QUADERNI DEL DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

FINANCIAL PERFORMANCE IN MANUFACTURING FIRMS: A COMPARISON BETWEEN PARAMETRIC AND NON PARAMETRIC APPROACHES

Eleonora Bartoloni - Maurizio Baussola

Serie Rossa: Economia – Quaderno N. 82 maggio 2012

UNIVERSITÀ CATTOLICA DEL SACRO CUORE
PIACENZA

Financial performance in manufacturing firms: a comparison between parametric and non parametric approaches

Eleonora Bartoloni*
Maurizio Baussola†

Abstract

This paper provides a methodological analysis of credit risk in manufacturing firms by using two different credit scoring approaches . The first is the traditional discriminant approach (DA) for bankruptcy prediction based on a logistic regression model, whereas the second, Data Envelopment Analysis (DEA), is a non-parametric approach for measuring firms' efficiency which does not require ex-ante information on bankrupted firms. By using a manufacturing sample of both healthy and bankrupted firms during the period 2003-2009 we provide an indepth comparison of DA and DEA and conclude that a correct evaluation of firms' credit worthiness is the result of successive fine tuning procedures requiring the use of multiple methodological tools.

JEL Classifications: G32, O32, L25

Keywords: Bankruptcy, Discriminant Analysis, Data Envelopment Analysis, Credit rating

*ISTAT Italian Institute of Statistics, Via Porlezza, 12, Milano 20123, Italy. E-mail: bartolon@istat.it.

†UCSC Catholic University of Sacred Heart, Via Emilia Parmense 12, Piacenza 29100, Italy. E-mail: maurizio.baussola@unicatt.it.

1 Introduction

The study of a firm's financial performance is relevant in the context of the present economic downturn, as it allows us to understand whether significant threats to economic recovery do exist and whether investment decisions by firms may stimulate and sustain economic growth in the medium to long term. Firms can be ranked according to their degree of financial constraint which, in turn, may depend upon macroeconomic factors (the cycle or structural characteristics of the economy) and individual characteristics related to the economic and financial position of each firm. A firm's decision to invest may crucially be affected by its rank which reflects its level of financial constraint. Thus, an understanding of the distribution of such financial constraints is particularly relevant with respect to innovative investment, which represents the key to business success.

A firm which is willing to seize growing opportunities by investing in an innovative project may be defined as financially constrained when the amount of internally-generated funds is not sufficient to finance investment activity and it cannot access an adequate amount of external resources (debt and/or equity).

Indeed, although several definitions of credit or financial constraint have been proposed by the relevant literature - Kaplan and Zingales (1997) refer to a wedge between the internal and external cost of funds, while Hall (2002) refers to a situation in which there is a wedge, sometimes large, between the rate of return required by an entrepreneur investing his own funds and that required by external investors - there is currently no general agreement on how financially-constrained firms can be identified empirically.

The debate concerning the measurement of financial friction at the firm level may gain interesting input from the field of business failure prediction. The main goal here is to predict bankruptcy risk, i.e. to develop models of financial failure at the firm level before this actually happens.

Although business failure has long been debated in both economic and accountancy research, accurate credit risk analysis has become even more important today than it was in the past due to the recent global financial crisis, which has demonstrated how difficult it is to measure and manage business distress.

In this contribution we provide a thorough analysis of credit risk in manufacturing firms during the period 2003-2009 by using two alternative approaches: the first is the traditional discriminant analysis (DA) approach for bankruptcy prediction, based on a logistic regression model, while the second, Data Envelopment Analysis (DEA), is a non-parametric approach for measuring efficiency. We propose a combination of these two complementary tools, which can be used in synergy in order to derive a more accurate prediction of business failure.

2 Measuring financial constraints: a review of the empirical literature

Several methods have been proposed in the empirical literature for measuring financial constraints and the debate is still controversial.

A consistent part of the empirical investigation of firms' investment has adopted the analytical framework proposed by Fazzari, Hubbard, and Petersen (1988) arguing that a positive and significant investment-to-cashflow sensitivity signals financial constraint. However, a number of studies have found that differences in cashflow responsiveness between constrained and unconstrained firms is insignificant, or that the investment of unconstrained firms is actually more responsive (Gilchrist and Himmelberg 1995; Kaplan and Zingales 1997; Cleary 1999; Allayannis and Mozumdar 2004).

Following this line of research, Almeida et al. (2004) suggested that a better measure of financial constraint is given by the sensitivity of cash to cashflow. Using a sample of manufacturing firms between 1971 and 2000, they demonstrated that financially-constrained firms have a positive cash-to-cashflow sensitivity while unconstrained firms do not show any systematic pattern. This is explained on the grounds that in contrast with the liquidity irrelevance that characterizes the unconstrained firm, the constrained firm may be forced to save cash today in order to finance future investment opportunities.

In other works the assessment of the existence of credit constraints is based on qualitative-type information when a firm's subjective perception of its particular financial position is available (Canepa and Stoneman, 2008; Giudici and Paleari, 2000; Silva, 2011). The main problem here is represented by possible misreporting when a credit-demand point of view is considered.

A demonstration of problems associated with self-reported information may be found in the 10th Unicredit survey on manufacturing firms (2009), according to which in contrast with the most established empirical evidence, one would conclude that only a small fraction of manufacturing firms (about 4% out of a sample of 5,000 firms) faced some type of financial hindrance during the year 2006.

As an alternative, empirical researchers have proposed a sorting approach, which is based on the idea that a firm's financial status may be categorized on the basis of its specific characteristics.

Following this approach, Kaplan and Zingales (1997) and Lamont et al. (2001) proposed indexes of financial constraint estimated by using ordered logit models. Whited and Wu (2006) developed an alternative index based on GMM estimations of a standard intertemporal investment augmented model to account for financial frictions. In these models a firm's financial status is a function of various quantitative explanatory variables.

Hadlock and Pierce (2010) exploited an alternative approach based on qualitative information in order to categorize firms. Annual letters to shareholders and management statements from financial filings provided the necessary information for classifying firms in different risk categories. Using this qualitative categorization, order logit models of quantitative information were estimated in order to test the validity of alternative indexes of financial constraint proposed by the empirical literature.

In the field of business failure prediction, while originally traditional methods were essentially subjective, i.e. based on an expert's evaluation, as for example in the so-called five Cs credit analysis (Saunders and Allen, 2002, pp. 5-9), as of the 1960s a variety of techniques have been proposed in the empirical literature.

Discriminant analysis has been widely used for failure prediction. This approach is essentially based on the idea that a firm's probability of default may be estimated by using a set of key variables. These variables, appropriately combined together, produce a range of quantitative scores, which can be used as a classification tool when combined with an appropriate cut-off point. We refer to the seminal work by Altman (1968) and further developments (Deakin, 1972; Altman et al., 1977), which employ a linear discriminant model based on accounting data of failed and non-failed firms in order to determine a firm's bankruptcy risk.

Ohlson (1980) proposed a conditional logistic model that has the advantage of overcoming problems associated with the linear discriminant model, that is the assumption of normality and equal covariances for both failed and non-failed groups.

The peculiar feature of this approach is the way a model's precision is tested for by considering both classification and future prediction accuracy. Classification accuracy is assessed on the original database, that is the data-set used in order to specify the model. Following this, prediction accuracy is tested for by using a new data set, in order to assess how well the model works for future predictions.

In evaluating prediction accuracy there is no way of adjusting the cut-off point for the distribution in order to reduce simultaneously the two types of classification errors, that is the error of classifying a sound firm as unsound (Type I error) and the error of classify an unsound firm as sound (Type II error).

