

# Identifying excessive credit growth and leverage

Lucia Alessi and Carsten Detken\*

## Abstract

This paper aims at providing policymakers with a set of early warning indicators helpful in guiding decisions on when to activate macroprudential tools targeting excessive credit growth and leverage. To robustly select the key indicators we apply the “Random Forest” method, which bootstraps and aggregates a multitude of decision trees. On these identified key indicators we grow a binary classification tree which derives the associated optimal early warning thresholds. By using credit to GDP gaps, credit to GDP ratios and credit growth rates, as well as real estate variables in addition to a set of other conditioning variables, the model not only predicts banking crises but is also able to give an indication on which macro-prudential policy instrument would be best suited to address specific vulnerabilities.

**Keywords:** Early Warning Systems, Banking Crises, Macroprudential Policy, Decision Trees, Random Forest.

**JEL Classification** C40 · G01 · E44 · E61 · G21.

---

\*Both European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany. Emails: lucia.alessi@ecb.europa.eu; carsten.detken@ecb.europa.eu. We thank Nadya Jahn for excellent research assistantship in the construction of debt service ratios.

# 1 Introduction

Past financial crises and in particular the global financial crisis have shown that excessive credit growth often leads to the build-up of systemic risks to financial stability, which may materialize in the form of systemic banking crises. As mitigating systemic financial stability risks is the objective of macroprudential policy, several macroprudential tools have been designed to curb excessive leverage and/or build-up buffers against likely future losses.<sup>1</sup> Such instruments include the countercyclical capital buffer, the systemic risk buffer as well as a potentially time-varying leverage ratio, and instruments directly targeting borrowers such as loan-to-value (LTV) and loan-to-income (LTI) caps.<sup>2</sup>

However, the application of macroprudential policy is still at an early stage and much effort is currently being devoted to providing policymakers with concrete advice on how to actually design macroprudential instruments. Indeed, the macroprudential policy strategy has been defined by the European Systemic Risk Board (ESRB) with reference to the *guided discretion*

---

<sup>1</sup>As it is common in the macro-financial literature (see Section 2), this paper defines leverage as the ratio of a credit aggregate to GDP at the country level, while the micro-financial concept of leverage corresponds to debt divided by equity. Leverage in banking is the ratio of lending to equity and is indeed affected by some macroprudential measures. The broader definition of leverage used in this paper covers non-financial-corporations and household debt, i.e. a country's total private sector leverage. We use this definition of leverage to indicate the level of debt, as opposed to the concept of credit growth (and gap).

<sup>2</sup>In Europe, the countercyclical and the systemic risk buffers are regulated at the EU level while LTV and LTI limits as well as the leverage ratio are currently based on national law.

principle, whereby the exercise of judgement is complemented by quantitative information derived from a set of selected indicators and associated ‘early warning’ thresholds. In particular, with respect to the countercyclical capital buffer, already the Basel Committee on Banking Supervision (BCBS) identified the aggregate private sector credit-to-GDP gap as a useful buffer guide, as this variable would have performed well in signalling the build-up of excessive leverage in the past.<sup>3</sup> However, policymakers should supplement the signal coming from credit-to-GDP trend deviations with judgement based on a broader information set, as implicitly suggested also in the current Capital Requirements Directive (CRD IV), which tasks the ESRB to provide recommendations on other variables which should inform the policy decision. Taking into account other conditioning variables is necessary because not all credit expansions are bad for financial stability, and the heroic task of identifying credit bubbles in real time requires assessing whether conjunctural credit developments might be disconnected from fundamentals or reflect excessive risk taking and overly optimistic expectations.

Against this background, we propose an early warning model to be used for identifying those periods in which the build-up of leverage can be defined as excessive and may warrant the activation of relevant macroprudential instruments. In our analysis we consider several variables as a policy guide, select the most relevant ones on the basis of a robust quantitative assess-

---

<sup>3</sup>See also Detken et al. (2014) providing evidence for the good performance of the credit-to-GDP gap for the EU as a whole.

ment of their predictive power, and propose a fully-fledged system where the key indicators and the respective early warning thresholds are considered in a unified framework. The benchmark model we derive is a transparent tool which would also enable the public at large to understand and possibly anticipate macroprudential decisions.

We achieve our objective by using decision tree learning, a statistical methodology which retains the advantages of the two approaches traditionally used in the Early Warning literature, i.e. the signalling and the discrete choice approach. The model we develop aims at identifying whether the European financial system is in a given period particularly vulnerable, a situation in which the increased likelihood and importance of a subsequent banking crisis would suggest to build-up capital buffers and/or to curb credit growth. The paper is structured as follows. The next section reviews the related literature on macroprudential tools, in particular the countercyclical capital buffer, and economic applications of recursive trees. In Section 7 we define our target variable, i.e. broadly speaking banking crises in the European Union in the last 40 years. Section 4 describes our candidate early warning indicators. Section 5 outlines the Classification Tree approach and its extension to Random Forests. The results of the empirical analysis using the Random Forest approach are presented in Section 6 and compared to the results from logit models in Section 7, while 8 illustrates the benchmark Early Warning Tree. Section 9 reports for which countries the tree would issue early warning signals and why, while Section 10 describes the results of

an out-of-sample exercise using only pre-2007 information. The policy implications of our findings are discussed in 11. Section 12 summarizes the main conclusions.

## 2 Review of the Literature

The literature on Early Warning Systems for banking crises has a long tradition (see e.g. Eichengreen and Rose (1998)). However, it has so far focused mostly on emerging markets and on identifying banking crises determinants without an explicit focus on the policy tools intended to reduce the likelihood and severity of such crises. The recent financial crisis and the subsequent policy responses have spurred the efforts towards providing policymakers with concrete indications on how to actually design macroprudential instruments.

Countercyclical capital buffers (CCBs) are one of the main tools envisaged by Basel III and the one on which the analytical framework is most advanced. The countercyclical capital buffer is designed to increase the resilience of the banking sector and smooth the credit cycle, e.g. in ensuring that the flow of credit is not unnecessarily reduced due to pro-cyclical supply side constraints during a bust phase. BCBS (2010) states that the authorities responsible for operating CCBs should follow a common reference guide, based on the aggregate private sector credit-to-GDP gap. Indeed, Drehmann et al. (2010) and Drehmann et al. (2011) show that deviations of the credit to GDP ratio from a long term trend actually outperform other candidate early warning

indicators such as GDP and credit growth, their ratio as such, as well as indicators based on asset prices or measures of banking sector performance. The credit-to-GDP gap, however, suffers from some shortcomings: among others, it may provide misleading signals in real-time as it is prone to large revisions (Edge and Meisenzahl (2011)). This is mainly due to the end-point bias affecting the one-sided Hodrick-Prescott filter, which is widely used to extract the long-term trend. Moreover, this filter is sensitive to the choice of the smoothing parameter, and adjusts very slowly following during a reversal after a prolonged period of negative credit growth. Finally, positive deviations from trend could be due to either excessive credit growth or low or negative output growth, two scenarios which arguably require different policy responses (Repullo and Saurina (2011)).<sup>4</sup> Owing to the limitations of the credit-to-GDP gap, it is advisable to complement it with other early warning indicators, ideally in a multivariate framework.

Other capital-based instruments targeting excessive leverage are the leverage ratio and the systemic risk buffer. The former aims at addressing risks directly linked to excessive leverage, namely losses occurring in the wake of fire sales and adjustments in asset valuation. The latter is envisaged to increase resilience in the banking sector by addressing structural systemic risks like the size of the banking sector compared to the rest of the economy. Hardly any applied research is available on the use of the leverage ratio for

---

<sup>4</sup>For a discussion of the measurement problems related to the credit-to-GDP gap, see Drehmann and Tsatsaronis (2014).

macroprudential purposes or on the systemic risk buffer. With respect to this latter, one of the biggest challenges is related to the notion of structural systemic risk itself, which is in practice open to interpretations and difficult to measure in an empirical exercise (see Borio and Drehmann (2009)).

With respect to instruments targeting borrowers, the literature suggests some indicators which could be taken into consideration when deciding whether to impose limits to loan-to-value and loan-to-income ratios, e.g. to prevent a credit boom fuelling an asset price bubble. Quite naturally, these indicators are mainly related to house prices (see e.g. Barrell et al. (2010), Borio and Drehmann (2009) and Mendoza and Terrones (2008)). Due to poor commercial property price data coverage and quality and owing to cross-country comparability issues with respect to LTI and LTV ratios themselves, assessing the ‘early warning’ performance of these promising indicators has been so far very challenging.

The multivariate methodology we propose to adopt to support decisions on the macroprudential instruments described above is decision tree learning, a greatly underutilized technology in economics. Indeed, while Classification and Regression Trees (CARTs, see L. Breiman and J. Friedman and R. Olshen and C. Stone (1984)) are extensively used in other disciplines from biology to chemometrics, their economic applications are rare. The Early Warning literature, in particular, has so far almost uniquely relied on two approaches, namely the signalling approach and the categorical dependent variable regression. The signalling approach has the advantage of being ex-

tremely straightforward.<sup>5</sup> Indeed, the early warning signal is issued when the considered indicator breaches a pre-specified threshold, set by optimizing the past predictive performance. The downside of this approach is that it considers early warning indicators separately. Logit/probit regression, contrary to the signalling approach, offers a multivariate framework within which one can assess the relative importance of several factors.<sup>6</sup> However, while a desirable feature of an early warning system is to provide clear early warning thresholds for the considered indicators, the logit/probit model offers only an estimate of the contribution of each factor to the increase in the overall probability of a crisis, rather than a threshold value for each regressor. The early warning threshold for the estimated crisis probability is eventually set in a second step and outside of the logit/probit model itself. Moreover, this framework, the way it is commonly applied, is unable to handle unbalanced panels and missing data, which is a serious issue in particular with credit data, with the result that the regression can ultimately be estimated only on a relatively short sample. Decision trees, and classification trees in particular, retain the advantages of both approaches as they are on the one hand very easy to explain and use, and on the other hand able to provide an early warning *system* where the relevant indicators are considered in a unitary framework. Moreover, decision trees are not sensitive to outliers and can handle nonstationary time series, as the time dimension is not relevant in

---

<sup>5</sup>See e.g. Kaminsky and Reinhart (1999) and more recently Alessi and Detken (2011).

