A New Classifier based on The Reference Point Method with Application in Bankruptcy Prediction

Jamal Ouenniche (Jamal.Ouenniche@ed.ac.uk)
University of Edinburgh, Business School
29 Buccleuch Place, Edinburgh EH8 9JS, United Kingdom

Kais Bouslah (kbhb@st-andrews.ac.uk)
University of St Andrews, School of Management,
Gateway Building, North Haugh, St Andrews KY16 9RJ, United Kingdom

Jose Manuel Cabello (jmcabello@uma.es)
University of Malaga, Department of Applied Economics (Mathematics),
Calle Ejido 6, 29071 Malaga, Spain

Francisco Ruiz (rua@uma.es)
University of Malaga, Department of Applied Economics (Mathematics),
Calle Ejido 6, 29071 Malaga, Spain

Abstract: The finance industry relies heavily on the risk modelling and analysis toolbox to assess the risk profiles of entities such as individual and corporate borrowers and investment vehicles. Such toolbox includes a variety of parametric and non-parametric methods for predicting risk class belonging. In this paper, we expand such toolbox by proposing an integrated framework for implementing a full classification analysis based on a reference point method; namely, in-sample classification and out-of-sample classification. The empirical performance of the proposed reference point method-based classifier is tested on a UK dataset of bankrupt and non-bankrupt firms. Our findings conclude that the proposed classifier can deliver a very high predictive performance, which makes it a real contender in industry applications in banking and investment. Three main features of the proposed classifier drive its outstanding performance; namely, its non-parametric nature, the design of our RPM score-based cut-off point procedure for in-sample classification, and the choice of a k-Nearest Neighbour as an out-of-sample classifier which is trained on the in-sample classification provided by the reference point method based classifier.

Keywords: In-Sample Prediction, Out-of-Sample Prediction, Reference Point Method Classifier, k-Nearest Neighbour Classifier, Bankruptcy, Risk Class Prediction

1. Introduction

Decision making under multiple and often conflicting criteria or objectives is common in a variety of real-life settings or applications. Formally, these multi-criteria problems are classified into several categories; namely, selection problems, ranking problems, sorting problems, classification problems, clustering problems, and description problems. A variety of multi-criteria decision-aid (MCDA) methodologies have been designed to address each of these problems. Ranking problems have received considerable attention in both academia and industry. One popular class of ranking methods consists of the so-called reference point methods (RPMs). To the best of our knowledge, there are no classifiers based on RPMs. In this paper, we extend the risk analytics toolbox by proposing a first RPM-based classifier and test its performance in risk class prediction with application in bankruptcy prediction.

Conceptually, a reference point method makes use of one or several reference points or benchmarks to assess the relative performance of a set of entities, where the performance of each entity is measured by an index that aggregates the individual distances between the entity and the reference point(s) with respect to multiple dimensions, criteria, or objectives. Such indices are typically used to rank entities from best to worse depending on how close or far they are from the benchmarks or reference points. RPMs are a family of methods for solving multi-objective optimization problems and thus belong to the multi-objective programming toolbox. By design, RPMs make use of aggregate distance functions or indexes commonly referred to as scalarizing functions (e.g., Miettinen, 1999, Steuer, 1986), achievement functions (e.g., Benayoun et al., 1971; Romero, 2001, 2004; Rodríguez-Uría et al., 2002), or achievement scalarizing functions (e.g., Wierzbicki, 1980; Steuer and Choo, 1983; Lewandowski and Wierzbicki, 1989; Buchanan, 1997; Nakayama and Sawaragi, 1984; Ogryczak and Lahoda, 1992; Miettinen and Mäkelä, 1995; Wierzbicki et al., 2000). Regardless of how these functions are called, they all model deviations from the reference point(s). The main differences however between these functions lie in how entities are rewarded or penalised for being close to or far from the reference point(s). As pointed out by Wierzbicki (1979), these functions can be viewed as ad hoc approximations of the decision maker utility function. A comparative study of different achievement scalarizing functions can be found in Miettinen and Mäkelä (2002), while a survey of the different weighting schema used in achievement scalarizing functions is available in Ruiz et al. (2008). On the other hand, the connection between the scalarizing functions of goal programming and reference point schema have been studied in Ogryczak (2001). Some of the above-mentioned functions have been used in a variety of applications. Examples include paper industry (Diaz-Balteiro et al., 2011), tourism industry (Blancas et al., 2010), finance industry (Cabello et al., 2014), sustainability of municipalities (Ruiz et al., 2011), wood manufacturing industry (Voces et al., 2012), computer networks (Granat and Guerriero, 2003), procurement auctions (Kozlowski and Ogryczak, 2011), and inventory management (Ogryczak et al., 2013).

