Ann Oper Res (2014) 217:535–563
DOI 10.1007/s10479-014-1565-y

Quantile–DEA classifiers with interval data

Quanling Wei · Tsung-Sheng Chang · Song Han

Published online: 2 March 2014

Abstract This research intends to develop the classifiers for dealing with binary classification problems with interval data whose difficulty to be tackled has been well recognized, regardless of the field. The proposed classifiers involve using the ideas and techniques of both quantiles and data envelopment analysis (DEA), and are thus referred to as quantile–DEA classifiers. That is, the classifiers first use the concept of quantiles to generate a desired number of exact-data sets from a training-data set comprising interval data. Then, the classifiers adopt the concept and technique of an intersection-form production possibility set in the DEA framework to construct acceptance domains with each corresponding to an exact-data set and thus a quantile. Here, an intersection-form acceptance domain is actually represented by a linear inequality system, which enables the quantile–DEA classifiers to efficiently discover the groups to which large volumes of data belong. In addition, the quantile feature enables the proposed classifiers not only to help reveal patterns, but also to tell the user the value or significance of these patterns.

Keywords Data envelopment analysis · Classifier · Quantile · Production possibility set · Interval data

1 Introduction

Lack of data used to be a big challenging problem faced by large numbers of researchers and practitioners in a great many domains. However, nowadays, they quite often face a

Q. Wei · S. Han
Institute of Operations Research and Mathematical Economics, Renmin University of China, Beijing, China

T.-S. Chang (✉)
Department of Transportation and Logistics Management, National Chiao Tung University, 1001 University Rd., Hsinchu 30010, Taiwan
e-mail: tsc@nctu.edu.tw

new challenge of generating useful information from an explosive growth of data volume due to the advanced technologies for generating and collecting data. Therefore, it has become necessary to develop new technologies and tools to intelligently and rapidly process massive amounts of data into useful information and knowledge, which has resulted in data mining becoming increasingly important and receiving much attention in many fields. Data mining involves the use of sophisticated data analysis tools such as statistical models, mathematical algorithms, and machine learning methods to search for valuable information in large volumes of data (Seifert 2004). To date, data mining has been used for a variety of purposes in both the private and public sectors such as association, sequence or path analysis, classification, clustering, and forecasting (see, e.g., Han and Kamber 2007). In addition, as pointed out by Seifert (2004), data mining has not only been used by different industries, e.g., banking, insurance, medicine, and retailing, to reduce costs, enhance research, and increase sales, but has also been used in the public sector to detect fraud and waste, and measure and improve program performance. However, there are some limitations to the capability of data mining. For instance, data mining helps reveal patterns and relationships, but it does not tell the user the value or significance of these patterns; the user, therefore, still needs specialists to interpret the created output (Seifert 2004). That is, more user-friendly and efficient data mining tools are called for. It is noteworthy that operations research methods have been shown to be promising for improving data mining techniques (Corne et al. 2012).

As indicated above, data mining includes several main functions. This research focuses on the function of classification that is to judge whether a piece of data belongs to a particular group by evaluating a set of characteristic values. Of particular interest in this study are binary classification problems, which are also referred to as two-group discriminant analysis problems. The problems, in which a piece of data belongs to one of two groups (more specifically, inside or outside an acceptance domain), has been applied to a wide variety of fields such as economics, finance, insurance and risk for credit scoring, bankruptcy prediction, insurance underwriting, management fraud detection and so on (Sinha and Zhao 2008). Up to now, there have been a few popular algorithms for data classification such as decision tree induction, Bayesian classification, rule-based classification and support vector machines (see, e.g., Han and Kamber 2007). In addition, data envelopment analysis-(DEA-)based approaches seem to have been receiving attention very recently in academia; the relevant works, which show that DEA-based methods are quite promising in practice, are reviewed as follows. It is well known that the function of conventional DEA theories, models and methods is to evaluate the relative efficiency among a given number of decision making units (DMUs) with multiple inputs and multiple outputs (Cooper et al. 2006). Nonetheless, Troutt et al. (1996) have pioneered the use of DEA models in binary classification (more precisely, by developing an acceptance boundary); they propose a sample-based decision system to make a decision on whether or not to accept or reject a credit risk based on samples predetermined by experts. Seiford and Zhu (1998) extend the work to develop a DEA-type linear programming model to decide whether a new case is acceptable; the model also determines the location of the case corresponding to the previously classified samples. Pendharkar et al. (2000) apply the method of Troutt et al. to discover the breast cancer pattern; their empirical results show that the DEA-based approach outperforms statistically linear discriminant analysis. Instead of determining the acceptability of a new case such as in Troutt et al. (1996) and Seiford and Zhu (1998), Pendharkar (2002) deals with an inverse classification problem where the objective is to find out how to change a DMU's inputs so that it can be classified into another class. Pendharkar (2011) integrates the DEA model with the radial basis function network (RBFN) to develop a hybrid RBFN–DEA neural network for a binary classification problem with negative inputs and non-linearly separable classes. Pendharkar (2012) uses DEA for fuzzy classification where the classification output is a fuzzy membership function. He empirically shows that his DEA-based fuzzy classification system outperforms the adaptive neuro fuzzy inference system, fuzzy rule-based classification system and logistic regression. However, it is noted that the research deals with exact data instead of fuzzy or interval data. Finally, Yan and Wei (2011) propose a DEA classification machine with exact data that includes an acceptance domain (i.e., a production possibility set under the DEA framework) and a classification function. The acceptance domain is constructed by an explicit system of linear inequalities, which makes the classification process very efficient.

To our knowledge, the DEA-based methods for data classification proposed in the literature assume that data are measured by exact values. However, as pointed out in Cooper et al. (1999), it is quite common in many applications for some data to be known only within specified intervals while other data may be known only in terms of ordinal relations; such a data type is commonly referred to as "imprecise data." So far, there have been a few works that incorporate imprecise data into DEA models in the literature (see, e.g., Cooper et al. 1999; Despotis and Smirlis 2002; Zhu 2003; Kao 2006). It is important to note that the DEA models dealing with imprecise data in the literature are used to evaluate the relative efficiency among DMUs, but not to perform data classification. That is, there is no research, to our knowledge, working on developing DEA-based methods for classifying imprecise data so far in the literature. Therefore, this research intends to propose DEA classifiers to deal with binary classification problems with classification data that are known to have either exact values or values only within bounded intervals; note that it is well recognized, regardless of the field, that it is difficult to tackle interval data.

The proposed classifiers involve using the concepts and techniques of both quantiles and DEA, and are thus referred to as quantile–DEA classifiers. Here we briefly introduce the proposed classifiers, and elaborate on the classifiers in the succeeding sections. It is well known that the conventional way for dealing with a binary classification problem with exact data is to construct the corresponding acceptance domain such that classification data are either inside (i.e., accepted by) or outside (i.e., rejected by) the acceptance domain. A classifier embedded with a single acceptance domain while efficient is usually unable to provide the user with the degree of acceptance or rejection. In addition, due to the inherent complexity of interval data, it seems to be necessary to construct the classifiers that are embedded with multiple acceptance domains for handling binary classification problems with interval data. Therefore, we adopt the idea of quantiles in statistics to tackle the issue of multiple acceptance domains. That is, given n training interval data, we create n exact data by specifying a quantile for each of the original interval data; it follows that, by specifying t quantiles, we can obtain t exact-data sets with each containing n exact data. Then, we construct t acceptance domains with each corresponding to one of the t exact-data sets by applying the concept and technique of a production possibility set in the DEA framework. Clearly, whether or not a classifier with multiple acceptance domains is efficient and effective largely depends on the construction and presentation of the acceptance domains. In this research, we first use the idea and technique of a conventional sum-form production possibility set to construct the acceptance domains. Then, we adopt the techniques proposed in Wei and Yan (2001) and Yan and Wei (2000) to transform the sum-form acceptance domains into the ones with an intersection form that is a linear inequality system. The intersection-form acceptance domain enables the quantile–DEA classifiers to efficiently discover the groups to which a huge amount of classification data belong. In addition, the feature of quantiles makes the proposed classifiers not only able to reveal patterns, but also able to tell the user the value or significance of these patterns.

The remainder of this paper is organized as follows: Sect. 2 shows how the acceptance domains are constructed; Sect. 3 introduces the quantile–DEA classifier for dealing with exact data; Sect. 4 elaborates the quantile–DEA classifier for tackling interval data; Sect. 5 extends the classifiers proposed in Sects. 3 and 4 to deal with general cases; and, finally, Sect. 6 concludes the paper.

2 Construction of acceptance domains

The technique of acceptance domains lies at the heart of the proposed classifiers. Hence, this section introduces how we construct the acceptance domains by using n training interval data. Assume that each of the training interval data denoted as interval DMU-$\bar{x}_j$ ($j = 1, \ldots, n$) is associated with m inputs. We consider the setting where all training data are known only within specified bounds with values that are drawn from uniform distributions. Denote the interval-DMUs as

$$\bar{x}_j = \left(\bar{x}_{1j}, \bar{x}_{2j}, \ldots, \bar{x}_{mj}\right), \quad j = 1, \ldots, n,$$

where

$$\bar{x}_{ij} \in \left[a_{ij}, b_{ij}\right], \quad i = 1, \ldots, m; \quad j = 1, \ldots, n.$$

That is,

$$\bar{x}_j = \left(\left[a_{1j}, b_{1j}\right], \left[a_{2j}, b_{2j}\right], \ldots, \left[a_{mj}, b_{mj}\right]\right), \quad j = 1, \ldots, n.$$

Define

$$a_j = \left(a_{1j}, a_{2j}, \ldots, a_{mj}\right)^T, \quad j = 1, \ldots, n,$$
$$b_j = \left(b_{1j}, b_{2j}, \ldots, b_{mj}\right)^T, \quad j = 1, \ldots, n.$$

Furthermore, denote the training data set as

$$\bar{T} = \left\{\bar{x}_j | j = 1, \ldots, n\right\}.$$

Moreover, define

$$\bar{x}_j^{\beta} = \left(\bar{x}_{1j}^{\beta}, \bar{x}_{2j}^{\beta}, \ldots, \bar{x}_{mj}^{\beta}\right), \quad j = 1, \ldots, n,$$

where

$$\bar{x}_{ij}^{\beta} = a_{ij} + \beta\left(b_{ij} - a_{ij}\right) > 0, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n,$$

and $\beta \in (L, \ +\infty)$ with

$$L = \max_{1 \le i \le m; 1 \le j \le n}\left(\frac{-a_{ij}}{b_{ij} - a_{ij}}\right).$$

It is noted that, since the values of all training data are uniformly distributed within specified bounds, then, theoretically, $\beta \in [0, 1]$. However, it is quite common in practice that some of the classification data are positioned outside the range formed by the

n training interval data. It follows that the acceptance domains formed by the n training interval data based on the condition that $\beta \in [0, 1]$ cannot classify those classification data. Hence, we need to extend the value of β to construct the acceptance domains that can handle the classification data with values spreading outside the range formed by the n training interval data. In actual fact, since $\bar{x}_{ij}^{\beta} > 0\,\forall i, j$, we can derive the lower bound of β, i.e., L, which is obviously less than 0; however, there is no systematic way to confine β from above, and thus the upper bound of β is defined as $+\infty$. The purpose and usefulness of extending the value of β will become clearer later.

In addition, we make the following assumptions for formally defining the considered binary classification problem. First, all training data $\bar{x}_j, j = 1, \ldots, n$ in training data set $\bar{T}$ are accepted to, however, different degrees. Second, the bigger the values of the data, the higher the probabilities that the data will be accepted, which is referred to as the conditional monotonicity/non-satiety assumption by Pendharkar and Troutt (2011). We will relax this assumption to deal with more general cases in Sect. 5. Third, $\bar{x}_{ij}^{\beta} > 0, i = 1, \ldots, m; j = 1, \ldots, n$ for practical applications; it follows that $\beta > L$. Fourth, if two data are accepted (rejected), then a data that can be represented by the convex combination of the two data is also accepted (rejected). As a result, let T_β represent the acceptance domain constructed by $x_j^{\beta} = (x_{ij}^{\beta}, x_{2j}^{\beta}, \ldots, x_{mj}^{\beta})^T, j = 1, \ldots, n$ given a specified $\beta \in (L, +\infty)$ that satisfies the above assumptions. It is easy to check that T_β satisfies the following postulates (it is noted that an acceptance domain is uniquely determined by the system of postulates):

Postulate 1 (Ordinary postulate) the observed $x_j^{\beta} \in T_\beta$ for all $j = 1, \ldots, n$.

Postulate 2 (Convexity postulate) If $x \in T_\beta$, and $\hat{x} \in T_\beta$, then $\lambda x + (1 - \lambda)\hat{x} \in T_\beta$, for $\lambda \in [0, 1]$.

Postulate 3 (Monotonicity postulate) if $x \in T_\beta$, and $\hat{x} \geq x$, then $\hat{x} \in T_\beta$.

