Available online at www.sciencedirect.com

ScienceDirect

Procedia Economics and Finance 39 (2016) 634 – 641

Procedia Economics and Finance

www.elsevier.com/locate/procedia

3rd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, 26-28 November 2015, Rome, Italy

The logistic lasso and ridge regression in predicting corporate failure

Jose Manuel Pereira[a]*, Mario Basto[a], Amelia Ferreira da Silva[b]

[a]*IPCA - Polytechnic Institute of Cavado and Ave, Campus do IPCA, 4750-810 Barcelos, Portugal*
[b]*IPP - Polytechnic Institute of Oporto, Rua Dr. Roberto Frias, 4200-465 Oporto, Portugal*

Abstract

The prediction of corporate bankruptcy is a phenomenon of interest to investors, creditors, borrowing firms, and governments alike. Many quantitative methods and distinct variable selection techniques have been employed to develop empirical models for predicting corporate bankruptcy. For the present study the lasso and ridge approaches were undertaken, since they deal well with multicolinearity and display the ideal properties to minimize the numerical instability that may occur due to overfitting. The models were employed to a dataset of 2032 non-bankrupt firms and 401 bankrupt firms belonging to the hospitality industry, over the period 2010-2012. The results showed that the lasso and ridge models tend to favor the category of the dependent variable that appears with heavier weight in the training set, when compared to the stepwise methods implemented in SPSS.

Peer-review under responsibility of the Organizing Committee of BEMTUR- 2015
Keywords: Corporate Bankruptcy; Prediction Models; Lasso; Ridge Regression

1. Introduction

There are several undesirable consequences of business failures. Its economic and social cost can be significant. So, it is quite natural that this issue has occupied a significant part of researcher's agenda. In spite of recent growing interest on non-financial attributes in explaining business failures, traditionally investigation on this issue has been focused on financial attributes. In most of the works statistical or artificial intelligence techniques were applied to the accountancy data of the companies, aiming at obtaining prediction models that would indicate whether the

* Jose Manuel Pereira. Tel.: +351-963-621-955; fax: +351-253-821-111.
E-mail address: jpereira@ipca.pt

Peer-review under responsibility of the Organizing Committee of BEMTUR- 2015
doi:10.1016/S2212-5671(16)30310-0

company would or would not reach a bankruptcy situation in the future (Beaver, 1966; Altman, 1968; Martin, 1977; Tam and Kiang, 1992).

In a study on corporate bankruptcy prediction, one of the aspects we immediately need to clarify is the concept of bankruptcy we shall use. In specialized literature the term has been used in different ways by different authors: legal bankruptcy, insolvency, inability to do payments or continued losses. As we lack a general theory on corporate bankruptcy, there is also no unique definition for this concept. This is an important limitation, since the sample's selection, both in terms of firms that have and have not "bankrupt", depends on the definition of corporate bankruptcy used.

Throughout the last four decades several techniques were used to design models regarding this issue. The Altman model is the classical model on prediction of corporate bankruptcy most referred to in literature. It was developed in the end of the 70's using the discriminant analysis. The purpose of this technique was to obtain an indicator or «Z» score (variable dependent on a function) that was the result of the linear combination of several independent variables (ratios or financial indicators).

After the publication of Altman's Z-Score in 1968, most studies published in the decade that followed also used the discriminant analysis (Blum, 1974; Elam, 1975; Altman et al., 1977; Moyer, 1977; Norton and Smith, 1979). The emergence of critiques emphasizing the limitations of this theory may have influenced researchers to try new techniques, namely logit and probit. With the application of logit models to corporate bankruptcy prediction, it is possible to estimate the probability that a certain event will happen, as well as the probability of failure or corporate bankruptcy, considering the values of certain indicators of the company (Ohlson, 1980; Keasey and Watson, 1987).

