



HAL
open science

Bank insolvency risk and Z-score measures: caveats and best practice

Vincent Bouvatier, Laetitia Lepetit, Pierre-Nicolas Rehault, Frank Strobel

► To cite this version:

Vincent Bouvatier, Laetitia Lepetit, Pierre-Nicolas Rehault, Frank Strobel. Bank insolvency risk and Z-score measures: caveats and best practice. 2018. hal-01937929

HAL Id: hal-01937929

<https://unilim.hal.science/hal-01937929v1>

Preprint submitted on 28 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bank insolvency risk and Z-score measures: caveats and best practice

Vincent Bouvatier* Laetitia Lepetit[†]
Pierre-Nicolas Rehault[‡] Frank Strobel[§]

July 18, 2018

Abstract

We highlight caveats arising in the application of traditional ROA-based Z-scores for the measurement of bank insolvency risk, develop alternative Z-score measures to resolve these issues, and make recommendations for best practice for the US/Europe based on the experience of the financial crisis of 2007-2008. Using a probabilistic approach (i) our novel regulatory capital Z-score dominates traditional Z-score measures for both US/Europe; (ii) Z-scores computed with exponentially weighted moments dominate those with moving moments for the US sample, but not for Europe. For both US/Europe, using a multivariate logit approach (i) allows computation of augmented Z-scores that provide probabilities of distress that better discriminate between distressed/surviving banks than the probabilistic approach; (ii) suggests that the ROA-based Z-score using current values of the capital-asset ratio is best, calculated either with moving or exponentially weighted moments.

Keywords: Z-score; bank; insolvency risk; risk measure

*Université Paris Est Créteil, ERUDITE, Créteil, France. E-mail: vincent.bouvatier@u-pec.fr

[†]Université de Limoges, LAPE, Limoges, France. E-mail: laetitia.lepetit@unilim.fr

[‡]Université de Limoges, LAPE, Limoges, France. E-mail: pierre-nicolas.rehault@unilim.fr

[§]University of Birmingham, Department of Economics, Birmingham, UK. E-mail: f.strobel@bham.ac.uk (corresponding author)

1 Introduction

The Z-score is widely used in the empirical banking literature to reflect a bank's probability of insolvency; it is, e.g., one of the indicators used by the World Bank in their Global Financial Development Database to measure financial institutions' soundness.¹ The widespread use of Z-scores can be explained by their relative simplicity and the fact that they require only accounting information for their computation; as such, they can be applied to both listed and unlisted financial institutions. However, both their simplicity and reliance on accounting data result in potential challenges in their application; our paper highlights these caveats, develops alternative Z-score measures to address these issues, and then proceeds to develop recommendations regarding best practice in the application of Z-score measures for the measurement of bank insolvency risk, based on the experience of the financial crisis of 2007-2008 for the US and Europe.

The traditional return-based Z-scores commonly used at present, designed as an indicator for the likelihood that a bank's equity might be wiped out through losses, are faced with several potential challenges in their practical application; we consider improvements for each of these issues. First, return-based Z-score measures disregard the importance of regulatory capital constraints faced by banks; we take this aspect into account by constructing an alternative "regulatory capital" Z-score as an indicator for the likelihood that a bank's regulatory capital drops below a given threshold. Secondly, the construction of time-varying Z-score measures more generally introduces questions of appropriate estimation of relevant moments (see e.g. the survey by Lepetit and Strobel (2013)); in addition to a thorough evaluation of alternative approaches to using moving moments we also consider exponentially weighted moments in this paper. Lastly, while the return-based and regulatory capital Z-score measures can be related to unconditional probabilities of distress, the link between Z-scores and probability of distress could be conditional on observable bank characteristics (e.g. bank size) and the macroeconomic environment (i.e. a reduction in capital might be more challenging for a bank during a financial crisis period than during normal times). We address this issue by constructing a multivariate logit model which allows for heterogeneity in the relationship between the Z-score and the probability of distress, allowing an improved assessment of the probability of distress of banks and the construction of an "augmented" Z-score formula for out-of-sample use.

We compare the performance of these different Z-score measures to determine the best Z-score(s) to use in practice. We pursue comparisons drawing on a range of alternative testing procedures, first at a probabilistic level, where our Z-score measures can be related to predicted probabilities of distress using sound statistical foundations. We then progress to a multivariate logit model in

¹For some recent papers using this methodology, see e.g. Caiazza et al. (2018), Pino and Sharma (2018), Tsionas and Mamatzakis (2017), Berger et al. (2016), Fukuyama and Matousek (2017), Han et al. (2016), Doumpos et al. (2015), Vazquez and Federico (2015), Hakenes et al. (2015), Berger et al. (2014), Delis et al. (2014), Fang et al. (2014), Fu et al. (2014), Beck et al. (2013), Bertay et al. (2013), DeYoung and Torna (2013).

order to link the Z-score to the probability of distress, which allows us to include key observable characteristics at the bank level as well as changes in economic conditions. This parametric econometric model approach allows additional testing procedures to assess the probabilities of distress and evaluate the different Z-score measures used. We then finally compare the multivariate logit and probabilistic approaches in order to determine if there is a "best" Z-score measure to use in empirical analysis.

We exploit for our empirical analysis the substantial number of banks that became distressed during the financial crisis of 2007-2008, predominantly in the US and Europe, which will allow us to compare and evaluate the different approaches to the construction of Z-score measures developed.

Our results firstly show that our novel regulatory capital Z-score, which we introduce to reflect the relevance of regulatory capital constraints faced by banks, dominates the two versions of traditional return-based Z-score measures we examine for the majority of the test criteria used in our probabilistic approach, for both the US and European samples. Focussing next on how best to allow for time-variation in Z-score measures, our analysis using the probabilistic approach shows that Z-scores computed with exponentially weighted moments are better than those with moving moments for the US sample, but do not present a clear advantage for the European one. We further demonstrate in this context that moving moment Z-score measures based on 3-year windows outperform Z-scores using 5-year windows for both the US and Europe. Moving on to the multivariate logit approach, including selected observable characteristics at the bank level and changes in economic conditions allows us to compute augmented Z-scores that provide probabilities of distress that better discriminate between distressed and surviving banks than the probabilistic approach, for both the US and European samples. Using the multivariate logit approach, our results suggest that the best Z-score to use, in both the US and Europe, is the ROA-based Z-score using current values of the capital-asset ratio, calculated either with moving or exponentially weighted moments.

The remainder of the paper is organized as follows: Section 2 presents the traditional and regulatory Z-score measures considered, links them to predicted probabilities of distress and presents the data; Section 3 tests the relative performance of the different Z-scores using the probabilistic approach; Section 4 develops a parametric econometric model to assess probabilities of distress and applies it to evaluating the different Z-score measures; Section 5 compares the performance of the parametric econometric model approach with that of the probabilistic one, and develops an augmented Z-score formula for out-of-sample use; finally, Section 6 concludes the paper.

2 Z-score measures and data

2.1 Z-score definitions

2.1.1 Traditional ROA-based Z-score

Let us first restate the traditional justification for using Z-scores as a risk measure reflecting a bank's probability of insolvency. For these purposes, bank insolvency is commonly defined as a state where $(CAR + ROA) \leq 0$, with CAR the bank's capital-asset ratio and ROA its return on assets. Lepetit and Strobel (2015) showed that if ROA is a random variable with finite mean μ_{ROA} and variance σ_{ROA}^2 , the one-sided Chebyshev inequality gives an upper bound of the probability of insolvency as a nonlinear function of the Z-score $Z \equiv (CAR + \mu_{ROA}) / \sigma_{ROA} > 0$ as follows

$$P(ROA \leq -CAR) \leq (1 + Z^2)^{-1} < 1$$

In this sense, a high Z-score can thus be argued to correspond to a low probability of insolvency.

2.1.2 A new "regulatory" Z-score

We develop a new "regulatory" Z-score that is based on a bank's regulatory rather than standard capital-asset ratio. In analogy to the definition of bank distress in Section 2.1.1, one could argue that a bank is in "regulatory" distress when a particular regulatory capital ratio $RCAR$ drops below a given regulatory threshold TR , so that, e.g., its total capital ratio drops below 8% or its Common Equity Tier 1 ratio drops below 4.5% (under Basel III). This then allows us to relate a "regulatory capital" Z-score, defined as $Z_{RCAP} \equiv (\mu_{RCAR} - TR) / \sigma_{RCAR}$, to an upper bound of the (unconditional) probability of regulatory distress, as follows

Proposition 1 *If the regulatory capital ratio $RCAR$ is a random variable with finite mean $\mu_{RCAR} > TR$ and variance $\sigma_{RCAR}^2 > 0$, an upper bound of the bank's probability of not satisfying the regulatory threshold tr is given by $P(RCAR \leq TR) \leq (1 + Z_{RCAP}^2)^{-1} < 1$, where the "regulatory capital" Z-score is defined as $Z_{RCAP} \equiv (\mu_{RCAR} - TR) / \sigma_{RCAR} > 0$.*

Proof. See Appendix A. ■

2.1.3 Time-variation - moving moments

The time-varying implementation of the two Z-score measures introduced above can be carried out in different ways; common approaches to this, for the traditional Z-score, are discussed in Lepetit and

Strobel (2013).² In common with the empirical literature to date, we first calculate "rolling" versions of the Z-scores. For this we initially consider the most commonly used 3-year rolling windows for the calculation of moving moments. For the traditional Z-score, we calculate two alternative versions: (i) the "moving capital" Z-score Z^{r3} , using moving mean and standard deviation estimates for both ROA and CAR , with window widths of three observations; and (ii) the "current capital" Z-score $Z^{r3,c}$, using moving mean and standard deviation estimates for ROA , with window widths of three observations, and combining these with current values of CAR . For the regulatory capital Z-score, we calculate Z_{RCAP}^{r3} using moving mean and standard deviation estimates for the total capital ratio (TCR), with window widths of three observations, and a regulatory threshold of 8%. Alternatively, we also consider equivalently constructed Z-scores Z^{r5} , $Z^{r5,c}$ and Z_{RCAP}^{r5} that use 5-year rolling windows for the calculation of moving moments instead.

