Comparison between Statistical Approaches and Linear Programming for Resolving Classification Problem

Slah Ben Youssef ¹ and Abdelwaheb Rebai ²

¹Faculté des Sciences Economiques et de Gestion de Sfax

Tunisie, route Aéroport km 4

BP N° 1088, 3018 Sfax, Tunisia

benyoussef_slah@yahoo.fr

²Institut Supérieur d'Administration des Affaires, Sfax

Tunisie, route Aéroport km 4

BP N° 1013, 3018 Sfax, Tunisia

Abdelwaheb.rebai@fsegs.rnu.tn

Abstract

The applications that are related to classification problem are wide-ranging. In fact, differentiating between patients with strong prospects for recovery and those highly at risk, between good credit risks and poor ones, or between promising new firms and those likely to fail, are among the most known of these applications. To solve such classification problem, several approaches have been applied. In this paper, on one hand, we dealt with the parametric approach illustrated by the use of Fisher's linear discriminant function, Smith's quadratic discriminant function and the logistic discriminant model. On the other hand, we studied the non-parametric approach such as linear programming models as alternatives to the parametric

approach when hypotheses of this later are not satisfied. Such hypotheses are: the distribution's normality, the homogeneity of variance-covariance matrix, the sample's size and the absence of outliers in data. Our study showed that, being more flexible in such a way to allow the analyst to incorporate some a priori information in the models, the non-parametric approach outperforms the parametric one. Nevertheless, this does not exclude the use of statistical techniques once the required hypotheses are satisfied.

Keywords: Fisher's linear discriminant function, Smith's quadratic discriminant function, logistic regression, classification problem and linear programming models

Introduction

Many practical problems can be reduced to the assignment of various objects to different classes. For example in the case of the medical diagnosis, it is a question of recognizing the pathology of a given patient; the purposes correspond to the patients and the classes with various pathologies. In the economy field, a bank wants to know if a customer applying for a loan is a good or bad customer while being based on several variables like the age, the profession, former fidelity, the required credit.

Indeed, there are three parametric discriminant methods. The first is the linear discriminant function who is the oldest discriminant method initiated by Fischer in 1936. It is the optimal combination which separates the averages from two groups. This method of discrimination requires that the sample be distributed normally and that the variances-covariances matrices of the two groups are homogeneous. The second method is the quadratic function suggested by Smith in 1947. This method supposes the normality of the sample with heterogeneous variances-covariances matrices. The last parametric method of discrimination is the logistic regression, which is an econometric method whose endogenous variable is binary; it requires neither the normality of the sample nor the homogeneity of the variances-covariances matrices of the groups.

Recently, several linear programming models were proposed for resolving the classification problems by various authors such as Freed and Glover (1981), Hand (1981), Bajgier and Hill (1982), Glorfeld and Gaither (1982), Markowski and Markowski (1985), Gehrlein 1986, Glover and *al.* (1988), Nath and Jones (1988), Koehler (1990) and Stam and Ragsdale (1991).

These approaches were justified by the violation of certain assumptions in the traditional discriminant methods such as: the normality of distributions, the sample size and the homogeneity of variances-covariances matrices. This violation made these traditional methods inadequate for the complex discriminant tasks. As several empirical studies showed, the mathematical programming approaches can produce more powerful classification rules than the traditional discriminant methods.

1 Discriminant analysis methods

The methods of discriminant analysis were largely studied. The literature on this subject is very abundant. We present in this paragraph a short description of these methods. The aim of these methods is to produce decisions concerning the membership or not of an object to a class by using discriminant functions also called decision functions. This kind of methods is based on Fisher's work (1936). According to the forms of the classes, we can find various types of discrimination:

1.1 Linear discriminant Function (LDF)

This method consists in separating the classes by linear frontiers in order to group the points to be classified around the centre of gravity of the class (the average of the class) and to create a linear hyperplane between the classes. The application of this method requires certain assumptions: the normality distribution of the sample, homogeneity of the variances-covariances matrices and an important size of the basic sample.

If we have n variables, the discriminant function is:

$$d(a) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1}x_{n+1} \quad (1)$$

This function depends on parameters $w_1, \dots, w_n, w_{n+1}$. The determination of these parameters is done by an algorithm of training, which aims at satisfying the criterion associated with the model. According to the data, the criterion more used to adjust these parameters is that which generally aims at minimizing the error of classification.

