European Journal of Business and Management
ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online)
Vol.7, No.2, 2015

Selection of Determinants in Corporate Financial Distress

Luca Sensini
University of Salerno, Italy, Via Giovanni Paolo II, 132 – 84084 Fisciano (SA)
E-mail: lsensini@unisa.it

Abstract
This paper investigates the capability of forecasting models for bankruptcy prediction referring to annual balance sheet information of Italian firms in the limited liability sector. The performance of default risk models in terms of forecast accuracy is mainly related to the selection of the set of best predictors. Therefore our main research question refer to the identifications of the determinants of corporate financial distress, comparing the performance of innovative selection techniques. Furthermore, several issues involved in default risk analysis are considered, such as the structure of the data-base and the sampling procedure The predictive performance of the proposed default risk model has been assessed by means of different accuracy measures. The results of the analysis, carried out on a data-set of financial ratios expressly created from a sample of industrial firms annual reports, give evidence in favour of the proposed model over traditional ones.
Keywords: Default Risk, Bankruptcy, Variable Selection, Lasso.

1. Introduction

Business failure is one of the most investigated topics in corporate finance and the empirical approach to bankruptcy prediction has recently gained further attention from financial institutions, mainly due to the increasing availability of financial information (Agarwal and Taffler, 2008; Alberici, 1975; Altman, 2006; Amendola et al., 2011; Becchetti, Sierra, 2003; Bisogno, 2012, Hotchkiss, 2006; Platt & Platt, 2002; Quagli, 1990; Teodori, 1994).

Starting from the seminal paper of Beaver (1966), that first proposes to use financial ratios as failure predictors in a univariate context, and from the following paper of Altman (1968), that suggests a multivariate approach based on discriminant analysis, there have been many contributions to this field (Balcæn and Ooghe, 2006; Ohlson, 1980, Poddighe and Madonna, 2006; Ravi Kumar and Ravi, 2007).

In addition to the Multivariate Discriminant Analysis (MDA), different statistical approaches have been declared throughout the years, such as Logit and Probit models (Ohlson, 1980; Zmijewski, 1984; Lennox, 1999), classification trees and artificial neural network (Wilson and Sharda, 1994; Serrano, 1997; Charalambous et al., 2000; Perez, 2006). Furthermore, the development of computer intensive methods has led to the use of machine learning techniques (Hardle et al., 2009).

In spite of numerous empirical findings, significant issues still remain unsolved, such as arbitrary definition of failure; non-stationarity and instability of data; choice of the optimization criteria; sample design and variable selection. Furthermore, despite the increasing number of data warehouse, it is not an easy task to collect data on a specific set of homogeneous firms related to a definite geographic area or a small economic district.

The focus of this paper is to investigate different aspects of bankruptcy prediction, focusing in particular on the variable selection problem.

In corporate failure prediction, the purpose is to have a methodological approach which discriminates firms with a high probability of future failure from those which could be considered to be healthy, using a large number of financial indicators as potential predictors. In order to select the relevant information, several selection methods can be applied, leading to different optimal predictions set.

In Amendola et al. (2011) we have proposed to use modern selection techniques based on penalized regression models and compare their performance over traditional variable selection methods.

In this paper, the analysis have been extended on a larger sample of industrial firms of the South of Italy, aims at evaluating the capability of a regional model to improve the forecasting performance over different optimal prediction sets and different sampling approaches. An out-of-sample validation procedure has been implemented by means of properly chosen accuracy measures.

The structure of the paper is as follows. The next section introduces sample characteristics and data-set. Section 3 briefly illustrates the variable selection techniques. The proposed models are described in section 4, while the results of the prediction power's comparison of the different models are reported in Section 5. The final section will give some concluding remarks.

2. The data

In the literature business failure has been defined in many different ways, although it there is not a widely accepted definition (Crutzen and van Caillie, 2007).