Postulate 4 (Minimum extrapolation postulate) T_β is the intersection set of all $\tilde{T}$ satisfying Postulates 1–3.

In actual fact, the acceptance domain that satisfies Postulates 1–4 defined above can be represented as follows:

$$T_\beta = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\beta}\lambda_j \leq x,\ \sum_{j=1}^{n} \lambda_j \geq 1,\ \lambda_j \geq 0, j = 1, \ldots, n \right\}.$$

Note that acceptance domain T_β has the same structure as the production possibility set corresponding to the classical CCR model (Charnes et al. 1978) with reference set $\{(x_j^{\beta}, 1) | j = 1, \ldots, n\}$ in DEA research. Hence, in this study, the boundary of T_β is, for convenience, also referred to as the frontier of T_β.

Definition 1 Let $\beta \in (L, +\infty)$ and

$$T_\beta = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\beta}\lambda_j \leq x,\ \sum_{j=1}^{n} \lambda_j \geq 1,\ \lambda_j \geq 0, j = 1, \ldots, n \right\}.$$

T_β is referred to as the *acceptance domain with β-quantile*.

Here, a piece of data on the frontier of T_β is regarded as acceptance if $\beta \geq 0$. The value of β is referred to as the acceptance degree; the larger the value of β, the higher the acceptance degree of the data. On the contrary, a piece of data on the frontier of T_β is considered to be a rejection if $L < \beta < 0$. The rationale for using β to represent the degree of acceptance is as follows. Recall that $\bar{x}_{ij}^{\beta} = a_{ij} + \beta(b_{ij} - a_{ij}) > 0$, $i = 1, \ldots, m$; $j = 1, \ldots, n$, and that the values within interval $[a_{ij}, b_{ij}]$ are drawn from uniform distributions. Therefore, for any $\beta \in [0, 1]$, $\beta = \Pr\{a_{ij} \le \hat{x}_{ij} \le \bar{x}_{ij}^{\beta}\} = \frac{\bar{x}_{ij}^{\beta} - a_{ij}}{b_{ij} - a_{ij}}$, which represents the probability that $\hat{x}_{ij}$ falls into interval $[a_{ij}, \bar{x}_{ij}^{\beta}]$. It follows that, here, β can be naturally used to represent the degree of acceptance. On the other hand, if $\beta \in (L, 0) \cup (1, +\infty)$, then β is not associated with the above probability property. However, based on the assumption that the bigger the values of the data, the higher the probability that the data are accepted (i.e., the monotonicity postulate), it is appropriate to extend the property of the $\beta \in [0, 1]$ to the $\beta \in (L, +\infty)$. That is, $\beta \in (0, +\infty)$ and $|\beta|$ such that $\beta \in (L, 0)$ can be used to represent the degrees of acceptance and rejection, respectively.

Theorem 1 *Let* $L < \bar{\beta} < \hat{\beta}$, *and*

$$T_{\bar{\beta}} = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\bar{\beta}} \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n \right\},$$

and

$$T_{\hat{\beta}} = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n \right\}.$$

Then,

(i) $T_{\hat{\beta}} \subset T_{\bar{\beta}}$;
(ii) *There is no intersection between the frontiers of* $T_{\hat{\beta}}$ *and* $T_{\bar{\beta}}$.

Proof See Appendices 1 and 4 for the proofs of (i) and (ii), respectively.

Example 1 Consider a sample training-data set of $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$ and $\bar{x}_4$ in which $m = 2$. Their corresponding characteristic values are as follows:

$$\bar{x}_1 = ([1, 6], [4, 7]), \quad \bar{x}_2 = ([2, 6], [2, 9]),$$
$$\bar{x}_3 = ([4, 10], [1, 4]), \quad \bar{x}_4 = ([4, 7], [4, 8]).$$

It is easy to obtain that $L = -0.2$, and

$$\begin{aligned} \bar{x}_{11}^{\beta} &= 1 + 5\beta; & \bar{x}_{21}^{\beta} &= 4 + 3\beta; \\ \bar{x}_{12}^{\beta} &= 2 + 4\beta; & \bar{x}_{22}^{\beta} &= 2 + 7\beta; \\ \bar{x}_{13}^{\beta} &= 4 + 6\beta; & \bar{x}_{23}^{\beta} &= 1 + 3\beta; \\ \bar{x}_{14}^{\beta} &= 4 + 3\beta; & \bar{x}_{24}^{\beta} &= 4 + 4\beta. \end{aligned}$$

The acceptance domain with β-quantile, T_β, is given as follows:

$$T_\beta = \left\{ x = (x_1, x_2) \,\middle|\, \left(\sum_{j=1}^{4} x_{1j}^{\beta} \lambda_j,\ \sum_{j=1}^{4} x_{2j}^{\beta} \lambda_j \right) \le (x_1, x_2),\ \sum_{j=1}^{4} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, 4 \right\}.$$

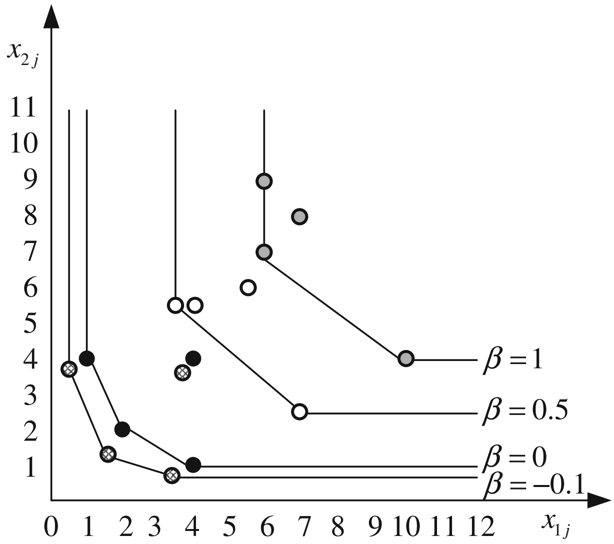


Fig. 1 T_β wrt different values of β

Let the values of β be -0.1, 0, 0.5 and 1; the corresponding $T_{-0.1}$, T_0, $T_{0.5}$ and T_1 are depicted in Fig. 1, which clearly shows that $T_{-0.1} \supset T_0 \supset T_{0.5} \supset T_1$, and that there is no intersection between each pair of frontiers.

3 Quantile–DEA classifier with exact data

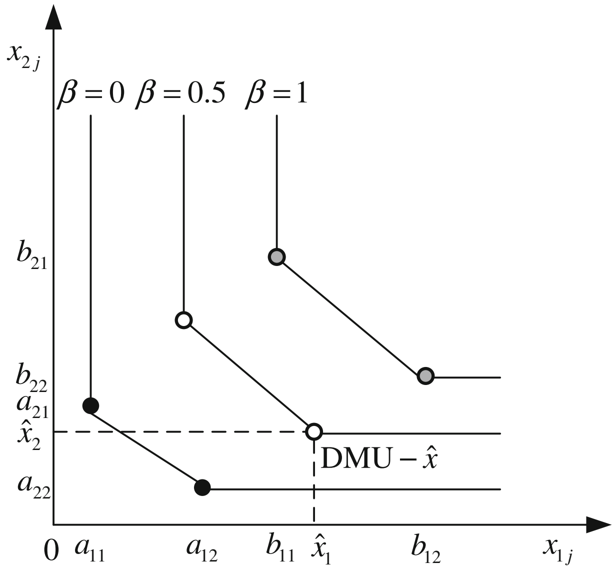
The focus of this study is on developing a quantile–DEA classifier for dealing with binary classification problems with interval data. Here, however, we first introduce the classifier for dealing with exact data for the following two reasons. First, we do not rule out the possibility that some classification data may have crisp values. Second, the methods for developing the classifier for handling exact data can be the building blocks for developing the one for tackling interval data. Nonetheless, it is important to note that the training data sets with respect to (wrt) both types of classifiers consist of only interval data.

Denote an exact data $\hat{x}$ as DMU-$\hat{x}$, and consider the following linear program wrt DMU-$\hat{x}$ (i.e., data $\hat{x} \in \hat{T}$, a set of classification data) with a specified $\beta \in (L, +\infty)$, where $\hat{T} \subset \Re^m_+$.

$$
(P_\beta) \quad \text{s.t.} \quad
\begin{aligned}
&\hat{\theta}(\beta) = \min \theta,\\
&\sum_{j=1}^{n} x_j^{\beta} \lambda_j \le \theta \hat{x},\\
&\sum_{j=1}^{n} \lambda_j \ge 1,\\
&\lambda_j \ge 0, \quad j = 1, \ldots, n.
\end{aligned}
$$

In actual fact, $\hat{\theta}(\beta)$, the optimal objective function value of problem (P_β) given a specified value of $\beta \in (L, +\infty)$, is a function of β. Here, the function is directly denoted

Fig. 2 Quantile function of DMU-$\hat{x}$



as $\hat{\theta}(\beta)$ and is referred to as the *quantile function* of DMU-$\hat{x}$. It is not difficult to show that $\hat{\theta}(\beta) > 0$, and that it is possible that $\hat{\theta}(\beta) > 1$ (see Lemma 2 and its proof in Appendix 2). Figure 2 demonstrates an example in which $0<\hat{\theta}(0)<1$ and $\hat{\theta}(1) > 1$. It is noted that, in Fig. 2, the acceptance domain between T_0 and T_1 (including the boundaries corresponding to T_0 and T_1) can be represented by the set $\{x|x \in T_0,\ x \notin \operatorname{Int} T_1\}$. The following theorem defines the properties of $\hat{\theta}(\beta)$.

Theorem 2 *Let* $\hat{x} \in \hat{T} \cap \operatorname{Int}\left\{x|\sum_{j=1}^{n} x_j^L \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j=1,\ldots,n\right\}$, and $\hat{\theta}(\beta)$ *be the quantile function of DMU-*$\hat{x}$*. Then,*

(i) $\hat{\theta}(\beta)$ *is a continuous function defined over* $(L, +\infty)$.
(ii) $\hat{\theta}(\beta)$ *is a strictly monotonically decreasing function over* $(L,\ +\infty)$.

Proof See Appendix 2. □

Definition 2 Let $\hat{x} \in \hat{T} \cap \operatorname{Int}\{x|\sum_{j=1}^{n} x_j^L \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j=1,\ldots,n\}$, and $\hat{\theta}(\beta)$ be the quantile function of DMU-$\hat{x}$. The $\beta^* \in (L, +\infty)$ that satisfies $\hat{\theta}(\beta^*) = 1$ is referred to as the *quantile* of DMU-$\hat{x}$ (the existence and uniqueness of β^* are shown in Appendices 3 and 4, respectively); to facilitate subsequent discussion, the quantile of DMU-$\hat{x}$ is further represented as $\beta^*(\hat{x})$.

The quantile here actually denotes the degree of acceptance. For instance, in Fig. 2, the quantile of DMU-$\hat{x}$ is 0.5, which also indicates that the degree of acceptance corresponding to DMU-$\hat{x}$ is 0.5. Likewise, if the quantile of a DMU is on the boundary of the acceptance domain wrt $T_0(T_1)$, then the degree of acceptance corresponding to the DMU is 0(1). It is clear that there is an infinite possible number of acceptance degrees that a DMU might take since $\beta^* \in (L, +\infty)$. It follows that, to determine the acceptance domain T_{β^*} corresponding to DMU- $\hat{x}$, we may need to solve, according to Theorem 2, an infinite number of linear programs (P_β) with one corresponding to a $\beta \in (L, +\infty)$. Fortunately, in practice, it may not be necessary to find the exact value of β^*; that is, a close value to β^* is usually sufficient. Therefore, we consider only t, instead of an infinite number of, different values of β, i.e., $\beta = \beta_1, \beta_2, \ldots, \beta_{t'-1}, \beta_{t'}, \ldots, \beta_{t''-1}, \beta_{t''}, \ldots, \beta_{t-1}, \beta_t$, such that $L < \beta_1 < \beta_2 < \cdots < \beta_{t'-1} < \beta_{t'} < \cdots < \beta_{t''-1} < \beta_{t''} < \cdots < \beta_{t-1} < \beta_t$, $\beta_{t'} = 0$ and $\beta_{t''} = 1$. However, it is noted that the larger the value of t, the stronger is the classification power of the classifiers and the more operation time that is needed.