<sup>6</sup>Among the latest works, see e.g. Lo Duca and Peltonen (2013) on systemic risks and Behn et al. (2013) on CCBs.



such a framework. We are aware of only a handful of papers using binary recursive trees for assessing vulnerabilities in relation to financial crises: Gosh and Gosh (2002) and Frankel and Wei (2004) analyze the determinants of currency crises, Manasse and Roubini (2009) and Savona and Vezzoli (2014) deal with sovereign crises, while Dutttagupta and Cashin (2011) and Manasse et al. (2013) study banking crises in emerging markets. Similarly to this latter paper and to Savona and Vezzoli (2014), the present study grows the benchmark tree on the solid ground of a preliminary analysis based on bootstrapping and aggregating a multitude of trees. However, our explicit objective is to provide a set of triggers for macroprudential policy instruments in the European Union, therefore our crisis episodes and the countries considered are carefully selected accordingly. Moreover, as we adopt a strict policy perspective, we aim at a model that allows for timely decision making and therefore focus on identifying pre-crisis periods rather than crisis periods (see Section 7).

### **3 The Banking Crises Dataset**

The basis for the banking crises dataset used in this paper is provided by the dataset assembled by Babecky et al. (2012). This quarterly dataset covers, inter alia, banking crisis episodes in EU countries over 1970-2010. The authors do not provide a unique definition of banking crisis: rather, they derive banking crisis episodes by aggregating the information about crisis

occurrence coming from other works and an ad-hoc survey among country experts mainly in national central banks. The definitions of banking crisis in the source papers cover the following: i) ‘episodes in which much or all of bank capital was exhausted’ (Caprio and Klingebiel (2003)); ii) ‘bank runs that lead to the closure, merger, or takeover by the public sector of one or more financial institutions’ as well as ‘the closure, merging, takeover, or large-scale government assistance of an important financial institution (or group of institutions) that marks the start of a string of similar outcomes for other financial institutions’ (Kaminsky and Reinhart (1999)); iii) ‘significant signs of financial distress in the banking system (as indicated by significant bank runs, losses in the banking system, and/or bank liquidations)’ as well as ‘significant banking policy intervention measures in response to significant losses in the banking system’, where the considered measures include extensive liquidity support, bank restructuring costs, significant bank nationalizations, significant guarantees put in place, significant asset purchases, deposit freezes and bank holidays (Laeven and Valencia (2008), (2010), (2012)).

Neither of the above definitions of banking crisis, however, is fully aligned with the objective and operation of the macroprudential tools targeting credit, as they aim to avoid a broader array of circumstances than simply a banking crisis as defined in these terms alone. Therefore, we use an updated and slightly amended dataset with respect to the one constructed by Babecky et al. (2012), which has been built in the framework of a broader project by the European Systemic Risk Board on the basis of country ex-

perts' judgement (see Detken et al. (2014)). In this dataset, a banking crisis is defined by significant signs of financial distress in the banking system as evidenced by bank runs in relevant institutions or losses in the banking system (nonperforming loans above 20% or bank closures of at least 20% of banking system assets); or significant public intervention in response to or to avoid the realization of losses in the banking system (see above). Most importantly, non-systemic crises have been excluded, as well as systemic banking crises that had no association with a domestic credit/financial cycle. Moreover, a value of 1 to the binary crisis variable has been assigned to those periods in which 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 credit cycle.<sup>7</sup> The target variable used in this analysis thus captures: (i) systemic banking crises associated with a domestic credit/financial cycle; (ii) periods in which in the absence of policy action or of an external event that dampened the credit cycle a crisis as in (i) would likely have occurred.

The data cover all 28 EU members from 1970Q1 till 2012Q4. However, we have extended the coverage to 2013Q4, while limiting our analysis to euro area countries together with the UK, Denmark and Sweden. We excluded

---

<sup>7</sup>In particular, four episodes of near-crisis events have been added, namely: Bulgaria Q4/2004-Q2/2007, Netherlands Q1/2002-Q3/2003, Portugal Q1/1999-Q1/2000 and Germany Q1/2000-Q4/2003 due to strong credit cycles during these periods. 15 banking crises have been deleted from the original databank: one in Austria, Belgium, Czech Republic, Ireland, Luxembourg and Slovakia; two in Estonia, Latvia and the UK; and three in Germany. Among the latter is included e.g. the 1974 Herstatt failure, which was due to settlement risk materialising. We refer to Detken et al. (2014) for further details.

Central and Eastern European transition economies as their data series are generally relatively short, implying that the overall results would be driven by the evidence linked mainly to the global financial crisis, and in some cases exhibit peculiar patterns which would warn against pooling these countries together with the ones under study. The coverage of banking crises dataset constructed by the ESRB prevented us from extending the analysis to other advanced economies. Over the considered period, 25 separate crisis episodes are recorded for euro area countries, the UK, Denmark and Sweden. They are marked in black in Chart 1. While the incidence of crises shows a marked increase for the current financial crisis, only slightly more than half of the 21 country experts thought that for their country the current crisis met one of the above criteria. Moreover, some countries (Austria, Belgium, Luxembourg, Malta and Slovakia) did not record any crisis consistent with the above criteria over the sample period. Of the remaining countries, 8 experienced one crisis, 7 experienced two crises while the UK experienced three crises.

Finally, in constructing our binary target variable we take into account policy lags. For example, with respect to CCBs, banks should usually be given at least one year time to meet the additional capital requirements before any increases in the buffer take effect. An early warning signal leading the inception of the crisis by less than one year, or once the crisis is already in place, would be late. At the same time, we do not aim at building a model which predicts exactly when a banking crisis will materialize. Rather, we propose an Early Warning System signalling that financial imbalances

are building up and the risk of a systemic crisis in the not-so-far future is increasing. Therefore, we define as correct any warning signals issued in the four years preceding the start of a crisis, excluding from the analysis the three quarters immediately preceding the crisis and the crisis period itself. The pre-crisis periods are marked in red in Chart 1, while the periods excluded from the analysis are marked in grey. We do not remove from the sample the quarters following the crisis because our model is not expected to suffer from any post-crisis bias.<sup>8</sup> With the exception of the Spanish and Cypriot crises, the period after 2009Q1 is de-facto not taken into account while optimizing the early warning thresholds because the dataset ends in 2012Q4 and ignores whether a crisis happened in any of the countries in 2013.

## 4 Early Warning Indicators

For the reasons described in Section 2, it makes sense to monitor a broader set of variables for macroprudential decisions. In this paper, we examine a battery of indicators which could contain valuable information. In particular, we consider financial and macroeconomic variables, as well as real-estate based indicators.

With respect to credit related indicators, the key aggregate is broad credit. In this respect, we use a broad credit aggregate compiled by the

---

<sup>8</sup>See Bussière and Fratzscher (2006), who show that the econometric results of binomial logit early warning models are at least in part explained by the behavior of the independent variables during and directly after a crisis, i.e. periods which are often disorderly and volatile corrections towards longer-term equilibria.

BIS (see Dembiermont et al. (2013)), which covers credit from all sources, including debt securities, to the non-financial private sector. We consider the y-o-y rate of growth, as well as the ratio to GDP and the deviations of such ratio from its trend (i.e. the ‘gap’), computed with a backward-looking slowly-adjusting ( $\lambda = 400000$ ) HP filter. This latter transformation assumes that the financial cycle is four times as long as the business cycle and has been suggested by BCBS (2010) - we’ll therefore refer to it as the “Basel gap”. However, such an HP trend might be adjusting too slowly following a prolonged period of negative credit growth, therefore we also consider an alternative gap computed with  $\lambda = 26000$ , corresponding to a financial cycle which is twice as long as the business cycle. We also look at the narrower bank credit aggregate, which we analogously consider as y-o-y rate of growth, ratio to GDP and gap.<sup>9</sup> The level of bank loans as a ratio to GDP is one of the indicators Schularick and Taylor (2012) take as evidence of a *story of decades of slowly encroaching risk on bank balance sheets*: by including it in our model we aim at exploiting the panel dimension in order to pin-down an ‘early warning’ level of aggregate leverage.<sup>10</sup> With respect to the time

---

<sup>9</sup>Rates of growth are deflated by subtracting the y-o-y CPI changes. Gaps have been constructed by taking a standard HP filter for the first 5 years of available data and then a recursive HP filter. Although it is advisable to only use gaps after 5-10 years of data due to the start point problem affecting HP trend estimates (see Borio and Lowe (2002)), such an approach would have yielded too short time series. As a result, the evaluation of the predictive performance of gap measures would have been driven mainly by the recent global financial crisis. Also based on the results by Drehmann and Tsatsaronis (2014), who analyze the potential practical consequences of the start point bias, we decided in favor of keeping the longest possible time series.

<sup>10</sup>Other indicators studied by Schularick and Taylor (2012) are e.g. the ratios of bank assets to GDP and money, which we do not analyze owing to lack of long enough quarterly

dimension, it could be argued that such an ‘early warning level’ does not make sense for nonstationary series. However, we would argue that the ratio of credit to GDP is theoretically bounded, hence stationary in the long run. Furthermore, our statistical procedure is not affected by ‘spurious regression’ problems. For this reason, we do include credit to GDP levels in the analysis as they serve as conditioning variables for other indicators. Sectoral credit aggregates, namely credit to households and non-financial corporations, are transformed into y-o-y rates of growth, deflated by CPI inflation, and ratios to GDP. The real rate of growth of housing loans is also considered.<sup>11</sup> Global liquidity is included in the form of global credit growth and gaps.<sup>12</sup> We also consider debt service costs. In particular, we use extended debt service ratio (DSR) series with respect to those in Drehmann and Juselius (2012), computed on high-quality (and sometimes confidential) data.<sup>13</sup> We include the bank balance sheet observations.

<sup>11</sup>The source for loans to households for house purchase is the ECB.

<sup>12</sup>Global credit variables are computed as GDP (at PPP) weighted averages of broad credit growth rates and gaps. In particular, global credit growth is constructed by averaging the y-o-y credit growth rates across countries, deflated by subtracting the y-o-y changes of the national CPI. The countries considered for the construction of the global credit variables are the ones under study together with Brazil, Canada, China, Hong Kong, India, Indonesia, Japan, Korea, Mexico, Norway, Russia, Singapore, South Africa, Switzerland, Thailand and the US.

