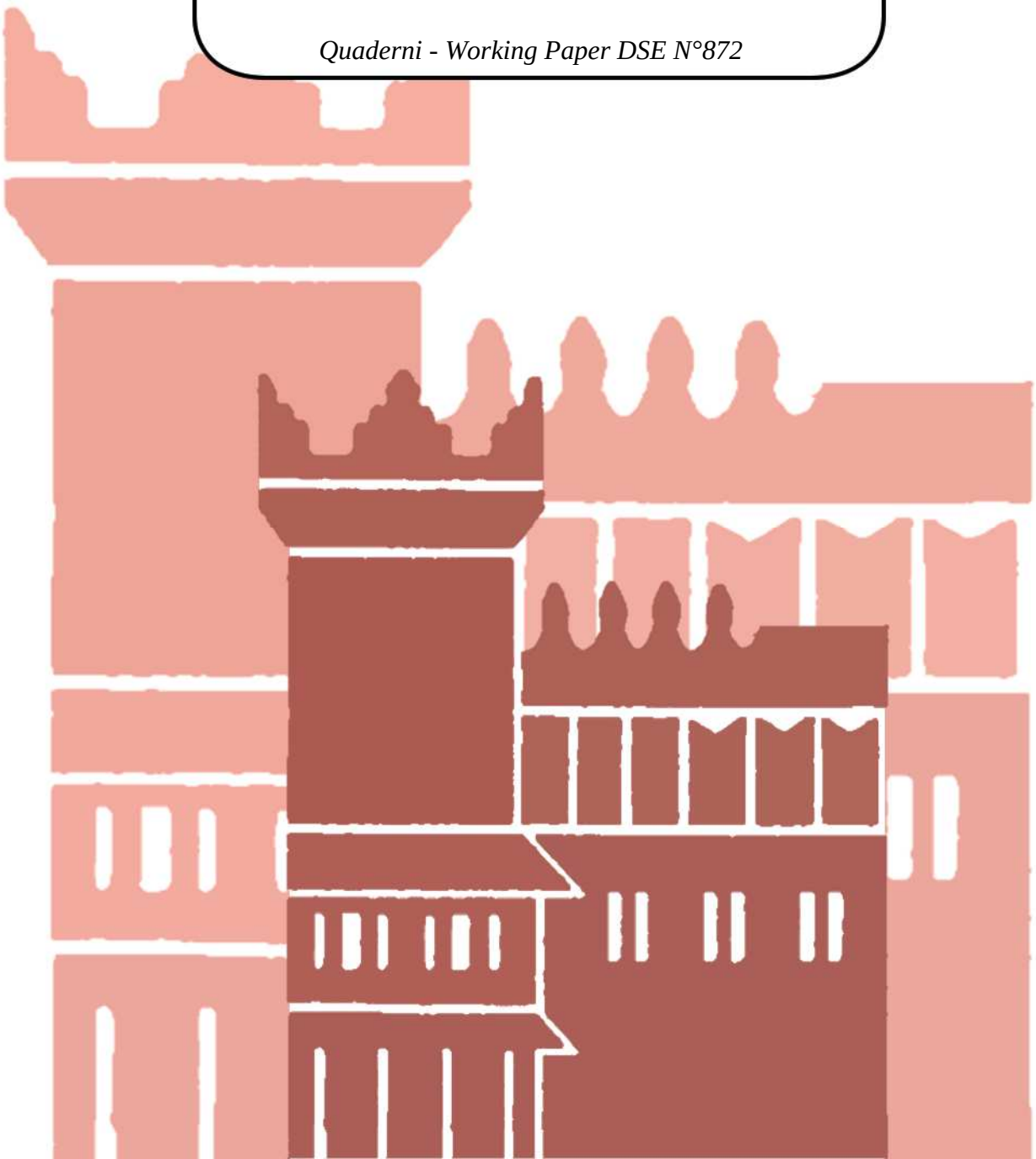
Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Rules of Thumb for Banking Crises in Emerging Markets

Paolo Manasse
Roberto Savona
Marika Vezzoli

Quaderni - Working Paper DSE N°872



Rules of Thumb for Banking Crises in Emerging Markets

by

Paolo Manasse[*], Roberto Savona[†], Marika Vezzoli[‡]

Abstract: This paper employs a recent statistical algorithm (CRAGGING) in order to build an early warning model for banking crises in emerging markets. We perturb our data set many times and create "artificial" samples from which we estimated our model, so that, by construction, it is flexible enough to be applied to new data for out-of-sample prediction. We find that, out of a large number (540) of candidate explanatory variables, from macroeconomic to balance sheet indicators of the countries' financial sector, we can accurately predict banking crises by just a handful of variables. Using data over the period from 1980 to 2010, the model identifies two *basic types* of banking crises in emerging markets: a "Latin American type", resulting from the combination of a (past) credit boom, a flight from domestic assets, and high levels of interest rates on deposits; and an "Asian type", which is characterized by an investment boom financed by banks' foreign debt. We compare our model to other models obtained using more traditional techniques, a Stepwise Logit, a Classification Tree, and an "Average" model, and we find that our model strongly dominates the others in terms of out-of-sample predictive power.

JEL: E44, G01, G21

Keywords: Banking Crises, Early Warnings, Regression and Classification Trees, Stepwise Logit

[*] Economics Department, University of Bologna; Address: Strada Maggiore 45, 40100 Bologna, Italy; email: paolo.manasse@unibo.it.
[†] Department of Economics and Management, University of Brescia; Address: C/da S. Chiara 50, 25122 Brescia, Italy; email: savona@eco.unibs.it.
[‡] Department of Economics and Management, University of Brescia; Address: C/da S. Chiara 50, 25122 Brescia, Italy; email: vezzoli@eco.unibs.it.

1. Introduction

The recent wave of banking crises stemming from developed countries, from the sub-prime-Lehman crisis in the US, to the European cases of Iceland, Ireland and Spain, have spurred a renewed interest in the quest for "early warning" indicators, signals of troubles ahead that may, with some confidence and anticipation, flag the need for preemptive action. In the European case, banks' exposure to collapsing real estate prices have threatened the joint solvency of the banking sector as well as of the sovereign, as in the classical "twin" paradigm described by Kaminsky and Reinhart (1999) and experienced in many episodes of the past three decades in emerging markets (the "Asian crisis" comes to mind). As in the past, the new wave of crises took most international financial institutions, such as the EU and the ECB, as well as rating agencies and academics, by surprise, as most forecasters grossly under-estimated the likelihood of such crises. The costs of the bailouts, their implication for government finances and sovereign solvency are so large that the failure to anticipate such events and to respond adequately are today the most serious threats to the very existence of the Euro area.

There has been a significant amount of research regarding early warnings of banking crises, particularly for emerging markets, and there is agreement on the presence of common pattern of economies that end up in a crisis: for example, credit booms and poor banking supervision, together with "unsustainable" current account imbalances, very often lead to real estate and stock market bubbles and crashes, and ultimately to banking crises, see for example Reinhart and Rogoff (2009). However, the interdependence between the financial institutions' balance sheets, interest rates and the macroeconomic outlook is still not properly analyzed. Most analysts and practitioners continue to use simple regression analysis for forecasting, despite the fact that these models are ill-suited for uncovering the joint effects of these indicators.

Our aim in this paper is to provide answers to the following questions. What set of macroeconomic and balance sheet conditions for the banking sector are jointly associated to a country's exposure to the risk of a banking crisis? Can we derive "thresholds" for vulnerability/resilience indicators that effectively identify high/low banking risk profiles?

We employ a state-of-the-art empirical technique borrowed from the statistical literature on "perturbation and combination" algorithms (Breiman, 2001; Friedman and Popescu, 2003) in order to build an early warning model for banking crises in emerging markets. The methodology condensates the relevant information on future crises into the monitoring of a small set of predictors, their interactions, and "threshold" values. Technically, our approach offers an interesting solution to the trade-off between the "in sample" fit and the "out-of-sample" predictions. Typically, there is a trade-off between the model's ability in fitting the data, and its out-of-sample predictive power. This goal is achieved by applying the CRAGGING algorithm of Vezzoli and Stone (2007). The algorithm effectively mimics the possibility of extracting multiple random samples from the original data, so that the resulting model performs well when applied to new data. We compare our approach with other

commonly used techniques, such as Stepwise Logit and Classification Trees, and we rank them in terms of their out-of-sample prediction accuracy. The main findings of the paper are as follows.

First, not all banking crises are equal: our early warning model identifies two *basic types* of banking crises in emerging markets. The first (that we label "Latin American crisis") is the result of a (past) credit boom and a flight from domestic assets, followed by very high levels of interest rates on bank deposits, possibly reflecting fears of a run, and/or serious banks' illiquidity. The second type ("Asian crisis") is preceded by a boom in real investment, which is typically financed by banks' foreign debt.

Second, out of a large number (540) of possible predictors, ranging from macroeconomic to financial sector's balance sheet indicators, we show that banking crises can be accurately predicted by just a handful of variables: the interest rate on deposits, the net accumulation of foreign asset/liabilities of the banking sector, the change the ratio of domestic credit to GDP, the ratio of real investment to GDP.

Third, while all the examined competing econometric models present similar prediction capabilities in-sample, our model clearly dominates in terms of out-of-sample predictive power and robustness.

The plan of the paper is as follows. Section 2 reviews the empirical literature on early warnings for banking crisis; Section 3 describes the data set, and Section 4 presents the methodology. Section 5 discusses the model as well as three alternative benchmarks. Section 6 contains the horse race between the different models and Section 7 concludes.

2. Review of the Literature

Our study complements the vast empirical literature on Early Warning Systems (EWS) for banking crisis. This literature mainly concerns emerging markets, and has applied a variety of empirical methodologies, spanning from parametric (such as Logit or Probit regression) to non-parametric models (such as the Signal-to-Noise approach or Classification and Regression Trees). Among the papers employing the former methodology, Eichengreen and Rose (1998) use multivariate Probit regression for understanding the role of international shocks in determining banking crises in developing and emerging economies. They find that interest rate shocks stemming from OECD countries as well as their GDP growth rate have a significant and strong effect on bank fragility in developing countries. Hutchison (2002) estimates multivariate Probit models linking the probability of banking risks to domestic macroeconomic variables and institutional characteristics. He finds that the following variables are strongly and significantly associated to banking crisis: inflation, GDP growth, exchange rate turbulence [1] as well as variables describing the characteristics of the financial and the regulatory

[1] Such a measure is constructed according to Glick and Hutchison (2001) from "large" changes in a weighted average index of monthly real exchange rate changes and monthly reserve losses. The weights are inversely related to the variance of

environments. González-Hermosillo (1999) is one of the first papers to exploit banks' balance sheets at the micro level in order to explore the determinants of banking risks in a Panel-Logit framework. She finds that non-performing loans and capital asset ratios appear to be the best predictors of banking crisis and that individual bank failures are not significantly affected by the systemic banking fragility, so that, despite balance sheet linkages, she finds little evidence of "contagion" between banks.