The probit model is associated to the cumulative function of normal probability. Although this model is not as popular as the logit one in this area of research, there are several studies that have used this methodology (Zmijewski, 1984; Lennox, 1999) with similar outcomes as with other techniques. The evolution of ICT has created adequate conditions for the development and application of other techniques that, despite their limitations, do not demand certain conditions, unlike statistical techniques. Among artificial intelligence techniques, the most used ones have been neural networks and the induction of rules and decision trees. During the last few years, studies that use the theory of Rough Sets have emerged. Several neural networks models have been used throughout the last decade in studies related to corporate solvency prediction. Among them we can highlight Bell et al. (1990) and Koh and Tan (1999), using the multilayer perceptron model; Coats and Fant (1993) and Lacher et al. (1995), using the Cascor method (cascade correlation) as a learning algorithm; and Serrano and Martin (1993), using a multilayer perceptron net and Kohonen's self-organizing maps. Min and Lee (2005) applied support vector machines to the bankruptcy prediction problem. Some of the recent research on predicting corporate failure focuses on ensemble models (Kim and Kang, 2010) and hybrid models (Ahn and Kim, 2009).

The least absolute shrinkage and selection operator (Lasso) is a variable selection technique that has been recently applied on corporate bankruptcy forecasts (Tian et al., 2015). For the present study the lasso and ridge approaches were undertaken, since they deal well with multicolinearity and display the ideal properties to minimize the numerical instability that may occur due to overfitting. The models were employed to a dataset of 2032 non-bankrupt firms and 401 bankrupt firms belonging to the hospitality industry, over the period 2010-2012.

2. The Ridge and Lasso logistic regression

The task of determining which predictors are associated with a given response is not a simple task. When selecting the variables for a linear model, one generally looks at individual p-values. This procedure can be misleading. For instance, if the variables are high correlated, the p-values will also be high, driving the researcher to mistakenly deduce that those variables are not important predictors. On the other hand, irrelevant variables may be included in the model that are not associated with the response, adding an unnecessary complexity and interpretability to the model. Also, if the number of observations is not much larger than the number of variables then there can be a lot of variability, resulting in overfitting (increased likelihood by adding more parameters but poorer predictions on future observations not used in the model training). There are some approaches for automatically performing variable selection.

Linear regression by least squares is not applied for a binary response coded as zero and one, but although some of the estimates of the binary response might be outside the interval (0,1), the output obtained can be seen as a crude

estimate of the response's probability taking the value one. Hence linear regression could be applied, but others much better classification methods for binary responses are available, like logistic regression, *k*-nearest neighbors or discriminant analysis (James et al., 2013). However, linear regression can be used as a first step to select the best variables, so that afterwards they can be used to build another more appropriate model.

To select the best set of explanatory variables when the response variable is binary some procedures that can be undertaken (James et al., 2013). This paper aims to build a logistic model to predict enterprise failure, by resorting on two kinds of approaches: stepwise or best subset selection methods, and the ridge regression or the lasso, procedures less known, since they are not usually available in most commercial software. A comparison is made between those procedures.

Basically, the stepwise or best subset selection, aims to identify a subset of the predictors that one believes to be the best one related to the response. This can be done automatically by stepwise selection procedures. Ideally, one would like to try out a lot of different models or even all the possible combinations of predictors, each containing a different subset of the predictors, and then select the one with the lower test error. This approach may suffer from computational limitations. Stepwise methods, which explore only a restricted set of models, a sequence of nested models, manifests as good alternatives to the best subset selection. For *p* predictors, forward stepwise method starts from the null model, and in each step *k*, chooses the best model with one additional predictor. At each step, the predictor satisfying the entry criterion is added to the model. Backward stepwise selection is similar but begins with the full model containing all the *p* predictors. At each step, the predictor that contributes the least is removed from the model, until all of the predictors in the model are significant. None of these two approaches guarantees to find the best possible model. Backward selection, like the best subset selection, requires that the number of samples *n* is larger than the number of variables *p*. Hybrid versions of forward and backward stepwise selection procedures present as fair alternatives. All these methods create a set of models with different variables. The following variable selection methods are available in SPSS: forward conditional, forward Likelihood Ratio, forward Wald, backward conditional, backward Likelihood Ratio, and backward Wald (IBM, 2013).

Another approach less used, is by shrinkage or regularization, which involves fitting a model with all the predictors, but where the estimated coefficients are shrunken towards zero relatively to the classical estimates (James *et al.*, 2013). As consequence, the variance model reduces and by estimating some of the coefficients to be zero, variable selection is also performed. To accomplish this task, some approaches are available such as ridge regression (Hoerl and Kennard, 1970; Cessie and Houwelingen, 1992; Tibshirani, 1996) and the lasso (Park and Casella, 2008; Tibshirani, 1996).

