Journal of Banking & Finance 45 (2014) 225–241

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Predicting distress in European banks

Frank Betz [a], Silviu Oprică [b], Tuomas A. Peltonen [c], Peter Sarlin [d,e,*]

[a] European Investment Bank, Luxembourg
[b] Goethe-Universität Frankfurt, Germany
[c] European Central Bank, Germany
[d] Center of Excellence SAFE at Goethe University Frankfurt, Germany
[e] RiskLab at IAMSR, Åbo Akademi University (Turku) and Arcada University of Applied Sciences (Helsinki), Finland

ARTICLE INFO

Article history:
Available online 7 December 2013

JEL classification:
E44
E58
F01
F37
G01

Keywords:
Bank distress
Early-warning model
Prudential policy
Signal evaluation

ABSTRACT

The paper develops an early-warning model for predicting vulnerabilities leading to distress in European banks using both bank and country-level data. As outright bank failures have been rare in Europe, the paper introduces a novel dataset that complements bankruptcies and defaults with state interventions and mergers in distress. The signals of the early-warning model are calibrated not only according to the policymaker's preferences between type I and II errors, but also to take into account the potential systemic relevance of each individual financial institution. The key findings of the paper are that complementing bank-specific vulnerabilities with indicators for macro-financial imbalances and banking sector vulnerabilities improves model performance and yields useful out-of-sample predictions of bank distress during the current financial crisis.

1. Introduction

The global financial crisis has had a significant impact on the health of the European banking system and on the soundness of individual banks. Data from the European Commission shows that government assistance to stabilise the EU banking sector peaked at EUR 1.5 trl at the end of 2009, amounting to more than 13% of EU GDP. Though large, the immediate bailout costs account only for a moderate share of the total cost of a systemic banking crisis. As shown in Dell'Ariccia et al. (2008) and Laeven and Valencia (2008, 2010, 2011), among others, the output losses of previous banking crises have averaged around 20–25% of GDP. In addition, the interplay of fiscally strained sovereigns and weak banking systems that characterize the ongoing sovereign debt crisis show the crucial role of the euro area banking sector for the stability of the entire European Monetary Union. The rationale behind an early-warning model for European banks is thus clear.

To derive an early-warning model for European banks, this paper introduces a novel dataset of bank distress events. As bank defaults are rare in Europe, the dataset complements bankruptcies, liquidations and defaults by also taking into account state interventions, and mergers in distress. State interventions comprise capital injections and forms of asset relief (asset protection and guarantees). A distressed merger occurs if (i) a parent receives state aid within 12 months after the merger or (ii) if a merged entity has a coverage ratio (capital equity and loan reserves minus non-performing loans to total assets) smaller than 0 within 12 months before the merger.

The outbreak of a financial crisis is notoriously difficult to predict (e.g. Rose and Spiegel, 2011). Recently, the early-warning model literature has therefore focused on detecting underlying vulnerabilities, and finding common patterns preceding financial crises (e.g. Reinhart and Rogoff, 2008, 2009). Thus, this paper focuses on predicting vulnerable states, where one or multiple triggers could lead to a bank distress event. The early-warning model applies a micro-macro perspective to measure bank vulnerability. Beyond bank-specific and banking-sector vulnerability indicators, the paper uses measures of macroeconomic and financial imbalances from the EU Alert Mechanism Report related to the EU Macroeconomic Imbalance Procedure (MIP).

The models are estimated to derive probabilities of banks being in vulnerable states, but a policy maker needs to know when to act. Following Sarlin (2013), the signals of the model are evaluated taking into account the policymaker's preferences between type I and type II errors, the uneven frequency of tranquil and distress events, and the systemic relevance of the bank. This paper presents the

* Corresponding author at: RiskLab at IAMSR, Åbo Akademi University (Turku) and Arcada University of Applied Sciences (Helsinki), Finland. Tel.: +35 8405727670.
E-mail address: psarlin@abo.fi (P. Sarlin).

0378-4266/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jbankfin.2013.11.041

first application of the evaluation framework to a bank-level model and represents a bank's systemic relevance with its size. Thus, the early-warning model can also be calibrated to focus on predicting systemic banking failures.

The results provide useful insights into determinants of banking sector fragility in Europe. We find that complementing bank-specific vulnerabilities with indicators of macro-financial imbalances and banking sector vulnerabilities improves model performance as e.g. in González-Hermosillo (1999) and in Hernandez Tinoco and Wilson (2013). The results also confirm the usefulness of the vulnerability indicators introduced recently as part of the EU MIP as well as findings in the earlier literature. Moreover, the paper shows that an early-warning exercise using only publicly available data yields useful out-of-sample predictions of bank distress during the global financial crisis (as also e.g. Cole and White (2012) in the case of US).

Finally, the results of the evaluation framework show that a policymaker has to be substantially more concerned of missing bank distress than issuing false alarms for the model to be useful. This is intuitive if we consider that an early-warning signal triggers an internal in-depth review of fundamentals, business model and peers of the bank predicted to be in distress. Should the analysis reveal that the signal is false, there is no loss of credibility for the policy authority. The evaluations also imply that it is important to give more emphasis to systemically important and large banks for a policymaker concerned with systemic risk. At the same time, vulnerabilities and risks of large financial institutions are more complex, as the models show poorer performance when accounting for the size of the banks.

The paper is organized as follows. Section 2 provides a brief review of the related literature. Section 3 describes the data used to define bank distress events as well as the construction of the vulnerability indicators. Section 4 describes the methodological aspects of the early-warning model. Section 5 presents results on determinants of distress and predictive performance, and Section 6 discusses their robustness. Finally, Section 7 concludes the paper. Technical aspects, such as variable definitions, data sources and summary statistics, are found in Appendix A.

2. Related literature

The paper is linked to two strands of literature. First, it relates to papers predicting failure or distress at the bank level, and second, to studies on optimal early-warning signals for policymakers.

The literature on individual bank failures draws heavily on the Uniform Financial Rating System, informally known as the CAMEL ratings system, introduced by the US regulators in 1979, where the letters refer to Capital adequacy, Asset quality, Management quality, Earnings, Liquidity. Since 1996, the rating system includes also Sensitivity to Market Risk (i.e. CAMELS). The CAMELS rating system is an internal supervisory tool for evaluating the soundness of financial institutions on a uniform basis and for identifying those institutions requiring special supervisory attention or concern. Several studies find that banks' balance-sheet indicators measuring capital adequacy, asset quality, and liquidity are significant in predicting bank failures in accounting-based models (e.g. Thomson (1992) and Cole and Gunther (1995, 1998)). Other studies augment the pure accounting-based models with macroeconomic indicators and asset prices. Several papers, mainly based on US bank data, suggest that both macroeconomic and market price-based indicators contain useful predictive information not contained in the CAMELS indicators (see e.g. Flannery (1998), González-Hermosillo (1999), Jagtiani and Lemieux (2001), Curry et al. (2007), Bharath and Shumway (2008), Campbell et al. (2008) and Arena (2008)). A comprehensive survey is provided by Demyanyk and Hasan (2010), who review the empirical results obtained in several economics, finance and operations research papers that attempt to explain or predict financial crises or bank defaults.

Several studies, mainly focusing on US banks, have recently emerged to analyse bank failures during the global financial crisis. All studies reviewed report a high success in predicting US bank failures by using traditional proxies for CAMELS indicators, particularly, when complemented with some information about banks' internal controls on risk-taking (Jin et al., 2013), audit quality (Jin et al., 2011), income from nontraditional banking activities (DeYoung and Torna, 2013) or real estate investments (Cole and White, 2012). Moreover, Cole and Wu (2009) show that a simple and parsimonious probit model estimated using US data from the 1980s is highly accurate in predicting US bank failures occurring during 2009–2010. This result provides strong support for the use of simple static binary choice models in early-warning exercises. Beyond binary choice models, Jordan et al. (2010) use proxies for CAMELS and multiple discriminant analysis to predict US bank failures during the global financial crisis, while López-Iturriaga et al. (2010) use proxies of CAMELS and an artificial neural network for the same purpose. Both studies find a high degree of predictability of US bank failures during the global financial crisis.

Moreover, Beltratti and Stulz (2012) use a large sample of banks for 32 countries to examine how the stock price performance of banks during the global financial crisis relates to governance, regulation, balance sheet composition, and country characteristics other than regulation. According to their results, large banks with more Tier 1 capital, more deposits, less exposure to US real estate, and less funding fragility performed better in terms of stock prices. Banks from countries with current account surpluses fared significantly better during the crisis, while banks from countries with banking systems more exposed to the US fared worse. These latter results show that macroeconomic imbalances and the traditional asset contagion channel were related to bank performance during the crisis. Finally, the authors find no important role for bank governance nor that stronger regulation led to better performance of banks during the crisis.

As shown above, most papers analyzing individual bank failures or distress events focus on US banks or a panel of banks across countries, while there are only a few studies dealing with European banks. The data limitations arising from relatively few direct bank failures in core Europe are illustrated by some recent works: Männasoo and Mayes (2009) focus on Eastern European banks, Ötker and Podpiera (2010) create distress events using Credit Default Swaps (CDS), and Poghosyan and Cihák (2011) create events by keyword searches in news articles. All these studies suffer, however, from three respective limitations: no focus on the entire European banking system, in particular the core European countries, the use of CDS data limits the sample to banks with CDS prices, and data from news articles are inherently noisy. Using a different approach, Haq and Heaney (2012) analyse factors determining European bank risk over 1996–2010 and find evidence of a convex (U-shaped) relation between bank capital and bank systematic risk. The authors also find a positive association between off-balance sheet activities and bank risk.

Finally, the literature on country-level banking crises is broad and has most often focused on continents, if not pursuing a fully global approach. Demirgüç-Kunt and Detragiache (2000), Davis and Karim (2008a,b) and Sun (2011) analyse banking crises with a global country coverage, whereas Hutchinson (2002) and Mody and Sandri (2012) focus on European countries, where the latter study concerns the recent crisis.

Turning to the second strand of literature to which this paper is related, namely calibration and evaluation of model signals, Kaminsky et al. (1998) introduce the so-called "signal" approach to evaluate the early-warning properties of univariate indicator signals when they exceed a predefined threshold. The threshold is set to minimize the noise-to-signal ratio, given by the number of false alarms relative to the correct calls. Many later studies, such as Berg and Pattillo (1999a) and Edison (2003), while introducing a discrete-choice model, do not adopt a structured approach to evaluate model performance. An issue addressed by Demirgüç-Kunt and Detragiache (2000) is the introduction of a loss function of a policymaker that considers costs for preventive action and relative preferences between missing crises (type I errors) and false alarms (type II errors). The authors also show that optimising model thresholds on the basis of the noise-to-signal ratio can lead to sub-optimal results under some preference schemes.[1]

Alessi and Detken (2011) apply the loss function of a policymaker in a univariate signal approach to asset price boom/bust cycles and extend it by introducing a measure that accounts for the loss of disregarding the signals of a model. LoDuca and Peltonen (2013) apply the evaluation framework of Alessi and Detken (2011) in a multivariate logit model, while Sarlin (2013) further extends it by amending the policymaker's loss function and usefulness measure to include unconditional probabilities of the events and also computes a measure called relative Usefulness. By computing the percentage share of available Usefulness that a model captures, the relative Usefulness facilitates interpretation of the measure. Sarlin (2013) also adapts the Usefulness measures to account for observation-specific weights. This paper presents the first application of the multivariate evaluation framework based on policymaker's loss function to a bank-level model, taking into account bank-specific systemic relevance (here proxied with a bank's size).