Example 2 Consider an example in which there is only one training data with $m = 2$, i.e., $\bar{x}_1 = ([3, 5], [3, 5])$, and thus $\bar{x}_{11}^{\beta} = 3 + 2\beta$, $\bar{x}_{21}^{\beta} = 3 + 2\beta$, and $L = \frac{-3}{2}$. In addition, let $t = 5$, $t' = 2$, and $t'' = 4$, and assume $\beta_1 = \frac{-1}{2}$, $\beta_2 = 0$, $\beta_3 = \frac{1}{2}$, $\beta_4 = 1$ and $\beta_5 = \frac{3}{2}$. Thus,

$$\begin{aligned}
\bar{x}_1^{\beta_1} &= \left(\bar{x}_{11}^{\beta_1}, \bar{x}_{21}^{\beta_1}\right) = (2, 2);\\
\bar{x}_1^{\beta_2} &= \left(\bar{x}_{11}^{\beta_2}, \bar{x}_{21}^{\beta_2}\right) = (3, 3);\\
\bar{x}_1^{\beta_3} &= \left(\bar{x}_{11}^{\beta_3}, \bar{x}_{21}^{\beta_3}\right) = (4, 4);\\
\bar{x}_1^{\beta_4} &= \left(\bar{x}_{11}^{\beta_4}, \bar{x}_{21}^{\beta_4}\right) = (5, 5);\\
\bar{x}_1^{\beta_5} &= \left(\bar{x}_{11}^{\beta_5}, \bar{x}_{21}^{\beta_5}\right) = (6, 6).
\end{aligned}$$

In addition,

$$\begin{aligned}
T_\beta &= \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\beta} \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0, j = 1, \ldots, n \right\}\\
&= \left\{ x \,\middle|\, \begin{pmatrix} 3 + 2\beta \\ 3 + 2\beta \end{pmatrix} \lambda_1 \le x,\ \lambda_1 \ge 1 \right\}.
\end{aligned}$$

Hence, if the value of β is taken as β_1, β_2, β_3, β_4, and β_5, then

$$\begin{aligned}
T_{\beta_1} &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,\middle|\, x_1 \ge 2,\ x_2 \ge 2 \right\};\\
T_{\beta_2} &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,\middle|\, x_1 \ge 3,\ x_2 \ge 3 \right\};\\
T_{\beta_3} &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,\middle|\, x_1 \ge 4,\ x_2 \ge 4 \right\};\\
T_{\beta_4} &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,\middle|\, x_1 \ge 5,\ x_2 \ge 5 \right\};\\
T_{\beta_5} &= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \,\middle|\, x_1 \ge 6,\ x_2 \ge 6 \right\}.
\end{aligned}$$

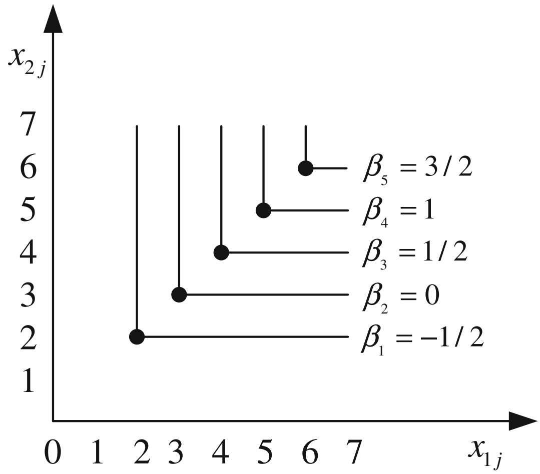
Figure 3 graphically demonstrates $T_{\beta_1}, T_{\beta_2}, \ldots, T_{\beta_5}$.

Now, consider a classification data set

$$\hat{T} = \left\{ \hat{x} \middle| \hat{x} \in \Re_+^m \right\},$$

and define the following *approximate quantile* of DMU-$\hat{x}$:

Fig. 3 $T_{\beta_1}, T_{\beta_2}, \ldots, T_{\beta_5}$



$$\beta^{**}(\hat{x}) = \begin{cases} \beta_1, & \text{if } \hat{x} \notin T_{\beta_1}, \\ \beta_i, & \text{if } \hat{x} \text{ is located on the frontier of } T_{\beta_i},\ 1 \le i \le t, \\ \frac{\beta_i + \beta_{i-1}}{2}, & \text{if } \hat{x} \in \left(\operatorname{Int} T_{\beta_{i-1}}\right) \backslash T_{\beta_i},\ 2 \le i \le t, \\ \beta_t, & \text{if } \hat{x} \in \operatorname{Int} T_{\beta_t}. \end{cases}$$

Here, if $\beta^{**}(\hat{x}) < 0$, then $\beta^{**}(\hat{x})$ represents the rejection degree wrt DMU-$\hat{x}$; otherwise, $\beta^{**}(\hat{x})$ denotes the acceptance degree wrt DMU-$\hat{x}$.

It is clear that, to implement the above defined approximate quantile $\beta^{**}(\hat{x})$, we need to repeatedly check the following four classification conditions, which could be quite time-consuming: (a) $\hat{x} \notin T_{\beta_1}$; (b) $\hat{x}$ is located on the frontier of T_{β_i}, $1 \le i \le t$; (c) $\hat{x} \in (\operatorname{Int} T_{\beta_{i-1}}) \backslash T_{\beta_i}$, $1 \le i \le t$; and (d) $\hat{x} \in \operatorname{Int} T_{\beta_t}$. Hence, to efficiently classify the data included in the classification data set $\hat{T}$ by using conditions (a)–(d), we transform the sum-form acceptance domain into the intersection-form one. The transformation method is detailed in Wei and Yan (2001) and Yan and Wei (2000). Recall that sum-form acceptance domain T_β is as follows:

$$T_\beta = \left\{ x \middle| \sum_{j=1}^{n} x_j^\beta \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n \right\}.$$

On the other hand, intersection-form acceptance domain T_β is as follows:

$$T_\beta = \left\{ x \middle| \left(\omega_\beta^k\right)^T x - \mu_\beta^k \ge 0,\ k = 1, \ldots, l_\beta \right\},$$

where $\omega_\beta^k \ge 0$, $\omega_\beta^k \ne 0$, $\mu_\beta^k \ge 0$, $k = 1, \ldots, l_\beta$. Note that, in both types of acceptance domain, $\beta = \beta_1, \ldots, \beta_{t-1}, \beta_t$. It follows that $\operatorname{Int} T_\beta = \{x | (\omega_\beta^k)^T x - \mu_\beta^k > 0,\ k = 1, \ldots, l_\beta\}$. Furthermore, it is noticed that intersection-form acceptance domain T_β is actually a linear inequality system, and that the number of linear inequalities is less than or equal to $m \times n$, which happens when all n DMUs are extreme points of T_β such that each extreme point is the single intersection point of m (the number of inputs associated with the DMUs) hyper-planes in $\Re^m$. It follows that in practical applications, it takes a reasonable amount time to perform the procedure of transforming the sum-form acceptance domain into the intersection-form one because the main task of the transformation procedure is to search for the linear inequalities. Moreover, it is important to note that the classification power of the proposed classifiers that are described below and in the next section is not affected by the values of both m and n. That is, once the intersection-form acceptance domains T_{β_i}, $i = 1, \ldots, t$ are constructed, the classification data can be efficiently and effectively classified. In short, the values of both m and n affect the time needed to transform the sum-form acceptance domains into the intersection-form ones, but do not affect the classification power of the quantile–DEA classifiers.

The quantile–DEA classifier with exact data can be formally described as follows:

Step 1 Select training data set $\bar{T} = \{[a_j, b_j], j = 1,\ldots,n\}$, where $a_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$, $j = 1, \ldots, n$ and $b_j = (b_{1j}, b_{2j}, \ldots, b_{mj})^T$, $j = 1, \ldots, n$.

Step 2 Set first the value of t $(t \geq 1)$ and then the value of β such that $\beta_1 < \beta_2 < \cdots < \beta_{t'} < \cdots < \beta_{t''} < \cdots < \beta_{t-1} < \beta_t$, $t' = 0$, and $t'' = 1$. Compute $x_j^{\beta_i} = a_j + \beta_i(b_j - a_j)$, $i = 1,\ldots,t; j = 1,\ldots,n$.

Step 3 Construct intersection-form acceptance domains $T_{\beta_i} = \{x | (\omega^k_{\beta_i})^T x - \mu^k_{\beta_i 0} \geq 0, k = 1,\ldots,l_{\beta_i}\}$, and $\operatorname{Int} T_{\beta_i} = \{x | (\omega^k_{\beta_i})^T x - \mu^k_{\beta_i 0} > 0, k = 1,\ldots,l_{\beta_i}\}$, where $i = 1, \ldots, t$.

Step 4 Implement the approximate quantile $\beta^{**}(\hat{x})$ defined above to classify (accept or reject) and at the same time give the corresponding degree of every piece of data $\hat{x} \in \hat{T}$.

Example 3 Consider again the training data set given in Example 1; that is, the training data set $\bar{T} = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4\}$. Set $t = 4$, and let $\beta_1 = -0.1$, $\beta_2 = 0$, $\beta_3 = 0.5$ and $\beta_4 = 1$. Based on the values of x_{ij}^{β}, $i = 1, 2$; $j = 1, 2, 3, 4$ (see Example 1), we can construct T_{β_1}, T_{β_2}, T_{β_3} and T_{β_4} in both sum and intersection forms.

(i) $\beta_1 = -0.1$:

$$T_{\beta_1} = \left\{ \begin{array}{l} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \left| \begin{pmatrix} 0.5 \\ 3.7 \end{pmatrix}\lambda_1 + \begin{pmatrix} 1.6 \\ 1.3 \end{pmatrix}\lambda_2 + \begin{pmatrix} 3.4 \\ 0.7 \end{pmatrix}\lambda_3 + \begin{pmatrix} 3.7 \\ 3.6 \end{pmatrix}\lambda_4 \leq \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \right. \\ \sum_{j=1}^{4} \lambda_j \geq 1,\ \lambda_j \geq 0, j = 1,\ldots,4 \end{array} \right\} \text{(sum-form)}$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \left| \begin{array}{l} 240x_1 + 110x_2 - 527 \geq 0,\ x_1 \geq 0.5, \\ 2x_1 + 6x_2 - 11 \geq 0,\ x_2 \geq 0.7 \end{array} \right. \right\} \text{(intersection-form)}.$$

(ii) $\beta_2 = 0$:

$$T_{\beta_2} = \left\{ \begin{array}{l} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \left| \begin{pmatrix} 1 \\ 2 \end{pmatrix}\lambda_1 + \begin{pmatrix} 2 \\ 2 \end{pmatrix}\lambda_2 + \begin{pmatrix} 4 \\ 1 \end{pmatrix}\lambda_3 + \begin{pmatrix} 4 \\ 4 \end{pmatrix}\lambda_4 \leq \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \right. \\ \sum_{j=1}^{4} \lambda_j \geq 1,\ \lambda_j \geq 0, j = 1,\ldots,4 \end{array} \right\} \text{(sum-form)}$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \left| \begin{array}{l} 2x_1 + x_2 - 6 \geq 0,\ x_1 \geq 1, \\ x_1 + 2x_2 - 6 \geq 0,\ x_2 \geq 1 \end{array} \right. \right\} \text{(intersection-form)}.$$

(iii) $\beta_3 = 0.5$:

$$T_{\beta_3} = \left\{ \begin{array}{l} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \begin{pmatrix} 3.5 \\ 5.5 \end{pmatrix}\lambda_1 + \begin{pmatrix} 4 \\ 5.5 \end{pmatrix}\lambda_2 + \begin{pmatrix} 7 \\ 2.5 \end{pmatrix}\lambda_3 + \begin{pmatrix} 5.5 \\ 6 \end{pmatrix}\lambda_4 \leq \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \\ \sum_{j=1}^{4} \lambda_j \geq 1,\ \lambda_j \geq 0,\ j = 1,\ldots,4 \end{array} \right\} \text{(sum-form)}$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \begin{array}{l} 12x_1 + 14x_2 - 119 \geq 0, \\ x_1 - 3.5 \geq 0,\ x_2 - 2.5 \geq 0 \end{array} \right\} \text{(intersection-form)}.$$

(iv) $\beta_4 = 1$:

$$T_{\beta_4} = \left\{ \begin{array}{l} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \begin{pmatrix} 6 \\ 7 \end{pmatrix}\lambda_1 + \begin{pmatrix} 6 \\ 9 \end{pmatrix}\lambda_2 + \begin{pmatrix} 10 \\ 4 \end{pmatrix}\lambda_3 + \begin{pmatrix} 7 \\ 8 \end{pmatrix}\lambda_4 \leq \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \\ \sum_{j=1}^{4} \lambda_j \geq 1,\ \lambda_j \geq 0,\ j = 1,\ldots,4 \end{array} \right\} \text{(sum-form)}$$

$$= \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \begin{array}{c} 3x_1 + 4x_2 - 46 \geq 0, \\ x_1 - 6 \geq 0,\ x_2 - 7 \geq 0 \end{array} \right\} \text{(intersection-form)}.$$

