Research Article

An Application of Monte-Carlo-Based Sensitivity Analysis on the Overlap in Discriminant Analysis

S. Razmyan and F. Hosseinzadeh Lotfi

Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran

Correspondence should be addressed to S. Razmyan, s.razmyan@qhttc.com

Discriminant analysis (DA) is used for the measurement of estimates of a discriminant function by minimizing their group misclassifications to predict group membership of newly sampled data. A major source of misclassification in DA is due to the overlapping of groups. The uncertainty in the input variables and model parameters needs to be properly characterized in decision making. This study combines DEA-DA with a sensitivity analysis approach to an assessment of the influence of banks' variables on the overall variance in overlap in a DA in order to determine which variables are most significant. A Monte-Carlo-based sensitivity analysis is considered for computing the set of first-order sensitivity indices of the variables to estimate the contribution of each uncertain variable. The results show that the uncertainties in the loans granted and different deposit variables are more significant than uncertainties in other banks' variables in decision making.

1. Introduction

The classification problem of assigning observations to one of different groups plays an important role in decision making. When observations are restricted to one of two groups, the Binary classification has wide applicability in business environments.

Discriminant analysis (DA) is a classification method that can distinguish the group membership of a new observation. A group of observations for which the memberships have already been identified is used for the estimation of a discriminant function by some criteria, such as the minimization of misclassification. A new sample is classified into one of the groups based on the gained results [1].

Mangasarian [2] identified that linear programming (LP) could be used to determine separating hyperplanes, namely, when two set of observations are linearly separable use linear discriminant function. Freed and Glover [3] and Hand [4] using objectives such as minimization of the sum of deviations (MSD) or maximization of the minimum

deviation (MMD) of misclassified observations from the separating hyperplane, when sets of observations that are not necessarily linearly separable, proposed LP methods for generating linear discriminant function.

Then a model is based on the goal programming (GP) extension of LP by choosing different criteria, such as minimizing the maximum deviation, maximizing the minimum deviation, minimizing the sum of interior deviation, minimizing the sum of deviations, minimizing misclassified observations, minimizing external deviations, maximizing internal deviations, maximizing the ratio of internal to external, and hybrid models for which there are both advantages and deficiencies [5–9].

In DA, LP and other mathematical programming (MP) based approaches are non-parametric and more flexible than statistical methods [6, 10]. Retzlaff-Roberts [11, 12] and Tofallis [13] proposed the use of DEA-ratio model for DA. Sueyoshi [14] using a data envelopment analysis (DEA) additive model, described a goal programming formulation of DA in which the proposed model is more directly linked to minimizing the sum of deviations from the separating hyperplane; this method was named DEA-DA to distinguish it from other DA and DEA approaches. The original GP version of DEA-DA could not deal with negative data. Therefore, Sueyoshi [15] extended DEA-DA to overcome this deficiency. This approach was designed to minimize the total distance of misclassified observations and formulated by two-stage GP formulations. The number of misclassifications can, however, be considered as measures of misclassification, in which binary variables indicate whether observations are correctly or in correctly classified. Bajgier and Hill [16] proposed a Mixed Integer Programming (MIP) model that included the number of misclassifications in the objective function for the two-group discriminant problem. Gehrlein [17] and Wilson [18] introduced a MIP approach for minimizing the number of misclassified observations in multigroup problems. Chang and Kuo [19] proposed a procedure based on benchmarking model of DEA to solve the two-group problems. Sueyoshi [20] reformulated DEA-DA by MIP to minimize the total number of misclassified observations. When an overlap between two groups is not a serious problem, dropping the first stage of the two-stage MIP approach simplifies the estimation process [21].

Sensitivity analysis provides an understanding of how the model outputs is affected by changes in the inputs, therefore it can assist to increase the confidence in the model and its predictions. sensitivity analysis can use in deciding whether inputs estimates are sufficiently precise to give reliable predictions or we can find the model parameters that can be eliminated.

Two classes in sensitivity analysis have been distinguished [22].

Local Sensitivity Analysis

Studies how some small variations of inputs around a given value, change the value of the output. This approach is practical when the variation around a baseline of the input-output relationship to be assumed linear.

Global Sensitivity Analysis

Takes into account all the variation range of the inputs, and has for aim to apportion output uncertainty to inputs' ones. It quantified the output uncertainty due to the uncertainty in the input parameters. Global sensitivity analysis apportions the output uncertainty to the uncertainty in the input factors, described typically by probability distribution functions that cover the factors' ranges of existence.

Local methods are less helpful when sensitivity analysis is used to compare the effect of various factors on the output, as in this case the relative uncertainty of each input should be weighted. A global sensitivity analysis technique thus incorporates the influence of the whole range of variation and the form of the probability density function of the input. The variance-based methods can be considered as a quantitative method for global sensitivity analysis. In this Study, the Sobol' decomposition in the framework of Monte Carlo simulations (MCS), [22], which is from the family of quantitative methods for global sensitivity analysis, is applied to study of the effect of the variability in DA due to the uncertainty in the variables. The results of the sensitivity analysis can determine which of the variables have a more dominant influence on the uncertainty in the model output.

This paper is organized as follows: Section 2 briefly introduces the DEA-DA model; Section 3 describes the sensitivity analysis based on a Monte-Carlo simulation. Section 4 contains an example and the conclusion is provided in Section 5.