In practice, as there is a trade-off between the two types of error, a pragmatic rule is adopted depending on the specific aim of the classification and, therefore, on the characteristics of the users of such financial information. Indeed, a bank which is evaluating a firm's financial position is probably more interested in minimizing the cost of making a bad investment (Type II error) due to lending funds to a potentially defaulting customer, whereas a shareholder in an innovative firm may be willing to reduce the cost of under-investment (Type I error) resulting from not taking advantage of an investment opportunity.

One criticism that has been made of traditional approaches is that they are essentially based on accounting ratios, thus omitting the influence of sectoral and macroeconomic conditions. Another criticism concerns the fact that these models are essentially static and inappropriate for predicting a rare event, such as bankruptcy, due to their reliance on data from an arbitrary period before the extreme event occurs.

In more recent years, which have been characterized by a structural increase in bankruptcy worldwide (probably due to even more intense global competition), new approaches have been explored when appropriate longitudinal data are available.

Among the most significant contributions, Shumway (2001) proposed a survival analysis approach, which is able to correct for time spent by a firm in the healthy (non bankruptcy) group and uses time-varying regressors. By using a panel of quoted firms during the period 1962-1992 for a total of more than 3,000 firms including 300 firms which went bankrupt, the author estimated a hazard model, based on maximum likelihood estimates of a particular logit model. Among the regressors the model includes not only traditional accounting ratios but also market-driven variables derived from information at the firm level on market capitalization and stock returns. Of course the use of these additional explanatory variables is constrained by the availability of quoted firms to analyse.

In another study Linde and Jacobson (2011) studied a firm's probability of default by using a logistic specification with a panel of almost seventeen million quarterly observations of Swedish firms during the period 1990-2009. In order to evaluate the effects of macroeconomic conditions, four aggregate variables (output gap, yearly inflation rate, nominal interest rate and real exchange rate) are incorporated into the model, together with a set of financial ratios. Results support the view that although firm-specific variables are important for ranking firms according to their relative risk propensity, macroeconomic conditions do affect the average default level, and thus are important determinants of a firm's risk level.

It is worth noting that discriminant procedures have been criticized on the grounds that they suffer from some of the failings which typically characterize the parametric approach. One of the criticisms concerns possible endogeneity problems affecting financial distress estimations based on accounting information. Endogeneity arises when the financial indexes used as explanatory variables are instead the effects of a particular situation of distress.

Another criticism concerns the selection of the appropriate proportion of failed firms in the final sample, given that bankruptcy is a rare event and, thus, difficult to predict. It has been argued that because the link function is symmetric, a logistic regression tends to underestimate bankruptcy probabilities. As a result, more flexible skewed link functions have been indicated as being more suitable for analysing binary response data (Stuckel, 1988; Wang and Day, 2010). Using a sample of Italian SMEs drawn from the AIDA - Bureau van Dijk database over the years 2005-2009, Calabrese and Osmetti (2011) propose a Generalized Extreme Value (GEV) regression for analysing default probabilities and find that its predictive performance is better than that of the logistic regression predictive model.

In recent years non-parametric techniques such as neural networks and decision trees have been proposed in the empirical literature. These techniques are based on the machine learning approach, i.e. the design and development of algorithms that allow computers to predict behaviour based on empirical data.

Although neural networks have been widely used for failure prediction, no clear demonstration of their superiority has been provided so far. The major criticism of this methodology is that it is a black-box approach, as it is not possible to check for the internal structure of the networks or have information on the relative importance of the variables used.

An application to Italian data developed by Altman et al. (1994) and based on a sample of one thousand industrial firms demonstrated that neural networks do not outperform traditional discriminant analysis in their ability to classify sound and unsound firms correctly.

An alternative non-parametric approach to credit scoring is based on the DEA (Data Envelopment Analysis) methodology (Charnes, Cooper and Rhodes, 1978), which has the advantage of not depending on the availability of ex-ante information on bankruptcy events. The DEA scoring approach is essentially a mathematical programming method to evaluate the relative efficiency of "decision making units" (DMUs). By converting multiple inputs into multiple outputs, DEA computes the relative efficiency scores of each decision making unit (i.e. a firm or a bank). The DEA approach has been widely applied in different frameworks; examples of applications to banking and finance are given by Yeh (1996), Troutt et al. (1996), Simak (1999), Cielen and Vanhoof (1999) and, more recently, Min and Lee (2008).

In the present work we adopt a credit scoring procedure. Our main interest is to provide different methods which may be used as complementary approaches for predicting firms' economic and financial performance. Thus, we first perform a discriminant analysis (DA) based on a sample of both failed and non-failed firms, in order to derive an empirical measure of financial worthiness and, implicitly, financial constraint. Default probabilities are estimated by using a logistic model which includes both firm-specific characteristics and financial indexes. We then apply the DEA approach to the same database used for the logistic discriminant in 2003. By using an appropriate set of financial ratios, firms' credit worthiness is estimated by exploring the relative efficiency of the complete set of firms (both failed and non-failed). Both methodologies are then applied to the sample of firms in 2009 in order to define appropriate credit scoring suitable for comparison.

3 The data

3.1 Innovation and Accounting information

Our main sample of firms is derived from the 10th Unicredit Survey on Manufacturing Firms (2009). This sample is composed of more than five thousand firms representative of the manufacturing sector and extracted from the AIDA data base. A rich set of information is collected by this survey, including firm-specific characteristics and investment and innovative activities.

This starting sample has been inflated with a rich set of accounting data. The economic and financial information derived from firms' balance sheets has allowed us to derive the financial indexes (see Table 1) used in the credit scoring procedures which will be described in the following sections.

3.2 Bankruptcy data

Bankruptcy data have been collected from the AIDA data base. We extracted a sample of 150 firms which went bankrupt during the years 2005 and 2006. Balance sheet information refers to years 2003 and 2004 in order to have an adequate time span difference (not less than one year) between the last relevant balance sheet and the bankruptcy date.

The sample size was fixed by taking into account two important conditions in order to derive reliable default predictions. Firstly, although firm default is a rare event - estimations by the Cerved Group (Cerved, 2007) show that Italy, with 18 cases per 10,000 firms in 2006, has an insolvency ratio which is far below the European average (63 cases)[1] it is important to supplement the sample of non-defaulting firms with an adequate number of defaulting firms in order to derive a reliable discriminant rule for predicting a "rare" event. Another reason is that firms which are close to bankruptcy may present abnormal accounting data which should be removed, given that discriminant techniques are particularly sensitive to outliers, thus determining a further reduction in bankruptcy observations.

Secondly, the probability of default for the firms on the Italian business register is significantly affected by specific characteristics, such as age, size and localization. In general, smaller and younger firms localized in southern regions show a higher probability to default compared to older and larger firms localized in northern regions. In order to take these differences into account we decided to stratify the sample so as to increase the representativeness of our set of bankrupt firms. Appendix A shows how the sample

[1] However, it is worth noting that international comparisons should be interpreted with caution, due on the one hand to different insolvency regimes and on the other hand, to differences in firms' structural characteristics across Europe. In fact, in countries such as Spain, Greece and Italy, where the proportion of small businesses is higher, the low levels of the insolvency ratio may be justified on the ground that insolvent firms do not opt for the orderly insolvency proceedings, voluntary abandonment being the preferred option.

