Improving EWIs for banking crises - satisfying policy requirements[1]

Mathias Drehmann and Mikael Juselius

Bank for International Settlements

September 2012

Abstract

Constructing a meaningful forecast requires that its purpose is clearly defined. In this paper, we discuss the decision problem of macroprudential policymakers and use it to outline the ideal statistical properties of early warning indicators (EWIs) for banking crises. Specifically, we show that uncertainty with respect to the costs and benefits of policy actions, and the need for temporal consistency, have direct implications for the procedures which are used to evaluate EWIs. Based on these procedures, we assess the relative performance of both standard indicators, as well as two recently suggested indicators, the debt service ratio (DSR) and non-core deposits. We find that the credit-to-GDP gap and the DSR are more closely aligned with policymakers' preferences than other competing EWIs. While the former performs consistently well over longer horizons, the latter issues almost perfect signals around one year ahead of before a crisis.

JEL classification:

Keywords: EWIs, ROC

---

[1] The views expressed in the paper are those of the authors and do not necessarily represent the views of the BIS. Mathias Drehmann: Bank for International Settlements, Centralbahnplatz 2, CH-4002 Basel, Switzerland, mathias.drehmann@bis.org. Mikael Juselius: Bank for International Settlements, Centralbahnplatz 2, CH-4002 Basel, Switzerland, mikael.juselius@bis.org.

1. Introduction

Macroprudential policy can only succeed if emerging financial vulnerabilities are detected early enough for preventive action to be taken. How this can be done has receive much attention recently, with early warning indicators (EWIs) playing a prominent role.[2] More often than not, though, the forecasting performance of EWIs and other financial stability models is evaluated predominantly based on various statistical criteria without much regard for their policy relevance. The objective of this paper is to bridge this gap by explicitly relating the choice of statistical procedures for constructing and evaluating EWIs to various policy requirements. In doing so, we propose a number of novel extensions to existing techniques and EWIs.

We proceed in a straightforward manner: We begin by outlining the ideal properties that EWIs for systemic banking crises should have from the point of the policymakers' decision problem. Specifically, discuss uncertainty with respect to the costs and benefits of policy action, as well as the timing and consistency of the signals. We then draw the implications of each of these properties for the choice of estimation procedure and EWI.

Ideally, policymakers would know the cost and benefits of macroprudential intervention, which in turn would determine their relative aversion of type I versus type II errors. In a world where costs of macroprudential interventions are low, but benefits high, a policymaker will have a high aversion for failing to predict a crisis – a type I error – whilst placing less emphasis on type II errors. Indeed, if no reliable EWIs are available, this type of policymaker may prefer structural measures, such as high capital requirements, to safeguard against financial turmoil. However, in such cases, even relatively imprecise EWIs may help reduce reliance on the structural measures, thereby lowering costs of intervention. Of course, the opposite holds if benefits of interventions are low relative to costs, so that policymakers a high aversion against type II errors.

The costs and benefits of macroprudential interventions are, however, to a large extent unknown. Given reasonable ranges for unobservable policy parameters, Drehmann (2012) finds in a simulation study that the scope for policymakers' preferences is surprisingly wide. Even extreme cases, where policymakers essentially care only about type I or type II errors fit into this range. This finding has direct implications for the choice of evaluation criteria. In particular, they should not be heavily dependent on the precise preferences of the policymakers. For this reason we apply a new technique – the receiver operating characteristic (ROC) curve – to assess the performance of potential EWIs.

The ROC curve summarizes all potential type I – type II trade-offs that a prediction model for binary variables generates. Clearly, a model with fewer classification errors is better, and this is reflected in a larger area under the ROC curve (AUROC). As such, AUROC can be used as a summary measure to judge the forecast performance of any binary prediction model. An additional benefit is that it has several convenient statistical properties. For example, confidence bands for AUROC can easily be constructed. In contrast to other methods for evaluating binary predictors, such as minimum noise-to-signal ratios or log probability scores, this method does not assume an underlying loss function. Hence, evaluations based on the ROC curve are robust to different policy preferences and thus is an ideal approach in a situation where these are unknown.

Despite its heavy use in other sciences like medicine, engineering or meteorology, the AUROC curve has not seen many applications in economics. Recent exceptions include Berge and Jorda (2011) who evaluate business cycle indicators, Taylor and Jorda (2010) who study different investment strategies, and Jorda et al (2011) who evaluate a crisis

[2] For a recent survey see e.g. Bisias et al (2012).

model. We expand on this literature by looking more closely at a large range of potential EWIs, as well as by providing confidence intervals and hypothesis tests.

The second issue that we raise concerns the timing and consistency of the EWI. On the one hand, it needs to signal impeding crises sufficiently early. Because of long lags between the times that a signal is observed, policy action is taken, and the impact is felt in the economy (see e.g. Basel Committee (2010)), the signal should be received at least one and a half to two years before a crisis in order to be effective. On the other hand, EWIs should not signal crises too early, more than five years ahead, say. For one, this incurs unnecessary costs of regulation. Second, the political pressure, which may weaken the impact of macroprudential interventions, has a tendency to build-up as booms progress (eg Caruana (2010)). For the same reason, EWIs need to persistently issue signals as not to provide conflicting signals whether to intervene or not.

We investigate the temporal performance of established EWIs (see Hahm et al (2012) and Drehmann et al (2011)) in much more systematic way than has hitherto been conducted in the literature. In particular, we calculate the AUROC separately for each date in the forecast interval and compare the resulting time profiles. It turns out that the indicator variables, which issue the most consistent signals, are also the ones that are both smooth and highly persistent – ie each display (near) double unit-root behaviour.[3]

The finding of near I(2) dynamics with respect to the signal variables has implications for the choice of estimation procedure. In particular, the statistical properties of standard regression based models for binary choice are still to a large extent unknown under this degree of persistence.[4] Hence, inference from such models is likely to be highly misleading. Moreover, these models are estimated to maximize a specific likelihood function that, to the extent to which it is subject to misspecification, can perform arbitrarily bad at specific points of the policymaker's loss function (Elliott and Lieli (2010)). For these reasons we adopt a non-parametric signal extraction approach in the spirit of Kaminsky and Reinhart (1999). This does not impact on the applicability of the ROC curve, as it can be used for all binary classifiers, whether they are derived from a regression model or a non-parametric approach.

Summarizing, the policy perspective that we have adopted here seems to call for an approach to constructing and evaluating EWIs which: (i) uses non-parametric estimation, (ii) evaluates forecast performance based on AUROC, which is independent of policymakers preferences, and (iii) explicitly takes timing and consistency into account.

We apply our approach to assess the performance of nine selected EWIs on a sample of 27 economies, covering quarterly time-series starting in the early 1980's. The set of potential EWIs includes standard variables, such are gaps and growth rates of credit-to-GDP, property prices, and equity prices. We also include two new measures: The debt service ratio (DSR) recently suggested by Drehmann and Juselius (2012) and bank's non-core liabilities as proposed by Hahm et al (2012). The DSR is defined as the proportion of interest payments and mandatory repayments of principal to income. It can be interpreted as capturing incipient liquidity constraints of private sector borrowers. Banks' non-core liabilities consist of foreign liabilities and liabilities to the non-bank financial sector. This measure captures increased reliance by banks on alternative funding sources during credit booms.

We find that gap transformations and levels of variables' generally outperform growth rates and deliver more consistent signals over time. Moreover, the credit-to-GDP gap and the DSR dominate the other EWIs at most horizons. These two variables have somewhat different time profiles. While the credit gap-to-GDP gap performs consistently well, even over

[3] Auto-regressive processes which contain double unit-roots in their characteristic polynomial, typically display smooth patterns. We do not attach any structural meaning to such roots here. Instead we view them as a convenient way of characterizing the relevant order of persistency within a given sample.

[4] Park and Phillips (2000) develop an asymptotic theory for binary choice models where the conditioning variables are allowed to be I(1).

horizons of up to five years ahead of crises, the DSR is extremely precise during the last two years. In the last year before a crisis it nearly has perfect classification ability. Of the remaining indicators, only the property price gap comes close to having equally good performance. By contrast, equity prices and non-core deposits are not very informative as EWIs.

The remaining paper is organized as follows: Section 2 outlines the statistical requirements on EWIs based on the constraints on policymakers during the decision process. Section 3 introduces the potential EWIs. Section 4 evaluates and compares their relative signalling quality based on the criteria laid down the pervious sections, and Section 5 concludes.

2. Constructing EWIs based on policy requirements

Constructing a good forecast requires that its purpose is clearly defined, as different objectives can emphasize different properties of the forecast (Lawrence et al (2000)). For example, researchers who are interested in testing whether certain theories hold out of sample may adopt conventional probabilities of making type I errors (e.g. the 5% significance level). Moreover, the dynamic properties of the forecast may not be of particular interest in such a context, as economic theory is often silent about precise (realistic) adjustment paths. In contrast, supervisors who need EWIs to rally support for prompt corrective actions may be much more adverse to type I errors and are, moreover, critically dependent on the timing and consistency of the signals. In this section, we discuss forecasting objectives from the perspective of marcoprudential policy, which is concerned with the real economy effect of system-wide financial risks (IMF, FSB and BIS, 2009). Specifically, we limit attention to EWIs of baking crises which, for example, can be used guide the build-up of bank's capital buffers in "good times" to absorb losses in "bad times" (Basel Committee on Banking Supervision (2010b))

The macroprudential policy objective has several implications for the choice of empirical strategy. First, it provides a framework to analyse costs benefits of interventions and can, thus, help to determine the policymaker's relative aversion for type I and type II errors. Second, it has several important consequences for the desired temporal properties of the EWIs. In this section, we discuss both these issues in turn.

2.1 Unknown policy trade-offs and the ROC curve

Information about the expected costs and benefits from taking a certain policy action is clearly valuable for designing and evaluating forecasts. But it is possible to obtain such information in the macroprudential context?

To make matters concrete, consider a simple economy where there are two states: a boom which is inevitably succeeded by a crisis (C=1) or a normal period without a crises (C=0). Policymakers are risk neutral and can either impose buffers (B=1) or not (B=0).

