EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Jan Hannes Lang, Tuomas A. Peltonen,
Peter Sarlin

A framework for early-warning modeling with an application to banks

No 2182 / October 2018

Abstract

This paper proposes a framework for deriving early-warning models with optimal out-of-sample forecasting properties and applies it to predicting distress in European banks. The main contributions of the paper are threefold. First, the paper introduces a conceptual framework to guide the process of building early-warning models, which highlights and structures the numerous complex choices that the modeler needs to make. Second, the paper proposes a flexible modeling solution to the conceptual framework that supports model selection in real-time. Specifically, our proposed solution is to combine the loss function approach to evaluate early-warning models with regularized logistic regression and cross-validation to find a model specification with optimal real-time out-of-sample forecasting properties. Third, the paper illustrates how the modeling framework can be used in analysis supporting both micro- and macro-prudential policy by applying it to a large dataset of EU banks and showing some examples of early-warning model visualizations.

Keywords: Early-warning models; Financial crises; Bank distress; Regularization; Micro- and macro-prudential analysis

JEL classification: G01, G17, G21, G33, C52, C54

Non-Technical Summary

The recent global financial crisis highlighted the large costs for societies that the unravelling of macro-financial imbalances can have. The corresponding policy response in the aftermath of the financial crisis has been to strengthen micro-prudential regulation of financial institutions and provide new macro-prudential mandates and tools to competent authorities with the aim of dampening financial cycles and making the financial system more resilient to adverse shocks. A key issue for implementing macro-prudential policy is to identify the build-up of macro-financial vulnerabilities with a sufficient lead time so that policy action can still be effective in preventing severe financial crises. The policy interest in so-called early-warning models has therefore increased considerably in recent years, especially in the context of guiding the activation of macro-prudential policy tools (see for example Detken et al. (2014)). At the same time the academic interest in early-warning models has also increased considerably recently, as various papers have shown that there indeed seem to be common patterns in the data that often precede financial crises (see for example Borio and Lowe (2004) or Reinhart and Rogoff (2008)).

Despite many previous efforts, building an early-warning model is a complex task that involves numerous assumptions and practical choices that need to be made. For instance, the purpose of the model, i.e. whether it is used to predict potential future crises or whether it is used to understand past crises episodes, should guide several choices that need to be made regarding model complexity, model evaluation and real-time information lags. Similarly, depending on the ultimate policy use of a model, be it for instance to try and dampen the financial cycle or to increase resilience to already existing imbalances, the choice regarding the relevant forecast horizon would probably need to differ. The complexity of building an early-warning model is indeed well reflected in the various recent contributions to the early-warning literature that employ a multitude of econometric methods, prediction horizons, evaluation approaches and datasets.[1]

With this background in mind, the paper extends the literature in the following ways. First, we propose a conceptual framework to guide the building of early-warning models, highlighting the key assumptions and decisions that need to be made, and linking them to the various approaches in the existing literature. Second, we propose a flexible modeling solution to the conceptual early-warning framework

[1]For some recent contributions to the early-warning literature see for example Borio and Drehmann (2009), Alessi and Detken (2011), Lo Duca and Peltonen (2013), Behn et al. (2013), Betz et al. (2014), Alessi and Detken (2014) and Holopainen and Sarlin (2015). A detailed discussion of the various approaches found in the literature is contained in Section 2.

that facilitates model selection in real-time for forecasting purposes. Specifically, our proposed solution is to combine the loss function approach to evaluate early-warning models with regularized logistic regression and cross-validation to find a model specification with optimal real-time out-of-sample forecasting properties.[2] The third contribution of the paper is to illustrate how the modeling framework can support both micro- and macro-prudential policy, by applying it to a large dataset of EU banks with the aim to predict bank distress. We show that a parsimonious model with only 11 risk drivers has good in-sample and out-of-sample signalling properties for bank distress events with a lead time of 1-8 quarters, and illustrate how model output can support policy analysis by providing some examples of early-warning model visualizations.

This paper therefore adds to the existing literature both from a conceptual and an empirical point of view. One of the advantages of our bank-level early-warning model is that it allows for the analysis of the build-up of vulnerabilities at the micro and the macro level. The model can therefore be used for the analysis of systemic risk in both the cross-sectional and the time dimension. More specifically, the model can be used to identify vulnerabilities at a given point in time for systemically important institutions, as well as the build-up of banking-sector vulnerabilities over time at the country or regional level. For the analysis of the build-up of vulnerabilities over time an aggregation method is proposed, while for both the cross-sectional and time dimension of systemic risk, a decomposition procedure of vulnerabilities into contributing factors is proposed that adds additional value for policy purposes, as it allows to identify at a high level in which areas possible vulnerabilities are emerging. Concrete examples of how model output can be used for risk-identification in the macro-prudential policy process are provided throughout the paper.

[2]For work on the evaluation of early-warning models, see Kaminsky et al. (1998), Demirgüç-Kunt and Detragiache (2000), Alessi and Detken (2011) and Sarlin (2013b). For regularized logistic regression, we make use of the LASSO (Least Absolute Shrinkage and Selection Operator) approach introduced by Tibshirani (1996).

1 Introduction

The recent global financial crisis highlighted the large costs for societies that the unravelling of macro-financial imbalances can have. The corresponding policy response in the aftermath of the financial crisis has been to strengthen micro-prudential regulation of financial institutions and provide new macro-prudential mandates and tools to competent authorities with the aim of dampening financial cycles and making the financial system more resilient to adverse shocks. A key issue for implementing macro-prudential policy is to identify the build-up of macro-financial vulnerabilities with a sufficient lead time so that policy action can still be effective in preventing severe financial crises. The policy interest in so-called early-warning models has therefore increased considerably in recent years, especially in the context of guiding the activation of macro-prudential policy tools (see for example Detken et al. (2014)). At the same time the academic interest in early-warning models has also increased considerably recently, as various papers have shown that there indeed seem to be common patterns in the data that often precede financial crises (see for example Borio and Lowe (2004) or Reinhart and Rogoff (2008)).

Despite many previous efforts, building an early-warning model is a complex task that involves numerous assumptions and practical choices that need to be made. For instance, the purpose of the model, i.e. whether it is used to predict potential future crises or whether it is used to understand past crises episodes, should guide several choices that need to be made regarding model complexity, model evaluation and real-time information lags. Similarly, depending on the ultimate policy use of a model, be it for instance to try and dampen the financial cycle or to increase resilience to already existing imbalances, the choice regarding the relevant forecast horizon would probably need to differ. The complexity of building an early-warning model is indeed well reflected in the various recent contributions to the early-warning literature that employ a multitude of econometric methods, prediction horizons, evaluation approaches and datasets.[3]

With this background in mind, the paper extends the literature in the following ways. First, we propose a conceptual framework to guide the building of early-warning models, highlighting the key assumptions and decisions that need to be made, and linking them to the various approaches in the existing literature. Second, we propose a flexible modeling solution to the conceptual early-warning framework

[3]For some recent contributions to the early-warning literature see for example Borio and Drehmann (2009), Alessi and Detken (2011), Lo Duca and Peltonen (2013), Behn et al. (2013), Betz et al. (2014), Alessi and Detken (2014) and Holopainen and Sarlin (2015). A detailed discussion of the various approaches found in the literature is contained in Section 2.

that facilitates model selection in real-time for forecasting purposes. Specifically, our proposed solution is to combine the loss function approach to evaluate early-warning models with regularized logistic regression and cross-validation to find a model specification with optimal real-time out-of-sample forecasting properties.[4] The third contribution of the paper is to illustrate how the modeling framework can support both micro- and macro-prudential policy, by applying it to a large dataset of EU banks with the aim to predict bank distress. We show that a parsimonious model with only 11 risk drivers has good in-sample and out-of-sample signalling properties for bank distress events with a lead time of 1-8 quarters, and illustrate how model output can support policy analysis by providing some examples of early-warning model visualizations.

The remainder of the paper is structured into four parts. Section 2 presents the conceptual framework for building early-warning models, while Section 3 presents the proposed modeling solution to the early-warning framework. In Section 4, the modeling framework is applied to build a model for predicting distress in EU banks. Finally, Section 5 concludes.

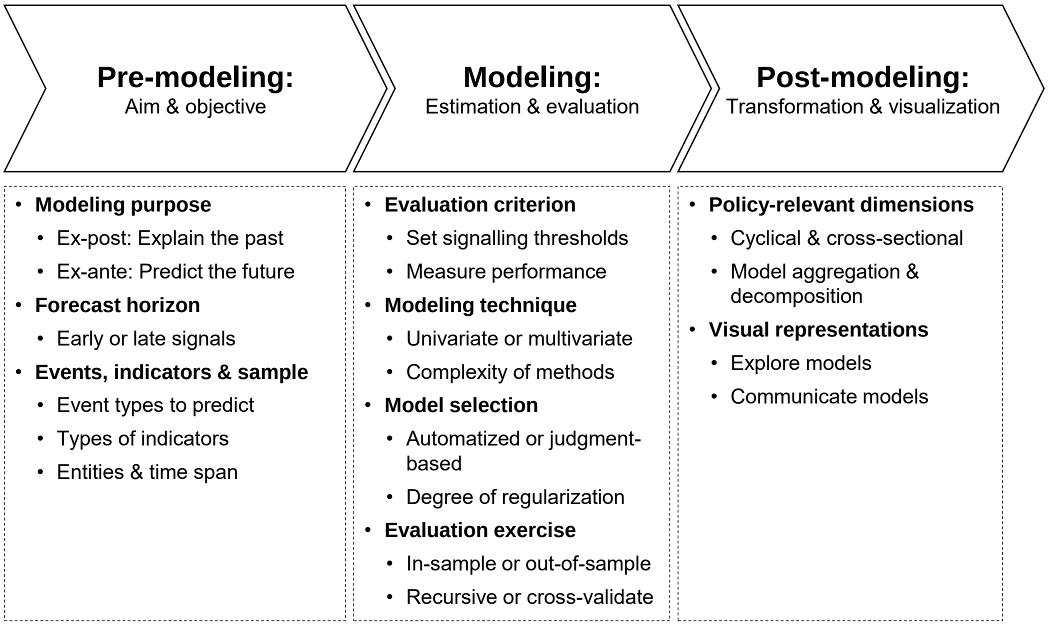
2 A conceptual framework for early-warning modeling

The basic principle in early-warning modeling is to distinguish underlying vulnerabilities which make crises more probable from potential triggers that cause the actual materialisation of crises. By analyzing common patterns in the data prior to historical crisis episodes, it is (to some extent) possible to identify imbalances that could lead to a crisis given a suitable trigger, i.e. to identify vulnerable states. In contrast, the exact timing of a trigger leading to the unravelling of imbalances is much more difficult or even impossible to predict with high precision. Therefore, early-warning models are generally concerned with identifying vulnerable states prior to financial crises, which can also be viewed as a standard two-class classification task, where the key objective is to separate the vulnerable from non-vulnerable states.

Despite being a standard modeling task, the derivation of an early-warning model requires a large number of underlying modeling choices and free parameters to be specified. This complexity of choices is well represented in the varying approaches

[4]For work on the evaluation of early-warning models, see Kaminsky et al. (1998), Demirgüç-Kunt and Detragiache (2000), Alessi and Detken (2011) and Sarlin (2013b). For regularized logistic regression, we make use of the LASSO (Least Absolute Shrinkage and Selection Operator) approach introduced by Tibshirani (1996).