2.1.4 Time-variation - exponentially weighted moments

We then go further than the existing literature by also implementing exponentially weighted moments in the construction of alternative time-varying Z-scores; exponentially weighted moments have the advantage of always drawing on the full data history available, while weighting more recent observations more heavily. We thus construct the "exponential capital" Z-score Z^{exp} , using exponential mean and standard deviation estimates for both ROA and CAR , the "current capital" Z-score $Z^{\text{exp},c}$, using exponential mean and standard deviation estimates for ROA combined with current values of CAR , and the regulatory capital Z-score Z_{RCAP}^{exp} , using exponential mean and standard deviation estimates for the total capital ratio (TCR). For each of our relevant variables, the exponentially weighted moving average $EWMA$, and exponentially weighted moving variance $EWVAR$, are calculated recursively as³

$$\begin{aligned} \text{mean: } EWMA_t(x) &= \alpha x_t + (1 - \alpha)EWMA_{t-1}(x) \\ \text{variance: } EWVAR_t(x) &= (1 - \alpha)(EWVAR_{t-1}(x) + \alpha(x_t - EWMA_{t-1}(x))^2) \end{aligned}$$

where the recursion for the (naive) exponentially weighted variance follows Finch (2009) and West (1979). We determine the smoothing parameter α optimally, carrying out a grid search over the range 0.01 to 0.99 with increments of 0.01. This produces 99 competing exponential Z-score can-

²Lepetit and Strobel (2013) further examine, for commercial, cooperative and savings banks in G20 countries from 1992–2009, which of the various mean and standard deviation estimates commonly used in the literature to compute traditional time-varying Z-score measures "best fit the data", which, however, is only interpreted narrowly in the sense of producing minimum one-period-ahead forecast errors regarding the relevant moments. As a consequence, in contrast to our present paper, their assessment does not extend to examining the relative performance of different time-varying Z-score measures in relation to actual, observed instances of bank distress.

³Stata code implementing these exponentially weighted moments and associated Z-score measures is available from the corresponding author on request.

didates; amongst those, the one maximizing the AUROC curve is then chosen (see Section 3.1 for details of our testing procedure). For our sample (described in Section 2.2), determining the optimal smoothing parameter α optimally for CAR shows consistently that current values of CAR are best; hence we consider only the exponentially weighted Z-scores $Z^{\text{exp},c}$ and Z_{RCAP}^{exp} throughout the paper.

2.2 Bank distress and sample construction

In order to determine which one of the different Z-score measures introduced in Section 2.1 is best, we focus on the financial crisis of 2007-2008. A good Z-score measure should allow us to discriminate between banks that survived the financial crisis and banks that fell into distress during it. Using annual financial statement data for US commercial banks drawn from SNL Financial LC, and European commercial, savings and cooperative ones extracted from BvD BankScope,⁴ we consider for our empirical investigation banks existing in 2006 and still operating in 2014, and banks existing in 2006 but falling into distress between 2007 and 2014. Thus, we exclude non-distressed banks existing in 2006 but disappearing between 2007 and 2014 due to mergers & acquisitions, and all banks created between 2007 and 2014.

We classify a bank as distressed for the US and Europe, respectively if: (i) it failed according to the FDIC or ECB; or (ii) it asked for Troubled Asset Relief Program (TARP) funds for the US, or represents a "rescued bank" following Molyneux et al. (2014) and Fratianni and Marchionne (2013) for Europe; or (iii) it was "weakly" capitalized and was subsequently acquired by a safer bank. We consider a bank as weakly capitalized if the tangible common equity/tangible assets ratio is lower than 2%, or the Tier 1 capital ratio is lower than 4%, or the total capital ratio is lower than 8%. When a bank is classified as distressed during a given year, it is dropped from the sample in the following and all subsequent years.

Our initial US sample consists of 5931 commercial banks; amongst those, 545 fell into distress between 2007 and 2014. Dropping banks with a net loans to total assets ratio lower than 5%, we retain a final US sample allowing us to calculate 3-year rolling Z-score measures of 5823 banks, of which 543 fell into distress between 2007 and 2014. The initial European sample, on the other hand, has 3073 commercial, savings and cooperative banks; 157 of these fell into distress in the period considered. When we similarly drop banks with a net loans to total assets ratio lower than 5%, we obtain a final sample allowing us to calculate 3-year rolling Z-score measures of 2866 European banks, of which 142 became distressed between 2007 and 2014. Not all European banks publish their total capital ratio, reducing the sample that can be used to examine the regulatory Z-score to

⁴Our European sample consists of Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom. Data coverage of European banks in SNL is limited in both number of banks and years of data available.

2138 banks, of which only 60 fell into distress; this unavailability of total capital ratios is consistent across BvD BankScope, SNL, Bloomberg as well as annual reports. Table 1 provides summary statistics for both our US and European samples. As expected, the unconditional probability of distress shows considerable fluctuation, and is naturally highest in 2008–2010.

As regards the construction of the time-varying Z-score measures we consider, those using rolling windows are calculated with the window widths considered from 2007 onwards. The Z-scores using exponentially weighted moments are similarly constructed from 2007 onwards, but draw on the maximum amount of historical data available for each bank in our sample; the earliest usable observations are in 1990 and 1999 for the US and European samples, respectively. Table B1 (in Appendix B) presents summary statistics for the main Z-score measures computed for our samples of distressed and non-distressed banks. As expected, Z-score measures of distressed banks are significantly lower than those of non-distressed banks. Table B2 further gives correlation matrices for the main Z-score measures. We observe that the regulatory capital Z-score is only weakly correlated with the total capital ratio (TCR), suggesting that it may indeed be of interest to consider our novel regulatory capital Z-score as a measure of default risk, rather than simply the TCR itself.

3 Comparing predicted probabilities of bank distress to actual bank distress

The different Z-scores developed in Section 2.1 can be considered as rival binary classifiers, with each classifier providing predicted probabilities that banks become distressed. These predicted probabilities can then be compared to observed discrete outcomes of distressed banks, in order to evaluate the different approaches to constructing Z-score measures, using the techniques outlined in the following section. In order to retain observations with negative Z-scores in the calculation of predicted probabilities, negative Z-scores are rescaled to lie between zero and the minimum observed Z-score in the sample. This is necessary as negative Z-scores do not fit into the probabilistic framework drawn on in Section 2.1, where a Z-score of zero already corresponds to a probability of insolvency of 100%; however, negative Z-scores can nevertheless occur when calculated based on actual accounting data when a bank is in fact in distress.

3.1 Testing procedure

Various criteria have been developed to compare formulas that predict the probabilities of having observed a positive outcome ($Y_{i,t} = 1$) of a corresponding binary response ($Y_{i,t} = \{0, 1\}$). Here, we consider several complementary criteria: three criteria that focus on the relative ranking of the

predicted probabilities provided by each classifier (AUROC curve, AUPR curve and H measure); and a criterion that takes into account the numerical values of the predicted probabilities (Tjur R^2). The latter three are not commonly applied in economics/finance, but more widely used in different scientific disciplines such as biostatistics and machine learning (e.g. Hand (2012), Saito and Rehmsmeier (2015)).

The first category of criteria are defined and computed from the confusion matrix (Hand (2012)).⁵ We first consider the area under the receiver operating characteristic (AUROC) curve. The receiver operating characteristic (ROC) curve plots the true positive rate against the false positive rate at various threshold settings. Therefore, the area under the ROC curve, considering all possible thresholds, has the attractive property of providing a summary measure of classification ability. In addition, the AUROC curve has an intuitive interpretation: it equals the concordance probability (also called the c -index) that is equivalent to the probability that a randomly chosen positive outcome is ranked higher than a randomly chosen negative outcome. Realistic values for the AUROC curve range thus from 0.5 (random ranking) to 1 (perfect ranking).

The AUROC curve is a widely used measure of classifier performance, but can be a misleading indicator when positive outcomes are sparse in the dataset (i.e. for imbalanced datasets). This issue is related to the fact that the ROC curve reports the false positive rate on the x-axis, calculated as the ratio of false positives to overall negatives (i.e. false positives and true negatives). When positive outcomes are sparse, it becomes easier to predict negative outcomes and, consequently, true negatives are much more numerous than false positives (i.e. the false positive rate is generally weak). Therefore, if we care more about the positive outcomes, the AUROC curve can overstate the overall performance of the classifier. The area under the precision-recall (AUPR) curve can be considered as a more appropriate evaluation metric in such situations (Saito and Rehmsmeier (2015)). The precision-recall (PR) curve plots precision against recall at various threshold settings, where precision is defined as the ratio of true positives to overall positives, and recall is a synonym for the true positive rate. As a consequence, the PR curve and the ROC curve both use the true positive rate, but the PR curve considers precision instead of the false positive rate. In other words, in contrast to the ROC curve, the PR curve does not consider the true negatives; this gives a more informative picture of the performance of the classifier in an imbalanced dataset. However, the AUPR curve does not have an intuitive interpretation, unlike the AUROC curve.