Variable	Description	N firms		Mean	Median	Min	Max
ACID	Acid test - Liquid Assets/Current Liabilities	4,166	all	1.01	0.88	0.19	4.20
		4,088	non bankrupted	1.01	0.88	0.19	4.20
		78	bankrupted	0.67	0.59	0.25	2.39
LEV	Leverage - Total Debts/Net Capital	4,166	all	7.20	3.21	0.21	103.01
		4,088	non bankrupted	6.92	3.15	0.21	103.01
		78	bankrupted	21.71	12.84	0.92	85.06
CN_A	Net Capital/Total Assets	4,166	all	0.25	0.22	0.01	0.80
		4,088	non bankrupted	0.26	0.22	0.01	0.80
		78	bankrupted	0.11	0.07	0.01	0.52
CL_S	Current Liabilities/Sales	4,166	all	0.52	0.47	0.13	2.02
		4,088	non bankrupted	0.51	0.47	0.13	2.02
		78	bankrupted	0.72	0.61	0.13	1.92
CL_A	Current Liabilities/Total Assets	4,166	all	0.57	0.58	0.09	1.60
		4,088	non bankrupted	0.57	0.57	0.09	1.60
		78	bankrupted	0.76	0.82	0.16	1.06
ROS	Return on Sales	4,166	all	0.05	0.04	-0.20	0.26
		4,088	non bankrupted	0.05	0.04	-0.20	0.26
		78	bankrupted	0.02	0.02	-0.14	0.16
IR	Interest Payment/Sales	4,166	all	0.02	0.02	0.00	0.18
		4,088	non bankrupted	0.02	0.02	0.00	0.18
		78	bankrupted	0.03	0.02	0.00	0.11
TA	Total Assets (Euros)	4,166	all	15,193,402	4,938,010	249,212	343,369,462
		4,088	non bankrupted	15,345,403	4,984,758	249,212	343,369,462
		78	bankrupted	7,226,973	1,550,833	254,972	52,655,037
AGE	Firm Age in years	4,166	all	29	25	0	256
		4,088	non bankrupted	29	25	0	256
		78	bankrupted	9	6	0	25

Table 1: Descriptive statistics - Years 2003

was allocated to each of the strata according to the stratification variables. Stratification was first determined by area, then by age and size[2]. Firms were selected randomly in each stratum, while the allocation of the sample across strata was assessed on the basis of a system of weights which were applied to the default probability observed in the the Italian Business Register (the reference population) (Cerved, 2011).

4 The Discriminant Approach

4.1 The logistic discriminant model

We estimate the default probability of a firm by using a logistic discriminant function defined as follows:

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k \tag{1}$$

where:

$$p = Prob(D = 1/X) \tag{2}$$

D is our binary dependent variable, which assumes the value of 1 if we observe a default event between years 2005 and 2006 and 0 otherwise and X is the vector of covariates, i.e. firm-specific characteristics and financial indexes which are observed in years 2003 and 2004.

We have included a set of variables which are commonly considered good predictors of the outcome event in the relevant literature:

- a measure of a firm's leverage (LEV), the ratio of total debts to net capital, which is expected to affect the default probability positively, as a highly-leveraged structure may worsen the perceived financial risk;
- a measure of short-term indebtedness (CL_S), the ratio of current liabilities to sales, whose expected sign is positive, given that a firm with a high short-term debt may find it difficult to borrow additional resources to finance its short run activities and, thus, may be close to insolvency;

[2]Cerved Group (February 2011 - Rapporti Flash) has estimated that one of the most relevant determinants of the default probability is a firm's localization, followed by age and firm size, while the sector of activity is not among the most relevant factors

- another similar indicator, the ACID ratio; this measures the extent to which short-term debt is covered by short term liquidity. Creditors prefer a high ACID ratio as it reduces their risk. We thus expect a negative sign;

- firm operating profitability (ROS), proxied by the ratio of operating margins to sales. We expect a negative effect on the default risk, as the higher a firm's profitability the higher the flow of internal resources available to cover debt exposure should be;

- the firm's interest burden, proxied by the interest payment to sales (IR) ratio, which is expected to positively affect the default probability given that a high interest burden may worsen the financial risk associated with external finance. We have used a dummy variable assuming the value of 1 when a firm shows an interest burden ratio higher than 5%, which identifies the last 5% of the IR distribution, and 0 otherwise, in order to capture the effect of those firms which are potentially financially constrained;

- finally, structural characteristics, captured by variables AGE (years) and SIZE, proxied by a firm's total assets (logarithmic values). We expect a negative effect of both these variables, as agency costs related to indebtedness are expected to be higher for those firms with a low reputation or contractual power, such as those which are smaller or less well established.

Estimation results are shown in Table 2. We estimate default probabilities within one and two years. In the first case the model is computed by using predictors observed in the year 2004, while in the second case we use information for the year 2003. In both models our variables present the expected signs, although it is worth noting that the explanatory power is higher when information two years before bankruptcy is used. This evidence suggests that the choice of an adequate lead time span is a relevant point and needs to be taken into account. In our case, by using accounting information from two years prior the default event, we can build a more accurate prediction model.

4.2 Classification accuracy

Classification accuracy is evaluated by using the samples of firms used to predict default probabilities in the years 2003 and 2004. As the classification table shows, different cutoff points are associated with a tradeoff between Type I Error (False Positives) and Type II Error (False Negatives). If a cutoff point of 0.5 is selected from the 2003 model, only 5 firms out of 78 are correctly classified as bankrupt. In 2004 the same cutoff produces an even worse prediction (1 out of 67).

The Receiver Operating Characteristic (ROC) curve is then used as a diagnostic test for accuracy. It plots the true positive rate (sensitivity) against the false positive rate (1-specificity) for different cutoff points (Figure 1). The area under the curve describes how well the classification rule works: a ROC curve which passes through the upper left-hand corner would indicate an optimal discrimination (100% sensitivity and 100% specificity), while the 45° degree line indicates a situation of irrelevance, as for each cutoff point one would observe the same error for both types (no trade off). Thus, the closer the ROC curve is to the upper left-hand corner the higher the accuracy of the discrimination rule.

By overlapping the ROC curves for years 2003 and 2004 it is possible to show the better performance of the 2003 model with respect to the 2004 model. For each cutoff point, classification based on the 2003 model yields

year 2003: N=4,100			
Percent Concordant: 93.5%			
LR chi2 (10)	**284.2**		
year 2004: N=4,607			
Percent Concordant: 90.8%			
LR chi2 (107)	**201.9**		
	coefficient sign	**P-value** *(year 2003)*	**P-value** *(year 2004)*
constant	+	0.0009	0.1085
ACID	-	0.0172	0.0149
LEV	+	0.0002	0.0423
CL_S	+	0.1473	0.0607
ROS	-	0.0009	<.0001
dIR	+	0.0085	0.6346
L_TA	-	<.0001	<.0001
AGE	-	0.0004	0.0151
dNW	-	<.0001	0.0002
dNE	-	0.0002	0.0657
dC	-	0.113	0.2481

