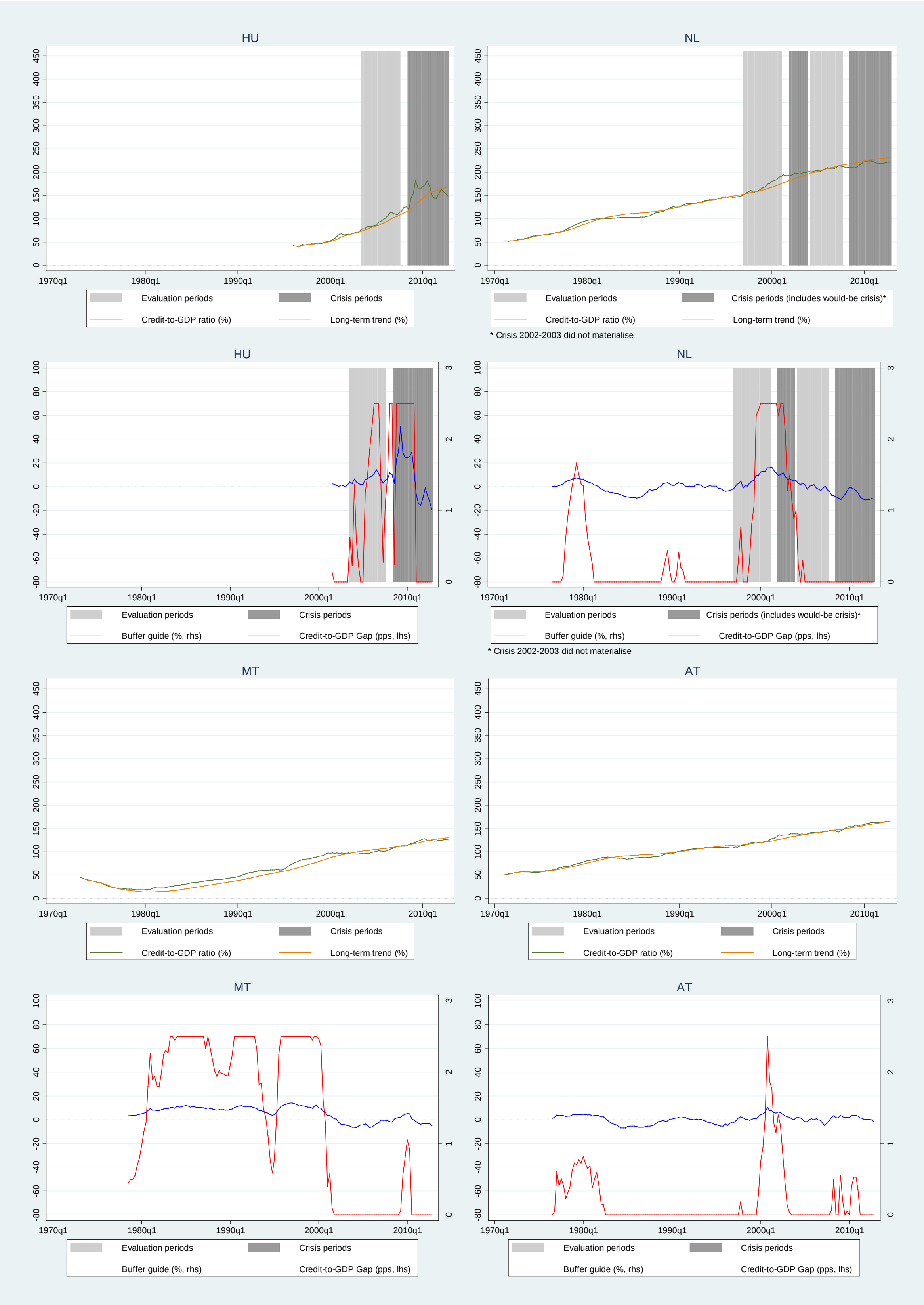

Real Residential property prices gap, nrw
Nominal Residential property prices relative gap, rw60
Real Residential property prices relative gap, nrw
Real Residential property prices gap, rw60
Nominal Residential property prices gap, nrw
Residential property prices-to-rent ratio gap, nrw
Nominal Residential property prices relative gap, nrw
y-o-y growth of nominal commercial property prices
AUROC
1
.8
.6
.4
.2
0
-20
-15
-10
-5
0
AUROC
99% confidence interval

