Omega 38 (2010) 345–357

Contents lists available at ScienceDirect

Omega

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

Assessing bank soundness with classification techniques

Christos Ioannidis [a,*,1], Fotios Pasiouras [b,1], Constantin Zopounidis [b]

[a] Department of Economics, University of Bath, UK
[b] Financial Engineering Laboratory, Department of Production Engineering and Management, Technical University of Crete, Greece

ARTICLE INFO

Article history:
Received 4 March 2009
Accepted 27 October 2009
Processed by B. Lev
Available online 13 November 2009

Keywords:
Bank
Classification
Integration
Soundness

ABSTRACT

The recent crisis highlighted, once again, the importance of early warning models to assess the soundness of individual banks. In the present study, we use six quantitative techniques originating from various disciplines to classify banks in three groups. The first group includes very strong and strong banks; the second one includes adequate banks, while the third group includes banks with weaknesses or serious problems. We compare models developed with financial variables only, with models that incorporate additional information in relation to the regulatory environment, institutional development, and macroeconomic conditions. The accuracy of classification of the models that include only financial variables is rather poor. We observe a substantial improvement in accuracy when we consider the country-level variables, with five out of the six models achieving out-of-sample classification accuracy above 70% on average. The models developed with multi-criteria decision aid and artificial neural networks achieve the highest accuracies. We also explore the development of stacked models that combine the predictions of the individual models at a higher level. While the stacked models outperform the corresponding individual models in most cases, we found no evidence that the best stacked model can outperform the best individual model.

1. Introduction

Bank soundness is a central theme in the agenda of policy makers. After a relatively stable period between the Second World War and the early 1970s, several countries experienced a banking crisis over the last thirty years. Caprio and Klingebiel [1] provide information on 117 systemic banking crises that occurred in 93 countries and 51 borderline and smaller banking crises in 45 countries since the late 1970s. These crises have both direct and indirect costs for the economy. First, as documented in Caprio and Klingebiel [1] the costs for restructuring and recapitalisation can reach 10–20% and occasionally 40–55% of GDP (e.g. Argentina, Indonesia). Second, the crises have adverse effects on the efficient operation of the market economy due to the central role of banks as financial intermediates. Such adverse developments result in reduction in investment and consumption, increases in unemployment, and disturb the flow of credit to individuals and firms causing an overall economic slowdown.

To reduce the likelihood of financial instability several countries have introduced prudential regulation frameworks, making banking one of the most heavily regulated industries. Possibly the most renowned example is the 1988 Basel Accord (i.e. Basel I) that established the capital adequacy requirements and Basel II that introduced additional pillars in relation to supervisory monitoring and market discipline. Furthermore, institutions like the International Monetary Fund and the World Bank have developed and promoted checklists of "best practices" for banking regulation and supervision in an attempt to achieve financial stability and economic development [2,3]. However, the ongoing crisis that started in the US in 2007 revealed that despite these regulatory efforts, crises can still occur and spread rapidly around the world. The recent events generated a new round of discussions regarding the adequacy of the regulatory environment as well as numerous studies that attempt to explain the reasons behind the crises and how they could be avoided in the future.

The recent crisis highlighted, once again, the importance of early warning models to forecast banking crises and assess the soundness of individual banks (See Demyanyk and Hasan [60] for a review of the literature). The first strand of the literature that deals with early warning models examines systemic banking crisis at the country level (e.g. [4,5]). However, there are a number of problems associated with these studies. First, owing to data availability they focus on the 1980s and the 1990s, when we experienced the bulk of banking crises, their results may not be applicable to the modern financial environment. Second, these studies concentrate on emerging market economies due to the higher frequency of crises in these economies in the past [6] whilst the current crisis started from developed countries like the US and the UK. In addition, there are notable differences in the

* Corresponding author.
E-mail addresses: C.Ioannidis@bath.ac.uk (C. Ioannidis), pasiouras@dpem.tuc.gr (F. Pasiouras), kostas@dpem.tuc.gr (C. Zopounidis).

[1] An earlier version of the manuscript was prepared while both C. Ioannidis and F. Pasiouras were affiliated with the University of Bath School of Management.

0305-0483/$ - see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.omega.2009.10.009

dates attributed to the banking crises [6], making their empirical modelling problematic. Finally, dating is also problematic when there are successions of crises episodes as later crises can be extensions or re-emergences of previous financial distress rather than individual events [4,7].

The second strand of the literature focuses on quantitative models that predict individual bank failures (e.g. [8,9]). These studies have the advantage that bank level can provide more rich datasets and additional information compared to aggregate data used in country studies. Nevertheless, a drawback that is also applicable to the country level studies, is that they concentrate on the classification of banks in two groups, failed and non-failed. Obviously, this classification of banks as "bad" or "good" reduces the usefulness of the model.