Figure 1: Overview of the conceptual early-warning framework



to early-warning modeling that have been employed in the literature to date.[5] In order to structure these complex choices and make the underlying assumptions for the early-warning model explicit, a conceptual framework is hence called for. Rather than an all-encompassing guidebook for setting the parameters of an early-warning exercise, our aim is to illustrate and structure the numerous choices that need to be made when building early-warning models.

At the highest level, building an early-warning model can be described as a sequential process that involves three steps: pre-modeling, modeling and post-modeling. These three steps include decisions on (i) the model aims and objectives, (ii) the model estimation and evaluation, and (iii) the appropriate representations of model output. A schematic overview of the the conceptual framework is illustrated in Figure 1, while the following subsections illustrate each of these steps in more detail by relating them to the existing literature.

2.1 Pre-modeling: Aim and objective

The starting point for building any type of early-warning model lies in the ultimate aim and objective the model is supposed to serve. This involves decisions about whether the model is mainly used for explanatory or forecasting purposes, the desired prediction horizon, and the relevant entities, risk indicators and event types for

[5]The various approaches are described in detail in the following subsections.

building the model.

2.1.1 Modeling purpose

A decisive question when deriving an early-warning model relates to the general task it is supposed to solve. The main choice in this respect is whether to focus on explaining the past (ex-post analysis) or on predicting the future (ex-ante analysis). As pointed out in the famous article by Shmueli (2010), distinguishing between statistical modeling for causal explanation and for prediction has a large impact on decisions at each step of the statistical modeling process. Depending on the chosen purpose, models are to be derived with an entirely different objective function. For explaining the past, we are mainly concerned with maximizing fit to historical data (in-sample analysis), whereas models for predictive purposes should optimize performance on future data (out-of-sample analysis). The fundamental decision on the modeling purpose therefore impacts many of the other decisions that need to be made when building an early-warning model. For example, the appropriate modeling technique (see section 2.2.2), as well as the method for model selection (see Section 2.2.3) and the appropriate evaluation exercise (see Section 2.2.4) should all be heavily influenced by the decision on whether the purpose of the model is causal analysis or pure forecasting. Moreover, the desirability of using lagged data to replicate real-time information availability should depend on whether the modelling purpose is to explain the past or to use the model to predict future events (in real-time).

2.1.2 Forecast horizon

Early-warning models aim at signaling distress events early on, but the specific forecast horizon will depend on the application at hand. Even though the forecast horizon could be treated as a parameter to be optimized (as e.g. in Bussiere and Fratzscher (2006)), we argue that it ought to be specified as an ex-ante decision to support the task at hand rather than optimizing the fit to data. While early studies on currency crises used an 8-quarter horizon (Kaminsky et al., 1998), Lo Duca and Peltonen (2013) used a 6-quarter horizon in the context of predicting systemic events. Longer horizons of 5 to 12 quarters (Behn et al., 2013) and even 5 to 16 quarters (Detken et al., 2014) have been applied in recent contributions focusing on banking crises. While macro-prudential purposes clearly require long horizons[6], there is no consensus on one correct horizon, which also becomes apparent from the common

[6] E.g. the CRD IV regulation in the EU specifies a 12-month implementation period for the Counter-cyclical Capital Buffer

testing of multiple horizons (e.g., Schudel (2013); Behn et al. (2013); Lainà et al. (2015)). On the other hand, micro-prudential purposes may require much shorter forecast horizons, as the risks of concern are more concurrent in nature (e.g., up to eight quarters ahead as in Gropp et al. (2006) and Betz et al. (2014)).

2.1.3 Events, indicators and sample

Once the modeling purpose and forecast horizon have been specified, the next key decisions relate to data, which can have a large impact on the modeling outcome. These relate to the type of distress events that are supposed to be predicted, the pre-selection of the relevant risk indicators, the choice of relevant entities to include in the cross-section, as well as the appropriate time sample.

First, it is essential to define the crisis events in a manner that reflects the vulnerabilities that we are interested in modeling. For instance, when using a financial stress index, a key choice is to specify a threshold above which we define periods to be crisis events (e.g., mean exceedance above three standard deviations in Kaminsky et al. (1998) or the 90th country-specific percentile in Lo Duca and Peltonen (2013)). Moreover, multiple approaches also exist when it comes to defining binary crisis variables based on rules or expert-judgement (e.g. the BIS initiative (e.g. Borio and Lowe (2002) and Borio and Drehmann (2009)), the IMF initiative (Laeven and Valencia, 2012) or the ESCB initiative (Babecky et al., 2012)). Likewise, one needs to define the type of failure one is concerned with when assessing bank distress, such as direct failures, state-aid cases and forced mergers as in Betz et al. (2014) or news-based events as in Poghosyan and Cihák (2011).

Another key decision relates to the pre-selection of potential risk indicators, as well as relevant observations across entities and over time. This step will affect the modeling outcome to a large extent, as the potential early-warning indicators often have varying coverage and as such variable pre-selection can affect the sample heavily (See for example tables A2 - A3 in Detken et al. (2014) for varying data availability across different early warning indicators in EU countries). The trade-off between the number of variables and number of observations descends from the fact that most multivariate modeling approaches cannot process observations with missing values. This also highlights the fact that new indicators cannot be included in some models if their time series dimension is too short. Hence, there is a two-way feedback process whereby the decisions on indicators and the desired entity and time coverage should affect each other.

2.2 Modeling: Estimation and evaluation

Once the aim of the early-warning exercise is defined, the next step is to set-up an approach for modeling and evaluation. This involves the steps of defining an evaluation criterion, deciding on a modeling technique, selecting an optimal model complexity and specification, and setting up an evaluation exercise.

2.2.1 Evaluation criterion

One of the key modeling decisions for any early-warning exercise relates to the measurement of its performance or goodness of fit, independent of whether the objective is explanatory or predictive. Most approaches to evaluating early-warning models proposed in the literature are based on the notion that in a two-class classification problem, a model may conduct Type I errors (miss crisis events) or Type II errors (issue false alarms). Both types of errors can be assumed to be costly and the various evaluation approaches in the literature mainly differ regarding the assumptions about the relevant trade-off between these types of errors.

The first early-warning models based upon the signaling approach used the noise-to-signal ratio to evaluate models and set signalling thresholds (Kaminsky et al., 1998). The noise-to-signal ratio measures the ratio of false alarms to the share of crises that are correctly predicted. In recent years, methods dating back to signal detection in World War II, such as the Receiver Operating Characteristics (ROC) curve and the area under the ROC curve (AUC), have been used to measure classification performance of early-warning models[7] (e.g. Peltonen (2006), Marghescu et al. (2010); Jorda and Taylor (2011)). While the ROC curve and AUC provide performance measures summarizing all possible signalling thresholds, a recent variation of this approach limits the relevant parts of the AUC using so-called partial AUCs (Detken et al., 2014). Moreover, Peltonen (2006) introduced other goodness of fit measures, such as Cramer's V, Brier Quadratic Probability Score, and Spiegelhalter's z-statistic in the early-warning context, which was further extended in a later overview by Candelon et al. (2012).

An important contribution by Demirgüç-Kunt and Detragiache (2000) was to introduce a loss function based upon type I and type II errors and preferences between them, while Alessi and Detken (2011) put forward a further refined usefulness measure to evaluate the model. This framework was extended by Sarlin (2013b) to a loss function that does not assume equal class distributions and to incorporate a

[7]The ROC curve traces out all combinations of true positive and false positive rates that are associated with all possible signalling thresholds.

more interpretable relative usefulness measure. Further, in the vein of Drehmann and Juselius (2014), one might also include other performance-related factors in the evaluation of early-warning models, such as the timing and stability of signals.

2.2.2 Modeling technique

The task of an early-warning model is nothing else than a text book example of two-class classification, as defined in machine learning (e.g., Ripley (1996), Bishop (2006)). The literature has applied a range of methods for the task of classifying between vulnerable and tranquil periods, some of the methods descending from classical statistics and others from the later strand of machine learning approaches. As proposed by Holopainen and Sarlin (2015), we can categorize early-warning modeling techniques into four subgroups that vary in their complexity.

First, the most common group of modeling techniques relies on the covariance matrix to estimate model coefficients for belonging to one class, such as linear discriminant analysis and logistic regression (Demirgüc-Kunt and Detragiache, 1998; Berg and Pattillo, 1999; Schularick and Taylor, 2012; Lo Duca and Peltonen, 2013; Behn et al., 2013). Second, a large number of methods rely directly on the contingency matrix that classifies observations into true positives, false positives, true negatives and false negatives, such as signal extraction (Kaminsky et al., 1998; Borio and Lowe, 2002, 2004; Borio and Drehmann, 2009; Alessi and Detken, 2011), decision trees or random forests (Alessi and Detken, 2014) and naive Bayes. The third group includes approaches based upon similarity functions, such as k-nearest neighbors, while the fourth category includes various other, more complex, ways for estimating the likelihood of belonging to one class or the other, including Artificial Neural Networks (e.g. Peltonen (2006); Sarlin (2013a)) and Support Vector Machines. As Holopainen and Sarlin (2015) discuss in relation to their results from comparing all above approaches, there is no single best modeling technique. The modeling technique of choice should be influenced to a large extent by the purpose that the model is supposed to serve, as set out in Section 2.1.1.

2.2.3 Model selection and complexity

Once a modeling techniques has been chosen, decisions on how the "best" model specification should be selected and how complex such an optimal model should be need to be made. The purpose that the model is supposed to serve, as set out in Section 2.1.1, should be again a key factor for these decisions. Purely predictive purposes will mostly require objective model-selection procedures (such as Bayesian model averaging or LASSO with cross-validation) and could possibly feature less explanatory variables that are however robust for real-time out-of-sample predictive purposes. On the other hand, causal analysis will most likely require more complex models so as to control for all possible explanatory variables and avoid any omitted variables bias. Hence, the model-selection procedure is also likely to be more subjective and incorporates more expert judgement.

However, the dividing lines between these approaches are not always clear, and depending on the policy context for which the model is designed, one may still have preferences regarding model complexity / interpretability and regarding either automatic or expert-driven model selection procedures. Hence, in line with the discussion in Drehmann and Juselius (2014), we treat model complexity and model selection as a preference related to parsimony and interpretability of the model, which will always be specific to the context at hand.

2.2.4 Evaluation exercise

Depending on the ultimate purpose of the early-warning model, an evaluation exercise based on in-sample data (for inferential analysis) or out-of-sample data (for predictive purposes) might be more appropriate. In the latter case various approaches are conceivable including splitting the sample, recursive out-of-sample evaluations or cross-validation. For example the approach by Kaminsky et al. (1998) and Borio and Drehmann (2009) is to split the data into two sub-samples with a specific cut-off point that usually occurs prior to an important cluster of crisis events. While the common approach is to use the earlier sample for estimation and the latter for evaluation, one could in the same vein also estimate models on the later sub-sample and evaluate on the earlier one (e.g. Behn et al. (2013); Lainà et al. (2015)) or leave out specific entities from the estimation sample altogether.