<sup>13</sup>The DSR at time  $t$  is calculated using the standard formula for the fixed debt service costs ( $DSC_t$ ) of an instalment loan and dividing it by income ( $Y_t$ ):

$$DSR_t = \frac{DSC_t}{Y_t} = \frac{i_t D_t}{(1 - (1 + i_t)^{-s_t}) Y_t}$$

where  $D_t$  denotes the aggregate stock,  $i_t$  denotes the average interest rate per quarter on the stock,  $s_t$  denotes the average remaining maturity on the stock and  $Y_t$  denotes quarterly aggregate income. The source for credit aggregates is the BIS, income data are sourced from Eurostat, while lending rates and the average loan maturity are sourced from the ECB (MFI Interest Rate statistics and MFI Balance Sheet Items statistics, respectively).

aggregate DSR as well as sectoral DSRs for non-financial corporations and households. Finally, we include public debt, as a ratio to GDP, in the pool of credit-related indicators.<sup>14</sup>

The macroeconomic variables we examine are real GDP y-o-y growth and the current account in percentage of GDP (on the properties of the current account as an early warning signal for banking crises, see Kauko (2012)). We also consider the M3 money aggregate, in terms of real y-o-y rate of growth and gap, and the real effective exchange rate.<sup>15</sup>

With respect to property prices, house price growth (y-o-y, consumer price deflated) is considered, as well as gap measures. Moreover, we include in the dataset two standard property valuation measures, namely the house price to income ratio and the house price to rent ratio.<sup>16</sup>

Finally, the market-based indicators included in our pool are the long (10 years) and short (3 months) interest rates, both deflated by subtracting the y-o-y CPI changes, as well as the deflated y-o-y growth rate of equity prices.<sup>17</sup>

The dataset goes from 1970:Q1 to 2013:Q4; however, the last 4 years

---

The interest rate is the 3 month average money market interest rate from Eurostat.

<sup>14</sup>Eurostat data.

<sup>15</sup>The main source for real and nominal GDP data is the OECD; Eurostat data have been used whenever OECD series were not available or shorter (i.e. for Cyprus, Estonia, Greece, Latvia, Malta, Slovakia and Slovenia). The source for the current account balance is Eurostat. M3 is provided by the ECB. The real effective exchange rate is sourced from the IMF's IFS and from Eurostat for Estonia, Latvia and Slovenia.

<sup>16</sup>These valuation measures are provided by the OECD in its house price database as indexes and are transformed by subtracting the long-term mean.

<sup>17</sup>Interest rates are sourced from Eurostat, while the source for the stock price indexes is the OECD Main Economic Indicators database.



of data are excluded from the analysis (see previous section). To proxy for publication lags and taking a conservative stand, we lag all the variables by one quarter. In other words, the model aims at classifying the current quarter as pre-crisis or tranquil on the basis of data referring to no later than the last quarter, although some information on conjunctural developments from higher-frequency indicators would already be available in real time.

## 5 Classification Trees and the Random Forest

A binary classification tree is a partitioning algorithm which recursively identifies the indicators and the respective thresholds which are able to best split the sample into the relevant classes, say pre-crisis and tranquil periods. The output of the predictive model is a tree structure like the one shown in Figure 4, with one root node, only two branches departing from each *parent node* (hence “binary” classification tree), each entering into a *child node*, and multiple *terminal nodes* (or “leaves”). Starting by considering all available indicators and threshold levels, the procedure selects the single indicator and threshold yielding the two purest subsamples in terms of some impurity measure. A standard impurity measure, which we also employ, is the Gini index:

$$GINI(f) = \sum_{i=1}^n f_i(1 - f_i) = 1 - \sum_{i=1}^n f_i^2 = \sum_{i \neq j} f_i f_j$$

where  $f_i$  is the fraction of periods belonging to each category  $i$  in a given node, with  $i = 1, 2$  in our case, i.e. pre-crisis and tranquil. The value of the

Gini index will be 0 for a node which contains only observations belonging to the same class. The more mixed a sample is, the higher the Gini index will be, reaching a maximum of 0.25 in the case of two categories. It is possible to generalize the above expression for the Gini index in order to take into account different misclassification costs  $C_{ij}$  for the various classes. The Gini index can then be written as follows:

$$GINI(f) = \sum_{i,j} C_{ij} f_i f_j$$

with  $C_{ii} = 0$  and  $C_{ij}$  reflecting the cost of assigning an observation belonging to category  $i$  to category  $j$ . In our case, for example, it could make sense to be conservative and assume that misclassifying a pre-crisis quarter as tranquil would yield more serious consequences than vice-versa, implying that the cost of a banking crisis is in general larger than the cost of prudential pre-emptive measures. In other words, this would amount to assuming unbalanced policymakers' preferences against missing crises. Asymmetric misclassification costs will also impact the classification of the tree leaves.<sup>18</sup>

Once the first best split is selected, the algorithm proceeds recursively by further partitioning the two subsamples, i.e. finding the best split for each of them. The whole logical structure of the tree is then constructed recursively and the algorithm stops when either some stopping rule becomes binding (e.g. a minimal terminal node size) or there is no further gain from splitting

---

<sup>18</sup>See e.g. Tuffery (2011).

nodes. The resulting tree can be used in real time to map the current value of a set of indicators into a single prediction, expressed as the probability of being in each of the classes. Indeed, rare leaves will contain only observations (i.e. country-quarters in our case) all belonging to the same class. On the contrary, several observations from different classes typically end up in the same leaf.<sup>19</sup> The probability that an out-of-sample observation belongs to a particular class can therefore be computed as the frequency of in-sample observations actually belonging to that class, which ended up in that same leaf while growing the tree. For early warning purposes, it is therefore enough to go down the classification tree, according to the current values of the relevant indicators, to see whether the model foresees an incoming banking crisis. If the policymaker's preferences between missing a crisis (type 1 error) and issuing a false alarm (type 2 error) are balanced, an early warning will be issued if the relevant leaf is associated with a frequency of pre-crisis periods larger than 50%. However, policymakers' preferences after the global financial crisis are likely to have become biased against missing crises, implying a lower threshold.

The main drawback of the tree technology is that, while it can be very good in-sample, it is known not to be particularly robust when additional predictors or observations are included. We overcome this problem by using

---

<sup>19</sup>Theoretically, one can always grow a tree which has enough branches to yield pure leaves, i.e. correctly classify all sample data, unless the data is contradictory in some dimension. However, to avoid overfitting, such a tree should be *pruned* by replacing some parent nodes with leaves.

the Random Forest method proposed by Breiman (2001). This framework is a popular machine learning technique which involves *bagging*, i.e. bootstrapping and aggregating, a multitude of trees. Each of the trees in the forest is grown on a randomly selected set of indicators and country quarters.<sup>20</sup> Analogously to the tree, the forest allows for interaction across the various indicators, is able to handle large datasets, is not influenced by outliers and does not require distributional or parametric assumptions. Once a new quarter of data is available, the prediction of the forest will be based on how many trees in the forest classify it as a pre-crisis or tranquil period, and it will also reflect policymakers' preferences. Each of the trees in the forest is in itself an out-of-sample exercise, as the observations that are not used to grow the tree (so called *out-of-bag* observations) can be put down the tree to get a classification. It is therefore possible to compute the total misclassification error of the forest.

Together with being an extremely powerful predictor, the Random Forest allows to measure the importance of each of the input variables by evaluating the extent to which it contributes to improve the prediction. This is done in practice by randomly permuting the values of the  $n$ -th indicator in the out-of-bag cases, and comparing these tree predictions to those obtained by not permuting the values. If the error rate increases substantially by permuting

---

<sup>20</sup>Following the Random Forest literature, the number of indicators selected for each tree is equal to  $\sqrt{N}$ , where  $N$  is the total number of indicators. At each repetition, 70% of the observations are sampled with replacement. However, the Forest is not very sensitive to the value of these parameters.

the values of an indicator, that means that the indicator does convey relevant information for an accurate classification. If, on the contrary, there is no difference between the two error rates, the indicator is useless.

## 6 Results from the Random Forest

The Random Forest could be used as a regular tool for policy purposes. Indeed, based on the error rate of a 100,000-tree forest we have grown on all of the indicators, the chance of misclassifying an incoming quarter of data is 6%. A standard metrics for the evaluation of the performance of a classifier across a range of preferences is the Area Under the Receiver Operating Characteristic curve (AUROC), the ROC curve plotting the combinations of true positive rate (TPR) and false positive rate (FPR) attained by the model. It is constructed by varying the forest ‘early warning’ threshold, i.e. the required fraction of trees classifying a particular observation as pre-crisis, beyond which that observation will be actually classified as pre-crisis. The ROC curve of a random classifier will tend to coincide with a 45 degree line, corresponding to an AUROC of 0.5, while the AUROC of a good classifier will be closer to 1 than to 0.5. Chart 2 shows the ROC curve of the Random Forest, corresponding to an AUROC above 0.9 (0.94). This result is derived assuming biased policymaker’s preferences against missing crises - in particular, we set misclassification costs such that the cost of misclassifying a pre-crisis quarter is twice as large as the cost of misclassifying a tranquil

quarter - and is robust to assuming balanced preferences.

Notwithstanding the remarkably good performance of the Random Forest, we acknowledge that this is a black-box model and its predictions would be hard to defend, in particular if they would support the activation of a macroprudential instrument. Therefore, in this paper we rely on the Random Forest in order to identify the key indicators, on which we construct our benchmark tree. By doing so, we ensure that the variables selected to grow the tree are truly the most important ones in the pool and we rule out the possibility that the tree selects a relatively weak indicator which just happens to seem useful in-sample but would not survive an out-of-sample robustness check. Chart 3 shows the ranking of the indicators in the forest, with the bars representing a measure of the increase in the classification error associated with randomly permuting the values of the considered indicator across the out-of-bag cases. This measure is computed for every tree, then averaged and divided by the standard deviation over all of the trees.<sup>21</sup>

Not surprisingly, since the model is designed to predict banking crises associated with a domestic credit boom, the most important indicator turns out to be bank credit in the form of its ratio to GDP, followed closely by the gap derived with a very slowly adjusting trend. The level of broad credit

---

<sup>21</sup>Given that the Forest includes an element of randomness, multiple runs of the algorithm on the same dataset won't necessarily yield the same indicators' ranking, in particular if the error associated with different indicators is similar. A robustness check based on several 1000-tree forests indicates that there could be a difference of at most two positions with respect to the ranking illustrated here. The exact ranking is anyway not the focus here, as we are only interested in telling the good indicators from the bad ones.

and the Basel gap rank lower than the narrow credit counterparts, though still in the top half of all the indicators. The Lucas critique however applies: economic agents' decisions are indeed not policy-invariant, therefore one could expect that with increasing bank lending regulation, such activities will more and more shift to the non-banking sphere - supporting the use of the total credit aggregate as a more comprehensive indicator for the future. In general, credit to GDP ratios appear helpful in assessing how vulnerable a country is because of excessive structural leverage rather than conjunctural developments, and are therefore useful in conditioning the information provided by gaps and rates of growth. Global liquidity - in the form of both the global credit gap and the growth rate - turns out to be another key concept, ranking among the five most important indicators. The remaining two indicators among the top five are the level of household credit and the aggregate debt service ratio. Immediately following the top six indicators there are some measures relating to house prices, namely the house price to income ratio, the house price gap and house price growth. Equity price growth ranks a little lower. Indeed, heated asset price growth might be associated with excessive credit growth fuelling a growing bubble. After considering the housing market, the Random Forest suggests that the real short term rate should be looked at next, most likely because a low rate may encourage risk-taking in a search-for-yield behavior. Also among the top half of all the indicators are the household debt service ratio, bank credit growth, the NFC credit to GDP ratio and M3 gaps.

