

## COMBINING THE DISCRIMINANT ANALYSIS AND THE DATA ENVELOPMENT ANALYSIS IN VIEW OF MULTIPLE CRITERIA DECISION MAKING: A NEW MODEL

Hasan BAL, H.Hasan ÖRKÜ\*

Gazi Üniversitesi Fen-edebiyat Fakültesi, İstatistik Bölümü  
Teknikokullar, 06500, Ankara, TÜRKİYE, e-mail:hhorkcu@gazi.edu.tr

### ABSTRACT

In this paper, a new classification model which combines the discriminant analysis and the data envelopment analysis and bases on on the multicriteria decision making is developed. Our suggested model utilizes the relative efficiency concept of the data envelopment analysis in predicting group membership of units. The study is supported with an application which examines a few selected socio-economic indicators of some member and candidate countries of European Union.

**Key Words:** Classification, Discriminant Analysis, Data Envelopment Analysis, multiple criteria decision making, efficiency.

## ÇOK KRİTERLİ KARAR VERME AÇISINDAN VERİ ZARFLAMA ANALİZİ İLE DİSKİRİMİNANT ANALİZİNİN BİRLEŞTİRİLMESİ: YENİ BİR MODEL

### ÖZET

Bu çalışmada çok kriterli karar vermeye dayalı olarak, Diskriminant Analizi ile Veri Zarflama Analizi tekniklerinin birleştirildiği yeni bir sınıflama modeli geliştirilmiştir. Önerilen bu model, birimlerin grup üyeliklerinin belirlenmesinde Veri Zarflama Analizi tekniğindeki göreceli etkinlik kavramından yararlanmaktadır. Çalışma, Avrupa Birliğine üye ve üye olmayan ülkelerin seçilmiş bazı sosyo-ekonomik göstergelerini inceleyen bir uygulama ile desteklenmiştir.

**Anahtar Kelimeler:** Sınıflandırma, Diskriminant Analizi, Veri Zarflama Analizi, çok kriterli karar verme, etkinlik

### 1. GİRİŞ

Diskriminant Analizi (DA), birimlerin gözlenen nitelik skorlarına göre uygun sınıflarına atanması işlemi ile uğraşır. Diskriminant Analizini uygulamanın temel amaçlarından biri sınıflandırılması istenen birimlerin grup üyeliğini kestirmektir. Birimleri iki ya da daha fazla gruba sınıflandırmak için en çok kullanılan istatistiksel yöntem, Fisher'in geliştirdiği yöntemdir. Fisher (1), iki ya da daha fazla gruptan gözlenmiş birimleri gruplardan birine sınıflandırmak için değişkenler üzerinden tanımlanacak doğrusal fonksiyonları önermiştir. Bu doğrusal fonksiyonlar, gruplar arası farklılığı maksimum yapacak şekilde alınırlar. Doğrusal programlama teknikleri hiç bir varsayım gerektirmeyen parametrik olmayan yöntemler iken istatistiksel DA yöntemleri verinin normal dağıldığını varsayarlar.

Doğrusal programlama teknikleri ile sınıflandırma probleminin incelenmesi ilk defa Fred ve Glover (2, 3) tarafından yapılmıştır. Fred ve Glover, sapmalar toplamının minimizasyonuna dayanan bir sınıflandırma modeli önermişlerdir. Bajgier ve Hill (4) sınıflandırma probleminde, doğrusal programlama teknikleri ile istatistiksel tekniklerin deneysel bir karşılaştırmasını vermiştir. Bu çalışmalardan sonra, sapmalar toplamının minimizasyonuna, yanlış sınıflandırılmış birimlerin

### 1. INTRODUCTION

Discriminant Analysis (DA) deals with the problem of classifying objects into their groups based on the observed variable scores of these objects. One of the basic aims of applying Discriminant Analysis is to predict group membership of the objects which we want to classify. Fisher's Linear Discriminant Method is an important statistical approach to solve two or multiple group discriminant problem. Fisher (1) recommended linear function approach which is defined on variables for classifying observed scores of objects in to the their appropriate groups. These linear functions are established in such a way that the variation between groups is maximized. Recently, instead of the solving DA problem with these statistical methods, many new efficient Linear Programming (LP) approaches have been studied. In statistical DA the data is assumed to have normal distributions, whereas LP methods are non-parametric and make no distributional assumptions.

Firstly, Fred and Glover suggested the LP approach to classification problems in DA (2, 3) Fred and Glover suggested a classification model which bases on minimizing the sum of deviations. It has been given by Bajgier and Hill (4) an experimental comparison of between statistical DA and LP approaches to the

sayısının minimizasyonuna ve gruplar arası uzaklıkların maksimizasyonu gibi sınıflandırma kriterlerine dayalı bir çok teknik geliştirilmiştir (5, 6, 7).

Veri Zarflama Analizi (VZA) ile Doğrusal Diskriminant Analizini inceleyen ilk çalışmalar Retzlaff-Roberts (8) tarafından verilmiştir. Bu çalışmalarda DA ve VZA'nın benzerlik ve farklılıkları tartışılmış ve her iki tekniğinde uygulandığı veri seti problemlerinin özellikleri tanımlanmıştır. Suoyeshi (9,10) çalışmalarında her iki tekniği de Hedef programlama ışığında incelemiş ve iki aşamadan oluşan yeni bir sınıflandırma modeli önermiştir. Suoyeshi'nin çalışmaları iki gruba da ait olan birimlerin (kesişme durumu) belirleme ve doğrusal olmayan kısıtları da içeren farklı bir bakış açıdır. Retzlaff-Roberts'in VZA'daki etkinlik kısıtlarına dayalı kurduğu modeli genel bir etkinlik kavramını yansıtmaktadır. Yani birimler için elde edilen sınıflandırma skorları, VZA kısıtlarını içeren fakat sapmalar toplamının minimizasyonu modelinden elde edilmiştir (11). Önerdiğimiz model, her birimin sınıflandırma skorlarını kendi etkinlik skorları ile ilişkilendiren ve sapmalar toplamının minimizasyonunu da garanti eden bir sınıflandırma modelidir.