The sample can also be randomly split into many sub-samples or so-called folds, where the model is then repeatedly estimated on all but one fold and predictions are made for the fold left out. This method is commonly referred to as cross-validation (e.g., Sarlin (2013a); Holopainen and Sarlin (2015)) and becomes particularly relevant when making use of more complex non-linear function approximators that may be prone to overfitting the underlying data. Another out-of-sample evaluation approach is to mimic real-time modeling by performing recursive estimations and predictions by only using the information set that would have been available at the time (e.g., Betz et al. (2014)).

2.3 Post-modeling: Transformation and visualization

Once a suitable early warning model has been specified and estimated, it is of key importance to decide on how to best analyze, transform, represent and visualize model output in order to support interpretation and communication of results. This is particularly important for policy use of early-warning models. Key concerns ought to be the policy-relevant dimensions to focus on and how model output is to be visualized.

2.3.1 Policy-relevant dimensions

Generally, we can divide the use of early-warning models according to the cross-sectional and cyclical dimensions of systemic risk. The cross-sectional dimension relates to the distribution of vulnerabilities at given a point in time, while the cyclical dimension relates to the build-up of aggregate risk over time. In the former case we would be mainly interested in an overall ranking of vulnerable entities at a specific point in time or the clustering of vulnerabilities within a sub-group of entities, while in the latter case the evolution of suitable aggregate model output over time would be of greater interest. Moreover, one might also be concerned with a network perspective that links the vulnerability of one entity to other interlinked entities for the cross-sectional dimension of systemic risk (e.g., Peltonen et al. (2015)).

Depending on the granularity of the underlying entity data and the ultimate policy-use of the model, various aggregations and transformations of model output are therefore possible, such as over countries or firm-level entities, to support either the analysis of cross-sectional or cyclical systemic risk. Likewise, when using multi-variate techniques, a decomposition of vulnerabilities into more granular risk-driver categories can be useful to better understand the driving factors of vulnerabilities.

2.3.2 Model visualization

Once the relevant dimensions of the model and suitable transformations are decided, their visualisation constitutes a final, albeit still important step. As reviewed by Flood et al. (2014) and Sarlin (2014), visualization in general and visual analytics in particular constitute natural aids for monitoring systemic risk and financial stability. For example Dattels et al. (2010) disentangle the sources of risk by a mapping of six composite indices with a spider chart visualization. Similarly, Sarlin and Peltonen (2013) use a Financial Stability Map that projects the state of financial stability from high-dimensional indicators to a two-dimensional map.

With more standard graphs, the cross-sectional dimension is natural to visualize as bar charts, while the cyclical dimension can be represented with time-series plots of model output. An additional concept to be considered is the uncertainty in estimated probabilities and thresholds through confidence bands as introduced by Holopainen and Sarlin (2015). Further, if non-linear models are used, we might also need specific visualization techniques for interpreting contributing factors. For instance, Support Vector Machines for supervised learning algorithms and Self-Organizing Maps for unsupervised learning algorithms require additional means for representing model output, such as representations of classifier hyperplanes in high-dimensional data. Moreover, in the vein of the VisRisk platform for visual systemic risk analytics (Sarlin, 2014), the visualization of early-warning models might also require interactive means for better accessing the available data and model output.

3 A modeling solution to the early-warning framework

This section provides a flexible modeling solution to the conceptual early-warning framework described above. In particular, we propose a model selection procedure based on regularization techniques in combination with cross-validation and the loss function approach to evaluate early-warning models, that facilitates the selection of optimal out-of-sample forecasting models in real-time. This modeling solution still leaves important decisions to the modeler, such as the desired forecast horizon, the events to predict or the set of possible explanatory variables to consider. However, the proposed modeling solution provides a number of default choices for key decisions that need to be made when building an early-warning model, as highlighted in Section 2. For example, the modeling solution is mainly designed for prediction purposes, the loss function specification by Sarlin (2013b) is used to set signalling thresholds and evaluate models, the modeling technique and model selection procedure is the logistic LASSO with cross-validation, the evaluation of models is done through a recursive real-time out-of-sample exercise and an approach to aggregation and decomposition of model output is put forward. Each of the building blocks of the modeling solution are described in greater detail in the following subsections.

3.1 The policymaker's loss function

As set out in Section 2.2.1, an essential part of building an early-warning model is to decide on an evaluation criterion to measure its performance and to select an optimal signalling threshold that allows to classify observations into vulnerable and

Table 1: Contingency matrix for model signals and crises

	Crisis	**No Crisis**
Signal	True Positive (TP)	False Positive (FP)
No Signal	False Negative (FN)	True Negative (TN)

non-vulnerable states. In what follows, we use the loss function approach proposed by Sarlin (2013b) in order to set optimal signalling thresholds and measure the usefulness of a model for a policymaker who is concerned about missing vulnerable states as well as issuing false alarms about underlying vulnerabilities.

Let us represent the occurrence of a crisis with a binary state variable $I_{i,t} \in \{0,1\}$, where i refers to a given entity and t to a given time period. Yet, in order to focus on vulnerabilities prior to crises, we are usually concerned with pre-crisis periods $I^h_{i,t} \in \{0,1\}$ when building early-warning models, where h indicates the relevant forecast horizon to define pre-crisis episodes. The pre-crisis indicator is equal to one in all of the h time periods up to and including a given crisis event and zero otherwise. Formally we can represent this as $I^h_{i,t} = 1,\ if\ I_{i,t+s} = 1,\ \forall\, 0 \leq s \leq h$. Further, let $p_{i,t}$ be a univariate risk driver or an estimated probability of being in a vulnerable state. In order to be able to classify observations into vulnerable and non-vulnerable states, $p_{i,t}$ needs to be turned into a binary signal $P_{i,t} \in \{0,1\}$ that equals one if $p_{i,t}$ exceeds a specified threshold τ and zero otherwise. This allows us to assess the correspondence between $P_{i,t}$ and $I^h_{i,t}$ using a so-called contingency matrix, which is displayed in Table 1.

Given a forecast horizon h, a policymaker ought to choose a threshold τ for probabilities $p_{i,t}$ of a model so as to minimize her loss. In a two-class classification problem, she can be assumed to be concerned about conducting two types of errors: issuing false alarms and missing pre-crisis periods. Type I error rates represent the proportion of missed pre-crisis periods relative to the total number of pre-crisis periods in the sample ($T_1(\tau) = FN/(TP + FN) \in [0,1]$), while type II error rates represent the proportion of false alarms relative to the number of tranquil periods in the sample ($T_2(\tau) = FP/(FP+TN) \in [0,1]$). We compute the loss of a policymaker as a weighted average of T_1 and T_2 according to her relative preferences μ between missing crises and issuing false alarms. However, as weighting only with relative preferences does not consider imbalances in class size, and the errors are defined in relation to the size of each class, we also account for unconditional probabilities of pre-crisis events ($P_1 = \Pr(I^h_{i,t} = 1)$) and tranquil periods ($P_2 = \Pr(I^h_{i,t} = 0)$). Thus, the policymaker's loss function can be written as:

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

With this definition of the loss function we can compute the usefulness of a given model and signalling threshold to a policymaker. The policymaker could achieve a loss of μP_1 by never issuing a crisis signal or a loss of $(1 - \mu)P_2$ by always issuing a crisis signal. Thus, the loss equals $\min[\mu P_1, (1 - \mu)P_2]$ when ignoring the model. The absolute usefulness U_a of a model can therefore be computed by subtracting the policymaker's loss when using the model from the loss achieved when the model is ignored, similar in spirit to Alessi and Detken (2011):

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

In addition, the relative usefulness U_r of a model can be defined by putting the absolute usefulness in relation to the maximum available usefulness. It reports U_a as a percentage of the usefulness that a policymaker would gain with a perfectly performing model and provides a means for representing usefulness in relative terms which should be easier to interpret:[8]

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

For a given policy preference parameter there will be a corresponding optimal signalling threshold $\tau^*(\mu)$ that is obtained by minimizing the loss function above. Let the policymaker's loss evaluated at the optimal signalling threshold be denoted by $L^*(\mu)$. The corresponding maximum possible usefulness measures for a given model and preference parameter can therefore be denoted as $U_a^*(\mu)$ and $U_r^*(\mu)$.

3.2 Logistic LASSO with cross-validation

Once an evaluation criterion has been decided upon, further key decisions that need to be made when building an early-warning model relate to the choice of modeling technique and procedure for model selection, as set out in Section 2.2.3. Here we propose an integrated approach where the chosen modeling technique allows for an easy implementation of automatized model selection. Specifically, we propose to

[8]Beyond usefulness measures, it can also be beneficial to compute other performance metrics such as the Noise-to-Signal ratio or the Area Under the Receiver Operating Characteristics Curve (AUROC), even if these measures are not used for model selection and evaluation purposes or to set optimal signalling thresholds.

use the logistic LASSO method[9] as the modeling technique in combination with cross-validation to set the penalty parameter that determines the complexity of the model. While the logistic LASSO method allows for a simultaneous variable selection and classifier estimation, the cross-validation is in effect an automatic model selection device that helps to build models with optimal out-of-sample forecasting properties.[10] This integrated framework for model selection is described in greater detail in the following paragraphs.

A common approach to model selection is to tap into regularization methods. Their particular aim is to prevent overfitting, which is most often accomplished by penalizing models with extreme parameter values. Instead of an estimation minimizing a loss function $E(X, Y)$, the most general cases of L_1 and L_2 regularization refer to a penalization of the loss function through $E(X, Y) + \lambda|\beta|$, where β are the estimated coefficients, $|\cdot|$ is the L_1 norm or the squared L_2 norm, and λ is a free parameter specifying the size of penalization. The key difference between L_1 and L_2 norms is that the former shrinks variables towards zero, giving sparse estimates. The size of penalization is commonly specified empirically, such as through AIC, BIC or cross-validation. For the case of standard linear regression, the approaches described above would result in Least Absolute Shrinkage and Selection Operator (LASSO) regression and ridge regression respectively. However, regularization is also widely used in other approaches, such as artificial neural networks (ANNs) and support vector machines (SVMs) among many others. For instance, the approach used in ANNs is most often denoted as weight decay.

Out of the two most common norms, we make use of L_1 regularization in order to produce sparse models, and thus to also use the modeling technique as a simultaneous variable selection device. Furthermore, as we are dealing with a two-class classification problem, we need to make use of binary-choice methods instead of linear regression. Hence, from the family of regularization techniques we make use of the logistic LASSO regression (Tibshirani, 1996). An additional argument for using the LASSO method is that it provides a means for considering interaction terms in models in a hierarchical manner, which is computationally less costly (Bien et al., 2013). In our framework, the probability of being in a vulnerable state is modeled via the following logit model, where x is a vector of risk drivers plus an intercept and β is a coefficient vector of size $q + 1$:

[9] LASSO is the acronym for Least Absolute Shrinkage and Selection Operator (Tibshirani, 1996).