3. Data

We construct the sample based on the availability of balance-sheet and income-statement data in Bloomberg. The observation period starts in 2000Q1 and ends in 2013Q2. We obtain data on 546 banks with a minimum of EUR 1bn in total assets during the period under consideration (in total 28,832 observations). We therefore focus on large banks with significance for systemic instability. The sample covers banks in all EU countries but Cyprus, Estonia, Lithuania and Romania. We seek to reconstruct the information set that would have been available to investors at each point in time. Thus, for instance, if a bank reports its accounts at annual frequency, we use this information in four subsequent quarters. Likewise, publication lags of data are taken into account to the extent possible. To reconstruct the information set at each reference period, all bank balance-sheet and house price indicators are lagged by 2 quarters, the structural MIP variables from the EU Alert Mechanism Report are lagged by 6 quarters, whereas other macro-financial variables and banking sector indicators are lagged by 1 quarter. The dataset consists of two parts, bank distress events and vulnerability indicators, which are described next.

3.1. Distress events

Given that actual bank failures are rare in Europe, identification of bank distress events is challenging. Thus, in addition to bankruptcies, liquidations, and defaults, the paper also takes into account state interventions and forced mergers to represent bank distress.

First, we use data on bankruptcies, liquidations and defaults to capture direct bank failures. A bankruptcy is defined to occur if the net worth of a bank falls below the country-specific guidelines, whereas liquidations occur if a bank is sold as per the guidelines of the liquidator in which case the shareholders may not receive full payment for their ownership. We define two types of defaults as follows: a default occurs (i) if a bank has failed to pay interest or principal on at least one financial obligation beyond any grace period specified by the terms or (ii) if a bank completes a distressed exchange, in which at least one financial obligation is repurchased or replaced by other instruments with a diminished total value. The data on bankruptcies and liquidations are retrieved from Bankscope, while defaults are obtained from annual compendiums of corporate defaults by Moody's and Fitch. We define a distress event to start when the failure is announced and to end when the failure de facto occurs. This method leads to 13 distress events at the bank-quarter level, of which most are defaults.

Second, we use data on state support to detect banks in distress. A bank is defined to be in distress if it receives a capital injection by the state or participates in asset relief programmes (asset protection or asset guarantees).[2] This definition focuses on assistance on the asset side and hence does not include liquidity support or guarantees on banks' liabilities. The state aid measures are based on data from the European Commission as well as data collected by the authors from market sources (Reuters and Bloomberg). Events in this category are defined to last from the announcement of the state support to the execution of the state support programme. This approach leads to 153 distress events, which shows the extent to which state interventions are more common than outright defaults.

Third, mergers in distress capture private sector solutions to bank distress. The merged entities are defined to be in distress if (i) a parent receives state aid within 12 months after the merger or (ii) if a merged entity has a coverage ratio smaller than 0 within 12 months before the merger. The coverage ratio is commonly used in the literature to define distressed banks (e.g. González-Hermosillo, 1999). The rationale for applying the rule only on mergers is that we want to capture banks that are forced to merge due to distress. A bank may have a negative coverage ratio, but still survive without external support. Data on mergers are obtained from Bankscope, whereas the coverage ratio is defined as the ratio of capital equity and loan reserves minus non-performing loans to total assets and computed using data from Bloomberg. While these definitions should thoroughly cover distressed mergers, a caveat is a possible mismatch in the sample coverage of the two data sources. The events identified using these definitions of distressed mergers are, however, also cross-checked using market sources (Reuters and Bloomberg). We define the two types of distressed mergers to start and end as follows: (i) to start when the merger occurs and end when the parent bank receives state aid and (ii) to start when the coverage ratio falls below 0 (within 12 months before the merger) and end when the merger occurs. Based on this approach, we identify 35 mergers in distress.

In total, we obtain 194 distress events at the bank-quarter level. This figure is smaller than the sum of events across the categories as they are not mutually exclusive. As a bank that experiences two distress events within one year is likely to be in distress also in between those events, we modify the bank-specific time series accordingly. While potentially being a question of interest, we do not distinguish between the different types of distress events in this paper as does e.g. Kick and Koetter (2007). The low frequency of direct failures and distressed mergers hinders robust estimations of determinants for all three distress categories.

Fig. 1 shows the number of banks and distress events by country. Given the chosen sample and data availability, Italy is the country with the largest number of banks, followed by Spain,

[1] If banking crises are rare events and the cost of missing a crisis is high relative to that of issuing a false alarm, minimising the noise-to-signal ratio could lead to many missed crises. As a consequence, the selected threshold could be sub-optimal from the point of view of the preferences of policymakers.

[2] See Stolz and Wedow (2010) for a comprehensive overview of state support measures for the financial sector in the EU and the US.

Germany, and France. In the case of Greece, Ireland, and Belgium, the number of distress events exceeds the number of banks, which is feasible as a bank can experience multiple distress periods. This paper focuses on vulnerable states, or pre-distress events, which can be defined from the dates of the distress events. In our benchmark case, a binary pre-distress variable is defined to take the value 1 in 8 quarters prior to the earlier defined distress events, and otherwise 0.

The number of the distress and pre-distress events per category is illustrated in detail in Table 1. As mentioned earlier, the occurrence of the distress and pre-distress events in various categories are not mutually exclusive. Hence, the categories do not sum up. The table illustrates that most distress events, and thus also pre-distress periods, are state interventions and a large share of them is capital injections. The unconditional probabilities of the events show that distress events represent only a small share (less than 1%) of the observations in the dataset. This imbalance in class size will be taken into account in the model evaluation framework.

3.2. Vulnerability indicators

The paper uses three categories of indicators in order to capture various aspects of a bank's vulnerability to distress. First, indicators from banks' income statements and balance sheets measure bank-specific vulnerabilities. Following the literature, we use indicators to account for all dimensions in the CAMELS rating system (e.g. Flannery, 1998; González-Hermosillo, 1999; Poghosyan and Cihák, 2011). The indicators to proxy the CAMELS dimensions are as follows. The equity-to-assets ratio (capital ratio) and Tier 1 capital ratio represent Capital adequacy (C) and are used to proxy the level of bank capitalization. In both cases, higher level of capital acts as a buffer against financial losses protecting a bank's solvency and is expected to reduce the probability of a bank failure.

Asset quality (A) is represented by return on assets (ROA), the share of non-performing assets to total assets, reserves for loan losses as a share of non-performing assets, and the share of loan loss provisions to total average loans. Overall, weaker asset quality is expected to be positively associated with bank distress. In both cases, the higher share of non-performing assets to total assets and the higher share of loan loss provisions to total average loans are expected to increase the probability of failure. However, the effect of reserves for loan losses as a share of non-performing assets is potentially ambiguous. Whereas higher reserves should correspond to a higher cover for expected losses, they could also proxy for higher expected losses.

The cost-to-income ratio represents Management quality (M), which is expected to reduce the probability of bank failure. Similarly, both indicators measuring Earnings (E), return on equity (ROE) and net interest margin are expected to be negatively associated with bank distress. Liquidity (L) is represented by the share of interest expenses to total liabilities, the deposits-to-funding ratio and the ratio of net short-term borrowing to total liabilities. Given that deposits are usually considered as a more stable funding source than the interbank market or securities funding, a higher deposits-to-funding ratio is expected to be negatively associated with bank distress. On the other hand, a higher share of interest expenses to total liabilities and the higher ratio of net short-term borrowing to total liabilities are both expected to be positively associated with a bank failure.

The share of trading income proxies for Sensitivity to market risk (S). Again, the relation of this variable with respect to bank distress is ambiguous. On the one hand, higher trading income could be associated with a riskier business model as trading income is a volatile source of earnings. On the other hand, investment securities are more liquid than e.g. loans, and thus allow a bank to minimize fire sale losses in case of a changing macro-financial environment. Thus, the expected sign could also be negative as in Cole and Gunther (1998). All bank-level indicators are constructed using Bloomberg data.

Finally, in contrast to studies like Agarwal and Taffler (2008), we do not consider market-based indicators for the following two reasons. First, we aim at predicting underlying vulnerabilities 1–3 years prior to distress, whereas market-based signals tend to have a shorter horizon (see e.g. Bongini et al., 2002; Milne, forthcoming); and second, we aim at using a broad sample of banks, rather than only listed banks.

Second, country-specific banking sector indicators proxy for imbalances at the level of banking systems. These indicators are often cited as key early-warning indicators for banking crises (e.g. Demirgüç-Kunt and Detragiache, 1998, 2000; Kaminsky and Reinhart, 1999; Borio and Lowe, 2002; Hahm et al., 2013). The indicators proxy the following types of imbalances: booms and rapid increases in banks' balance sheets are proxied by total assets to GDP and growth in non-core liabilities; banking-system leverage by debt-to-equity and loans-to-deposits ratios; securitization by debt securities to liabilities; and property booms by the ratio of mortgages to loans. All indicators are constructed using the ECB's statistics on the Balance Sheet Items (BSI) of the Monetary, Financial Institutions and Markets (MFI).

Finally, country-specific macro-financial indicators identify macro-economic imbalances and control for conjunctural variation in asset prices and business cycles. To control for macro-economic imbalances, the paper uses selected internal and external indicators from the EU Macroeconomic Imbalance Procedure (MIP), such as private sector credit flow, government debt, and international investment position (EC, 2012). Moreover, asset prices (stock and house prices) and business cycle indicators (real GDP growth and CPI inflation) capture conjunctural variation. The macro-financial indicators are retrieved from Eurostat and Bloomberg with the exception of the house price indicators that are from the ECB.

Table A in Appendix A describes the indicators used, their definitions and transformations, while Table B provides their

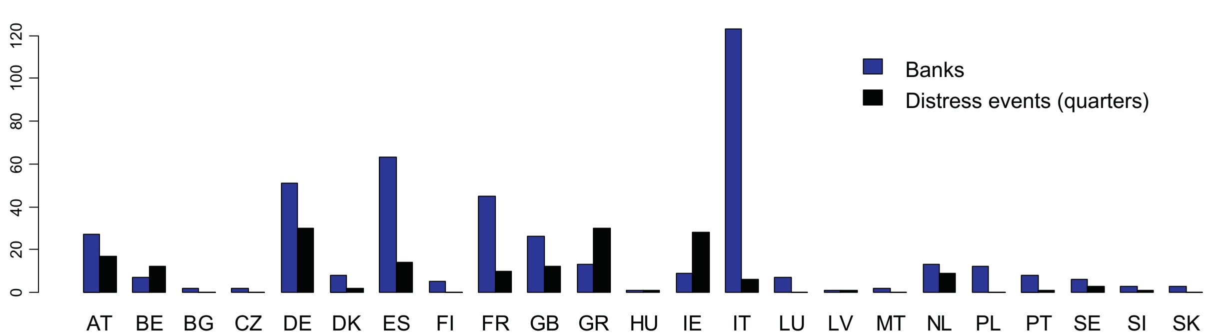

Fig. 1. The number of banks and distress events by country.

Table 1
The number of distress and pre-distress events by category.

Distress categories	Distress events uncond.		Pre-distress events uncond.	
	Freq.	Prob. (%)	Freq.	Prob. (%)
Direct failure	13	0.05	110	0.41
Bankruptcy	1	0.00	8	0.03
Liquidation	2	0.01	16	0.06
Defaulted by Moody's	11	0.04	75	0.28
Defaulted by Fitch	2	0.01	21	0.08
Distressed mergers	35	0.13	228	0.85
Merger with state intervention	28	0.10	179	0.67
Merger with coverage ratio <0	13	0.05	105	0.39
State intervention	153	0.57	892	3.32
Capital injection	113	0.42	763	2.84
Asset protection	33	0.12	180	0.67
Asset guarantee	23	0.09	127	0.47
Total	194	0.72	1000	3.72

Note: The statistics are derived from the entire sample with 28,832 observations and 546 banks and the pre-distress period is defined to start 8 quarters prior to the distress events.

summary statistics. Statistical tests applied show that the data are non-normally distributed and exhibit most often a positive skew with a leptokurtic distribution. Table C in Appendix A shows the discriminatory power of the indicators between tranquil ($C = 0$) and pre-distress events ($C = 1$) through mean-comparison tests. The t-test results indicate that most variables are good candidates for discriminating between tranquil and vulnerable periods. Among bank-specific indicators, the cost-to-income ratio, deposits-to-funding ratio, net interest margin, and the share of trading income do not hold promise to discriminate between the classes. The ratio of loans to deposit is the poorest discriminator among banking-sector indicators, whereas CPI inflation is the poorest among macro-financial indicators.

