Contribution of macroeconomic factors to the prediction of small bank failures

Contribution of macroeconomic factors to the prediction of small bank failures

Davide Salvatore Mare[a]

[a] Business School, The University of Edinburgh, U.K.

Abstract

This paper uses bank-level data and macroeconomic indicators to assess the following two questions: (1) Does the external environment influence the occurrence of failure of small banks? (2) To what extent macroeconomic indicators improve bankruptcy prediction? These issues are addressed for the Italian Cooperative Banks using an unbalanced panel model. The results show that local economic environment is a significant determinant of model bank failures.

JEL-Classification: C23, G21, G28

Keywords: financial crises, banking failures, default prediction model, Italian Cooperative Banks, macroeconomic factors

* *Corresponding author: The University of Edinburgh Business School, 29buccleuch Place,EH89JU, Edinburgh, UK, phone: +44 (0)1316501000; fax: +44 (0)1316501000, e-mail: davidemare@hotmail.com*

1 Introduction

The recent financial turmoil has renewed the attention of governments and financial authorities on the endemic risks associated to banking operations. The reluctance of banks to lend to each other together with less liquid financial markets has brought to an increase in the number of failures of financial institutions. As a consequence, it is of utmost importance to further explore the causes behind bank distress.

This paper presents an innovative study related to the analysis of the determinants of failure for small banks. The investigation clarifies the relationship between the specific characteristics of small banks and their probability of failure. The issue is of particular relevance for the socio-economic role of this type of financial institutions and for the relevant output losses generated at local level. Thus, the present analysis develops a default predictive model using the panel data technique.

This study focuses on small banks for two main reasons. First, these credit institutions play an important role at a local and national level. For instance, European cooperative banks serve more than 176 million clients, manage over EUR 5 trillion in assets and hold a deposits and credits market share of 19% and 20% respectively[1]. Second, a few studies provide a comprehensive picture of the main determinants of risk for small banks. In particular, there is a little knowledge about the relationship between their probability of default[2] and the local economic conditions.

From a theoretical point of view, banks are affected by the economy in which they do business. The relevant economic conditions for small banks are local due to the limited geographic market and legal restrictions. Cooperative model - as a distinct business model - exposes the banks to the cyclical fluctuations of the national and local economy (Caratelli et al., 2008). Hence, the research explores whether nonbank economic data can be used to improve forecasts of small bank health.

---

[1] Source of data: European Association of Cooperative Banks (EACB), key statistics as on 31-12-2009.

[2] In this article the term default will, except where otherwise noted, refer to the definition adopted in the estimation of the model (see §3).

The analysis employs the panel data technique to show that cooperative failures prediction is statistically related to macroeconomic variables and bank-level fundamentals. The focus is on Italian cooperative banks (CBs)[3] since this case is particularly interesting to analyze the link between small banks and the local economic environment. Hence the study shows that the inflation rate and the growth in personal income are statistically significant and affects positively the probability of default. As a result, this finding will contribute to the growing literature of explaining the causes of banking distress and it will allow formulating related policies in order to head off or limit direct and indirect costs caused by banking distress.

The remainder of the paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 deals with the method and provides the description of the procedure utilized for the model specification. Section 4 details the data used for the estimation and the sample description. Section 5 reports the results of the analysis and Section 6 illustrate the robustness check. Section 7 concludes and gives the final remarks.

2 Selected literature

Although there are a handful of studies on bank failures and banking crisis, there seems to be two separate streams in the empirical literature[4]: the "micro" and the "macro" camps. The "micro" approach focuses in general on individual banks' balance sheet data, possibly augmented with market data, to predict bank failure. The "macro" approach explains the banking crises examining the macroeconomic determinants. These studies typically analyze a large sample of countries trying to find out which macroeconomic variables signal the happening of a banking crisis in advance. Several works have been done in both areas but a few have tried to combine the two approaches together.

Bankruptcy predictions aiming at detecting individual failures using financial ratios have become important research topics after the pioneeristic research of Altman (1968) and

---

[3] For the sake of simplicity, the term "Italian Cooperative Banks" or "Cooperatives" or "CBs" stands for:

1. Banche di credito cooperativo;
2. Casse rurali;
3. Casse Raiffeisen.

[4] Gonzalez-Hermosillo (1999).

Beaver (1966). These studies try to discriminate between sound and unsound institutions using accounting data. Soon after their introduction, several studies test the predictive ability of financial ratios to detect the financial health of bank operations (among others, Meyer and Pifer, 1970; Sinkey, 1975; Santomero and Visno, 1977; Estrella et al. , 2000). In the same vein many other works aim at testing the superiority of a specific technique with respect to another (Martin, 1977; Espahbodi, 1991; Shumway, 2001; Glennon et al. , 2002; Boyacioglu et al. ,2009). Other researchers focus on the usage of specific variables to predict bank failures. Gonzalez-Hermosillo (1999) analyzes the role of both micro and macro factors in the occurrence of banking system distress in the United States, Mexico and Colombia in the 1980s and 1990s. Using panel data and duration model, the author argues that bank-specific variables seem to capture the fundamental sources of ex-ante risk. Then, the introduction of the macroeconomic or regional variables enhances the predictive power of the models based on bank-specific data only. Furthermore, Ioannidis et al. (2010) find that the use of country-level variable significantly improve the classification accuracies of the models. Also Arena (2008) suggests that systemic macroeconomic and liquidity shocks not only destabilize the banks that were already weak before the crises, but also the relatively stronger banks ex-ante. This result suggests that even strong banks can be particularly affected by the negative effects triggered by systemic crises.

The usage of macroeconomic information in default predicting models is still under debate. Several works have attempted to determine if and under which conditions environmental influences affect banks' likelihood of failure (Daly et al., 2004; Yeager, 2004; Nuxoll et al., 2003). Just a few of those studies have introduced some macroeconomic/regional factors as explanatory variables for individual bank failures (i.e. Nuxoll, 2003; Porath, 2006). Furthermore, studies that look at the banking system as a whole often employ a heterogeneous set of control variables. For instance, Demirguc-Kunt and Detragiache (1998) underline that elements of the macroeconomic environment, such GDP growth, excessively high real interest rates and high inflation, significantly increase the likelihood of systemic banking crises. Also Männasooa and Mayes (2009) test a theoretical framework in which a combination of macroeconomic, structural and bank-specific factors is able to predict banking distress in the European transition countries.

