Probability of Default Estimation for Construction Firms

Marco Muscettola[1]

[1] Credit risk manager, Independent researcher, Italy

Correspondence: Marco Muscettola, Via Generale Scattaglia, 30/B, 70010 Adelfia (Ba), Italy. E-mail: marcomuscettola@hotmail.com

Abstract

The purpose of this work is to create one statistical model for a future behaviour of construction companies. The model, therefore, aims to quantify the potential for progress and durability of construction companies operating in the Italian territory, as well as may assign ratings. Paper summarizes the typical rating process with the identification of definite financial paradigms suitable to rank the companies by solvency criteria in scale credit ratings.

Keywords: construction firms, default estimation, modelling credit risk

1. Introduction

The purpose of this work is to create a new predictive model of construction firms by lengthening the time frame of the projections, to better adapt to the specific undertakings of the statistical predetermined objectives. The model, therefore, aims to quantify the growth potential and durability of Italian construction companies, to which will be assigned a rating. The focus will be both in the identification of financial paradigms that are able to order the companies to solvency and the investigation of connections between capital structure and all the variables that affect the profitability and, in general, the lifetime of the firm.

The sample taken in this examination is based on financial statements of 1,338 Italian construction firms with a percentage, within the sample, more than 10% of insolvency cases.

The paper contributes to the existing literature on probability of default estimation for two reasons: because the analyzed firms are construction companies, and because the time horizon of analysis includes the financial statements for the three years preceding the assignment of the judgment of risk.

Construction firm is particular to see, in fact, especially for the production cycles and cash conversion cycles that last more than a year accounting, and, therefore, is absolutely not easy to achieve the expression of creditworthiness limiting the analysis to a single annual financial statements. The production process in the construction industry is also distinguished by the use of moderately low technology and high intensity of labor. The other peculiarity is related to the historical moment of economic downturn that Italy, specifically the present research, is experiencing. This economic crisis is felt particularly in the construction sector. That segment is a good trolling motor for the economy during the economic expansion but it can be a really bad sector in times of economic difficulty, like a litmus test of the eventual trend of the times.

For direct and indirect reflections on the economy of the territory, for the absence of previous empirical studies of this type of firms as well as for the absence of literature on analysis of the financial budget of construction firms except for some few studies carried out in non-EU countries, this paper aims to turn cues reflection for future in-depth theoretical and practical.

The work is divided into seven parts over the introduction: Second paragraph shows the definitions of research, presents the sample used and the statistical methodology. Third section lists the most important works on the topic. Fourth section describes the construction firms used in the sample by descriptive analysis illustrating how firms are structured. Fifth section describes the empirical analysis. With the use of logistic regression function, the statistical model will be able to distinguish companies that will become insolvent by healthy firms. Sixth section shows the results of the analysis, the validation's model and the distribution of insolvent firms within the rating scale. Last paragraph is the concluding remarks of the paper.

2. Literature Review

Certain exacting distinctiveness of the construction industry makes it significantly different from other sectors, mainly the fact that its final result shows uniqueness, large assets and long cash conversion cycle. Construction output is the consequence of a project production system in which the result is tailored to demanding locations. The same value of buildings constructed is strongly influenced by demand and place. For these reasons, the success of a construction company depends a lot on factors unrelated to the accounting business and this inimitability of the output disheartens the use of analysis of financial statements. Activities that construction firms carry out are heterogeneous not only in style or type but also in geographical location. For these reasons, it is difficult to analyze the value of the construction firms through a single standardized method.

On the other hand, financial ratio models, by their character, may not be able to judge every aspect of an insolvent company's distinctiveness. Subjective assessment has an important responsibility to play in predicting default in construction industry and any forecast forms should not employed as the exclusive judgment tool.

For all the motivations listed, academic studies that have analyzed the construction companies are really very few. Among the major works carried out on the construction sector, see Kangari (1988), Padget (1991), Kangari et al (1992), Russell and Jaselskis (1992), Landford et al 1993), Kale and Arditi (1999), Davidson and Maguire (2003), Koksal and Arditi (2004), Singh and Tiong (2006) or Huang (2009).

3. Data and Methodology

To achieve the purpose of research it is essential to analyze all the variables that affect the solvency of the firms and to classify them for significance, importance and usefulness. Only after doing this, it will be possible to sort the same construction firms in discrete classes pre-ordered by probability of default.

In this essay was deemed indispensable to broaden the definition of "default". In addition to the failure to fulfil a bank obligation, appearing in a law court, and piling up *past-due (Note 1)*, also some other events of insolvency arising from the imperfect payment of debts (especially to repay a loan) were included in the definition. This very extensive definition was chosen as, on the other hand, it almost entirely gets rid of the behavioural analysis of the firms.

Unlike the more common and regulated rating models, this archetype is directed to be more "forward looking" (Muscettola & Pietrovito, 2012b). In this logic the time frame is fixed at three years following the analysis even if it is estimated an evident deterioration of the outcomes in robustness, verifiability and replicability. The variables, by which has been derived the forecast of insolvency, are related to a period up to three years earlier. This point borders (but not entirely) the influence of the economic cycle on the inherent risk of the construction firm observed. In addition to the macroeconomic impacts on the companies, extending the period of analysis is best expressed the potential typical of the construction company that, in fact, has a structure of multiyear financial statements that reflect the phases of production and trade that usually continue for more than one accounting year.

