At night all cats are gray, but at day they are not: Default (PD) forecasts capturing Italian banks' idiosyncrasy

Gianpiero Chironna[a], Giuseppe Orlando[1a], Henry Penikas[b]

[a] *Università degli Studi di Bari "Aldo Moro", Department of Economics and Finance, Largo Abbazia S. Scolastica, Bari, I-70124 Italy,*

[b] *Bank of Russia, P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, Russia*

Abstract

Although the probability of default (PD) modeling has reached a great maturity in both academia and business, for the Italian case we demonstrate that banks' available PD models would be misleading if today applied directly to Italian banks. We argue that what determines the PD of Italian banks, rather than the liquidity, are the return on assets (ROA), the financial leverage and the BCC category of the bank. Furthermore, we demonstrate that the conventional approach dominates the more trendy machine learning (ML). Finally, we demonstrate that model's performance could be used as a supervisory tool for retrospective analysis of the bank's position. Moreover, we bring positive evidence on the BCC 2016 reform in Italy.

Keywords: bank failure, adaptive lasso, logistic regression, CART, probability of default, random forest, machine learning, model selection.
JEL Codes: C52, C53, D22, G21, G28, G33.

[1] Corresponding author, giuseppe.orlando@uniba.it.

Electronic copy available at: https://ssrn.com/abstract=4141518

1. Introduction

In this article, we address the probability of default (PD) of Italian banks because the country is of systemic importance not only for Europe (see Heimberger (2020)) but for the world as a member of the G7 and contributing to multiple initiatives to the global scale. These include "Investing in cultural heritage", "Economic growth in Sub-Saharan Africa", "The role of Italy and other donors in the global migration crisis" (see Runde (2018)).

Despite widespread concerns about Italy's prospects and its debt-servicing capacity, Daveri and Verona (2020) recall that the Italian government fulfilled its obligations despite exorbitant market rates and CDS spreads in 2009 during the crisis of the sovereign debt. The basis of Italian resilience is the leading position in some sectors such as fashion, pharmaceutical, mechatronic, food, tourism, etc. The Fortis-Corradini Index (FCI) indicates that "for 932 products Italy was either first, second or third worldwide in terms of foreign trade surplus in 2012". "Furthermore, the FCI reveals, for example, that only three countries (China, Germany and the United States) surpassed Italy in 2012 in terms of the number of first, second and third places in their trade balance worldwide" Fortis et al. (2016). S&P thus confirmed the aforementioned solidity of the Italian banking sector, despite the economic consequences of the pandemic. In fact, "on average, probability of default for financial services companies is below the average for Italian companies in all sectors, which stands at 1.4% and 2.6% over one and two years, respectively" Standard&Poor's (2020).

In describing the Italian economy within the World Economic Forum initiative, Daveri and Verona (2020) underline the large share of small and medium-sized enterprises (SMEs) operating in Italy compared to other European countries. Consequently, many studies on the probability of default (PD) focus on Italian SMEs (e.g. Marotta et al. (2005), Fantazzini and Figini (2009), Provenzano and Arnone (2015), Orlando and Pelosi (2020)). Less studied, conversely, is the segment of large companies in Ciampi and Gordini (2008), Altman et al. (2013), Micheli (2015).

Notwithstanding, there is little literature on the determinants of Italian bank default models. For this reason, the purpose of the present work is to fill this gap and show that, due to the specificities of Italian banks, the PD models currently available do not appear entirely appropriate. This is why we propose a well interpretable model with high predictive power and rapid applicability thanks to its simplicity and transparency.

Before introducing the model and getting to the heart of the analytical treatment, Section 2 provides a reader with a brief review of the literature on PD modeling (in general, and with applications for banks in particular) along with the motivations

Electronic copy available at: https://ssrn.com/abstract=4141518

of the present work. The dataset is described in Section 3. Section 4 lays down the proposed methodology and compares advanced ML tools with conventional approaches. Section 5 summarizes the modeling results and Section 6 concludes with policy implications on supervision and treatment effect estimate of the BCC reform in Italy.

2. Literature Review and Motivations

Probability of default (PD) modeling dates back to the works by Beaver (1966), Altman (1968), Ohlson (1980). They started with the application of discriminant analysis for relatively small datasets. Those were often limited to several dozens of observations. Modern studies (like in Moscatelli et al. (2020)) actively employ machine learning (ML) techniques processing dozens of millions of retail credit data. The interested reader may find a comprehensive review in Qu et al. (2019).

Herein lies a challenge. Machine learning (ML) techniques are gaining popularity at an increasing pace as regulators require banks' PD models to be transparent and reluctant to accept ML models for regulatory purposes as discussed in EBA (2021). A possible compromise solution has been envisaged by the Bank of Spain which suggests using conventional models for regulation and ML for the validation of the first in Alonso and Carbo (2020).

PD models are available for many countries: France - Jabeur and Fahmi (2018), China - Chen et al. (2006), Liu et al. (2022), UK - Altman et al. (2008), Japan - Tian and Yu (2017), EU - Bisogno et al. (2018), and Hungary - Kristóf and Virág (2020).

Aside from its banking heritage which dates back to the Middle Ages, Italy is known for its expertise in default modeling. For example, before launching the second Basel agreement (Basel II), the Basel Committee studied the existing results in PD modeling. In his report on the millennium (BCBS, 2000, p. 110), the Committee cites the study of the Italian Credit Register (CSFD) dating back to 1997. We could not find the study as the reference does not have enough details. Instead, we found an equivalent document of Laviola and Trapanese (1997). However, we suspect that these are two different documents as they present a different number of observations: 1885 in the Italian CSFD study from BCBS (2000) and 1274 in Laviola and Trapanese (1997).

That said, PD models for banks are scarce. Relevant reviews are available at Kumar and Ravi (2007), Citterio (2020). This is because the banks refer to a so-called *low default portfolio (LDP)*, see Penikas (2020). It means that the default rate - the proportion of the defaulting class within the whole set of observations - is

Electronic copy available at: https://ssrn.com/abstract=4141518

almost negligible. However, as we recall in the Lehman Brothers case, such a single default could have triggered a bank run on the entire US banking system and their complete withdrawal within two days, according to P. Krugman's statement. This was the reason for more than double the increase in the state deposit insurance limit of USD 100k to USD 250k within two days after the Lehman accident. This example highlights that a negligible fraction of bank insolvencies could be very important for national and sometimes even global financial stability. This raises the importance of continuing to offer another PD model for a specific segment of borrowers - for banks.

Among the studies focusing on bank defaults and distress, we mention Bräuning et al. (2020), Durand et al. (2021) for the EU, Yuksel et al. (2015) for Turkey, Shrivastava et al. (2020) for India, Kočenda and Iwasaki (2022) for Japan, Cole et al. (2020) for the USA, Obeid (2021) for the Persian Gulf region, and Cheong and Ramasamy (2019), Kristóf (2021) for some others.

Notable contributions to banking PD modeling relate to the RiskCalc model developed by Moody's. USA banks are covered in the document by Kocagil et al. (2002), while an equivalent model for banks globally was made available only a decade later in Moody's Analytics (2016).

Interestingly enough, Russian bank data are attracting attention as a source for the development of PD models, disregarding the fact that the Russian banking system is not the most developed one in the world. Examples are scholars from Finland: Fungáčová and Solanko (2009), Fungáčová et al. (2010), Fungáčová and Weill (2013), Fungáčová et al. (2020), The Netherlands: Karas and Vernikov (2019), Switzerland: Goncharenko et al. (2022) and, finally, Russia: Shibitov and Mamedli (2019). This is due to the fact that The Central Bank of the Russian Federation publishes detailed monthly financial statements of local credit institutions. Such disclosures of the bank financial statements also exist for the American banks, though those seem to be more difficult to process:

- For existing banks Federal Financial Institutions Examination Council (2022), though earlier it was available at Federal Reserve Bank of Chicago (2022);
- For failed ones Federal Deposit Insurance Corporation (2022).

On the contrary, surveys of Italian banks in Italy are scarce. A working paper by Ferriani et al. (2019) is a rare exception, but it suffers from some limitations. First, it uses non-public (supervisory) data to assign a default (financial distress) flag. Second, it is limited to less significant banks, thus not capturing the largest banks. The research novelty thus is the introduction of co-operative banks (BCC) dummy by Ferriani et al. (2019). Italy has undertaken a special reform six years ago

Electronic copy available at: https://ssrn.com/abstract=4141518

by merging all co-operative banks under a single entity for supervision, according to EACB (2016). The 367 Italian mutual banks, in fact, constitute an SSM significant banking group, representing the third largest Italian banking institution and the first in terms of capital.

The question that arises at this point is whether there is a need to develop a new PD model or whether something already available can be used, such as the well-established Moody's Analytics (2016) model developed from a dataset of 90 countries. In anticipation of the results obtained, we show that Italy has peculiar characteristics that are not adequately covered neither by the aforementioned Moody's model, nor by the nearest benchmarks from Ferriani et al. (2019), Penikas (2022).

3. Data Analysis and Preparation

3.1. Definition of Default

In this section, we aim to predict the state of *one year before default* for Italian banks starting from the default flag assignment (which is the cornerstone of any PD model development). To do this, we first take the standard approach and then compare the current year's default with the previous year's financial data. The financial data for the current year undoubtedly reflects the deterioration in creditworthiness in the event of default this year. This is because financial results at a given moment are useful to test the goodness of the forecast. The interested reader can find a graphic illustration of the process in (Burova et al., 2021, p. 54, Figure 2).

3.2. Classification of Outliers

Data cleansing is the first step. Since we take the logarithms of the variations, the first operation is to remove observations with zero value.

Next, we pay close attention to the best candidates for the PD drivers. From Figure 1, visual inspection indicates that outliers are numerous especially concerning the profitability ratio of the ROA and the change of the ROA. However, these values are linked to the specificity of Italian banks. Therefore, even if it is normal practice to eliminate outliers, in this case, some values correspond to banks in the process of default and, therefore, must be preserved.

After having specified the above, for the definition of outlier we adopt the criteria reported in Table 1. Since some banks have more than one default flag, the total number of outliers is 14 (instead of the 21 resulting from Table 1) so that Figure 2 shows the distribution of data after outliers' removal.