Table 1 summarizes the state and decision dependent utilities ($U_{B,C}$) of the policymakers. If there is no crisis in the next period and policymakers do not impose buffers ($U_{B=0,C=0}$) welfare losses are zero, whereas if buffers are implemented ($U_{B=1,C=0}$) the economy faces the costs of regulation. If the economy is in the boom state, however, and no buffers are imposed the economy has to bear the full cost of a crisis ($U_{B=0,C=1}$). Finally, if buffers are imposed during a boom ($U_{B=1,C=1}$) , we assume that the costs of a crisis are reduced by a factor α but that the economy still has to bear the costs of regulation.[5]

[5] Alternatively, we could assume that buffers act to reduce the unconditional probability of a crisis or that the cost of regulation are included in α. While such changes alter the policymakers' decision problem, what

Table 1: Costs and benefits of policy interventions

	No crisis next period	Crisis next period
No buffers	$U_{B=0,C=0}$ =0	$U_{B=0,C=1}$ =-Cost of crisis
Impose buffers	$U_{B=1,C=0}$ =-Cost of regulation	$U_{B=1,C=1}$ = -(1- α)Cost of crisis - cost of regulation

α is between zero and one and summarizes the effectiveness of the macroprudential tool

Policy makers observe an indicator variable $S \in R$, which issues a noisy signal about the state of the economy. In particular, we assume that the true positive *(TP)* and false positive rate *(FP)* of the indicator variable are related to the threshold $\theta \in R$ with

$TP(\theta)= prob(S> \theta| C=1)$

$FP(\theta)= prob(S > \theta| C=0)$

which summarize the signaling quality of the indicator. For any given value of θ, the trade-offs between true and false positives are contained in the upper half above a 45° line within a unit square.[6] Obviously, $TP \to 0$ and $FP \to 0$ as $\theta \to \infty$, and $TP \to 1$ and $FP \to 1$ as $\theta \to -\infty$. For uninformative indicators, these trade-offs will move along the 45° line as θ goes from minus to plus infinity, whereas the trade-offs will move closer to the upper boundary of the unit square for highly informative ones. Graph 1 depicts the TP/TF trade-offs for three hypothetical signals.

The decision problem for the policy maker is therefore to determine the threshold θ so to minimize the welfare costs of crises and regulation. Given a particular signal, the expected utility of the policy maker can be written as:

$$U(\theta)= p(C=1)* [TP(\theta) * U_{B=1,C=1}+(1-TP(\theta))*U_{B=0,C=1}] \\ + (1-p(C=1))*[FP(\theta) * U_{B=1,C=0}+ (1-FP(\theta))*U_{B=0,C=0}] \qquad (1)$$

where *p(C=1)* is the (unconditional) probability that the economy is in the crises generating state. For given state and decision dependent utilities and an unconditional probability of a crisis, as well as a specific signal, the policymakers maximizes (1) with respect to *θ*. This yields an optimal threshold for the signal, *S*, and associated true positive and false positive rates.

The optimal choice of the policy maker can also be analyzed graphically. The relative trade-off rates of two different policymakers are indicated by the dashed straight lines in Graph 1. The higher the costs of an intervention is compared to expected costs of a crisis, the steeper the slope of the line, indicating that the policymaker is less willing to increase the rate of true positives at the expense of false positives. Put differently, the policy maker will worry more about making type II errors than type I errors. The converse happens when the relative cost of policy is low. In the extreme case where the indicator is completely uninformative,

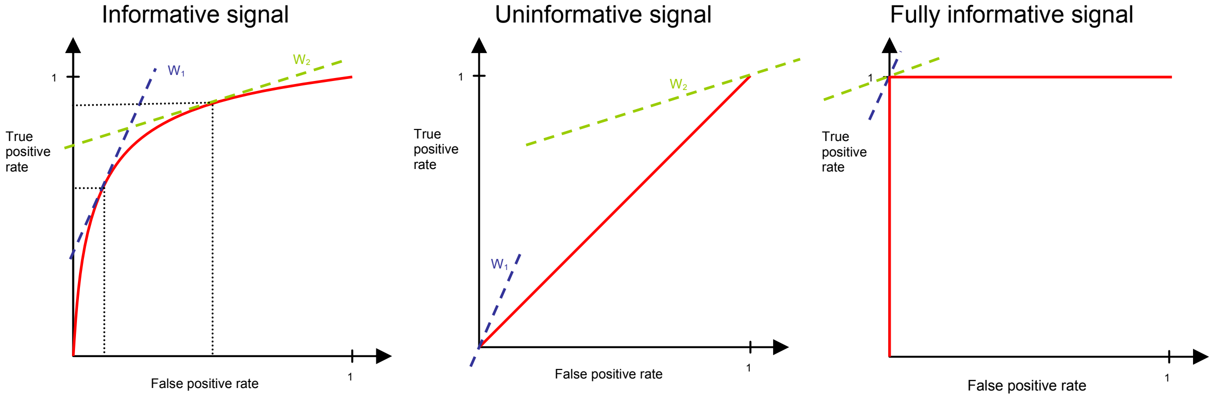
matters from our perspective is that the signalling quality of the indicator variable, *S*, does not change (see definitions in the main text).

6 It case the trade-offs are in the lower half, a simple transformation of the indicator variable will ensure that they are in the upper half.

$\partial ROC / \partial FP = 1$, implying that the policy maker will always take action if *Cost of regulation* < *α*p(C=1)*Cost of crisis*, and vice versa.[7] Intervening in every period is equivalent to setting minimum standards, such as minimum capital requirements. Similarly, a fully informative indicator touches the point (0, 1) implying that there is a θ such that $TP(\theta) = 1$ and $FP(\theta) = 0$ and the correct policy action will always be taken.

Graph 1

Signal quality and welfare



More generally, the optimal threshold is given by the point where marginal trade-off rate of the signal equals the marginal rate of substitution between expected benefits from taking action if there is a crisis versus the expected costs of imposing the buffer in case there is no crisis Yet, that trade-off is specific to each signal variable, which implies that information about the policymakers' marginal trade-off between true and false positives should ideally be incorporated into the assessment of EWIs.

To calculate the policymaker relative trade-off requires data on the probability of crises and the expected costs and benefits of interventions. The unconditional probability of a crisis, the cost of a crisis, and the costs of regulation can to some extent be gauged from data on past events. For example based one official sector study assessing the impact of tighter capital requirements (Basel Committee on Banking Supervision (2010a)) finds that the unconditional probability of a crisis ranges from 1-5% depending on the level of capital in the system. Similarly, the study finds for costs of interventions that a one percentage point increase in capital decreases long-run output by 0.09% on average, whereas the Macroeconomic Assessment Group (2010) finds that at its maximum these output costs are around 0.19% in the short run. However, very little is known about the benefits of macroprudential policy measures, such as countercyclical capital buffers. Drehmann (2012) therefore implements a simulation study based on a similar policy problem as described above, analyzing different values for the probability of crises, the costs of interventions and the benefits of policy interventions as described by α. He finds that the scope for policymakers' relative trade-offs is surprisingly wide comprising even extreme cases, where policymakers essentially care only about type I or type II.

In the absence of more detailed information about policymakers' preferences, the second best solution is to use assess EWIs based on procedures which are less dependent on the

[7] Technically, there is also a degenerate case where policy makers are indifferent between intervention and not and the slope of the utility curve is 1.

policymaker's trade-off between true and false positives. One such procedure is to compare the areas under the red curves in Graph 1. These curves are called receiver operating characteristic (ROC) curves, which describe the mapping from FP to TP, formally TP = ROC(FP(θ)), and capture the signalling quality of the indicator. The area under the curve is denoted AUROC and is given by

$$AUROC = \int_0^1 ROC(FP)dFP \qquad (2)$$

and provides a summary measure of the classification ability of the indicator.

AUROC is increasing with the predictive power of the indicator across all possible thresholds θ and lies between 0.5 for uninformative predictors, and 1 for and prefect predictors. AUROC can be estimated parametrically or non-parametrically and has convenient large sample properties so that hypothesis testing, for example whether AUROC is significantly different from 0.5, can easily be implemented. While AUROC based assessments have been used extensively in other sciences like engineering or medicine, such applications have only recently found their way into economics (see eg Berge and Jorda (2011)). Given the independence of AUROC from policymakers' preferences, we use it as the measure to evaluate the forecast performance of EWIs in this paper.

2.2 Timing, persistency, and consistency

Next, we discuss the appropriate timing of an ideal EWI. This issue is more complex from a policy perspective than from a purely statistical point of view. First, EWIs need to signal impeding crises early enough. For monetary policy, the typical lead-lag relationship between changes in the interest rates and inflation is around one to two years. While evidence on lead-lag relationships for macroprudential policies are currently not available in the literature, they are likely to be at least as long. For instance, banks have one year to comply with increased capital requirements under the countercyclical capital buffer framework of Basel III (Basel Committee (2010)). In addition, data are reported with lags and policymakers generally do act immediately on data developments, but observe trends for some time before they change policies (e.g. Bernanke (2004)).

Second, ideal EWIs should not signal crises too early. Booms are popular as money is made, output is growing and more and more households may get a foot on the property market ladder. Macroprudential policies responding to risks building-up in the background inevitable draw criticism. Combined with the argument that "this time it's different", this can undermine the effectiveness of macroprudential measures if they are introduced too early (e.g. Caruana (2010)). Taken together, these two requirements suggest that an ideal EWI should signal crises early enough, but not too early. For our empirical analysis, we argue that good EWIs issue a crisis signal at least 1 year before a crisis. Judging what it "too early" is more difficult. But to be conservative, we use at most a 5 year horizon.

An important additional requirement which has largely been overlooked in the literature concerns the persistency of EWI signals. As already discussed, in practice policymakers tend to observe data developments for some time before they gradually change policy instruments (e.g. Bernanke (2004)). In the context of monetary policy it also theoretically optimal for policymakers to be more cautious and respond less to new information, the noisier this information is (e.g. Orphanides (2003)).[8] Rephrasing this for EWIs means that an indicator which fluctuates frequently between crises and no-crises signals is worse than an indicator delivering persistent signals.

It seems evident that the persistence of the signals is directly tied to the persistence of the underlying conditioning variables. This is, for instance, in line with Park and Phillips (2000),

[8] Gradual changes in interest rates are also theoretically optimal for other reasons….

who find, in the context of regression based binary choice models, that policy is likely to manifest streams of little or intensive intervention when the explanatory variables are difference stationary. This suggests that variables which in past studies have been found to be useful for macroprudential policy may also be ones which display a high degree of persistence.[9] And this is what we find for all the variable analysed here (see Section 3.3) While this type of persistence may be benign from the policymakers' perspective, it can nevertheless have implications for both correct inference and the choice of estimator.

The preceding discussion on timings related issues has several direct implications for our preferred approach. First, we assume that a signal is correct if it forecasts a crisis in an interval over the next five years. This also addresses the uncertainty of dating crisis correctly and the difficulty of predicting the actual timing when a crisis will materialize. Moreover, we do not take any signals in the two years after the beginning of a crisis into account, as binary EWIs become biased if the immediate post-crisis period is included in the analysis (Brussiere and Fratzscher (2006)).[10] This also has sound economic rationale: it makes no sense to predict another crisis immediately after one has materialized.