4. Methodology

The methodology presented in this section consists of two building blocks. First, a framework for evaluating signals of early-warning models, and second, the estimation and prediction methods.

4.1. Evaluation of model signals

Early-warning models require evaluation criteria that account for the nature of the underlying problem. Distress events are oftentimes outliers in three regards: the dynamics of the entity differ significantly from tranquil times, they are often costly, and they occur rarely. Given these properties, an evaluation framework that resembles the decision problem faced by a policymaker is of central importance. Designing a comprehensive evaluation framework for early-warning model signals is challenging as there are several political economy aspects to be taken into account. For instance, the frequency and optimal timing when the policymaker should signal a crisis might depend on potential inconsistencies between the maximisation of the policymaker's own utility vs. social welfare. While important, these types of considerations are beyond the scope of this study. Therefore, the signal evaluation framework focuses only on a policymaker with a relative preference between type I and II errors and the usefulness that she gets by using a model vs. not using it. Thus, it is implicitly assumed that the policymaker internalises the expected costs of a banking crisis and a false alarm into her preferences between type I and II errors.

As the focus is on detecting vulnerabilities and risks prior to distress, the ideal leading indicator can be represented by a binary state variable $C_j(h) \in \{0, 1\}$ for observation j, where $j = 1, 2, \ldots, N$ with a specified forecast horizon h. Let $C_j(h)$ be a binary indicator that is one during pre-crisis periods and zero otherwise. For detecting events C_j using information from indicators, discrete-choice models can be used for estimating crisis probabilities $p_j \in [0, 1]$. To mimic the ideal leading indicator, the probability p is transformed into a binary prediction P_j that is one if p_j exceeds a specified threshold $\lambda \in [0, 1]$ and zero otherwise. The correspondence between the prediction P_j and the ideal leading indicator C_j can be summarized by a so-called contingency matrix (see Table 2).

While the elements of the matrix (frequencies of prediction-realization combinations) can be used for computing a wide range of measures,[3] a policymaker can be thought of mainly being concerned about two types of errors: giving false alarms and missing crises. The evaluation framework in this paper follows Sarlin (2013) for turning policymakers' preferences into a loss function, where the policymaker has relative preferences between type I and II errors.[4] Type I errors represent the proportion of missed crises relative to the number of crises in the sample $T_1 \in [0, 1] = \frac{FN}{TP+FN}$, and type II errors the proportion of false alarms relative to the number of tranquil periods in the sample $T_2 \in [0, 1] = \frac{FP}{FP+TN}$. Given probabilities p of a model, the policymaker should choose a threshold λ such that her loss is minimized. The loss of a policymaker consists of T_1 and T_2, weighted according to her relative preferences between missing crises μ and giving false alarms $1 - \mu$. By accounting for unconditional probabilities of crises $P_1 = P(D = 1)$ and tranquil periods $P_2 = P(D = 0) = 1 - P_1$, the loss function is as follows:

$$L(\mu) = \mu T_1 P_1 + (1 - \mu) T_2 P_2, \tag{1}$$

where $\mu \in [0, 1]$ represents the relative preferences of missing events, $1 - \mu$ the relative preferences of giving false alarms, T_1 the type I errors and T_2 the type II errors. P_1 refers to the relative size of the crisis class and P_2 to the relative size of the tranquil class. Using the loss function $L(\mu)$, the Usefulness of a model can be defined in two ways. First, the absolute Usefulness U_a is given by:

$$U_a = \min(\mu P_1, (1 - \mu) P_2) - L(\mu), \tag{2}$$

which computes the extent to which a model performs better than no model at all. As the unconditional probabilities are commonly

[3] Some of the commonly used simple evaluation measures are as follows. Recall positives (or TP rate) = TP/(TP + FN), Recall negatives (or TN rate) = TN/(TN + FP), Precision positives = TP/(TP + FP), Precision negatives = TN/(TN + FN), Accuracy = (TP + TN)/(TP + TN + FP + FN), FP rate = FP/(FP + TN), and FN rate = FN/(FN + TP).

[4] In the literature of bank early-warning models, Cole and Gunther (1998), for instance, assessed their model performance by graphically plotting type I and II errors.

Table 2
Contingency matrix.

	Actual Class C_j	
	1	0
Predicted Class P_j		
1	True positive (TP)	False positive (FP)
0	False negative (FN)	True negative (TN)

imbalanced and the policymaker may be more concerned about one class, a policymaker could achieve a loss of $\min(\mu P_1, (1-\mu)P_2)$ by either always or never signalling an event. It is thus worth noting that already an attempt to build an early-warning model for events with imbalanced events implicitly assumes a policymaker to be more concerned about the rare class. With a non-perfectly performing model, it would otherwise easily pay-off for the policymaker to always signal the high-frequency or highly preferred class. Second, relative Usefulness U_r is computed as follows:

$$U_r = \frac{U_a}{\min(\mu P_1, (1-\mu)P_2)}, \tag{3}$$

where the absolute Usefulness U_a of the model is compared with the maximum possible usefulness of the model. That is, U_r reports U_a as a percentage of the usefulness that a policymaker would gain with a perfectly performing model.

A policymaker may further want to enhance the representation of preferences by accounting for observation-specific differences in costs. In bank early-warning models, the bank-specific misclassification costs are highly related to the systemic relevance of an entity for the policymaker. While this relevance can be measured with network measures such as centrality, a simplified measure of relevance for the system in general is the size of the entity (e.g. assets of a financial institution) relative to the financial system's size. Let w_j be a bank-specific weight that approximates the importance of correctly classifying observation j. In addition, let TP_j, FP_j, FN_j, and TN_j be binary vectors of combinations of predictions and realizations rather than only their sums. By multiplying each binary element of the contingency matrix by w_j, we can derive a policymaker's loss function with bank and class-specific misclassification costs. Let T_1 and T_2 be weighted by w_j to have weighted type I and II errors:

$$T_{w_1} \in [0,1] = \frac{\sum_{j=1}^{N} w_j FN_j}{\sum_{j=1}^{N} w_j TP_j + \sum_{j=1}^{N} w_j FN_j}, \tag{4}$$

$$T_{w_2} \in [0,1] = \frac{\sum_{j=1}^{N} w_j FP_j}{\sum_{j=1}^{N} w_j FP_j + \sum_{j=1}^{N} w_j TN_j}. \tag{5}$$

As T_{w_1} and T_{w_2} are ratios of weights rather than ratios of binary values, the errors T_{w_1} and T_{w_2} can replace T_1 and T_2 in Eqs. (1)–(3), and thus weighted counterparts of the loss function $L(\mu, w_j)$, and absolute and relative usefulness $U_a(\mu, w_j)$ and $U_r(\mu, w_j)$ for given preferences can be derived.

Receiver operating characteristics (ROC) curves and the area under the ROC curve (AUC) are also viable measures for comparing performance of early-warning models. The ROC curve shows the trade-off between the benefits and costs of a certain λ. When two models are compared, the better model has a higher benefit (TP rate on the vertical axis) at the same cost (FP rate on the horizontal axis).[5] Thus, as each FP rate is associated with a threshold, the measure shows performance over all thresholds. In this paper, the size of the AUC is computed using trapezoidal approximations. The AUC measures the probability that a randomly chosen distress event is ranked higher than a tranquil period. A perfect ranking has an AUC equal to 1, whereas a coin toss has an expected AUC of 0.5.

4.2. Estimation and prediction

The early-warning model literature has utilized a wide range of conventional statistical methods for estimating distress probabilities. The obvious problem with most statistical methods (e.g. discriminant analysis and discrete-choice models) is that all assumptions on data properties are seldom met. By contrast, the signals approach is univariate in nature. We turn to discrete-choice models, as methods from the generalized linear model family have less restrictive assumptions (e.g. normality of the indicators). Logit analysis is preferred over probit analysis as its assumption of more fat-tailed error distribution corresponds better to the frequency of banking crises and bank distress events (van den Berg et al., 2008). Hazard models would hold promise for these inherently problematic data by not having assumptions about distributional properties, such as shown in Whalen (1991) in a banking context. However, the focus of hazard models is on predicting the timing of distress, whereas we aim at predicting vulnerable states, where one or multiple triggers could lead to a bank distress event.

Typically, the literature has preferred the choice of a pooled logit model (e.g. Fuertes and Kalotychou, 2007; Kumar et al., 2003; Davis and Karim, 2008b; LoDuca and Peltonen, 2013; Sarlin and Peltonen, 2013). Fuertes and Kalotychou (2006) show that accounting for time- and country-specific effects leads to better in-sample fit, while decreasing the predictive performance on out-of-sample data. Further motivations of pooling the data vs. using panel methods are the relatively small number of crises in individual countries and the strive to capture a wide variety of vulnerable states. Country-specific effects are, to some extent, still taken into account as country-level explanatory variables are included in the model. Rather than using lagged explanatory variables, the dependent variable is defined as a specified number of quarters prior to the event (8 quarters in the benchmark case). The early-warning model is a recursive logit model that makes a prediction at each quarter $t = 1, 2, \ldots, T$ with an estimation sample that grows in an increasing-window fashion and functions according to the following steps:

1. Estimate the model on in-sample data using the information that would have been available from the beginning of the sample up to quarter $t-1$ (in-sample period).
2. Collect the probabilities p of the model for the in-sample period and compute the Usefulness for all thresholds $\lambda \in [0, 1]$.
3. Choose the λ that maximizes in-sample Usefulness, estimate distress probabilities p for the out-of-sample data (quarter t), apply λ to the out-of-sample data and collect the results.
4. Set $t = t + 1$ and recursively re-estimate the model starting from Step 1 at each quarter t, while $t \leqslant T$.

In practice, we estimate a model at each quarter t with all available information up to that point, evaluate the signals to set an optimal threshold, and provide an estimate of the current vulnerability of each bank with the same threshold as on in-sample data. The algorithm is based on recursive increasing windows for the in-sample period and rolling windows (one quarter at a time) for the out-of-sample period. These recursive changes in in-sample and out-of-sample data enable testing the performance of the model in real-time use.

The estimation strategy accounts for post-crisis and crisis bias, as proposed by Bussière and Fratzscher (2006), by not including periods when a bank distress event occurs or the 4 quarters thereafter. However, post-distress periods are included in the sample if

[5] In general, the ROC curve plots, for the whole range of measures, the conditional probability of positives to the conditional probability of negatives: $ROC = \frac{P(P=1|C=1)}{1-P(P=0|C=0)}$.

they are also pre-distress periods. The excluded observations are not informative regarding the transition from tranquil times to distress events, as they can neither be considered "normal" periods nor vulnerabilities prior to distress. While the above recursive estimation includes only the Usefulness measure for optimizing the models, all measures introduced in Section 4.1 are computed for evaluating model performance.

5. Results

This section presents the results, focusing on two key issues: what are the main sources of bank vulnerabilities and to what extent do indicators, or groups of them, predict bank vulnerabilities. Table 3 presents the estimates of the benchmark logit model, which aims at predicting bank vulnerability 8 quarters ahead of distress. The coefficients refer to the full estimation sample (2000Q1–2011Q2). The ending date depends on the availability of full information on bank vulnerabilities. That is, with a forecast horizon of 8 quarters, the binary pre-crisis indicator $C_j(h)$ can only be created up to two years prior to the current date (i.e. 2013Q2). The predictions use recursive increasing windows for the in-sample data (2000Q1–2011Q2), starting with data until 2007Q1, and rolling windows for the out-of-sample data (2007Q1–2011Q2).