## 2. DİSKİRİMİNANT ANALİZİ

Doğrusal programlama yöntemleri ile doğrusal diskriminant fonksiyonunun belirlenmesi, ilk olarak Fred ve Glover (3) tarafından ileri sürülmüştür. Doğrusal programlama ile  $k$  değişkenli bir sınıflama problemi incelendiğinde, istatistiksel DA'ya benzer olarak  $X$ ;  $G_1$  ve  $G_2$  gruplarından alınmış  $n$  tane bireyden oluşan bir örneğin değişken değerlerini gösteren  $(n \times p)$  boyutlu bir matrisi göstermektedir. Bu durumda  $x_{ij}$ ;  $i$ . bireyin  $j$ . değişken değerini gösterir. İkili sınıflandırmada en çok kullanılan ve sapmalar toplamının minimizasyonuna dayanan doğrusal programlama modeli (MSD-Minimum Sum of Distance) aşağıda verilmiştir (3,5).

MSD

$$\min \sum_{i=1}^n d_i$$

$$\sum_{j=1}^k w_j x_{ij} - d_i \leq c \quad i \in G_1$$

$$\sum_{j=1}^k w_j x_{ij} + d_i \geq c \quad i \in G_2$$

$$d_i \geq 0, \quad i=1,2, \dots, n$$

[1]

$w, c$  kısıtsız.

$d_i$  dış sapma değişkeni, yanlış sınıflandırılan bir birimin kendi grubu dışında bulunduğu uzaklıktır.  $c$  kesme değeri (kritik değer, eşik değeri)  $w$  ile beraber iki grubu ayıran bir hiperdüzlem tanımlar. Herhangi bir örneğin ayırma skoru  $\sum_{j=1}^k w_j x_{ij}$ ,  $c$ 'ye eşit ya da küçükse 1. Gruba ( $G_1$ ), diğer durumda 2. Gruba ( $G_2$ ) sınıflandırılır.

Bu model ( $w=0, c=0$ ) elde edilmesi gibi anlamsız

classification problems. Later on these studies, many LP approaches based on the classification criteria such as, minimizing the sum of the deviations, minimizing the sum of misclassified objects and maximizing the distance of intra groups have been developed (5, 6, 7).

Firstly, Retzlaff-Roberts (8) proposed a study relating DA and DEA to one another. In these studies, It has been discussed the similarities and the dissimilarities of both DA and DEA, and defined the properties of data set problems for applying these two methods. Suoyeshi (9,10) addressed both of the DA and DEA method in view of goal programming, and proposed a new classification model which involves two phases. Suoyeshi's approach has a different point of view which determines the overlapping objects in the groups, and involves nonlinear constraints. Retzlaff-Roberts approach which is based on efficiency model constraints reflects a concept of common efficiency in DEA in that approach. Classification scores obtained for each objects can be obtained by a model involving DEA constraints and minimizing the sum of deviations (11). Our proposed classification model ensures that classification scores for each objects are related to their efficiency scores, and is considered as minimizing the sum of deviations.

## 2. DISCRIMINANT ANALYSIS

Firstly, Fred and Glover suggested the LP approach to the classification problems in DA (3). In considering a classification problem with  $k$  variables for two groups by LP methods like as statistical DA method, let  $X$  denote the matrix of dimension  $n \times p$  representing the values of known sample of  $n$  objects from the  $G_1$  and  $G_2$ .

Hence  $x_{ij}$  denote the value of  $j$ . variable for the  $i$ . object in the sample. The minimization of the sums of individual deviations model (MSD) can be formulated as follows (3,5).

MSD

$$\min \sum_{i=1}^n d_i$$

$$\sum_{j=1}^k w_j x_{ij} - d_i \leq c \quad i \in G_1$$

$$\sum_{j=1}^k w_j x_{ij} + d_i \geq c \quad i \in G_2$$

$$d_i \geq 0, \quad i=1,2, \dots, n$$

[1]

$w, c$  unrestricted variables.

where  $d_i$  is known to as the external deviation variable representing the amount by which the object is misclassified;  $w$  are defined as variables for weights and  $c$  is the cut-off value (critical value, threshold value), which together with  $w$ , defines a hyperplane that attempts to separate the two groups. Any object will be classified into group 1 ( $G_1$ ) if its classification score  $\sum_{j=1}^k w_j x_{ij}$  is less than or equal to  $c$ , otherwise it will be

sonuçlar verebilir. Anlamli sonuçlar elde edebilmek için küçük bir  $\mathcal{E}$  pozitif sayısına ya da bir normalizasyon kısıtı ayarlamasına ihtiyaç duyulur.  $\mathcal{E}$  pozitif sayısı modele ilave edilerek [1] modeli [2] olarak yeniden yazılabilir (11).

$$\begin{aligned} & \min \sum_{i=1}^n d_i \\ & \sum_{j=1}^k w_j x_{ij} - d_i \leq c - \mathcal{E} \quad i \in G_1 \\ & \sum_{j=1}^k w_j x_{ij} + d_i \geq c \quad i \in G_2 \\ & d_i \geq 0, \quad i=1,2, \dots, n \\ & W, C \text{ kısıtsız} \end{aligned} \quad [2]$$

En bilinen yaklaşım, anlamsız sonuçlardan kaçınmak için modele bir normalizasyon kısıtı eklenmesini önerir. Bu normalizasyon kısıtının kurulması için bir çok farklı yol önerilmiştir. Glover (12) tarafından verilen Diskriminant analizi için doğrusal programlama formülasyonunda kullanılan normalizasyon kısıtlı model; [3]'de verilmiştir.

$$\begin{aligned} & \min h_0 d_0 + \sum_{j=1}^n h_j d_j - k_0 \beta_0 - \sum_{j=1}^n k_j \beta_j \\ & \sum_{i=1}^m w_{ij} x_{ij} - d_0 - d_j + \beta_0 + \beta_j = c \quad j \in G_1 \\ & \sum_{i=1}^m w_{ij} x_{ij} + d_0 + d_j - \beta_0 - \beta_j = c \quad j \in G_2 \\ & 2n_1 n_2 (\beta_0 - d_0) + n_2 \sum_{j \in G_1} (\beta_j - d_j) + n_1 \sum_{j \in G_2} (\beta_j - d_j) = 1 \\ & d_0, d_j, \beta_0, \beta_j \geq 0, \quad i=1,2, \dots, n \\ & W, C \text{ kısıtsız} \end{aligned} \quad [3]$$