Second, we evaluate each indicator in terms of its ability to issue persistent signals. We do this by calculating the AUROC for each date in the forecast interval separately…

Third, we adopt a non-parametric signal extraction approach in the spirit of Kaminsky and Reinhart (1999) which may be more robust under double unit root variables than standard regression based models for binary choice. Also, the latter models are estimated to maximize a specific likelihood function that, to the extent to which it is subject to misspecification, can perform arbitrarily bad at specific points of the policymaker's loss function (Elliott and Lieli (2010)).

Perspective: interpretability and logic of the EWI signal

From a policy perspective, an ideal EWI does not only need to fulfil these statistical criteria, but has to be backed-up by a coherent, analytical framework which policymakers understand. Policymakers never rely purely on statistical tools, even for macroeconomic forecasting where the theoretical and empirical literature is far more advance. Instead they analyse a range of models and indicators and supplement it by judgement.[11] Findings in the literature show that practitioners value the sensibility of forecasts more than accuracy (Huss, 1987) and adjust forecasts if they lack justifiable explanations (Önkal-Atay et al (2009)). For an empirical strategy to find EWIs, this implies that simply data driven indicators, for example derived by a general to specific approach, are not suitable for policy purposes.

Ideally, the analytical framework would be based on one or several well established theoretical models, which are however not yet available for financial stability purposes. Most of the advanced models cannot account for crises, the main event we are chiefly interested in (e.g. Gertler and Karadi (2011) or Angeloni and Faia (2012)). Instead the more appropriate analytical framework in our view is in the tradition of Kindleberger (2000) and Minsky (1982), which see financial crises as the result of mutually reinforcing processes between the financial and real sides of the economy. In this view, financial imbalances are driven by, but also feed, an unsustainable economic expansion, which manifests itself in unusually rapid growth of credit and asset prices. As the economy grows, cash flows, incomes and asset prices rise, risk appetite increases and external funding constraints

[9] This conjecture, is consistent with the evidence in Borio et al. (2012), for instance, which suggests that the financial cycle is much longer than the conventional business cycle

[10] Cecchetti et al (2009) find that crises last nearly three years on average. In our sample, the minimum time between two crises in one country is five years. Thus by assuming that crises last only two years we bias our noise-to-signal ratio upwards, as only type 2 errors are issued during the quarters immediately following the end of the second year after an episode.

[11] See Lawrence et al (2006) for a survey on judgemental forecasting and its importance in practice.

weaken. This, in turn, facilitates risk-taking. The financial system typically does not build up sufficient capital and liquidity buffers during benign economic conditions, when it is easier and cheaper to do so, in order to deal with more challenging times. At some point, imbalances have to unwind, potentially causing a crisis, characterised by large losses, liquidity squeezes and possibly a credit crunch.

3. Potential EWIs

In this section, we propose a range of potential early warning indicators which fit the discussed policy requirements, which we then test in the next section. As a first step, we focus on single variable indicators since using these to anchor potential policy actions has the advantage of being more transparent (see Drehmann et al (2011)). As will be shown, single indicators already provide very good guidance, leaving limited scope for incremental improvement through the use of multivariate approaches.

Drehmann et al (2011) cover a wide range of potential indicators which could be used to anchor countercyclical capital buffers as one particular macroprudential tool. They consider macroeconomic variables, indicators of banking sector conditions and market indicators. While market indicators perform very well to signal the right timing for releasing macroprudential tools – a question not analysed in this paper – they are poor indicators during the build-up phase. In addition, data availability is often limited. Equally, banking sector conditions such as aggregate profits and losses do not signal crises well and data availability in a cross-country sample is challenging. We therefore focus on a small set of macroeconomic indicator variables, potentially capturing the build-up of vulnerabilities in the domestic economy. In total we assess nine different variables.

In line with a Minsky (1982), Drehmann et al (2011) find that the best performing EWIs for banking crises are measures of credit and asset price booms. In particular credit developments are important, in line with the literature (see eg Reinhart and Rogoff (2009), Gourinchas and Obstfeld (2012),orJorda et al (2011)). The single best indicator is the credit-to-GDP gap measuring deviations of the credit-to-GDP ratio from a long run trend (see next section on data discussion). This variable acts as the starting point of discussions about the level of countercyclical capital buffer charges according to the Basel Committee (2010). Real credit growth also performs well, which we therefore include in the analysis as well.

The second class of useful EWIs are indicators of asset price booms. We look at property as well as equity markets and assess the signalling performance of growth rates and deviations from long term trends.

More recently, Drehmann and Juselius (2012) propose the aggregate debt service ratio (DSR) as a useful early warning indicator. The DSR is an aggregate measure of the proportion of interest payments and mandatory repayments of principals relative to income. If DSRs are high, it is a clear sign that households and firms are overextended, so that even small income shortfalls prevent them from smoothing consumption or making new investments. Larger shortfalls could even trigger a rise in defaults and ultimately crisis. In a world with constant lending rates and constant maturities, the DSR and established leverage measures, such as the debt-to-GDP ratio, would provide the same information. Yet, Drehmann and Juselius (2012) show that this assumption – in particular for lending rates – is not fulfilled, so that the DSR is able to capture the burden imposed by debt on borrowers more accurately.

Hahm et al (2012) argue that lending booms can only be sustained if non-core liabilities, defined as wholesale and cross-border funding, increase rapidly as traditional retail-deposits (core liabilities) adjust only sluggishly. As they find an empirical relationship between the ratio of non-core to core liabilities with the likelihood of crisis, we include this measure as well.

Last, as a macroeconomic benchmark we assess the signalling quality GDP growth, even though Drehmann et al (2011) have already shown that it performs quite poorly.

3.1 Data

We analyse quarterly time-series data from 27 countries. The sample starts in 1980 for most countries, and at the earliest available date for the rest. Table X in the Annex (to be completed) provides detailed definitions and sample coverage of the data.

For dating systemic banking crises we follow Laeven and Valencia (2012), but ignore three crises which were driven by cross-border exposures.[12] In addition, we adjust some dates following discussions with central banks. In section 4.4.1, we however undertake robustness checks with respect to alternative crisis dates based on Reinhart and Rogoff (2008) and Drehmann and Borio (2009). (A list of all crisis dates is given in Table A1 in the Annex).

For most of our results we consider a balanced sample, ie we include only dates for which all indicator variables are available.[13] In this this case, we observe around 1500 data points and 17 systemic crises. We also consider an unbalanced sample with 26 crises and more than 2600 data point..

Macroeconomic variables are collected from national authorities, the IMF international financial statistics and the BIS database. Residential real estate property prices are based on BIS statistics which are only available for a subset of countries and do generally not cover the full sample period.[14]

To measure credit in the economy, the literature generally relies on monetary statistics measuring bank credit to the private-non financial sector, such as the IMF-IFS. For several countries, this excludes important sources of credit to the economy, such as bond markets or cross-border loans. In the US, for example, bank credit accounts on average only for less than 40% of total credit in the last 10 years. And even in more bank based systems such as France, this fraction amounts to only 70%. Drawing on a new BIS data-base we therefore use measures of total credit to the private non-government sector based on a flow-of-funds concept where available.

We derive gap measures by subtracting a one-sided Hodrick-Prescott filtered trend from the level of a series.[15] This is achieved by recursively extending the sample by one period and retaining the difference between the actual value of the variable and the value of the trend at the new point. Thus, a property price trend calculated in, say, 1988q1 only takes account of information up to 1988q1, and the GDP trend in 2008q4 takes account of all information up to 2008q4. This is an important practical constraint, as policymakers have to take decisions

[12] Crises of the latter type were identified for three countries (Germany, Sweden and Switzerland in 2007 and 2008) through information provided to us by national central banks.

[13] The most constraining variables are property prices and the ratio of non-core to core deposits.

[14] As a robustness check we also assess the signalling performance of a property price index combining residential and commercial real-estate. There is no significant difference between both indicators and as less data on commercial real estate are available, we use only residential property prices. Results are available on request.

[15] For asset price gaps, the difference between the actual data and the trend at each point in time is normalized by the trend in that period. For the credit-to-GDP gap, we simply take the difference between the actual data and the trend at each point in time.

in real time and rely on data that are available at that point.[16] Before using any trend, we require at least eight years of information.[17]

The calculation of the Hodrick-Prescott filter involves a key smoothing parameter λ. Following Hodrick and Prescott (1981)it has become standard to set the smoothing parameter λ to 1600 for quarterly data. Ravn and Uhlig (2002) show that for series of other frequencies (daily, annual etc) it is optimal to set λ equal to 1600 multiplied by the fourth power of the observation frequency ratio. We set lambda for all the gaps to 400,000, implying that financial cycles are four times longer than standard business cycles. This seems appropriate, as crises occur on average once in 20 to 25 years in our sample.Drehmann et al (2011) explore different lambdas but find that lambda equal to 400.000 works best for early warning indicators. Equally, we could have used a time trend such as Gourinchas and Obstfeld (2012), but our approach is in line with the suggestion for calculating credit-to-GDP gaps as indicator variables for the countercyclical capital buffers in Basel III (Basel Committee on Banking Supervision (2010b)).

Debt service ratios (DSR) are taken from Drehmann and Juselius (2012). Even though levels are surprisingly similar across countries and time despite different levels of financial development, some country differences persist due to different rates of homeownership or different industrial structures. To account for this, we subtract 15-year rolling averages from the DSRs for our analysis.

Hahm et al (2012) construct a range of different ratios of non-core to core deposits. Empirically, the best performing measure is when non-core deposits are proxied by the cross-border borrowing of banks taken from the IMF International Financial Statistics and core deposits by M2, which is the ratio included in our analysis. To adjust for cross-country differences we also subtract 15-year rolling averages to normalize this ratio.

3.3 The behaviour of candidate variables around systemic crises

Before conducting our statistical tests, we look at the time profile for all indicator variables around systemic banking crises. Graph 2 summarises the behaviour of the variables during a window of 20 quarters before and after the onset of a crisis (time 0 in the graphs). For each variable, we show the median (solid line) as well as the 25th and 75th percentiles (dashed lines) of the distribution across episodes.

The graph reveals that the DSR, the credit related variables and property price indicators all could be useful indicators for signalling impeding crises. Their timing seems, however, to be very different. The median DSR starts from a relatively low base and triples during the four years before a crisis, at which point it peaks. The credit-to-GDP gap and real credit growth, on the other hand, are already very high three to five years ahead of a crisis and rises much more slowly. Finally, property prices seem to reach their peak around 2 years before crises, after which they start to decrease.