If we have k classes, we define k discriminant functions:

$$d_i(a) = W_i X^t \quad \text{avec } W_i = (w_{i1}, \dots, w_{in}) \text{ et } X = (x_1, \dots, x_n, 1)$$

The rule of assignment is given as follows:

$$\text{If } d_i(a) > 0 \quad \text{then} \quad a \in C^i \quad \text{for } i = 1, \dots, k$$

For better illustrating the method, the majority of the authors consider the case of two groups G_1 and G_2 , which are defined by two variables x_1 and x_2 . The LDF procedure is illustrated by the Diagram below:

Figure 1. Illustrative diagram of the LDF procedure

The LDF method consists in identifying the hyperplane, which is crossing the double scatter of points representative of the two groups, will have the best discriminating powers of the two groups:

- The distance $\bar{z}_1$ and $\bar{z}_2$ are maximized;
- The area of uncertainty is the possible weakest;
- The high number of good classified

1.2 Quadratic discriminant function (QDF)

The quadratic discriminant function was proposed by Smith in 1947, supposes the normality of the sample with heterogeneous variances-covariances matrices. This procedure is given by the following equation:

$$f(X) = X' \Gamma X + \beta' X - \alpha$$

Where

- $\Gamma = S_2^{-1} - S_1^{-1}$;
- $\beta = 2(S_1^{-1} m_1 - S_2^{-1} m_2)$;
- $\alpha = \ln \left(\frac{|S_1|}{|S_2|} \right) + m_1' S_1^{-1} m_1 - m_2' S_2^{-1} m_2$;
- S_k is the variances-covariances matrices of group k , $k = 1, 2$;
- $|S_k|$ is the determinant of variances-covariances matrices of group k , $k = 1, 2$;
- m_k is the average vector of group k , $k = 1, 2$.

This quadratic function minimizes the total probability of misclassification, the assignment rule assigns X to group 1 if and only if:

$$(X - \bar{X}_2)S_2^{-1}(X - \bar{X}_2)' - (X - \bar{X}_1)S_1^{-1}(X - \bar{X}_1)' \geq \ln|S_1| - \ln|S_2|$$

and to group 2 otherwise

1.3 Logistic regression (LGD)

Logistic regression is a widely used statistical modelling technique in which the probability of a dichotomous outcome is related to a set of potential predictor variables in the form:

$$\log[p/(1-p)] = \beta_0 + \beta_1 x_1 + \dots + \beta_i x_i \quad (3)$$

Where p is the probability of the outcome of interest, β_0 is the intercept term, and $\beta_i (i = 1, \dots, n)$ represents the coefficient associated with the corresponding explanatory variable $x_i (i = 1, \dots, n)$ (Cox, 1970; Cox & Snell, 1989; Hosmer & Lemeshow, 1989; Pampel, 2000). The dependent variable is the logarithm of the odds, $\{\log[p/(1-p)]\}$ which is the logarithm of the ratio of two probabilities of the outcome of interest. These variables are usually selected for inclusion by using some form of backward or forward stepwise regression technique (Neter, Kutner, Nachtsheim, & Wasserman, 1996; Pampel, 2000) even though these selection techniques may be prone to problems. In addition, the maximization of the likelihood function is usually applied as the convergent criterion to estimate the coefficients of corresponding parameters when the logistic regression models are utilized.

The logistic regression model does not necessarily require the assumptions of LDA. However, Harrell and Lee (1985) found that logistic regression is as efficient and accurate as LDA even though the assumptions of LDA are satisfied. One advantage of LDA is that ordinary least square estimation procedure can be implemented to estimate the coefficients of the linear discriminant function, whereas maximum likelihood methods are required for the estimation of logistic regression models. Another advantage of LDA over logistic regression is that prior probabilities and misclassification costs can easily be incorporated into the LDA approach (Desai et al., 1996).

Logistic regression models have been widely used in social research, medical research, biological research, food science, design, control, bankruptcy prediction, market segmentation, and customer behaviors (Kay, Warde, & Martens, 2000; Laitinen & Laitinen, 2000; Suh, Noh, & Suh, 1999; Vellido, Lisboa, & Vaughan, 1999; Wong, Bodnovich, & Selvi, 1997). Henley (1995), Joanes (1993), Laitinen (1999), Westgaard, has also explored logistic regression and van der Wijst (2001), Wiginton (1980) in building credit scoring models for personal loan, business loan, and credit card applications.

