Università degli Studi di Salerno
DIPARTIMENTO DI SCIENZE ECONOMICHE E STATISTICHE

Alessandra Amendola [1] Marialuisa Restaino [2]
Luca Sensini [3]

VARIABLE SELECTION IN FORECASTING MODELS FOR CORPORATE BANKRUPTCY

WORKING PAPER n. 3.216

[1]Department of Economics and Statistics, Università di Salerno, Italy, email: alamendola@unisa.it.
[2]Department of Economics and Statistics, Università di Salerno, Italy, email: mlrestaino@unisa.it (Correspondent author).
[3]Department of Management Research, Università di Salerno, Italy, email: lsensini@unisa.it.

Contents

Abstract[1]

In this paper we develop statistical models for bankruptcy prediction of Italian firms in the limited liability sector, using annual balance sheet information. Several issues involved in default risk analysis are investigated, such as the structure of the data-base, the sampling procedure and the influence of predictors. In particular we focus on the variable selection problem, comparing innovative techniques based on shrinkage with traditional stepwise methods. The predictive performance of the proposed default risk model has been evaluated 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 favor of the proposed model over traditional ones.

Keywords

Forecasting, Default Risk, Variable Selection, Shrinkage, Lasso.

[1]The paper is the result of the joint collaboration of all the authors; § 1 and § 4 were written by Alessandra Amendola, § 3, § 6 and § 7 were written by Marialuisa Restaino, § 2 e § 5 were written by Luca Sensini.

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.

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 (recent reviews are Balcæn and Ooghe, 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 lead to recent contributions to the use of machine learning techniques (Härdle *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.

Our aim 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. We proposed to use modern selection techniques based on shrinkage and compare their performance with traditional variable selection methods.

The analysis, carried out on a sample of industrial firms throughout the Campania region, 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 on panel and cross-sectional data sets 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 at different horizons are reported in Section 5. The final section will give some concluding remarks.

2 Data Base and Predictors

The notion of business failure has been defined in many different ways in literature although it is not easy to agree on 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, we have chosen the juridical concept, 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 considered data-set includes industrial companies that had undertaken the juridical procedure of bankruptcy in Campania in a given time period, t. The information on the legal status, as well as the financial information for the analysis, were 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 Campania at t=2004, for a total of 93 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 2.1) 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).

Table 2.1: *Failed firms sample.*

	1999	2000	2001	2002	2003
Published Statement	72	72	70	62	39
Total firms	93	93	93	93	93
Percentage	77.42%	77.42%	75.27%	66.67%	41.94%

We chose the year 2004 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 Campania industrial firms according to the following criteria: were still active at time t; have not incurred in any kind of bankruptcy procedures between 2004 and 2009; 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 (2000, 2001, 2002).

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).

Our sampling procedure for selecting the panel data set is based on both balance and unbalanced cluster[2] sampling designs. We also use as benchmark a cross-sectional approach, that is widely applied in the empirical literature.

The predictors data-base for the three years of interest (2000, 2001, 2002) was elaborated starting from the financial statements of each firm included in the sample for a total of 522 balance sheets. We computed $nv = 55$ indicators (Tables .1-.2 in Appendix) selected as potential bankruptcy predictors among the most relevant in highlighting current and prospective conditions of operational unbalance (Altman, 2000; Dimitras *et al.*, 1996). The explanatory variables considered in the analysis have been chosen on

[2]The cluster scheme refers to the geographical distribution of the industrial firms within the region.

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 2.2.

Table 2.2: *Financial predictors.*

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

A pre-processing procedure was performed on the original data set. The results of exploratory data analysis indicates that there are some accounting data observations which are severe outliers. These observations would seriously distort the estimation results, if they were to be included in the default risk model. Therefore, those firms that show values of the financial predictors outside the 3^{th} and 97^{th} percentiles have been excluded from the analysis. In order to achieve stability, we applied a modified logarithmic transformation, defined for non-positive argument (Perederiy, 2009).

The final sample dimensions have been reported in Table 2.3.

Table 2.3: *Sampling Designs.*

	Unbalanced	Balanced	Cross-Sectional
Failed	50	50	150
Healthy	124	50	372

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 Variable selection

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 technique in this domain is the *subset regression*, which aims at choosing the set of the most important regressors removing the noise regressors from 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, e.g. the smallest residual sum of squares, 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 *shrinkage 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 first proposal in linear regression estimation was the *Ridge Regression* (Miller, 2002; Draper and Smith, 1998), which focused on coefficients estimation. Ridge Regression includes all candidate predictors, but with typically smaller coefficients compared to ordinary least squares.

Suppose we have n independent observations $(x_{i1}, x_{i2}, \ldots, x_{ip}; y_i) = (\mathbf{x}; \mathbf{y})$ with $i = 1, \ldots, n$ from a multiple linear regression model:

$$y_i = \mathbf{x_i}'\boldsymbol{\beta} + \epsilon_i, \ \ \forall i$$

with $\mathbf{x_i}$ a p-vectors of covariates and y_i the response variable for the n cases, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ the vector of regression coefficients and the error term, ϵ_i, assumed to be i.i.d. with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2 > 0$.

