THE USE OF PCA IN REDUCTION OF CREDIT SCORING MODELING VARIABLES: EVIDENCE FROM GREEK BANKING SYSTEM

PANAGIOTA GIANNOULI, CHRISTOS E. KOUNTZAKIS

ABSTRACT. In this paper, we use the Principal Components Logistic Regression as a technique to reduce the variables being used in Credit Scoring Modeling. Specifically, we construct two models in which greek enterprises are classified, through their credit behavior and we evaluate them, relying on real data. In general, we propose a general way to use PC Regression, in case that we have high correlations and categorical variables in the sample.

1. INTRODUCTION: MOTIVATION FOR THE USE OF PCA -LOGISTIC REGRESSION IN CREDIT SCORING MODELING

1.1. **Summary for what is proposed and why.** The existence of high correlations between the variables being used in Credit Scoring Models (CSM), both with the use of categorical variables, which are the columns in a credit rating database, may lead to the use of principal components for the reduction of the variables being chosen for the final model. The Principal Components' vectors are linearly independent, hence we may select which of them may enter a model of Credit Rating. For this reason, the discrimination between the good and the bad credit behavior of an enterprise is made through Logistic Regression (LR), since this is the more effective and widely known in credit scoring industry. The reduction of dimensions is however a topic of interest under the frame of 'Big Data Analysis', where the question is which of the variables in a large database of credit rating variables are really significant. In logistic regression, we use a logit function:

$$P(1) = \frac{e^y}{1 + e^y}, P(0) = \frac{1}{1 + e^y},$$

where y is a linear combination consisted by some of the initial variables, after choosing the appropriate number of principal components. 0 in the samples below denotes the good credit behavior, while 1 denotes the bad credit behavior, respectively. Since the Credit Scoring Models are usually tested on samples being simulated in labs, but are not tested on real data, For we test the proposed PCA reduction algorithm, in two cases: the first sample contains 40 Variables and 1889 enterprises, and the second sample contains 53 Variables and 2690 enterprises. The first sample is the sample of the small enterprises and the second one is the sample of the great enterprises, while this classification was made according to the annual revenues of them, as it is explained below. The use of financial ratios in such models, which actually may come from accounting practice, is also an idea appearing in this paper, coming from the seminal work [2], which also appears in the present paper. Below, the symbol / dentes division, and 9 from the 53 initial variables are the more semantic in

the PCA-LR Model concerning the Great Enterprises' Credit Behavior, which is presented in the final paragraph. These variables are the following:

(1) $C1$: Quick ratio is an indicator of a company's short-term liquidity, and measures a company's ability to meet its short-term obligations with its most liquid assets. Because we're only concerned with the most liquid assets, the ratio excludes inventories from current assets. Quick ratio is calculated as follows: Quick ratio = (current assets and inventories) / current liabilities, or Quick ratio = (cash and equivalents + marketable securities + accounts receivable) / current liabilities

(2) $C5$: Total liabilities/ Total Assets.

(3) $C20$: Current Ratio =Current Assets / Current Liabilities. The current ratio is called current because it incorporates all current assets and liabilities.

(4) $C22$: Working capital turnover ratio=Working capital turnover is a measurement comparing the depletion of working capital used to fund operations and purchase inventory, which is then converted into sales revenue for the company. The working capital turnover ratio is used to analyze the relationship between the money that funds operations and the sales generated from these operations.

(5) $C23$: Net working capital/total assets.

(6) $C35$: Short-term liabilities.

(7) $C41$: Maximum revolving loans=Maximum Percent Credit Utilization -Payments to Primal/Joint Lenders - Revolving-SME - Updated in Last 12 Months

(8) $C42$: Worst Payment Status of Loans last month/ Worst Payment Status of Loans last 24 months.

(9) $C44$: Worst Payment Status of Loans last 3 months=
Worst Payment Status - SME - Payments to Primal/Joint Lenders During Last 3 Months

On the other hand, the reduction of the Variables is needed in order to keep a number of Variables which are more significant and keeping the Credit Scoring Model as informative as the Model including the total number of Variables allows, or the specific use of Pearson's χ^2 in LR Models, we refer to the paper [1]. We notice that among the Credit Behavior semantic variables, a set of pure accounting variables and ratios is included as it was expected. The significant variables in the model do provide a positive relation between both of the current liquidity and the short-term liabilities of the great Greek enterprises and the good credit characterization of them in the period of the sample selection, which it was a sub-period period of the Greek Sovereign Debt Crisis. In this paper, the period of 12 months (01/01/2014 to 31/12/2014) was a performance period and 24 months (01/01/2012 to 31/12/2013) as an observation period, as it often occurs in creating similar models. Specifically, while in the accepted model in the accepted model concerning great enterprises all the enterprises are GOOD in the observation period, in the performance period we notice that they are separated into GOOD and BAD, which is directly related to Debt Crisis. In order to understand that these variables are common in order to build credit scoring scoring models, we also refer to the explanation of the variables that seem to be more important for the credit behavior of the small enterprises:

(1) $C12$: Net Profit Margin-The ratio of the net profits to revenues for a company or business segment.