$d_j$  dış sapma değişkeni, yanlış sınıflandırılan bir birimin kendi grubu dışında bulunduğu uzaklığı ve  $\beta_j$  iç sapma değişkeni ise kendi grubu içerisinde bulunan doğru sınıflandırılmış bir birimin kendi grubuna olan uzaklığıdır. Bu iki sapma değişkeni, örneklerin grupların dışında ya da içinde olduğunun bir büyüklüğünü (derecesini) yansıtır.  $d_0$  ve  $\beta_0$  sırasıyla, maksimum dış ve minimum iç sapmayı temsil eder.  $h_j$  ve  $k_j$  ise sapmalara karşı atanan ağırlıklardır. Böylelikle; iç sapmalar maksimize dış sapmalar minimize edilmeye çalışılır. Bu modelin dezavantajı ise parametre değerlerinin belirlenmesinin zorluğu yanında çok karmaşık olmasıdır (8,11).

Hedef Programlama ışığında Suoyeshi' nin sınıflandırmayı iki aşamada yapan modeli (DEA-DA) ise aşağıdaki gibi verilebilir (9).

classified into group 2 ( $G_2$ ).

This model may give the trivial solution ( $w=0, c=0$ ) which is meaningless. In order to obtain meaningful optimal solution to the model [1], a small positive number  $\mathcal{E}$  or a normalization constraint condition is often needed. When the small positive number  $\mathcal{E}$  is introduced into model [1], the model can then be rewritten as model [2], (11).

$$\begin{aligned} & \min \sum_{i=1}^n d_i \\ & \sum_{j=1}^k w_j x_{ij} - d_i \leq c - \mathcal{E} \quad i \in G_1 \\ & \sum_{j=1}^k w_j x_{ij} + d_i \geq c \quad i \in G_2 \\ & d_i \geq 0, \quad i=1,2, \dots, n \\ & W, C \text{ unrestricted variables} \end{aligned} \quad [2]$$

Most common approach to avoid trivial solution is to append a normalization constraint to the model [1]. There are several different ways to construct a normalization constraint as given. The general formulation of LP for DA with the normalization constraint given by Glover (12) is given in model [3].

$$\begin{aligned} & \min h_0 d_0 + \sum_{j=1}^n h_j d_j - k_0 \beta_0 - \sum_{j=1}^n k_j \beta_j \\ & \sum_{i=1}^m w_{ij} x_{ij} - d_0 - d_j + \beta_0 + \beta_j = c \quad j \in G_1 \\ & \sum_{i=1}^m w_{ij} x_{ij} + d_0 + d_j - \beta_0 - \beta_j = c \quad j \in G_2 \\ & 2n_1 n_2 (\beta_0 - d_0) + n_2 \sum_{j \in G_1} (\beta_j - d_j) + n_1 \sum_{j \in G_2} (\beta_j - d_j) = 1 \\ & d_0, d_j, \beta_0, \beta_j \geq 0, \quad i=1,2, \dots, n \\ & W, C \text{ unrestricted variables.} \end{aligned} \quad [3]$$

where  $d_j$  and  $\beta_j$  stand for the external and the internal deviation variables, respectively, which refer to magnitudes by which the samples lie outside or inside their groups.  $d_0$  and  $\beta_0$  represent the maximum external and the minimum internal deviations, respectively.  $h_j$  and  $k_j$  are weights assigned to corresponding deviations. Thus, external deviations are maximized and internal deviations are minimized. The disadvantage of this model in addition to difficulty of determining the parameter values, should be very complex (8,11).

In view of the goal programming Suoyeshi's DEA-DA model which tries to classify objects in two stage can be given as follow (11).

$$\begin{aligned} \min \quad & \sum_{j \in G_1} s_j^+ + \sum_{j \in G_2} s_j^- \\ \sum_{i=1}^k \alpha_i z_{ij} + s_j^+ - s_j^- &= d \quad j \in G_1 \\ \sum_{i=1}^k \alpha_i z_{ij} + s_j^+ - s_j^- &= d - \eta \quad j \in G_2 \\ s_j^+, s_j^- &\geq 0, \end{aligned} \tag{4}$$

$\alpha_i, d$  serbest,  $\eta$  çok küçük pozitif bir sayı

İlk aşamada bu model çözüldükten sonra  $\alpha_i^*$  ve ayırma değeri  $d^*$  bulunur. İkinci aşamada ise,

$$\begin{aligned} \sum_{i=1}^k \alpha_i^* z_{ij} &\geq d^* \quad j \in G_1 \\ \sum_{i=1}^k \alpha_i^* z_{ij} &< d^* \quad j \in G_2 \end{aligned}$$

kriterleri ile de bireyler uygun gruplarına sınıflandırılır.

Hem sapmalar toplamının minimizasyonuna hem de Veri Zarflama Analizine dayanan, Retzlaff-Roberts' in tek aşamalı sınıflandırma modeli (R-R) ise aşağıdaki gibi formüle edilebilir (8).

$$\begin{aligned} \min \quad & \sum_{j=1}^n d_j \\ \sum_{r=1}^s u_r y_{rj} - T \sum_{i=1}^m v_i x_{ij} - d_j &\leq 0 \quad j \in G_1 \\ \sum_{r=1}^s u_r y_{rj} - T \sum_{i=1}^m v_i x_{ij} + d_j &\geq 0 \quad j \in G_2 \\ \sum_{j=1}^n \left( \sum_{r=1}^s u_r y_{rj} \right) &= n \\ u_r, v_i, d_j &\geq 0 \end{aligned} \tag{5}$$

burada,  $T$  önceden belirlenen herhangi bir pozitif sayıdır.

### 3. VERİ ZARFLAMA ANALİZİ

Veri Zarflama Analizi ilk olarak Charnes, Cooper ve Rhodes (13) tarafından, ürettikleri mal veya hizmet açısından birbirlerine benzer ekonomik KVB (Karar Verme Birimi) 'lerinin göreceli etkinliklerinin ölçülmesi amacı ile geliştirilmiş olan parametrik olmayan bir etkinlik yöntemidir. Bu yöntemin sahip olduğu özellikleri kısaca özetlersek; her KVB'deki etkinsizlik miktarını ve kaynaklarını tanımlayabilmesi, her bir KVB'nin etkinlik değeri diğerlerine göre hesaplandığından hesaplanan etkinliklerin göreceli etkinlikler olması ve değişkenler üzerinde herhangi bir fonksiyonel varsayım öne sürmemesidir (13).

