Centre Emile Bernheim
Research Institute in Management Sciences

Solvay Brussels School
Economics & Management

Partial Least Square Discriminant Analysis (PLS-DA) for bankruptcy prediction

C. Serrano-Cinca and B. Gutiérrez-Nieto

This paper uses Partial Least Square Discriminant Analysis (PLS-DA) for the prediction of the 2008 USA banking crisis. PLS regression transforms a set of correlated explanatory variables into a new set of uncorrelated variables, which is appropriate in the presence of multicollinearity. PLS-DA performs a PLS regression with a dichotomous dependent variable. The performance of this technique is compared to the performance of 8 algorithms widely used in bankruptcy prediction. In terms of accuracy, precision, F-score, Type I error and Type II error, results are similar; no algorithm outperforms the others. Behind performance, each algorithm assigns a score to each bank and classifies it as solvent or failed. These results have been analyzed by means of contingency tables, correlations, cluster analysis and reduction dimensionality techniques. PLS-DA results are very close to those obtained by Linear Discriminant Analysis and Support Vector Machine.

Keywords: Bankruptcy, financial ratios, banking crisis, solvency, data mining, PLS-DA

CEB Working Paper N° 11/024
2011

Université Libre de Bruxelles - Solvay Brussels School of Economics and Management
Centre Emile Bernheim
ULB CP114/03 50, avenue F.D. Roosevelt 1050 Brussels BELGIUM
e-mail: ceb@admin.ulb.ac.be Tel. : +32 (0)2/650.48.64 Fax : +32 (0)2/650.41.88

ULB

Partial Least Square Discriminant Analysis (PLS-DA) for bankruptcy prediction

C. Serrano-Cinca*
Department of Accounting and Finance
University of Zaragoza, Spain.

B. Gutiérrez-Nieto
Department of Accounting and Finance
University of Zaragoza, Spain.

ABSTRACT

This paper uses Partial Least Square Discriminant Analysis (PLS-DA) for the prediction of the 2008 USA banking crisis. PLS regression transforms a set of correlated explanatory variables into a new set of uncorrelated variables, which is appropriate in the presence of multicollinearity. PLS-DA performs a PLS regression with a dichotomous dependent variable. The performance of this technique is compared to the performance of 8 algorithms widely used in bankruptcy prediction. In terms of accuracy, precision, F-score, Type I error and Type II error, results are similar; no algorithm outperforms the others. Behind performance, each algorithm assigns a score to each bank and classifies it as solvent or failed. These results have been analyzed by means of contingency tables, correlations, cluster analysis and reduction dimensionality techniques. PLS-DA results are very close to those obtained by Linear Discriminant Analysis and Support Vector Machine.

KEY WORDS

Bankruptcy, financial ratios, banking crisis, solvency, data mining, PLS-DA

ACKNOWLEDGEMENTS

The work reported in this paper was supported by grant ECO2010-20228 of the Spanish Ministry of Education and Science, and the European Regional Development Fund and by grant Ref. S-14 (3) of the Government of Aragon.

* **Carlos Serrano-Cinca**: *Department of Accounting and Finance, Fac. Economía y Empresas, Univ. Zaragoza, Gran Vía 2, Zaragoza (50.005)* ***Spain****, serrano@unizar.es http://ciberconta.unizar.es/charles.htm*

Partial Least Square Discriminant Analysis (PLS-DA) for bankruptcy prediction

1. INTRODUCTION

Bankruptcy prediction from financial ratios using mathematical models is a classical in data mining research. Since the pioneer work by Beaver (1966), that used univariate ratio analysis, many different techniques have been employed. Altman (1968) used Linear Discriminant Analysis (LDA); Ohlson (1980) used Logistic Regression (LR); Marais et al (1984) used Decision Trees such as Id3, C4.5 and Random Trees; Tam and Kiang (1992) used Multilayer Perceptron (MLP), a neural network model and K-Nearest Neighbors (KNN); Serrano-Cinca (1996) and du Jardin and Séverin (2011) applied Self Organizing Feature Maps; Fan and Palaniswami (2000) used Support Vector Machine (SVM) and Sarkar and Sriram (2001) applied Naive Bayes (NB). Techniques of ensembles, such as Boosting or Bagging, have been applied by Foster and Stine (2004), who combined C4.5 and Boosting; while Mukkamala et al (2006) combined Bagging and Random Tree (BRT). This paper applies Partial Least Square Discriminant Analysis (PLS-DA). To the best of our knowledge, this technique has not previously been applied to bankruptcy prediction.

PLS regression combines features from Principal Component Analysis (PCA) and Multiple Linear Regression (Wold, 1966). PLS-DA is based on the PLS model, being the dependent variable a categorical one, which is useful for classification tasks. PLS-DA is a standard tool in Chemometrics, because of its capability to deal with multicollinearity (Barker and Rayens, 2003; Westerhuis et al, 2008). Financial ratios commonly used for failure prediction suffer from multicollinearity. This is a severe problem, usually approached using stepwise selection procedures, or principal components, which are orthogonal by definition.

The paper compares PLS-DA with 8 algorithms (LDA, LR, MLP, KNN, NB, SVM, C4.5, and BRT). 2008-2011 USA banking crisis data have been used. Many studies analyze bankruptcy, and compare the results obtained by the different techniques, see Ravi Kumar and Ravi (2007) and Demyanyk and Hasan (2010) for a revision. It can be concluded that no technique is clearly better than others, because it depends on the problem analyzed and the performance measure chosen, Caruana and Niculescu-Mizil (2006). LDA is optimum if data satisfied some statistical properties. Unfortunately, financial ratios are not normally distributed, as shown by Ezzamel and Mar-Molinero (1987), which justifies the use of alternative techniques. LR needs fewer requisites, and being regression based, results can be interpreted in a straightforward way. When relationships are not linear, MLP performs well, because neural networks are universal approximators, as Hornick et al (1989) showed. But results by MLP are difficult to interpret. By contrast, decision tree algorithms,

such as C4.5, have the advantage that they generate rules useful for the design of expert systems. Boosting and Bagging techniques improve the performance of decision trees. The strength of PLS-DA is its capability to deal with multicollinearity, because it transforms original variables into orthogonal components. It is also robust to missing data and skew distributions, see Cassel et al (1999). It can also deal with the 'too few cases/too many variables' problem. These problems are common in financial information; a priori, PLS-DA seems to be a promising technique.