(2) $C13$: Pretax Return on Equity- The amount of net income returned as a precentage of shareholders equity.

(3) $C32$: Income Tax
(4) $C34$: Maximum percent credit utilization : Payments to Primal/ Joint Lenders-Non Revolving SME updated in last 12 months
(5) $C37$: Maximum percent credit utilization : Payments to Primal/ Joint Lenders-Revolving SME updated in last 12 months

From a financial point of view, the fact that these variables were selected in order to describe a good credit behavior of the small enterprises during a sub-period of the crisis, implies that the Greek Financial System as a whole, has a impressive stability since the weights of the variables in this equations are positive. In the first Appendix, we provide the Algorithm of PCA-LR in a condensed form. In the second Appendix, we provide some performance measures that assure that the accepted LR model conserning the credit behavior of the great enterprises is well-fitted on the data of the performance period, which is also a sub-period of the greek credit crisis. We remind that the performance period for this model, as it is determined below, is between 01/01/2014 and 31/12/2014. The performance measures being used is the Kolmogorov -Smirnov and the Gini Index. Through Gini Index for the model of the great enterprises, we conclude that the specific model being accepted is a good discriminator of the good and the bad behavior in the performance period.

1.2. **Review of the literature.** The Altman Model, which introduces the use of Discriminant Analysis is a specific answer to another seminal paper for the Credit Rating Modeling as a subject of interest in Finance and Banking Science, which is [6]. This excuses the presence of such a set of variables in real-data models, and the presence of them in the databases we examine below. A review of the problems in the application of Discriminant Analysis in Credit Scoring Models appear in the paper [7] and they refer to the violation of the assumption about the underlying distributions of the variables, the use of linear discriminant functions instead of quadratic functions when the group dispersions are unequal, the improper interpretation of the role of individual variables in the analysis, reductions in dimensionality, problems in the definition of the groups, use of inappropriate a priori probabilities and/or costs of misclassification, problems in the estimation of classification error rates to assess the performance of the model. By the present paper, we establish the definition of the 'bad' and the 'good' credit behavior and we contribute to the problem of the reduction of the variables. We insist on using Logistic Regression (LR), as a general methodology for Credit Scoring Model fitting, because it gives a prompt answer about the fitting of a CSM model, including specific Variables. Moreover, Logistic Regression provides a direct estimation of the 'probability of default' both for an enterprise and for the whole Finance System, as well. This preference of us related to LR on the accuracy of the fitted Credit Scoring Model and the predictive probability of being 'bad', is something which is a research subject for a long time in the topic of CSM, though a lot of alternative ways are present in Credit Rating Modeling. For example, an alternative way to the problems of Discriminant Analysis, an alternative way of separation of the groups of 'bad' and 'good', appears instead of Discriminant Analysis, in [8]. A recent paper, in which Neural Networks (NN) are compared to linear regression Credit Rating Modeling if the distribution of the dependent variable is 'skew', is [9]. Another paper which refers to the predictive ability of Neural Networks in CSM, is [4]. Another paper comparing NN and Logistic Regression is [13]. Sometimes, like in cases described in [10], the predictive power of LR comparing to this of the Neural Networks, relies on specific characteristics of subgroups existing in the same sample. If we would like to refer to a paper for the use of financial variables alike the ones

which are included in the model which describes the credit behavior of the great enterprises (such as $C1, C5, C20, C35$) for credit risk modelling appears in recent bibliography, this is [5]. Also, recent papers concerning the robustness and the predictive power between different statistical techniques used in prediction purposes and classification problems in credit scoring are [6], [11].