The preliminary choices of the essay were oriented to have a scheme of analysis transferable (in time and space) to other samples of firms and empirically verifiable at any time. For these details it is crucial to have clear and valid sources of information and data accessibility in time (Note 2).

A more detailed consideration, however, deserves the step of organization and cleanliness of the test sample. Before starting to work on the data, therefore, it is required to clean up the dataset by eliminating from the sample firms that seem to have the data irrational or reliant on exceptional cases as the case, for example, of a start-up company or a firm controlled exclusively by another company.

Concerning the treatment of the outliers it is chosen to follow a rigid acceptance. This is due to the strong volatility of the financial statements of construction companies, although the sample is focused on firms with revenues very similar. Many atypical cases, in fact, were recognized in a sufficiently large test sample though this collection is specific of the construction industry.

It is not easy to define what data is non-standard, because there is not an universal regulation that explains the "normal" values (Muscettola & Pietrovito, 2012a). In the present pattern the distributions of the values that exceeded two times the standard deviation of the distance from the first quartile to the bottom and from the third quartile to the top have been truncated.

As regards the study of the distribution and dispersion around the mean and the median values, so that the indicators of the financial statements as explanatory variables would be accepted, a table expedient was set for understanding the asymmetry of the statistical distribution and the significance of code extremes. For some ratios

the table is also led to a comparable revision on the trend of indicators using statistical techniques such as the standardization of the elements (Note 3). By way of stress-testing, to verify the logical behaviour of firms and the functionality of variables taken, the indicators have also been assessed comparing the results with the most known and used ratios in the literature of ratings.

The companies analyzed are small, medium and large enterprises with not less than five million euro as sum of total revenues from sales of the three yearly statements under examination. The building firms included in the sample are operating in Italy for at least 8 years old (Note 4). Off the analysis sample also are all those firms that have shown significant shareholdings in other companies or, on the other hand, are dependent by a parent company.

In order to balance the sample, 1,338 construction firms were extracted, with the characteristics specified in Table 1.

Table 1. Characteristics of the sample used in the research

DATASET	Whole sample		Insolvent firms		Healthy firms	
	NR	%	NR	%	NR	%
Northern Italy	626	100.00	72	11.50	554	88.50
Central Italy	394	100.00	53	13.45	341	86.55
Southern Italy	318	100.00	46	14.47	272	85.53
Whole sample	1.338	100.00	171	12.78	1.167	87.22

The period of analysis stands from 2007 to 2011. All firms without insolvencies at least until the year 2012 (including that year) are defined "healthy firms". Have been eliminated, on the other hand, companies that have become insolvent in a year within the period of observation (Note 5). In this way it is held tightly the optical time of three years.

4. Descriptive Analysis

Descriptive analysis exposed below is smoothed with respect to certain insights. Specifically, in fact, owing to the specificity of the construction company, in some statistical distributions there is a high variance around the average, which does not allow classifying the firms into the clusters. These phenomena are very evident as regards the efficiency and rotation ratios and, not less, as regards some indices of net and operating profitability.

In order to represent the characteristics of the sample, with the remarkable distinction of the group of firms that become insolvent after three years from the healthy firms, it is preferred to observe the accounting data of firms expressed as three-year averages or sum of three years. In other words, everything that follows will concern the expression of financial ratios formed by the arithmetic mean over three years preceding the year of expectation value. It also points out that, even for insolvent firms shall not be never judged the specific year of insolvency but three previous years.

Table 2. Summary statistics: sums of three years and averages of three years for goof firms and for insolvent firms

Explanatory variables	Sums of three years		Averages of three years	
	Good firms	Bad firms	Good firms	Bad firms
Total fixed assets / Total assets %	110,92	109,06	36,97	36,35
Cash and bank deposits / Total assets %	11,28	7,92	3,76	2,64
Total receivables / Total assets %	109,79	104,91	36,60	34,97
Inventory / Total assets %	68,64	78,79	22,88	26,26
Trade receivables / Total assets %	92,57	87,15	30,86	29,05
Prepaid expenses and current financial / Total assets %	3,06	4,37	1,02	1,46
Intangible fixed assets / Total assets %	6,41	13,76	2,14	4,59
Long term liabilities/ Total assets %	45,66	48,25	15,22	16,08