Electronic copy available at: https://ssrn.com/abstract=4141518

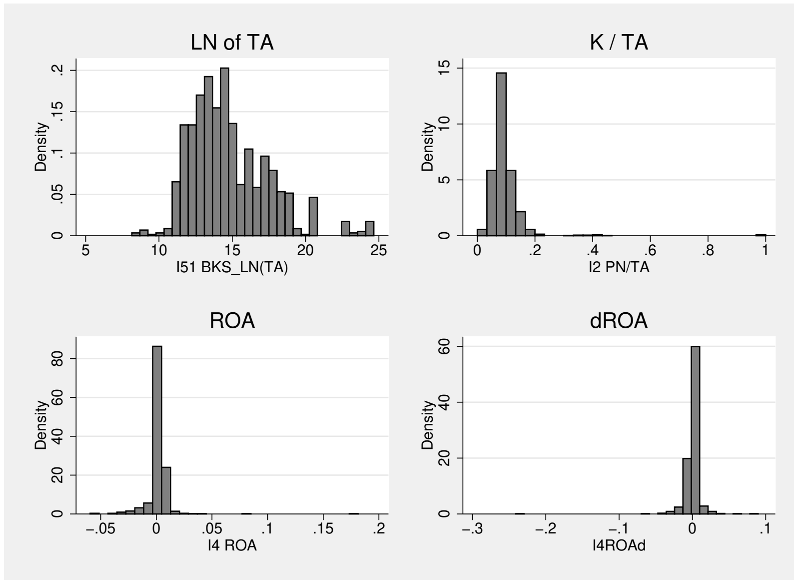

Figure 1: There is need to filter for outliers.

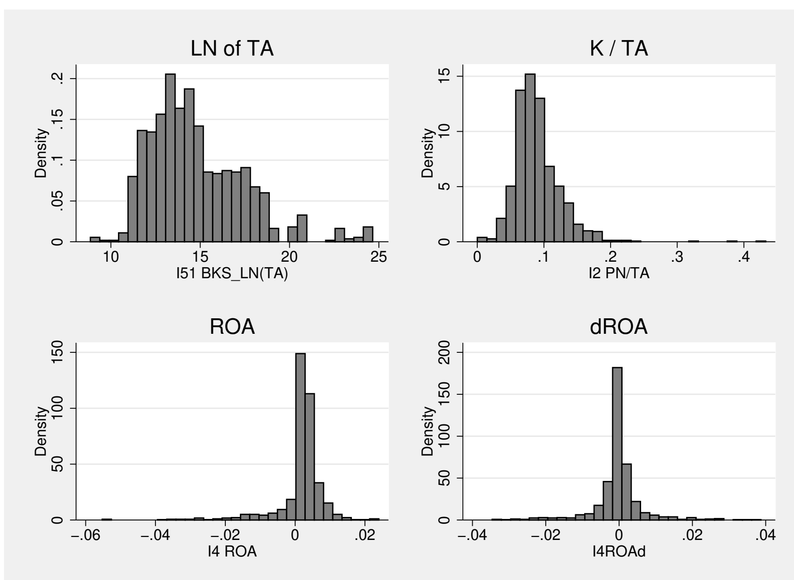

Figure 2: After filtering for outliers the distributions approaches bell-shaped (though not Gaussian) form.

Electronic copy available at: https://ssrn.com/abstract=4141518

Table 1: Classification of Outliers

Criteria	# of resulting outliers
1. ROA (I4ROA) too low (below -6% p.a.) or too high (above +6% p.a.)	2
2. ROE (I1ROE) too low (below -100% p.a.)	2
3. Operating profit to brokerage margin (I43RGMID) too low (below -2) or too high (above +2)	5
4. Exorbitant total capital ratio (I11TCR) above 90% to capture the risk of assets	5
5. Exorbitant leverage ratio (I2PNTA) above 90% to capture (non adjusted) risk	3
6. Extreme change in ROA (more than 5 pp. p.a. in absolute terms)	4
Total	21

3.3. Analysis of Data

The default rate for the sample considered is 2.6 % per annum. This corresponds to the so-called low default portfolio (LDP). Notice that Figure 3 (right) hints a bimodal distribution of the default rate. This implies a significant default correlation for the LDP segment. Furthermore, observing the growth of Italian GDP, we can hypothesize that a lag of 4 years may have importance for the PD of Italian banks. However, as shown in Figure 4 - such a trend is not quite pronounced. Adding such a lag in GDP growth as a PD driver does not produce a statistically significant coefficient.

Furthermore, it is worth recalling that there is an interrelation between bank insolvencies and the current economic situation which is reflected in the GDP growth numbers. Economic downturns affect both bank balance sheets and GDP. There is a loop mechanism whereby the economic crisis triggers bank insolvencies, while bank insolvencies limit the available credit financing needed for a rapid recovery. To avoid a long and often unsuccessful search for adequate tools to counter the problem of the endogeneity of GDP-Bank failures, we refrain from adding GDP to the model.

A quick look at conventional PD drivers in Figure 5 suggests that the Italian bank is more stable if (a) it is negligibly smaller in size (the 'too big to fail (TBTF)' issue); (b) the more capital the Italian bank has; (c) the more profitable it is; (d) the

Electronic copy available at: https://ssrn.com/abstract=4141518

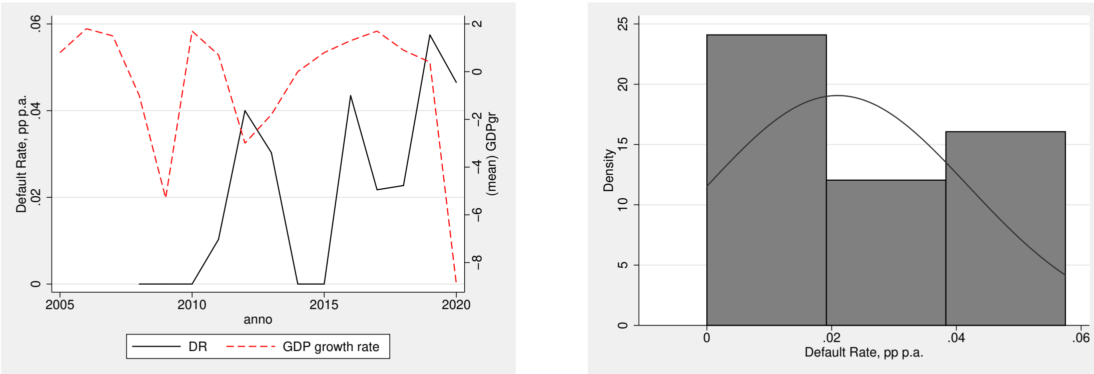

Figure 3: The world financial crisis of 2007-09 triggered defaults among Italian banks (left). The bimodal DR distribution signals for the presence of positive default correlation (right).
Note: DR - the default rate (pp. per annum, p.a.).

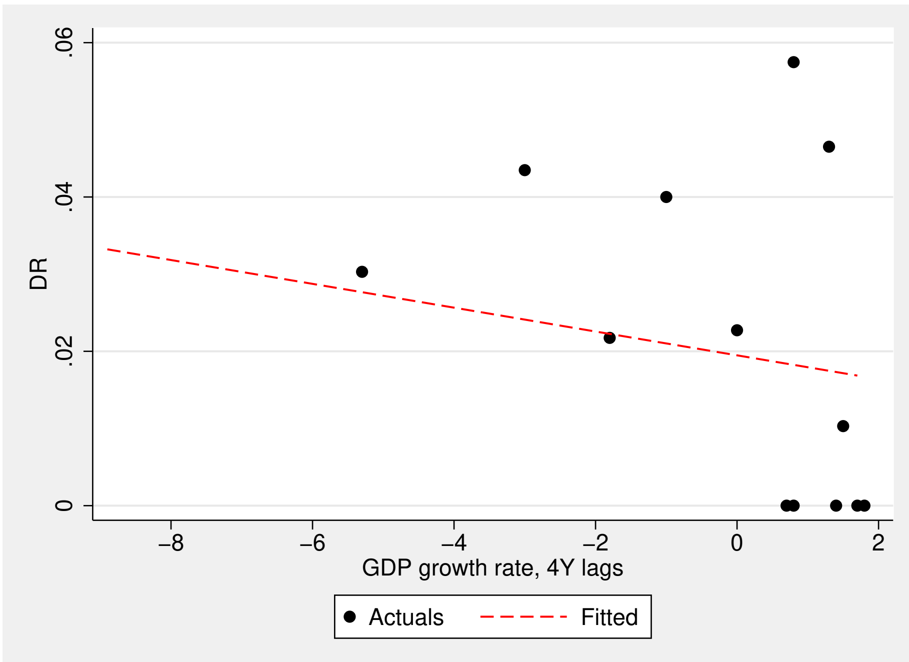

Figure 4: Higher pace of economic growth is associated with lower bank defaults in Italy, though the path is not quite robust.
Note: DR - the default rate (pp. per annum, p.a.).

Electronic copy available at: https://ssrn.com/abstract=4141518

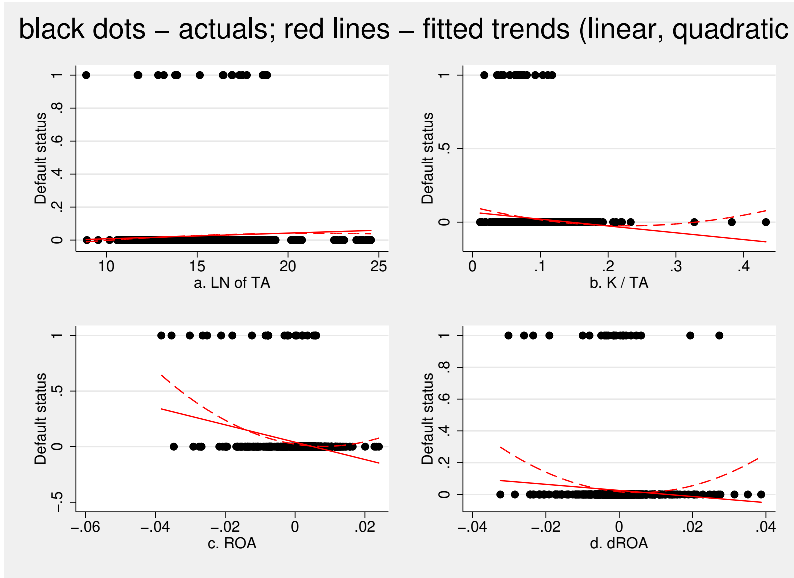

Figure 5: Higher capital and higher profitability is associated with low PD of Italian banks.
Note: black dots stand for actual observations; solid red line is the linear fit, while the red dashed line is the parabolic trend.

larger its profit grows. These patterns correspond to the stylized facts on banking sustainability.