[10] However, the modeling solution can also be used to understand risk drivers for past crisis episodes, for example if the optimal penalty parameter for the LASSO is not set via cross-validation, as described in Section 3.2, but in a way that allows for the inclusion of the most important variables to explain the data in-sample.

$$\Pr\left(I_{i,t}^{h}=1|X_{i,t}=x_{i,t}\right)=p_{i,t}=\frac{e^{\beta' x_{i,t}}}{1+e^{\beta' x_{i,t}}} \tag{4}$$

Instead of simply maximizing the usual log likelihood to estimate the parameters for this model, the logistic LASSO performs the maximization step with an added penalty term that depends on the coefficient estimates. Formally, the following penalized negative binomial log-likelihood is minimized, where λ is the LASSO penalty parameter that determines the complexity of the model:[11]

$$\min_{\beta\in\mathbb{R}^{q+1}} -\left[\frac{1}{N\cdot T}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{i,t}\cdot(\beta' x_{i,t})-\log\left(1+e^{\beta' x_{i,t}}\right)\right]+\lambda\left\|\beta\right\|_1 \tag{5}$$

A key question is of course how to choose the LASSO penalty parameter in practice. In effect, each value of the penalty parameter will yield a different model so that the choice of the penalty parameter boils down to selecting a particular model specification. Given that our proposed modeling solution focuses on early-warning models for prediction purposes, the parameter should be chosen so as to maximize the model's forecasting performance. In this context, the resampling method of cross-validation (Stone, 1977) is commonly used in machine learning to assess the generalization performance of a model on out-of-sample data and to prevent overfitting on in-sample data. Hence, for model selection purposes and to obtain an optimal λ we propose to use cross-validation with the relative usefulness of equation 3 as the relevant performance measure. The scheme that we use is commonly referred to as K-fold cross-validation and functions as follows:

1. Randomly split the set of observations into K folds of approximately equal size.

2. For the k^{th} out-of-sample validation fold, fit a model on the remaining $K-1$ folds, also called the in-sample data, and compute an optimal signalling threshold τ_{-k}^{*} by maximizing the relative usefulness measure $U_r^{-k}(\mu,\tau)$ on all but the k^{th} fold.

3. Apply the optimal in-sample threshold to the observations contained in the k^{th} fold and compute its out-of-sample usefulness $U_r^{k}(\mu,\tau_{-k}^{*})$.

4. Repeat Steps 1, 2 and 3 for $k=1,2,...,K$, and collect the out-of-sample performance measures for all K validation sets as $U_r^{K}(\mu)=\frac{1}{K}\sum_{k=1}^{K}U_r^{k}(\mu,\tau_{-k}^{*})$.

[11]The higher the penalty parameter that is applied, the more coefficient estimates will be shrunk towards zero and the fewer variables will be included in the model.

This cross-validation exercise is then performed for all possible values of the LASSO penalty parameter. The optimal penalty parameter is then the one that is associated with the highest cross-validated out-of-sample relative usefulness. Equipped with this optimal penalty parameter it is straightforward to estimate the logistic LASSO model and obtain an automatically selected model specification with optimal out-of-sample forecasting properties.

3.3 Recursive out-of-sample evaluation

While the logistic LASSO in combination with cross-validation helps to select an optimal forecasting model, we are often also interested in how a model would have performed in real-time. Hence, an evaluation exercise that accounts for features of real-time modeling that a policymaker faces in reality is extremely useful. We therefore follow Betz et al. (2014) in performing a real-time recursive out-of-sample exercise using in each time period only the information set that would have been available to a policymaker at the time. Yet, beyond only accounting for the use of historical data in an expanding-window fashion and publication lags in the data used, the logistic LASSO method also allows to re-estimate optimal models in each period. This provides a true real-time exercise, as the model specification can vary over time and is not fixed ex-ante based on full-sample information.

The early-warning model evaluation exercise can therefore be expressed as a recursive logistic LASSO regression that makes a prediction in each time period $t = 1, 2, ..., T$ with an estimation sample that grows in an expanding-window fashion and functions according to the following steps:

1. Estimate the LASSO model on in-sample data using the information set that would have been available up to period $t - 1$.

2. Collect model probabilities p for the in-sample period and compute the usefulness for all thresholds $\tau \in [0, 1]$.

3. Choose the τ^* that maximizes in-sample Usefulness, estimate distress probabilities p for the out-of-sample data (period t), apply τ^* 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. Repeat this as long as $t \leq T$.

The above algorithm hence estimates a logistic LASSO model with a varying model specification in each time period t using all available information up to the previous period. Then, model probabilities are used to set an optimal threshold τ^* on in-sample data, and provide an estimate of the current vulnerability of each entity by applying τ^* to the estimated probability. This is repeated in a recursive fashion. Hence, the estimation sample changes as the window length expands and the out-of-sample data is also treated in a rolling-window fashion (i.e. one period at a time). These recursive changes in in-sample and out-of-sample data enable not only to test model performance in real-time use, but allows also models to adapt over time.

The recursive out-of-sample evaluation exercise with a model specification that is allowed to change constantly over time is mainly meant to establish that the proposed early warning model would have worked without using information from the future. This evaluation exercise is therefore even more challenging than choosing a model specification based on full-sample information and evaluating this model by a recursive out-of-sample exercise. For actual policy use of an early warning model, a constantly changing model specification would probably not be desirable. A re-estimation of the logit LASSO model at only lower frequencies could therefore be beneficial in the context of using the proposed early warning framework for policy purposes.

3.4 Aggregation and decomposition of model output

Once an optimal model has been selected and evaluated it is important to decide how model output should be best analyzed and presented. This section puts forward two approaches to aggregate and decompose model output with the ultimate aim of increasing usefulness for policy purposes. First, the distress probability for each entity is decomposed into the contributions stemming from different risk-driver categories, in order to help a policymaker to determine which factors are driving the build-up of vulnerabilities at the entity-level. Second, the distress probabilities are aggregated across the cross-sectional dimension, which can help a policymaker to identify the build-up of systemic risk over time at the aggregate level. Examples of an aggregate may refer to country-level output for micro data or global output for macro data. Each of these analytical exercises are described in turn from a conceptual view point.

Given that the logistic function is non-linear in nature, a decomposition of distress probabilities into contributing factors is not trivial and there does not exist one unique or "correct" method for such a transformation. There nevertheless exists an intuitive way to decompose the probabilities that is similar in spirit to using marginal effects at the means of the variables. The proposed probability decomposition consists of three steps. First, a counterfactual probability for each factor is computed by assuming that all other factors are at their mean values. Second, the probability share of each factor is calculated as the ratio of the counterfactual probability of that factor to the sum of the counterfactual probabilities of all factors. Third, the probability shares of each factor are multiplied with the distress probability from the model to arrive at the respective probability contributions. If we define the logit function f as in equation 4, then the probability contribution of factor m for entity i at time t can be expressed in the following way:

$$P^c(x_{i,t}^m) = \frac{f(x_{i,t}^m | x_{i,t}^{-m} = E_{i,t}(x_{i,t}^{-m}))}{\sum_n f(x_{i,t}^n | x_{i,t}^{-n} = E_{i,t}(x_{i,t}^{-n}))} f(x_{i,t}) \tag{6}$$

For the aggregation of probabilities at the country or regional level, there are again various possible ways to proceed. Ideally, network connections between entities and interlinkages through common exposures should be taken into account in the aggregation procedure. However, due to data limitations, this might not always be a possible approach. The proposed feasible alternative for the probability aggregation is therefore to use the size of each entity in relation to the full sample. For instance, the share of each bank in the total assets of all banks that are part of the country or region for which the aggregation is desired, or stock-market capitalization if applied to country-level data. Such an aggregation is easy to implement and makes intuitive sense as long as we assume that systemic relevance of an entity-level distress event is increasing in its size. Formally, the distress probability for aggregate j at time t can be expressed as the weighted average of the distress probabilities of all of the N_j entities that are located in that aggregate:

$$ADP_{j,t} = \sum_{i=1}^{N_j} f(x_{i,t}) \frac{a_{i,t}}{\sum_k a_{k,t}} \tag{7}$$

The aim of these approaches for aggregating and decomposing model output is to support the use of early-warning models for policy purposes. In the remainder of the paper we illustrate how the entire proposed modeling solution can be applied in practice.

4 A micro-macro application to banks

In this section we apply our modeling framework to build an early-warning model for European banks that can be used to analyze the build-up of vulnerabilities at the micro (i.e. bank) as well as at the macro level (i.e. country or region). The overarching aim of the model that we construct is to identify the build-up of vulnerabilities with a sufficient lead time and in real-time so as to be useful for micro- and macro-prudential policy purposes. This overarching goal will guide many of the modeling choices outlined in the conceptual framework of Section 2.

The application is meant to illustrate the usefulness and flexibility of the modeling solution that was put forward in section 3, as well as the need to make the modeling choices presented in section 2 explicit. For example, it is shown in section 4.5 that the optimal model specification for the application at hand depends on the policy preferences between type I and type II errors, as well as on the chosen forecast horizon. Different aims and objectives for the final early warning model should therefore lead to different optimal early warning models. The proposed modeling solution offers one way of dealing with this multitude of optimal models in a flexible way through the use of automated model selection techniques. Hence, compared to existing bank-level early warning models such as Betz et al. (2014) the main contribution of this part of the paper is to illustrate that the set of optimal predictors for bank distress events depends crucially on the aim and objective of the early warning model.

4.1 Overview of the dataset

The dataset that we employ for our empirical application of the proposed modeling framework consists of a large unbalanced panel of EU banks covering the period 1999Q1 - 2014Q4 and is based on publicly available information only. The three main building blocks of the dataset are i) a collection of bank-level distress events; ii) a large number of bank-specific variables derived from publicly available financial statements; iii) various country-level macro-financial indicators and aggregate banking-sector variables. The dataset builds upon and extends the bank-level dataset described in Betz et al. (2014) by adding recent observations, additional banks, as well as a larger set of bank-specific and country-level variables. In total, the dataset covers 625 banks from 27 EU countries, but data availability varies widely across banks, variables and time periods. The dataset is constructed in a way to reflect real-time information availability, by applying varying time lags to all of the variables. The following subsections describe the building blocks of the dataset in greater detail.

Distress events

Given that outright bank failures are rare within the EU, we employ a number of surrogate bank-level events that proxy for bank distress. In total, four types of bank distress events are considered: i) state aid cases; ii) distressed mergers; iii) defaults; iv) bankruptcies.

State aid cases are defined to comprise direct capital injections, asset protection measures and loans/guarantees other than guaranteed bank bonds. Distress events based on state aid cases are defined to last from the announcement date up to the implementation date of a given measure. The underlying data for state aid cases is taken from the European Commission and complemented with information from national authorities and market sources (Reuters and Bloomberg).

A distressed merger is defined to occur if either the target bank had a negative coverage ratio within the year prior to the merger or if the acquiring bank received state aid within one year after the merger.[12] Distressed mergers are defined to last from the date when the coverage ratio turned negative (within one year prior to the merger) up to the merger date in the former case and from the merger date to the date that state aid was received in the latter case. Information on mergers and coverage ratios is obtained from Bloomberg, while state aid cases are defined and sourced as described in the previous paragraph.

Finally, defaults on financial obligations are taken from annual compendiums of defaults by Moody's and Fitch, while bankruptcies are obtained from Bankscope (See Betz et al. (2014) for a more detailed discussion of the definition of distress events). Figure 2 shows the numbers of banks by country that had at least one distress event and how distress events are distributed across the various subcategories of events.