A few studies have assessed the probability of default of the Italian cooperative banks. For instance, Fiordelisi and Mare (2011) investigate the link between efficiency and the probability of bank survival. However, as far as the author is aware, no one has assessed the contribution of the local macroeconomic environment on bank operations. This fact gives room to the present research on the light of the systemic threat represented by the failure of individual banks.

3 The empirical study

Bank distress is often related to a set of different elements that affect bank operations. In the present study, the failure depends on internal (i.e. managerial risks) and external (i.e. economic environment) conditions that trigger the event of the default. Moreover, the focus is on the factors that are not directly under the control of the bank management.

3.1 Definition of failure

In line with previous studies (Arena, 2008; Gonzalez-Hermosillo et al., 1997; Männasooa and Mayes, 2009), bank failure is associated with public intervention. The Italian insolvency regime establishes that major companies (groups) experiencing financial distress, might be subject to both the Extraordinary Administration and the Liquidation Procedure. The former procedure is a going-concern contingent measure that aims at restructuring and reorganizing the enterprise while protecting the company from creditor action. The latter is a gone-concern action in which the insolvent bank has to be shut down.

A bank is classified as being in default if it underwent any of the following procedures between January 1st 1999 and December 31st 2006:

1. it entered the extraordinary administration;

2. it entered liquidation.

In the analysis, a bank is considered to have failed if it is opened one of the two aforementioned procedures. Moreover, for the sake of simplicity, it is assumed that default can only occur at discrete points in time (t = year 1, year 2, etc.). The definition constitutes an objective indication of a bank's inability to continue its operations or a temporary instability. Moreover, it permits to have a reasonable number of observations (i.e. 34) in order to draw the statistical inference from the data.

< Insert Table 1 >

As it is noticed from Table 1, the number of BCCs have been diminishing from 1997 to 2004. This effect is mainly due to the process of concentration witnessed in recent years in the Italian banking sector. In fact, considering the total of number of banks disappeared from 1997 to 2004 (146), 65% have merged or have been acquired by other banks. Moreover, of the 50 cases of default, 28 ended with a merger or acquisition favored by the intervention of the regulatory supervisor (Bank of Italy).

3.2 Methodology

The risk of failure is assessed using an unbalanced panel binary response model. The choice seems the more appropriate because of the nature of bankruptcy data (Shumway, 2001). The specific feature of panel data is the possibility of following the same individuals over time, which facilitates the analysis of dynamic responses and the control of unobserved heterogeneity (Arellano, 2003)[5]. Moreover, this technique permits to utilise bank-specific and macroeconomic variables simultaneously and it helps to forecast future financial condition and to focus on risk categories.

The estimated probabilities are obtained using the traditional random effects model[6]:

$$y_{it} = S_{it} + u_{it} \tag{1}$$

With:

$$S_{it} = \beta_0 + \sum_{j=1}^{p} \beta_j X_{it} + c_i \tag{2}$$

Where S_{it} is the score that constitutes an order of the banks according to their riskiness and u_{it} is the unobserved, individual specific heterogeneity. X_{it} is a vector (p x 1) that can contain a variety of factors, including lagged variables (Wooldridge, 2002). β is the correspondent (1 x p) vector of coefficients and it measures the effect on the probability

[5] Heterogeneity arises since we are dealing with different individuals (banks).

[6] Wooldridge, 2002.

of bank failure of a unit change in the corresponding independent variables. C_i is the (unobserved) heterogeneity.

The estimated response probabilities are given by:

$$P(y_{it} = 1 | x_{it}) = \lambda_{it} = F(S_{it}) \quad (3)$$

where λ_{it} is the probability of default (PD) of bank i at time t. The link function F transforms the score into the PD. This model explicitly assumes that some omitted variables may be constant over time but vary between cases, and others may be fixed between cases but vary over time.

The choice of F(·) determines how the coefficients β_j are estimated. Two different link factions are used (logistic and standard normal). We have then the unobserved effects probit and its logit counterpart. A fixed effects probit analysis is not possible since it leads to incidental parameters problem (Wooldridge, 2002). Then, the traditional random effects model is estimated by using the procedure employed by Butler and Moffitt (1982). The conditional MLE in this context is called the random effects probit estimator (Wooldridge, 2002). The unobserved logit model has an important advantage over the probit model since it is possible to obtain $\sqrt{N}$ -consistent estimator of $\boldsymbol{\beta}$. In this case, using the procedure implemented by Chamberlain (1984), we can estimate the fixed effects logit estimator. However, as already noticed in some studies (Porath, 2006; Davis and Karim, 2008), this procedure cannot be utilized since only defaulted banks would contribute to the log likelihood and excluding non default banks would generate a biased sample and biased coefficients. Therefore, also for the logit model, we obtain the estimated coefficients by using the random effects logit estimator.

Following the ex-post empirical approach employed in previous bank failure studies (Espahbodi, 1991; Glennon et al., 2002; Martin, 1977), the explanatory variables (X_{it}) are drawn from data for a time period prior to failure. The characteristics of the predetermined groups (sound and defaulted banks) are compared considering a time lag permitting to examine the dynamic behaviour. According to the procedure utilized by Gilbert et al. (1999), it is assumed a two year lag (X_{it-2}) as default events are often the result of the balance sheet audit (Porath, 2006). Thus, the resulting function relates the probability of default in period t to the control variables of period t-2. The econometric model then predicts the likelihood that a bank, currently considered safe and sound, enters default in a period between 12 and 24 months. This procedure permits to have at least a one-year forecast horizon and to include in the analysis a minimum number of failures (34)[7]. The model for two years before failure could be used to predict whether a given bank will fail in the future.

A vector of explanatory variables X_{it} (X_{1t},X_{2t},...,X_{pt}) corresponds to each dependent variable (Y_{it}). The database of the independent variables contains the information (macroeconomic and accounting data) for different individuals (banks) at a given point in time across time (different years).

3.3 Determinants of small bank failures

The set of potentially explanatory variables is chosen considering previous similar empirical studies or taking into account the specific characteristics of the cooperative banks. The ratios belong to two broad categories:

1. macroeconomic determinants;
2. bank-level fundamentals.