Annex E: Credit-to-GDP ratios, gaps and buffer guides

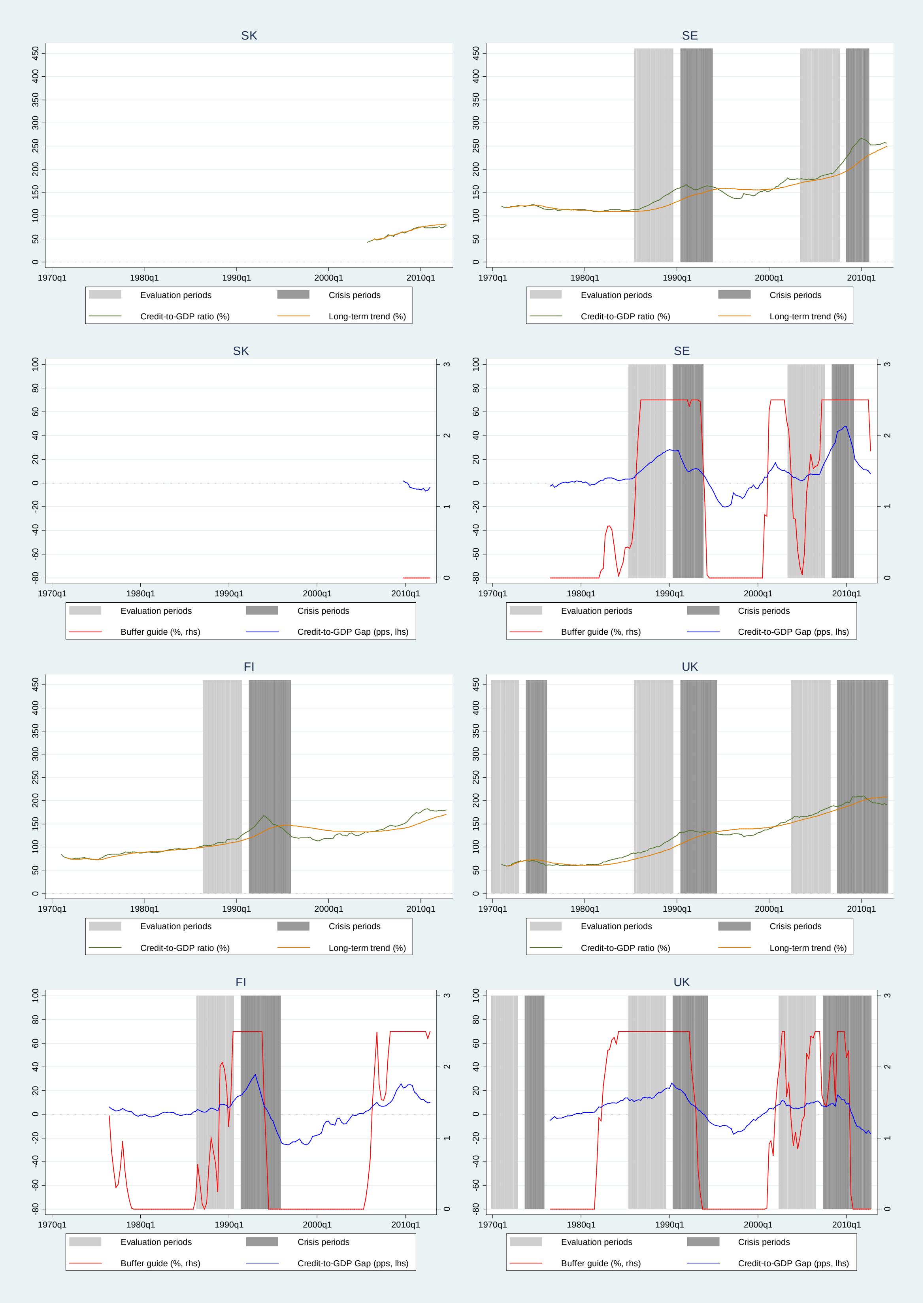

DE
IE
1970q1
1980q1
1990q1
2000q1
2010q1
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
EE
GR


ES
HR
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
ES
HR
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
FR
IT
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
FR
IT
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
1970q1
1980q1
1990q1
2000q1
2010q1

CY
LT
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
CY
LT
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
LV
LU
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
LV
LU
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
1970q1
1980q1
1990q1
2000q1
2010q1


HU
NL
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
Crisis periods (includes would-be crisis)*
* Crisis 2002-2003 did not materialise
HU
NL
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
Crisis periods (includes would-be crisis)*
* Crisis 2002-2003 did not materialise
MT
AT
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
MT
AT
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)


PL
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
RO
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
PL
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
RO
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
PT
Evaluation periods
Crisis periods (includes would-be crisis)*
Credit-to-GDP ratio (%)
Long-term trend (%)
* Crisis 1999-2000 did not materialise
SI
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
PT
Evaluation periods
Crisis periods (includes would-be crisis)*
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
* Crisis 1999-2000 did not materialise
SI
Evaluation periods
Crisis periods
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)


SK
SE
Evaluation periods
Crisis periods
Credit-to-GDP ratio (%)
Long-term trend (%)
Buffer guide (%, rhs)
Credit-to-GDP Gap (pps, lhs)
FI
UK
1970q1
1980q1
1990q1
2000q1
2010q1

Annex F: Multivariate logit model – robustness tests

In order to assess the extent to which multivariate models can improve crisis prediction power compared with the best univariate model, two additional tests are carried out. The tests are based on samples which cover, for each country, the same time span for all explanatory variables – i.e. in contrast to the previous analysis they are based on a balanced panel.[24]

The first test preserves a relatively large time span and country coverage. Some housing variables, which are on average only available for the last 20 years across all countries, are thus not considered. Instead, four variables from the ranking in Table 5 are chosen to be included in the estimated models: the bank credit-to-GDP gap, equity price growth, the debt service ratio and the change in the house price-to-income ratio. The resulting sample covers 20 countries and an average time span of 14 years. Table F1 reports the results. While the best multivariate models, i.e. models (11), (13) and (15), have a higher AUROC than the best univariate model, i.e. model (1), the difference is not statistically significant at the 5% level.