In many studies, business failure is defined as a series of different situations that lead to the closing down of the firm due to relevant financial problems (Morris, 1997). However, this definition only concentrates on the financial disease without taking into account other difficulties that can affect the firms' health in the early stages of the failure process (Argenti, 1976).

Therefore, it is necessary to clarify the meaning of business failure our study refers to. In a predictive prospective, the empirical literature distinguishes between two main aspects of the definition of business failure: economic and juridical.

In this paper the juridical concept have been considered, focusing on those companies that have experienced permanent financial disease, not including companies with temporary financial problems or companies which, for any reasons, have voluntarily chosen liquidation.

The data-set includes industrial companies that had undertaken the juridical procedure of bankruptcy in the South of Italy in a given time period, t. The information on the legal status and the annual reports was extracted from the Infocamere database and the AIDA database of Bureau Van Dijk (BVD).

In particular, the disease set is composed of those industrial firms that had entered the juridical procedure of bankruptcy in the South of Italy at t=2008, for a total of 486 failed firms and five years of financial statement information prior to failure (t- i; i = [1; 5]). Not all the firms in the dataset provide full information suitable for the purpose of our analysis. In order to evaluate the availability and the significance of the financial data, a preliminary screening was performed (Table 1) dividing, for each year of interest, the whole population of failed firms into two groups: firms that provided full information (i.e. have published their financial statements) and firms with incomplete data (i.e. did not present their financial statements, presented an incomplete report or stopped their activity) (Amendola et al., 2010).

Table 1. Failed firms sample

	2003	2004	2005	2006	2007
Published Statement	416	367	306	245	171
Total Firms	486	486	486	486	486
Percentage	85,60%	75,51%	62,96%	50,41%	35,19%

We chose the year 2008 as a reference period, t, in order to have at least 4 years of future annual reports (at t + i; i = [1; 4]) to assure that the company selected as healthy at time t does not get into financial problems in the next 4 years.

The healthy set was randomly selected among the South Italy industrial firms according to the following criteria: were still active at time t; have not incurred in any kind of bankruptcy procedures between 2008 and 2012; had provided full information at time (t - i; i = [1; 4]) and (t + i; i = [0; 4]).

In order to have a panel of full information, i.e. each firm provides complete financial data for each time period t, the analysis has been limited to the three years of interest (2004, 2005, 2006).

One of our aims is to investigate the performance of the developed default risk models over different sample designs. The relation between forecasting performance and sample choice has been debated in the literature without ending up with a clear evidence in favor of a unique solution.

A common approach is to adopt a balanced-sample, by choosing the same sample size for both classes of failure and healthy firms. The reason is that the population proportion significantly favours active firms and so a non-balanced sample would select a reduced number of failed firms and might lead to a biased estimator. In addition, the true proportion among the two conditions is not easy to calculate in practice (Cortes et al., 2008).

However, there are also reasons in favour of different choices, such as oversampling the failing companies with unbalanced proportion (Back, 1997).

The sampling procedure for selecting the panel data set is based on both balance and unbalanced cluster sampling designs. The cluster scheme refers to the geographical distribution of the industrial firm. A cross-sectional approach is considered as benchmark.

The predictors data-base for the three years of interest (2004, 2005, 2006) was elaborated starting from the financial statements of each firm included in the sample. We computed nv = 55 indicators selected as potential bankruptcy predictors among the most relevant in highlighting current and prospective conditions of operational unbalance (Altman, 2000; Dimitras et al., 1996).