The benchmark model (Model 1) contains vulnerability indicators that are drawn from the three groups introduced in Section 3: bank-level indicators, country-specific banking sector indicators and country-level macro-financial indicators. The model is chosen based on two considerations. On the one hand, the model should be encompassing and contain a wide-range of potential vulnerabilities. On the other hand, bank-specific items that have a comparatively short history in available data sources limit the number of observations. Model 2 (Benchmark+) in Table 3 illustrates the trade-off between the number of observations and the number of indicators. For instance, including additional variables, such as the Tier 1 capital ratio, impaired assets and net interest margin, reduces not only the number of available banks from 298 to 238 and the observations from 8340 to 6088, but especially the beginning of the sample, which hinders early prediction of the crisis. More importantly, it does not improve the predictive usefulness of the model.

Table 3 presents the estimated coefficients for the benchmark model. Among bank-specific indicators, a high capital ratio (total equity to total assets) is estimated to lower the probability of bank distress. Concerning asset quality, a high return on assets (ROA) as well as high loan loss provisions increase the probability of distress, while reserves to impaired assets is not statistically significant at the 10% level. Also the cost-to-income ratio, a proxy for bank management efficiency, is not statistically significant at the 10% level. Regarding measures of profitability, high interest expenses are estimated to increase a bank's vulnerability, while the coefficient for return on equity is not statistically significant at the 10% level. Among liquidity indicators, a high dependency on short-term borrowing is estimated to increase the probability of bank distress, while deposits to funding ratio is not statistically significant at the 10% level. Finally, the coefficient for the share of trading income, a proxy for sensitivity to market risk, is not statistically significant at the 10% level.

With two exceptions, most statistically significant coefficients have a sign consistent with the discussion in Section 3.2 and the earlier literature. However, while the literature commonly finds a negative sign for return on assets (to proxy for profitability), the estimated positive association can potentially be explained by risks in the bank's business model that are not represented by high leverage, which the specifications control for. Second, the negative sign for the share of trading income is also somewhat counterintuitive as trading income is a volatile source of income for the bank and could also be related to a riskier business model of a bank. However, as in Cole and Gunther (1998), investment securities are more liquid than e.g. loans, and thus allow bank to minimize fire sale losses in case of a changing macro-financial environment. Thus, the expected sign could also be negative.

Among the country-level banking-sector indicators, almost all are estimated to be statistically significant. In particular, rapid growth in non-core liabilities, a high debt-to-equity ratio, as well as a large banking sector, measured as total assets to GDP, are associated with higher probabilities of bank distress. This is in line with e.g. Hahm et al. (2013), who find that a lending boom can be detected from the composition of bank liabilities when traditional retail deposits (core liabilities) cannot keep pace with asset growth and banks turn to other funding sources (non-core liabilities) to finance their lending. In contrast, the ratio of debt securities to liabilities, a measure of securitization, is estimated to decrease bank vulnerability. Both the share of mortgages among loans, a proxy for property booms, as well as the ratio of loans-to-deposits are not statistically significant at the 10% level.

Among the country-specific macro-financial indicators, all statistically significant coefficients have the expected sign. Starting with conjunctural variables, low real GDP growth increases bank vulnerability, while CPI inflation is not statistically significant at 10% level. Regarding asset prices, decreasing house prices are positively associated with bank distress, while stock prices are not statistically significant at the 10% level. For indicators of internal imbalances, the estimated coefficients for government debt and private sector credit flow are positive and thus positively correlated with bank distress. Regarding external competitiveness, a high international investment position is estimated to decrease bank vulnerability. Finally, long-term government bond yields are not statistically significant at 10% level. Thus, our results are overall in line with the literature in that combining bank-level accounting, market, and macroeconomic data improves model performance (see e.g. González-Hermosillo (1999), Hernandez Tinoco and Wilson (2013) and references in the literature review in Section 2).

Table 3 also evaluates the predictive performance of the models based upon the recursive estimation procedure presented in Section 4.2 for each quarter in 2007Q1–2011Q2 (out-of-sample) conditional on the policymaker's preference parameter μ. Given that the threshold λ for classifying signals is a time-varying parameter that is chosen to optimize in-sample usefulness at each t, the table does not report the applied λ. As discussed above, we assume that the policymaker is substantially more interested in correctly calling bank distress events than tranquil periods. This is intuitive if we consider that an early-warning signal triggers an internal in-depth review of fundamentals, business model and peers of the bank predicted to be in distress. Should the analysis reveal that the signal is false, there is no loss of credibility for the policy authority. Hence, in the benchmark case, preferences are set to $\mu = 0.9$. Table 3 reports both the absolute and the relative Usefulness measures as well as the unconditional probability of pre-distress events (0.06 and 0.07). The benchmark model's absolute Usefulness U_a equals 0.04, which translates into a relative Usefulness U_r equal to 42%, in contrast to 40% for Model 2, which includes more bank-specific indicators.

Table 4 provides information on the predictive power of the three indicator groups. Conditional on a preference parameter $\mu = 0.9$, Model 4 based on macro-financial indicators outperforms the other models by achieving a U_r of 22%. Model 2, which includes only bank-specific indicators, achieves a U_r of 19% compared to 21% for the banking-sector in Model 3. It is an interesting finding

Table 3
Logit estimates on bank distress and their predictive performance.

Estimates	(1) Benchmark		(2) Benchmark +	
Bank-specific indicators				
C[a]				
Intercept	−3.46***		−3.26***	
Capital ratio	−0.76***		−1.37***	
Tier 1 ratio			−5.91	
Impaired assets			0.14$^{\cdot}$	
A[a]				
Reserves to impaired assets	−0.19		−0.15	
ROA	0.12*		0.56***	
Loan loss provisions	0.09$^{\cdot}$		0.18$^{\cdot}$	
M[a]				
Cost to income	0.09		0.22*	
ROE	−0.06		−0.28	
E[a]				
Net interest margin			0.23	
Interest expenses to liabilities	0.14***		0.50**	
L[a]				
Deposits to funding	0.01		−0.33**	
Net-short term borrowing	0.18**		0.48	
S[a]				
Share of trading income	−0.14		−0.27$^{\cdot}$	
Country-specific banking sector indicates				
Total assets to GDP	0.71***		1.71***	
Non-core liabilities	0.32***		0.28***	
Debt to equity	0.30***		0.37***	
Loans to deposits	0.14		0.05	
Debt securities to liabilities	−0.22*		−0.19*	
Mortgages to loans	0.03		0.21*	
Country-specific macro-financial indicators				
Real GDP	−0.10$^{\cdot}$		−0.06	
Inflation	0.06		0.15**	
Stock prices	0.02		−0.05	
House prices	−0.38***		−0.28**	
Long-term government bond yield	0.04		0.12	
International investment position to GDP	−0.50***		−0.46***	
Government debt to GDP	0.50***		0.43***	
Private sector credit flow to GDP	0.36***		0.23***	
R2[b]	0.27		0.24	
No. of banks	298		238	
No. of observations	8340		6088	
Predictive performance	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)
Usefulness for a policy maker[c]				
$\mu = 0.6$	0.01	12	0.01	14
$\mu = 0.7$	0.01	16	0.02	22
$\mu = 0.8$	0.03	31	0.03	29
$\boldsymbol{\mu = 0.9}$	**0.04**	**42**	**0.03**	**40**
$P(I_j(h) = 1)$[d]	0.06		0.07	

Note: For standardized coefficients, the explanatory variables have been transformed to have zero mean and unit variance. Bold entries correspond to the benchmark preferences.

[a] The letters of CAMELS refer to Capital adequacy, Asset quality, Management, Earnings, Liquidity and Sensitivity to market risk.

[b] *R2* refers to the Nagelkerke's pseudo *R*-squared.

[c] The Usefulness for a policy maker is computed with absolute and relative usefulness $U_a(\mu)$ and $U_r(\mu)$ as described in Section 4.1.

[d] $P(I\,j(h) = 1)$ refers to the unconditional probability of pre-distress events.

$^{\cdot}$ Signif. codes: 0.10.

* Signif. codes: 0.05.

** Signif. codes: 0.01.

*** Signif. codes: 0.001.

that country-level indicators turn out to be more useful for predicting vulnerabilities at the bank level than bank-specific indicators. However, the latest crisis clearly demonstrated the importance of country-specific factors driving bank failures. While the macro-financial indicators consist of those featured in the MIP, they follow the earlier literature on country-level imbalances and crises (e.g. Demirgüç-Kunt and Detragiache, 1998; Kaminsky and Reinhart, 1999; Borio and Lowe, 2002).

Models 5–6 confirm the benefit of combining bank-level data with country-level banking sector and macro-financial indicators.

Table 4
Logit estimates on bank distress and their predictive performance – models using different sets of indicators.

Estimates	(1) Benchmark		(2) BS Model		(3) BSI Model		(4) MF Model		(5) BS & BSI Model		(6) BS & MF Model	
Bank-specific indicators												
C[a]												
Intercept	−3.46***		−3.05***		−3.02***		−3.06***		−2.91***		−3.47***	
Capital ratio	−0.76***		−2.20***						−1.40***		−2.21***	
Reserves to impaired assets	−0.19		−0.15						−0.15		−0.15	
A[a]												
ROA	0.12*		0.28·						0.16		0.26·	
Loan loss provisions	0.09·		0.36						0.22**		0.21*	
M[a]												
Cost to income	0.09		0.18*						0.26**		0.14	
E[a]												
ROE	−0.06		−2.29*						−1.84*		−1.67***	
Interest expenses to liabilities	0.14***		0.91						0.82***		0.88***	
L[a]												
Deposits to funding	0.01		0.38***						0.09		−0.01	
Net-short term borrowing	0.18**		1.20***						0.68**		1.03***	
S[a]												
Share of trading income	−0.14		−0.20						−0.27*		−0.15	
Country-specific banking sector indicators												
Total assets to GDP	0.71***				1.53***				2.33***			
Non-core liabilities	0.32***				0.22***				0.40***			
Debt to equity	0.30***				0.49***				0.08			
Loans to deposits	0.14				−0.21**				−0.60***			
Debt securities to liabilities	−0.22*				−0.45***				−0.08			
Mortgages to loans	0.03				0.40***				0.30***			
Country-specific macro-financial indicators												
Real GDP	−0.10·						−0.16***				−0.13**	
Inflation	0.06						0.10*				0.10*	
Stock prices	0.02						−0.04				−0.03	
House pric	−0.38***						−0.62***				−0.56***	
Long-term government bond yield	0.04						0.19*				0.10	
International investment position to GDP	−0.50***						−0.25***				−0.42***	
Government debt to GDP	0.50***						−0.02				0.10·	
Private sector credit flow to GDP	0.36***						0.52***				0.50***	
R2[b]	0.27		0.10		0.15		0.12		0.20		0.21	
No. of banks	298		298		298		298		298		298	
Predictive performance	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu)$	$U_r(\mu)$ (%)
Usefulness for a policymaker[c]												
$\mu = 0.6$	0.01	12	0.00	0	0.00	3	0.00	0	0.00	0	0.00	1
$\mu = 0.7$	0.01	16	0.00	0	0.00	4	0.00	4	0.00	5	0.01	7
$\mu = 0.8$	0.03	31	0.01	5	0.01	13	0.01	14	0.02	17	0.02	24
$\boldsymbol{\mu = 0.9}$	**0.04**	**42**	**0.02**	**19**	**0.02**	**21**	**0.02**	**22**	**0.03**	**30**	**0.03**	**36**
$P(I_j(h) = 1)$[d]	0.06		0.06		0.06		0.06		0.06		0.06	

Note: For standardized coefficients, the explanatory variables have been transformed to have zero mean and unit variance. Bold entries correspond to the benchmark preferences.
[a] The letters of CAMELS refer to Capital adequacy, Asset quality, Management, Earnings, Liquidity and Sensitivity to market risk.
[b] *R2* refers to the Nagelkerke's pseudo *R*-squared.
[c] The Usefulness for a policymaker is computed with absolute and relative usefulness $U_a(\mu)$ and $U_r(\mu)$ as described in Section 4.1.
[d] $P(I_j(h) = 1)$ refers to the unconditional probability of pre-distress events.
· Signif. codes: 0.10.
* Signif. codes: 0.05.
** Signif. codes: 0.01.
*** Signif. codes: 0.001.