Table 2: Default probability - Logistic discriminant

	Model 2003				Model 2004			
Cutoff point	default firms correctly classified	sound firms wrongly classified	Type I Error	Type II Error	default firms correctly classified	sound firms wrongly classified	Type I Error 98.5	Type II Error .
0	78	4022	98.1	.	67	4540	93.2	0.3
0.02	66	660	90.9	0.4	54	738	90	0.6
0.04	58	385	86.9	0.5	42	379	87.8	0.8
0.06	53	270	83.6	0.7	31	223	84.4	0.9
0.08	48	192	80	0.8	27	146	80.8	1
0.1	42	144	77.4	0.9	24	101	80.2	1.1
0.12	39	120	75.5	1	18	73	78.1	1.1
0.14	38	97	71.9	1	16	57	76.7	1.2
0.16	37	73	66.4	1	14	46	70.2	1.2
0.18	34	67	66.3	1.1	14	33	64.1	1.2
0.2	30	55	64.7	1.2	14	25	61.8	1.2
0.22	27	47	63.5	1.3	13	21	58.1	1.2
0.24	25	43	63.2	1.3	13	18	56	1.2
0.26	20	38	65.5	1.4	11	14	55	1.3
0.28	18	33	64.7	1.5	9	11	50	1.3
0.3	17	29	63	1.5	9	9	56.3	1.3
0.32	16	27	62.8	1.5	7	9	57.1	1.3
0.34	14	25	64.1	1.6	6	8	70	1.4
0.36	14	20	58.8	1.6	3	7	70	1.4
0.38	12	18	60	1.6	3	7	70	1.4
0.4	11	13	54.2	1.6	3	7	70	1.4
0.42	10	11	52.4	1.7	3	7	75	1.4
0.44	8	8	50	1.7	2	6	71.4	1.4
0.46	8	7	46.7	1.7	2	5	71.4	1.4
0.48	6	6	50	1.8	2	5	83.3	1.4
0.5	5	5	50	1.8	1	5	80	1.4
0.52	4	4	50	1.8	1	4	80	1.4
0.54	3	4	57.1	1.8	1	4	80	1.4
0.56	2	4	66.7	1.9	1	4	80	1.4
0.58	2	3	60	1.9	1	4	75	1.4
0.6	2	3	60	1.9	1	3	75	1.4
0.62	2	2	50	1.9	1	3	75	1.4
0.64	2	2	50	1.9	1	3	75	1.4
0.66	2	2	50	1.9	1	3	75	1.4
0.68	2	2	50	1.9	1	3	50	1.4
0.7	2	2	50	1.9	1	1	50	1.4
0.72	2	2	50	1.9	1	1	50	1.4
0.74	2	2	50	1.9	1	1	100	1.5
0.76	1	2	66.7	1.9	0	1	100	1.5
0.78	1	2	66.7	1.9	0	1	100	1.5
0.8	0	2	100	1.9	0	1	100	1.5
0.82	0	2	100	1.9	0	1	100	1.5
0.84	0	1	100	1.9	0	1	100	1.5
0.86	0	1	100	1.9	0	1	100	1.5
0.88	0	1	100	1.9	0	1	.	1.5
0.9	0	1	100	1.9	0	0	.	1.5
0.92	0	1	100	1.9	0	0	.	100
0.94	0	1	100	1.9	0	0	.	100
0.96	0	0	.	1.9	0	0	.	100

Table 3: Classification Table

Type I and Type II Errors which are lower than for the 2004 model.

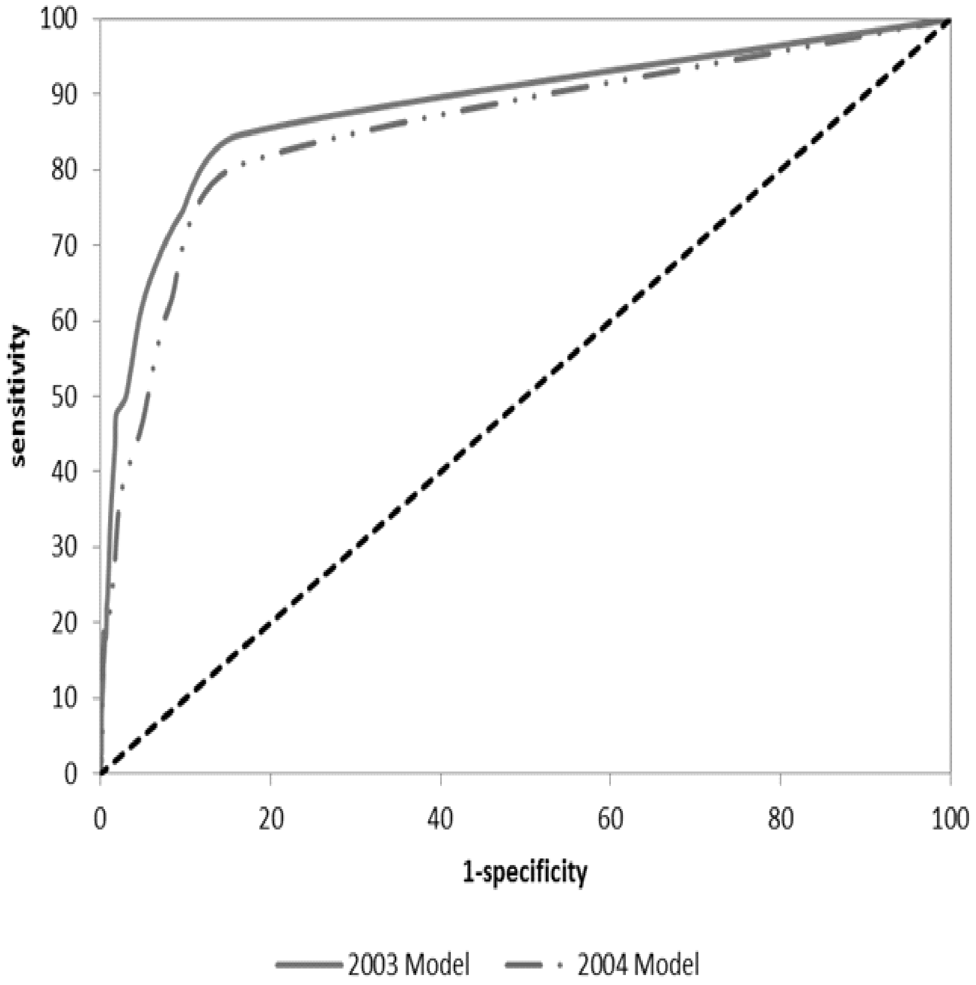

Figure 1: ROC Curve for Logistic Discriminant model - Area under the curve: 0.9351 (2003 Model) and 0.9083 (2004 Model)

If a cutoff point of 0.02 is fixed, a Type II Error of 0.5 is obtained with the 2003 model (66 out of 78 bankruptcy cases correctly predicted). However, as at this cutoff point we also wrongly classify as unsound 660 out of 4022 healthy firms, we prefer to accept a small increase in Type I Error in order to reach a better classification for the group of healthy firms. Thus, a cutoff point of 0.04 seems to be a reasonable compromise (58 out of 78 bankruptcy cases correctly predicted and 3637 out of 4022 sound firms correctly classified) to be used for future prediction.

The set of estimated coefficients from the logistic discriminant together with the adjusted cutoff point will be used to predict business failure. We perform a new logistic discriminant based on the previously-saved set of rules and on a new dataset. The new dataset comprises the Unicredit sample of firms observed in 2009 and the same accounting variables as for 2003.

4.3 Future prediction

We are interested in predicting business failure. In order to do this, we decided to divide firms into four risk classes according to estimated probability intervals and relative frequency distributions. By considering the sub-sample of Unicredit firms operating in 2009 (3,424 firms), only 200 firms presented an estimated default probability greater than the fixed cutoff point (Table 4). By splitting the 2009 sample into two sub-samples, the sample with default probabilities higher than 0.04 was further divided into two additional sub-samples: the first group, representing the last 90th percentile, can be regarded as the group of "risky" firms (20 firms in 2009), the rest of the distribution (179 firms) can be regarded as "critical" firms. The other sub-sample with default probabilities lower than 0.04 was divided into "good" firms (the last 75th percentile corresponding to 2,419 firms) and "excellent" firms (the first 25th percentile corresponding to 806 firms).

We also applied the same classification to the 2003 sample of firms in order to derive a cross tabulation with frequency distributions across the four risk classes at time T and T+6. This representation allows us to investigate persistence patterns and transition probabilities across risk classes during a six-year time span. Table 5 shows high degrees of persistence in the normal and excellent classes: 91% and 84% of firms which had been classified, respectively, as normal and excellent in 2003 were still in the same category in 2009. Persistence in the critical group of firms is much lower but not negligible: 30% of firms which were classified as critical in 2003 were still in the same situation (69 out of 209 firms) in 2009. Finally, only 12 firms in the critical group in 2003 were still present in 2009. 8 of them were still in an unsafe condition (critical or risky), but the most interesting consideration here is that 73% of the risky firms in 2003 had ceased business before 2009 (Table 6). Although the exit rate may be affected by factors other than business management (i.e. data availability in both years and/or mergers and acquisitions), it is worth noting that the exit rate is much lower in the other classes (26%) in both the excellent and normal groups of firms and 47.5% in the critical group).