1.3. **The definition of good and the bad credit behavior for an enterpise.** The gradual development of financial risk research, leads to the need for a high level of CSM, in order to forecast this kind of Credit Risk. The principal aim of this paper is to develop credit risk models for the Greek Financial System, concerning small and big companies (according to their revenues) by using a combination of financial data and credit behavior data. Credit behavior data was taken from three reliable inter-bank systems (RCS, DFO and MPS) developed by Tiresias S.A. (an independent authority founded by almost all banks in Greece and its resposiblity is Credit Risk Rating and Monitoring). Credit Consolidation System (RCS) contains corporate and personal loans and credit cards and its purpose is credit risk assessment. Default Financial Obligation System (DFO) contains bounced checks, 'protested' collateral bills, denounced contracts, court derogatory data, etc. and its purpose is the assessment of solvency. MPS is a system which contains mortgages and prenotations and its purpose is the liens on assets. The data sources of Tiresias S.A. are banks and financial institutions, courts of first instance, credit companies, funding companies, leasing and card managing companies. This fact indicates that the models presented below, are tested on real data coming from the Greek banking system. This is important, because it indicates which variables are included as interpretive at times when data on the banking system and business changes rapidly, as it happens in cases of crises, hence stability is an important factor from the aspect of useful of such a model in practice. In this paper, the period of 12 months (01/01/2014 to 31/12/2014) was a performance period and 24 months (01/01/2012 to 31/12/2013) as an observation period, as it often occurs in creating similar models (see for example [12]). These models are intended to discriminate the bad from good behavior in the performance period. First of all, we have to explain the terms of 'bad' and 'good' credit behavior for an enterprise:

    (i) An enterprise is classified in the set of the ones having good credit behavior, ($y = 0$) if it belongs to the set of the enterprises with no delinquency or it belongs to the enterprises with maximum delinquency in the last 12 months from 0 to 29 days past due either to the credit limit utilization over 102 per cent from 0 to 29 days, including SME Overdrafts.

    ((ii)) An enterprise is classified in the set of ones having bad credit behavior, ($y = 1$, if it is an enterprise showing severe delinquency, which denotes:

        (i) they own SME Contracts, not Overdrafts with maximum delinquency in the last 12 months greater or equal to 90 days past

        (ii) they SME Overdrafts with maximum delinquency in the last 12 months, greater or equal to 90 days past due either to credit limit utilization over 102 per cent for time period greater or equal to 90 days with over limit amount greater than 100

    (iii) In case where there is some Guarantor for the enterprise, this enterprise is classified in the set of the ones having bad credit behavior, in the following cases:

        ((i)) totally owned SME Contracts, not Overdrafts with maximum delinquency in the last 12 months greater or equal to 150 days past

(ii) totally SME Overdrafts with maximum delinquency in the last 12 months greater or equal to 150 days past or credit limit utilization over 102 for time period greater or equal to 90 days.

Also, a company is included in the ones with bad credit behavior when there is a new DFO (loan denunciation), within performance period. The term 'utilization' is the following Financial Ratio: Current Balance of the Enterprise/Credit Limit of it. This information and data obtained from the Web of www.tiresias.gr. Also, small companies are those whose annual revenues are less than 700.000 Euros and big companies are those whose annual revenues are greater than 700.000 Euros. From this definition of bad and good credit behavior, we may understand that a new entry in the finally fitted model relies mainly on accounting variables, which are related to delinquency of it.

2. Logistic Regression in Practice and PCA

We show below the steps being followed for the use of PCA jointly with Logistic Regression algorithm, accompanied by the appropriate comments:

(1) The optimal number of the PC finally chosen to enter in the model of Principal Components Logistic Regression is considered by comparing the value of R^2_{adj} between the model including all the variables (the so-called 'total' model) and the R^2_{adj} including these PC. The PC included in the test of the Algorithm are m, where m is defined in the Appendix, in order to achieve either a satisfactory level of variable reduction, or to abandon the application of the algorithm.

(2) The same comparison is the one that we have to follow between the value of AIC on the 'total' model and the value of AIC on the Principal Components finally chosen to enter the model of Logistic Regression. If at least one of R^2_{adj} and AIC of an increased number of Principal Components are much less than the equivalent values of these statistics calculated on the total model, we abandon the use of PCA. If both of these statistics are close to the values of the total model for a specific number of principal components, which is a threshold in order to apply the dimensional reduction, we apply the next steps.

(3) Since we do not use principal components in practice but some of the initial variables, and since the principal components are linear combinations of the initial variables, we go back to the chosen principal components, and we have to choose which of initial variables seem to be more significant than the others.

