risks

MDPI

Article

Die Hard: Probability of Default and Soft Information

Giampaolo Gabbi [1,*], Michele Giammarino [2] and Massimo Matthias [3]

1 Banking and Insurance Department, SDA Bocconi School of Management, 20136 Milan, Italy
2 Banca Monte dei Paschi di Siena, 53100 Siena, Italy; m.giammarino@yahoo.it
3 Department of Business and Law, University of Siena, 53100 Siena, Italy; matthias@unisi.it
* Correspondence: giampaolo.gabbi@sdabocconi.it

Received: 28 March 2020; Accepted: 6 May 2020; Published: 13 May 2020

check for updates

Abstract: The research aims to verify whether the credit risk of small and medium-sized enterprises can be estimated more accurately using qualitative variables together with financial information from reports. In our paper, we select qualitative variables within the conceptual framework of the balanced scorecard to assess the credit quality of Italian companies of various sizes, from micro to medium. Data were collected to estimate the company's resilience following the shock of the financial crisis of 2007–2008. The analysis based on customer size, processes, knowledge, and corporate finance, synthesized with balanced scorecard methodology, allows us to estimate the resilience of companies in a period of crisis. The research highlights the important contribution of qualitative variables for the estimation of credit risk. The implications concern both financial intermediaries and their supervisory functions, and regulators for rating models based on soft forward and countercyclical variables.

Keywords: modeling credit risk; qualitative variables; rating; balanced scorecards; SME

1. Introduction

Recent events, such as the financial turmoil and the Covid19 pandemic, affected the resilience of businesses and radically changed the way to assess the credit risk of borrowers. When analyzing the most frequently used models in credit risk management, the most commonly used variables are quantitative variables based on historical balance sheet data and trends, both for loans and bonds (Gabbi and Sironi 2005). Precisely because many (if not all) of these data are historical, they can often generate procyclicality effects (Gabbi and Vozzella 2013). The supervisory authorities themselves have acknowledged that the Basel II framework has caused undesirable effects on the stability of the system during the crisis, generating credit crunch phenomena from which small and medium firms (SMEs) in particular have suffered, whose access to credit may be affected by regulation (Vozzella and Gabbi 2020).

Our contribution aims to examine the extent to which the integration of qualitative variables into default risk estimation models can be useful for providers of financial resources.

There is interesting research that highlights the ability of soft variables to approximate future business dynamics (Brunner et al. 2000; Morales et al. 2000; Grunert et al. 2005), management plans and companies' perspectives. As several studies (Lehmann 2003; Grunert et al. 2005; Godbillon-Camus and Godlewski 2005) show, one of the main problems in assessing the credit risk of loans to SMEs is the opacity of the information that banks can process.

The use of forward looking information allows SMEs to rebalance information asymmetries compared to large companies and reduce credit crunch risk (Grunert and Norden 2012; Howorth and Moro 2012).

This is not the first study dedicated to the implementation of risk management internal models with both quantitative and qualitative inputs (Mistrulli and Casolaro 2008; Agarwal and Hauswald

2010; Bartoli et al. 2013). Some contributions showed that banks that had already implemented this information in the models were medically more resilient in terms of credit portfolio during the financial crisis of 2007–2008 (Beccalli et al. 2009).

Our contribution differs from those previously published because we propose to develop a solution for the integration of qualitative variables that harmonizes the database typically used by risk managers and allows a more robust integration of the results obtained. This imposes a process of soft information hardening that often weakens the construction of adequate risk estimation models.

Once the problem of information collection has been solved, which is a privilege in particular of banks that manage to strengthen a close relationship with loan beneficiaries (Gabbi et al. 2020), knowledge must be transformed into statistically negotiable data.

The problem of how to transform soft variables into hard variables has been debated for some time: for Petersen (2004) this information cannot be transformed into a number without losing its implicit content. Other researchers propose solutions based on the use of fictitious variables (Keasey and Watson 1987) or by giving the answers certain scores (Grunert et al. 2005). The solution we propose is instead able to maintain the informative power of soft variables in terms of approximation of the probability of default.

The structure of our paper is as follows: Section 2 contains the description of the database, the sampling design and the methodology; Section 3 shows and comments our empirical findings. In Section 4, we conclude.

2. Data, Research Design and Methodology

The research is based on an archive of information recovered from 16,850 Italian companies, belonging to the sectors of industry (about 30%), construction (about 10%) and services (about 60%).