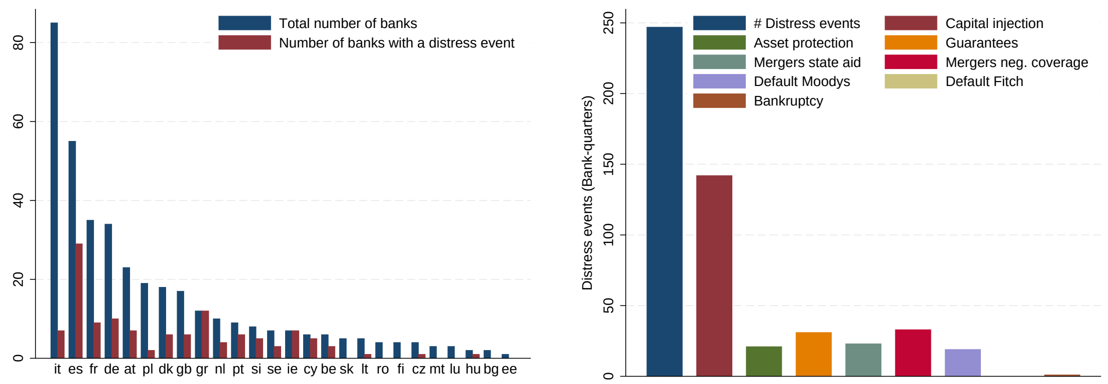
Risk drivers

A large set of around 175 potential bank-specific risk drivers was collected and constructed based on publicly available financial statements that are obtained through Bloomberg. The set of indicators covers all categories of the CAMELS framework and is based on quarterly financial statements whenever available and annual financial statements otherwise. In order to replicate a real-time information structure

[12] The coverage ratio is defined as the ratio of total capital and reserves for non-performing assets minus non-performing assets to total assets.

Figure 2: Distributions of banks by country and distress events by type

(a) Banks by country (b) Distress events by type



Notes: The distribution of banks across EU countries and the distribution of distress events across different types are for the data underlying the optimal model that is presented in Section 4.3 and that is estimated on the entire sample of available data and not restricted to the common sample that feeds into the logit LASSO.

quarterly financial statements data is lagged by two quarters, while annual financial statements data for a given year is used from the second quarter of the following year onwards.

Given that the environment in which a bank operates will affect its vulnerability to a large extent, a number of country-level banking-sector and macro-financial indicators were also collected and constructed from various sources such as the ECB Balance Sheet Items (BSI) Statistics, Eurostat National Accounts, ECB Statistical Data Warehouse (SDW), and the European Commission Macroeconomic Imbalances Procedure (MIP) Scoreboard. All market data related variables such as government bond yields or stock prices are lagged by one quarter, as are all of the banking sector indicators, while macro and macro-financial indicators such as GDP, house prices or the MIP variables are lagged by two quarters.

4.2 Choice of key parameters

The conceptual framework that was outlined in Section 2 highlighted the key choices that need to be made when building an early-warning model. The modeling framework that we proposed in Section 3 already provides some of the key choices by default, whereas others still need to be decided. For example, the chosen modeling technique is logistic LASSO, the evaluation metric is the relative usefulness measure based on the loss function specification by Sarlin (2013b) and the evaluation exercise is chosen to be recursive out-of-sample. However, the proposed modeling framework still leaves many choices to the model developer, such as the policy preference parameter, the forecasting horizon, the complexity of the model, and the data and sample pre-selection.

For the application at hand a preference parameter of $\mu = 0.9$ is chosen due to the fact that missing bank distress events should be considerably more costly to society than issuing false alarms about banks being vulnerable. While lower values of μ could also be conceivable, assigning a considerably higher weight to Type I errors than Type II errors is standard in the bank early warning literature (See e.g. Betz et al. (2014)). Moreover, in section 4.5 we show how the optimal model changes when the preference parameter is set to $\mu = 0.8$. A forecast horizon of 1-8 quarters prior to bank distress events is chosen for our application so as to have a model that identifies the build-up of vulnerabilities with a sufficient lead time. All bank distress events and the subsequent four quarters are excluded from the estimation in order to account for a possible crisis and post-crisis bias as highlighted by Bussiere and Fratzscher (2006). The recursive evaluation exercise for the model is chosen to start in 2006Q1 in order to gauge how the model would have performed in predicting the global financial crisis.

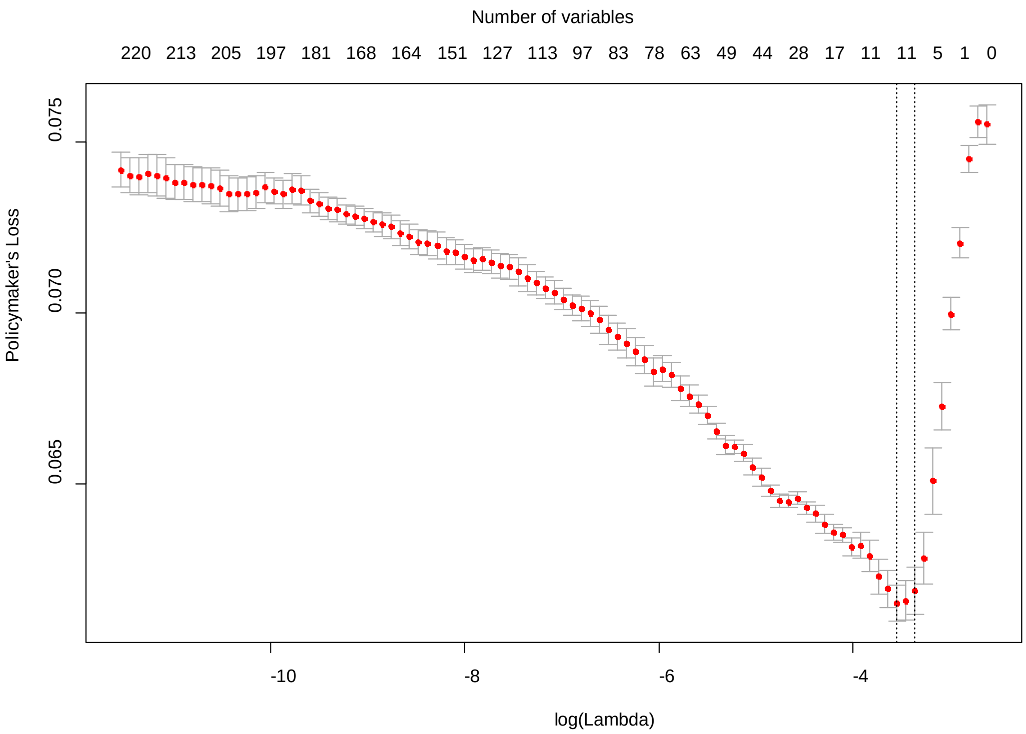
As the aim for the model is to provide good forecasts in real-time, we opt for a model with low complexity and optimal out-of-sample performance by choosing the shrinkage parameter for the logit LASSO that minimizes the cross-validated loss of the policymaker. The cross-validation is performed on 10 folds. Given that data availability varies considerably across the different bank-specific indicators, variable pre-selection for the logit LASSO is of key importance as it will determine the common sample on which the cross-validation is performed. There is a clear trade-off between considering as many potential risk drivers as possible and having a large sample of banks and quarters to feed into the logit LASSO. For our application we opt for having at least 5,000 bank-quarter observations in the common sample. The relevant bank-specific variables for the LASSO are chosen recursively by always adding the variable that reduces the common sample by the fewest observations. This leaves us with 176 bank-specific variables and 232 banks.[13]

4.3 Model specification and performance

Based on the chosen bank sample and key parameters described above, the optimal cross-validated logit LASSO penalty parameter translates into a model with eleven

[13]Data availability for some indicator categories is highly constrained, which unfortunately limits the number of possible risk drivers that can be included in the model. For example, no variable related to non-performing loans can be included in the model given extremely limited data availability.

Figure 3: Illustration of the cross-validation for the LASSO



Notes: The cross-validation for the LASSO is performed on 10 folds and the policymaker's loss is derived for a preference parameter of $\mu = 0.9$.

risk drivers. This is illustrated in Figure 3 which shows that there is a u-shaped relationship between the out-of-sample cross-validated policymaker's loss and the LASSO penalty parameter or equivalently model complexity. Both, very simple and very complex models have worse out-of-sample performance than medium-sized models with a handful of risk drivers. The final logistic LASSO model with the optimal cross-validated penalty parameter contains three bank-specific, four banking-sector and four macro-financial variables. The corresponding logit model estimates for the LASSO sample and the full sample of available observations are displayed in Table 2 including coefficient standard errors and significance levels, which are not available for LASSO models.[14]

The optimally chosen bank-specific variables are the ratio of tangible equity to total assets, the ratio of interest expenses to total liabilities and the NPL reserves to total assets ratio. All three variables have highly significant coefficients and signs that are in line with intuition. A lower leverage ratio increases a bank's vulnerability, as do higher interest rates on liabilities as well as higher reserves in relation to the bank's balance sheet size. Interestingly, a simple leverage ratio appears to be a better indicator to gauge a bank's vulnerability than a risk-weighted capital adequacy measure. Regarding the reserves to assets measure, there is an intuitive explanation for its inclusion in the model. Given that no indicators related to non-performing loans were included in the LASSO sample due to data limitations, reserves probably simply proxy for non-performing loans, which would also explain the positive coefficient sign.

The four banking-sector indicators that are chosen by the LASSO are the ratio of financial assets to GDP, the loan-to-deposit ratio (1-year change), the mortgages to loans ratio (1-year change) and the ratio of issued debt securities to total liabilities (1-year change). While the first two chosen banking-sector variables are estimated to increase the vulnerability of a bank, the latter two variables are estimated to decrease a bank's vulnerability over a two-year horizon. The fact that a larger banking sector relative to the size of the economy and a larger increase in the loan-to-deposit ratio raise banks' vulnerabilities appears in line with basic economic intuition. In contrast, the fact that an increase in the share of mortgages and an increase in market-based funding reduce banks' vulnerabilities could seem counterintuitive at first sight. However, given that the prediction horizon is chosen to be one to eight

[14]Given that the final model includes only eleven variables, the full sample of available observations is almost twice as large as the common sample used for the LASSO to select the relevant variables for the model. Although there are some differences in the magnitudes and significance levels of the estimated coefficients between the two samples, all estimated coefficients have the same sign across the two samples.

quarters and house prices and housing related activity are known to peak well in advance of financial crises (See e.g. Schudel (2013)) the negative coefficient on the mortgage share can easily be rationalised. In addition, the share of market-based funding can also be expected to decrease as potential problems in the banking-sector become more and more apparent, so that a negative estimated coefficient can indeed make sense given the prediction horizon.

Table 2: Estimated coefficients for the optimal model as of 2014Q4

Variable	(1) Full sample	(2) Lasso sample
Intercept	-3.039***	-3.111***
	(0.153)	(0.217)
Bank-specific variables		
Tangible equity / Total assets, lag 2	-0.306***	-0.296***
	(0.0445)	(0.0669)
Interest expenses / Total liabilities, lag 2	0.127**	0.256***
	(0.0541)	(0.0890)
Reserves for NPLs / Total assets, lag 2	0.165***	0.349***
	(0.0593)	(0.0956)
Banking-sector variables		
Financial assets / GDP, lag 2	0.000769	0.00295***
	(0.000600)	(0.00111)
Loans / Deposits (1-year change), lag 1	0.00911	0.0151
	(0.00611)	(0.0106)
Mortgages / Loans (1-year change), lag 1	-0.413***	-0.261*
	(0.0851)	(0.135)
Issued debt / Total liabilities (1-year change), lag 1	-0.153**	-0.136*
	(0.0628)	(0.0800)
Macro-financial variables		
Total credit / GDP (3-year change), lag 2	0.0166***	0.0171**
	(0.00635)	(0.00846)
House price gap (lambda = 1,600), lag 2	-0.0458***	-0.0526**
	(0.0172)	(0.0265)
MIP International Investment Position, lag 2	-0.0104***	-0.00695
	(0.00372)	(0.00574)
10-year yield (1-year change), lag 1	0.340***	0.402*
	(0.102)	(0.221)
Observations	8,195	4,293
Total number of banks	384	232
Number of SBGs	106	69
Number of LCBGs	23	20
Number of distressed banks	124	81
Number of pre-distress events	803	385
Pseudo R2	0.239	0.260
AUROC	0.847	0.850

Notes: Coefficient estimates refer to the logit model with the same specification as the logit LASSO model. Robust standard errors are in parentheses. Stars indicate the level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. SBGs refers to significant banking groups and LCBGs refers to large and complex banking groups.