A more fundamental criticism has been expressed towards the AUROC curve by Hand (2009, 2010). Considering optimal thresholds, these papers demonstrate that optimising the AUROC

⁵The confusion matrix reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) for a given threshold. For instance, for a given threshold $t \in [0, 1]$, the number of positive outcomes with predicted probabilities higher than t gives the TP. Regarding the criteria computed from the confusion matrix, some simpler fit statistics like accuracy (i.e., $(TP+TN)/(FP+TP+FN+TN)$) require specifying a particular threshold, while some more complex measures provide an average performance covering all possible thresholds.

curve is equivalent to minimising expected minimum loss (resulting from misclassification costs), showing that the latter depends on the score densities, and hence on the classifier. Therefore, the interpretation of the AUROC curve is argued to be model-specific, leading to the conclusion that the AUROC curve is not a coherent measure to compare rival classifiers. Hand (2009) proposes the H measure, which is not classifier-dependent, as a coherent alternative to the AUROC curve.⁶ However, Flach et al. (2011) challenges the argument developed by Hand (2009, 2010), contending that considering only optimal thresholds is over-optimistic from a practical perspective. When all scores that have been assigned to data points are considered as possible thresholds, the AUROC curve is arguably not model-specific, and thus would remain a coherent measure of classification performance.

The AUROC curve, AUPR curve and H measure focus only on the relative ranking of the predicted probabilities. However, the numerical values of the predicted probabilities can also be considered as meaningful. As a result, we use an additional criterion in order to also measure the accuracy of the predicted probabilities, the Tjur (2009) R^2 , which is defined as the difference between the mean of the predicted probabilities of positive outcomes and the mean of the predicted probabilities of negative outcomes.

Adopting an agnostic approach as to which of these criteria might be best suited for our purposes, we apply all four criteria to empirically compare the performance of the different Z-scores as classifiers providing predicted probabilities that banks become distressed. A "best" Z-score would ideally present the highest values for the AUROC curve, the AUPR curve, the H measure and the Tjur R^2 . As these criteria assess different aspects of performance, it is, however, unclear whether one should expect one of the alternative Z-score measures to dominate all others for the entire set of criteria. Therefore, we simply aim to identify a hierarchy in which a subgroup of Z-score measures is dominated on average by other more efficient ones.

3.2 Results

Our test results are reported in Table 2, with testing carried out separately for the US and European samples (see Panels A and B, respectively). Testing by region is done, firstly, due to the significant difference in sample size, and secondly to allow us to examine whether results differ between the US and European samples.

We first evaluate the performance of the different Z-scores when calculated using moving moments, using 3-year windows. The two traditional, widely used ROA-based Z-scores, $Z^{r3,c}$ and Z^{r3} , computed with either current values or moving mean of the capital-asset ratio, are compared with our novel regulatory capital Z-score Z_{RCAP}^{r3} . We observe throughout that our results are quali-

⁶Flach et al. (2011) show that the H measure is actually a variation of the area under the cost curve developed by Drummond and Holte (2006).

tatively similar for the US and European samples. The regulatory capital Z-score dominates the two traditional ROA-based Z-scores for all test criteria but the AUPR curve for both the US and the European sample. When choosing more narrowly only between the two traditional ROA-based Z-scores, the one computed with current values of the capital-asset ratio ($Z^{r3,c}$) is strictly preferable throughout.

We next investigate whether time-varying Z-scores computed using exponentially weighted moments perform better than those using simple moving moments. Results in Table 2 indicate that, for the US sample, exponentially weighted Z-scores dominate corresponding moving moments Z-scores using 3-year windows for all criteria, with the only exception of the AUROC curve for the regulatory Z-score. Test results are, however, less clear-cut as to whether the exponentially weighted "current capital" Z-score ($Z^{\text{exp},c}$) or the exponentially weighted regulatory capital Z-score (Z_{RCAP}^{exp}) are best, with each performing well for certain criteria in the US sample. For the European sample, on the other hand, exponentially weighted Z-scores do not outperform their corresponding moving moments Z-scores based on all criteria but the Tjur R^2 .

Additionally, we carried out equivalent comparisons using Z-scores based on moving moments using 5-year windows instead of the more common 3-year ones; this, however, marginally reduces the number of usable observations. We find that Z-score measures based on 3-year windows outperform the Z-scores using 5-year windows for all the criteria we use to compare their performance, for both the US and Europe (see Table B3 in Appendix B).

Overall, our analysis suggests that for US data, it is better to use Z-scores computed with exponentially weighted moments rather than those with moving moments, and that the two best Z-scores to use are the exponentially weighted regulatory capital Z-score and the exponentially weighted "current capital" Z-score. If only simple moving moments are applied, however, it is better to use the regulatory capital Z-score than traditional Z-scores for US data, using 3-year windows. The latter recommendation also applies to European data; however in this case using exponentially weighted moments furthermore does not offer any relevant significant advantages, with the moving moment regulatory capital Z-score, using 3-year windows, remaining best overall. This result notwithstanding, as the required regulatory data is much less available for European banks it may be preferable to use the moving moment "current capital" Z-score instead.

Whereas our analysis allows the comparison of these different Z-scores measures, drawing on the predicted probabilities that banks become distressed they imply, it does not offer guidance for what particular transformation of those Z-score measures is the most appropriate to use in applied empirical work. As Lepetit and Strobel (2015) emphasize, log-transformed Z-scores may be more appropriate in applied work due to the skewness of Z-scores in levels; the log of the Z-score can additionally be shown to be negatively proportional to the log odds of insolvency, giving it a sound probabilistic foundation. However, a practical inconvenience of log-transformed Z-scores is the

fact that calculated Z-scores can be negative by construction. This issue could be resolved either through appropriate rescaling of negative Z-scores into the positive domain (as detailed above), or by resorting to alternative transformations such as the log-modulus transformation, which we introduce in the following section.

Next, we move beyond the probabilistic approach to the evaluation of different Z-score measures that links the Z-score to the probability of bank distress, and introduce an alternative approach based on a parametric econometric model.

4 A multivariate logit model for the probability of bank distress

We now develop a multivariate logit model to link the Z-score to the probability of distress, allowing us to include selected control variables in the specification that represent key observable characteristics at the bank level as well as changes in economic conditions. This parametric econometric model then allows us to apply an alternative range of statistical tests to further assess the different Z-score measures introduced in Section 2.1 in relation to their ability to predict bank insolvency/distress.

4.1 Econometric model specification

The baseline model for the multivariate-model approach corresponds to a pooled logit specification given by

$$\begin{cases} P(Y_{i,t} = 1 \mid \mathbb{Z}_{i,t}^K, X_{i,t}) = \Lambda(\gamma \mathbb{Z}_{i,t}^K, X_{i,t} \beta) \\ P(Y_{i,t} = 0 \mid \mathbb{Z}_{i,t}^K, X_{i,t}) = 1 - \Lambda(\gamma \mathbb{Z}_{i,t}^K, \mathbb{Z}_{i,t} \beta) \end{cases} \quad (1)$$

where the subscripts refer to bank i ($i = 1, \dots, N$) in period t ($t = 1, \dots, T$). The variable $Y_{i,t}$ is a binary variable that is equal to 1 if the bank falls into distress and 0 otherwise, $\mathbb{Z}_{i,t}^K$ is a Z-score measure of type K , $X_{i,t}$ is a set of control variables, γ is a parameter estimate and β a vector of parameter estimates. The cumulative density function $\Lambda(\cdot)$ is given by

$$\Lambda(\gamma \mathbb{Z}_{i,t}^K, \mathbb{Z}_{i,t} \beta) = \frac{1}{1 + \exp[-\gamma \mathbb{Z}_{i,t}^K - X_{i,t} \beta - c]} \quad (2)$$

where c is the intercept. The multivariate logit model allows us to compute alternative augmented (model-based) probabilities of distress, denoted $P(\mathbb{Z}_{i,t}^K, X_{i,t})$.

The control variables ($X_{i,t}$) take into account selected observable characteristics at the bank level and changes in economic conditions. We select bank-level control variables that are publicly available and provided by most, if not all, banks to ascertain that they can be used universally. We firstly control for bank size by including the log of total assets ($Size_{i,t}$), but also examine

additional inclusion of the square of the log of total assets ($(Size_{i,t})^2$). Economies of scale and scope in information production might imply that larger banks should be less prone to falling into distress and be more efficient than smaller ones (e.g. Diamond (1984), Williamson (1986), and Boyd and Prescott (1986)). On the other hand, large banks might also have incentives to pursue higher asset risk relative to smaller banks in response to "too-big-to-fail" subsidies and government bailouts (e.g. Farhi and Tirole (2012)). We additionally take into account if a bank is subject to market scrutiny by including the dummy variable $Listed_{i,t}$ equal to 1 if bank i is listed on a public stock market, and 0 otherwise. Market discipline has the potential to curb incentives for excessive risk taking, by rendering risk-taking more costly (Kwan (2004), Nier and Baumann (2006)). Lastly, we include the VIX (VIX_t) to account for changes in economic conditions; this is preferred to time dummies, as those would not be appropriate for out-of-sample use of the estimates.

The Z-score measure ($Z_{i,t}^K$) used here is the log-modulus transformation (John and Draper (1980)) of the Z-score variable $Z_{i,t}^K$ defined as

$$Z_{i,t}^K = \text{sign}(Z_{i,t}^K) \cdot \ln(|Z_{i,t}^K| + 1)$$

where $\text{sign}(Z_{i,t}^K)$ is equal to -1 if the variable $Z_{i,t}^K$ is negative and 1 otherwise. The log-modulus transformation allows us to both handle the skewedness of the distribution of the Z-score and the potential negative values taken by the Z-score when returns are negative and equity drops to sufficiently low levels. The latter aspect may be particularly relevant in studies like ours dealing specifically with bank distress, as it would lead to exclusion of such observations, or require some rescaling, when applying a simple log-transformation.

4.2 Testing procedure

The comparison of different Z-score measures (for a given model specification) is equivalent to the comparison of non-nested logit models. For this we rely on the Vuong (1989) test and the Clarke (2003) test. These two procedures focus on the relative strength of two rival models and exploit the differences in the log-likelihoods of two competing models. The non-nested tests are more appropriate than familiar model selection criteria such as the McFadden R^2 or the AIC and BIC criteria because non-nested tests provide probabilistic statements regarding model selection. In addition, the non-nested tests include information from the rival model in the selection procedure, unlike the common model selection criteria.