The first group seeks to gauge the impact of the economic environment on the riskiness of the banks and is the objective of the study. The underlying assumption is that economic variables proxy the increase in the risk of the environment in which Cooperatives operate. Since diversification is not an option due to specific restrictions to Cooperatives' business activities, adverse local economic conditions increase the vulnerability of Cooperatives to local exogenous financial shocks.

Bank specific factors are used to control for the effect of other elements that provide information on the early warning of distress. These ratios, derived from bank accounting information, give forewarning of safety-and-soundness problems. To ensure coverage of the most important aspects of bank vulnerability, CAMEL-type[8] variables are used as

[7] By considering a one-year lag, the number of default decreases to 27; in the case of a three-year lag the figure is 25. The decrease in the number of default is given by the unavailability of the financial statements (especially when considering a one-year lag).

[8] CAMELS is the acronym referring to the following six factors traditionally examined by US baking regulators: "C" stands for Capital adequacy, "A" for Asset quality, "M" for the quality of Management ,"E" refers to Earnings and "L" to Liquidity.

input variables, which are in turn associated to five critical dimensions of banks' operations: liquidity (i.e. liquidity risk), performance (i.e efficiency), risk (i.e credit risk), profitability (i.e. return on equity investment), balance sheet structure (i.e. sources of funding) and capital adequacy (i.e adequacy). This classification is drawn in order to control for bank "internal elements" that affect its probability of default. Furthermore, it permits to give an explicit indication and economic meaning to the numerical output of the model.

The absence of a clear theory on banking financial crisis makes the choice of the input variables arbitrary. In fact, rather than the causes, financial ratios provide information about the symptoms of financial difficulty (Arena, 2008). Nevertheless, accounting data is used as a measure of the potential to failure giving the "tangible" results expressed by the bank financial statements.

The list of the 57 covariates is built from the above mentioned categories[9]. Quality factors are omitted (i.e. management quality) from the analysis since they could not be measured adequately with the available data. From the original list of selected accounting indicators, a subset of variables is chosen using quantitative and qualitative criteria. The goal of this analysis is to select the "most informative" subset of independent variables which has the best "discriminatory ability". In addition, in a multivariate framework, it is desirable that each ratio conveys as much additional information as possible. Finally, a subset selection guarantees some properties of the estimators and permits to draw some inference on the relationship between each category and the probability of default (Espahbodi, 1991). It also enhances the simplicity of the model throughout the parsimony of the parameters.

< Insert Table 2 >

4 Data and sample description

The initial time window considers the period 1997-2009. Year 2005 has been excluded since during 2007 no events of default were observed. Years 2006, 2007, 2008 and 2009 have not been considered in the estimation since from 1/01/2006 CBs have adopted the

[9] The variable names and definitions, along with some descriptive statistics, are available upon request to the author.

International Accounting Standards (IAS). The adoption of IAS changes key financial measures and the value relevance of financial statement information, making the data prior and after the introduction of the new standards not homogeneous. Therefore, the sample set covers the period 1997 through 2004.

The original data set (3,832 observations) has been cleaned and organized to get a more homogeneous sample in order to be able to apply the proposed methodology. Hence, the resulting total number of observations included in the estimation is 3,748.

The sample consists of 604 individuals (banks). The number of observations per group ranged from 1 to 8. In particular, the participation pattern of the cross-sectional time-series data denotes that for each individual in the sample, data are not available for all the years.

<Insert Table 3 >

Looking at the geographical distribution of the CBs, almost half of them are located in a single geographical area (North-West 44%) and almost one-fourth in only one region (Trentino Alto-Adige). The total number of CBs is decreasing across years due mostly to operation of merger and acquisition.

<Insert Table 4 >

In terms of average total assets and number of branches, the Cooperatives in the North-West have a bigger size than CBs located in the other areas, especially compared to CBs in the South. The average total assets show a positive trend across the areas over the years. CBs in the North-East present the highest growth (138%) whilst CBs in the South show lower growth (121%) and lower average number of branches over the period.

< Insert Figure 1 >

The data set for the explanatory variables combines accounting with macroeconomic information. Market information is not considered since CBs are not publicly traded and there is very little other market information available[10]. The financial ratios that are investigated as potential leading indicators of failure are drawn from the banks' financial statements. Data are publicly available for most key items - the liquidity position, balance

[10] A few CBs are provided with credit ratings by credit rating agencies.

sheet, profit and loss, off-balance sheet items, large depositors and large exposures. The major gaps are information on the sectoral pattern of their lending (including exposures to the property sector) and the interest rates on their liabilities and assets. Macroeconomic information is obtained from the Italian National Institute of Statistics (Istat), the Bank of Italy and the Ministry of Interior.

5 Results

The model takes into account the sample unobserved heterogeneity. The purpose is to analyze the relationship between the macroeconomic variables and the probability of default. The coefficients obtained from regressing the dummy variable on macroeconomic and accounting information shed some light on the significance of the relationship between the regressors and the probability of failure. Moreover, the usage of different techniques of estimation guarantees more robust results.

< Insert Table 6 >

The results are pretty similar using the logit and the probit model giving more robust inference.

The signs of the coefficients are the same across models. Both the macroeconomic variables are positively related with the probability of default. The positive sign of the inflation rate denotes that banking crises tend to be high in number when the macroeconomic environment is weak, particularly when the inflation rate is high (Canbas et al., 2005; Demirguc-Kunt and Detragiache, 1998). The relationship between growth in personal income and the event of default is not of immediate perception. Nevertheless, losses incurred in the banking sector con be larger when a downturn is preceded by particular favorable macroeconomic conditions (i.e. excess credit growth[11]). Moreover, if this environmental variable serves as a proxy for the local demand structure (Hahn, 2007), banks which operate in richer region face an external environment which is likely to foster banking efficiency, in that lowering the probability of default. On the other hand, external competition raises costs[12] and a positive sign of the coefficient

[11] The so called Basel III regulatory framework (Basel III: A global regulatory framework for more resilient banks and banking systems,2010) explicitly takes into account specific measures to avoid excess credit growth and procyclicality.