Finally, the four optimally chosen macro-financial variables are the credit to GDP ratio (3-year change), the house price gap (real-time HP-filtered with $\lambda = 1,600$), the MIP international investment position and the 10-year government bond yield (1-year change). All four estimated coefficients are statistically significant at the 1 % level for the full sample estimation and the coefficient signs are in line with economic intuition, taking into account the fact that house prices peak well in advance of financial crises. The 3-year change in the credit to GDP ratio and the annual change in the 10-year government bond yield both increase banks' vulnerabilities, while the house-price-gap and the international investment position decrease banks' vulnerabilities over a two-year prediction horizon.

Table 3: In-sample and out-of-sample performance for the optimal model

	In-sample	**Out-of-sample Constant specification**	**Out-of-sample Changing specification**
Signaling threshold	0.083		
AUROC	0.847		
Relative usefulness	0.533	0.346	0.184
Noise-2-Signal ratio	0.352	0.320	0.505
Type I Error rate	0.167	0.310	0.235
Type II Error rate	0.293	0.221	0.386
Conditional pre-distress probability	0.236	0.326	0.251
Unconditional pre-distress probability	0.098	0.134	0.169
Probability difference	0.138	0.192	0.082
True positives	0.082	0.093	0.111
False positives	0.264	0.191	0.330
True negatives	0.638	0.674	0.525
False negatives	0.016	0.042	0.034

Notes: In-sample performance is computed for the logit model with the same specification as the logit LASSO model on the full sample of available data. The out-of-sample performance measures are for a recursive exercise starting in 2006Q1, where in each quarter predictions are made based on a model and signalling threshold that are estimated with data up to the previous quarter. The first out-of-sample exercise assumes a constant model specification across time in line with the model presented in Table 2. The second out-of-sample exercise applies the optimal cross-validated LASSO penalty parameter at each point in time to allow for a changing model specification in real-time. True positives, false positives, true negatives, and false negatives are expressed as a share of the total number of observations.

In order to assess how well the optimal parsimonious model explains and predicts the data, Table 3 displays a number of performance measures for the full data sample and for two different recursive out-of-sample prediction exercise starting in 2006Q1. Starting with the in-sample fit of the model, we see that the parsimonious model seems to explain the data reasonably well. The AUROC[15], a measure of the global signalling performance of the model independent of policy preferences, is fairly high at 0.847. In addition, the relative usefulness for a policymaker with $\mu = 0.9$, is around 53 %, indicating that the model could have considerable benefit

[15]AUROC stands for the Area Under the Receiver Operating Characteristics Curve. A perfect indicator has an AUROC of 1, while an uninformative indicator has an AUROC of 0.5.

for a policymaker who is relatively concerned about bank failures. In terms of more tangible numbers, the model only fails to signal less than 17 % of pre-distress events, while less than one third of calm periods are incorrectly classified as pre-distress events. The early warning performance is comparable to other state-of-the art bank early warning models (see e.g. Betz et al. (2014)), but based on a more parsimonious set of explanatory variables that were chosen based on objective model selection criteria (LASSO with cross-validation).

In terms of the out-of-sample performance, the final LASSO model specification also attains a fairly high relative usefulness of 34.6 % which is associated with 31 % of missed pre-distress events and 22.1 % of false distress alarms. Hence, the conditional out-of-sample distress probability is fairly high at almost 33 % compared to an unconditional distress probability of 13.4 %.[16] The out-of-sample performance of the recursive LASSO with a changing model specification at each point in time, as described in Section 3.3, is somewhat lower with a relative usefulness of 18.4 %. However, this is not surprising given that the changing model specifications are chosen solely based on the information set that was available at the time. Given that there are hardly any bank distress events in the dataset before 2008, the positive usefulness that the model yields can be interpreted as an encouraging sign that the proposed method works out-of-sample.

In summary, the parsimonious optimal early-warning model contains all relevant risk-driver categories, displays coefficient signs that are in line with economic reasoning and has good in-sample and out-of-sample signalling properties.

4.4 Model output for micro-macro analysis

One of the advantages of our bank-level early-warning model is that it allows for the analysis of the build-up of vulnerabilities at the micro and the macro level. The model can therefore be used for the analysis of systemic risk in both the cross-sectional and the time dimension. More specifically, the model can be used to identify systemically important banks that are vulnerable at a given point in time, as well as the build-up of banking-sector vulnerabilities over time at the country or regional level. For the analysis of the build-up of vulnerabilities over time the aggregation method proposed in Section 3.4 is particularly useful, while for both the cross-sectional and time dimension of systemic risk, the decomposition of vulnerabilities into contributing factors adds additional value for policy purposes, as it allows to identify at a high level in which areas possible vulnerabilities are emerging.

[16] The conditional distress probability is defined as the share of true pre-distress events whenever the model issued a signal.

Concrete examples of how model output can be used for risk-identification in the macro-prudential policy process are given below.

Starting with the cross-sectional dimension of systemic risk, the model can be used to display possible vulnerabilities for systemically important institutions either at a given point in time or how vulnerabilities have developed over a certain period of time. Beginning with the latter case, Figure 4 provides a number of bank-specific examples of recursive out-of-sample predictions from the optimal model. These examples qualitatively illustrate that the model performs generally well in signalling vulnerable banks early on, while inevitably also showing that the model misses some vulnerable states and issues a few false alarms. More specifically, the examples illustrate that the model can capture reasonably well bank vulnerability differences both within a country and across countries.

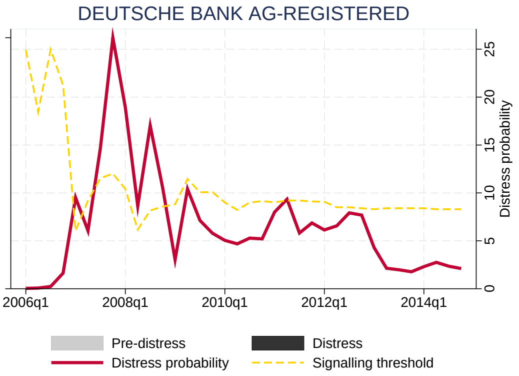
Moving to a snapshot view of cross-sectional systemic risk, Figure 5 shows a bar chart of the most vulnerable significant institutions in the euro area that are identified by the optimal model in 2014Q2. The model predictions for 2014Q2 are based on public bank financial statement data as of end-2013, and therefore make for a meaningful comparison to the results of the Comprehensive Assessment (CA) that were published by the Single Supervisory Mechanism (SSM) at the end of October 2014. The model signals nine banks in 2014Q2 that subsequently "failed" the CA. Moreover, seven CA "failures" are below the signalling threshold in 2014Q2 but among the 20 most vulnerable banks.[17] Interestingly, the model also signals Espirito Santo Financial Group, which failed in summer 2014, as one of the most vulnerable banks in 2014Q2 (with using bank-specific information until 2013Q4). These results underscore the potential usefulness of the proposed model to identify vulnerable systemically important banks.

For the analysis of the build-up of systemic risk over time, the evolution of the aggregate vulnerability for banks in the euro area can be useful, which is illustrated in Figure 6(a). In addition, the evolution of the aggregate vulnerability for banks in a given country is illustrated in Figure 6(b)-(f) for Germany, Austria, Italy, Spain and Ireland. Even though the proposed aggregation method has some shortcomings, as it disregards interconnectedness and possible non-linear effects of size on systemic importance as highlighted in Section 3.4, the aggregated micro-level vulnerabilities seem to provide a fairly good approximation of the build-up of systemic risk at the aggregate levels. For the euro area as a whole the model starts to signal vulnerable states starting in 2006Q2 and stops signalling in 2013Q2. For the given country ex-

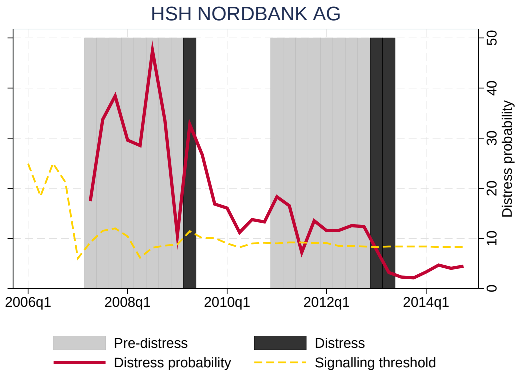
[17] It should be noted that for six out of the 25 banks that "failed" the CA the model cannot issue any signal due to a lack of data. In total, sufficient data to perform model predictions is available for 72 banks out of the 130 banks that are directly supervised by the SSM.

Figure 4: Examples of recursive out-of-sample predictions for the optimal model

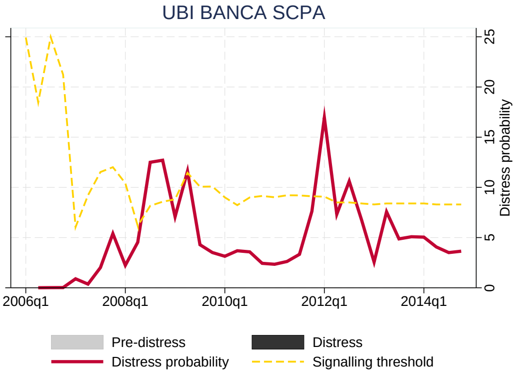
(a)



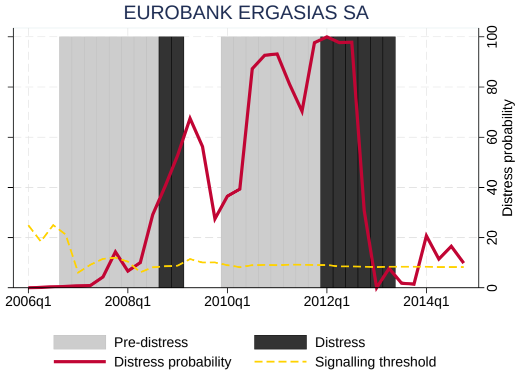
(b)



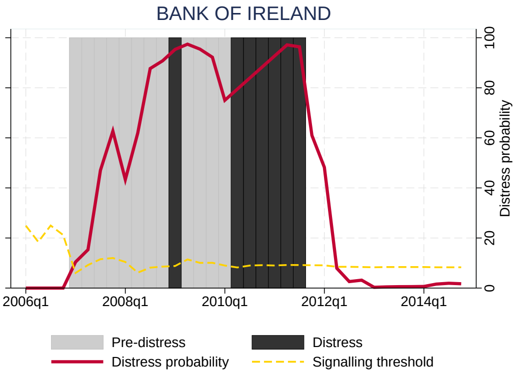
(c)