Combining bank-level data and macro-financial indicators produces a model that not only outperforms a model with only bank-level data, but also one with bank-level and banking sector data. As the benchmark model still improves predictive performance compared to that of Model 6, it is justified to use all three groups of indicators also from a statistical point of view. Finally, Table 4 confirms the overall relative stability of the estimates and that in addition to the highest Usefulness, the benchmark model also obtains the highest R^2 (0.27).

Table 5 shows the predictive performance of the benchmark model for different policymaker preferences between type I and II errors. The models are calibrated with respect to non-weighted absolute Usefulness U_a, but we also compute weighted absolute and relative Usefulness $U_a(\mu, w_j)$ and $U_r(\mu, w_j)$ for each preference

Table 5
The predictive performance of the benchmark specification for different policymakers' preferences.

Preferences	TP	FP	TN	FN	Positives		Negatives		Accuracy	FP rate	FN rate	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu, w_j)$	$U_r(\mu, w_j)$ (%)	AUC
					Precision	Recall	Precision	Recall								
$\mu = 0.0$	0	0	3672	499	–	0.00	0.88	1.00	0.88	0.00	1.00	0.00	–	0.00	–	0.83
$\mu = 0.1$	0	0	3672	499	–	0.00	0.88	1.00	0.88	0.00	1.00	0.00	0	0.00	0	0.83
$\mu = 0.2$	0	0	3672	499	–	0.00	0.88	1.00	0.88	0.00	1.00	0.00	0	0.00	0	0.83
$\mu = 0.3$	0	0	3672	499	–	0.00	0.88	1.00	0.88	0.00	1.00	0.00	0	0.00	0	0.83
$\mu = 0.4$	33	20	3652	466	0.62	0.07	0.89	0.99	0.88	0.01	0.93	0.00	0	0.00	0	0.83
$\mu = 0.5$	57	56	3616	442	0.50	0.11	0.89	0.98	0.88	0.02	0.89	0.00	4	0.00	0	0.83
$\mu = 0.6$	127	97	3575	372	0.57	0.25	0.91	0.97	0.89	0.03	0.75	0.01	12	0.00	0	0.83
$\mu = 0.7$	206	292	3380	293	0.41	0.41	0.92	0.92	0.86	0.08	0.59	0.01	16	0.00	0	0.83
$\mu = 0.8$	288	529	3143	211	0.35	0.58	0.94	0.86	0.82	0.14	0.42	0.03	31	0.00	1	0.83
$\mu = 0.85$	324	703	2969	175	0.32	0.65	0.94	0.81	0.79	0.19	0.35	0.04	40	0.00	4	0.83
$\mu = 0.9$	**380**	**1055**	**2617**	**119**	**0.26**	**0.76**	**0.96**	**0.71**	**0.72**	**0.29**	**0.24**	**0.04**	**42**	**0.01**	**13**	**0.83**
$\mu = 0.95$	437	1645	2027	62	0.21	0.88	0.97	0.55	0.59	0.45	0.12	0.01	23	0.00	0	0.83
$\mu = 1.0$	499	3672	0	0	0.12	1.00	–	0.00	0.12	1.00	0.00	0.00	–	0.00	–	0.83

Notes: The table reports results for real-time out-of-sample predictions of a logit model with optimal thresholds w.r.t. Usefulness with given preferences. Bold entries correspond to the benchmark preferences. Thresholds are calculated for $\mu = \{0.0, 0.1, \ldots, 1.0\}$ and the forecast horizon is 8 quarters. The table also reports in columns the following measures to assess the overall performance of the models: TP = True positives, FP = False positives, TN = True negatives, FN = False negatives, Precision positives = TP/(TP + FP), Recall positives = TP/(TP + FN), Precision negatives = TN/(TN + FN), Recall negatives = TN/(TN + FP), Accuracy = (TP+TN)/(TP + TN + FP + FN), absolute and relative usefulness U_a and U_r (see formulae 1–3), and AUC = area under the ROC curve (TP rate to FP rate). See Section 4.1 for further details on the measures.

Table 6
The predictive performance of the benchmark specification by size quartile.

Bank size	Size range	TP	FP	TN	FN	Positives		Negatives		Accuracy	FP rate	FN rate	$U_a(\mu)$	$U_r(\mu)$ (%)	AUC
						Precision	Recall	Precision	Recall						
Very small	[1.0, 6.2]	17	133	870	26	0.11	0.40	0.97	0.87	0.85	0.13	0.60	0.01	13	0.68
Small	[6.2, 19.0]	50	114	867	15	0.30	0.77	0.98	0.88	0.88	0.12	0.23	0.05	60	0.89
Large	[19.0, 84.0]	114	194	682	58	0.37	0.66	0.92	0.78	0.76	0.22	0.34	0.03	37	0.77
Very large	[84.0, 1258.5]	145	277	540	84	0.34	0.63	0.87	0.66	0.65	0.34	0.37	0.02	21	0.77

Notes: Given a break down of the benchmark specification results by quartiles of total asset, the table reports real-time out-of-sample prediction results with $\mu = 0.9$ and a forecast horizon of 8 quarters. The first column shows labels representing quartiles of total assets, whereas the second column refers to the range of bank size in billions. The rest of the table reports in columns the following measures to assess the overall performance of the models: TP = True positives, FP = False positives, TN = True negatives, FN = False negatives, Precision positives = TP/(TP + FP), Recall positives = TP/(TP + FN), Precision negatives = TN/(TN + FN), Recall negatives = TN/(TN + FP), Accuracy = (TP + TN)/(TP + TN + FP + FN), absolute and relative usefulness U_a and U_r (see formulae 1–3), and AUC = area under the ROC curve (TP rate to FP rate). See Section 4.1 for further details on the measures.

Table 7
The predictive performance of the benchmark specification for different countries.

Country	TP	FP	TN	FN	Positives		Negatives		Accuracy	FP rate	FN rate	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu, w_j)$	$U_r(\mu, w_j)$ (%)	AUC
					Precision	Recall	Precision	Recall								
Austria	9	16	74	21	0.36	0.30	0.78	0.82	0.69	0.18	0.70	0.00	−3	−0.01	−8	0.72
Belgium	15	0	4	5	1.00	0.75	0.44	1.00	0.79	0.00	0.25	0.06	69	0.07	78	0.98
Czech Republic	0	0	36	0	–	–	1.00	1.00	1.00	0.00	–	–	–	–	–	–
Denmark	11	20	81	13	0.35	0.46	0.86	0.80	0.74	0.20	0.54	0.01	14	0.07	76	0.67
Finland	0	13	77	0	0.00	–	1.00	0.86	0.86	0.14	–	–	–	–	–	–
France	23	69	553	27	0.25	0.46	0.95	0.89	0.86	0.11	0.54	0.02	23	0.00	3	0.91
Germany	11	32	128	13	0.26	0.46	0.91	0.80	0.76	0.20	0.54	0.01	14	−0.01	−7	0.85
Greece	66	40	2	21	0.62	0.76	0.09	0.05	0.53	0.95	0.24	−0.02	−25	−0.01	−15	0.73
Hungary	0	3	2	6	0.00	0.00	0.25	0.40	0.18	0.60	1.00	−0.07	−82	−0.07	−82	0.60
Ireland	25	0	2	3	1.00	0.89	0.40	1.00	0.90	0.00	0.11	0.08	87	0.08	87	0.98
Italy	7	102	1164	17	0.06	0.29	0.99	0.92	0.91	0.08	0.71	0.00	5	0.02	28	0.67
Lithuania	0	0	13	0	–	–	1.00	1.00	1.00	0.00	–	–	–	–	–	–
Netherlands	1	8	65	0	0.11	1.00	1.00	0.89	0.89	0.11	0.00	0.08	89	0.07	79	0.50
Poland	0	20	214	0	0.00	–	1.00	0.91	0.91	0.09	–	–	–	–	–	–
Portugal	5	96	48	0	0.05	1.00	1.00	0.33	0.36	0.67	0.00	0.03	33	0.03	32	0.60
Slovakia	0	0	36	0	–	–	1.00	1.00	1.00	0.00	–	–	–	–	–	–
Slovenia	0	0	18	0	–	–	1.00	1.00	1.00	0.00	–	–	–	–	–	–
Spain	120	231	322	29	0.34	0.81	0.92	0.58	0.63	0.42	0.19	0.03	34	0.02	26	0.71
Sweden	3	15	32	12	0.17	0.20	0.73	0.68	0.56	0.32	0.80	−0.03	−30	−0.02	−19	0.51
United Kingdom	20	53	83	16	0.27	0.56	0.84	0.61	0.60	0.39	0.44	0.01	7	0.00	3	0.68

Notes: Given a break down of the benchmark specification results by countries, the table reports real-time out-of-sample prediction results with $\mu = 0.9$ and a forecast horizon of 8 quarters. The table also reports in columns the following measures to assess the overall performance of the models: TP = True positives, FP = False positives, TN = True negatives, FN = False negatives, Precision positives = TP/(TP + FP), Recall positives = TP/(TP + FN), Precision negatives = TN/(TN + FN), Recall negatives = TN/(TN + FP), Accuracy = (TP + TN)/(TP + TN + FP + FN), absolute and relative usefulness U_a and U_r (see formulae 1–3), and AUC = area under the ROC curve (TP rate to FP rate). See Section 4.1 for further details on the measures.

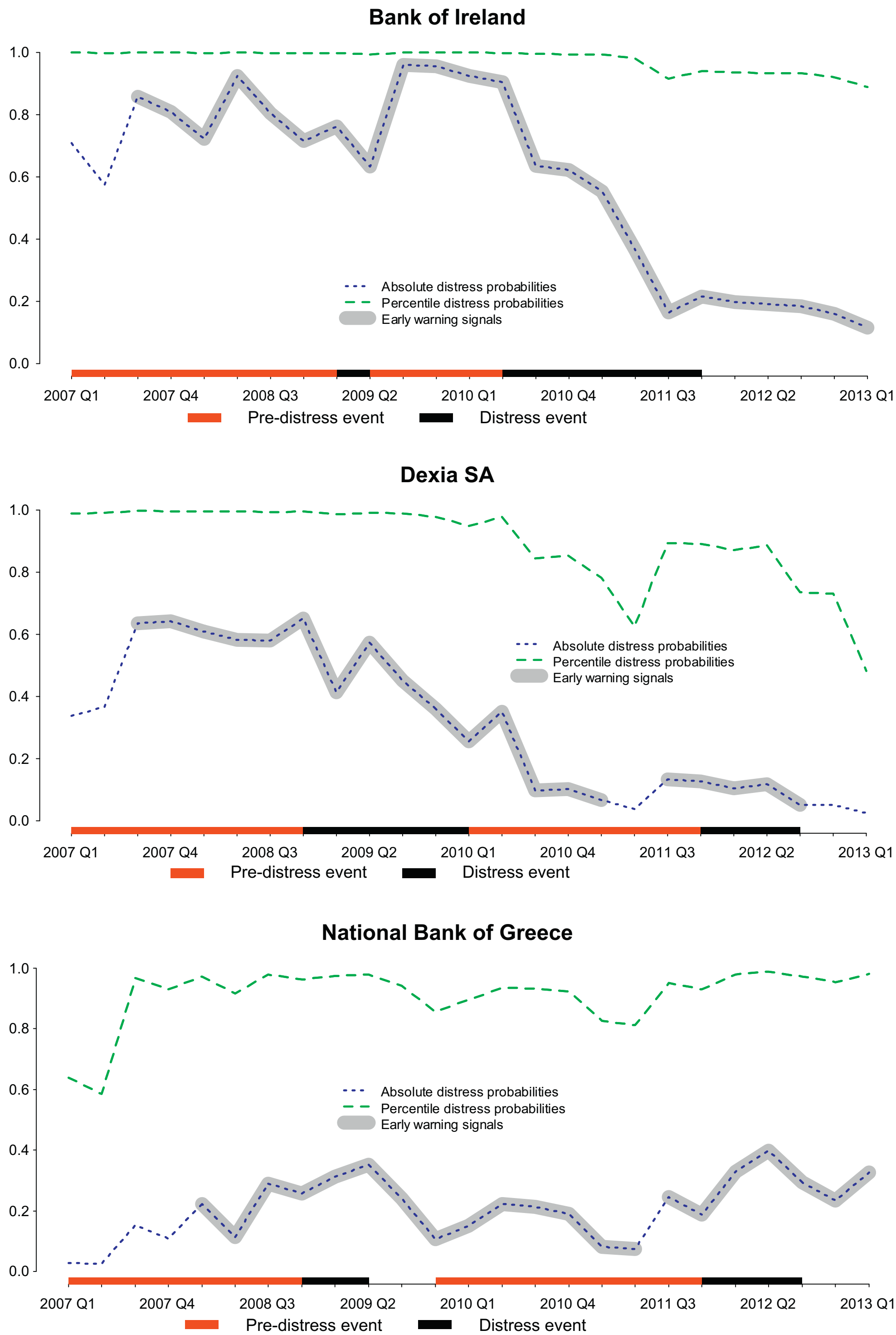

Fig. 2. Case studies on Bank of Ireland, Dexia and National Bank of Greece. Out-of-sample prediction of bank distress (8 quarters ahead) from 2007Q1 to 2013Q1.

parameter, where weights w_j represent bank size.[6] As mentioned earlier, from the systemic risk monitoring point of view, the policymaker may want to attach a higher weight for early-warning signals from systemically important banks.