[24] A panel is said to be balanced if there are the same time periods for each cross-section observation. Here, this only holds for a given country, but this paper still refers to a balanced panel.

Table F1 – Results of logit model estimations – balanced panel

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Bank credit-to-GDP gap, nrw	0.142***				0.149***	0.146***		0.128***			0.153***	0.135***	0.133***		0.140***
	(0.018)				(0.016)	(0.018)		(0.017)			(0.016)	(0.015)	(0.017)		(0.015)
Year-on-year growth rate of equity prices		0.010***			0.016***		0.010***		0.010***		0.016***	0.015***		0.010***	0.016***
		(0.003)			(0.003)		(0.003)		(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
Debt service ratio			1.384***			2.223***	1.432***			1.603***	2.384***		2.394***	1.665***	2.575***
			(0.233)			(0.245)	(0.244)			(0.264)	(0.275)		(0.251)	(0.276)	(0.275)
Year-on-year change in house price-to-income ratio				0.074***				0.046***	0.074***	0.076***		0.045***	0.050***	0.076***	0.050***
				(0.010)				(0.010)	(0.010)	(0.010)		(0.010)	(0.010)	(0.010)	(0.010)
Constant	-1.709***	-1.179***	-1.327***	-1.307***	-1.972***	-2.176***	-1.485***	-1.817***	-1.450***	-1.659***	-2.479***	-2.080***	-2.340***	-1.823***	-2.653***
	(0.104)	(0.126)	(0.107)	(0.119)	(0.111)	(0.093)	(0.113)	(0.105)	(0.124)	(0.114)	(0.106)	(0.120)	(0.101)	(0.126)	(0.128)
Observations	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198	1,198
Crises	314	314	314	314	314	314	314	314	314	314	314	314	314	314	314
Pseudo R-Squared	0.185	0.013	0.011	0.057	0.209	0.204	0.024	0.203	0.069	0.071	0.229	0.227	0.225	0.083	0.250
AUROC	0.816	0.600	0.609	0.668	0.821	0.828	0.626	0.826	0.702	0.682	0.832	0.829	0.830	0.710	0.837
std(AUROC)	0.013	0.018	0.020	0.019	0.013	0.012	0.018	0.012	0.017	0.018	0.012	0.012	0.012	0.017	0.013
True positive rate	0.841	0.780	0.506	0.516	0.764	0.898	0.688	0.892	0.640	0.596	0.796	0.882	0.783	0.688	0.869
False positive rate	0.337	0.550	0.213	0.230	0.234	0.387	0.455	0.352	0.308	0.290	0.279	0.348	0.281	0.353	0.331
Loss	0.248	0.385	0.353	0.357	0.235	0.244	0.383	0.230	0.334	0.347	0.242	0.233	0.249	0.333	0.231
Usefulness	0.252	0.115	0.147	0.143	0.265	0.256	0.117	0.270	0.166	0.153	0.258	0.267	0.251	0.167	0.269

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The table reports the results from a pooled logit regression. The dependent variable is set to one for the period twenty to four quarters prior to the start of a banking crisis in the respective country. Robust standard errors adjusted for clustering at the quarterly level are reported in parentheses. Coefficients marked with ***, **, * are significant at the 1%, 5% and 10% level, respectively. The true positive rate, false positive rate, loss and usefulness measures reported correspond to balanced preferences between missing crises and issuing false alarms, i.e. the preference parameter was set to 0.5.

The second test covers all 15 indicator variables and aims to determine the best two, three and four-variable models. As this includes the four housing variables, the sample covers only 12 countries and an average time span of 6 years. The results are thus largely driven by the global financial crisis. Table F2 reports the results on the models' inclusion of indicator variables as in Table 5 but based on a balanced panel. The results are qualitatively similar: the bank credit-to-GDP gap, equity price growth, the change in the house price-to-disposable income ratio and the current account-to-GDP ratio are again among those variables most often included in the estimated models. The estimated coefficient of the debt service ratio, for this particular balanced panel, has a higher standard error and is hence less often included in the estimated models compared with the unbalanced panel results reported in Table 5. Once the debt service ratio is included in a model it increases the AUROC.

Table F2 – Models' inclusion of indicator variables – balanced panel

Variable	Included in % of models	Average AUROC	Average # of RHS variables	Average coefficient estimate
Annual absolute change in house price-to-income ratio	99.6%	0.87	5.2	0.19
Year-on-year growth rate of equity prices	99.1%	0.87	5.4	0.02
Bank credit-to-GDP gap, nrw	98.2%	0.88	5.1	0.14
Year-on-year growth rate of real GDP	94.2%	0.88	5.6	* -0.38
Public debt-to-GDP ratio	94.1%	0.88	5.4	* -0.04
Current account-to-GDP ratio	91.8%	0.87	5.7	-0.13
Year-on-year growth rate of real commercial property prices	81.7%	0.87	5.6	0.05
Total credit-to-GDP gap, nrw, (Basel gap)	57.3%	0.86	5.4	0.00
Year-on-year growth rate of real residential property prices	55.1%	0.86	5.2	0.05
Year-on-year growth rate of real bank credit	54.2%	0.86	5.6	0.11
Year-on-year growth rate of real total credit	46.8%	0.86	5.9	* -0.12
Year-on-year growth rate of real M3	36.4%	0.84	5.5	0.07
Real three-month money market interest rate	32.3%	0.86	6.0	* 0.28
Annual absolute change of house price-to-rent ratio	21.0%	0.86	5.6	0.04
Debt service ratio	20.7%	0.90	6.3	9.33

