Global Journal of Accounting and Economy Research
ISSN: 2319-443X • Vol. 4, No. 1, 2023, pp. 33-46

https://DOI:10.47509/GJAER.2023.v04i01.03

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

LOGIT RIDGE AND LASSO IN PREDICTING BUSINESS FAILURE

Vera Rabaça[1], José M. Pereira[2] and Mário Basto[3*]

[1]Higher School of Management, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
[2]CICF - Research Center on Accounting and Taxation, Polytechnic Institute of Cávado and Ave
[3]Higher School of Technology, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal
(Corresponding author; email: mbasto@ipca.pt)

Received: 5 December 2023; Revised: 2 February 2023;
Accepted: 15 February 2023; Published: 10 April, 2023

ABSTRACT

The issue of organisational insolvency is very important to the economies of all countries, and it is especially important in light of unforeseen events like the COVID-19 pandemic that broke out in 2020. For bankruptcy forecasting, the creation of statistical models for predicting insolvency is essential, and many models have been presented and assessed for the widest range of circumstances. Using data from 2017, 2018, and 2019, this study applies the logistic regression model, along with its ridge and lasso variations, to Portuguese SMEs in the textile industry and examines how well each model predicts the viability of the companies in 2020. The results show comparable forecasting abilities when using 2019 data; however, whereas the predictions for the conventional logistic regression model deteriorate as the distance to 2020 increases, for the other two models, they are comparable when using 2018 data and improve when using 2017 data, which was unexpected. Additional research is required to discover whether this propensity remains in other circumstances when unexpected occurrences, like the COVID-19 outbreak, have not occurred. This behaviour may, at least in part, be a result of the COVID-19 epidemic.

Keywords: bankruptcy, prediction models, logistic regression, ridge regression, lasso

1. INTRODUCTION

The issue of corporate insolvency is crucial for the economies of all nations. This is particularly valid in view of the 2020 COVID-19 pandemic. According to Lev (1978), a company is insolvent if it is unable to make timely debt payments or if its assets are insufficient to meet its obligations. Portugal has shown a decline in insolvencies since 2015, with a little increase in 2018. While studies conducted up to this point indicated a reduction in comparison to the year 2019, one had thought that 2020 would be a year of major increases.

When a business fails, it can have a number of unfavourable effects. Both the financial and social tolls can be high. As a result, it is not surprising that this topic has been high on the list of priorities for researchers. Though research on this topic has historically concentrated on financial factors, there has been a recent increase in attention on non-financial attributes in understanding business failures. Throughout the last five decades, several techniques were used to design models regarding this issue. Most of these studies used statistical or artificial intelligence methods to analyse corporate financial records in an effort to forecast whether or not the businesses under study would go bankrupt in the

Altman's Z-Score is perhaps the most frequently cited classical model for predicting company bankruptcy in the literature. Altman's Z-Score is a numerical metric used to estimate the likelihood that a company will fail during the next two years. Using discriminant analysis, it was created in 1968. In the subsequent decade, discriminant analysis was widely employed in published studies (Blum, 1974; Elam, 1975; Altman *et al.*, 1977; Moyer, 1977; Norton & Smith, 1979). It is suspected that researchers explored other techniques such as logit and probit because of criticisms highlighting this theory's weaknesses (Ohlson, 1980; Keasey & Watson, 1987; Zmijewski, 1984; Lennox, 1999).

The most widely used artificial intelligence techniques are decision trees, neural networks, induction of rules, and rough groups of ideas (Koh & Tan, 1999; Min & Lee, 2005; Kim & Kang, 2010).

Forecasts for company insolvency have been subjected to the Lasso variable selection technique, a variation of logistic regression. By using Lasso, Tian *et al.* (2015) examined the relative significance of numerous bankruptcy predictors widely used in the literature at the time. They discovered that the reduced model chosen using the Lasso technique outperformed the models used in the earlier research in terms of out-of-sample prediction. The lasso and ridge techniques were also employed by Pereira *et al.* (2016). The models were applied to a dataset of 2032 hotel industry companies that were not in bankruptcy and 402 companies that were, from 2010 to 2012. In comparison to stepwise techniques, the results showed that the lasso and ridge models tended to favour the category of the dependent variable that appeared with a greater weight in the training set.