To compare PLS-DA with the rest of the techniques, their performance was first analyzed. Several performance measures have been proposed. Ferri et al (2009) analyze the behavior of 18 different performance metrics; being accuracy, precision or F-score the most popular. One technique can obtain better results than another by using a given performance measure; but it can obtain poorer results by using another performance measure, Caruana and Niculescu-Mizil (2006). But it can also be the case that two techniques obtain the same performance, but one of them correctly classifies some cases and the other one correctly classifies different cases. The paper analyzes coincidences and divergences in the classification produced by the 9 techniques, by means of contingency tables and Phi correlations. Each technique obtains a score for each bank, which can be interpreted as a solvency measure. A table with banks in rows and the 9 scores in columns can be obtained. This table can be analyzed with multivariate analysis, by using Cluster Analysis (CA) and Categorical Principal Components Analysis (CATPCA). This approach will produce a taxonomy of techniques, it being specially interesting to know the techniques closest to PLS-DA.

The paper analyzes the banking crisis. Financial data are taken from the Federal Deposit Insurance Corporation (FDIC), an independent agency created by the US Congress to maintain stability and public confidence in the banking system. The crisis has caused the failure of an important number of USA banks. 140 banks failed in 2009, 157 in 2010, and there are still banks in difficulties, since the crisis is still not over. The FDIC (2011) affirms that there are a good number of banks, 884, on their "problem list". The procedure followed by this paper, based on pattern recognition, will allow the identification of these banks.

The following section presents the conceptual background, with an emphasis on presenting the PLS-DA technique. The third section presents the empirical study, focusing on performance comparison. Finally, conclusions and bibliography are presented.

2. CONCEPTUAL BACKGROUND

Failure prediction from the analysis of financial ratios has been a fruitful research line, as shown by the state of the art revisions (Zavgren, 1983; Ravi Kumar and Ravi, 2007; and Demyanyk and Hasan, 2010). The use of financial ratios as a tool to analyze the financial health of the

companies has a long pedigree; the work by Beaver (1966) marks a starting point from an academic point of view. Beaver uses univariate analysis, but companies' health has a multivariate nature, which is not captured by this approach. Altman (1968) starts using multivariate techniques, particularly Linear Discriminant Analysis (LDA). LDA derives a linear combination of ratios which best discriminate between failed and non-failed firms. An overall score, known as Altman's Z-Score, can be calculated from LDA. LDA depends on several restrictive assumptions, such as linearity, normality, or independence among predictors, whose violation has been studied by Eisenbeis (1977). Ohlson (1980) applies Logistic Regression (LR). In common with LDA, LR weights financial ratios and obtains a score. LR and LDA share the same underlying linear model, but LR optimizes the conditional likelihood. This supposes that it is not necessary for some of the restrictive assumptions of LDA to take place. Despite these differences, both LDA and LR obtain very similar results in practical applications with financial information. According to the empirical study by Lo (1986), the null hypothesis that LDA and LR are equivalent may not be rejected.

A different approach is followed by decision trees used by Marais et al (1984). Decision trees employ recursive partitioning algorithm to induce rules on a given data set. Widespread algorithms in failure prediction are Id3 and C4.5 by Quinlan (1993). Decision trees have been successful in obtaining useful bankruptcy prediction rules. However, decision tree algorithms may suffer from overfitting: the algorithm reduces training set error at the cost of an increased test set error.

Tam and Kiang (1992) employed Multilayer Perceptron (MLP), a neural network model to predict bankruptcy. Hornick et al (1989) proved that under certain weak conditions, multilayer feedforward networks perform as a class of universal approximators. This explains their powerful capability for classification. A meta-analysis performed by Adya and Collopy (1998) reveals that neural networks outperformed alternative approaches in 19 out of the 22 analyzed studies; but they warn that the bias against publication of negative results may mean that successful applications are over-represented in the published literature. It is difficult to obtain an optimal combination of parameters that produces the best prediction performance: there are a large number of controlling parameters, such as the number of hidden layers, the number of hidden nodes, the learning rate, the momentum term, epochs, and transfer functions (Shin et al 2005).

k-Nearest Neighbors (KNN) is one of the most straightforward machine learning algorithms (Fix and Hodges, 1951), which is very useful when there is no prior knowledge about the distribution of the data. Employed by Tam and Kiang (1992) and Park and Han (2002) in bankruptcy prediction, KNN is the fundamental for many Case Base Reasoning (CBR) developments, widely used in credit scoring. When evaluating a new application, CBR retrieves similar cases in the case base.

The aim of Support Vector Machine (SVM) is to find the best hyperplane that represents the

largest separation between the two classes, failed and non-failed firms. It has been applied to loan default prediction by Fan and Palaniswami (2000). SVM are based on few restrictive assumptions, and they are obtaining very promising results in failure prediction (Ravi Kumar and Ravi, 2007).

Naive Bayes (NB) algorithm is based on conditional probabilities. NB looks at the historical data and uses Bayes' Theorem to calculate a probability by counting the frequency of values and combinations of values. Sarkar and Sriram (2001) used NB for predicting failure. Among the advantages of NB, Sun and Shenoy (2007) highlight that it does not have any requirements on the underlying distributions of variables, it does not require complete information for observations and is easy to understand because the relationships among variables are explicitly represented. They empirically compare LR and NB, finding that there is no significant difference (at the 5% level) between both models' performance.

In recent years, techniques of ensembles combining multiple classifiers, like Boosting or Bagging (or bootstrap aggregating) have been developed. They have been applied to bankruptcy prediction by Foster and Stine (2004) and Alfaro et al (2008), that blend C4.5 and Boosting, and Mukkamala et al (2006) that blend Bagging with Random Tree (BRT). Both algorithms are appropriate to improve the performance of the classification algorithms. The main difference between them is that Bagging trains each individual classifier independently while Boosting trains each individual classifier dependently (Hung and Cheng, 2009).