Real GDP growth shows a markedly different time-profile. It is around 4% four years prior to a crisis. It then starts to decline, with a slowdown gathering momentum in the year leading up to the crisis. Once the crisis materialises, GDP growth turns negative. After around two years, on average, the economy returns to its pre-crisis growth rate, suggesting that this

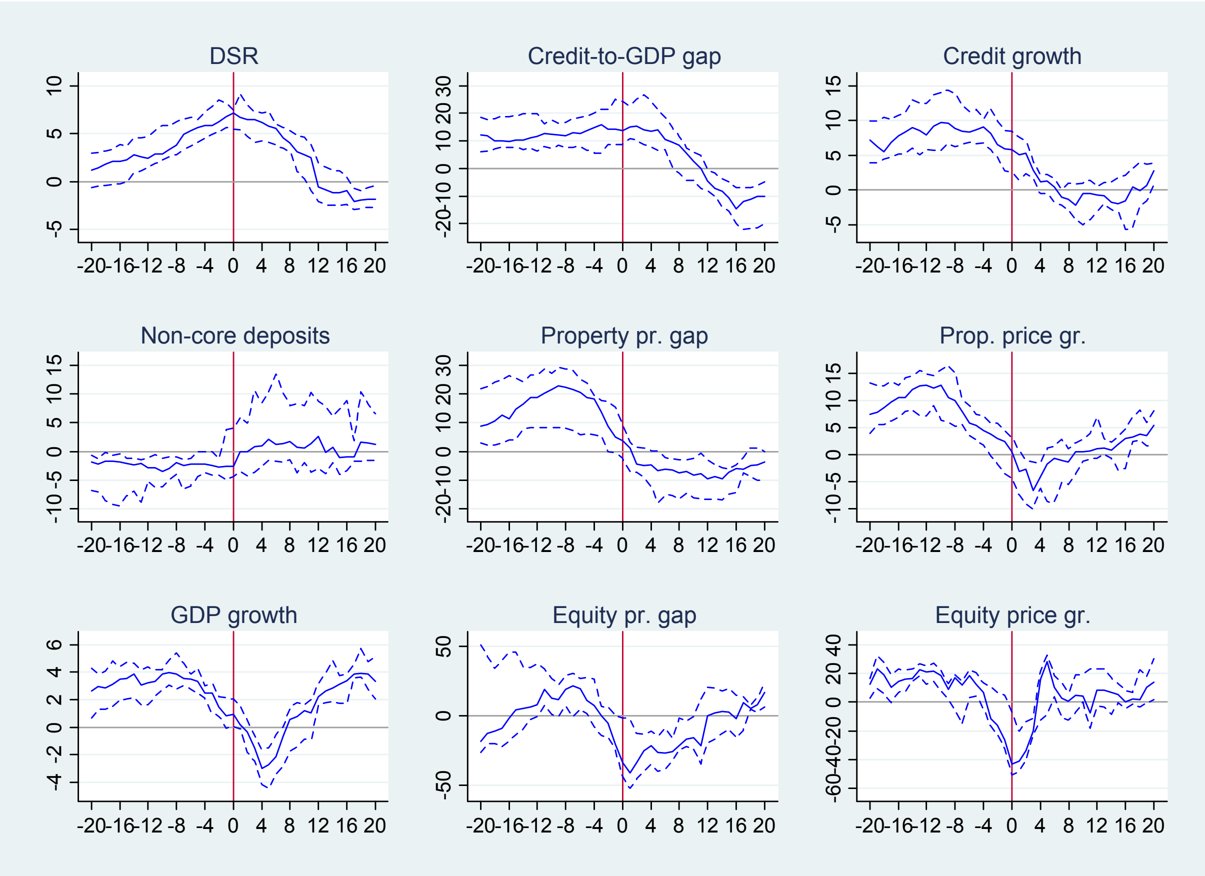
[16] To test real time EWI properties, we would ideally use only the vintage of data policymakers had available at a particular moment. However, these data are not available for the large set of countries and the historical period we need to cover to establish robust indicators. It has also been shown that data revisions, at least for credit and GDP series in the US, are not of first order importance for our type of analysis (see Edge and Meisenzahl (2011)).

[17] Ideally, ten years of data would be better (eg, Borio and Lowe (2002)). But given that data are limited for some series in some countries, we chose a 8-year window to ensure sufficient observations.

growth rate is not particularly unusual. Interestingly, the 75th percentile shows that many crises are not preceded by any slowdown in output.[18]

Graph 2 clearly indicates that the ratio of non-core to core deposits is unlikely to be a good early warning indicator. During the 5 years in the run-up to crises, it even tends to be below its 15 year average. Only during the crisis, the ratio increases rapidly, suggesting that it could be a useful indicator for the release phase.[19] Similarly, the dispersion around the median for the equity price indicators is rather large, suggesting that they may not be ideal EWIs.

Graph 2: Indicator variables around crises[1]



1 The horizontal axis depicts plus\minus 20 quarters around a crisis, which is indicated by the vertical line. The historical dispersion (median (solid line), 25th and 75th percentiles (dashed lines)) of the relevant variable is taken at the specific quarter across all crisis episodes in the sample.

3.2 Persistence of candidate variables

The time profiles in Graph 2 suggest that the DSR, the credit to GDP gap, the property price gap, property price growth and possibly GDP growth have the most consistent time profiles ahead of crises, i.e. they show less fluctuation over time and less dispersed across

[18] This provides a clear indication that models linking fragilities in the banking sector to weak macroeconomic fundamentals, such as macro stress tests, do not capture the dynamics of many crises (Alfaro and Drehmann (2009)).

[19] This patter may also explain why Hahm et al (2012) find it to be a useful indicator. In their analysis they include 6 months around the crisis date as “ones” in their empirical probit analysis. They also use somewhat different crisis dates.

countries. Hence, it seems reasonable that they are also more persistent than non-core deposits and the equity price gap (growth).

To assess the persistence of the potential indicator variables analysed in this paper, we estimate AR(k) processes for the levels and first differences of each variable and apply standard unit root test, as well as calculate the sum of autoregressive coefficients (see for instance, CITATION).[20] Table (to-be-added) summarizes the results which are reported in full in the Annex. As can be seen from the table, credit, property price, and GDP growth display a high level of persistency which is in many cases statistically indistinguishable from unit-root processes. This is particularly the case for credit growth and property price growth and would imply that the levels of these variables are close to I(2) processes, ie they display double unit-root behaviour. In contrast, while equity price growth and core deposits is less persistent and the unit-root hypothesis is rejected in most cases.

To check whether or not the pronounced persistence of these variables remains in their corresponding gap transforms, we conduct similar testing for the latter variables.[21] We also include the first difference if the DSR and the non-core deposits. As can be seen from the table, the first difference of credit and property price gaps, as well as the DSR, still displays dynamics which is consistent with unit-roots in most cases. The equity price gap and the non-core deposits on the other hand are much less persistence.

These results suggest that persistency and consistency go hand in hand. Hence, while growth rates, for example, have the advantage of not requiring statistical pre-filtering, they are nevertheless likely to be dominated by gap transforms or levels of variables (to the extent that these don't contain deterministic trends). This pattern is visible in the formal analysis below. However, a higher level of persistency also comes with a price: standard regression based inference has a tendency to become unreliable. This is particularly true when variables are close to I(2) as statistical theory is often unavailable for this case. Hence, the benefits of applying parametric likelihood based estimation procedures to estimating the classification ability of individual variables is limited compared with the risks of spurious inference. For this reason we apply a non-parametric estimation procedure.

4 The signalling quality of different EWIs

In this section, we formally test the signalling quality of the nine proposed EWIs. We first start by discussing our empirical methodology to construct ROC curves. We then show estimated ROC curves for a two year horizon before analysing the timing and consistency of EWIs more thoroughly. We end the section with some robustness checks.

4.1 Estimating ROC curves

All ROC curves are estimated non-parametrically. Thus for any possible threshold and particular forecast horizon, we calculate the empirical fraction of true-positives and false-positives. We then compute AUROC in line with equation 2. Standard errors are derived via bootstraps using 1000 replications.

As discussed, for policy purposes EWIs need to signal impeding crisis sufficiently early and we therefore take a five year forecast horizon. However, rather than using a flexible forecast horizon in line with Kaminsky and Reinhart (1999),[22] we want to better understand the

[20] It is well known that the sum of autoregressive coefficients has a downward bias in small samples (CITATION). Hence, the actual persistence may be even larger than what is reported here.

[21] The unit-root hypothesis cannot be rejected for any of gaps and the DSRs in their levels.

[22] With a flexible forecast horizon of eg three years, a signal is correct if a crisis occurs at any point within the following three years. Otherwise, it is a false positive.