Farrell (13)'in fikirlerini geliştiren Charnes, Cooper ve Rhodes tek bir çıktının tek bir girdiye oranlanmasıyla elde edilen etkinlik değerini, çoklu çıktılardan çoklu girdilere

$$\begin{aligned} \min \quad & \sum_{j \in G_1} s_j^+ + \sum_{j \in G_2} s_j^- \\ \sum_{i=1}^k \alpha_i z_{ij} + s_j^+ - s_j^- &= d \quad j \in G_1 \\ \sum_{i=1}^k \alpha_i z_{ij} + s_j^+ - s_j^- &= d - \eta \quad j \in G_2 \\ s_j^+, s_j^- &\geq 0, \end{aligned} \tag{4}$$

$\alpha_i, d$  unrestricted variables and,  $\eta$  very small positive number

In the first stage, after the solution to this model, we can find out  $\alpha_i^*$  values and  $d^*$  cut off value. Then, in the second stage, we should solve the following formulation.

$$\begin{aligned} \sum_{i=1}^k \alpha_i^* z_{ij} &\geq d^* \quad j \in G_1 \\ \sum_{i=1}^k \alpha_i^* z_{ij} &< d^* \quad j \in G_2 \end{aligned}$$

and after the solution to this formulation, objects can be classified in to group 1 ( $G_1$ ) and group 2.

Retzlaff-Roberts' one- phased model which is based on minimizing of the sum of deviations and the DEA can be as follows (8):

$$\begin{aligned} \min \quad & \sum_{j=1}^n d_j \\ \sum_{r=1}^s u_r y_{rj} - T \sum_{i=1}^m v_i x_{ij} - d_j &\leq 0 \quad j \in G_1 \\ \sum_{r=1}^s u_r y_{rj} - T \sum_{i=1}^m v_i x_{ij} + d_j &\geq 0 \quad j \in G_2 \\ \sum_{j=1}^n \left( \sum_{r=1}^s u_r y_{rj} \right) &= n \\ u_r, v_i, d_j &\geq 0 \end{aligned} \tag{5}$$

where,  $T$  is a any predetermined positive number.

### 3. DATA ENVELOPMENT ANALYSIS

DEA, non parametric method of evaluating relative efficiencies for groups of similar units in point of view of the produced product and service, was introduced by Charnes, Cooper and Rhodes (13). The summary of the main characteristics of DEA method are to be able to identify the sources and the level of inefficiency for each DMU and their evaluated efficiencies are relative efficiencies since the level of efficiency of each DMU is obtained with respect to the other units, and making no assumptions on the variables (13).

Charnes, Cooper and Rhodes who developed Farrell's (13) idea extended the single-output/input ratio measure of efficiency to the multiple output / input measure of efficiency. Then the relative efficiency of any DMU is calculated by forming the ratio of a weighted sum of

oranlanmasına genişletmişlerdir. Bu sayede her bir KVB için yapay bir çıktı ve yapay bir girdi bulunmakta ve bu yapay çıktı ve girdiler vasıtasıyla KVB'lerin etkinlik değerleri bulunabilmektedir. Burada ağırlıklar, etkinlik değerlerini 1'den büyük yapmayacak şekilde seçilirler.

Veri Zarflama Analizinde etkinlik ölçümünde kullanılan çeşitli modeller vardır ve bu modeller etkinliğin çıktıların ağırlıklı toplamının, girdilerin ağırlıklı toplamına oranı olarak ölçüldüğü oran modelinden türetilir (13).  $x_{ij}$  ( $i=1,2, \dots, m$ ) ve  $y_{rj}$  ( $r=1,2, \dots, s$ ) olarak  $j$ . örneğin (birimin)  $i$ . girdisini ve  $r$ . çıktısını tanımlamak üzere oran formu aşağıdaki gibi tanımlanabilir:

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} / \sum_{i=1}^m v_i x_{io} \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \leq 1 \quad j=1,2, \dots, n \\ & u_r, v_i \geq 0 \end{aligned} \quad [6]$$

Etkinliği ölçülen KVB'nin ağırlıklı girdi toplamı ( $\sum_{i=1}^m v_i x_{io}$ ) 1'e eşitlenir ise CCR modeli olarak bilinen temel etkinlik modeli elde edilmiş olur.

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j=1,2, \dots, n \\ & \sum_{i=1}^m v_i x_{io} = 1 \\ & u_r, v_i \geq 0 \quad i=1,2, \dots, m ; r=1,2, \dots, s \end{aligned} \quad [7]$$

Bu model ile KVB'lerin etkinliği ölçülürken modelin her KVB için yani  $n$  defa çözülmesi gerekmektedir. Optimal amaç fonksiyonu ilgili KVB'nin etkinlik skorunu vermektedir. Her bir KVB için farklı  $u_r, v_i$  ağırlık kümesi seçilecektir. Bu sayede optimal ağırlık seti her KVB için bir düzlem tanımlar. Etkinlik skoru 1'e eşit olan her KVB etkin olarak değerlendirilir. Etkinlik skoru 1'in altında olan her birim de etkin olmayan olarak değerlendirilecektir.

VZA' da değişkenlerin girdi ve çıktı olarak ayrılması gerekir. Değişkenlerin, girdi ve çıktı olarak ayrılması birim üzerindeki etkilerine bağlıdır. Retzlaff-Roberts, girdi ve çıktı değişkenleri yerine birimler üzerinde pozitif ve negatif etkili değişkenler kavramını kullanmayı uygun buldu. Artışı birimin daha iyi olarak değerlendirilmesini sağlayan değişkenlerin *pozitif etkili*, tersine düşüşü birimin daha iyi olarak değerlendirilmesini sağlayan değişkenlerin ise *negatif etkili* olarak alınmasını önermiştir (8,11).

outputs to a weighted sum of inputs, where the weights for both outputs and inputs are to be selected in a manner that calculates efficiency measure of each DMU subject to the constraint that no DMU can have relative efficiency score greater than unity.

