Bank i Kredyt 49(2), 2018, 115-144

BANK I KREDYT
BANK & CREDIT

Paths of glory or paths of shame? An analysis of distress events in European banking

Małgorzata Iwanicz-Drozdowska[*], Erkki K. Laitinen[#], Arto Suvas[‡]

Submitted: 16 September 2017. Accepted: 22 January 2018.

Abstract

This paper sheds some new light on banks' distress in Europe, with special attention paid to the period of the global financial crisis (GFC). Unlike in previous research we investigate non-distress ("glory") and distress ("shame") paths of banks from 1 to 4 years prior to a distress event to test how different they are. This approach allows us to outline guidelines for supervisors on how to detect banks generating higher risk of distress several years before its occurrence. We use a balanced panel of data, applying factor and cluster analysis for extraction of distress processes and a logistic regression for distress prediction. We conclude that the differences between distressed and non--distressed banks become more visible 1 and 2 years prior to the distress event. However, liquid assets and loans to assets ratios are significant and stable predictors of banks' distress even 3–4 years in advance.

Keywords: distressed bank, distress process, prediction, variables

JEL: G21, G33

* Warsaw School of Economics; miwani@sgh.waw.pl.
[#] University of Vaasa.
[‡] University of Vaasa.

1 Introduction

During the global financial crisis (GFC), many banks, both in Europe and in the US, faced significant financial troubles and were bailed out by their governments. The question of how to predict a bank's financial distress has been raised many times, starting with Meyer and Pifer (1970). Much research has been carried out for the US banking sector (e.g., Sinkey 1975, 1978; Martin 1977; Santomero, Vinso 1977; Pettway, Sinkey 1980; Cole, Gunther 1995; Peek, Rosengren 1996; Alam et al. 2000; Wheelock, Wilson 2000; Kolari et al. 2002; Cole, White 2012; Shaffer 2012; De Young, Torna 2013; Cox, Wang 2014; López Iturriaga, Sanz 2015; Covas, Rump, Zakrajšek 2014; Kapinos, Mitnik 2016 for stress testing) due to fairly numerous failures of small or medium-sized banks. For Europe, the research is much more limited, e.g. Berg and Hexenberg (1994) and Clare and Priestley (2002) for Norway, Mannasoo and Mayes (2005, 2009) for Central and Eastern Europe, Poghosyan and Cihak (2011) for the EU, Altman, Cizel and Rijken (2014) for the period of GFC (Europe and the US). Single regional or country studies for emerging markets are also available to a limited extent, e.g. Venezuela (Molina 2002), Turkey (Canbas, Cabuk, Kilic 2005; Boyacioglu, Kara, Baykan 2009), Russia (Lanine, Vander Vennet 2006), Turkey, Spain and the US (Chauhan, Ravi, Karthik Chandra 2009), and Gulf Cooperation Council countries (Maghyereh, Awartani 2014).

Predicting bank failures is important for policy makers and financial safety net players. Our study expands research on banks' distress in Europe with special attention paid to the distress and non--distress paths of banks, from 1 to 4 years prior to a distress event. Our goal is to find out how different both paths are and to diagnose if distressed banks follow different processes. To this aim, we build up clusters to account for bank heterogeneity. Against this background we provide evidence that adjusting the model to a distress path improves prediction accuracy in certain cases. Such an approach allows outlining guidelines for supervisors on how to detect banks of various profiles generating a higher risk of distress several years before its occurrence.

This paper is organized as follows. In the next section we present a review of relevant literature on bank distress and distress processes. The third section explains the variables, data and methodology, while the fourth one presents empirical results. The final section concludes.

2 Review of literature

Our review of literature is selective and based on two criteria. First, we wish to show the literature of failure "paths" developed for banks. Second, we want to show how experiences for non-financial companies can be applied to banks. Moreover, a separate review of banks' bankruptcy prediction studies had been conducted to select variables for this study. However, it is not presented in this section (see: Section 3.1).

According to our knowledge, there is a shortage of literature on banks' distress processes, except for two recent studies by Kolari et al. (2002) and Hambusch and Shaffer (2016), both for US banks. Kolari et al. (2002) analysed the models' performance one and two years prior to failures. They found that logit model performance deteriorated over time, while trait recognition results were quite stable. Hambusch and Shaffer (2016) indicated that prediction performance deteriorates over longer forecast horizons. This study accounted for the heterogeneity of banks. Hambusch and Shaffer (2016) attempted to tackle the problem of bank failures by using the leverage ratio (equity to assets ratio) as a continuous variable to predict US banks' problems. For 2000–2011 they registered 441 bank failures, which is much more than distress events occurring among the European banks in that period. Their model offered a reasonable forecasting ability and was capable of using different regressors, estimation techniques and macroeconomic data. However, forecasts for larger banks were less accurate than those for smaller ones. Moreover, the prediction accuracy for the crisis year was lower than for other periods. This divergence implies different distress patterns between large and small banks and at turning points. A different perspective of banks' distress has been presented by Kapinos and Mitnik (2016) through the lens of US banks' stress testing for resilience of individual banks and the banking sector to macroeconomic shocks. Taking into account banks' heterogeneity improved the prediction capacity of their models. Therefore in our study we account for banks' heterogeneity by extracting clusters (see: Section 3.2) to check if this improves model accuracy.

Although in many respects financial companies are different from non-financial firms, their distress prediction research reveals multiple similarities. One of the first studies by Meyer and Pifer (1970) shared the same philosophy as Altman (1968) to apply multiple discriminant analysis (MDA) to a sample of failed and non-failed entities. However, to analyse banks' distress one should apply different variables to grasp their specific features. Mayes and Stremmel (2014) reviewed relevant banking failure literature showing the evolution of techniques applied for that purpose from MDA (4 out of 43 studies), through logit analysis (18/43) to neural networks (6/43) and hazard models (9/43),[1] which follows the practice of non-financial sector distress prediction. Therefore, we decided to search for inspiration regarding banks distress paths in the studies of the non-financial sector.

The pioneering research on failure processes for the non-financial sector was conducted by Argenti (1976), who outlined three processes (trajectories) for failing firms using illustrative case studies. Argenti portrayed a newly founded firm that will never become successful (1), firms that grow quickly but then suddenly decline (2), and gradually declining mature firms (3). These kinds of processes were found and developed in several further management studies (D'Aveni 1989; Laitinen 1991, Richardson, Nwankwo, Richardson 1994; Ooghe, de Prijcker 2008). D'Aveni (1989) identified 3 types of declining firms: lingerers, gradual decliners and sudden decliners. Laitinen (1991) found three financial failure processes (chronic failure, revenue financing failure, and acute failure firms) having obvious similarities with those described by D'Aveni (1989). Richardson, Nwankwo and Richardson (1994) distinguished four types of firms: boiled frogs, drowned frogs, bullfrogs, and tadpoles that never become frogs (failed start-ups). Ooghe and de Prijcker (2008) identified four types of failing firms, namely ambitious growth company, dazzled growth company, apathetic established companies and unsuccessful start-ups. Those analyses were based to a large extent on case studies of failed firms and management behaviour. Different approach based on advanced modelling techniques (Kohonen maps) was used by du Jardin and Severin (2011, 2012) showing "trajectories" of firm failure. Du Jardin (2015) presented seven different terminal failure processes using Kohonen maps combined with different estimation techniques. Overall, regardless of the methodology, firms presented different characteristics and followed different processes either leading them to a "survival" or a "death". Processes followed by banks have not been studied in such a detail so far, one of the obstacles being a low number of bankruptcies or other distress events.

[1] Probit and data envelopment analysis (DEA) were used respectively in the 2nd and 4th study.

In general, the turnaround literature suggests that multiple processes exist for both firm survival and failure which depend on the severity of decline, external conditions, and management actions (Mellahi, Wilkinson 2004; Trahms, Ndofor, Sirmon 2013). These results refer to organizations in general and can thus be applied to bank failures. As various processes are described by varying behaviour of financial ratios, it is expected that the accuracy of a generic failure prediction model (estimated for all firms in the sample) differs in groups of firms with different processes. The processes described by D'Aveni (1989) and Laitinen (1991) seem to fit the best the potential description of distress processes of banks.

Against this background, we will extract failure processes for distressed banks to show that they are different from those for non-distressed banks and check if the accuracy of the distress prediction model is associated with different distress processes.

3 Data and methodology

3.1 Variables and data

Since the 1970s the CAMEL[2] has been one of the most popular approaches to assess the banks' financial positions, both in research and supervisory practice (e.g., Lopez 1999), because it covers the most important aspects of bank risks and performance. The CAMEL-like variables are also widely used in recent studies (e.g., Mayes, Stremmel 2014; Poghosyan, Čihak 2011; Betz et al. 2014; Altman, Cizel, Rijken 2014). Therefore we decided to apply the same approach. In Table 1 we provide a list of the final set of variables that were chosen for both the cluster analyses (identification of processes) and the logistic regression models (distress prediction).

There are only two ratios in the "C" category. Total capital ratio (TCR) was not selected for two reasons: (1) due to too many missing values and (2) due to the fact that Mayes and Stremmel (2014) and Hambusch and Shaffer (2016) found out that the simple leverage (or equity to total assets) ratio outperformed the risk-based capital measures in detecting problems. Moreover, in the international context the TCR is less comparable than the simple leverage ratio due to different methods used to assess the risk-weighted assets.

The data for European banks were extracted from the Bankscope database for commercial, cooperative and savings banks. We marked the type of bank to account for banks' heterogeneity (more complex = commercial vs. less complex = cooperative and savings). Due to a low number of actual bankruptcies, the cases of bailouts and forced mergers have been included as distress categories as in Arena (2008), Betz et al. (2014) and Altman, Cizel and Rijken (2014), but we extended this approach by adding the event of bank's negative equity without any bailout or state aid. The distress status and the year of distress were determined using the database from Iwanicz-Drozdowska, Smaga and Witkowski (2016) supplemented by new distress events identified in the European Commission's communication and the press. We then additionally searched the data for banks with negative equity. The identified cases were assigned to the distressed bank group and the first year with negative equity in the data series was coded as the distress year for that bank.