Note: The second column reports the share of models in which a given variable was included. Only coefficients significant at the 5% level were considered based on HAC standard errors. For robustness, the Bonferroni adjustment was used to account for a potential alpha-inflation. The results, which are available upon request, are qualitatively similar. Due to the more conservative significance level the average model contains only two variables. The third and fourth columns report the average AUROC and average number of right-hand-side variables, respectively, of all models in which a given variable was included. The fifth column reports the average coefficient estimate of all models in which a given variable was included. Average coefficients marked with an asterisk have a sign which differs from the expectations.

Based on these results it is possible to identify the best *K*-variable models. The results, reported in Table F3, include one to four-variable models with the highest AUROC and variables whose coefficients have the expected signs. The models include the bank credit-to-GDP gap, equity price growth, the debt service ratio and the change in the house price-to-disposable income ratio. The last column reports the best four-variable model using a measure of broad credit.

Table F3 – One to four-variable models with the highest AUROC (including broad and bank credit) – balanced panel

VARIABLES	(1)	(2)	(3)	(4)	(5)
Total credit-to-GDP gap, nrw, (Basel gap)					0.056***
					(0.014)
Bank credit-to-GDP gap, nrw	0.123***	0.132***	0.139***	0.127***	
	(0.013)	(0.012)	(0.013)	(0.013)	
Debt service ratio		8.580**	8.976*	10.590	9.787
		(4.318)	(4.837)	(6.564)	(6.297)
Year-on-year growth rate of equity prices			0.021***	0.027***	0.029***
			(0.005)	(0.006)	(0.007)
Annual absolute change in house price-to-income ratio				0.136***	0.132***
				(0.018)	(0.019)
Constant	-1.221***	-3.106***	-3.451***	-4.519***	-4.057***
	(0.171)	(0.782)	(0.873)	(1.247)	(1.104)
Observations	380	380	380	380	380
Crises	154	154	154	154	154
Pseudo R-Squared	0.189	0.299	0.327	0.391	0.293
AUROC	0.774	0.854	0.865	0.900	0.855
std(AUROC)	0.023	0.019	0.018	0.015	0.019
True positive rate	0.734	0.883	0.844	0.844	0.779
False positive rate	0.301	0.279	0.279	0.190	0.217
Loss	0.284	0.198	0.217	0.173	0.219
Usefulness	0.216	0.302	0.283	0.327	0.281

Notes: The table reports the results from a pooled logit regression. The dependent variable is set to one for the period twenty to four quarters prior to the start of a banking crisis in the respective country. Robust standard errors adjusted for clustering at the quarterly level are reported in parentheses. Coefficients marked with ***, ** and * are significant at the 1%, 5% and 10% levels, respectively. The true positive rate, false positive rate, loss and usefulness measures reported correspond to a policy-maker's preferences of 0.5.

Annex G: List of members of the Expert Group

This paper has benefited from the input of the entire team which is listed below.†

Mr	Carsten Detken (Chair)	European Central Bank
Mr	Olaf Weeken (Secretary)	ESRB Secretariat
Ms	Diana Bonfim/ Mr Miguel Medonca Boucinha*	Banco de Portugal
Mr	Peter Brun	Danmarks Nationalbank
Mr	Maciej Brzozowski	Narodowy Bank Polski
Mr	Christian E. Castro	Banco de España
Mr	Saša Cerovac	Hrvatska narodna banka
Mr	Conn Creedon / Mr Eoin O'Brien*	Central Bank of Ireland
Mr	Stijn Ferrari	Nationale Bank van België
Ms	Julia Giese / Mr Oliver Bush*	Bank of England
Mr	Gaston Andres Giordana	Banque centrale du Luxembourg
Mr	Karsten Gerdrup**	Norges Bank
Ms	Matilda Gjirja	Finansinspektionen
Mr	Philipp Hochreiter	Austrian Financial Market Authority (FMA)
Ms	Anna Jernova	Bank of England, (PRA)
Mr	Jan Kakes	De Nederlandsche Bank
Mr	Karlo Kauko	Suomen Pankki
Ms	Anna Kelber / Mr Benjamin Klaus***	Banque de France
Ms	Claire Labonne	Autorité de Contrôle Prudentiel
Mr	Massimo Libertucci	Banca d'Italia
Mr	Stan Maes	European Commission
Mr	Ola Melander	Sveriges Riksbank
Mr	Florian Neagu	Banca Naţională a României
Mr	Benjamin Neudorfer	Oesterreichische Nationalbank
Mr	Peer Osthoff	BaFin
Ms	Evangelia Rentzou	ESRB Secretariat
Mr	Štefan Rychtárik	Národná banka Slovenska
Ms	Sofia Savvidou	Bank of Greece
Mr	Jakub Seidler / Mr Miroslav Plašil*	Česká národní banka
Ms	Ingrid Stein / Ms Natalia Puzanova	Deutsche Bundesbank
Mr	Wolfgang Strohbach	European Banking Authority
Mr	Balázs Világi	Magyar Nemzeti Bank
Ms	Ianna Yordanova / Mr Emilio Hellmers*	Finanstilsynet
Mr	Balázs Zsamboki	European Central Bank

*Alternate member, ** Observer; ***ECB since October 2013 (see †)

† The Expert Group also benefited from contributions by Nuno Monteiro (Bank of Portugal) [programming], Mr Tuomas Peltonen (ECB) [containing crises probabilities], Mr Willem Schudel (ECB Graduate Programme) [containing crises probabilities and initial dataset] and Mr Thorsten Wezel (ECB/IMF) [unexplained residual losses].