Borrowings / Total assets %	102,47	127,70	34,16	42,57
Trade payables / Total assets %	76,25	72,84	25,42	24,28
Total shareholders' equity / Short term debt %	160,91	110,30	53,64	36,77
Total shareholders' equity / Total assets %	72,16	53,02	24,05	17,67
Shareholder debt / Total assets %	4,95	2,66	1,65	0,89
Quick ratio %	227,43	188,40	75,81	62,80
Total shareholders' equity / Fixed assets %	435,83	398,06	145,28	132,69
Current ratio %	368,43	335,36	122,81	111,79
Long term debts and equity / Fixed assets %	251,86	218,48	83,95	72,83
Net working capital / Total investment %	25,21	5,42	8,40	1,81
Inventory / Trade receivables %	327,46	382,32	109,15	127,44
Interest expense / Total debt %	9,34	11,82	3,11	3,94
Total debt / Sales %	263,22	347,88	87,74	115,96
Current liabilities / Total debt %	236,10	237,08	78,70	79,03
Interest expense / Financial debt %	24,25	23,50	8,08	7,83
Interest expense / Sales %	8,05	12,64	2,68	4,21
Intercompany and shareholder debt / Assets %	15,16	9,51	5,05	3,17
Debt ratio %	208,17	228,23	69,39	76,08
Account receivable turnover	20,58	13,63	6,86	4,54
Inventory turnover	19,59	12,19	6,53	4,06
Investment turnover	2,88	2,41	0,96	0,80
Trade payables turnover	10,78	14,55	3,59	4,85
Current liabilities / Sales %	179,21	235,25	59,74	78,42
Fixed assets turnover	138,44	174,10	46,15	58,03
Operating cash flow / Total investment %	14,11	7,78	4,70	2,59
Operating cash flow / Current liabilities %	26,98	15,14	8,99	5,05
Operating cash flow coverage %	46,67	28,37	15,56	9,46
Operating cash flow / Sales %	13,68	10,84	4,56	3,61
Roe %	-2,51	-36,88	-0,84	-12,29
Operating profit / Sales %	23,50	-6,23	7,83	-2,08
Ros %	11,90	-1,26	3,97	-4,21
Gross profit / Sales %	71,95	-1,56	23,98	-4,19
Ebitda / Cost for employees %	218,14	237,96	72,71	79,32
Ebitda / Interest expense %	20,50	4,79	6,83	1,60
Ebit / Total liabilities %	17,02	10,72	5,67	3,57
Cost for employees %	44,97	46,45	14,99	15,48
Ebitda / Total investment %	23,86	19,35	7,95	6,45
Depreciation and amortization / Sales %	12,57	14,72	4,19	4,91
Roi %	12,46	8,14	4,15	2,71
Roa %	10,88	7,45	3,63	2,48
Cash and bank deposits / Sales %	13,61	12,51	4,54	4,17
Net Working capital / Sales %	25,50	-2,04	8,50	-6,78
Total stakeholders' equity / Sales %	93,09	87,26	31,03	29,09
Financial leverage	7,39	11,68	2,46	3,89
Ebitda / Net financial position %	0,42	0,16	0,14	0,05
Gearing	1,58	2,16	0,53	0,72

Among the most interesting aspects to deepen there is certainly the perfect understanding of the composition of the assets. As you can easily realize, in fact, is very important for a construction firm the mixture of corporate assets and their elasticity (Muscettola, 2014a). Undoubtedly relevant might be the juxtaposed point of view of the averages of the healthy firms and of firms which then will become insolvent.

The first figure describes specifically the typical composition of the assets detailed in the financial statements opposing the pie graph refers to the sample of healthy firms to the group of insolvent companies. The graph shows that healthy firms hold fewer inventories (22.9% versus 26.3% of insolvent firms) and only a slightly greater portion of trade receivables and cash. Very similar is the impact of fixed assets to total assets for the two groups of firms.
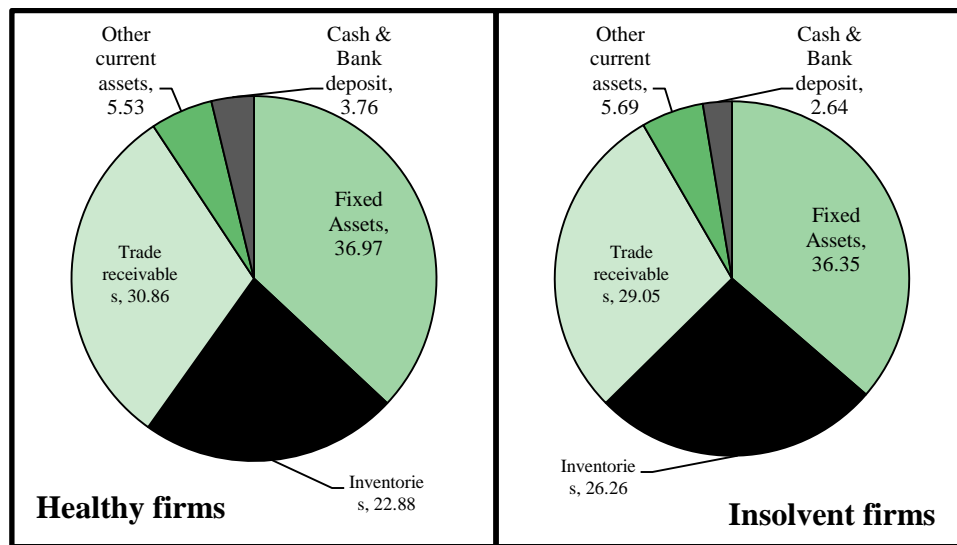