In general, when the relationship between the logit in a dichotomous response and the predictors is close to linear, the maximum likelihood estimates will have low bias but may have high variance as when the number of covariates is large compared to the number of observations or the predictors are highly correlated. This way a small change in the training data may cause a large change in the coefficient estimates. Ridge regression and lasso perform by trading off a small increase in bias for a large decrease in variance of the predictions, hence they may improve the overall prediction accuracy.

Ridge logistic regression (Hoerl and Kennard, 1970; Cessie and Houwelingen, 1992; Schaefer et al., 1984), is obtained by maximizing the likelihood function with a penalized parameter applied to all the coefficients except the intercept. The ordinary logistic regression with binary response is given by the probability of the response success:

$$P(y_i = 1) = \pi_i = \frac{e^{x_i\beta}}{1+e^{x_i\beta}}$$

where x_i is the *i*-th row of an matrix of *n* observations with *p* predictors and a column of ones to accommodate the intercept, and β is the column vector of the regression coefficients. The parameters estimates are obtained by maximizing the log-likelihood function:

$$l(\beta) = \sum_{i=1}^{n}\left[y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)\right] = \sum_{i=1}^{n}\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \log(1-\pi_i)\right] = \sum_{i=1}^{n}\left[y_i x_i \beta - \log\left(1+e^{x_i\beta}\right)\right]$$

The logistic ridge regression estimator depends on the choice of a tuning parameter $\lambda \geq 0$, to be determined separately. The coefficients estimates are the values that maximize the following slightly different log-likelihood function where a L_2 ridge penalty is added to the function (Duffy and Santner, 1989; Cessie and Houwelingen, 1992), resulting in the following constrained maximization equation:

$$l_{\lambda}^{R}(\beta) = \sum_{i=1}^{n}\left[y_i x_i \beta - \log\left(1 + e^{x_i\beta}\right)\right] - \lambda \sum_{j=1}^{p} \beta_j^2$$

This way, ridge regression shows substantial computational advantages over the best subset selection, which has a very heavy computational effort which turn out to be almost prohibitive to apply in many cases, since it requires searching through 2^p models. For any fixed value of λ, ridge regression only fits a single model, which can be performed quite quickly (James et al., 2013).

As shrinkage penalty λ increases, the ridge coefficient estimates will tend to approach zero. However, the penalty introduced in the log-likelihood function will shrink all of the coefficients towards zero, but it will not set any of them exactly to zero. Hence, ridge regression has the disadvantage over model selection, of including all the predictors in the final model (James et al., 2013). This way the model interpretation when the number of variables p is large turns out to be more problematic.

The lasso is another alternative of regularization that overcomes the disadvantage of ridge regression inability of reducing the number of predictors in the final model. The penalized version of the log-likelihood function to be maximized takes now the form (Hastie, 2009):

$$l_{\lambda}^{L}(\beta) = \sum_{i=1}^{n}\left[y_i x_i \beta - \log\left(1 + e^{x_i\beta}\right)\right] - \lambda \sum_{j=1}^{p} \left|\beta_j\right|$$

Comparing to ridge regression the lasso uses a L_1 penalty instead of a L_2 used in ridge regression. The L_1 penalty used in the lasso is used for both variable selection and shrinkage, since it has the effect, when the λ is sufficiently large, of forcing some of the coefficient estimates to be exactly equal to zero (James et al., 2013).

Lasso has an advantage over ridge regression, since the final model may involve only a subset of the predictors, which in turn improves model interpretability. Concerning prediction accuracy, usually when only a small number of predictors have substantial coefficients, one can expect lasso to perform better, while when all coefficients are roughly of equal size, one expects a better performance of ridge regression. Cross-validation can be used in order to determine which approach is better on a particular data set (James et al., 2013).

For both lasso and ridge regression, generally one do not penalize the intercept term, and standardize the predictors for the penalty to be meaningful (Hastie *et al.*, 2009).

Another regularization and variable selection method proposed by Zou and Hastie (2005), called elastic net, includes a tuning parameter $\alpha \geq 0$, being the penalty a mixture of the previous two approaches:

$$\alpha \sum_{j=1}^{p} \beta_j^2 + (1-\alpha) \sum_{j=1}^{p} \left|\beta_j\right|$$

This approach is particularly useful when the number of predictors is much larger than the number of observations (Zou and Hastie, 2005).