5 An alternative approach to credit scoring: Data Envelopment Analysis (DEA)

5.1 Introduction

In the previous sections we have described the use of discriminant analysis (DA) to develop adequate credit scoring indexes.

We have underlined how the results crucially depend on the availability of a sufficiently large amount of information on bankrupted firms. Typically, such an approach implies that the number of bankrupted firms is relatively small compared with the overall number of firms under investigation.

This fact may crucially affect the results of DA, which may underestimate default probabilities. Our task is therefore to develop a methodology which can be used as a complement to DA, to help assess the financial and economic

Total n. of firms: 3,424		**N. of firms above the cutoff point: 200**		**N. of firms below the cutoff point: 3,224**	
Quantile	Prob. Estimate	Quantile	Prob. Estimate	Quantile	Prob. Estimate
100% Max	0.856470587	100% Max	0.8564706	100% Max	0.039837629
99%	0.214908159	99%	0.7020042	99%	0.034542941
95%	0.047518534	95%	0.39709	95%	0.021215554
90%	0.02223708	90%	0.3027572	90%	0.012868728
75% Q3	0.005199702	75% Q3	0.1405008	75% Q3	0.003673321
50% Median	0.000746195	50% Median	0.0780178	50% Median	0.000596812
25% Q1	0.000075912	25% Q1	0.054063	25% Q1	0.000063037
10%	0.000005003	10%	0.0465096	10%	0.000004221
5%	0.000000628	5%	0.0430681	5%	0.000000526
1%	0	1%	0.0407516	1%	0
0% Min	0	0% Min	0.0404833	0% Min	0

Table 4: frequency distributions

Class_discr03	Class_discr09				
	risky	**critical**	**normal**	**excellent**	**Total**
risky	3	5	4	0	12
	0.1	0.17	0.14	0	0.41
	25.0	41.67	33.33	0	
	23.08	4.39	0.19	0	
critical	7	69	133	0	209
	0.24	2.37	4.57	0	7.19
	3.35	**33.01**	63.64	0	
	53.85	60.53	6.38	0	
normal	3	40	1840	128	2011
	0.1	1.38	63.27	4.4	69.15
	0.15	1.99	**91.5**	6.36	
	23.08	35.09	88.25	18.39	
excellent	0	0	108	568	676
	0	0	3.71	19.53	23.25
	0	0	15.98	**84.02**	
	0	0	5.18	81.61	
Total	13	114	2085	696	2908
	0.45	3.92	71.7	23.93	100

Note. In each cell: frequency, percent, row percent, column percent

Table 5: Association between risk classes - Years 2003 and 2009

Class_discr	**firms observed in 2003**	**firms observed in 2003 and in 2009**	**exit rate**
risky	45	12	73%
critical	398	209	47%
normal	2742	2011	27%
excellent	915	676	26%
total	4100	2908	29%

Table 6: Firms by risk class - Exit rates

position of a firm more accurately. We therefore apply Data Envelopment Analysis (DEA) in order to rank firms according to a financial score derived from a non-parametric methodology.

DEA has been used widely to analyze the efficiency and productivity of firms in the economy since the seminal contribution by Charnes, Cooper and Rhodes (1978), with applications to many different sectors, contexts and activities. As a nonparametric approach, it can easily be applied to different frameworks, particularly when comparisons between firms or decision making units (DMU) - to use the terminology of the DEA literature - is fundamental either for policy analysis or for other economic choices.

It is therefore worth recalling the hypothesis underlying such a methodology before discussing the results. DEA is a non-parametric approach to measuring the relative performance of a firm (or any other organizational unit) with respect to other competitors. Charnes et al. (1978) proposed the basic DEA model (CCR), which has since been extended to a variety of different hypotheses.

The basic model (CCR) implies that there are n decision making units (DMU) which convert the same m inputs into the same s outputs. In general terms, the *j-th* firm (DMU) uses an input vector x_{ij} *(i=1,2,....,m)* to produce an s-dimensional output vector y_{rj}. This implies the following maximization problem:

$$Max\ \theta_0 = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \tag{3}$$

subject to:

$$Max\ \theta_o = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \ \leq\ 1$$

$$u_r \geq\ 0,\ v_i \geq\ 0\ \ \forall\ r;\ i$$

where subscript o indicates the particular firm (DMU) being evaluated, and u_r $(r = 1, 2..., s)$ and v_i $(v = 1, 2..., m)$ are respectively output and input weights, which are required to be non-negative. The previous definition of the DEA problem relies on the concept of input and output variables which may be minimized or maximized; in particular, one can think of inputs as being minimized while satisfying at least the given output levels, or outputs being maximized without requiring more of any of the given inputs. The first approach refers to the so called "input oriented" model, while the second refers to the "output oriented" model.

In our analysis we have adopted the first problem-setting hypothesis, which can in our opinion be better applied to a financial problem, in that a firm is trying to minimize financial expenses for a given output variable [3].

One should also recall that the CCR model implies constant returns to scale: in other words, one can think of the existence of a linear and significant relationship between input and output variables. We therefore tested for the existence of a significant linear relationship with respect to the input and output variables, before deciding to apply such a model setting.

The application of the DEA methodology to credit scoring is relatively recent, thus suggesting that this is an applied field of research which has not yet been exploited.

We decided to use some of the financial and economic ratios described in Table 1. Some of these indexes have been selected for the logistic discriminant model described in the previous section (*ACID, LEV, ROS, IR*), although we decided to include the Net Capital to Total Assets Ratio (*CN_A*) and the Short Term Debt to Total Assets ratio (*DEBT*) which may better represent the DEA problem set.

The inputs to be minimized are *ACID, LEV* and *DEBT*, while the operating profitability ratio (*ROS*) and the Net Capital to Total Assets Ratio (*CN_A*) are set as output variables.

The solution of the optimization problem described in equation 3 determines n scores which may be thought of as financial stability scores.

The distribution of DEA scores in 2003 and 2009 is shown in Table 7. In order to classify firms according to their score, we analyzed the quintile distribution of the scores and we propose a classification which defines the top 25% of the distribution as excellent and the bottom 5% as risky. Having defined the extreme scores, one should attempt to define the intermediate quantiles. Therefore we adopt a classification which implies that a firm which

[3] This problem setting is also used by Min and Lee (2008)

lies between the 25^{th} (excluded) and the 5^{th} (included) percentile is classified as critical, whereas a firm which lies between the 75^{th} (excluded) percentile and the the 25^{th} (included) is classified as normal.

The analysis of the data confirms the soundness of such a classification: the ex-post evaluation of the scores attached to the subsample of bankrupted firms used for the discriminant analysis confirms that all firms have been correctly classified as risky.

Quantile	Scores Estimate	
	Year 2003 N. of firms: 4,166	*Year 2009 N. of firms: 3,845*
100% Max	1.000	1.000
99%	0.975	1.000
95%	0.666	0.582
90%	0.547	0.443
75% Q3	0.364	0.286
50% Median	0.228	0.172
25% Q1	0.143	0.091
10%	0.094	0.044
5%	0.069	0.022
1%	0.029	0.000
0% Min	0.011	0.000

Table 7: frequency distributions

This classification implies that almost 5% are risky, 20% are critical, 50% are, instead, good and 25% excellent. This distribution is stable, as it holds in both years (Table 8). However, although the distribution is stable, we can observe movements within these financial states during the time span considered. We can analyze such movements by looking at Table 9, which shows the flows of firms from one state to another between the two periods. Thus, more than 55% remain excellent, while more than 62% remain normal, and almost 42% and almost 14% stay critical and risky respectively.