More specifically, Vuong (1989) proposes a procedure to discriminate between two rival models. He puts forward that the first model should be selected over the second model if the average log-likelihood of the first model is significantly greater than the average log-likelihood of the second model, and conversely. Thus, the Vuong test uses the average difference in the log-likelihoods of two

competing specifications. The null hypothesis is that the average difference is zero, suggesting that the two rival models are equally close to the true model. The test statistic, based on an appropriate normalization of the estimated average difference, is normally distributed under the null hypothesis. Furthermore, when the null hypothesis is rejected, the first model should be selected if the test statistic is positive, while the second model should be selected if the test statistic is negative.

Additionally, Clarke (2003) shows that some conditions affect the performance of the Vuong test. In particular, a small sample size, a strong canonical correlation between the sets of covariates in the two rival models and the presence of outliers in the (individual) log-likelihood ratios reduce the performance of the Vuong test. Clarke (2003) proposes a more robust selection procedure in such conditions. The Clarke test, also called the distribution-free test, is based on the median difference in the log-likelihoods of two competing specifications. The null hypothesis is that the two specifications are equally close to the true model. Therefore, under the null hypothesis, half of the differences of the log-likelihoods should be greater than zero. The number of positive differences is used to compute the test statistic that is distributed binomial $(n, 0.5)$, where n is the number of observations.⁷ Furthermore, when the null hypothesis is rejected, the number of positive differences is significantly different from 0.5, suggesting that the median log-likelihood ratio is statistically different from zero. A number of positive differences significantly larger than 0.5 means that the first model should be selected, whereas a number of positive differences significantly lower than 0.5 means that the second model should be selected.

Finally, we also rely on the AUROC curve which is reliable for assessing whether one of the Z-score measures performs better for the detection of distressed banks. More precisely, we consider the statistical test proposed by DeLong et al. (1988) to compare two AUROC curves.

4.3 Results

Tables 3 and 4 present the results of the multivariate logit regressions for the Z-scores we introduced in Section 2.1, for the US and European sample, respectively.

For the European sample, the square of the log of total assets $((Size_{i,t})^2)$ is not significant throughout, hence we resort to a specification only including the log of total assets $(Size_{i,t})$ in this case (see Table 4). This is plausibly due to the fact that our sample of distressed banks in Europe is dominated by large banks, contrary to the US sample. Furthermore, as inclusion of the regulatory capital Z-score Z_{RCAP}^3 when carrying out the tests outlined in Section 4.2 would dramatically reduce the size of the available European sample (as discussed above), we focus only on ROA-based Z-scores when examining the European sample.

⁷The competing specifications we consider have the same number of coefficients; hence, we do not need to rely on a correction for the degrees of freedom.

Our main results show that the addition of control variables in the four multivariate models we consider results in a statistically significant improvement in model fit, obtaining higher Log-Likelihoods than with univariate models, with large LR test statistics associated with very low p-values. The different multivariate models also have larger McFadden R^2 and lower values for AIC and BIC compared to the univariate versions.

Using the multivariate logit approach, we then re-examine the performance of the different Z-score measures introduced in Section 2.1, with results presented in Tables 5 and 6, for the US and Europe, respectively. We observe for both the US and Europe that the "current capital" Z-score $Z^{r3,c}$ dominates the "moving capital" Z-score Z^{r3} throughout in the multivariate context. Similarly, again for both the US and European samples, while the exponentially weighted "current capital" Z-score ($Z^{\text{exp},c}$) strictly dominates the "moving capital" Z-score Z^{r3} in the multivariate context, its performance relative to the "current capital" Z-score $Z^{r3,c}$ is far from conclusive. As far as the regulatory capital Z-scores Z_{RCAP}^{r3} and Z_{RCAP}^{exp} are concerned for the US sample, both are strictly dominated by the "current capital" Z-score $Z^{r3,c}$ in the multivariate context.

Overall, our results for both the US and European samples suggest that in a multivariate context, which also takes into account selected observable characteristics at the bank level and changes in economic conditions, the "current capital" Z-score $Z^{r3,c}$ and exponentially weighted "current capital" Z-score ($Z^{\text{exp},c}$) jointly outperform all other considered Z-score measures, without however a clear "winner" between the two of them.

In the next section, we compare the probabilistic approach linking the Z-score to the probability of bank distress, as discussed in Section 3, with the alternative approach based on a multivariate logit model developed here, allowing us to develop an "augmented" Z-score measure as a consequence.

5 Towards an augmented Z-score

We first compare the performance of the multivariate logit approach with that of the probabilistic one, using the same criteria as in Section 3.2 (AUROC curve, AUPR curve, H statistic and Tjur R^2). The results displayed in Table 7, for both the US and European sample, show that the multivariate logit approach outperforms the probabilistic approach for three of the four criteria considered (AUROC curve, AUPR curve and H measure). The fact that the probabilistic approach seemingly outperforms the multivariate model approach according to the Tjur R^2 should be treated with caution, as it is mainly driven by the properties of the link functions between the Z-score measure and the probability of distress.⁸ The dominant performance of the multivariate logit

⁸Note that the probability of distress increases faster with decreasing Z-scores with the link function used in the probabilistic approach than with the one used in the model-based approach (see Figure C1 in the Appendix). As a result, if a given Z-score performs well in discriminating between distressed and surviving banks, the mean of the predicted probabilities of positive outcomes will be higher under the probabilistic approach than the model-

approach shows that the inclusion of selected observable characteristics at the bank level and changes in economic conditions can give more accurate probabilities of distress to discriminate between distressed banks and survivor banks.

In terms of practical application, we can draw on these results by constructing a formula defining an "augmented" Z-score that can be used in out-of-sample applications, as an alternative to the "unconditional" Z-score commonly used in empirical work in the banking and financial stability related literature. Formally, our multivariate model framework allows the computation of a formula for an augmented Z-score measure $\hat{Z}_{i,t}^K$, defined as

$$\hat{Z}_{i,t}^K = -\hat{\gamma}Z_{i,t}^K - X_{i,t}\hat{\beta}$$

For the best Z-score measure we identified in the multivariate model approach, i.e. the ROA-based Z-score using current values of the capital-asset ratio, we can retrieve the relevant estimated parameters $\hat{\gamma}$ and $\hat{\beta}$ from Tables 3 and 4. The corresponding augmented ROA-based Z-score measure $Z_{i,t}^{r3,c}$, computed using 3-year rolling windows (for simplicity) and current values of the capital-asset ratio, can then be given as

$$\begin{aligned} \text{for the US: } \hat{Z}_{i,t}^{r3,c} &= 1.6181Z_{i,t}^{r3,c} - 0.4014Listed_{i,t} - 1.1543Size_{i,t} + 0.0432(Size_{i,t})^2 - 0.1178VIX_t \\ \text{for Europe: } \hat{Z}_{i,t}^{r3,c} &= 1.1306Z_{i,t}^{r3,c} - 0.6654Listed_{i,t} - 0.2603Size_{i,t} - 0.0585VIX_t \\ \text{where: } Z_{i,t}^{r3,c} &= sign(Z_{i,t}^{r3,c}) \cdot \ln(|Z_{i,t}^{r3,c}| + 1) \end{aligned} \quad (3)$$

The associated probability of distress for this augmented Z-score is given by the respective logit functions $[1 + \exp(\hat{Z}_{i,t}^{r3,c} + 11.1360)]^{-1}$ for the US, and $[1 + \exp(\hat{Z}_{i,t}^{r3,c} + 4.7536)]^{-1}$ for Europe. Thus, the estimated coefficients can be used as optimal weights in an out-of-sample perspective that make the Z-score measure conditional on economic conditions and banks' observable characteristics, and thereby provides an improved assessment of a bank's risk of distress.

We can use two graphical representations to illustrate the main properties of this augmented Z-score. In Figure 1, the Z-score measure is on the horizontal axis, whereas the vertical axis gives the probability of distress. Figure 1 shows that, for a given Z-score, the probability of distress is not unique in the multivariate logit model as economic conditions and banks' observable characteristics change for each observation. In particular, Figure 1 shows that the differences in the probabilities of distress for a given Z-score can be substantial when Z-scores are low, i.e. when banks face a greater risk of distress. Therefore, the augmented Z-score measures computed from the multivariate

based approach; hence, the Tjur R^2 will conclude that the probabilistic approach outperforms the model-based approach. This result highlights that for users that are sensitive to the numerical values of the predicted probabilities, the probabilistic approach has the interesting feature of generating on average a large gap between the respective predicted probabilities of distressed and surviving banks.

logit model allow the introduction of heterogeneity in the relationship between the Z-score and the probability of distress. This heterogeneity between banks and over time ensures an improved assessment of the probability of distress of banks.

We can also illustrate the main advantage of the augmented Z-score measure in the time dimension. More precisely, we can use the multivariate logit model to construct a Z-score map that illustrates how the relationship between the Z-score and the probability of distress is affected by changes in economic conditions over time (i.e. the VIX). Figures 2 and 3 show to what extent the probability of distress changes when both the Z-score measure and the VIX vary. In other words, the non-linearity of the logit model implies that the marginal effect of the Z-score on the probability of distress is conditional on the level of the control variables, including the economic conditions proxied by the VIX. Thus, Figures 2 and 3 show, for the US and Europe, that in a high volatility environment (i.e. $VIX > 40$) the probabilities of distress start to increase when the Z-score measure goes below 5, whereas this increase only starts when the Z-score measure drops below three in a low volatility environment (i.e. $VIX < 15$). In addition, the slopes of the surfaces in Figures 2 and 3 show that a decline in the Z-score measure can lead rapidly to a high probability of distress in a high volatility environment, whereas a similar decline will not have a substantial effect on the risk of distress in a low volatility environment.