The survey conducted on Italian companies has a relevant value since it is a representative sample of a sector of companies that is more relevant for the economy of the country than elsewhere, accounting for about 98.5% of the total number of productive organizations. The company information was collected by a large Italian commercial bank operating throughout Italy and whose internal credit risk model has been validated by the national regulator. The Bank carries out commercial and credit activities, focused on the distribution of credit products and services; it also operates in the various segments of banking and financial activities, from traditional banking to special credit, asset management, bancassurance and investment banking. It is a retail-oriented bank, with about 30,000 employees and over 3000 branches.

The enterprises were divided into micro enterprises (ME) and small and medium firms (SMEs) following the EU Recommendation 2003/361, according to which the characterizing elements are the workforce and turnover. The thresholds refer to turnover (ME ≤ EUR 2 million; SME ≤ EUR 50 million) and workforce (ME < 10 employees; SME < 250 employees). For the definition of the default, has been adopted the one contained in the EU Regulation n. 575/2013 art. 178. The geographical and sectorial data show a diffusion of enterprises to represent the Italian economy in a sufficiently complete way.

We collected data from 2008 until 2013 to detect the robustness of risk management models during the most critical years after the beginning of the financial crisis and the public debt shock of the Eurozone. The observation of default events (in our database 578 companies) allowed us to compute a default rate of 3.43% (Table 1).

The information used in the study is of two types: on the one hand, quantitative information based on information processed within an internal rating based model validated by the national supervisory authority (Bank of Italy). Since the model is validated by the Italian Banking Supervisory Authority (Bank of Italy), it is confidential and cannot be published in its analytical structure. Nevertheless, it can be specified that the quantitative score is obtained with a logit model that elaborates the income, equity and financial ratios published in company's financial reports, together with internal and external historical credit behaviors. On the other hand, qualitative information were collected directly from debtor companies through a questionnaire whose characteristics are described in Table 2.

The qualitative data were collected by the bank's relationship managers, using a standard questionnaire designed by the Authors and shared with the Bank.

Table 1. Database features.

Companies' Size	Number of Observations	Number of Default Firms	Probability of Default
Micro Companies (ME)	11,301	377	3.34%
Small Medium Enterprises (SME)	5549	201	3.62%
Total	**16,850**	**578**	**3.43%**

Table 2. Strategic and Business Variables.

Scorecard Category	Variable	Issue	Answers
Internal rocess	PROD_DIV	Productive diversification	a—the company operates in more than one industry
			b—the company operates in one industry with flexible processes
			c—the company operates in one industry with inflexible production processes
Customers	BUS_DIV	Business diversification	a—high diversification of customers and turnover
			b—good diversification of turnover and few big customers
			c—low sales diversification to few big customers
Internal process	MANAG_EX	Management experience	a—>10 years
			b—>5 years and <10 years
			c—<5 years
Learning	STRAT_VIS	Strategic vision and management quality	a—outstanding vision with long term goals and excellent track record
			b—good vision with long term goals and adequate track record
			c—satisfactory vision with mixed previous results
			d—poor vision with former hazardous decisions
Internal process	ORG_STR	Organizational structure	a—organization with efficient processes and well-structured operations
			b—structured and managed managerial positions
			c—inadequate organizational structure with some ungoverned management positions
Customers	MARK_TREND	Demand side trend	a—growing
			b—reinforcing
			c—stable
			d—declining
Learning	R&D	The firm invests in research and development	a—yes
			b—no
Learning	COMP	Competition intensity	a—low
			b—medium
			c—high

Table 2. *Cont.*

Scorecard Category	Variable	Issue	Answers
Internal process	QUAL_CERT	Quality certification	a—the company has obtained quality awards or accreditations
			b—the company has applied for one or more quality accreditations
			c—the company never applied for any quality certification or failed to achieve it
Financial	BANK_REL	Bank-firm relationship	a—the company gets more than enough credit lines
			b—the company gets enough credit lines, few overdrafts were recorded
			c—some illiquid events were recorded
			d—frequent unauthorized financial overdrafts have been recorded
Financial	BOR_REQ	Credit requirements outlook	a—consistent with the company's income and equity dynamics
			b—not in line with the company's income and equity dynamics

Table 2 shows the questions contained in the questionnaire, broken down by Balanced Scorecard category (Kaplan and Norton 1992). The customer perspective was used to verify the company's ability to meet demand requirements. To this end, we did collect the information of business diversification (BUS_DIV) and market prospects (MARK_TREND). This information has been considered in relation to the size of the companies and therefore calibrated on the basis of average Italian data.

