



**POLITECNICO**  
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE



EXECUTIVE SUMMARY OF THE THESIS

# Default prediction of SMEs in the Italian Minibond market: an analysis under a new human perspective

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

**AUTHOR: MATTEO SPEDICATI, TOMMASO TREGROSSO**

**ADVISOR: GIANCARLO GIUDICI**

**ACADEMIC YEAR: 2022-2023**

## 1. Literature Analysis

As a starting point of the work, an overview of main themes in discussion is given to introduce the specific analysis that is the core of the dissertation. With this, the related literature review is presented for each topic.

Firstly, it's described the important role played by SMEs at European and Italian level, in terms of percentage of SMEs out of all firms, value added and employment, highlighting their importance in the whole economic system. Then, a critical review of the literature related to SMEs' access to capital markets and SMEs' financing structure is done, to better understand which are the main determinants. In the second part, an analysis of the current financing situation in Europe and in Italy is made up, emerging that Italian SMEs are more dependent on bank loans. Finally, following the research results summarized in *Rapporto Cerved 2022*, the focus shifts to Italian SMEs' performances, to understand which are the conditions after Covid crisis and current energy shock.

Then, it's analysed in deep the mini-bond world, addressing the reasons that led to its birth, the

characteristics of the instrument (advantages and disadvantages) and the regulatory framework (which facilitates its spread into SMEs market and providing data and statistics on Italian and foreign mini-bond market). In addition, some numbers for the Italian mini-bond market are given, resulting that the 2022 was a year of further diffusion of awareness among companies with the record of 190 new issuers. Looking also at the issuances' perspective, it is confirmed that minibonds hold their own "stable" market since the reforms, initiated by the 2012 Development Decree, came into force. Starting from 2013, we have a total sample of 1461 placements (268 in 2022).

Lastly, it is presented a deep overview of literature situation of the topic "*business failure prediction*", from traditional studies to new approaches for SMEs. The recent global financial crisis has resulted in numerous company failures in many countries, renewing the literature's interest in default risk forecasting models. Even though these models have been developed since the 1960s, a growing number of studies have been published in recent years, either proposing new approaches or contrasting various models that already exist to determine which has the highest predictive power. Nonetheless, the significance of implementing early warning systems has been highlighted by the

current financial crisis. While warning system implementation and business failure forecasting are conceptually different, there is a risk of overlapping these concepts. Lots of model have been reviewed and most of them use financial ratios to predict bankruptcy, rather than corporate governance indicators. On the other hand, it is underlined how SMEs have specific characteristics that led to construct innovative model for their default prediction (financial ratios are no longer accurate enough). So, many models have been reviewed here too and lots of non-traditional quantitative methodologies and non-financial predictive variables that improve prediction accuracy have been briefly described.

## 2. Objectives and Data Collection

The objective of this dissertation is finding out if there is a relation between the default probability in the companies that issued minibonds focusing, on one hand, on the coexistence in the model of the classical predictors like the Assets or the Annual Interest Rate and ,on the other, new type of variable (the registration of intellectual properties and the composition of the Board of Directors).

For the analysis it has been taken in consideration the emissions from the 8<sup>th</sup> of April 2013 to the 31<sup>st</sup> of December 2020, the second type of filter is based on the elimination from the sample of the companies belonging to the financial sector (ATECO code K) and the emissions that exceed the amount of fifty millions of Euros.

The research questions that guided the research are:

**RQ1:** *Is it possible to observe any pattern between the different variables and the default of a firm?*

**RQ2:** *Are there any new types of variables that can perform a default probability prediction that are different from the more known indicators?*

The process of data collection consists in five distinct phases that brought to the classification of all the companies in the sample as defaulted or not. The phases are the following:

1. Query in the AIDA platform of all the companies and recognize those without any type of problem as “non default”.

2. Research on Telemaco database for those firms that are not present in AIDA, the objectives are the same of the previous phase.
3. Check for all the companies that show problems and classification of all the observations as Default or No Default.
4. Discovery new variables for each company regarding three types of IPs: trademarks, designs, and patents. The research has been conducted through the database of the EUIPO (for trademarks and designs that are collected as a single variable) and EPO (for patents applications).
5. Research on the new variables in the field of the composition of the Board of Directors from the databases of AIDA and Telemaco, the new variables found regard the age of the members their gender distribution and the total number of members.

## 3. Data Analysis and Results

The core of this research is founding a predictive model that can represent the default probability of the firms using different sets of variables. Before the building of the model, univariate statistics on all the variables that will be necessary later, considering that there are both numeric, dichotomic and multilevel variables, have been conducted.

After the completion of this preliminary phase in which it is possible to observe differences between the distribution of the different variable, the next step is the multivariate analysis in order to create a logistic regression classifier. A stepwise procedure has been conducted that, starting from three models composed by all the variables in the sets, deletes the less significant variable until each model reach a stable solution, a solution has been considered as stable when all the variables that compose the set are significative in:

- Absence of randomness.
- Coefficient statistically different from zero.
- Collinearity that is non-significant from a statistical point of view.

The three models are composed from different sets of variables. for the detail of the variables that compose each set please refer to the Table 3.4 in the

main body of the dissertation, here will be named as conventional variables and non-conventional variables. The first model is composed only by conventional variables (e.g. the annual interest rate, the revenues, and the assets), the second one only by non-conventional variables (e.g. Number of patents applications and dimension of the Board of Directors) and the last one that include both of conventional and non-conventional variables. At the end of this first phase the statistics of the model are the following (Table 1: Statistics of the models.):

Table 1: Statistics of the models.