The remainder of this paper unfolds as follows. In section 2, we provide a detailed description of the proposed integrated in-sample and out-of-sample framework for RPM classifiers and discuss implementation decisions. In section 3, we empirically test the performance of the proposed framework in bankruptcy prediction of companies listed on the London Stock Exchange (LSE) and report on our findings. Finally, section 4 concludes the paper.

2. An Integrated Framework for Designing and Implementing Reference Point Method-based Classifiers

Nowadays, prediction models – whether designed for predicting a continuous variable (e.g., the level or volatility of the price of a strategic commodity such as crude oil) or a discrete one (e.g., risk class belonging of companies listed on a stock exchange) – have to be implemented both in-sample and out-of-sample to assess their ability to reproduce or forecast the response variable in the training sample and to forecast the response variable in the test sample, respectively. In principle, a properly designed prediction model fitted to in-sample data should be able to reproduce or predict the response variable with a high level of accuracy. However, in real-life settings, in-sample performance is not enough to qualify a prediction model for actual use in predicting the future. Because the future is unknown, out-of-sample implementation and evaluation frameworks are required to simulate the future. Therefore, an integrated framework for implementing a full classification analysis based on a reference point method; namely, in-sample classification and out-of-sample classification, is proposed in this paper.

Hereafter, we shall present our integrated RPM-based classification framework – see Figure 1 for a graphical representation of the process.

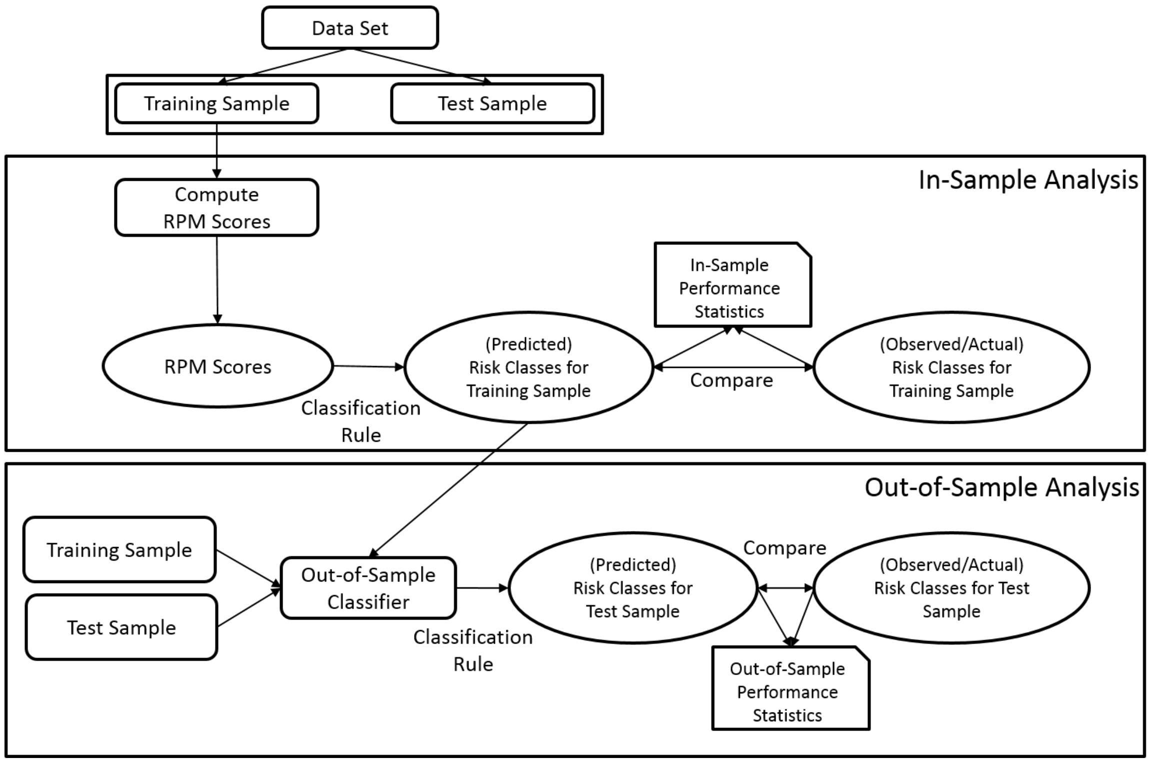

Figure 1: Generic Design of In-Sample and Out-of-Sample Analyses of RPM Classifiers

For illustration purposes, we shall customize the presentation of the proposed framework to a bankruptcy application where we reproduce or rework a classical bankruptcy prediction model; namely, the multivariate discriminant analysis (MDA) model of Taffler (1984), within a RPM classifier framework.