2. Data Envelopment Analysis-Discriminant Analysis (DEA-DA)

The two-stage MIP approach [20] is used in this study to describe DEA-DA. We considered two groups (G_1 and G_2) for which the sum of the two groups has n observations ($j = 1, \ldots, n$). Each observation has m independent factors ($i = 1, \ldots, m$), denoted by r_{ij}. It is necessary to identify the group membership of each observation before its computation. In the two stage approach, the computation process consists of classification and overlap identification, and handling overlap. The first stage is formulated as follows [20].

Stage 1.

$$
\begin{aligned}
&\text{Minimize} \quad s, \\
&\text{Subject to} \quad \sum_{i=1}^{m} (\lambda_i^+ - \lambda_i^-) r_{ij} - d + s \geq 0, \quad j \in G_1, \\
&\qquad\qquad\quad \sum_{i=1}^{m} (\lambda_i^+ - \lambda_i^-) r_{ij} - d - s \leq -\varepsilon, \quad j \in G_2, \\
&\qquad\qquad\quad \sum_{i=1}^{m} (\lambda_i^+ + \lambda_i^-) = 1, \\
&\qquad\qquad\quad \xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+, \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, \quad i = 1, \ldots, m, \\
&\qquad\qquad\quad \xi_i^+ + \xi_i^- \leq 1, \quad i = 1, \ldots, l, \\
&\qquad\qquad\quad \sum_{i=1}^{m} (\xi_i^+ + \xi_i^-) = l, \\
&\qquad\qquad\quad d, s : \text{unrestricted}, \ \xi_i^+, \xi_i^- \in \{0, 1\} \text{ and all other variables} \geq 0.
\end{aligned}
$$

(2.1)

Here "d" indicates a discriminant score for group classification and "s" indicates the size of an overlap between two groups.

Let λ_i^* ($=\lambda_i^{+*} - \lambda_i^{-*}$) and d^* and s^* be an optimal solution of the model (2.1). Then, the original data set (G) is classified into the following subsets $G = G_1 \cup G_2 = C_1 \cup N_1 \cup C_2 \cup N_2$, where

$$C_1 = \left\{ j \in G_1 \middle| \sum_{i=1}^{m} \lambda_i^* \, r_{ij} > d^* + s^* \right\},$$

$$C_2 = \left\{ j \in G_2 \middle| \sum_{i=1}^{m} \lambda_i^* \, r_{ij} < d^* - s^* \right\}, \qquad (2.2)$$

$$N_1 = G_1 - C_1, \qquad N_2 = G_2 - C_2.$$

Then, we determine that observations in C_1 belong to G_1 and the observations of C_2 belong to G_2 because their location is identified from model (2.1). The two subsets N_1 and N_2 consist of the observations have not yet been classified in the first stage.

Stage 2. If $s^* \geq 0$, then the existence of an overlap is identified in the fist stage. In this stage, we reclassify all of the observations belonging to the overlap ($N_1 \cup N_2$) because the group membership of these observations is still undetermined. The second stage is reformulated as follows [20]:

$$\text{Minimize} \quad \sum_{j \in D_1} l_j + w \sum_{j \in D_2} l_j,$$

$$\text{Subject to} \quad \sum_{i=1}^{m} (\lambda_i^+ - \lambda_i^-) r_{ij} - c + M l_j \geq 0, \quad j \in N_1,$$

$$\sum_{i=1}^{m} (\lambda_i^+ - \lambda_i^-) r_{ij} - c - M l_j \leq -\varepsilon, \quad j \in N_2,$$

$$\sum_{i=1}^{m} (\lambda_i^+ + \lambda_i^-) = 1, \qquad (2.3)$$

$$\xi_i^+ \geq \lambda_i^+ \geq \varepsilon \xi_i^+, \quad \xi_i^- \geq \lambda_i^- \geq \varepsilon \xi_i^-, \quad i = 1, \ldots, m,$$

$$\xi_i^+ + \xi_i^- \leq 1, \quad i = 1, \ldots, m,$$

$$\lambda_i^+ + \lambda_i^- \geq \varepsilon, \quad \text{for some } i,$$

$$c : \text{unrestricted}, \ \xi_i^+, \xi_i^-, y_j \in \{0, 1\}, \ \text{and all other variables} \geq 0.$$

Here, the binary variable (l_j) counts the number of observations classified incorrectly. The objective function minimizes the number of such misclassifications. The weight (w) identifies the importance between G_1 and G_2 in terms of the number of observations. In the presented model (2.3), it is necessary to prescribe a large number (M) and a small number (ε). The equation $\lambda_i^+ + \lambda_i^- \geq \varepsilon$ (for some i) indicates that some pairs avoid the occurrence of $\lambda_i^+ = 0$ and $\lambda_i^- = 0$.

After gaining an optimal solution on λ_i^* and c^*, the second stage classifies observations in the overlap as follows: if $\sum_{i=1}^{m} \lambda_i^* \, r_{ij} \geq c^*$, then the jth observation belongs to G_1, or if

$\sum_{i=1}^{m} \lambda_i^* \, r_{ij} \leq c^* - \varepsilon$, then it belongs to G_2. Thus, all of the observations in G are classified into G_1 or G_2 at the end of the second stage.

3. Sensitivity Analysis Based on Monte-Carlo Simulation (MCS)

Sensitivity analysis was created to deal simply with uncertainties in the input variables and model parameters [22]. The results of an sensitivity analysis can determine which of the input parameters have a more dominant influence on the uncertainty in the model output [23]. A variance-based sensitivity analysis, which addresses the inverse problem of attributing the output variance to uncertainty in the input, quantifies the contribution that each input factor makes to the variance in the output quantity of interest. A global sensitivity analysis of complex numerical models can be performed by calculating variance-based importance measures of the input variables, such as the Sobol' indices. These indices are calculated by evaluating a multidimensional integral using a Monte-Carlo technique. This approach allows analyzing the influence of different variables and their subsets, the structure of $f(x)$, and so forth.