6 Conclusion

We examine traditional and alternative versions of Z-score measures, a commonly used, accounting data based bank insolvency risk measure. After highlighting various caveats arising in the practical application of traditional ROA-based Z-scores, we develop alternative Z-score measures that aim to resolve these issues, and then proceed to recommendations for best practice in the application of Z-score measures for the measurement of bank insolvency risk in the empirical banking and financial stability related literature. Our empirical analysis is focussed on the financial crisis of 2007-2008, using annual financial statement data for US and European banks for the period 2006-2014. We carry out comparisons of the different Z-score measures considered using a range of alternative testing procedures, drawing on both a probabilistic approach as well as a multivariate logit approach; the latter further conditions on selected observable bank characteristics and the macroeconomic environment.

Our results demonstrate that our novel regulatory capital Z-score dominates the two versions of traditional ROA-based Z-score measures we examine using the probabilistic approach, for both the US and European samples. Examining how best to allow for time-variation in Z-score measures, we find using the probabilistic approach that Z-scores computed with exponentially weighted moments are better than those with moving moments for the US sample, but have no clear advantage

for the European one. Furthermore, moving moment Z-score measures based on 3-year windows outperform Z-scores using 5-year windows for both the US and Europe. Using a multivariate logit approach, which includes selected observable characteristics at the bank level and changes in economic conditions, we are able to compute augmented Z-scores that provide probabilities of distress that better discriminate between distressed and surviving banks than the probabilistic approach, for both the US and European samples. Based on the multivariate logit approach, our results suggest that the best Z-score to use, for both the US and Europe, is the ROA-based Z-score using current values of the capital-asset ratio, calculated either with moving or exponentially weighted moments.

References

- Beck, T., O. De Jonghe, and G. Schepens (2013). Bank competition and stability: cross-country heterogeneity. *Journal of Financial Intermediation* 22(2), 218–244.
- Berger, A. N., C. H. Bouwman, T. Kick, and K. Schaeck (2016). Bank liquidity creation following regulatory interventions and capital support. *Journal of Financial Intermediation* 26, 115–141.
- Berger, A. N., W. Goulding, and T. Rice (2014). Do small businesses still prefer community banks? *Journal of Banking & Finance* 44, 264–278.
- Bertay, A. C., A. Demirgüç-Kunt, and H. Huizinga (2013). Do we need big banks? Evidence on performance, strategy and market discipline. *Journal of Financial Intermediation* 22(4), 532–558.
- Boyd, J. H. and E. C. Prescott (1986). Financial intermediary-coalitions. *Journal of Economic Theory* 38(2), 211–232.
- Caiazza, S., M. Cotugno, F. Fiordelisi, and V. Stefanelli (2018). The spillover effect of enforcement actions on bank risk-taking. *Journal of Banking & Finance* 91, 146 – 159.
- Clarke, K. A. (2003). Nonparametric model discrimination in international relations. *Journal of Conflict Resolution* 47(1), 72–93.
- Delis, M. D., I. Hasan, and E. G. Tsionas (2014). The risk of financial intermediaries. *Journal of Banking & Finance* 44, 1–12.
- DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3), 837–845.
- DeYoung, R. and G. Torna (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation* 22(3), 397–421.

- Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies* 51(3), 393–414.
- Doumpos, M., C. Gaganis, and F. Pasiouras (2015). Central bank independence, financial supervision structure and bank soundness: An empirical analysis around the crisis. *Journal of Banking & Finance* 61, S69–S83.
- Drummond, C. and R. C. Holte (2006). Cost curves: an improved method for visualizing classifier performance. *Machine Learning* 65(1), 95–130.
- Fang, Y., I. Hasan, and K. Marton (2014). Institutional development and bank stability: evidence from transition countries. *Journal of Banking & Finance* 39, 160–176.
- Farhi, E. and J. Tirole (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *The American Economic Review* 102(1), 60–93.
- Feller, W. (1971). *Probability Theory and its Applications, vol. II*. New York, NY: John Wiley & Sons.
- Finch, T. (2009). Incremental calculation of weighted mean and variance. Mimeo (University of Cambridge Computing Service).
- Flach, P. A., J. Hernández-Orallo, and C. Ferri (2011). A coherent interpretation of AUC as a measure of aggregated classification performance. In L. Getoor and T. Scheffer (Eds.), *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, New York, NY, USA, pp. 657–664. ACM.
- Fратиanni, M. and F. Marchionne (2013). The banking bailout of the subprime crisis: Was the bang worth the buck? *Journal of International Financial Markets, Institutions and Money* 23, 240–264.
- Fu, X. M., Y. R. Lin, and P. Molyneux (2014). Bank competition and financial stability in Asia Pacific. *Journal of Banking & Finance* 38, 64–77.
- Fukuyama, H. and R. Matousek (2017). Modelling bank performance: a network DEA approach. *European Journal of Operational Research* 259(2), 721–732.
- Hakenes, H., I. Hasan, P. Molyneux, and R. Xie (2015). Small banks and local economic development. *Review of Finance* 19(2), 653–683.

- Han, I., H.-Y. Liang, and K. C. Chan (2016). Locational concentration and institutional diversification: evidence from foreign direct investments in the banking industry. *The North American Journal of Economics and Finance* 38, 185–199.
- Hand, D. J. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning* 77(1), 103–123.
- Hand, D. J. (2010). Evaluating diagnostic tests: the area under the ROC curve and the balance of errors. *Statistics in Medicine* 29(14), 1502–1510.
- Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review* 80(3), 400–414.
- John, J. A. and N. R. Draper (1980). An alternative family of transformations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29(2), 190–197.
- Kwan, S. H. (2004). Risk and return of publicly held versus privately owned banks. *FRBNY Economic Policy Review* 10(2), 97–107.
- Lepetit, L. and F. Strobel (2013). Bank insolvency risk and time-varying Z-score measures. *Journal of International Financial Markets, Institutions and Money* 25, 73–87.
- Lepetit, L. and F. Strobel (2015). Bank insolvency risk and Z-score measures: a refinement. *Finance Research Letters* 13, 214–224.
- Molyneux, P., K. Schaeck, and T. M. Zhou (2014). ‘Too systemically important to fail’ in banking – Evidence from bank mergers and acquisitions. *Journal of International Money and Finance* 49, 258–282.
- Nier, E. and U. Baumann (2006). Market discipline, disclosure and moral hazard in banking. *Journal of Financial Intermediation* 15(3), 332–361.
- Pino, G. and S. C. Sharma (2018). On the contagion effect in the US banking sector. *Journal of Money, Credit and Banking (Early View)*, doi.org/10.1111/jmcb.12489.
- Ross, S. M. (1997). *A First Course in Probability* (5th ed.). Upper Saddle River, N.J.: Prentice Hall.
- Saito, T. and M. Rehmsmeier (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* 10(3), 1–21.
- Tjur, T. (2009). Coefficients of determination in logistic regression models - a new proposal: the coefficient of discrimination. *The American Statistician* 63(4), 366–372.

- Tsionas, E. G. and E. C. Mamatzakis (2017). Adjustment costs in the technical efficiency: an application to global banking. *European Journal of Operational Research* 256(2), 640–649.
- Vazquez, F. and P. Federico (2015). Bank funding structures and risk: evidence from the global financial crisis. *Journal of Banking & Finance* 61, 1–14.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2), 307–333.
- West, D. (1979). Updating mean and variance estimates: an improved method. *Communications of the ACM* 22(9), 532–535.
- Williamson, S. D. (1986). Costly monitoring, financial intermediation, and equilibrium credit rationing. *Journal of Monetary Economics* 18(2), 159–179.

Table 1: Sample characteristics

Panel A: USA

Year	2007	2008	2009	2010	2011	2012	2013	2014	All sample
Number of observations	5,692	5,780	5,589	5,447	5,373	5,327	5,304	5,285	43,806
Number of distressed banks	43	191	132	84	46	23	19	5	543
Percentage of distressed banks	0.76	3.30	2.36	1.54	0.86	0.43	0.36	0.09	1.24

Panel B: Europe

Year	2007	2008	2009	2010	2011	2012	2013	2014	All sample
Number of observations	2,655	2,694	2,636	2,636	2,640	2,643	2,623	2,598	21,125
Number of distressed banks	15	46	22	18	13	15	10	3	142
Percentage of distressed banks	0.56	1.71	0.83	0.68	0.49	0.57	0.38	0.12	0.67

Table 2: Comparison of Z-score measures (probabilistic approach)

Panel A: USA

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	$Z^{\text{exp},c}$	Z_{RCAP}^{exp}
AUROC curve	0.9256	0.9376	0.9379	0.9438	0.9371
AUPR curve	0.3025	0.4250	0.2668	0.4400	0.2728
H measure	0.6341	0.6682	0.6760	0.6709	0.6867
Tjur R^2	0.3171	0.5456	0.5875	0.6257	0.7418
Observations	43,806	43,806	43,806	43,806	43,806
Number of banks	5,823	5,823	5,823	5,823	5,823
Number of distressed banks	543	543	543	543	543

Panel B: Europe

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	$Z^{\text{exp},c}$	Z_{RCAP}^{exp}
AUROC curve	0.8595	0.8684	0.9151	0.8664	0.9111
AUPR curve	0.1263	0.2013	0.1746	0.1947	0.1214
H measure	0.4012	0.4336	0.6759	0.4079	0.5944
Tjur R^2	0.1646	0.2277	0.4718	0.2642	0.5849
Observations	21,125	21,125	9,524	21,125	9,524
Number of banks	2,866	2,866	2,138	2,866	2,138
Number of distressed banks	142	142	60	142	60

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): Z^{r3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z_{RCAP}^{r3} = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments: $Z^{\text{exp},c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.28$ for the US and 0.54 for Europe; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha = 0.82$ for the US and 0.53 for Europe. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model. Four criteria are used to compare the performance of the different rival classifiers (AUROC curve, AUPR curve, H measure and Tjur R^2). Best Z-scores present the highest values for the AUROC curve, the AUPR curve, the H measure and the Tjur R^2 .