Input: A set of n entities (e.g., LSE listed firm-year observations) to be assessed on m pre-specified criteria (e.g., financial criteria) along with their measures (e.g., financial ratios), where the performance criteria could belong to one or several, say nc, categories (e.g., liquidity, asset utilization, operating performance, cash flow, capital structure and solvency, return on investment, market performance), and the measure of each criterion could either be minimized or maximized. Thus, each entity, say i $(i = 1, \dots, n)$, is represented by an m-dimensional vector of (observed) measures of the criteria under consideration, say $x_i = (x_{ijk})$, where x_{ijk} denotes the observed measure of criterion j of category k for entity i. Let the set of x_is be denoted by X and M^- (respectively, M^+) denote the set of measures for which lower (respectively, higher) values are better. An observed risk (e.g., distress, bankruptcy) status or risk class belonging, say Y, is also available for all entities. The "historical" sample X is divided into a training sample, say X^E, and a test sample, say X^T. Let training (respectively, test) sample observations be denoted by $x_{ijk}^{E}, i = 1, \dots, \#X^{E},\ k = 1, \dots, nc, j = 1, \dots, \#C_k$ (respectively, $x_{ijk}^{T}, i = 1, \dots, \#X^{T},\ k = 1, \dots, nc, j = 1, \dots, \#C_k$), where $\#X^{E}$ (respectively, $\#X^{T}$) denote the cardinality of the training (respectively, test) sample and $\#C_k$ denotes the cardinality of the kth category of performance criteria C_k.

Phase 1: In-Sample Analysis

Step 1: Choice and Computation of Double Reference Points or Benchmarks

Choose the approach and method according to which double reference points (reservation, aspiration) or virtual benchmarks, say r^{-} and r^{+}, are determined and compute them for each criterion or measure j ($j = 1, \dots, m$). There are several approaches to setting up these reference points; namely, the neutral scheme where, for example, all aspiration and reservation levels are chosen as percentages of criteria ranges; the statistical scheme is based on some statistical measures of each criterion; and the voting scheme, which requires input from a group of decision makers, experts, or stakeholders – for details on these schema, see Wierzbicki et al (2000). In our risk application, we have chosen the so-called statistical scheme to show the merit of the proposed classifier empirically. Note that the use of neutral or statistical reference points results in a final comparative measure, whereas the use of the voting scheme results in an absolute measure. When the size of the training set is large enough, it seems reasonable to assume that a comparative measure is enough, and we do not need to add subjective reference points to the model. Within the statistical scheme, the best, the worst, and the average observed measure of each criterion j ($j = 1, \dots, m$), say r_j^{best}, r_j^{worst} and $r_j^{average}$ respectively, are first computed. Then, the double reference points (reservation, aspiration), r^{-} and r^{+}, are computed as follows:

$$r_j^{-} = r_j^{average} - \frac{\left(r_j^{average} - r_j^{worst}\right)}{2}; j = 1, \dots, m$$

$$r_j^{+} = r_j^{average} + \frac{\left(r_j^{best} - r_j^{average}\right)}{2}; j = 1, \dots, m$$

where

$$r_j^{best} = \begin{cases} \min\limits_{i=1,\dots,\#X^E} x_{ijk}^{E}\ IF\ j \in M^{-} \\ \max\limits_{i=1,\dots,\#X^E} x_{ijk}^{E}\ \ IF\ j \in M^{+} \end{cases}; j = 1, \dots, m$$

$$r_j^{worst} = \begin{cases} \max\limits_{i=1,\dots,\#X^E} x_{ijk}^E \; IF\; j \in M^- \\ \min\limits_{i=1,\dots,\#X^E} x_{ijk}^E \; IF\; j \in M^+ \end{cases}; j = 1, \dots, m$$

$$r_j^{average} = \frac{1}{\#X^E} \sum_{i=1}^{\#X^E} x_{ijk}^E; \; j = 1, \dots, m$$

Notice that, for each criterion or its measure, the reservation reference point r_j^- is placed half way between the worst and the average observed values of the measure, and the aspiration reference point r_j^+ is placed half way between the average and the best values of the measure. Note that one might, in principle, choose to use a single reference point. However, from an application perspective, individual achievement functions have mainly been designed for the double reference point method. As will be seen in the next step, these functions are crucial for operationalizing these methods. Note that a major advantage of using two reference points is that they allow one to take account of the whole range of data or its distribution better than a single reference point and are likely to enhance the predictive performance of an RPM classifier.