It is assumed that a mathematical model having n input parameters gathered in an input vector with a joint probability density function (pdf) can be presented as a model function:

$$s = f(r), \quad r = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m) \in U^n \equiv [0;1]^n, \quad \text{that } r_{ij} \in U^n \equiv [0;1]^n \quad i = 1, \ldots, m; \ j = 1, \ldots, n, \tag{3.1}$$

where $\mathbf{r}_i = (r_{1i}, r_{2i}, \ldots, r_{ni})^T$. Because of the variables are affected by several kinds of heterogeneous uncertainties that reflect the imperfect knowledge of the system, it is assumed that input variables are independent and that the probability density function $p(r)$ is known, even if the r_{ij} are not actually random variables.

The Sobol' sensitivity method explores the multidimensional space of the unknown input parameters x with a certain number of MC samples. The sensitivity indices are generated by a decomposition of the model function f in an n-dimensional factor space into summands of increasing dimensionality [22]:

$$f(r) = f(\mathbf{r}_1, \ldots, \mathbf{r}_m) = f_0 + \sum_{i=1}^{m} \sum_{q_1 < \cdots < q_i} f_{q_1, \ldots, q_i}(\mathbf{r}_{q_1}, \mathbf{r}_{q_2}, \ldots, \mathbf{r}_{q_i}). \tag{3.2}$$

where the constant f_0 is the mean value of the function, and the integral of each summand over any of its independent variables is zero. Due to this property, the summands are orthogonal to each other in the following form:

$$\int_{U^n} f_{i_1, \ldots, i_t} f_{j_1, \ldots, j_q} dr = 0, \quad (i_1, \ldots, i_t) \neq (j_1, \ldots, j_q). \tag{3.3}$$

The sensitivity index, SI_i, represents the fractional contribution of a given factor $\mathbf{r}_i$ to the variance in a given output variable, s. To calculate the sensitivity indices, the total

variance, V_s, in the model output, s, is apportioned to all of the input factors, $\mathbf{r}_1, \ldots, \mathbf{r}_m$, as follows:

$$V_s = \int_{U^n} f^2(r)\,dr - f_0^2. \tag{3.4}$$

By integrating the square of (2.2) and with (2.3), it is possible to decompose the total variance (3.1) as follows [24]:

$$V_s = \sum_i V_i + \sum_{i<h} V_{ih} + \sum_{i<h<k} V_k + \cdots + V_{12\cdots o}, \tag{3.5}$$

where $V_i = V[E(s|\mathbf{r}_i = \mathbf{r}_i^*)]$,

$$V_{ih} = V\left[E\left(s\big|\mathbf{r}_i = \mathbf{r}_i^*, \mathbf{r}_h = \mathbf{r}_h^*\right)\right] - V_i - V_h, \tag{3.6}$$

and so on. $V[E(s \mid \mathbf{r}_i = \mathbf{r}_i^*)]$ is referred to as the variance of the conditional expectation and is the variance over all of the values of $\mathbf{r}_i^*$ in the expectation of s given that $\mathbf{r}_i$ has a fixed value $\mathbf{r}_i^*$. This is an intuitive measure of the sensitivity of s to a factor $\mathbf{r}_i$, as it measures the amount by which $E(s \mid \mathbf{r}_i = \mathbf{r}_i^*)$ varies with the value of $\mathbf{r}_i^*$ whilst averaging over the $\mathbf{r}_h$'s, $h \neq i$. Following the above definition for the partial variances, the sensitivity indices are defined as

$$\mathrm{SI}_i = \frac{V_i}{V_s} \tag{3.7}$$

Higher order indices can be calculated with a similar approach. With regard to (3.2), the decomposition of the sensitivity indices can be written in the following form:

$$\sum_{i=1}^{m} \mathrm{SI}_i + \sum_{1 \leq i \leq h \leq m} \mathrm{SI}_{ih} + \cdots + \mathrm{SI}_{1,2,\ldots,m} = 1. \tag{3.8}$$

The Sobol' indices are usually computed with a MC simulation. The mean value and total and partial variance can be derived with N_{sim} samples in the following [22]:

$$\widehat{f_0} = \frac{1}{N_{\mathrm{sim}}} \sum_{p=1}^{N_{\mathrm{sim}}} f(\mathbf{r}_p),$$

$$\widehat{V_s} = \frac{1}{N_{\mathrm{sim}}} \sum_{p=1}^{N_{\mathrm{sim}}} f^2(\mathbf{r}_p) - \widehat{f_0^2}, \tag{3.9}$$

$$\widehat{V_i} = \frac{1}{N_{\mathrm{sim}}} \sum_{p=1}^{N_{\mathrm{sim}}} f^2\left(\mathbf{r}_{(\sim i)p}^{(1)}, \mathbf{r}_{ip}^{(1)}\right) f\left(\mathbf{r}_{(\sim i)p}^{(2)}, \mathbf{r}_{ip}^{(1)}\right) - \widehat{f_0^2}.$$

In the later equations, $\mathbf{r}_m$ is a sampled variable in U^n, and

$$r_{(\sim i)p} = \left(r_{1p}, r_{2p}, \ldots, r_{(-1)p}, r_{(i+1)p}, \ldots, r_{mp} \right).$$ (3.10)

The superscripts (1) and (2) indicate that two different samples are generated and mixed.