In this study, a model for predicting the insolvency of SMEs in the textile industry was constructed using logistic regression. The ridge and lasso models, two versions of the logistic regression model, were used to create two extra models that were compared to the standard logistic regression model. The addition of these two models was intended to determine whether a more precise forecast could be produced than with the traditional logistic regression.

Small and medium-sized enterprises (SMEs) predominate in the Portuguese business sector. Nearly 99.90% of all firms in 2019 were SMEs. The size of the businesses considered in the current analysis was influenced by this factor. The textile industry was chosen because it employs a sizable fraction of Portuguese SMEs, it has one of the highest employment rates in Portugal, and there has been very little study on forecasting models using samples from this sector and nation.

2. MATERIAL AND METHODS

2.1. Sample and explanatory ratios

Financial information was extracted from the SABI (Sistemas de Análisis de Balances Ibéricos) database. The sample consists of insolvent and healthy SMEs in the textile industry. Based on the most relevant prior studies, such as Beaver (1966), Altman (1968), Altman *et al.* (1977), Altman *et al.* (1979), Ohlson (1980), and Altman and Sabato (2007), the explanatory variables were chosen for their ability to anticipate insolvency. Economic-financial ratios and a corporate characteristic variable (age) were examined. The logistic regression model (LOGIT) and its ridge and lasso variations were used to predict insolvency and identify the economic-financial and macroeconomic indicators with the greatest capacity to differentiate between healthy and insolvent businesses.

2.2. Logistic Regression Model

Logistic regression is a type of regression meant to predict and explain a categorical dependent variable with a binary value. The fundamental goals of logistic regression are the identification of the independent variables that explain the membership in each category of the dependent variable as well as the development of a classification system.

This form of analysis is frequently favoured over discriminant analysis due to the latter's reliance on the verification of more stringent assumptions, such as multivariate normality and covariance matrix equality between groups. When these assumptions are not fulfilled, logistic regression is significantly more robust, requiring neither a specified distribution of the independent variables nor considerations such as homocedasticity. Logistic regression is comparable to multiple regression in terms of statistical tests, methods for including nonnumerical independent variables by encoding them as dummy variables, non-linear effects, and diagnostic procedures.

In possession of the sample, one of the most significant tasks is to identify the independent variables that best explain and forecast the binary response variable. The selection of the optimal set of independent variables relies heavily on p-values, or the significance of the model variables. Thus, the stepwise method can be employed to determine which independent variables best explain the response variable. The stepwise method explores only a restricted number of models to identify the independent variables that best explain and forecast the binary response variable. The selection of the optimal set of independent variables relies heavily on p-values, or the significance of the model variables. Thus, the stepwise method can be employed to determine which independent variables best explain the response variable. The stepwise method explores only a restricted number of models (a series of models) and is a viable alternative to picking the optimal subset of independent variables, which is computationally intensive. However, if the correlation between variables is high, the respective p-values will also be high, which can lead to the exclusion of crucial variables in the explanation and prediction of the dependent variable as well as the inclusion of irrelevant variables (Pereira *et al.*, 2016). By using logistic regression modifications such as ridge and lasso regressions, this problem can be overcome. To prevent overfitting, cross-validation of the model can be accomplished by randomly dividing the sample into a training sample and a test sample.

The present study's sample consists of 340 insolvent SMEs and 1953 healthy SMEs for the year 2019, 270 insolvent SMEs and 1874 healthy SMEs for the year 2018, and 220 insolvent SMEs and 1831 healthy SMEs for the year 2017.

2.2.1. Math behind the logistic regression model and its ridge and lasso variations

The logistic regression model is determined by the probability of success of the dependent variable, with this category coded as one and the other category coded as zero. For k independent variables, the model is translated into the probability of the dependent variable being equal to one. For observation i:

$$P(y_i = 1) = \frac{e^{\beta_0 + \Sigma_{j=1}^{k} \beta_j x_{ji}}}{1 + e^{\beta_0 + \Sigma_{j=1}^{k} \beta_j x_{ji}}}$$

The logit (linearization) is defined by the logarithm of the odds:

$$\text{logit}\,[P(y_i = 1)] = \log \frac{P(y_i = 1)}{P(y_i = 0)} = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ji}$$

For *n* observations and *k* independent variables, the parameters are estimated by maximizing the corresponding log likelihood function (Pereira *et al.*, 2016):

$$l(\beta) = \sum_{i=1}^{n} \left[y_i\beta_0 + \sum_{j=1}^{k} y_i\beta_j x_{ji} - \log\left(1 + e^{\beta_0 + \Sigma_{j=1}^{k} \beta_j x_{ji}}\right)\right]$$

The maximum likelihood estimates will typically have low biases when the relationship between the logit of a dichotomous response and the predictors is close to linear, but they could have high variances when the number of covariates is high compared to the number of observations or when the variables are overly correlated. Because of this, a slight change in the training sample data can result in a significant change in the coefficient estimations.