References

Acharya, V.V., Pedersen, L.H., Philippon, T. and Richardson, M. (2012), "Measuring systemic risk", *CEPR discussion paper*, No. 8824.

Acharya, V. and Merrouche, O. (2013), "Precautionary Hoarding of Liquidity and Inter-Bank Markets: Evidence from the Sub-prime Crisis", *Review of Finance*, 17(1), pp. 107-160.

Adalid, R. and Detken, C. (2007), "Liquidity shocks and asset price boom/bust cycles", *ECB Working Paper*, No. 732.

Adrian, T. and Brunnermeier, M.K. (2011), "CoVaR", *NBER Working Paper*, No. 17454.

Agur, I. and Sharma, S. (2013), "Rules, discretion and macro-prudential policy", *IMF Working Paper*, No 13/65.

Aikman, D., Haldane, A.G. and Nelson, B. (2011), "Curbing the credit cycle", prepared for the Columbia University Centre on Capital and Society Annual Conference, New York, November 2010, available at:
http://www.bankofengland.co.uk/archive/Documents/historicpubs/speeches/2010/speech463.pdf

Anguren Martin, R. (2011), "Credit cycles: evidence based on a non linear model for developed countries", *Documentos de Trabajo del Banco de España*, No 1113.

Alessi, L. and Detken, C. (2011), "Real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity", *European Journal of Political Economy*, 27 (3), pp. 520-533.

Alessi, L. and Detken, C. (2014), "Identifying excessive credit growth and leverage", ECB Working Paper, forthcoming.

Arsov, I., Canetti, E., Kodres, L.E., Mitra, S. (2013), "Near-coincident' indicators of systemic stress", *IMF Working Paper*, No 13/115.

Babecky, J., Havranek, T., Mateju, J., Rusnak, M., Smidkova, K. and Vasicek, B. (2011), "Early warning indicators of economic crises: Evidence from a panel of 40 developed countries", *Czech National Bank Working Paper*, No 8.

Bank for International Settlements (2008), "Addressing Financial System Procyclicality: a Possible Framework", *Note for the FSF Working Group on Market and Institutional Resilience*, Basel, September.

Bank for International Settlements (2013), "Long series on credit to private non-financial sectors", available at http://www.bis.org/statistics/credtopriv.htm

Basel Committee on Banking Supervision (2010a), "An assessment of the long-term economic impact of stronger capital and liquidity requirements", available at
http://www.bis.org/publ/bcbs173.pdf

Basel Committee on Banking Supervision (2010b), “Guidance for national authorities operating the countercyclical capital buffer”, available at http://www.bis.org/publ/bcbs187.htm

Basel Committee on Banking Supervision (2011), “Basel III: A global regulatory framework for more resilient banks and banking systems”, December 2010, rev. June 2011, available at http://www.bis.org/publ/bcbs189.pdf

Barrell, R, Davis, E.P., Karim, D. and Liadze, I. (2010a), “Bank regulation, property prices and early warning systems for banking crises in OECD countries”, *NIESR Discussion Paper*, No 330.

Barrell R., Davis, E.P., Liadze, I. and Karim, D. (2010b), “Does the Current Account Balance Help to Predict Banking Crises in OECD Countries?”, NIESR and Brunel University, mimeo.

Behn, M., Detken, C., Peltonen, T. and Schudel, W. (2013a), “Setting countercyclical capital buffers based on early warning models: Would it work?”, *ECB Working Paper Series*, No 1604, available at http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1604.pdf

Behn, M., Groß, M., Peltonen, T. and Schudel, W. (2013b), “Calibrating countercyclical capital buffers with an integrated early warning GVAR model”, ECB, mimeo.

Berg, A., Borensztein, E. and Patillo, C. (2005), “Assessing early warning systems: How have they worked in practice?”, *IMF Staff Papers*, 52(3), pp. 462-502.

Berg, A. and Pattillo, C. (1999), “Are currency crises predictable? A test”, *IMF Staff Papers*, 46 (2), pp. 107-138.

Berge, T.J. and Jordà, Ò. (2011), “Evaluating the classification of economic activity into recessions and expansions”, *American Economic Journal: Macroeconomics*, 3 (2), pp. 246-277.

Betz, F., Opricǎ, S., Peltonen, T.A. and Sarlin, P. (2013), “Predicting distress in European banks”, *ECB Working Paper Series*, No 1597.

Blum, J. (1999), “Do capital adequacy requirements reduce risks in banking?”, *Journal of Banking and Finance*, 23 (5), pp. 755-771.