The perspective of business processes concerns the efficiency and effectiveness of operations. The variables that make this perspective explicit are: (a) the management experience of the general manager or a similar top management figure (MANAG_EX); (b) the acquisition of one or more quality certifications (QUAL_CERT) of the ISO and OHSAS type; (c) the organizational business model (ORG_STR), which depends on the degree of efficiency and structural coherence of the organization as well as the supervision of the main business functions; (d) the diversification of the products/services offered (PROD_DIV) according to the number of sectors in which it operates and the evaluation of the flexibility/rigidity of the production processes. The goal of diversification of business activities can be achieved either through a combination of outsourcing processes or through unrelated business lines (Ferretti et al. 2019). The third dimension taken into consideration is one which focuses on knowledge within the organization. We have therefore gone into three elements in depth: (a) the strategy of the firm (STRAT_VIS), the intensity of competition in the market in which the company operates (COMP) and research and development (R&D). Finally, the fourth dimension is the financial dimension considered from a qualitative and relational point of view with lenders. The variables analyzed are the internal perception of financial availability with respect to growth prospects (BOR_REQ) and any financial tensions that may have emerged in the bank-business relationship (BANK_REL).

In the estimation models, the dependent variable (default; non default) assumes a value of 1 if the company went into default during the analysis period and 0 if it is not. On the basis of each answer given in the questionnaire, the frequencies of the defaults were estimated and, consequently, the distance from the average default rate of 3.4%, a rate of non-performing loans observed in 2007 in Italy. Our dataset consists of 16,850 questionnaires completed during the years 2008 to 2012. In detail, during this period, each company that has made a loan application (and has been granted) the relationship manager of the same has completed the questionnaire. During the 12 months following the granting of the credit line, and thus the conclusion of the questionnaire, it was observed whether the company had gone into default or not. The latter aspect constitutes the variable depending on the models for estimating and calculating the default rates. Therefore, the probability of default estimated by the logistic models has a time horizon of one year.

Each cluster of responses has its own PD, which has been compared with the system average: as expected, in almost all cases a growing PD is observed as the quality of the response worsens. Our method gives weight to this difference. The weight is the distance between the PD of the cluster of each response and the average PD of the Italian economic system.

These values are reported in Table 3.

Table 3. Strategic Variables.

Variables	Average PD by Answer (%)				Distance from Sample Average PD (%)			
	a	b	c	d	a	b	c	d
PROD_DIV	3.9	3.1	3.9		0.5	−0.3	0.5	
BUS_DIV	2.8	3.7	5.6		−0.6	0.3	2.2	
MANAG_EX	3.1	4.4	12.4		−0.3	1.0	9.0	
STRAT_VIS	0.8	2.4	5.3	16.8	−2.6	−1.0	1.9	13.4
ORG_STR	2.0	4.0	16.7		−1.4	0.6	13.3	
MARK_TREND	3.2	2.9	4.4	6.8	−0.2	−0.5	1.0	3.4
R&D	3.4	3.3			0.0	−0.1		
COMP	1.6	3.0	4.0		−1.8	−0.4	0.6	
QUAL_CERT	2.8	3.6	3.7		−0.6	0.2	0.3	
BANK_REL	0.8	1.5	8.0	13.3	−2.6	−1.9	4.6	9.9
BOR_REQ	2.7	11.2			−0.7	7.8		

Numbers are the distance between the average PD for each group of the responses (a–d in Table 2) and the average PD of the whole database (3.4%).

In order to validate the use of soft information, after hardening it, we have estimated correlation coefficients (Table 4). The correlation coefficients have been computed using the differences of PDs as shown in Table 3, after hardening the soft information. The low coefficients confirm the robustness of the analytical framework of the balanced scorecard that intercepts mutually independent dimensions.

Table 4. Correlation Matrix for Soft Facts.

Correlation Matrix	PROD _DIV	BUS _DIV	MANAGE _EX	STRAT _VIS	ORG _STR	MARK _TREND	R&D	COMP	QUAL _CERT	BANK _REL	BOR _REQ
PROD_DIV	1.00										
BUS_DIV	0.05	1.00									
MANAGE_EX	0.03	0.05	1.00								
STRAT_VIS	0.02	0.06	0.06	1.00							
ORG_STR	0.02	0.04	0.05	0.46	1.00						
MARK_TREND	0.05	−0.02	−0.02	0.18	0.17	1.00					
R&D	0.11	−0.05	0.08	0.04	0.06	0.08	1.00				
COMP	−0.05	0.06	0.04	−0.04	−0.04	−0.10	−0.05	1.00			
QUAL_CERT	0.09	−0.08	0.11	0.06	0.09	0.07	0.37	−0.06	1.00		
BANK_REL	0.01	0.07	0.06	0.39	0.32	0.12	0.02	−0.03	0.05	1.00	
BOR_REQ	0.02	0.02	−0.01	0.24	0.18	0.14	−0.01	−0.04	−0.01	0.23	1.00

The default probability models using quantitative and qualitative variables were run with logit regressions, where the dependent variable is the DEFAULT of borrowers. In this model, we checked for liquidity, economic and equity characteristics recovered from financial reports (QUANT), to which we added qualitative variables (QUAL).