Ridge (Cessie & Houwelingen, 1992) and lasso (Tibshirani, 1996) regressions work by trading a slight bias increase for a substantial reduction in predicted variance. They are more adept at handling multicollinearity and have the appropriate qualities to avoid the numerical instability caused by overfitting. The ridge and the lasso fit the model with all independent variables; however, the predicted coefficients tend to be closer to zero than the conventional estimates (James *et al.* 2013). The estimates in ridge and lasso regression are contingent upon the selection of a nonnegative fit parameter. The logistic regression's log likelihood function is slightly modified by the addition of a parcel involving this parameter (Cessie & Houwelingen, 1992).

For the ridge regression, a single model is fitted by maximizing the following function (James, 2013):

$$l_{\lambda}^{R}(\beta) = \sum_{i=1}^{n} \left[y_i\beta_0 + \sum_{j=1}^{k} y_i\beta_j x_{ji} - \log\left(1 + e^{\beta_0 + \Sigma_{j=1}^{k} \beta_j x_{ji}}\right)\right] - \lambda \sum_{j=1}^{k} \beta_j^2$$

As the value of λ increases, the estimated coefficients tend toward zero, but none will be quite zero. Thus, no variable selection is undertaken, as all independent variables are included in the model (James *et al.*, 2013), making its interpretability challenging when the number of variables *k* is big.

Lasso regression, on the other hand, is an alternative to traditional logistic regression that allows for a reduction in the number of independent variables by having some coefficients take exactly the value zero (James *et al.* 2013), making the model more interpretable. The function to be maximized takes the following form (Hastie *et al.*, 2009):

$$l_{\lambda}^{L}(\beta) = \sum_{i=1}^{n}\left[y_i\beta_0 + \sum\nolimits_{j=1}^{k} y_i\beta_j x_{ji} - \log\left(1 + e^{\beta_0 + \Sigma_{j=1}^{k}\beta_j x_{ji}}\right)\right] - \lambda\sum_{j=1}^{k}|\beta_j|$$

In both situations, the model's constant is unaffected, and the independent variables are often standardized so that the parameter λ makes sense.

Depending on the various values of the parameter λ, the values of the estimates will vary. For determining the optimal value of the parameter, k-fold cross-validation (Hastie *et al.*, 2009) is expected to be used. In this approach, the data is separated into k subgroups of about similar size, with each subset thereafter serving as the validation set. As training data, the remaining subsets are employed. This process is done k times, and the optimal value is determined so that the log likelihood is maximized (Goeman, 2010).

3. RESULTS

To find multivariate outliers, the data were first visually evaluated. According to the company's circumstances, the Mahalanobis distance was applied to each of the categories (active or bankrupt). This is a statistical distance that accounts for the variation and correlation of the different variables. As a result, data with greater values and p-values below 0.001 were disregarded because they might have a significant impact on the model's predictions.

Training samples and test samples were created from the sample. The number of active and inactive firms in the training sample was chosen at random to be equal. If the company had failed, positive results were allocated to the dichotomous dependent variable; if the company was still operating during the assessment year, negative values were assigned. An example of a positive result is a bankrupt corporation, whereas a negative result is represented by zero (active company).

Adjustment of the classic logistic regression equation was accomplished using the minitab software's stepwise procedure, significance of the variables, VIF values (variance inflation factors), and interpretability of the coefficients. The acquired results were compared to those of ridge and lasso regression performed with the R software and package glmnet (Friedman *et al.*, 2015).

The factor that was discovered to have the greatest relative importance for defining the company's condition in the three years under consideration was the ratio "Total Assets/Total Liabilities" (highest absolute value of the standardised coefficient).