2 Linear programming models

To simplify, we will consider only the discrimination problem with two groups, although there are other models being able to be applied to the problems with more than two groups. In this section, we will expose only three types of the classification models: the MMD model, the MSD model and the RS model.

Let n_k observations belonging to the group k , $k = 1, 2$, with each observation is defined by p variables (attributes).

Let X_{kij} the value of the variable j , $j = 1, 2, \dots, p$, for the observation i , $i = 1, 2, \dots, n_k$, of the group k , $k = 1, 2$.

The linear programming models for the resolution of the classification problems are used to determine the coefficients w_j $j = 1, 2, \dots, p$, for variable j and the constant term b . These different models enable to determine the discriminant function, which will be used to classify new observations in one of the two groups. For example, for a new observation i , x_{ij} , $j = 1, 2, \dots, p$, is the value of the variable j if:

- (i) $\sum_{j=1}^p w_j x_{ij} > b$: the observation is classified in group 1;
- (ii) $\sum_{j=1}^p w_j x_{ij} < b$: the observation is classified in group 2;
- (iii) $\sum_{j=1}^p w_j x_{ij} = b$: the observation can not be classified.

2.1 The MMD model

Freed and Glover in 1981 and 1986 suggest the MMD model (Maximize the Minimum of Deviations) to solve the classification problem. Their formulation is as follows:

$$\begin{aligned}
 & \text{Max } d \\
 & \quad w, d \\
 & \text{St} \\
 \text{(MMD)} \quad & \sum_{j=1}^p w_j x_{ij} - d \geq b, \quad \forall i \in G_1 \quad (4) \\
 & \sum_{j=1}^p w_j x_{ij} + d \leq b, \quad \forall i \in G_2 \quad (5) \\
 & \sum_{j=1}^p w_j = 1 \quad (6)
 \end{aligned}$$

Where w_j for $j = 1, \dots, p$, and b are unrestricted in sign. The equation (6) is a normalization constraint to avoid the trivial solution (that is, an all zero solution).

2.2 The MSD model

The “Minimize the Sums of Deviations” model (MSD) was proposed by Bajgier and Hill [12] and Freed and Glove [4, 5]. The objective of this formulation is to minimize the total group overlap. This model can be formulated as follows:

$$\begin{aligned}
 & \text{Min } \sum_{i=1}^n d_i \\
 & \text{St} \\
 (\text{MSD}) \quad & \sum_{j=1}^p w_j x_{ij} + d_i \geq b, \quad \forall i \in G_1 \quad (7) \\
 & \sum_{j=1}^p w_j x_{ij} - d_i \leq b, \quad \forall i \in G_2 \quad (8) \\
 & \sum_{j=1}^p w_j = 1 \quad (9)
 \end{aligned}$$

Where w_j for $j = 1, \dots, p$, and b are unrestricted in sign, and $d_i \geq 0$ for all i . A normalization constraint (equation (9)) is needed to avoid a trivial solution. The classification rule of MSD model is:

$$\begin{aligned}
 d_i &= 0 && \text{if applicant } i \text{ is correctly classified} \\
 d_i &= \sum_{j=1}^p w_j x_{ij} && \text{if applicant } i \text{ is misclassified}
 \end{aligned}$$

2.3 The RS model

To solve the classification problems, which contain outliers, Ragsdale and Stam [2] propose the RS model. The RS LP formulation can be set up as follows:

$$\begin{aligned}
 & \text{Min } \sum_{i=1}^n d_i \\
 & \text{St} \\
 (\text{RS}) \quad & \sum_{j=1}^p w_j x_{ij} + d_i \geq b_1, \quad \forall i \in G_1 \quad (10) \\
 & \sum_{j=1}^p w_j x_{ij} - d_i \leq b_2, \quad \forall i \in G_2 \quad (11)
 \end{aligned}$$

Where w_j is unrestricted in sign, for $j = 1, \dots, p$, $d_i \geq 0$ for all i . b_1 and b_2 are two predetermined constants and $b_1 > b_2$. If the classification score is between b_1 and b_2 the observation is misclassified.