In DEA there are many models which can be used to measure of efficiency, and these models are derived from the ratio models in which the weighted sum of efficiency outputs are measured as the ratio to the weighted sum of inputs (13). Considering as  $n$  units each of which has  $m$  inputs denoted by  $x_{ij}$  ( $i=1,2, \dots, m$ ) and  $s$  outputs denoted by  $y_{rj}$  ( $r=1,2, \dots, s$ ), the mathematical programming problem of ratio form can be given as follows:

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} / \sum_{i=1}^m v_i x_{io} \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \leq 1 \quad j=1,2, \dots, n \\ & u_r, v_i \geq 0 \end{aligned} \quad [6]$$

If the sum of weighted inputs of DMU whose efficiency is measured is made equal to 1 (i.e.  $\sum_{i=1}^m v_i x_{io} = 1$ ), then the CCR model is obtained as follows.

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j=1,2, \dots, n \\ & \sum_{i=1}^m v_i x_{io} = 1 \\ & u_r, v_i \geq 0 \quad i=1,2, \dots, m ; r=1,2, \dots, s \end{aligned} \quad [7]$$

While it has been measured efficiency of DMUs by this model, it is necessary to solve model  $n$ -times for each DMU. The optimum value of objective function gives the efficiency score of the interested DMU in the model. It will be selected different set of weights  $u_r, v_i$  for each DMU. Therefore, the set of optimum weight identifies a hyperplane for each DMU. Any DMU whose efficiency score equals to one is defined as efficient, otherwise inefficient.

It is necessary to separate the variables as input and output in DEA. This separation depends on their effects related to DMU. Instead of input and output variables, Retzlaff-Roberts appropriated to using the concept of positive effect variables and negative effect variables on units. Any factor whose increase (while others are held constant) leads to unit considered better or more efficient is defined as *positive*. On the contrary, any factor whose decrease (while others are held constant) leads to unit considered better or more efficient is defined as *negative* (8,11).

#### 4. DİSKİRİMİNANT ANALİZİ VE VERİ ZARFLAMA ANALİZİNİN BİRLEŞTİRİLMESİ

VZA, KVB'lerin göreceli etkinliklerinin ölçülmesi amacıyla geliştirilmiş olan parametrik olmayan bir etkinlik yöntemidir. DA ve VZA doğrusal programlama kullanılarak bir grup benzer birim ya da gözlemin performanslarını hesaplama yöntemleri olarak düşünülebilir. Bir gruba ait olanı ayırt etmek için her iki yöntem de bir takım ağırlık faktörlerini kullanırlar. DA' da her bir birimin hangi grubun üyesi olduğu başlangıçta bilinirken, iki grup arasında en iyi ayırımı sağlayan faktör ağırlıkları ve eşik değerleri araştırılır. Genelde VZA iki gruba sınıflama tekniği olarak bilinmekle birlikte, birimleri etkin olmayan ve VZA-etkin olan olarak iki gruba ayırır. VZA' da başlangıçta birimlerin hangi gruba ait oldukları bilinmezken, iki grubu ayıran eşik değeri bilinir; etkinlik skoru 1'in altında olan birim etkin olmayan olarak değerlendirilirken, diğerleri VZA-etkin olarak değerlendirilir.

Bir Diskiriminant analizi probleminde  $m$  tane girdi değişkeni ve  $S$  tane çıktı değişkeni sınıflama değişkenleri olarak göz önüne alınsın.  $x_{ij}$  ( $i=1,2, \dots, m$ ) ve  $y_{rj}$  ( $r=1,2, \dots, s$ ) ile  $j$ . örneğin (birimin)  $i$ . girdisini ve  $r$ . çıktısını gösterelim. Veri Zarflama Analizinin avantajlarından biri çoklu çıktı ve çoklu girdileri kullanabilmesidir. Sınıflama kriteri olarak çoklu girdiler ve çoklu çıktılar üzerinden tanımlanan etkinlik kavramı kullanıldığında 1 değeri (kesme değeri, eşik değer) iki grubu birbirinden ayırabilir:

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} / \sum_{i=1}^m v_i x_{io} \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \geq 1 \quad j \in G_1 \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \leq 1 \quad j \in G_2 \\ & u_r, v_i \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \quad [8]$$

Bu model doğrusal programlamaya dönüştürüldüğünde [9] ile verilen modele ulaşılır.

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \geq 0 \quad j \in G_1 \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j \in G_2 \\ & \sum_{i=1}^m v_i x_{io} = 1 \quad j = 1, 2, \dots, n \\ & u_r, v_i \geq 0 \quad i = 1, 2, \dots, m ; r = 1, 2, \dots, s \end{aligned} \quad [9]$$

Yanlış sınıflandırma oranının ölçüsü olan  $d_j$  dış sapma değişkenleri alınarak, bu dış sapmalar toplamının minimizasyonu kriterinin yukarıdaki modele eklenmesiyle aşağıdaki çok kriterli karar modeli elde edilir.

#### 4. COMBINING DISCRIMINANT ANALYSIS AND DATA ENVELOPMENT ANALYSIS

DA and DEA can be considered as methods to calculate the performance of for the purpose of measuring the DMU's by using LP. In order to distinguish the members of a group, each method uses some factor weights. In DA it is initially known which group each unit is a member of, and a set of factor weights and threshold value are sought which produce the best possible separation of two groups. While DEA is not generally thought of as a two-group classification technique, it does classify units into the two groups inefficient and DEA-efficient. In DEA it is not initially known to which group units belong, but threshold that separates the two groups is known, an efficiency score below 1 categorizes a unit as inefficient and others are deemed DEA-efficient.

Consider a DA problem, whose classification attributes consist of  $m$  input attributes and  $S$  output attributes. We denote  $x_{ij}$  ( $i=1,2, \dots, m$ ) and  $y_{rj}$  ( $r=1,2, \dots, s$ ) as the values of the  $i$ . input and  $r$ . output attributes of the  $j$ . sample. One of the advantages of DEA is its capability of dealing with multiple inputs and multiple outputs. When the efficiency of multiple inputs and multiple outputs is used as the classification criterion, it is naturally hoped that there can exist an efficiency cut of value 1 (threshold, boundary value) that separates two groups.