PLS regression (PLS) combines features from Principal Component Analysis (PCA) and multiple regression. PLS tries to provide a dimension reduction strategy in a situation where we want to relate a set of independent variables to a set of dependent variables. But PLS is a supervised method because it uses the independent as well as the dependent variables, whereas PCA is a non supervised method that considers only independent variables. PLS was developed by Wold (1966), trying to solve the multicollinearity problem. This is its main advantage facing Ordinary Least Square (OLS) regression, because the PLS components are orthogonal by definition. PLS is a useful method when there is a large number of explanatory variables, Garthwaite (1994). PLS is also robust to missing data and skew distributions, Cassel et al (1999). These are frequent problems when dealing with financial information, Deakin (1976). Yang et al (2011) use feature selection based on PLS and later apply SVM to predict bankruptcy.

Partial Least Squares Discriminant Analysis (PLS-DA) is a PLS regression of a set of binary variables on a set of predictor variables (Pérez-Enciso and Tenenhaus, 2003). PLS and PLS-DA are standard tools in Chemometrics, and widely used in Genetics and other knowledge areas, see Wold et at (2001) and Kjeldahl and Bro (2010). However, to the best of our knowledge, it has not been used before in failure prediction, credit scoring or similar task that use financial information.

3. EMPIRICAL STUDY

3.1 Sample and data

The USA banking crisis, rooted in 2008, and still not over in 2011, is presented in this paper. The database comes from the Federal Deposit Insurance Corporation (FDIC), publicly available on the Internet. 2008 accounting statements from 8,293 banks are taken. For each bank 17 financial ratios were extracted, trying to capture the main issues in banking financial health. Ratios and their definition can be found in Table 1.

***Table 1 ***

The banking business consists of buying and selling money. Because money is a standardized product, it is said that banking is a business of intermediation, where margins are crucial. The first six ratios relate incomes and expenses to assets. They are yield on earning assets (INTINCY), cost of funding earning assets (INTEXPY), net interest margin (NIMY), noninterest income to earning assets (NONIIY), noninterest expense to earning assets (NONIXY), and net operating income to assets (NOIJY). The following set are profitability ratios, that relate profits to equity or assets, such as return on assets (ROA), pretax return on assets (ROAPTX), return on equity (ROE), and retained earnings to average equity (ROEINJR). Next is the efficiency ratio, EEFFR, a key aspect in banking. Berger and Humphrey (1997) review 130 studies on efficiency issues in financial institutions. Notice that the higher the EEFFR value, the lower the bank's efficiency. Delinquency is measured with the ratio NPERFV, that relates assets that are past due 90 to assets. The last set is solvency ratios. They are relevant, because they measure the capital strength of the bank: net loans and leases to deposits (LNLSDEPR), equity capital to assets (EQV), and three ratios that include risk-weighted assets (RBC1AAJ, RBC1RWAJ and RBCRWAJ), whose definitions come from Bank for International Settlements (2004) and Federal Register (2007).

An exploratory data analysis was performed, with results presented in Table 2. Mean, median and standard deviation of the financial ratios are displayed into two groups: 320 failed and 7,973 non-failed banks. Several univariate mean and median tests were performed, to assess their discriminatory power. An outlier analysis was also performed, but no entity was dropped. The only ratio that presented extremely high values was LNLSDEPR, because there were 6 non-failed banks that hardly collect deposits. These 6 banks are in the test sample, so they not affect the parameter estimation performed by the algorithms.

Table 2

In bankruptcy studies, it is common to take a sample of failed and solvent firms, to apply cross validation, jackknife or bootstrap. Lau (1987) criticized some of these early studies because holdout

samples were drawn from the same time period as the original samples, which is not an intertemporal validation. This is not our case, because it reproduces a real world situation. Another strength is that it analyzes the whole USA banking sector. Algorithms were trained as if being financial analysts situated at the end of 2009. At that time, 2008 annual statements were available, and the 140 failed banks in 2009 were identified. For each failed bank, a same sized non-failed bank was matched. The paired matched sample technique is commonly used in bankruptcy studies (Beaver 1966, and Zmijewski 1984). To sum up, the trained sample had 140 failed banks and 140 non-failed banks.

The test sample consisted of 180 banks that failed after 2009 (157 failed in 2010 and 23 failed until February 2011, when data were collected), and 7,833 non-failed banks. The crisis is still not over, which differs from other studies. FDIC (2011) claims that there are 884 banks in their "problem list", but they do not make the details of this list publicly available, because that could provoke a run bank. Pattern recognition algorithms will obtain a list of banks with similar financial features to those already failed. When analyzing the test results, Type I error rate is relevant, because the model identifies as solvent banks that are already failed. But Type II error rate is not relevant because non-failed banks, according to the algorithm, could still go bankrupt.

3.2 Performance comparison

There are many performance measures for supervised algorithms. Ferri et al. (2009), Martens et al. (2011), and Niculescu-Mizil (2006) provide a description of the most frequently used. This study has calculated 6 of them: accuracy, Type I and Type II error rates, precision of failed and precision of non-failed, and F-score. Most bankruptcy studies emphasize accuracy. Type I error rate is also important, because the misclassification of a failed firm is considered more costly than the opposite. 1 minus Type II error rate is defined as recall. Often there is an inverse relationship between precision and recall, where it is possible to increase one at the cost of reducing the other (Chen and Lin, 2010). For this reason, it is important to calculate the F-score (van Rijsbergen, 1979), defined as the weighted harmonic mean of precision and recall.

Table 3 shows the results obtained with each algorithm. It has two panels: train and test. There are different commercial software to implement PLS-DA, like Unscrambler or PLS_Toolbox, or R libraries. Most of the PLS-DA software is intended for Chemometrics. We have chosen the Tanagra software by Rakotomalala (2005), for its data mining orientation, and because is research driven. An additional advantage of Tanagra is that not only implements PLS-DA, but also the rest of the algorithms. For each algorithm, different parameters were tried. Table 3 shows, for each algorithm, the configuration that obtained the best result.