The ridge coefficients minimizes a penalized residual sum of squares:

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

$$\text{subject to} \sum_{j=1}^{p} \beta_j^2 \leq \delta.$$

This is equivalent to:

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

where $\lambda \geq 0$ is a parameter that controls the amount of shrinkage corresponding to the tuning parameter δ.

A variation of ridge regression that modifies coefficients estimation, so as to reduce some coefficients to zero, effectively performing variable selection, is the *Least Absolute Shrinkage and Selection Operator*, LASSO (Tibshirani, 1996), defined as:

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

$$\text{subject to} \sum_{j=1}^{p} |\beta_j| \leq \delta.$$

This is equivalent to:

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

The Lasso allows for simultaneous execution of both parameter estimation and variable selection. It shrinks some coefficients 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.

A related model-building algorithm is the *Forward Stagewise Regression*, an incremental version of stepwise regression that appears to be very different from the Lasso, but turns out to have similar behavior. This procedure originates from the need to mitigate the negative effects of the greedy behavior of stepwise regression. In stepwise regression, the most useful predictor is added to the model at each step, and the coefficient jumps from zero to the the least-squares value. Forward stagewise picks the same first variable as forward stepwise, but it changes the corresponding coefficient only by a small amount. The algorithms start fitting $r = y - \hat{y}$, with centered prediction and coefficients $\beta_1, \beta_2, \ldots, \beta_p = 0$. At each step, it picks the variable showing the highest correlation to the current residuals and takes a small step for that variable computing the simple linear regression coefficient of the residual of this variable, and then adds it to the current coefficient for that variable. As a consequence, Forward Stagewise can take many steps for reaching the final model, and the resulting coefficients are more stable than those for stepwise.

A more recent proposal by Efron *et al.* (2004), is the *Last Angle Regression*, LAR. The LAR procedure can be easily modified to efficiently compute the LASSO and Forward Stagewise solutions (LARS algorithm) (Friedman *et al.*, 2009), enlarging the gain in application context. Least Angle Regression can be viewed as a version of stagewise that uses mathematical formulas to accelerate computations. Rather than taking many tiny steps with the first variable, the appropriate number of steps is determined algebraically, until the second variable begins to enter the model.

The LAR selection is based on the correlation between each variable and the residuals. It starts with the predictor x_j most correlated with the residual $r = y - \bar{y}$. Put $r = y - x_1$, where γ is determined such that:

$$|\text{cor}(r, x_1)| = \max_{j \neq i} |\ \text{cor}(r, x_j)|$$

Select x_2 corresponding to the maximum above. Continue until all p predictors have been entered.

Briefly, traditional methods have some limits and drawbacks that can be avoided with modern procedures, in terms of stability and prediction. The computational effort in implementing such procedures is overcome by the availability of fast and efficient algorithms.

4 The default-risk models

Our main interest 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 equation can be written as:

$$\text{ln}\left(\frac{p(y)}{1-p(y)}\right) \equiv \text{logit}(p(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \tag{4.1}$$

and

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n} \left\{ y_i \text{ln} p(y_i) + (1-y_i)\text{ln}(1-p(y_i)) \right\}. \tag{4.2}$$

It is modified adding a L_1 norm penalty term in the Regularized Logistic Regression:

$$\hat{\beta}_{lasso} = \underset{\beta}{\text{argmin}} \left[\sum_{i=1}^{n} \left\{ y_i \text{ln} p(y_i) + (1-y_i)\text{ln}(1-p(y_i)) \right\} - \lambda \sum_{i=1}^{p} |\beta_i| \right]. \tag{4.3}$$

In order to generate the maximum likelihood solution, we need to properly choose the tuning parameter λ. Therefore, we use a Cross Validation approach partitioning the training data N into K separate sets of equal size, $N = (N_1, N_2, \dots, N_K)$, for each $k = 1, 2, \dots, K$, fit the model to the training set excluding the k_{th}-fold N_k, and select the value of λ that reached the minimum cross-validation error (CVE).

5 Accuracy Measures

Classification techniques, based on the analysis of financial information, have been used for the predictions and diagnosis of the risk of bankruptcy. 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 5.1). 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*: an healthy firm classified as healthy;
- *False Positive*: an healthy firm classified as failed.

Table 5.1: *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 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.

6 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 [3].

For the unbalanced sample (Table 6.1-6.2), 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 2000 to 2001, but decreases from 2001 to 2002, 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 6.3-6.4), 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. This statement is confirmed by the graphs of the three models' ROC curves, given in Figure 6.1 and 6.2 respectively for the unbalanced and balanced sample designs, and in Figure 6.3 for the cross-sectional data.