3 Empirical study

To compare the parametric approaches and the non-parametric approaches we consider the data set of 30 applicants. This data set is divided into two equal groups of applicants: applicants who were accepted, and were good credit (Group 1 “good loans”); applicants who were accepted, but defaulted on their loans obligations (Group 2 “bad loans”). Further, the two groups are characterized by four

quantitative variables (income (x_1), a number of cards (x_2), age (x_3) and number of children (x_4)).

3.1 The results of statistical approaches

- Result of LDF procedure

The aim of this method is to select, among the set of variables, the most significant variables and to determine the Fisher's linear discriminant function. The table 1 illustrates the most significant variables given by the stepwise procedure using the Lambda of Wilks

Table 1 Stepwise procedure									
Lambda of Wilks									
Step	Introduce	Statistic	df1	df2	df3	Statistic	df1	df2	Sig
1	Income	0.391	1	1	28	43.592	1	28	0.000
2	Number of cards	0.314	2	1	28	29.537	2	27	0.000

This table shows that the most significant variables are the income and the number of cards. Now, we must determine the coefficients of these variables and the constant term to give the Fisher's linear discriminant function. The Table 2 below gives the coefficients of linear discriminant function:

Table 2 Coefficients of linear discriminant function	
	Function
Income	1.988
Number of cards	1.324
Constant	-22.724

Therefore, the linear discriminant function is given by the following equation:

$$Z = -22.724 + 1.988 \times \text{Income} + 1.324 \times \text{Number of cards} \quad (10)$$

By using the equation (10) we can classify the new observation into group 1 or group 2, also we can determine the hit ratio of classification. The classification result is given by the following table 3:

Table 3 Classification result of linear discriminant function

			Assignment classes		
			1	2	Total
Original	Effectif	Group			
		1	14	1	15
	2	1	14	15	
	Hit ratio	1	93.3	6.7	100
2		6.7	93.3	100	
Cross-validation	Effectif	1	14	1	15
		2	1	14	15
	Hit ratio	1	93.3	6.7	100
		2	6.7	93.3	100

According to table 3, we can notice that there are only two observations that are misclassified. Therefore, the hit ratio given by the LDF procedure in the original sample and the cross-validation sample is 93.3%.

This value indicates that the equation (10) gives a good discrimination between the two applicants groups and can be used to classify the new observations.

To confirm the result given by table 3 we consider the following figure (cf. Fig2):

Figure 2. The scatter plot of two groups (LDF procedure)

By examining this figure, we can note that only observation n°1 and n°30 are misclassified. Indeed, observation n°1 belongs to group1 but it is classified into

group 2; and observation n°30 belongs to group 2 and is classified into groupe1 (the red observations). Hence, the analytical result was confirmed graphically by the figure 2.

• *Result of Quadratic discrimination function*

To apply the quadratic discriminant function we must determine the variances-covariances matrices of the two groups. Let S_1 and S_2 be the variances-covariances matrices of the G_1 and G_2 , respectively such as:

$$S_1 = \begin{pmatrix} 16.353 & -2.885 & 4.519 & -2.550 \\ -2.885 & 6.924 & 7.690 & -0.105 \\ 4.519 & 7.690 & 34.810 & -1.452 \\ -2.550 & -0.105 & -1.452 & 2.552 \end{pmatrix} \text{ and } S_2 = \begin{pmatrix} 2.831 & 1.592 & 4.333 & 0.06905 \\ 1.592 & 3.981 & 0.405 & -0.810 \\ 4.333 & 0.405 & 20.952 & 0.167 \\ 0.06905 & -0.810 & 0.167 & 1.952 \end{pmatrix}$$

The determinant of the matrix S_1 and S_2 is equal, respectively, to 4981.7 and 183.578. From where, $\ln(|S_1|) = 8.513$ and $\ln(|S_2|) = 5.213$.