On the whole almost 22% of firms show an upward shift in the ranking, whereas almost 24% show a downward shift, thus implying a downgrade of their financial condition. These ratios are derived by dividing the sum of the

Class_DEA	firms observed in 2003	%	firms observed in 2009	%
risky	208	5.0	192	5.0
critical	833	20.0	769	20.0
normal	2083	50.0	1922	50.0
excellent	1042	25.0	962	25.0
total	4166	100.0	3845	100.0

Table 8: firms by risk class

Class_dea03	Class_dea09				
	risky	**critical**	**normal**	**excellent**	**Total**
risky	18	68	40	6	132
	0.56	2.13	1.26	0.19	4.14
	13.6	51.52	30.3	4.55	
	14.17	10.76	2.42	0.78	
critical	40	250	282	28	600
	1.26	7.84	8.85	0.88	18.83
	6.67	**41.67**	47	4.67	
	31.5	39.56	17.03	3.63	
normal	64	277	1003	270	1614
	2.01	8.69	31.47	8.47	50.64
	3.97	17.16	**62.14**	16.73	
	50.39	43.83	60.57	34.97	
excellent	5	37	331	468	841
	0.16	1.16	10.39	14.68	26.39
	0.59	4.4	39.36	**55.65**	
	3.94	5.85	19.99	60.62	
Total	127	632	1656	772	3187
	3.98	19.83	51.96	24.22	100

Note. In each cell: frequency, percent, row percent, column percent

Table 9: Association between risk classes - Years 2003 and 2009

lower (upper) off-diagonal values of the matrix represented in Table 9 by the total number of firms. In particular, 51.4% of firms remain either critical or risky, while 84.4% stay normal or excellent, thus suggesting that persistence does characterize firms' financial condition. We will return to this point in the following section.

5.2 Scoring performance: a comparison between DA and DEA approaches

These results underline a significant difference between the two proposed scoring methodologies, in that DA is a parametric procedure, whose outcomes crucially depend on the choice of the Type I and Type II Error classification one is willing to accept. In our sample of firms, we decided to choose a probability threshold (0.04) which enabled us to identify more than 74% of bankrupted firms correctly. On the other hand, DEA is a non-parametric methodology which implies an optimization problem, and thus does not depend on an a priori hypothesis concerning the model being estimated or simulated.

We also provide a more in-depth examination of DA and DEA results. Table 10 and 11 present cross classifications which enable us to verify and test for the degree of association of the two methodology. If we consider the values on the main diagonal of the *4X4* matrices which compare DA and DEA classifications, we note that 49% of firms in 2003 and more than 51% are accordingly classified by the two methodologies. If one considers a less restrictive classification, say, good firm (excellent or normal) and bad firm (critical or risky), these percentages significantly increase to 74.5% and 77.8% respectively in 2003 and 2009. This evidence is then reflected in the Chi-Square tests on the degree of association of the two classifications and the Spearman Correlation Coefficient.

This comparison between DA and DEA scores highlights the different methodological foundations of the two approaches, and suggests that they can be used as complements in the analysis of firms' financial worthiness.

Also, another interesting difference becomes apparent if one considers class-movements between the reference years recorded according to the two different approaches (Table 12). Firm performance seems to be more conservative according to the discriminant rule: 85.3% of the firms do not change risk class during the observed period, and only 9.3% and 5.4% of the sample

Class_dea03	Class_discr03				
	risky	critical	normal	excellent	Total
risky	15	63	116	13	207
	0.37	1.54	2.83	0.32	5.05
	7.3	30.43	56.04	6.28	
	33.33	15.83	4.23	1.42	
critical	14	122	583	103	822
	0.34	2.98	14.22	2.51	20.05
	1.7	**14.84**	70.92	12.53	
	31.11	30.65	21.26	11.26	
normal	14	172	1467	394	2047
	0.34	4.2	35.78	9.61	49.93
	0.68	8.4	**71.67**	19.25	
	31.11	43.22	53.5	43.06	
excellent	2	41	576	405	1024
	0.05	1	14.05	9.88	24.98
	0.2	4	56.25	**39.55**	
	4.44	10.3	21.01	44.26	
Total	45	398	2742	915	4100
	1.1	9.71	66.88	22.32	100

Note. In each cell: frequency, percent, row percent, column percent

Measures of associations			
Statistic	**DF**	**Value**	**Prob**
Chi-Square (a)	9	470.6916	<.0001
Spearman Correlation Coefficient (b)		-0.35655	<.0001

(a) Association between risk classes (4 classes).

(b) Measure of association based on the ranks of the firms' scores. Prob > |r| under H0: Rho=0. The correlation is negative as firms are inversely ranked: DEA scores imply that a value close to one is related to a good financial condition, whereas the DA scoring rule implies that such a condition is realized when the score tend to zero.

Table 10: Association between DA and DEA approaches - Year 2003

Class_dea09	Class_discr09				
	risky	critical	normal	excellent	Total
risky	5	11	35	2	53
	0.15	0.32	1.02	0.06	1.55
	9.4	20.75	66.04	3.77	
	25	6.15	1.45	0.25	
critical	9	85	556	74	724
	0.26	2.49	16.27	2.17	21.19
	1.24	**11.74**	76.8	10.22	
	45	47.49	23.02	9.22	
normal	5	73	1337	393	1808
	0.15	2.14	39.13	11.5	52.91
	0.28	4.04	**73.95**	21.74	
	25	40.78	55.36	48.94	
excellent	1	10	487	334	832
	0.03	0.29	14.25	9.77	24.35
	0.12	1.2	58.53	**40.14**	
	5	5.59	20.17	41.59	
Total	20	179	2415	803	3417
	0.59	5.24	70.68	23.5	100

Note. In each cell: frequency, percent, row percent, column percent

Measures of associations			
Statistic	**DF**	**Value**	**Prob**
Chi-Square (a)	9	383.5786	<.0001
Spearman Correlation Coefficient (b)		-0.40018	<.0001

(a) Association between risk classes (4 classes).

(b) Measure of association based on the ranks of the firms' scores. Prob > |r| under H0: Rho=0. The correlation is negative as firms are inversely ranked: DEA scores imply that a value close to one is related to a good financial condition, whereas the DA scoring rule implies that such a condition is realized when the score tend to zero.

Table 11: Association between DA and DEA approaches - Year 2009

experiment, respectively, experience an upgrading or a downgrading.

Conversely, firms classified according to the DEA approach show a higher sensitivity to movements between classes: as we showed in the previous section, firms remaining in the same class represent 54.6% of the sample, while downward and upward movements involve, respectively, 23.6% and 21.8% of the sample.

In addition, our approach contrasts previous validations and comparisons of the two methodologies (Min and Lee 2008), as we do not use regression analysis to validate DEA scores either by simply regressing such scores with respect to the input and output variables used in the DEA optimization procedure, or by applying a logit (probit) regression to a dichotomous variable, derived from the application of a given cut-off point (e.g. the median value) to the distribution of DEA scores, dependent on the same explanatory variables used in the linear regression.

Indeed, such an approach is self-reinforcing and self validating as DEA scores are derived from an optimization process which uses the same variables then used in the regression analysis.

On the contrary, our validation approach compares the raw outcomes of the two procedures and thus enables us to state the advantages (strengths) and disadvantages (weaknesses) of the two methodologies clearly.