Several others regularization methods are described in literature (Yuan and Lin 2007; Meier et al. 2008; Park and Hastie 2007; Fan and Li 2005; Friedman et al. 2008; Hastie et al. 2004).

Both, ridge regression and lasso produce a set of coefficient estimates whose values depend on the different values of λ. Choosing a good value of λ is a critical step for both methods. To accomplish this task, different approaches are described in literature. Cule and De Iorio (2012) propose a semi-automatic method to choose the ridge parameter for data with very high dimensions and many more covariates than observations. Another useful way of determining the best value for the tuning parameter λ is by making recourse on a technique known as cross-

validation, which is a method of assessing how well a model can be generalized to an independent data set. One method is by using the so called *k*-fold cross-validation (Hastie et al., 2009), whereby the data are partitioned into *k* subsets of approximately equal size and one of the subsets becomes the validation set. The remaining *k*-1 subsets are used as training data. This procedure is repeated *k* times, each time with a different validation set, and the optimum value of λ is estimated such that the cross-validated log-likelihood is maximized (Goeman, 2010).

3. Methodology

For the present study the lasso and ridge approaches were undertaken, since they deal well with multicolinearity and display the ideal properties to minimize the numerical instability that may occur due to overfitting. Therefore, improve prediction accuracy can be achieved. Lasso will shrink parameter estimates towards zero and, in some cases, equate parameters to be exactly zero and thus allows the exclusion of some of the variables from the model. These solutions were compared in terms of accuracy of predictions to the stepwise methods available in SPSS.

To perform the lasso and ridge regression, one resort on software R and package glmnet (Friedman *et al.*, 2015). The package allows to fit generalized linear models with different penalties from the L_1 regularization from lasso to the L_2 regularization from ridge regression, or the elastic net regularization penalty (Zou and Hastie, 2005), for generalized linear models via cyclical coordinate descent algorithm (Friedman *et al.*, 2010). To estimate the parameter λ a *k*-fold cross-validation procedure implemented in package glmnet (Friedman *et al.*, 2015), was undertaken.

The data was initially explored in order to detect situations requiring correction prior to the accomplishment of the statistical procedures, such outliers, missing data, or variables with reduced variability. To detect multivariate outliers one resorted on a procedure implemented in SPSS. This procedure creates a clustering model and anomaly indices for each case to measure how unusual the case is with respect to its cluster. At the end no cases were eliminated.

Concerning missing values, all the cases with more than 30% of missing values were eliminated (3.86 % of the cases). For the remaining cases none of the variables had more than 15% of missing data. Also, no signs of data missing not at random was detected. If data is missing at random or missing completely at random, missing data can be estimated. Therefore, the decision to deal with missing values was between elimination or estimation. The elimination of cases would significantly reduce the sample size of failed enterprises, a reduction of about 45.9%, so the process of the *k* neighbors was applied to fill in the missing data, a process of identification of the *k* more similar cases.

To construct, validate and compare all the methods, the data was divided in two sets, a training set and a test set. The quality has been evaluated by applying the estimated models to the test set. Four divisions of the data were undertaken. One with the training set with failed and good enterprises of equal sizes, another with the training set with failed enterprises half smaller than the healthy ones, the third with about 40% of the overall data used as training set and the last one with the training set of failed enterprises one third larger than the number of good ones.

4. Results

The type I error (percentage of failed enterprises predicted good by the model), the type II error (percentage of good enterprises predicted failed by the model) and the overall error are displayed in the following four tables. For the model with the training set with failed and good enterprises of equal sizes (table 1), ridge regression showed the least type II and overall errors, but the differences were not substantial between SPSS stepwise methods and lasso and ridge regressions.

Table 1. Model with the training set with failed and good enterprises of equal sizes.

	Forward Condition	Forward LR	Forward Wald	Backward Condition	Backward LR	Backward Wald	Lasso	Ridge
Type I error	37,65%	37,65%	37,65%	46,47%	46,47%	37,65%	38,82%	40,59%

Type II error	29,21%	29,21%	29,21%	27,55%	27,55%	29,21%	28,73%	25,79%
Overall error	29,92%	29,92%	29,92%	29,13%	29,13%	29,92%	29,57%	27,02%

For the model with the training set with failed enterprises half smaller than the good ones (table 2), lasso and ridge regressions showed the least type II and overall errors, but the reduction of the overall error was achieved at the expence of an increase in type I error.