[2] This methodology requires knowledge about the bank's capital adequacy (C), asset quality (A), management (M), earnings (E) and liquidity (L).

We have an initially large dataset of European banks, including data from 3,566 non-distressed banks and 163 distressed banks. These data consist of independent samples and are very heterogeneous with respect to the bank type, bank size, country, and the years considered. The objective of the study is to identify different paths or processes prior to the distress event. Therefore, the starting point for the analysis is the time-series of the distressed banks before the distress events. We require that the distressed bank should have a complete four-year time-series of financial statements for the years 1, 2, 3, and 4 prior to the distress to enable us to extract reasonable processes. All in all, 132 banks fulfilled this basic criterion. There were many banks that had missing values for one or several variables for at least one of the four years. The final dataset includes 99 distressed banks with complete financial statement series and with no missing values for any of the variables. The financial data of those banks enable to identify different distress processes in the European context. However, it is difficult to say how these processes differ from those of similar non-distressed banks. Therefore, we use the original data of 3,566 non-distressed banks to select, with the use of paired sampling, a non-distressed mate for each distressed bank. We require that although selected randomly, the mate should be similar to its distressed mate bank with respect to the bank type, bank size, country, and calendar years. If the processes identified for the distressed banks and their non-distressed mates are similar, these processes are not directly associated with the distress event. Furthermore, as the number of banks in both status classes is equal (99 + 99), the research data set is completely balanced.

The use of paired samples in failure research was pioneered by Beaver (1966). He used the paired-sample design to help provide a measure of control over factors that otherwise might blur the relationship between the ratios and the failure. In our study, we use the paired sample t-test as a statistical procedure to determine whether the mean difference in research variables between distressed banks and their paired non-distressed mate banks is zero. In this test, we first subtract the value of the variable for each distressed bank from the value for its non-distressed mate, and then compare these differences to zero. The paired sample t-test can only compare the means for two related (paired) units for a continuous variable that is normally distributed.

Table 2 cross-tabulates the distressed banks by country and by the identified first year of distress. These years cover the period of 1992–2014, with emphasis on 2008–2012 – a period that comprises the subprime financial crisis, and the Eurozone crisis. We did not use older distress cases due to the evidence provided by Shaffer (2012) that statistical linkages vary over time.

We also identified multiple distress events, meaning that the distress in a bank was indicated more than once, e.g. due to state aid provided in consecutive years (e.g. Allied Irish Bank, Erste, Hypo Alpe Adria, Lloyds, Bank of Cyprus, Dexia, Hypo Real Estate, Bankia, Alpha Bank, EFG Eurobank, National Bank of Greece, Piraeus Bank, Nova Ljubljanska Banka). In almost all the cases, the first (oldest) year coded as “distress” was recognized as the distress event in the final dataset. The inclusion of multiple distress events was possible if, following a distress code, at least the next four years are non-distressed. There are only two such cases (Agricultural Bank of Greece in 2004 and 2009; SNS Reaal NV in 2008 and 2013).

3.2 Methodology

We have used several statistical methods. First, we applied factor analysis separately to distressed and non-distressed groups of banks. In this way we got independent input variables that conform better to the normal distribution. Secondly, we applied cluster analysis separately to distressed and non-distressed groups of banks to extract processes for both groups. Thirdly, we estimated logistic regression models using the original financial ratios for each year separately to discriminate between distressed and non-distressed banks (0 = non-distressed bank; 1 = distressed bank). The idea was to examine how accurately bank failure can be estimated in different time horizons. Fourthly, we analysed how the processes affect prediction accuracy.

The paired sampling design limits the scope of analysis because it virtually eliminates the predictive power of the confounders. Therefore, we use a multivariate analysis to predict the distress event in this setting. We actually analyse how significantly financial variables of a distressed bank in a particular year differ from those of a non-distressed bank that is similar with respect to the type, size, country, and calendar year. In this study, we apply (binary) logistic regression analysis (e.g., Hosmer jr, Lemeshow, Sturdivant 2013) to estimate the distress prediction models using the 12 variables separately for various years (from 1 to 4 years prior to the distress event). For this estimation, the dependent variable $Y = 0$ when the bank is non-distressed and $Y = 1$ when it is distressed. LRA creates a score (logit) L for every bank. It is assumed that the independent variables are linearly related to L. This score is used to determine the conditional probability of becoming a distressed bank as follows:

$$p\left(Y=1|X\right)=\frac{1}{1+e^{-L}}=\frac{1}{1+e^{-\left(b_0+b_1x_1..+b_nx_n\right)}} \tag{1}$$

where b_i ($i = 0, \ldots, n$) are coefficients and n is the number of independent variables x_i ($i = 1, \ldots, n$); since we use a paired sampling design, the critical cut-off probability is 0.50.

The SAS programme was used to estimate the parameters of LR models (1) using the maximum likelihood method. The variables are selected into the model separately from the original set of the 12 annual variables (see Table 1 for details) using the (forward) stepwise method.[3] For each variable, the Wald Chi-square statistics[4] is presented for each year of estimation, provided that the variable in question is selected to the stepwise model. The goodness of fit of the LR model is measured by the R-square (coefficient of determination), which similarly as in the linear regression model reflects the amount of information gained when including the predictors into the model in comparison with the "null" model. Because this measure cannot reach the maximum value of 1, also the Nagelkerke max-rescaled R-square is reported (Nagelkerke 1991).[5]

When interpreting the results, the multivariate prediction (classification) accuracy for different years plays the key role. For each annual model, the original and jack-knife (leaving-one-out) cross-validated classification accuracies for the distressed and non-distressed banks are reported.[6]

[3] This method starts with no variables in the model, tests the addition of each variable using a model fit criterion, adds the variable (if any) whose inclusion gives the most statistically significant improvement of the fit, and repeats this process until none variable improves the model to a statistically significant extent. We use the level of significance of 0.05.

[4] Wald Chi-square statistic involves testing the null hypothesis that an individual predictor's regression coefficient is zero, given the other predictor variables in the model. The Chi-square test statistic is the squared ratio of the estimate to the standard error of the respective predictor.

[5] See also: http://www2.sas.com/proceedings/sugi25/25/st/25p256.pdf.

[6] The original classification accuracy is here the classification accuracy obtained in the estimation sample. Leave-one-out cross validation is an N-fold cross validation where N is the number of observations in the sample. That means that N separate times, the LR model is estimated on the entire sample except for one observation and a prediction is made for that observation. Finally, the classification accuracy for all N observations is calculated.

The results are also cross-validated between the four years in question. This means that the classification accuracy of the LR model estimated for year i (i = 1, 2, 3, 4) is assessed using the data from other years j (j = 1, 2, 3, 4 and $j \neq i$).

The classification accuracy of the LR models is also measured by the AUC (area under curve) measure extracted from the ROC (Receiver Operating Characteristic curve). ROC curve presents the cumulative distribution function of the true positive rate TPR (probability of true-alarm; sensitivity) on the y-axis versus the cumulative distribution function of the false positive rate FPR (probability of false-alarm; specificity) on the x-axis.

Since banks (and their paired mates) may behave differently prior to distress, the prediction accuracy of LR models may be associated with the distress (and non-distress) processes. Therefore, we extract processes for both the distressed and non-distressed banks. The more the processes differ from each other, the stronger the processes extracted for the distressed banks are associated with the event. Distress processes largely reflect correlations between financial ratios from different periods before the distress event. The extraction of the bank distress processes (and processes for non-distressed banks) is performed in two stages.

First, factor analysis was applied (e.g., Mulaik 2009; the FACTOR procedure in SAS) separately to the samples of distressed and non-distressed banks using altogether 45 variables accounting respectively for the first, second, third, and fourth year before the event. Since we want to compare the clusters in the sample of distressed banks with the control group of non-distressed banks, we used factor analysis in the same way in both samples to make the input variables uncorrelated, standardized and conforming better to the normal distribution, which are all favourable properties for the clustering procedure.[7] Since cluster analysis is carried out for both samples separately, also the variables are factored separately in both samples. Second, we apply k-means clustering (e.g., Everitt et al. 2009), which is a method of vector quantization, originally from signal processing, that aims to partition n observations (here n = 99) into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster that characterizes the elements of the cluster (the FASTCLUS Procedure in SAS). The main problem in clustering is the determination of k. There are no fully satisfactory methods for determining the number of population clusters for any type of the cluster analysis. However, Milligan and Cooper (1985) and Cooper and Milligan (1988) compared thirty methods for estimating the number of clusters and showed that the pseudo F-statistic (F) and the cubic clustering criterion (CCC) performed best. Thus, we use these criteria together to select k.