## 7 Comparison with logit models

In this section we aim at testing the predictive performance of competing regression models, namely discrete choice models.<sup>22</sup> In particular, we estimate logistic regressions, where a logit mapping function takes the explanatory variables into a continuous indicator variable between 0 and 1, which indicates the (crisis) probability. The pooled logit model specification is as follows:

$$Prob(y_{it} = 1|X_{it-1}) = \frac{e^{\alpha_i + X'_{it-1}\beta}}{1 + e^{\alpha_i + X'_{it-1}\beta}} \quad (1)$$

The experimental design is the same as the one adopted for the Random Forest:  $Prob(y_{it} = 1|X_{it-1})$  denotes the probability that a given country  $i$  in a given quarter  $t$  is in a pre-crisis state, with pre-crisis states defined as described in Section , i.e. five to one year before the outbreak of a crisis; also in this case, the 3 quarters immediately preceding the crisis and the crisis quarters themselves are excluded from the sample. The crisis probability at time  $t$  depends on the information set at time  $t - 1$ , as all regressors are lagged by one quarter. The logit models we estimate may also include a set of country dummy variables,  $\alpha_i$ . The selection of the regressors  $X_{it-1}$  is unfortunately heavily affected by data availability. Indeed, ideally one would like to compare competing models on the same sample, both in terms of time

---

<sup>22</sup>For a comparison of logit models and decision trees in the context of credit risk assessment see Joos et al. (2001), while Savona and Vezzoli (2014) test the performance of logit models and a similar algorithm to the Random Forest in predicting sovereign debt crises.



and cross-sectional dimension. In this respect, more than 30 variables for 21 countries are included in the Random Forest model, each of which for the whole time-span for which observations for that indicator are available. On the contrary, the estimation of a pooled logit regression requires a balanced panel. In the selection of the relevant sample, a trade-off emerges between maximizing the time dimension and the cross-sectional dimension.

The first of the two models we estimate aims at maximizing the time dimension, namely keeping observations as of 1970. This requires dropping a number of countries (Cyprus, Estonia, Latvia, Luxembourg, Malta, Slovakia and Slovenia), and restricting the regressors to a handful of credit variables. In particular, we estimate model 3 in Behn et al. (2013), which includes domestic broad credit growth, the domestic Basel gap, a term representing the interaction of the two, global credit growth, the global credit gap, an interaction term for global liquidity variables (i.e. global credit growth\*global credit gap), an interaction term for broad domestic and global credit growth, and an interaction term for the domestic and the global gap. Country dummies are also included.<sup>23</sup> Table 1 shows the estimation results: with respect to the predictive power of this model, the AUROC is 0.84, i.e. 10 percentage points lower than the Random Forest AUROC.

The second model we estimate aims at extending the types of regressors included, e.g. by considering also asset prices. This requires restricting even

---

<sup>23</sup>This amounts to excluding countries which did not experience a crisis, i.e. Austria and Belgium. See Behn et al. (2013) for a discussion of the bias arising from including country dummies and from not including them.

further the number of countries considered, limiting them to Finland, France, Italy, the Netherlands, Sweden, Spain and the UK, while keeping observations as of 1980. The model includes the domestic broad credit growth, the Basel gap, the DSR, equity price growth, global credit growth and gap, the house price to income ratio, and real GDP growth. The estimation results are also shown in Table 1: this model does much better than the first one, yielding an AUROC of 0.93, which is comparable to the Random Forest AUROC. However, it should be noticed that restricting the sample to a handful of countries with similar economic and financial systems, and excluding from the analysis the banking crises that took place in the 70's (in Spain and the UK), when the financial system was arguably quite different, makes the task easier for the logit than for the Random Forest.

## 8 The early warning tree

Chart 4 shows the benchmark tree grown on the best indicators described in Section 6, assuming that the underlying preferences of the policymakers with respect to missing crises and issuing false alarms are biased against missing crises.<sup>24</sup> The Random Forest and the associated early warning tree grown by assuming balanced preferences between type 1 and type 2 errors are described in the Appendix. The indicator appearing in the root node is the

---

<sup>24</sup>To avoid overfitting, this tree has been grown by imposing a minimum parent node size of 8 country/periods and a minimum leaf size of 4 country/periods, while some of the terminal branches have been pruned.

DSR, associated with a threshold of 18%. According to end-2012 data, this threshold splits the sample equally, with around half of the countries ending up in the right branch and the other half in the left branch. The next node along the right branch of the tree corresponds to the bank credit to GDP ratio with a threshold of 92%. If this threshold is breached, the next relevant indicator is household credit as a percentage of GDP with a threshold of 54.5%. At the end of 2012 a relatively large number of countries breached all of these thresholds, ending up in the ‘warning’ leaf associated with a 90% in-sample crisis frequency. As cyclical developments might be less relevant along this branch of the tree, one could consider employing macroprudential instruments like the systemic risk buffer to increase resilience in the system given the elevated leverage identified by the model. However, this estimate of the probability of a crisis should be interpreted with caution for the following two reasons. The first is that the better the tree is at fitting in-sample data, the purer the leaves it will yield, with associated in-sample frequencies close to 1 or 0. However, in assessing a country’s situation one should consider whether the relevant indicators only marginally exceed (or not) the respective thresholds. The second caveat relates to country specificities, which cannot be captured by the model. With respect to this leaf, for example, the concept of the DSR could be misleading for specific countries that for reasons not harmful for financial stability have structurally high private sector debt. In such a case, a net debt concept taking into account accumulated private sector wealth would be more suitable.

If the bank credit to GDP threshold of 92% is not breached, the next relevant indicator is the bank credit gap with a threshold of 3.6 p.p.. If this threshold is breached, the crisis probability increases to above 60%. In this case, there would be a role for macroprudential tools such as the countercyclical capital buffer as the credit gap can be associated with cyclical systemic risk.

Looking at the left branches of the tree, the main messages are as follows. If the DSR is below 10.6% the crisis probability is negligible. A relatively large number of countries, however, are in the middle range, with a DSR between 10.6% and 18%. For these countries, essentially depending on the sign of the M3 gap, different variables become relevant. These indicators relate to the following: i) house prices, in the form of house price growth and gap and in relation to income; ii) equity prices; iii) the Basel gap; iv) the short term real interest rate; v) bank credit level and growth; and vi) household credit. As an example, a country falling in the ‘warning’ leaf associated with a house price to income ratio 27 points above its long term average might consider adopting measures such as caps to loan-to-value and loan-to-income ratios.

With respect to the in-sample predictive performance of this benchmark tree, the true positive rate and the false positive rate (or share of type 2 errors) are equal to 85% and 4%, respectively, while the share of type 1 errors is 15%. The noise to signal ratio is 5%. A more sophisticated measure of the usefulness of the model, taking into account the policymaker’s greater

aversion towards type 1 errors, indicates that a policymaker using this tree increases his/her utility by 65% compared with ignoring it.<sup>25</sup>

## 9 Country classification

According to end-2012 data, the countries above the DSR threshold of 18% are Belgium, Cyprus, Denmark, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the UK. Almost all of these end up in the ‘warning’ leaf associated with bank credit at more than 92% of GDP and household credit at more than 54.5% of GDP, characterized by a 90% crisis probability. However, it should be noticed that Italy and Greece breach the first and the second threshold, respectively, by only a couple of percentage points. Based on the available data, the probability of a banking crisis in Cyprus would be 35%, which is the in-sample crisis frequency associated with the bank credit to GDP node. Belgium breaches neither the bank credit to GDP 92% threshold, nor the 3.6 p.p. bank credit to GDP gap threshold, ending up in a leaf characterized by a zero crisis probability. With respect to the countries for which the DSR is below 18%, Luxembourg and Slovakia end up in the ‘tranquil’ leaf associated with a DSR lower than 10%. France, Slovenia, Austria, Finland and Latvia do not breach the -0.24 p.p. M3 gap threshold, while the real short term rate is in all of these countries below -0.5%. Due to missing data on bank credit for Slovenia and Latvia, these two coun-

---

<sup>25</sup>See Sarlin (2013).

tries remain associated with the parent node characterized by a 26% crisis probability, while Austria, Finland and France do not breach the 7.3% y-o-y growth threshold and therefore end up in a ‘tranquil’ leaf. Germany breaches the M3 gap threshold only very marginally, as the gap is still negative, while it does not breach any of the housing-market related nodes, ending up in a leaf characterized by zero crisis probability. The M3 gap in Estonia and Malta is positive. Due to data availability issues, these two countries cannot be classified into any terminal node; the crisis probability associated with the parent nodes they end up in is 8% (house price to income node) and 13% (equity price growth node), respectively.

## 10 Out-of-sample exercise

An out-of-sample exercise testing the predictive performance of the model with respect to the global financial crisis is a heroic task, as only slightly more than half of the crisis episodes are left in the sample and some data series become extremely short. Nevertheless, the credibility of any early warning model of this sort crucially depends on whether the model would have been of any help in detecting in real time the build-up of financial imbalances in the run-up to the crisis. Therefore, in this section we describe what the suggestions of the model would have been in mid-2006, based on data up to the second quarter of 2006 only and ignoring whether the period

starting in mid-2001 would later be classified as a pre-crisis period.<sup>26</sup>

A 100,000-tree Random Forest grown on this information set indicates that the global credit, the bank credit and the Basel gaps would have turned out to be the key variables back in 2006, as well as the level of bank credit (see Figure 5). The M3 gaps would have ranked immediately lower, followed by house price valuation measures. Among the best performing indicators there would have been also other global liquidity indicators, as well as the DSR, the level of broad credit, household and NFC credit, bank and broad credit growth and the house price gap.

The tree built on the indicators listed above (excluding global liquidity) would have had the M3 gap at the root node (see Figure 6). Germany and Greece would have ended up in the same ‘tranquil’ leaf, as at that time the M3 gaps, the Basel gap and the house price gap were all rather low in these countries. No warning signal would have been issued for Portugal, notwithstanding its large Basel gap. Despite a relatively low M3 gap, a warning signal would have been issued for Denmark, while the Netherlands would have been assigned a zero crisis probability due to its bank credit gap not breaching the relevant threshold. Considering the countries characterized by a relatively large M3 gap, Belgium and Luxembourg would have been assigned a zero crisis probability owing to low bank credit gap and ratio to

---

<sup>26</sup>For this exercise, gaps have been constructed by taking a standard HP filter for the first year and a half of available data and then a recursive HP filter, while the long term average of house price to income and house price to rent ratios is computed on observations up to the first quarter of 2006.