Bonfim, D. and Monteiro, N. (2013), “The implementation of the countercyclical capital buffer: rules versus discretion”, *Financial Stability Report*, Banco de Portugal, November 2013.

Borgy, V., Clerc, V. and Renne, J.-P. (2009), “Asset-price boom-bust cycles and credit: what is the scope of macroprudential regulation?”, *Working Paper Series*, No 263, Banque de France.

Borio, C. (2008), “The financial turmoil of 2007-?: a preliminary assessment and some policy considerations”, *BIS Working Papers*, No 251.

Borio, C., Furfine, C. and Lowe, P. (2001), “Procyclicality of the Financial System and Financial Stability: Issues and Policy Options,” *BIS Papers,* No 1.

Borio, C. (2012), “The financial cycle and macroeconomics: what have we learnt?”, *BIS Working Papers*, No 395.

Borio, C. and Drehmann, M. (2009), “Assessing the risk of banking crises – revisited”, *BIS Quarterly Review*, March, pp. 29-46.

Borio, C. and Lowe, P. (2002), “Assessing the risk of banking crises”, *BIS Quarterly Review*, pp. 43-54.

Brunnermeier, M. (2009), “Deciphering the Liquidity and Credit Crunch 2007–08,” *Journal of Economic Perspectives*, Vol. 23 (Winter), No 1, pp. 77-100.

Bussière, M. and Fratzscher, M. (2006), “Towards a new early warning system of financial crisis”, *Journal of International Money and Finance*, 25 (6), pp. 953-973.

Busch, U. (2012), “Credit cycles and business cycles in Germany: A comovement analysis”, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2015976

Bush, O, Guimaraes, R and Stremmel H, (forthcoming), “Beyond the credit gap: bank balance sheets, credit and the price of risk”, mimeo

Candelon, B., Dumitrescu, E. and Hurlin, C. (2012), “How to evaluate an early-warning system: Toward a unified statistical framework for assessing financial crises forecasting methods”, *IMF Economic Review*, 60 (1), pp. 75-113.

Caprio, G. and Klingebiel, D. (2003) “Episodes of Systemic and Borderline Financial Crises”, World Bank Research Dataset.

Ceron, J. and Suarez J. (2006). “Hot and Cold Housing Markets: International Evidence”, *CEMFI Working Paper*, No. 0603.

Choi, I. (2001), “Unit root tests for panel data”, *Journal of International Money and Finance*, 20 (2), pp. 249-272.

Claessens, S., Kose, M.A. and Terrones, M.E. (2011a), “Financial cycles: What? How? When?”, *IMF Working Paper*, No 11/76.

Claessens, S., Kose, M.A. and Terrones, M.E. (2011b), “How do business and financial cycles interact?”, *IMF Working Paper*, No 11/88.

Cogley, T. and Nason, J. (1995), “Effects of the HP Filter on Integrated Series”, *Journal of Economic Dynamics and Control*, 19, pp. 253-278.

Committee on the Global Financial System (2012), “Operationalizing the selection and application of macroprudential instruments”, *CGFS Publications*, No 48.

Cortes, C. and Mohri, M. (2005), “Confidence intervals for the area under the ROC curve”, in *Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference,* Vol. 17, MIT Press, p.305.

Davis, E.P. and Karim, D. (2008a), “Comparing Early Warning Systems for Banking Crises”, *Journal of Financial Stability*, 4 (2), pp. 89-120.

Davis, E.P. and Karim, D. (2008b), "Could Early Warning Systems Have Helped to Predict the Sub-Prime Crisis", *National Institute Economic Review*, 206, pp. 35-47.

DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988), "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach", *Biometrics*, pp. 837-845.

Dembiermont, C., Drehmann, M. and Muksakunratana, S. (2013), "How much does the private sector really borrow? A new database for total credit to the private non-financial sector", *BIS Quarterly Review*, March, available at http://www.bis.org/publ/qtrpdf/r_qt1303h.pdf

Demirgüç-Kunt, A. and Detragiache, E. (1998), "The determinants of banking crises: Evidence from industrial and developing countries", *IMF Staff Papers*, 45 (1), pp. 81-109.

Demirgüç-Kunt, A. and Detragiache, E. (1999), "Monitoring banking sector fragility: A multivariate logit approach with an application to the 1996-97 crisis", *World Bank Policy Research Working Paper*, No 2085.

Demirgüç-Kunt, A. and Detragiache, E. (2005), "Cross-Country Empirical Studies of Systemic Bank Distress: A Survey", *IMF Working Papers*, No 05/96.

Deutsche Bundesbank (2012), *Financial Stability Review*, pp. 87-88.

De Nederlandsche Bank (2010), *Quarterly Bulletin*, December, pp. 45-49.

Dorfman, D.D. and Alf, E. (1969), "Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data", *Journal of Mathematical Psychology*, *6*(3), pp. 487-496.

Drehmann, M., Borio, C., Gambacorta, L., Jiminez, G. and Trucharte, C. (2010), "Countercyclical capital buffers: exploring options", *BIS Working Papers*, No 317, available at http://www.bis.org/publ/work317.pdf

Drehmann, M., Borio, C. and Tsatsaronis, K. (2011), "Anchoring countercyclical capital buffers: the role of credit aggregates", *International Journal of Central Banking*, 7 (4), pp. 189-240.