Let classification data set $\hat{T} = \{\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5, \hat{x}_6, \hat{x}_7\}$, where

$$\hat{x}_1 = \begin{pmatrix} \hat{x}_{11} \\ \hat{x}_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \hat{x}_2 = \begin{pmatrix} \hat{x}_{12} \\ \hat{x}_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}, \quad \hat{x}_3 = \begin{pmatrix} \hat{x}_{13} \\ \hat{x}_{23} \end{pmatrix} = \begin{pmatrix} 2 \\ 10 \end{pmatrix},$$
$$\hat{x}_4 = \begin{pmatrix} \hat{x}_{14} \\ \hat{x}_{24} \end{pmatrix} = \begin{pmatrix} 3.5 \\ 8 \end{pmatrix}, \quad \hat{x}_5 = \begin{pmatrix} \hat{x}_{15} \\ \hat{x}_{25} \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \end{pmatrix}, \quad \hat{x}_6 = \begin{pmatrix} \hat{x}_{16} \\ \hat{x}_{26} \end{pmatrix} = \begin{pmatrix} 12 \\ 4 \end{pmatrix},$$
$$\hat{x}_7 = \begin{pmatrix} \hat{x}_{17} \\ \hat{x}_{27} \end{pmatrix} = \begin{pmatrix} 10 \\ 10 \end{pmatrix}.$$

The resulting classification wrt each element in $\hat{T}$ is as follows:

(1) $\hat{x}_1$: since $240\hat{x}_{11} + 110\hat{x}_{21} - 527 < 0$, $\hat{x}_1 \notin T_{\beta_1}$ and thus $\beta^{**}(\hat{x}_1) = -0.1$.
(2) $\hat{x}_2$: since $2\hat{x}_{12} + \hat{x}_{22} - 6 > 0$, $\hat{x}_{12} + 2\hat{x}_{22} - 6 > 0$, $\hat{x}_{12} = 1$ and $\hat{x}_{22} > 1$, $\hat{x}_2$ is located on the frontier of T_{β_2} and thus $\beta^{**}(\hat{x}_2) = 0$.
(3) $\hat{x}_3$: since $2\hat{x}_{13} + \hat{x}_{23} - 6 > 0$, $\hat{x}_{13} + 2\hat{x}_{23} - 6 > 0$, $\hat{x}_{13} > 1$ and $\hat{x}_{23} > 1$, $\hat{x}_3 \in \operatorname{Int} T_{\beta_2}$. In addition, since $\hat{x}_{13} - 3.5 < 0$, $\hat{x}_3 \notin T_{\beta_3}$. Hence, $\beta^{**}(\hat{x}_3) = \frac{\beta_2+\beta_3}{2} = 0.25$.
(4) $\hat{x}_4, \hat{x}_5, \hat{x}_6$: similar to the analysis in (1)–(3), we can obtain that

$$\hat{x}_4 \text{ is located on the frontier of } T_{\beta_1} \text{ and thus } \beta^{**}(\hat{x}_4) = 0.5;$$
$$\hat{x}_5 \in \operatorname{Int} T_{\beta_1}, \quad \beta^{**}(\hat{x}_5) = \frac{\beta_1 + \beta_2}{2} = 0.75;$$
$$\hat{x}_6 \text{ is located on the frontier of } T_{\beta_2} \text{ and thus } \beta^{**}(\hat{x}_6) = 1.$$

(5) $\hat{x}_7$: since $2\hat{x}_{17} + \hat{x}_{27} - 6 > 0$, $\hat{x}_{17} + 2\hat{x}_{27} - 6 > 0$, $\hat{x}_{17} > 1$ and $\hat{x}_{27} > 1$, $\hat{x}_7 \in \operatorname{Int} T_{\beta_2}$. Furthermore, since $12\hat{x}_{17} + 14\hat{x}_{27} - 119 > 0$, $\hat{x}_{17} - 3.5 > 0$ and $\hat{x}_{27} - 2.5 > 0$, $\hat{x}_7 \in \operatorname{Int} T_{\beta_1}$. Moreover, since $3\hat{x}_{17} + 4\hat{x}_{27} - 46 > 0$, $\hat{x}_{17} - 6 > 0$ and $\hat{x}_{17} - 7 > 0$, $\hat{x}_7 \in \operatorname{Int} T_{\beta_3}$. As a consequence, $\beta^{**}(\hat{x}_7) = 1$.

Figure 4 graphically shows the resulting classification wrt each element in $\hat{T}$.

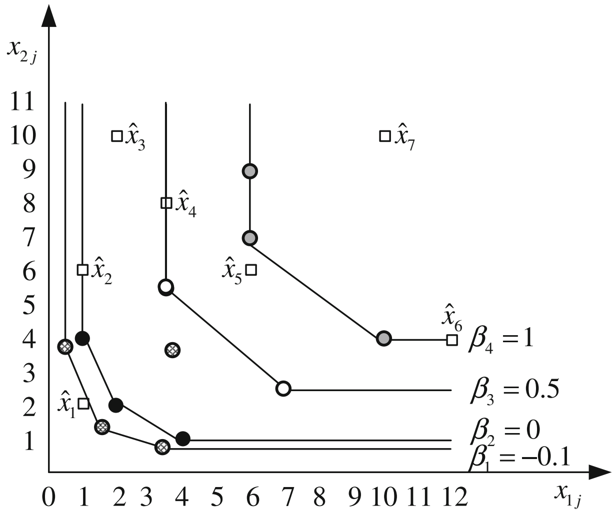


Fig. 4 Resulting classification with respect to each piece of data

4 Quantile–DEA classifier with interval data

This section introduces the quantile–DEA classifier with interval data (interval DMUs) that is mainly built on the methods proposed in the preceding section for developing the classifier for handling exact data. Let interval DMU-$\hat{x} \in \Re_{+}^{m}$, where

$$\begin{aligned}
\hat{x} &= \left(\left[\hat{a}_1, \hat{b}_1\right], \left[\hat{a}_2, \hat{b}_2\right], \ldots, \left[\hat{a}_m, \hat{b}_m\right]\right);\\
\hat{a} &= (\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_m)^T;\\
\hat{b} &= \left(\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_m\right)^T.
\end{aligned}$$

Recall that the training data set, $\bar{T}$, that is used to construct acceptance domains is defined as $\bar{T} = \{[a_j, b_j], j = 1, \ldots, n\}$, where $a_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$, $j = 1, \ldots, n$ and $b_j = (b_{1j}, b_{2j}, \ldots, b_{mj})^{\mathrm{T}}$, $j = 1, \ldots, n$. Here, we consider t different values of $\beta \in (L, +\infty)$ such that $L < \beta_1 < \beta_2 < \cdots < \beta_{t'-1} < \beta_{t'} < \cdots < \beta_{t''-1} < \beta_{t''} < \cdots < \beta_{t-1} < \beta_t$, $\beta_{t'} = 0$ and $\beta_{t''} = 1$. The corresponding approximate quantiles of $\hat{a}$ and $\hat{b}$, i.e., $\beta^{**}(\hat{a})$ and $\beta^{**}(\hat{b})$, respectively, can be calculated by applying the following formula, which is defined in the preceding section.

$$\beta^{**}(\hat{a}\ (\text{or}\ \hat{b})) = \begin{cases} \beta_1, & \text{if}\ \ \hat{a}\,(\text{or}\ \hat{b}) \notin T_{\beta_1}, \\ \beta_i, & \text{if}\ \ \hat{a}\,(\text{or}\,\hat{b})\ \text{is located on the frontier of}\ T_{\beta_i}, \quad 1 \le i \le t, \\ \frac{\beta_i + \beta_{i-1}}{2}, & \text{if}\ \ \hat{a}\,(\text{or}\,\hat{b}) \in \left(\text{Int}\,T_{\beta_{i-1}}\right)\backslash T_{\beta_i}, \quad 2 \le i \le t, \\ \beta_t, & \text{if}\ \ \hat{a}\,(\text{or}\,\hat{b}) \in \text{Int}\,T_{\beta_t}. \end{cases}$$

It is noted that the frontier of T_0, i.e., $\beta = 0$, is constructed by the minimums of $[a_{1j}, a_{2j}, \ldots, a_{mj}]$, $j = 1, \ldots, n$ in $\bar{T}$; therefore, if $\hat{x}$ is located outside T_0, then we can confidently consider $\hat{x}$ as being “rejected”. By contrast, the frontier of T_1, i.e., $\beta = 1$, is constructed by the minimums of $[b_{1j}, b_{2j}, \ldots, b_{mj}]$, $j = 1, \ldots, n$ in $\bar{T}$; hence, if $\hat{x}$ is located inside T_1, then we can be confident of considering $\hat{x}$ as being “accepted”. However, if $\hat{x}$ is located inside T_0 and outside T_1, then we can only consider $\hat{x}$ as being “accepted with risk”. It follows that we can divide $\Re_{+}^{m}$ based on T_0 and T_1 into three regions, i.e., “rejection region” I, “risky

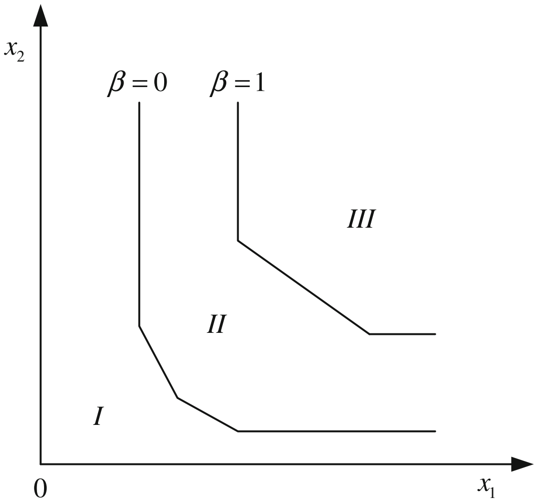


Fig. 5 Three classification regions separated by T_0 and T_1

acceptance region" II and "acceptance region" III, which are graphically demonstrated in Fig. 5. As a result, we can classify each interval DMU-$\hat{x}$ by the three degrees of "rejection", "risky acceptance" and "acceptance" that are represented by $\hat{z}_1(\hat{x})$, $\hat{z}_2(\hat{x})$ and $\hat{z}_3(\hat{x})$, respectively. In what follows, we consider six possible scenarios corresponding to interval DMU-$\hat{x}$, and show the formulas that are used to calculate its $\hat{z}_1(\hat{x})$, $\hat{z}_2(\hat{x})$ and $\hat{z}_3(\hat{x})$ wrt each scenario.

Scenario 1. Let $\hat{b} \notin \mathrm{Int} T_0$ (see Fig. 6), and thus $\beta^{**}(\hat{a}) < 0$ and $\beta^{**}(\hat{b}) \leq 0$. Then, define

$$\begin{aligned} \hat{z}_1(\hat{x}) &= |\beta^{**}(\hat{a})| + |\beta^{**}(\hat{b})|, \\ \hat{z}_2(\hat{x}) &= 0, \\ \hat{z}_3(\hat{x}) &= 0. \end{aligned}$$

Scenario 2. Let $\hat{a} \in T_1$ (see Fig. 7), and thus $\beta^{**}(\hat{a}) \geq 1$ and $\beta^{**}(\hat{b}) > 1$. Then, define

$$\begin{aligned} \hat{z}_1(\hat{x}) &= 0, \\ \hat{z}_2(\hat{x}) &= 0, \\ \hat{z}_3(\hat{x}) &= \beta^{**}(\hat{a}) + \beta^{**}(\hat{b}). \end{aligned}$$

Scenario 3. Let $\hat{a} \in T_0$ and $\hat{b} \notin \mathrm{Int}\, T_1$ (see Fig. 8), and thus $\beta^{**}(\hat{a}) \geq 0$ and $\beta^{**}(\hat{a}) < \beta^{**}(\hat{b}) \leq 1$. Then, define

$$\begin{aligned} \hat{z}_1(\hat{x}) &= 0, \\ \hat{z}_2(\hat{x}) &= \beta^{**}(\hat{a}) + \beta^{**}(\hat{b}), \\ \hat{z}_3(\hat{x}) &= 0. \end{aligned}$$

Scenario 4. Let $\hat{a} \notin T_0$ and $\hat{b} \in (\mathrm{Int}\, T_0 \backslash T_1$ (see Fig. 9), and thus $\beta^{**}(\hat{a}) < 0$ and $0 < \beta^{**}(\hat{b}) < 1$. It follows that the line segment connecting points $\hat{a}$ and $\hat{b}$ intersects the efficient frontier of T_0 (see Fig. 9); note that the point on the above defined line segment can be denoted as

$$\hat{x}_\alpha = (1 - \alpha)\hat{a} + \alpha\hat{b}, \quad \alpha \in [0, 1].$$