[12] For instance, the cost associated to high-quality personnel (Hahn, 2007).

suggests that an increase in the environmental variable generates an increase in costs that raises the probability of default (Glass and McKillop, 2006).

In all equations, bank-level data that proxy balance-sheet structure, capital adequacy, performance and liquidity are significant at the 5% level. The profitability and risk ratios are not statistically significant in all regressions. This result should be related to the extent of which profitability influences the management of these small banks. The BCC ownership structure together with the mutualistic nature of these banks imply that profits are not the only main target for the management. This relaxes the traditional view that sees shareholders setting financial goals that management must achieve. Also the risk ratio is dropped from the analysis as not statistically significant. This can be attributing to the relationship between the worsening of the real economy conditions and the ability of debtors to repay their debts. An increase in the loan-loss provisions relative to the amount of riskiest loans generates an increase in riskiness, thus the relationship with the event of default is positive. An increase in the profit after tax relative to the total assets indicates an increase in profitability, thus the relationship with the event of default is negative. The capital structure gives an index of the fraction of the total assets that provides interest income. The higher the portion of interest bearing assets, the lower will be the probability of default. A high percentage of regulatory free capital[13] is again associated with low risk. This relative measure indicates that the the larger the capital buffer against losses, the lower the probability of failure. A higher level of staff costs is correlated with a lower level of risk as more qualified employees add value to the banks. The liquidity ratio represents an indication of a company's ability to meet short-term debt obligations; the higher the ratio, the lower is the ability to meet unexpected liquidity needs.

The results of the estimation are in line with previous studies as underlined in the following table.

< Insert Table 7 >

[13] In this case, percentage of regulatory free capital refers to the capital in excess with respect to the regulatory capital.

6 Model performance

The analysis of the goodness-of-fit permits to address how well the statistical model fits the observed phenomena. Several measures are employed for measuring the success rate and the accurateness of the estimation in-sample. In addition, an analysis of robustness is performed to test some general assumptions of the model. Without loss of generality, such measures are calculated for the models in which all the explanatory variables are significantly related with the event of default.

The estimated probability of default is utilized to distinguish failed from non-failed banks. A given bank is classified as failed if its posterior probability of failure is greater than an optimum cut-off point[14]. It is then possible the comparison between this classification and the actual outcome of the event of default. In particular, sensitivity and specificity are the statistical measures employed to evaluate the response of the model. Sensitivity measures the proportion of actual positives which are correctly identified as such (the percentage of observed defaulted banks that are classified in default by the model). Specificity quantifies the proportion of negatives which are correctly identified (the percentage of healthy banks that are identified as being such). The overall accurateness of the estimate is evaluated through the percentage of correctly classified observations[15] and the McFadden's pseudo R-squared[16]. In addition, the ROC area, the Brier score[17] and the Wilks' Lambda[18] are utilized for the assessment of discriminatory power.

---

[14] This threshold was fixed at the level of the given sample prior probability of failure. In general, the optimum cut-off point depends on the prior probability of failure, the decision context of interest and on an appropriate pay-off function (Espahbodi 1991).

[15] Percentage of times the predicted y_{it} matched the actual y_{it}.

[16] The McFadden's pseudo R^2 compares the likelihood for the intercept only model to the likelihood for the model with the predictors:

$$R^2_{McF} = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Interceptl})}$$

where M_{Full} is the model with the predictors, $M_{Intercept}$ is the model with only the intercept and $\hat{L}$ is the estimated likelihood. This measure provides information on the level of improvement over the intercept model offered by the full model. The higher the ratio the higher the total variability explained by the model.

[17] The Brier score evaluates the quality of the forecast of a probability. It is defined as:

< Insert Table 8 >

Since the choice of the cut-off point is arbitrary, it is meaningful to understand how the model classified the defaulted banks two years prior failure by ranking the banks. From the panel regressions we obtain the individual probability of failure of each bank. Banks are then ranked using their fitted probability values from the less risky to the riskiest. The following step consider the actual occurrence of failure to see if the model classify the defaulted banks into one of the five highest probability deciles two years prior failure. The following table reports the results.

< Insert Table 9 >

6.1 Robustness test

The results obtained through the estimation process depend on the data utilized and the method employed. Therefore, in order to get more generalized results, robustness tests are run to check for the predictive capacity of the model.

Out-of-sample data would permit to tackle the data-dependency issue. However, the model is estimated on the whole population therefore no out-of-sample data are available. In order to tackle the problem, rolling window estimation is employed. The method starts with determining the sample window to estimate the probability of default. Since a fair number of defaults must be considered for the estimation , a sample window of seven years is chosen. Then, moving up the window by one year, it is possible to get the risk forecast for the second period. For the ease of exposure, the following

---

$$B = \frac{1}{n}\sum_{j=1}^{n}(p_j - \theta_j)^2$$

where p_0, p_1,..., p_n, are the estimated default probabilities of the banks and θj is the actual outcome of the event of default (it equals 1 if obligor j defaults and 0 otherwise). It follows that the Brier score is always between zero and one. The closer the Brier score is to zero the better is the forecast of default probabilities.

[18] The Wilks' Lambda gives a measure of the success rate of a model:

$$\Lambda = \frac{\sum_{i \in Nd}(z_i - z_{Nd})^2 + \sum_{i \in D}(z_i - z_D)^2}{\sum_{i=1}^{n}(z_i - \bar{z})^2}$$

where $\bar{Z}$ represents the mean of z_i in the entire sample of healthy and defaulted banks. The higher the ratio, the lower the discriminant capacity of the model.

tables present the results per period of the employed technique using only the traditional random effects logit estimator and all the available information.

< Insert Table 10 >

The results confirm that, even considering a different estimation window, the covariates that are significant in the previous regressions remain as such, what brings to more robust results. Moreover, looking at the goodness-of fit measure, the ROC area statistic is fairly high compared with previous estimations. It is also important to notice that the main variable of the analysis (the local environmental indicator) keep being statistically significant under different time windows.

The reason of the methodological choice derives from the available data and from the procedures employed in previous studies. As a consequence, it is interesting to check whether the results are still valid when some aspects of the methodology change. Furthermore, this test provides a useful assessment of the specific results obtained through the model and they offer more support to the inference. As a consequence, it is assumed a different lag in the observation of explanatory variables. By changing the lag it is possible to establish if the regressors are still significant and accurate with a different time to default. Following the general approach employed in similar previous studies[19] ,a one-year lag (t-1) and a three-year lag (t-3) are considered in the test of robustness[20].