The research compared the quality of the estimates of three empirical models: in the first model the probability of default was estimated only with quantitative information, taken essentially from financial reports; in the second model the probability of default was, on the contrary, estimated only with qualitative information; finally, the third model integrated the independent quantitative and qualitative variables. The latter model was further detailed to understand which of the four dimensions of the balanced scorecard provides the most significant contribution to explain the dependent variable.

3. Results

We reported the empirical evidence in order to compare the quality of the results of the different models (Tables 5 and 6). Subsequently, in order to highlight the robustness of the models with indicators more consistent with the aims of risk management, we estimated the errors of the first and second type (Table 7) comparing them with what emerged from previous studies (Table 8).

Table 5. Logit Regression.

Variables	MICRO Model 1	SME Model 1	MICRO Model 2	SME Model 2	MICRO Model 3	SME Model 3
Constant	−3.827 *** (0.0644)	−3.798 *** (0.0908)	−3.789 *** (0.0932)	−3.686 *** (0.123)	−3.971 *** (0.0969)	−3.878 *** (0.128)
QUANT	10.19 *** (0.521)	17.02 *** (1.196)			6.682 *** (0.551)	10.41 *** −1.223
BUS_DIV			15.64 ** (6.603)	29.20 *** (9.041)	13.24 * (6.958)	26.88 *** (9.563)
MARK_TREND			8.115 * (4.322)	−5.051 (6.898)	7.509 (4.572)	−0.234 (7.090)
PROD_DIV			28.58 * (15.02)	−10.86 (21.47)	30.61 ** (15.57)	−13.31 (22.35)
MANAG_EX			15.25 *** (2.415)	9.048 * (4.785)	15.24 *** (2.494)	4.629 (5.507)
ORG_STR			2.708 (1.782)	2.241 (2.660)	3.493 * (1.901)	4.165 (2.726)
QUAL_CERT			29.13 ** (14.54)	31.18 * (18.85)	38.77 ** (15.27)	34.67 * (19.79)
STRAT_VIS			4.659 *** (1.473)	6.549 *** (2.043)	2.534 ** (1.611)	5.304 ** (2.156)
R&D			189.3 (132.8)	19.66 (182.6)	202.3 (137.9)	81.17 (192.2)
COMP			24.29 *** (9.215)	26.36 ** (13.03)	27.91 *** (9.601)	20.95 (13.41)
BANK_REL			15.99 *** (1.082)	18.98 *** (1.463)	11.54 *** (1.208)	13.26 *** (1.696)
BOR_REQ			7.804 *** (1.676)	3.952 (2.518)	5.400 *** (1.817)	2.996 (2.650)
Observation	11,301	5549	11,301	5549	11,301	5549
LR Chi-squared	433.200	245.686	501.464	286.239	664.263	369.883
Prob. > Chi-squared	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.131	0.142	0.152	0.166	0.201	0.214

The table reports the estimates of the logit regressions. The dependent variable is a dummy with value 1 if the firm is in default, zero otherwise. The table shows the coefficients and statistics used to assess their significance. Model 1 estimates firms' default running only the quantitative variables (QUANT). Model 2 estimates firms' default running the qualitative facts (QUAL). Model 3 estimates firms' default running all the explanatory variables (both QUANT and QUAL). Statistics to evaluate the goodness of fit of the regression are the Likelihood Ratio (LR), Chi-squared and Pseudo R2 are computed. *, **, *** denotes a statistical significance at 10, 5, and 1 percent levels, respectively; the standard error is in brackets.

A first result worth highlighting is related to the pseudo R^2 test, which appears higher when using only qualitative variables compared to the model based only on quantitative measurements. This evidence seems to justify the hypothesis that qualitative variables are fundamental and often more forward looking than quantitative ones. The pseudo-R^2 and Chi-square LR tests of model 4 in Table 6 (determined on the basis of the variables related to the quality of the company's financing conditions) are higher than the other models, both for SMEs and micro enterprises. Further evidence suggests that the model based only on qualitative variables offers better results than the model entirely calibrated on quantitative quantities; this is the case for both size categories considered. This result appears unexpected in view of the fact that banking practice is to estimate the probability of default by prioritizing balance sheet information and past behavior, which in our model are summarized in the variable QUANT. More precisely, the pseudo-R^2 and Chi-square LR tests of Model 1 are lower than the other models (0.152 vs. 0.131 and 0.166 vs. 0.142).