We have checked whether there are any quadratic patterns, but the parabolic trends mostly overlap with the linear ones in Figure 5. That is why we did not include squared factor values in our PD model.

4. Methodology

4.1. PD Model Development and Validation

4.1.1. Machine Learning (ML)

After having analyzed and prepared the data, the next step is to model them through the use of some widespread ML and conventional tools such as a) LASSO and ridge regression to search for the most impacting PD drivers, b) logistic regression, c) classification and regression tree (CART), d) random forest.

For the ensuing analysis, we consider two subsets:

- DS(B100) made of both types of financial statements for 100 banks: consolidated and non-consolidated. This subset of data contains 21 default cases.

Electronic copy available at: https://ssrn.com/abstract=4141518

- DS(B73) based on non-consolidated reports for 73 banks and containing 11 defaults.

It should be noted that the larger sample of DS (B100) is no less significant from the point of view of the development of the PD model. Double counting due to the use of consolidated and non-consolidated financial statements has been appropriately avoided. Consolidated reports are considered only for those banks for which we do not have unconsolidated financial statements.

ML Fitting and Sampling. Regarding fitting and data sampling, the ML models aforementioned have been used to fit over a) imbalanced data (i.e., data as is); b) oversampled data (i.e., data with added historical defaults), c) over-under sampling (where added copies of smaller class and the same time reduce the non-defaulted observations) d) random over-sampling (ROSE), proposed in Menardi and Torelli (2014) where the new artificial observations are generated through the estimation of a kernel density function. The illustrative comparison of the mentioned four approaches to sampling formation is available in Table 2.

Robustness Check. To ensure the robustness of the results, each method is used by performing a 10-fold cross-validation using the common validation sample as presented in row 'Validation' of Table 2. This means that the algorithm randomly chooses 10 % of a sample and treats it as a test set, while the remaining 90 % of a sample is treated as a training set. Then further extractions of non-overlapping data from the 10 % volume are performed and the model is run again. The results, in terms of goodness-of-fit (classification) metrics, of these ten predictions, are then averaged.

To feel the data, Table 2 offers the breakdown of samples used for training and validation. In gross total, we have 1059 observations, of which there are 14 outliers as indicated in subsection 3.2 and we arrive at 1045 observations presented in Table .5. By subtracting the one-year lag for 100 banks - i.e. 100 observations - we arrive at 946 observations for the one-year shifted Default flag as demonstrated in Table .5.

Classification and Precision Metrics. For the benchmarking of the results obtained from the models, we apply a series of classification accuracy metrics such as: a) accuracy, (i.e. the proportion of real detection in a class); b) sensitivity (i.e. the rate of true positive predictions (TPR)), c) p-value for testing the null of accuracy exceeding the No-Information Rate (NIR) from (Kuhn, 2008, p. 15), d) area under the Receiver Operating Curve (AUC, AUROC), e) kappa of Cohen (1960), f) J of (Youden, 1950, p. 33).

Electronic copy available at: https://ssrn.com/abstract=4141518

Table 2: Benchmarking Sample Formation Approaches for Training and Validation

Set Type	Approach	ND	D	Total = ND+D	D, % of Total
Training	a) as is	780	14	794	1.76
	b) over	780	120	900	13.33
	c) over-under	723	177	900	19.66
	d) ROSE	406	388	794	48.87
Validation		258	7	265	2.64
Training + Validation	as is	1038	21	1059	1.98

Note: ND - non-defaulted observations; D - defaulted observations; Validation - the same number of observations was used for testing independent of the sample composition during training (that is why we present four rows for Training and a single row for Validation); the paragraph 'ML Fitting and Sampling' from the above Subsection 4.1.1 contains the description of the four approaches a)-d) used to form the initial training sets.

4.1.2. Conventional Approaches

As an alternative to the ML approach, we propose a conventional approach to PD model development. To do this, let us start with a single factor analysis. There are 72 financial indicators at our disposal whose characteristic descriptions are available in Table .5.

For further analysis, we take into account the variables that are statistically significantly correlated with the default flag at the significance level of 5 % (marked with a single star in Table .5). Next, we trace whether the chosen separate PD drivers could be related thus causing multicollinearity. Therefore, if we observe a correlation between X factors greater than 50 % we keep the indicator which is the simplest. For example, if the return on assets (ROA) is related to some cash flow-based indicator, we keep the ROA. This also ensures the comparability of the model with the previously mentioned studies in which we used the ROA. Third, we perform the multifactorial analysis using the *Probit* model specification.

In terms of conventional approaches, we consider eight different models as described in Table 3. The results of the performed analyses are reported in Table .12.

Following Diebolt (2015) and in contrast to the ML approach of model validation, we estimate model parameters using the entire (full) data sample without any decomposition into training and testing parts. Diebolt argues that in retrospect one could always strategically fit the best model on the subset of out-of-sample (out-of-time) historical data. Furthermore, the very Bayesian updating principle incorporated in

Electronic copy available at: https://ssrn.com/abstract=4141518

Table 3: Model specifications

Model	Description
Pr01	From the entire set of variables excludes the statistically insignificant drivers in a step-wise fashion
Pr02	Starts with an empty set and step-wise new statistically significant variables are added
Pr03	Includes all variables from single-factor analysis and after exclusion of the collinear drivers
Pr04	Excludes statistically insignificant variables from Pr03 specification
Pr05	From Pr04 interactions with small bank dummy indicator (BCC) are excluded
Pr06	Specification mostly one-to-one as suggested by Moody's Analytics (2016) for comparison/benchmarking where: Return on assets (ROA) enters directly as I4ROA; Equity to assets is also available as I2PNTA; Non-performing assets is proxied by I25RRV_130TA; Change in ROA is I4ROAd; Provision to loans is proxied by I14RRV_CCVC; Loan-to-Deposit ratio comes from I17CVCDVC.
Pr07	Improvement version of Pr05 in which significant variables are taken from Pr06
Pr08	Includes all variables from specifications Pr01-Pr07

ML tools assumes a full sample estimate, as Diebolt (2015) shows. Therefore, we choose the best model according to the statistical significance of the coefficient estimates and their stability across specifications. As we will show, the departure from this stability is a consequence of the multicollinearity which does not allow for correctly interpreting the separate coefficient, as it happens instead in Moody's Analytics (2016).

Electronic copy available at: https://ssrn.com/abstract=4141518

4.2. Defining Factor Weights and Contributions

At this point, we proceed to explicitly compare the results obtained with those previously available to show the novelty contribution of this study.

Most of the ML PD models are mostly "black boxes". For example, model properties (such as discriminatory power etc.) are reported, but the importance of the factor is not available, as in Shibitov and Mamedli (2019). However, there is one notable exception (see (Dendramis et al., 2020, p. 39, Table 5)) when, in the case of a PD model for the SME segment, the authors report factorial sensitivities (the so-called 'covariate effects', which are equivalent to marginal effects).

Aside from the exception mentioned above, "white box" PD models in which estimated coefficients, marginal effects and the like can be seen are rarely disclosed. This is especially true of banks. Possible candidates for comparison are the Moody's models for the USA in Kocagil et al. (2002), the world banks in Moody's Analytics (2016), the Italian banks in Ferriani et al. (2019), and for the Russian ones Penikas (2022). Although these works describe the importance of the factor, they are difficult to compare with each other. Kocagil et al. (2002) presents some factor contributions to the PD model which are always non-negative and in total amount to 100 %, while Moody's Analytics (2016) shows the weights of the model which can assume negative values and in total do not add up to 100 %. Furthermore, neither Kocagil et al. (2002), nor Moody's Analytics (2016) establish the operational implementation on how they arrived at the contributions and factor weights respectively.

Since the articles of Ferriani et al. (2019), Penikas (2022) have no commercial sales targets unlike the papers by Moody's, they both contain indications on the coefficients. The dilemma concerns the form of the benchmarking representation of our PD model for Italian banks with the four 'white box' models mentioned.

To undertake an adequate comparative analysis, we suggest the possible way to arrive at the contributions and factor weights to make our model comparable to Kocagil et al. (2002), Moody's Analytics (2016). First, we explain our understanding of contributions and weights. Since the weights can vary in sign, they should correspond to the estimated coefficients and/or marginal effects. Recall that the weights of the model in Moody's Analytics (2016) do not reach 100 % (their sum is equal to approximately 30 % for the one-year PD model and -8 % for the five-year PD model). However, it is interesting to note that the absolute values of those weights add up precisely to 100 %.

So the question is how to define factor weights after estimating coefficients from nonlinear models such as logit in Ferriani et al. (2019) or probit in Penikas (2022).

We suppose that we have $\hat{\beta}_{ik}$ as an estimated coefficient on i-th PD driver that forms part of k-th PD factor.

Electronic copy available at: https://ssrn.com/abstract=4141518

Let us denote W_j a weight of the j-th PD factor (it may include multiple PD drivers as separate i variables), and C_j as a contribution of the j-th PD factor. We impose that contributions must sum up to 100%, while we let weights reflect the direction (sign) of impact.

We depart from the *probit* functional form. Equations (1) and (2) tell us that the portfolio average PD forecast ($\hat{PD}$) can be proxied as the Gaussian CDF N value at the point $\hat{y}$ which equals to the linear combination of mean driver values $\bar{x}_{ik}$ multiplied by the estimated coefficients $\hat{\beta}_{ik}$ plus the intercept $\hat{\beta}_0$.

$$\hat{PD} \approx N\big(\hat{y}\big), \tag{1}$$

where

$$\hat{y} = \hat{\beta}_0 + \sum_{k=1}^{J} \sum_{i=1}^{I} \hat{\beta}_{ik} \cdot \bar{x}_{ik}, \tag{2}$$

where $\bar{x}_{ik}$ is the sample mean of the i-th variable falling into the k-th PD factor (we compute the marginal effects $\frac{dF}{dx_{ik}}$ exactly at the point of $\bar{x}_{ik}$); β_0 is the intercept.