GDP. Despite a more elevated level of bank credit, Austria would have also been assigned a zero crisis probability due to its Basel gap being relatively small, i.e. not breaching the 2.4 p.p. threshold. The UK would have ended up in a leaf associated with a 100% crisis probability as both its bank credit level and Basel gap breached their respective thresholds in 2006. Finland, France, Ireland, Italy, Spain and Sweden would have all ended up in a leaf characterized by a 79% crisis probability due to rather elevated M3 and bank credit gap, with the house price to income ratio breaching its threshold at the same time. Finally, due to lack of data for Estonia, Cyprus, Slovakia, Latvia, Malta and Slovenia, all of these countries would have remained associated with parent nodes characterized by a low crisis probability (up to 17%).

As summarized in the matrix below, six of the eight countries for which the model would have issued a warning actually experienced a crisis in the five subsequent years. Overall, the crisis would have been correctly predicted for all of the large EU economies that did indeed later undergo one. A prompt policy reaction, assuming the current macroprudential legislation were already in place, would have allowed, for example, to have counter-cyclical capital buffers in place in these countries already for one year before the Lehman collapse. Considering type 2 errors and taking the size of the financial system as a proxy for the costs incurred by the economy as a consequence of the misclassification, the only large country for which the indication would have been to implement pre-emptive macroprudential measures when no credit related systemic banking crisis actually followed is Italy. Though



one could argue that the Italian banking sector and thus the Italian economy would also have benefited from higher capital buffers during the post-Lehman crisis years. No warning signal would have been issued for the majority of the countries (in some cases due to data availability issues). Notably, no warning signal would have been issued for Germany, which indeed did not experience a crisis afterwards. Considering type 1 errors, it should be noted that for some of these countries later crises were not due only, or mainly, to credit and asset price developments, but also to e.g. developments in the sovereign debt sphere, making it relatively difficult for the model to make a correct prediction.

	Crisis	No crisis
Warning	FR, IE, ES, SE, DK, UK	FI, IT
No warning	GR, PT, LV, SI, NL	AU, BE, LU, DE, EE, SK, MT, CY*

\*Crisis started beyond prediction horizon

Finally, following Drehmann and Juselius (2013), we check the stability of the signals from the early warning tree going forward. Indeed, policy decisions are generally based on persistent indications for action, and an early warning system which gives contradictory messages quarter after quarter would be unreliable. Therefore, keeping the tree fixed, we investigate whether warning signals would have continued to be issued for the countries flagged in mid-2006. As shown in Chart 7, the early warning tree would have been

an extremely stable model, as it would have suggested to continue closely monitor all of the countries for which a warning would have been issued in mid-2006, in all of the subsequent quarters until the outbreak of the global financial crisis.

## 11 Policy implications

Policy makers at the national designated authorities becoming responsible for macro-prudential policies in the EU as well as at the European level, i.e. at the ECB and ESRB, will have to use their judgement in setting the macro-prudential policy stance for the respective countries. Tools like our proposed early warning tree and Random Forest can serve several purposes in this process. First, the good out-of-sample performance of such analytical models should help to overcome the possible inaction bias on the part of policy makers. In case risks are emerging which have in the past led to systemic banking crises, the onus is on those who aim to use judgement alone to justify why macro-prudential policy tools are not activated. Second, the intuitive nature of a decision tree model and its easy visualization is likely to increase acceptance of an analytical approach as a starting point for policy discussions. As section 7 has shown, the approach can be used to also trigger discussions on country specificities affecting the risk assessment. Third, a further advantage of the tree model is that depending on the characteristics of the leaf associated with a certain crisis probability, the nature of the

vulnerability can also be identified, which in many cases would then suggest the use of a specific policy instrument over another.

## 12 Conclusions

We build an early warning system aiming at supporting policy decisions on when to activate macroprudential tools targeting excessive credit growth and leverage. Together with total credit to GDP deviations from trend (the so-called ‘Basel gap’) we consider a battery of indicators as a policy guide, including credit ratios and real estate indicators.

By using decision trees, we build a multivariate predictive model which is at the same time extremely accurate and very easy to interpret. Based on the experience of EU countries over the last 40 years, it applies decision tree learning to the problem of identifying excessive credit growth and leverage with a sufficient lead time to allow policy reactions. One of the main advantages of the presented approach is that it takes into account the conditional relations between various indicators when setting early warning thresholds. At the same time, the model is able to give an indication on which macroprudential tool could be best suited to address specific vulnerabilities.

The proposed early warning system can be regarded as a useful common reference point informing policy makers when using their judgement. Indeed, it is crucial that the use of judgement be firmly anchored to a clear set of principles to promote sound decision-making in the operationalization of

macroprudential instruments.

## Appendix

The ranking of the indicators derived by assuming balanced preferences between missing crises and issuing false alarms is very similar to that described in 6 and is shown in Figure 8. The top two indicators remain the level of bank credit and the global credit gap, while the main differences relate to global credit growth and the Basel gap, which turn out to be relatively less important than in the biased preferences case.

The early warning tree derived on the best half indicators, excluding global liquidity and assuming balanced preferences between Type 1 and Type 2 errors is shown in Figure 9. By and large, the same key variables appear in both the trees derived with biased and balanced preferences. When preferences are balanced, the root node is associated with the bank credit to GDP gap and a threshold of 3.4 p.p.. Along the right branch, we find the DSR with an almost identical threshold compared to the one relevant for the benchmark tree presented in Section 8, i.e. 17%. The lower level nodes in this part of the tree are associated with house price growth and the ratio of household credit to GDP, the M3 gap and government debt. The warning threshold for this latter, which is absent in the benchmark tree, is 60% of GDP. Along the left-hand side branch of the tree we find again house price based measures, namely gaps and the house price to income ratio, the DSR, the ratio of bank credit to GDP in two different nodes, the short term rate and household credit growth.

With respect to the in-sample predictive performance, this tree yields a true positive rate of 88% and a false positive rate of 2%, while the share of missed crises is 12%. Notice that, although the benchmark tree described in Section 8 is constructed by placing a higher weight on Type 1 errors, it still yields a higher share of missed crises compared to the balanced-preferences tree due to the fact that some branches have been pruned and therefore both trees are in some sense ‘suboptimal’. Finally, the noise to signal ratio associated with this tree is 2% while the relative Usefulness measure, i.e. the gain by using this model compared to ignoring it, is equal to 86%.

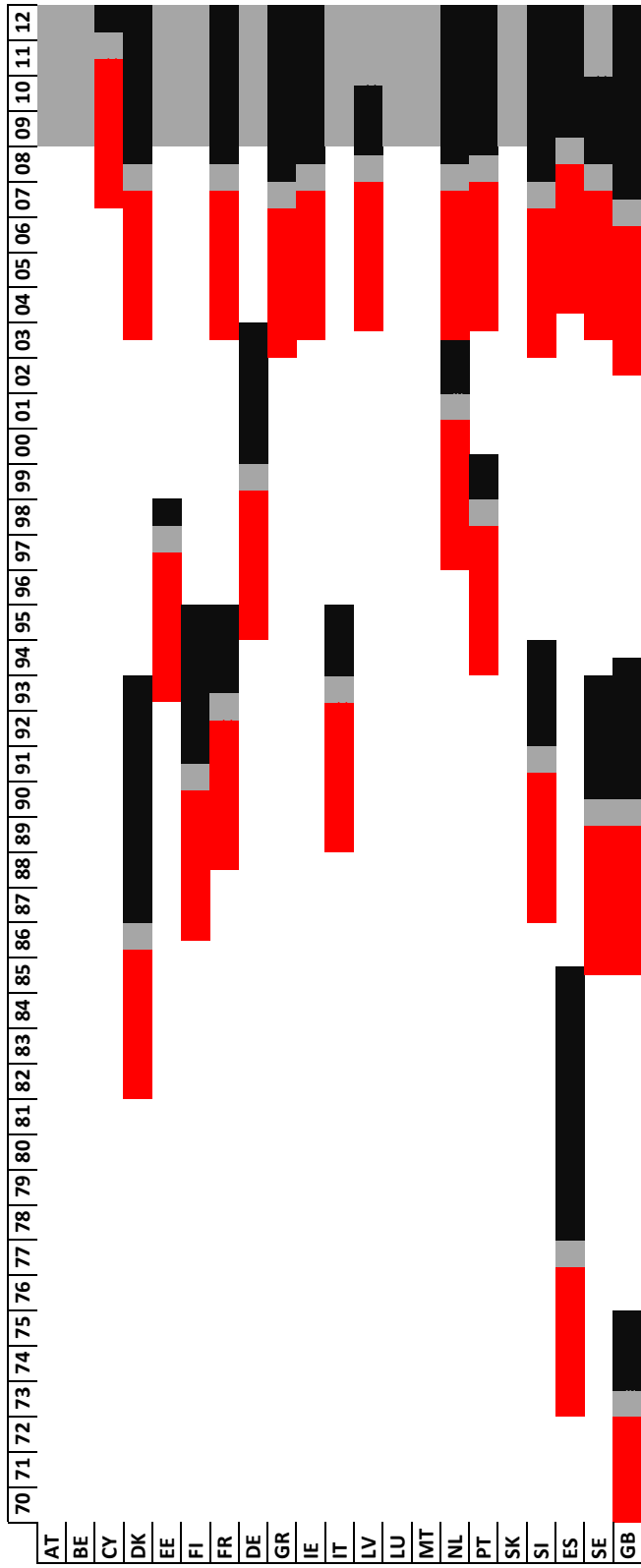


Figure 1: Identified crises (in black), pre-crisis periods (in red) and periods excluded from the analysis (in grey).

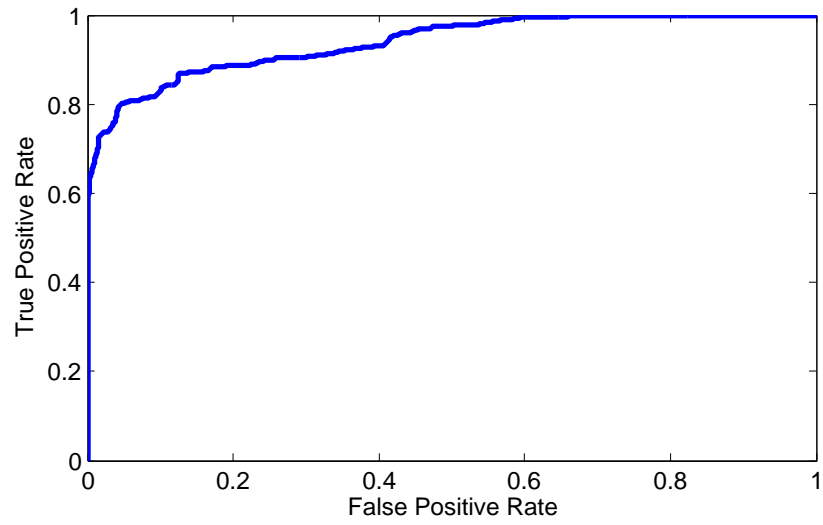


Figure 2: ROC curve associated with the Random Forest.