Step 2: Choice and Computation of Each Category of Criteria-dependent Performance Scores

For each entity i ($i = 1, \dots, n$) and each category of performance criteria C_k ($k = 1, \dots, nc$), compute the strong or non-compensating performance score, say S_{ik}^{strong}, which does not allow for any compensation between criteria, as follows:

$$S_{ik}^{strong} = \min_{j \in C_k} \{ a_{ijk}(x_{ijk}^E, r_j^-, r_j^+) \},$$

compute the weak or compensating performance score, say S_{ik}^{weak}, which allows for full compensation between criteria, as follows:

$$S_{ik}^{weak} = \frac{1}{\#C_k} \sum_{j=1}^{\#C_k} a_{ijk}(x_{ijk}^E, r_j^-, r_j^+),$$

and compute the entity i individual achievement function, $a_{ijk}(x_{ijk}^E, r_j^-, r_j^+)$, on criterion j with respect to the benchmarks or reference points r_j^- and r_j^+, first proposed by Ogryczak et al (1992) as a piecewise function, as follows:

$$a_{ijk}\left(x_{ijk}^{E}, r_j^{-}, r_j^{+}\right) = \begin{cases} 1 + \dfrac{x_{ijk}^{E} - r_j^{+}}{r_j^{best} - r_j^{+}}, \ IF\ r_j^{+} \le x_{ijk}^{E} \le r_j^{best} \\ \dfrac{x_{ijk}^{E} - r_j^{-}}{r_j^{+} - r_j^{-}}, \ IF\ r_j^{-} \le x_{ijk}^{E} < r_j^{+} \\ \dfrac{x_{ijk}^{E} - r_j^{-}}{r_j^{-} - r_j^{worst}}, \ IF\ r_j^{worst} \le x_{ijk}^{E} < r_j^{-} \end{cases}$$

Note that this achievement function takes values between -1 and 2. To be more specific, when entity i performance on criterion j is below the reservation level r_j^{-}, $a_{ijk}\left(x_{ijk}^{E}, r_j^{-}, r_j^{+}\right)$ takes values between -1 and 0; when entity i performance on criterion j is between the reservation level $r_j^{-}$and the aspiration level r_j^{+}, $a_{ijk}\left(x_{ijk}^{E}, r_j^{-}, r_j^{+}\right)$ takes values between 0 and 1; and when entity i performance on criterion j is above the aspiration level r_j^{+}, $a_{ijk}\left(x_{ijk}^{E}, r_j^{-}, r_j^{+}\right)$ takes values between 1 and 2. Note also that higher values of S_{ik}^{weak} and S_{ik}^{strong} indicate better performance.

Step 3: Choice and Computation of Overall Performance Scores

Choose a weighting scheme for the nc categories of criteria, which reflects their relative importance for the decision maker, say μ, so that $\sum_{k=1}^{nc} \mu_k = 1$. For each entity i ($i = 1, \dots, n$), aggregate each of the category of criteria-dependent performance scores computed in the previous step into a single one as a convex combination of the weighted sum across categories of weak performance scores and the maximum across categories of weighted strong performance scores, commonly referred to as the mixed performance score, S_{ik}^{mixed} as follows:

$$S_i^{mixed} = \alpha \sum_{k=1}^{nc} \mu_k S_{ik}^{weak} + (1-\alpha) \min_{k=1,\dots,nc}\left\{\mu_k S_{ik}^{strong}\right\}; (0 \le \alpha \le 1).$$

Note that the mixed performance score allows for varying degrees of compensation between criteria ranging from none to full compensation depending on the choice of α. In fact, the choice of a value for α determines the choice of the type of performance score most relevant to the application at hand. To be more specific, $\alpha = 0$ corresponds to choosing the non-compensating performance score; $\alpha = 1$ corresponds to choosing the compensating performance score; and values of $\alpha \in (0,1)$ correspond to choosing a degree of compensation between the former two extremes. Finally, note that higher values of S_i^{mixed} indicate better performance.

Step 4: Use the mixed performance scores computed in the previous step to classify entities in the training sample X_E according to a *user-specified classification rule* into risk classes, say $\hat{Y}_E$. Then, compare the RPM-based classification of entities in X_E into risk classes (i.e., the predicted risk classes $\hat{Y}_E$) with the observed risk classes Y_E of entities in the training sample, and compute the relevant in-sample performance statistics. The choice of a decision rule for classification depends on the nature of the classification problem; that is, a two-class problem or a multi-class problem. In this paper, we are concerned with a two-class problem; therefore, we shall provide a solution that is suitable for these problems. In fact, we propose a RPM score-based cut-off point procedure to classify entities in X_E. The proposed procedure involves solving an optimization problem whereby the RPM score-based cut-off point, say ρ, is determined so as to optimize a given classification performance measure, say π (e.g., Type I error, Type II error, Sensitivity, Specificity), over an interval with a lower bound, say ρ_{LB}, equal to the smallest RPM score of entities in X_E and an upper bound, say ρ_{UB}, equal to the largest RPM score of entities in X_E. Any derivative-free unidimensional search procedure could be used to compute the optimal cut-off score, say ρ^* – for details on derivative-free unidimensional search procedures, the reader is referred to Bazaraa et al. (2006). The optimal cut-off score ρ^* is used to classify observations in X_E into two classes; namely, bankrupt and non-bankrupt firms. To be more specific, the predicted risk classes $\hat{Y}_E$ is determined so that firms with RPM scores less than ρ^* are assigned to a bankruptcy class and those with RPM scores greater than or equal to ρ^* are assigned to a non-bankruptcy class. Note that a novel feature of the design of our RPM score-based cut-off point procedure for classification lies in the determination of a cut-off score so as to optimise a specific performance measure of the classifier.