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} / \sum_{i=1}^m v_i x_{io} \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \geq 1 \quad j \in G_1 \\ & \sum_{r=1}^s u_r y_{rj} / \sum_{i=1}^m v_i x_{ij} \leq 1 \quad j \in G_2 \\ & u_r, v_i \geq 0 \quad j = 1, 2, \dots, n \end{aligned} \quad [8]$$

This model can be transformed into an equivalent linear programming model, obtained in [9].

$$\begin{aligned} \max \quad & \sum_{r=1}^s u_r y_{ro} \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \geq 0 \quad j \in G_1 \\ & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j \in G_2 \\ & \sum_{i=1}^m v_i x_{io} = 1 \quad j = 1, 2, \dots, n \\ & u_r, v_i \geq 0 \quad i = 1, 2, \dots, m ; r = 1, 2, \dots, s \end{aligned} \quad [9]$$

Considering the deviation variable  $d_j$  which is a measure of missclassification ratio, we introduce the external deviation variables and its minimum sum criterion into the above model, which results in a multi objective linear programming model shown as follows:

$$\begin{aligned}
 & \max \sum_{r=1}^s u_r y_{ro} \\
 & \min \sum_{j=1}^n d_j \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - d_j \geq 0 \quad j \in G_1 \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + d_j \leq 0 \quad j \in G_2 \\
 & \sum_{i=1}^m v_i x_{io} = 1 \quad j = 1, 2, \dots, n \\
 & u_r, v_i, d_j \geq 0 \quad i = 1, 2, \dots, m ; r = 1, 2, \dots, s
 \end{aligned}
 \tag{10}$$

$$\begin{aligned}
 & \max \sum_{r=1}^s u_r y_{ro} \\
 & \min \sum_{j=1}^n d_j \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} - d_j \geq 0 \quad j \in G_1 \\
 & \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + d_j \leq 0 \quad j \in G_2 \\
 & \sum_{i=1}^m v_i x_{io} = 1 \quad j = 1, 2, \dots, n \\
 & u_r, v_i, d_j \geq 0 \quad i = 1, 2, \dots, m ; r = 1, 2, \dots, s
 \end{aligned}
 \tag{10}$$

[10] modeli dış sapmalar toplamının minimizasyonunu ile VZA’ da ilgili KVB için ağırlıklı çıktı toplamının maksimizasyonuna dayalı bir sınıflandırma modelidir. Bu modeli ÇKVZA-DA (Çok Kriterli Veri Zarflama Analizine dayalı sınıflandırma modeli) olarak isimlendirelim. ÇKVZA-DA modeli, çok kriterli (amaçlı) doğrusal programlama modelidir. Çok amaçlı doğrusal programlama modellerinin çözümü için çok sayıda çözüm yöntemi mevcuttur. Stuer’ in Daralan Koni Yöntemi, Çok Kriterli Simpleks Yöntemi, Uzlaşık Programlama ve Hedef Programlama gibi (14). ÇKVZA-DA çok amaçlı doğrusal programlama modelinin çözümü için Öncelikli Hedef Programlama yöntemi kullanılacaktır. Model, çıktıların ağırlıklı toplamının maksimum yapılmasına birinci öncelik, sapmalar toplamının minimum yapılmasına da ikinci öncelik verilerek öncelikli Hedef programlama tekniği ile çözülür. Yukarıdaki modelin çözümünden sonra, her birimin sınıflandırma skoru elde edilecektir. Bu skorlar 1 ile karşılaştırılır, eğer skor 1’ den büyük ise birim 1. Gruba, aksi takdirde 2. Gruba atanır.

Aşağıdaki hipotetik bir veri olarak, önerdiğimiz model ve diğer modellerin performansını göstermeye çalışalım.

Önce iki girdi ve bir çıktı değişkeni bakımından 13 bireyin (ilk altı elemanı birinci grup, diğerleri ikinci grup) önerdiğimiz model ve diğer modelleri kullanarak sınıflandırma performansını görelim.

The model [10] is a classification model based on both minimizing sum of external deviations and maximizing sum of weighted outputs for the related DMUs. We named this model as MCDEA-DA (classification model which bases on multiple criteria data envelopment analysis). The model MCDEA-DA is a multi criteria (objective) linear programming model. There are several solution methods in order to solve the multiple objective linear programming model as Stuers’s narrowing cone, multiple criteria simplex method, compromise programming and goal programming. We used preemptive goal programming in order to solve the MCDEA-DA multi objective linear programming model (14). This model can be solved by preemptive goal programming, where first priority is to maximize the sum of weighted outputs and second priority is to minimize the sum of deviation variables. After the solution to the above model, it will be obtained the classification score for each unit. These scores are compared with 1, if it is greater than 1 we assign the unit to first group, and otherwise to second group.

Let us consider the following a hypotheticalal data, and try to compare the performance of our model and other models.

Firstly, we attempt to show the classification performance of our model and of other models for 13 objects with two input variables and one output variable (first six member of objects denote G1 group, the others denote G2 group)

**Table1.** The values of variables for a hypothetical data and obtained results by 5 methods  
**Çizelge 1.** Hipotetik örneğe ilişkin değişken değerleri ve 5 yöntemden elde edilen sonuçlar

	$x_1$	$x_2$	$y_1$	FLDF	MSD	ÇKVZA-DA MCDEA-DA	DEA-DA	R-R
1	300	210	506	2.68	1.67	2.28	3.78	2.78
2	398	252	412	0.67	1.18	3.62	4.96	3.26
3	325	264	756	1.12	0.94	3.78	3.64	1.44
4	537	206	612	0.98	0.83	1.46	2.54	1.69
5	452	390	582	2.24	1.75	1.57	3.12	3.64
6	400	512	961	3.68	1.87	2.96	2.45	2.57
7	102	64	261	-1.67	0.24	0.78	1.57	0.73
8	245	269	476	0.46*	0.67*	0.94	2.32*	1.24*
9	157	108	206	-0.96	0.34	0.64	1.13	0.88
10	96	190	262	-0.78	0.37	0.68	0.96	0.76
11	189	200	384	0.12*	0.58*	0.82	1.23	1.06*
12	76	164	192	-1.23	0.08	0.45	0.68	0.47
13	232	168	301	0.08*	0.41	0.57	1.04	0.51

Burada (\*) bireyin yanlış sınıflandırıldığını göstermektedir. Küçük boyutlu bu örnekte önerdiğimiz model, diğer modellere göre daha iyi bir performans göstermiştir. Çünkü yanlış sınıflanan birimi bulunmamaktadır. Oysa diğer modellerde yanlış sınıflanan birimlere rastlanmıştır. Her durumda modellimizin çok üstün olduğunu söylemek erkendir. Böyle durumlarda modellerin gerçek performansı simülasyon çalışması sonucunda ortaya çıkarılabilir. Aşağıda da gerçek bir veri üzerinde bütün modellerin performansını incelemeye çalışalım.