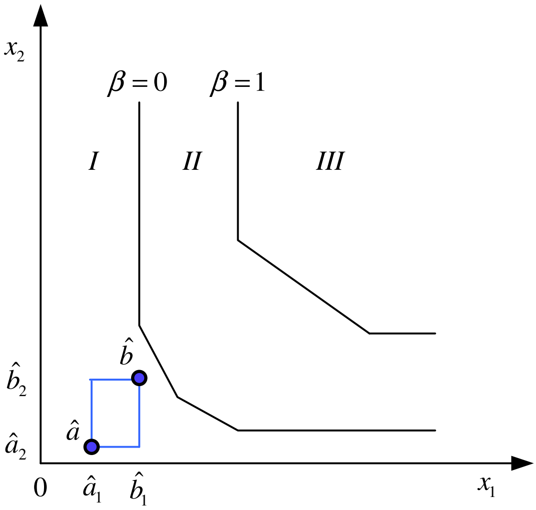


Fig. 6 Scenario 1 wrt interval DMU-$\hat{x}$

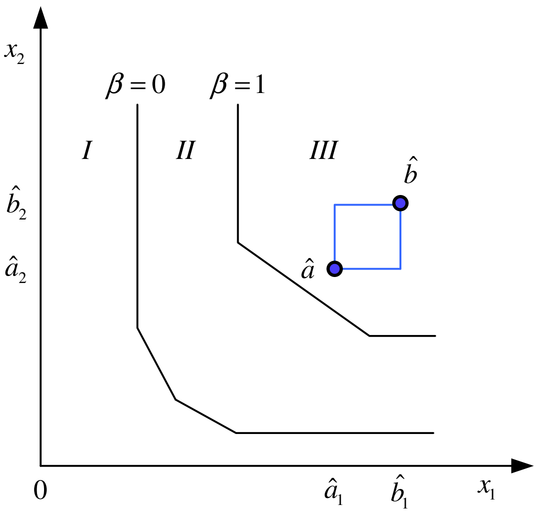


Fig. 7 Scenario 2 wrt interval DMU-$\hat{x}$

Let $\hat{x}_{\alpha^*}$ represent the intersection point; thus, $\beta^{**}(\hat{x}_{\alpha*}) = 0$. In addition, define

$$\hat{z}_1(\hat{x}) = |\beta^{**}(\hat{a})| + \beta^{**}(\hat{x}_{\alpha^*}) = |\beta^{**}(\hat{a})|,$$
$$\hat{z}_2(\hat{x}) = \beta^{**}(\hat{x}_{\alpha^*}) + \beta^{**}(\hat{b}) = \beta^{**}(\hat{b}),$$
$$\hat{z}_3(\hat{x}) = 0.$$

Scenario 5. Let $\hat{a} \in (\operatorname{Int} T_0)\backslash T_1$ and $\hat{b} \in \operatorname{Int} T_1$ (see Fig. 10), and thus $0 < \beta^{**}(\hat{a}) < 1$ and $\beta^{**}(\hat{b}) > 1$. It follows that the line segment connecting points $\hat{a}$ and $\hat{b}$ intersects the efficient frontier of T_1 (see Fig. 10).

Let $\hat{x}_{\alpha^{**}}$ represent the intersection point; thus, $\beta^{**}(\hat{x}_{\alpha^{**}}) = 1$. Define

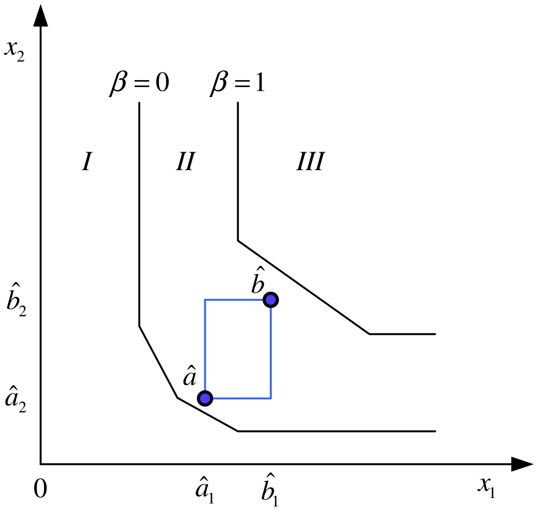


Fig. 8 Scenario 3 wrt interval DMU-$\hat{x}$

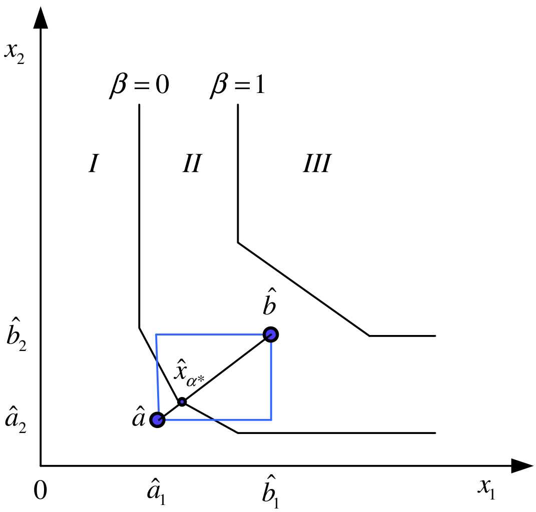


Fig. 9 Scenario 4 wrt interval DMU-$\hat{x}$

$$\hat{z}_1(\hat{x}) = 0,$$
$$\hat{z}_2(\hat{x}) = \beta^{**}(\hat{a}) + \beta^{**}(\hat{x}_{\alpha^{**}}) = \beta^{**}(\hat{a}) + 1,$$
$$\hat{z}_3(\hat{x}) = \beta^{**}(\hat{x}_{\alpha^{**}}) + \beta^{**}(\hat{b}) = 1 + \beta^{**}(\hat{b}).$$

Scenario 6. Let $\hat{a} \notin T_0$ and $\hat{b} \in \operatorname{Int} T_1$ (see Fig. 11), and thus $\beta^{**}(\hat{a}) < 0$ and $\beta^{**}(\hat{b}) > 1$. It follows that the line segment connecting points $\hat{a}$ and $\hat{b}$ intersects both the efficient frontier of T_0 and the efficient frontier of T_1 (see Fig. 11). Let $\hat{x}_{\alpha^*}$ and $\hat{x}_{\alpha^{**}}$ represent the intersection points wrt T_0 and T_1, respectively; we thus have $\beta^{**}(\hat{x}_{\alpha^*}) = 0$ and $\beta^{**}(\hat{x}_{\alpha^{**}}) = 1$. Define

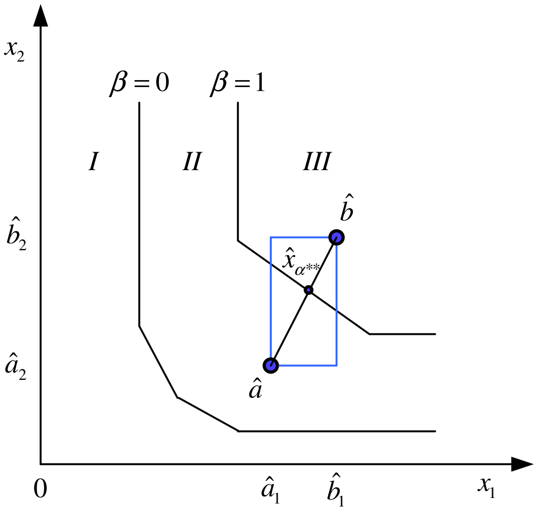


Fig. 10 Scenario 5 wrt interval DMU-$\hat{x}$

$$\hat{z}_1(\hat{x}) = |\beta^{**}(\hat{a})| + \beta^{**}(\hat{x}_{\alpha^*}) = |\beta^{**}(\hat{a})|,$$
$$\hat{z}_2(\hat{x}) = \beta^{**}(\hat{x}_{\alpha^*}) + \beta^{**}(\hat{x}_{\alpha^{**}}) = 1,$$
$$\hat{z}_3(\hat{x}) = \beta^{**}(\hat{x}_{\alpha^{**}}) + \beta^{**}(\hat{b}) = 1 + \beta^{**}(\hat{b}).$$

Note that it is easy to check, according to the formulas of $\hat{z}_1$, $\hat{z}_2$ and $\hat{z}_3$ in scenarios 1–6, that $\hat{z}_1(\hat{x}) \in (0, 2|L|)$, $\hat{z}_2(\hat{x}) \in (0, 2)$ and $\hat{z}_3(\hat{x}) \in (2, \infty)$; it is defined that the larger the value of $\hat{z}_1(\hat{x})$, $\hat{z}_2(\hat{x})$ and $\hat{z}_3(\hat{x})$, the higher the degree of rejection, risky acceptance, and acceptance, respectively. However, due to that, as indicated above, only the β that belongs to [0, 1] is associated with a well-defined probability property, $\hat{z}_1(\hat{x})$, $\hat{z}_2(\hat{x})$ and $\hat{z}_3(\hat{x})$, which are respectively derived from $\beta \in (L, 0)$, $\beta \in [0, 1]$ and $\beta \in (1, +\infty)$, can be compared only to themselves, but not to each other.

Based on the six scenarios and their corresponding formulas for calculating the quantile (degree) of classification data $\hat{x}$ that are described above, we can formally define the quantile–DEA classifier with interval data as follows:

Step 1 Select training data set $\bar{T} = \{\bar{x}_j | j = 1, \ldots, n\}$, where $\bar{x}_j = ([a_{1j}, b_{1j}], [a_{2j}, b_{2j}], \ldots, [a_{mj}, b_{mj}])$, $j = 1, \ldots, n$; $a_j = (a_{1j}, a_{2j}, \ldots, a_{mj})^T$, $j = 1, \ldots, n$; and $b_j = (b_{1j}, b_{2j}, \ldots, b_{mj})^T$, $j = 1, \ldots, n$.

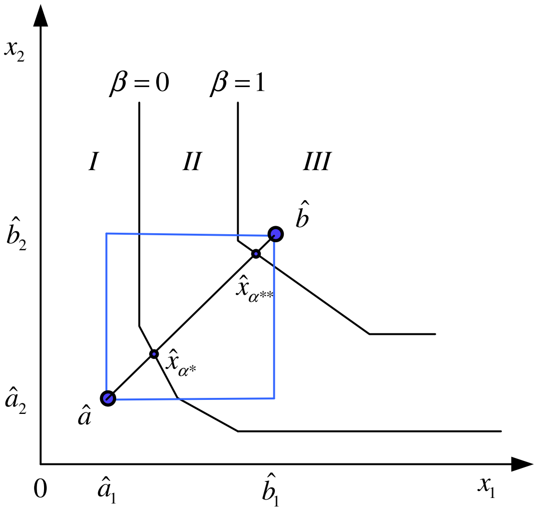
Step 2 Set first the value of t ($t \geq 1$), and then the value of β such that $L < \beta_1 < \beta_2 < \cdots < \beta_{t'} < \cdots < \beta_{t''} < \cdots < \beta_{t-1} < \beta_t$, $t' = 0$, and $t'' = 1$. Compute $x_j^{\beta_i} = a_j + \beta_i(b_j - a_j)$, $i = 1, \ldots, t$; $j = 1, \ldots, n$.

Step 3 Construct intersection-form acceptance domains $T_{\beta_i} = \{x | (\omega_{\beta_i}^k)^T x - \mu_{\beta_i 0}^k \geq 0, k = 1, \ldots, l_{\beta_i}\}$, and Int $T_{\beta_i} = \{x | (\omega_{\beta_i}^k)^T x - \mu_{\beta_i 0}^k > 0, k = 1, \ldots, l_{\beta_i}\}$, where $i = 1, \ldots, t$.

Step 4 Check the corresponding scenario wrt classification data $\hat{x} \in \hat{T}$.

Step 5 Implement the approximate quantile $\beta^{**}(\hat{x})$ to calculate $\hat{z}_1$, the "rejection degree", $\hat{z}_2$, the "risky acceptance degree", and $\hat{z}_3$, the "acceptance degree" by using the formulas corresponding to the scenario that $\hat{x}$ is involved.