For the sake of the argument, it is of utmost importance to remark that by changing the time lag of the relationship between the covariates and the dependent variable, we are changing the forecast horizon.

< Insert Table 11 >

The local economic variable is statistically significant in all the regressions. The variable performs very well considering every time period. The capital adequacy ratio seems to be fairly stable and it is not dropped from any specification either. Nevertheless, the model does not perform very well when the time lag is one year. Only three variables are not dropped from the analysis. There are a number of possible explanations for such

[19] Among the others, Vulpes (1999), Guidi (2005), Canbas et al. (2005).

[20] As noted by Varetto (1999), time lags greater than t-3 (like t-4 and t-5) cast some doubts on the effective degree of realism of the models.

result. One of these is given by the fact that the low number of defaults (27 when we consider a one-year lag) can have some negative effect on the discriminatory power of the variables. In addition, data mining can be a very common phenomenon when the time to default gets closer, in that decreasing the reliability of the accounting information.

Looking at the three years lag estimation, the general macroeconomic variable is also dropped from the model. The significance of the other variables confirms the previous results. The sample default rate is again low (0.007) but possibly the data mining phenomenon does not take place. The local macroeconomic variable keeps being significant denoting a strong relationship with the occurrence of the event of default.

7 Conclusions

The research developed a descriptive analysis of the main drivers of risk for the Italian Cooperative Banks. A panel binary model is proposed since the main goal is to examine the relationship between the available information (macroeconomic and accounting information) and the probability of default across unit (banks) over time. Moreover, macroeconomic information is designed to take into account the specific characteristic of the CBs' operating environment. The model is estimated with a set of banks' default observed over eight years.

The study brings different results. The most important finding is that the local economic environment affects CB probability of default. The variable employed as a proxy of the local economic conditions performs very well both in univariate context[21] and multivariate analysis. Moreover, these results are robust to different timing specifications and a different characterization of the employed method. In addition, the local economic variable synthesise the mechanism related to the boom and bust cycle. Finally, the specific legal and environmental conditions under which CBs operate permit to gauge the heterogeneity of Italian regions through the usage of nonbank data.

The capital adequacy indicator has the most significant relationship with the probability of default. This finding confirms, as already noticed in previous studies, the fact that Italian Cooperative banks are well capitalized financial institution and that this feature

[21] Data are available from the author upon request.

helps in lowering their probability of default. Also the balance-sheet structure, the performance and the liquidity ratios are fairly significant. Nonetheless, these variables present unstable results when they are subject to the test of robustness.

There are some potential drawbacks in terms of interpreting the results. Since the variables do not have the same unit measure, it is difficult to understand which one contributes more to the probability of default. In addition, it is not possible to determine if a significant variable is more useful in identifying failed banks as opposed to non-failed banks (i.e., no information is available about a variable's ability to reduce Type I versus Type II errors). Moreover, the model does not determine which of the variables is "out of line" for a particular bank, which must be discerned in a univariate context by comparison of mean values of variables in failed versus non-failed banks. In addition, the model should be tested out-of-sample to check for its predictive power. Finally, the absence of a clear theory of bank-failure makes arbitrary the choice of the covariates explaining the event of default. Nevertheless, the study represents a step forward in determining the main drivers of risk that affect the probability of default of small banks and in assessing the possible usage of nonbank data to assess the impact of adverse environmental shocks.

In conclusion, further adjustments should be done in order to utilize macroeconomic factors as control variables. The results show that regional macroeconomic time series data contribute to the explanation of the happening of small banks' default. Since the lack of bank failures did not permit to carry on extensive analysis on the issue, the use of other external data sources may be of help to test the predictive ability of the model in holdout samples. In particular, a cross country analysis could support and generalize the hypothesized relationship between the local environment and the risk of default of small banks. Another further development regards the usage of a different technique of analysis. One possible development is the usage of asymmetric distributions (i.e. the Weibull distribution), in order to take into account for the intrinsic features of systemic risk in the banking sector. Another possibility considers that if from one hand CBs' operations are geographically limited by the banking law, on the other hand the protection scheme (Fondo Garanzia Depositanti, Fondo Garanzia Obbligazionisti and Fondo di Garanzia Istituzionale) guarantees that the influence of local market risk is softened by the presence of this "safety net". Then it would be interesting to analyze CB's performance using a portfolio approach and see whether local market risk is still a relevant driver of risk.

Acknowledgement

The author wish to thank Franco Fiordelisi, Maria Paula Gerardino and Andrea Petrella who kindly provided comments on earlier versions of this paper as well as Federcasse for providing the data necessary for the estimation of the model and for the helpful collaboration, especially in the persons of Alessandra Appennini and Roberto Di Salvo. All errors and omissions as usual rest with the author.

8 References

Albert, P. S., Liang, K.-Y. and Zeger, S. L. (1988), ‘‘Models for longitudinal data: a generalized estimating equation approach’’, *Biometrics*, No. 44,pp. 1049–1060.

Altman, E.I. (1968), “Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”, *The Journal of Finance*, No. 23, pp. 589-609

Arellano, M. (2003), “*Panel data econometrics*”, Oxford University Press

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 & Finance*, Vol. 32, Issue 2, pp. 299-310

Beaver, W.H. (1966), “Financial ratios as predictors of failures”, in supplement of *Journal of Accounting Research*, Vol. 4,pp. 71-127

Boyacioglu, M.A., Kara, Y. and Baykan, O.K. (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*, Vol. 36,pp. 1685-1694.