4 Empirical results

4.1 Descriptive statistics

The descriptive statistics of the original 12 variables in four years (45 year-variables) for the paired samples are presented in Table 3. The paired sample t-test that tests the null hypothesis assuming

[7] In our study, the factor scores are made uncorrelated with the application of the Varimax rotation that is considered the best orthogonal rotation (Fabrigar et al. 1999). The advantage of this rotation is that it makes the factor solution easier to interpret. For the interpretation of the distress processes it is important that the processes are independent of each other. The number of factors is determined by the-eigenvalue-greater-than-one rule (K1 or Kaiser criterion) (Kaiser 1960). This criterion means that the number of factors k is determined by the kth eigenvalue greater than one when k + 1th eigenvalue is equal to or less than one. The kth eigenvalue is equal to the variance accounted for by the kth most important factor.

that the true mean difference between the paired samples is zero, shows that there are statistically significant differences between the distressed and non-distressed banks. The most significant differences between them are found in the time-series of the equity to total assets ratio (capital adequacy) and the liquid assets to funding; at the same time these time-series showed that that capital adequacy and liquidity in the distressed banks were lower. Capital adequacy and liquidity are dominant in the regulatory framework. There is also a statistically significant difference in the return on assets ratio (ROA) in the first year prior to distress, reflecting the low profitability of the distressed banks just before the event. Figures 1, 2, and 3 show the development of the median of the ratios in both groups of banks. The decline in ROA is remarkably quicker in the distressed banks than in the non-distressed ones, although it is declining in both groups (Figure 1). Moreover, 3 and 2 years prior to the event, the ROA of the distressed banks was higher than that of the non-distressed ones. This may reflect the fact that they took a higher risk and were rewarded for that. For the non-distressed banks, the time-series of equity to total assets (EQ_to_TA) remains quite steady, but for the distressed banks, it is clearly declining (Figure 2) and the level of the ratio is lower, starting from the third year prior to the distress event. The development of the median of liquid assets to funding is declining for both groups of banks, but the level of the median is remarkably lower for the distressed ones (Figure 3). The overall picture of a distress bank 4 years before the distress is as follows: higher profitability, similar capital adequacy and lower liquidity. All those ratios decline over time. One year prior to the distress, all those ratios are lower than in the non-distressed peers.

4.2 Extracting factors[8]

For distressed banks the application of the Kaiser criterion results in an eleven-factor solution explaining 85.9% of the total variance of the 45 variables. The eleven factors are in a clear way associated with the year-variables, constructing a set of processes. The majority of these factors directly reflects the time--series of the variables in the following way: (F1) liquid assets to funding ratio, (F2) deposits to gross loans, (F3) net interest margin, (F4) loans to funding ratio, (F6) equity to total assets ratio, (F7) cost to income ratio (CI), and (F8) loan impairment charge to gross loans ratio. However, F5 is associated with the return on assets ratio (ROA) and growth rates of: equity, gross loans and total assets two years prior to the distress event. F9 reflects growth rates three years prior and F10 one year prior to the distress event. F11 is negatively associated with profitability (ROA) one year prior and positively with profitability four years prior to distress reflecting a long-term change in ROA.

For non-distressed banks the Kaiser criterion leads to a nine-factor solution which explains 83.0% of the total variance of the 45 year-variables. This nine-factor solution has a number of direct similarities with that of the distressed banks: F1 of the non-distressed banks is equal to factor 1 for the distressed banks, F4 to factor 4, F5 to factor F2, and F7 to factor 7. Furthermore, F2 is close to F6 (with a slight impact from profitability) and F8 to factor 8. However, F3 is a mixed factor related to factors 8 and 3 of distressed banks linking NIM and impairment charges (both constitute components of banks' pricing policy). The growth factor F6 is also comparable with factor 5 of the distressed banks. F9 is a mixed factor of profitability and growth of equity and loans three years prior to the distress event.

[8] Eigenvalues for factors and Varimax-rotated factor patterns estimated separately for the distressed (n = 99) and non--distressed banks (n = 99) are available from authors on request.

The mixed factors of profitability and growth for both groups emphasize the close empirical relationship between profitability and growth. Although the factor solutions for the distressed and non--distressed banks have much in common, they differ from each other in many respects. One of important differences is that there are more factors related to growth and overall (changing) profitability (ROA) in the case of distressed banks. This shows that the growth rates and changes in profitability should help identify potentially distressed banks. Non-distressed banks are generally less aggressive in growth and their profitability is more stable.

4.3 Extracting clusters of banks

Table 4a shows the cluster solution for distressed banks (n = 99). This solution includes four clusters with a very small cluster 1 (n = 2). Cluster 1 consists of two peculiar distressed banks (2%),[9] while clusters 2, 3, and 4 are the main clusters referring to the most important distress processes that we will comment on.

The second cluster consists of 60 larger banks (60.6%). These banks typically show a steady low net interest margin, but are mainly characterized by gradually declining growth rates, profitability, and capital adequacy. At the same time, the loan impairment charges to gross loans ratio, although low, increases continuously. We refer the reader at this point to the literature on non-financial companies in order to explain our observations and results better. Obviously, banks in the second cluster have numerous similarities with the chronic failure firms (Laitinen 1991) and the gradual decliners (D'Aveni 1989). Taking into account their characteristics, the second cluster may be called "low margin-decliners".

The third cluster includes 10 small banks (10.1%) which have an exceptionally high net interest margin but at the same time they report a very high loan impairment charge to gross loans ratio. The most notable characteristic of the process is that the banks, three years prior to the distress event have very high growth rates. Thus, these banks partly share the characteristics of the lingerers (D'Aveni 1989), although the horizon for collapse is quite short. The third cluster may be called "high margin--lingerers".

The fourth cluster has 27 banks (27.2%) of average size. The cost to income ratio of these banks tends to increase gradually, while the profitability and the capital adequacy continuously decline. These banks report very high growth rates two years before distress. Therefore, these banks share characteristics of both chronic failure (gradual decliners) and acute failure (sudden decliners) firms (Laitinen 1991; D'Aveni 1989). This cluster may be called "high costs-decliners".

Table 4b presents the resulting cluster solution for non-distressed banks (n = 99). The solution includes six clusters with three very small clusters consisting of one or two banks. These small clusters again include outlier banks where the growth and financial ratios do not exhibit any regularity.

The last three clusters are the main clusters revealing slight similarities with those of the distressed banks. The fourth cluster consists of 68 banks (68.7%). This big group of banks has a very slow but systematic tendency of declining growth rates, profitability, and liquidity. However, their capital adequacy, on average, stays constant over time. Therefore, these banks share the characteristics of gradual decliners (D'Aveni 1989), but do not suffer from any capital adequacy or profitability crisis as chronic failure firms do (Laitinen 1991). This group may be dubbed "solid-decliners".

[9] These two small banks share the characteristics of acute failure firms (Laitinen 1991) and the sudden decliners (D'Aveni 1989).

The fifth cluster has 11 banks (11.1%) which report very steady development. The only systematic tendency is that the net interest margin slowly increases. These banks show a high loans to total assets ratio but, especially in the last year, low liquidity ("solid-steady growth"). Finally, the sixth cluster includes 16 banks (16.2%) with very slow growth rates that have been declining continually. These banks have been able to maintain stable capital adequacy, despite a steady increase in their net interest margin and profitability ("solid-slow growth"). Thus, the banks in this cluster share some characteristics of gradual decliners (D'Aveni 1989).

4.4 Prediction of distress event

Table 5 presents the stepwise logistic regression models separately estimated for each of the four years prior to the distress event. The most statistically significant variable of the models in each year is the liquid assets to funding ratio, emphasizing the long-term importance of liquidity for entrance into distress situation. The coefficient and the significance of the ratio remains quite constant over the four-year timeframe. The second important stable predictor in LR models over the horizon is the loans to total assets ratio. In the early years the loans to funding ratio also plays an important role, but it loses its significance in the first year prior to the event. In the first and second years prior to the event the loan impairment charges to gross loans ratio also shows notable statistical significance. In the first year, the significance of the capital adequacy ratio is high. However, in the earlier years its coefficient gets a positive sign that is not consistent with expectations. In summary, in the four-year horizon the liquid assets to funding ratio and the loans to total assets ratio are the most significant predictors of a distress event. Moreover, the sequence – in the given years – of ratios with predictive abilities represents a deterioration process of sorts starting with ratio of loans to funding (showing, among others, an aggressive lending policy), then loan impairment charges (decreasing quality of the loan portfolio) and, finally, capital adequacy (showing the negative impact of the latter on the capital position).

Table 6 presents the statistics for the performance of the stepwise LR models. The R-Square measures in Panel 1 generally indicate that the performance of the models is not high and it rapidly deteriorates, especially in the third and second years of prediction. Panel 2 shows that even one year prior to distress, the performance of the cross-validated model, as measured by AUC, is poorer than 0.750, leading to AR (accuracy ratio)[10] of less than 0.5 (0.469), a lower figure than for an average model. Panel 3 indicates that all the models based on annual data report low performance 3 and 4 years prior to distress. Panels 4–6 present percentages of correct classification for different models and different years. The jack-knife cross-validated accuracies are significantly lower than the accuracies measured from the estimation sample (Panels 4 and 5). The percentage of correctly classified non-distressed banks is low one year prior to the event but the percentage stays quite stable over time (for the constant cut-off value of 0.5). Panel 5 shows that the overall percentage of correct classification is low, when the estimated models based on annual data are applied to alternative years in cross-validation. In summary, the statistics of accuracy shows that when classifying banks as distressed and non-distressed banks using an LR model, the accuracy is low if the non-distressed banks are of the same size, of the same type, and from the same country as their distressed mate banks.

[10] Accuracy ratio is defined as follows: AR = 2 × AUC − 1. AR equals 0 for a random model, 1 for a perfect model, and 0.5 for a model with an average classification performance.

It is expected that an explanation for the level of prediction accuracy can be found in the characteristics of the clusters.[11] Table 7 reports the rates of correct classifications by year and cluster for the one-year LR models estimated separately for each year prior to the event. For the small clusters with several banks the results in the estimation sample (Panel 1) are directly comparable with the jack-knife cross-validated results (Panel 2). Due to a limited number of observations in those clusters we refrain from offering an interpretation of these results.