Drehmann, M., Borio, C. and Tsatsaronis, K. (2012), "Characterising the financial cycle: don't lose sight of the medium term!", *BIS Working Papers*, No 380.

Drehmann, M. and Juselius, M. (2012), "Do debt service costs affect macroeconomic and financial stability?", *BIS Quarterly Review*, September, available at http://www.bis.org/publ/qtrpdf/r_qt1209e.pdf

Drehmann, M. and Juselius, M. (2014), "Evaluating early warning indicators of banking crises: Satisfying policy requirements", International Journal of Forecasting, 22(3), pp. 493-518.

Drehmann, M. and Tsatsaronis, K. (2014), "The credit-to-GDP gap and countercyclical capital buffers: questions and answers", *BIS Quarterly Review*, March, pp. 55-71.

Dufrénot, G., Klaus, B., Malik, S. and Vardoulakis, A.P. (2012), *Credit standards and financial institutions' leverage*, Banque de France, mimeo.

Edge, R. and Meisenzahl, R.R. (2011), "The unreliability of the credit-to-GDP ratio gaps in real-time: Implications for countercyclical capital buffers", *Finance and economics discussion series*, Federal Reserve Board, 2011–37.

Efron, B. and Tibshirani, R. (1993), *An introduction to the bootstrap,* CRC press.

Eickmeier, S., Gambacorta, L. and Hofmann, B. (2013), "Understanding global liquidity", *BIS Working Papers*, No 402.

Eidenberger, J., Neudorfer, B., Sigmund, M. and Stein, I. (2013), "Quantifying Financial Stability in Austria – new tools for Macroprudential Supervision", *Financial Stability Report 26*, Österreichische Nationalbank.

Federal Reserve Bank of St. Louis (2009), "What the Libor-OIS spread says", *Economic synopses*, No 24.

Frankel, J. and Rose, A. (1996). "Currency Crashes in Emerging Markets: An Empirical Treatment", *International Finance Discussion Paper*, 534, Board of Governors of the Federal Reserve System, Washington D.C.

Gerdrup, K., Kvinlog, A.B. and Schaanning, E. (2013), "Key indicators for a countercyclical capital buffer in Norway – Trends and uncertainty", *Staff Memo*, No 13, Norges Bank.

Geršl, A. and Seidler, J. (2011), "Credit growth and capital buffers: empirical evidence from Central and Eastern European Countries", *Research and Policy Notes*, 2011/02, Czech National Bank.

Giese, J., Andersen, H., Bush, O., Castro, C., Farag, F. and Kapadia, S. (2014), "The credit-to-GDP gap and complementary indicators for macroprudential policy: evidence from the UK", *International Journal of Finance Economics*, 19, pp. 25-47

Hanley, J.A. and McNeil, B. (1982), "The meaning and use of the Area under a Receiver Operating Characteristic (ROC) Curve", *Radiology*, 143 (1), pp. 29-36.

Hanley, J.A. and McNeil, B. (1983), "A method of comparing the Areas under a Receiver Operating Characteristic Curves derived from the same cases", *Radiology*, 148 (3), pp. 839-843.

Hardy, D. and Pazarbasioglu, C. (1998), "Leading Indicators of Banking Crises; Was Asia Different?", *IMF Working Paper*, 98/91.

Harvey, A.C. and Jaeger, A. (1993), "Detrending Stylized Facts and the Business Cycle", *Journal of Applied Econometrics*, 8, pp. 216-237.

Harmsen, M.P.P. (2010), "Basel III: Macroprudential regulation by means of countercyclical capital buffers", *Monetary Review*, 4th Quarter, Danmarks Nationalbank, pp. 97-114.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999), "Bayesian model averaging: A tutorial", *Statistical Science*, 14 (4), pp. 382-401.

Holló, D., Kremer, M. and Lo Duca, M. (2012), "CISS - a composite indicator of systemic stress in the financial system", *ECB Working Paper Series*, No 1426.

Jahn, N. (2014), "Timing the Release Phase of the Countercyclical Capital Buffer", mimeo.

Jiang, Y., Metz, C.E. and Nishikawa, R.M. (1996), "A receiver operating characteristic partial area index for highly sensitive diagnostic tests", *Radiology*, 201 (3), pp. 745-750.

Jordà, Ò. and Taylor, A.M. (2011), "Performance evaluation of zero net-investment strategies", *NBER Working Paper*, No 17150.

Jordà, Ò. (2012), "Discussion of 'Anchoring countercyclical capital buffers: The role of credit aggregates'", *International Journal of Central Banking*, 7 (4), pp. 241-259.

Juks, R. and Melander, O. (2012), "Countercyclical capital buffers as a macroprudential instrument", *Riksbank Studies*, December.

Kaminsky, G.L, Lizondo, S and Reinhart, C.M. (1998), "Leading indicators of currency crises", *IMF Staff Papers*, 45 (1), pp. 1-48.

Kaminsky, G.L. and Reinhart, C.M. (1999), "The twin crises: the causes of banking and balance-of-payments problems", *American Economic Review*, 89 (3), pp. 473-500.