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 compu-

[3] The estimate results for the fitted models have been reported in Table from .3 to .14 in Appendix.

tational 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 2001. The indicators selected as predictors for the three estimated models (Table9a:LOGVarUnbalanced1 to .14 in Appendix) are in line with those included, at different levels, in many other empirical studies (Amendola *et al.* , 2010; Dimitras *et al.* , 1996).

Table 6.1: *Unbalanced sample: Accuracy measures for training set.*

	Model1 LR	Model2 Lasso	Model3 LDA
	2000		
Correct Classification Rate	0.83607	0.89344	0.81967
Miss Classification Rate	0.16393	0.10656	0.18033
Type I Error	0.34286	0.37143	0.57143
Type II Error	0.09195	0.00000	0.02299
AUC	0.87685	0.94713	0.80887
AR	0.75369	0.89425	0.61773
	2001		
Correct Classification Rate	0.84426	0.91803	0.87705
Miss Classification Rate	0.15574	0.08197	0.12295
Type I Error	0.40000	0.22857	0.34286
Type II Error	0.05747	0.02299	0.03448
AUC	0.86404	0.96814	0.92118
AR	0.72808	0.93629	0.84237
	2002		
Correct Classification Rate	0.93443	0.94262	0.88525
Miss Classification Rate	0.06557	0.05738	0.11475
Type I Error	0.14286	0.14286	0.28571
Type II Error	0.03448	0.02299	0.04598
AUC	0.96289	0.96880	0.94844
AR	0.92578	0.93760	0.89688

Table 6.2: *Unbalanced sample: Accuracy measures for test set.*

	Model1 LR	Model2 Lasso	Model3 LDA
	2000		
Correct Classification Rate	0.75000	0.86538	0.78846
Miss Classification Rate	0.25000	0.13462	0.21154
Type I Error	0.46667	0.40000	0.73333
Type II Error	0.16216	0.02703	0.00000
AUC	0.70631	0.91171	0.67748
AR	0.41261	0.82342	0.35496
	2001		
Correct Classification Rate	0.86538	0.88462	0.80769
Miss Classification Rate	0.13462	0.11538	0.19231
Type I Error	0.26667	0.26667	0.53333
Type II Error	0.08108	0.05405	0.05405
AUC	0.92793	0.97297	0.83604
AR	0.85586	0.94595	0.67207
	2002		
Correct Classification Rate	0.92308	0.98077	0.90385
Miss Classification Rate	0.07692	0.01923	0.09615
Type I Error	0.06667	0.06667	0.33333
Type II Error	0.08108	0.00000	0.00000
AUC	0.96757	0.99456	0.96757
AR	0.93513	0.98919	0.93514

Table 6.3: *Balanced sample: Accuracy measures for training set.*

	Model1 LR	Model2 Lasso	Model3 LDA
		2000	
Correct Classification Rate	0.84286	0.87143	0.78571
Miss Classification Rate	0.15714	0.12857	0.21429
Type I Error	0.11429	0.14286	0.17143
Type II Error	0.20000	0.11429	0.25714
AUC	0.91510	0.94122	0.88571
AR	0.83020	0.88244	0.77143
		2001	
Correct Classification Rate	0.75714	0.88571	0.87143
Miss Classification Rate	0.24286	0.11429	0.12857
Type I Error	0.22857	0.11429	0.14286
Type II Error	0.25714	0.11429	0.11429
AUC	0.85633	0.94531	0.89531
AR	0.71265	0.89061	0.79062
		2002	
Correct Classification Rate	0.92857	0.97143	0.95714
Miss Classification Rate	0.07143	0.02857	0.04286
Type I Error	0.08571	0.00000	0.05714
Type II Error	0.05714	0.05714	0.02857
AUC	0.97551	0.99265	0.98367
AR	0.95102	0.98531	0.96735

Table 6.4: *Balanced sample: Accuracy measures for Test set.*

	Model1 LR	Model2 Lasso	Model3 LDA
		2000	
Correct Classification Rate	0.76667	0.80000	0.73333
Miss Classification Rate	0.23333	0.20000	0.26667
Type I Error	0.26667	0.26667	0.33333
Type II Error	0.20000	0.13333	0.20000
AUC	0.76889	0.92444	0.74667
AR	0.53778	0.84889	0.49333
		2001	
Correct Classification Rate	0.80000	0.90000	0.83333
Miss Classification Rate	0.20000	0.10000	0.16667
Type I Error	0.13333	0.13333	0.06667
Type II Error	0.26667	0.06667	0.26667
AUC	0.88444	0.96444	0.89778
AR	0.76889	0.92889	0.79556
		2002	
Correct Classification Rate	0.83333	0.93333	0.90000
Miss Classification Rate	0.16667	0.06667	0.10000
Type I Error	0.20000	0.06667	0.13333
Type II Error	0.13333	0.06667	0.06667
AUC	0.89333	0.99556	0.94222
AR	0.78667	0.99111	0.88444