Demirgüç-Kunt and Detragiache (1998) find that low GDP growth, high real interest rates, and high inflation are significantly correlated with the occurrence of a banking crisis. Demirgüç-Kunt and Detragiache (2000) compare the results with the signal approach of Kaminsky and Reinhart (1999), and show that their model performs better in terms of in-sample prediction. They find that banking crises are associated to low GDP growth, high real interest rates, and high inflation, strong growth of past bank credit and a large ratio of broad money to reserves. Hardy and Pazarbasioglu (1999) have similar findings and add real depreciation of the exchange rate and trade balance deterioration to the list of early warning predictors of banking crises. Kalotychou and Staikouras (2005) focus on Latin American and Asian countries. They find that debt and trade developments are correlated to the solvency of banks (but only for Latin American countries). More recently, Kraft and Galac (2007) examine the banking crises occurred in Croatia using Logit models, and find that high deposit interest rates were the most significant variable predicting bank failures. This finding is consistent with Keeley (1990), who argues that the increase in risk-taking following deregulation would lead to higher deposit interest rates, and with Hellman, Murdoch and Stiglitz (2000), who provide a theoretical connection between financial liberalization and the degree of the moral-hazard problem in the banking system, which leads towards financial crises (they find that under freely determined deposit rates, banks offer Pareto-inefficiently high deposit rates in an effort to steal share from their rivals, thus increasing the moral-hazard problem). The role of high interest rates, particularly on deposits, as a crucial predictor of banking crises, emerges quite clearly in the literature and is confirmed, and made precise, by our own analysis

Davis and Karim (2008) employ an eclectic approach for predicting banking crises. They show that a multinomial Logit model is better suited for providing early warnings for systemic/global crises, whereas the signal extraction approach of Kaminsky and Reinhart (1999) performs better for country-specific ones. In terms of individual leading indicator performance, the Logit approach selects the ratio of M2 to official reserves and budget balance as the best indicators, while the signal approach identifies real GDP growth and the changes in the terms of trade as the best leading indicators. Finally, Lund-Jensen (2012) derives risk factor thresholds in the Logit framework and estimates the time varying conditional probability of a systemic banking crisis. He argues that this approach dominates the popular Signal-to-Noise approach in terms of type I and type II errors, and obtains the following list of leading indicators: banking sector leverage, equity price growth, the credit-to-GDP gap, real effective exchange rate appreciation, changes in the banks' lending premium and the degree of banks

---

changes in each component over the sample for each country. Large changes in exchange rate pressure are defined as changes in the index that exceed the mean plus 2 times the country-specific standard deviation.

interconnectedness as measured by the ratio of non-core to core bank liabilities.

On the non-parametric front, the prediction of banking crisis mainly employs the signal extraction approach and the Classification and Regression Trees methodologies. The former was firstly applied to banking crises by Kaminsky and Reinhart (1999), who focus on the 'twin crises' (currency and banking crises). They find that problems in the banking sector typically precede a currency crisis and such a link is bi-directional, since the collapse of the currency often generates a banking default, if assets are denominated in domestic currency and liabilities in foreign currency. Both types of crisis often occur during recessions, induced by either a worsening of the terms of trade, an overvalued exchange rate, a credit crunch or/and a sharp contraction in exports. Furthermore, financial liberalization and/or increased access to international capital markets fuels are often associated to credit booms and expose the banking sector to stock market and real estate crashes. Kaminsky (1999) and Goldstein, Kaminsky and Reinhart (2000) refine this approach by combining several indicators into an aggregate composite index. Also, the same signal-based approach is used by Borio and Lowe (2002) and Borio and Drehmann (2009). In the first study, the authors conclude that credit aggregates and asset prices indicators together with the real exchange rate and the rate of growth are leading indicators of banking distress in both developed and emerging markets. In the second study, the authors add property prices to the list (Borio and Drehmann, 2009).

Duttagupta and Cashin (2011) employ Classification and Regression Trees to study the determinants of banking crises in 50 emerging and developing countries during 1990-2005. The authors identifies three key crisis-prone conditions: (i) macroeconomic instability, with high annual inflation and relatively low terms of trade growth; (ii) low bank profitability, proxied by the spread between lending and deposit rates; (iii) high foreign exchange risk, with liability dollarization (foreign exchange deposits to official foreign exchange reserves) combined with either (a) relatively high depreciation, or (b) low bank liquidity (private credit to deposits). This is the paper which is closest, in terms of methodology and results, to ours. We improve on this paper, and, we think, on the literature, by introducing the state-of-the-art algorithm developed by Vezzoli and Stone (2007). The main advantages of our approach is that it extends the merits of the Classification and Regression Tree methodology, by preserving the information structure contained in the panel dataset and by substantially improving the robustness and the reliability of the model for generating out-of-sample predictions and vulnerability assessment.

3. The Data Set

We use the dataset on banking crises put together at the IMF by Laeven and Valencia (LV, 2010), which over the period of interest, 1980-2010, covers 85 emerging markets. The choice to focus on emerging markets, despite the recent wave of banking crises in the US and Europe (Iceland, Ireland and Spain, and the Scandivian episodes in the 1990s) is due to the fact that we want to test our methodology on a long and possibly homogenous set of

banking crisis episodes and want to compare our results with those of a well-established literature. We leave an application of our methodology to the European crises to a following paper.

The in-sample analysis (1980-2007) features 79 episodes classified as "crisis". For the more recent years, 2007-2010 we combine the LV data with the Reinhart and Rogoff (RR, 2010) crisis periodization, and we obtain five extra episodes (Hungary, Kazakhstan, Latvia, Russia, Ukraine) all occurring in 2008. LV define an episode of banking crisis to occur when:

"… a country's corporate and financial sectors experience a large number of defaults and financial institutions and corporations face great difficulties repaying contracts on time. As a result, non-performing loans increase sharply and all or most of the aggregate banking system capital is exhausted. This situation may be accompanied by depressed asset prices (such as equity and real estate prices) on the heels of run-ups before the crisis, sharp increases in real interest rates, and a slowdown or reversal in capital flows. In some cases, the crisis is triggered by depositor runs on banks, though in most cases it is a general realization that systemically important financial institutions are in distress".

The definition of RR is quite similar, since they identify a banking crisis by the following two events: (1) a bank runs that leads to the closure, merging, or takeover by the public sector of one or more financial institutions; or, in the absence of bank runs (2) a closure, merging, takeover, or large-scale government assistance of one or more important financial institution. The episodes of banking crises (country-year) that appear in our dataset are reported in Table 1. Over the period 1980-2010 there are 84 banking crises, with an average duration of nearly 3 years. They are clustered over the period 1995-1998, when one third of the total events (31 banking crises) takes place. More information on the data set is available in the LV paper. In constructing our "banking crisis" dummy variable we use only the first year of "crisis" and ignore the observations that occur during the following years in order to avoid endogeneity problems. The source for the explanatory variables is the Economist Intelligence Unit, which collects data from the OECD, the IMF, the BIS and the World Bank.

In order to use our algorithm in a computationally efficient way we need the number of our candidate explanatory variables not to exceed about one hundred. Therefore we implement the following three step procedure. First, we start with an initial list of 105 variables, taken from the literature (we constraint the choice by excluding variables for which missing values in the sample period exceed 30% of the observations). By computing, where appropriate, GDP ratios, rates of changes (prefix "D"), first differences (prefix "DIF"), and considering two lags for each variable, we end up with a total of 529 candidate explanatory variables. Second, we employ the Random Forest algorithm (see Manasse and Roubini, 2009 for a discussion and an application) as a pre-selection device, and we retain only the variables with significant explanatory power[2]. This reduces the number of candidate predictors to 74. The final list of candidate predictors is in Table 2, which reports the

[2] Specifically, we used the Variable Importance measure (Mean Decrease in Accuracy) retaining all variable showing a measure greater than zero.

variables names (excluding transformations). The list includes: (1) standard macroeconomic and national accounts variables; (2) variables from the balance of payments statistics; (3) external and domestic debt; (4) interest and exchange rates; (5) fiscal accounts and monetary aggregates; (6) banking sector balance sheet data; (7) global factors such as the Oil price, the TED spread (three-months Libor minus three-month US T-bill rate) and the Libor (with three-months maturity).

4. Methodology

Our methodology is a refinement of the Classification and Regression Trees (CART) technique developed by Breiman et al. (1984) and widely employed in different research fields such as genetics, engineering, marketing, finance. Manasse and Roubini (2009) apply CART for predicting sovereign debt crises, while Chamon, Manasse and Prati (2007) apply it to "sudden stops". In short, CART is an algorithm that partitions the sample space of covariates, **X**, into homogeneous subsets with respect to the dependent variable *Y*. For example, in a binary problem where the *Y*'s are "crisis/non-crisis" indicators, the aim is to find "criteria" (expressed as inequalities) for the **X**, such that as many "crisis" as possible fall in one subset, and as many "non-crisis" fall in a different one. The aim is to identify the joint characteristics (the variables in **X** and their thresholds) that are associated with a "frequent occurrence" of the phenomenon of interest. For example, in the field of medicine, the likelihood of a heart attack may be found to be associated with a particular combination of individual characteristics, such as sex (= male), age (= above 45), blood pressure (= systolic pressure below 20), so that a patient jointly satisfying these traits can be considered "at risk". Thus CART iteratively uses the actual values of predictors as "thresholds" that identify cases of disproportionate occurrence of crises. This technique presents numerous advantages (it is non-parametric, effectively deals with non-linearity, selects thresholds and interactions out of possibly a very large set of predictors), but it suffers from three main drawbacks (which, to be fair, also affect most standard econometric methods); see, among others, Chamon, Manasse, and Prati (2007). The first is model instability. The variables selected by the algorithm in the "top nodes" clearly have a disproportionate effect on the following choice of predictors, so that even a relatively small change in the data set (new observations on existing variables, or the addition of new variables to the data set) can give rise to very different models. The second is that CART's predictive accuracy "out-of-sample" is often worse than that of other "hybrid" methods, those that combine different statistical techniques or different models (see Steinberg and Cardell, 1998). Finally, CART is designed for cross-sectional data, and therefore makes an inefficient use of information in panel data sets, since the observations relating to a particular unit, say a country, are effectively treated as if they concerned different units.