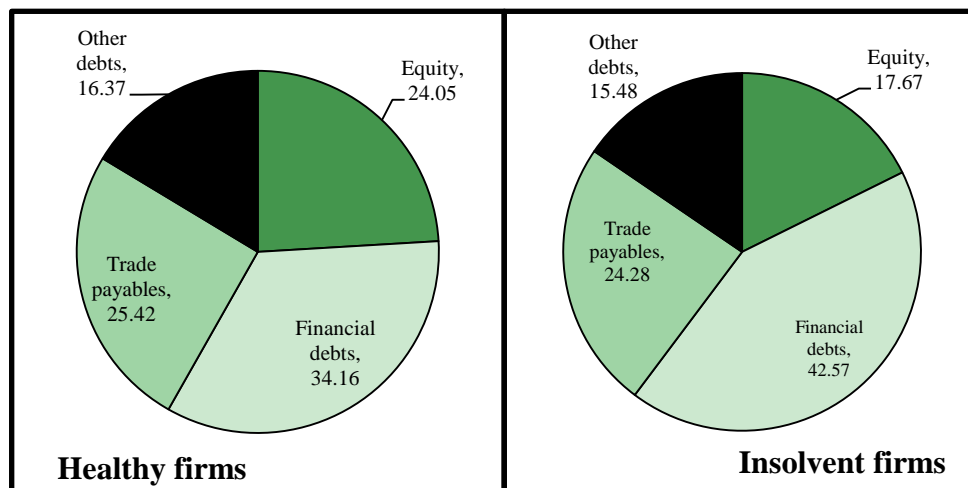Figure 1. Composition of assets

Figure 2. Capital structure

The second figure describes the capital structure of construction firms. The borrowings exceed 42% of the total sources for insolvent firms against a more acceptable 34% for healthy firms. Considering, then, that the impact of trade payables is not very different between the two types of firms, the foregoing suggests that viable firms outweigh the less use of financial debts with a larger portion of equity (the shareholders' equity is 24.0% for good firms and 17.7% for insolvent companies).

The considerations concerning the level of corporate debt are also a consequence of the type of activity. It is sufficiently obvious, in fact, that a manufacturing firm, especially in comparison with a service firm, in order to

pursue its mission (fabrication) generally need a larger amount of investment and infrastructure (Muscettola & Gallo, 2008). On the other side, therefore, the production firm will need higher capital resorting consequently to higher payables to support the heaviest cash conversion cycle (Muscettola, 2014b). Similarly, a construction firm, with, by definition, long and outsized production cycles, has the extreme necessity to obtain more capital compared to other types of firms.

Looking at the two sub-samples in the two graphs it must be noted that the insolvent firms are characterized by their relative unbalanced between rigidity of assets and level of indebtedness. In particular insolvent firms have greater financial exposure, for the same assets, compared to healthy firms, which show instead a greater liquidity of the assets in contrast to liabilities.

As regards to the capitalization of the company that will become insolvent, as noted, it is fairly logical and generally accepted that firms in difficulty have a lower strength of auto-capitalization of income created, especially if the onset of the crisis is determined by its an profitable problematic (low sales and low Ebit). On the other hand, the losses lead to a worsening of the balance sheet primarily affecting the reserves and, then, the net worth. In this respect, regardless of the cause or consequence (Note 6), excluding a few exceptions, it would seem that insolvent firms in 2012 were more indebted during previous years than firms that have shown durability.

Although financial leverage may provide more opportunities for productive investment of construction firms, it improves the probability of default in a more than proportional, as will be seen also by the empirical analysis. The strong financial indebtedness induces a whole series of negative effects that impact both on business management, and on pricing assigned by banking system.

As regards the latter, is not disputed that more debt means higher costs to be incurred in the statement of income of the company entrusted by the bank. Taking a process of generalization, then, we can say that the construction firms who have not passed the economic downturn are those who had a heavy indebtedness and, consequently, a burden of liabilities (Muscettola, 2014c) that eroded the income produced.

5. Empirical Analysis

To associate the peculiarities of the construction firms to the probability of default after three years was used the logistic regression. The functional procedure selected for this research was, therefore, the logit model. The event insolvency is the dependent variable of the model and it is binary discrete variable (Note 7) (dichotomous) that designates whether firm "i" has defaulted or not after 3 years. The explanatory variables are the indices of financial statements shaped as averages of three years, and as sums of three years before the evaluation of the dependent variable (see table 2). The other indicators (financial ratios) available in the database will serve as control variables in order to verify the correlation of distributions with the event insolvency.

In order to form a ladder of discrete values that are able to sort the construction firms for probability of default was used the statistical technique of logistic regression advance by the approach called "forward stepwise" to reduce the number and the redundancy of the independent variables (Muscettola & Naccarato, 2013). This heuristic scheme is usually used for low levels of computational complexity like the present essay. Consequently to the above, the process will determine as a first step the creation of a function with the support of each explanatory variable. Subsequently it goes on through a progression repeated until the entry of a new variable in the function does not cause an improvement in the performance of model. In this way each explanatory variable taken by the dataset of financial ratios has worked in a massive mode, starting at the most significant by itself to conclude with all the components that maximize the contribution of selection.