(4) For this purpose, we replace the initial variables in the equation of Logistic Regression under the Principal Components, these components correspond to some of the initial variables, because each of the principal component is actually a linear combination of them. Hence, by replacing the linear equation of the Logistic Regression, by each of this 'inverse' equations, indicates which of the initial variables are significant in a new Logistic Regression Model. These ones are those which have the greatest absolute value in the expansion of the linear equation with the principal components.

(5) After the selection of these initial variables, we create a new Logistic Regression Model containing them.

(6) The rejection or the approvement of the candidate final PCR model (including the initial variables we decided to incorporate by the greatest absolute value in the expansion of the linear combination of the principal components), is tested by the value of the fraction χ^2/Df, where by χ^2 we denote the Pearson's one. If this value is greater

than p-value, then the model is statistically approved, if this model's R^2_{adj} and AIC are close to the ones of the total model's. Otherwise, it is rejected.

(7) For the performance of the model, we specify a period in which we collect the sample, and a period in which we observe the fitted model.

In the first Appendix, we show the diagram of the above algorithm and we also quote on its application.

3. Application of PCA-LR on Real Data obtained from Greek Enterprises

We apply the Steps of the Algorithm described above on the two samples described in Introduction (The Data Analysis is was made for both of the samples on Minitab 17).

3.1. **Small enterprises:** The R^2_{adj} of the total model of Logistic Regression is 45,02, while the Value of AIC for this model is 1327,47. Also, the value of R^2_{adj} for the model having 5 Principal Components is 42,61, while the value of AIC for this model is 1350,00. The fact that the AIC and the R^2_{adj} of the model containing the 5 Principal Components and the total model are close, implies that the number of Principal Components that we have to choose is 5. Model Equation for the PCR:

$$y = -1,590 - 0,4707W_1 + 0,3628W_2 + 0,4123W_3 + 0,2294W_4 + 0,5308W_5, \ P(1) = \frac{e^y}{1+e^y},$$

where the W_j denote the first five components, $j = 1,2,3,4,5$. The probability for some of the enterprises of this sample to be classified as 'good' is estimated- without the error term - by $P(0) = \frac{1}{1+e^y}$. After some calculations, we conclude that the above PCR Model for the small enterprises, implies that the initial variables having the higher absolute weight in the above expansion of the five principal components are

$$C12, C13, C32, C34, C37.$$

Hence, we go on with testing the fitting of the Logistic Regression, on these 9 selected variables. For the fitting of this model, $R^2_{adj} = 18,54$ and AIC=1929,64, which implies that the selection of these variables is not satisfactory, since these values are far from the values for both of R^2_{adj} and AIC, either of the total model or for the model with the 5 principal components. The value of the (Pearson's) $\chi^2/\text{Df}=1895,05/1879 > 0,393$, hence we could accept the specific model, but due to the high AIC and the low R^2_{adj}, the model is rather rejected.

3.2. **Great enterprises:** The R^2_{adj} of the total model of Logistic Regression is 47,65, while the Value of AIC for this model is 467,22. Also, the value of R^2_{adj} for the model having 6 Principal Components is 43,51, while the value of AIC for this model is 457,10, which is due to the high correlations of the variables. The fact that the AIC and the R^2_{adj} of the model containing the 6 Principal Components and the total model are close, implies that the number of Principal Components that we have to choose is 6. Model Equation for the PCR:

$$y = -1,736 - 0,4592W_1 - 0,1938W_2 - 0,2695W_3 - 0,5449W_4 + 0,4283W_5 + 0,2056W_6,$$

$$P(1) = \frac{e^y}{1+e^y},$$

where the W_j denote the first six components, $j = 1,2,3,4,5,6$, The probability for some of the enterprises of this sample to be classified as 'good' is estimated- without the error term

- by $P(0) = \frac{1}{1+e^y}$. After some calculations, we conclude that the above PCR Model for the great enterprises, implies that the initial variables having the higher absolute weight in the above expansion of the six principal components are

$$C1, C5, C20, C22, C23, C35, C41, C42, C44.$$

Hence, we go on with testing the fitting of the Logistic Regression, on these 9 selected variables. For the fitting of this model, $R_{adj}^2 = 41,05$ and AIC=479,69, which implies that the selection of the variables is satisfactory. Also, since the value of (Pearson's) χ^2 / Df=0,94720 > p-value =0,834, we may accept the specific model. The Model Equation for the final model is:

$$y = -1,265 - 0,00144C1 + 0,00161C5 + 0,00218C20 + 0,00297C22 - 0,00636C23$$

$$-0,00718C35 - 0,008756C41 - 0,00238C42 - 0,00537C44,$$

where the description for the initial variables entered into the final model of LR, is explained in the Introduction.