Fig. 11 Scenario 6 wrt interval DMU-$\hat{x}$



Example 4 Consider the intersection-form acceptance domains $T_{\beta_1},\ldots,T_{\beta_5}$ that are constructed in Example 2, where $\beta_1=\frac{-1}{2}$, $\beta_2=0$, $\beta_3=\frac{1}{2}$, $\beta_4=1$, and $\beta_5=\frac{3}{2}$. Let classification data set

$$\hat{T}=\{\hat{x}_1,\hat{x}_2,\hat{x}_3,\hat{x}_4,\hat{x}_5,\hat{x}_6\},$$

where interval DMU-$\hat{x}_i$, $i=1,\ldots,6$ are as follows:

$$\begin{aligned}
&\hat{x}_1=\left(\left[\hat{a}_{11},\hat{b}_{11}\right],\left[\hat{a}_{21},\hat{b}_{21}\right]\right)=([1,2],[1,2]),\quad \hat{a}_1=(1,1)^T,\quad \hat{b}_1=(2,2)^T;\\
&\hat{x}_2=\left(\left[\hat{a}_{12},\hat{b}_{12}\right],\left[\hat{a}_{22},\hat{b}_{22}\right]\right)=([6,7],[6,7]),\quad \hat{a}_2=(6,6)^T,\quad \hat{b}_2=(7,7)^T;\\
&\hat{x}_3=\left(\left[\hat{a}_{13},\hat{b}_{13}\right],\left[\hat{a}_{23},\hat{b}_{23}\right]\right)=([4,4.5],[4,4.5]),\quad \hat{a}_3=(4,4)^T,\quad \hat{b}_3=(4.5,4.5)^T;\\
&\hat{x}_4=\left(\left[\hat{a}_{14},\hat{b}_{14}\right],\left[\hat{a}_{24},\hat{b}_{24}\right]\right)=([2,4],[2,4]),\quad \hat{a}_4=(2,2)^T,\quad \hat{b}_4=(4,4)^T;\\
&\hat{x}_5=\left(\left[\hat{a}_{15},\hat{b}_{15}\right],\left[\hat{a}_{25},\hat{b}_{25}\right]\right)=([4,6],[4,6]),\quad \hat{a}_5=(4,4)^T,\quad \hat{b}_5=(6,6)^T;\\
&\hat{x}_6=\left(\left[\hat{a}_{16},\hat{b}_{16}\right],\left[\hat{a}_{26},\hat{b}_{26}\right]\right)=([2,7],[2,7]),\quad \hat{a}_6=(2,2)^T,\quad \hat{b}_6=(7,7)^T.
\end{aligned}$$

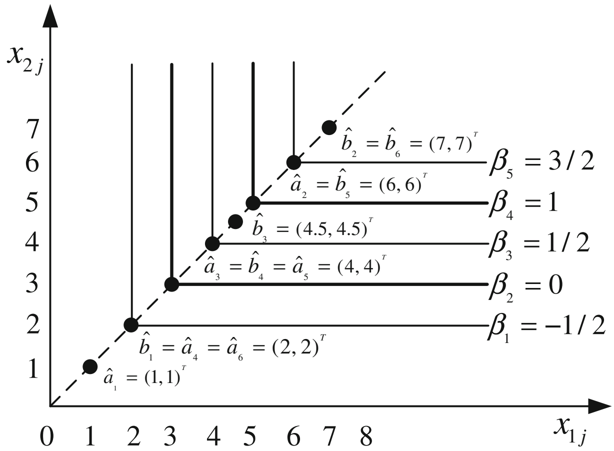
The positions of $\hat{a}_i$, $\hat{b}_i$, $i=1,\ldots,7$ corresponding to β_i, $i=1,\ldots,5$ are graphically shown in Fig. 12.

In what follows, we calculate the values of $\hat{z}_1(\hat{x})$, $\hat{z}_2(\hat{x})$ and $\hat{z}_3(\hat{x})$ wrt each classification data $\hat{x}\in\hat{T}$ in sequence.

(i) The condition of $\hat{x}_1$ satisfies scenario 1. Its corresponding $\hat{z}_1(\hat{x}_1)$, $\hat{z}_2(\hat{x}_1)$ and $\hat{z}_3(\hat{x}_1)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_1)=|\beta^{**}(\hat{a}_1)|+\left|\beta^{**}\left(\hat{b}_1\right)\right|=|\beta_1|+|\beta_1|=1,\\
&\hat{z}_2(\hat{x}_1)=0,\\
&\hat{z}_3(\hat{x}_1)=0.
\end{aligned}$$

Fig. 12 Positions of $\hat{a}_i, \hat{b}_i, i = 1, \ldots, 7$



Hence, interval DMU-$\hat{x}_1$ can obviously be classified as "rejection" with degree 1.

(ii) The condition of $\hat{x}_2$ satisfies scenario 2. Its corresponding $\hat{z}_1(\hat{x}_2)$, $\hat{z}_2(\hat{x}_2)$ and $\hat{z}_3(\hat{x}_2)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_2) = 0,\\
&\hat{z}_2(\hat{x}_2) = 0,\\
&\hat{z}_3(\hat{x}_2) = \beta^{**}(\hat{a}_2) + \beta^{**}(\hat{b}_2) = \beta_5 + \beta_5 = 3.
\end{aligned}$$

Hence, interval DMU-$\hat{x}_2$ can evidently be classified as "acceptance" with degree 3.

(iii) The condition of $\hat{x}_3$ satisfies scenario 3. Its corresponding $\hat{z}_1(\hat{x}_3)$, $\hat{z}_2(\hat{x}_3)$ and $\hat{z}_3(\hat{x}_3)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_3) = 0,\\
&\hat{z}_2(\hat{x}_3) = \beta^{**}(\hat{a}_3) + \beta^{**}(\hat{b}_3) = \beta_3 + \frac{1}{2}(\beta_4 + \beta_3) = \frac{5}{4},\\
&\hat{z}_3(\hat{x}_3) = 0.
\end{aligned}$$

Hence, interval DMU-$\hat{x}_3$ can be classified as "risky acceptance" with degree $\frac{5}{4}$.

(iv) The condition of $\hat{x}_4$ satisfies scenario 4. Its corresponding $\hat{z}_1(\hat{x}_4)$, $\hat{z}_2(\hat{x}_4)$ and $\hat{z}_3(\hat{x}_4)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_4) = |\beta^{**}(\hat{a}_4)| = |\beta_1| = \frac{1}{2},\\
&\hat{z}_2(\hat{x}_4) = \beta^{**}(\hat{b}_4) = \beta_3 = \frac{1}{2},\\
&\hat{z}_3(\hat{x}_4) = 0.
\end{aligned}$$

Since interval DMU-$\hat{x}_4$ crosses over conflicting regions of "rejection" and "risky acceptance", a classification on $\hat{x}_4$ may inevitably result in a "Type I" or "Type II" error. Hence, we do not provide the user with a definite classification, but rather provide the user with both the "rejection" and "risky acceptance" degrees, i.e., $\hat{z}_1(\hat{x}_4)$ and $\hat{z}_2(\hat{x}_4)$, for his/her reference.

(v) The condition of $\hat{x}_5$ satisfies scenario 5. Its corresponding $\hat{z}_1(\hat{x}_5)$, $\hat{z}_2(\hat{x}_5)$ and $\hat{z}_3(\hat{x}_5)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_5) = 0,\\
&\hat{z}_2(\hat{x}_5) = \beta^{**}(\hat{a}_5) + 1 = \beta_3 + 1 = \frac{3}{2},\\
&\hat{z}_3(\hat{x}_5) = 1 + \beta^{**}\left(\hat{b}_5\right) = 1 + \beta_5 = \frac{5}{2}.
\end{aligned}$$

The range associated with interval DMU-$\hat{x}_5$ crosses over two regions, which may cause difficulty in classifying it. However, since the two of them are "risky acceptance" and "acceptance" regions, we should have confidence in classifying interval DMU-$\hat{x}_5$ as "acceptance with low risk".

(vi) The condition of $\hat{x}_6$ satisfies scenario 6. Its corresponding $\hat{z}_1(\hat{x}_6)$, $\hat{z}_2(\hat{x}_6)$ and $\hat{z}_3(\hat{x}_6)$ are as follows:

$$\begin{aligned}
&\hat{z}_1(\hat{x}_6) = \beta^{**}(\hat{a}_6) = |\beta_1| = \frac{1}{2},\\
&\hat{z}_2(\hat{x}_6) = 1,\\
&\hat{z}_3(\hat{x}_6) = 1 + \beta^{**}\left(\hat{b}_6\right) = 1 + \beta_5 = \frac{5}{2}.
\end{aligned}$$

In actual fact, an interval data with a range crossing over three regions as shown in scenario 6 is unusual if not impossible. Here we consider such an interval data, e.g., interval DMU-$\hat{x}_6$, to be un-classifiable due to the inherently uninformative values of $\hat{z}_1(\hat{x}_6)$, $\hat{z}_2(\hat{x}_6)$ and $\hat{z}_3(\hat{x}_6)$, and suggest that the user should re-check the data.

5 Extensions

Recall that in Sect. 2, we assumed that each training interval data denoted as interval DMU-$\bar{x}_j$ ($j = 1, \ldots, n$) is solely associated with m inputs, and that the bigger the values of the data, the higher the probability that the data are accepted. Based on the above setting, we construct acceptance domain T_β in both sum-form and intersection-form. In this section, we consider two new settings: (1) each interval DMU-$\bar{y}_j$ ($j = 1, \ldots, n$) is solely associated with s outputs, and (2) each interval DMU-($\bar{x}_j$, $\bar{y}_j$) ($j = 1, \ldots, n$) is associated with both m inputs and s outputs. Accordingly, we construct acceptance domain T_β in both sum-form and intersection-form wrt each new setting.

First, consider the setting in which each interval DMU-$\bar{y}_j$ ($j = 1, \ldots, n$) is associated with s outputs, and the smaller the values of the data, the higher the probability that the data will be accepted. Denote the interval-DMUs as

$$\bar{y}_j = \left(\bar{y}_{1j}, \bar{y}_{2j}, \ldots, \bar{y}_{sj}\right), \quad j = 1, \ldots, n,$$

where

$$\bar{y}_{rj} \in \left[c_{rj}, d_{rj}\right], \quad r = 1, \ldots, s; \quad j = 1, \ldots, n,$$

and

$$d_{rj} > c_{rj} > 0, \quad r = 1, \ldots, s; \quad j = 1, \ldots, n.$$

In addition, denote the training data set as

$$\bar{T} = \left\{\bar{y}_j | j = 1, \ldots, n\right\},$$

and define

$$\bar{y}_j^{\beta} = \left(\bar{y}_{1j}^{\beta}, \bar{y}_{2j}^{\beta}, \ldots, \bar{y}_{sj}^{\beta}\right), \quad j = 1, \ldots, n,$$

where

$$\bar{y}_{rj}^{\beta} = d_{rj} - \beta\left(d_{rj} - c_{rj}\right) > 0, \quad r = 1, \ldots, s; \quad j = 1, \ldots, n,$$

and $\beta \in (-\infty, R)$ with

$$R = \min_{1 \le r \le s; 1 \le j \le n} \left(\frac{d_{rj}}{d_{rj} - c_{rj}}\right) > 1.$$

Let T_β represent the acceptance domain constructed by $y_j^{\beta} = (y_{1j}^{\beta}, y_{2j}^{\beta}, \ldots, y_{sj}^{\beta})^T, j = 1, \ldots, n$ given a specified $\beta \in (-\infty, R)$. It follows that that T_β satisfies the following postulates:

Postulate 1 (Ordinary postulate) the observed $y_j^{\beta} \in T_\beta$ for all $j = 1, \ldots, n$.
Postulate 2 (Convexity postulate) if $y \in T_\beta$ and $\hat{y} \in T_\beta$, then $\lambda y + (1 - \lambda)\hat{y} \in T_\beta$, for $\lambda \in [0, 1]$.
Postulate 3 (Monotonicity postulate) if $y \in T_\beta$, and $\hat{y} \le y$, then $\hat{y} \in T_\beta$.
Postulate 4 (Minimum extrapolation postulate) T_β is the intersection set of all $\tilde{T}$ satisfying Postulates 1–3.

The sum-form acceptance domain that satisfies Postulates 1–4 defined above can be represented as follows:

$$T_\beta = \left\{ y \left| \sum_{j=1}^{n} y_j^{\beta} \lambda_j \ge y, \sum_{j=1}^{n} \lambda_j \le 1, \lambda_j \ge 0, j = 1, \ldots, n \right. \right\}.$$

Note that sum-form acceptance domain T_β has the same structure as the production possibility set corresponding to the classical CCR model with reference set $\{(1, y_j^{\beta}) | j = 1, \ldots, n\}$ in DEA research. In addition, intersection-form acceptance domain T_β is as follows:

$$T_\beta = \left\{ y \left| \omega_\beta^k - \left(\mu_\beta^k\right)^T y \ge 0, k = 1, \ldots, l_\beta \right. \right\},$$

where $\omega_\beta^k \geq 0$, $\mu_\beta^k \geq 0$, $\mu_\beta^k \neq 0$, $k = 1, \ldots, l_\beta$. Denote an exact classification data $\hat{y}$ as DMU-$\hat{y}$, and consider the following linear program wrt DMU-$\hat{y}$ with a specified $\beta \in (-\infty, R)$.

$$
(P_\beta) \quad \text{s.t.} \quad \begin{aligned} &\hat{\varphi}(\beta) = \max \varphi, \\ &\sum_{j=1}^{n} y_j^\beta \lambda_j \geq \varphi \hat{y}, \\ &\sum_{j=1}^{n} \lambda_j \leq 1, \\ &\lambda_j \geq 0, \quad j = 1, \ldots, n. \end{aligned}
$$

Second, consider the setting in which each interval DMU-$(\bar{x}_j, \bar{y}_j)$ $(j = 1, \ldots, n)$ is associated with both m inputs and s outputs. Denote the interval-DMUs as

$$
\begin{aligned} \bar{x}_j &= \left(\bar{x}_{1j}, \bar{x}_{2j}, \ldots, \bar{x}_{mj}\right), \quad j = 1, \ldots, n, \\ \bar{y}_j &= \left(\bar{y}_{1j}, \bar{y}_{2j}, \ldots, \bar{y}_{sj}\right), \quad j = 1, \ldots, n, \end{aligned}
$$

where

$$
\begin{aligned} \bar{x}_{ij} &\in \left[a_{ij}, b_{ij}\right], \quad i = 1, \ldots, m; \quad j = 1, \ldots, n, \\ \bar{y}_{rj} &\in \left[c_{rj}, d_{rj}\right], \quad r = 1, \ldots, s; \quad j = 1, \ldots, n. \end{aligned}
$$