4. Illustrative Examples

Classification methods are widely used in economic and finance. They are useful for classifying the sectors based upon their performance in different groups and predict the group memberships of new firms. Most of researchers used classification methods to classifying the firms based upon their performance assessment. DA is a classification method that is used in this study. The purpose of the first stage in DEA-DA is to determine whether there is an overlap between the two groups. The existence of an overlap is the main source of misclassification in DA. By identification of the overlap between two groups, it is possible to increase the number of observations classified correctly. If there is no overlap, any DA method may produce an almost perfect classification. However, if there is an overlap, an additional computation process is needed to deal with such an overlap [20]. So, there is a tradeoff between computational effort/time and a high level of classification capability.

Misclassification can result as a consequence of an intersection between two groups. Many researchers have proposed approaches that try to reveal the advantage of identifying the minimized overlap of two groups for risk management on the classification problem [19, 20, 25, 26]. Given the importance of the banking sector for, in general, the whole economy and, in particular, for the financial system, in this section, we present an application of the sensitivity analysis to overlaps, s, on data from a commercial bank of Iran. This assertion is illustrated numerically for bank branches that have more than 20 and 30 personnel, in two different examples.

If we wish to take into account the inherent randomness with respect to what the criteria might experience, we have to bring stochastic characterization into play. The stochastic efficiency assessment of banking branches normally requires performing a set of analyses on DMUs with a suit of variables as criteria.

At first, we use the additive model to discriminate banking branches. Most models need to examine both a DEA efficiency score and slacks or an efficiency score measured by them, which depends upon input-based or output-based measurement. The additive model [27] aggregates input-oriented and output-oriented measures to produce the efficiency score. Consequently, the efficiency status is more easily determined by the additive model than by the radial model.

In two different examples, the real data set consists of 78 and 18 banking branches. This study selects 31 and 8 branches as inefficient branches, and 47 and 10 branches as efficient branches, respectively, in example 1 and 2, documented in Tables 1 and 2. For determining the classifications based on the additive model, three variables of personnel, payable interest, and non-performing loans are considered as inputs. Nine variables of loans granted, long-term deposit, current deposit, non-benefit deposit, short-term deposit, received interest, and received fee are assumed as outputs.

Table 1

Personnel	Loans granted	Long-term deposits	Current deposits	Non-benefit deposit	Short-term deposits	Payable interest	Received interest	Received fee	Non-performing loans
					Efficient branches				
20	$311E+08$	$90E+08$	$62E+08$	$85E+08$	$71E+08$	$8E+08$	$48E+07$	$117E+06$	$386E+07$
39	$737E+08$	$135E+08$	$228E+08$	$74E+08$	$225E+08$	$14E+08$	$234E+07$	$152E+06$	$9E+07$
26	$377E+08$	$100E+08$	$126E+08$	$28E+08$	$154E+08$	$11E+08$	$158E+07$	$52E+06$	$44E+07$
20	$507E+08$	$172E+08$	$153E+08$	$42E+08$	$135E+08$	$16E+08$	$221E+07$	$59E+06$	$110E+07$
34	$538E+08$	$105E+08$	$173E+08$	$434E+08$	$158E+08$	$10E+08$	$22E+07$	$161E+06$	$1226E+07$
27	$1383E+08$	$246E+08$	$508E+08$	$92E+08$	$190E+08$	$23E+08$	$59E+07$	$145E+06$	$221E+07$
28	$1740E+08$	$590E+08$	$253E+08$	$86E+08$	$242E+08$	$47E+08$	$445E+07$	$294E+06$	$1075E+07$
28	$960E+08$	$100E+08$	$987E+08$	$120E+08$	$109E+08$	$95E+08$	$122E+07$	$515E+06$	$460E+07$
29	$1036E+08$	$225E+08$	$148E+08$	$118E+08$	$199E+08$	$18E+08$	$49E+07$	$186E+06$	$895E+07$
27	$1068E+08$	$97E+08$	$1066E+08$	$68E+08$	$125E+08$	$10E+08$	$123E+07$	$245E+06$	$3121E+07$
27	$1593E+08$	$109E+08$	$1977E+08$	$92E+08$	$103E+08$	$11E+08$	$42E+07$	$301E+06$	$1135E+07$
20	$734E+08$	$366E+08$	$142E+08$	$203E+08$	$350E+08$	$33E+08$	$40E+07$	$176E+06$	$537E+07$
24	$572E+08$	$189E+08$	$239E+08$	$87E+08$	$268E+08$	$23E+08$	$197E+07$	$628E+06$	$1530E+07$
22	$454E+08$	$168E+08$	$34E+08$	$64E+08$	$143E+08$	$17E+08$	$565E+07$	$102E+06$	$5130E+07$
27	$350E+08$	$144E+08$	$609E+08$	$95E+08$	$236E+08$	$17E+08$	$30E+07$	$423E+06$	$140E+07$
20	$268E+08$	$158E+08$	$111E+08$	$49E+08$	$163E+08$	$15E+08$	$16E+07$	$120E+06$	$100E+07$
21	$624E+08$	$78E+08$	$125E+08$	$98E+08$	$164E+08$	$97E+08$	$21E+07$	$193E+06$	$362E+07$
21	$288E+08$	$19E+08$	$1367E+08$	$29E+08$	$499E+08$	$27E+08$	$13E+07$	$265E+06$	$31E+07$
20	$86E+08$	$274E+08$	$314E+08$	$32E+08$	$210E+08$	$24E+08$	$31E+07$	$36E+06$	$26E+07$
25	$126E+08$	$235E+08$	$163E+08$	$104E+08$	$477E+08$	$27E+08$	$14E+07$	$75E+06$	$37E+07$
20	$137E+08$	$162E+08$	$169E+08$	$27E+08$	$208E+08$	$16E+08$	$14E+07$	$116E+06$	$409E+07$
25	$224E+08$	$356E+08$	$491E+08$	$107E+08$	$309E+08$	$32E+08$	$1E+07$	$160E+06$	$59E+07$
20	$369E+08$	$110E+08$	$131E+08$	$49E+08$	$190E+08$	$13E+08$	$24E+07$	$81E+06$	$306E+07$
21	$358E+08$	$241E+08$	$109E+08$	$129E+08$	$279E+08$	$22E+08$	$23E+07$	$94E+06$	$45E+07$
24	$108E+08$	$696E+08$	$46E+08$	$38E+08$	$389E+08$	$60E+08$	$12E+07$	$63E+06$	$5E+07$

Table 1: Continued.