Table 3: Multivariate logit models (USA)

	(1)	(2)	(3)	(4)	(5)
$Z_{i,t}^{r3}$	-1.7386*** (0.0662)				
$Z_{i,t}^{r3,c}$		-1.6181*** (0.0528)			
$Z_{RCAP,i,t}^{r3}$			-2.3353*** (0.1177)		
$Z_{i,t}^{exp,c}$				-1.6200*** (0.0800)	
$Z_{RCAP,i,t}^{exp}$					-1.2787*** (0.0615)
$Listed_{i,t}$	0.4731*** (0.1352)	0.4014*** (0.1421)	0.4429*** (0.1424)	0.3613*** (0.1190)	0.3345** (0.1299)
$Size_{i,t}$	1.1526*** (0.4778)	1.1543*** (0.4394)	0.8073** (0.3773)	1.6876*** (0.5103)	1.3042*** (0.4085)
$(Size_{i,t})^2$	-0.0650** (0.0189)	-0.0432** (0.0173)	-0.0293** (0.0140)	-0.0648*** (0.0201)	-0.0440*** (0.0153)
VIX_t	0.1266*** (0.0096)	0.1178*** (0.0102)	0.1886*** (0.0123)	0.1127*** (0.0082)	0.1420*** (0.0103)
<i>Intercept</i>	-13.0309*** (3.0375)	-11.1360*** (2.8142)	-11.4991*** (2.5573)	-14.5717*** (3.2416)	-15.1030*** (2.7231)
N obs.	43,806	43,806	43,806	43,806	43,806
N banks	5,823	5,823	5,823	5,823	5,823
N failures.	543	543	543	543	543
Log-likelihood	-1700.29	-1524.30	-1703.00	-1606.13	-1726.80
LR test [p-value]	264.40 [0.0000]	206.63 [0.0000]	492.26 [0.0000]	232.42 [0.0000]	355.86 [0.0000]
McFadden R ²	0.4184	0.4786	0.4175	0.4506	0.4094
AIC	3412.59	3060.60	3418.01	3224.27	3465.61
BIC	3464.71	3112.73	3470.13	3276.40	3517.74

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): Z^{r3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z_{RCAP}^{r3} = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments: $Z^{exp,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha=0.28$; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha=0.82$. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model. Control variables: $Listed_i = 1$ if bank i is listed on a public stock market, and 0 otherwise; $Size = \log$ of total assets; VIX = the CBOE Volatility Index. ***, ** and * indicate significance respectively at the 1%, 5% and 10% levels. Cluster (at the bank level) robust standard deviations are in brackets.

Table 4: Multivariate logit models (Europe)

	(1)	(2)	(3)	(4)	(5)	(6)
$Z_{i,t}^3$	-1.1247*** (0.0791)			-1.1265*** (0.0784)		
$Z_{i,t}^{3,c}$		-1.1297*** (0.0701)			-1.1306*** (0.0697)	
$Z_{i,t}^{exp,c}$			-0.6943*** (0.0455)			-0.6927*** (0.0448)
$Listed_{i,t}$	0.5982** (0.2725)	0.6041** (0.2779)	0.5283* (0.2744)	0.6523** (0.2625)	0.6654** (0.2675)	0.5937** (0.2641)
$Size_{i,t}$	0.0138 (0.2087)	-0.0106 (0.2068)	0.0330 (0.1963)	0.2593*** (0.0387)	0.2603*** (0.0381)	0.3207*** (0.0385)
$(Size_{i,t})^2$	0.0133 (0.0115)	0.0147 (0.0114)	0.0157 (0.0110)			
VIX_t	0.0603*** (0.0132)	0.0572*** (0.0131)	0.0572*** (0.0134)	0.0614*** (0.0131)	0.0585*** (0.0131)	0.0588*** (0.0133)
<i>Intercept</i>	-3.6943*** (1.0491)	-3.5908*** (1.0271)	-4.6155*** (1.0021)	-4.7485*** (0.6135)	-4.7536*** (0.5761)	-5.8613*** (0.5983)
N obs.	21,125	21,125	21,125	21,125	21,125	21,125
N banks	2,866	2,866	2,866	2,866	2,866	2,866
N failures.	142	142	142	142	142	142
Log-likelihood	-623.16	-597.72	-606.83	-624.04	-598.77	-608.10
LR test [p-value]	104.45 [0.0000]	101.11 [0.0000]	131.31 [0.0000]	102.69 [0.0000]	99.01 [0.0000]	128.76 [0.0000]
McFadden R ²	0.2685	0.2983	0.2876	0.2674	0.2971	0.2861
AIC	1258.32	1207.45	1225.66	1258.08	1207.55	1226.21
BIC	1306.07	1255.20	1273.41	1297.88	1247.35	1266.00

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): Z^{r3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio. Z-scores computed with exponentially weighted moments: $Z^{exp,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.54$. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model. Control variables: $Listed_i = 1$ if bank i is listed on a public stock market, and 0 otherwise; $Size = \log$ of total assets; $VIX =$ the CBOE Volatility Index. ***, ** and * indicate significance respectively at the 1%, 5% and 10% levels. Cluster (at the bank level) robust standard deviations are in brackets.

Table 5: Comparison of Z-score measures with non-nested tests (USA)

$Model\ 2 \setminus Model\ 1$	Z-score: Z^{r3} Log-likelihood=-1700 AUROC curve=0.9415	Z-score: $Z^{r3,c}$ Log-likelihood=-1524 AUROC curve=0.9499	Z-score: Z_{RCAP}^3 Log-likelihood=-1703 AUROC curve=0.9514	Z-score: $Z^{exp,c}$ Log-likelihood=-1606 AUROC curve=0.9532
Z-score: $Z^{r3,c}$ Log-likelihood=-1524 AUROC curve=0.9499	Vuong = -12.89 [0.0000] Clarke = 0.33 [0.0000] Auroc = 69.38 [0.0000]			
Z-score: Z_{RCAP}^3 Log-likelihood=-1703 AUROC curve=0.9514	Vuong = 0.08 [0.9360] Clarke = 0.52 [0.0000] Auroc = 5.40 [0.0201]	Vuong = 5.23 [0.0000] Clarke = 0.54 [0.0000] Auroc = 0.14 [0.7119]		
Z-score: $Z^{exp,c}$ Log-likelihood=-1606 AUROC curve=0.9532	Vuong = -4.01 [0.0000] Clarke = 0.23 [0.0000] Auroc = 11.35 [0.0000]	Vuong = 3.54 [0.0003] Clarke = 0.25 [0.0000] Auroc = 0.98 [0.3231]	Vuong = -2.71 [0.0067] Clarke = 0.31 [0.0000] Auroc = 0.22 [0.6362]	
Z-score: Z_{RCAP}^{exp} Log-likelihood=-1726 AUROC curve=0.9479	Vuong = 0.72 [0.4730] Clarke = 0.52 [0.0000] Auroc = 1.56 [0.2124]	Vuong = 5.44 [0.0000] Clarke = 0.54 [0.0000] Auroc = 0.18 [0.6746]	Vuong = 1.23 [0.2152] Clarke = 0.51 [0.0334] Auroc = 1.64 [0.2006]	Vuong = 3.57 [0.0003] Clarke = 0.68 [0.0000] Auroc = 1.70 [0.1922]

Note: The null hypotheses of the Vuong test and the Clarke test indicate that the two Z-score measures perform similarly. Positive and significant values for the Vuong test indicate that the first measure (model 1) performs better than the second measure (model 2); and conversely for negative and significant values. Values significantly higher than 0.5 for the Clarke test indicate that the first measure (model 1) performs better than the second measure (model 2); and conversely for values significantly lower than 0.5. The null hypothesis of the AUROC test is that the AUROC curves are equal in the two specifications (model 1 and model 2).

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z^{r3} = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments: $Z^{exp,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha=0.28$; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha=0.82$. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model.

Table 6: Comparison of Z-score measures with non-nested tests (Europe)

$Model\ 2 \setminus Model\ 1$	Z-score: Z^{r^3} Log-likelihood=-624.04 AUROC curve=0.8987	Z-score: $Z^{r^3,c}$ Log-likelihood=-598.77 AUROC curve=0.9037
Z-score: $Z^{r^3,c}$ Log-likelihood=-598.77 AUROC curve=0.9037	Vuong = -3.91 [0.0001] Clarke = 0.10 [0.0000] Auroc = 6.44 [0.0112]	
Z-score: $Z^{exp,c}$ Log-likelihood=-608.10 AUROC curve=0.9179	Vuong = -2.10 [0.0356] Clarke = 0.34 [0.0000] Auroc = 8.22 [0.0041]	Vuong = 1.21 [0.2236] Clarke = 0.39 [0.0000] Auroc = 4.26 [0.0389]

Note: The null hypotheses of the Vuong test and the Clarke test indicate that the two Z-score measures perform similarly. Positive and significant values for the Vuong test indicate that the first measure (model 1) performs better than the second measure (model 2); and conversely for negative and significant values. Values significantly higher than 0.5 for the Clarke test indicate that the first measure (model 1) performs better than the second measure (model 2); and conversely for values significantly lower than 0.5. The null hypothesis of the AUROC test is that the AUROC curves are equal in the two specifications (model 1 and model 2).

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): $Z^{r^3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio. Z-scores computed with exponentially weighted moments: $Z^{exp,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.54$. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model.