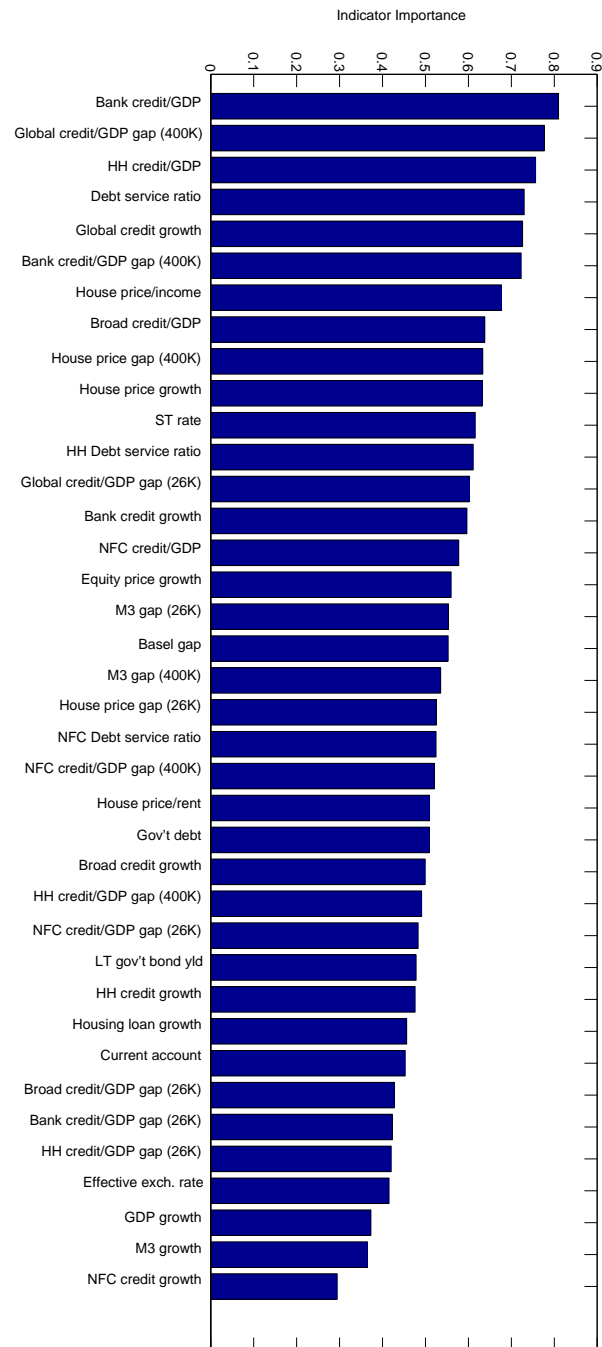


Figure 3: Ranking of the indicators according to the conveyed amount of useful information.

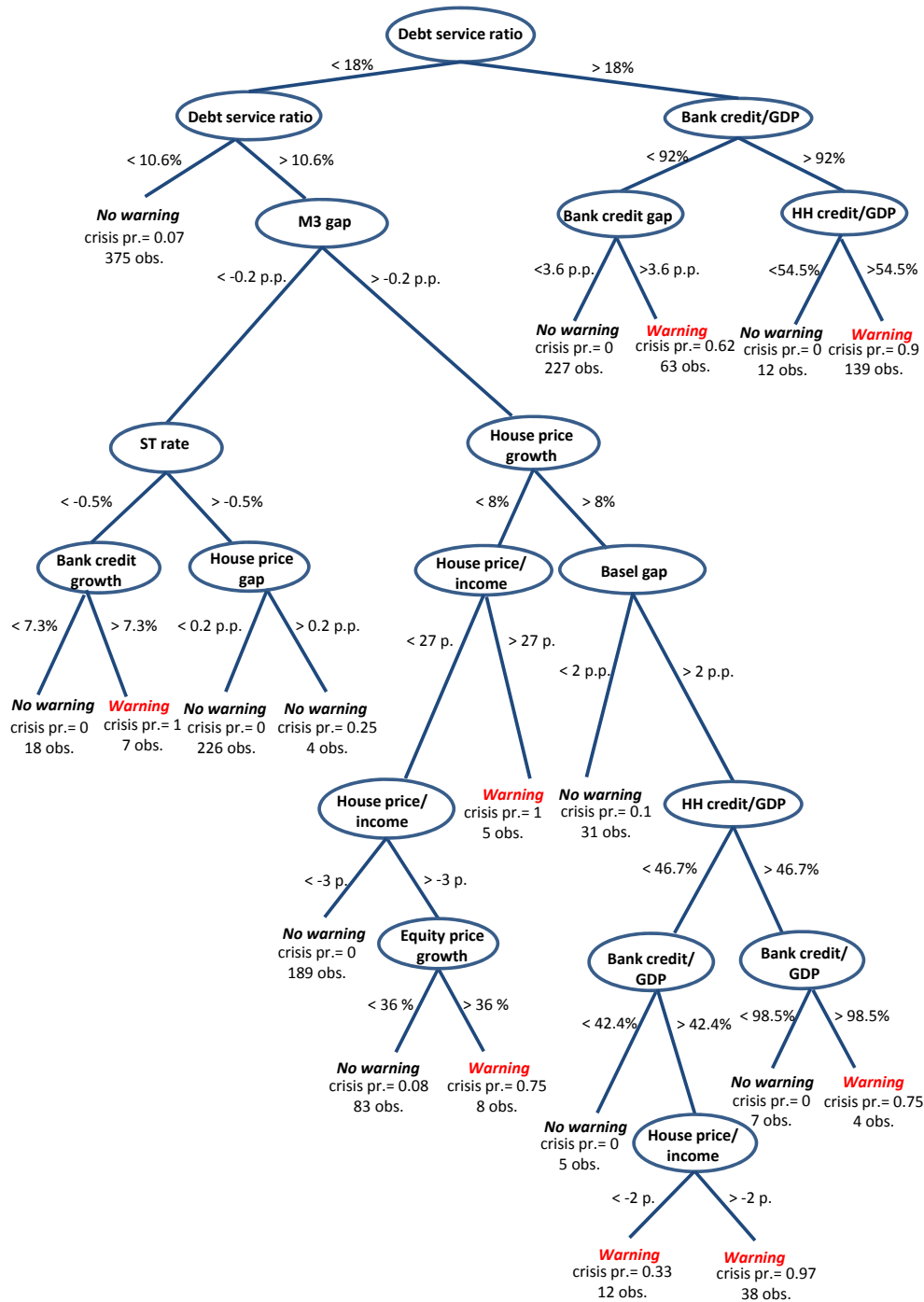


Figure 4: The benchmark early warning tree. The threshold for the house price to income ratio is in terms of index points above/below its long term average, while p.p. stands for percentage points. In each terminal node (leaf) of the tree the crisis probability is indicated, based on the frequency of pre-crisis quarters ending up in that particular leaf, considering the historical data on which the tree has been grown. The total number of country/quarters ending up in each leaf is also indicated. When the crisis probability associated with a leaf exceeds 30% the leaf is labelled as a ‘warning’ leaf.

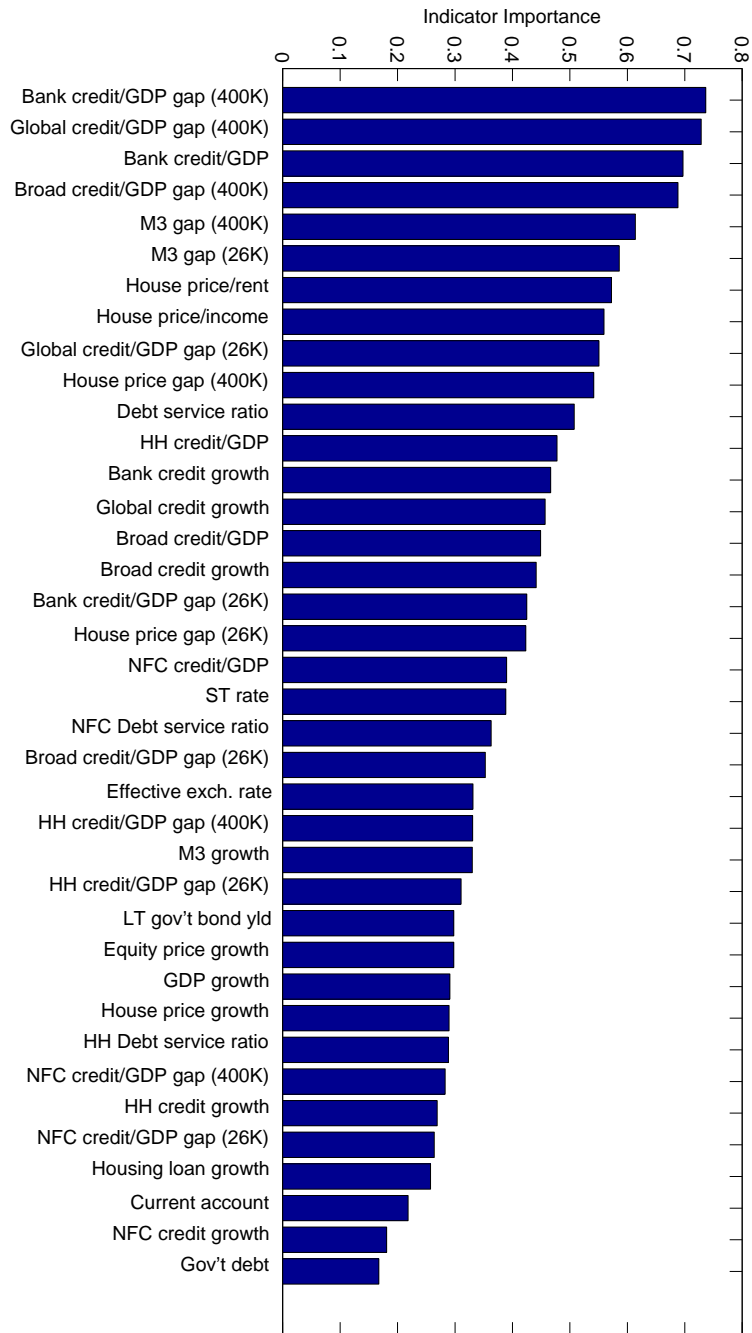


Figure 5: Ranking of the indicators according to the conveyed amount of useful information, using data available in mid-2006.