The issue of predictive accuracy has been addressed by a large literature in statistics and econometrics. Recent contributions in statistics have developed the "perturbation and combination" approach. The idea is to take the

original data set and perturb it by creating many "artificial" samples, so that one can estimate multiple models and can generate multiple predictions which are can then averaged. This approach is followed in the "bagging" algorithm (Breiman, 1996), where estimation samples (training sets) are randomly generated by bootstrapping. The aim is to reduce the potential instability of forecasts and also to control for the problem of over-fitting. A similar approach is that of the "random forest" algorithm (Breiman, 2001), where many classification trees are calculated, by randomly selecting variables in different combinations as well as by generating different sub-samples; in the "boosting" approach of Freund and Schapire (1996), the idea is to generate multiple simple "rule-of-thumb" models for random portion of the data and then to combine them. All these statistical approaches generate predictions (a crisis probability for each observation, country-year) by averaging the outcomes of thousands of possibly different models. Thus they rely on black-box model which allow for no "economic intuition" and are impossible to interpret. In other words, predictive accuracy is obtained at the expenses of interpretability. Our approach here exploits the basic idea of "perturbation and combination" for achieving predictive accuracy without sacrificing the economic intuition.

The other improvement concerns the efficient use of the panel data structure of the data. The CART algorithm was designed for cross-sectional data, where each observation pertains to a particular individual. By converse, the CRAGGING algorithm preserves the information contained in the data by treating all observations pertaining to an individual country as particular item (see below).

4.1. The CRAGGING algorithm

The empirical methodology, called CRAGGING (CRoss-validation AGGregatING) is formally presented in Vezzoli and Stone (2007) and Savona and Vezzoli (2012). Let (*Y*, *X*) be a panel data with *N* observations containing *j* units (in our case countries), with $j = 1,\ldots, J$, observed at time *t,* with $t = 1,\ldots, T$, covering a period from 1980 to 2007. Let $\mathcal{L} = \{1,\ldots, j,\ldots,85\}$ be the set of the units and $x_{jt-1} = \left(x_{1jt-1},\ldots, x_{rjt-1},\ldots, x_{Rjt-1}\right)$ be the vector of predictors relative to the unit *j* measured at time $t-1$ with $j \in \mathcal{L}$. Letting $R = 74$ denote the number of explanatory variables for each country; consequently *X* is a $N \times R$ matrix. The "perturb and combine" technique consists in a two steps procedure.

In the *first step*, $\mathcal{L}$ is randomly partitioned into $V = 17$ subsets, denoted by $\mathcal{L}_v$ (test set) with $v = 1,\ldots\ldots,17$, each one containing $J_v = J/V = 5$ different countries (see Figure 1). The number of observations pertaining to each test set is denoted by N_v. We then randomly select a test set (e.g. $v = 1$: Brazil, Argentina, Egypt, Hungary, Latvia in Figure 2) that is taken out of the observations used for estimation and reserved for testing. Thus the corresponding training (estimation) set is denoted by $\mathcal{L}_v^c$, obtained as $\mathcal{L}_v^c = \mathcal{L} - \mathcal{L}_v$ and containing $J_v^c = 80$ units

and N_v^c observations. It follows that $\mathcal{L}_v^c$ includes $17-1=16$ test sets. By removing one unit (country) ℓ from $\mathcal{L}_v^c$ we estimate a regression tree[3] on the perturbed training set $\left(\mathcal{L}_{v\setminus\ell}^c\right)$ (see step A in Figure 2). After the perturbation, $\mathcal{L}_{v\setminus\ell}^c$ contains 79 countries $(80-1)$. Next, we test the regression tree on the set $\mathcal{L}_v$. At this point, the ℓ-th country is reinserted in the training set and the same procedure is repeated for each ℓ in $\mathcal{L}_v^c$ (see step B in Figure 2), thus obtaining 80 predictions of default probabilities for each country-year belonging to the test set (from $\hat{y}_1$ to $\hat{y}_{80}$ in Figure 2). These $\hat{y}_i$, with $i=1,\ldots\ldots,80$, are averaged for each unit of the test set (see step C in Figure 2) and a single predicted probability of crisis is obtained for each country-year of the test set . This particular type of perturbation, conceived with the end of maintaining the hierarchical structure of the panel data, is repeated for the remaining 16 test sets. In this way, each unit always appears either in training set used for estimation, or in the test set used for validation, but never in both at the same time. More precisely, we estimate $V\times J_v^c=17\times 80=1360$ regression trees, adopting what statistician call "double cross-validation": the l*eave-one unit out* cross-validation and the *v-fold* cross-validation (when sample observations are randomly split into *v* sub-samples and one, in turn, is reserved for validation purposes). To increase the accuracy of predictions, we repeat the original partition of the set $\mathcal{L}$ 40 times (for a total of $1360\times 40=54400$ regression trees), thus generating different groupings of the original 85 countries. We thus simulate situations in which the available data for estimation are different. We end up by averaging all the predicted crisis probabilities for each observation in the sample computing $80\times 40=3200$ predictions of crisis probabilities for each country-year observation.

The *second step* takes care of interpretation. We replace the original binary (0, 1) definition of crisis by the probabilities generated in step one, so that the original "coarse" definition of a crisis is replaced by a continuous variable, approximating the "degree of vulnerability" to a crisis. We use this as a dependent variable of regression tree and obtain the best predictors, thresholds and interactions, as usual. Implicit here is the idea that not all banking crises are alike, *ex ante*: they stem from a combination of different fundamentals' vulnerabilities and random events. Thus the binary indicator represents a necessary simplification, a coarse description of situations which *ex-ante* had different likelihoods of ending up in a crisis, all though *ex*-post all did. We exploit the continuous measure derived in step 1 in order to estimate a (Final) Model that can be interpreted and used for prediction.

5.1. The (Final) Model

[3] The regression tree is estimated in correspondence of a tuning parameter, called α, that modulates the trade-off between the complexity and the interpretability of the results.

The procedure outlined in the previous section generates the model depicted in Figure 3. The model selects the interest rate on bank deposit (lagged one year, RAT2) as the most important indicator of crisis vulnerability. In fact, in episodes when this rate exceeds 16.2 percent, the probability of observing a crisis in the following period rises from 4 percent, the ex-ante sample frequency, to 10.1 percent, as we move from the top node down to the right in the figure. Conversely, a deposit rate below this threshold is associated to a crisis probability of only 2.6 percent, as we move to the left of the root node.

When the interest rate is above the threshold, two other variables are significantly associated to a higher probability of a banking crisis. The first is an abnormal year-to-year jump the banking sector's net acquisition of foreign assets, occurring two years before the crisis (the rate of change of net foreign assets DNFAS, lagged twice, is above 80 percent), so that either domestic banks sharply raise their investments abroad and/or foreign residents reduce their domestic deposits, or both; this early capital flight is associated to a 10.5 percent probability of crisis. The second is a credit boom two years before the crisis (an *acceleration* of 11.62 percentage points in the rate of growth of domestic credit growth, DIFSODD), which pushes the default probability to 42.4 percent. It is interesting to note that, as we saw in the literature review section, there is a large consensus on the importance of interest rates, credit booms and capital flights as predictors of EM's banking crises. High deposit rates signal lack of confidence in the banking system, a prelude to a run, and/or a large liquidity problem forcing banks to raise their deposits rates in order to contrast a flight of deposits. Similarly, capital flights are known to be associated to currency collapses and balance sheet mismatches, while credit booms fuels stock market and real estate bubbles. The characteristics of node 8 fit the stylized facts of many Latin American episodes in the mid-nineties (for example Costa Rica and Venezuela, 1994 share all the characteristics of this node).

The model also identifies a somewhat different type of risky environment, one of a real investment boom financed by foreign borrowing of the banking sector, see node 4 on the left side of the tree. Here, despite lower interest rates on deposits, we observe three signs of vulnerability. The first is that the banking sector has large foreign net liabilities (net foreign assets, NFAS are negative and below 5% of GDP). Also, these liabilities are not being reduced at a sufficient pace (the lagged rate of change of banks' foreign liabilities is above -77% year-to-year). These two features are associated to a crisis probability of 7.8 percent. The other early sign of weakness is an investment boom that pushes the ratio of investment to GDP above the threshold of 39.06 percent. Under this scenario the probability of a banking crisis moves from the unconditional 4 percent in the root node, to 36.8 percent in node 4. Here we find episodes epitomized by the "Asian crisis" label, which were characterized by a real investment boom fuelled by domestic banks foreign borrowing, and where a currency crash lead to a "twin" bank and sovereign default. Episodes such as Malaysia and Thailand 1997, indeed belong to node 4.

5.2. The Competitor Models

Since the proof of the pudding is in the eating, the merits of our model rely in its ability to accurately predict banking crises. For making a comparison we need some benchmark model(s). We could pick models in the literature, but that would not be very useful as forecasting performances depend on the actual data used for estimation and forecast. We prefer to build, to our best knowledge, three alternative models using the same sample for estimation and prediction. The first model is a Stepwise Logit (Slogit), the second is an ordinary Classification and Regression Tree (RT), and the third model is an average of the predictions of all the models, namely our "Final" Model (FM), the Slogit, and the RT.

5.2.1. The Stepwise Logit

The Logit approach is very common for estimating banking and sovereign debt crisis probabilities, as discussed in Section 2. We build our own model by following this procedure. We start from the list of the 74 variables selected for our FM. The Stepwise procedure includes one variable at a time into the regression and uses the criterion of adding additional variables as long as this reduces the AIC (Akaike Information Criterion), starting from the variable that achieves the largest reduction, until it is not possible to obtain further improvements in the AIC. By this automatic procedure we obtain Model 3 in Table 3. Model 2 is the most parsimonious model, and it is obtained by eliminating from Model 3 all statistically insignificant variables as well as variables which, despite being statistically significant, show the "wrong" sign. This is the case of the change in the total external debt/GDP ratio (DIFTED), the change in the US T-bill rate (DIFTBILL), which ("wrongly") appear to be negatively associated to the probability of a crisis, as well of the percentage change of commercial banks' net foreign assets (DNFAS), which appears with a counter-intuitive positive coefficient. Our preferred model is Model 1 which differs from Model 2 because it adds in two variables, the deposit interest rate (RAT2) and the percentage variation in BIS banks' total liabilities (DBSDT), which are a priori plausible (have the "right" positive sign), despite not being statistically significant. In the end, this model features the following explanatory variables: the ratio of exports to GDP lagged 2 years (PEXPLAG2), and the price of oil (OILLAG), which captures the impact on the financial health of oil producers, with expected negative sign; a positive association to the probability of a crisis is found for the rate on deposits (RAT2LAG), the average interest cost of external debt (EFIRLAG), the acceleration in BIS banks' total liabilities (DBSDTLAG), the change in the ratio of short-term banks' liabilities with respect to GDP (DIFBSD1_GDPLAG), the level of the US T-bill rate (TBILLLAG), the TED spread lagged 2 years (TEDLAG2), a proxy of banks' borrowing costs, and finally the percentage change in commercial banks' net foreign assets (DNFASLAG).