Table 7: Comparing probabilistic and multivariate logit approach using non model-based criteria
 Panel A : USA

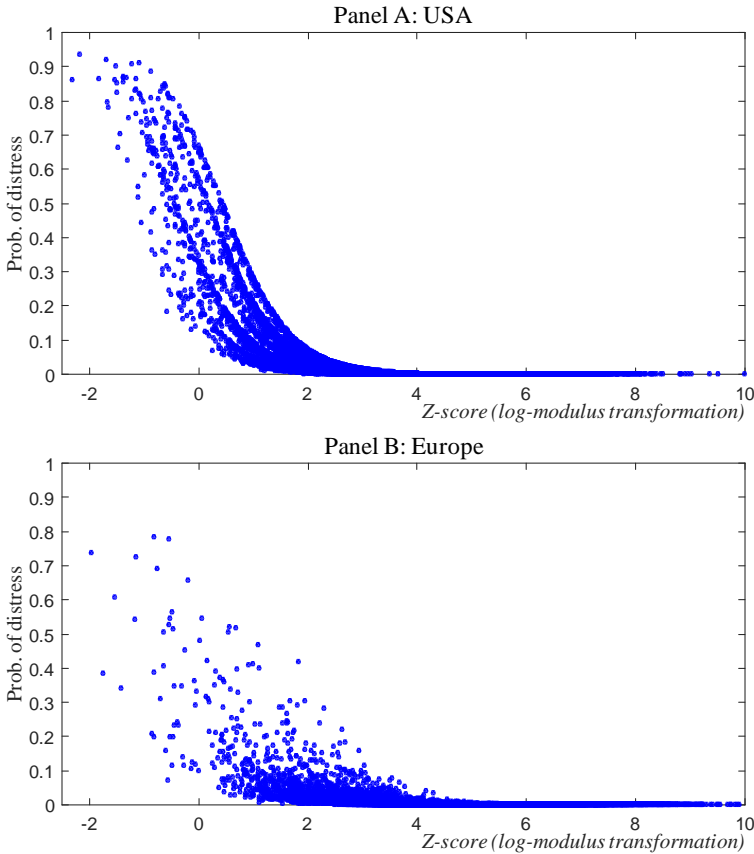
Criteria	Approach	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	$Z^{\text{exp},c}$	Z_{RCAP}^{exp}
AUROC curve	Probabilistic	0.9256	0.9376	0.9379	0.9438	0.9371
	Multivariate logit	0.9415	0.9499	0.9514	0.9532	0.9479
AUPR curve	Probabilistic	0.3025	0.4250	0.2668	0.4404	0.2728
	Multivariate logit	0.3657	0.4693	0.3389	0.4306	0.2890
H measure	Probabilistic	0.6341	0.6682	0.6760	0.6709	0.6867
	Multivariate logit	0.6657	0.6998	0.6993	0.7072	0.7124
Tjur R^2	Probabilistic	0.3171	0.5456	0.5875	0.6257	0.7418
	Multivariate logit	0.2248	0.3102	0.2181	0.2387	0.1870

Panel B: Europe

Criteria	Approach	Z^{r3}	$Z^{r3,c}$	$Z^{\text{exp},c}$
AUROC curve	Probabilistic	0.8595	0.8684	0.8664
	Multivariate logit	0.8987	0.9037	0.9179
AUPR curve	Probabilistic	0.1263	0.2013	0.1947
	Multivariate logit	0.1513	0.2168	0.1572
H measure	Probabilistic	0.4012	0.4336	0.4079
	Multivariate logit	0.5414	0.5708	0.5705
Tjur R^2	Probabilistic	0.1646	0.2277	0.2642
	Multivariate logit	0.0895	0.1232	0.0925

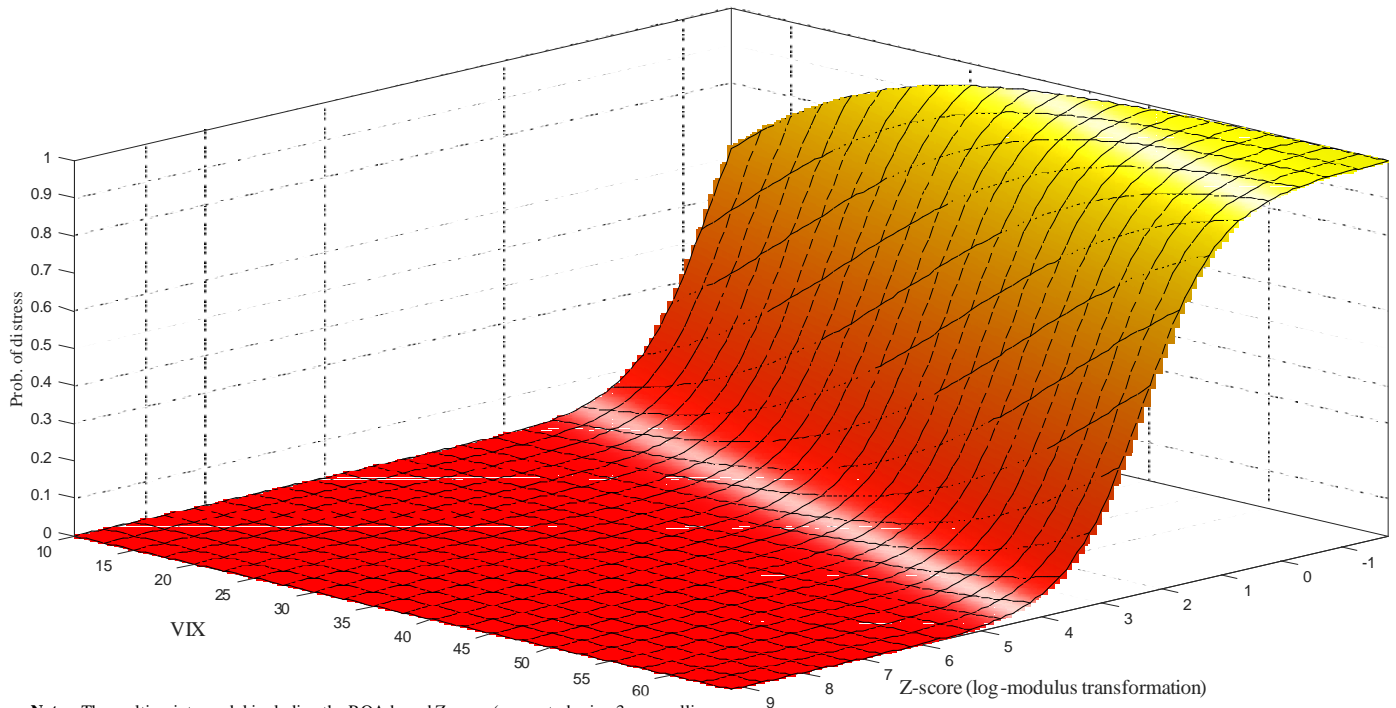
Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details):
 Z^{r3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z_{RCAP}^{r3} = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments:
 $Z^{\text{exp},c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.28$ for the US and 0.54 for Europe; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha = 0.82$ for the US. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model.

Figure 1: Probabilities of distress in the multivariate logit models



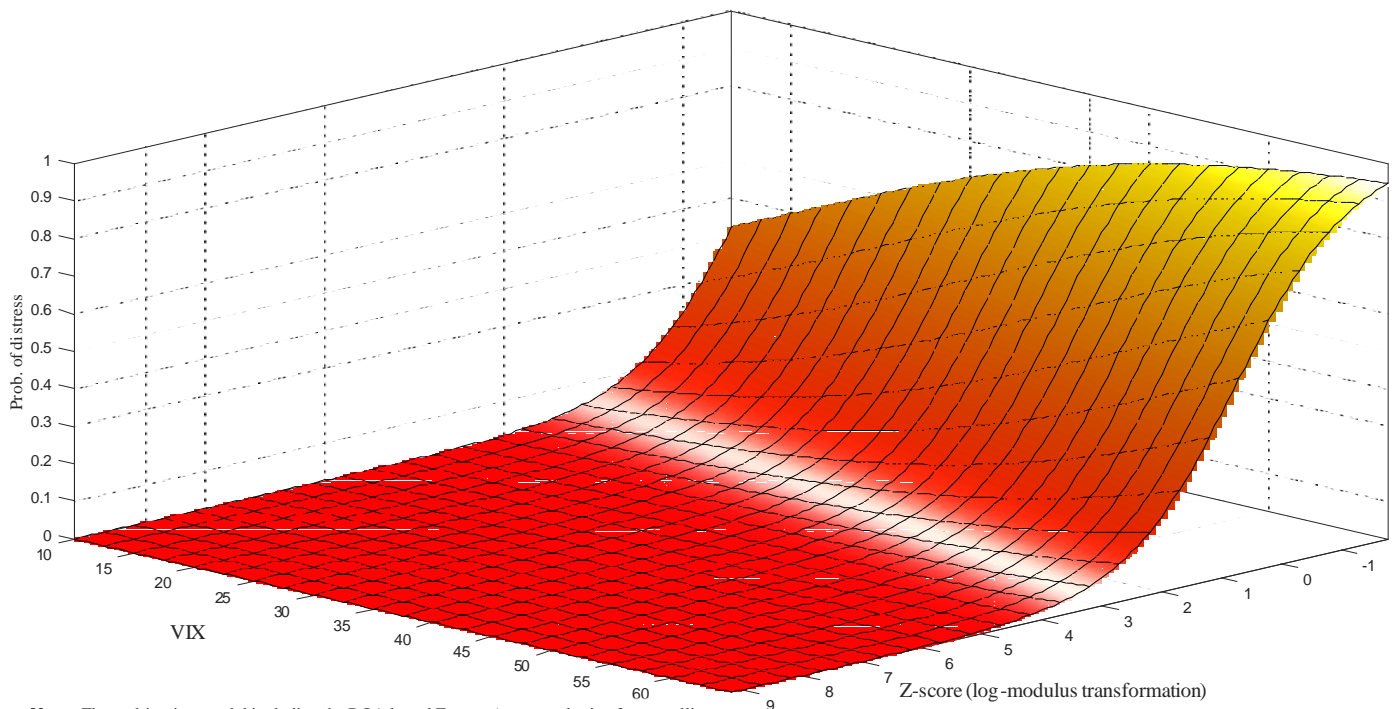
Note: The multivariate model including the ROA-based Z-score (computed using 3-year rolling windows and current values of the capital-asset ratio) is used to estimate the probabilities of distress.