Table 2. Financial indicators and financial area

N.	Financial Indicator	Area
1	Net Proceeds/Invested Capital	Profitability
2	Return on Equity	Profitability
3	Return on Investment	Profitability
4	Return on Assets	Profitability
5	Return on Sales	Profitability
6	Net Proceeds/Current Assets	Profitability
7	Leverage	Profitability
8	Liquidity/Total Assets	Liquidity
9	Current Ratio I	Liquidity
10	Current Ratio II	Liquidity
11	Quick Ratio	Liquidity
12	Equity Ratio	Size and Capitalization
13	Net Worth/Capital Stock	Size and Capitalization
14	Equity - Intangible Assets	Size and Capitalization
15	Gross Income/Financial Charges	Profitability
16	Net Capital - Net Capital Assets	Size and Capitalization
17	Net Worth/Sales	Size and Capitalization
18	Capital Stock/Sales	Profitability
19	Inventory/Sales	Turnover ratios
20	Total Debts/Total Assets	Size and Capitalization
21	Net Worth/Fixed Assets	Size and Capitalization
22	Capital Stock/Fixed Assets	Size and Capitalization
23	Current Assets/Fixed Assets	Liquidity
24	Inventory/Current Assets	Liquidity
25	Gross Working Capital/Total assets	Liquidity
26	Capital assets/Total Assets	Size and capitalization
27	Liquid Assets/Total Assets	Liquidity
28	Net Worth/Total Assets	Size and capitalization
29	Capital Stock/Total Assets	Size and Capitalization
30	Net Worth/Total Debts	Size and Capitalization
31	Capital Stock/Total Debts	Size and Capitalization
32	Financial Debt /Total Assets	Size and Capitalization
33	Cash Flow	Liquidity
34	Cash Flow/Sales	Profitability
35	Cash Flow/Total Assets	Liquidity
36	Cash Flow/Net Worth	Liquidity
37	Cash Flow/Capital Stock	Liquidity
38	Cash Flow/Total Debts	Liquidity
39	Cash/Sales	Liquidity
40	Account Receivable/Sales	Turnover ratios
41	Total Debts/Sales	Turnover ratios
42	Net Income/Sales	Profitability
43	Net Income/Total Assets	Profitability
44	Net Income/Total Debts	Profitability
45	Sales/Fixed Assets	Profitability
46	Sales/Advances from Customers	Turnover ratios
47	Sales/Inventory	Turnover ratios
48	Sales/Total Assets	Profitability
49	Labour Cost/Production Cost	Operating structure
50	Labour Cost/Production Value	Operating structure
51	Labour Cost/Net Sales	Operating structure
52	Finance Charges/Debt	Operating structure
53	Finance Charges/Financial Debt	Operating structure
54	Finance Charges/Production Value	Profitability
55	Finance Charges/Net Sales	Profitability

The explanatory variables considered in the analysis have been chosen on the basis of a few different criteria. They have a relevant financial meaning in a failure context, and have been commonly used in failure

predictions literature, and also the information needed to calculate these ratios is available.
Furthermore, the selected indicators reflect different aspects of the firms' structure, as synthesized in Table 3.

Table 3. Financial Predictors

Area	Nv
Liquidity	14
Operating structure	5
Profitability	17
Size and Capitalization	14
Turnover	5

For each sample set, the 70% of the observations has been included in the training data set used for estimating the forecasting models, while the remaining 30% has been selected for the test set used for evaluating the predictive power of those models.

3. Selection techniques

A relevant problem, for the analysts who attempt to forecast the risk of failure, is to identify the optimal subset of predictive variables. This has been perceived as a real challenge since Altman (1968) and largely debated both in the financial literature and in the more general context of variable selection.

Different selection procedures have been proposed over the years, mainly based on: personal judgment; empirical and theoretical evidence; meta-heuristic strategies; statistical methods. We focused our attention on the last group developed in the context of regression analysis. Goals in variable selection include: accurate predictions, predictors easily to interpret and scientifically meaningful, robustness (i.e. small changes in the data should not result in large changes in the subset of predictors used).

One of the widely used techniques in this domain is the subset regression, which aims at choosing the set of the most important predictors to be included in the model. In this class we can allow different methods: all-subset; forward (backward) selection; stepwise selection.