Note that the smaller the values of the outputs and the larger the values of the inputs, the higher the probability that the data will be accepted. In addition, define

$$
\begin{aligned} \bar{x}_j^\beta &= \left(\bar{x}_{1j}^\beta, \bar{x}_{2j}^\beta, \ldots, \bar{x}_{mj}^\beta\right), \quad j = 1, \ldots, n, \\ \bar{y}_j^\beta &= \left(\bar{y}_{1j}^\beta, \bar{y}_{2j}^\beta, \ldots, \bar{y}_{sj}^\beta\right), \quad j = 1, \ldots, n, \end{aligned}
$$

where

$$
\begin{aligned} \bar{x}_{ij}^\beta &= a_{ij} + \beta\left(b_{ij} - a_{ij}\right) > 0, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n, \\ \bar{y}_{rj}^\beta &= d_{rj} - \beta\left(d_{rj} - c_{rj}\right) > 0, \quad r = 1, \ldots, s; \quad j = 1, \ldots, n. \end{aligned}
$$

and $\beta \in (L, R)$ with

$$
\begin{aligned} L &= \max_{1 \leq i \leq m; 1 \leq j \leq n} \left(\frac{-a_{ij}}{b_{ij} - a_{ij}}\right) < 0, \\ R &= \min_{1 \leq r \leq s; 1 \leq j \leq n} \left(\frac{d_{rj}}{d_{rj} - c_{rj}}\right) > 1. \end{aligned}
$$

Let T_β represent the acceptance domain constructed by (x_j^β, y_j^β), $j = 1, \ldots, n$ given a specified $\beta \in (L, R)$. It follows that that T_β satisfies the following postulates:

Postulate 1 (Ordinary postulate) the observed $(x_j^\beta, y_j^\beta) \in T_\beta$ for all $j = 1, \ldots, n$.

Postulate 2 (Convexity postulate) if $(x, y) \in T_\beta$, and $(\hat{x}, \hat{y}) \in T_\beta$, then $\lambda(x, y) + (1 - \lambda)(\hat{x}, \hat{y}) \in T_\beta$, for $\lambda \in [0, 1]$.

Postulate 3 (Monotonicity postulate) if $(x, y) \in T_\beta$, $\hat{x} \geq x$ and $\hat{y} \leq y$, then $(\hat{x}, \hat{y}) \in T_\beta$.

Postulate 4 (Ray unbounded postulate) if $(x, y) \in T_\beta$, then $\alpha(\hat{x}, \hat{y}) \in T_\beta$ for all $\alpha \geq 0$.

Postulate 5 (Minimum extrapolation postulate) T_β is the intersection set of all $\tilde{T}$ satisfying Postulates 1–4.

The sum-form acceptance domain that satisfies Postulates 1–5 defined above can be represented as follows:

$$T_\beta = \left\{ (x,\, y) \middle| \sum_{j=1}^{n} x_j^\beta \lambda_j \le x,\ \sum_{j=1}^{n} y_j^\beta \lambda_j \ge y,\ \lambda_j \ge 0,\, j = 1, \ldots, n \right\}.$$

Note that sum-form acceptance domain T_β has the same structure as the production possibility set corresponding to the classical CCR model with reference set $\{(x_j^\beta,\, y_j^\beta) | j = 1, \ldots, n\}$ in DEA research. In addition, intersection-form acceptance domain T_β is as follows:

$$T_\beta = \left\{ (x,\, y) \middle| \left(\omega_\beta^k\right)^T x - \left(\mu_\beta^k\right)^T y \ge 0,\, k = 1, \ldots, l_\beta \right\},$$

where $\begin{pmatrix} \omega^k \\ \mu^k \end{pmatrix} \ge 0$, $\begin{pmatrix} \omega^k \\ \mu^k \end{pmatrix} \ne 0$. Denote an exact classification data $(\hat{x},\, \hat{y})$ as DMU-$(\hat{x},\, \hat{y})$, and consider the following linear program wrt DMU-$(\hat{x},\, \hat{y})$ with a specified $\beta \in (L,\, R)$.

$$\begin{aligned} & & & \hat{\theta}(\beta) = \min \theta, \\ & (P_\beta) & \text{s.t.} \quad & \sum_{j=1}^{n} x_j^\beta \lambda_j \le \theta \hat{x}, \\ & & & \sum_{j=1}^{n} y_j^\beta \lambda_j \ge \hat{y}, \\ & & & \lambda_j \ge 0, \quad j = 1, \ldots, n. \end{aligned}$$

It is easy to check that once the intersection-form acceptance domains T_β corresponding to the two new settings are constructed, the quantile–DEA classifiers introduced in Sects. 3 and 4 can be directly used to discover the groups to which a huge amount of classification data belong. That is, the proposed quantile–DEA classifiers can deal with the settings in which each interval DMU-$\bar{x}_j$, ($j = 1, \ldots, n$) is associated with solely inputs, solely outputs, and both inputs and outputs.

6 Conclusions

This research proposes, to our knowledge, the first DEA-based classifiers, quantile–DEA classifiers, for dealing with binary classification problems with classification data that are known to have either exact values or values only within bounded intervals. The technique of multiple acceptance domains that is derived from the ideas and methods of both quantiles in statistics and the intersection-form production possibility set in the DEA framework enables the quantile–DEA classifiers not only to promptly classify a large volume of data, but also to provide the degrees associated with the patterns. It is noteworthy that the proposed classifier simply classifies an exact piece of data into “acceptance” or “rejection” with a corresponding degree. However, due to the inherent complexity of interval data, the proposed classifier outputs three types of degrees associated with classification data: the “rejection”, “risky acceptance”, and “acceptance” degrees.

In addition, it is worth mentioning that, in the proposed quantile–DEA classifiers, it is assumed that the m characteristic values are equally important. However, in practice, the decision makers may value them differently. Hence, there are proposed DEA models in the literature that apply the preference cone to reflect the different importance of the characteristic values (see, e.g., Charnes et al. 1989; Yu et al. 1996; Wei and Yu 1997). In actual fact, by applying the ideas in the articles, it is not difficult to construct the preference-cone restricted quantile–DEA classifiers that can deal with the classification problems with different weighted characteristic values.

In short, this study sheds some light in extending the function of traditional DEA models from evaluating to classifying. The proposed quantile–DEA classifiers are both efficient and quite user-friendly in terms of detailed output information. Therefore, they have great potential in practical applications and can thus be effective complementary approaches for data mining.

Acknowledgments This paper has benefited from the suggestions offered by the reviewers, and this assistance is gratefully acknowledged. In addition, the first and the third authors are partially supported by the National Natural Science Foundation of China, NNSF 71271208.

Appendix 1: Proof of Theorem 1(i)

Theorem 1 *Let* $L < \bar{\beta} < \hat{\beta}$, *and*

$$T_{\bar{\beta}} = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\bar{\beta}} \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n \right\},$$

and

$$T_{\hat{\beta}} = \left\{ x \,\middle|\, \sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n \right\}.$$

Then, $T_{\hat{\beta}} \subset T_{\bar{\beta}}$.

Proof Since $b_j > a_j$, $j = 1, \ldots, n$, if $L < \bar{\beta} < \hat{\beta}$, then

$$x_j^{\bar{\beta}} = a_j + \bar{\beta}(b_j - a_j) < a_j + \hat{\beta}(b_j - a_j) = x_j^{\hat{\beta}}, \quad j = 1, \ldots, n.$$

It follows that if $\sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n$, then

$$\sum_{j=1}^{n} x_j^{\bar{\beta}} \lambda_j < \sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j.$$

Thus, if $x \in T_{\hat{\beta}}$, then $x \in T_{\bar{\beta}}$; that is, $T_{\hat{\beta}} \subset T_{\bar{\beta}}$. □

Appendix 2: Proof of Theorem 2

To prove Theorem 2, we first present the following two lemmas:

Lemma 1 *If* $L<\bar{\beta}<\hat{\beta}$, *and* $\tilde{x}\in T_{\hat{\beta}}=\left\{x\middle|\sum_{j=1}^{n}x_j^{\hat{\beta}}\lambda_j\le x,\ \sum_{j=1}^{n}\lambda_j\ge 1,\ \lambda_j\ge 0, j=1,\ldots,n.\right\}$, *then the optimal objective function value of the following linear program is less than one*; *i.e.*, $\hat{\theta}(\bar{\beta})<1$.

$$\left(P_{\bar{\beta}}\right)\quad \begin{aligned}&\hat{\theta}(\bar{\beta})=\min\theta,\\ &\text{s.t. } \sum_{j=1}^{n}x_j^{\bar{\beta}}\lambda_j\le\theta\tilde{x},\\ &\sum_{j=1}^{n}\lambda_j\ge 1,\\ &\lambda_j\ge 0,\quad j=1,\ldots,n.\end{aligned}$$

Proof Let $T_{\hat{\beta}}=\left\{x\middle|\sum_{j=1}^{n}x_j^{\hat{\beta}}\lambda_j\le x,\ \sum_{j=1}^{n}\lambda_j\ge 1,\ \lambda_j\ge 0, j=1,\ldots,n\right\}$. Since $\tilde{x}\in T_{\hat{\beta}}$, there exist $\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_n$ that satisfy

$$\begin{aligned}&\sum_{j=1}^{n}x_j^{\hat{\beta}}\tilde{\lambda}_j\le\tilde{x},\\ &\sum_{j=1}^{n}\tilde{\lambda}_j\ge 1,\\ &\tilde{\lambda}_j\ge 0,\quad j=1,\ldots,n.\end{aligned}$$

Furthermore, since $\sum_{j=1}^{n}\tilde{\lambda}_j\ge 1$, $(\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_n)\ne 0$. Moreover, since $a_j<b_j$, $0<x_j^{\bar{\beta}}<x_j^{\hat{\beta}}$, $j=1,\ldots,n$. In summary,

$$\sum_{j=1}^{n}x_j^{\bar{\beta}}\tilde{\lambda}_j<\sum_{j=1}^{n}x_j^{\hat{\beta}}\tilde{\lambda}_j\le\tilde{x}.$$

It follows that there exist solutions to the following system of inequalities:

$$\begin{aligned}&\sum_{j=1}^{n}x_j^{\bar{\beta}}\lambda_j<\tilde{x},\\ &\sum_{j=1}^{n}\lambda_j\ge 1,\\ &\lambda_j\ge 0,\quad j=1,\ldots,n.\end{aligned}$$

As a result, $\hat{\theta}(\bar{\beta})<1$ (i.e., the optimal objective function value of $(P_{\bar{\beta}})$ is less than one). □

Lemma 2 *If* $L<\bar{\beta}$, $\hat{x}>0$ *and* $\hat{x}\notin T_{\beta}$, *then the optimal objective function value of the following linear program is greater than one*; *i.e.*, $\hat{\theta}(\beta)>1$.

$$\left(P_{\beta}\right)\quad \text{s.t.}\quad \begin{aligned}&\hat{\theta}(\beta)=\min\theta,\\ &\sum_{j=1}^{n}x_j^{\beta}\lambda_j\le\theta\hat{x},\\ &\sum_{j=1}^{n}\lambda_j\ge 1,\\ &\lambda_j\ge 0,\quad j=1,\ldots,n.\end{aligned}$$

Proof Let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_n$ denote the optimal solution to (P_β) and $\hat{\theta}(\beta) = \hat{\theta}$. If $\hat{\theta}(\beta) = \hat{\theta} \leq 1$, then

$$\sum_{j=1}^{n} x_j^{\beta} \hat{\lambda}_j \leq \hat{\theta}\hat{x} \leq \hat{x},$$
$$\sum_{j=1}^{n} \hat{\lambda}_j \geq 1,$$
$$\hat{\lambda}_j \geq 0, \quad j = 1, \ldots, n.$$

That is, $\hat{x} \in T_\beta$, which is a contradiction. □

In what follows, we give the proof to Theorem 2, first to (i) and then to (ii).

Theorem 2 *Let* $\hat{x} \in \hat{T} \cap \mathrm{Int}\left\{x \middle| \sum_{j=1}^{n} x_j^L \lambda_j \leq x,\ \sum_{j=1}^{n} \lambda_j \geq 1,\ \lambda_j \geq 0, j = 1, \ldots, n\right\}$, *and* $\hat{\theta}(\beta)$ *be the quantile function of DMU-*$\hat{x}$*. Then,*

(i) $\hat{\theta}(\beta)$ *is a continuous function defined over* $(L, +\infty)$.
(ii) $\hat{\theta}(\beta)$ *is a strictly monotonically decreasing function over* $(L, +\infty)$.

