FINANCIAL MARKETS, INSTITUTIONS & INSTRUMENTS
Topics in Financial Intermediation
NEW YORK UNIVERSITY SALOMON CENTER
NYU STERN

How Accurately Can Z-score Predict Bank Failure?

By Laura Chiaramonte, (Frank) Hong Liu, Federica Poli, and Mingming Zhou

Bank risk is not directly observable, so empirical research relies on indirect measures. We evaluate how well Z-score, the widely used accounting-based measure of bank distance to default, can predict bank failure. Using the U.S. commercial banks' data from 2004 to 2012, we find that on average, Z-score can predict 76% of bank failure, and additional set of other bank- and macro-level variables do not increase this predictability level. We also find that the prediction power of Z-score to predict bank default remains stable within the three-year forward window.

Keywords: Z-score, bank failure, financial crisis.
JEL Classification: E37, G01, G21.

I. INTRODUCTION

This paper assesses the validity of Z-score proposed by Boyd and Graham (1986) as a bank risk measure. Z-score has been widely applied as an indicator of bank's distance-to-default in both academic research and practice. It is calculated as the sum of bank's return on assets and equity to assets ratio divided by the standard deviation of return on assets. It is an estimate of the number of standard deviations below the mean that bank's profits would have to fall to make the bank's equity negative. Higher values of Z-score are thus indicative of low probability of insolvency and greater bank stability. The attractiveness of Z-score relies on the fact that it does not require strong assumptions about the distribution of returns on assets (Boyd and Graham, 1986; Hannan and Hanweck, 1988; Strobel, 2011), which represents an especially interesting advantage from the practitioner's point of view. The popularity of Z-score also originates from its relative simplicity and the capability to compute it using solely accounting information. Contrary to market-based risk measures which are computable just for listed financial institutions and may raise estimation concerns stemming from the size of available samples, Z-score is applicable when dealing with an extensive number of unlisted as well as listed entities.

Despite the advantages attributable to the Z-score, however, it is not immune from some caveats. First, its reliability depends on the quality of underlying accounting and auditing framework. Such an issue is more prominent in cross-country studies due to the degree of each country's institutional development. Second, as banks may smooth out accounting data over time, the Z-score may offer an excessively positive assessment of the risk of bank insolvency. Third,

Corresponding author: Federica Poli, Banking at the Università Cattolica del Sacro Cuore, Milan, Italy. Email: federica.poli@unicatt.it.

by definition, Z-score is highly sensitive to the standard deviation of ROA.[1] In addition, given the tendency of the dominance of equity to assets ratio in calculating bank's Z-score, the magnitude of the differences in Z-scores may not correspond linearly to the differences in bank risk, since the variation of ROA is only a minor part of the calculation in the numerator.[2] Furthermore, as suggested by Huizinga and Laeven (2012), banks tend to overstate their value of distressed assets and regulatory capital during the U.S. mortgage crisis, and the calculation of Z-score based on the accounts reported by the bankers may thus be biased upward towards a safer ratio. Hence, despite the popularity of Z-score in banking literature as a proxy for distance-to-default given its soundness in theory, how well it perms in forecasting default is still unknown.

In this study, we examine two research questions. First, we analyze whether Z-score is a sufficient statistic to predict bank failure. Second, we investigate whether the predicting power of bank failures could significantly increase by adding additional bank-specific and macro variables in the forecasting model. We test these empirical questions in the following ways. We incorporate various versions of Z-score into a complementary log-logistic (clog-log) model that determines US bank failure from 2004 through 2012. Considering both Type I and I errors, we compare the performance of three bank failure prediction models that: (i) include Z-score as the only predictor, (ii) include a set of bank- (other than Z-score) and macro-level variables as the predictors, and (iii) include only the combination of the set of bank- and macro-level variables as the predictors. Further, we compare the short-term, out-of-sample forecasting ability of Z-score to that of the combination of Z-score and a set of other bank- and macro-level variables. Finally, we examine the ability of Z-score to explain Merton Distance-to-default, a market based bank risk measure.

We find strong empirical evidence to provide affirmative answer to both questions. First, we find that on average, Z-score together with time fixed effects are able to predict bank failures with the accuracy of 76% (based on Type I errors), while adding a set of other bank-specific and macro variables do not increase the predictability accuracy. Besides, the out-of-sample forecasting performance of Z-score shows that the lowest two deciles of Z-score can predict on average 74% of bank failures across the whole sample. We also find that Z-score is a significant determinant factor of Merton DD measure, indicative of high correlation between the two widely used bank risk measures. Finally, we show that the prediction power of Z-score remains stable within the forward three-year window.

[1]For example, consider two banks A and B, both with equity ratio being 0.04. Bank A has average ROA being 0.01 and standard deviation of ROA being 0.001, hence the Z-score for Bank A is 50. While Bank B has higher ROA of 0.02, however, its standard deviation of ROA is also significantly higher, with being 0.002. Thus Bank B's Z-score is 30. Although both banks have proportional ROAs (0.01 vs. 0.02) and its standard deviations (0.001 vs. 0.002), Z-score shows that Bank A is twice as safe as Bank B.

[2]Our data shown in Table 2 indicates that average equity to assets ratio is 11% while average ROA is only 0.9%. Therefore, unless a bank has consistently considerable loss over time, Z-score is more likely to be dominated by changes in equity to asset ratios than changes in ROA.

Assessing the Z-score's accuracy in measuring bank risk is important for several reasons. First, since a bank's risk is not directly observable, the empirical literature finds itself having to rely on indirect proxies, which should be sound both theoretically and empirically. Even though Z-score is a widely used bank risk measure among many researchers and practitioners, its statistical properties are not yet known. It is hence important to demonstrate the validity of this measure, and whether it can indeed reflect the underlying bank risk. Second, given the simplicity and transparency of the calculation of Z-score, establishing its predictive power for bank failures would have extensive implications for both policy makers and practitioners, who are currently looking for effective measure of bank risk in their policy making process or risk management of the banking sector. Third, given that our measures of Z-score does not rely on whether the bank is publicly traded, it can be widely applied to both publicly listed banks and private banks, and this is an important advantage over most systemic risk measures proposed so far that are heavily based on stock price information of the bank (see, e.g., Acharya *et al.*, 2012; Billio *et al.*, 2012). Fourth, establishing Z-score as an effective predictor for bank failure in our empirical study also implies that the disclosure quality regarding bank's earnings and equity is crucial to improve information environment for banks, and that any managerial incentives or regulations that give rise to earnings smoothing in the banking industry might lead to under-estimation of default risk by outsiders.[3]

Our paper also contributes to the current surging literature on various factors that may lead to bank failure. These literature examine both micro-level factors such as bank ownership and corporate governance, subprime lending and loan securitization, as well as macro-level factors such as bank competition and regulations (see, e.g., Akins *et al.*, 2014; Beck *et al.*, 2013; Brown and Dinc, 2011; DeYoung and Torna, 2013; Erkens, 2012; Gorton and Metrick, 2012; Ivashina and Scharfstein, 2010; Martinez-Miera and Repullo, 2010; Repullo and Suarez, 2013).

Finally, our research also complements to Altman's (1968) Z-score based on multiple discriminant analysis (Balcaen and Ooghe, 2006). Altman proposes a model of five variables to predict bankruptcy up to "two years prior to distress and that accuracy diminishes substantially as the lead time increases" (Altman, 2000).[4] However, as well spelled out in these studies, the Altman's (1968) Z-score

[3] In this sense, our study is also related to Jin *et al.* (2011) who develop six and ten accounting and audit quality variables to predict whether banks failed during the financial crisis starting from 2007. For recent studies on managerial incentives that give rise to earnings smoothing for financial industries, see Cheng *et al.* (2011) and Eckles *et al.* (2011), and for discussions on how regulations could change earnings smoothing incentives for bank managers, see Kilic *et al.* (2012).

[4] The variables used in his 1968 seminal study are: (1) working capital/total assets, (2) retained earnings/total assets, (3) earnings before interest and taxes/total assets, (4) market value equity/book value of total liabilities, and (5) sales/total assets. Given that the initial model was developed to predict failure of publicly traded listed manufacturing firms, later in Altman (2000), Altman modified his original model to predict failures in private and in publicly traded listed non-manufacturing firms (1984), known as the "revised" or "alternative" Z-score model.