***Table 3 ***

To minimize the multicollinearity problem, that affects both LDA and LR, a stepwise selection process, with both forward search and backward search, was performed. One of the problems of MLP is the selection of the appropriate architecture, as well as the learning parameters. The number of neurons was varied from 1 to 10. The error rate threshold was ranged from 0.1 to 0.01. Maximum iteration parameters of 100 and 200 were tried to select the best network.

As for K-NN, Euclidean and Heterogeneous Euclidean Overlap Metric (HEOM) distance for continuous attributes was used. Neighbourhood size was varied from 2 to 10. When using NB algorithm, a discretization process using the Minimum Description Length Principle Cut criterion (MDLPC) was first performed. Then, the algorithm was executed using Laplacian probability estimate with Lambda = 1. Two feature selection processes were chosen: Mutual Information Feature Selector (MIFS) and Fast Correlation Based Filter (FCBF).

SVM also allows different configurations and parameters. The following kernels were carried out: polymon, normalized polynom and Radial Basis Function (RBF). Polynom exponents, gamma values and epsilon parameters were varied.

Boosting C4.5 was estimated using several options. Boosted trees after 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 and 2048 replications were built. As for BRT, trees after 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 and 2048 replications were bagged. The number of PLS axes used in PLS-DA was 2.

Confusion matrix in Table 3 shows true positives, true negatives, Type I errors and Type II errors. As for the train sample, the accuracy ratio of the 9 analyzed algorithms ranges from 90.70% for SVM and 100% for Boosting C4.5. Test results are very similar: F-score ranges from 96.55% for BRT and 97.89% for SVM. However, it has been already remarked that neither the accuracy nor the F-score or Type II error are relevant, because the crisis is still not over. The only truth is Type I error. In general, it varies between 61 banks erroneously classified as non-failed by NB (33.89% error rate) and 78 by SVM (43.33% error rate). To extract more conclusions from this table is not prudent; what really matters is the comparison of coincidences and divergences among algorithms in the following section.

3.3 Comparison of coincidences, divergences and scores

Two different techniques can obtain the same accuracy ratio, but they can assign every bank to a different class. The same happens with two teachers that pass the same percentage of students, but they pass different students. For this reason, we propose that coincidences and divergences obtained for every pair of algorithms should be analyzed. A table was obtained with banks as cases

and 9 dichotomous variables indicating if the bank has been classified as failed or non-failed, according to the 9 techniques. This is a 0-1 table. It has been analyzed with contingency tables, which compare the 9 techniques in a paired way (Table 4). This table shows the number of banks that have been classified as failed by both techniques, as non-failed by both techniques, and the divergences. For example, both LDA and LR classify 436 banks as failed and 7,666 as non-failed. This is to say, they classify 8,102 banks in the same way. LDA and MLP coincide in 415 and 7,642 banks; these total 8,057 banks. It can be concluded that LDA is more similar to LR than to MLP. The last column allows the comparison between PLS-DA and the rest of the techniques. LDA is the technique that has more coincidences with PLS-DA.

***Table 4 ***

Having a 0-1 table, another synthetic indicator that can be obtained is the correlation coefficient among techniques. Being dichotomous variables, it will be a Phi-correlation. Table 5 displays the correlation values. As can be seen, they range from 0.60 to 0.85. PLS-DA has the highest correlation with LDA (0.81).

***Table 5 ***

A Cluster Analysis has been obtained from these 0-1 data. The analysis was performed using Ward's method. Figure 1 shows the dendrogram. Different clusters are displayed. One of them contains the two decision tree techniques (C4.5, BRT) with NB. LR and MLP are very similar, which is not surprising, because LR can be considered a particular case of MLP without hidden layer (Sarle, 1994). LDA is similar to SVM, and this cluster also contains PLS-DA. KNN joins them to form one of the large clusters. The use of Cluster Analysis for algorithm classification is not new. Michie et al (1994, p 190) use it as a part of their Statlog study, one of the most comprehensive studies that compares techniques. They use standardized error rates as variables and recognize that "there are some surprising errors in the clusterings". In our opinion, the method used here for clustering, based on coincidences rather than on error rates, is more revealing.

Figure 1

Beyond the failed/non failed classification, each technique provides a score. Two banks can be classified as solvent, but a 0.9 score is better than a 0.6 score; the same way as for two students who passed, a high mark is better than "just passed". This score can be interpreted as a bank's wealth indicator, and after transforming it, as a default probability. Niculescu-Mizil and Caruana (2005) examine the relationship between the predictions made by different algorithms and true posterior probabilities. They conclude that some techniques, like MLP or bagged trees predict well-calibrated probabilities, while others such as NB or SVM have biases. To avoid this problem, the Spearman's

rank correlation coefficient among scores was calculated, which is non-parametric. Table 5 shows Spearman's correlations in brackets. Results of the score analysis are coherent with those obtained from the 0-1 table. In the case of PLS-DA the technique more correlated is, again, LDA [0.93].

The scores table can also be analyzed with multivariate techniques. To avoid bias problems, the scores have been treated as ordinary variables, by using Categorical Principal Components Analysis (CATPCA). Figure 2 displays the two first principal components. The first one explains most of the variance (86.4%) and can be interpreted as bank's solvency. The more to the right, the higher the solvency of a given bank. The figure also summarizes the results of the 9 classification techniques. Every technique is represented as a vector in the figure. The second principal component allows differentiating among groups of techniques. C4.5, BRT and NB point upwards to the second principal component. PLS-DA, LDA and SVM point downwards. In the middle MLP, LR and KNN are placed. These results are coherent with Cluster Analysis' results.

Figure 2

Figure 2 represents failed banks with triangles and non-failed banks with circles. It can be clearly appreciated that failed banks are on the left, in a low scores area; while non-failed banks are on the right. Type I errors can be perceived, that is, failed banks on the right. It is not appropriate to mention Type II errors, because the crisis is not over, and many banks currently classified as non-failed can still go bankrupt. The FDIC (2011) affirms that there are 884 banks on the "problem list". But their names remain unknown, because "the FDIC does not, at any time, comment on open financial institutions". Among the 7,973 banks that have not failed so far, there are, according to our study, 885 classified as failed at least by one of the techniques. This figure is almost the same as the one by the FDIC. There are 586 banks classified as failed by at least two of the techniques; 452 by at least for three; 360 by at least for four; 297 by at least for five; 236 by at least for six; 185 by at least for seven; and 149 by at least for eight. There are 108 non-failed banks so far, but classified as failed by all the 9 techniques. These 108 banks are placed on the left side of Figure 2, and it can be concluded that they are risking failure.