Table 6.5: *Cross-Sectional sample: Accuracy measures for training set.*

	Model1 LR	Model2 Lasso	Model3 LDA
Correct Classification Rate	0.87671	0.94795	0.88767
Miss Classification Rate	0.12329	0.05205	0.11233
Type I Error	0.27619	0.15238	0.32381
Type II Error	0.06154	0.01154	0.02692
AUC	0.92919	0.97927	0.91641
AR	0.85839	0.95853	0.83282

Table 6.6: *Cross-Sectional sample: Accuracy measures for test set.*

	Model1 LR	Model2 Lasso	Model3 LDA
Correct Classification Rate	0.82803	0.96815	0.85987
Miss Classification Rate	0.17197	0.03185	0.14013
Type I Error	0.31111	0.06667	0.37778
Type II Error	0.11607	0.01786	0.04464
AUC	0.83591	0.98651	0.87937
AR	0.67182	0.97301	0.75873

Figure 6.1: *Accuracy measure for unbalanced sample.*

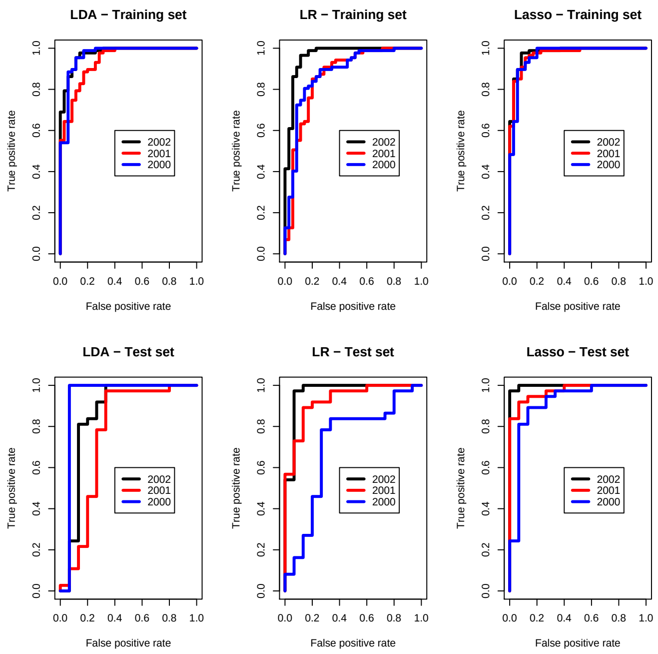


Figure 6.2: *Accuracy measure for Balanced sample.*

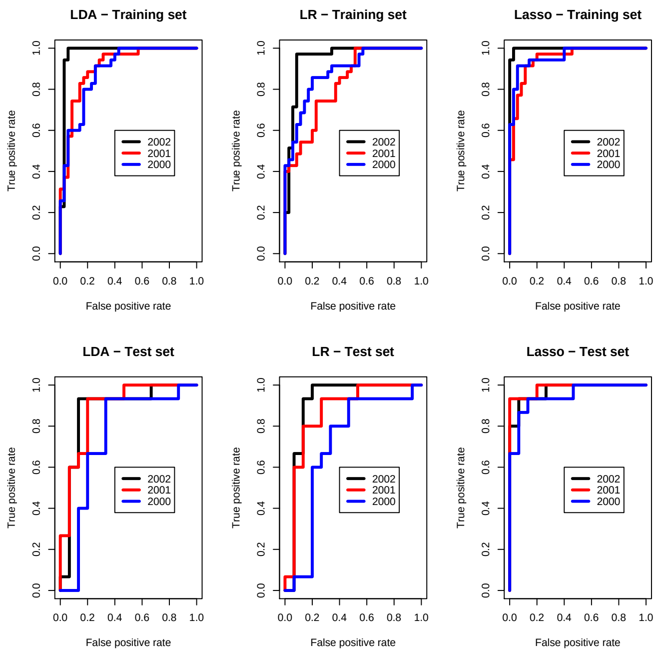
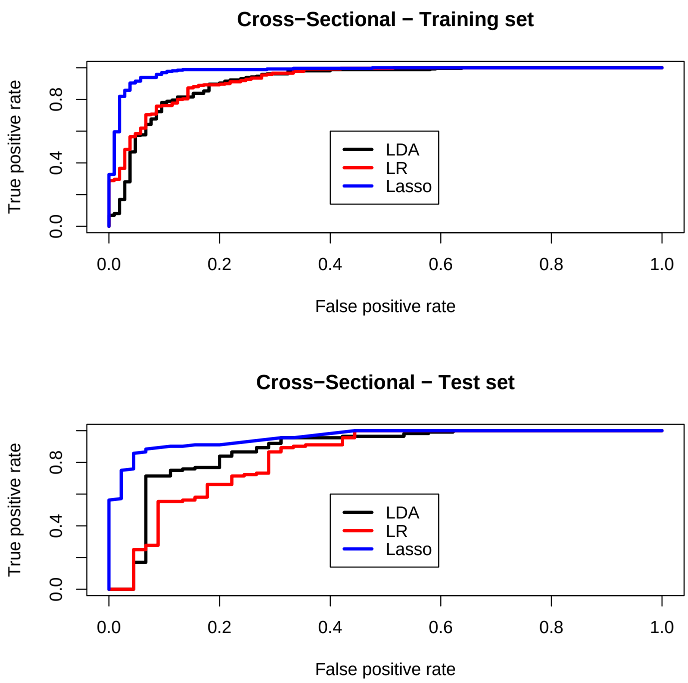