(along with the Altman *et al.*'s (1977) Zeta credit risk model, or the 2000 modified Z-score) is mostly applicable to industrial corporations instead of banks.

We organize the remainder of this paper as follows. Section II describes the data sample and how we identify failure events. Section III discusses the methodology as well as the variables used in our paper and their descriptive statistics. Sections IV and V present empirical results and robustness tests. Section VI concludes.

II. DATA

We obtain fourth-quarter data from 2003 to 2012 on private and public commercial banks in the U.S. from the Reports on Condition and Income ("Call Reports") submitted by insured banks to the Federal Reserve.[5] Following Berger *et al.* (2004), we study only commercial banks and exclude savings banks, savings and loan associations, credit unions, investment banks, mutual banks, and credit card banks We use bank-level data and treat each individually chartered bank as a separate entity.

Our final sample consists of 8,478 unique banks (there are totally 58,017 bank-year observations), out of which 552 failed and 7,926 are active. The information on bank failure is obtained from the inactive bank data provided by Federal Deposit Insurance Corporation (FDIC). The FDIC lists all banks that were closed owing to bankruptcy, merger and acquisition (M&A) and change of charter among other causes of closure, and provides a structural change coding for the reason for closure and the date of closure. We define these bank closures as failure. Table 1 presents the sample distribution by bank status (active versus failed banks) in each year during the sample period 2004–2012. It shows that the majority of bank failure events in the U.S. took place during the 2007–09 financial crisis. Specifically, in our sample, more than 400 commercial banks under FDIC supervision failed after (or during) 2007 compared to less than 80 between 2004 and 2006. In light of the numerous bank failure events in the recent years, in our empirical analysis we investigate the suitability of the Z-score as a measure of bank failure not only in the whole period (2004–2012), but also on the crisis and post-crisis period (2007–2012).

III. METHODS

Discrete-Time Proportional Hazards Model

To empirically investigate whether and to what extent the Z-score is an informative measure of bank risk, we use a discrete-time representation of a continuous-time proportional hazards model, the so-called complementary log-log model where the dependent variable (the failed bank dummy) is a binary variable that takes value 0 when a bank is still active and 1 when it failed.

[5] We use yearly data instead of quarterly data to minimize the seasonal effects of bank performance.

Table 1: Distribution of Failed and Active Banks Over the Sample Period From 2004 to 2012

	Bank-year observation		
Year	Failed	Active	Total
2004	9	6,985	6,994
2005	4	6,779	6,783
2006	26	6,607	6,632
2007	40	6,453	6,493
2008	134	6,350	6,484
2009	154	6,265	6,419
2010	96	6,178	6,274
2011	68	6,936	6,004
2012	22	5,912	5,934
Total	552	57,465	58,017

This table shows the sample distribution by bank status (active banks versus failed banks) in each year. The numbers reported in the table refers only to those banks with data available to compute our main variable of interest (the natural logarithm of the Z-score). We obtain fourth-quarter data from 2004 to 2012 on private and public commercial banks in the US from the Reports on Condition and Income ("Call Reports") submitted by insured banks to the Federal Reserve.

Complementary log-log model is frequently used when the probability of an event is very small or very large, as the logit and probit models are inappropriate under such circumstances. Complementary log-log model belongs to the discrete-time functional specifications applied when survival occurs in continuous time, but spell length are observed only in interval as it is the case for bank failure recorded on annual basis in our sample. Guo (1993) observes that time-varying covariates offer an opportunity to examine the relation between the failure probability and the changing conditions under which the failure happens. The complementary log-log model with time-varying covariates has the following form (Männasoo and Mayes, 2009):

$$log(-log[1 - h_j(X)]) = \gamma_j + \beta' X \quad (1)$$

where X contains time-varying covariates for each bank at time $t - 1$. Traditional complementary log-log model assumes duration independence, i.e., the probability of surviving or failing at any point in time is always the same. In order to deal with time dependency problems arising when using these models, we use robust standard errors clustered on the unit of analysis and include in the vector X temporal dummy variables for each period or 'spell'. In addition, the complementary log-log model yields estimates of the impact of the indicators on the conditional probability of failure, which means that we obtain failure probabilities, conditional on surviving to a certain point in time.

In order to examine whether the model is able to correctly identify failed banks, we compute two types of errors: Type I and Type II errors. Type I error occurs when the model fails to identify the failed banks (that is a missed failure). It is computed as the ratio of false negative (FN) events to the sum of false negative and true positive (TP) events. Type II error occurs when a healthy bank is falsely identified as failure (that is a false alarm). It is computed as the ratio of false positive (FP) events to the sum of false positive and true negative (TN) events.

To assign a particular bank into one of the two categories (failed versus active), we set up a cut-off point in terms of the probability of bank failure. All banks above (below) that cut-off point are considered as failed (healthy) banks. A higher cut-off point results in a lower number of banks on the blacklist of failed banks, which tends to increase the Type I errors. Setting a lower cut-off point can reduce the Type I errors, but at the expense of generating more Type II errors. The optimal cut-off point depends on the relative weights that an advisable puts on Type I and Type II errors. From a prudential perspective, it is considerate to put a larger weight on Type I errors (Persons, 1999), because supervisors are primarily concerned about missing a failed bank (Poghosyan and Čihák, 2011). This implies a preference for relatively low cut-off points, which limit the Type I errors at the expense of relatively long blacklists (and potentially more Type II errors). For these reasons, we primarily focus on the Type I error results obtained using the cut-off point equal to 1%.

The analysis based on Type I and II errors is based on the arbitrary decision of the cut-off point. To overcome this problem, we also assess the accuracy of failure forecasts using the empirical distribution of the predicted probabilities of failure generated by complementary log-log model. We assign each observation to a decile of this empirical distribution, and we count how many genuine failure events fall into each decile. The accuracy of the model increases when a high fraction of failure events fall in the deciles associated to high predicted probabilities of failure.

The Estimation of Z-score

Despite various shortcomings of Z-score, a number of approaches have been developed for the Z-score's construction, and abundant empirical studies employ Z-score as proxy for bank risk (see, e.g., Boyd and Graham, 1986; De Nicolò, 2000; Stiroh, 2004; Beck and Laeven, 2006; Laeven and Levine, 2009; Beck *et al.*, 2013; Chiaramonte *et al.*, 2015; DeYoung and Torna, 2013; Liu *et al.*, 2013).

We compute the Z-score following different approaches developed by the literature for its construction (see the variable definition in the Appendix). On the basis of the most common approach (Boyd and Graham, 1986; Hannan and Hanweck, 1988), the first Z-score used in our analysis (hereafter 'Z-score 1') is calculated as the sum of equity to total assets (ETA) and return on assets (ROA) divided by the three-year standard deviation of ROA (σROA). Following Maecheler *et al.* (2007), we also compute the Z-score using the three-year moving return of assets (A_ROA) plus the three-year moving average of equity to total assets (A_ETA) over the three-year standard deviation of A_ROA (σA_ROA). We label this type of Z-score as 'Z-score 2'. The third way of estimation of the Z-score follows Boyd *et al.* (2006) and is calculated as the sum of three-year moving average of equity to total assets (A_ETA) and current values of return on assets (ROA) divided by the three-year standard deviation of ROA (σROA). We label this type of Z-score as 'Z-score 3'. Finally, following Laeven and Lavine (2009) and Dam and Koetter (2012), we compute the Z-score as the sum of tier 1 ratio (TIER 1 RATIO) and return on risk weighted assets (R_RWA) divided by the three-year standard deviation of R_RWA (σR_RWA). We label this type of Z-score as 'Z-score 4'. Since the Z-score is usually highly skewed, we use the natural logarithm of the Z-score, which is more likely to follow normal distribution (Laeven and Levine, 2009; Liu *et al.*, 2013). We label the natural logarithm of Z-score as lnZ.