When focusing on non-weighted $U_a(\mu)$, the results in Table 5 indicate that it is optimal to disregard the model for $\mu \leqslant 0.3$. For $\mu \leqslant 0.3$, given the uneven distribution of bank distress events compared to tranquil periods, the policymaker is better off by not signalling at all. In addition, Table 5 shows that the model performance significantly decreases for $\mu = 0.9$ when augmenting the Usefulness measure with bank-specific weights w_j. This supports the view that vulnerabilities and risks of large financial institutions are oftentimes more complex than those of smaller ones, while

[6] Systemic relevance of a bank is approximated by computing its share of total assets to the sum of total assets of all banks in the sample at quarter t. Possible extensions to proxy for the systemic relevance of a bank are to take into account its degree of interconnectedness in the interbank network or its estimated Value-at-Risk.

Table 8
Robustness of the model with respect to out-of-sample forecast horizon.

Forecast horizon	TP	FP	TN	FN	Positives		Negatives		Accuracy	FP rate	FN rate	$U_a(\mu)$	$U_r(\mu)$ (%)	$U_a(\mu, w_j)$	$U_r(\mu, w_j)$ (%)	AUC
					Precision	Recall	Precision	Recall								
12 months	191	626	3272	82	0.23	0.70	0.98	0.84	0.83	0.16	0.30	0.03	44	0.01	20	0.86
24 months	**380**	**1055**	**2617**	**119**	**0.26**	**0.76**	**0.96**	**0.71**	**0.72**	**0.29**	**0.24**	**0.04**	**42**	**0.01**	**13**	**0.83**
36 months	482	1413	2182	94	0.25	0.84	0.96	0.61	0.64	0.39	0.16	0.03	37	0.01	8	0.82

Notes: The table reports results for real-time out-of-sample predictions of a logit model with optimal thresholds w.r.t. Usefulness with given preferences. Bold entries correspond to the benchmark horizon and thresholds are calculated for $\mu = \{0.0, 0.1, \ldots, 1.0\}$. The table also reports in columns the following measures to assess the overall p erformance of the models: TP = True positives, FP = False positives, TN = True negatives, FN = False negatives, Precision positives = TP/(TP + FP), Recall p ositives = TP/(TP + FN), Precision negatives = TN/(TN + FN), Recall negatives = TN/(TN + FP), Accuracy = (TP + TN)/(TP + TN + FP + FN), absolute and relative usefulness U_a and U_r (see formulae 1–3), and AUC = area under the ROC curve (TP rate to FP rate). See Section 4.1 for further details on the measures.

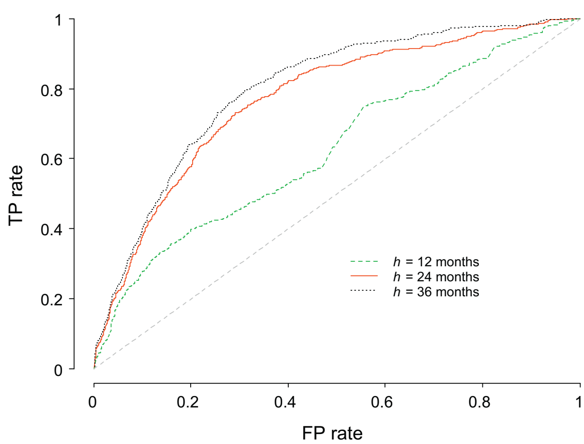

Fig. 3. Robustness of the model with respect to λ for different forecast horizons.

bank size has been shown to be positively correlated with the probability of bank failure (see e.g. Cole and White, 2012; Jin et al., 2011).

Breaking down the results of the benchmark specification by size confirms the above results. Table 6 presents results by quartiles of total assets and shows that the model performs best for the banks in the second quartile of the size distribution, (labelled "small", with total assets of 6.2–19.0 bln euros) according to both the usefulness measures and the area under ROC curve. The model performance comes in second for banks classified as "large" (total assets 19.0–84.0 bln euros). However, at 37%, the relative usefulness for large banks is barely more than half of that for small banks (60%). For banks in the fourth quartile of the size distribution, the model achieves a relative usefulness of 21% and for those in the first quartile of 13%.

Domestic macro-financial factors played an important factor during the global financial crisis, which is reflected in the significant explanatory power of banking sector and macro-financial indicators. This can also explain why the model performs better for comparatively small banks than for large ones. Large banks' business models are generally more complex and diversified, both in terms of income sources and geography. Thus, one would expect them to be less dependent on the fortunes of their home market. The banking sector and macro-financial indicators used in this study, however, characterize developments in the home market, and are thus likely to be less effective in predicting the vulnerabilities of large banks with diversified income sources from various countries.

Table 7 shows the predictive performance of the model for banks in different countries. It can be observed that the model performs best for banks in the Netherlands, Ireland, and Belgium when evaluating with the relative usefulness measure, while the next best performance can be observed for banks in Spain and Portugal. In some of these cases, the model performance is superior when the signals are weighted by the bank size, as is the case of Belgium. However, in the cases of banks in the Netherlands, Spain and Portugal, the model yields a higher level of relative usefulness when the unweighted signals are considered (the signals for banks in Ireland are similar across the two methods). Using the relative usefulness measure, the model performance is the weakest for banks in Hungary, Sweden and Greece.

Owing to the fact that the model is estimated using pooled data among all banks and countries (i.e. it does not account for country- or bank-specific effects, except for the use of country-level data), the model is also capable of discriminating between pre-distress and tranquil periods for banks in countries which have not experienced any distress events. The use of country-level data enables the model to capture the clustering of distress events at the country level. Thus, the model can distinguish between crisis and non-crisis countries. Model performance in the Czech Republic, Lithuania, Slovakia, and Slovenia reflects this. In these countries, the model achieves perfect accuracy by never signalling.

To illustrate the out-of-sample performance of the benchmark model, Fig. 2 shows how the benchmark model would have performed predicting out of sample the failures of three example banks, namely Bank of Ireland, Dexia and National Bank of Greece. The figure shows blue[7] (short dashes) and green (long dashes) lines for absolute and percentile distress probabilities, and highlights in grey the periods when the model signals a vulnerable state for the bank in question. The black lines on top of the x-axis represent the distress events and the red lines the vulnerable states (or pre-distress periods) that the model aims to correctly call. Starting with the example of Bank of Ireland, the model signals vulnerability in 2007Q3, when the distress event occurs in 2009Q1, and throughout the pre-distress period before the distress event that started in 2010Q2. The middle panel of Fig. 2 shows a similar case study on Dexia. In the run up to the first distress event in 2008Q4, the model signals early on and consistently ranks Dexia as one of the most risky banks in the sample (as shown by the percentile probabilities). Later, the model correctly signals most of the quarters of vulnerability before the second distress event in 2011Q4. Likewise, the model signals correctly well in advance both distress periods of National Bank of Greece. Moreover, as shown in Fig. 2, in the cases of Bank of Ireland and National Bank of Greece, the model signals a potential vulnerability in the final quarter 2013Q1. To further illustrate the model performance, Fig. A in Appendix A shows three additional case studies on large euro area banks that did not experience a distress event over the sample period: Banco Santander, Deutsche Bank and UniCredit. Whereas the model issues a few false alarms for Banco Santander, which to a large extent stem from country-specific

[7] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

vulnerabilities, the figure illustrates that the model, in the cases of Deutsche Bank and UniCredit, does not issue early-warning signals.

6. Robustness

We test the robustness of the early-warning model in several ways. As Table 3 shows, the results are, in a broad sense, robust to omitting some key CAMELS indicators with weak data coverage, such as Tier 1 capital ratio and impaired assets. Similarly, Table 4 shows that complementing bank-specific vulnerabilities with indicators of macro-financial imbalances and banking sector vulnerabilities substantially improves model performance. Further, as partly discussed in Section 5 (Tables 3–5), the out-of-sample performance of the model is sensitive to the policymaker's preferences due to the imbalanced frequencies of distress events and tranquil periods. The model is useful for preference parameters of $\mu > 0.5$ (see Table 5).

In addition, we study the out-of-sample performance for different forecast horizons. As shown in Table 8, the absolute Usefulness

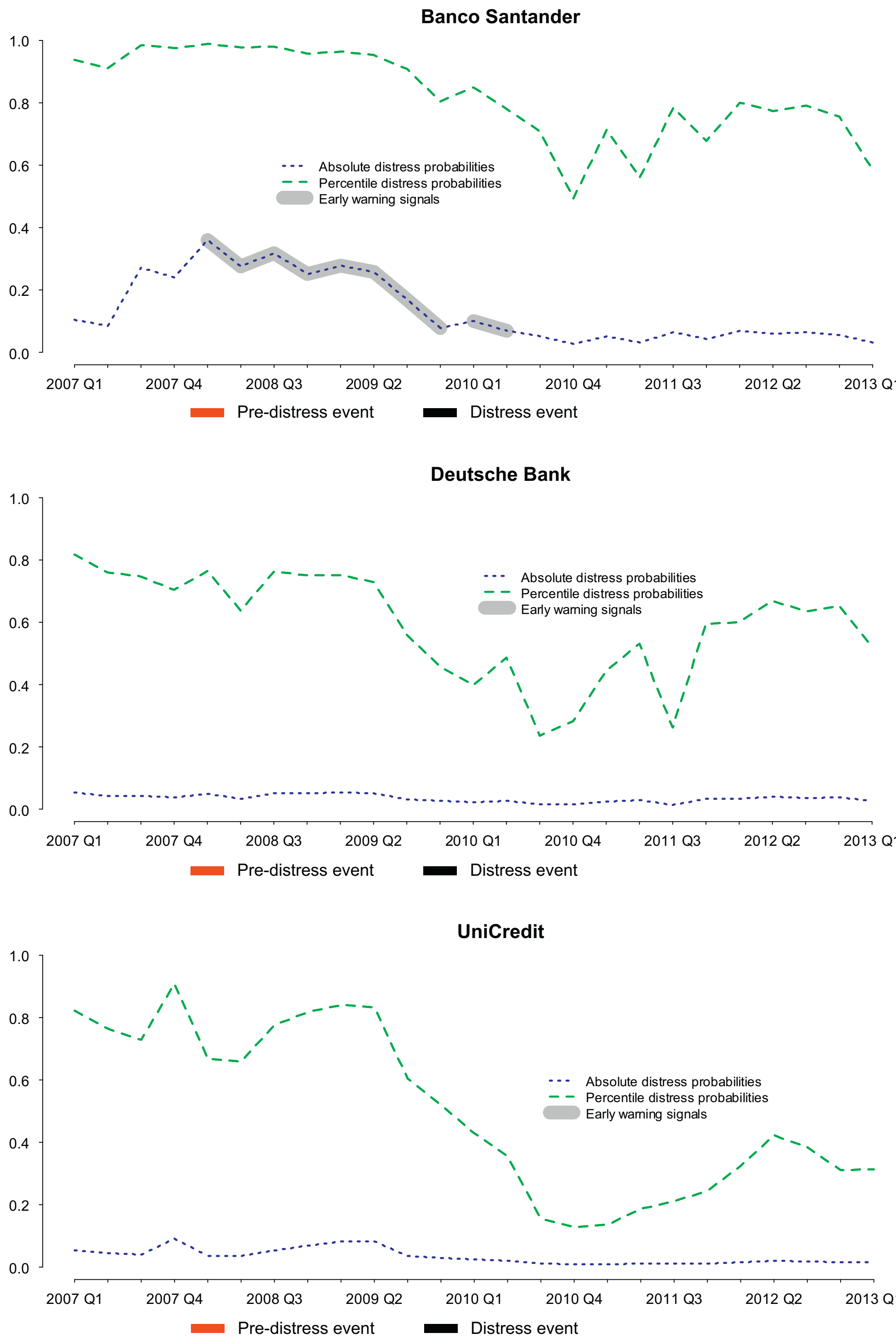

Fig. A. Case studies on Banco Santander, Deutsche Bank and UniCredit. Out-of-sample prediction of bank distress (8 quarters ahead) from 2007Q1 to 2013Q1.