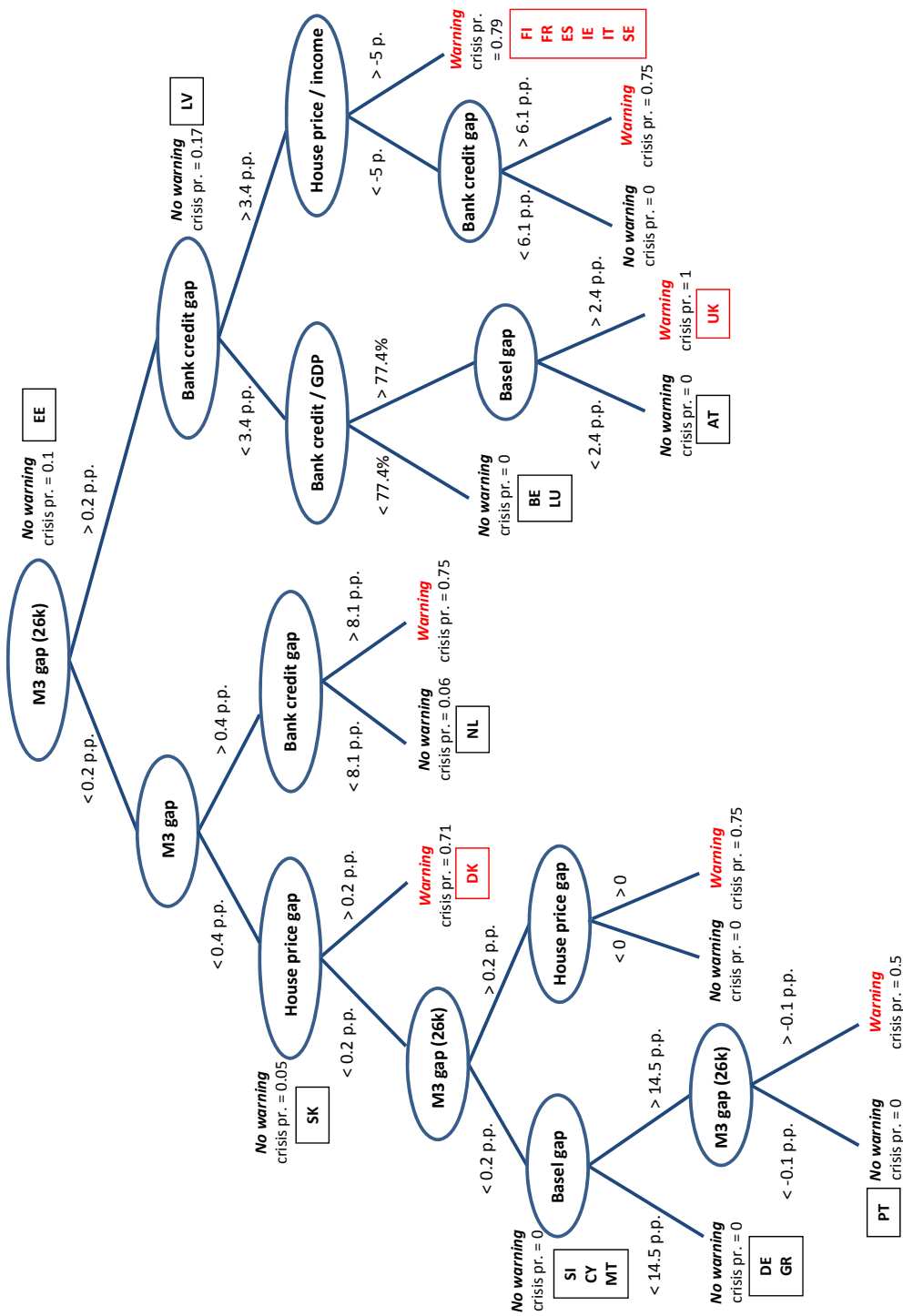


Figure 6: The early warning tree derived with data as of 2006Q2. Gaps are computed by setting  $\lambda = 40000$  unless otherwise indicated. In each terminal node (leaf) of the tree the crisis probability is indicated, based on the frequency of pre-crisis quarters ending up in that particular leaf, considering the historical data on which the tree has been grown. When the crisis probability associated with a leaf exceeds 30% the leaf is labelled as a 'warning' leaf.

	AT	BE	CY	DK	EE	FI	FR	DE	GR	IE	IT	LV	LU	MT	NL	PT	SK	SI	ES	SE	GB
2006Q3				x		x	x			x	x								x	x	x
2006Q4				x		x	x			x	x								x	x	x
2007Q1				x		x	x			x	x								x	x	x
2007Q2				x		x	x			x	x								x	x	x
2007Q3				x		x	x		x	x	x								x	x	
2007Q4				x		x	x		x	x	x					x			x	x	
2008Q1				x		x	x			x	x		x			x			x	x	
2008Q2				x		x	x			x	x		x			x			x	x	
2008Q3						x					x		x				x		x		

Figure 7: Stability of the signals issued by the early warning tree in the period between mid-2006 and the outbreak of the global financial crisis. Crosses indicate issued warnings, while dark cells denote crisis periods.

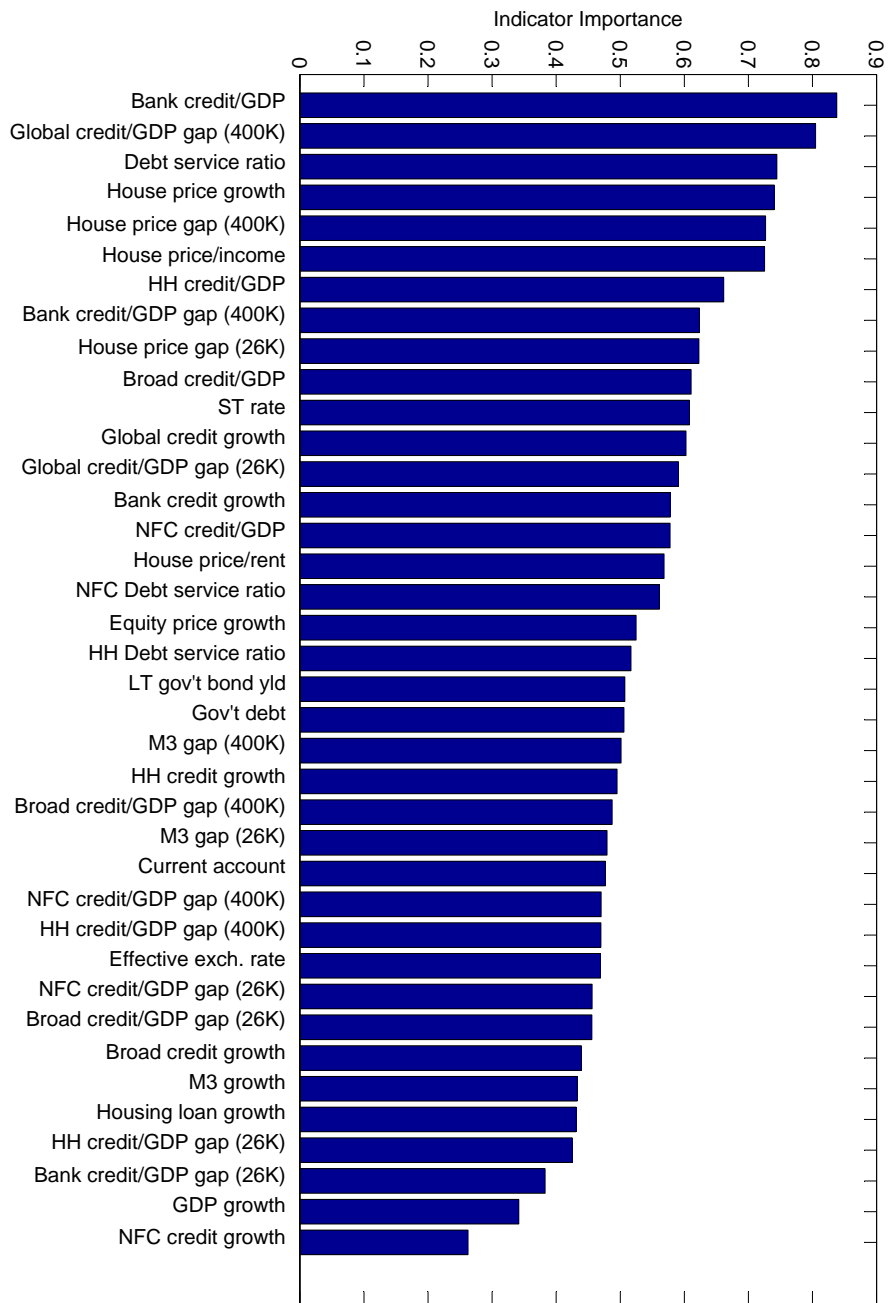


Figure 8: Ranking of the indicators according to the conveyed amount of useful information, assuming the policymaker has balanced preferences between Type 1 and Type 2 errors.