Figure 6.3: *Accuracy measure for cross-sectional sample.*



7 *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 Campania for a given time period have been analyzed. 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 2004. Thus, we have at least four future reports to evaluate the real status of the selected firms. 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. This leads to further investigation by taking into account the time dimension and the evolutionary behavior of the financial variables.

Acknowledgements
We would like to thank the Chamber of Commerce of Salerno for providing part of the data.

References

Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, New York: John Wiley & Sons.

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*.

Amendola A., Bisogno M., Restaino M., Sensini L. (2010), Forecasting corporate bankruptcy: an empirical analysis on industrial firms in Campania, *Rirea*, 2, 229-241.

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.

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.

Draper N.R., Smith H. (1998), *Applied Regression Analysis*, John Wiley & Sons, INC.

Efron B., Hastie T., Johnstone I., Tibshirani R. (2004), Least Angle Regression, *Annals of Statistics*, 32, 2, 407-499.

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.

Miller A. (2002), *Subset Selection in Regression*, Chapman & Hall, CRC.

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.

Pa rk 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.

Perederiy V. (2009), Bankruptcy Prediction Revisited: Non-Traditional Ratios and Lasso Selection, European University Viadrina Frankfurt, *Working Paper*.

Perez M. (2006), Artificial neural networks and bankruptcy forecasting: a state of the art, *Neural Computing and Application*, 15, 2, 154-163.

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),Feedforward neural networks in the classification of financial information, *European Journal of Finance*, 3, 3, 183-202.

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.

Appendix

Table .1: *Financial indicators and financial area.*

	Financial Indicators	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

Table .2: *Financial indicators and financial area.*

	Financial Indicators	Area
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

Table .3: *Model 1 - Unbalanced sample: variables of interest and their coefficients and standard error.*

Variables	Coefficients (s.e.) 2002	Coefficients (s.e.) 2001	Coefficients (s.e.) 2000
Intercept	2.69377 (0.80707)	1.2765 (0.2927)	1.8375 (0.4026)
Leverage	0.22790 (0.30344)	0.6230 (0.2844)	-0.3097 (0.3141)
Current ratio I	- 0.61808 (0.60516)	-0.5704 (0.4958)	0.3295 (0.5020)
Current ratio II	1.32328 (0.55794)	1.0101 (0.4594)	1.5581 (0.6291)
Current Assets/Fixed Assets	- 0.77579 (0.64118)	0.2335 (0.5722)	1.5445 (0.6849)
Net Worth/Total debts	4.65996 (1.50682)	1.2531 (0.4086)	1.2012 (0.3806)
Account receivable/Sales	- 1.71743 (0.68778)	-1.2675 (0.4163)	-1.3879 (0.4962)
Net income/Sales	2.30509 (0.92426)	-0.5243 (0.6136)	-0.7715 (0.4161)
Sales/Fixed Assets	0.41780 (0.71410)	-0.7800 (0.5327)	-1.8642 (0.6930)
Sales/Inventory	0.09978 (0.46274)	0.6600 (0.3130)	0.8564 (0.3509)

Table .4: *Model 1 - Unbalanced sample: Estimates of Odd Ratios and their confidence intervals.*

Variables	Year 2002	Year 2001	Year 2000
Intercept	14.7872756 [3.04013561; 71.9255809]	3.5842432 [2.0196370; 6.360945]	6.2805075 [2.85298060; 13.8258122]
Leverage	1.2559560 [0.69291450; 2.2765080]	1.8645066 [1.0677367; 3.255845]	0.7336682 [0.39638368; 1.3579494]
Current ratio I	0.5389772 [0.16460688; 1.7647887]	0.5653012 [0.2139186; 1.493865]	1.3902554 [0.51970751; 3.7190344]
Current ratio II	3.7557220 [1.25825166; 11.2103549]	2.7459900 [1.1159594; 6.756931]	4.7496972 [1.38414509; 16.2985973]
Current Assets/Fixed Assets	0.4603399 [0.13100834; 1.6175523]	1.2630749 [0.4115192; 3.876752]	4.6858582 [1.22408176; 17.9377455]
Net Worth/Total debts	105.6319715 [5.51016119; 2025.0067127]	3.5011258 [1.5718802; 7.798229]	3.3241102 [1.57641422; 7.0093941]
Account receivable/Sales	0.1795266 [0.04663143; 0.6911602]	0.2815443 [0.1244960; 0.636705]	0.2496041 [0.09438282; 0.6601010]
Net income/Sales	10.0250993 [1.63811778; 61.3524970]	0.5919513 [0.1778248; 1.970514]	0.4623409 [0.20451987; 1.0451751]
Sales/Fixed Assets	1.5186094 [0.37461981; 6.1560397]	0.4583891 [0.1613504; 1.302263]	0.1550217 [0.03985391; 0.6029958]
Sales/Inventory	1.1049319 [0.44611192; 2.7367000]	1.9347440 [1.0476428; 3.573006]	2.3545706 [1.18353215; 4.6842858]