For the distressed banks, the classification accuracy of the largest cluster 2 ("low margin-decliners") is low one year before the event but the accuracy stays quite constant over time. This kind of result is typical for chronic failure firms (Laitinen 1991). In contrast, for the banks in cluster 3 ("high margin – lingerers") the accuracy is very high one and two years before distress, but before that it is extremely low. This results from the fact that these banks share the characteristics of lingerers with a collapse several years before the event (D'Aveni 1989). The classification accuracy for the banks in cluster 4 ("high costs – decliners") is high one year prior to distress, but very low earlier on. This kind of result is typical of acute failure firms (Laitinen 1991) and sudden decliners (D'Aveni 1989). This means that the detection of distress is not possible within the same time horizon prior to the event for banks with varying characteristics. For banks with high growth rates the chances for detection are quite high one ("high margin" and "high costs") and two ("high margin") years before the event. For "low margin--decliners" problems are more visible 3–4 years before the event, but they do not become as clear-cut just before the distress event itself.

For non-distressed banks, the largest cluster 4 ("solid-decliners") shows quite low but stable classification accuracy over the four years. This result can be explained by the stability (solidity) with only a very slow tendency for gradual decline (profitability and low liquidity). For cluster 5 ("solid-steady growth") the number of correctly classified non-distressed banks is very limited (5–6) 2–4 years before the distress of the mates. The classification accuracy of cluster 6 ("solid-slow growth") is very low in the last year but increases rapidly when the prediction time horizon becomes longer. The banks in this cluster are typically gradual decliners (D'Aveni 1989), which explains the increasing classification accuracy one and two years before the event (as occuring for their mates). For non-distressed banks the prediction accuracy is overall lower in the short term than for the distressed ones. This shows that it is more difficult to achieve a low level of false alarms in that group of banks regardless of the cluster.

5 Conclusions

The empirical results for the development paths of banks show that there are different processes banks follow before the distress event. We found three main categories that the distressed banks fall into. These are: (1) "low margin-decliners", (2) "high margin-lingerers" and (3) "high costs-decliners". For the non-distressed banks, we also found three main processes which, however, are different from those extracted for the distressed banks. We call them: (1) "solid-decliners", (2) "solid-steady growth", and (3) "solid-slow growth". The growth rates play a special role in development paths, however, the most numerous clusters consist of stagnant banks with low and/or declining growth rates. Moreover, for those banks the prediction accuracy of cluster models is lower than in other cases.

[11] Please see Table 4a and 4b.

Four or three years prior to the distress event, the differences between the distressed banks and their mates seem not to be palpable in most of the cases. However, 1 or 2 years prior to the distress, they become more visible and this information still allows supervisors to intervene in due course. Special attention should be paid to measures of liquidity (in our study liquid assets to funding ratio) and to the business model (loans to assets ratio in our study). The latter is linked to credit risk, however, in more sophisticated banks it is also generated by other assets and off-balance sheet items. More detailed supervisory data should allow a more detailed analysis of this issue. Also equity to total assets ratio and impairment charges provide some guidelines on how to detect differences in "the paths of glory" and "the paths of shame" in a shorter time horizon. All these ratios should be included in banks' recovery and restructuring plans, allowing both banks and supervisors to monitor their development paths. So far, they are not on the list of preferred ratios in the European Banking Authority guidelines (EBA/GL/2015/02) on recovery plan indicators.

Higher performance of models based on clusters, especially for "growth" banks, calls for monitoring in homogenous groups based on a wide set of characteristics, not just on banks' legal forms or sizes. The main challenge in building a supervisory monitoring system is to identify groups of banks with various specifics which are not as obvious as the size or the legal form.

All in all, it is difficult to predict the distress events with the use of a set of CAMEL-like variables, although they are widely used in academic literature and in practice. Therefore, one needs to search for an alternative set of ratios. Ratios based on market data (e.g., prices of shares) may be treated as a good choice, however, not too many banks are listed on the stock exchange. Another opportunity is to use ratios implemented within Basel 3 regulatory reform, but such a solution is applicable only in a more distant future due to the lack of time series for the new ratios. Moreover, new distress events would have to occur to enable further analysis since finding good proxies for Basel 3 ratios on the basis of historical data is – in our view – not viable. One also needs to consider including lagged macroeconomic variables in distress prediction.

References

Alam P., Booth D., Lee K., Thordarson T. (2000), The use of fuzzy clustering algorithm and self--organizing neural networks for identifying potentially failing banks: an experimental study, *Expert Systems with Applications*, 18(3), 185–199.

Altman E.I. (1968), Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23, 589–609.

Altman E.I., Cizel J., Rijken H.A. (2014), *Anatomy of bank distress: the information content of accounting fundamentals within and across countries*, http://dx.doi.org/10.2139/ssrn.2504926.

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

Argenti J. (1976), *Corporate Collapse: the Causes and Symptoms*, McGraw-Hill.

Beaver W.H. (1966), Financial ratios as predictors of failure. Empirical research in accounting. Selected studies, *Journal of Accounting Research,* 4, 71–111.

Berg S., Hexenberg B. (1994), *Early warning indicators for Norwegian banks: a logit analysis of the experiences from the banking crisis*, Norges Bank Working Papers, 1.

Betz F., Oprica S., Peltonen T., Sarlin P. (2014), Predicting distress in European banks, *Journal of Banking and Finance*, 45, 225–241.

Boyacioglu M.A., Kara Y., Baykan O. (2009), Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey, *Expert Systems with Applications*, 36(2), 3355–3366.

Canbas S., Cabuk A., Kilic S.B. (2005), Prediction of commercial bank failure via multivariate statistical analysis of financial structures: the Turkish case, *European Journal of Operational Research*, 166(2), 528–546.

Chauhan N., Ravi V., Karthik Chandra D. (2009), Differential evolution trained wavelet neural networks: application to bankruptcy prediction in banks, *Expert Systems with Applications*, 36(4), 7659–7665.

Clare A., Priestley R. (2002), Calculating the probability of failure of the Norwegian banking sector, *Journal of Multinational Financial Management*, 12(1), 21–40.

Cole R.A., Gunther J.W. (1995), Separating the likelihood and timing of bank failure, *Journal of Banking and Finance*, 19(6), 1073–1089.

Cole R.A., White L.J. (2012), Déjà vu all over again: the causes of U.S. commercial bank failures this time around, *Journal of Financial Services Research*, 42(1), 5–29.

Cooper M.C., Milligan G.W. (1988), The effect of error on determining the number of clusters. Proceedings of the international workshop on data analysis, *Decision Support and Expert Knowledge Representation in Marketing and Related Areas of Research*, 319–328.

Covas F.B., Rump B., Zakrajšek E. (2014), Stress-testing US bank holding companies: a dynamic panel quantile regression approach, *International Journal of Forecasting*, 30(3), 691–713.

Cox R.A.K., Wang G.W. (2014), Predicting the US bank failure: a discriminant analysis, *Economic Analysis and Policy*, 44(2), 202–211.

D'Aveni R. (1989), The aftermath of organizational decline: a longitudinal study of the strategic and managerial characteristics of declining firms, *Academy of Management Journal*, 32(3), 577–605.

Du Jardin P. (2015), Bankruptcy prediction using terminal failure processes, *European Journal of Operational Research*, 242(1), 286–303.

Everitt B.S., Landau S., Leese M., Stahl D. (2009), *Cluster Analysis*, John Wiley & Sons.

Fabrigar L.R., Wegener D.T., MacCallum R.C., Strahan E.J. (1999), Evaluating the use of exploratory factor analysis in psychological research, *Psychological Methods*, 4(3), 272–299.

Hosmer D.W., Lemeshow S. (1989), *Applied logistic regression*, Wiley.

Hosmer D.W. jr., Lemeshow S., Sturdivant R.X. (2013), *Applied Logistic Regression*, John Wiley & Sons.

Iwanicz-Drozdowska M., Smaga P., Witkowski B. (2016), Bank restructuring in the EU. Which way to go?, *Journal of Policy Modeling*, 38(3), 572–586.

Kaiser H.F. (1960), The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, 20, 141–151.

Kapinos P., Mitnik O. (2016), A top-down approach to stress-testing banks, *Journal of Financial Services Research*, 49(2), 229–264.

Kolari J., Glennon D., Shin H., Caputo M. (2002), Predicting large US commercial bank failures, *Journal of Economics and Business*, 54(4), 361–387.

Laitinen E. (1991), Financial ratios and different failure processes, *Journal of Business Finance and Accounting*, 18(5), 649–673.

Lanine G., Vander Vennet R. (2006), Failure predictions in the Russian bank sector with logit and trait recognition models, *Expert Systems with Applications*, 30, 463–478.

Lopez J.A. (1999), *Using CAMELS ratings to monitor bank conditions*, Federal Reserve Bank of San Francisco Economic Letter, 19.

López Iturriaga F.J., Sanz I.P. (2015), Bankruptcy visualization and prediction using neural networks: a study of U.S. commercial banks, *Expert Systems with Applications*, 42(6), 2857–2868.

Maghyereh A.I., Awartani B. (2014), Research in international business and finance bank distress prediction: empirical evidence from the Gulf Cooperation Council countries, *Research in International Business and Finance*, 30, 126–147.

Mannasoo K., Mayes D. (2005), *Investigating the early signals of banking sector vulnerabilities in Central and East European emerging markets*, Working Papers of Eesti Pank, 2005-08.

Mannasoo K., Mayes D. (2009), Explaining bank distress in Eastern European transition economies, *Journal of Banking and Finance*, 33(2), 244–253.

Martin D. (1977), Early warning of bank failure, *Journal of Banking and Financ*e, 1(3), 249–276.

Mayes D., Stremmel H. (2014), The effectiveness of capital adequacy measures in predicting bank distress, *SUERF Study*, 2014/1.

Mellahi K., Wilkinson A. (2004), Organizational failure: a critique of recent research and a proposed integrative framework, *International Journal of Management Reviews,* 5(1), 21–41.