5.2.2. Classification and Regression Tree and the Average Model

The second benchmark model is a Classification Tree, obtained from the same 74 variables obtained in the pre-selection stage described before. The main difference with respect to the FM is that the latter is built to enhance its robustness to out-of-sample data. Overall, RT choice of variables is not that different from that of the FM: high deposit rates, credit booms, large foreign debt and capital flight are confirmed as most relevant early warnings of banking crises. The model is shown in Figure 4.

Interestingly, the top variable selected by RT, the interest rate on banks' deposit, and its threshold value, 16%, are the same as those picked by the FM, which confirms this variable as the most critical early warning for banking crises. The consequences of the "perturb and aggregate" procedure however show up in the subsequent "splits" of the data. Unlike our model, RT identifies two risky sets of characteristics, both occurring when banks' deposit rates exceed 16.24%. The first, see node 5, occurs when the country has suffered a large liquidity problem (when its "external financial requirements", FNRQ, the sum of current account deficit and debt coming due, exceed 10% of GDP), it has experienced a large drop in Net Income flows in percentage of GDP (IGDP), and when the country has substantially increased its medium and long term borrowing from abroad (including commercial bank loans, official guaranteed loans and international bond issues). These circumstances are associated to a 28.4 percent probability of bank default. The second risky combination is found in node 4, where in addition to high deposit rates and financial requirements, the banking sector has expanded its lending very rapidly (the rate of change in bank lending, SODC, exceeds 15%).

Finally, the last benchmark model is built by following the averaging approach common in the EWS literature (e.g. Fuertes and Kaloutychou, 2007), which argues that often the best predictions are obtained by combining the predictions of many competing models. Thus, we take the simple averages of the crisis probabilities generated for each observation by our three models (FM, Slogit, RT), and we label this black-box model as Average Model (Ave).

6. Comparing the Models' Performances

In this section we assess the performances of the different models in predicting banking crises, both "in sample", over the same observations used to estimate the model, 1980-2007, and out-of-sample, 2008-2010. The results are sensitive to the choice of probability threshold above which one decides to classify an observation as "crisis". This choice determines the trade-off between type I error (probability of "missing" a crisis) and type II error (probability of erroneously "calling" a crisis, or "false alarm"). Should one choose a very low threshold (e.g. zero) the models would always "call a crisis" (and correctly predict all crises, with zero type I error) at the expenses of always misclassifying "non-crises" (100% type II error). Therefore we report each model's

predictive accuracy, for both crisis and non-crisis episodes, at different cut-off points.

6.1. In-Sample Accuracy

We start from the in-sample performance. Figure 5 reports the percentage of crisis episodes correctly classified (on the y-axis), for each value of the cut-off point (x-axis). The FM is represented by a dot, the Slogit by a trait, the RT by a white square, the Ave by a triangle. Clearly, at very low thresholds, below 2 percent, all model perfectly predict crises (they *always* predict crises), while at very high levels, above 50 percent, all models always fail to predict crises (they always predict a non-crisis). For the reason previously discussed, the percentage of correct predictions of crisis falls as we raise the threshold. Models, however, behave very differently for thresholds in the range 2-5 percent. It makes sense to focus on what happens around the cut-off point of 3.55 percent, which is the sample frequency of actual defaults in the sample (79 defaults over a total of 2,222 pooled observations), so that we may decide to classify an observation as "crisis" whenever the model assigns to it a probability higher than the sample frequency. First note that that the ranking of the models' performances changes with the threshold: for example, at 3.55 percent the Slogit is the best model (it predicts 70% of crisis) and the RT is the worse (only 42%), with our FM in the middle. However, when we raise the cut off to 5 percent, the Ave becomes the most accurate; the Slogit is only third-best while the FM keeps its position. Figure 6 shows the accuracy in predicting non-crises. As we raise the cut-off probability for calling a crisis, all models predicts more tranquil episodes. Relative to the previous case, there is less differentiation among models because tranquil times represent about 96.5 percent of the observation and are much easier to predict. At the benchmark cut-off of 3.55 percent, we have a perfect reversal of the previous models ranking: the Slogit, which was the best (70%) for crisis prediction, becomes the worst for non-crises (60%) because of the large number of "false alarms" it implies, and the opposite occurs for the RT.

A general way to summarize the models' performances is to consider both Type I (missing crisis) and Type II (false alarms) errors. Let *LF* denote a loss function, defined as a weighted average of the two probability errors :

$$LF(\delta) = \delta \cdot Type\ I\ error + (1-\delta) \cdot Type\ II\ error$$

By letting the weight δ vary between 0 (zero cost assigned to missing a crisis relative to missing a tranquil episode) and 1 (very high relative cost assigned to missing a crisis), we can measure the extent of the prediction errors of each model, for each cut-off point. Figure 7 shows the LF associated to each model, for different cut-off probabilities. Each model is represented by a vertical segment and a box. The segment's extremes, outside the box, correspond to the largest/lowest values of the LF taken as δ varies on the unit interval. The boxes' floor and ceiling correspond to the loss evaluated, respectively, at $\delta = 1$ and $\delta = 0$. For example, the Slogit at 3.55% cut-off correctly predicts 70% of crises (see Figure 5) so that the box's floor $\delta = 1$ corresponds to

$LF(1) = Type\ I\ error = 1 - 70\% = 30\%$. Similarly, the Slogit predicts about 60% of tranquil episodes (see Figure 6), so that at $\delta = 0$, the box's ceiling corresponds to $LF(0) = Type\ II\ error = 1 - 60\% = 40\%$. The dot inside the box shows the value of the Loss for $\delta = 0.5$, when the two errors have equal weights, $LF(0.5)$ The diagram gives a visual representation of both the model "average" performance (the dot) and of its robustness to different preferences on δ (the height of the box). Figure 7 shows that the models do not substantially differ in terms of their average in-sample accuracy, although they vary considerably in terms of their robustness to type I and type II errors. In particular, using the cut-off 3.55%, the Slogit and FM models are much less sensitive to changes in the preference parameter δ than the other models.

6.2. Out-of-Sample Accuracy

Predicting out-of sample is much more difficult, since there are only 5 banking crises (Hungary, Kazakhstan, Latvia, Russia, and Ukraine), among the 255 observations of the period 2008-10. Looking at crisis prediction, Figure 8, we see that most models either always call a crisis (at low cut-off point), or never call a crisis, at high cut-offs. Note, however, that for thresholds crisis probabilities between 3.55% and 5%, the FM model is the only one which consistently correctly predicts *all* the five crises, while the others predict none. As Figure 9 shows, this does not come at the cost of many false alarms: while correctly classifying all crises, the FM correctly classifies 80% of tranquil episodes. Figure 10 presents the Loss Function plots for cut-off points of 2, 3.55, and 5%. For the latter two values, the FM has an average loss function about one fifth of the competing models; moreover, FM is by far the most robust to changes in the risk preference parameter, so that it clearly outperforms all other models for most preference weights assigned to the different error types.

7. Conclusions

This paper has employed a state-of-the-art algorithm (CRAGGING) borrowed from the statistical literature in order to build an early warning model for banking crises in emerging markets. The methodology condensates the relevant information on future crises into the monitoring of a small set of predictors, interactions, and "threshold" values. The idea is to take the original dataset and perturb it many times by creating "artificial" samples, so that we can construct a flexible model which may perform well when applied to new data, while at the same time preserving the model's interpretability. We find that, out of a large number (540) of possible predictors, from macro to balance sheet indicators of the country's banking sector, we can accurately predict banking crises by just a handful of variables: the interest rates on bank deposits, the net accumulation of foreign asset/liabilities of the banking sector, the change the ratio of domestic credit to GDP, the ratio of investment to GDP. Our model identifies two *basic types* of banking crises in emerging markets. The first ("Latin American type") is the result

of a (past) credit boom, a flight from domestic assets, followed by very high levels of interest rates on deposits, possibly due to a bank run, and/or serious banks' illiquidity. The second ("Asian type") is characterized by an investment boom financed by banks' foreign debt. We also compare our model to a Stepwise Logit, a Classification Tree, and an "Average" model, built on the same data. While all the econometric models show similar prediction capabilities in-sample, our model strongly dominates the others in terms of out-of-sample predictive power, both because it delivers more accurate predictions of crisis and less false alarms, and because it is more robust with respect to different preferences regarding the costs of missing a crisis or sending out a false alarm.

References

Borio, C. and P. Lowe (2002). "Assessing the Risk of Banking Crises", *BIS Quarterly Review*, December, 43-54.

Borio, C. and M. Drehmann (2009). "Assessing the Risk of Banking Crises – revisited", *BIS Quarterly Review*, March, 29-46 .

Breiman, L. (1996). "Bagging Predictors", *Machine Learning*, 26, 123-140

Breiman, L. (2001). "Random Forests", *Machine Learning*, 45, 5-32.

Breiman, L., J. Friedman, R. Olshen and C. Stone (1984). *Classification and Regression Trees*. Wadsworth Inc., California.

Chammon, C., P. Manase and P. Prati (2007). "Can We Predict the Next Capital Account Crisis?", *IMF Staff Papers*, 54(2).

Davis, E. and D., Karim (2008). "Comparing Early Warning Systems for Banking Crisis", *Journal of Financial Stability*, 4, 89-120.

Duttagupta R. and P. Cashin (2011). "Anatomy of Banking Crises in Developing and Emerging Market Countries", *Journal of International Money and Finance*, 30, 354-376.

Demirgüç-Kunt, A. and E. Detragiache (1998). 'The Determinants of Banking Crises: Evidence from Developing and Developed Countries", *IMF Staff Papers*, 45, 81-109.

Demirgüç-Kunt, A. and E. Detragiache. (2000). "Monitoring Banking Sector Fragility: A Multivariate Logit", *World Bank Economic Review*, 14(2), 287-307.

Eichengreen, B. and A. Rose (1998). "Staying Afloat When the Wind Shifts: External Factors and Emerging-Market Banking Crises", NBER Working paper no. 6370, January.

Freund, Y. and R. Schapire (1996). "Experiments with a New Boosting Algorithm", *Machine Learning: Proceedings of the Thirteenth International Conference*, Morgan Kauffman, San Francisco, 148-156.

Friedman, J. and J. Popescu (2003). Importance Sampled Learning Ensembles. Technical Report, Department of Statistics, Stanford University.

Fuertes, A. and E. Kalotychou (2007). "Optimal Design of Early Warning Systems for Sovereign Debt Crises", *International Journal of Forecasting*, 23(1), 85-100.