of the model is slightly poorer for out-of-sample forecast horizons of 12 and 36 months. The models' relative Usefulness is, however, slightly better for shorter horizons. This is mainly due to the fact that decreases in unconditional probabilities of pre-distress events lead to decreases in the available Usefulness, and thus the captured share of the Usefulness increases.

Finally, we show the sensitivity of the early-warning model to variation of the thresholds with an ROC curve. The curve plots the benefit (TP rate) to the cost (FP rate) of a certain model for each threshold λ, as noted in Section 4.1. While not accounting for imbalanced data and misclassification costs, the ROC curve's area above the diagonal line represents the benefit of a model in relation to a coin toss. Fig. 3 not only shows that the ROC curves are above those of a coin toss, but also that curves of 24 and 36-month horizons are similar, while that of a 12-month horizon is somewhat poorer. This exercise is, however, imprecise in nature. While the models issue signals based upon time-varying thresholds such that in-sample Usefulness is optimized, the ROC statistics treat all probabilities as similar. Another common limitation of the ROC curve, especially the AUC, is that parts of it, which are not policy relevant, are included in the computed area.

7. Conclusions

The paper presents an early-warning model for predicting individual bank distress in the European banking sector, using both bank-level and country-level indicators of vulnerabilities. As outright bank failures have been rare in Europe, we introduce a novel dataset that complements bankruptcies, liquidations and defaults by also taking into account state interventions and mergers in distress. Moreover, the signals of the early-warning model are calibrated not only according to policymakers' preferences between type I and II errors, but also accounting for the potential systemic relevance of each individual financial institution, proxied by its size.

The paper finds that complementing bank-specific vulnerabilities with indicators for country-level macro-financial imbalances and banking sector vulnerabilities improves model performance. In particular, our results confirm the usefulness of the vulnerability indicators introduced recently via the EU Macroeconomic Imbalance Procedure (MIP). In addition, the results show that an early-warning model based on publicly available data yields useful out-of-sample predictions of bank distress during the global financial crisis. While one could argue for the Lucas critique, the aim of more accurate early-warning signals relates to few undesirable

Table A
Indicators, definitions, transformations and data sources.

Variable	Definition and transformation	Source
Bank-specific indicators		
C[a]		
Capital ratio	Total Equity/Total Assets	Bloomberg
Tier 1 ratio	Tier1 Capital Ratio	Bloomberg
Impaired assets	Non Performing Assets/Total Assets	Bloomberg
A[a]		
Reserves to impaired assets	Reserves for Loan Losses/Non Performing Assets	Bloomberg
ROA	Return on Assets	Bloomberg
Loan loss provisions	Provisions for Loan Losses/Total Average Loans	Bloomberg
M[a]		
Cost to income	Operating Costs/Operating Income	Bloomberg
ROE	Return on Equity	Bloomberg
E[a]		
Net interest margin	Net Interest Margin	Bloomberg
Interest expenses to liabilities	Interest Expenses/Total Liabilities	Bloomberg
L[a]		
Deposits to funding	Deposits/Funding	Bloomberg
Net-short term borrowing	Short-term borrowing-Cash/Total Liabilities	Bloomberg
S[a]		
Share of trading income	Trading Income/Operating Income	Bloomberg
Country-specific banking sector indicators		
Total assets to GDP	Total Assets/GDP	ECB MFI statistics
Non-core liabilities	Growth rate of (Total Liabilities – Capital and Reserves -Deposits)	ECB MFI statistics
Debt to equity	(Total Liabilities – Capital and Reserves)/Capital and Reserves	ECB MFI statistics
Loans to deposits	Total Loans/Deposits	ECB MFI statistics
Debt securities to liabilities	Debt securities to Liabilities	ECB MFI statistics
Mortgages to loans	Mortgages to Total Loans	ECB MFI statistics
Country-specific macro-financial indicators		
Real GDP	Growth rate of real GDP	Euroostat
Inflation	Growth rate of the HICP index	Eurostat
Stock prices	Growth rate of the stock price index	Bloomberg
House prices	Growth rate of the house price index	ECB
Long-term government bond yield	10-year government bond yield	Bloomberg
International investment position to GDP	Net International Investment Position as a % of GDP	Eurostat/Alert Mechanism Report
Government debt to GDP	General government debt as % of GDP	Eurostat/Alert Mechanism Report
Private sector credit flow to GDP	Private sector credit flow as % of GDP	Eurostat/Alert Mechanism Report

Notes: The growth rates are quarterly growth rates. To account for publication lags, bank balance sheet and house price indicator are lagged by 2 quarters, whereas the MIP variables from the Alert Mechanism Report are lagged by 6 quarters. Banking sector indicators and all other macro-financial indicators are lagged by 1 quarter.
[a] The letters refer to Capital adequacy (C), Asset quality (A), Management (M), Earnings (E), Liquidity (L), and Sensitivity to market risk (S).

Table B
Summary statistics.

Variable	Obs	Min	Max	Mean	Std. dev.	Skew	Kurt
Bank-specific indicators							
C[a]							
Capital ratio	20432	0.00	0.30	0.07	0.04	2.20	8.05
Tier 1 ratio	11999	0.04	0.32	0.10	0.04	2.18	6.71
Impaired assets	13079	0.00	0.17	0.02	0.03	2.62	8.79
A[a]							
Reserves to impaired assets	12735	0.00	19.19	1.52	2.38	5.61	36.78
ROA	20608	−0.04	0.03	0.01	0.01	−1.15	8.48
Loan loss provisions	17010	0.00	0.06	0.01	0.01	3.60	16.62
M[a]							
Cost to income	20122	−24.05	43.27	2.84	6.43	2.20	19.24
ROE	20351	−0.86	0.44	0.07	0.14	−3.55	20.52
E[a]							
Net interest margin	17501	0.00	0.06	0.02	0.01	0.85	1.42
Interest expenses to liabilities	19793	0.00	0.10	0.03	0.02	1.85	5.06
L[a]							
Deposits to funding	19321	0.00	0.98	0.55	0.24	−0.30	−0.58
Net-shor0t term borrowing	19950	−0.41	0.72	0.08	0.19	0.73	1.77
S[a]							
Share of trading income	19761	−3.60	4.74	0.26	0.84	0.89	13.18
Total assets to GDP	26562	1.40	135.92	12.27	12.05	7.61	67.02
Country-specific banking sector indicators							
Non-core liabilities	26517	−0.11	0.05	0.00	0.01	−0.16	8.03
Debt to equity	26578	3.87	38.70	14.38	4.23	0.60	−0.34
Loans to deposits	26578	1.00	7.42	2.39	0.75	1.73	5.79
Debt securities to liabilities	26578	0.00	0.51	0.17	0.09	0.66	1.97
Mortgages to loans	26438	0.01	0.41	0.17	0.07	0.65	0.16
Country-specific macro-financial indicators							
Real GDP	27297	−0.13	0.08	0.00	0.01	−1.29	11.24
Inflation	27401	−0.03	0.11	0.01	0.01	1.72	16.65
Stock prices	27279	−0.55	0.89	0.00	0.12	−0.16	1.14
House prices	26628	−0.40	1.59	0.01	0.03	8.01	275.70
Long-term government bond yield	26919	0.01	0.34	0.04	0.02	6.57	81.79
International investment position to GDP	27401	−1.75	1.40	−0.19	0.33	−0.26	2.18
Government debt to GDP	27401	0.04	1.71	0.71	0.28	0.34	−0.61
Private sector credit flow to GDP	27401	−0.38	1.76	0.09	0.09	4.50	70.48

Note: The statistics are derived from the entire sample with 28,832 observations.
[a] The letters refer to Capital adequacy (C), Asset quality (A), Management (M), Earnings (E), Liquidity (L), and Sensitivity to market risk (S).

effects in terms of changes in behavior and expectations, as is also noted by Bisias et al. (2012). They relate it to two arguments: (i) the more accurate the risk measures are, the more accurate are the inputs to policy and (ii) one key intent of early-warning exercises is to encourage agents to take actions on their own, rather than only relying on governments. Hence, this relates mainly to challenges in ex post evaluations of predictive accuracy, as terminally false predictions might have been presumptively true.

Finally, the results of the evaluation framework show that a policymaker has to be substantially more concerned with missing bank distress than issuing false alarms for the model to be useful. This is intuitive if we consider that an early-warning signal triggers an internal in-depth review of fundamentals, business model and peers of the bank predicted to be in distress. Should the analysis reveal that the signal is false, there is no loss of credibility on behalf of the policy authority as the model results are not published. The evaluations also imply that it is important to give more emphasis to systemically important and large banks for a policymaker concerned with systemic risk.

Acknowledgments

The authors want to thank Riina Vesanen for excellent research assistance and colleagues at the ECB, particularly Carsten Detken, Jean-Paul Genot, Marco Gross, Philipp Hartmann, Urszula Kochanska, Markus Kolb, Bernd Schwaab, Joseph Vendrell as well as Kostas Tsatsaronis from the BIS, both for discussions and sharing data. Thanks also to participants at the ECB Financial Stability seminar on 16 May 2012 and the ECB Finan-cial Stability conference: Methodological advances and policy issues on 14–15 June 2012 in Frankfurt am Main, Germany, particularly to Dilruba Karim, the third MAFIN conference on 19–21 September 2012 in Genoa, Italy, CEQURA Conference on Advances in Financial and Insurance Risk Management on 24–26 September 2012 in Munich, Germany, the second conference of the ESCB Macro-prudential Research Network on 30–31 October 2012 in Frankfurt am Main, Germany, particularly to Philip Davis, Federal Reserve Bank of Cleveland and the Office of Financial Research Financial Stability Conference on 30–31 May 2013 in Washington, DC, United States, and the INFINITI Conference on International Finance – "The Financial Crisis, Integration and Contagion" on 10–11 June 2013 in SciencesPo Aix, Aix-en-Provence, France, and the International Conference C.R.E.D.I.T. 2013 on 26–27 September 2013 in Venice, Italy. All remaining errors are our own. Peter Sarlin gratefully acknowledges the financial support of the Academy of Finland (Grant no. 127592). The views presented in the paper are those of the authors only and do not necessarily represent the views of the European Central Bank, the Eurosystem or the European Investment Bank.

Table C
Mean-comparison tests.