Variables

We include several bank- and macro-level factors as control variables to capture differences in bank risk profiles that are associated with other bank characteristics, macroeconomic conditions or banking market structures. These different categories of indicators represent various determinants of a bank's vulnerability (see Betz *et al.*, 2014). In the Appendix, we describe the control variables outlined below and summarize their hypothesized relationships with the probability of bank failure.

The first control variable we consider is the natural logarithm of a bank's total assets as a proxy for bank size (SIZE). Existing literature indicates that the sign linking SIZE to the probability of bank failure could be uncertain. The relationship can be negative when growth of bank size leads to efficiency gains and superior ability of diversification, which would result in higher bank stability. On the other hand, the relationship may become positive when diversification strategies followed by large banks do not make them safer and may exacerbate the risk of a system-wide breakdown (Allen and Jagtiani, 2000) or result in higher earnings volatility while relying on the implicit guarantee associated with the too-big-to-fail argument (DeYoung and Roland, 2001; DeJonghe, 2010, Demirguc-Kunt and Huizinga, 2010).

Next, we include bank diversification (DIV) as another control variable and measure it by the ratio of non-interest income to total operating income following Stiroh (2004). We expect a negative sign between DIV and the probability of bank failure because diversification leads to risk reduction and therefore lower the likelihood of failure.

In addition, we employ the ratio of the sum of cash, available-for-sale securities and federal funds sold to total assets (LIQ) as a proxy for bank liquidity. The relationship linking LIQ to bank failure is expected to be negative. The more liquid the bank is and the less vulnerable to a classic run. An increase in LIQ should therefore correspond to a reduction in probability of bank default. In addition, we include the ratio of non-performing loans to total assets (NPL) as a proxy for asset quality. The higher ratio of NPL indicates the lower quality of the bank loan portfolio. Hence, an increase in NPL should lead to an increase in probability of bank failure. Furthermore, we employ the cost-to-income ratio (CIR) as a proxy for bank operational efficiency. Since low values of CIR indicate better managerial quality, the relationship between CIR and profitability of bank failure is expected to be positive.

Finally, within the bank-specific factors, we include the Bank Holding Company (BHC) dummy variable, which takes the value of 1 if the bank is owned by a BHC and 0 otherwise. We expect a negative sign between BHC dummy and bank failure. A bank that is a part of a BHC may be subject to more complex risk management and stricter monitoring because BHCs boards have more committees and meet more frequently than other boards (Adams and Mehran, 2003). The increased corporate governance may thus reduce the likelihood of bank failure.

In our empirical analysis, we also consider the most commonly used macroeconomic indicators: the annual percentage change of gross domestic product (GDPC) and the annual inflation rate (INF). We expected that low GDP growth and high inflation increase bank vulnerability (see Betz *et al.*, 2014). Hence, we hypothesize a negative sign for GDPC and a positive sign for INF.

To measure the degree of banking system concentration, we determine the Herfindahl–Hirschman index (hereafter HHI). The HHI is calculated as the sum of the squared market share value (in term of total assets) of all banks in the country. The theoretical relationship linking HHI to bank survival is uncertain based on the previous studies. The competition-fragility view expects a positive sign as competitive markets limit the ability of banks to gain informational advantages from their relationships with borrowers, reducing their incentives to properly screen borrowers, thus increasing the risk of default (Allen and Gale, 2000, 2004; Carletti, 2008; Beck *et al.*, 2013). Contrary to this view, the competition-stability view (Boyd an De Nicolò, 2005) predicts a negative sign and maintains that highly competitive banking systems (i.e., lower HHI) result in more stability. If competition reduces the cost of financing, bank borrowers would be better able to repay their loan obligations, thus reducing the risk of bank failure due to credit risk. Given the unsolved contradictions of predictions from the existing theories, we leave the sign for the coefficient of the HHI variable to empirical testing.

Summary Statistics

Table 2 reports descriptive statistics of the variables used in our U.S. sample for the whole sample period from 2004 to 2012, tabulated by bank status (active or failed). To mitigate the effect of outliers, we winsorize observations in the outside 1% of each tail of each explanatory variable, with the exception of SIZE.

As expected, active banks show higher values for the average lnZ than failed banks for all types of Z-score in the time period considered. This result can be largely explained both by a lower volatility of returns (proxied by the standard

Table 2: Summary Statistics of Variables By Failed and Active Banks

	ACTIVE BANKS			FAILED BANKS			ACTIVE and FAILED	FULL SAMPLE		
Variables	N. of observation	Mean	Standard Deviation	N. of observation	Mean	Standard Deviation	BANKS differences	N. of observation	Mean	Standard Deviation
lnZ, (Z-score 1)	57,571	3.938	1.178	552	2.254	1.714	1.684^{***}	59,784	3.916	1.196
lnZ, (Z-score 2)	57,571	3.963	1.121	552	2.604	1.363	1.359^{***}	59,784	3.945	1.131
lnZ, (Z-score 3)	57,571	3.959	1.129	552	2.495	1.467	1.464^{***}	59,784	3.940	1.143
lnZ, (Z-score 4)	57,571	3.867	1.190	552	2.133	1.690	1.734^{***}	59,784	3.844	1.209
ETA	57,571	11.224	7.161	552	9.902	8.802	1.322^{***}	59,784	11.212	7.206
ROA	57,571	0.883	2.317	552	−1.086	8.271	1.969^{***}	59,784	0.858	2.439
σ_{ROA}	57,571	0.440	1.008	552	1.607	5.008	-1.167^{***}	59,784	0.453	1.122
SIZE	57,571	11.860	1.311	552	12.373	1.396	-0.513^{***}	59,784	11.869	1.319
DIV	57,567	16.831	12.386	552	15.505	16.759	1.326^{*}	59,777	16.843	12.467
LIQ	57,571	12.797	12.519	552	10.307	10.550	2.490^{***}	59,784	12.768	12.532
NPL	57,571	0.157	0.318	552	0.264	0.492	-0.107^{***}	59,784	0.157	0.320
CIR	57,567	46.598	13.616	552	53.830	23.058	-7.232^{***}	59,777	46.631	13.830
GDPC								59,784	1.641	1.946
INF								59,784	2.344	0.779
HHI								59,784	452.979	88.277

This table reports the summery statistics of the four different measures of the natural logarithm of the Z-score (i.e., our main variable of interest), of its components and of the control variables (bank-specific and macro factors) used in our analysis. We report only the descriptive statistics for the components of the Z-score 1 given that the components of the other different types of Z-score show a similar trend. The estimates are done by bank status and on the full sample, with the sole exception of the macro variables that are observed only with reference to the whole sample. The ‘full sample’ includes the failed and active banks. The descriptive statistics are referred to the whole period (2004–2012). To mitigate the effect of outliers, we winsorize observations in the outside 1% of each tail of each variable, except for SIZE, GDPC, INF and HHI. All the variables, except SIZE and HHI, are in percentage.The numbers reported in the table refers only to those banks with data available to compute our main variable of interest (the natural logarithm of the Z-score). See the Appendix for the description of different Z-score and of the control variables used in the paper. ***, ** and * are referred to the two-sided unpaired t-test statistical significance at 1%, 5%, and 10% respectively.

deviation ROA) and by higher average ROA values of active banks compared to failed banks. Failed banks also show lower level of capitalization (ETA) compared to active banks. Overall, the difference in terms of the mean test between active and failed banks for the Z-score and its components is statistically significant at the 1% level during the whole period.

With regard to bank-specific characteristics observed by bank status, it emerges that failed banks are larger in size than surviving banks. This finding is in line with that of Jin *et al.* (2011). Additionally, banks that experienced a failure showed poorer quality loans portfolio, lower efficiency, less diversified into non-interest income activities and holding less liquidity. All these characteristics helped healthy banks to survive during the period of analysis. The latter results are confirmed by the more recent U.S. bank failure literature (Jin *et al.*, 2011; Dam and Koetter, 2012). Overall, the differences in terms of mean test between active and failed banks for the bank-specific variables are statistically significant at the 1% level during the period 2004–2012.