Phase 2: Out-of-Sample Analysis

Step 5: Use an appropriate *algorithm* to classify entities in X_T into, for example, risk or bankruptcy classes, say $\hat{Y}_T$. Then, compare the predicted risk classes $\hat{Y}_T$ with the observed ones Y_T and compute the relevant out-of-sample performance statistics. Note that entities i in the test sample X_T could be classified using a decision rule similar to the one used for classifying entities in the training sample, where ρ^* is the optimal cut-off score determined in step 4 which is based on the training sample. This naïve classification rule might fail to predict the right class belonging for an entity $i \in X_T$, because the performance of entity i on some criterion j might be better (respectively, worse) than r_j^{best} (respectively, r_j^{worst}) which would make the reference points inappropriate; instead, we propose an instance of case-based reasoning; namely, the k-nearest neighbour (k-NN) algorithm, which could be described as follows:

Initialization Step
Choose the Case Base as X_E and the Query Set as X_T;
Choose a *distance metric* d_{k-NN} to use for computing distances between entities. In our implementation, we tested several choices amongst the following: Euclidean, Cityblock, and Mahalanobis;
Choose a *classification criterion*. In our implementation, we opted for the most commonly used one; that is, the majority vote;
Iterative Step
// Compute distances between queries and cases
FOR $i_1 = 1$ to $|X_T|$ **{**
  FOR $i_2 = 1$ to $|X_E|$ **{**
    Compute $d_{k-NN}\left(entity_{i_1}, entity_{i_2}\right)$; **}}**
// Sort cases in ascending order of their distances to queries and classify queries
FOR $i_1 = 1$ to $|X_T|$ **{**
  Sort the list $L_{i_1} = \left\{\left(i_2, d_{k-NN}\left(entity_{i_1}, entity_{i_2}\right)\right);\ i_2 = 1, \ldots |X_E|\right\}$ in ascending order of distances and use the first k entries in the list $L_{i_1}(1{:}\,k, .)$ to classify $entity_{i_1}$ according to the chosen criterion; that is, the majority vote; **}**

Output: In-sample and out-of-sample classifications or risk class belongings of entities along with the corresponding performance statistics.

Finally, we would like to stress out that, when the decision maker is not confident enough to provide a value for α in step 3 above, one could automate the choice of α. In fact, an optimal value of α with respect to a specific performance measure (e.g., Type 1 error, Type 2 error, Sensitivity, or specificity) to be optimized either in-sample or out-of-sample could be obtained by using a derivative-free unidimensional search procedure, which calls either a procedure that consists of step 1 through step 4 to optimize in-sample performance, or the whole procedure; that is, step 1 through step 5, to optimize out-of-sample performance.

In the next section, we shall report on our empirical evaluation of the proposed framework.

3. Empirical Results

In order to assess the performance of the proposed framework, we considered a sample of 6605 firm-year observations consisting of non-bankrupt and bankrupt UK firms listed on the London

Stock Exchange (LSE) during 2010-2014 excluding financial firms and utilities as well as those firms with less than 5 months lag between the reporting date and the fiscal year. The source of our sample is DataStream. The list of bankrupt firms is however compiled from London Share Price Database (LSPD) – codes 16 (Receivership), 20 (in Administration) and 21 (Cancelled and Assumed valueless). Information on our dataset composition is summarised in Table 1. As to the selection of the training sample and the test sample, we have chosen the size of the training sample to be twice the size of the test sample. The selection of observations was done with random sampling without replacement to ensure that both the training sample and the test sample have the same proportions of bankrupt and non-bankrupt firms. A total of thirty pairs of training sample-test sample were generated.