The confusion matrices for the three models are shown in Tables 1 through 3, where the false positive rate is the type II error, or the percentage of active companies that the model predicts will go bankrupt, and the false negative rate is the type I error, or the percentage of bankrupt companies that the model predicts will go active.

Table 1: Confusion matrix for the classic logistic regression model

			Predicted active	*Predicted bankrupt*
Year 2017	Training sample	Active	Specificity = 63.6%	Type II error = 36.4%
		Bankrupt	Type I error = 20.0%	Sensitivity = 80.0%
		Overall accuracy = 71.82%		
	Test sample	Active	Specificity = 64.4%	Type II error = 35.6%
		Bankrupt	Type I error = 33.7%	Sensitivity = 66.3%
		Overall accuracy = 64.55%		
Year 2018	Training sample	Active	Specificity = 75.6%	Type II error = 24.4%
		Bankrupt	Type I error = 21.5%	Sensitivity = 78.5%
		Overall accuracy = 77.04%		
	Test sample	Active	Specificity = 69.4%	Type II error = 30.6%
		Bankrupt	Type I error = 27.5%	Sensitivity = 72.5%
		Overall accuracy = 69.64%		
Year 2019	Training sample	Active	Specificity = 77.6%	Type II error = 22.4%
		Bankrupt	Type I error = 22.9%	Sensitivity = 77.1%
		Overall accuracy = 77.35%		
	Test sample	Active	Specificity = 75.1%	Type II error = 24.9%
		Bankrupt	Type I error = 29.9%	Sensitivity = 70.1%
		Overall accuracy = 74.65%		

Concerning 2019 data, the global overall accuracy rate is comparable for all three models; however, in 2017 and 2018, the ridge and lasso regressions produced a higher global overall accuracy rate than the conventional logistic regression, with an unexpected rise in the global overall accuracy rate in 2017.

Similarly, the discriminant ability to discriminate between active and bankrupt companies provided by the ROC curves was generated for the three

Table 2: Confusion matrix for the ridge regression model

			Predicted active	*Predicted bankrupt*
Year 2017	Training sample	Active	Specificity = 84.5%	Type II error = 15.5%
		Bankrupt	Type I error = 19.1%	Sensitivity = 80.9%
		Overall accuracy = 82.73%		
	Test sample	Active	Specificity = 78.0%	Type II error = 22.0%
		Bankrupt	Type I error = 26.9%	Sensitivity = 73.1%
		Overall accuracy = 77.72%		
Year 2018	Training sample	Active	Specificity = 81.5%	Type II error = 18.5%
		Bankrupt	Type I error = 20.0%	Sensitivity = 80.0%
		Overall accuracy = 80.74%		
	Test sample	Active	Specificity = 72.9%	Type II error = 27.1%
		Bankrupt	Type I error = 29.0%	Sensitivity = 71.0%
		Overall accuracy = 72.79%		
Year 2019	Training sample	Active	Specificity = 78.2%	Type II error = 21.8%
		Bankrupt	Type I error = 21.8%	Sensitivity = 78.2%
		Overall accuracy = 78.24%		
	Test sample	Active	Specificity = 74.2%	Type II error = 25.8%
		Bankrupt	Type I error = 34.1%	Sensitivity = 65.9%
		Overall accuracy = 73.53%		

Confusion matrix for the lasso regression model

			Predicted active	*Predicted bankrupt*
Year 2017	Training sample	Active	Specificity = 90.0%	Type II error = 10.0%
		Bankrupt	Type I error = 16.4%	Sensitivity = 83.6%
		Overall accuracy = 86.82%		
	Test sample	Active	Specificity = 81.4%	Type II error = 18.6%
		Bankrupt	Type I error = 27.9%	Sensitivity = 72.1%
		Overall accuracy = 80.83%		
Year 2018	Training sample	Active	Specificity = 80.0%	Type II error = 20.0%
		Bankrupt	Type I error = 23.0%	Sensitivity = 77.0%
		Overall accuracy = 78.52%		
	Test sample	Active	Specificity = 74.0%	Type II error = 26.0%
		Bankrupt	Type I error = 30.4%	Sensitivity = 69.6%
		Overall accuracy = 73.69%		
Year 2019	Training sample	Active	Specificity = 80.0%	Type II error = 20.0%
		Bankrupt	Type I error = 20.6%	Sensitivity = 79.4%
		Overall accuracy = 79.71%		
	Test sample	Active	Specificity = 75.1%	Type II error = 24.9%
		Bankrupt	Type I error = 34.7%	Sensitivity = 65.3%
		Overall accuracy = 74.30%		