We also observe low values of inflation ratio (INF) and bank concentration (HHI) with low variations throughout the period while the annual GDP growth (GDPC) shows relevant changes. Finally, Table 3 presents the correlation matrix for our main variables of interest (the four measures of Z-score), its components and the control variables. It shows that all the four Z-scores we construct are highly correlated with one another as expected. It also shows that though many of the pairwise correlation coefficients are statistically significant, the correlation magnitudes are in general low.

IV. MAIN RESULTS

Regression Analysis and Prediction Results

Table 4 shows the complementary log-log models estimations results and also displays the relationship between model predictions and actual failure events (see Type I and II errors) using a cut-off point equals to 1%. In order to investigate to what extent Z-score is a sufficient statistic of bank failure, for each measures of Z-score, we test the model on Z-score alone, and the combination of Z-score and the common bank- and macro-level control variables. In the final column, we also test the predictive power of control variables without the inclusion of Z-score. We also include time fixed effects in all our regressions. The bottom of Table 4 displays the relationship between model predictions and actual failure events for the complementary log-log model for the entire sample period (2004-2012) using a cut-off point of 1%.

Table 4 shows that on average Z-score can accurately predict 76–77% of bank failures. For example when Z-score 1 is the only independent variable included in the hazard model (and with year fixed effects added), the Type I error is 23.9% while the Type II error is 21.8%, indicating that 23.9% of the time Z-score 1 fails to identify the failed banks and 21.8% of the time a healthy bank is falsely

Table 3: Correlations

	lnZ, (Z-score 1)	lnZ, (Z-score 2)	lnZ, (Z-score 3)	lnZ, (Z-score 4)	ETA	ROA	σ_{ROA}	SIZE	DIV	LIQ	NPL	CIR	GDPC	INF	HHI
lnZ (Z-score 1)	1.000														
lnZ (Z-score 2)	0.989*	1.000													
lnZ (Z-score 3)	0.992*	0.999*	1.000												
lnZ (Z-score 4)	0.950*	0.935*	0.939*	1.000											
ETA	0.076*	0.063*	0.065*	0.048*	1.000										
ROA	0.223*	0.185*	0.200*	0.223*	0.283*	1.000									
σ_{ROA}	−0.503*	−0.489*	−0.490*	−0.503*	0.309*	0.180*	1.000								
SIZE	−0.032*	−0.210*	−0.020*	−0.028*	−0.034*	−0.037*	−0.058*	1.000							
DIV	−0.022*	0.351*	0.241*	0.205*	−0.036*	−0.034*	−0.034*	0.172*	1.000						
LIQ	0.024*	0.264*	0.045*	0.087*	0.027*	0.027*	0.043*	−0.280*	0.114*	1.000					
NPL	−0.037*	−0.019*	−0.013*	0.011*	−0.036*	−0.040*	−0.031*	−0.057*	−0.015*	−0.004	1.000				
CIR	−0.264*	−0.006	−0.331*	0.116*	−0.228*	−0.233*	−0.271*	−0.254*	−0.103*	0.063*	−0.035*	1.000			
GDPC	0.207*	−0.253*	0.107*	−0.105*	0.188*	0.199*	0.200*	−0.051*	0.010*	0.006	−0.039*	−0.089*	1.000		
INF	0.243*	−0.006	0.096*	−0.122*	0.234*	0.239*	0.226*	−0.018*	−0.001	−0.048*	−0.029*	−0.040*	0.654*	1.000	
HHI	−0.182*	0.048*	−0.086*	0.091*	−0.173*	−0.179*	−0.173*	0.095*	−0.043*	−0.034*	−0.010*	0.098*	−0.462*	−0.127*	1.000

This table shows the correlation matrix for the explanatory variables used in the empirical analysis: the four different measures of the natural logarithm of the Z-score (lnZ), our main variable of interest; the components of lnZ; and the control variables. We report only the results for the components of the Z-score 1 given that the correlations for the components of the other types of Z-score show the correct sign and are very low. See the Appendix for the description of the explanatory variables used in the paper. Data in the table are referred to the whole period: 2004–2012 (latest data available). The numbers reported in the table refers only to those banks with data available to compute our variable of interest (the natural logarithm of the Z-score). * indicates statistically significance at the 10% level.

Table 4: Complementary Log-log Model Estimations Results and Type I and II Errors

	Z-score 1		Z-score 2		Z-score 3		Z-score 4		
Variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	Control variables only
lnZ (−1)	−0.862***	−0.838***	−0.986***	−0.933***	−0.967***	−0.917***	−0.821***	−0.809***	
	(0.029)	(0.032)	(0.041)	(0.044)	(0.037)	(0.040)	(0.033)	(0.030)	
SIZE (−1)		0.171***		0.192***		0.181***		0.153***	0.252***
		(0.030)		(0.029)		(0.029)		(0.029)	(0.030)
DIV (−1)		−0.005		−0.007*		−0.006		−0.006	−0.001
		(0.003)		(0.004)		(0.004)		(0.003)	(0.005)
LIQ (−1)		−0.009*		−0.008		−0.009*		−0.008	−0.011*
		(0.005)		(0.005)		(0.005)		(0.005)	(0.004)
NPL (−1)		0.366***		0.380***		0.358***		0.392***	0.651***
		(0.093)		(0.092)		(0.092)		(0.093)	(0.092)
CIR (−1)		0.001		0.005**		0.004*		−0.0004	0.013***
		(0.002)		(0.002)		(0.002)		(0.002)	(0.001)
GDPC (−1)		−0.224***		−0.270***		−0.243***		−0.215***	−0.197***
		(0.032)		(0.033)		(0.032)		(0.032)	(0.031)
INF (−1)		1.009***		1.073***		1.023***		0.955***	0.651***
		(0.110)		(0.111)		(0.110)		(0.110)	(0.108)
HHI (−1)		0.020***		0.022***		0.021***		0.020***	0.017***
		(0.002)		(0.002)		(0.002)		(0.002)	(0.002)
BHC dummy		−0.052		−0.100		−0.137		−0.015	−0.259*
		(0.113)		(0.113)		(0.113)		(0.113)	(0.114)
Year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N. of Obs.	58,123	58,119	58,123	58,119	58,123	58,119	58,123	58,119	75,197

(*Continued*)

Table 4: (*Continued*)

	Z-score 1		Z-score 2		Z-score 3		Z-score 4		
Variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	Control variables only
Type Errors:									
TP	420	414	423	415	422	412	417	416	439
FN	132	138	129	137	130	140	135	136	173
FP	12,563	12,130	13,419	12,827	13,016	12,503	12,571	12,171	20,856
TN	45,008	45,437	44,152	44,740	44,555	45,064	45,000	45,396	53,729
Type I	0.239	0.250	0.233	0.248	0.235	0.253	0.244	0.246	0.282
Type II	0.218	0.210	0.233	0.222	0.226	0.217	0.218	0.211	0.279

This table shows a comparison of the complementary log-log model results obtained using alternatively the four different Z-score measures (i.e., our main variable of interest) alone and with the control variables. Finally, we also test the complementary log-log model on the control variables only (see last column). Each regressions is tested on the whole period, 2004–2012 (latest data available). The different types of the Z-score and the control variables used in this paper are described in the Appendix. Year dummy variables are also incorporated in the model. The robust standard errors of the estimated coefficients are reported in parentheses. ***, **, and * denote coefficients statistically different from zero at the 1%, 5%, and 10% levels in two-tailed tests, respectively. This table also displays the relationship between model predictions and actual failure events on the full sample for the whole period (see Panel A) using a cut-off point equals to 0.01. TP stands for 'True Positive'; FN stands for 'False Negative'; FP stands for 'False Positive'; TN stands for 'True Negative'. Type I error occurs when the model fails to identify the failed bank. It is computed as: FN/(FN+TP). Type II error occurs when a healthy bank is falsely identified as failed (i.e., a false alarm). It is computed as: FP/(FP+TN).

identified as a failing bank by using the information of Z-score 1 only[6]. In the last column of Table 4, we report the results by considering alternative set of other bank-specific and macro variables, and find that both Type I error (28.2%) and Type II error (27.9%) are higher than those when Z-score alone is considered, suggesting a better predictability using Z-score alone in comparison to using the set of other bank-specific and macro variables as we defined earlier as independent variables. For each Z-score variable, we also report the results by combining the Z-score and the other bank-specific and macro variables, and we find that the latter leads to slightly higher Type I errors while slightly lower Type II errors. These results suggest that by adding a set of other bank-specific and macro variables to the Z-score does not significantly improve the predictability of our hazard model.