Observations (2010-2014)	Nb.	%
Bankrupt Firm-Year Observations	407	6.16%
Non-Bankrupt Firm-Year Observations	6198	94.38%
Total Firm-Year Observations	6605	100%

Table 1: Dataset Composition

In our experiment, we reworked a standard and well known parametric model within the proposed RPM framework; namely, the multivariate discriminant analysis (MDA) model of Taffler (1984), to provide some empirical evidence on the merit of the proposed framework. Recall that Taffler's model makes use of four explanatory variables or bankruptcy drivers which belong to the same category; namely, liquidity. These drivers are current liabilities to total assets, number of credit intervals, profit before tax to current liabilities, and current assets to total liabilities. Note that lower values are better than higher ones for Current Liabilities to Total Assets and Number of Credit Intervals, whereas higher values of Current Assets to Total Liabilities and Profit Before Tax to Current Liabilities are better than lower ones. We report on the performance of the proposed framework using four commonly used metrics; namely, Type I error (T1), Type II error (T2), Sensitivity (Sen) and Specificity (Spe), where T1 is the proportion of bankrupt firms predicted as non-bankrupt, T2 is the proportion of non-bankrupt firms predicted as bankrupt, Sen is the proportion of non-bankrupt firms predicted as non-bankrupt, and Spe is the proportion of bankrupt firms predicted as bankrupt.

Since both the RPM classifier and the k-NN classifier, trained on the classification done with RPM, require a number of decisions to be made for their implementation, we considered several combinations of decisions to find out about the extent to which the performance of the proposed framework is sensitive or robust to these decisions. Recall that, for the RPM classifier, the analyst must choose (1) the type of similarity score to use or equivalently a value for α, and (2) the classification rule. On the other hand, for the k-NN classifier, the analyst must choose (1) the metric to use for computing distances between entities, d_{k-NN}, (2) the classification criterion, and (3) the size of the neighbourhood k. Our choices for these decisions are summarised in Table 2.

RPM	
Decision	**Options Considered and Justification, if relevant**
Type of Similarity Score or Value for α	We performed tests for $\alpha = 0, 0.5, 1$.
Classification Rule	RPM score-based cut-off point procedure, where the choice of the cut-off point optimises a specific performance measure (i.e., T1, T2, Sen, Spe)
k-NN	
Decision	**Options Considered and Justification, if relevant**
Metric d_{k-NN}	Euclidean, Cityblock, Mahalanobis.
Classification Criterion	Majority vote. Several criteria could have been used such as a Weighted Vote, but once again our choice is made so as to avoid any personal (subjective) preferences.
Size of the neighbourhood k	$k = 3; 5; 7$. The results reported are for $k = 3$ since higher values delivered very close performances but required more computations.

Table 2: Implementation Decisions for RPM and k-NN

Hereafter, we shall provide a summary of our empirical results and findings. Table 3 provides a summary of in-sample and out-of-sample statistics on the performance of the MDA model of Taffler (1984) reworked within our proposed framework, which is an integrated in-sample – out-of-sample framework for RPM classifiers, for $\alpha = 1$; that is, the weak or compensating performance score, which allows for full compensation between criteria, is used for classification. The performance of the classifier in-sample is outstanding; in fact, none of the bankrupt and non-bankrupt firms are misclassified. On the other hand, the performance of the classifier out-of-sample is also outstanding; however, it is slightly affected by the choice of the distance. To be more specific, the Mahalanobis distance seem to slightly reduce the predictive power of k-NN trained on the risk classes predicted by RPM. The same outstanding predictive performance is also obtained with $\alpha = 0.5$; that is, when the mixed performance score is used for classification – see Table 4. As to the predictive performance of the proposed classifier when using the strong or non-compensating performance score (i.e., $\alpha = 0$), it is lower than the ones obtained with the weak and the mixed scores – see Table 5. This lower performance is expected, as the strong scheme is a much stricter one in that it only identifies the worst performance of each unit. Therefore, as far as bankruptcy prediction is concerned, we recommend the implementation of our framework using the weak performance score as long as the relevant stakeholders are comfortable with a full compensation scheme; otherwise the mixed performance score should be used with α values closer to 1 than to 0. Note that, in risk class prediction applications, a mixed scheme may help to identify risky units that otherwise would not have been identified by a purely weak one. Note, however, that regardless of the choice of the decision maker with regards the compensation scheme, the predictive performance of the proposed classification framework is by far superior to the predictive performance of multivariate discriminant analysis – see Table 6.

	In-sample Performance				
	Statistics	T1	T2	Sen.	Spe.
	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
	Out-of-Sample Performance				
Distance Metric	Statistics	T1	T2	Sen.	Spe.
Euclidean	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
Cityblock	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
Mahalanobis	Min	0%	0%	100%	97.7941%
	Max	2.2059%	0%	100%	100%
	Average	0.2941%	0%	100%	99.7059%
	Std. Dev.	0.4961%	0%	0%	0.4961%

Table 3: Summary Statistics of The Performance of The Proposed Framework for $\alpha = 1$