With regard to the independent variables selected to estimate the models, it emerges that business diversification (BUS_DIV) shows a sign consistent with that expected and a significant contribution in all regressions.

The dynamics of the industry (MARK_TREND) does not contribute significantly to explain borrowers' quality (Table 5), but contributes significantly to the explanation of the corporate default for all company sizes analyzed (in particular for micro enterprises) within the customer-based model (Table 6, Model 3a). This confirms that the environment where firms compete with each other explains the achievability of the business decisions and their resilience in case of stress (Anderson and Narus 2005).

With regard to internal controls, PROD_DIV is significant for MICRO companies. This depends on the fact that a medium-sized company is more frequently able to diversify its production processes, as well as having greater ease in surviving even by dedicating itself to a single economic sector, with the possibility of becoming a leader and integrating horizontally or vertically (Penrose 1959).

The managerial skills (MANAG_EX) are statistically relevant, especially for micro enterprises. This is explained by the fact that as the size of the company shrinks, management experience becomes a critical factor in determining the resilience of companies.

As demonstrated by Flynn et al. (1995), mainly smaller companies with an efficient quality management system not only ensure better earnings performance (Samson and Terziovski 1999) but also more stable financial balances (Kaynak 2003). Companies claiming to have achieved quality certification (QUAL_CERT) are more resilient, regardless of size. Our findings support the literature exploring the correlation between market advantages and customer satisfaction (Anderson et al. 1995). The last factor related to internal controls (ORG_STR) is more sensitive and significant for SMEs; our finding depends on the de-structured organization of micro companies.

The survey shows that the variable STRAT_VIS is much more sensitive for SMEs: this would show that the robustness of a small business requires a more robust strategy (Miles et al. 1978; Porter 2001).

An element that, on the contrary, does not appear statistically significant in the models is the contribution of research and development (R&D): the reason for this is that in the analyzed database, it is a measure that shows very little variability for micro enterprises and SMEs.

A decisive factor for the profitability of companies is the possibility that a sector may be particularly attractive to competitors (Porter 1987). As suggested by Montgomery and Porter (1991), we focus on the conflicts among agents competing in the market. Our variable (COMP) shows that competition adversely affects the resilience of companies, even more significantly for micro-enterprises, due to their structural weakness when competing with major players.

The last dimension considered is the financial dimension for which two questions were asked to companies, respectively on the quality of the bank-business relationship and the perception of capital needs (BANK_REL and BOR_REQ). From the results that emerged, this is the most significant dimension that can be used to explain the probability of default of enterprises, in line with what Castelli et al. (2006) stated for SMEs.

The ability to balance the company's cash flow and the relative perception of financial needs (BOR_REQ) is also decisive in the model, especially for micro-enterprises. In this case, the borrowers best able to manage treasury and optimize the financial structure are clearly distinguished in the model for estimating insolvency (Hamilton and Fox 1998).

Table 6. Regression of Model 3.

Variables	MICRO MOD_3a	SME MOD_3a	MICRO MOD_3b	SME MOD_3b	MICRO MOD_3c	SME MOD_3c	MICRO MOD_3d	SME MOD_3d
Constant	−3.839 *** (0.0652)	−3.810 *** (0.0920)	−3.913 *** (0.0686)	−3.781 *** (0.0942)	−3.845 *** (0.0894)	−3.847 *** (0.112)	−4.011 *** (0.0740)	−3.921 *** (0.101)
PD	9.950 *** (0.526)	16.77 *** (1.196)	9.719 *** (0.522)	16.49 *** (1.201)	9.308 *** (0.537)	15.54 *** (1.182)	6.793 *** (0.540)	10.74 *** (1.203)
BUS_DIV	15.76 ** (6.752)	23.60 *** (9.113)						
MARK_TREND	19.24 *** (4.280)	16.97 *** (6.415)						
PROD_DIV			35.39 ** (15.37)	−3.942 (22.02)				
MANAG_EX			15.12 *** (2.399)	5.149 (5.458)				
ORG_STR			10.01 *** (1.557)	12.74 *** (2.100)				
QUAL_CERT			33.35 ** (14.49)	32.46 * (17.94)				
STRAT_VIS					9.464 *** (1.271)	11.50 *** (1.673)		
R&D					50.29 (130.0)	−28.24 (175.6)		
COMP					29.52 *** (9.471)	21.86 * (13.20)		
BANK_REL							12.80 *** (1.166)	14.69 *** (1.620)
BOR_REQ							7.233 *** (1.683)	6.073 ** (2.424)
Observation	11,301	5549	11,301	5549	11,301	5549	11,301	5549
LR Chi-squared	456.379	257.502	516.750	279.060	493.117	289.109	589.567	341.826
Prob. > Chi-squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.138	0.149	0.156	0.161	0.149	0.167	0.178	0.198