Table 4 also shows that during the period 2004–2012, the natural logarithm of the Z-score (lnZ) enters the regressions significantly at 1% level and negatively in all the cases considered, indicating that the significance of Z-score as a predictor of bank failure does not disappear once the other variables are controlled. The negative sign of the coefficient means that higher values of Z-score are indicative of lower likelihood of bank failure.

In Table 4, we display that the empirical results of the control variables are in general consistent with our expectations. The positive sign of SIZE implies that larger banks take on higher risk which may endanger their probability of survival. Similarly, more concentrated banking markets result to increase the probability of bank default. Positive relationship is also found between the non-performing ratio (NPL) as a measure of asset quality and the probability of default. This result is consistent with those reported in Poghosyan and Čihák (2011) and Betz *et al.* (2014), who find that failure probabilities are influenced by the deterioration of the loan portfolio. Diversification (DIV) is found to have significant negative impact on the probability of bank failure when Z-score 2 (but not the other Z-scores) is considered, indicating that diversification leads to risk reduction and therefore lower the likelihood of bank failure. The bank's level of liquidity (LIQ) is found to have significant negative impact on the probability of bank failure when Z-score 1 and Z-score 3 (but not Z-score 2) are considered, indicating that banks with more liquidity are less vulnerable to bank failure. Cost-to-income ratio (CIR) as a measure of managerial inefficiency is also found to have a positive relationship with the likelihood of bank failure when Z-score 2 and 3 are considered.

The two macro-variables, INF and GDPC, show positive and negative signs, respectively. Hence, high inflation and low real GDP growth increase bank vulnerability, confirming the results of Betz *et al.* (2014).

Overall, Table 4 indicates that the Z-score, in all its computations, is a key determinant of the probability of bank survival, and the additional contribution of the bank-specific and macro variables to predict bank default is marginal at best.

[6]We also exclude from the model the time fixed effects to examine the predictive power of Z-score on its own. We find that on average the exclusion of time fixed effects increases the Type 1 error by 10% while the Type 2 error remains unchanged to that reported for the models with time fixed effects.

Default Forecasts

The predictive accuracy of the Z-score relative to the control variables with or without the Z-score is further confirmed by the failure forecasts in Table 5. Following Bharath and Shumway (2008), we assess the accuracy of our complementary log-log model by sorting banks in deciles based on the predicted probabilities and calculating the percentage of defaults by decile of the sole forecast variable (Z-score), the combination of Z-score and bank-specific and macro variables, and the set of control variables alone. Table 5 shows that the highest percentage of failure is in the tenth and ninth deciles (i.e., banks with the largest probability of failure or lowest value of Z-score) for all the specifications. By adding the other set of bank-specific and macro variables to the Z-score, however it is measured, will increase the predictability power of the tenth decile (for example, 64.31% vs. 61.59% for Z-score 1). However, the overall predictability of both tenth and ninth deciles will remain similar (for example, 73.91% vs. 73.54% for Z-score 1). Both these results with the inclusion of Z-scores report significant higher predictability power than that of control variables only. These results confirm that the Z-score alone is a good predictor of bank failure.

Z-Score Versus Merton Distance-To-Default Measure

In addition to the examination of the predictability of Z-scores to bank failure, we also examine to what extent Z-score, the accounting measure of bank distance-to-default, is consistent with the market price based Merton distance-to-default (DD), which is based on Merton's (1974) bond pricing model. Studies have demonstrated the ability of DD measures to predict default risk (Elton *et al.*, 2001; Gropp *et al.*, 2002; Vassalou and Xing, 2004). Kato and Hagendorff (2010) analyze the extent to which distance to default based on market data can be explained using accounting-based indicators of risk for a sample of U.S. bank holding companies. They show that a large number of bank fundamentals help to predict default for institutions that issue subordinated debt. However, they do not study the impact of Z-score on Merton DD. Gropp *et al.* (2002) empirically test European banks' distances-to-default and subordinated bond spreads in relation to their capability of anticipating a material weakening in banks' financial conditions. They use two different econometric models: a logit-model and a proportional hazard model. They find support in favor of using both indicators as leading indicators of bank fragility, regardless of the econometric specification. The predictive performance of the distance-to-default indicator is found to be robust between 6 to 18 months in advance, its predictive properties are quite poor closer to default.

We follow Bharath and Shumway's (2008) method to estimate the Merton DD model.[7] We examine all U.S. banks in the CRSP/Compustat Merged Database from 2003 to 2012, and then merged with CRSP to obtain stock price data. To examine the correlation between Z-scores and DD measure, we run a series of regressions

[7] The SAS commands for estimating the DD model can be found in Bharath and Shumway (2008).

Table 5: Default Forecasts

	Z-score 1		Z-score 2		Z-score 3		Z-score 4		
Deciles	lnZ only	lnZ and control variables	lnZ only	lnZ and control variables	lnZ only	lnZ and control variables	lnZ only	lnZ and control variables	Control variables only
10	0.6159	0.6431	0.5923	0.6304	0.5996	0.6394	0.6195	0.6394	0.4362
9	0.1195	0.0960	0.1340	0.1014	0.1340	0.0942	0.1159	0.1068	0.1944
8	0.0797	0.0815	0.0851	0.0923	0.0742	0.0905	0.0797	0.0778	0.0996
7	0.0452	0.0489	0.0416	0.0434	0.0489	0.0434	0.0507	0.0452	0.0915
6	0.0380	0.0326	0.0398	0.0271	0.0380	0.0307	0.0489	0.0434	0.0522
1–5	0.1014	0.0978	0.1068	0.1050	0.1050	0.1014	0.0851	0.0869	0.1258

This table reports the frequencies of default events by deciles of the distribution of the predicted probabilities for the complementary log-log model on the whole period (2004-2012), presented in Table 4. Decile 10 (1) is the decile with the highest (lowest) predicted probabilities of failure events. The model is tested on our main variable of interest, i.e., the natural logarithm of the Z-score (lnZ), the natural logarithm of the Z-score plus the control variables and on the control variables only (see last column).

Table 6: Comparison With Merton Distance Default (DD) Model

	(1)	(2)	(3)
lnZ (Z-score 1)	0.737*		
	(0.093)		
lnZ (Z-score 2)		0.900**	
		(0.017)	
lnZ (Z-score 3)			0.846**
			(0.021)
Constant	1.384	0.694	0.920
	(0.407)	(0.626)	(0.501)
N. of Obs.	5,689	5,795	5,795
Hansen	0.47	0.34	0.38
AR (2)	0.98	0.96	0.97

This table compares the Z-score by Merton (1974) distance default model. The three different types of the Z-score used are described in the Appendix. We follow Bharath and Shumway's (2008) method to estimate the DD model. We examine all banks in the CRSP/Compustat Merged Database from 2004 to 2012, and then merged with CRSP for stock price data. We use System GMM estimator with Windmeijer correction to all the regressions to address the potential endogeneity between the two bank stability measures. Hansen is the p-value of Hansen test statistic of over-identifying restrictions, while AR(2) is the p-value the second order autocorrelation test statistic.***, **, and * denote the statistical significance level at 1%, 5% and 10% respectively.

with the dependent variable being the DD measure, while the independent variable being different measures of Z-scores. Since both are bank risk measures, we use system generalized method of moments (GMM) estimator to treat the potential endogeneity issue between them. The results are reported in Table 6, where we observe that all our Z-score measures are significantly and positively correlated with the DD measure, which indicates that the accounting and market based bank risk measures are consistent with one another. This is the first attempt, to the authors' best knowledge, to examine the consistency of the accounting and market based bank risk measures and it strengthens the results in the previous sections that Z-score is an informative and reliable measure for bank risk.