The last step concerns, however, the explanation of the results looking for, if it's possible, a logical interpretation of the same outputs. In this regard it should be noted that, unlike to other statistical techniques, the regression coefficients are not always able to convey a perfect rationality just because the result must be read as a summary of the contributions of all the independent variables taken, in fact, all together. The search for the meaning of logical cause and effect must be considered for both the sign of the coefficient and the significance of the contribution.

Indicating p_i as the probability that the i-th firm may join the group of the defaulting firms after three years, x_{1-108} as the set of the financial ratios used, β_{1-108} as the coefficients of the variables and k as constant, it is conceivable to compose the function of the logit model, in which the odds is a linear function of the explanatory variables and the archetypal parameters (β_{1-108}) are calculated via the maximum likelihood approximation, in the following manner:

$$p_i = f(k + \beta_1 x_1^i + \beta_2 x_2^i + \ldots + \beta_{108} x_{108}^i) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1^i + \beta_2 x_2^i + \ldots + \beta_{108} x_{108}^i)}}$$

In Table 3 there are the outcomes of the forward stepwise regression that considers the 108 financial ratios designated by determining the probability of default on the sample of the costruction firms The table exposes the coefficients (β) of the factors with a significance level contained between 1% and 5% where "*β*" is the estimated coefficient of the logit regression, "S.E." denotes the standard error of the assessed coefficients (Note 8), "wald" is the Wald chi-square test (Note 9), "sig." identifies the *p-value* of the predictable coefficient (Note 10), "Exp(*β*)" is the variation in the odds ratio related with a 1 unit alteration in the forecaster variable.

Table 3. Stepwise logistic regression: functions calculated on costruction firms

Variables	β	S.E.	Wald	Sig.	Exp(B)
Σ Intangibles / Total Assets	0.03895	0.00725	28.83311	0.00000	1.03971
Σ Return on Debt	0.15690	0.02689	34.03752	0.00000	1.16988
Σ Investment turnover	-1.31687	0.13684	92.61197	0.00000	0.26797
Σ Gross profit / Sales	-0.00001	0.00003	0.06017	0.80623	0.99999
µ Net worth / Total Assets	-0.07804	0.01032	57.17362	0.00000	0.92492
µ Current liabilities / Total debts	0.03436	0.00729	22.24067	0.00000	1.03496
µ Current liabilities / Sales	-0.01142	0.00313	13.33839	0.00026	0.98864
µ Ebitda / Interest expense	0.01940	0.00348	31.06900	0.00000	1.01959
Constant	-1.02599	0.67347	2.32085	0.12765	0.35844

The outcome of the regression, exposed in the table, expresses a function of separation of the insolvent firms after three years by firms that will remain solvent throughout the period of analysis. Considering the estimated parameters it is possible to formulate the model of bankruptcy prediction of construction firms as follows:

z = - 1,02599 + 0,03895 x (Σ Intangibles / Total Assets) + 0,15690 x (Σ Return on Debt) -1,31687 x (Σ Investment turnover) - 0,00001 x (Σ Gross profit / Sales) - 0,07804 x (µ Net worth / Total Assets) + 0,03436 x (µ Current liabilities / Total debts) -0,01142 x (µ Current liabilities / Sales) + 0,01940 x (µ Ebitda / Interest expense)

The first parameter considered in the function is the sum (Note 11) of the three years of the index formed by the relation of the intangible assets and the total assets recorded in the balance sheet ("Σ Intangibles / Total Assets"). It has a positive coefficient. This means that higher portion of intangibles on total assets leads to greater probability that the firm becomes insolvent after three years.

Below, always with a positive sign, there is ratio called "Return on Debt", formed by the quotient in percentage between interest expenses and total debt. This variable should be considered as the sum of three years and also this factor rises with increasing cases of bankruptcy.

Third factor is "investment turnover" formed by the sum of the three-year of quotient between total capital invested and revenues of sales. This variable, unlike others, is not expressed as a percentage and, mainly, has a negative coefficient. This means that when the sum of these indicators grows the probability of finding a construction firm insolvent decreases. Also for this explanatory variable there is a considerable statistical significance even if the coefficient *β* is really small.

Fourth index used in the function is the whole three-year ratio formed by corporate value added and revenues of sales ("Σ Gross profit / Sales"). Equally to the previous factor it has a negative correlation with event occurring (independent variables). In other words, the growth of the value added as a percentage of net sales decreases the probability of default of the construction firm. Unlike the previous indicators, however, the statistical validity of this parameter has been quite low, as can be seen by the relatively high assessment of the p-value.

The following indicators in the regression function are formed, instead, by the average (Note 12) of the related variables within the three years under examination. The fifth index ("µ Net worth / Total Assets"), in fact, is formed by the arithmetic mean (for each firm comparative to the time horizon of three years) of the quotient made by equity and firm's total assets recognized in the financial statements. As you might guess, in this case the

parameter has a negative correlation with the probability of default of the construction firm: greater company capitalization, therefore, follows a smaller default cases.

The sixth factor is the three-year average of the ratio formed by debt due within one year and the total debts incurred by the firm (" µ Current liabilities / Total debts"). Moreover, it also obtains a good statistical significance and it shares a justifiable and logical relationship with event studied. As provided by a positive coefficient, in fact, with the growth of short-term debt it increases, too, the risk of insolvency of the construction firm.