Model 1	
N.Obs	969
AIC	483.89
BIC	503.40
Fisher Scoring	8
Model 2	
N.Obs	969
AIC	404.17
BIC	428.56
Fisher Scoring	6
Model 3	
N.Obs	969
AIC	460.12
BIC	489.38
Fisher Scoring	10

In all the statistics of the model it is possible to see how the second model results as the most performing, having the lowest AIC, BIC and reach the best likelihood in the lowest number of iterations.

The Table 2 represents the coefficient of the three logistic regressions.

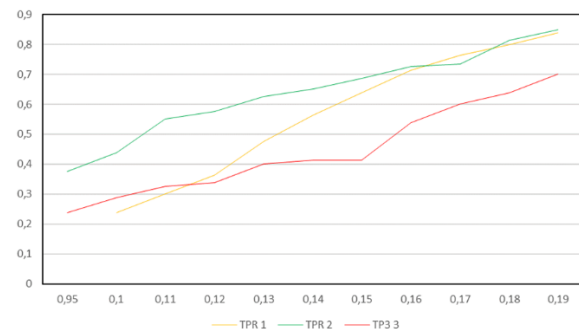
Table 2: Betas of the Logistic Regressions.

Statistic	Beta	Beta	Beta
Intercept	-3.533	-3.603	-5.914
Age_BoD		0.041	0.0495
Listed	-2.345		-
Patents		-0.149	-0.335
Rate	34.725		37.63
Revenues	-2.01e-8		-1.46e-8
TM_D		-0.085	-
Tot_BoD		-0.321	-0.264

The following step is the analysis of the global performance of the three different models in terms of True Positive Rate and True Negative Rate. The three models perform similarly for the negatives rate, while the difference is quite prominent for the true positive rates. In this statistic we can see how the second model clearly represent the best fitting model, in fact, looking at the graph below it is possible to see how the second model reach strongly greater performance for low level of threshold while for the higher performs similarly to the first model.

Here we have another check of the correct choice that would be the second model.

Figure 1: performance of True Positive Rate to vary of threshold.



The last step in order to create an effective prediction model is determining the best threshold to be used.

From now on the lender point of view, called generally the bank, will be considered; the assumption, quite strong, needed only for the purpose of this paper is the following: in the market there is only one financial institution (the bank) that can grant all the debit required by the borrower. The objective of the bank is the maximization of the profits intended as the difference between the revenues and costs. Therefore, the bank is interested in the maximization of the following equation:

$$\pi = TP \times (r_{TP} - c_{TP}) + FP \times (r_{FP} - c_{FP}) + TN \times (r_{TN} - c_{TN}) + FN \times (r_{FN} - c_{FN})$$

The total number of True Positives, False Positives, True Negatives, and False Negatives is a function of the threshold exclusively (for the legend of the symbols, is recommended the redirection to the section 3.6.2. of the main body).

In this paper proxies have been created using the means of the interest rates and of the total amount lent, these type of indicators poorly fit the model and, therefore, the precise maximum results, has

not been chosen, but rather the point of elbow in which the profits start to become more stable and the advantages of the bank would be no more significant if compared to the strict policy that would be applied.

The thresholds that best fit the model and that could be the most useful in terms of prediction are:

- For the first model: 0.19.
- For the second model: 0.18.
- For the third model: 0.24.

In order to assess the goodness of the model obtained from the logistic regression a robustness test has been conducted using the classification trees. Trees are definitely clearer but poorer in performances if compared to the Logistic Regression.

The analysis conducted using classification trees shows comparable results in the types of correlations between the variables and the default probability.

It is possible to answer positively to both the research questions of this dissertation:

1. The correlation between the variables exists and it is shown in Table 2, in fact the presence of statistically significant coefficient means that the correlation exists, and the positive or negative sign give information about the type of correlation.
2. There is effectively a set of variable that performs better than the model built only on the more economic variables.

#### 4. Comments on the results

The variables in this research have been grouped in two subset: the conventional variables and the non-conventional. From the first group the revenues and the interest rate resulted as the more significant variables. These two variables represent respectively the lender and the borrower perspective, higher interest rates lead to higher default probabilities, higher revenues strongly reduce this probability.

The second set of variables regards the innovation propension of the companies and the structure of their board of directors. The first aspect shows a negative correlation between the number of Intellectual Properties and the default probability, the second aspect shows positive correlation between the age of a board and the default, however, it is more interesting the strong

positively correlation given by the total number of members in the board of directors: wider is the board clearer the firm success. The second important aspect on these variables is that the gender does not results significative in the model, this statement could be misleading because it might seem that there is no difference from a gender perspective, however, going in deep, it is possible to note that the differences in the gender gap are still incredibly strong, more that 80% of the members in the boards of directors are male, at the same point this high disparity makes more difficult to detect statistically the patterns.

#### 5. Conclusions

The final output of this dissertation is the reasoning behind all the statistics developed previously. The most valuable resource on which a company can relies on is again the human brain, its inventing abilities, and the aptitude to cooperate brings better results than all the economic numbers that can be reached.

This paper has not at all the claim of being a whole comprehensive work, it is rather a starting point for research that put the human as the main resource also nowadays.

#### References

Angelino, A., Balda, F., Oliverio, F., Sampoli, L., & Schivardi, F. (n.d.). *Rapporto Cerved PMI 2022*.

#### Acknowledgements

We would like to express our gratitude, first of all, to our Professor Giancarlo Giudici, for the support that gives us during all the journey of this thesis.

Secondly, we would like to thank all the community of the Politecnico di Milano for all the competences and the development of our mind in these years.

Lastly, we would like to thank our beloved families, friends, and all people supporting us from the first day.