Meyer P.A., Pifer H.W. (1970), Prediction of bank failures, *The Journal of Finance*, 25, 853–868.

Milligan G.W., Cooper M.C. (1985), An examination of procedures for determining the number of clusters in a data set, *Psychometrika,* 50, 159–179.

Molina C.A. (2002), Predicting bank failures using a hazard model: the Venezuelan banking crisis, *Emerging Markets Review*, 3(1), 31–50.

Mulaik S.A. (2009), *Foundations of Factor Analysis*, Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences.

Ooghe H., de Prijcker S. (2008), Failure processes and causes of company bankruptcy: a typology, *Management Decision*, 46(2), 223–242.

Peek J., Rosengren E. (1996), The use of capital ratios to trigger intervention in problem banks: too little, too late, *New England Economic Review*, September/October, 49–58.

Pettway R., Sinkey J.F. jr (1980), Establishing on-site bank examination priorities: an early warning system using accounting and market information, *Journal of Finance*, 35(1), 137–150.

Poghosyan T., Čihak M. (2011), Determinants of bank distress in Europe: evidence from a new data set, *Journal of Financial Services Research*, 40(3), 163–184.

Richardson B., Nwankwo S., Richardson S. (1994), Understanding the causes of business failure crises. Generic failure types: boiled frogs, drowned frogs, bullfrogs and tadpoles, *Management Decision*, 32(4), 9–22.

Santomero A., Vinso J. (1977), Estimating the probability of failure for commercial banks and the banking system, *Journal of Banking and Finance*, 1(2), 185–205.

Shaffer S. (2012), Bank failure risk: different now? *Economics Letters*, 116(3), 613–616.

Sinkey J.F. jr (1975), A multivariate statistical analysis of the characteristics of problem banks, *Journal of Finance*, 30(1), 21–36.

Sinkey J.F. jr (1978), Identifying "problem" banks, *Journal of Money, Credit and Banking*, 10(2), 184–193.

Trahms C.A., Ndofor H.A., Sirmon D.G. (2013), Organizational decline and turnaround: a review and agenda for future research, *Journal of Management*, 39(5), 1277–1307.

Wheelock D.C., Wilson P.W. (2000), Why do banks disappear? The determinants of U.S. bank failures and acquisitions, *The Review of Economics and Statistics*, 82(1), 127–138.

Wong J., Wong T., Leung P. (2010), Predicting banking distress in the EMEAP economies, *Journal of Financial Stability*, 6(3), 169–179.

Appendix

Table 1
Description of the variables

	Variable name	Description and CAMEL category	Examples of references
1	Growth_TA	one-year total assets (TA) growth rate; A, M in CAMEL	Kolari et al. (2002), Kapinos, Mitnik (2016)
2	Growth_EQ	one-year equity (EQ) growth rate; C in CAMEL	Similar to Kolari et al. (2002), change in EQ_to_TA
3	Growth_G_Loans	one-year gross loans (GL) growth rate; A in CAMEL	Cox, Wang (2014); Kapinos, Mitnik (2016)
4	ROA	return on total assets ratio; E in CAMEL	Cole, Gunther (1995); Wheelcock, Wilson (2000); Kolari et al. (2002); Arena (2008); Mayes, Stremmel (2012); Shaffer (2012); Cox, Wang (2014); Hambusch, Shaffer (2016)
5	EQ_to_TA	equity to total assets ratio; C in CAMEL	Cole, Gunther (1995); Wheelcock, Wilson (2000); Kolari et al. (2002); Arena (2008); Poghosyan, Čihak (2011); Mayes, Stremmel (2012); Shaffer (2012); Cox, Wang (2014); Hambusch, Shaffer (2016)
6	Deposits_to_G_Loans	total customer deposits to (gross) loans ratio; L in CAMEL	similar to Mayes, Stremmel (2012) and Kapinos, Mitnik (2016)
7	L_Imp_to_G_Loans	loan impairment charge (LI) = loan loss provisions (LLP) to gross loans ratio; A in CAMEL	Kolari et al. (2002); Arena (2008); Poghosyan, Čihak (2011); Cox, Wang (2014); Altman, Cizel, Rijken (2014)
8	NIM	net interest margin to interest--earning assets; E in CAMEL	Kolari et al. (2002); Altman, Cizel, Rijken (2014); Covas, Rump, Zakrajšek (2014)
9	CI	operating cost (expense) to operating income ratio; E, M in CAMEL	Poghosyan, Čihak (2011) as managerial quality proxy (M); similar to Covas, Rump, Zakrajšek (2014) in compensation expenses
10	Loans_to_TA	net loans to total assets ratio; A, M in CAMEL	Wheelcock, Wilson (2000); Kolari et al. (2002); Arena (2008); Shaffer (2012); Hambusch, Shaffer (2016)
11	Loans_to_Funding	net loans to customer and short-term funding ratio; L in CAMEL	Altman, Cizel, Rijken (2014); similar to Kapinos, Mitnik (2016)
12	Liquid_A_to_Funding	liquid assets to deposits and short--term funding ratio; L in CAMEL	Poghosyan, Čihak (2011); Altman, Cizel, Rijken (2014); similar to Arena (2008)

Table 2
Distribution of the distressed banks by distress year and country in the empirical data (*n* = 99)

Year	AT	BE	CY	DE	DK	ES	FR	GR	IE	IS	IT	LV	NL	PT	SE	SI	UK	Total
1992	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
1993	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	3
1996	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4
1998	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
2001	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2003	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
2004	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
2007	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2
2008	3	0	1	2	0	1	1	0	0	1	0	0	2	0	1	0	1	13
2009	3	0	0	2	0	2	0	6	3	0	4	2	0	0	1	0	0	23
2010	0	0	0	0	2	14	0	0	0	0	0	0	0	0	0	0	0	16
2011	0	2	0	1	2	6	1	0	1	0	0	1	0	0	0	1	0	15
2012	1	0	2	0	1	2	0	0	0	0	0	0	0	3	0	1	0	10
2013	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	3	0	6
2014	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Total	7	3	4	6	6	26	12	7	4	2	4	3	3	4	2	5	1	99

Table 3
Descriptive statistics of the distressed banks and their non-distressed paired mates ($n = 99 + 99$)

Variable	Distressed banks			Non-distressed banks			Paired sample t-test		
	median	mean	std dev.	median	mean	std dev.	t-value	p-value	significance
Growth_TA	0.0137	0.0531	0.1979	0.0594	0.0803	0.1684	1.12	0.2649	
Growth_TA_L1	0.0771	0.1173	0.1800	0.0713	0.1254	0.2115	0.26	0.7968	
Growth_TA_L2	0.1151	0.1538	0.1864	0.0792	0.1203	0.2217	-1.35	0.1786	
Growth_EQ	-0.0005	-0.0316	0.3903	0.0357	0.0223	0.2428	1.22	0.2264	
Growth_EQ_L1	0.0419	0.0775	0.2674	0.0663	0.1551	0.3048	1.93	0.0569	*
Growth_EQ_L2	0.0895	0.1822	0.4203	0.0785	0.2126	0.5126	0.43	0.6656	
Growth_G_Loans	0.0130	0.0815	0.2041	0.0412	0.1031	0.2367	0.70	0.4854	
Growth_G_Loans_L1	0.0746	0.1329	0.2368	0.0571	0.1300	0.3006	-0.16	0.8755	
Growth_G_Loans_L2	0.1324	0.1742	0.2059	0.1075	0.1426	0.2555	-1.08	0.2813	
ROA	0.0013	-0.0011	0.0102	0.0027	0.0030	0.0092	3.15	0.0022	***
ROA_L1	0.0035	0.0036	0.0102	0.0045	0.0057	0.0088	1.90	0.0599	*
ROA_L2	0.0058	0.0057	0.0090	0.0050	0.0062	0.0080	0.51	0.6087	
ROA_L3	0.0063	0.0070	0.0071	0.0060	0.0073	0.0080	0.37	0.7106	
EQ_to_TA	0.0474	0.0506	0.0287	0.0639	0.0700	0.0434	4.18	<0.0001	***
EQ_to_TA_L1	0.0531	0.0590	0.0304	0.0656	0.0725	0.0397	3.35	0.0011	***
EQ_to_TA_L2	0.0592	0.0619	0.0312	0.0639	0.0725	0.0439	2.54	0.0127	**
EQ_to_TA_L3	0.0630	0.0655	0.0352	0.0626	0.0755	0.0620	1.71	0.0910	*
Deposits_to_G_Loans	0.7102	0.8514	1.0199	0.7742	1.0320	1.2540	1.12	0.2494	
Deposits_to_G_Loans_L1	0.7071	0.8484	1.0371	0.7697	1.0286	1.3338	1.05	0.2967	
Deposits_to_G_Loans_L2	0.6944	0.8436	0.9899	0.7769	1.0126	1.2127	1.08	0.2844	
Deposits_to_G_Loans_L3	0.7185	0.8741	0.9999	0.7837	1.0298	1.3496	0.90	0.3703	
L_Imp_to_G_Loans	0.0105	0.0141	0.0151	0.0067	0.0106	0.0142	-1.94	0.0555	*
L_Imp_to_G_Loans_L1	0.0066	0.0122	0.0148	0.0057	0.0095	0.0125	-1.56	0.1211	
L_Imp_to_G_Loans_L2	0.0050	0.0091	0.0135	0.0045	0.0079	0.0114	-0.84	0.4030	
L_Imp_to_G_Loans_L3	0.0042	0.0071	0.0110	0.0034	0.0062	0.0114	-0.62	0.5353	
NIM	0.0186	0.0207	0.0115	0.0193	0.0216	0.0150	0.55	0.5853	
NIM_L1	0.0202	0.0215	0.0112	0.0204	0.0229	0.0160	0.84	0.4035	
NIM_L2	0.0203	0.0215	0.0113	0.0197	0.0232	0.0161	0.98	0.3302	
NIM_L3	0.0215	0.0223	0.0120	0.0203	0.0235	0.0163	0.69	0.4930	
CI	0.6418	0.6526	0.2070	0.6056	0.6264	0.1847	1.01	0.3135	
CI_L1	0.5978	0.5944	0.1721	0.5786	0.5989	0.1756	0.38	0.7019	
CI_L2	0.5657	0.5867	0.1831	0.5787	0.5944	0.1848	0.26	0.7953	
CI_L3	0.5866	0.5963	0.1847	0.5953	0.6161	0.1979	0.83	0.4076	