Figure 2: Z-score map : VIX, Z-score and probability of distress (USA)



Note: The multivariate model including the ROA-based Z-score (computed using 3-year rolling windows and current values of the capital-as set ratio) is used to generate this graph. The variables used in the multivariate model other than Z-score and VIX are fixed at their median level.

Figure 3: Z-score map : VIX, Z-score and probability of distress (Europe)



Note: The multivariate model including the ROA-based Z-score (computed using 3-year rolling windows and current values of the capital-as set ratio) is used to generate this graph. The variables used in the multivariate model other than Z-score and VIX are fixed at their median level.

Appendix

A Proof of Proposition 1

Proof. This is an application of the one-sided Chebyshev inequality (see Ross, 1997, p. 414, or previously, Feller, 1971, p. 152): it states that for a random variable X with finite mean μ and variance σ^2 , it holds for any $a > 0$ that $P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$. Our result then follows setting $X = RCAR$ and $a = \mu_{RCAR} - TR > 0$, and dividing both numerator and denominator of the right hand side of the inequality by σ_{RCAR}^2 ; we observe that $\lim_{Z \rightarrow 0} (1 + Z^2)^{-1} = 1$. ■

B Additional tables

Table B1: Descriptive statistics of Z-score measures

Panel A: USA

Z-score	Banks	Obs.	Mean	Median	Std. Dev.	Min	Max
Z^{r^3}	Non-dist.	43263	102.39	55.03	234.25	-3.26	20785.18
	Distressed	543	9.57	1.93	27.05	-5.23	354.75
$Z^{r^3,c}$	Non-dist.	43263	102.68	55.27	236.54	-7.76	21610.53
	Distressed	543	7.56	0.69	26.33	-9.18	337.75
$Z_{RCAP}^{r^3}$	Non-dist.	43263	13.44	8.15	21.76	-47.45	1015.53
	Distressed	543	1.18	0.56	4.30	-23.55	65.97
$Z^{exp,c}$	Non-dist.	43263	383.25	141.98	797.43	-1.34	36131.03
	Distressed	543	14.74	0.29	80.16	-3.21	1180.86
Z_{RCAP}^{exp}	Non-dist.	43263	38.20	14.47	90.00	-60.69	4946.20
	Distressed	543	2.06	0.00	10.85	-12.21	107.83

Panel B: Europe

Z-score	Banks	Obs.	Mean	Median	Std. Dev.	Min	Max
Z^{r^3}	Non-dist.	20,983	286.29	73.70	812.86	-2.29	53000.73
	Distressed	142	24.55	13.03	54.44	-4.59	563.63
$Z^{r^3,c}$	Non-dist.	20,983	300.85	74.29	875.55	-3.10	57114.23
	Distressed	142	23.24	11.47	56.70	-27.84	586.92
$Z_{RCAP}^{r^3}$	Non-dist.	9,464	14.71	9.06	35.13	-154.06	2468.94
	Distressed	60	2.30	1.01	3.98	-0.62	19.46
$Z^{exp,c}$	Non-dist.	20,983	9983.95	511.70	56657.41	-0.54	2719910
	Distressed	142	110.63	21.16	362.92	-47.00	3978.66
Z_{RCAP}^{exp}	Non-dist.	9,464	42.61	9.66	661.65	-14947.13	51995.07
	Distressed	60	2.46	0.30	8.74	-0.20	65.16

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): Z^{r^3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r^3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; $Z_{RCAP}^{r^3}$ = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments: $Z^{exp,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.28$ for the US and 0.54 for Europe; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha = 0.82$ for the US and 0.53 for Europe. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model. Four criteria are used to compare the performance of the different rival classifiers (AUROC curve, AUPR curve, H measure and Tjur R^2).

Table B2: Correlation of Z-score measures (log transformations) and total capital ratio

Panel A: USA

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	$Z^{\text{exp},c}$	Z_{RCAP}^{exp}	TCR
Z^{r3}	1					
$Z^{r3,c}$	0.9938	1				
Z_{RCAP}^{r3}	0.4309	0.4393	1			
$Z^{\text{exp},c}$	0.8257	0.8348	0.4543	1		
Z_{RCAP}^{exp}	0.3806	0.4070	0.7981	0.4960	1	
TCR	0.2204	0.2277	0.2188	0.1838	-0.0227	1

Panel B: Europe

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	$Z^{\text{exp},c}$	Z_{RCAP}^{exp}	TCR
Z^{r3}	1					
$Z^{r3,c}$	0.9964	1				
Z_{RCAP}^{r3}	0.2435	0.2380	1			
$Z^{\text{exp},c}$	0.8676	0.8731	0.2301	1		
Z_{RCAP}^{exp}	0.2349	0.2301	0.7999	0.2459	1	
TCR	0.0231	0.0256	0.1070	-0.0231	-0.1711	1

Variable definitions, Z-scores computed with 3-year rolling windows (see Section 2 for more details): Z^{r3} = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z^{r3,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z_{RCAP}^{r3} = Regulatory capital Z-score (using TCR and 8% threshold). Z-scores computed with exponentially weighted moments: $Z^{\text{exp},c}$ = ROA-based Z-score computed with current values of the capital-asset ratio with $\alpha = 0.28$ for the US and 0.54 for Europe; Z_{RCAP}^{exp} = Regulatory capital Z-score (using TCR and 8% threshold) with $\alpha = 0.82$ for the US and 0.53 for Europe. The α parameter is provided by a grid search and maximizes the AUROC curve in the multivariate model. Four criteria are used to compare the performance of the different rival classifiers (AUROC curve, AUPR curve, H measure and Tjur R^2).

Table B3: Comparison of Z-score measures (3 and 5 year rolling windows, probabilistic approach)
 Panel A: USA

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	Z^{r5}	$Z^{r5,c}$	Z_{RCAP}^{r5}
AUROC curve	0.9364	0.9429	0.9377	0.9266	0.9401	0.9328
AUPR curve	0.3207	0.4479	0.2711	0.2273	0.4324	0.2217
H measure	0.6882	0.7117	0.6889	0.6416	0.6937	0.6552
Tjur R^2	0.3477	0.5871	0.6140	0.1672	0.5095	0.4964
Observations	43,205	43,205	43,205	43,205	43,205	43,205
Number of banks	5,769	5,769	5,769	5,769	5,769	5,769
Number of distressed banks	489	489	489	489	489	489

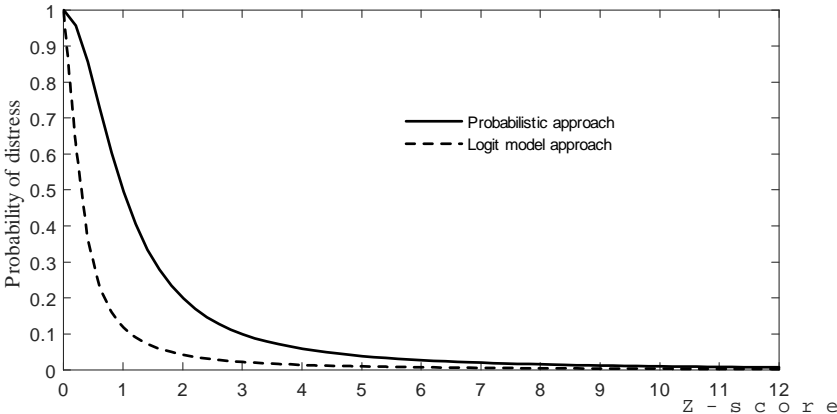
Panel B: Europe

	Z^{r3}	$Z^{r3,c}$	Z_{RCAP}^{r3}	Z^{r5}	$Z^{r5,c}$	Z_{RCAP}^{r5}
AUROC curve	0.8653	0.8747	0.9210	0.8616	0.8748	0.9552
AUPR curve	0.1314	0.2049	0.2782	0.0950	0.1982	0.2235
H measure	0.4172	0.4528	0.7487	0.4034	0.4375	0.7215
Tjur R^2	0.1700	0.2307	0.5222	0.1129	0.2052	0.4453
Observations	20,405	20,405	5,565	20,405	20,405	5,565
Number of banks	2,846	2,846	1,787	2,846	2,846	1,787
Number of distressed banks	136	136	47	136	136	47

Variable definitions, Z-scores computed with n-year rolling windows (with n=3,5; see Section 2 for more details):
 Z_{ROA}^n = ROA-based Z-score computed with moving mean of the capital-asset ratio; $Z_{ROA}^{n,c}$ = ROA-based Z-score computed with current values of the capital-asset ratio; Z_{ROE}^n = ROE-based Z-score; Z_{RCAP}^n = Regulatory capital Z-score (using TCR and 8% threshold). Four criteria are used to compare the performance of the different rival classifiers (AUROC curve, AUPR curve, H measure and Tjur R^2).

C Properties of the link functions

Figure C1. Link function and probabilities of distress



Note: The probability of distress is given by $1/(1+Z\text{-score}^2)$ in the probabilistic approach. The probability of distress in the model-based approach is obtained from an univariate logit model estimated for the US sample and given by $1/(1+\exp(1.6162*\ln(Z\text{-score}+1)+0.3781))$.