Distance Metric	In-sample Performance				
	Statistics	T1	T2	Sen.	Spe.
	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
	Out-of-Sample Performance				
	Statistics	T1	T2	Sen.	Spe.
Euclidean	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
Cityblock	Min	0%	0%	100%	100%
	Max	0%	0%	100%	100%
	Average	0%	0%	100%	100%
	Std. Dev.	0%	0%	0%	0%
Mahalanobis	Min	0%	0%	100%	97.7941%
	Max	2.2059%	0%	100%	100%
	Average	0.2941%	0%	100%	99.7059%
	Std. Dev.	0.4961%	0%	0%	0.4961%

Table 4: Summary Statistics of The Performance of The Proposed Framework for $\alpha = 0.5$

Distance Metric	In-sample Performance				
	Statistics	T1	T2	Sen.	Spe.
	Min	13.2841%	0%	100%	77.8598%
	Max	22.1402%	0%	100%	86.7159%
	Average	18.6101%	0%	100%	81.3899%
	Std. Dev.	1.7728%	0%	0%	1.7728%
	Out-of-Sample Performance				
	Statistics	T1	T2	Sen.	Spe.
Euclidean	Min	11.0294%	0%	100%	77.2059%
	Max	22.7941%	0%	100%	88.9706%
	Average	16.8873%	0%	100%	83.1127%
	Std. Dev.	3.1992%	0%	0%	3.1992%
Cityblock	Min	9.5588%	0%	100%	77.9412%
	Max	22.0588%	0%	100%	90.4412%
	Average	16.6422%	0%	100%	83.3578%
	Std. Dev.	3.3970%	0%	0%	3.3970%
Mahalanobis	Min	9.5588%	0%	100%	76.4706%
	Max	23.5294%	0%	100%	90.4412%
	Average	17.0098%	0%	100%	82.9902%
	Std. Dev.	3.6522%	0%	0%	3.6522%

Table 5: Summary Statistics of The Performance of The Proposed Framework for $\alpha = 0$

Statistics	In-sample Performance			
	T1	T2	Sen.	Spe.
Min	97.0500%	0.1900%	99.3700%	0%
Max	100%	0.6300%	99.8100%	2.9500%
Average	98.8200%	0.2600%	99.7400%	1.1800%
Std. Dev.	0.6700%	0.0900%	0.0900%	0.6700%
Statistics	Out-of-Sample Performance			
	T1	T2	Sen.	Spe.
Min	0%	0%	0.1500%	0%
Max	100%	99.8500%	100%	100%
Average	82.2000%	17.0100%	82.9900%	17.8000%
Std. Dev.	37.4300%	37.6600%	37.6600%	37.4300%

Table 6: Summary Statistics of The Performance of MDA

4. Conclusions

The analytics toolbox of risk management is crucial for the financial industry amongst others. In this paper, we proposed a new MCDM classifier for predicting risk class belonging as a new addition to the existing prediction methods available. The proposed classification framework performs both in-sample and out-of-sample predictions. From a design perspective, in-sample predictions of risk class belonging are performed using a new double reference point method-based classifier, whereas out-of-sample predictions are performed with a case-based reasoning algorithm; that is, k-nearest neighbour, which is trained on the in-sample predictions obtained with the double reference point method-based classifier. The proposed classification framework has many important features that drive its outstanding performance such as its non-parametric nature, the design of our RPM score-based cut-off point procedure for in-sample classification, and the choice of a k-Nearest Neighbour as an out-of-sample classifier which is trained on the in-sample classification provided by the new reference point method based classifier. In addition, the basic concepts behind both the reference point method and case-based reasoning are easy to explain to managers. We assessed the performance of the proposed framework using a UK dataset of bankrupt and non-bankrupt firms. Our results support its outstanding predictive performance. In addition, the outcome of the proposed framework is robust to a variety of implementation decisions. Last, but not least, the proposed classification framework delivers a high performance similar to the DEA-based classifier proposed by Ouenniche and Tone (2017).

References

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (2006). Nonlinear Programming: Theory and Algorithms. (3^{rd} edition), New Jersey: John Wiley & Sons Inc.

Benayoun, R., de Montgolfier, J., Tergny, J. and Laritchev, O. (1971). Linear Programming with Multiple Objective Functions: Step method (STEM). *Mathematical Programming*, 1: 366-375.

Blancas, F., Caballero, R., González, M., Lozano-Oyola, M. and Pérez, F. (2010). Goal programming synthetic indicators: an application for sustainable tourism in Andalusian coastal countries. *Ecological Indicators*, 69: 2158-2172.

Buchanan, J. (1997). A naive approach for solving MCDM problems: The GUESS method. *Journal of the Operational Research Society*, 48: 202-206.

Cabello, J., Ruiz, F., Pérez-Gladish, B. and Méndez-Rodríguez, P. (2014). Synthetic indicators of mutual funds' environmental responsibility: An application of the reference point method. *European Journal of Operational Research*, 236: 313-325.

Diaz-Balteiro, L., Voces, R. and Romero, C. (2011). Making sustainability rankings using compromise programming. An application to European paper industry. *Silva Fennica*, 45(4): 761-773.