## 5. UYGULAMA

Bu bölümde Avrupa Birliği üyesi 15 ülke ( $G_1$ ; 1.grup) ve birliğe üye olmayan 8 ülkenin ( $G_2$ ; 2.grup) seçilmiş sosyo-ekonomik göstergeleri kullanılarak, Fisher'in geliştirdiği istatistiksel yöntem (FLDF), Fred ve Glover'in geliştirdiği sapmalar toplamının minimizasyonuna dayanan doğrusal programlama yaklaşımı (MSD), Suoyeshi'nin DEA-DA modeli, Retzlaff-Roberts' in (R-R) modeli ve çalışmada önerilen Çok Kriterli VZA tekniğine dayalı sınıflandırma modeli (ÇKVZA-DA) karşılaştırılacaktır. Çalışmada kullanılan değişkenler (15) ve (16)' den derlenmiştir ve aşağıda gösterildiği gibidir.

$x_1$  : İşsizlik oranı (%)

$x_2$  : Bebek ölüm oranı (1000 kişi başına)

$y_1$  : Kişi başına düşen GSYİH (cari fiyat, \$)

$y_2$  : Eğitime ayrılan pay (\$)

Bölüm 3' te ifade edildiği üzere VZA' da değişkenlerin girdi-çıktı veya pozitif etkili-negatif etkili olarak ayrılması gerekir. Burada  $x_1$  ve  $x_2$  negatif etkili,  $y_1$  ve  $y_2$  ise pozitif etkili değişkenler olarak alınmıştır. Tablo 2, Avrupa Birliği üyesi 15 ülke ve birliğe üye olmayan 8 ülkeye ait standartlaştırılmış değişken değerlerini ve beş yöntemle elde edilen sınıflandırma skorlarını göstermektedir.

Here (\*) denotes misclassified objects. In this small dimensional example, our model has a better classification performance than the other models, because there is no misclassified objects in our model, but, there are some misclassified objects in the other models. We do not want to say that our model is better in all cases. In those cases it can be revealed the real classification performance of models by a simulation study. For this reason, we try to investigate the classification performance of all models on real data in following

## 5. APPLICATION

In this section, using selected socio-economic variables for 15 europe union members country ( $G_1$ , group 1) and 8 non union members country ( $G_2$ , group 2) we will compare the Fisher's Linear Discriminant Function (FLDF), Fred ve Glover's (MSD), Suoyeshi's DEA-DA model, Retzlaff-Roberts' (R-R) model and our suggested underlying multiple criteria and DEA model (MCDEA-DA). In this study that we have used variables are obtained in references (15) and (16) and shown as bellow:

$x_1$  : unemployment rate (%)

$x_2$  : infant deat rate ( per thousand people)

$y_1$  : per capita gross national product (current price, \$)

$y_2$  : expenditure on education(\$)

As is expressed in section 3, it is neccessary that the variables should be seperated as input-output or positive efficient and negative efficient variables in DEA.Here  $x_1$ ,

$x_2$  and  $y_1$ ,  $y_2$  defined respectively, as negative efficient variables and positive efficient variables. In Table 2, it is shown the values of standardized variables and classification scores by using all methods for 15 Europe union members country and 8 non union members country.



**Table 2.** The values of standardized input and output variables, and the obtained classification scores by using all methods for 15 Europe union members country and 8 non union Europe members country**Çizelge 2.** Avrupa Birliğine üye 15 ülke ve üye olmayan 8 ülkeye ait standartlaştırılmış değişken değerleri ve tüm yöntemlerin kullanılması ile elde edilen sınıflandırma skorları

		$x_1$	$x_2$	$y_1$	$y_2$	FLDF	MSD	ÇKVZA-DA MCDEA-DA	DEA- DA	R-R
1	Belgium/Belçika	0.504	0.147	0.567	0.373	1.641	0.476	2.346	9.623	1.962
2	Denmark/Danimarka	0.288	0.220	0.764	0.975	1.555	0.427	4.043	10.762	3.712
3	Germany/Almanya	0.506	0.123	0.604	0.578	1.827	0.410	2.981	8.632	3.621
4	Greece/Yunanistan	0.583	0.156	0.271	0.373	1.312	0.514*	1.069	4.212	1.968
5	Spain/İspanya	1	0.135	0.346	0.602	1.451	0.449	1.554	5.614	1.044
6	France/Fransa	0.636	0.128	0.557	0.722	0.862	0.434	2.648	6.321	2.762
7	Ireland/İrlanda	0.418	0.155	0.546	0.722	0.712	0.517*	2.215	7.826	1.672
8	Italy/İtalya	0.627	0.138	0.479	0.590	1.492	0.457	2.116	5.621	3.012
9	Luxembourg/Lüksemburg	0.138	0.120	1	0.481	1.912	0.381	6.017	14.231	4.621
10	Holland/Hollanda	0.225	0.134	0.580	0.614	1.613	0.444	2.994	8.621	3.021
11	Austria/Avusturya	0.246	0.121	0.605	0.650	1.441	0.409	3.321	8.761	2.125
12	Portugal/Portekiz	0.297	0.154	0.259	0.698	-0.09*	0.513*	1.130	4.312	1.171
13	Finland/Finlandya	0.618	0.101	0.575	0.903	1.981	0.359	3.458	8.625	2.912
14	Sweden	0.457	0.087	0.632	1	2.014	0.318	4.432	9.968	3.962
15	England/İngiltere	0.349	0.151	0.549	0.638	1.352	0.500	2.368	6.764	4.621
16	Turkey/Türkiye	0.345	1	0.068	0.204	-	2.104	0.121	1.621	0.264
17	Estonia/Estonya	0.551	0.263	0.079	0.879	5.721	-	0.302	2.101	0.528
18	Latvia/Letonya	0.816	0.396	0.054	0.698	0.241	1.262	0.156	1.562	0.142
19	Lithuania/Litvanya	0.729	0.271	0.062	0.650	-	0.872	0.249	1.832	0.452
20	Czech. Reb./Çek Cumh.	0.318	0.146	0.121	0.602	1.171	-	0.642	2.301	0.861
21	Slovenia/Slovenya	0.405	0.133	0.218	0.506	0.146	0.473*	1	2.626	1.126
22	Rumania/Romanya	0.340	0.563	0.037	0.433	0.011	0.436*	0.092	0.961	0.082
23	Bulgaria/Bulgaristan	0.832	0.409	0.031	0.313	-0.01	1.766	0.101	1.062	0.268
						3.421	1.283			
						5.491				