Table 3, cont'd

Variable	**Distressed banks**			**Non-distressed banks**			**Paired sample t-test**		
	median	**mean**	**std dev.**	**median**	**mean**	**std dev.**	**t-value**	**p-value**	**signifi-cance**
Loans_to_TA	0.6541	0.6082	0.1685	0.6773	0.5998	0.2317	-0.40	0.6887	
Loans_to_TA_L1	0.6482	0.6041	0.1712	0.6668	0.6006	0.2338	-0.23	0.8223	
Loans_to_TA_L2	0.6551	0.6088	0.1812	0.6712	0.6060	0.2330	-0.17	0.8619	
Loans_to_TA_L3	0.6521	0.6088	0.1827	0.6643	0.5980	0.2297	-0.49	0.6741	
Loans_to_Funding	0.8925	0.9335	0.3726	0.8423	0.8798	0.4394	-1.08	0.2812	
Loans_to_Funding_L1	0.9039	0.9757	0.4271	0.8617	0.8971	0.4439	-1.41	0.1618	
Loans_to_Funding_L2	0.9284	0.9534	0.4268	0.8563	0.9064	0.4545	-0.83	0.4110	
Loans_to_Funding_L3	0.9208	0.9527	0.4243	0.8884	0.8997	0.4714	-0.92	0.3614	
Liquid_A_to_Funding	0.1498	0.2257	0.2494	0.1757	0.3244	0.3470	2.71	0.0079	***
Liquid_A_to_Funding_L1	0.1800	0.2632	0.2641	0.1918	0.3530	0.3658	2.46	0.0157	**
Liquid_A_to_Funding_L2	0.1913	0.2380	0.2092	0.2223	0.3464	0.3582	3.17	0.0021	***
Liquid_A_to_Funding_L3	0.1877	0.2520	0.2050	0.2439	0.3735	0.3648	3.59	0.0005	***

Notes:
* ≤ 0.10, ** ≤ 0.05, *** ≤ 0.01. For variables see Table 1.
Time index:
no index = one year prior to distress event,
_L1 = two years prior to distress event,
_L2 = three years prior to distress event,
_L3 = four years prior to distress event.

Table 4a
Descriptive statistics of the extracted clusters for the sample of banks. Distressed banks (n = 99 + 99)

Variable	Cluster 1 (n = 2)		Custer 2 (n = 60)		Cluster 3 (n = 10)		Cluster 4 (n = 27)	
	median	mean	median	mean	median	mean	median	mean
Growth_TA	-0.081	-0.081	0.014	0.031	-0.026	0.000	0.039	0.131
Growth_TA_L1	0.575	0.575	0.061	0.075	0.027	0.134	0.137	0.173
Growth_TA_L2	0.102	0.102	0.114	0.125	0.269	0.372	0.102	0.140
Growth_EQ	0.067	0.067	0.004	-0.039	0.165	0.419	-0.137	-0.188
Growth_EQ_L1	-0.015	-0.015	0.033	0.062	0.103	0.003	0.122	0.145
Growth_EQ_L2	0.138	0.138	0.080	0.091	0.934	0.918	0.082	0.115
Growth_G_Loans	0.105	0.105	0.002	0.036	0.059	0.117	0.080	0.168
Growth_G_Loans_L1	0.071	0.071	0.056	0.072	0.062	0.108	0.249	0.281
Growth_G_Loans_L2	-0.024	-0.024	0.135	0.151	0.261	0.346	0.123	0.176
ROA	0.019	0.019	0.002	0.001	0.000	-0.005	0.000	-0.005
ROA_L1	0.019	0.019	0.004	0.003	-0.002	-0.005	0.006	0.007
ROA_L2	0.019	0.019	0.006	0.005	0.002	-0.001	0.009	0.008
ROA_L3	0.009	0.009	0.006	0.006	0.002	0.001	0.010	0.011
EQ_to_TA	0.036	0.036	0.051	0.056	0.042	0.041	0.048	0.044
EQ_to_TA_L1	0.031	0.031	0.055	0.061	0.036	0.037	0.061	0.066
EQ_to_TA_L2	0.056	0.056	0.059	0.063	0.035	0.041	0.067	0.067
EQ_to_TA_L3	0.055	0.055	0.061	0.067	0.043	0.049	0.071	0.070
Deposits_to_G_Loans	7.170	7.170	0.706	0.740	0.718	0.652	0.710	0.705
Deposits_to_G_Loans_L1	7.381	7.381	0.680	0.715	0.790	0.690	0.758	0.720
Deposits_to_G_Loans_L2	6.640	6.640	0.679	0.725	0.626	0.620	0.710	0.761
Deposits_to_G_Loans_L3	6.377	6.377	0.711	0.757	0.589	0.633	0.741	0.815
L_Imp_to_G_Loans	0.006	0.006	0.008	0.009	0.029	0.027	0.015	0.021
L_Imp_to_G_Loans_L1	0.014	0.014	0.006	0.009	0.026	0.029	0.006	0.013
L_Imp_to_G_Loans_L2	0.003	0.003	0.004	0.005	0.018	0.023	0.007	0.014
L_Imp_to_G_Loans_L3	0.005	0.005	0.004	0.004	0.019	0.020	0.005	0.009
NIM	0.011	0.011	0.018	0.019	0.036	0.033	0.023	0.021
NIM_L1	0.013	0.013	0.018	0.019	0.034	0.032	0.024	0.023
NIM_L2	0.013	0.013	0.019	0.019	0.033	0.035	0.022	0.022
NIM_L3	0.015	0.015	0.019	0.019	0.040	0.040	0.024	0.023
CI	0.605	0.605	0.632	0.664	0.746	0.675	0.657	0.622
CI_L1	0.674	0.674	0.598	0.621	0.624	0.623	0.544	0.519
CI_L2	0.681	0.681	0.581	0.615	0.615	0.612	0.506	0.508
CI_L3	0.794	0.794	0.600	0.633	0.613	0.579	0.489	0.507

Table 4a, cont'd

Variable	**Cluster 1 (*n* = 2)**		**Custer 2 (*n* = 60)**		**Cluster 3 (*n* = 10)**		**Cluster 4 (*n* = 27)**	
	median	**mean**	**median**	**mean**	**median**	**mean**	**median**	**mean**
Loans_to_TA	0.105	0.105	0.650	0.627	0.551	0.540	0.680	0.629
Loans_to_TA_L1	0.095	0.095	0.648	0.629	0.534	0.519	0.659	0.618
Loans_to_TA_L2	0.142	0.142	0.665	0.636	0.667	0.578	0.647	0.594
Loans_to_TA_L3	0.165	0.165	0.652	0.627	0.736	0.651	0.643	0.586
Loans_to_Funding	0.123	0.123	0.948	1.015	0.729	0.790	0.909	0.866
Loans_to_Funding_L1	0.107	0.107	0.953	1.049	0.673	0.764	0.958	0.955
Loans_to_Funding_L2	0.158	0.158	0.989	1.047	0.801	0.819	0.869	0.855
Loans_to_Funding_L3	0.182	0.182	0.983	1.039	0.935	0.905	0.841	0.835
Liquid_A_to_Funding	0.184	0.184	0.142	0.205	0.277	0.344	0.131	0.232
Liquid_A_to_Funding_L1	0.621	0.621	0.178	0.231	0.287	0.343	0.180	0.279
Liquid_A_to_Funding_L2	0.416	0.416	0.171	0.225	0.251	0.252	0.175	0.249
Liquid_A_to_Funding_L3	0.557	0.557	0.169	0.240	0.178	0.223	0.206	0.266
TA (size effect)	964.5	964.5	30,106.0	120,337.4	1,103.5	10,364.1	5,939.0	61,270.1

Notes:
For variables see Table 1.
Time index:
no index – one year prior to distress event,
_L1 – two years prior to distress event,
_L2 – three years prior to distress event,
_L3 – four years prior to distress event.
Statistical information of clustering:
Pseudo F Statistic = 7.17,
Approximate expected over-all R-squared = 0.2122,
Cubic clustering criterion (CCC) = -2.687.