Granat, J. and Guerriero, F. (2003). The interactive analysis of the multicriteria shortest path problem by the reference point method. *European Journal of Operational Research*, 151: 103-118.

Kozlowski, B. and Ogryczak, W. (2011). On ordered weighted reference point model for multiattribute procurement auctions, in Computational Collective Intelligence. Technologies and Applications, eds. P. Jedrzejowicz, N. Nguyen and K. Hoang, Lecture Notes in Computer Science, 6922, 294-303. Springer, Berlin.

Lewandowski, A. and Wierzbicki, A.P. (Eds.) (1989). Aspiration Based Decision Support Systems. Theory, Software and Applications. Lecture Notes in Economics and Mathematical Systems, 331. Springer, Berlin.

Miettinen, K. and Mäkelä, M.M. (1995). Interactive bundle-based method for nondifferentiable multiobjective optimization: NIMBUS. *Optimization*, 34: 231-246.

Miettinen K. and Mäkelä M.M. (2002). On scalarizing functions in multiobjective optimization. *OR Spectrum*, 24: 193–213.

Miettinen, K. (1999). Nonlinear Multiobjective Optimization. Kluwer, Boston.

Nakayama, H. and Sawaragi, Y. (1984). Satisficing trade-off method for multiobjective programming. In: Grauer, M., Wierzbicki, A.P (Eds.) Interactive Decision Analysis, pp. 113-122. Springer, Heildelberg.

Ogryczak, W. and Lahoda, S. (1992). Aspiration/reservation-based decision support – a step beyond goal programming. *Journal of Multi-Criteria Decision Analysis*, 1: 101-117.

Ogryczak, W., Studzinski, K. and Zorychta, K. (1992). DINAS: A computer-assisted analysis system for multiobjective transshipment problems with facility location. *Computers and Operations Research*, 19, 637-647.

Ogryczak, W. (2001). On goal programming formulations of the reference point method. *Journal of the Operational Research Society*, 52: 691-698.

Ogryczak, W., Perny, P. and Weng, P. (2013). A compromise programming approach to multiobjective markov decision processes. *International Journal of Information Technology & Decision Making*, 12, 1021-1053.

Ouenniche J. and Tone K. (2017), An Out-of-Sample Evaluation Framework for DEA with Application in Bankruptcy Prediction, *Annals of Operations Research*, doi:10.1007/s10479-017-2431-5

Rodríguez-Uría, M., Caballero, R., Ruiz, F., and Romero, C. (2002). Meta-Goal Programming. *European Journal of Operational Research*, 136: 422-429.

Romero, C. (1991). Handbook of Critical Issues in Goal Programming. Pergamon Press, Oxford.

Romero, C. (2001). Extended lexicographic goal programming: A unifying approach. Omega. *The International Journal of Management Science*, 29: 63-71.

Ruiz, F., Luque, M., Miguel, F. and Munoz, M.D.M. (2008). An additive achievement scalarizing function for multiobjective programming problems. *European Journal of Operational Research*, 188: 683–694.

Ruiz, F., Luque, M. and Cabello, J. M. (2009). A classification of the weighting schemes in reference point procedures for multiobjective programming. *Journal of the Operational Research Society*, 60: 544-553.

Ruiz, F., Cabello, J. M. and Luque, M. (2011). An application of reference point techniques to the calculation of synthetic sustainability indicators. *Journal of the Operational Research Society*, 62: 189-197.

Steuer, R.E. and Choo, E.U. (1983). An interactive weighted Tchebycheff procedure for multiple objective programming. *Mathematical Programming*, 26: 326-344.

Steuer, R.E. (1986). Multiple Criteria Optimization. Theory, Computation & Applications. Wiley, New York.

Taffler, R.J. (1984). Empirical models for the monitoring of UK corporations. *Journal of Banking and Finance*, 8(2), 199–227.

Voces, R., Diaz-Balteiro, L. and Romero, C. (2012). Characterization and explanation of the sustainability of the European wood manufacturing industries: A quantitative approach. *Expert Systems with Applications*, 39: 6618-6627.

Wierzbicki, A. P. (1979). A Methodological Guide to Multiobjective Optimization. IIASA Working Paper, WP-79-122, Laxenburg, Austria.

Wierzbicki, A. P. (1980). The use of reference objectives in multiobjective optimization. In G. Fandel and T. Gal, editors, Lecture Notes in Economics and Mathematical Systems, vol. 177, pp. 468-486, Springer-Verlag, Berlin.

Wierzbicki, A. P., Makowski, M. and Wessels, J., (Eds.) (2000). *Model-Based Decision Support Methodology with Environmental Applications*. Kluwer Academic Publishers, Dordrecht.