González-Hermosillo, B. (1999). "Determinants of Ex-Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes", IMF Working Paper 99/33, March.

Glick, R. and M. Hutchison (2001). "Banking and Currency Crises: How Common Are Twins?" in R. Glick, R. Moreno and M. Spiegel (eds.), *Financial Crises in Emerging Markets*, Cambridge University Press, Cambridge, UK.

Hutchinson, M. (2002). "European Banking Distress and EMU: Institutional and Macroeconomic Risks", *Scandinavian Journal of Economics*, 104(3), 365-389.

Kalotychou, E. and S.K. Staikouras (2005). "The Banking Exposure to International Lending: Regional Differences or Common Fundamentals", *Financial Markets, Institutions and Instruments* (NYU), 14, 187-214.

Kaminsky, G. (1999). "Currency and Banking Crises: the Early Warnings of Distress", IMF Working Paper, No. 99/178.

Kaminsky, G. and C. Reinhart (1998). "Financial Crises in Asia and Latin America: Then and Now", *American Economic Review*, 88 (2), 444-448.

Kaminsky, G. and Reinhart, C.M. (1999), 'The Twin Crises: the Causes of Banking and Balance of Payments Problems', *American Economic Review*, 89, 473-500.

Goldstein, M., G. Kaminsky and C. Reinhart (2000). "Assessing Financial Vulnerability. An Early Warning System for Emerging Markets", Institute for International Economics, Washington, DC.

Hardy, D. and C. Pazarbasioglu (1999), "Determinants and Leading Indicators of Banking Crises: Further Evidence", *IMF Staff Papers*, 46, 3, 247-58.

Hellmann, T., K. Murdock and J. Stiglitz (2000). "Liberalization, Moral Hazard in Banking and Prudential Regulation: Are Capital Requirements Enough?", *American Economic Review*, 90 (1), 147-165.

Keeley, M. (1990). "Deposit Insurance, Risk and Market Power in Banking", *American Economic Review*, 80 (5), 1184-1200.

Kraft, E. and T. Galac (2007). "Deposit Interest Rates, Asset Risk and Bank Failure in Croatia", *Journal of Financial Stability*, 2 (4), 312-336

Lund-Jensen, K. (2012). "Monitoring Systemic Risk based on Dynamic Thresholds", IMF Working Paper, No 12/159.

Manasse, P. and N. Roubini (2009). "Rules of Thumb for Sovereign Debt Crises". *Journal of International Economics* 78 (2), 192-205.

Laeven, L. and F. Valencia (2010). Resolution of Banking Crises: The Good, the Bad, and the Ugly, ", IMF Working Paper, No 10/146.

Reinhart, C. and K. Rogoff (2009). *This Time is Different: Eight Centuries of Financial Folly*. Princeton: Princeton University Press.

Reinhart, C. and K. Rogoff (2010). From Financial Crash to Debt Crisis. NBER Working Paper no 15795, Cambridge, MA.

Savona, R. and M. Vezzoli (2012). "Fitting and Forecasting Sovereign Defaults Using Multiple Risk Signals." Cà Foscari University of Venice, Department of Economics Working Paper No. 26.

Steinberg, D. and N.S. Cardell, (1998), "Improving Data Mining With New Hybrid Methods", Salford Systems, http://docs.salford-systems.com/newhybridmethods.pdf

Vezzoli, M. and C. Stone (2007). "Cragging", manuscript, Department of Statistics, University of California, Berkeley also published in Book of Short Papers CLADAG 2007, University of Macerata, September 12-14, 2007.

Table 1: Banking Crisis 1980-2010

Start Year	Average Duration	Nobs of crises	Countries
1980	3.0	3	Argentina-Egypt-Morocco
1981	5.0	3	Chile-Mexico-Uruguay
1982	3.0	5	Colombia-Ecuador-Ghana-Kuwait-Turkey
1983	1.8	4	Equadorial Guinea-Peru-Philippines-Thailand
1985	1.0	1	Kenya
1986	1.0	1	Bolivia
1987	3.6	5	Bangladesh-Cameroon-Costa Rica-Mozambique-Tanzania
1988	3.7	3	Ivory Coast-Panama-Senegal
1989	2.8	4	Argentina-El Salvador-Jordan-Sri Lanka
1990	4.3	6	Brazil-Algeria-Burkina Faso-Lebanon-Nicaragua-Romania
1991	3.7	3	Hungary-Nigeria-Tunisia
1992	3.0	3	Estonia-Poland-Kenya
1993	1.5	2	India-Togo
1994	3.1	7	Bolivia-Brazil-Haiti-Costa Rica-Mexico-Venezuela-Uganda
1995	2.7	9	Argentina-Azerbaijan-Latvia-Lithuania-Cameroon-Paraguay-Swaziland-Zambia-Zimbabwe
1996	2.0	3	Bulgaria-Jamaica-Yemen
1997	3.3	6	Indonesia-Malaysia-Philippines-Korea Rep.-Thailand-Vietnam
1998	2.3	6	China-Colombia-Ecuador-Croatia-Russia-Ukraine
2000	2.0	2	Nicaragua-Turkey
2001	3.0	1	Argentine
2002	4.0	1	Uruguay
2003	2.0	1	Dominican Republic
2008	ongoing	5	Hungary-Kazakhstan-Latvia-Russia-Ukraine
Average	**2.806**	**3.000**	

The table reports the number of banking crises occurred over the period 1980-2010. The crises are grouped based on the start date. Average duration is computed in years and it is calculated using the banking crises with the same start date.

Table 2: Banking Crisis Predictors

Code	Description
BEXP	Central government outlays (state budget) (% of GDP)
BREP	Central government receipts (state budget) (% of GDP)
BSD1	BIS banks' liabilities 0-1 year
BSD 1_GDP	BIS banks' liabilities 0-1 year over GDP
BSD3_GDP	BIS banks' liabilities over 2 years over GDP
BSDT	BIS banks' total liabilities
BSDT_GDP	BIS banks' total liabilities over GDP
CARA	Current account balance (% GDP)
COBL	Commercial bank loans
COBL_GDP	Commercial bank loans over GDP
DCPI	Consumer prices (% change in local currency)
DIMP	Imports of Good & Services (% real change)
DTPH	Total external debt divided by population
EFIR	Effective interest rate (%): interest payments made on medium- and long-term debt in current year as a percentage of medium- and long-term external debt at the end of the previous year.
EXCR	Export credits: stock of official export credits, suppliers' credits and bank credits officially guaranteed or insured by an export credit agency.
EXPS	Total exports (free-on-board) over GDP
FNRQ	Financing requirement: current-account balance plus principal due on public and private medium- and long-term debt and IMF debits
FRAS	Commercial banks' foreign assets
FRAS_FRLI	Commercial banks' foreign assets over Commercial banks' foreign liabilities
FRES	Foreign-exchange reserves (excluding gold)
FRLI	Commercial banks' foreign liabilities
FRLI_GDP	Commercial banks' foreign liabilities over GDP
GFDT	Public medium & long-term external debt
IGDP	Net income flows over GDP
IMLT	Interest payments made on medium- and long-term debt
IMPS	Total imports of goods (cost-insurance-freight basis)
LIBOR3	London Interbank Bank Offer Rate 3 Month (%)
LRAT	Commercial banks average lending rate to non-financial enterprises (%)
MCOV	Import cover (months): total international reserves divided by imports of goods and non-factor services expressed in months
MLTD	Medium & LT debt inflows: capital inflows generating medium- and long-term debt, consisting of commercial bank loans, official guaranteed loans and international bond issues
MULC	Official medium- and long-term debt owed to multilateral institutions, excluding the IMF
NFAS	Commercial banks' net foreign assets
NFAS_GDP	Commercial banks' net foreign assets over GDP
NTDT	Net debt: total external debt less total international reserves
OFFC	Medium- and long-term debt owed to official creditors
OFGL	Officially guaranteed loans: public and publicly guaranteed loans, excluding international bonds
OIL	Oil price (in US dollars per Barrel)
PEXP	Exports of Good & Services (% of GDP)
PFDT	Private medium & long-term external debt
PFDT_GDP	Private medium & long-term disbursed external debt over GDP
PFIN	Gross fixed investment (% of GDP)
PGCE	Government consumption expenditure at current market prices (% of GDP)

PIMP	Imports of Good & Services (% of GDP)
PRAM	Principal repayments made on medium- and long-term debt owed to private creditors
RAT2	Deposit interest rate (%)
SMN1_GDP	Stock of money M1 over gdp
SODC	Stock of domestic credit: Bank lending to public and private sectors, plus bank lending in domestic currency overseas.
SODD	Domestic credit growth (%): Percentage change in bank lending to public and private sectors, plus bank lending in domestic currency overseas.
TBILL	U.S. T-bill rate (%)
TDBT	Total external debt stock, comprising public and publicly guaranteed long-term debt, private non-guaranteed debt, use of IMF credit and short-term debt, at end-period
TED	Ted Spread (Libor minus U.S. T-bill)
TFDS	Total foreign debt service over GDP
TSPX	Debt-service ratio: total external debt service paid (% of exports of goods, non-factor services, income and workers' remittances)
TSPY	Debt-service over GDP
TSTD	Short term external debt owed by all sectors
UNEM	Recorded official unemployment as a percentage of total labour force.

The table reports the variables used to estimate predicting models in the empirical analysis.