More specifically, forward stepwise regression begins by selecting a single predictor variable which produces the best fit given a collection of possible predictors. Another predictor, which produces the best fit in combination with the first, is then added, and so on. This process continues until some stopping criteria are reached. The process is aggressive and unstable, in that may eliminate useful predictors in the early steps and relatively small changes in the data might cause one variable to be selected instead of another, after which subsequent choices may be completely different.

In contrast, all-subsets regression is exhaustive, considering all subsets of variables of each size, limited by a maximum number of best subsets (Furnival and Wilson, 1974). The advantage over stepwise procedure is that the best set of two predictors does not include the predictor that was best by itself. The disadvantage is that biases in inference are even greater, because it considers a much greater number of possible models.

These traditional methods focus on variable selection, rather than estimating coefficients. A different approach is given by the penalized regression methods.

They allow a variable to be partly included in the model via constrained least squares optimization. That is, the variable is included but with a shrunken coefficient. Shrinkage often improves prediction accuracy, trading off decreasing variance for increased bias (Hastie, Tibshirani and Friedman, 2009).

Among this frame, a widely used approach is the Least Absolute Shrinkage and Selection Operator, LASSO proposed by Tibshirani (1996) defined as:

$$\hat{\beta}_{lasso} = \text{argmin} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\, \beta_j \right)^2$$

Subject to $\sum_{j=1}^{p} \left| \beta_j \right| \leq \delta$.

The Lasso allows for simultaneous execution of both parameter estimation and variable selection. It shrinks some coefficients in the linear regression and sets others to 0, and hence tries to retain the good features of both subset selection and ridge regression. Since a small value of the threshold δ or a large value of the penalty term λ will set some coefficients to be zero, therefore the Lasso performs a kind of continuous subset selection. Correlated variables still have a chance to be selected. The Lasso linear regression can be generalized to other models, such as GLM, hazards model, etc. (Park and Hastie, 2007). In the early stage, when it was first proposed, the Lasso techniques have not had a large diffusion because of the relatively complicated computational algorithms. This has been overcome by more recent proposals.

4. Default-risk models and performance evaluation

The main aim of the analysis is in developing forecasting models for the predictions and diagnosis of the risk of bankruptcy, addressing the capability of such models of evaluating the discriminant power of each indicator and selecting the best optimal set of predictors.

For this purpose we compared different selection strategies, evaluating their performances in terms of predicting the risk that an industrial enterprise would incur in legal bankruptcy, for different sample sets and at different time points.

In particular, we considered the traditional Logistic Regression with a stepwise variable selection (Model 1) and the regularized Logistic Regression with a Lasso selection (Model 2). As benchmark we estimated a Linear Discriminant Analysis with a stepwise selection procedure (Model 3).

The logistic regression can be written as:

$$ln\left(\frac{p(y)}{1-p(y)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

The Regularized logistic Regression consider the penalty term, as illustrated in the previous session, via the Lasso.

For the evaluation of the estimation and forecasting performance of the considered models we refer to the literature on Classification techniques.

The classification results can be summarized in a two-by-two confusion matrix (also called a contingency table) representing the dispositions of the set of instances (Table 4). In particular, given a classifier and an instance (firm), there are four possible outcomes:

- True Positive: a failed firm classified as failed;
- False Negative: a failed firm classified as healthy;
- True Negative: a healthy firm classified as healthy;
- False Positive: a healthy firm classified as failed.

Table 4. Confusion Matrix

		Predicted Class	
		Failed	Healthy
Actual	*Failed*	True Positive	False Negative
Class	*Healthy*	False Positive	True Negative

From this framework two types of error can be defined: the Type I error rate, i.e. a failing firm is misclassified as a non-failing firm, and the Type II error rate, i.e. a non-failing firm is wrongly assigned to the failing group. An overall index, the Correct Classification Rate, (CCR), i.e. correct classified instances over total instances, can be computed.

The results of this matrix are the input data for some accuracy measures, widely used in a bankruptcy prediction study (Engelmann et al., 2003; Fawcett, 2006). A first approach is based on the Cumulative Accuracy Profile (CAP) and its summary statistic, the Accuracy Ratio, calculated by relating the area under the CAP plot to the area under the CAP of a hypothetical "perfect" rating system.