Table .5: *Model 2 - Unbalanced sample: Variables of interest and their coefficients.*

Variables	Coefficients 2002	Coefficients 2001	Coefficients 2000
Intercept	1.018476966	5.06205149	-0.4474469
Net Proceeds/Invested Capital			-0.0006996651
Return on Equity	2.764453606		
Return on Investment			-1.7796030
Return on Assets	6.096886873	37.77743437	5.4252000
Return on Sales	3.495008909		-5.244069
Net Proceeds/Current Assets	-2.032063547		0.2592640
Leverage		0.38481125	
Liquid Assets/Total Assets			2.406223
Current ratio I	-0.001365445		
Current ratio II			0.6434563
Quick Ratio	1.865601830	0.83431157	
Equity ratio	6.500826209		
Inventory/sales	-1.064546745		
Total debts/Total assets		-15.20842828	-0.7084720
Net Worth/Fixed Assets	0.106065746		
Capital Stock/Fixed Assets	-1.056310720	-1.53768744	
Current Assets/Fixed Assets	-0.441267698	-0.03174854	-2.121996
Inventory/Current Assets		2.32730434	
Liquid Assets/Total Assets			-1.311061
Net Worth/Total Assets			1.0399990
Net Worth/Total debts	4.997617294		
Financial bebt/Total Assets		-0.10235802	-0.7748252
Cash Flow/Net Worth		0.49374825	-1.474363
Cash/Sales	-4.436903486		
Account receivable/Sales	-4.357356706	-3.16875379	-3.050993
Total debts/Sales			1.949349
Net income/Sales		-8.73093211	
Sales/Fixed Assets		-1.21822300	
Sales/Advances from customers		-0.91995013	-0.5030433
Sales/Inventory	1.085483263	1.09306620	0.8840715
Sales/Total Assets			-0.001119065
Labour cost/Production cost			-1.392899
Labour cost/Net sales		-0.95442366	
Financial charges/Financial Debt			-3.854328
Financial charges/Production value	-24.705635387		
Financial charges/Net sales		-16.12453127	

Table .6: *Model 3 - Unbalanced sample: variables of interest and their coefficients.*

Variables	Coefficients 2002	Coefficients 2001	Coefficients 2000
Return on Equity	0.6156659	0.6189447	0.44093883
Return on Investment	0.2012200	0.1159395	- 0.26687136
Equity ratio	22.3216383	1.1837203	-107.42964440
Total debts/Total assets	0.0787034	-0.6757346	0.62661690
Net Worth/Fixed Assets	0.3321355	-0.4751470	- 0.11571557
Gross Working Capital/Total assets	2.5328962	1.4839388	- 0.26188687
Capital assets/Total Assets	2.9720155	1.7347579	- 0.04421868
Net Worth/Total Assets	-22.1273447	-0.8441356	109.34202405
Net Worth/Total debts	0.3398322	-0.3274019	- 0.55678145
Cash Flow/Total debts	- 0.3200035	0.4067501	0.49227050
Net income/Sales	0.3748707	-0.3131388	- 0.19717462

Table .7: *Model 1 - Balanced sample: variables of interest and their coefficients and their standard error.*

Variables	Coefficients (s.e.) 2002	Coefficients (s.e.) 2001	Coefficients (s.e.) 2000
Intercept	1.4108 (0.6709)	0.4272 (0.3556)	1.3150 (0.5257)
Return on Assets	7.1910 (2.6357)	8.4110 (3.0367)	2.4581 (1.0013)
Net Proceeds / Current Assets	0.1400 (0.5876)	0.1260 (0.3276)	-2.3374 (0.8970)
Liquid Assets/Total Assets	-0.3608 (0.5849)	-0.3125 (0.3217)	-0.4737 (0.5022)
Quick Ratio	2.6866 (1.3955)	0.1494 (0.3403)	-0.1667 (0.4071)
Net Worth/Total debts	3.2004 (1.3346)	1.1165 (0.5367)	1.5500 (0.6407)
Net Income/Total Assets	-4.3733 (2.0591)	2.1386 (1.0855)	-0.4849 (0.3347)
Sales/Inventory	1.0005 (0.6276)	0.4449 (0.3663)	3.1370 (1.1391)

Table .8: *Model 1 - Balanced sample: Estimates of Odd Ratios and their confidence intervals.*