\* Yanlış sınıflandırılmış birimler

\* misclassified units

Her model ile elde edilen sınıflama skorları incelenirse, FLDF tekniğinin doğru sınıflandırma oranı %95, MSD tekniğinin doğru sınıflandırma oranı %78, Suoyeshi' nin DEA-DA modelinin doğru sınıflandırma oranı %100, Retzlaff-Roberts' in (R-R) modelinin doğru sınıflandırma oranı %95 ve önerdiğimiz ÇKVZA-DA ile elde edilen doğru sınıflandırma oranının ise %100 olarak bulunduğu görülür. Bu uygulamada ÇKVZA-DA ile DEA-DA modellerinin performansları aynı olmakla birlikte, sınıflandırma skorları oldukça farklıdır.

## 6. SONUÇ

Bu çalışmada, iki gruplu sınıflandırma problemi, sapmalar toplamının minimizasyonuna dayanan model ile Veri Zarflama Analizindeki göreceli etkinlik kavramının birleştirildiği bir sınıflandırma modeli ile incelenmiştir. Uygulama olarak seçilen veri setinde grup üyelikleri önceden bilindiğinden (Avrupa Birliği' ne üye ve üye olmayan ülkeler), çalışmada önerilen model, grup üyeliklerinin bilindiği durum göz önünde tutularak uygulanmıştır. Grup üyelikleri bilinmeseydi bile Veri Zarflama Analizi ile birimler iki gruba ayrılabilirdi. Uygulama sonuçlarından da görüleceği üzere, önerilen

In considering the classification scores obtained by each model, the hit ratio of FLDF, MSD, DEA-DA, (R-R), Suoyeshi's DEA-DA and our model MCDEA-DA is respectively %95, %78, %100, %95 and %100. Hit ratio is the ratio of correctly classified units to the total number of units to be classified. The classification performance of Suoyeshi's DEA-DA and our model MCDEA-DA are the same but their classification scores are quite different.

## 6. CONCLUSION

In this study, the classification model for two groups is examined by means of a classification model which combines the model minimizing the sum of deviations and the model utilizing the relative efficiency concept (i.e. DEA). In the application, since the group memberships (15 countries of European union and 8 non-union countries) are known initially, the proposed model of this study is applied for this case in which the group memberships are not know, the units could be separated in two groups using DEA. As can be seen from the results of application, our model separates from the groups

model iki grubu tam olarak birbirinden ayırmıştır. Kuşkusuz her uygulamada böyle bir sonuç ile karşılaşılması beklenilemez. Bir simülasyon çalışması ile önerilen modelin tam performansı ve diğer modellerden üstün olup olmadığı görülebilir. Bu ilerde yapacağımız çalışmalar arasındadır.

completely. It provides a good separation from units (between groups). Undoubtedly It can not be expected to obtain like this result in each time and every application. It might have been shown real performance of our model by using simulation study. As a further study we will be interested this dimension.

#### KAYNAKLAR/ REFERENCES

1. Tatlıdil, H., Uygulamalı Çok Değişkenli İstatistiksel Analiz”, **Cem Web Ofset**, Ankara, 1996.
2. Fred, N., Glover, F., “Simple but powerful goal programming formulations for the statistical discriminant problem”, **European Journal of Operational Research**, 7: 44-60, (1981).
3. Fred, N., Glover, F., “A linear programming approach to the discriminant problem”, **Decision Sciences**, 12: 68-74, (1981).
4. Bajgier, S.M., Hill, A.V., "An experimental comparison of statistical and linear programming approaches to the discriminant problem", **Decision Sciences**, 13: 604-618, (1982).
5. Bal, H., "Çok Gruplu Ayırma Probleminde Doğrusal ve Tamsayı Programlama Yaklaşımı ve Yeni Bir Model", **Gazi Üniversitesi Fen Bilimleri Dergisi**, 12(4): 957-962, (1999).
6. Erenguc, S.S, Koehler, G.J., “Minimizing Misclassifications in Linear Discriminant Problem”, **Decision Sciences**, 21: 63-85, (1990).
7. Lam, KF., Choo, E.U., Moy, J.W., “Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem”, **European Journal of Operational Research**, 88, 358-367, (1996).
8. Retzlaff-Roberts, D.L., “Relating discriminant analysis and data envelopment analysis to one another”, **European Journal of Operational Research** , 23: 311-322, (1996).
9. Sueyoshi, T., "DEA-Discriminant Analysis in the view of goal programming", **European Journal of Operational Research** , 115: 564-582, (1999).
10. Sueyoshi, T., "Extended DEA-Discriminant Analysis", **European Journal of Operational Research** , 131: 324-351, (2001).
11. Retzlaff-Roberts, D.L., “A Data Envelopment Analysis approach to Discriminant Analysis”, **Annals of Operational Research** , 73: 299-321, (1997).
12. Glover, F., “Improved linear programming models for discriminant analysis”, **Decision Sciences**, 21: 771-785, (1990).
13. Cooper, W.W., Seiford, L.M., Tone, K., "Data Envelopment Analysis", **Kluwer Academic Publishers**, Boston, USA (2000).
14. Zeleny, M., “Multiple Criteria Decision Making”, **McGraw-Hill**, Newyork, USA, (1982).
15. Sayılarla Türkiye ve Avrupa Birliği, 1994-2000, **D.İ.E Matbaası**, Yayın No: 2516, Aralık 2001, Ankara.
16. World Development Report (2002).