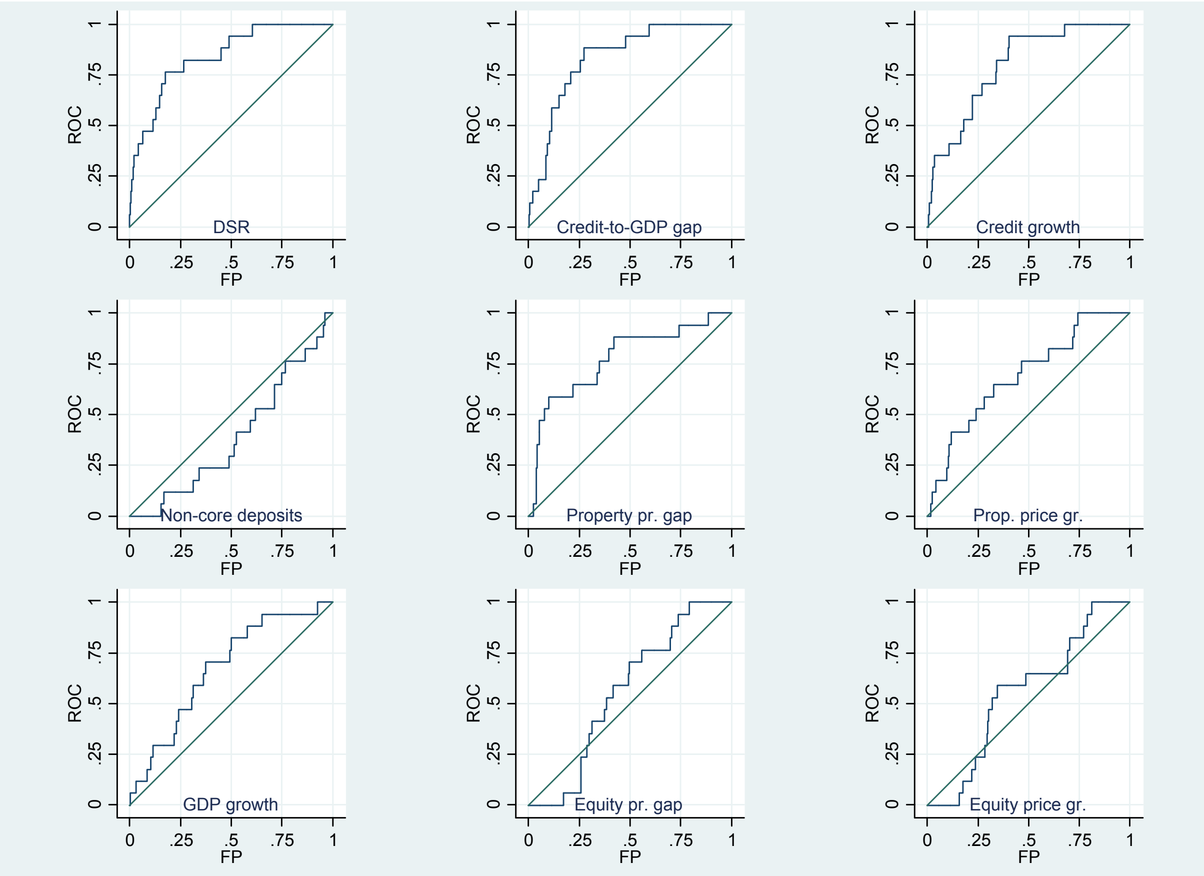
persistency of signals. Therefore we implement the following procedure for each of the 20 quarters within the 5 year horizon: In case a (no) crisis occurs in X quarters, the indicator will provide a correct signal if it breaches (does not) breach the threshold θ. Otherwise a type I error (false positive) is recorded. Except for the X quarter horizon, we do not take account of any other signals 5 years prior to crises to obtain an accurate picture of the time-profile of EWIs.

As binary EWIs become biased if the immediate post-crisis period is included in the analysis (Bussiere and Fratzscher (2006)), we also do not take any signals in the two years after the beginning of a crisis into account.[23] This also has sound economic rationale: it makes no sense to predict another crisis immediately after one has materialized.

4.2 ROC curves with a two year horizon

As an example, Graph 3 shows the estimated ROC curves for signals issued 8 quarter ahead of crises, whereas Table A2 in the Annex reports estimated AUROCs as well as the upper and lower 95% confidence intervals.

Graph 3: ROC curves for a 2 year forecast horizon



Unsurprisingly, the graph mirrors the message from Graph 2. In particular, the DSR and the credit-to-GDP gap, but also credit growth and the property price gap, issue quite powerful

[23] Cecchetti et al (2009) find that crises last nearly three years on average. In our sample, the minimum time between two crises in one country is five years. Thus by assuming that crises last only two years we bias our noise-to-signal ratio upwards, as only type 2 errors are issued during the quarters immediately following the end of the second year after an episode.

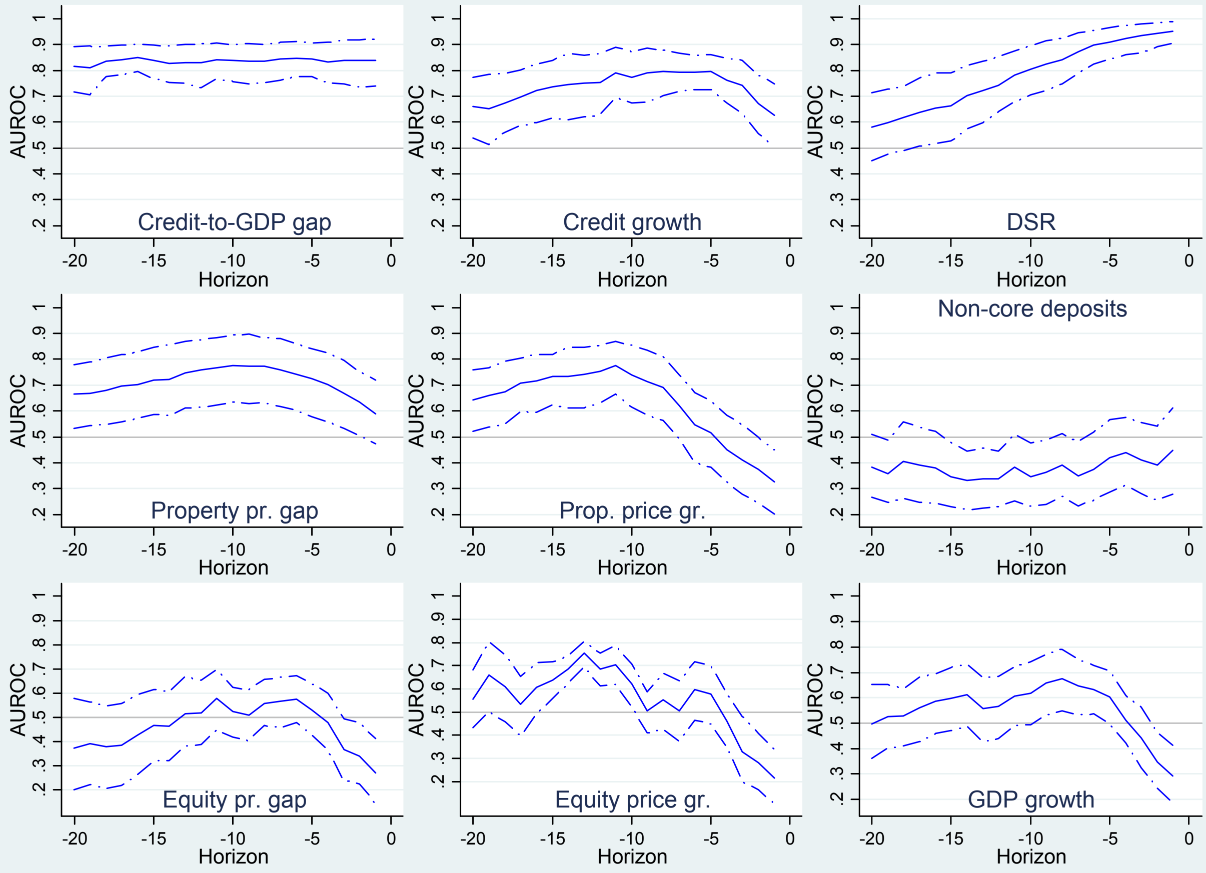
signals. Their AUROCs are 0.83, 0.84, 0.80 and 0.77, respectively (see Table A2). This is quite high. Jorda (2011) cites other studies showing that a widely used prostate-specific antigen (PSA) blood test has an AUROC of around 0.8 and that the S&P 500 has an AUROC of 0.86 for detecting in current time whether the economy is in recession or not. At the other extreme are non-core deposits and equity prices only provide signals which are indistinguishable from an uninformative signal.

4.3 AUROCs for all forecast horizons

Graph 3 is only indicative of the performance of different indicator variables as it only looks at one particular forecast horizon. In this sub-section, we therefore derive AUROC for each quarter individually. Such an approach provides clear information about the time-profile and the persistency of each indicator variable.

By doing so, we do not want, however, to suggest that the revealed average time pattern should be used to anchor policy very specifically, by for example suggesting that policy makers should for the first 4 quarters follow indicator X, then for 2 quarters indicator Y and so on. Future crisis will certainly not play out like this. Rather, our aim is to document broad patterns, which help policy makers in determining the appropriate stance for macroprudential policies.[24]

Graph 4: AUROC over time



Note: Solid line: AUROC, dotted lines: 95% confidence intervals. Horizon: quarters before a crisis.

[24] Would our results be used to guide policy decisions they would also be subject to the usual Lucas or Goodhart critiques. As Drehmann et al (2011) argue, though, the leading EWI properties would disappear by definition, if EWIs are well specified, their use would force banks to build up buffers to withstand the bust. Moreover, if, in addition, the scheme acted as a brake on risk taking during the boom, the bust would be less likely in the first place. However, the loss of predictive content per se would be no reason to abandon the scheme.

Graph 4 highlights that the time-profile of the different indicator variables differs substantially. DSR's early warning properties are especially strong in the two last years preceding crises. One quarter before crisis, its AUROC is 0.95% and its upper 95% confidence interval is 0.99 (see Table A2, Annex 1). Thus, it is nearly a perfect indicator. AUROC then drops continuously the longer the forecast horizon gets. But the performance in year 2 and 3 is still very good, with an AUROC on average of 0.88 and 0.78 respectively. Only in the fifth year, the signals from the DSR become uninformative.

The credit-to-GDP gap shows a markedly different pattern. Across the 5 years, AUROC fluctuates between 0.83 and 0.85. Only in the last two quarters (ie quarter -19 and -20) does it drop marginally (to 0.81). Given the length of the forecast horizon, this performance is remarkable.

The informational content of the property price indicators and credit growth is broadly similar. They are somewhat informative in predicting crises one two three years in advance. However, as seen in Graph 3 they tend to decrease ahead of crisis somewhat and thus their reliability as crisis indictor decreases as well.

Last, GDP growth, non-core deposits and equity price indicators are essentially ineffective as EWIs. There are a few quarters where one or the other is significant at the 95% confidence interval. However, policymakers would unlikely want to rely on these variables as they are clearly dominated by the other EWIs. More importantly, as discussed signals should be persistent. And in this regard an indicator variable which is only issuing a noisy signal in one particular period is not useful as policymakers will not be able to distinguish whether vulnerabilities subsided or whether the signal strength is decreasing because a crisis is coming soon.

What is interesting about these results is that the most persistent indicator variables (the credit-to-GDP gap, the DSR and the property price gap) dominate the less persistent indicators in all periods. Already by itself, this would suggest that policymakers should prefer these variables as indicators. The high level of persistency also imply that they deliver more consistent signals over longer periods which makes them even more attractive.

In summary, the results suggest that the credit-to-GDP gap is a highly useful indicator for identifying vulnerabilities which may lead to crises in five or less years. In the short horizon, the DSR performs even better and in the immediate quarters before crises it is nearly the perfect indicator.