Robustness Tests

In light of the numerous failure events that characterized the U.S. banking industry during the recent years, we investigate the suitability of the Z-score as a measure of bank risk during and after the crisis period of 2007–2012. Table 7 presents the complementary log-log models estimation results and displays the relationship between model predictions and actual failure events (see Type I and II errors) using a cut-off point equals to 1%. We test the model on the Z-score alone, the model with the combination of Z-score and the common bank- and macro level control variables and the model with the sole control variables.

Table 7: Complementary Log-log Model Estimations Results and Type I and II Errors in the Financial Crisis Period

	Z-score 1		Z-score 2		Z-score 3		Z-score 4		
Variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	lnZ only	lnZ and the control variables	Control variables only
lnZ (−1)	−0.879***	−0.850***	−1.020***	−0.960***	−0.994***	−0.935***	−0.837***	−0.820***	
	(0.030)	(0.033)	(0.042)	(0.046)	(0.038)	(0.042)	(0.027)	(0.031)	
SIZE (−1)		0.170***		0.190***		0.179***		0.152***	0.304***
		(0.032)		(0.031)		(0.031)		(0.031)	(0.031)
DIV (−1)		−0.009*		−0.011**		−0.010*		−0.010*	−0.018*
		(0.004)		(0.004)		(0.004)		(0.004)	(0.007)
LIQ (−1)		−0.012*		−0.011*		−0.011*		−0.010*	−0.016**
		(0.005)		(0.005)		(0.005)		(0.005)	(0.004)
NPL (−1)		0.388***		0.400***		0.379***		0.412***	0.686***
		(0.094)		(0.093)		(0.093)		(0.094)	(0.094)
CIR (−1)		0.001		0.005**		0.005*		−0.0003	0.015***
		(0.002)		(0.002)		(0.002)		(0.002)	(0.001)
GDPC (−1)		−0.224***		−0.272***		−0.243***		−0.215***	−0.196***
		(0.032)		(0.033)		(0.032)		(0.032)	(0.031)
INF (−1)		1.012***		1.083***		1.027***		0.956***	0.642***
		(0.111)		(0.112)		(0.111)		(0.111)	(0.109)
HHI (−1)		0.020***		0.022***		0.021***		0.020***	0.017***
		(0.002)		(0.002)		(0.002)		(0.002)	(0.002)
BHC dummy		0.066		0.015		−0.023		0.107	0.001
		(0.120)		(0.121)		(0.121)		(0.121)	(0.128)

(*Continued*)

Table 7: (*Continued*)

Variables	Z-score 1: lnZ only	Z-score 1: lnZ and the control variables	Z-score 2: lnZ only	Z-score 2: lnZ and the control variables	Z-score 3: lnZ only	Z-score 3: lnZ and the control variables	Z-score 4: lnZ only	Z-score 4: lnZ and the control variables	Control variables only
Year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
N. of Obs.	37,714	37,710	37,714	37,710	37,714	37,710	37,714	37,710	39,103
Type Errors:									
TP	413	410	415	409	413	407	412	414	426
FN	101	104	99	105	101	107	102	100	111
FP	11,884	11,475	12,580	11,999	12,250	11,765	11,930	11,570	19,617
TN	25,316	25,721	24,620	25,197	24,950	25,431	25,270	25,626	18,949
Type I	0.196	0.202	0.192	0.204	0.196	0.208	0.198	0.194	0.206
Type II	0.319	0.308	0.338	0.322	0.329	0.316	0.320	0.311	0.508

This table shows a comparison of the complementary log-log model results obtained using alternatively the four different Z-score measures (i.e., our main variable of interest) alone and with the control variables. Finally, we also test the complementary log-log model on the control variables only (see last column). Each regressions is tested on the crisis period, 2007–2012 (latest data available). The different types of the Z-score and the control variables used in this paper are described in the Appendix. Year dummy variables are also incorporated in the model. The robust standard errors of the estimated coefficients are reported in parentheses. ***, **, and * denote coefficients statistically different from zero at the 1%, 5%, and 10% levels in two-tailed tests, respectively. This table also displays the relationship between model predictions and actual failure events on the full sample for the whole period (see Panel A) using a cut-off point equals to 0.01. TP stands for 'True Positive'; FN stands for 'False Negative'; FP stands for 'False Positive'; TN stands for 'True Negative'. Type I error occurs when the model fails to identify the failed bank and is computed as: FN/(FN+TP). Type II error occurs when a healthy bank is falsely identified as failed (i.e., a false alarm) and is computed as: FP/(FP+TN).

Table 8: Complementary Log-log Model Estimations Results (Components of lnZ and its Lagged)

	(1)				(2)				(3)			
Variables	Z-score 1	Z-score 2	Z-score 3	Z-score 4	Z-score 1	Z-score 2	Z-score 3	Z-score 4	Z-score 1	Z-score 2	Z-score 3	Z-score 4
lnZ (−2)					−0.581***	−0.630***	−0.628***	−0.562***				
					(0.033)	(0.041)	(0.039)	(0.030)				
lnZ (−3)									−0.438***	−0.457***	−0.456***	−0.437***
									(0.038)	(0.042)	(0.041)	(0.034)
ETA	−0.099***											
	(0.028)											
ROA	0.059***		−0.095**									
	(0.013)		(0.034)									
σ_{ROA}	0.286***		0.233***									
	(0.027)		(0.056)									
A_ETA		−0.102*	−0.052**									
		(0.041)	(0.017)									
A_ROA		−0.197***										
		(0.025)										
σ_{A_ROA}		0.380***										
		(0.057)										
TIER 1 RATIO				−0.0006								
				(0.001)								
R_RWA				−0.012*								
				(0.005)								
σ_{R_RWA}				0.0007*								
				(0.0003)								

(*Continued*)

Table 8: (*Continued*)

	(1)				(2)				(3)			
Variables	Z-score 1	Z-score 2	Z-score 3	Z-score 4	Z-score 1	Z-score 2	Z-score 3	Z-score 4	Z-score 1	Z-score 2	Z-score 3	Z-score 4
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. of Obs.	58,125	42,555	58,125	58,125	50,247	50,247	50,247	50,247	42,654	42,654	42,654	42,654
Type Errors:												
TP	432	391	443	442	407	406	404	407	389	388	388	388
FN	120	105	119	120	118	119	121	118	111	112	112	112
FP	19,236	18,596	20,477	20,460	17,256	17,687	17,400	17,201	19,086	19,168	19,110	18,925
TN	38,337	23,463	37,096	37,103	32,466	32,035	32,322	32,521	23,068	22,986	23,044	23,229
Type I	0.217	0.211	0.211	0.213	0.224	0.226	0.230	0.224	0.222	0.224	0.224	0.224
Type II	0.334	0.442	0.355	0.355	0.347	0.355	0.349	0.345	0.452	0.454	0.453	0.448

This table reports the results obtained testing the complementary log-log models over the whole period 2004–2012 for: (1) the components of the natural logarithm of the Z-score (lnZ) and (2) the second lag and (3) the third lag of the Z-score. The different types of Z-score and their components are described in the Appendix. The dependent variable (the defaulted bank dummy variable) that takes the value of 1 if bank i becomes failed at time t (the year in progress) and 0 otherwise. All explanatory variables are lagged by one year. To mitigate the effect of outliers, we winsorize observations in the outside 1% of each tail of each variable. Year dummy variables are also incorporated in the model. These findings were obtained using unconsolidated bank statements. The robust standard errors of the estimated coefficients are reported in parentheses. ***, **, and * denote coefficients statistically different from zero at the 1%, 5%, and 10% levels in two-tailed tests, respectively. This table also displays the relationship between model predictions and actual failure events on the full sample for the whole period, using a cut-off point equals to 0.01. We also tested the regressions using a cut-off point equal to 0.10 rather than 0.01 and we obtained very similar results.TP stands for 'True Positive'; FN stands for 'False Negative'; FP stands for 'False Positive'; TN stands for 'True Negative'. Type I error occurs when the model fails to identify the failed bank and is computed as: FN/(FN+TP). Type II error occurs when a healthy bank is falsely identified as failed (i.e., a false alarm) and is computed as: FP/(FP+TN).