Personnel	Loans granted	Long-term deposits	Current deposits	Non-benefit deposit	Short-term deposits	Payable interest	Received interest	Received fee	Non-performing loans
42	$339E+08$	$645E+08$	$157E+08$	$53E+08$	$238E+08$	$42E+08$	$28E+07$	$48E+06$	$667E+07$
23	$199E+08$	$307E+08$	$205E+08$	$31E+08$	$155E+08$	$26E+08$	$17E+07$	$162E+06$	$10E+07$
37	$914E+08$	$397E+08$	$241E+08$	$61E+08$	$298E+08$	$34E+08$	$116E+07$	$242E+06$	$4950E+07$
33	$1685E+08$	$153E+08$	$117E+08$	$166E+08$	$255E+08$	$17E+08$	$17E+07$	$40E+06$	$200E+07$
21	$634E+08$	$173E+08$	$125E+08$	$120E+08$	$287E+08$	$18E+08$	$44E+07$	$83E+06$	$270E+07$
34	$288E+08$	$490E+08$	$86E+08$	$121E+08$	$549E+08$	$47E+08$	$3E+07$	$151E+06$	$368E+07$
20	$235E+08$	$112E+08$	$159E+08$	$99E+08$	$489E+08$	$9E+08$	$12E+07$	$42E+06$	$622E+07$
24	$486E+08$	$101E+08$	$233E+08$	$208E+08$	$988E+08$	$9E+08$	$15E+07$	$107E+06$	$363E+07$
20	$1361E+08$	$1194E+08$	$354E+08$	$78E+08$	$346E+08$	$94E+08$	$82E+07$	$242E+06$	$69E+07$
23	$712E+08$	$369E+08$	$102E+08$	$43E+08$	$106E+08$	$29E+08$	$43E+07$	$81E+06$	$190E+07$
28	$1574E+08$	$249E+08$	$204E+08$	$86E+08$	$132E+08$	$21E+08$	$64E+07$	$192E+06$	$1963E+07$
36	$2163E+08$	$323E+08$	$781E+08$	$132E+08$	$184E+08$	$27E+08$	$394E+07$	$323E+06$	$995E+07$
21	$352E+08$	$231E+08$	$108E+08$	$93E+08$	$107E+08$	$17E+08$	$20E+07$	$64E+06$	$261E+07$
22	$255E+08$	$558E+08$	$188E+08$	$46E+08$	$178E+08$	$41E+08$	$48E+07$	$68E+06$	$551E+07$
20	$191E+08$	$171E+08$	$90E+08$	$88E+08$	$159E+08$	$13E+08$	$27E+07$	$104E+06$	$250E+07$
27	$1373E+08$	$1002E+08$	$393E+08$	$45E+08$	$597E+08$	$76E+08$	$269E+07$	$482E+06$	$5347E+07$
23	$542E+08$	$538E+08$	$144E+08$	$15E+08$	$238E+08$	$39E+08$	$2E+07$	$113E+06$	$918E+07$
21	$1454E+08$	$431E+08$	$336E+08$	$34E+08$	$563E+08$	$41E+08$	$312E+07$	$124E+06$	$20E+07$
24	$1141E+08$	$723E+08$	$367E+08$	$28E+08$	$558E+08$	$56E+08$	$32E+07$	$193E+06$	$81E+07$
21	$1796E+08$	$1188E+08$	$416E+08$	$38E+08$	$801E+08$	$93E+08$	$442E+07$	$726E+06$	$103E+07$
26	$361E+08$	$223E+08$	$236E+08$	$75E+08$	$441E+08$	$21E+08$	$8E+07$	$58E+06$	$201E+07$
24	$1367E+08$	$730E+08$	$129E+08$	$28E+08$	$105E+08$	$55E+08$	$213E+07$	$538E+06$	$496E+07$
Inefficient branches									
33	$1088E+08$	$509E+08$	$283E+08$	$50E+08$	$183E+08$	$49E+08$	$172E+07$	$95E+06$	$2198E+07$
39	$744E+08$	$282E+08$	$248E+08$	$341E+08$	$282E+08$	$25E+08$	$69E+07$	$188E+06$	$1373E+07$
31	$1723E+08$	$565E+08$	$313E+08$	$55E+08$	$481E+08$	$61E+08$	$11E+07$	$578E+06$	$523E+07$
29	$1020E+08$	$196E+08$	$255E+08$	$120E+08$	$171E+08$	$18E+08$	$15E+07$	$164E+06$	$713E+07$

Table 1: Continued.