Given the above, we model bank soundness at the bank level; however, we follow the approach of Gaganis et al. [10] and classify banks in three groups. The first group contains very strong or strong banks; the second one contains adequate banks, while the third group contains banks with weaknesses or serious problems. By focusing on non-failed banks and distinguishing between these three groups the model can be useful in reducing the expected cost of bank failure, either by minimizing the costs to the public or by taking actions to prevent failure. Ravi Kumar and Ravi [11] also mention that "As a bank or firm becomes more and more insolvent, it gradually enters a danger zone. Then, changes to its operations and capital structure must be made in order to keep it solvent" (p. 1). Obviously, the models developed in the present study can be used to monitor changes in the status of banks from one year to another and provide especially an early warning system when a bank gradually deteriorates from the group of strong banks to the one with serious problems.

We differentiate our work from Gaganis et al. [10] and other studies in three important respects. First, we compare, to the best of our knowledge, for the first time the classification accuracy of models that include indicators of the regulatory framework such as restrictions on bank activities and the three pillars of Basel II (i.e. capital requirements, supervisory monitoring, market discipline) with the accuracy of models developed with financial variables only. Second, we compare various advanced techniques such as artificial neural networks, multi-criteria decision aid, classification and regression trees, and nearest neighbours. Third, we investigate the use of a meta-classifier that combines the estimation of the individual models in an integrated model. Applications in other problems in finance such as the default of non-financial firms and approval of credit cards (e.g. [12–14]) have shown that this approach can provide promising results. However, these studies focus on the two-group classification and non-banking institutions. Our problem may be considered as more complex, both due to its three-group dimension as well as the dynamic nature of banking. Thus, the results obtained in past studies are not necessarily applicable to bank soundness, and we aim to examine the effectiveness of this approach in the present study.

The rest of the paper is as follows: Section 2 presents the sample and the variables used in the study, while Section 3 outlines the classification techniques. Section 4 discusses the empirical results, and Section 5 concludes the study.

2. Sample and variables

2.1. Sample

Following Gaganis et al. [10] and Demirguc-Kunt et al. [15] we measure bank soundness using financial strength ratings. As Demirguc-Kunt et al. [15] mention, ratings provide a comprehensive measure of the ability of a bank to meet its obligations to depositors and other creditors and it can be a more accurate indicator of bank soundness than individual measures such as non-performing loans or Z-scores. In principle, this relates our work to a limited number of studies that examine the determinants of bank ratings by international agencies (e.g. [15–18]). However, the aforementioned studies use probit or logistic regression techniques and are of a more explanatory nature. More detailed, they focus on the determinants of ratings rather on the correct out-of-sample classification of the banks. One exception is the study by Pasiouras et al. [19], which attempts to model the ratings of Fitch. However, the authors examine Asian banks only, and they do not consider environmental factors other than an overall index of restrictions in the banking sector.

In the present study, we use the Fitch Individual bank ratings which are based on an A to E scale and represent Fitch's view on the likelihood that the bank would fail, and therefore require support to prevent it from defaulting. As our purpose is not to explain or replicate the ratings of Fitch, but rather to use them as the basis for the development of a general model to assess the soundness of banks, we classify the banks in three broad groups. The first consists of banks with ratings A and B, the second with banks with rating C, and the third with banks rated D and E. Hence, banks in Group 1 can be characterized as very strong or strong banks, banks in Group 2 can be characterized as adequate banks, and those in Group 3 can be characterized as banks with weaknesses or serious problems.

We are not interested in replicating all the ratings of Fitch for two reasons. First, this approach allows us to avoid (at least to some extent) problems associated with the timely adjustment of ratings. For instance, a delay in a downgrade from A to B or from D to E would have no impact in assessing the overall soundness of a bank as we do. Furthermore, small errors of judgment in the assignment of ratings such as rating an A/B or B bank as A would also had no impact on our model. Obviously, large errors of judgment could make a difference but we have no reason to believe that Fitch would classify let us say an E bank as A and *visa versa*. Second, the heterogeneous sample used in our study, consisting of numerous banks from various countries, could have an adverse effect on the classification ability of the model. As discussed in the introduction, the developed model could be useful in several occasions. Furthermore, it could be useful in assessing the overall soundness of banks not rated by Fitch.

Our dataset consists of 944 banks from 78 countries with available data and Fitch individual bank ratings in Bankscope database. The ratings were obtained in end 2008, while the bank specific characteristics correspond to end 2007 or March 2008 depending on the date of publication of the annual reports. The distribution of banks in the three groups is as follows: 447 (Group 1), 275 (Group 2), and 222 (Group 3). To ensure the proper estimation and validation of the models, we randomly select two thirds from each group for training purposes (i.e. a total of 629 banks) and we keep the remaining banks for out-of-sample evaluation (i.e. total of 315 banks). The definitions of Fitch, along with the coding used in the present study and the number of banks in the training and holdout samples appear in Table 1.

2.2. Variables

Credit agencies, researchers, and bank regulators tend to evaluate banks' performance on the basis of the CAMEL model that stands for the acronyms of Capital, Asset quality, Management, Earnings, and Liquidity. We follow the same approach