Let S_1^{-1} and S_2^{-1} be, respectively, the inverse matrices of S_1 and S_2 such as:

$$S_1^{-1} = \begin{pmatrix} 0.0913 & 0.0640 & -0.0226 & 0.0810 \\ 0.0640 & 0.2355 & -0.0587 & 0.0403 \\ -0.0226 & -0.0587 & 0.0446 & 0.4747 \\ 0.0810 & 0.0403 & 0.0004 & 0.4747 \end{pmatrix} \text{ and } S_2^{-1} = \begin{pmatrix} 0.8132 & -0.3763 & -0.1595 & -0.1712 \\ -0.3390 & 0.4306 & 0.0603 & 0.1855 \\ -0.1735 & 0.0847 & 0.0817 & 0.0343 \\ -0.1546 & 0.1848 & 0.0237 & 0.5924 \end{pmatrix}$$

According to the quadratic discriminant function, the observation is classified into group 1 if and only if:

$$(X - \bar{X}_2)S_2^{-1}(X - \bar{X}_2)' - (X - \bar{X}_1)S_1^{-1}(X - \bar{X}_1)' \geq 3.3$$

Where $\bar{X}_1$ and $\bar{X}_2$ represent, respectively, the average of group1 and group 2 such as:

$$\bar{X}_1 = (18.213 \ 5.933 \ 40.666 \ 2.533) \text{ and } \bar{X}_2 = (10.7467 \ 2.8667 \ 31.3333 \ 2.3333)$$

The classification result of the different observations by the quadratic discriminant function is given in the following table 4:

Table 4 Classification result of quadratic discrimination function				
	Group	Assignment classes		Total
		1	2	
Original	1	15	0	15
	2	2	13	15
Hit ratio	1	100	0	100
	2	13.4	86.6	100

According to table 4, only two observations of group 2 (observations n°26 and n°30) are misclassified considering their classification score, which is equal,

respectively, to 5 and 5.76 exceeds the critical level of quadratic discrimination function (3.3). From where, the hit ratio given by this procedure of discrimination is equal to 93.3%. We notice that the two procedures of discrimination (LDF and QDF) give the same hit ratio, but they haven't the same misclassified observations.

- *Result of logistic regression*

If the assumptions of LDF procedure and QDF procedure aren't verified, we recourse to the logistic regression. This type of procedure doesn't require any assumption. The aim of the logistic regression is to estimate the parameters of the following equation:

$$Y_i = \beta_0 + \beta_1 \times \text{income} + \beta_2 \times \text{nbcards} + \beta_3 \times \text{age} + \beta_4 \times \text{nbchildren} + \varepsilon_i$$

For observation i , we suppose that:

$$Y_i = \begin{cases} 1 & \text{If applicant } i \text{ recives a credit} \\ 0 & \text{Otherwise} \end{cases}$$

The estimation of parameters as well as the significativity of variables are given in the following table5:

Table 5 The parameters estimation				
Variables	Coefficients	Std.Error	z-Statistic	Prob
β_0	-1.10	0.3069	-3.586	0.001
Income	0.064135	0.01467	4.364	0.000
Nbcards	0.047345	0.02384	1.986	0.048
Age	0.00918	0.0114	0.806	0.4278
Nbchildren	0.054206	0.0367	1.473	0.1532

For a risk of $\alpha = 5\%$ the table 5, shows that only the two variables *Income* and *Nbcards* and the constant term β_0 are significant. From where, the logistic function can be written in the following form:

$$\hat{Y}_i = -1.10_0 + 0.064135 \times \text{income} + 0.0443745 \times \text{Nbcards}$$

Let $p_i = P(Y_i = 1 / X_i) = \frac{1}{1 + e^{-\hat{Y}_i}}$ and $1 - p_i = P(Y_i = 0 / X_i) = \frac{1}{1 + e^{\hat{Y}_i}}$ represent,

respectively, the probabilities to belong to G_1 and G_2 . Therefore, an observation x will be classified in the group having the greatest probability. The classification result of logistic regression is given by the following table 6:

Table 6 Classification result of logistic regression

	Group	Assignment classes		Total
		1	2	
Original	1	14	1	15
	2	2	13	15
Hit ratio	1	93.3	6.7	100
	2	13.4	86.6	100

According to the table 6 and the probabilities, we notice that there are only 3 misclassified observations (n°1, n°25 and n°30). From where, the hit ratio given by this procedure of discrimination is equal to 89.95%.