4. Conclusion

From a statistical point of view, we may say that if we have a great number of jointly high- correlated and categorical variables, PCA -Logistic Regression is a methodology in the way that we describe in the second section, is a way to reduce the variables and keeping the ones that we need, under a rational loss of information. On the other hand, we use Logistic Regression, because is an effective method, and widely known in financial industry. The combination of these statistical tools, lead to the use of PCA-LR Algorithm, which is analyzed in the Appendix From a financial point of view, the use of PCR and the consequent variable reduction, leads to a more efficient design of credit scoring models, either concerning small, or concerning great enterprises. This happens, because we may know the 'risk profile' of them, under real data, which arise by selection and collective processing of them by the whole of Greek Financial System.

5. Appendix -Presenting and Quoting the PCA-LR Algorithm

Here is a concentrated form of the algorithm:

(i) If d denotes the number of the variables of the design matrix X, then for $k = 1, ..., m$, Until $m = \frac{d}{2}$ if d is even, or $m = \frac{d+1}{2}$ if d is odd, repeat the following steps:

(ii) complete the matrix of results:

	Steps		Results	
		Total Model LR		LR Model with k PC
	Compare	R_{adj}^2		R_{adj}^2
	Compare	AIC		AIC

(iii) If there exists some k_1, such that the values of both criteria are close, then we go on to the next step

(iv) The LR having k_1 PC is a model, which finally is a linear combination of all the initial variables. The ones which finally are selected to enter the model, are the ones which have the greater absolute weight in this linear combination.

(v) For the LR including these initial variables we specified, we calculate the Pearson's χ^2 Goodness of Fit for LR: χ^2/Df. If its value is greater than the p-value of the model, then this model is accepted.

The memory needed in case of the application of the above algorithm is $O(m)$, because 2 calculations are needed for the Total Model in the matrices at the step (ii). In the same matrix we may store the closer results, till we should find some other result more closer. Also, for the selection of the variables at the step (iv), we need approximately m memory positions.

6. Appendix -Test of the Performance for the Model of the Great Enterprises

We separate the scores of the linear part of the model being accepted on the sample of the perfromance period. The results created values of the y, which may be classified in the classes appearing in the following matrix:

$$\begin{bmatrix} 3 & 241 & 244 & 1,2 & 11,3 & - \\ 10 & 239 & 249 & 4,0 & 21,4 & 0,01 \\ 12 & 247 & 259 & 4,6 & 31,6 & 0,01 \\ 28 & 444 & 472 & 5,9 & 48,9 & 0,04 \\ 50 & 438 & 488 & 10,2 & 62,5 & 0,11 \\ 67 & 178 & 245 & 27,3 & 60,8 & 0,17 \\ 87 & 158 & 245 & 35,5 & 54,9 & 0,25 \\ 168 & 77 & 245 & 68,6 & 32,5 & 0,51 \\ 217 & 26 & 243 & 89,3 & 0,0 & 0,67 \\ 642 & 2048 & 2690 & 23,9 & 62,5 & 0,76 \end{bmatrix}$$

The values of the above matrix refer to the classes of the linear part of the accepted model for the great enterprises, at the performance period. The second column refers to the enterprises which are characterized 'BAD' through this model and the third column refers to the enterprises which are characterized 'GOOD' by this model. The third column presents the sum of GOOD and BAD, which belong to the same score class group. The 4th column is the percentage of BAD RATE, the 5th column corresponds to the (K-S) test value for each class, and the 6th column is the Gini Index, which arise from each of the specific classes (scores) of the model at the performance period. The intervals of the scores, or else the values of the linear part appearing at the equation 3.2, which correspond to the lines of the above matrix are the following: The first interval contains enterprises, whose score is less than $-3,42$. The second interval contains enterprises, whose score is between $-3,41$ and $-−3,07$. The third interval contains enterprises, whose score is between $-3,06$ and $-−2,85$. The 4th interval contains enterprises, whose score is between $-2,84$ and $-−2,33$. The 5th interval contains enterprises, whose score is between $-2,32$ and $-1,58$. The 6th interval contains enterprises, whose score is between $-1,57$ and $-−0,93$. The 7th interval contains enterprises, whose score is between $-0,92$ and $-0,12$. The 8th score interval contains enterprises, whose score is between $-0,11$ and $0,85$. The last score interval contains the enterprises, whose score is $\geq 0,86$. The first column contains the enterprises, which are GOOD at the observation period, with respect to the same model, while with respect to the same model are BAD in the performance period. At the last line we show their sum in thw whole sample. The second column contains the enterprises, which were GOOD at the observation period, while at the performance period are GOOD, with respect to the same model. At the last line,