Table 4b
Descriptive statistics of the extracted clusters for the sample of banks. Non-distressed banks (n = 99 + 99)

Variable	Cluster 1 (n = 1)		Custer 2 (n = 1)		Cluster 3 (n = 2)		Cluster 4 (n = 68)		Cluster 5 (n = 11)		Cluster 6 (n = 16)	
	median	mean	median	mean	median	mean	median	mean	median	mean	median	mean
Growth_TA	-0.282	-0.282	0.923	0.923	0.011	0.011	0.067	0.093	0.048	0.064	0.018	0.016
Growth_TA_L1	0.916	0.916	0.923	0.923	0.429	0.429	0.069	0.129	0.055	0.051	0.016	0.024
Growth_TA_L2	0.923	0.923	0.915	0.915	-0.120	-0.120	0.085	0.116	0.124	0.133	0.050	0.058
Growth_EQ	-0.156	-0.156	1.484	1.484	0.014	0.014	0.047	0.031	0.050	0.003	-0.020	-0.080
Growth_EQ_L1	0.382	0.382	1.758	1.758	0.010	0.010	0.068	0.118	0.036	0.128	0.023	0.235
Growth_EQ_L2	1.010	1.010	0.375	0.375	-0.015	-0.015	0.088	0.255	0.063	0.128	0.037	0.058
Growth_G_Loans	0.674	0.674	1.200	1.200	0.441	0.441	0.040	0.106	0.071	0.052	0.007	-0.021
Growth_G_Loans_L1	1.200	1.200	1.078	1.078	-0.245	-0.245	0.069	0.136	0.057	0.039	0.020	0.087
Growth_G_Loans_L2	-0.254	-0.254	0.776	0.776	0.680	0.680	0.112	0.129	0.099	0.126	0.043	0.131
ROA	-0.020	-0.020	0.037	0.037	0.001	0.001	0.003	0.004	0.004	0.004	0.000	-0.003
ROA_L1	-0.020	-0.020	0.037	0.037	0.002	0.002	0.005	0.007	0.004	0.005	0.001	0.002
ROA_L2	0.003	0.003	0.030	0.030	0.002	0.002	0.006	0.008	0.005	0.005	0.003	0.000
ROA_L3	0.000	0.000	0.037	0.037	0.004	0.004	0.007	0.008	0.003	0.005	0.005	0.004
EQ_to_TA	0.254	0.254	0.137	0.137	0.026	0.026	0.067	0.071	0.053	0.057	0.066	0.064
EQ_to_TA_L1	0.188	0.188	0.128	0.128	0.027	0.027	0.066	0.073	0.057	0.060	0.070	0.073
EQ_to_TA_L2	0.261	0.261	0.108	0.108	0.036	0.036	0.068	0.074	0.053	0.055	0.060	0.067
EQ_to_TA_L3	0.530	0.530	0.165	0.165	0.031	0.031	0.067	0.076	0.052	0.054	0.057	0.061
Deposits_to_G_Loans	2.184	2.184	0.513	0.513	7.567	7.567	0.843	0.895	0.456	0.381	0.789	1.207
Deposits_to_G_Loans_L1	0.565	0.565	0.510	0.510	9.270	9.270	0.830	0.923	0.409	0.370	0.800	0.963
Deposits_to_G_Loans_L2	1.400	1.400	0.508	0.508	7.193	7.193	0.792	0.971	0.437	0.369	0.816	0.868
Deposits_to_G_Loans_L3	0.500	0.500	0.068	0.068	9.286	9.286	0.851	0.954	0.492	0.388	0.802	0.857
L_Imp_to_G_Loans	0.063	0.063	0.063	0.063	0.002	0.002	0.006	0.007	0.003	0.006	0.019	0.025
L_Imp_to_G_Loans_L1	0.063	0.063	0.058	0.058	-0.004	-0.004	0.005	0.006	0.003	0.005	0.024	0.024
L_Imp_to_G_Loans_L2	-0.026	-0.026	0.063	0.063	-0.001	-0.001	0.004	0.005	0.003	0.002	0.023	0.023
L_Imp_to_G_Loans_L3	-0.026	-0.026	0.063	0.063	-0.013	-0.013	0.003	0.005	0.002	0.002	0.015	0.016
NIM	0.004	0.004	0.072	0.072	0.007	0.007	0.019	0.022	0.017	0.015	0.021	0.026
NIM_L1	0.015	0.015	0.072	0.072	0.005	0.005	0.021	0.023	0.014	0.015	0.024	0.027
NIM_L2	0.011	0.011	0.072	0.072	0.004	0.004	0.020	0.024	0.012	0.014	0.028	0.028
NIM_L3	0.002	0.002	0.072	0.072	0.003	0.003	0.022	0.024	0.012	0.013	0.028	0.030
CI	0.802	0.802	0.424	0.424	0.724	0.724	0.614	0.635	0.490	0.515	0.622	0.656
CI_L1	0.350	0.350	0.487	0.487	0.723	0.723	0.579	0.596	0.530	0.591	0.580	0.624
CI_L2	1.191	1.191	0.603	0.603	0.854	0.854	0.581	0.585	0.472	0.520	0.578	0.615
CI_L3	1.400	1.400	0.375	0.375	0.645	0.645	0.591	0.603	0.555	0.560	0.636	0.673

Table 4b, cont'd

Variable	**Cluster 1 (*n* = 1)**		**Custer 2 (*n* = 1)**		**Cluster 3 (*n* = 2)**		**Cluster 4 (*n* = 68)**		**Cluster 5 (*n* = 11)**		**Cluster 6 (*n* = 16)**	
	median	**mean**	**median**	**mean**	**median**	**mean**	**median**	**mean**	**median**	**mean**	**median**	**mean**
Loans_to_TA	0.040	0.040	0.724	0.724	0.035	0.035	0.686	0.628	0.690	0.688	0.538	0.516
Loans_to_TA_L1	0.015	0.015	0.725	0.725	0.025	0.025	0.697	0.628	0.716	0.696	0.544	0.518
Loans_to_TA_L2	0.005	0.005	0.781	0.781	0.047	0.047	0.689	0.634	0.715	0.679	0.592	0.533
Loans_to_TA_L3	0.038	0.038	0.821	0.821	0.020	0.020	0.686	0.624	0.706	0.681	0.531	0.525
Loans_to_Funding	0.055	0.055	1.758	1.758	0.040	0.040	0.829	0.825	1.747	1.694	0.614	0.655
Loans_to_Funding_L1	0.019	0.019	1.834	1.834	0.029	0.029	0.856	0.839	1.703	1.747	0.670	0.663
Loans_to_Funding_L2	0.007	0.007	1.817	1.817	0.058	0.058	0.871	0.847	1.834	1.769	0.734	0.669
Loans_to_Funding_L3	0.150	0.150	2.200	2.200	0.023	0.023	0.883	0.829	1.702	1.763	0.666	0.683
Liquid_A_to_Funding	1.244	1.244	0.511	0.511	1.013	1.013	0.154	0.272	0.228	0.313	0.411	0.401
Liquid_A_to_Funding_L1	1.188	1.188	0.457	0.457	1.025	1.025	0.161	0.304	0.312	0.384	0.404	0.398
Liquid_A_to_Funding_L2	1.294	1.294	0.292	0.292	1.012	1.012	0.173	0.280	0.309	0.443	0.382	0.424
Liquid_A_to_Funding_L3	1.500	1.500	1.068	1.068	1.107	1.107	0.217	0.297	0.308	0.467	0.356	0.428
TA (size effect)	912.0	912.0	1647.0	1647.0	12557.5	12557.5	7452.0	65517.5	13977.0	40387.2	978.5	13808.3

Notes:
For variables see Table 1.
Time index:
no index – one year prior to distress event,
_L1 – two years prior to distress event,
_L2 – three years prior to distress event,
_L3 – four years prior to distress event.
Statistical information of clustering:
Pseudo F statistic = 11.21,
Approximate expected over-all R-squared = 0.3522,
Cubic clustering criterion = 2.060.

Table 5
The final stepwise logistic regression models estimated separately for each of the four years before the distress event (n = 99 distressed and 99 non-distressed banks)

Variable	Estimate	Standard error	Wald Chi-square statistic	p-value
Panel 1. Estimated year -1 model				
Intercept	5.3422	1.3893	14.79	0.0001
EQ_to_TA	-22.5017	5.7128	15.51	<0.0001
Deposits_to_G_Loans	-0.3709	0.1933	3.68	0.0549
L_Imp_to_G_Loans	28.3772	12.1465	5.46	0.0195
Loans_to_TA	-4.8094	1.6092	8.93	0.0028
Liquid_A_to_Funding	-4.1037	1.0730	14.63	0.0001
Panel 2. Estimated year -2 model				
Intercept	4.2347	1.2491	11.49	0.0007
Growth_EQ	-1.1094	0.6161	3.24	0.0718
Growth_G_Loans	1.6375	0.7525	4.74	0.0295
EQ_to_TA	18.2335	5.6075	10.57	0.0011
L_Imp_to_G_Loans	33.8642	12.5405	7.29	0.0069
Loans_to_TA	-6.1087	1.8633	10.75	0.0010
Loans_to_Funding	1.5977	0.5434	8.65	0.0033
Liquid_A_to_Funding	-4.2888	1.1377	14.21	0.0002
Panel 3. Estimated year -3 model				
Intercept	4.6988	1.2133	15.00	0.0001
EQ_to_TA	10.7103	4.8260	4.93	0.0265
Loans_to_TA	-5.6987	1.6454	11.99	0.0005
Loans_to_Funding	0.9085	0.4974	3.34	0.0678
Liquid_A_to_Funding	-4.7904	1.2007	15.92	<0.0001
Panel 4. Estimated year -4 model				
Intercept	3.6193	1.0451	11.99	0.0005
Loans_to_TA	-5.7880	1.6316	12.58	0.0004
Loans_to_Funding	1.2945	0.5141	6.34	0.0118
Liquid_A_to_Funding	-4.2972	1.0733	16.03	<0.0001

Table 6
Performance of the stepwise logistic regression models estimated separately for the four years before the distress event