models. The ROC curve stands for Receiver Operator Characteristic curve, and it is a graphic representation of a classification binary model across all possible cutoff points. As a measure of the model's quality, the area under the ROC curve was computed, and its values are presented in Tables 4 to 6. According to Hosmer and Lemeshow (2000), the models show a satisfactory to high capacity to differentiate between active and bankrupt companies (area greater than 0.7 and 0.8).

Table 4: Area under ROC curve for the classic logistic regression model

	Area under ROC curve (training sample)	*Area under ROC curve (test sample)*
Year 2017	0.7883	0.7311
	CI 95% = (0.729, 0.847)	CI 95% = (0.685, 0.777)
Year 2018	0.8454	0.7846
	CI 95% = (0.799, 0.892)	CI 95% = (0.745, 0.825)
Year 2019	0.8604	0.8107
	CI 95% = (0.822, 0.899)	CI 95% = (0.776, 0.845)

Table 5: Area under ROC curve for the ridge regression model

	Area under ROC curve (training sample)	*Area under ROC curve (test sample)*
Year 2017	0.8989	0.8231
	CI 95% = (0.859, 0.939)	CI 95% = (0.777, 0.869)
Year 2018	0.8614	0.7922
	CI 95% = (0.818, 0.905)	CI 95% = (0.752, 0.832)
Year 2019	0.8700	0.7781
	CI 95% = (0.833, 0.907)	CI 95% = (0.740, 0.817)

Table 6: Area under ROC curve for the lasso regression model

	Area under ROC curve (training sample)	*Area under ROC curve (test sample)*
Year 2017	0.9036	0.8461
	CI 95% = (0.863, 0.944)	CI 95% = (0.800, 0.893)
Year 2018	0.8526	0.7916
	CI 95% = (0.807, 0.898)	CI 95% = (0.751, 0.832)
Year 2019	0.8747	0.7802
	CI 95% = (0.839, 0.911)	CI 95% = (0.742, 0.818)

When utilising data from 2019, the results show that the classic logistic regression has a little better ability to discriminate between operational and bankrupt businesses than the other two models, but there are no appreciable differences when using data from 2018. Ridge and lasso regression perform noticeably better with 2017 data. In contrast to the traditional logistic regression, the ridge and lasso regressions get better at discriminating as we move away from 2020.

For the three models spanning the three years (2017, 2018, and 2019), the differences between the training and test samples were within the predicted range, which is pretty good.

4. DISCUSSION

Many studies on business insolvency and enterprise failure have been conducted throughout the years, and numerous conclusions have been reached. To be able to make timely decisions and thereby reduce the likelihood of failure in the future, it is crucial to have models with a strong predictive capability.

As a prediction model, the present work employs logistic regression as well as two variants of this one, ridge and lasso regressions, applied to Portuguese SMEs in the textile industry. Although there are few studies in Portugal using logistic regression in this sector, the findings of this investigation were, by comparison, pretty good. Thus, the findings were superior to those obtained by Leal and Machado-Santos (2007) for the external validation sample using the logit model, who obtained an accuracy rate of 65 percent for the previous year in the same sector. In addition, the accuracy rate was higher than what Pacheco (2015) observed with his four models, whose accuracy ranged from 63 to 69.7 percent for a sample of companies from the same country but from a different industry. Nevertheless, more research is needed to verify the precision of the predictions achieved in this study for the Portuguese textile industry.

5. CONCLUSION, LIMITATIOS & FUTURE WORK

5.1. Conclusion

It is noteworthy that the "total assets/total liabilities" ratio demonstrated to have the greatest relative impact in determining the company's status over the three years examined, despite the fact that the goal of this study was not to determine the significance of the variables included in the logistic model (2017, 2018 and 2019). which follows logically from the fact that the likelihood of the business going bankrupt increases with the total liabilities to assets ratio.