Variables	Year 2002	Year 2001	Year 2000
Intercept	4.099295 [1.1005328172; 1.5269170]	1.5329534 [0.7635986; 3.077463]	3.7245810 [1.32916251; 10.4370261]
Return on Assets	1.3274280 [7.5758619; 2.32589500]	2.0276030 [8.785801; 2.32503500]	11.6821588 [1.64141479; 83.1434169]
Net Proceeds / Current Assets	1.1502790 [0.3636039386; 3.638961]	1.1342500 [0.5968196; 2.155631]	0.0965814 [0.01664697; 0.5603401]
Liquid Assets/Total Assets	0.6971472 [0.2215381689; 2.1938170]	0.7316501 [0.3894523; 1.374525]	0.6226850 [0.23269360; 1.6662966]
Quick Ratio	1.4682240 [0.9526533725; 2.26281700]	1.1611375 [0.5959526; 2.262328]	0.8464485 [0.38115798; 1.8797324]
Net Worth/Total debts	2.4541360 [1.7940511266; 3.35708700]	3.0541130 [1.0666339; 8.744900]	4.7115932 [1.34224485; 16.5387936]
Net income/Total Assets	0.01260902 [0.0002228001; 0.7135871]	8.4872864 [1.0110314; 71.248067]	0.6157749 [0.31952103; 1.1867098]
Sales/Inventory	2.719668 [0.7948201643; 9.305995]	1.5603835 [0.7610072; 3.199440]	23.0348400 [2.47030519; 214.7928334]

Table .9: *Model 2 - Balanced sample: variables of interest and their coefficients.*

Variables	Coefficients 2002	Coefficients 2001	Coefficients 2000
Intercept	0.2167890	4.0727306	-0.7975193
Return on Equity	2.4368434	2.0266548	
Return on Assets	5.8881014	36.0929763	20.0728773
Net Proceeds / Current Assets	-2.7022697		0.1510361
Liquid Assets/Total Assets		-0.9866630	4.4923211
Current ratio II	0.2514802		0.3795167
Quick Ratio	1.8791706	0.4900981	
Equity ratio	4.8271847		
Inventory/sales	-0.7717858	0.8997890	
Total debts/Total assets		-14.8398978	
Capital Stock/Fixed Assets	-0.8549952	-2.1723391	-2.5259083
Current Assets/Fixed Assets	-0.2786438		
Inventory/Current Assets		0.8632208	
Gross Working Capital/ Total assets			-2.5151231
Net Worth/Total Assets			9.3992578
Net Worth/Total debts	8.5850875		
Financial bebt/Total Assets			-0.3034599
Cash Flow/Net Worth		0.0849717	-0.2634682
Cash/Sales	-16.4574529		
Account receivable/Sales	-3.4847106	-2.2245920	-0.3450945
Total debts/Sales	-0.2775671		
Sales/Fixed Assets		-0.4454478	-0.3038662
Sales/Advances from customers		-1.0147509	
Sales/Inventory	1.1213101	0.8339302	
Labour cost/Production cost			-0.5767321
Financial charges/Financial Debt			-24.9593016
Financial charges/Production value	-25.3661137		
Financial charges/Net sales		-10.9637343	

Table .10: *Model 3 - Balanced sample:variables of interest and their coefficients.*

Variables	Coefficients 2002	Coefficients 2001	Coefficients 2000
Return on Equity	0.60406524	0.26667686	0.05179017
Return on Assets	2.88390145	4.75747968	0.63560734
Equity ratio	-0.09862876	-0.08822949	0.25521060
Net Worth/ Sales	-0.56958870	0.16404338	0.06285994
Total debts/Total assets	-0.51301002	-0.03277058	0.38831096
Net Worth/Total debts	1.11815705	0.47569371	1.22517715
Cash Flow/Total debts	0.50697332	0.78507097	-0.52258623
Net income/Sales	-0.05142747	-0.07686993	-0.20221589
Net income/Total Assets	-3.02337229	-4.58993882	-0.16198233
Sales/Inventory	0.23824798	0.07727323	0.50967583

Table .11: *Model 1 - Cross-Sectional sample: variables of interest and their coefficients and their standard error.*

Variables	Coefficients (s.e.)
Intercept	1.4014 (0.2506)
Return on Investment	2.4960 (0.8100)
Return on Sales	-0.2832 (0.5795)
Leverage	0.2906 (0.3158)
Quick Ratio	0.5798 (0.2575)
Equity ratio	1.9050 (0.9515)
Net Worth/ Fixed Assets	0.7838 (0.5978)
Capital Stock/Fixed Assets	-0.6343 (0.3228)
Current Assets/Fixed Assets	-0.6548 (0.3783)
Net Worth/Total debts	-0.7377 (0.7302)
Financial bebt/Total Assets	-0.1535 (0.3850)
Cash Flow/Total Assets	0.4890 (0.7729)
Cash Flow/Total debts	0.4780 (1.0677)
Account receivable/Sales	-1.0259 (0.2500)
Net income/Total debts	-0.7048 (0.7903)
Sales/Inventory	1.2043 (0.3145)
Sales/Total Assets	-1.5384 (0.3408)
Labour cost/Production cost	-1.6460 (0.5726)
Labour cost/Production value	1.5548 (0.5834)
Financial charges/Production value	-0.6404 (0.2402)