Performance measure	Estimation year			
Panel 1. General performance measures				
Performance measure	Year -1	Year -2	Year -3	Year -4
Aikaike information criterion (AIC)	246.1	251.9	257.3	259.4
Schwarz criterion (SC)	265.8	278.2	273.8	272.5
R-square	0.185	0.177	0.128	0.110
Max-rescaled R-square	0.246	0.236	0.171	0.147
Panel 2. Areas under the ROC curve (AUCs)				
Model type	Year -1	Year -2	Year -3	Year -4
Estimated model	0.768	0.768	0.714	0.690
Jack-knife cross-validated model	0.734	0.719	0.677	0.659
Panel 3. AUC when the year's model is applied to alternative years[a]				
Applied to year	Year -1	Year -2	Year -3	Year -4
Year -1	0.734	0.746	0.752	0.692
Year -2	0.733	0.719	0.708	0.678
Year -3	0.688	0.699	0.677	0.688
Year -4	0.651	0.672[b]	0.689	0.659
Panel 4. Estimation year correct classifications (%)				
Classified banks	Year -1	Year -2	Year -3	Year -4
Percent of correctly classified distressed banks	77.78	67.68	64.65	59.60
Percent of correctly classified non-distressed banks	65.66	69.70	64.65	64.65
Percent of correctly classified (overall)	71.72	68.69	64.65	62.13
Panel 5. Estimation year (jack-knife) cross-validation results (%)				
Classified banks	Year -1	Year -2	Year -3	Year -4
Percent of correctly classified distressed banks	74.75	62.63	61.62	58.59
Percent of correctly classified non-distressed banks	63.64	62.63	63.64	64.65
Percent of correctly classified banks (overall)	69.20	62.63	62.63	61.62

Table 6, cont'd

Panel 6. Correct classifications (%) when the year's model is applied to alternative years[a]				
Applied to year	Year -1	Year -2	Year -3	Year -4
Year -1				
Distressed banks	74.75	81.82	79.80	75.76
Non-distressed banks	63.64	56.57	53.54	56.57
All banks	69.20	69.20	66.67	66.17
Year -2				
Distressed banks	57.58	62.63	69.70	68.69
Non-distressed banks	75.76	62.63	63.64	55.56
All banks	66.67	62.63	66.67	62.13
Year -3				
Distressed banks	48.48	59.60	61.62	66.67
Non-distressed banks	74.75	67.68	63.64	62.63
All banks	61.62	63.64	62.63	64.65
Year 4				
Distressed banks	41.41	51.81[b]	57.58	58.59
Non-distressed banks	74.75	70.51[b]	68.69	64.65
All banks	58.08	61.16[b]	63.14	61.62

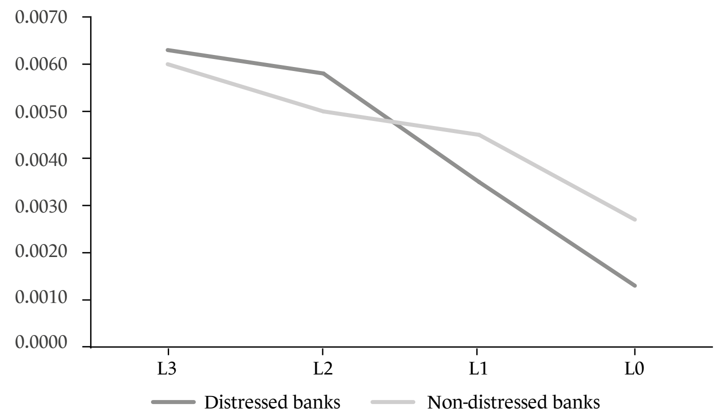
Notes:
Logistic models (Equation 1) estimated stepwise (forward selection) for the 12 independent variables in Table 1 separately for each year. The data from four years before the event does not include the growth variables.
Year -1 – one year prior to the event,
Year -2 – two years prior to the event,
Year -3 – three years prior to the event,
Year -4 – four years prior to the event.
[a] When the model is applied to the data from the year of its estimation, the cross-validated AUC is shown.
[b] LR model estimated for the data from two years before the event contains growth variables, which are missing in data from four years before the event for 16 distressed and 21 non-distressed banks.

Table 7
Correct classifications of the estimated one-year models by year and status across clusters

Period	Panel 1. Estimation data results									
	Distressed banks				**Non-distressed banks**					
	Clu-ster 1	**Clu-ster 2**	**Clu-ster 3**	**Clu-ster 4**	**Clu-ster 1**	**Clu-ster 2**	**Clu-ster 3**	**Clus-ter 4**	**Clus-ter 5**	**Clus-ter 6**
Year -1										
Number: correct/total	2/2	44/60	9/10	22/27	1/1	1/1	2/2	48/68	8/11	5/16
Percent correct	100.0	73.3	90.0	81.5	100.0	100.0	100.0	70.6	72.7	31.3
Year -2										
Number: correct/total	2/2	39/60	10/10	16/27	1/1	0/1	2/2	50/68	6/11	10/16
Percent correct	100.0	65.0	100.0	59.3	100.0	0.0	100.0	73.5	54.6	62.5
Year -3										
Number: correct/total	2/2	41/60	6/10	15/27	1/1	1/1	2/2	43/68	6/11	11/16
Percent correct	100.0	68.3	60.0	55.6	100.0	100.0	100.0	63.2	54.6	68.8
Year -4										
Number: correct/total	1/2	41/60	3/10	14/27	1/1	1/1	2/2	44/68	5/11	11/16
Percent correct	50.0	68.3	30.0	51.9	100.0	100.0	100.0	64.7	45.5	68.8
Period	**Panel 2. Jack-knife (Lachenbruch) cross-validation results**									
	Distressed banks				**Non-distressed banks**					
	Clus-ter 1	**Clus-ter 2**	**Clus-ter 3**	**Clus-ter 4**	**Clus-ter 1**	**Clus-ter 2**	**Clus-ter 3**	**Clus-ter 4**	**Clus-ter 5**	**Clus-ter 6**
Year -1										
Number: correct/total	1/2	42/60	9/10	22/27	1/1	1/1	2/2	46/68	8/11	5/16
Percent correct	50.0	70.0	90.0	81.5	100.0	100.0	100.0	67.7	72.7	31.3
Year -2										
Number: correct/total	1/2	36/60	10/10	15/27	1/1	0/1	2/2	46/68	5/11	8/16
Percent correct	50.0	60.0	100.0	55.6	100.0	0.0	100.0	67.7	45.5	50.0
Year -3										
Number: correct/total	2/2	40/60	5/10	14/27	1/1	1/1	2/2	42/68	6/11	11/16
Percent correct	100.0	66.7	50.0	51.9	100.0	100.0	100.0	61.8	65.6	68.8
Year -4										
Number: correct/total	1/2	40/60	3/10	14/27	1/1	1/1	2/2	44/68	5/11	11/16
Percent correct	50.0	66.7	30.0	51.9	100.0	100.0	100.0	64.7	45.5	68.8

Figure 1

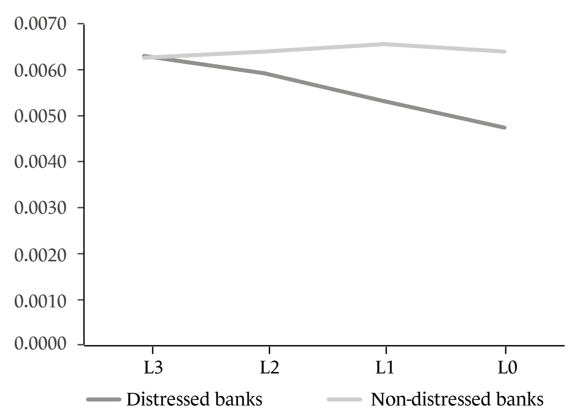
The development of the median of ROA towards the distress event year



Notes:
L0 – one year before the event,
L1 – two years before the event,
L2 – three years before the event,
L3 – four years before the event,
ROA – return on total assets ratio.

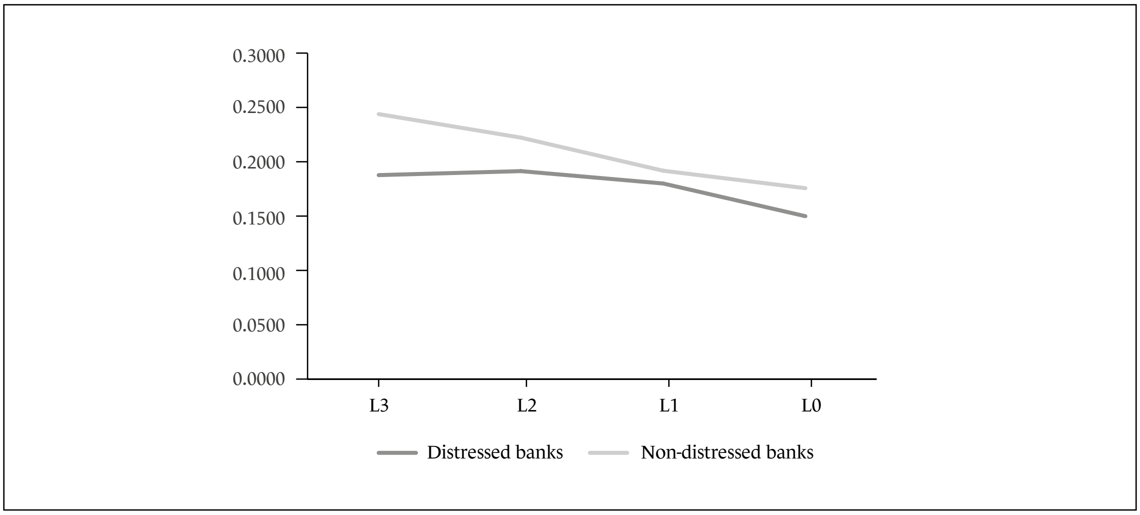
Figure 2

The development of the median of EQ_to_TA towards the distress event year



Notes:
L0 – one year before the event,
L1 – two years before the event,
L2 – three years before the event,
L3 – four years before the event,
EQ_to_TA – equity to total assets ratio.

Figure 3
The development of the median of Liquid_A_to_Funding towards the distress event year



Notes:
L0 – one year before the event,
L1 – two years before the event,
L2 – three years before the event,
L3 – four years before the event,
Liquid_A_to_Funding – liquid assets to deposits and short-term funding ratio.