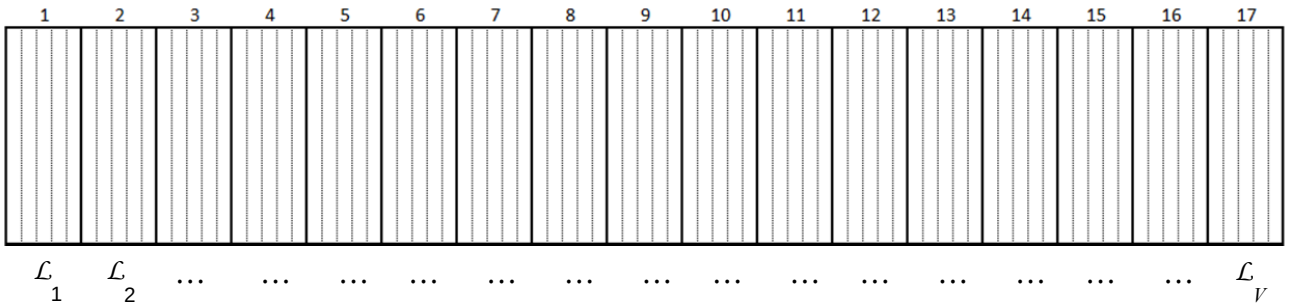
Table 3: Stepwise Logit estimates

Variable	Model 1	Model 2	Model 3
Constant	-2.1517	-2.8823	**-3.0492**
	(-3.9925***)	(-5.8731***)	(-6.7074***)
PGCELAG2	-0.0680		
	(-2.7391**)		
PEXPLAG	-0.0964	-0.0092	
	(-4.1641***)	(-1.1747)	
PEXPLAG2	0.0888	-0.0471	**-0.0533**
	(3.9725***)	(-1.9593*)	(-2.3673**)
DCPILAG2	-0.0009		
	(-1.6868*)		
LRATLAG	-0.0045	-0.0018	
	(-2.1726**)	(-1.5401)	
LRATLAG2	0.0023		
	(2.7271**)		
RAT2LAG	0.0018	0.0019	**4.92E-06**
	(2.3833**)	(2.4926**)	(-0.1505)
TSPYLAG2	-0.0846		
	(-1.5821)		
EFIRLAG	0.0815	0.0573	**0.0455**
	(2.6931**)	(2.2519**)	(1.7952*)
DBSDTLAG	0.2121	0.1766	**0.1098**
	(-1.5146)	(1.7077*)	(-1.269)
DIFBSD1_GDPLAG	6.5886	5.1303	**4.7400**
	(2.5737**)	(2.0918**)	(1.9418*)
TFDS_GDPLAG	7.3240	0.4793	
	(2.0452**)	(-0.1555)	
DPRAMLAG	-0.2362	-0.1135	
	(-2.0367**)	(-1.3314)	
TBILLLAG	2.2808	0.2138	**0.2262**
	(3.6598***)	(3.1565**)	(3.4136***)
LIBOR3LAG2	-2.0063		
	(-3.4205***)		
TEDLAG2	2.0765	0.0988	**0.1104**
	(3.6419***)	(1.6754*)	(1.9083*)
DIFTBILLLAG	-2.1564		
	(-3.3896***)		
DIFTEDLAG	-0.2691		
	(-2.4948**)		
DIFTEDLAG2	-0.2390		
	(-3.1287**)		

DFRASLAG	-0.3618		
	(-1.2602)		
DNFASLAG	0.0152	0.0136	**0.0133**
	(2.0897**)	(1.9098*)	(1.855*)
DNFASLAG2	0.0377		
	(2.1341**)		
OILLAG	-0.0641	-0.0360	**-0.0379**
	(-3.9877***)	(-3.4594***)	(-3.7043***)
FRLI_GDPLAG	0.0993		
	(-0.9202)		
Log likelihood	-279.7018	-306.6920	-317.0375
MSE	0.0302	0.0322	0.0331
Obs with Dep=0	2143		
Obs with Dep=1	79		
McFadden R-squared	0.1802	0.1011	0.0708
Akaike info criterion	0.2743	0.2887	0.2944
Schwarz criterion	0.3385	0.3246	0.3200

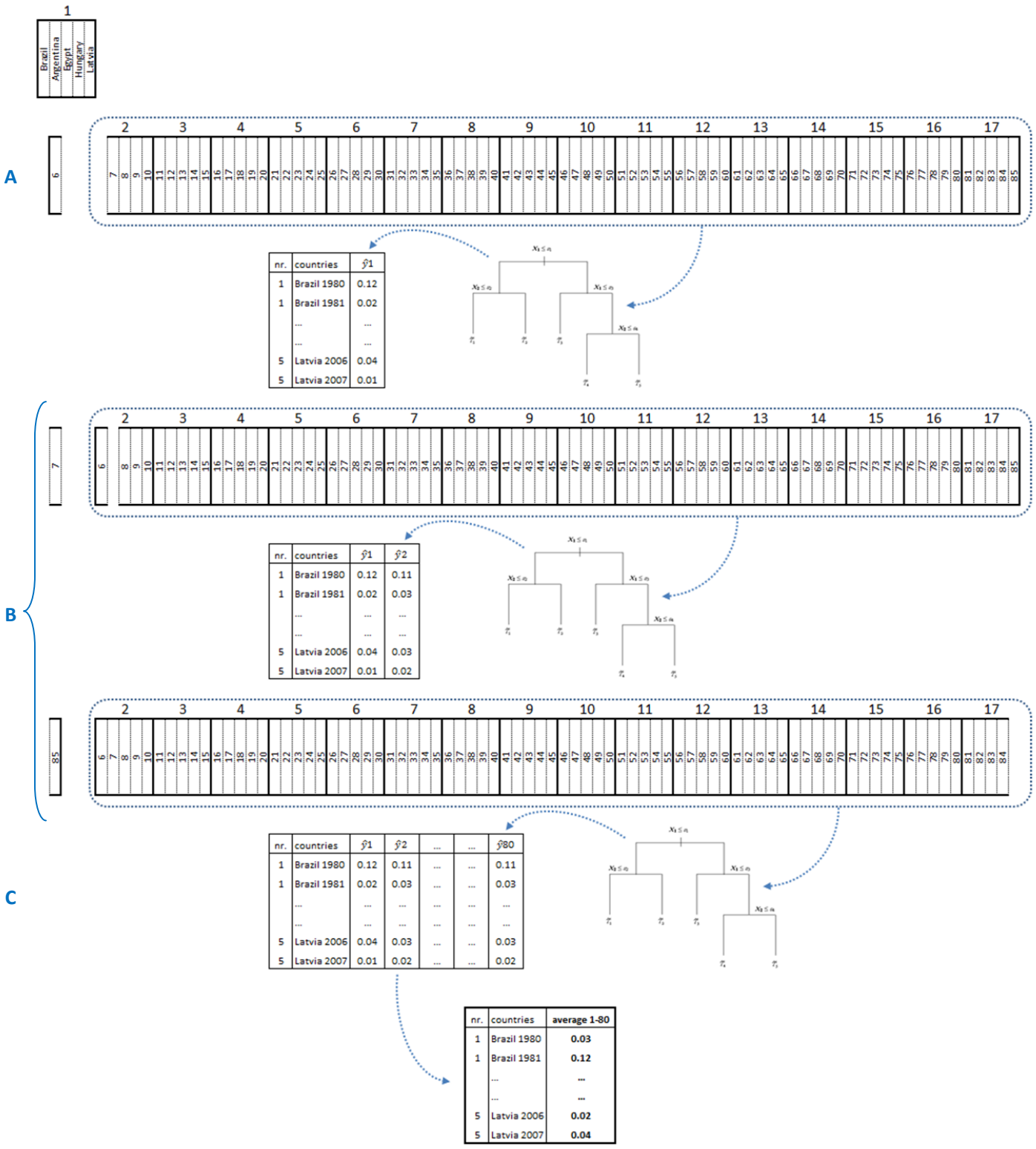
The table reports Stepwise Logit estimates. z-statistics are reported in parentheses. Variables that are significant at the 1, 5 and 10 percent significance level are denoted by ***, ** and * respectively. Model 1 is the pure Stepwise Logit estimation run over the 74 variables used to estimate the FM and RT. Model 2 is obtained by re-estimating Model 1 letting out the insignificant variables and those having the "wrong" sign (i. e. contrasting the economic expectation). Model 3 follows the same construction rule but starting from Model 2. Model 3 is our final Stepwise Logit model with selected variables and corresponding coefficients denoted in bold font.

Figure 1: Partitioning the set $\mathcal{L}$ into V groups



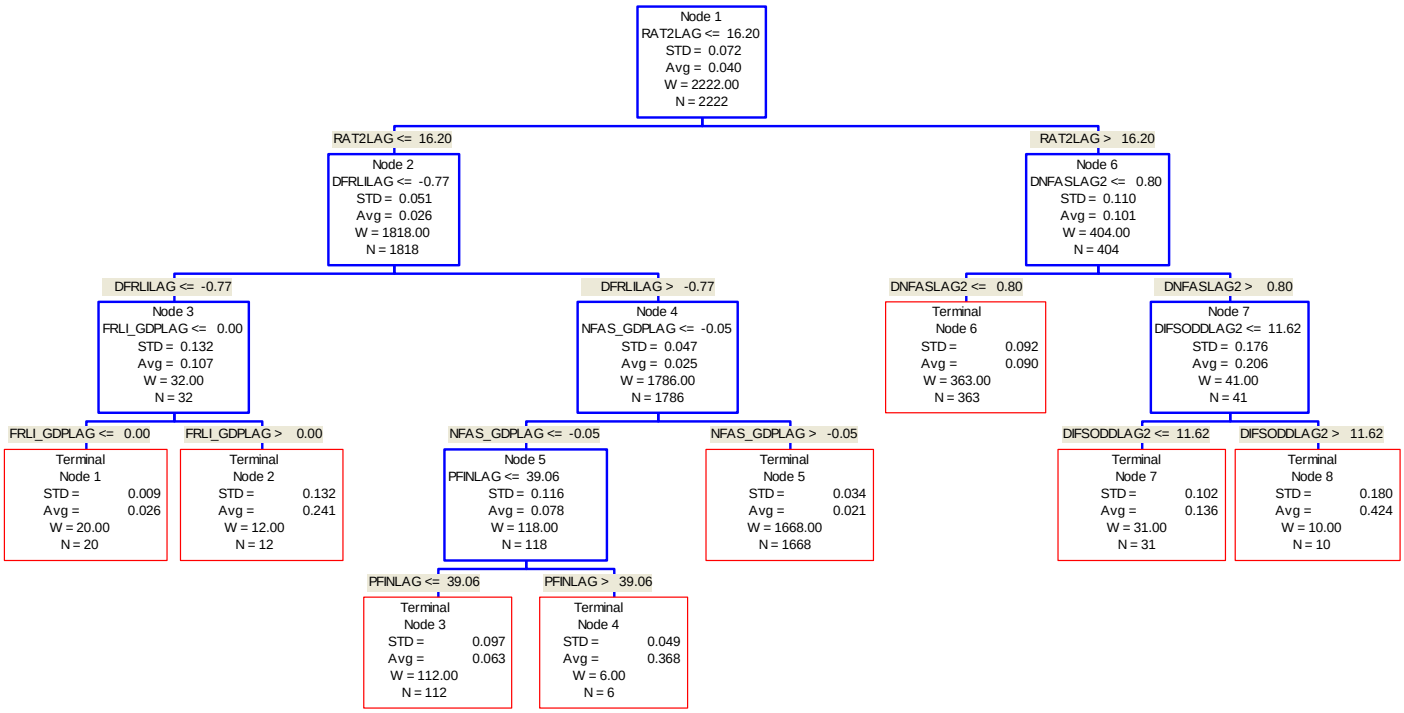
The figure depicts the randomly partition of the set $\mathcal{L}$ in $V = 17$ groups, each one containing 5 different countries.