Then we may compute the weights and contributions in Eq. (3), (4), respectively. In essence, we are decomposing the value of the latent factor $\hat{y}$ and not the average PD prediction ($\hat{PD}$). Here might be a difference in the marginal effect of a particular factor due to the non-linear transformation from $\hat{y}$ to $\hat{PD}$ in probit (likewise it is also non-linear in logit, as well in the designed ML models). However, given the undisclosed Moody's approach to computing factor weights and contributions, the described approach seems to be the simplest feasible and transparent solution.

$$W_j = \frac{\sum_{i=1}^{i=I} (\hat{\beta}_{ik} \cdot \bar{x}_{ik})|_{k=j}}{\sum_{k=1}^{k=J} \big(\sum_{i=1}^{i=I} (abs(\hat{\beta}_{ik} \cdot \bar{x}_{ik}|_{k=j}) \big)}. \tag{3}$$

$$C_j = \frac{\sum_{i=1}^{i=I} (abs(\hat{\beta}_{ik} \cdot \bar{x}_{ik})|_{k=j})}{\sum_{k=1}^{k=J} \big(\sum_{i=1}^{i=I} (abs(\hat{\beta}_{ik} \cdot \bar{x}_{ik}|_{k=j}) \big)}. \tag{4}$$

As we can see from Eq. (4), the denominator for the contribution of the PD factor (as well as for its weight in Eq. (3)) may differ from the value of $\hat{y}$ in the case at least a mean driver value can have a negative value, or a coefficient estimate can take a negative sign. We support the proposed weight calculation method to avoid the problem of compensation factor contributions (self-compensating). Suppose there is a simplistic PD model consisting of two factors. Let them have equal sample means of 1 and estimated coefficients of +1 and -1. Consequently, the point estimate for the overall PD factor will be zero ($\hat{y} = (1 \cdot +1) + (1 \cdot -1) = 0$) and the PD forecast at 50 % assuming Gaussian CDF. Hence each factor's contribution cannot be defined.

Electronic copy available at: https://ssrn.com/abstract=4141518

For example, when dividing a coefficient estimate by zero-sum of $\hat{y}$ we get values of infinity. It is not an interpretable outcome. However, our suggested approach gives factor contributions of 50 % and 50 %, while the factor weights according to Eq. (3) are +50 % and -50 %.

The nice property of the illustrated approach in Eq. (3), (4) is the comparability with the previous studies. However, we are bounded in part by the work by Ferriani et al. (2019) as they report the estimated coefficients $\hat{\beta}_{ik}$ and do not report the corresponding sample means $\bar{x}_{ik}$. We will assume $\bar{x}_{ik} = 1$ when otherwise is not stated. The advantage of the Ferriani et al. (2019) paper is the model calibration for 4 specifications for the PD prediction horizons of one year (4 quarters) and one year and a half (6 quarters). Unfortunately, the authors in Ferriani et al. (2019) do not explain which of 4 specifications per horizon should be considered as the best to proceed with. That is why we take the mean coefficient estimates by 4 specifications and focus on the one-year prediction horizon for comparability with the US banks in Kocagil et al. (2002), with the world banks in Moody's Analytics (2016) and the Russian banks in Penikas (2022).

5. Findings

5.1. Machine Learning

The LASSO method suggests that there are three most material PD drivers:

- Net losses/recoveries on impairment (ex IAS 39 Item 130) / Total Assets (I25RRV_130TA);
- Net value adjustments/write-backs on property, plant and equipment (IS.I.210) / Total Assets (I26RRV_AMTA);
- Net interest margin (IS.I.30) / Total Assets (I33MITA).

The problem with the chosen drivers is that they are co-dependent on construction since they use total assets as the denominator. Furthermore, the indicators are not comparable with previous studies. It may be that the Italian banks have their idiosyncrasies. However, according to the visual analysis from Figure 5, we saw the importance of profitability and the capital base as the PD driver of Italian banks.

To conclude, using the small sample of 73 banks without consolidated reporting [DS (B73)] we could not find an optimal model.

In contrast, the larger dataset of 100 banks favours the conventional logistics model. The best performance refers to the ROSE algorithm. It yields Youden's J = 0.521 and AUC = 0.761. The next best model is also the logistic regression using the

Electronic copy available at: https://ssrn.com/abstract=4141518

original data (unbalanced approach) with Youden's J = 0.4247 and AUC = 0.712, see details in Tables .6-.11.

Therefore, we have shown that the conventional model outperforms the random forest and CART. To confirm this, we perform a robustness check using the probit model in the conventional configuration.

5.2. Conventional Approach

Table .12 summarizes the coefficient estimates for various specifications. First, we note that similar to Ferriani et al. (2019) the dummy for small cooperative banks (BCCs) is statistically significant, even if we find a negative coefficient, while Ferriani et al. (2019) reports a positive one. This could be an expected discrepancy as Ferriani et al. (2019) paper suggests focusing more on the less significant banks. By contrast, our dataset additionally contains the largest banks in Italy. Furthermore, it is also quite likely that the 2016 reform of cooperative banks had a positive impact on the sustainability of cooperative banks in Italy as the BCC dummy had a positive sign until 2016 in Ferriani et al. (2019) (although it is negative for our case until 2020 presumably due to some delay effect or the positive BCC reform impact that we are going to elaborate on further in subsection 6.2).

Second, we find that half of the variables used in Moody's Analytics (2016) are not important for Italian banks, such as the loan/deposit ratio, asset or credit quality. This means that the Moody's Analytics (2016) model cannot be blindly applied to Italian banks.

Third, we observe that the 'Change in ROA' (coined in Moody's Analytics (2016)) is statistically significant in the specifications Pr01, Pr02, Pr03, Pr04, Pr06, Pr07.

As what is left is Pr05 specification, we should look closely at the respective coefficient estimate and further compare the results. From one side, the 'Change in ROA' coefficient estimates is non-stable. They vary from 22 in Pr03 to 40 in Pr02, i.e., mostly twice or 200 % (compare it to the stability of the I2PNTA coefficient estimate that varies only from 7.1 in Pr08 to 9.3 in Pr04, i.e. by 30 % only). On another side, adding 'Change in ROA' changes 'ROA' coefficient estimates from -59 in Pr05 to -77 in Pr07. This is a typical example of multicollinearity. In total, using both factors of 'ROA' and 'Change in ROA' results in similar overall PD predictions.

However, the separate interpretation of the coefficients incorrectly indicates a higher contribution of the ROA. For this reason, we consider the Pr05 specification to be the best, i.e.

$$Pr(D = 1) = N\big(-1.3 - 7.5 \cdot LR - 59 \cdot ROA - 0.6 \cdot BCC\big), \tag{5}$$

Electronic copy available at: https://ssrn.com/abstract=4141518

where: D corresponds to the flag of the of default; N is the Normal (Gaussian) CDF; LR is the leverage ratio (the ratio of total capital to assets); ROA indicates return on assets; BCC is the indicator whether the bank is a small cooperative one or not.

At this point, we are able to compare our findings with the following models:

- Italy (Best, 2022) - our best model from Eq. (5);
- Italy (Ferriani et al., 2019) - model from Ferriani et al. (2019);
- Italy (Moody's 2016) - model from specification Pr06 in Table .12. It has the very same PD drivers as mentioned in Moody's Analytics (2016), but recalibrated on our data;
- Russia (2022) - model from Penikas (2022) on Russian banks;
- World (2016, Moody's) - model weights here are taken precisely from Moody's Analytics (2016) that is based on a sample of 90 countries around the world;
- USA (2002, Moody's) reflects the factor contributions from Kocagil et al. (2002) on American banks.

The comparison of the PD factor weights for the first five specifications abovementioned is presented in Figure 6, while the benchmarking of the PD factor contributions for the entire set of six specifications is shown in Figure 7.

As we can see from Figure 6, the model weights mostly signal in the same direction of impact. The higher profitability (ROA), as well as the related capital base (Equity to Assets), have a negative impact on PD.

At the same time, important differences can be noted in the PD drivers of Italian banks. The capital base was much more important for Italian banks than for those of other banks in the world. Furthermore, the significant reduction in PD for Italian banks stems from other factors. This is the intercept and the BCC dummy. Contrary to Ferriani et al. (2019), Penikas (2022), we do not find statistical evidence that the size of the bank is important for predicting Italian idiosyncratic PD. We can partly say that here we are in line with Kocagil et al. (2002), Moody's Analytics (2016), but we do not know whether Moody's RiskCalc model rejected the size factor as insignificant or not considered at all, leaving the model with relative ratios only.

Regarding the absolute contribution of the PD factor, we can also compare our results with those of Kocagil et al. (2002) on USA banks, although they date back mainly two decades ago. We can see from Figure 7 that liquidity could be a fairly common PD factor contributing around 15-20 % of the PD forecast for Italian, Russian, US and other banks, although we do not find it as significant in our best model.

Electronic copy available at: https://ssrn.com/abstract=4141518

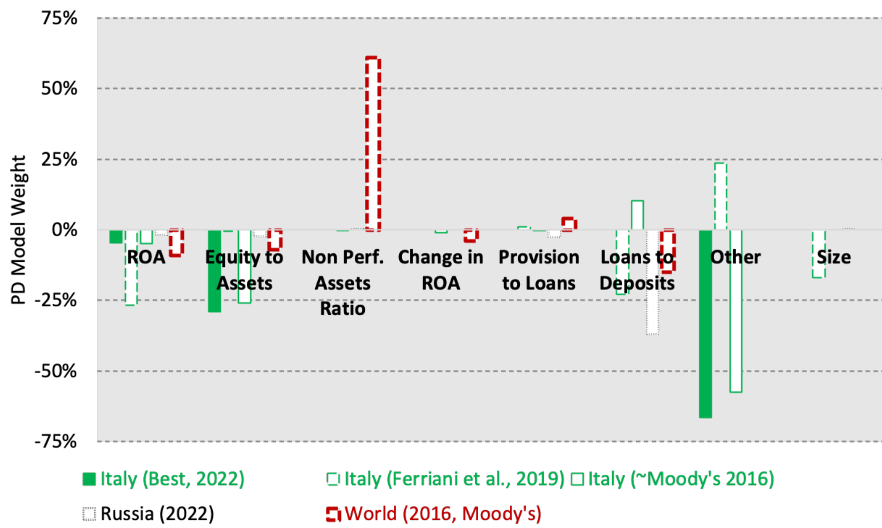

Figure 6: Italian banks' PD mostly depend to its relative earnings (ROA and change in it), being much less impacted by asset and loan book quality compared to the US and Russian banks.