4.4 Robustness

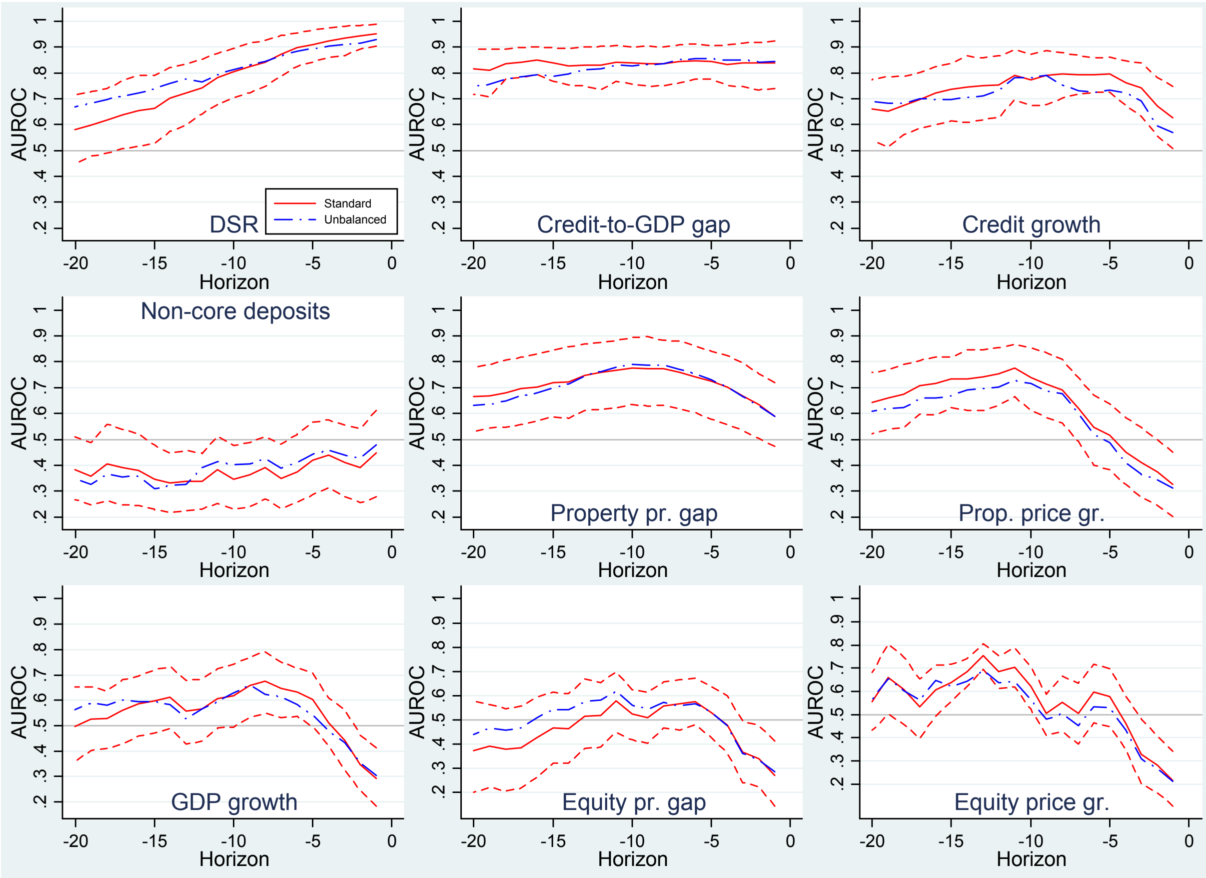
In this section we check the robustness of our results with respect to different crisis dating and different time periods and whether results change when only partial ROC curves are estimated

4.4.1 Unbalance samples

So far we have used a sample, which only included data point when all nine indicator variables are observable. However, this constraints our data sample considerably as it is reduce to around 1500 data points rather than more than at most 2600 otherwise. This also implies that only 17 instead of 26 crises are included in the previous sample. In this section we therefor re-estimate AUROCs including all possible observations for the different indicator variables.

Graph 5 highlights clearly that our results are independent of whether the balanced or unbalanced sample is used.

Graph 5: AUROC over time: unbalanced samples



Note: dotted red lines: 95% confidence intervals of AUROC for the standard estimation. Horizon: quarters before a crisis.

4.4.2 Crisis dating

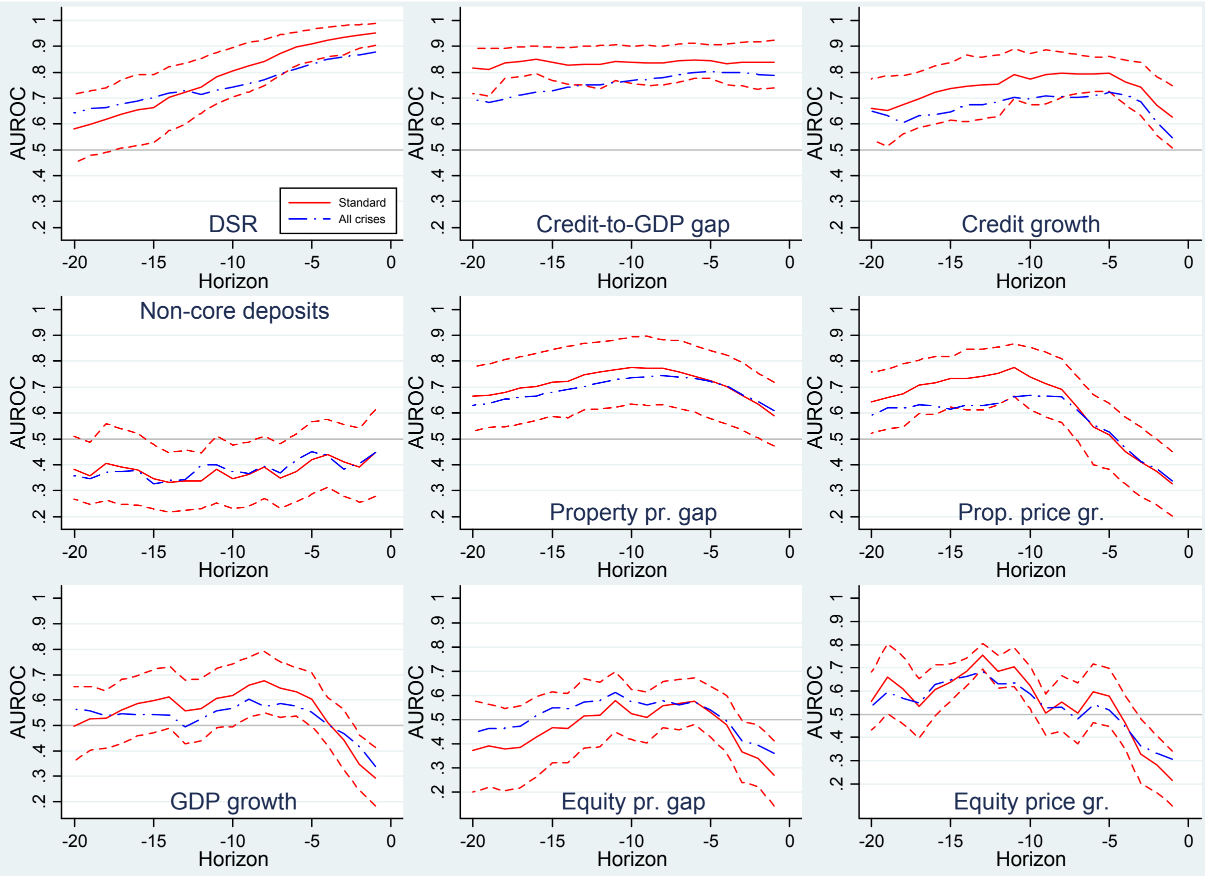
For the results presented so far, we use the stringent crisis dating provided by Laeven and Valencia (2012) as well as discussion with central banks. Laeven and Valencia follow a very strict definition what constitutes a crisis, as they require that two conditions are met: First, there are significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations), and second, there are significant banking policy intervention measures in response to significant losses in the banking system

Other authors have suggested weaker crisis definitions. In this section we therefore include these crises, which were not necessarily systemic as well and continue with an unbalanced sample. At most we therefore observe 36 crises. We follow the dating of Reinhart and Rogoff (2008) and for the most recent episode the broadest crises definition as used in Borio and Drehmann (2009). To fully check the robustness of the results, we also do not excluded crises which were only driven by cross border exposures in the last global crisis episode (Germany, Switzerland and Sweden). In these cases, it is however hard to argue that they should be included. Why for instance should an indicator of domestic vulnerabilities predict the near failure of one systemic relevant bank in Switzerland because of the banks' business in the US?

Graph 6 highlights that our results are very robust to changing the crisis dating. As we include cross-border crises, it is unsurprising that the predictive ability of most indicators decreases somewhat. However, even for the DSR, where this is most apparent, the results remain extremely strong. The ARUOC is still close to 0.9 in the quarters immediately before crises.

What is also interesting from Graph 6 is that differences for the residential property price indicators and equity price indicators are smallest. One explanation could be that in all cases asset prices were exuberant, but crises only turned really systemic if leverage and the DSR reach very high levels.

Graph 6: AUROC over time with different crisis dates



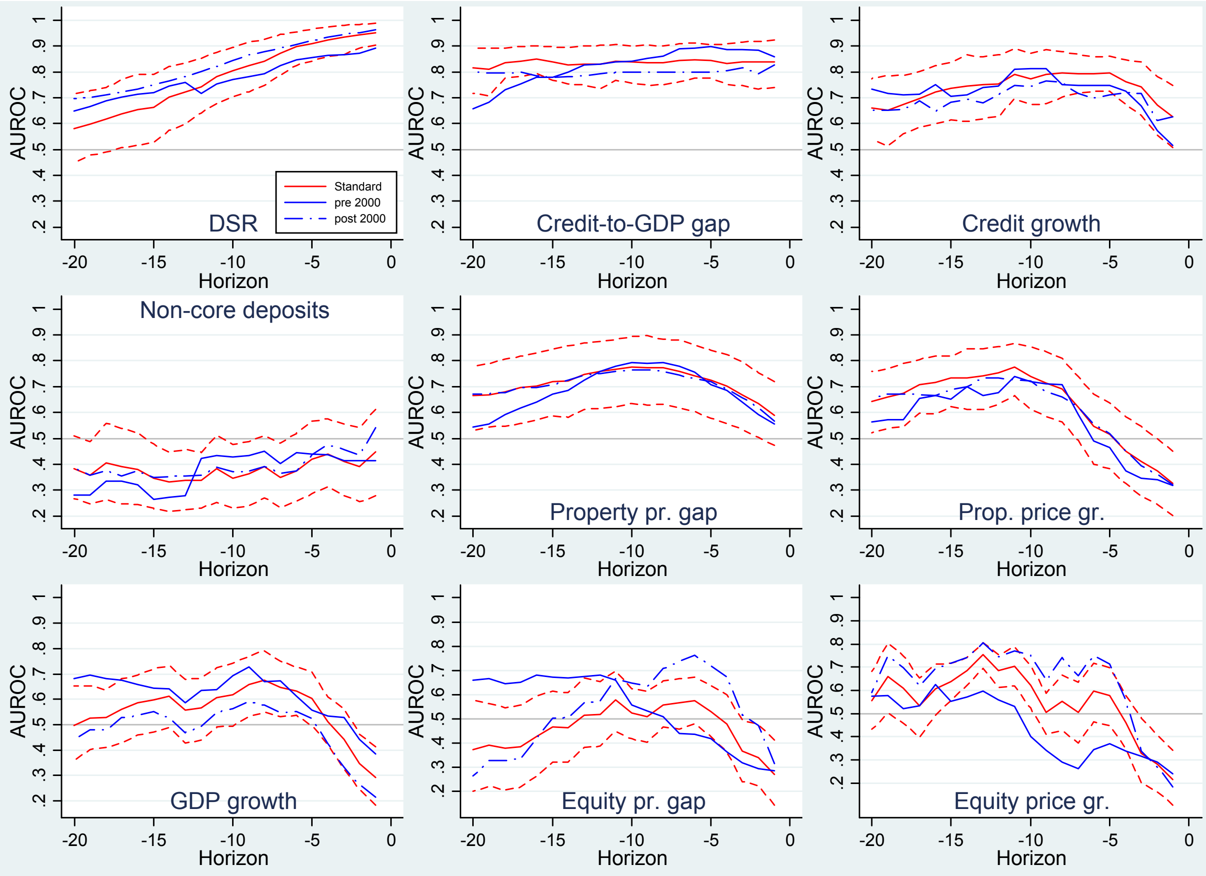
Note: dotted red lines: 95% confidence intervals of AUROC for the standard estimation. Horizon: quarters before a crisis.