Figure 2: Perturb and Combine



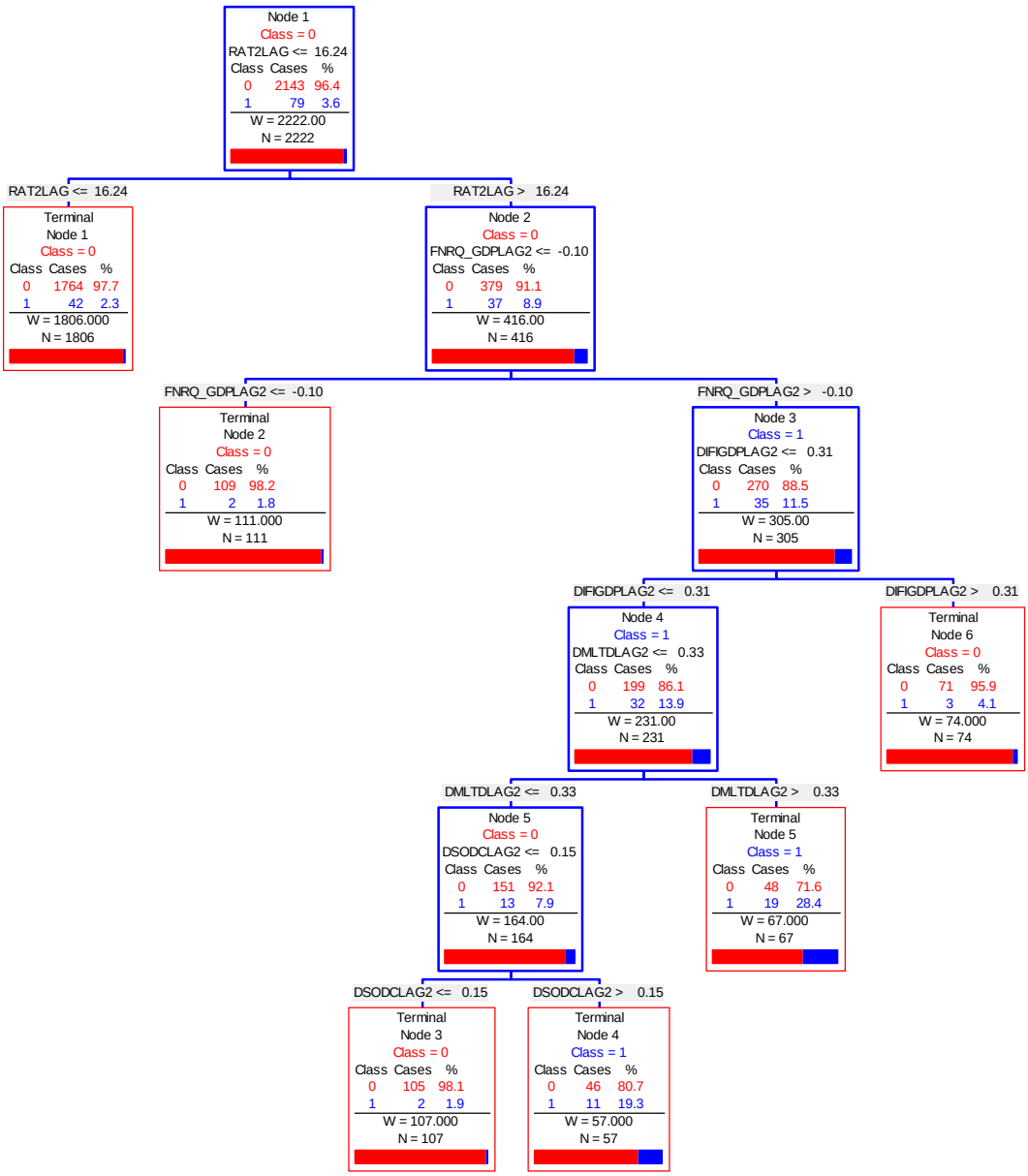
The figure depicts the CRAGGING procedure in order to obtain the predicted probability of a banking crisis through averaging the probability estimates realized by perturbing and combining the panel data.

Figure 3: Final Model



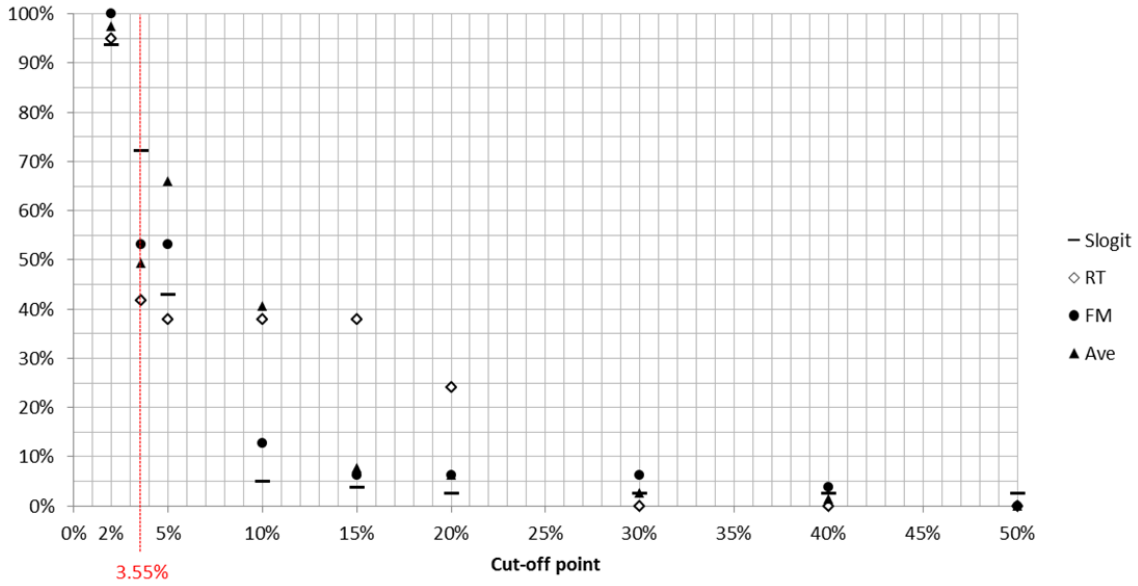
The figure depicts the Final Model (FM) structure with corresponding splitting rules computed over the period 1980-2007. STD and Avg are the standard deviation and the arithmetic average of the dependent variable within each node, N is the number of observations within each node and W is the number effectively used in computing STD and Avg, possibly using different weighting scheme (see Breiman et al., 1984).

Figure 4: CART



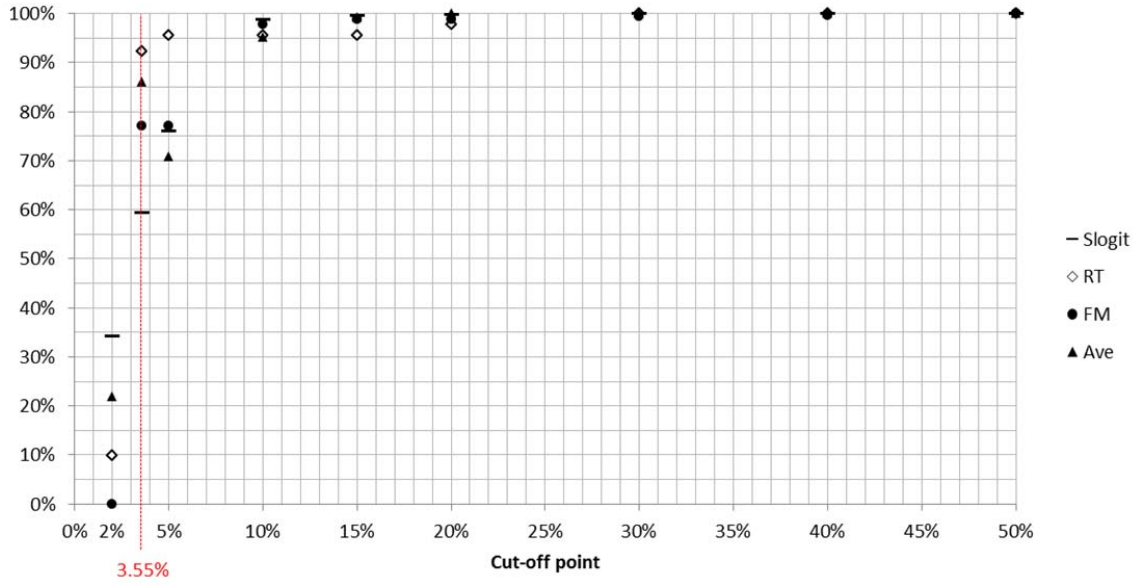
The figure depicts the Classification Tree with corresponding splitting rules computed over the period 1980-2007. Each node reports the 0-1 Classification (Class) with corresponding number and percentage of 0 (non-crisis) and 1 (crisis). N is the number of observations within each node and W is the number effectively used in computing 0-1 statistics (counting and percentage) also reported in graphical terms (the coloured rectangle at the bottom of each node shows the 0(red)-1(blue) composition).