The following figure (cf. Fig3) confirms the above result:

Figure 3. The scatter plot of two groups (logistic regression procedure)

This figure shows that the observations n°1, n°25 and n°30 (red observations) are three misclassified observations. Indeed, observation n°1 is located on the left of the hyperplane, while the observations n°25 and n°30 are located on the right of the hyperplane. Hence, the analytical result was confirmed graphically by the figure 3.

3.2 The results of linear programming models

• Result of the MSD model

The resolution of the MSD model by LINDO with the cut-off value $b = 8$ gave the following weightings: $(\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4) = (0.535, 0.371, -0.060, 0.154)$. Thus, the classification rule is:

$$d_i = 0 \quad \text{if applicant } i \text{ is correctly classified}$$

$$d_i = 8 - \sum_{j=1}^4 w_j x_{ij} \quad \text{if applicant } i \text{ is misclassified}$$

The table 7 below gives the classification result of the MSD model:

Table 7 Classification result of the MSD model				
	Group	Assignment classes		Total
		1	2	
Original	1	15	0	15
	2	0	15	15
Hit ratio	1	100	0	100
	2	0	100	100

According to this table, we can notice that the hit ratio of the MSD model is equal to 100%. From where, the misclassification percentage is equal to zero because the different observations have the null deviations ($d_i = 0$). Indeed, the hyperplane $\sum_{j=1}^4 w_j x_{ij} = 8$ gives the hard discrimination between the two applicants groups. This result shows that the MSD model is a very effective discrimination procedure to resolve the classification problems.

• Result of the RS model

The weightings given by the RS model after resolution by LINDO software with the cut-off values $b_1=11$ and $b_2=10$ are equal to:

$$(\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4) = (0.60, 0.167, 0.028, 0.225).$$

Therefore, the classification rule of the RS model is:

- If $\sum_{j=1}^4 w_j x_{ij} \geq 11$ the applicant belongs to G_1 , and
- If $\sum_{j=1}^4 w_j x_{ij} \leq 10$ the applicant belongs to G_2 .

The “good and badly” classified observations and the hit ratio are given by the following table 8:

Table 8 Classification result of the RS model

	Group	Assignment classes		Total
		1	2	
Original	1	15	0	15
	2	2	13	15
Hit ratio	1	100	0	100
	2	13.4	86.6	100

The table 8 shows that there are only two observations of group 2 (observations n°26 and n°30), which are misclassified considering their score of classification (11.13 and 10.32, respectively) exceeds the cut-off value $b_2 = 10$. Whereas, for group 1 all observations are good classified. From where, the hit ratio of RS model is equal to 93.3%.

• *Result of the MMD model*

The weightings and the cut-off value given by the MMD model are equal to $(\hat{w}_1, \hat{w}_2, \hat{w}_3, \hat{w}_4) = (0.5407, 0.34946, -0.051224, 0.16107)$ and $b = 8.1209451$.

Therefore, the classification rule of MMD model is:

- If $\sum_{j=1}^4 w_j x_{ij} > 8.1209451$ the applicant belongs to G_1 , and
- If $\sum_{j=1}^4 w_j x_{ij} < 8.1209451$ the applicant belongs to G_2 .

The scores and assignment group of applicants are given in the following table 9:

Table 9 Scores and assignment group of applicants (MMD model)

Applicant	Score of applicant	Assignment group	Applicant	Score of applicant	Assignment group
1	8.16	G_1	16	12.79	G_1^*
2	9.03	G_1	17	16.76	G_1^*
3	8.16	G_1	18	3.84	G_2
4	8.88	G_1	19	5.56	G_2
5	24.18	G_1	20	4.95	G_2
6	10.83	G_1	21	5.86	G_2
7	9.90	G_1	22	4.43	G_2
8	11.68	G_1	23	4.14	G_2
9	12.89	G_1	24	6.31	G_2
10	8.96	G_1	25	6.02	G_2
11	10.21	G_1	26	8.09	G_2
12	8.16	G_1	27	6.46	G_2
13	12.04	G_1	28	4.18	G_2
14	14.61	G_1	29	5.45	G_2
15	11.18	G_1	30	8.09	G_2

* misclassified applicant

According to the table 9, we can notice that only applicant number 16 and applicant number 17 are misclassified. Indeed, their classification scores (12.79 and 16.76) exceed the cut-off value ($b \approx 8.121$). Therefore, the hit ratio is equal to 93.3% and the classification result of MMD model is given by the following table 10:

Table 10 Classification result of the MMD model

		Assignment classes		Total
		1	2	
Original	Group			
	1	15	0	15
	2	2	13	15
Hit ratio	1	100	0	100
	2	13.4	86.6	100

To recapitulate, we consider the following table 11, which summarizes the different classification results found by the parametric and non-parametric methods

Table 11 The summary table of different methods

		Assignment classes												
		LDF		QDF		LGD		MSD		RS			MMD	
Original	Group	1	2	1	2	1	2	1	2	1	2	1	2	Total
	1	14	1	15	0	14	1	15	0	15	0	15	0	15
	2	1	14	2	13	2	13	0	15	2	13	2	13	15
Hit ratio	1	93.3	6.7	100	0	93.3	6.7	100	0	100	0	100	0	100
	2	6.7	93.3	13.4	86.6	13.4	86.6	0	100	13.4	86.6	13.4	86.6	100
Percentage of good classified		93.3		93.3		89.95		100		93.3		93.3		

This table shows the effectiveness of the non-parametric methods compared to the parametric methods. Indeed, the hit ratio of linear programming models is, in most cases, greater than that's of parametric methods. Furthermore, the linear programming models are more flexible than the parametric methods; this is due to the presence of the constraints in the linear programming models.

4 Conclusion

The aim of this paper is the comparison between parametric methods and non-parametric methods to resolve the classification problems. Indeed, the recourse to the linear programming models is justified by the violation of certain assumptions in the statistical discriminant methods. Such assumptions are the distribution's

normality, the homogeneity of variance-covariance matrix, the sample's size and the absence of outliers in data.

The classification results of the different methods applied on our data set concerning the credit scoring for the private individuals showed that, in the majority of cases, the classification result of linear programming models is more efficient than the result of statistical methods. Moreover, our study showed that, the linear programming models are more flexible in such a way to allow the analyst to incorporate some a priori information in the models. Nevertheless, this does not exclude the use of statistical techniques once the required hypotheses are satisfied.

References

- [1] A. Stam, Nontraditional approaches to statistical classification: some perspectives on L_p -norm methods, *Annals of Operation research*, 74 (1997), 1-6
- [2] C. T. Ragsdale & A. Stam, Mathematical programming formulations for the discriminant problem: an old dog does new tricks, *Decision Sciences*, 22 (1991), 296-307.
- [3] F. Glover, Improved linear programming models for discriminant analysis, *Decision Sciences*, 21 (1990), 771-785.
- [4] F. Glover & N. Freed, A linear programming approach to the discriminant problem, *Decision Sciences*, 12 (1981), 68-74.
- [5] F. Glover & N. Freed, Evaluating alternative linear programming models to solve the two-group discriminant problem, *Decision Sciences*, 17 (1986), 151-162.
- [6] F. Lam & J. W. Moy, Improved linear programming formulations for the multi-group discriminant problem, *Journal of the Operation Research Society*, 47 (1996), 1526-1529.
- [7] F. Lam & J. W. Moy, An experimental comparison of some recently developed linear programming approaches to the discriminant problem, *Computers Ops Res.*, 24 (1997), 593-599.
- [8] J. W. Moy & Kim Fung Lam, Combining discriminant methods in solving classification problems in two-group discriminant analysis, *European Journal of Operational Research*, 138 (2002), 294-301.
- [9] Joachimsthaler & A. Stam, Solving the classification problem in discriminant analysis via linear and nonlinear programming methods, *Decision Sciences*, 20 (1989), 284-293.
- [10] M. Doumpos and al., Assessing Country Risk Using a Multi-Group Discrimination Method: a comparative analysis, *Managerial Finance* 27 (2001), 16-34.
- [11] P. A. Rubin, A comparison of linear programming and parametric approaches to the two group discriminant problem, *Decision Sciences*, 21 (1990), 373-386.

- [12] S. Bajgier & A. Hill, An experimental comparison of statistical and linear programming approaches to the discriminant problem, *Decision Sciences*, 13 (1982), 604-618.
- [13] T. M. Cavalier, J. P. Ignizio & A. L. Soyster, Discriminant analysis via mathematical programming: Certain problems and their causes, *Computers Ops Res.*, 16 (1989), 353-362.

Received: May 21, 2007