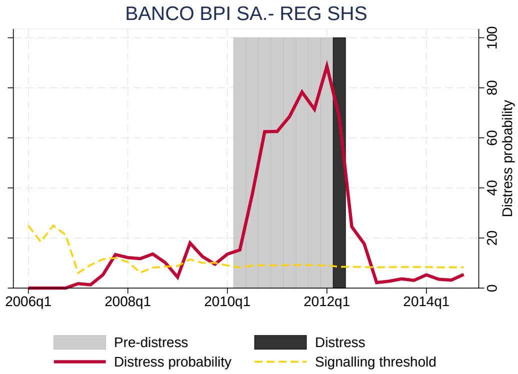
(d)



(e)

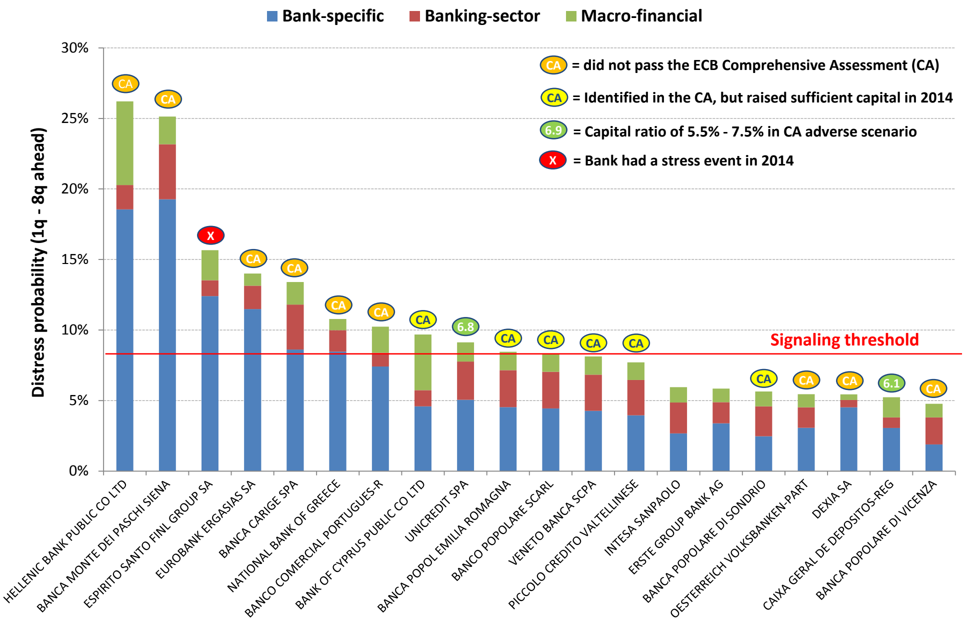


(f)



Notes: The banks are chosen to illustrate the method and no policy conclusions should be drawn from the examples. The recursive out-of-sample exercise is performed for the logit model with the same specification as the optimal logit LASSO model. In each quarter predictions are made based on a model and signalling threshold that are estimated with data up to the previous quarter. The signalling thresholds are derived for a policy preference parameter of $\mu = 0.9$. The grey areas in the charts indicate pre-distress periods that span 1-8 quarters prior to bank distress events (black bars) as defined in Section 4.

Figure 5: Most vulnerable banks in 2014Q2 for the optimal model

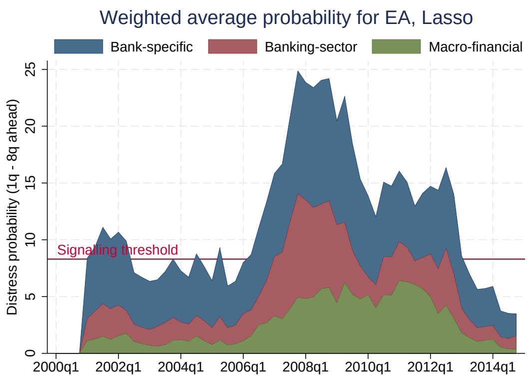


Notes: The output is chosen to illustrate the method and no policy conclusions should be drawn from the examples. The figure displays the 20 most vulnerable significant SSM banks in 2014Q2 based on the predictions of the logit model with the same specification as the logit LASSO model estimated on the full sample of available data. The signalling threshold is derived for a policy preference parameter of $\mu = 0.9$. The coloured bars illustrate the risk factor decomposition (bank-specific, banking-sector and macro-financial) as defined in Section 3.4.

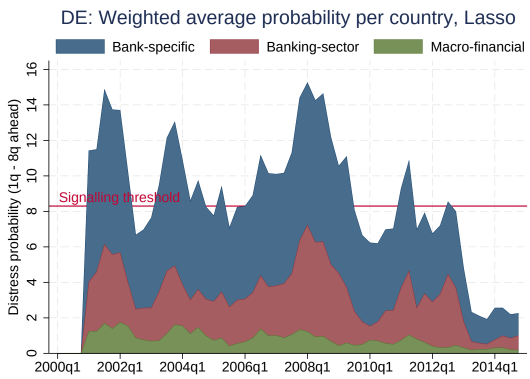
amples, it can be seen that the model captures different magnitudes of vulnerabilities across countries as well as different driving factors that are important.

Figure 6: Aggregate vulnerability for selected countries for the optimal model

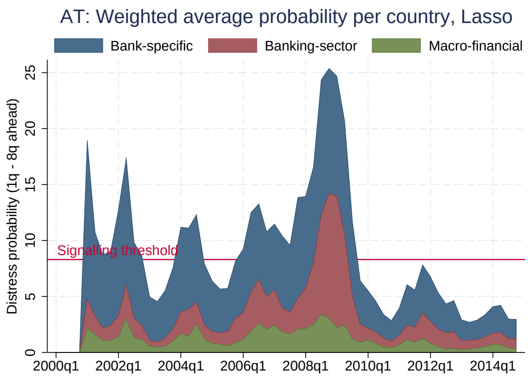
(a) Euro area



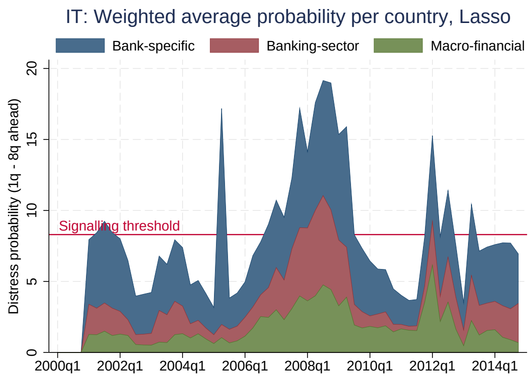
(b) Germany



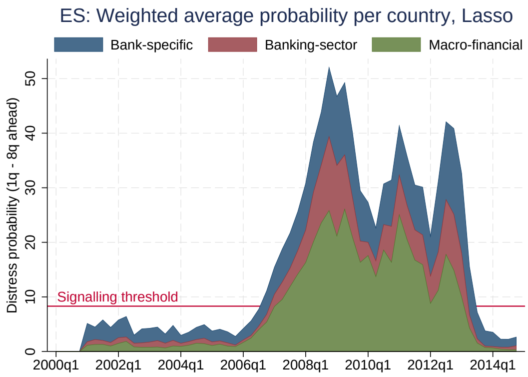
(c) Austria



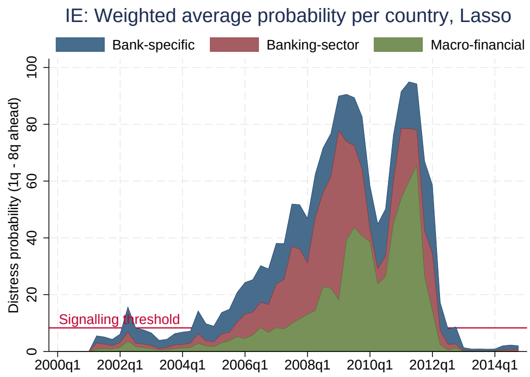
(d) Italy



(e) Spain



(f) Ireland



Notes: The countries are chosen to illustrate the method and no policy conclusions should be drawn. Aggregation is done for the logit model with the same specification as the logit LASSO model estimated on the full sample of available data. The coloured areas illustrate the risk factor decomposition (bank-specific, banking-sector and macro-financial) as defined in Section 3.4. The signalling threshold is derived for a policy preference parameter of $\mu = 0.9$.

4.5 Illustration of the flexibility of the modeling solution

The previous subsections have highlighted how the proposed modeling solution can be used to obtain an optimal early warning model specification for a given set of pre-modeling choices. However, the optimal early warning model will always depend on the specific aim it is to serve. As will be illustrated below, the proposed modeling solution is one possible way to deal with this multitude of context-specific optimal early warning models in a flexible way. The robustness exercises presented below also illustrate the importance of making the pre-modeling and modeling choices presented in section 2 explicit, as they will influence the optimal early warning model specification in a meaningful way.

Table 4 illustrates how the optimal cross-validated LASSO shrinkage parameter and therefore the optimal model complexity and specification changes when the forecast horizon, policy preference parameter and variable pre-selection are changed. Specifically, a shorter forecast horizon (1-4 quarters), a longer forecast horizon (1-12 quarters), a lower preference for not missing bank distress events ($\mu = 0.8$) and a smaller set of pre-selected variables (i.e. a larger requirement on the sample size of 10,000 observations) compared to the benchmark model are tested. In the benchmark specification a forecast horizon of 1-8 quarters before distress events, a preference parameter of $\mu = 0.9$ and a recursive variable pre-selection procedure that resulted in 5,000 observations were used.

Table 4 clearly shows that the optimal early warning model specification changes, when some key modeling choices are altered. For example, a shorter prediction horizon leads to a more complex model (15 variables) to be selected by the LASSO with cross-validation than in the baseline (11 variables), while a longer prediction horizon leads to a less complex model (8 variables). For our example, a lower preference for not missing distress events also leads to a less complex model with 7 variables, as does the case when one pre-selects fewer variables in order to cover a larger bank sample over time. Table 4 also illustrates that for the shorter prediction horizon of 1-4 quarters, more bank-specific variables get selected as relevant predictors for bank distress. These examples illustrate the need to be explicit about the modeling choices that are made, as set out in section 2, because these modeling choices will influence what type of early warning model is optimal. One benefit of our proposed modeling solution from section 3 is that it allows to derive an optimal early warning model for a given set of such choices in a straightforward way.