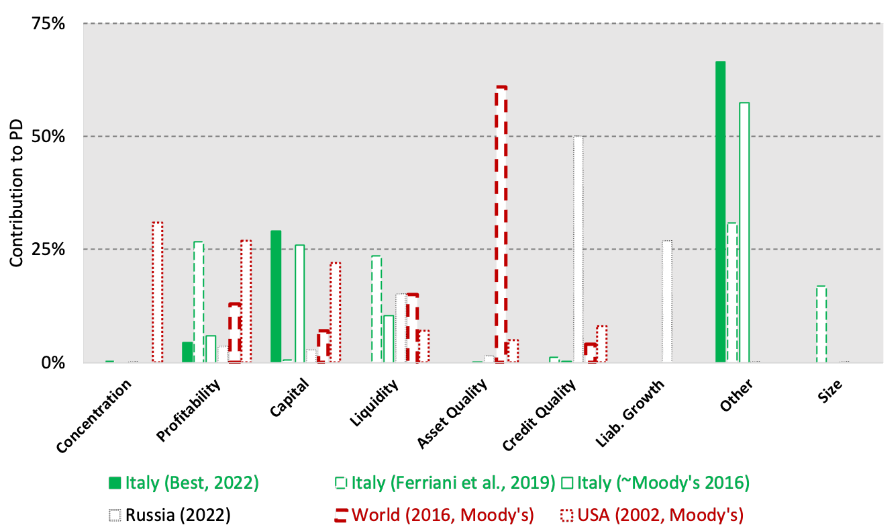

Figure 7: Italian banks are equally sensitive to leverage (equity to assets), as the US and Russian banks.

Electronic copy available at: https://ssrn.com/abstract=4141518

Our key conclusion is that Italian banks have their idiosyncrasies to be considered when designing a PD model. Moody's Analytics (2016) model for world banks is the closest to our model, although it has redundant factors that do not matter for Italian banks (Change in ROA, NPL). At the same time, other important factors that provide up to 50 % weight in a PD model are missing.

This is a systematic factor captured by the intercept and by the Italian specificities of the small banks marked by the BCC indicator. This means that the previous models, when extrapolated to Italian banks, produce misleading PD predictions. The PD model developed here has its strong advantages for future PD forecasts as well as for retrospective analysis. We will cover the latter in the next subsection 6.1.

6. Discussion and Policy Implications

Based on the developed PD model for the Italian banks we have two policy implications, concerning the approach to supervision and the impact of the 2016 BCC reform.

6.1. Modern Economic History of Italian Banks

As we have already mentioned in Alonso and Carbo (2020), Basel II and the Internal-Ratings-Based (IRB) approach fostered the demand for the development of the PD models. However, the PD model, if properly developed, can also provide valuable information about the past in the following way.

We developed a PD model using known facts about the particular bank default. After estimating a PD model, we predicted the PD not only for the future and calculate the current IRB capital requirements, but also retrospectively, i.e. for any bank, including defaulting ones as in Figure 8.

The supervisors can remember what has happened to the banking sector in past years and conclude whether particular supervisory action probably should have been taken sooner. For example, consider the defaulting Carige bank in 2019 (see dotted blue line in Figure 8). We see that the PD in the year of insolvency was close to 70 %.

However, the striking point is that in 2013 the retrospective PD forecast exceeded 80 %. This means that perhaps if the supervisor had imposed stricter supervisory restrictions on Carige in 2013, it could have continued to exist and not defaulted in 2019.

The respective analysis can be performed for so-called healthy banks, as shown in Figure 9. We have already mentioned an LDP term that is relevant for Italian banks as a common segment (or portfolio). A typical challenge for an LDP segment is to estimate the PD given the fact that there have been no defaults in the segment.

Electronic copy available at: https://ssrn.com/abstract=4141518

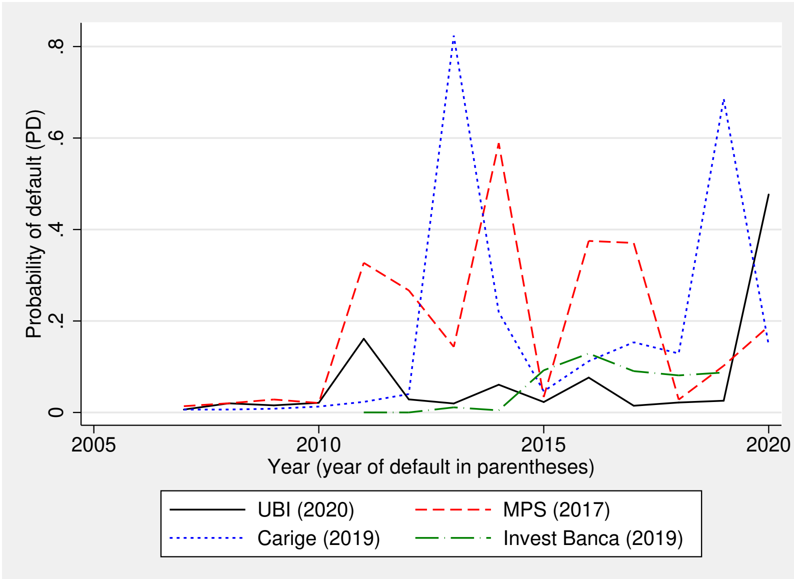

Figure 8: Developed PD model allows to trace back the roots when the default causes emerged.

In other words, if applied to our case, one wonders what the PD is for banks that have never defaulted, such as the major Italian banks like Intesa, Unicredit, Iccrea, etc.

The developed PD model allows us to answer the question by performing a relative ranking of non-defaulting banks with an indication of the potential value of the PD. For example, although Iccrea has never defaulted, its PD is the highest in 2020 and growing slightly (see blue dotted line in Figure 9). Furthermore, Unicredit is quite healthy with a PD forecast of around 5 % in 2020 (see dashed red line in Figure 9). However, in 2013 and 2016, the bank was in quite a difficult time judging by the model when the PD jumped above 20 %.

On the contrary, Intesa performs better than Unicredit, although both have had a comparable increase in the PD of up to 15 % in 2011 (see solid black line in Figure 9), but, while the PD of Intesa was continuously decreasing, the PD of Unicredit was rather volatile.

A final note concerns the Banco di Asti as it appears to be the most stable of all those considered. In this case, the lowest PD forecast is around 2 % and is fairly stable over the last decade (see long dashed green line in Figure 9), regardless of the defaults of the unhealthy banks and the peaks of PD for the healthy ones.

Electronic copy available at: https://ssrn.com/abstract=4141518

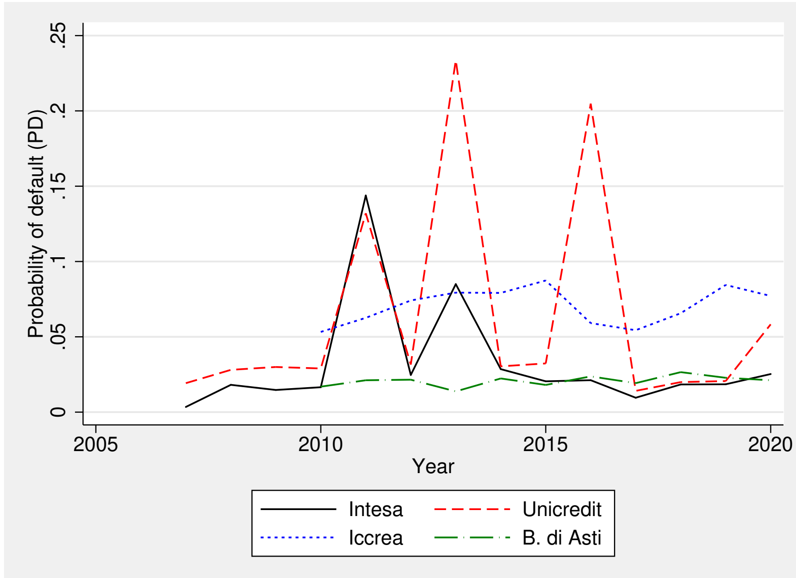

Figure 9: Developed PD model enables to differentiate credit risk with respect to liabilities of non-defaulted banks.

6.2. 2016 BCC Reform

We already referred to the reform of the small cooperative (BCC) banks in Italy launched in 2016, see EACB (2016). Following the paper by Ferriani et al. (2019) we introduced a BCC dummy. However, the curious reader might challenge us by asking why not performing a treatment effect estimate of the BCC reform. To do this, the widespread approach is to apply the difference-in-differences (DiD) setting. For instance, recent studies Duong (2021) use DiD to evidence that inflation targeting (IT) does not improve the well-being of the emerging economies, though during pandemics countries following IT had lower inflation rates, but paid a high cost for such disinflation.

We tried to improve the above specification Pr05 from Eq. (5) by adding before-after $Y2016$ dummy and the interaction one $BCC\_2016 = BCC \cdot Y2016$. However, the model omitted the interaction dummy of our core interest as it perfectly predicted non-defaults. As table 4 shows, we have a more or less proportionate allocation of observations within DiD matrix. However, the default rate (DR) for the BCC banks after 2016 equals to zero. It fell from 1% in the 'Before' 2016 period. Non-BCC banks demonstrate quite an opposite trend. The DR for the non-BCC banks in Italy rose from 2.4% in the 'Before' 2016 period to 7.4% in the aftermath.

Thus, the BCC 2016 treatment effect on the DR is the difference of differences

Electronic copy available at: https://ssrn.com/abstract=4141518

Table 4: Difference-in-Differences Matrix for BCC Reform in Italy

BCC Dummy	Indicator	Before-After Dummy		Total
		Before 2016	After 2016	
non-BCC	N	334	208	542
	DR	.0242	.0745	.0407
BCC	N	278	225	503
	DR	.0109	0.0000	.0066
Total	N	612	433	1045
	DR	.01818	.0352	.0243

Note: N - number of observations; DR - default rate (share of total N).

in DR equaling to -6 pp. as follows:

$$DR_{A-B}^{BCC} - DR_{A-B}^{Non-BCC} = \big(0.0\% - 1.1\%\big) - \big(7.4\% - 2.4\%\big) = -1.1\% - 5.0\% = -6.1\%$$

where $A - B$ stands for the 'After' minus 'Before' difference.

We are aware of the limitations of our approach as we are unable to test for the parallel trend assumption, as is done in (Mäkinen, 2021, p.15-16) or discussed in (Duong, 2021, p. 9). Nevertheless, disregarding the known limitations, we are able to offer a second-best estimate and conclude that the BCC reform had an improving impact on the propensity to default of small cooperative (BCC) banks in Italy, i.e. BCC banks in Italy became less prone to default after they started been supervised by the Single Supervisory Mechanism (SSM). We attribute this to the positive effects of the SSM under which BCCs are governed since 2016. In that, we affirm the earlier support to SSM raised in studies by Loipersberger (2018), Tziogkidis et al. (2020).