Logistic regression and its ridge and lasso variants achieved remarkably similar global accuracy rates in 2019 (74.6 percent versus 73.53% and 74.30%, respectively). The relative global accuracy rates in 2018 were 69.64%, 72.79%, and 73.69%, and in 2017, they were 64.55%, 77.72%, and 80.83%, respectively. Two and three years in advance (years 2018 and 2017), the global accuracy rates differed somewhat between the three models.

Given that the amount of uncertainty grows as one gets further away from the year 2020, the global accuracy rate for the traditional logistic regression fell. The total accuracy rates in 2019 and 2018 using the ridge and lasso regression, however, were very similar, with an increase in 2017 (to 77.72 and 80.83%, respectively), or three years in advance.

In short, the global overall accuracy rate is comparable for all three models in 2019, but in 2017 and 2018, the ridge and lasso regressions achieved a greater global overall accuracy rate than the classic logistic regression, with an unexpected increase in 2017.

The ROC curve's results and other outcomes were quite similar. The findings indicate that using data from 2019, the classic logistic regression has a marginally stronger discriminant ability than the other two models; however, using data from 2018, there are no discernible changes. With 2017 data, lasso and ridge regression perform significantly better.

These findings might be partially explained by the COVID-19 epidemic, but more study is needed to determine whether this inclination persists in situations when unanticipated events, like the COVID-19 epidemic, have not taken place.

5.2. Limitations

As previously suggested, the COVID-19 pandemic broke out at the beginning of the year 2020, causing mandatory lockdowns and the closure of various enterprises, presumably to stop the virus' spread. Due to the forecast data only extending to 2019, this element, which may have contributed to the insolvency outcome of some organisations, is not taken into consideration in the model.

The majority of the sampled organisations did not have their accounts audited since they were not obligated to do so, which is another flaw in the study because it means that the accounting information might not fully reflect the actual financial and economic situation.

Last but not least, the SABI database lacks information on a substantial fraction of Portuguese SMEs, mostly insolvent businesses.

5.3. Future Work

It is suggested that in further studies, samples from the same sector in other nations or from several economic sectors in the same nation be used to test the model's resilience. The comparison of this methodology with other statistical and artificial intelligence models and techniques that have been applied in this area of research will be another extension of this work.

Acknowledgement

The authors are grateful to the anonymous reviewers and the editor for their comments and for editing this paper.

Declaration of conflict of interest

There exist no ethical issues bothering the study and sponsorship regarding funding and related issues of contradictions.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance,* 23(4): 589-609.

Altman, E. I., Baidya, T., & Dias, L. (1979). Assessing potential financial problems of firms in Brazil, *Journal of International Business Studies,* 10(2): 9-24.

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.

Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market, *Abacus-a Journal of Accounting Finance and Business Studies,* 43 (3). 332-357.

Beaver, W. H. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research*, 4(3), 71–111.

Elam, R., (1975). The effect of lease data on the predictive ability of financial ratios, *The Accounting Review,* 50(1). 25–43.

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.

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., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* , (2nd edition), Springer.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd edition), New York: John Wiley.

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

Leal, C. P., & Machado-Santos, C. (2007). Insolvency prediction in the Portuguese textile industry, *European Journal of Finance and Banking Research,* 1(1). 16-28.

Lennox, C. (1999). Identifying failing companies: A e-evaluation of the Logit, Probit and DA approaches, *Journal of Economics and Business,* 51(4). 347–364.

Lev, B. (1978). *Análisis de los Estudos Financieros: Un Nuevo Enfoque*, Madrid: Ediciones ESIC.

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. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research,* 18(1). 109-131.

Pacheco, L. (2015). SMEs probability of default: The case of the hospitality sector, *Tourism & Management Studies,* 11(1) : 153-159.

Pereira, J. M., Basto, M., & Silva, A. F. (2016). The logistic lasso and ridge regression in predicting corporate failure, *Procedia Economics and Finance,* 39. 634-641.

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.

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.

Tam, K., & Kiang, M. (1992). Managerial applications of neural networks: The case of bank failure predictions, *Management Science,* 38(7): 926–947.

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

Zmijewski, M. E. (1984) Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22, (Supplement): 59–82.

To cite this article:

Vera Rabaça, José M. Pereira and Mário Basto (2023). Logit Ridge and Lasso in Predicting Business Failure. *Global Journal of Accounting and Economy Research*, Vol. 4, No. 1, 2023, pp. 33-46.