Butler, J. S. and Moffitt, R. A. (1982), ‘‘A computationally efficient quadrature procedure for the one-factor multinomial probit model’’, *Econometrica*, Vol. 50,pp. 761-764

Canbas, S., Cabuk, A. and Kilic S. B. (2005), “Prediction of commercial bank failure via multivariate statistical analysis of financial structures: the Turkish case”, *European Journal of Operational Research*, No 166,pp. 528–546

Caratelli, M., Carretta, A., Fiordelisi, F. and Martelli D. (2008), *“La mia banca è differente. Performance e determinanti del localismo bancario”*, Bancaria Editrice

Chamberlain, G. (1984), *‘‘Panel Data’’*, in Handbook of Econometrics, Volume 2, ed. Z. Griliches and M. D. Intriligator, North Holland, 1247–1318

Crowley, F. D. and Loviscek, A. L. (1990), “New directions in predicting bank failures: the case of small banks”, *North American Review of Economics & Finance*, Vol. 1,pp. 145-162

Daly, M., Krainer, J. and Lopez, J. A. (2004), “Does regional economic performance affect bank conditions? New analysis of an old question”, *Working Paper*, Federal Reserve Bank of San Francisco

Davis E. P., Karim D. (2008), "Comparing early warning systems for banking crises", *Journal of Financial Stability*, Vol. 4, 89–120

Demirguc-Kunt A.,Detragiache E. (1998), "The determinants of banking crises in developing and developed countries", *IMF Stuff papers*

Espahbodi, P. (1991) "Identification of problem banks and binary choice models", *Journal of Banking & Finance*, No. 15, pp. 53-71

Estrella A., Park S., Peristiani S. (2000), "Capital ratios as predictors of bank failure", *Economic Policy Review*, Vol. 6, No 2

European Association of Cooperative Banks, *"Key statistics as on 31-12-2009"*

Fiordelisi F. and Mare D. (2011), "Efficiency and Probability of Default in Cooperative Banking", *Working Paper*, INFINITI Conference on "International Finance", Dublin, 13-14 June 2011

Gilbert R. A., Meyer A. P., Vaughan M. D. (1999), "The role of supervisory screens and econometric models in off-site surveillance", *Federal Reserve Bank of St. Louis Review*, 31–56

Glass J. C., McKillop D. G. (2006), "The impact of differing operating environments on US Credit Union Performance, 1993-2001", *Applied Financial Economics*, Vol. 16, Issue 17, 1285 - 1300

Glennon D., Kolari J., Shin H., Caputo M. (2002), *"Predicting large US commercial bank failures"*, in Journal of Economics and Business, No 54(4), 361-387

Gonzalez-Hermosillo B. (1999), *"Determinants of Ex-Ante Banking System Distress: A Macro-Micro Empirical Exploration of Some Recent Episodes"*, IMF Working Paper No 33

Hahn F.R. (2007), *"Environmental determinants of banking efficiency in Austria"*, in Empirica, No 34, 231-245

Ioannidis C., Pasiouras F., Zopounidis C . (2010), *"Assessing bank soundness with classification techniques"*, in Omega, Vol. 38, Issue 5, 345-357

Logan. A. (2001), *"The United Kingdom's small banks' crisis of the early 1990s: what were the leading indicators of failure?"*, Bank of England Discussion Paper

Männasooa, K. and Mayes, D. G. (2009), “Explaining bank distress in Eastern European transition economies”, *Journal of Banking & Finance*, Vol. 33, Issue 2, pp. 244-253

Martin D. (1977), “Early warning of bank failure”, *The Journal of Banking & Finance*, No 1, 249-276

Meyer, P. A. and Pifer, H. W. (1970), “Prediction of Bank Failures”, *Journal of Finance*, Vol. 25 (4), pp. 853 -868

Nadotti, L. (1995), *“Il bilancio delle banche”*, Il Mulino

Nuxoll D. A. (2003), *“The contribution of economic data to bank-failure models”*, Working Paper

Nuxoll D. A., O'Keefe J., Samolyk K. (2003), *“Do local economic data improve off-site bank-monitoring models?”*, FDIC Banking Review Series, Vol. 15, No 2

Porath D. (2006), *“Estimating probabilities of default for German savings banks and credit cooperatives”*, SBR 58, 214-233

Sahajwala R., Van den Bergh P. (2000), *“Supervisory risk assessments and early warning systems“,* Basel Committee on Banking Supervision working papers, No 4

Santomero A., Vinso J. (1977), *“Estimating the probability of failure for commercial banks and the banking system”*, in Journal of Banking and Finance, Vol. 1, 185 - 206

Shumway T. (2001), *“Forecasting bankruptcy more accurately: a simple hazard model”*, in The Journal of Business Vol. 74, No 1, 101-124

Sinkey J. (1975), *“A multivariate statistical analysis of the characteristics of problem banks”*, in The Journal of Finance No 30, 21-36

Tutino F., Bastianini G., Parascandolo M. (2005), *“La performance delle Banche - Flussi informativi, analisi di gestione, valutazione”*, Bancaria Editrice

Vulpes G. (1999), *“L’impiego di early warning systems per la previsione delle crisi bancarie. Un’applicazione agli indicatori del Fondo Interbancario di Tutela dei Depositi”*, Working Papers, No 3, Fondo Interbancario di Tutela dei Depositi

Wheelock, D.C. and Wilson, P.W. (2000), “Why do banks disappear? The determinants of US bank failures and acquisitions”, *The Review of Economics and Statistics*, Vol. 82, No. 1, pp. 127-138

Wooldridge, J. M. (2002), *“Econometric analysis of cross section and panel data”*, The MIT Press

Wooldridge, J. M. (2008), *“Introductory econometrics: a modern approach”*, 2nd edition

Yeager, T. J. (2004), “The demise of community banks? Local economic shocks are not to blame”, *Journal of Banking and Finance*, Vol. 28, pp 2135–2153

9 Tables

Table 1:"Number of banks and sample default rates"

	SAMPLE			HISTORICAL		
Year	**Number**	**# Default***	**Default rate**	**Number of banks**	**# Default****	**Default rate**
1997	531	4	0.75%	591	8	1.35%
1998	512	5	0.98%	583	8	1.37%
1999	485	4	0.82%	562	8	1.42%
2000	476	6	1.26%	531	6	1.13%
2001	458	7	1.53%	499	5	1.00%
2002	439	2	0.46%	474	6	1.27%
2003	425	4	0.94%	461	7	1.52%
2004	422	2	0.47%	445	2	0.45%
Total	3748	34	0.91%	4.146	50	1.21%

Source: own calculations using data from Federcasse

* The number of defaults refer to the predetermined lag structure (i.e. 1997 failures are banks underwent in default in 1999).
** Banks failed in the correspondent year.