Personnel	Loans granted	Long-term deposits	Current deposits	Non-benefit deposit	Short-term deposits	Payable interest	Received interest	Received fee	Non-performing loans
21	$417E+08$	$194E+08$	$90E+08$	$82E+08$	$180E+08$	$17E+08$	$18E+07$	$129E+06$	$773E+07$
29	$757E+08$	$151E+08$	$218E+08$	$288E+08$	$204E+08$	$14E+08$	$34E+07$	$101E+06$	$1599E+07$
34	$461E+08$	$319E+08$	$183E+08$	$64E+08$	$309E+08$	$29E+08$	$65E+07$	$199E+06$	$144E+07$
25	$461E+08$	$120E+08$	$184E+08$	$70E+08$	$168E+08$	$12E+08$	$4E+07$	$114E+06$	$699E+07$
29	$1519E+08$	$521E+08$	$382E+08$	$80E+08$	$215E+08$	$44E+08$	$61E+07$	$236E+06$	$1960E+07$
22	$226E+08$	$245E+08$	$821E+08$	$90E+08$	$188E+08$	$21E+08$	$5E+07$	$73E+06$	$172E+07$
24	$769E+08$	$258E+08$	$72E+08$	$28E+08$	$182E+08$	$22E+08$	$6E+07$	$56E+06$	$446E+07$
24	$629E+08$	$310E+08$	$289E+08$	$28E+08$	$439E+08$	$33E+08$	$85E+07$	$291E+06$	$217E+07$
30	$67E+08$	$505E+08$	$201E+08$	$51E+08$	$464E+08$	$49E+08$	$53E+07$	$184E+06$	$94E+07$
21	$1182E+08$	$428E+08$	$298E+08$	$63E+08$	$452E+08$	$48E+08$	$132E+07$	$66E+06$	$197E+07$
32	$331E+08$	$331E+08$	$122E+08$	$67E+08$	$378E+08$	$33E+08$	$30E+07$	$105E+06$	$424E+07$
36	$376E+08$	$306E+08$	$102E+08$	$126E+08$	$202E+08$	$26E+08$	$18E+07$	$138E+06$	$377E+07$
25	$275E+08$	$355E+08$	$333E+08$	$115E+08$	$475E+08$	$40E+08$	$62E+07$	$150E+06$	$3097E+07$
24	$367E+08$	$218E+08$	$66E+08$	$117E+08$	$227E+08$	$21E+08$	$59E+07$	$49E+06$	$59E+07$
31	$257E+08$	$147E+08$	$380E+08$	$74E+08$	$237E+08$	$16E+08$	$22E+07$	$548E+06$	$182E+07$
20	$170E+08$	$220E+08$	$74E+08$	$30E+08$	$198E+08$	$21E+08$	$7E+07$	$63E+06$	$329E+07$
27	$452E+08$	$156E+08$	$323E+08$	$62E+08$	$314E+08$	$17E+08$	$23E+07$	$90E+06$	$768E+07$
27	$406E+08$	$227E+08$	$176E+08$	$110E+08$	$251E+08$	$22E+08$	$22E+07$	$119E+06$	$237E+07$
26	$334E+08$	$166E+08$	$78E+08$	$80E+08$	$146E+08$	$14E+08$	$10E+07$	$49E+06$	$246E+07$
34	$381E+08$	$218E+08$	$469E+08$	$129E+08$	$228E+08$	$21E+08$	$10E+07$	$250E+06$	$1167E+07$
28	$565E+08$	$133E+08$	$186E+08$	$225E+08$	$156E+08$	$12E+08$	$11E+07$	$168E+06$	$530E+07$
40	$675E+08$	$382E+08$	$121E+08$	$52E+08$	$109E+08$	$28E+08$	$208E+07$	$106E+06$	$1694E+07$
20	$313E+08$	$467E+08$	$98E+08$	$31E+08$	$199E+08$	$40E+08$	$73E+07$	$30E+06$	$133E+07$
24	$121E+08$	$849E+08$	$154E+08$	$23E+08$	$455E+08$	$73E+08$	$31E+07$	$24E+06$	$682E+07$
25	$327E+08$	$109E+08$	$237E+08$	$46E+08$	$196E+08$	$13E+08$	$44E+07$	$251E+06$	$547E+07$
24	$1001E+08$	$672E+08$	$444E+08$	$81E+08$	$419E+08$	$56E+08$	$133E+07$	$585E+06$	$1007E+07$
34	$482E+08$	$192E+08$	$173E+08$	$53E+08$	$199E+08$	$17E+08$	$94E+07$	$87E+06$	$252E+07$