6.3. Conclusions

To summarize, we have illustrated the advantage of the new PD model for Italian banks that outperforms the previous banking PD models of Kocagil et al. (2002), Moody's Analytics (2016), Ferriani et al. (2019), Penikas (2022) when applied to Italian banks. Furthermore, we have demonstrated the application of the PD model developed to the analysis of the Italian banking system and confirmed the positive effect of BCC reform in Italy as it led to a decline in BCC default rate and approve the SSM performance.

Electronic copy available at: https://ssrn.com/abstract=4141518

Acknowledgements

Authors acknowledge Olexei Glavatiy for assistance in coding the PD model for Russian banks presented in Penikas (2022).

Disclaimer

Opinions expressed in the paper are those of the authors and may not reflect the official position of the affiliated institutions.

References

Alonso, A. and Carbo, J. M. (2020). Machine Learning in Credit Risk: Measuring the Dilemma Between Prediction and Supervisory Cost. Banco de Espana Documentos de Trabajo N.º 2032. `https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/20/Files/dt2032e.pdf`. Online; accessed 08 November 2020.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4):589–609.

Altman, E. I., Danovi, A., and Falini, A. (2013). Z-score models' application to italian companies subject to extraordinary administration. *Journal of Applied Finance (Formerly Financial Practice and Education)*, 23(1).

Altman, E. I., Sabato, G., and Wilson, N. (2008). The value of qualitative information in sme risk management. `https://pages.stern.nyu.edu/~ealtman/SME_EA_GS_NW.pdf`.

BCBS (2000). Credit Ratings and Complementary Sources of Credit Quality Information. `https://www.bis.org/publ/bcbs_wp3.pdf`. Working Paper No. 3.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, pages 71–111.

Bisogno, M., Restaino, M., and Di Carlo, A. (2018). Forecasting and preventing bankruptcy: A conceptual review. *African journal of business management*, 12(9):231–242.

Bräuning, M., Malikkidou, D., Scalone, S., and Scricco, G. (2020). A new approach to early warning systems for small european banks. In *International Conference on Machine Learning, Optimization, and Data Science*, pages 551–562. Springer.

Electronic copy available at: https://ssrn.com/abstract=4141518

Burova, A., Penikas, H., and Popova, S. (2021). Probability of Default Model to Estimate Ex Ante Credit Risk. *Russian Journal of Money and Finance*, 80:49–72.

Chen, J., Marshall, B., Zhang, J., and Ganesh, S. (2006). Financial distress prediction in China. *Review of Pacific Basin Financial Markets and Policies*, 9:317–336.

Cheong, C. W. and Ramasamy, S. (2019). Bank failure: A new approach to prediction and supervision. *Asian Journal of Finance & Accounting*, 11:111–140.

Ciampi, F. and Gordini, N. (2008). Using economic-financial ratios for small enterprise default prediction modeling: An empirical analysis. In *2008 Oxford Business & Economics Conference Proceedings, Association for Business and Economics Research (ABER)*, pages 1–21.

Citterio, A. (2020). Bank failures: review and comparison of prediction models. https://ssrn.com/abstract=3719997.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20:37–46.

Cole, R. A., Taylor, J., and Wu, Q. (2020). Predicting bank failures using a simple dynamic hazard model. *Available at SSRN 1460526*.

Daveri, F. and Verona, G. (2020). Why Italy is doing much better than you think. https://www.weforum.org/agenda/2020/01/italy-economy-growth-instability-exports/.

Dendramis, Y., Tzavalis, E., and Cheimarioti, A. (2020). Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning. https://www.dept.aueb.gr/sites/default/files/allWP-12-20-Dendram-Tzaval-Cheimar-12-11-20_0.pdf. Athens University of Economics and Business, School of Economic Sciences Working Paper No. 12-2020.

Diebolt, F. X. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebolt-Mariano Tests. *Journal of Business & Economic Statistics*, 33:1–9.

Duong, T. H. (2021). Inflation targeting and economic performance over the crisis: evidence from emerging market economies. *Asian Journal of Economics and Banking*. 10.1108/AJEB-05-2021-0054.

Electronic copy available at: https://ssrn.com/abstract=4141518

Durand, P., Le Quang, G., et al. (2021). What do bankrupcty prediction models tell us about banking regulation? Evidence from statistical and learning approaches. https://xtra.economix.fr/pdf/dt/2021/WP_EcoX_2021-2.pdf?1.0.

EACB (2016). BCC: The Reform of the co-operative Banks in Italy is now law. https://www.eacb.coop/en/news/members-news/bcc-the-reform-of-the-co-operative-banks-in-italy-is-now-law.html. Online; accessed 03 June 2022.

EBA (2021). Discussion paper on machine learning for IRB models. https://www.eba.europa.eu/regulation-and-policy/model-validation/discussion-paper-machine-learning-irb-models. Online; accessed June 17, 2022.

Fantazzini, D. and Figini, S. (2009). Random Survival Forests Models for SME Credit Risk Measurement. *Methodology and Computing in Applied Probability*, 17:29–45.

Federal Deposit Insurance Corporation (2022). Balance Sheet Summaries from Failed Banks. [Online; accessed 20. Jun. 2022].

Federal Financial Institutions Examination Council (2022). Financial Data Download - National Information Center. [Online; accessed 20. Jun. 2022].

Federal Reserve Bank of Chicago (2022). Commercial Bank Data. [Online; accessed 20. Jun. 2022].

Ferriani, F., Cornacchia, W., Farroni, P., Ferrara, E., Guarino, F., and Pisanti, F. (2019). An early warning system for less significant Italian banks. https://www.bancaditalia.it/pubblicazioni/qef/2019-0480/QEF_480_19.pdf. Bank of Italy Occasional Paper No. 480.

Fortis, M., Corradini, S., and Carminati, M. (2016). Italy's Top Products in World Trade. The Fortis-Corradini Index. In *The Pillars of the Italian Economy*, pages 1–81. Springer, Cham, Switzerland.

Fungáčová, Z., Schoors, K., Solanko, L., and Weill, L. (2020). Political cycles and bank lending in Russia. https://helda.helsinki.fi/bof/bitstream/handle/123456789/16987/dp0820.pdf. BOFIT Discussion Papers No. 8 - 2020.

Fungáčová, Z. and Solanko, L. (2009). The Russian banking industry after the financial crisis – where to next? https://helda.helsinki.fi/bof/handle/123456789/9754. Bank of Finland Bulletin No. 2.

Electronic copy available at: https://ssrn.com/abstract=4141518

Fungáčová, Z. and Weill, L. (2013). Does competition influence bank failures? evidence from Russia. *Economics of Transition*, 21:301–322.

Fungáčová, Z., Weill, L., and Zhou, M. (2010). Bank capital, liquidity creation and deposit insurance. *Journal of Financial Services Research*, 51:97–123.

Goncharenko, R., Mamonov, M., Ongena, S. R. G., Popova, S., and Turdyeva, N. (2022). Quo Vadis? Evidence on New Firm-Bank Matching and Firm Performance Following 'Sin' Bank Closures. https://ssrn.com/abstract=4069801. CEPR Discussion Paper No. DP17015. Online; accessed 16 May 2022; registration required.

Heimberger, P. (2020). Italy is of systemic importance – European solutions are needed. https://wiiw.ac.at/italy-is-of-systemic-importance--european-solutions-are-needed-n-438.html.

Jabeur, S. and Fahmi, Y. (2018). Forecasting financial distress for French firms: a comparative study. *Empirical Economics*, 54:1173–1186.

Karas, A. and Vernikov, A. (2019). Russian bank data: Birth and death, location, acquisitions, deposit insurance participation, state and foreign ownership. *Data in Brief*, 27:104560.

Kocagil, A., Reyngold, A., Stein, R., and Ibarra, E. (2002). Moody's RiskCalc™ Model for Privately-Held U.S. Banks. http://www.rogermstein.com/wp-content/uploads/riskcalc-usbanks.pdf.

Kočenda, E. and Iwasaki, I. (2022). Bank survival around the world: A meta-analytic review. *Journal of Economic Surveys*, 36:108–156.

Kristóf, T. (2021). Bank failure prediction in the COVID-19 environment. *Asian Journal of Economics and Finance*, 3(1):157–171.

Kristóf, T. and Virág, M. (2020). A comprehensive review of corporate bankruptcy prediction in Hungary. *Journal of Risk and Financial Management*, 13(2):35.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28:1–26.

Kumar, P. R. and Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–a review. *European Journal of Operational Research*, 180(1):1–28.

Electronic copy available at: https://ssrn.com/abstract=4141518

Laviola, S. and Trapanese, M. (1997). Previsione delle insolvenze delle imprese e qualità del credito bancario: un'analisi statistica [in Italian]. https://www.bancaditalia.it/pubblicazioni/temi-discussione/1997/1997-0318/tema_318.pdf. Banca d'Italia Temi di discussione del Servizio Studi Numero 318.

Liu, Y., Yang, M., Wang, Y., Li, Y., Xiong, T., and Li, A. (2022). Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. *International Review of Financial Analysis*, 79:101971.

Loipersberger, F. (2018). The effect of supranational banking supervision on the financial sector: Event study evidence from Europe. *Journal of Banking & Finance*, 91:34–48.

Marotta, G., Pederzoli, C., and Torricelli, C. (2005). Forward-looking estimation of default probabilities with Italian data. http://morespace.unimore.it/giuseppemarotta/wp-content/uploads/sites/12/2014/11/MAR-PED-TOR-IFC3.pdf.

Menardi, G. and Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28(1):92–122.

Micheli, A. P. (2015). The default risk: An empirical analysis on the automotive companies. management studies. *International Journal of Financial Studies*, 3(7–8):169–178.

Moody's Analytics (2016). RiskCalcTM Banks 4.0. https://www.moodysanalytics.com/-/media/products/riskcalc-banks-4.pdf. The publication year is not explicitly disclosed, though we may refer to the copyright year.

Moscatelli, M., Parlapiano, F., Narizzano, S., and Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications*, 161:113567.

Mäkinen, M. (2021). Does a financial crisis change a bank's exposure to risk? A difference-in-differences approach. https://helda.helsinki.fi/bof/handle/123456789/17984. BOFIT Discussion Papers 8 • 2021; Online; accessed December 01, 2021.