Table .12: *Model 1 - Cross-Sectional sample: estimates of odd ratios and their confidence intervals.*

Variables	Odd Ratio	C.I.
Intercept	4.0609734	[2.48475110; 6.6370853]
Return on Investment	12.1339964	[2.48016359; 59.3645799]
Return on Sales	0.7533747	[0.24196924; 2.3456430]
Leverage	1.3372880	[0.72009344; 2.4834823]
Quick Ratio	1.7857559	[1.07803638; 2.9580857]
Equity ratio	6.7192968	[1.04092165; 43.3740131]
Net Worth/ Fixed Assets	2.1897438	[0.67848524; 7.0671811]
Capital Stock/Fixed Assets	0.5303187	[0.28167023; 0.9984653]
Current Assets/Fixed Assets	0.5195455	[0.24749500; 1.0906382]
Net Worth/Total debts	0.4781916	[0.11429462; 2.0006819]
Financial bebt/Total Assets	0.8576720	[0.40328005; 1.8240457]
Cash Flow/Total Assets	1.6307552	[0.35848089; 7.4184221]
Cash Flow/Total debts	1.6127859	[0.19893526; 13.0749996]
Account receivable/Sales	0.3584884	[0.21961614; 0.5851751]
Net income/Total debts	0.4942202	[0.10500052; 2.3262139]
Sales/Inventory	3.3343749	[1.80028788; 6.1757101]
Sales/Total Assets	0.2147288	[0.11009735; 0.4187971]
Labour cost/Production cost	0.1928292	[0.06276809; 0.5923885]
Labour cost/Production value	4.7339855	[1.50877970; 14.8534733]
Financial charges/Production value	0.5270582	[0.32913433; 0.8440029]

Table .13: *Model 2 - Cross-Sectional sample: variables of interest and their coefficients.*

Variables	Coefficients
Intercept	-5.2860012
Return on Equity	2.3272130
Return on Investment	-35.0399664
Return on Assets	141.9112220
Return on Sales	-1.4442691
Net Proceeds/Current Assets	-3.4739327
Leverage	2.3971550
Liquid Assets/Total Assets	13.8012351
Current ratio I	-3.5004547
Current ratio II	3.9410912
Equity ratio	22.2468427
Net Worth/Sales	-6.5662836
Capital Stock/Sales	7.1267543
Inventory/sales	-7.2988978
Total debts/Total assets	9.4205597
Net Worth/ Fixed Assets	6.3156994
Capital Stock/Fixed Assets	-6.3699754
Current Assets/Fixed Assets	-5.2779610
Inventory/Current Assets	22.6516571
Gross Working Capital/Total assets	4.5387650
Capital assets/Total Assets	4.9973221
Capital Stock/Total Assets	3.9189603
Net Worth/Total debts	-1.9496457
Capital Stock/Total Debts	-4.3953657
Financial bebt/Total Assets	-12.9650440
Cash Flow/Sales	-1.8718048
Cash Flow/Total Assets	-98.6081327
Cash Flow/Net Worth	-1.9577303
Cash Flow/Total debts	92.5191261
Cash/Sales	7.6164762
Account receivable/Sales	-10.2639420
Total debts/Sales	8.1128739
Net income/Sales	5.6262195
Net income/Total debts	-54.7531428
Sales/Fixed Assets	1.6026422
Sales/Advances from customers	-1.0221275
Sales/Inventory	2.1476772
Sales/Total Assets	-8.8094377
Labour cost/Production cost	-55.3656139
Labour cost/Production value	93.8415969
Labour cost/net sales	-42.9034142
Financial charges/Total Debts	39.9062459
Financial charges/Financial Debt	-0.6172331
Financial charges/Production value	-97.8336349
Financial charges/Net sales	11.5536873

Table .14: *Model 3 - Cross-Sectional sample: Variables of interest and their coefficients.*

Variables	Coefficients
Return on Equity	0.4674687
Return on Investment	-0.2052834
Return on Assets	1.7422374
Equity ratio	-0.5540270
Total debts/Total assets	0.6431632
Current Assets/Fixed Assets	-0.6996048
Capital assets/Total Assets	-0.2774203
Net Worth/Total Assets	1.3806180
Net Worth/Total debts	0.4656836
Cash Flow/Total debts	-0.1961435
Net income/Total Assets	-1.4853683
Net income/Total debts	0.4341946
Sales/Inventory	0.3394478

Dipartimento di Scienze Economiche e Statistiche
Università degli Studi di Salerno Depositato ai sensi di legge