The seventh parameter is formed by the three-year average of the quotient designed by short-term debt and revenues of sales (" µ Current liabilities / Sales "). The statistical significance of the explanatory variable is less strong than the other indicators, with the mere exception of the rapport between value added and sales. An issue is the negative sign of the coefficient. As anticipated, in fact, it may often happen that few indicators in a logistic regression lose a part of their "predictive logic" if taken so assembled. This effect is certainly more visible than a more reasonable univariate statistics. In this sense, in fact, the logistic function designates an increased risk of insolvency when short-term debt declines in relation to the amount of sales.

The last factor of the regression function is the three-year average of the ratio between EBITDA and financial expense (" µ Ebitda / Interest expense"). In this circumstance there is a strong statistical significance and even a logical difficulty in justifying the positive coefficient of the parameter. With the increase of margin, compared to interest expenses, in fact, also increases the probability of default of the firm. High burden of debts, on the other hand, with respect to the profitability produced by the company, should be associated to viable firms.

6. Results

By applying logistic model to the sample of construction firms in our database is possible to predict the probability of default in a time frame of three years. Aside from a few illogic, mostly related to the latest indicators in logistics function, below is shown the excellent performance of the model used. In Table 4, therefore, is illustrated the validation of the model. In particular, is evaluated the prediction errors of Type I and Type II. First type of error refers to the forecast of false default, while Type II errors are the provisions of fake healthy (Note 13).

As for the number of cases may seem irrelevant cases have more than one type of error with respect to the other, from the economic point of view, however, the weight of the two errors is obviously different. It is not sufficient therefore to reduce the total number of misclassifications but, instead, should advantage a methodology that minimizes the economic impact of the errors. For this reason is indispensable to take up a differentiate methods of handling errors by moving the cut-off value (Note 14) that split the two groups. An approach neutral, with cut-off equidistant from the average values of the two sub-samples, is expressed in first column of table 4. In the second column are shown the errors with a methodology to minimize the economic impact of misclassification.

In this way, the correct classification of total sample, by the aforementioned regression function, would reach the 80.94% of the total population for the first method (52 companies erroneously evaluated healthy and 203 enterprises judged erroneously insolvent) and 77.6 % for the second methodology.

Table 4. Error matrix

	Neutral approach		Economic approach	
	No.	%	No.	%
Error Type I	52	30.41	44	25.73
Error Type II	203	17.40	255	21.85
Hit (true default)	119	69.59	127	74.27
True (true positive)	964	82.60	912	78.15
Accuracy	**80.94 %**		**77.65 %**	

The difference between the two choices is related exclusively to the displacement of the cut-off in consideration of the computation, a priori, of the evaluation and for probability of error. To maximize the economic performance of the forecasting model, in fact, to make it more applicable (Note 15) the same statistical model, it is necessary to quantify the difference in cost between the two errors (Note 16). In this case, to standard practice, it is possible to weigh up twenty times more expensive assessing solvent firm which may become insolvent firm,

with respect to the contrary. The probability to find an insolvent company in the macro-sample of companies object of study is taken by the statistical default: 12.78%. With the above statistics will be able to move forward with the criterion the cut-off separation in order to have fewer cases of Type I error (false positive error) (Note 17).

In this study, therefore, the measure of the performance of the model is described by the amount (percentage) of firms classified correctly. As you can see (table 4), at least with regard to the neutral methodology, table 3 measures more than 80% firms have been correctly evaluated. This result can be judged very admirable especially if it emphasizes the longest period of time of analysis (3 years) and the specific sector (construction). It also adds that was not used an additional method for error handling based on its minimization of the number of the error. In that last case, in fact, the cut-off values should be shifted in the opposite direction. Unlike the economic approach it generates more Type I Errors but many fewer cases of Type II error, having regard to the dissimilar distribution of the two sub-samples. The matrix of error, ultimately, gives a sense of precision to the degree of classification used and shows the type of classification error more frequently encountered.

Developing the regression function will be possible to split the distribution of results in ten classes numerically equi-distributed in consideration of the decile position. Starting from the lowest value and up to the maximum value for each firm will be assigned a result of the regression function, therefore, belong to one of the ten clusters of risk. For each cluster correspond to an equal frequency of cases. The nine thresholds (there are nine cut-off values because there are ten clusters) are shifted according to the above methodology to minimize the cost of the error (economic approach) and divide accordingly the dependent variable.

Before showing the distribution of default within risk classes built by the function of logistic regression, and upon completion of the practice of validating model, set out below are the measurements of the quality model as the Pseudo R-squared. The approach called McFadden explains the accuracy of prediction evaluated in terms of correspondence between the calculated probability and the actual value taken by the dependent variable. This means, therefore, also give an appreciation to the model both for ability to correctly predict what happened and for adaptability of the model.

Taking as comparison a system that encompasses only a constant as explanatory variable, with log LX indicating the value of the maximum likelihood of the model to analyze and LY with the maximum value of the likelihood function when all parameters, except the intercept, are equal to zero, the Pseudo R-squared is: 1 - (log LX / LY).