Table 2. Model with the training set with failed enterprises half smaller than the good ones.

	Forward Condition	Forward LR	Forward Wald	Backward Condition	Backward LR	Backward Wald	Lasso	Ridge
Type I error	71,18%	71,18%	71,18%	65,88%	65,88%	71,18%	88,24%	81,76%
Type II error	6,17%	6,17%	6,17%	8,26%	8,26%	6,17%	1,34%	2,91%
Overall error	12,02%	12,02%	12,02%	13,45%	13,45%	12,02%	9,16%	10,01%

For the model with about 40% of the overall data used as training data (table 3), lasso and ridge regression showed the least type II and overall errors, with larger type I errors, but the differences with SPSS stepwise methods were not so substantial as in the previous case.

Table 3. Model with about 40% of the overall data used as training set.

	Forward Condition	Forward LR	Forward Wald	Backward Condition	Backward LR	Backward Wald	Lasso	Ridge
Type I error	91,80%	91,80%	91,80%	90,71%	90,71%	90,71%	97,81%	93,44%
Type II error	1,81%	1,81%	1,81%	1,81%	1,81%	1,64%	0,25%	0,90%
Overall error	13,57%	13,57%	13,57%	13,43%	13,43%	13,29%	13,00%	13,00%

The last case (table 4) where the training set had a number of failed enterprises larger than the number of good ones, lasso and ridge regression showed now the least type I error and larger type II and overall errors.

Table 4. Model with the training set with failed enterprises larger than the good ones.

	Forward Condition	Forward LR	Forward Wald	Backward Condition	Backward LR	Backward Wald	Lasso	Ridge
Type I error	32,50%	32,50%	32,50%	36,25%	36,25%	32,50%	28,13%	24,38%
Type II error	40,18%	40,18%	40,18%	38,76%	38,76%	40,18%	45,34%	45,29%
Overall error	39,58%	39,58%	39,58%	38,56%	38,56%	39,58%	44,00%	43,66%

5. Comments

Although the results showed an improvement of the overall classification for lasso and ridge regressions when compared to the stepwise methods implemented in SPSS in the second and third cases, that improvement was achieved with an increment in type I error, explained by a heavier percentage of healthy enterprises in the training data. For the model with the number of healthy enterprises equal to the number of failed ones in the training data, this behavior was not observed. An opposite pattern was observed in the last case.

From these results, the main conclusion is that ridge and lasso regressions behave not very distinctly from SPSS stepwise methods when the size of the healthy and failed enterprises in the training data is equal (although ridge regression showed the least type II and overall errors in that case, but with differences not very substantial), otherwise the lasso and ridge models tend to favor the category of the dependent variable that appears with heavier weight in the training set in a more outstanding way than what occurs in stepwise methods implemented in SPSS.

References

Ahn, H., Kim, K.-J., (2009). Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach. *Applied Soft Computing* 9, 599-607.

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., Haldeman, R. G., Narayanan P., (1977). Zeta Analysis: A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance* 1(1), 29–54.

Beaver, W. H., (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4(3), 71–111.

Bell, T. B., Ribar, G. S., Verchio, J. R., (1990). Neural Nets versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures. Deloitte & Touche, University of Kansas Simposium of Auditing Problems, Kansas City, 29–53.

Blum, M., (1974). Failing Company Discriminant Analysis. *Journal of Accounting Research* 12(1), 1–25.

Cessie, S., van Houwelingen, J. C., (1992). Ridge estimators in logistic regression. *Applied Statistics* 41(1), 191–201.

Coats, P. K., Fant, L. F., (1993). Recognizing Financial Distress Patterns Using a Neural Network Tool. *Financial Management* 22(3), 142–155.

Cule, E., De Iorio, M., (2012). A semi-automatic method to guide the choice of ridge parameter in ridge regression. arXiv:1205.0686v1 [stat.AP].