The table reports the estimates of Model 3, run separating qualitative variables by the BSC perspectives, that is customers (Model 3a); business processes (Model 3b); learning and growth (Model 3c); and financial perspectives (Model 3d). The dependent variable is a dummy with value 1 if the firm is in default, zero otherwise. The table shows the coefficients and statistics used to assess their significance. Statistics to evaluate the goodness of fit of the regression are the Likelihood Ratio (LR), Chi-squared and Pseudo R^2 are computed. *, **, *** denotes a statistical significance at 10, 5, and 1 percent levels, respectively; the standard error is in brackets.

The statistical test estimates in Tables 5 and 6 do not fully explain the goodness and robustness of a model for risk management, particularly credit risk. To this end, it is essential to estimate classification errors, distinguishing between forecasts of healthy companies compared to companies destined to default. The first type of error occurs when the model deems a company to be healthy and ex-post is observed to be in default; vice versa, the second type of error occurs when the model classifies a company that ex-post remains healthy as not performing.

The estimation of errors was based on the average default probability threshold of each enterprise group: for micro enterprises it was 3.3% and for SMEs it was 3.6%. When the model estimates an enterprise, based on the independent variables inserted in the model, lower than the average PD, the classification is that of a performing enterprise, vice versa the enterprise is classified as non-performing. Table 7 shows the errors and the correct classification rate.

Table 7. Error Estimation.

PANEL A—MICRO							
Statistics	**Model 1**	**Model 2**	**Model 3**	**Model 3a**	**Model 3b**	**Model 3c**	**Model 3d**
Type I error ratio	34.7%	31.8%	30.2%	40.8%	39.3%	38.7%	32.6%
Type II error ratio	12.9%	21.2%	18.8%	15.5%	11.2%	11.6%	17.9%
Correct Classification Rate	76.2%	73.5%	75.5%	71.8%	74.8%	74.8%	74.8%
PANEL B—SME							
Statistics	**Model 1**	**Model 2**	**Model 3**	**Model 3a**	**Model 3b**	**Model 3c**	**Model 3d**
Type I error ratio	39.3%	29.9%	25.9%	27.9%	32.8%	32.3%	25.9%
Type II error ratio	8.8%	19.0%	17.5%	19.1%	12.9%	13.4%	17.7%
Correct Classification Rate	75.9%	75.6%	78.3%	76.5%	77.1%	77.1%	78.2%

This table shows the ratios of the two types of classification errors and the correct classification rate in both the sample and out of sample analysis. The type I error ratio is the percentage of default firms classified as performing firm by the model (logit), while the type II error ratio is the percentage of performing firms classified as default firm by the model (logit). The Correct Classification Rate is determined as the weighted average between the correct classification rate of performing firms (complementary to one of the first type error ratio) and the correct classification rate of default firms (complementary to one of the second type error ratio). The discriminating cut-off corresponds to the threshold for each size cluster, respectively, 3.3% and 3.6% for micro and SMEs. Firms with a PD lower than the threshold level were classified as performing, and vice-versa, firms with a PD higher than the threshold were classified as non-performing.

To make the quality of our research results more evident, we have collected some of the main contributions on credit risk, by originality of the contribution and by the use, albeit in a different way from our approach, of soft and hard information. In Table 8, we have included the authors, the methodology used, classification errors and, finally, the correct classification rate.

Table 8. Classification Errors in Previous Literature.

Authors	**Methodology**	**Type I Error Ratio**	**Type II Error Ratio**	**Correct Classification Rate**
Altman (1968)	DA with old FR	28.00	6.00	83.00
Keasey and Watson (1987)	LG with FR, IS	23.30	23.30	76.70
	LG with QV, IS	24.70	24.70	75.30
	LG with FR and QV, IS	17.80	17.80	82.20
	LG with FR, OS	70.00	20.00	55.00
	LG with QV, OS	30.00	40.00	65.00
	LG with FR and QV, OS	20.00	50.00	65.00
Grunert et al. (2005)	PR with FR	55.81	4.25	69.97
	LR with FR and QV	41.86	4.25	76.95
Altman et al. (2010)	LG with old FR, OS	24.00	27.00	74.50
	LG with old QV, OS	24.00	25.00	75.50
	LG with new FR, OS	23.00	27.00	75.00
	LG with new FR and QV, OS	20.00	24.00	78.00