Proof

(i) Consider the following linear program (P_β) :

$$(P_\beta) \quad \text{s.t.} \quad \begin{aligned} &\hat{\theta}(\beta) = \min \theta, \\ &\sum_{j=1}^{n} x_j^{\beta} \lambda_j \leq \theta \hat{x}, \\ &\sum_{j=1}^{n} \lambda_j \geq 1, \\ &\lambda_j \geq 0, \quad j = 1, \ldots, n. \end{aligned}$$

Equivalently,

$$(P_\beta) \quad \text{s.t.} \quad \begin{aligned} &\hat{\theta}(\beta) = \min \theta, \\ &\sum_{j=1}^{n} \left[a_j + \beta\left(b_j - a_j\right)\right] \lambda_j \leq \theta \hat{x}, \\ &\sum_{j=1}^{n} \lambda_j \geq 1, \\ &\lambda_j \geq 0, \quad j = 1, \ldots, n. \end{aligned}$$

According to the stability of linear programming (Ying et al. 1975), the optimal objective function value of (P_β), $\hat{\theta}(\beta)$, is a continuous function defined over $(L, +\infty)$.

(ii) Let $L < \bar{\beta} < \hat{\beta}$, and consider the following problem $(P_{\hat{\beta}})$:

$$(P_{\hat{\beta}}) \quad \text{s.t.} \quad \begin{aligned} &\hat{\theta}(\hat{\beta}) = \min \theta, \\ &\sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j \leq \theta \hat{x}, \\ &\sum_{j=1}^{n} \lambda_j \geq 1, \\ &\lambda_j \geq 0, \quad j = 1, \ldots, n. \end{aligned}$$

It is clear that $\hat{\theta}(\hat{\beta})\hat{x} \in T_{\hat{\beta}}$. Consider also the following problem $(\tilde{P}_{\bar{\beta}})$:

$$\left(\tilde{P}_{\bar{\beta}}\right) \quad \text{s.t.} \quad \begin{aligned} &\hat{\theta}(\bar{\beta}) = \min \theta, \\ &\sum_{j=1}^{n} x_j^{\bar{\beta}} \lambda_j \le \theta(\hat{\theta}(\hat{\beta})\hat{x}), \\ &\sum_{j=1}^{n} \lambda_j \ge 1, \\ &\lambda_j \ge 0, \quad j = 1, \ldots, n. \end{aligned}$$

Let $\tilde{\theta}$, $\tilde{\lambda}_1$, $\tilde{\lambda}_2, \ldots, \tilde{\lambda}_n$ denote the optimal solution to $(\tilde{P}_{\bar{\beta}})$. It is easy to check that $\tilde{\theta} > 0$. Furthermore, since $\hat{\theta}(\hat{\beta})\hat{x} \in T_{\hat{\beta}}$ and $L < \bar{\beta} < \hat{\beta}$, from Lemma 1, $\tilde{\theta} < 1$. Moreover, since $\hat{\theta}(\bar{\beta})$ is the optimal objective function value of $(\tilde{P}_{\bar{\beta}})$, $\hat{\theta}(\bar{\beta}) \le \tilde{\theta}\hat{\theta}(\hat{\beta}) < \hat{\theta}(\hat{\beta})$. □

Appendix 3: Existence of β^*

The following Theorem 3 shows the existence of β^*.

Theorem 3 *Let* $\hat{x} \in \hat{T} \cap \operatorname{Int}\left\{x \middle| \sum_{j=1}^{n} x_j^L \lambda_j \le x,\ \sum_{j=1}^{n} \lambda_j \ge 1,\ \lambda_j \ge 0,\ j = 1, \ldots, n\right\}$, *and* $\hat{\theta}(\beta)$ *be the quantile function of DMU-*$\hat{x}$. *Then, there exists* $\beta^* \in (L, +\infty)$ *such that the optimal objective function value of the following problem* (P_β) *is equal to one*; *i.e.*, $\hat{\theta}(\beta^*) = 1$.

$$\left(P_{\beta}\right) \quad \text{s.t.} \quad \begin{aligned} &\hat{\theta}(\beta) = \min \theta, \\ &\sum_{j=1}^{n} x_j^{\beta} \lambda_j \le \theta \hat{x}, \\ &\sum_{j=1}^{n} \lambda_j \ge 1, \\ &\lambda_j \ge 0, \quad j = 1, \ldots, n. \end{aligned}$$

Proof

(i) If $\hat{x}$ is located on the frontier of T_1, then $\hat{\theta}(1) = 1$, i.e., $\beta^* = 1$.

(ii) If $\hat{x}$ is not located on the frontier of T_1, and $\hat{x} \in \operatorname{Int} T_1$, then there exist $\lambda_j^0 \ge 0$, $j = 1, \ldots, n$, $\sum_{j=1}^{n} \lambda_j^0 \ge 1$ such that

$$\sum_{j=1}^{n} \left[a_j + 1 \times \left(b_j - a_j\right)\right] \lambda_j^0 = \sum_{j=1}^{n} b_j \lambda_j^0 < \hat{x}, \tag{1}$$

and $\hat{\theta}(1) < 1$. Let

$$\hat{\beta} > \max\left\{\max_{1 \le i \le m; 1 \le j \le n} \left\{\left(\hat{x}_{ij} - a_{ij}\right) / \left(b_{ij} - a_{ij}\right)\right\},\ L\right\}.$$

Then,

$$a_j + \hat{\beta}\left(b_j - a_j\right) > \hat{x}, \quad j = 1, \ldots, n.$$

Therefore, for any $\lambda_j \geq 0$, $j = 1, \ldots, n$, $\sum_{j=1}^{n} \lambda_j \geq 1$, we have

$$\sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j > \hat{x}. \tag{2}$$

From (2), $\hat{x} \notin T_{\hat{\beta}}$, and from Lemma 2, $\hat{\theta}(\hat{\beta}) > 1$. As a result, since $\hat{\theta}(1) < 1$, $\hat{\theta}(\hat{\beta}) > 1$, $\hat{\beta} \in (L, +\infty)$, from Theorem 2(i), $\hat{\theta}(\beta)$ is a continuous function defined over $(L, +\infty)$. It follows that there exists $\beta^* \in (L, +\infty)$ such that $\hat{\theta}(\beta^*) = 1$.

(iii) If $\hat{x} \notin T_1$, from Lemma 2, $\hat{\theta}(1) > 1$. In addition, since

$$\hat{x} \in \mathrm{Int}\left\{ x \middle| \sum_{j=1}^{n} x_j^L \lambda_j \leq x, \sum_{j=1}^{n} \lambda_j \geq 1, \lambda_j \geq 0, j = 1, \ldots, n \right\},$$

there exist $\lambda_j^0 \geq 0$, $j = 1, \ldots, n$, $\sum_{j=1}^{n} \lambda_j^0 \geq 1$ such that

$$\sum_{j=1}^{n} \left[a_j + L \times \left(b_j - a_j\right)\right] \lambda_j^0 = \sum_{j=1}^{n} x_j^L \lambda_j^0 < \hat{x}.$$

Therefore, there exists $\hat{\beta}$ that satisfies $\hat{\beta} > L$ such that

$$\sum_{j=1}^{n} x_j^{\hat{\beta}} \lambda_j^0 < \hat{x}.$$

That is, $\hat{x} \in \mathrm{Int}\, T_{\hat{\beta}}$, and thus $\hat{\theta}(\hat{\beta}) < 1$. Consequently, since $\hat{\theta}(1) > 1$, $\hat{\theta}(\hat{\beta}) < 1$, $\hat{\beta} \in (L, +\infty)$, from Theorem 2(i), $\hat{\theta}(\beta)$ is a continuous function defined over $(L, +\infty)$. It follows that there exists $\beta^* \in (L, +\infty)$ such that $\hat{\theta}(\beta^*) = 1$. □

Appendix 4: Uniqueness of β*

The following Theorem 4 shows the uniqueness of β*.

Theorem 4 *Let $b_j > a_j$, $j = 1, \ldots, n$, $L < \bar{\beta} < \hat{\beta}$, and $\hat{x} \in \hat{T}$. Then*

(i) *There is no intersection between the frontiers of $T_{\hat{\beta}}$ and $T_{\bar{\beta}}$.*
(ii) *The quantile of DMU-$\hat{x}$, i.e., β*, is uniquely determined.*

Proof The proof to (i) is achieved by contradiction. If there exists $x^0 \in \Re_+^m$, and x^0 is located on the frontiers of both $T_{\hat{\beta}}$ and $T_{\bar{\beta}}$, then, from Theorem 2, $1 = \hat{\theta}(\bar{\beta}) < \hat{\theta}(\hat{\beta}) = 1$, which is a contradiction. That is, there is no intersection between the frontiers of $T_{\hat{\beta}}$ and $T_{\bar{\beta}}$.

The proof to (ii) is also achieved by contradiction. Assume that there exist two quantiles of DMU-$\hat{x}$, i.e., β_1^* and β_2^*. Without loss of generality, assume that $L < \beta_1^* < \beta_2^*$. Since both

β_1^* and β_2^* are the quantiles of DMU-$\hat{x}$, $\hat{\theta}(\beta_1^*) = \hat{\theta}(\beta_2^*) = 1$. However, from Theorem 2, $\hat{\theta}(\beta_1^*) < \hat{\theta}(\beta_2^*)$. That is, there is a contradiction. It follows that β^* is uniquely determined. □

References

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*, 429–444.

Charnes, A., Cooper, W. W., Wei, Q. L., & Huang, Z. M. (1989). Cone ratio data envelopment analysis and multiobjective programming. *International Journal of Systems Science, 20*(7), 1099–1118.

Cooper, W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-REA: Models for dealing with imprecise data in DEA. *Management Science, 45*, 597–607.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Introduction to data envelopment analysis and its uses: With DEA-solver software and references*. New York: Springer.

Corne, D., Dhaenens, C., & Jourdan, L. (2012). Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal of Operational Research, 221*, 469–479.

Despotis, D. K., & Smirlis, Y. G. (2002). Data envelopment analysis with imprecise data. *European Journal of Operational Research, 140*, 24–36.

Han, J., & Kamber, M. (2007). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufman Publishers.

Kao, C. (2006). Interval efficiency measures in data envelopment analysis with imprecise data. *European Journal of Operational Research, 174*, 1087–1099.

Pendharkar, P. C. (2002). A potential use of DEA for inverse classification problem. *Omega: An International Journal of Management Science, 30*, 243–248.

Pendharkar, P. C. (2011). A hybrid radial basis function and data envelopment analysis neural network for classification. *Computers and Operations Research, 38*, 256–266.

Pendharkar, P. C. (2012). Fuzzy classification using the data envelopment analysis. *Knowledge Based Systems, 31*, 183–192.

Pendharkar, P. C., Khosrowpour, M., & Rodger, J. A. (2000). Application of Bayesian network classifiers and data envelopment analysis for mining breast cancer patterns. *The Journal of Computer Information Systems, 40*(4), 127–132.

Pendharkar, P. C., & Troutt, M. D. (2011). DEA based dimensionality reduction for classification problems satisfying strict non-satiety assumption. *European Journal of Operational Research, 212*, 155–163.

Seifert, J. W. (2004). Data mining: An overview. CRS Report for Congress, The Library of Congress, Order Code RL31798. http://www.fas.org/irp/crs/RL31798.pdf.

Seiford, L. M., & Zhu, J. (1998). An acceptance system decision rule with data envelopment analysis. *Computers and Operations Research, 25*(4), 329–332.

Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems, 46*, 287–299.

Troutt, M. D., Rai, A., & Zhang, A. (1996). The potential use of DEA for credit applicant acceptance systems. *Computers and Operations Research, 23*(4), 405–408.

Wei, Q. L., & Yan, H. (2001). A method of transferring polyhedron between the intersection-form and the sum-form. *Computers and Mathematics with Application, 41*, 1327–1342.

Wei, Q. L., & Yu, G. (1997). Analyzing the properties of K-cone in generalized data envelopment analysis model. *Journal of Econometrics, 80*, 63–84.

Yan, H., & Wei, Q. L. (2000). A method of transferring cones of intersection-form to cones of sum-form and its applications in DEA models. *International Journal of Systems Science, 31*(5), 629–638.

Yan, H., & Wei, Q. L. (2011). Data envelopment analysis classification machine. *Information Science, 181*, 5029–5041.

Ying, M. Q., Xu, R. E., & Wei, Q. L. (1975). Stability of mathematical programming. *Acta Mathematical Sinica, 18*(2), 123–175.

Yu, G., Wei, Q. L., & Brockett, P. (1996). A generalized data envelopment analysis model: A unification and extension of existing methods for efficiency analysis of decision making units. *Annals of Operations Research, 66*, 47–89.

Zhu, J. (2003). Imprecise data envelopment analysis: A review and improvement with an application. *European Journal of Operational Research, 144*, 513–529.