Electronic copy available at: https://ssrn.com/abstract=4141518

Obeid, R. (2021). Bank failure prediction in the arab region using logistic regression model. https://www.amf.org.ae/sites/default/files/publications/2021-12/bank-failure-prediction-arab-region-using-logistic-regression-model.pdf. Arab Monetary Fund (Working Paper No. 7-2021), Available online.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, pages 109–131.

Orlando, G. and Pelosi, R. (2020). Non-Performing Loans for Italian Companies: When Time Matters. An Empirical Research on Estimating Probability to Default and Loss Given Default. *International Journal of Financial Studies*, 8(4):68.

Penikas, H. (2020). Low Default Portfolios in Basel II and Basel III as a Special Case of Significantly Unbalanced Classes in Binary Choice Models. *Russian Journal of Money and Finance*, 79:101–128.

Penikas, H. (2022). PD-LGD correlation for the banking lending segment: Empirical evidence from Russia. *Model Assisted Statistics and Applications*, 17:27–39.

Provenzano, V. and Arnone, M. (2015). Microcredit and Probability of Default for Small Business in Italy. https://euricse.eu/wp-content/uploads/2015/09/Provenzano-Arnone.pdf.

Qu, Y., Quan, P., Lei, M., and Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, 162:895–899.

Runde, D. E. (2018). The Role of Italy in Global Economic Development. https://www.csis.org/analysis/role-italy-global-economic-development.

Shibitov, D. and Mamedli, M. (2019). The finer points of model comparison in machine learning: forecasting based on Russian banks' data. http://www.cbr.ru/content/document/file/87572/wp43_e.pdf. Online; accessed 08 September 2020.

Shrivastava, S., Jeyanthi, P. M., and Singh, S. (2020). Failure prediction of indian banks using smote, lasso regression, bagging and boosting. *Cogent Economics & Finance*, 8(1):1729569.

Electronic copy available at: https://ssrn.com/abstract=4141518

Standard&Poor's (2020). Probability of default for Italian financials falls in July. https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/probability-of-default-for-italian-financials-falls-in-july-59079996.

Tian, S. and Yu, Y. (2017). Financial ratios and bankruptcy predictions: An international evidence. *International Review of Economics & Finance*, 51:510–526.

Tziogkidis, P., Philippas, D., and Tsionas, M. G. (2020). Multidirectional conditional convergence in European banking. *Journal of Economic Behavior & Organization*, 173:88–106.

Youden, W. (1950). Index for rating diagnostic tests. *Cancer*, 3:32–35.

Yuksel, S., Dincer, H., and Hacioglu, U. (2015). Camels-based determinants for the credit rating of turkish deposit banks. *International Journal of Finance & Banking Studies*, 4(4):1–17.

Electronic copy available at: https://ssrn.com/abstract=4141518

Table .5: Descriptive Statistics (after filtering for outliers)

Var	Var Description	Def	Obs	Mean	Std. Dev.	Min	Max
Def	Default Flag**	1	946	0.02	0.15	0	1
BCC	Dummy for small bank (Banca di Credito Cooperativo, BCC); see EACB (2016)	-0.1107*	1045	0.48	0.50	0	1
I1_ROE	Roe = Profit _ Net Equity	-0.3260*	1045	0.01	0.1	-0.89	0.28
I2_PN/TA	Net Equity _ Total Assets = Equity _ Total Assets	-0.1068*	1045	0.09	0.04	0	0.43
I3_PN/CVC	Net Equity _ Loans Vs Customers = Equity _ Total Loans	-0.0367	1045	0.18	0.34	0	10.09
I4_ROA	ROA = Profit _ Total Assets	-0.3050*	1045	0	0.01	-0.06	0.02
I4_ROAd	Change in ROA: ROA(t) - ROA(t-1)	-0.0779*	948	0	0.01	-0.03	0.04
I5_LA/TA	Liquidity = Liqudity Assets _ Total Assets	0.0071	1045	0.01	0.01	0	0.16
I6_P/TA	Loans = Loans _ Total Assets	0.0671*	1045	0.62	0.16	0	0.95
I7_D/TA	Deposits Vs Customers _ Total Assets	0.0175	1045	0.55	0.14	0	0.96
I8_MI/MID	Interest Margin _ Brokerage Margin	-0.0725*	1045	0.6	0.14	0	1.17
I9_RAPSI/TA	Result Before Extraordinary Income and Taxes _ Total Assets	-0.2862*	1045	0	0.01	-0.07	0.04
I10_T1CR	Tier 1 capital ratio (Tier 1 capital _ Risk weighted assets)	-0.0841*	1045	0.17	0.08	0	0.64
I11_TCR	Total capital ratio (Regulatory capital including TIER 3 _ Risk weighted assets)	-0.0761*	1045	0.18	0.07	0	0.69
I12_RRV_C/PN	Net Value Adjustments _ Write-backs Due to Impairment of: "Loans" _ Net Equity	-0.2602*	1045	-0.08	0.1	-1.31	0.1
I13_RLG	Rate of Loan Growth	-0.0115	1045	0.2	2.73	-1	85.95
I14_RRV_C/CVC	Net adjustments _ write-backs for deterioration of Loans	-0.0258	1045	0	0	-0.03	0.03
I15_CCN/TA	CCN (Interest-bearing assets - Interest-bearing liabilities) (NWC) _ Total Assets	-0.1393*	1045	0.07	0.04	-0.48	0.44
I16_CCN/PN	NWC _ Net Equity	-0.1543*	1045	0.8	0.29	-4.45	2.14
I17_CVC/DVC	Receivables from customers _ Deposits from customers	0.0422	1045	1.17	0.48	0	6.99
I18_RRV_C/CCN	Net adjustments _ write-backs for impairment of: "Loans" _ NWC	0.0073	1045	0.01	0.04	-0.57	0.56
I19_CVC/AF	Receivables from customers _ interest-bearing assets	0.0827*	1045	0.65	0.17	0	0.97
I20_TDD/AF	Debt Securities _ Interest-bearing Assets	-0.0957*	1045	0.24	0.14	0	0.84
I21_P/AF	Equity investments _ Interest-bearing activities	0.0042	1045	0	0.01	0	0.14
I22_DVB/PO	Payables to Banks _ Expensive Liabilities	-0.0036	1045	0.15	0.11	0	0.77
I23_DVC/PO	Payables to customers _ Onerous liabilities	0.0053	1045	0.63	0.17	0	1
I24_DT/PO	Payables for Securities _ Onerous Liabilities	0.0012	1045	0.18	0.12	0	0.56
I25_RRV_130/TA	Net adjustments _ write-backs for impairment of loans and other fin.items (V. 130 EC)_ TA	-0.2181*	1045	-0.01	0.01	-0.05	0
I26_RRV_AM/TA	Net adjustments _ write-backs on tangible assets (V.200 EC) _ Total Assets	-0.0322	1045	0	0	-0.01	0
I27_RRV_AI/TA	Net adjustments _ write-backs on intangible assets (V.210 CE) _ Total Assets	-0.052	1045	0	0	-0.01	0.01
I28_RVA/TA	Value adjustments of goodwill _ Total Assets	-0.0776*	1045	0	0	-0.02	0
I29_TRRV/TA	Total adjustments _ write-backs _ total assets	-0.2209*	1045	-0.01	0.01	-0.06	0
I30_TRRV/PN	Total adjustments _ write-backs _ shareholders' equity	-0.2585*	1045	-0.11	0.12	-1.5	0.06
I31_TA/PN	Total Assets _ Net Equity	0.1149*	1045	12.9	7	0	89.84
I32_TAT/PN	Total Tangible Assets _ Tangible Net Equity	0.1036*	1045	13.69	7.59	0	93.03
I33_MI/TA	Interest Margin _ Total Assets	-0.0745*	1045	0.02	0.01	0	0.06
I34_MID/TA	Brokerage Margin _ Total Assets	-0.0535	1045	0.03	0.01	0	0.1
I35_RG/TA	Operating Profit _ Total Activity	-0.1188*	1045	0.01	0.01	-0.03	0.04
I36_RAPSI/PNT	Result Before Extraordinary Income and Taxes _ Tangible Net Equity	-0.2418*	1045	0.02	0.12	-1.44	0.43
I37_RAI/TA	Pre-Tax Result _ Total Assets	-0.2744*	1045	0	0.01	-0.07	0.04
I38_RN/RIS(140)	Net Result _ Reserves (140)	0.0404	1045	-0.44	17.62	-565.37	53.62
I39_RG/PN	Operating Profit _ Net Equity	-0.0561	1045	0.12	0.07	-0.27	0.58
I40_RAPSI/PN	Result Before Extraordinary Income and Taxes _ Net Equity	-0.2969*	1045	0.02	0.12	-1.33	0.43
I41_RAI/PN	Pre-Tax Result _ Net Equity	-0.2859*	1045	0.02	0.12	-1.28	0.43
I42_ROTE	ROTE = Net Result _ Tangible Net Equity	-0.3088*	1045	0.01	0.11	-0.96	0.3
I43_RG/MID	Operating Profit _ Brokerage Margin	-0.1432*	1045	0.32	0.17	-1.35	1.5
I44_RAPSI/RG	Result Before Extraordinary Income _ Operating Result	-0.0898*	1045	0.31	2.89	-27.19	75.81
I45_RAI/RAPSI	Result Before Taxes _ Result Before Extraordinary Income and Taxes	-0.0073	1045	1.11	2.13	-2.69	68.22
I46_RN/RAI	Net Result _ Pre-Tax Result	-0.0132	1045	1.43	22.77	-66.67	731.98
I47_MI/SA	Interest Margin _ Administrative Expenses	0.1090*	1045	-0.88	0.26	-2.72	0
I48_MID/SA	Brokerage _ Administrative Expenses	0.0897*	1045	-1.45	0.33	-3.64	0
I49_MI/CO	Interest Margin _ Operating Costs	0.1161*	1045	-0.94	0.3	-2.72	1.23
I50_MID/CO	Brokerage Margin _ Operating Costs	0.0991*	1045	-1.55	0.39	-3.67	1.99
I51_BKS_LN(TA)	Bank Size (Proxy Variable) = Ln (TA)	0.06	1044	14.8	2.66	8.85	24.65
I52_RN/MID	Net Result _ Brokerage Margin	-0.3153*	1045	0.05	0.26	-3.22	0.74
I53_RN/CO	Net Income _ Operating Costs (Net Income _ Operating Costs)	0.2546*	1045	-0.12	0.35	-2.45	2.67
I54_CO/MID	Operating Costs _ Brokerage Margin	-0.1333*	1045	-0.67	0.19	-2.35	0.63
I55_RRVsC/MID	Net adjustments _ write-backs for impairment of: a) Loans _ Brokerage margin	-0.2492*	1045	-0.22	0.21	-1.91	0.19
I56_MID/SP	Brokerage margin _ Personnel expenses	0.057	1045	-2.47	1.16	-8.44	4.92
I57_Fc/Ac	Fc _ Ac = (PN + DvB + DvC) _ (DL + CvB + CvC)	-0.0651*	1045	1.14	0.37	0	4.14
I58_RAV/AI	RAV _ AI = Value adjustments of goodwill _ intangible assets	-0.0102	1045	-0.69	14.07	-440.76	0
I59_AM/TA	Tangible Assets _ TA	0.043	1045	0.01	0.01	0	0.06
I60_AF/(TA-PN)	Financial assets _ (TA-PN)	-0.1023*	1045	0.25	0.16	0	0.9
I61_DVC/(TA-PN)	Customer deposits _ (TA-PN)	0.0002	1045	0.61	0.16	0	0.98
I62_DVC/SA	Deposits Vs Customers _ Administrative Expenses	-0.0014	1045	-27.73	10.83	-149.74	0
I63_DVB/(PN+DVB+DVC)	Payables Vs Banks _(PN+Deb.Vs. Banks+Deb.Vs. Customers)	0.0067	1045	0.18	0.12	0	0.88
I64_P/(PN+DVB+DVC)	Loans _ (PN + Debt to Banks + Debt to Clients)	0.0592	1045	0.82	0.26	0	1.8
I65_CDL/DVC	Cash and cash equivalents _ payables to customers	0.0051	1045	0.02	0.02	0	0.25
I66_(CVC-RRV130)/DVC	(Receivables to Customers - Adj. _ Rep. To Net Values Per Det. (V.130)) _ Payables Vs Customers	0.0405	1045	1.2	0.47	0	7.03
I67_CVB/AF	Loans Vs Banks _ Interest-bearing Assets	0.0016	1045	0.09	0.1	0	0.79
I68_RRV130/CVC	Adjustments _ Rep. A Val. Net for Det. (V.130)) _ Credits Vs Customers	-0.0836*	1045	-0.01	0.02	-0.5	0.02
I69_RN/(TA-PN)	Net Result _ (TA-PN)	-0.2973*	1045	0	0.01	-0.06	0.03
I70_IP/(TA-PN)	Interest expense _ (TA-PN)	0.0173	1045	-0.01	0.01	-0.04	0
I71_RRV130/PNT	Adjustments _ Rep. A Val. Net for Det. (V.130)) _ Tangible Shareholders' Equity	-0.2455*	1045	-0.09	0.11	-1.45	0.1
I72_RRV130/MID	Adjustments _ Rep. A Val. Net for Det. (V.130)) _ Brokerage margin	-0.2497*	1045	-0.23	0.21	-1.96	0.17