Kauko, K. (2012a), "External deficits and non-performing loans in the recent financial crisis", *Economics Letter*, 115, pp. 196-199.

Kauko, K. (2012b) "Triggers for countercyclical capital buffers", *Bank of Finland Online*, 7/2012

Keller, J. (2011), "The BCBS proposal for a countercyclical capital buffer: an application to Belgium", *Financial Stability Review*, National Bank of Belgium, pp. 121-132.

Licchetta, M. (2011), "Common determinants of currency crises: the role of external balance sheet variables", *International Journal of Finance & Economics*, 16(3), pp. 237-255.

Libertucci, M and Quagliariello, M. (2010), "Rules vs discretion in macroprudential policies", VoxEu.org, available at http://www.voxeu.org/article/rules-vs-discretion-macroprudential-policies

Lund-Jensen, K. (2012), "Monitoring Systemic Risk based on Dynamic Thresholds", *IMF Working Paper*, 12/159.

Lo Duca, M. and Peltonen, T.A. (2013), "Macro-financial vulnerabilities and future financial stress: assessing systemic risks and predicting systemic events", *The Journal of Banking and Finance*, 7 (7), pp. 2183-2195.

Macskassy, S. and Provost, F. (2004), "Confidence bands for ROC curves: Methods and an empirical study", *Proceedings of the First Workshop on ROC Analysis in AI*, August.

Manasse, P., Savona, R. and Vezzoli, M. (2013), "Rules of thumb for banking crises in emerging markets", *Working paper DSE*, University of Bologna, No 872.

McClish, D.K. (1989), “Analyzing a portion of the roc curve”, *Medical Decision Making*, 9 (3), pp. 190-195.

Mendoza, E. and Terrones, M. (2008), “An anatomy of credit booms: evidence from macro aggregates and micro data”, *NBER Working Paper*, No 14049.

Orphanides, A. and van Norden, S. (2002), “The unreliability of output-gap estimates in real-time" *The Review of Economics and Statistics*, 84, pp. 569-583.

Pesaran, M., Schuermann., T. and Weiner, S. (2004), “Modelling Regional Interdependencies Using A Global Error-Correcting Macroeconometric Model”, *Journal of Business and Economics Statistics*, Vol. 22, pp. 129-162.

Pepe, M., Longton, G. M. and Janes, H. (2009), “Estimation and comparison of receiver operating characteristic curves”, *Stata Journal*, 9 (1), pp. 1-16.

Puzanova, N. and Düllmann, K. (2013), “Systemic risk contributions: a credit portfolio approach”, *The Journal of Banking and Finance*, 37 (4), pp. 1243-1257.

Ravan, M. O. and Uhlig, H. (2002), “On adjusting the Hodrick-Prescott filter for the frequency of observations”, *The Review of Economics and Statistics*, 84(2), pp. 371-380.

Repullo, R. and Saurina, J. (2011), “The countercyclical capital buffer of Basel III: A critical assessment”, *CEPR Discussion Papers*, No 8304.

Riiser, M. (2005), “House prices, equity prices, investment and credit – what do they tell us about banking crises? A historical analysis based on Norwegian data”, *Economic Bulletin*, 05 Q3, Norges Bank, pp. 145-154.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. and Müller, M. (2011), “pROC: An open-source package for R and S+ to analyze and compare ROC curves”, *BMC Bioinformatics*, 12, pp. 1-8.

Schularick, M., and Taylor, A.M. (2012), “Credit booms gone bust: monetary policy, leverage cycles and financial crises, 1870–2008”, *American Economic Review*, 102 (2), pp. 1029-1061.

Schwaab, B., Lucas, A. and Koopman, S.J. (2011), “Systemic risk diagnostics: coincident indicators and early warning signals”, *ECB Working Paper Series*, No 1327.

Segoviano, M. and Goodhart, C. (2009), “Banking stability measures”, *IMF Working Paper*, No 09/4.

Sveriges Riksbank (2011), *Financial Stability Report*, 2011:1, pp. 52-55.

Swiss National Bank (2013), “Implementing the countercyclical capital buffer in Switzerland: concretising the Swiss National Bank’s role”, February, available at: http://www.snb.ch/en/mmr/reference/CCB%20communication/source

Van Norden, S. (2011), “Discussion of ‘The unreliability of credit-to-GDP ratio gaps in real time: implications for countercyclical capital buffers’”, *International Journal of Central Banking*, 7 (4), pp. 299-303.

VM (2012), “Valtiovarainministeriö, Finanssimarkkinoiden makrotaloudellisten vaikutusten sääntely ja valvonta” (in Finnish), Finnish Ministry of Finance, available at http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/07_rahoitusmarkkinat/20121106Finans/Finanssimarkkinoiden.pdf.

***

Directive 2013/36/EU of the European Parliament and the Council of 26 June 2013 on access to the activity of credit institutions and the prudential supervision of credit institutions and investment firms, amending Directive 2002/87/EC and repealing Directives 2006/48/EC and 2006/49/EC, OJ L 176, 27.6.2013, p. 338

Regulation (EU) No 575/2013 of the European Parliament and the Council of 26 June 2013 on prudential requirements for credit institutions and investment firms and amending Regulation (EU) No 648/2012, OJ L 176, 27.6.2013, p. 1