Table 2:"Explanatory variables considered in the specification"

Ratio	Economic meaning	Y=0		Y=1	
		Mean	**St dev**	**Mean**	**St dev**
Profit after tax / Total assets	How profitable is a company relative to its total assets	0.008	0.008	0.006*	0.018
Interest-bearing assets / Total assets	Percentage of the total assets that provides interest income	0.940	0.023	0.919	0.036
Capital in excess of regulatory requirements / Minimum capital requirements	Percentage of capital in excess of regulatory requirements	1.747	1.407	1.286	1.968
Loan-loss provisions/ Non performing loans	Size of losses compared with the amount of riskiest loans	0.102	0.324	0.906	3.790
Staff costs / Number of staff	Overhead factor that each employee carries	55.108	6.931	49.104	11.988
Interest bearing liabilities / Cash and cash equivalents	Ability of the institution to meet projected obligations with the available liquidity	7.737	18.851	17.259	22.122
Inflation rate	Business cycle indicator	2.135	0.402	2.206	0.429
Growth in personal income (regional figures calculated on a two-year basis)	Regional environmental variable	8.216	6.258	13.177	6.438

Source: own calculations using data from Federcasse
* Multiplied by hundred

Table 3:"Pattern of cross-sectional time-series data"

Frequency	%	Cumul. %	Pattern
359	59%	59.44	11111111
38	6%	65.73	11……
35	6%	71.52	111…..
27	4%	75.99	1111….
26	4%	80.3	1…….
20	3%	83.61	11111…
18	3%	86.59	...11111
12	2%	88.58	..111111
10	2%	90.23	111111..
7	1%	91.39	1111
52	9%	100	(other patterns)
604	100%	-	-

Source: own calculations using data from Federcasse

Table 4:"Number of CB per year and geographical area"

Area	1997	1998	1999	2000	2001	2002	2003	2004	Total
Center	87	85	78	87	83	85	81	81	667
North-East	236	228	219	212	206	194	188	186	1669
North-West	78	73	68	64	61	59	58	57	518
South	130	126	120	113	108	101	98	98	894
Total	531	512	485	476	458	439	425	422	3748

Center: Abruzzo, Lazio, Marche, Toscana, Umbria
North-East: Emilia Romagna, Friuli Venezia Giulia, Trentino Alto Adige, Veneto
North-West: Liguria, Lombardia, Piemonte, Valle d'Aosta
South: Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, Sicilia

Table 5:"Number of CB per year and geographical area"

Area	1997	1998	1999	2000	2001	2002	2003	2004	Total
Center	87	85	78	87	83	85	81	81	667
North-East	236	228	219	212	206	194	188	186	1669
North-West	78	73	68	64	61	59	58	57	518
South	130	126	120	113	108	101	98	98	894
Total	531	512	485	476	458	439	425	422	3748

Center: Abruzzo, Lazio, Marche, Toscana, Umbria
North-East: Emilia Romagna, Friuli Venezia Giulia, Trentino Alto Adige, Veneto
North-West: Liguria, Lombardia, Piemonte, Valle d'Aosta
South: Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, Sicilia

Table 6: *"Traditional random effects model"*

VARIABLES	LOGIT		PROBIT	
	Dependent Variable: Event of Default			
	(1.1)	(2.1)	(1.2)	(2.2)
Profit after tax / Total assets	-29.581	-34.051	-17.498	
	(-1.66)	(-1.92)	(-1.72)	
Interest-bearing assets / Total assets	-23.986	-23.366	-11.832	-13.524
	(-3.66)**	(-3.58)**	(-2.78)**	(-3.05)**
Capital in excess of regulatory requirements / Minimum capital requirements	-0.672	-0.691	-0.374	-0.408
	(-3.17)**	(-3.27)**	(-2.51)*	(-2.51)*
Loan-loss provisions/ Non performing loans	0.899		0.551	0.677
	(1.57)		(1.63)	(1.99)*
Staff costs / Number of staff	-0.074	-0.072	-0.040	-0.045
	(-3.21)**	(-3.14)**	(-2.66)**	(-2.83)**
Interest bearing liabilities / Cash and cash equivalents	0.024	0.026	0.015	0.017
	(2.49)*	(2.8)**	(2.43)*	(2.72)**
Inflation rate	0.995	1.008	0.533	0.602
	(2.13)*	(2.17)*	(1.88)	(2.03)*
Growth in personal income (regional figures calculated on a two-year basis)	0.082	0.083	0.053	0.057
	(2.94)**	(3)**	(2.48)*	(2.53)*
Constant	19.093	18.518	8.822	10.218
	(3.05)**	(2.98)**	(2.26)*	(2.52)*
Observations	3748	3748	3748	3748
Number of banks	604	604	604	604

Notes: Absolute value of z statistics in parentheses.* and ** indicates that the individual coefficient is statistically significant at the 5 and 1% level respectively using a two-sided test. The number associated with each column (i.e. 3.1) indicates the specification (i.e. 3) and the model employed (i.e. 1. traditional random effects logit, 2. traditional random effects probit).

Table 7: *"Empirical and expected sign of the relationship between the explanatory variables and the response variable"*

Variables	Source	Exp. Effect	Set of covariates	
			1	2
Profit after tax / Total assets	Logan (2001), Crowley, Loviscek (1990)	-	-	-
Interest-bearing assets / Total assets	Tutino et al (2005)	-	-	-
Capital in excess of reg. requirements / Min capital requirements	Logan (2001)	-	-	-
Loan-loss provisions/ Non performing loans	Nadotti (1995)	+	+	+
Staff costs / Number of staff	Tutino et al (2005)	-	-	-
Interest bearing liabilities / Cash and cash equivalents	Nadotti (1995)	+	+	+
Inflation rate	Porath (2006), Demirguc-Kunt ,Detragiache (1998)	+	+	+
Growth in personal income (regional figures calculated on a two-year basis)	Nuxoll et al (2003), Glass, McKillop (2006), Hahn (2007)	-/+	+	+

Table 8: *"Measures of goodness-of-fit"*

MEASURE	LOGIT MODEL	PROBIT MODEL
	Random Effects	
	1.2	**2.2**
Sensitivity	0.706	0.412
Specificity	0.838	0.957
Correctly classified	0.837	0.952
ROC area	0.874	0.873
McFadden's R^2	0.196	0.209
Brier score*	0.868	0.884
Wilks' lambda	0.962	0.976

Source: own calculations
* Multiplied by hundred

Table 9: *"Fitted accuracy of the estimated probability of default on the actual bankruptcies*"*

RANDOM EFFECTS MODEL		
DECILE**	**LOGIT**	**PROBIT**
1 - 5	5.9	2.9
6	2.9	0.0
7	2.9	11.8
8	11.8	11.8
9	17.6	14.7
10	58.8	58.8

Source: own calculations
* Probability rankings versus actual bankruptcies; percent classified out of 34 possible.
** Deciles of the distribution of estimated probability of default.