As for PLS-DA, the analysis of performance measures, coincidences and divergences, and scores comparison reveals that it offers very similar results to LDA, SVM, LR and MLP. It has the advantage that it is tolerant to multicollinearity, because the components are orthogonal. But if a researcher uses LDA or LR with an appropriate procedure for variable selection, results will be equally good.

5. CONCLUSIONS

This paper compares Partial Least Square Discriminant Analysis (PLS-DA) with other 8

techniques widely used for classification tasks. It has performed an empirical study with data of the 2008-2011 USA banking crisis.

In performance terms, the techniques obtain different results depending on the performance measures chosen. Some techniques have more accuracy but less recall. This justifies the use of performance measures like the F-score, the weighted harmonic mean of precision and recall. The study examines what is behind performance, by analyzing how each bank is classified according to the 9 techniques. With this aim, a contingency table has been calculated to compare, in a paired way, the classifications of each technique. The paper has also analyzed the scores assign to each bank by all the different techniques. Results have been analyzed by means of correlation coefficients and Cluster Analysis. Results allow obtaining a taxonomy of techniques which classify banks in a similar way. Logistic Regression and Multilayer Perceptron are found to be close. The same happens to Boosting C4.5 and Bagging Random Tree. It was found that PLS-DA results resemble Linear Discriminant Analysis and Support Vector Machine results. PLS-DA has the advantage that is not affected by multicollinearity, because its components are orthogonal.

The USA banking crisis is still not over. The Federal Deposit Insurance Corporation does not comment on open financial institutions, but it recognizes that there are many banks in risk of failure. The score analysis of each of the 9 techniques, interpreted by means of Categorical Principal Components Analysis, allows identifying those firms with difficulties.

REFERENCES

Adya M and Collopy F (1998). How effective are neural networks at forecasting and prediction? A review and evaluation. International Journal of Forecasting 17(5-6): 48-495.

Alfaro E, García N, Gámez M and Elizondo D (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks, Decision Support Systems, 45(1):110-122

Altman EI (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance 23(4):589-609.

Barker M and Rayens W (2003). Partial least squares for discrimination. Journal of Chemometrics, 17(3): 166–173.

Beaver WH (1966). Financial Ratios as Predictors of Failure, Journal of Accounting Research 4(Supplement):71-111.

Bank for International Settlements (2004). International Convergence of Capital Measurement and Capital Standards, Basel Committee on Banking Supervision, Bank for International Settlements (BIS) Basel, Switzerland

Berger AN and Humphrey DB (1997). Efficiency of financial institutions: International survey and directions for future research. European Journal of Operational Research, 98(2):175-212.

Caruana R and Niculescu-Mizil A (2006). An Empirical Comparison of Supervised Learning Algorithms, Proceedings of the 23 rd International Conference on Machine Learning, Pittsburgh, PA

Cassel CM, Hackl P and Westlund AH (1999). Robustness of partial least-squares method for estimating latent variable quality structures, Journal of Applied Statistics, 26(4): 435-446.

Chen P-I and Lin S-J (2010). Automatic keyword prediction using Google similarity distance. Expert Systems with Applications 37(3):1928-1938

Deakin E. (1976). Distributions of financial accounting ratios: some empirical evidence. The Accounting Review 51(1):90-96.

du Jardin P and Séverin E (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. Decision Support Systems, 51(3):701-711

Demyanyk Y and Hasan I (2010). Financial crises and bank failures: A review of prediction methods. Omega 38(5):315-324

Eisenbeis RA (1977). Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics, Journal of Finance 32(3):875-900.

Ezzamel M and Mar-Molinero C (1987). On the distributional properties of financial ratios. Journal of Business Finance and Accounting 14(4):463-481.

Fan A and Palaniswami M (2000). A new approach to corporate loan default prediction from financial statements. Proceedings of the computational finance/forecasting financial markets conference, London (CD), UK.

Federal Register (2007). Risk-Based Capital Standards: Advanced Capital Adequacy Framework - Basel II. 72 FR 69288

FDIC (2011). Quarterly Banking Profile, Fourth Quarter 2010, Federal Deposit Insurance Corporation [http://www2.fdic.gov/qbp/2010dec/qbpall.html]

Ferri C, Hernandez-Orallo J and Modroiu R (2009). An experimental comparison of performance measures for classification. Pattern Recognition Letters 30(1):27-38

Fix E, Hodges JL (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.

Foster DP and Stine RA (2004). Variable selection in data mining: Building a predictive model for bankruptcy. Journal of the American Statistical Association 99(466):303-313.

Garthwaite PH (1994). An Interpretation of Partial Least Squares. Journal of the American Statistical Association 89(425):122-27.

Hornik K, Stinchcombe M and White H (1989). Multilayer Feedforward Networks are Universal Approximators. Neural Networks 2(5):359-366.

Hung C and Chen JH (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. Expert Systems with Applications 36(3):5297-5303

Kjeldahl K and Bro R (2010). Some common misunderstandings in chemometrics. Journal of Chemometrics 24(7-8):558-564.

Lau AHL (1987). A five-state financial distress prediction model. Journal of Accounting Research 25(1):127-138.

Lo AW (1986). Logit versus Discriminant Analysis. A Specification Test and Application to Corporate Bankruptcies. Journal of Econometrics 31(2):151-178.

Marais ML, Patel J and Wolfson M (1984). The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications. Journal of Accounting Research 22(Supplement): 87-113

Martens D, Vanthienen J, Verbeke W and Baesens B (2011). Performance of classification models from a user perspective, Decision Support Systems, In Press.

Michie D, Spiegelhalter DJ and Taylor CC (1994). Machine learning, neural and statistical classification, Ed Ellis Horwood.