Our variable of interest, lnZ, remains highly significant during the period of 2007–2012. The bottom of Table 7 highlights that during this period, the Z-score can predict bank failures with an accuracy of 81% (see Type I errors). The results for Type II errors also confirm the best predictive power of the Z-score, especially compared to the control variables alone.[8]

We further test whether, and to which extent, the single components of the natural logarithm of the Z-score affect the probability of bank failure (see results (1) of Table 8).[9] To this aim we re-estimate the complementary log-log model on the whole period (2003-2012), but only for our main variables of interest, the Z-scores, given that the contribution of the control variables is only marginal as shown in Table 4. Results (1) of Tables 8 show that, regardless of how the Z-score is computed, all the three components significantly affect the bank probability of failure, with the exception of the Tier 1 ratio being insignificant.

Finally, we check whether Z-score has predictive power two or three years before the failure (see results (2) and (3) of Table 8). Therefore, we test the complementary log-log model firstly on a two-year lag and then on a three-year lag of the natural logarithm of the Z-score. We find in the results (2) and (3), that lnZ is strongly significant both in two and three years before failure with the expected negative sign. These results indicate that Z-score has the ability to predict bank failure even two to three years before the failure events.

V. CONCLUSIONS

Understanding the accuracy of measures of bank soundness that are widely used in the empirical banking literature is an important theme. The numerous bank failures in modern times, especially those during the 2007–09 global financial crisis, highlight the urgency and need of effective, transparent and easy to implement predictors for bank failures.

In this empirical study, we examine the accuracy and the contribution of the Boyd and Graham (1986) Z-score in predicting bank failures, based on three main analyses and several robustness tests. First, we incorporate various versions of Z-score into a complementary log-logistic model to forecast bank failure from 2003 through 2012. We find that Z-score is able to predict bank failures with the accuracy of on average 76%, while adding a set of other bank- and macro-level variables can only marginally increase the model's predictability. Second, we compare the short-term, out-of-sample forecasting ability of Z-score and find that the lowest two deciles of Z-score can predict on average 74% of bank failures. We also examine whether the accounting value based distance-to-default measure Z-score is highly correlated with the market based Merton distance-to-default (DD) measure. We find that Z-score is a significant determinant factor of Merton

[8]Following Barath and Shumway (2008), we also assess the accuracy of our complementary log-log model for the 2007-09 financial crisis time period in an unreported analysis. Our main results hold.
[9]The components of the lnZ are lagged by one year.

DD measure, indicative of high correlation between the two widely used bank risk measures. Furthermore, we find that Z-score alone can predict bank default with three years in advance. Finally, our main results survive the several robustness checks including testing the predicting power of the Z-score for the crisis and post-crisis period (2007–2012) and testing the single components of the natural logarithm of the Z-score affect the probability of bank failures. Based on the consistent and strong empirical evidence documented in this study, we conclude that Z-score is a useful and sufficient predictor for forecasting bank failure.

Our research provides noteworthy contributions to the literature. The obtained empirical results justify the extensive use of this bank risk measure by both academic researchers and practitioners. The advantage of Z-score as a simple measure, and its non-reliance on the publicly traded status of the bank makes it widely applicable to both private and publicly listed banks, and suitable to improve information environment for both retail and institutional investors. In addition, our evidence of establishing Z-score as an effective predictor for bank failure also suggests that accounting quality of banks' earnings and equity is crucial for investors to derive unbiased judgment of bank failure risk. Thus our research calls for further studies aimed to investigate the effects of managerial incentives and various regulations on bank earnings management that could potentially lead to systemically underestimating bank risk.

VI. APPENDIX: VARIABLE DEFINITIONS

This appendix describes the natural logarithm of the Z-score (i.e., our main variable of interest) computed in our paper following the different approaches developed by the literature for its construction and the definition of the control variables used. The table summarizes also their hypothesized relationships with the dependent variable (the failed bank dummy variable).

Variables	Definition	Expected sign
Main variables of interest:		
lnZ (Z-score 1)	The sum of equity to total assets (ETA) and return on average assets (ROA) over the three-year standard deviation of ROA (σ_{ROA}). See Boyd and Graham (1986) and Hannan and Hanweck (1988).	
lnZ (Z-score 2)	The sum of the three-year moving average of equity to total assets (A_ETA) and the three-year moving return of average assets (A_ROA) over the three-year standard deviation of A_ROA ($\sigma_{\mathrm{A\_ROA}}$). See Maecheler *et al.* (2007).	

Variables	Definition	Expected sign
lnZ (Z-score 3)	The sum of the three-year moving average of equity to total assets (A_ETA) and the current values of return on average assets (ROA) over the three-year standard deviation of ROA (σ_{ROA}). See Boyd *et al.* (2006).	NEGATIVE
lnZ (Z-score 4)	The sum of tier 1 ratio (TIER 1 RATIO) and return on risk weighted assets (R_RWA) over the three-year standard deviation of R_RWA (σ_{R_RWA}). See Laeven and Levine (2009) and Dam and Koetter (2012).	
Control variables:		
SIZE	Natural logarithm of total assets (thousands of dollars)	POSITIVE/NEGATIVE
DIV	The ratio of Non-interest income to net operating revenue	NEGATIVE
LIQ	The ratio of the sum of cash, for sale securities and federal funds sold to total assets	NEGATIVE
NPL	The ratio of Non-performing loans to total assets	POSITIVE
CIR	The ratio of Operating expenses to operating income	POSITIVE
BHC dummy	1 if the bank is a member of a BHC; 0 otherwise	NEGATIVE
GDPC	Annual percentage change of gross domestic product	NEGATIVE
INF	Inflation rate (annual percentage change of GDP deflator)	POSITIVE
HHI	Sum of the squared market share value (in term of total assets) of all banks in a year	POSITIVE/NEGATIVE

VII. REFERENCES

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson. 2012. "Measuring Systemic Risk." Centre for Economic Policy Research (CEPR) Discussion Paper, No. 8824.

Adams, R. and H. Mehran. 2003. "Is Corporate Governance Different for Bank Holding Companies?" *Economic Policy Review* 9:123–142.

Akins, B., L. Li, J. Ng, and T.O. Rusticus. 2016. "Bank Competition and Financial Stability: Evidence from the Financial Crisis." *Journal of Financial and Quantitative Analysis* 51:1:1–28.

Allen, F. and D. Gale. 2000. “Financial Contagion.” *Journal of Political Economy* 108:1–33.

Allen, F. and D. Gale. 2004. “Competition and Systemic Stability.” *Journal of Money, Credit and Banking* 36:453–480.

Allen, L. and J. Jagtiani. 2000. “The Risk Effects of Combing Banking, Securities, and Insurance Activities.” *Journal of Economics and Business* 52:485–497.

Altman, E. 1968. “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy.” *Journal of Finance* 22:550–612.

Altman, E. 2000. “Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models.” Unpublished Paper, NYU Stern School of Business.

Altman, E., R. Haldeman, and P. Narayanan. 1977. “Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations.” *Journal of Banking and Finance* 1:29–54.

Balcaen, S. and H. Ooghe. 2006. “35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems.” *British Accounting Review* 38:63–93.

Beck, T. and L. Laeven. 2006. “Resolution of Failed Banks by Deposit Insurers: Cross-Country Evidence.” World Bank Policy Research Working Paper 3920.

Beck, T., O. De Jonghe, and G. Schepens. 2013. “Bank Competition and Stability: Cross-Country Heterogeneity.” *Journal of Financial Intermediation* 22:218–244.

Berger, A.N., S. Bonime, L.G. Goldberg, and L.J. White. 2004. “The Dynamics of Market Entry: They Effects of Mergers and Acquisitions on Entry in The Banking Industry.” *Journal of Business* 77:797–834.

Betz, F., S. Oprica, T.A. Peltonen, and P. Sarlin. 2014. “Predicting Distress in European Banks.”*Journal of Banking and Finance* 45:225–241.

Bharath, S.T. and T. Shumway. 2008. “Forecasting Default With the Merton Distance to Default Model.” *Review of Financial Studies* 21:1339–1369.

Billio, M., M. Getmansky, A.W. Lo, and L. Pelizzon. 2012. “Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors.”*Journal of Financial Economics* 104:535–559.