Variables	$C = 0$			$C = 1$			Mean t-test	
	Obs	Mean std.	Dev.	Obs	Mean	Std. dev.	t	Prob.
Bank-specific indicators								
C[a]								
Capital ratio	19585	0.07	0.04	847	0.05	0.03	22.13	0.00
Tier 1 ratio	11397	0.10	0.04	602	0.08	0.02	20.64	0.00
Impaired assets	12474	0.02	0.03	605	0.03	0.03	1.61	0.11
A[a]								
Reserves to impaired assets	12146	1.53	2.39	589	1.22	2.17	3.37	0.00
ROA	19761	0.01	0.01	847	0.00	0.01	5.95	0.00
Loan loss provisions	16292	0.01	0.01	718	0.01	0.01	4.42	0.00
M[a]								
Cost to income	19283	2.84	6.44	839	2.85	6.33	0.03	0.98
ROE	19504	0.07	0.14	847	0.06	0.19	1.67	0.10
E[a]								
Net interest margin	16768	0.02	0.01	733	0.02	0.01	1.27	0.20
Interest expenses to liabilities	18984	0.03	0.02	809	0.04	0.02	9.62	0.00
L[a]								
Deposits to funding	18508	0.55	0.24	813	0.55	0.23	0.12	0.90
Net-short term borrowing	19121	0.08	0.19	829	0.10	0.15	4.78	0.00
S[a]								
Share of trading income	18928	0.26	0.83	833	0.29	1.02	0.71	0.48
Country-specific banking sector indicators								
Total assets to GDP	25552	12.19	2.19	1010	14.40	7.37	9.06	0.00
Non-core liabilities	25507	0.00	0.01	1010	0.00	0.01	3.75	0.00
Debt to equity	25568	14.37	4.25	1010	14.83	3.92	3.66	0.00
Loans to deposits	25568	2.39	0.74	1010	2.42	1.02	0.97	0.33
Debt securities to liabilities	25568	0.18	0.09	1010	0.16	0.10	5.91	0.00
Mortgages to loans	25428	0.17	0.07	1010	0.20	0.07	15.25	0.00
Country-specific macro-financial								
Real GDP	26279	0.00	0.01	1018	0.00	0.01	10.34	0.00
Inflation	26383	0.01	0.01	1018	0.01	0.01	0.51	0.61
Stock prices	26261	0.01	0.12	1018	−0.03	0.13	7.20	0.00
House prices	25617	0.01	0.03	1011	0.00	0.03	12.89	0.00
Long-term government bond yield	25933	0.04	0.02	986	0.05	0.02	3.37	0.00
International investment position to GDP	26383	−0.19	0.32	1018	−0.38	0.42	14.48	0.00
Government debt to GDP	26383	0.72	0.28	1018	0.59	0.27	14.56	0.00
Private sector credit flow to GDP	26383	0.09	0.09	1018	0.16	0.11	18.60	0.00

Note: The statistics are derived from the entire sample with 28,832 observations.
[a] The letters refer to Capital adequacy (C), Asset quality (A), Management (M), Earnings (E), Liquidity (L), and Sensitivity to market risk (S).

Appendix A

See Fig. A and Tables A–C.

References

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. Journal of Banking and Finance 32, 1541–1551.

Alessi, L., 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, 520–533.

Arena, M., 2008. Bank failures and bank fundamentals: a comparative analysis of Latin America and East Asia during the nineties using bank-level data. Journal of Banking and Finance 32 (2), 299–310.

Beltratti, A., Stulz, R., 2012. The credit crisis around the globe: Why did some banks perform better? Journal of Financial Economics 105, 1–17.

Berg, A., Pattillo, C., 1999a. Predicting currency crises: the indicators approach and an alternative. Journal of International Money and Finance 18, 561–586.

Bharath, S., Shumway, T., 2008. Forecasting default with the Merton distance to default model. The Review of Financial Studies 21, 1339–1369.

Bisias, D., Flood, M., Lo, A., Valavanis, S., 2012. A survey of systemic risk analytics. Annual Review of Financial Economics 4, 255–296.

Bongini, P., Laeven, L., Majnoni, G., 2002. How good is the market at assessing bank fragility? A horse race between different indicators. Journal of Banking and Finance 26 (5), 1011–1028.

Borio, C., Lowe, P., 2002. Asset Prices, Financial and Monetary Stability: Exploring the Nexus. BIS Working Papers 114.

Bussière, M., Fratzscher, M., 2006. Towards a new early warning system of financial crises. Journal of International Money and Finance 25, 953–973.

Campbell, J., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. The Journal of Finance 58, 2899–2939.

Cole, R.A., Gunther, J., 1995. Separating the likelihood and timing of bank failure. Journal of Banking and Finance 19, 1073–1089.

Cole, R.A., Gunther, J., 1998. Predicting bank failures: a comparison of on- and off-site monitoring systems. Journal of Financial Services Research 13, 103–117.

Cole, R.A., White, L.J., 2012. Deja vu all over again: the causes of U.S. commercial bank failures this time around. Journal of Financial Services Research 42, 5–29.

Cole, R.A., Wu, Q., 2009. Is Hazard or Probit more Accurate in Predicting Financial Distress? Evidence from U.S. Bank Failures. Working Paper. <http://mpra.ub.uni-muenchen.de/24688/>.

Curry, T., Elmer, P., Fissel, G., 2007. Equity market data, bank failures and market efficiency. Journal of Economics and Business 59, 536–559.

Davis, E., Karim, D., 2008a. Could early warning systems have helped to predict the sub-prime crisis? National Institute Economic Review 206, 35–47.

Davis, E., Karim, D., 2008b. Comparing early warning systems for banking crisis. Journal of Financial Stability 4, 89–120.

Dell'Ariccia, G., Detragiache, E., Rajan, R., 2008. The real effects of banking crises. Journal of Financial Intermediation 17, 89–112.

Demirgüç-Kunt, A., Detragiache, E., 1998. The determinants of banking crises in developed and developing countries. IMF Staff Papers 45, 81–109.

Demirgüç-Kunt, A., Detragiache, E., 2000. Monitoring banking sector fragility. A multivariate logit. World Bank Economic Review 14, 287–307.

Demyanyk, Y., Hasan, I., 2010. Financial crises and bank failures: a review of prediction methods. Omega 38 (5), 315–324.

DeYoung, R., Torna, G., 2013. Nontraditional banking activities and bank failures during the financial crisis. Journal of Financial Intermediation 22 (3), 397–421.

Edison, H., 2003. Do indicators of financial crises work? An evaluation of an early warning system. International Journal of Finance and Economics 8, 11–53.

European Commission (2012). Scoreboard for the Surveillance of Macroeconomic Imbalances. European Economy Occasional Papers 92.

Flannery, M., 1998. Using market information in prudential bank supervision: a review of the US empirical evidence. Journal of Money, Credit and Banking 30, 273–305.

Fuertes, A., Kalotychou, E., 2007. Towards the optimal design of an early warning system for sovereign debt crises. International Journal of Forecasting 23, 85–100.

Fuertes, A.-M., Kalotychou, E., 2006. Early warning system for sovereign debt crisis: the role of heterogeneity. Computational Statistics and Data Analysis 5, 1420–1441.

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.

Hahm, J.-H., Shin, H.S., Shin, K., 2013. Noncore bank liabilities and financial vulnerability. Journal of Money, Credit and Banking 45 (s1), 3–36.

Haq, M., Heaney, R., 2012. Factors determining European bank risk. Journal of International Financial Markets, Institutions and Money 22 (4), 696–718.

Hernandez Tinoco, M., Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. International Review of Financial Analysis 30, 394–419.

Hutchinson, M., 2002. European banking distress and EMU: institutional and macroeconomic risks. Scandinavian Journal of Economics 104, 365–389.

Jagtiani, J., Lemieux, C., 2001. Market discipline prior to bank failure. Journal of Economics and Business 53, 313–324.

Jin, J., Kanagaretnam, K., Lobo, G., 2011. Ability of accounting and audit quality variables to predict bank failure during the financial crisis. Journal of Banking and Finance 35 (11), 2811–2819.

Jin, J., Kanagaretnam, K., Lobo, G., Mathieu, R., 2013. Impact of FDICIA internal controls on bank risk taking. Journal of Banking and Finance 37 (2), 614–624.

Jordan, D.J., Rice, D., Sanchez, J., Walker, C., Worth, D.H., 2010. Predicting Bank Failures: Evidence from 2007 to 2010. Working Paper. <http://papers.ssrn.com>.

Kaminsky, G., Lizondo, S., Reinhart, C., 1998. Leading indicators of currency crises. IMF Staff Papers 45, 1–48.

Kaminsky, G., Reinhart, C., 1999. The twin crises: the causes of banking and balance-of-payments problems. American Economic Review 89, 473–500.

Kick, T., Koetter, M., 2007. Slippery slopes of stress: ordered failure events in German banking. Journal of Financial Stability 3 (2), 132–148.

Kumar, M., Moorthy, U., Perraudin, W., 2003. Predicting emerging market currency crashes. Journal of Empirical Finance 10, 427–454.

Laeven, L., Valencia, F., 2008. Systemic Banking Crises: A New Database. IMF Working Paper 08/224.

Laeven, L., Valencia, F., 2010. Resolution of Banking Crises: The Good, the Bad, and the Ugly. IMF Working Paper 10/146.

Laeven, L., Valencia, F., 2011. The Real Effects of Financial Sector Interventions during Crises. IMF Working Paper 11/45.

LoDuca, M., Peltonen, T., 2013. Assessing systemic risks and predicting systemic events. Journal of Banking and Finance 37 (7), 2183–2195.

López-Iturriaga, F., López-de-Foronda, O., Pastor-Sanz, I., 2010. Predicting Bankruptcy using Neural Networks in the Current Financial Crisis: A Study of U.S. Commercial Banks. Working Paper. <http://papers.ssrn.com>.

Männasoo, K., Mayes, D., 2009. Explaining bank distress in Eastern European transition economies. Journal of Banking and Finance 33, 244–253.

Milne, A., forthcoming. Distance to default and the financial crisis. Journal of Financial Stability (2013).

Mody, A., Sandri, D., 2012. The eurozone crisis: how banks and sovereigns came to be joined at the hip. Economic Policy 27, 199–230.

Ötker, I., Podpiera, J., 2010. The Fundamental Determinants of Credit Default Risk for European Large Complex Financial Institutions. IMF Working Paper 10/153.

Poghosyan, T., Cihák, M., 2011. Distress in European banks: an analysis based on a new dataset. Journal of Financial Services Research 40, 163–184.

Reinhart, C., Rogoff, K., 2008. Is the 2007 US sub-prime financial crisis so different? An international historical comparison. American Economic Review 98, 339–344.

Reinhart, C., Rogoff, K., 2009. The aftermath of financial crises. American Economic Review 99, 466–472.

Rose, A., Spiegel, M., 2011. Cross-country causes and consequences of the crisis: an update. European Economic Review 55, 309–324.

Sarlin, P., 2013. On policymakers' loss functions and the evaluation of early warning systems. Economics Letters 119, 1–7.

Sarlin, P., Peltonen, T., 2013. Mapping the state of financial stability. Journal of International Financial Markets, Institutions and Money 26, 46–76.

Stolz, S., Wedow, M., 2010. Extraordinary Measures in Extraordinary Times. Public Measures in Support of the Financial Sector in the EU and the United States. ECB Occasional Paper 117/July 2010.

Sun, T., 2011. Identifying Vulnerabilities in Systemically-important Financial Institutions in a Macrofinancial Linkages Framework. IMF Working Paper 11/111.

Thomson, J., 1992. Modeling the regulator's closure option: a two-step logit regression approach. Journal of Financial Services Research 6, 5–23.

van den Berg, J., Candelon, B., Urbain, J., 2008. A cautious note on the use of panel models to predict financial crises. Economics Letters 101, 80–83.

Whalen, G., 1991. A proportional hazards model of bank failure: an examination of its usefulness as an early warning tool. Federal Reserve Bank of Cleveland Economic Review Q1, 21–31.