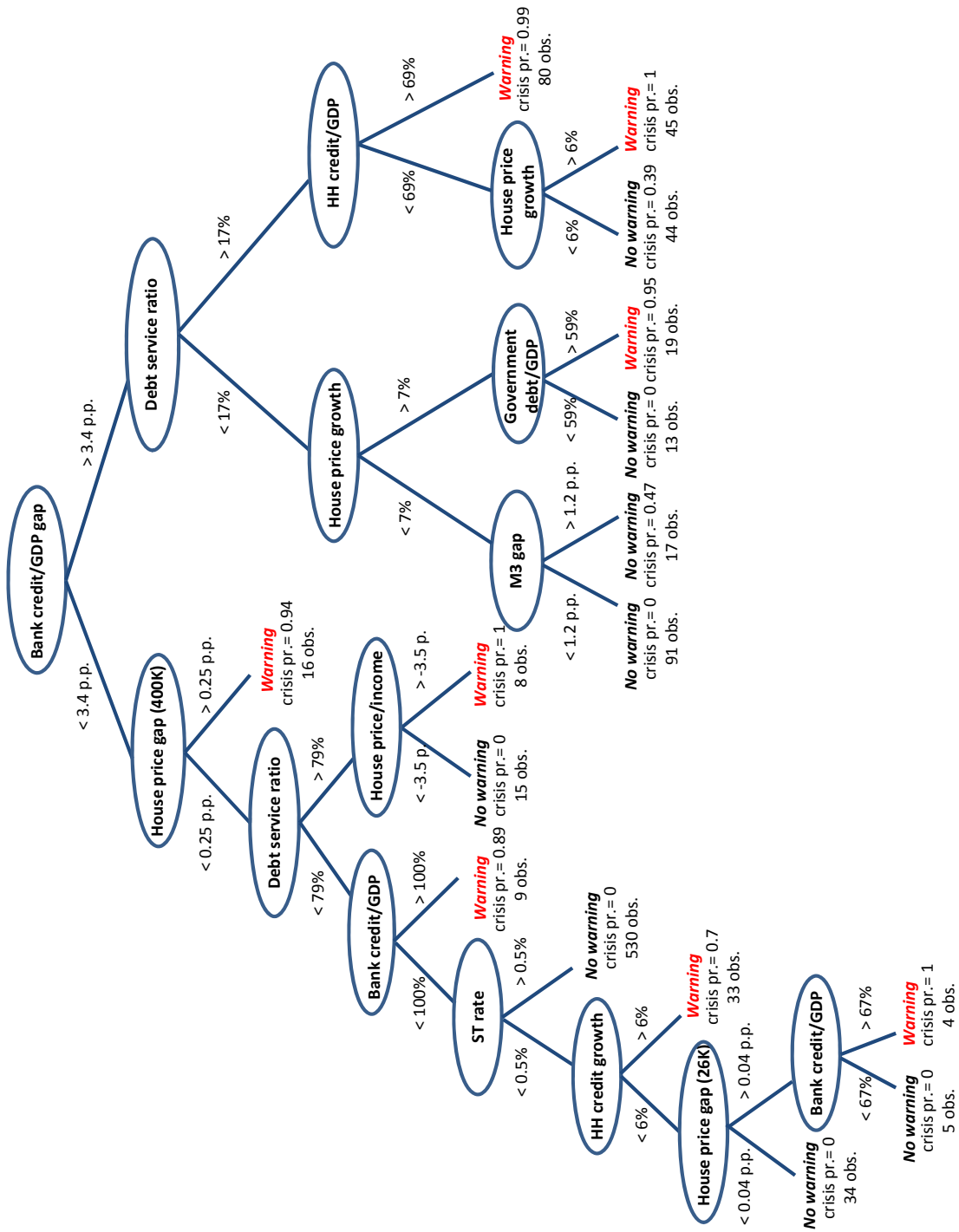


Figure 9: The early warning tree derived by assuming balanced preferences. The threshold for the house price to income ratio is in terms of index points above/below its long term average, while p.p. stands for percentage points. In each terminal node (leaf) of the tree the crisis probability is indicated, based on the frequency of pre-crisis quarters ending up in that particular leaf, considering the historical data on which the tree has been grown. The total number of country/quarters ending up in each leaf is also indicated. When the crisis probability associated with a leaf exceeds 50% the leaf is labelled as a 'warning' leaf.

	Behn et al. model 3	Extended model
Dom. broad credit growth (DC1)	-5.72 (4.94)	5.35 (4.35)
Dom. Basel gap (DC2)	17.86*** (2.32)	12.32*** (3.12)
Interaction (DC1*DC2)	-0.65*** (0.18)	
Glo. credit growth (GC1)	-2.74 (6.34)	-20.41*** (7.78)
Glo. credit gap (GC2)	-21.28** (10.56)	61.76*** (8.77)
Interaction (GC1*GC2)	6.81*** (1.63)	
Interaction (DC1*GC1)	1.68** (0.79)	
Interaction (DC2*GC2)	-0.88** (0.39)	
DSR		273.40*** (60.05)
Equity price growth		1.81*** (0.53)
House price to income		11.11*** (1.24)
Real GDP growth		35.19*** (9.72)
Country dummies	YES	NO
Observations	1296	727
Pseudo R-Squared	0.28	0.52
AUROC	0.84	0.93

Table 1: Estimation results for multivariate logit models. Standard errors are reported in parentheses. \* indicates statistical significance at the 10% level, \*\* at the 5% level, and \*\*\* at the 1% level.



## References

- Alessi, L. and Detken, C., 2011. Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy*, 27(3): 520–533.
- Babecky, J., Havranek, T., Mateju, J., Rusnak, M., Smidkova, K. and Vasicek, B., 2012. Banking, debt and currency crises: early warning indicators for developed countries. *European Central Bank Working Paper Series*, no 1485.
- Barrell, R., Davis, E. P., Karim, D. and Liadze, I., 2010. Bank regulation, property prices and early warning systems for banking crises in OECD countries. *NIESR Discussion Paper*, no 330.
- Basel Committee on Banking Supervision, 2010. *Guidance for national authorities operating the countercyclical capital buffer*.
- Behn, M., Detken, C., Peltonen, T. A. and Schudel, W., 2013. Setting Countercyclical Capital Buffers Based on Early Warning Models: Would it Work? *ECB Working Paper*, no 1604.
- Borio, C. and Drehmann, M., 2009. Towards an operational framework for financial stability: “fuzzy” measurement and its consequences. *BIS Working Paper No. 284*.
- Borio, C. and Drehmann, M., 2009. Assessing the risk of banking crises - revisited. *BIS Quarterly Review*, March 2009, pp. 29-46.
- Borio, C. and Lowe, P., 2002. Assessing the risk of banking crises. *BIS Quarterly Review*, December 2002, pp 43-54.
- Breiman, L., Friedman, J., Olshen, R. and Stone, C., 1984. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45(1):5 - 32.
- Bussière, M. and Fratzscher, M., 2006. Towards a new early warning system of financial crises. *Journal of International Money and Finance*, 25(6):953-973.
- Caprio, G. and Klingebiel, D., 2003. *Episodes of Systemic and Borderline Financial Crises*. World Bank, January 22.

- Committee on the Global Financial System, 2010. Macroprudential instruments and frameworks: a stocktaking of issues and experiences. CGFS Publications, no 38.
- Committee on the Global Financial System, 2012. Operationalising the selection and application of macroprudential instruments. CGFS Publications, no 48.
- 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 2013.
- Detken C., Weeken, O., Alessi, L., Bonfim, D., Boucinha, M. M., Castro, C., Frontczak, S., Giordana, G., Giese, J., Jahn, N., Kakes, J., Klaus, B., Lang, J. H., Puzanova, N. and Welz, P., 2014. Operationalizing the Countercyclical Capital Buffer. ESRB Occasional Paper, no 5.
- Demirgüç-Kunt, A. and Detragiache, E., 2000. Monitoring banking sector fragility: A multivariate logit approach with an application to the 1996-97 banking crisis. Policy Research Working Paper, no 2085, World Bank.
- Drehmann, M., Borio, C., Gambacorta, L., Jimenez, G. and Trucharte, C., 2010. Countercyclical capital buffers: exploring options. BIS Working Papers, 01/2010.
- Drehmann, M., Borio, C., and Tsatsaronis, K., 2011. Anchoring countercyclical capital buffers: the role of credit aggregates. International Journal on Central Banking, Vol. 7(4), 189-240.
- Drehmann, M. and Juselius, M., 2012. Do debt service costs affect macroeconomic and financial stability? BIS Quarterly Review, September 2012.
- Drehmann, M. and Juselius, M., 2013. Evaluating early warning indicators of banking crises: satisfying policy requirements BIS Working Paper No. 421, August 2013.
- Drehmann, M. and Tsatsaronis, K., 2014. The credit-to-GDP gap and countercyclical capital buffers: questions and answers. BIS Quarterly Review, March 2014.
- Duttagupta, R. and Cashin, P., 2011. Anatomy of banking crises in developing and emerging market countries. Journal of International Money and Finance, Vol. 30(2), 354-376.

- Edge, R. M. and Meisenzahl, R. R., 2011. The unreliability of credit-to-GDP ratio gaps in real time: implications for countercyclical capital buffers. *International Journal on Central Banking*, 7(4):261-298.
- Eichengreen, B. and Rose, A. K., 1998. Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking Crises. NBER Working Papers, no 6370.
- Frankel, J. A. and Wei, S., 2004. Managing Macroeconomic Crises. NBER Working Papers, no 10907.
- Gersl, A. and Seidler, J. 2012. Excessive Credit Growth and Countercyclical Capital Buffers in Basel III: An Empirical Evidence from Central and East European Countries. *ACTA VSFS*, 6(2):91-107.
- Ghosh, S. R. and Ghosh, A., 2002. Structural Vulnerabilities and Currency Crises. *IMF Staff Papers*, 50(3).
- Joos, P., Vanhoof, K., Ooghe, H. and Sierens, N., 1998. Credit classification: a comparison of logit models and decision trees. Proceedings notes of the workshop on application of machine learning and data mining in finance.
- 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), 473-500.
- Kauko, K., 2012. Triggers for countercyclical capital buffers. Bank of Finland, 7/2012.
- Laeven, L. and Valencia, F., 2008. Systemic Banking Crises: A New Database. *IMF Working Paper*, no 08/224.
- Laeven, L. and Valencia, F., 2010. Resolution of Banking Crises: The Good, the Bad, and the Ugly. *IMF Working Paper*, no 10/146.
- Laeven, L. and Valencia, F., 2012. Systemic Banking Crises Database: An Update. *IMF Working Paper*, no 12/163.
- Lo Duca, M. and Peltonen, T., 2013. Assessing systemic risks and predicting systemic events. *Journal of Banking and Finance*, 37 (7): 2183-2185.
- Manasse, P. and Roubini, N., 2009. "Rules of thumb" for sovereign debt crises. *Journal of International Economics*, 78: 192-205.

- Manasse, P., Savona, R. and Vezzoli, M., 2013. Rules of Thumb for Banking Crises in Emerging Markets. Università di Bologna DSE Working Paper, no 872.
- Mendoza, E and Terrones, M., 2008. An anatomy of credit booms: evidence from macro aggregates and micro data. NBER Working Paper, no 14049.
- Repullo, R. and Saurina, J. 2011. The countercyclical capital buffer of Basel III: A critical assessment. CEMFI Working Paper.
- Sarlin, P., 2013. On policymakers' loss function and the evaluation of early warning systems. *Economics Letters*, 119(1): 1–7.
- Savona, R. and Vezzoli, M., 2013. Fitting and Forecasting Sovereign Defaults using Multiple Risk Signals. *Oxford Bulletin of Economics and Statistics*, forthcoming.
- 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): 1029-61.
- Tuffery, S., 2011. *Data Mining and Statistics for Decision Making*. Wiley Series in Computational Statistics.