Figure 5: In-Sample Crisis Predictions



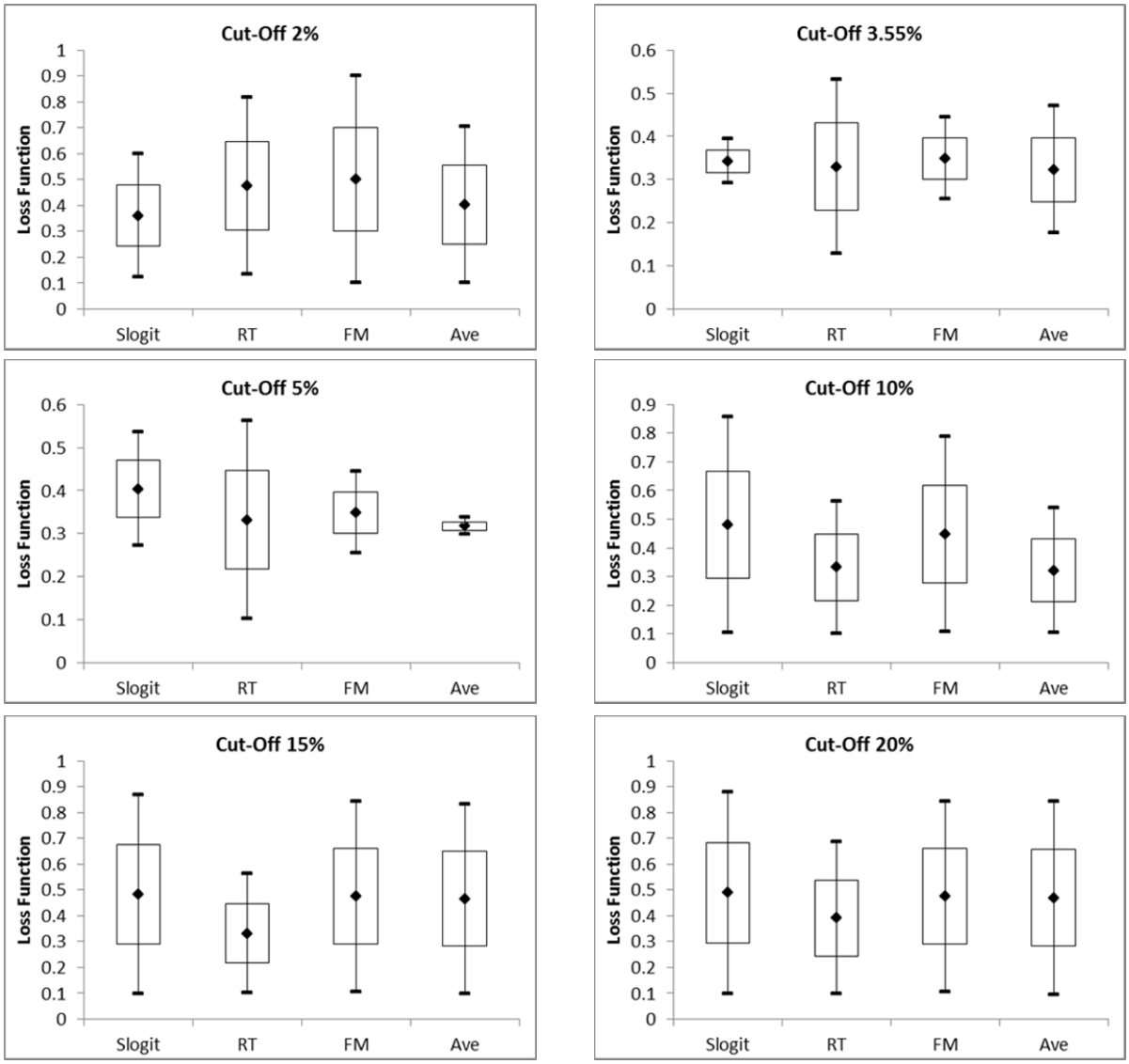
The figure reports the percentages of correcly classified crisis predictions for different cut-off points over the period 1980-2007. The value of 3.55% is the sample frequency of actual defaults in-sample and it is used as the "benchmark" cut-off point to be used as crisis/non-crisis classification.

Figure 6: In-Sample NoN-Crisis Predictions



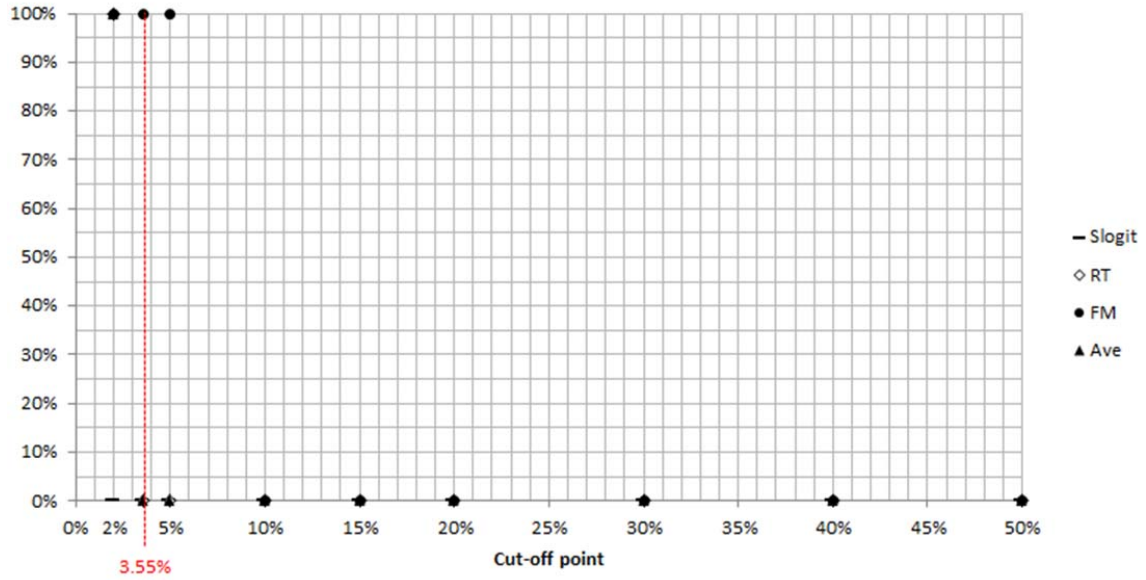
The figure reports the percentages of correcly classified non-crisis predictions for different cut-off points over the period 1980-2007. The value of 3.55% is the sample frequency of actual defaults in-sample and it is used as the "benchmark" cut-off point to be used as crisis/non-crisis classification.

Figure 7: In-Sample Loss Function Box Plots



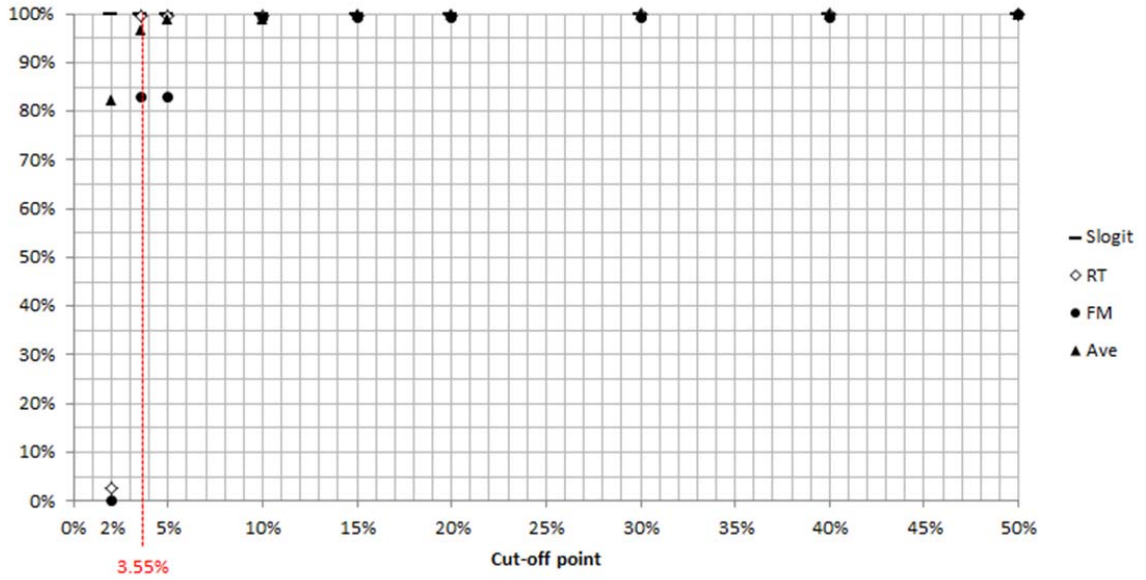
The figure reports the box plots based upon the loss function values in-sample obtained by changing the cut-off points. The bottom and top of the box are the 25th and 75th percentile, and the indicator in the middle (the rumble) of the box is the 50th percentile (the median). The minimum and the maximum of the the loss functions are indicated by the dashes.

Figure 8: Out-Of-Sample Crisis Predictions



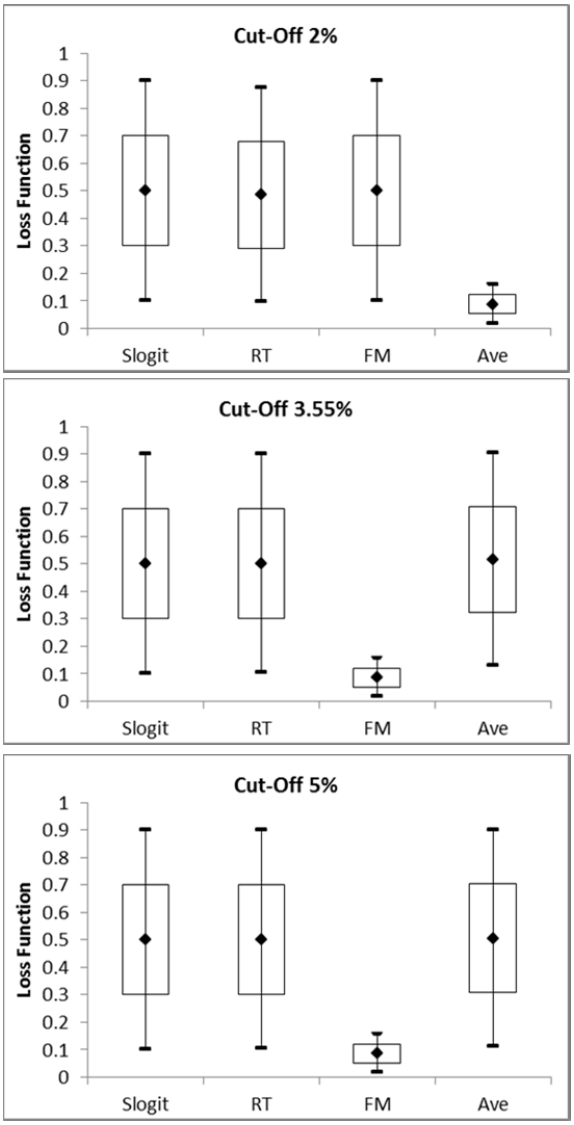
The figure reports the percentages of correcly classified crisis predictions for different cut-off points over the period 2008-2010. The value of 3.55% is the sample frequency of actual defaults in-sample and it is used as the "benchmark" cut-off point to be used as crisis/non-crisis classification out-of-sample.

Figure 9: Out-Of-Sample NoN-Crisis Predictions

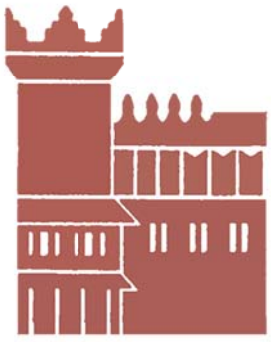


The figure reports the percentages of correcly classified non-crisis predictions for different cut-off points over the period 2008-2010. The value of 3.55% is the sample frequency of actual defaults in-sample and it is used as the "benchmark" cut-off point to be used as crisis/non-crisis classification out-of-sample.

Figure 10: Out-Of -Sample Loss Function Box Plots



The figure reports the box plots based upon the loss function values out-of-sample obtained by changing the cut-off points. The bottom and top of the box are the 25th and 75th percentile, and the indicator in the middle (the rumble) of the box is the 50th percentile (the median). The minimum and the maximum of the the loss functions are indicated by the dashes.



Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it