This table compares some relevant contributions to credit risk estimates, based on quantitative, qualitative or blended databases. Some of the qualitative variables are strategic factors. The first column shows the authors and the year of publication of their work. The second column shows the statistical methods used to determine the correct classification errors specifying whether the analysis and what data types used (financial, non-financial or both types). The last two columns show the Type I and the Type II error ratios, respectively. LR stands for linear regressions, LG stands for logit regression, PR, stands for probit regression, IS, in-sample; OS, out of sample. All values are percentage. The first column contains the bibliographical references of the study analyzed; the second column summarizes the methodology used to estimate the credit risk model (DA = discriminant analysis; LG = logit regression; PR = probit regression; IS = in-sample; OS = out of sample) and the typology of independent variables used (FR = financial ratios; SV = soft variables).

The variables used in our model, consistent with the dimensions of the balanced scorecard (Table 7), compared with the outcomes observed in the previous literature (Table 8) show a more robust contribution to explain the probability of default. This conclusion is supported by the consideration of the different cost associated to non-performing loans (type I) compared to the opportunity cost embedded in type II errors.

4. Conclusions

Our research aims to propose an approach to harden soft variables and to integrate them within credit risk models. We run the empirical analysis for micro, small and medium companies, showing that utilizing strategic perspectives approximated via soft variables, such as the managerial skills, business diversification, vision, competition intensity, quality accreditation, the bank-firm relationship and borrowing requirements expectations, it is feasible to estimate credit risk and by association the survival probability of firms (Gabbi et al. 2019).

Our research proposes a solution to the problem of soft information hardening, the selection of the same soft variables within a robust theoretical framework and the integration of quantitative and qualitative dimensions into the credit risk estimation models of micro and SME enterprises.

Selecting qualitative variables with scorecard logic allows us to interpret the results in a more consistent way with the strategies adopted by companies and make the collection process more standard.

Moreover, the hardening model is no longer discretionary but objective and based on the link between subgroups and their default probabilities.

With reference to the regulatory implications, our results offer a contribution to the debate on the revision of internal models (Basel IV), suggesting a validation that is also a function of the completeness of the information used by banks to estimate the risk weighted assets and consequently the capital absorption.

We hope that, with the aim of optimizing credit risk estimation and subsequent capital allocation (Gabbi and Levich 2019), the revision of internal models (Basel IV) will also further refine the credit risk estimation on the basis of our findings.

Author Contributions: Conceptualization, M.M. and G.G.; methodology, M.G.; validation, G.G.; formal analysis, M.M.; investigation, M.M.; resources, M.M.; data curation, M.M.; writing—original draft preparation, M.M. and M.G.; writing—review and editing, M.M.; supervision, G.G.; project administration, M.M.; data acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Agarwal, Sumit, and Robert Hauswald. 2010. Distance and private information in lending. *The Review of Financial Studies* 23: 2757–88. [CrossRef]

Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609. [CrossRef]

Altman, Edward I., Gabriele Sabato, and Nick Wilson. 2010. The value of non-financial information in SME risk management. *SSRN Electronic Journal*. [CrossRef]

Anderson, James C., and James A. Narus. 2005. *Business Market Management: Capire, Creare, Fornire Valore*. Milan: Etas Libri.

Anderson, John C., Manus Rungtusanatham, Roger G. Schroeder, and Saravan Devaraj. 1995. A path analytic model of a theory of quality management underlying the Deming management method: Preliminary empirical findings. *Decision Sciences* 26: 637–58. [CrossRef]

Bartoli, Francesca, Ferri Giovanni, Pierluigi Murro, and Rotondi Zeno. 2013. SME financing and the choice of lending technology in Italy: Complementarity or substitutability? *Journal of Banking & Finance* 37: 5476–85.

Beccalli, Elena, Paola Bongini, and Arturo Patarnello. 2009. Modeli di intermediazione delle banche europee e la valutazione del mercato. *Osservatorio Monetario-Assbb* 2: 23–40.

Brunner, Antje, Jan Pieter, and Martin Weber. 2000. Information Production in Credit Relationship: On the Role of Internal Ratings in Commercial Banking. Center for Financial Studies Working Paper, Center for Financial Studies, Frankfurt am Main, Germany. vol. 10.

Castelli, Annalisa, Gerald P. Dwyer Jr., and Iftekhar Hasan. 2006. Bank Relationships and Small Firms' Financial Performance. Working Paper, No. 2006-05, Federal Reserve Bank of Atlanta, Atlanta, GA, USA.