Table 2

| Personnel	Loans granted	Long-term deposits	Current deposits	Non-benefit deposit	Short-term deposits	Payable interest	Received interest	Received fee	Non-performing loans
					Efficient branches				
34	$538E+08$	$105E+08$	$173E+08$	$434E+08$	$158E+08$	$10E+08$	$22E+07$	$161E+06$	$1226E+07$
42	$339E+08$	$645E+08$	$157E+08$	$53E+08$	$238E+08$	$42E+08$	$28E+07$	$48E+06$	$667E+07$
37	$914E+08$	$397E+08$	$241E+08$	$61E+08$	$298E+08$	$34E+08$	$116E+07$	$242E+06$	$4950E+07$
34	$288E+08$	$490E+08$	$86E+08$	$121E+08$	$549E+08$	$47E+08$	$3E+07$	$151E+06$	$368E+07$
36	$2163E+08$	$323E+08$	$781E+08$	$132E+08$	$184E+08$	$27E+08$	$394E+07$	$323E+06$	$995E+07$
39	$744E+08$	$282E+08$	$248E+08$	$341E+08$	$282E+08$	$25E+08$	$69E+07$	$188E+06$	$1373E+07$
31	$1723E+08$	$565E+08$	$313E+08$	$55E+08$	$481E+08$	$61E+08$	$11E+07$	$578E+06$	$523E+07$
30	$67E+08$	$505E+08$	$201E+08$	$51E+08$	$464E+08$	$49E+08$	$53E+07$	$184E+06$	$94E+07$
31	$257E+08$	$147E+08$	$380E+08$	$74E+08$	$237E+08$	$16E+08$	$22E+07$	$548E+06$	$182E+07$
34	$381E+08$	$218E+08$	$469E+08$	$129E+08$	$228E+08$	$21E+08$	$10E+07$	$250E+06$	$1167E+07$
					Inefficient branches				
33	$1088E+08$	$509E+08$	$283E+08$	$50E+08$	$183E+08$	$49E+08$	$172E+07$	$95E+06$	$2198E+07$
34	$461E+08$	$319E+08$	$183E+08$	$64E+08$	$309E+08$	$29E+08$	$65E+07$	$199E+06$	$144E+07$
32	$331E+08$	$331E+08$	$122E+08$	$67E+08$	$378E+08$	$33E+08$	$30E+07$	$105E+06$	$424E+07$
36	$376E+08$	$306E+08$	$102E+08$	$126E+08$	$202E+08$	$26E+08$	$18E+07$	$138E+06$	$377E+07$
40	$675E+08$	$382E+08$	$121E+08$	$52E+08$	$109E+08$	$28E+08$	$208E+07$	$106E+06$	$1694E+07$
34	$482E+08$	$192E+08$	$173E+08$	$53E+08$	$199E+08$	$17E+08$	$94E+07$	$87E+06$	$252E+07$
39	$737E+08$	$135E+08$	$228E+08$	$74E+08$	$225E+08$	$14E+08$	$234E+07$	$152E+06$	$9E+07$
33	$1685E+08$	$153E+08$	$117E+08$	$166E+08$	$255E+08$	$17E+08$	$17E+07$	$40E+06$	$200E+07$
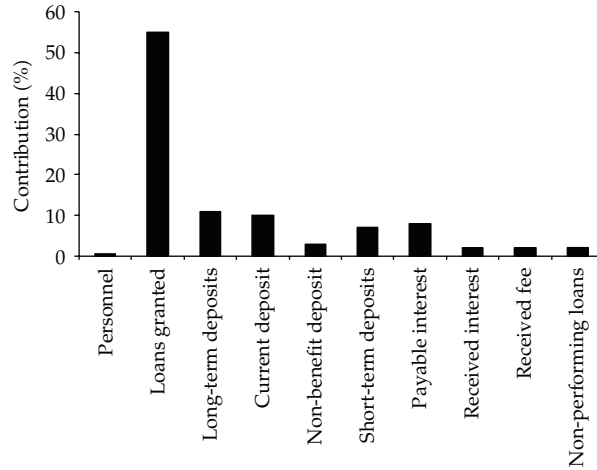
Figure 1: The relative contributions of different variables on the overlap in DEA-DA in example 1 that the real data set consists of 78 banking branches.

Then, for sensitivity analysis in DEA-DA, each observation is modeled as random parameters as follows [28]:

$$r_{ij} = \mu_{r_{ij}}\left(1 + V_{r_{ij}}\alpha_l\right), \tag{4.1}$$

where $\mu_{r_{ij}}$ and $V_{r_{ij}}$ are the mean value and the coefficients of variation (COV) of the random parameters, respectively, and α_l is generated random parameter with a zero mean that is used in MC simulation. The determination of the bank branches' parameters carries a high degree of uncertainty, and the specification of these parameters can involve a significant degree of expert judgment. Additionally, the COV of these variables plays an important role in the variation of the efficiency. Here, the COV of the all of the parameters are assumed to equal 0.05.

To compute the sensitivity indices, the Sobol' sampling scheme has been used. Sequences of Sobol' sampling vectors are essentially quasi-random sequences that are defined as sequences of points that have no intrinsic random properties. In this study, a sensitivity analysis is applied to assess the influence of the banks' variables on the overlap.

After one hundred estimates with a sample size of 5000, convergence was seen in the first-order Sobol' indices derived by a Sobol' sampling of the uniform criteria spaces for different banks' variables. The sensitivity indices, S_i, are depicted in Figures 1 and 2. These figures present the comparison of the first-order indices of the banks' variables. The total fraction of the variance captured by the first order functions is approximately 99%. This indicates that, for this problem, higher order contributions to the Sobol' series are relatively small. The overall variance in banks' efficiency is affected by the variances in each of the random variables. Figure 1 indicates that 54% and 33% of the overall variance in overlap in DEA-DA is attributable to the variance in loans granted and different deposits, respectively, while the personnel, received interest, fee, and non-performing loans variables have little effect. Also, Figure 2 indicates that uncertainties in loans granted and different deposits are the dominant variables in overlap in DEA-DA.
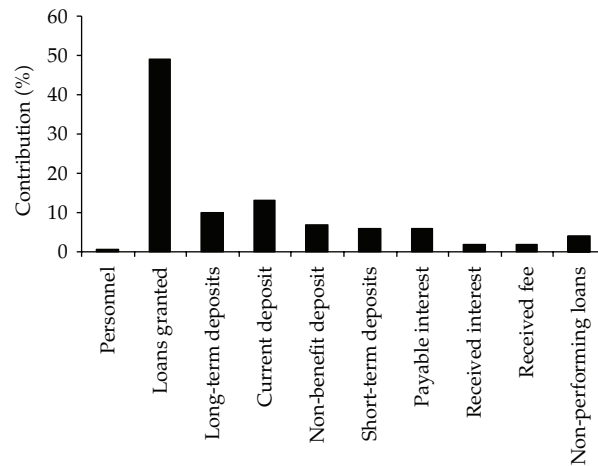
Figure 2: The relative contributions of different variables on the overlap in DEA-DA in example 1 that the real data set consists of 18 banking branches.