A different approach is based on the Receiver Operating Characteristics (ROC) analysis that shows the ability of the classifier to rank the positive instances relative to the negative instances. Although the construction of the ROC curve differs from the CAP approach, the summary measures of both curves essentially contain the same information. The Area under the ROC curve (AUC) can be defined as the probability that the classifier will rank a randomly chosen failed firm higher than a randomly chosen solvent company.

It can be shown that the Accuracy Ratio can be also calculated referring to the Area under the ROC curve with following equation:

$$AR = 2 * AUC - 1.$$

The Accuracy Ratio is normalized between -1 and 1, while the Area under the ROC curve lies between 0 and 1. The area is 1.0 for a perfect model. Testing the performance of a default model means to investigate its ability to discriminate between different levels of default risk. The outcomes of the performance measures strongly depend on the overall framework such as the structure of the true default probabilities in the underlying portfolio, the time of default, etc. Clearly, comparisons of different classification techniques have to be referred to the same point in time and for a given sample data.

5. Empirical Results

The predictive performance of the developed models has been evaluated by means of training and test sets, considering appropriate accuracy measures. Namely, we compare the results in terms of: Correct Classification Rate (CCR); Area under the ROC curve (AUC); Accurancy Ratio (AR).

The accuracy measures have been computed on the training and test sets for each forecasting model, previously described (Model 1, Model 2 and Model 3) and each sample design.

For the unbalanced sample (Table 5-6), the correct classification rate of the three models increases as approaching the bankruptcy year, both in training set and in test set. Looking at the Type I and II error rates, it can be noted that in the training set, the Type I error rate of Logistic Model has a non-steady trend. In fact, it increases from 2004 to 2005, but decreases from 2005 to 2006, while the Type II error rate has a constant progress. For the other two models (Lasso and Discriminant Analysis), in the training set, the trend of the two errors is steady, while in the test set they do not have a constant increasing or decreasing behavior. Though the two error rates do not have a uniform trend, the values of the AUC and the AR show an improvement in the prediction accuracy, as the failure time is approaching.

An exception is the values of the Logistic Regression model in training set. The effect of the sample design seems to be no so relevant, in fact the trend of the accuracy measures for the balanced sample (Table 7-8), is quite similar to that in the unbalanced sample. Looking at the error rates, the values for the balance sample are on average slightly worse than the unbalanced.

Now, comparing the performance of the three models, it can be noted that the Lasso has a better performance in each year, in both sets and for both samples, compared to Logistic Regression and Discriminant Analysis.

Thus, the forecasting accuracy of Model 2 (Lasso Regression) in both balanced and unbalanced settings, is higher if compared with Logit and Discriminant Analysis for almost all the time intervals considered.

The results give evidence in favor of forecasting models based on unbalanced sample and shrinkage selection methods. The Lasso procedure leads to more stable results and gives advantage also in terms of computational time and number of variables selected as predictors. Overall, the models performance increases, as the forecasting horizon decreases even if some drawbacks can be registered for the Logistic Regression in the year 2005. The indicators selected as predictors for the three estimated models are in line with those included, at different levels, in many other empirical studies (Amendola et al., 2010; Dimitras et al., 1996).

Table 5. Unbalanced sample: Accuracy measures for training set

	Model 1	Model 2	Model 3
		2004	
Correct Classification Rate	0.82651	0.87322	0.82695
Miss Classification Rate	0.17349	0.12678	0.17305
Type I Error	0.33264	0.36185	0.56124
Type II Error	0.09092	0.00253	0.02398
AUC	0.85692	0.92711	0.84567
AR	0.75469	0.87422	0.66773
		2005	
Correct Classification Rate	0.83362	0.89803	0.87702
Miss Classification Rate	0.16638	0.10197	0.12298
Type I Error	0.38365	0.23856	0.33286
Type II Error	0.04985	0.02199	0.03388
AUC	0.85402	0.95842	0.91120
AR	0.71804	0.91868	0.86238
		2006	
Correct Classification Rate	0.92443	0.93262	0.89425
Miss Classification Rate	0.07557	0.06738	0.10575
Type I Error	0.12285	0.12390	0.24518
Type II Error	0.34462	0.02088	0.04498
AUC	0.95269	0.95480	0.92344
AR	0.91894	0.92850	0.90688