5.3 Conclusions

Firms' financial performance is crucial as it determines future decisions and actions which, in turn, affect growth at the micro (company) and macro (economy-wide) levels.

status	**DA classification**		**DEA classification**	
2009 respect to 2003	***n. of firms***	***Percent***	***n. of firms***	***Percent***
stable	2480	85.3	1739	54.6
downgrading	158	5.4	754	23.6
upgrading	270	9.3	694	21.8
Total	2908	100.0	3187	100.0

Table 12: Movements between risk classes

In particular, we have emphasized and reviewed how firms' financial performance may affect their investment decisions and, more importantly, innovation. Thus, it is crucial to be able to determine and classify a firm's financial worthiness.

We have therefore analyzed the performance of a representative sample of Italian manufacturing firms, by applying a parametric (logistic discriminant) and a non-parametric approach (Data Envelopment Analysis). The comparison between the different approaches is necessary, as the evaluation of a firm's financial performance is the result of fine tuning procedures which require the use of multiple methodological tools.

Discriminant Analysis (DA) is based on the assumption of a given distribution of a firm's default probability, which is assumed to be logistic. Such a procedure enables one to estimate and then forecast a firm's default probability. However, we have emphasized that one significant drawback lies in the fact that in order to estimate such probabilities one needs to gather information on firms which are already bankrupt. Typically, the number of these latter firms is relatively small compared with that of non-bankrupt firms; this fact produces a bias in that the estimated default probabilities are underestimated.

Thus, we have proposed a methodology that does not require ex-ante information on bankrupt firms. Data Envelopment Analysis (DEA) is a non-parametric approach which enables one to rank firms according to their efficiency or other measures of financial worthiness, by applying appropriate linear programming models. In particular, we have chosen the CCR input oriented model, which implies the minimization of given input variables for given outputs. This choice is based on some experiments which have enabled us to verify that the relationship between the inputs and outputs we have chosen is linear and, therefore, the constant return to scale hypothesis implied by the CCR model is not ruled out. In addition, the input-oriented model seems better adapted to the setting of firms' financial and economic problems.

Our results enable us to achieve a more comprehensive picture of firms' financial performance, and we are able to predict defaults of those bankrupted firms whose balance sheets were used in the discriminant analysis. The discriminant analysis, on the contrary, correctly predicts default probability in 75% of cases.

The analysis presented in this study therefore represents a fundamental and necessary background for investigating the aforementioned relationship between financial worthiness and investment and firms' performance, in particular, with respect to investment and innovation decisions.

A Bankrupted firms: stratification by area, age and size

	21 years and more	11-20 years	1-10 years	Total
North East				
>10 mill. Euros	1	2	4	**7**
1-10 mill. Euros	2	3	5	**10**
100 k - 999 k Euros	3	4	6	**13**
Sub-total	**6**	**9**	**15**	**30**
North West				
>10 mill. Euros	2	2	4	**8**
1-10 mill. Euros	2	3	5	**10**
100 k - 999 k Euros	3	4	7	**14**
Sub-total	**7**	**9**	**16**	**32**
Centre				
>10 mill. Euros	2	3	4	**9**
1-10 mill. Euros	3	4	6	**13**
100 k - 999 k Euros	4	5	8	**17**
Sub-total	**9**	**12**	**18**	**39**
South				
>10 mill. Euros	3	3	4	**10**
1-10 mill. Euros	3	5	8	**16**
100 k - 999 k Euros	5	7	11	**23**
Sub-total	**11**	**15**	**23**	**49**
TOTAL	**33**	**45**	**72**	**150**

Stratification weights

Area: North-East 1; North-West 1.078; Centre 1.255; South 1.627

Age: 1-10 years 2.31; 11-20 years 1.44; more than 20 years 1

Size (sales): 100 k- 999 k Euros 1.80; 1-10 mill. Euros 1.29; >10 mill. Euros 1

B Descriptive statistics by risk class. DEA - year 2009

Variable	Description	N. firms		Mean	Median	Min	Max
ACID	Acid test - Liquid Assets/Current Liabilities	3,845	all	1.20	0.94	0.20	5.75
		192	risky	0.77	0.68	0.20	3.60
		961	critical	0.88	0.79	0.27	4.46
		2,883	normal	1.07	0.92	0.21	5.22
		962	excellent	1.80	1.62	0.24	5.75
LEV	Leverage - Total Debts/Net Capital	3,845	all	4.25	1.97	-12.70	73.06
		192	risky	17.76	22.59	-12.70	73.06
		769	critical	10.91	8.13	0.54	60.89
		1,922	normal	2.51	2.04	0.15	47.51
		962	excellent	0.72	0.60	0.05	4.57
CN_A	Net Capital/Total Assets	3,845	all	0.32	0.30	-2.04	0.92
		192	risky	-0.10	0.02	-2.04	0.05
		769	critical	0.11	0.10	0.01	0.61
		1,922	normal	0.32	0.30	0.02	0.87
		962	excellent	0.57	0.57	0.17	0.92
CL_S	Current Liabilities/Sales	3,845	all	0.55	0.48	0.12	2.49
		192	risky	0.83	0.72	0.14	2.49
		769	critical	0.67	0.62	0.17	2.45
		1,922	normal	0.56	0.50	0.12	2.46
		962	excellent	0.38	0.30	0.12	2.48
CL_A	Current Liabilities/Total Assets	3,845	all	0.48	0.46	0.04	2.70
		192	risky	0.78	0.76	0.19	2.70
		769	critical	0.65	0.68	0.07	0.96
		1,922	normal	0.47	0.47	0.08	0.95
		962	excellent	0.31	0.30	0.04	0.77
ROS	Return on Sales	3,845	all	0.02	0.03	-0.46	0.28
		192	risky	-0.12	-0.10	-0.46	0.04
		769	critical	0.00	0.02	-0.45	0.15
		1,922	normal	0.02	0.03	-0.43	0.25
		962	excellent	0.05	0.04	-0.40	0.28
IR	Interest Payment/Sales	3,845	all	0.018	0.013	0.000	0.398
		192	risky	0.029	0.022	0.000	0.104
		769	critical	0.026	0.021	0.000	0.147
		1,922	normal	0.020	0.014	0.000	0.209
		962	excellent	0.007	0.002	0.000	0.398
TA	Total Assets (Euros)	3,845	all	17,436,284	6,293,267	736,368	321,386,959
		192	risky	15,643,710	3,910,625	813,616	303,830,920
		769	critical	11,918,528	4,441,134	736,368	261,845,800
		1,922	normal	19,322,470	7,218,616	740,810	320,259,996
		962	excellent	18,293,980	7,073,901	793,142	321,386,959
AGE	Firm Age in years	3,845	all	31	28	2	259
		192	risky	29	22	5	259
		769	critical	26	23	2	259
		1,922	normal	32	29	2	259
		962	excellent	34	30	2	159

Note: observations below the 1th or above the 99th percentile excluded.