Duffy, D. E., Santner, T. J., (1989). On the small sample properties of norm-restricted maximum likelihood estimators for logistic regression models. Communications in Statistics - Theory and Methods 18, 959-980.

Elam, R., (1975). The Effect of Lease Data on the Predictive Ability of Financial Ratios. The Accounting Review 50(1), 25–43.

Fan, J., Li, R., (2005). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* 96, 1348–1360.

Friedman, J., Hastie, T., Simon, N., Tibshirani, R., (2015). glmnet: Lasso and elastic-net regularized generalized linear models. R package version 1.9-8, http://cran.r-project.org/web/packages/glmnet/index.html.

Friedman, J., Hastie, T., Tibshirani, R., (2008). Sparse Inverse Covariance Estimation with the Graphical Lasso. Biostatistics 9, 432–441.

Friedman, J., Hastie, T., Tibshirani, R., (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33(1), 1-22. http://www.jstatsoft.org/v33/i01/.

Goeman J. J., (2010). L1 and L2 Penalized Regression Models. Technical document for CRAN, http://cran.r-project.org/web/packages/penalized/penalized.pdf.

Hastie, T., Rosset, S., Tibshirani, R., Zhu, J., (2004). The Entire Regularization Path for the Support Vector Machine. *Journal of Machine Learning Research* 5, 1391–1415.

Hastie, T., Tibshirani, R., Friedman, J., (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Springer.

Hoerl, A. E., Kennard, R. W., (1970). Ridge Regression: Applications to Non-Orthogonal Problems. Technometrics 12(1), 69-82.

IBM Corporation, 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning with Applications in R. Springer.

Keasey, K., Watson, R., (1987). Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti´s Hypotheses. *Journal of Business Finance & Accounting* 14(3), 335–354.

Kim, M. J., Kang, D. K.,(2010). Ensemble with neural networks for bankruptcy. *Expert Systems with Applications* 37, 3373-3379.

Koh, H. C., Tan, S. S., (1999). A Neural Network Approach to the Prediction of Going Concern Status. Accounting and Business Research 29(3), 211–216.

Lacher, R. C., Coats, P. K., Sharma, S. C., Fant, L. F., (1995). A Neural Network for Classifying the Financial Health of a Firm. European Journal of Operational Research 85(1), 53–65.

Lennox, C., (1999). Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches. *Journal of Economics and Business* 51, 347–364.

Martin, D., (1977). Early warning of bank failure: a logit regression approach. *Journal of Banking and Finance* 1(3), 249–276.

Meier, L., van de Geer, S., Buhlmann, P., (2008). The Group Lasso for Logistic Regression. Journal of the Royal Statistical Society, Series B 70(1), 53-71.

Min, J. H., Lee, Y. C., (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications* 28(4), 603–614.

Moyer, R. C., (1977). Forecasting Financial Failure: A Re-Examination. *Financial Management* 6(1), 11–17.

Norton, C. L., Smith, R. E., (1979). A Comparison of General Price Level and Historical Cost Financial Statements in the Prediction of Bankruptcy. The Accounting Review 54(1), 72–86.

Ohlson, J. S., (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18(1), 109–131.

Park, M.Y., Hastie, T., (2007). L1-Regularization Path Algorithm for Generalized Linear Models. *Journal of the Royal Statistical Society*, Series B 69, 659-677.

Park, T., Casella, G., (2008). The Bayesian Lasso. *Journal of the American Statistical Association* 103 (482), 681-686.

Schaefer, R. L., Roi, L.D., Wolfe, R. A., (1984). A Ridge Logistic Estimator. *Communications in Statistics – Theory and Methods* 13(1), 99-113.

Serrano Cinca, C., Martin Del Brio, B., 1993. Prediccion de la Quiebra Bancaria Mediante el Empleo de Redes Neuronales Artificiales. Revista Espanola de Financiacion y Contabilidad 22(74), 153–176.

Tam, K., Kiang, M., (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science* 38(7), 926–947.

Tian, S., Yu, Y., Guo, H., (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance* 52, 89-100.

Tibshirani, R., (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, Series B 58(1), 267-288.

Yuan M., Lin Y., (2007). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society*, Series B 68(1), 49-67.

Zmijewski, M. E., (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research* 22, Supplement, 59–82.

Zou, H., Hastie, T., (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, Series B 67(2), 301-320.