Boyd, J. and D.E. Runkle. 1993. “Size and Performance of Banking Firms.”*Journal of Monetary Economics* 31:47–67.

Boyd, J. and G. De Nicolò. 2005. “The Theory of Bank Risk Taking and Competition Revisited.” *Journal of Finance* 60:1329–1343.

Boyd, J.H. and S.L. Graham. 1986. “Risk, Regulation, and Bank Holding Company Expansion into Nonbanking.” Quarterly Review – Federal Reserve Bank of Minneapolis 10.

Brown, C.O. and I.S. Dinc. 2011. “Too Many to Fail? Evidence of Regulatory Forbearance When the Banking Sector is Weak.” *Review of Financial Studies* 24:1378–1405.

Carletti, E. 2008. “Competition and Regulation in Banking.” In *Handbook of Financial Intermediation and Banking*, eds., A. Boot and A. Thakor. Elsevier.

Cheng, Q., T. Warfield, and M. Ye. 2011. "Equity Incentives and Earnings Management: Evidence from the Banking Industry." *Journal of Accounting, Auditing and Finance* 26:317–349.

Chiaramonte, L., F. Poli, and M.E. Oriani. 2015. "Are Cooperative Banks a Lever for Promoting Bank Stability? Evidence from the Recent Financial Crisis in OECD Countries." *European Financial Management* 21:491–523.

Dam, L. and M. Kotter. 2012. "Bank Bailouts and Moral Hazard: Evidence from Germany." *Review of Financial Studies* 25:2343–2380.

De Jonghe, O.G. 2010. "Back to the Basics in Banking? A Micro-Analysis of Banking System Stability." *Journal of Financial Intermediation* 19:387–417.

De Nicoló, G. 2000. "Size, Charter Value and Risk in Banking: An International Perspective." EFA 2001 Barcelona Meetings; FRB International Finance Discussion Paper N. 689.

Demirgüç Kunt, A. and H. Huizinga. 2010. "Bank Activity and Funding Strategies: the Impact on Risk and Returns." *Journal of Financial Economics* 98:626–650.

Demyanyk, Y. and I. Hasan. 2010. "Financial Crises and Bank Failures: A Review of Prediction Methods." *Omega* 38:315–324.

DeYoung, R. and K.P. Roland. 2001. "Product Mix and Earnings Volatility at Commercial Banks: Evidence from a Degree of Total Leverage Model." *Journal of Financial Intermediation* 10:54–84.

DeYoung, R. and G. Torna. 2013. "Nontraditional Banking Activities and Bank Failures During the Financial Crisis." *Journal of Financial Intermediation* 22:397–421.

Eckles, D. L., M. Halek, E. He, D.W. Sommer, and R. Zhang. 2011. "Earnings Smoothing, Executive Compensation, and Corporate Governance: Evidence from the Property–Liability Insurance Industry." *Journal of Risk and Insurance* 78:761–790.

Elton E. J., M.J. Gruber, D. Agrawal, and C. Mann. 2001. "Explaining the Rate Spread on Corporate Bonds." *Journal of Finance* 56:247–277.

Erkens, D. H., M. Hung, and P. Matos. 2012. "Corporate Governance in the 2007–2008 Financial Crisis: Evidence from Financial Institutions Worldwide." *Journal of Corporate Finance* 18:389–411.

Gorton, G. and A. Metrick. 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics* 104:425–451.

Gropp, R., J. Vesala, and G. Vulpes. 2002. "Equity and Bond Market Signals as Leading Indicators of Bank Fragility." ECB Working Paper No. 150, Frankfurt.

Guo, G. 1993. "Event-History Analysis for Left-Truncated Data." *Sociological Methodology* 23:217–243.

Hannan, T.H. and G.A. Hanweck. 1988. "Bank Insolvency Risk and the Market for Large Certificates of Deposit." *Journal of Money, Credit and Banking* 20:203–211.

Huizinga, H. and L. Laeven. 2012. "Bank Valuation and Accounting Discretion During a Financial Crisis." *Journal of Financial Economics* 106:614–634.

Ivashina, V. and D. Scharfstein. 2010. "Bank Lending During the Financial Crisis of 2008."*Journal of Financial economics* 97:319–338.

Jin, J. Y., K. Kanagaretnam and G.J. Lobo. 2011. "Ability of Accounting and Audit Quality Variables to Predict Bank Failure During the Financial Crisis." *Journal of Banking and Finance* 35:2811–2819.

Kato, P. and J. Hagendorff. 2010. "Distance to Default, Subordinated Debt, and Distress Indicators in the Banking Industry." *Accounting and Finance* 50:853–870.

Kilic, E., G.J. Lobo, T. Ranasinghe, and K. Sivaramakrishnan. 2012. "The Impact of SFAS 133 on Income Smoothing by Banks Through Loan Loss Provisions." *Accounting Review* 88:233–260.

Kumar, R.P. and V. Ravi. 2007. "Bankruptcy Prediction in Banks and Firms Via Statistical and Intelligent Techniques – A Review." *European Journal of Operational Research* 180:1–28.

Laeven, L. and R. Levine. 2009. "Bank Governance, Regulation and Risk Taking." *Journal of Financial Economics* 93:259–275.

Liu H., P. Molyneux, and J.O.S. Wilson. 2013. "Competition and Stability in European Banking: A Regional Analysis." The Manchester School N. 81:176–201.

Maecheler, A.M., M. Srobona, and W. DeLisle. 2007. "Decomposing Financial Risks and Vulnerabilities in Eastern Europe." International Monetary Fund Working Paper 248.

Männasoo, K. and D.G. Mayes. 2009. "Explaining Bank Distress in Eastern European Transition Economies." *Journal of Banking and Finance* 33: 244–253.

Martinez-Miera, D. and R. Repullo. 2010. "Does Competition Reduce the Risk of Bank Failure?" *Review of Financial Studies* 23:3638–3664.

Merton, R.C. 1974. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates." *Journal of Finance* 29:449–70.

Persons, O. 1999. "Using Financial Information to Differentiate Failed vs. Surviving Finance Companies in Thailand: An Implication for Emerging Economies." *Multinational Finance Journal* 3:127–145.

Poghosyan, T. and M. Čihák. 2011. "Distress in European Banks: An Analysis Based on a New Dataset." *Journal of Financial Services Research* 40: 163–184.

Repullo, R. and J. Suarez. 2013. "The Procyclical Effects of Bank Capital Regulation." *Review of Financial Studies* 26:452–490.

Stiroh, K.J. 2004. "Do Community Banks Benefit from Diversification?" *Journal of Financial Services Research* 25:135–160.

Strobel, F. 2011. "Bank Insolvency Risk and Z-score Measures with Unimodal Returns." *Applied Economic Letters* 18:1683–1685.

Vassalou, M. and Y. Xing. 2004. "Default Risk in Equity Returns." *Journal of Finance* 59:831–868.

VIII. NOTES ON CONTRIBUTORS

Laura Chiaramonte is a Lecturer of Banking at the Università Cattolica del Sacro Cuore, Milan (Italy). Her research interest include the role of bank CDS in the recent financial crisis, the new liquidity rules for banks (Basel III) and the contribution of cooperative banks to financial stability. She recently published on *The European Journal of Finance, Global Finance Journal, British Accounting Review* and *European Financial Management*.

(Frank) Hong Liu is a Senior Lecturer at Adam Smith Business School, University of Glasgow. His research areas focus on banking and security issuance. His recent publication include *Journal of Banking & Finance*, *European Financial Management*, and *International Review of Financial Analysis*.

Federica Poli is an Associate Professor of Banking at the Università Cattolica del Sacro Cuore, Milan (Italy). Her main research areas pertain to bank internationalization, bank organizational models, financial distribution channels and financial innovations. She is author of several publications including published chapters of books and manuals on banking and financial intermediation She recently published on *European Financial Management* and *Global Finance Journal*.

Mingming Zhou is an Assistant Professor at University of Colorado at Colorado Springs, and she has been working in the area of banking, corporate finance, and emerging markets. Her recent publications include *Journal of Banking and Finance, Journal of Investing, Journal of Financial Markets, Institutions, and Instruments, Economic Development Quarterly,* etc.