[a] * Significant at 5%.
[b] ** The number of observations for 'Def' is less by almost 100 observations (i.e. 1045-946) due to one year shift when predicting PD for 100 banks one year ahead.

Electronic copy available at: https://ssrn.com/abstract=4141518

Table .6: Logistic Regression DS(B100)

Solution to unbalanced class	Accuracy	Sensitivity (EII°T)	Specificity	P-Value	Kappa	AUC	Youden's J
Imbalanced	0,9811	0,9961	0,4286	0,2973	0,5366	0,712	0,4247
Over Sampling	0,9698	0,9845	0,4286	0,7309	0,4131	0,707	0,4131
Both (Over-Under Sampling)	0,966	0,9806	0,4286	0,8332	0,3826	0,705	0,4092
ROSE	0,9396	0,9496	0,5714	0,9992	0,3074	0,761	0,521
Mean	0,96	0,98	0,46	0,72	0,41	0,72	0,44
Median	0,97	0,98	0,43	0,78	0,40	0,71	0,42

[a] Logistic regression model results with the four DS(B100) dataset balancing solutions

Table .7: Logistic Regression DS(B73)

Solution to unbalanced class	Accuracy	Sensitivity (EII°T)	Specificity	P-Value	Kappa	AUC	Youden's J
Imbalanced	0,9894	1	0	0,4795	0	0,5	0
Over Sampling	0,9574	0,9677	0	0,9998	-0,0162	0,516	-0,0323
Both (Over-Under Sampling)	0,9734	0,9839	0	0,984	-0,0129	0,508	-0,0161
ROSE	0,9255	0,9355	0	1	-0,0186	0,532	-0,0645
Mean	0,96	0,97	0,00	0,87	-0,01	0,51	-0,03
Median	0,97	0,98	0,00	0,99	0,01	0,51	-0,02

[a] Logistic regression model results with the four DS(B73) dataset balancing solutions

Table .8: Classification and Regression Tree DS(B100)

Solution to unbalanced class	Accuracy	Sensitivity (EII°T)	Specificity	P-Value	Kappa	AUC	Youden's J
Imbalanced	0,9774	0,9661	0,2857	0,4477	0,3903	0,641	0,2518
Over Sampling	0,9698	0,9884	0,2857	0,7309	0,3183	0,637	0,2741
Both (Over-Under Sampling)	0,9887	0,9884	1	0,8179	0,8179	0,994	0,9884
ROSE	0,9698	0,9845	0,4286	0,7309	0,4131	0,707	0,4131
Mean	0,98	0,98	0,50	0,68	0,48	0,74	0,48
Median	0,97	0,99	0,36	0,73	0,40	0,67	0,34

[a] Cart model results with the four balancing solutions on DS(B100) datasets

Table .9: Classification and Regression Tree DS(B73)

Solution to unbalanced class	Accuracy	Sensitivity (EII°T)	Specificity	P-Value	Kappa	AUC	Youden's J
Imbalanced	0,9787	0,9892	0	0,9483	-0,0108	0,505	-0,0108
Over Sampling	0,9787	0,9892	0	0,9483	-0,0108	0,505	-0,0108
Both (Over-Under Sampling)	1	1	1	0,1339	1	1	1
ROSE	0,9415	0,9516	0	1	-0,0177	0,524	-0,0484
Mean	0,97	0,98	0,25	0,76	0,24	0,63	0,23
Median	0,98	0,99	0,00	0,95	-0,01	0,51	-0,01

[a] Cart model results with the four balancing solutions on DS(B73) datasets

Table .10: Random Forest DS(B100)

Solution to unbalanced class	Accuracy	Sensitivity (EII°T)	Specificity	P-Value	Kappa	AUC	Youden's J
Imbalanced	0,9736	0,9922	0,2857	0,5987	0,3512	0,639	0,2779
Over Sampling	0,9736	0,9922	0,2857	0,5987	0,3512	0,639	0,2779
Both (Over-Under Sampling)	0,9962	0,9961	1	0,006797	0,9314	0.998	0,9961
ROSE	0,9698	0,9845	0,4286	0,7309	0,4131	0,707	0,4131
Mean	0,98	0,99	0,50	0,48	0,51	0,75	0,49
Median	0,97	0,99	0,36	0,60	0,38	0,67	0,35

[a] Random Forest results with the four balancing solutions on DS(B100) datasets

Electronic copy available at: https://ssrn.com/abstract=4141518

Table .11: Random Forests DS(B73)

Solution to unbalanced class	— Accuracy	— Sensitivity (EII°T)	— Specificity	— P-Value	— Kappa	— AUC	— Youden's J
Imbalanced	0,9894	1	0	0,6767	0	0,5	0
Over Sampling	0,9894	1	0	0,6767	0	0,5	0
Both (Over-Under Sampling)	1	1	1	0,1339	1	1	1
ROSE	0,9681	0,9785	0	0,9957	-0,0144	0,511	-0,0215
Mean	0,99	0,99	0,25	0,62	0,25	0,63	0,24
Median	0,99	1,00	0,00	0,68	0,00	0,51	0,00

[a] Results of the random forests model with the four DS dataset balancing solutions (B73)

Table .12: Higher R-squared is often driven by multicollinearity and exorbitant scales of estimated coefficients

Variable	Pr01	Pr02	Pr03	Pr04	**Pr05**	Pr06	Pr07	Pr08
_cons	-1.945***	-1.562***	-0.469	-1.126***	-1.261***	-1.639***	-1.341***	1.337
I1ROE	-29.940***							-13.666
I2PNTA			-8.797*	-9.328*	-7.556*	-8.201*	-8.508**	-7.119
I2PNTA_BCC				5.317				13.169
I4ROA		-243.758***	-103.830***	-57.223***	-59.484***	-79.499***	-77.413***	-68.364
I4ROA_BCC				-6.983				42.706
I4ROAd	31.543**	40.077***	22.928*			28.737**	28.361**	37.866**
I6PTA			1.278					-0.488
I8MIMID			-0.329					-1.058
I14RRV_CCVC						-3.348		21.753
I16CCNPN			-0.155					-0.118
I17CVCDVC						0.254		0.357
I25RRV_130TA						3.845		-3.415
I28RVATA			102.192*					63.582
I36RAPSIPNT								3.173
I37RAITA		294.814***						-17.117
I40RAPSIPN								-13.314**
I41RAIPN	8.450***							21.825
I44RAPSIRG			0.012					0.019
I48MIDSA			-0.448					0.026
I51BKS_LNTA			-0.102*					-0.117*
I60AFTAPN		-2.840***						-2.798*
I68RRV130CVC			9.857					-17.853
BCC	-0.727**		-0.949***	-1.012	-0.568**			-1.792*
N	849	849	849	946	946	849	849	849
r2_p	0.291	0.307	0.316	0.255	0.253	0.24	0.236	0.403
AUC	0.8493	0.8742	0.8765	0.8610	0.8560	0.8330	0.8226	0.8900

[a] *** Significant at 1%; ** - at 5%; * - at 10%.
_cons - intercept; N - number of observations, r2_p - pseudo R-squared; AUC - area under ROC curve. Variable description is given in Table .5. Pr05 (in bold) is the best specification as it has comparable fit by pseudo R-squared value with alternatives, is not prone to multicollinearity (no inflation in estimated coefficients), has only statistically significant factors.

Electronic copy available at: https://ssrn.com/abstract=4141518