C Descriptive statistics by risk class. DA - year 2009

Variable	*Description*	*N. firms*		*Mean*	*Median*	*Min*	*Max*
ACID	*Acid test - Liquid Assets/Current Liabilities*	3,424	*all*	1.20	0.94	0.24	5.75
		20	*risky*	0.54	0.55	0.26	1.04
		179	*critical*	0.72	0.69	0.24	1.67
		2,419	*normal*	1.06	0.90	0.25	4.85
		806	*excellent*	1.73	1.42	0.32	5.75
LEV	***Leverage - Total Debts/Net Capital***	3,424	***all***	4.33	2.02	0.12	73.06
		20	***risky***	20.50	10.10	0.61	71.58
		179	***critical***	11.03	6.10	0.49	73.06
		2,419	***normal***	4.49	2.33	0.14	64.50
		806	***excellent***	1.97	0.93	0.12	35.09
CN_A	*Net Capital/Total Assets*	3,424	*all*	0.33	0.30	0.01	0.88
		20	*risky*	0.13	0.09	0.01	0.58
		179	*critical*	0.17	0.13	0.01	0.63
		2,419	*normal*	0.30	0.27	0.01	0.85
		806	*excellent*	0.46	0.46	0.02	0.88
CL_S	***Liabilities/Sales***	3,424	***all***	0.54	0.48	0.12	2.17
		20	***risky***	1.13	1.05	0.43	1.78
		179	***critical***	0.79	0.74	0.19	2.10
		2,419	***normal***	0.55	0.50	0.12	2.17
		806	***excellent***	0.43	0.36	0.12	1.79
CL_A	*Liabilities/Total Assets*	3,424	*all*	0.47	0.46	0.04	0.96
		20	*risky*	0.63	0.64	0.22	0.93
		179	*critical*	0.62	0.63	0.21	0.95
		2,419	*normal*	0.50	0.49	0.05	0.96
		806	*excellent*	0.36	0.34	0.04	0.89
ROS	***Return on Sales***	3,424	***all***	0.02	0.03	-0.33	0.28
		20	***risky***	-0.14	-0.13	-0.31	0.04
		179	***critical***	-0.02	0.01	-0.33	0.10
		2,419	***normal***	0.02	0.03	-0.33	0.28
		806	***excellent***	0.04	0.04	-0.31	0.28
IR	*Interest Payment/Sales*	3,424	*all*	0.018	0.013	0.000	0.287
		20	*risky*	0.052	0.054	0.013	0.097
		179	*critical*	0.033	0.025	0.000	0.116
		2,419	*normal*	0.018	0.013	0.000	0.287
		806	*excellent*	0.013	0.007	0.000	0.159
TA	***Total Assets (Euros)***	3,424	***all***	17,184,112	6,222,189	758,370	321,386,959
		20	***risky***	6,503,191	4,043,623	1,902,179	24,423,864
		179	***critical***	9,879,416	3,578,353	762,110	279,764,100
		2,419	***normal***	13,891,942	5,677,906	758,370	321,386,959
		806	***excellent***	28,952,001	10,977,225	820,690	303,485,000
AGE	*Firm Age in years*	3,424	*all*	31	28	2	259
		20	*risky*	11	11	2	23
		179	*critical*	11	9	2	29
		2,419	*normal*	24	24	3	69
		806	*excellent*	57	49	6	259

Note: observations below the 1th or above the 99th percentile excluded.

References

[1] Allayannis, G., and Mozumdar, A.(2004), *The impact of negative cash flow and influential observations on investment-cash flow sensitivity estimates*, Journal of Banking & Finance, 28, 5, 901-930.

[2] Almeida, H., Campello, M., and Weisbach, Michael S. (2004), *The Cash Flow Sensitivity of Cash*, Journal of Finance, 59, 4, 1777-1804.

[3] Altman, Edward I., Marco, G., and Varetto F. (1994), *Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)*, Journal of Banking & Finance, 18, 3, 505-529.

[4] Altman, Edward I., Haldeman Robert G., and Narayanan P. (1977), *A New Model to Identify Bankruptcy Risk of Corporation*, Journal of Banking & Finance, 1, 1, 29-54.

[5] Altman, Edward I. (1968), *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, Journal of Finance, 23, 4, 589-609.

[6] Calabrese, R., and Osmetti Silvia A. (2011), *Generalized Extreme Value Regression for Binary Rare Events Data: an Application to Credit Defaults*, UCD Geary Institute Discussion paper series 2011/20.

[7] Canepa, A., and Stoneman, P. (2008), *Financial constraints to innovation in the UK: evidence from CIS2 and CIS3*, Oxford Economic Papers, 60, 711-730.

[8] Cerved (2007), *I fallimenti in Italia e in Europa*, Focus - Analisi ed approfondimenti, luglio 2007.

[9] Cerved (2011), *Caratteristiche delle imprese, governance e probabilità di insolvenza*, Rapporti Flash, febbraio 2011.

[10] Charnes, A., Cooper, W. W., and Rhodes, E. (1978), *Measuring the efficiency of decision making units*, European Journal of Operational Research, 2, 6, 429-444.

[11] Cielen, A., and Vanhoof, K. (1999), *Bankruptcy prediction using a data envelopment analysis*, Manuscript, Limburg University, Diebenpeek.

[12] Cleary, M. (1999), *The Relationship between Firm Investment and Financial Status*, Journal of Finance, 54, 2, 673-692.

[13] Deakin, Edward B. (1972), *A Discriminant Analysis of Predictors of Business Failure*, Journal of Accounting Research, 10, 1, 167-179.

[14] Fazzari, S.M., Hubbard, R. G., and Petersen, B.C. (1988), *Financing Constraints and Corporate Investment*, Brookings Papers on Economic Activity, 1, 141-205.

[15] Gilchrist, S., and Himmelberg, Charles P. (1995), *Evidence on the role of cash flow for investment*, Journal of Monetary Economics, 36, 3, 541-572.

[16] Giudici, G., and Paleari, S. (2000), *The Provision of Finance to Innovation: A Survey Conducted among Italian Technology-based Small Firms*, Small Business Economics, 14, 37-53.

[17] Hall, B. (2002), *The financing of research and development*, Oxford Review of Economic Policy, 18, 35-51.

[18] Hadlock, Charles J., and Pierce, Joshua R. (2010), *New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index*, Oxford Review of Economic Policy, 23, 5, 1909-1940.

[19] Hadlock, Charles J., and Pierce, Joshua R. (2010), *New Evidence on Measuring Financial Constraints: Moving Beyond the KZ Index*, Oxford Review of Economic Policy, 23, 5, 1909-1940.

[20] Kaplan, Steven N., and Zingales, L.(1997), *Do Financing Constraints Explain Why Investment is Correlated with Cash Flows*, Quarterly Journal of Economics, 112, 1, 169-215.

[21] Lamont, O., Polk, C., and Saa´ -Requejo, J. (2001), *Financial Constraints and Stock Returns*, Review of Financial Studies, 14, 529-544.

[22] Linde, J., and Jacobson, T. (2011), *Firm Default and Aggregate Fluctuations*, FRB International Finance Discussion Paper No. 1029.

[23] Min, Jae H., and Lee, Y. (2008), *A practical approach to credit scoring*, Expert Systems with Applications: An International Journal, 35, 4, 1762-1770.

[24] Ohlson James A. (1980), *Financial Ratios and the Probabilistic Prediction of Bankruptcy*, Journal of Accounting Research, 18, 1, 109-131.

[25] Saunders A., and Allen L. (2002), *Credit Risk measurement*, Wiley and Sons, Inc., New York.

[26] Shumway , T.(2001), *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*, Journal of Business, 74, 1, 101-124.

[27] Simak, P.C.(1999), *DEA based analysis of corporate failure*, Manuscript, University of Toronto, Toronto.

[28] Unicredit (2009), *Decima Indagine sulle imprese manifatturiere italiane*, Rapporto Corporate, n.1.

[29] Whited, Toni N., and Wu, G. (2006), *Financial Constraints Risk*, Review of Financial Studies, 19, 2, 531-559.

[30] Yeh, Quey-jen (1996), *The Application of Data Envelopment Analysis in Conjunction with Financial Ratios for Bank Performance Evaluation*, The Journal of the Operational Research Society, 47, 8, 980-988.

[31] Troutt, M.D., Rai, A., and Zhang, A. (1996) *The potential use of DEA for credit applicant acceptance systems*, Computers and Operations Research, 23, 4, 405-408.