we see the sum of these enterprises at the whole sample. At the third column, we see the sum of GOOD and BAD, with respect to their score at the performance period. At the 4th column, we see the percentage of BAD enterprises in any of the score interval, at the perfomance period. The total percentage of BAD enterprises in the whole sample appears at the corresponding element of the 4th column. The maximum deviation between GOOD and BAD, relying on the specific sample is 62,5 per cent, which implies a satisfactory distinction accuracy via the PCA-LR model we proposed. Finally, the Gini index is equal to 0, 76. We compute Kolmogorov-Smirnov test (K-S) and Gini Index were used as performance indexes, in order to test models performance (our calculations are made here at the 'performance' period mentioned above). These indexes are used in order to verify if the model is capable to distinguish the two populations. A K-S value of zero would indicate that the model is unable to make any distinction between two populations, while a K-S score of 100 would indicate that the model is capable of perfect distinction between two populations. It is a way of verifying the chosen model as a K-S value of zero would indicate that the model is unable to make any distinction between two populations, while a K-S score of 100 would indicate that the model is capable of perfect distinction between two populations. The 62,5 per cent is the maximum deviation that the BAD enterprises seem to have from the GOOD enterprises, in this model. Also, the Area Under the ROC Curve (AUC), which shows the accuracy of the model, can be calculated by the Gini Index value as Gini Index=2*AUC-1. In this case, AUC is 0,88 (which is also implies that the equivalent value of the Gini Index=0,76), which implies that the specific model is 'fair'. Below, we may see by the traditional academic point system (a rough guide for classifying the accuracy of a diagnostic test):

(1) 0,90-1=excellent
(2) 0,80-0,90=good
(3) 0,70-0,80=fair
(4) 0,60-0,70=poor
(5) 0,50-0,60=fail

References

[1] ALISON, P.D. Measures of Fit for Logistic Regression. *SAS Global Forum* Paper 1485   (2014)

[2] ALTMAN, E.I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, **23** 589–609 (1968)

[3] ALTMAN, E.I. An emerging market credit scoring system for corporate bonds. *Emerging Market Review*, **6** 311–323 (2005)

[4] BORITZ, J.E., KENNEDY, D.A. Effectiveness of Neural Network Types for Prediction of Business Failure. *Expert Systems with Applications*, **9** 503–512 (1995)

[5] BOGUSLAUSKAS, V., MILERIS, R., ADLYTÊ, R. The Selection of Financial Ratios as Independent Variables for Credit Risk Assessment. *Economics and Management*, **16** 1032–1038 (2011)

[6] DURAND, D. Risk Elements in Consumer Installment Financing National Bureau of Economy Research, New York, pp. 189–201 (1941)

[7] EISENBEIS, R. Problems in applying discriminant analysis in credit scoring models. *Journal of Banking and Finance*, **2** 205–219 (1978)

[8] HARDY, W.E.Jr., ADRIAN, J.L. A linear programming alternative to discriminant analysis in credit scoring. *Agribusiness*, **1** 285–292 (1985)

[9] KUMAR, A.U. Comparison of neural networks and regression analysis: A new insight. *Expert Systems with Applications*, **29** (2): 424–430 (2005)

[10] LEE, T. H., JUNG, S. C. Forecasting creditworthiness: Logistic vs. artificial neural net. *The Journal of Business Forecasting Methods and Systems*, **18** (4): 28–30 (2000)

[11] PALEOLOGO, G., ELISSEEFF, A., ANTONINI, G. Subagging for credit scoring models. *European Journal of Operational Research* **201**, 490–499 (2010)

[12] SIDDIQI, N. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring.* New Jersey: John Wiley and Sons, Inc. (2006)

[13] TAM, K. Y., KIANG, M. Y. Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, **38** (7): 926–947 (1992)

DEPT. OF MATHEMATICS, DIVISION OF STATISTICS AND ACTUARIAL-FINANCIAL MATHEMATICS, UNIVERSITY OF THE AEGEAN, KARLOVASSI, GR-83 200 SAMOS, GREECE

E-mail address: giannouli@aegean.gr   chr_koun@aegean.gr .