Table 10: *"Rolling window estimation using random effects logit regression and all the available information"*

VARIABLES	RANDOM EFFECTS LOGIT	
	Dep. Variable: Event of Default	
	1997 – 2003	1998 - 2004
Profit after tax / Total assets	-28.042	-34.843
	(-1.59)	(-1.79)
Interest-bearing assets / Total assets	-22.520	-24.950
	(-3.37)**	(-3.40)**
Capital in excess of regulatory requirements / Minimum capital requirements	-0.583	-0.639
	(-2.83)**	(-3.06)**
Loan-loss provisions/ Non performing loans	1.052	0.797
	(1.87)	(1.36)
Staff costs / Number of staff	-0.061	-0.067
	(-2.58)**	(-2.64)**
Interest bearing liabilities / Cash and cash equivalents	0.024	0.029
	(2.58)**	(2.96)**
Inflation rate	0.939	0.837
	(2.00)*	(1.67)
Growth in personal income (regional figures calculated on a two-year basis)	0.082	0.070
	(3.02)**	(2.35)*
Constant	17.144	20.088
	(2.71)**	(2.8)**
Observations	3326	3217
Number of banks	601	578
Sensitivity	0.688	0.633
Specificity	0.808	0.839
Correctly classified	0.807	0.837
ROC area	0.882	0.881

Notes: Absolute value of z statistics in parentheses.* and ** indicates that the individual coefficient is statistically significant at the 5 and 1% level respectively using a two-sided test.

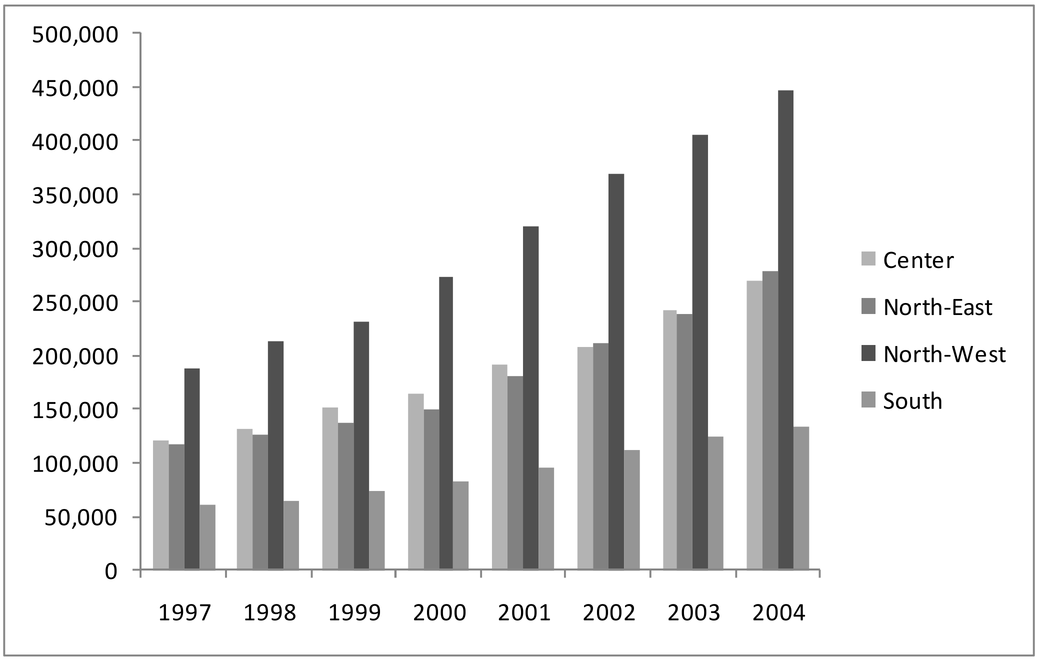
Table 11: *"Estimation results using random effects logit regression and considering a different lag structure"*

VARIABLES	LOGIT		
	Dependent Variable: Event of Default		
	(1 year)	(2 years)	(3 years)
Profit after tax / Total assets	-96.858	-29.994	-47.141
	(-4.06)**	(-1.68)	(-2.26)*
Interest-bearing assets / Total assets	-12.058	-23.987	-18.005
	(-1.46)	(-3.66)**	(-2.21)*
Capital in excess of regulatory requirements / Minimum capital requirements	-0.702	-0.665	-0.673
	(-2.53)*	(-3.15)**	(-2.75)**
Loan-loss provisions/ Non performing loans	0.623	0.908	-1.336
	(0.97)	(1.59)	(-0.99)
Staff costs / Number of staff	-0.038	-0.074	-0.059
	(-1.46)	(-3.19)**	(-2.17)*
Interest bearing liabilities / Cash and cash equivalents	0.008	0.024	0.014
	(0.64)	(2.48)*	(1.08)
Inflation rate	0.815	0.999	0.187
	(1.5)	(2.14)*	(0.36)
Growth in personal income (regional figures calculated on a two-year basis)	0.076	0.082	0.078
	(1.97)*	(2.97)**	(2.56)*
Constant	6.584	19.042	14.745
	(0.82)	(3.05)**	(1.88)
Observations	3775	3748	3714
Number of banks	608	604	596
Sample default rate	0.007	0.009	0.007

Notes: Absolute value of z statistics in parentheses.* and ** indicates that the individual coefficient is statistically significant at the 5 and 1% level respectively using a two-sided test.

10 Figures

Figure 1: *Average total* assets(*) *and number of branches per year and geographical area*



(*) Data in euro thousands