Table 6. Unbalanced sample: Accuracy measures for test set

	Model 1	Model 2	Model 3
		2004	
Correct Classification Rate	0.7500	0.85428	0.79436
Miss Classification Rate	0.2500	0.14572	0.20564
Type I Error	0.45556	0.42206	0.72223
Type II Error	0.15218	0.03705	0.00022
AUC	0.69632	0.90172	0.66746
AR	0.40266	0.83594	0.34896
		2005	
Correct Classification Rate	0.87526	0.89424	0.81765
Miss Classification Rate	0.12474	0.10576	0.18235
Type I Error	0.25667	0.25667	0.53333
Type II Error	0.07906	0.05205	0.05106
AUC	0.92496	0.96287	0.84602
AR	0.84576	0.92585	0.68205
		2006	
Correct Classification Rate	0.92506	0.94682	0.90685
Miss Classification Rate	0.07494	0.05318	0.09315
Type I Error	0.05987	0.05987	0.43333
Type II Error	0.07906	0.0000	0.0000
AUC	0.97654	0.99857	0.96852
AR	0.93813	0.99219	0.94215

Table 7. Balanced sample: Accuracy measures for training set

	Model 1	Model 2	Model 3
		2004	
Correct Classification Rate	0.84285	0.87142	0.79561
Miss Classification Rate	0.15715	0.12858	0.20439
Type I Error	0.10829	0.13986	0.16892
Type II Error	0.21125	0.11249	0.24845
AUC	0.90812	0.93182	0.87581
AR	0.84026	0.89284	0.77156
		2005	
Correct Classification Rate	0.78918	0.88543	0.86956
Miss Classification Rate	0.21082	0.11457	0.13044
Type I Error	0.21757	0.11457	0.13886
Type II Error	0.24876	0.11457	0.11457
AUC	0.86455	0.93681	0.88532
AR	0.72268	0.89184	0.79852
		2006	
Correct Classification Rate	0.93627	0.96854	0.96214
Miss Classification Rate	0.06373	0.03146	0.03786
Type I Error	0.08479	0.00000	0.05618
Type II Error	0.05692	0.05692	0.031847
AUC	0.97121	0.99165	0.98364
AR	0.95128	0.98564	0.95974

Table 8. Balanced sample: Accuracy measures for Test set

	Model 1	Model 2	Model 3
		2004	
Correct Classification Rate	0.77876	0.82505	0.74565
Miss Classification Rate	0.22124	0.17495	0.25435
Type I Error	0.27667	0.27667	0.35453
Type II Error	0.20125	0.13333	0.20125
AUC	0.75889	0.91425	0.75667
AR	0.52868	0.84446	0.48956
		2005	
Correct Classification Rate	0.82000	0.90000	0.83333
Miss Classification Rate	0.18000	0.10000	0.16667
Type I Error	0.13333	0.13333	0.06667
Type II Error	0.26667	0.09667	0.02667
AUC	0.89333	0.96444	0.90778
AR	0.77889	0.93789	0.80556
		2006	
Correct Classification Rate	0.84000	0.94333	0.90000
Miss Classification Rate	0.16000	0.05667	0.10000
Type I Error	0.19333	0.05667	0.13333
Type II Error	0.12333	0.05667	0.05667
AUC	0.90333	0.99456	0.95224
AR	0.78667	0.98911	0.89456