4.4.2 Samples

In this section , we assess whether the results are robust to using time periods. We split the sample in roughly equal parts before and after 2000. We return to our standard crisis definition but continue to use the unbalanced sample to have observations before 2000.

Graph 7 shows that for most variables the ROC curves virtually independent of the sample used. The only important differences arise for equity price indicators, where the gap prior to the crisis seemed to have some predictive power pre 2000 in line with the findings of Borio and Lowe (2002). However, this may reflect the fact that equity and property price cycles were more synchronised then than now (see Borio and McGuire (2004))

Graph 7: AUROC over time for different samples



Note: dotted red lines: 95% confidence intervals of AUROC for the standard estimation. Horizon: quarters before a crisis.

4.4.4 Partial AUROCs

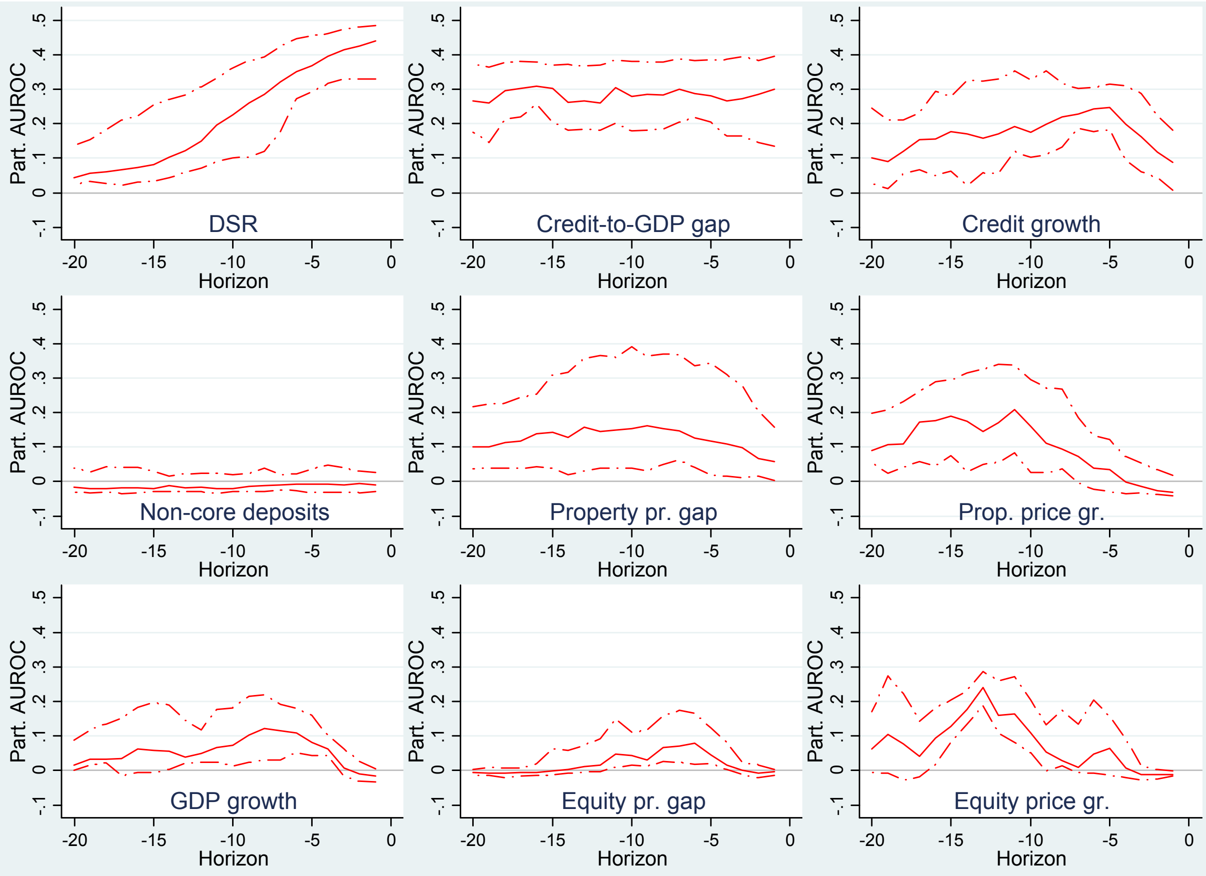
In some of our own previous work (eg Borio and Drehmann (2009)), we have argued that an ideal EWI should signal at least 66% of crisis, and conditional on this should have as little false positives as possible. The argument was primarily used for deriving the optimal threshold θ for different indicator variables and alternative approaches in the literature, such as minimising the noise-to-signal ratio as in Kaminsky and Reinhart (1999), weighting the sum of type I and type II errors (eg Demirgüc-Kunt and Detragiache (1998) or more judgmental based approaches (eg Borio and Lowe (2002)). As it is clear from the discussion in Section 2, an optimal utility based approach would deliver different θ for each variable. However, given that policymakers' preferences are not know and that past crises have been very costly in terms of GDP, a threshold of predicting at least 66% of the crisis was judged to provide the appropriate trade-off for the indicator variables considered.

Rather than focusing on one particular θ – which given the discussion in Section 2 would ultimately depend on policymakers preferences – we want to derive partial AUROCs in this section requiring that at least 66% of crises are predicted.[25] As the area under the 45 degree line is not always ½ but depends on the cut-off point, we only calculate the area under the ROC curve which is above the 45 degree line.

[25] Estimates are based on the balanced sample and use the tight definition of crises.

Unsurprisingly, Graph 8 shows that the messages of the partial ROC curves essentially mirror the full AUROC results. The DSR and the credit-to-GDP gap remain the best single best indicators.

Graph 8: Partial AUROCs



Note: Solid line: partial AUROC capturing at least 66% of crises and considering only the area between the 45 degree line and the ROC curve. Dotted lines: 95% confidence intervals. Horizon: quarters before a crisis.

4. Conclusions

In this paper, we argue that the statistical procedures used to construct and evaluate EWIs should be aligned with the requirements of policymakers. Among the considerations which seem particularly important are the relative trade-offs between type I and type II errors, as well as the timing and consistency of the EWI signals. Because data which could be used to pin down more narrow ranges for the policymakers' preferences are unavailable, we employ a technique which does not assume particular policy preferences – the ROC curve and its associated AUROC value – for assessing the performance of different EWIs. We also make novel use of these techniques to assess the timing and consistency of the EWIs over time.

We find that a new measure of private sector indebtedness, the DSR, as well as the credit-to-GDP gap, significantly dominate the other EWIs. Both of these variables also perform well over time, with the DRS dominating at shorter horizons and the credit-to-GDP gap dominating at longer horizons. The other EWIs have less stable temporal performance and are generally dominated by the DSR and the credit-to-GDP gap.

Our findings indicate that EWIs can have vastly different performance over time. Given the importance of consistent and timely signals for policymakers, this aspect cannot be ignored.

The ROC curve offers a convenient way of assessing the signalling quality of EWIs over time which is not dependent of knowledge about the exact preferences of policymaker's. Hence, we conclude that it constitute an additional useful tool for policymakers.

Bibliography

Alfaro, R and M Drehmann (2009): "Macro stress tests and crises: What can we learn?", *BIS Quarterly Review,* December.

Angeloni, I and E Faia (2009): "Capital regulation and monetary policy with fragile banks", *Kiel Working Papers*, no 1569.

Basel Committee on Banking Supervision (2010a): *An assessment of the long-term economic impact of stronger capital and liquidity requirements*.

Basel Committee on Banking Supervision (2010b): *Guidance for national authorities operating the countercyclical capital buffer*.

Berge, T J and O Jorda (2011): "Evaluating the classification of economic activity into recessions and expansions", *American Economic Journal: Macroeconomics,* 3, 246-277.

Bisias, D, M D Flood, A W Lo and S Valavanis (2012): "A survey of systemic risk analytics", *U.S. Department of Treasury, Office of Financial Research Working Paper* 1.

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

Borio, C and P Lowe (2002): "Asset prices, financial and monetary stability: Exploring the nexus", *BIS Working Papers* 114.

Borio, C and P McGuire (2004): "Twin peaks in equity and housing prices?", *BIS Quarterly Review*, March, pp 79–93.

Bussiere, M and M Fratzscher (2006): "Towards a new early warning system of financial crises", *Journal of International Money and Finance,* 25, 953-973.

Cecchetti, S G, M Kohler and C Upper (2009): "Financial crises and economic activity", *Paper persented at the Federal Reserve Bank of Kansas City symposium at Jackson Hole, August 2009*.

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

Drehmann, M (2012): "How often should macroprudential tools be used?", *Mimeo*.

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

Drehmann, M and M Juselius (2012): "Do debt service costs affect macroeconomic and financial stability?", *BIS Quarterly Review,* September, 21-34.

Edge, R M and R R Meisenzahl (2011): "The unreliability of credit-to-gdp ratio gaps in real-time: Implications for countercyclical capital buffers", *International Journal of Central Banking,* 7.