Mukkamala S, Tilve DD and Sung AH (2006). Computational intelligent techniques for financial distress detection. International Journal of Computational Intelligence Research, 2(1):60-65

Niculescu-Mizil, A and Caruana R (2005). Predicting Good Probabilities With Supervised Learning, pp 625–632 in L. De Raedt and S. Wrobel, editors. ICML ’05: Proceedings of the 22nd International Conference on Machine Learning, ACM, New York.

Ohlson JA (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18(1):109-31.

Park C-S and Han I (2002). A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction. Expert Systems with Applications 23(3):255-264

Pérez-Enciso M and Tenenhaus M (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. Human Genetics 112(5-6):581-92

Quinlan R (1993). C4.5: Programs for machine learning, Morgan Kaufmann Publishers, San Mateo, CA

Rakotomalala R (2005). TANAGRA: A free software for research and academic purposes. In: Proceedings of EGC’2005, RNTI-E-3, (2):697-702.

Ravi Kumar PR and 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

Sarkar S and Sriram RS (2001). Bayesian models for early warning of bank failures. Management Science 47(11):1457-1475.

Sarle WS (1994). Neural Networks and Statistical Models, Proceedings of the Nineteenth Annual SAS Users Group International Conference, April:1538-1550.

Serrano-Cinca, C. (1996). Self organizing neural networks for financial diagnosis. Decision Support Systems 17(3):227-238

Shin K-S, Lee T-S and Kim H-J (2005). An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications 28(1): 127-135

Sun L and Shenoy PP (2007). Using Bayesian networks for bankruptcy prediction: Some methodological issues. European Journal of Operational Research 180(2):738-753

Tam KY and Kiang MY (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. Management Science 38(7):926-947

Van Rijsbergen C (1979). Information Retrieval, London, 2nd edn. Butterworths

Westerhuis JA, Hoefsloot CJH, Smit S, Vis DJ, Smilde AK, van Velzen EJJ, Duijnhoven JPM, and van Dorsten FA (2008). Assessment of PLSDA cross validation. Metabolomics 4(1):81–89

Wold H (1966). Estimation of principal components and related models by iterative least squares. In: Krishnaiah PR, ed. Multivariate Analysis. New York: Academic Press 391–420

Wold S, Sjöström M and Eriksson L (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems 58(2):109-130

Yang Z, You W, and Ji G (2011). Using partial least squares and support vector machines for bankruptcy prediction. Expert Systems with Applications 38(7):8336-8342

Zavgren CV (1983). The Prediction of Corporate Failure: The State of the Art. Journal of Accounting Literature 2(Spring):1-35.

Zmijewski M (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research 22(1):59-82.

Variable	*Definition*

	All (N=82938)			*Failed (N=320)*			*Non-failed (N=7973)*		
	Mean	*Median*	*St dev*	*Mean*	*Median*	*St dev*	*Mean*	*Median*	*St dev*
INTINCY	6.26	6.27	1.19	6.44	6.44	0.74	6.25	6.26	1.20
INTEXPY	2.48	2.51	0.71	3.33	3.39	0.63	2.44	2.49	0.70

INTINCY	Yield on earning assets. Total interest income as a percent of average earning assets.
INTEXPY	Cost of funding earning assets. Annualized total interest expense on deposits and other borrowed money as a percent of average earning assets on a consolidated basis.
NIMY	Net interest margin. Total interest income less total interest expense as a percent of average earning assets.
NONIIY	Noninterest income to earning assets. Income derived from bank services and sources other than interest bearing assets as a percent of average earning assets.
NONIXY	Noninterest expense to earning assets. Salaries and employee benefits, expenses of premises and fixed assets, and other noninterest expenses as a percent of average earning assets.
NOIJY	Net operating income as a percent of average assets.
ROA	Return on assets (ROA). Net income after taxes and extraordinary items as a percent of average total assets.
ROAPTX	Pretax return on assets. Annualized pre-tax net income as a percent of average assets.
ROE	Return on Equity (ROE). Annualized net income as a percent of average equity on a consolidated basis.
ROEINJR	Retained earnings to average equity. Net income, less cash dividends declared, as a percent of average total equity capital.
EEFFR	Efficiency ratio. Noninterest expense, less the amortization expense of intangible assets, as a percent of the sum of net interest income and noninterest income.
NPERFV	Noncurrent assets plus other real estate owned to assets. Noncurrent assets are defined as assets that are past due 90 days or more plus assets placed in nonaccrual status plus other real estate owned (excluding direct and indirect investments in real estate).
LNLSDEPR	Net loans and leases to deposits. Loans and lease financing receivables net of unearned income, allowances and reserves as a percent of total deposits.
EQV	Total equity capital as a percent of total assets.
RBC1AAJ	Core capital (leverage) ratio. Tier 1 (core) capital as a percent of average total assets minus ineligible intangibles.
RBC1RWAJ	Tier 1 risk-based capital ratio. Tier 1 (core) capital as a percent of risk-weighted assets as defined by the appropriate federal regulator for prompt corrective action during that time period.
RBCRWAJ	Total risk-based capital ratio. Total risk based capital as a percent of risk-weighted assets.

Table 1. Variables employed and their definition. Acronyms and definitions are taken from the Federal Deposit Insurance Corporation (FDIC)

NIMY	3.78	3.77	1.12	3.11	3.08	0.96	3.81	3.79	1.12
NONIIY	1.82	0.61	22.46	0.62	0.34	1.71	1.86	0.62	22.90
NONIXY	4.42	3.25	20.05	4.28	3.55	3.29	4.43	3.24	20.44
NOIJY	0.28	0.64	3.30	-2.85	-2.16	3.45	0.40	0.67	3.23
ROA	0.22	0.60	3.35	-3.15	-2.46	3.42	0.36	0.64	3.27
ROAPTX	0.33	0.78	4.66	-3.50	-2.83	3.60	0.49	0.82	4.63
ROE	2.43	5.53	19.86	-39.07	-29.70	49.13	4.10	5.86	15.55
ROEINJR	-2.26	1.76	17.31	-40.86	-31.49	48.47	-0.72	1.98	12.46
EEFFR	92.39	70.94	801.48	124.66	89.64	190.05	91.10	70.59	816.50
NPERFV	1.92	1.07	2.87	8.74	6.76	7.04	1.64	1.01	2.16
LNLSDEPR	1628.91	85.69	87376.92	90.94	90.50	18.32	1690.64	85.41	89112.79
EQV	11.66	9.90	7.66	6.91	7.10	2.99	11.85	10.01	7.73
RBC1AAJ	11.34	9.40	8.56	6.69	7.03	2.86	11.53	9.49	8.66
RBC1RWAJ	18.90	12.91	70.67	8.40	8.83	3.94	19.32	13.18	72.04
RBCRWAJ	20.00	14.03	70.63	9.75	10.18	4.06	20.41	14.27	71.99

Table 2. Descriptive statistics of the two groups (failed and non-failed) and the whole sample.