This indicator will vary from 0 to 1 including the upper extreme, for the possibility that the two likelihoods can coincide. To decrease of the relationship between likelihoods follows the increase of the functionality of the model and the rise of R2. Pseudo R-squared assigned to the model, which measures the skill of the model used to explain the variability of default, is 0.17686. It indicates that the analytical scheme is adaptable and functional than the hypothetical intercept model.

To finish the following Table 5 illustrates the distributions of default cases in the risk classes in order to have a visually stunning on the functioning of the model.

Table 5. Distribution of insolvency cases within the classes of ranking built using logistic regression

Rating classes		Default cases	Frequency
Probability of default	Description		
1 - None	Remote: failure is unlikely	3	1.75
2 - Very minor	Low: failures are few and far between	4	2.34
3 - Minor	Low: relatively few failures	2	1.17
4 - Very low	Moderately low: infrequent failures	5	2.92
5 - Low	Moderate: occasional failures	2	1.17
6 - Moderate	Moderately high: frequent failures	14	8.19
7 - High	High: failures occur often	12	7.02
8 - Very high	High: repeated failures	19	11.11
9 - Extremely high	High: failures occur almost as often as not	39	22.81
10 - Dangerously high	Failure is almost inevitable	71	41.52
Total		171	100.00

The table 5 demonstrates that majority of insolvent firms is found in the higher risk classes. Over 75% of companies that will become insolvent parts in the worst three risk classes and about 41% of the total cases of default falls precisely in the last cluster of the ranking scale. On the other hand, only 16 cases of insolvent firms are positioned within the top 5 risk classes with a total attendance of about 9% of the total number of insolvent firms. Recalling that the ten classes contain the same abundance of firms, it is well understood what the predictive ability of the model is.

7. Conclusions

This manuscript empirically examines the predictive model of business failure using the sample of construction firms that went bankrupt during the period from 2011 to 2012 when a deep recession started in Italy and in many other place of world. The research has shown that defaults of the Italian construction firms can be predicted on the basis of balance sheet and income statement financial ratios. Even though the model was developed based on construction firms database, it cannot be applied to other type of firms for the unusual distinctiveness of building firms.

Best merit of this paper is that a bankruptcy prediction model is constructed based on the financial statement of three years assembled all together. Measures of firms' ability of servicing short-term debts, capital structure, current liabilities to sales, interest expense and investment turnover are variables that comprise the prediction model. Like other types of firms, the findings further imply that a high level of debt use, normal consequence of increased fixed assets, is unhealthy for the financial success of the construction firm.

While many financial theories have been concerning with the determinants of risk of default of manufacturing firms and optimality structure of commercial firms, existing theories are silent on quantitative implications that would provide rationale behind of the prevision of default of construction firms.

Logit maximum likelihood estimator is employed as the statistical technique. With logistic regression the results are consistent with the hypothesis that is possible to create a prevision model also for construction firms. The model demonstrated excellent prediction accuracy and robustness. The sensitivity of model (true positive rate) is 78.15 per cent and the specificity (true negative rate) is 74.27 per cent, accuracy of the model is 77.65 %.

In addition to building a bankruptcy prediction model for a sector firms often overlooked, this paper demonstrates that most of firms that went bankrupt during the economic crisis had shown signs of financial distress just before the crisis even the bankruptcy probabilities of the sample are consistently high during the period of analysis. The results can be interpreted as implying that in economic downturn, firms which had previously too much indebted prior to the crisis, got much more vulnerable than firms which did not touch their own capital structure. The crisis advent, essentially, much more easily hit even firms looking more inflexible and exposed to the market.

Aside from explaining just the sole Italian situation of construction firms, the evidence of this essay can be seen as complementary to the standpoint that draws economic crisis to the qualities of firms from other countries.

References

Davidson, R. A., & Maguire, M. G. (2003). Ten most common causes of construction contractor failures. *Journal of Construction Accounting and Taxation, 13*(1).

Huang, Y. L. (2009). Prediction of contractor default probability using structural models of credit risk: an empirical investigation. *Construction Management and Economics, 27*(6). http://dx.doi.org/10.1080/01446190902960474

Kale, S., & Arditi, D. (1999). Age-dependent business failures in the US construction industry. *Construction Management and Economics, 17*(4). http://dx.doi.org/10.1080/014461999371411

Kangari, R. (1988). Business failure in construction industry. *Journal of Construction Engineering and Management, ASCE, 114*(2). http://dx.doi.org/10.1061/(ASCE)0733-9364(1988)114:2(172)

Kangari, R., Farid, F., & Elgharib, H. M. (1992). Financial performance analysis for construction industry. *Journal of Construction Engineering and Management, ASCE, 118*(2). http://dx.doi.org/10.1061/(ASCE)0733-9364(1992)118:2(349)

Koksal, A., & Arditi, D. (2004). Predicting construction company decline. *Journal of Construction Engineering and Management, ASCE, 130*(6). http://dx.doi.org/10.1061/(ASCE)0733-9364(2004)130:6(799)

Landford, D., Iyagba, R., & Komba, D. M. (1993). Prediction of solvency in construction companies. *Construction Management and Economics, 11*(5).