Table 9. Cross-Sectional sample: Accuracy measures for training set

	Model 1	Model 2	Model 3
Correct Classification Rate	0.88651	0.93895	0.88467
Miss Classification Rate	0.11349	0.06105	0.11533
Type I Error	0.28619	0.16238	0.33286
Type II Error	0.06854	0.01246	0.02986
AUC	0.93219	0.97827	0.91861
AR	0.86219	0.96253	0.84182

Table 10. Cross-Sectional sample: Accuracy measures for Test set

	Model 1	Model 2	Model 3
Correct Classification Rate	0.83205	0.96821	0.86027
Miss Classification Rate	0.16795	0.03179	0.13973
Type I Error	0.33333	0.06667	0.35778
Type II Error	0.13607	0.01786	0.05464
AUC	0.84291	0.98251	0.88207
AR	0.68082	0.97257	0.74973

6. Concluding remarks

In this study the Regional industrial enterprise default risk models have been developed by investigating the role of variable selection procedures and sample designs in the overall forecasting performance. A data-set of financial statements of balanced and unbalanced samples of companies in South Italy for a given time period have been analysed. To select the two classes of healthy and failed firms, we used the concept of legal failure to include those firms which had gone bankrupt during the year 2008. In particular, the opportunity to implement shrinkage techniques in defining the optimal predictions set has been evaluated. The performance of the proposed forecasting models has been evaluated at different time horizons and by means of properly chosen accuracy measures. From the reached results, we find that models based on a Lasso selection procedure significantly outperform the traditional methods, specifically logistic regression and discriminant analysis, and are more stable in terms of error rates.

This can be observed for both balanced and unbalanced sample, highlighting the marginal effect of the sample design. Therefore, the proposed approach seems to be a promising and valid alternative. As expected by the dynamical nature of the problem, the overall performance depends on the time horizon. Taking into account the time dimension and the evolutionary behaviour of the financial variables may leads to better results in terms of forecast accuracy. Furthermore the empirical finding can be generalized extending the analysis to a larger data set including a wide geographic area such as the whole country.