5. Conclusions

Due to the inherent complexity and randomness of the data in DEA and problems involving unpredictable or stochastic variables, a probabilistic analysis may be the most rational method of analysis. Therefore, in a probabilistic-based approach, the results open the door to understanding the appropriate estimation of the deciding variables in DEA. In the overlap in DA, the analytical results show that the loans granted and different deposit variables are the main sources of uncertainty, while other variables have a relatively small effect. The main advantage of the used sensitivity analysis approach is that it provides quantified evaluation of the influence of individual variables in DA, and results may be used for decision making.

Acknowledgment

References

[1] T. Sueyoshi and M. Goto, "Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA-discriminant analysis," *European Journal of Operational Research*, vol. 196, no. 1, pp. 289–311, 2009.

[2] O. L. Mangasarian, "Linear and nonlinear separation of patterns by linear programming," *Operations Research*, vol. 13, pp. 444–452, 1965.

[3] N. Freed and F. Glover, "A linear programming approach to the discriminant problem," *Decision Sciences*, vol. 12, pp. 68–74, 1981.

[4] D. J. Hand, *Discrimination and Classification*, John Wiley & Sons, Chichester, UK, 1981.

[5] N. Freed and F. Glover, "Simple but powerful goal programming models for discriminant problems," *European Journal of Operational Research*, vol. 7, no. 1, pp. 44–60, 1981.

[6] N. Freed and F. Glover, "Evaluating alternative linear programming models to solve the two-group discriminant problem," *Decision Sciences*, vol. 17, pp. 151–162, 1986.

[7] W. J. Banks, Abad, and P. L, "An efficient optimal solution algorithm for the classification problem," *Decision Sciences*, vol. 22, pp. 1008–1023, 1991.

[8] F. Glover, "Improved linear programming models for discriminant analysis," *Decision Sciences*, vol. 21, pp. 771–785, 1990.

[9] D. L. Retzlaff-Roberts, "A ratio model for discriminant analysis using linear programming," *European Journal of Operational Research*, vol. 94, no. 1, pp. 112–121, 1996.

[10] E. A. Joachimsthaler and E. A. Stain, "Four approaches to the classification problem in discriminant analysis: an experimental study," *Decision Sciences*, vol. 19, pp. 322–333, 1988.

[11] D. L. Retzlaff-Roberts, "Relating discriminant analysis and data envelopment analysis to one another," *Computers and Operations Research*, vol. 23, no. 4, pp. 311–322, 1996.

[12] D. L. Retzla-Roberts, "A ratio model for discriminant analysis using linear programming," *European Journal of Operational Research*, vol. 94, pp. 112–121, 1996.

[13] C. Tofallis, "Improving discernment in DEA using profiling," *Omega*, vol. 24, no. 3, pp. 361–364, 1996.

[14] T. Sueyoshi, "DEA-discriminant analysis in the view of goal programming," *European Journal of Operational Research*, vol. 115, no. 3, pp. 564–582, 1999.

[15] T. Sueyoshi, "Extended DEA-discriminant analysis," *European Journal of Operational Research*, vol. 131, no. 2, pp. 324–351, 2001.

[16] S. M. Bajgier and A. V. Hill, "An experimental comparison of statistical and linear programming approaches to the discriminant problem," *Decision Sciences*, vol. 13, pp. 604–618, 1982.

[17] W. V. Gehrlein, "General mathematical programming formulations for the statistical classification problem," *Operations Research Letters*, vol. 5, no. 6, pp. 299–304, 1986.

[18] J. M. Wilson, "Integer programming formulations of statistical classification problems," *Omega*, vol. 24, no. 6, pp. 681–688, 1996.

[19] D. S. Chang and Y. C. Kuo, "An approach for the two-group discriminant analysis: an application of DEA," *Mathematical and Computer Modelling*, vol. 47, no. 9-10, pp. 970–981, 2008.

[20] T. Sueyoshi, "Mixed integer programming approach of extended DEA-discriminant analysis," *European Journal of Operational Research*, vol. 152, no. 1, pp. 45–55, 2004.

[21] T. Sueyoshi, "Financial ratio analysis of electric power industry," *Asia-Pacific Journal of Operational Research*, vol. 22, pp. 349–376, 2005.

[22] A. Saltelli, K. Chan, and E.M. Scott, *Sensitivity Analysis*, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, UK, 2000.

[23] A. Saltelli, M. Ratto, T. Andress, F. Campolongo, J. Cariboni, and F. Gatelli, *Global Sensitivity Analysis Guiding the Worth of Scientific Models*, John Wiley and Sons, New York, NY, USA, 2007.

[24] I. M. Sobol', "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modeling and Computational Experiment*, vol. 1, no. 4, pp. 496–515, 1993.

[25] J. J. Glen, "A comparison of standard and two-stage mathematical programming discriminant analysis methods," *European Journal of Operational Research*, vol. 171, no. 2, pp. 496–515, 2006.

[26] P. C. Pendharkar, "A hybrid radial basis function and data envelopment analysis neural network for classification," *Computers & Operations Research*, vol. 38, no. 1, pp. 256–266, 2011.

[27] A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics*, vol. 30, no. 1-2, pp. 91–107, 1985.

[28] A. Yazdani and T. Takada, "Probabilistic study of the influence of ground motion variables on response spectra," *Structural Engineering & Mechanics*, vol. 39, 2011.