Table 4: Optimal model specifications for different early warning choices

Variable	(1) Baseline specification	(2) Different horizon (shorter)	(3) Different horizon (longer)	(4) Different preferences	(5) Different pre-selection
Intercept	-3.111***	-3.928***	-2.528***	-2.906***	-3.144***
Bank-specific variables					
Tangible equity / Total assets, lag 2	-0.296***	-0.286***	-0.183	-0.227***	
Interest expenses / Total liabilities, lag 2	0.256***	0.226***	0.216***	0.213**	0.075*
Reserves for NPLs / Total assets, lag 2	0.349***	0.267**			
Non-operating losses / Net revenue, lag 2		0.001			
Other income / Net revenue (1-year change), lag 2		0.016**			
Common equity / Total assets, lag 2			-0.064		-0.235***
Banking-sector variables					
Financial assets / GDP, lag 2	0.003***	0.0004	0.001	0.003***	
Loans / Deposits (1-year change), lag 1	0.015	0.01			
Mortgages / Loans (1-year change), lag 1	-0.261*	-0.308*			-0.472***
Issued debt / Total liabilities (1-year change), lag 1	-0.136*	-0.160			
Financial liabilities / GDP, lag 2		0.002			
Total assets / GDP, lag 2			0.012*		
Macro-financial variables					
Total credit / GDP (3-year change), lag 2	0.017**		0.019*	0.019*	0.021***
House price gap (lambda = 1,600), lag 2	-0.053**	-0.093***		-0.084***	-0.039**
MIP International Investment Position, lag 2	-0.007	-0.009*			
10-year yield (1-year change), lag 1	0.402*	0.201*	0.566***	0.534**	0.412***
Stock prices (1-quarter growth), lag 1		-0.013*			
Stock prices (4-quarter growth), lag 1		-0.009**			
MIP Private sector debt, lag 2			0.003		
Total credit / GDP, lag 2				-0.002	
MIP Current account balance, lag 2					-0.061***
Preference parameter	0.9	0.9	0.9	0.8	0.9
Pre-crisis period	1 - 8	1 - 4	1 - 12	1 - 8	1 - 8
Variable pre-selection	5,000	5,000	5,000	5,000	10,000
LASSO penalty parameter	0.029	0.013	0.037	0.035	0.024
Number of variables	11	15	8	7	7
Pseudo R2	0.260	0.270	0.209	0.211	0.194
AUROC	0.850	0.874	0.810	0.831	0.820
Signalling Threshold	8.906	7.704	9.500	19.30	8.299
Relative Usefulness	0.515	0.467	0.422	0.281	0.442

Notes: Coefficient estimates refer to the logit model with the same specification as the optimal logit LASSO model estimated on the pre-selected sample for the cross-validation. Stars indicate the level of significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

As the examples provided above have illustrated, building an early warning model is a complex task with several choices that need to be made. Most importantly, decisions about the specific purpose of the early warning model should influence its optimal specification in a meaningful way. This highlights the need for a conceptual framework as presented in section 1 to guide the model building process and to make certain modeling choices explicit rather than implicit. Based on these explicit modeling choices, the proposed modeling solution presented in section 3 provides one possible way to easily obtain a model specification that suits the specific early warning purpose at hand.

5 Conclusion

The large economic costs brought about by severe financial crises have again become apparent in recent years. In order to avoid or at least mitigate the impact of future financial crises it is necessary to gain a deeper understanding of the driving factors that cause such crisis episodes and to devise models that help to identify the build-up of financial imbalances and systemic risk early on. The work on early-warning models has therefore gained prominence in recent years, both in the academic and policy sphere. However, the numerous complex choices that are involved in building such models and the various approaches that have been employed in the literature call for a structured modeling approach.

This paper has put forward a general-purpose framework for deriving early-warning models and has applied it to predicting distress in European banks. The paper therefore contributes to the existing literature in three main ways. First, the paper has introduced a conceptual framework to guide the process of building early-warning models, which highlights and structures the numerous complex choices that the modeler needs to make. Second, the paper has provided a flexible modeling solution to the conceptual framework that supports model selection in real-time. Specifically, our proposed solution combines the loss function approach to evaluate early-warning models with regularized logistic regression and cross-validation to find a model specification with optimal real-time out-of-sample forecasting properties. Finally, the paper has illustrated how the modeling framework can be used in analysis supporting both micro- and macro-prudential policy by applying it to a large dataset of EU banks and showing some examples of early-warning model visualizations.

References

Alessi, Lucia and Carsten Detken, "Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity," *European Journal of Political Economy*, September 2011, *27* (3), 520–533.

_ **and** _ , "Identifying excessive credit growth and leverage," Working Paper Series 1723, European Central Bank August 2014.

Babecky, Jan, Tomas Havranek, Jakub Mateju, Marek Rusnak, Katerina Smidkova, and Borek Vasicek, "Banking, debt and currency crises: early warning indicators for developed countries," Working Paper Series 1485, European Central Bank October 2012.

Behn, Markus, Carsten Detken, Tuomas A. Peltonen, and Willem Schudel, "Setting Countercyclical Capital Buffers based on Early Warning Models: Would it Work?," Working Paper Series 1604, European Central Bank November 2013.

Berg, Andrew and Catherine Pattillo, "Predicting currency crises:: The indicators approach and an alternative," *Journal of International Money and Finance*, August 1999, *18* (4), 561–586.

Betz, Frank, Silviu Oprica, Tuomas A. Peltonen, and Peter Sarlin, "Predicting distress in European banks," *Journal of Banking & Finance*, 2014, *45* (C), 225–241.

Bien, Jacob, Jonathan Taylor, and Robert Tibshirani, "A Lasso for hierarchical interactions," *The Annals of Statistics*, 2013, *41*, 1111–1141.

Bishop, Christopher M., *Pattern Recognition and Machine Learning*, Springer, 2006.

Borio, Claudio and Mathias Drehmann, "Assessing the risk of banking crises - revisited," *BIS Quarterly Review*, March 2009, pp. 29–46.

_ **and Philip Lowe**, "Asset prices, financial and monetary stability: exploring the nexus," BIS Working Papers 114, Bank for International Settlements 2002.

_ **and** _ , "Securing sustainable price stability: should credit come back from the wilderness?," BIS Working Papers 157, Bank for International Settlements July 2004.

Bussiere, Matthieu and Marcel Fratzscher, "Towards a new early warning system of financial crises," *Journal of International Money and Finance*, October 2006, *25* (6), 953–973.

Candelon, B., E. Dumitrescu, and C. Hurlin, "How to Evaluate an Early-Warning System: Toward a Unified Statistical Framework for Assessing Financial Crises Forecasting Methods," *IMF Economic Review*, 2012, *60(1)*, 75–113.

Dattels, P., R. McCaughrin, K. Miyajim, and J. Puig, "Can you Map Global Financial Stability?," IMF Working Paper, WP/10/145, International Monetary Fund 2010.

Demirgüc-Kunt, Asli and Enrica Detragiache, "The Determinants of Banking Crises in Developing and Developed Countries," *IMF Staff Papers*, March 1998, *45* (1), 81–109.

Demirgüç-Kunt, Asli and Enrica Detragiache, "Monitoring banking sector fragility. A multivariate logit approach," *World Bank Economic Review*, 2000, *14* (2), 287–307.

Detken, Carsten, Olaf Weeken, Lucia Alessi, Diana Bonfim, Miguel M. Boucinha, Christian Castro, Sebastian Frontczak, Gaston Giordana, Julia Giese, Nadya Jahn, Jan Kakes, Benjamin Klaus, Jan Hannes Lang, Natalia Puzanova, and Peter Welz, "Operationalising the countercyclical capital buffer: indicator selection, threshold identification and calibration options," ESRB Occasional Paper Series No. 5, ESRB June 2014.

Drehmann, Mathias and Mikael Juselius, "Evaluating early warning indicators of banking crises: Satisfying policy requirements," *International Journal of Forecasting*, 2014, *30(3)*, 759–780.

Flood, M.D., V. Lemieux, M. Varga, and Wong W., "The Application of Visual Analytics to Financial Stability Monitoring," 2014. Office of Financial Research Working Paper 2014-02 .

Gropp, Reint, Jukka Vesala, and Giuseppe Vulpes, "Equity and Bond Market Signals as Leading Indicators of Bank Fragility," *Journal of Money, Credit and Banking*, 2006, *38* (2), 399–428.

Holopainen, Markus and Peter Sarlin, "Toward Robust Early-Warning Models: A Horse Race, Ensembles and Model Uncertainty," BOF Discussion Paper 06/2015, Bank of Finland 2015.

Jorda, Oscar and Alan M. Taylor, "Performance Evaluation of Zero Net-Investment Strategies," NBER Working Papers 17150, National Bureau of Economic Research, Inc. 2011.

Kaminsky, Graciela, Saul Lizondo, and Carmen M. Reinhart, "Leading Indicators of Currency Crises," *IMF Staff Papers*, March 1998, *45* (1), 1–48.

Laeven, Luc and Fabian Valencia, "Systemic Banking Crises Database: An Update," IMF Working Papers 12/163, International Monetary Fund June 2012.

Lainà, P., J. Nyholm, and P. Sarlin, "Leading Indicators of Systemic Banking Crises: Finland in a Panel of EU Countries," *Review of Financial Economics*, 2015, *24*, 18–35.

Lo Duca, Marco and Tuomas A. Peltonen, "Assessing systemic risks and predicting systemic events," *Journal of Banking & Finance*, 2013, *37* (7), 2183 – 2195.

Marghescu, Dorina, Sarlin Peter, and Shuhua Liu, "Early Warning Analysis for Currency Crises in Emerging Markets: A Revisit with Fuzzy Clustering," *Intelligent Systems in Accounting, Finance and Management*, 2010, *17* (3-4), 134–165.

Peltonen, Tuomas A., "Are Emerging Market Currency Crises Predictable? A test," *ECB Working Paper Series*, 2006, *571.*

_ , Andreea Piloiu, and Peter Sarlin, "Network linkages to predict bank distress," 2015. ECB Working Paper No. 1828.

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

Reinhart, Carmen M. and Kenneth S. Rogoff, "Is the 2007 US Sub-prime Financial Crisis So Different? An International Historical Comparison," *American Economic Review*, May 2008, *98* (2), 339–44.

Ripley, B.D., *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.

Sarlin, Peter, "On biologically inspired predictions of the global financial crisis," *Neural Computing and Applications*, 2013, *24* (3–4), 663–673.

— , "On policymakers' loss functions and the evaluation of early warning systems," *Economics Letters*, 2013, *119* (1), 1–7.

— , "Macroprudential oversight, risk communication and visualization," 2014. London School of Economics, Systemic Risk Centre, Special Paper 4.

— **and Tuomas A. Peltonen**, "Mapping the state of financial stability," *Journal of International Financial Markets, Institutions and Money*, 2013, *26*, 46–76.

Schudel, Willem, "Shifting horizons: Assessing macro trends before, during, and following systemic banking crises," Mimeo, European Central Bank 2013.

Schularick, Moritz and Alan M. Taylor, "Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008," *American Economic Review*, April 2012, *102* (2), 1029–61.

Shmueli, Galit, "To explain or to predict?," *Statistical Science*, 2010, *25* (3), 289–310.

Stone, Mervyn, "An asymptotic equivalence of choice of model by cross-validation and Akaike criterion," *Journal of the Royal Statistical Society*, 1977, *39* (1), 44–47.

Tibshirani, Robert, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, 1996, *39* (58), 267–288.

Acknowledgements

The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank, the Eurosystem or the European Systemic Risk Board. All results are derived from publicly available information and do not imply any policy conclusions regarding individual banks or countries.

Jan Hannes Lang
European Central Bank, Frankfurt am Main, Germany; email: jan-hannes.lang@ecb.europa.eu

Tuomas A. Peltonen
European Systemic Risk Board, Frankfurt am Main, Germany; email: tuomas.peltonen@ecb.europa.eu

Peter Sarlin
Silo.AI and RiskLab at Hanken School of Economics and Arcada, Helsinki, Finland; email: peter@silo.ai

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-3287-5 ISSN 1725-2806 doi:10.2866/25674 QB-AR-18-062-EN-N