	Train (N=280; 140 failed; 140 non-failed)							Test (N=8013; 180 failed; 7833 non-failed)						
	Confusion matrix		*Accuracy*	*Type I error rate*	*Type II error rate*	*Precision of solvent*	*Precision of failed*	*F-score*	*Confusion matrix*		*Accuracy*	*Type I error rate*	*Type II error rate*	*Precision of solvent*	*Precision of failed*	*F-score*
Linear Discriminant Analysis LDA	134	6	91.79	12.14	4.29	88.74	95.35	92.10	7552	281	95.56	41.67	3.59	99.02	27.20	97.70
	17	123							75	105						
Logistic Regression LR	134	6	95.36	5.00	4.29	95.04	95.68	95.37	7535	298	95.42	38.33	3.80	99.09	27.14	97.62
	7	133							69	111						
Multilayer Perceptron MLP	132	8	93.93	6.43	5.71	93.62	94.24	93.95	7529	304	95.31	40.00	3.88	99.05	26.21	97.56
	9	131							72	108						
k-Nearest Neighbors KNN	130	10	93.57	5.71	7.14	94.20	92.96	93.53	7420	413	94.02	36.67	5.27	99.12	21.63	96.87
	8	132							66	114						
Naive Bayes NB	136	4	95.00	7.14	2.86	93.15	97.01	95.10	7416	417	94.03	33.89	5.32	99.18	22.20	96.88
	10	130							61	119						
Support Vector Machine SVM	132	8	90.71	12.86	5.71	88.00	93.85	91.03	7584	249	95.92	43.33	3.18	98.98	29.06	97.89
	18	122							78	102						
Boosting C4.5	140	0	100.00	0.00	0.00	100.00	100.00	100.00	7375	458	93.45	37.22	5.85	99.10	19.79	96.56
	0	140							67	113						
Bagging Random Tree BRT	139	1	99.64	0.00	0.71	100.00	99.29	99.64	7381	452	93.41	42.22	5.77	98.98	18.71	96.55
	0	140							76	104						
Partial Least Squares PLS-DA	131	9	91.07	11.43	6.43	89.12	93.23	91.29	7499	334	95.02	36.11	4.26	99.14	25.61	97.41
	16	124							65	115						

Confusion matrix	
a	b
c	d

True positive = a; True negative = d
Type I Error = b; Type II Error = c
Accuracy = (a+d)/(a+b+c+d)
Type I error rate = c/(c+d). Type II error rate = b/(a+b)
Precision of Solvent = a/(a+c); Precision of Failed = d/(b+d)
F-score = Harmonic mean of Precision of Solvent and (1- Type II error)

Table 3. Performance measures of the different techniques.

	Linear Discriminant Analysis LDA	*Logistic Regression LR*		*Multilayer Perceptron MLP*		*k-Nearest Neighbors KNN*		*Naive Bayes NB*		*Support Vector Machine SVM*		*Boosting C4.5*		*Bagging Random Tree BRT*		*Partial Least Squares PLS-DA*	
Linear Discriminant Analysis LDA		436	79	415	100	443	72	430	85	407	108	414	101	387	128	449	66
		112	7666	136	7642	226	7552	240	7538	74	7704	297	7481	310	7468	133	7645
Logistic Regression LR				474	74	460	88	455	93	399	149	465	83	445	103	435	113
				77	7668	209	7536	215	7530	82	7663	246	7499	252	7493	147	7598
Multilayer Perceptron MLP						462	89	451	100	395	156	471	80	456	95	432	119
						207	7535	219	7523	86	7656	240	7502	241	7501	150	7592
k-Nearest Neighbors KNN								476	193	433	236	475	194	435	262	434	235
								194	7430	48	7576	236	7388	234	7362	148	7476
Naive Bayes NB										398	272	512	158	483	187	449	221
										83	7540	199	7424	214	7409	133	7490
Support Vector Machine SVM												381	100	368	113	409	72
												330	7482	329	7483	173	7639
Boosting C4.5														551	160	433	278
														146	7436	149	7433
Bagging Random Tree BRT																419	278
																163	7433
Partial Least Squares PLS-DA																	

Contingency table	
w	x
y	z

w = number of banks classified as failed by both techniques
x = number of banks classified as non-failed by the above technique and as failed by the other technique
y = number of banks classified as failed by the above technique and as non-failed by the other technique
z = number of banks classified as non-failed by both techniques

Table 4. Crostabulation obtained from coincidences.

	Linear Discriminant Analysis LDA	*Logistic Regression LR*	*Multilayer Perceptron MLP*	*k-Nearest Neighbors KNN*	*Naive Bayes NB*	*Support Vector Machine SVM*	*Boosting C4.5*	*Bagging Random Tree BRT*	*Partial Least Squares PLS-DA*
Linear Discriminant Analysis LDA	-	0.81 [0.89]	0.76 [0.85]	0.74 [0.62]	0.71 [0.66]	0.81 [0.90]	0.66 [0.43]	0.62 [0.37]	0.81 [0.93]
Logistic Regression LR		-	0.85 [0.96]	0.74 [0.54]	0.73 [0.65]	0.76 [0.79]	0.72 [0.42]	0.70 [0.37]	0.75 [0.90]
Multilayer Perceptron MLP			-	0.74 [0.54]	0.72 [0.64]	0.75 [0.75]	0.73 [0.43]	0.71 [0.42]	0.75 [0.89]
k-Nearest Neighbors KNN				-	0.69 [0.65]	0.75 [0.66]	0.66 [0.51]	0.60 [0.50]	0.67 [0.62]
Naive Bayes NB					-	0.68 [0.71]	0.72 [0.62]	0.68 [0.62]	0.70 [0.66]
Support Vector Machine SVM						-	0.63 [0.46]	0.61 [0.40]	0.76 [0.86]
Boosting C4.5							-	0.76 [0.65]	0.65 [0.40]
Bagging Random Tree BRT								-	0.63 [0.36]
Partial Least Squares PLS-DA									-