References

Agarwal, V., Taffler, R. (2008), "Comparing the Performance of Market-based and Accounting based Bankruptcy Prediction Models", Journal of Banking and Finance, vol. 32, n° 8, 1541-1551.
Alberici A. (1975), Analisi dei bilanci e previsione delle insolvenze, Isedi.
Altman E.I. (1968), Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy, The Journal of Finance, 23, 4, 589-609.
Altman E.I. (2000), Predicting financial distress of companies: revisiting the Z-score and ZTM model, New York University, Working Paper.
Altman E.I., Kotchkiss (2006), Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distresses Debt, Wiley, 3rd ed.
Amendola A., Bisogno M., Restaino M., Sensini L. (2010), Forecasting corporate bankruptcy: an empirical analysis on industrial firms in Campania, Rirea, 2, 229-241.
Amendola A., Bisogno M., Restaino M., Sensini L. (2011), Forecasting corporate bankruptcy: empirical evidence on Italian data, EuroMed Journal of Business, Vol. 6, Issue 3, 294-312.
Amendola A., Restaino M., Sensini L. (2011), Variable Selection in Default Risk Models, The Journal of Risk Model Validation, Vo. 5, n.1, 3-19.
Argenti J. (1976), Corporate Collapse: the causes and symptomos, Holsted Press, McGraw-Hill, London.
Back B., Laitinen T., Hekanaho J., Sere K. (1997), The Effect of Sample Size on Different Failure Prediction Methods, Turku Centre for Computer Science, TUCS Technical Report, n. 155, 1-26.
Balcæn S., Ooghe H. (2006), 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems, The British Accounting Review, 38, 63-93.
Beaver W.H. (1966), Financial Ratios as Predictors of Failure, Journal of Accounting Research, Vol. 4, Empirical Research in Accounting: Selected Studies 1966, 71-111.
Becchetti L., Sierra J. (2003), Bankruptcy risk and productive efficiency in manufacturing firms, Journal of Banking and Finance, 27, 2099-2120.
Bisogno M. (2012), I modelli di previsione delle insolvenze. Profili teorici e applicazioni empiriche in ambito giuridico, Franco Angeli.
Charalambous C., Charitou A., Kaorou Y.F. (2000), Comparative analysis of artificial neural network models: applications in bankruptcy prediction, Annals of Operations Research, 99, 4, 403-425.
Cortés E.A., Martinez M.G., Rubio N.G. (2008), FIAMM return persistence analysis and determinants of the fees charged, Revista Española De Financiación Y Contabilidad, XXXVII, 13-32.
Crutzen N., Van Caillie D. (2007), The business failure process: towards an integrative model of the literature, Ecole de Gestion de l'Université de Liège Working Paper, 200710/03.
Dimitras A., Zanakis, S., Zopoudinis, C. (1996), A survey of business failures with an emphasis on failure prediction methods and industrial applications, European Journal of Operational Research 90, 3, 487-513.
Engelmann B., Hayden E., Tasche D. (2003), Testing rating accuracy, RISK, 16, 82-86.
Fawcett T. (2006), An introduction to ROC analysis, Pattern Recognition Letters, 27, 861-874.
Friedman J.H., Hastie T., Tibshirani R. (2009), Regularization Paths for Generalized Linear Models via Coordinate Descent, Journal of Statistical Software, 33, 1, 1-22.
Furnival G.M., Wilson R.W. (2000), Regression by Leaps and Bounds, Technometrics, 42, 1, 69-79.
Hastie T., Tibshirani R., Friedman J.H. (2009), The Elements of Statistical Learning: Data mining, inference and prediction, Springer.
Härdle W., Lee Y., Schafer D., Yeh Y. (2009), Variable Selection and Oversampling in the Use of Smooth Support Vector Machines for Predicting the Default Risk of Companies, Journal of Forecasting, 28, 6, 512-534.
Lennox C. (1999), Identifying Failing Companies: a revaluation of the Logit, Probit and DA approaches, Journal of Economics and Business, 51, 347-364.
Morris R. (1997), Early Warning Indicators of Corporate Failure: a critical review of previous research and further empirical evidence, Ashgate Publishing, Great Britain.
Ohlson J. (1980), Financial rations and the probabilistic prediction of bankruptcy, Journal of Accounting Research, 18, 1, 109-131.
Park M.Y., Hastie T. (2007), An L1 regularization path algorithm for generalized linear models, Journal of the Royal Statistical Society, Series B, 69, 4, 659-677.
Perez M. (2006), Artificial neural networks and bankruptcy forecasting: a state of the art, Neural Computing and Application, 15, 2, 154-163.
Platt H.D., Platt M.B. (2002), Predicting corporate financial distress: reflections on choice-based sample bias, Journal of Economics and Finance, Vol. 26, Issue 2, 184-199.
Poddighe F., Madonna S. (2006), I modelli di previsione delle insolvenze: possibilità e limiti, Giappichelli,

Torino.
Quagli A. (1990), Modelli di previsione delle insolvenze, in Amministrazione e Finanza, 17.
Ravi Kumar P., Ravi V. (2007), Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review, European Journal of Operational Research, 180, 1, 1-28.
Serrano C. (1997), Feed forward neural networks in the classification of financial information, European Journal of Finance, 3, 3, 183-202.
Teodori C. (1994), I modelli di previsione delle insolvenze e l'amministrazione controllata, Rivista dei dottori commercialisti, 45, 75-101.
Tibshirani R. (1996), Regression Shrinkage ad Selection via the Lasso, Journal of the Royal Statistical Society, Series B, 58, 1, 267-288.
Wilson R.L. Sharda R. (1994), Bankruptcy prediction using neural network, Decision Support Systems, 11, 545-557.
Zmijewsky M.E. (1984), Methodological issues related to the estimation of financial prediction models, Journal of Accounting Research, 22, 59-82.