Gourinchas, P-O and M Obstfeld (2012): "Stories of the twentieth century for the twenty-first", *American Economic Journal: Macroeconomics,* 4, 226-265.

Gertler, M and P Karadi (2011): "A model of unconventional monetary policy," *Journal of Monetary Economics*, no 58(1); pp 17–34.

Hahm, J-H, H S Shin and K Shin (2012): "Non-core bank liabilities and financial vulnerability", *mimeo*.

Hodrick, R J and E Prescott (1981): "Post-war u.S. Business cycles: An empirical investigation", *Northwestern University, Center for Mathematical Studies in Economics and Management Science, Discussion Papers: 451*.

IMF, BIS and FSB (2009): *Report to g20 finanice ministers and governors. Guidance to assess the systemic importance of financial institutions, markets and instruments: Initial considerations*.

Jorda, O (2011): "Anchoring countercyclical capital buffers: The role of credit aggregates: Discussion", *International Journal of Central Banking,* 7, 241-259.

Jorda, O, M Schularick and A M Taylor (2011): "When credit bites back: Leverage, business cycles and crises", *Federal Reserve Bank of San Francisco Working Paper Series* 2011-27.

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

Laeven, L and F Valencia (2012): "Systemic banking crises database: An update", *IMF Working Paper* WP/12/163.

Lawrence, M, M O'Connor and B Edmundson (2000): "A field study of sales forecasting accuracy and processes", *European Journal of Operational Research,* 122, 151-160.

Macroeconomic Assessment Group (2010): *Assessing the macroeconomic impact of the transition to stronger capital and liquidity requirements* Group established by the Basel Committee on Banking Supervision and Financial Stability Board.

Minsky, H P (1982): *Can it happen again? Essays on instability and finance*, M E Sharpe, Armonk.

Orphanides, A (2003): "Monetary policy evaluation with noisy information", *Journal of Monetary Economics,* 50, 605-631.

Ravn, M O and H Uhlig (2002): "On adjusting the hodrick-prescott filter for the frequency of observations", *Review of Economics and Statistics,* 84, 371-376.

Reinhart, C M and K Rogoff (2008): "Banking crises: An equal opportunity meace", *NBER Working Paper* 14587.

Reinhart, C M and K S Rogoff (2009): *This time is different: Eight centuries of financial folly*, Princeton University Press, Princeton and Oxford.

Annex: Additional tables

Table A1: Crisis dates

Country	Domestic systemic crises	Additional crises
Australia		1989q4
Australia		2008q4
Belgium	2008q4	
Czech Republic*	1996q1	
Denmark		1987q4
Denmark	2008q4	
Finland	1991q3	
France		1994q1
France	2008q4	
Germany		2007q3
Greece	2008q4	
Hungry*	1991q2	
Hungry*	2008q4	
Ireland	2008q4	
Italy	1992q3	
Italy	2008q4	
Japan*	1992q4	
Korea	1997q3	
Malaysia*	1997q3	
New Zealand*		1987q1
Norway	1990q4	
Portugal	2008q4	
South Africa		1989q4
Spain	2008q4	
Sweden	1991q3	
Sweden		2008q4
Switzerland*	1991q3	
Switzerland		2007q3
Thailand*	1983q4	
Thailand*	1997q3	
The Netherlands	2008q4	
UK	1990q2	
UK	2007q3	
US	1990q2	
US	2007q3	

* Crisis is not part of the balanced sample. Domestic systemic crises are dated in line with Laeven and Valencia (2012) and discussions with central banks. Additional crises are based on Reinhart and Rogoff (2008) and Borio and Drehmann (2009).

Table A2: AUROC for different horizons

			-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12	-13	-14	-15	-16	-17	-18	-19	-20
Credit-to-GDP gap	**AUROC**		**0.84**	**0.84**	**0.84**	**0.83**	**0.84**	**0.85**	**0.84**	**0.83**	**0.84**	**0.84**	**0.84**	**0.83**	**0.83**	**0.83**	**0.84**	**0.85**	**0.84**	**0.84**	**0.81**	**0.81**
	95% conf.	lower	0.74	0.73	0.75	0.75	0.78	0.78	0.76	0.75	0.75	0.76	0.77	0.73	0.75	0.75	0.77	0.8	0.78	0.78	0.71	0.72
	intervall	upper	0.92	0.92	0.92	0.91	0.91	0.91	0.91	0.9	0.9	0.9	0.91	0.9	0.9	0.9	0.9	0.9	0.9	0.89	0.89	0.89
Credit growth	**AUROC**		**0.63**	**0.67**	**0.74**	**0.76**	**0.8**	**0.79**	**0.79**	**0.8**	**0.79**	**0.77**	**0.79**	**0.75**	**0.75**	**0.75**	**0.74**	**0.72**	**0.7**	**0.68**	**0.65**	**0.66**
	95% conf.	lower	0.51	0.56	0.63	0.67	0.73	0.72	0.72	0.7	0.68	0.67	0.7	0.63	0.62	0.61	0.61	0.6	0.59	0.56	0.51	0.54
	intervall	upper	0.75	0.78	0.84	0.85	0.86	0.86	0.87	0.88	0.89	0.87	0.89	0.87	0.86	0.87	0.84	0.82	0.8	0.79	0.79	0.77
DSR	**AUROC**		**0.95**	**0.94**	**0.93**	**0.92**	**0.91**	**0.9**	**0.87**	**0.84**	**0.83**	**0.8**	**0.78**	**0.74**	**0.72**	**0.7**	**0.66**	**0.65**	**0.64**	**0.62**	**0.6**	**0.58**
	95% conf.	lower	0.9	0.89	0.87	0.86	0.84	0.83	0.79	0.75	0.72	0.7	0.68	0.64	0.6	0.57	0.53	0.52	0.51	0.49	0.48	0.45
	intervall	upper	0.99	0.98	0.98	0.97	0.97	0.95	0.95	0.92	0.92	0.89	0.88	0.85	0.83	0.82	0.79	0.79	0.77	0.74	0.73	0.71
Equity price gap	AUROC		**0.27**	**0.34**	**0.37**	**0.48**	**0.53**	**0.57**	**0.57**	**0.56**	**0.51**	**0.52**	**0.58**	**0.52**	**0.51**	**0.46**	**0.47**	**0.43**	**0.39**	**0.38**	**0.39**	**0.37**
	95% conf.	lower	0.14	0.22	0.24	0.36	0.43	0.48	0.46	0.47	0.4	0.42	0.45	0.39	0.38	0.32	0.32	0.26	0.22	0.21	0.22	0.2
	intervall	upper	0.41	0.48	0.49	0.6	0.64	0.67	0.67	0.66	0.61	0.62	0.7	0.65	0.67	0.61	0.61	0.6	0.56	0.55	0.56	0.58
Equity price growth	**AUROC**		**0.21**	**0.28**	**0.33**	**0.46**	**0.58**	**0.6**	**0.51**	**0.55**	**0.5**	**0.62**	**0.7**	**0.69**	**0.76**	**0.68**	**0.64**	**0.61**	**0.53**	**0.61**	**0.66**	**0.56**
	95% conf.	lower	0.11	0.16	0.2	0.35	0.45	0.46	0.37	0.43	0.41	0.52	0.62	0.61	0.7	0.62	0.56	0.49	0.4	0.46	0.5	0.43
	intervall	upper	0.34	0.41	0.48	0.58	0.7	0.72	0.63	0.67	0.59	0.71	0.79	0.75	0.8	0.74	0.72	0.71	0.65	0.75	0.8	0.68
GDP growth	**AUROC**		**0.29**	**0.35**	**0.44**	**0.51**	**0.6**	**0.63**	**0.65**	**0.67**	**0.66**	**0.62**	**0.61**	**0.57**	**0.56**	**0.61**	**0.6**	**0.59**	**0.56**	**0.53**	**0.53**	**0.5**
	95% conf.	lower	0.18	0.24	0.33	0.42	0.5	0.54	0.53	0.55	0.53	0.5	0.49	0.44	0.43	0.49	0.47	0.46	0.43	0.41	0.4	0.36
	intervall	upper	0.41	0.46	0.56	0.61	0.71	0.73	0.75	0.79	0.77	0.74	0.72	0.68	0.68	0.73	0.72	0.7	0.68	0.64	0.65	0.65
Non-core deposits	**AUROC**		**0.45**	**0.39**	**0.41**	**0.44**	**0.42**	**0.37**	**0.35**	**0.39**	**0.36**	**0.35**	**0.38**	**0.34**	**0.34**	**0.33**	**0.35**	**0.38**	**0.39**	**0.41**	**0.36**	**0.38**
	95% conf.	lower	0.28	0.25	0.28	0.31	0.29	0.26	0.23	0.27	0.24	0.23	0.25	0.23	0.22	0.22	0.23	0.24	0.25	0.26	0.25	0.27
	intervall	upper	0.61	0.54	0.55	0.58	0.57	0.52	0.48	0.51	0.49	0.48	0.51	0.45	0.46	0.45	0.48	0.52	0.54	0.56	0.49	0.51
Property price gap	**AUROC**		**0.59**	**0.63**	**0.67**	**0.7**	**0.72**	**0.74**	**0.76**	**0.77**	**0.77**	**0.77**	**0.77**	**0.76**	**0.75**	**0.72**	**0.72**	**0.7**	**0.7**	**0.68**	**0.67**	**0.67**
	95% conf.	lower	0.47	0.5	0.53	0.56	0.58	0.6	0.62	0.63	0.63	0.63	0.62	0.62	0.61	0.58	0.59	0.57	0.56	0.55	0.54	0.53
	intervall	upper	0.72	0.75	0.8	0.82	0.84	0.86	0.88	0.88	0.9	0.89	0.88	0.88	0.87	0.86	0.85	0.83	0.82	0.81	0.79	0.78
Property price growth	**AUROC**		**0.33**	**0.37**	**0.41**	**0.45**	**0.51**	**0.55**	**0.62**	**0.69**	**0.71**	**0.74**	**0.78**	**0.75**	**0.74**	**0.73**	**0.73**	**0.72**	**0.71**	**0.67**	**0.66**	**0.64**
	95% conf.	lower	0.2	0.25	0.28	0.33	0.38	0.4	0.49	0.56	0.58	0.61	0.66	0.63	0.61	0.61	0.62	0.59	0.6	0.55	0.54	0.52
	intervall	upper	0.45	0.5	0.55	0.58	0.64	0.67	0.74	0.81	0.83	0.85	0.87	0.85	0.85	0.85	0.82	0.82	0.81	0.79	0.77	0.76

Horizon: quarters before crisis