Table 5. Phi correlation coefficients among techniques from 0-1 table. [Spearman correlation coefficients among techniques from scores]

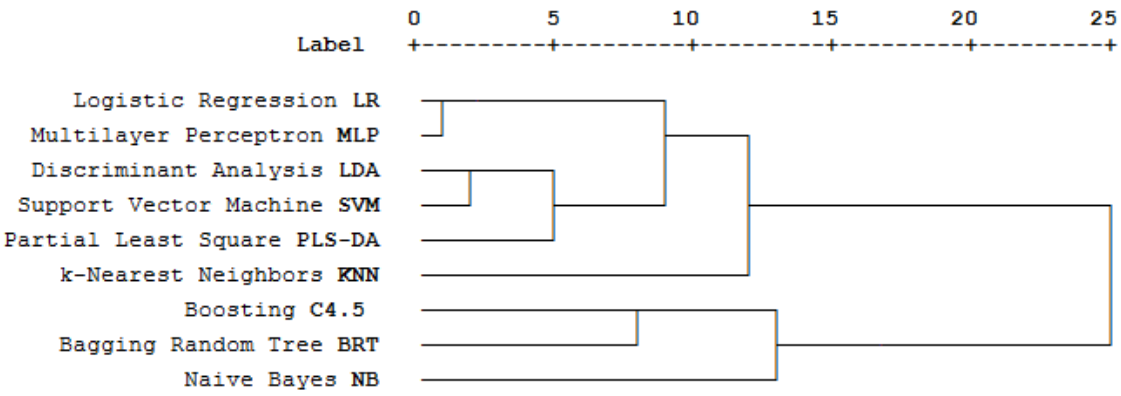

Figure 1. Cluster analysis from 0-1 table. Dendrogram using Ward Method

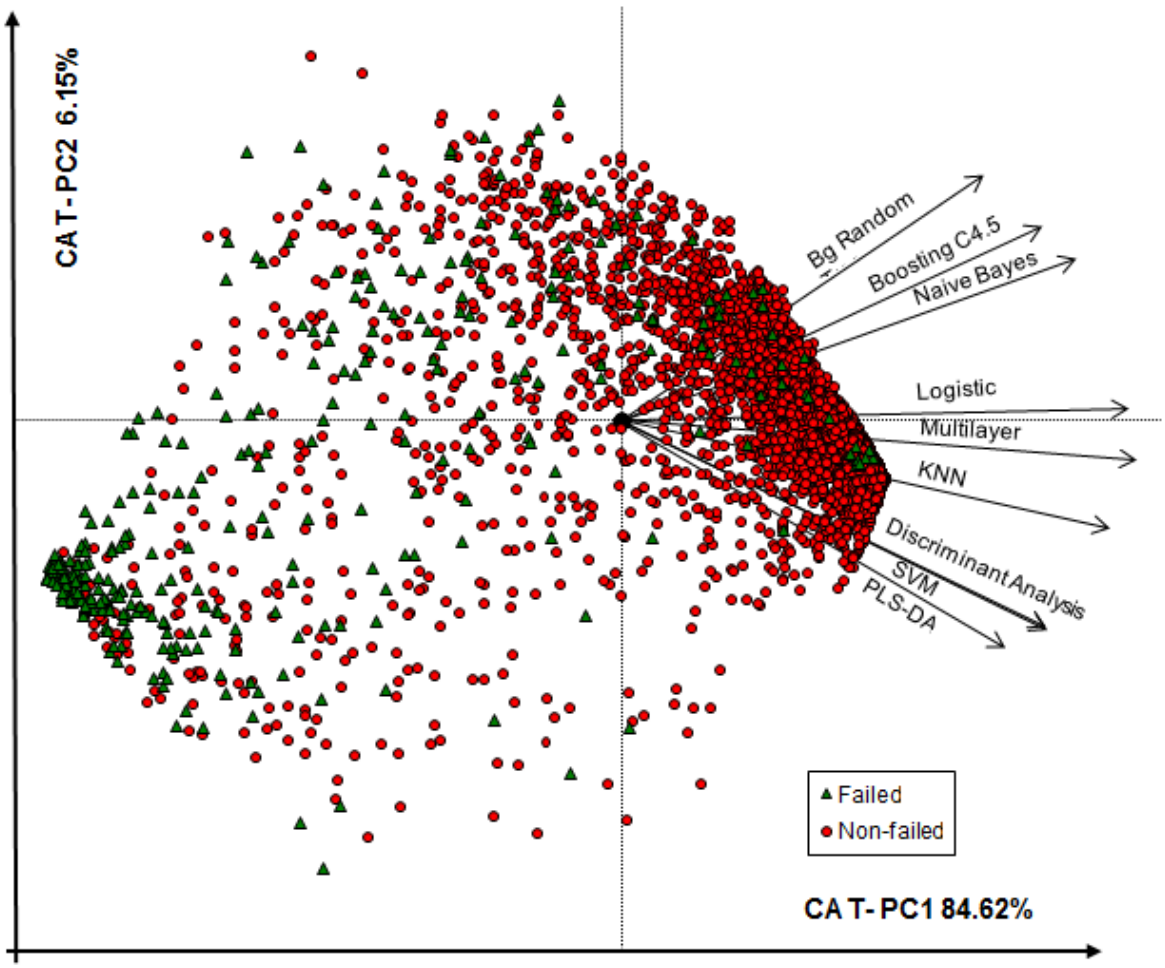

Figure 2. Component loadings from Categorical Principal Component Analysis. Obtained from the scores by the 9 techniques.