Muscettola, M. (2013). Leverage Risk. The weight of borrower capital distinguishes the solvency of firms: an empirical analysis on a sample of 4,500 Italian SMEs. *International Journal of Economics and Finance, 5*(12). http://dx.doi.org/10.5539/ijef.v5n12p24

Muscettola, M. (2014a). Structure of assets and capital structure. What are the relations with each other? An empirical analysis of a sample of Italy. *European Journal of Economics and Finance, 5*(12). http://dx.doi.org/10.5539/ijef.v5n12p24

Muscettola, M. (2014b). Cash conversion cycle and firm's profitability. An empirical analysis on a sample of 4,226 manufacturing SMEs of Italy. *International Journal of Business and Management, 9*(5). http://dx.doi.org/10.5539/ijbm.v9n5p25

Muscettola, M. (2014c). Effects of fixed capital investments in current economic downturn. *International Journal of Business and Management, 5*(9).

Muscettola, M., & Gallo, M. (2008). *Analisi e gestione del rischio di credito. Il progetto Mayflower*. FrancoAngeli Editore.

Muscettola, M., & Naccarato, F. (2013). Probability of Default and Probability of Excellence, an Inverse Model of Rating. One More Tool to Overcome the Crisis: an Empirical Analysis. *Business System Review, 2*(2). http://dx.doi.org/10.7350/BSR.BV06.2013

Muscettola, M., & Pietrovito, F. (2012a). Le caratteristiche delle imprese insolventi. *Sinergie Rapporti di Ricerca, 36*.

Muscettola, M., & Pietrovito, F. (2012b). La rilevanza delle variabili finanziarie nel rating: i risultati di un'analisi empirica sulle PMI italiane. *Sinergie Rapporti di ricerca, 36*.

Padget, P. (1991). *Accounting in the construction industry*. London: CIMA.

Russell, J. S., & Jaselskis, E. J. (1992). Predicting construction contractor failure prior to contract award. *Journal of Construction Engineering and Management, ASCE, 118*(4). http://dx.doi.org/10.1061/(ASCE)0733-9364(1992)118:4(791)

Singh, D., & Tiong, R. L. K. (2006). Evaluating the financial health of construction contractors. *Proceedings of the Institution of Civil Engineers, 159*(3). http://dx.doi.org/10.1680/muen.2006.159.3.161

Notes

Note 1. A loan is named "past-due" if the firm missed a principal or interest payment for more than 90 days.

Note 2. This means that the model, in order to have a permanent quality, it must be based on a clear and simple partition between the segments of the sample and must be constructed with a non-excessive number of variables that can gather data always sufficiently numerous and available.

Note 3. The normalization of data is made through division of all values with the relative standard deviation in order to detect the distribution of values around their average.

Note 4. A minimum age of firms was essential in order to avoid issues related to the interpretation of start-up firms.

Note 5. The analysis does not contemplate firms that have become insolvent during the year 2007, 2008, 2009 and 2010. The insolvent firms covered by the definition of "default" are exclusively those which have entered into default in 2011 (will be analyzed accounting data of the years 2008, 2009 and 2010) and in 2012 (will be used the respective financial statements of 2009, 2010 and 2011).

Note 6. It is not easy to distinguish if over-indebtedness is the cause or the effect of a lack of company capitalization. (Muscettola, 2013).

Note 7. The logistic model used is a binary choice where the dependent variable can assume only two values. To search the probability of default of construction firms *pi* is included in the range (0, 1) and denotes the probability of insolvency where the value "1" is assigned to the insolvent firm while the value "0" is given to healthy firms.

Note 8. It denotes the standard deviation of the difference between the estimated values and true values.

Note 9. It trials the null hypothesis that the constant equalizes 0. The supposition is excluded when the *p-value* is smaller than the critical *p-value*.

Note 10. P-value specifies how probable (high values) or unlikely (low values) is the option to detect precisely the "x value" of the test statistic x under the null hypothesis. The *p-value* designates the minimum level of significance for which the null hypothesis is rejected.

Note 11. The symbol "Σ" explains the algebraic sum of the indices of each firm during the three years of analysis.

Note 12. The symbol "µ" defines the arithmetic mean of the indices of each firm during the three years of analysis.

Note 13. A false positive for a specified examination shows, in fact, do not have the conditions that the model would imply.

Note 14. Establishing a point of separation between the two sets it is possible to estimate the performance of the model by quantifying the number of errors that occur.

Note 15. It is fairly clear that there is a non-symmetric cost structure in the event of Error of Type I or Error Type II error, obviously skewed towards the higher costs caused by incorrect classification of firms that will become insolvent. In other words it is not economically justifiable to treat the two types of error in the same way.

Note 16. Making a rough calculation, the cost of the evaluation of the errors of a healthy society can be estimated at about 3% of the loan (not available for the bad judgment), compared to about 60% of borrowing for cost of other type of error. The ratio between the two types of mistake will be, therefore, about 20 times.

Note 17. Unlike a rating model based solely on mathematics and the empirical evidence, therefore, these results will certainly be more stringent to allow for fewer mistakes more expensive.

Copyrights