Ferretti, Camilla, Giampaolo Gabbi, Piero Ganugi, Federica Sist, and Pietro Vozzella. 2019. Credit Risk Migration and Economic Cycles. *Risks* 7: 109. [CrossRef]

Flynn, Barbara B., Roger G. Schroeder, and Sadao Sakakibara. 1995. The impact of quality management practices on performance and competitive advantage. *Decision Sciences* 26: 659–91. [CrossRef]

Gabbi, Giampaolo, and Richard Levich. 2019. Controlling risks to ensure financial stability and reducing volatility. *Journal of International Financial Management & Accounting* 30: 183–87.

Gabbi, Giampaolo, and Andrea Sironi. 2005. Which factors affect corporate bonds pricing? Empirical evidence from eurobonds primary market spreads. *The European Journal of Finance* 11: 59–74. [CrossRef]

Gabbi, Gabbi, and Pietro Vozzella. 2013. Asset Correlation and Bank Capital Adequacy. *European Journal of Finance* 19: 55–74. [CrossRef]

Gabbi, Giampaolo, Massimo Matthias, and Michele Giammarino. 2019. Modeling Hard and Soft Facts for SMEs. Some International Evidence. *Journal of International Financial Management and Accounting* 30: 203–22.

Gabbi, Giampaolo, Michele Giammarino, Massimo Matthias, Stefano Monferrà, and Gabriele Sampagnaro. 2020. Does face-to-face contact matter? Evidence on loan pricing. *The European Journal of Finance* 26: 820–36. [CrossRef]

Godbillon-Camus, Brigitte, and Christophe J. Godlewski. 2005. Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation. Working Paper, University of Strasbourg, Strasbourg, France.

Grunert, Jens, and Lars Norden. 2012. Bargaining power and information in SME lending. *Small Business Economics* 39: 401–17. [CrossRef]

Grunert, Jens, Lars Norden, and Martin Weber. 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance* 29: 509–31.

Hamilton, Robert T., and Mark A. Fox. 1998. The financing preferences of small firm owners. *International Journal of Entrepreneurial Behavior & Research* 4: 239–48.

Howorth, Carole, and Andrea Moro. 2012. Trustworthiness and interest rates: An empirical study of Italian SMEs. *Small Business Economics* 39: 161–77. [CrossRef]

Kaplan, Robert S., and David P. Norton. 1992. The Balanced Scoreboard-Measures that Drive Performance. *Harvard Business Review* 70: 71–79.

Kaynak, Hale. 2003. The relationship between total quality management practices and their effects on firm performance. *Journal of Operations Management* 21: 405–35. [CrossRef]

Keasey, Kevin, and Robert Watson. 1987. Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses. *Journal of Business Finance & Accounting* 14: 335–54.

Lehmann, Bina. 2003. Is it worth the while? The relevance of qualitative information in credit rating. The Relevance of Qualitative Information in Credit Rating. Paper Presented at the EFMA Annual Meetings, Helsinki, Finland, June 25–28; pp. 1–25.

Miles, Raymond E., Charles C. Snow, Alan D. Meyer, and Henry J. Coleman. 1978. Organizational strategy, structure, and process. *The Academy of Management Review* 3: 546–62. [CrossRef]

Mistrulli, Paolo E., and Luca Casolaro. 2008. *Distance, Lending Technologies and Interest Rates*. Rome: Bank of Italy.

Montgomery, Cynthia A., and Michael E. Porter. 1991. *Strategy: Seeking and Securing Competitive Advantage*. Boston: Harvard Business School Press.

Morales, Ann, Rene Sacasas, and Paul Munter. 2000. 'Safe Harbor' under the Private Securities Litigation Reform Act of 1995. *The CPA Journal* 70: 66.

Penrose, Edith. 1959. *The Theory of the Growth of the Firm*. Oxford: Blackwell.

Petersen, Mitchell A. 2004. Information: Hard and Soft. Working Paper, Kellogg School of Management at Northwestern University, Evanston, IL, USA. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.8246&rep=rep1&type=pdf (accessed on 12 May 2020).

Porter, Michael. 1987. *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: The Free Press.

Porter, Michael. 2001. Strategy and the Internet. *Harvard Business Review* 79: 62–78. [PubMed]

Samson, Danny, and Mile Terziovski. 1999. The relationship between total quality management practices and operational performance. *Journal of Operations Management* 17: 393–409. [CrossRef]

Vozzella, Pietro, and Giampaolo Gabbi. 2020. What is good and bad with the regulation supporting the SME's credit access. *Journal of Financial Regulation and Compliance*. [CrossRef]