



## **ANÁLISE ENVOLTÓRIA DE DADOS APLICADA PARA AVALIAÇÃO DE JOGADORES DE FUTEBOL**

**Iago Cambre Añon**

FCA-Unicamp

R. Pedro Zaccaria, 1300

Caixa Postal 1068, CEP 13484-350 - Limeira - São Paulo

iago.anon@gmail.com

**Alcides Scaglia**

FCA-Unicamp

Caixa Postal 1068, CEP 13484-350 - Limeira - São Paulo

alcides.scaglia@fca.unicamp.br

**Cristiano Torezzan**

FCA-Unicamp

Caixa Postal 1068, CEP 13484-350 - Limeira - São Paulo

cristiano.torezzan@fca.unicamp.br

### **RESUMO**

Neste trabalho aplicamos a Análise Envoltória de Dados (DEA) para analisar as eficiências de jogadores de futebol. O trabalho foi realizado com base em dados reais, considerando os 23 melhores jogadores do mundo segundo o ranking FIFA de 2015. Além do estudo comparativo das eficiências, realizado através dos modelos CCR e BBC, foi proposto, também, uma ordenação de jogadores e clubes através da eficiência cruzada, que considera todos os pesos encontrados no DEA e não apenas os inerentes à cada DMU. Nos cenários analisados, verifica-se que os jogadores mais eficientes não são os mais caros, o que sugere que os altos valores investidos não estão retornando, na mesma escala, em termos de desempenho dos jogadores.

**PALAVRAS CHAVE. DEA, Futebol, Eficiência.**

**DEA – Análise Envoltória de Dados**

**ADM – Apoio à Decisão Multicritério**

### **ABSTRACT**

In this work we apply the Data Envelopment Analysis (DEA), CCR and BCC models, to analyze the relative efficiencies of soccer players. The work was carried out based on real data, considering the 23 best players in the world according to the FIFA ranking of 2015. In addition to the analysis of the efficiencies, a rankings through cross-efficiency was performed in order to analyze the players in the different weights found for each unit. In the analyzed scenarios, it turns out that the most efficient players are not the most expensive, which suggests that the high values invested are not returning on the same scale in player performance.

**KEYWORDS. DEA, Soccer, Efficiency.**

**DEA – Data Envelopment Analysis**

**ADM - Support for the Multicriteria Decision**



## 1. Introdução

O futebol é um esporte que movimenta cifras astronômicas em todo o mundo. Muitos dos negócios que envolvem o futebol são destinados à formação de equipes que possam obter grandes resultados desportivos.

Neste cenário, tem sido crescente a demanda por medidas de eficiência sobre os investimentos realizados, assim como para a avaliação comparativa das equipes adversárias, afim de se realizar um *benchmark* sobre gestão e negócios relacionados ao esporte.

Analisar a qualidade de um processo produtivo é uma atividade complexa, que demanda o estudo dos mais variados fatores que influenciam no processo. Esta dificuldade é agravada quando se tem um cenário de múltiplos recursos e/ou múltiplos resultados.

Desta forma, modelos matemáticos que analisam eficiência têm recebido atenção como ferramentas de apoio à tomada de decisão de gestores e comissão técnica, principalmente no período de investimentos financeiros. Dentre os modelos que analisam a eficiência, destaca-se a Análise Envoltória de Dados (DEA), que permite não apenas modelar problemas com múltiplos fatores, mas também, que cada unidade avaliada tenha liberdade para decidir os pesos (ou constantes de escala) de cada um dos fatores no momento de sua avaliação.

Originalmente restrito Engenharia de Produção e Administração, a aplicação da DEA vem ganhando força em diversas outras áreas e, no futebol, não é diferente. Diversos deles procuram analisar a eficiência de diferentes competições pelo mundo analisando jogadores e/ou clubes. Como exemplos podemos destacar a aplicação da DEA para a análise de eficiência de clubes na Bundesliga (Alemanha), com o estudo de Haas *et al.* [2004], na La Liga (Espanha) com os trabalhos de González-Gómez e Picazo-Tadeo [2010], García-Sánchez [2007] e Barros e Garcia-Del-Barrio [2011], na Premier League (Inglaterra) com Guzmán e Morrow [2007] e Haas [2003] e no Campeonato Brasileiro com Pestana Barros *et al.* [2010]. Já a aplicação para a análise de eficiência dos jogadores obtém destaque com o estudo na Bundesliga (Alemanha) realizado por Tiedemann *et al.* [2011].

Neste estudo aplicamos os modelos clássicos CCR (abreviatura dos nomes dos criadores do modelo, sendo eles: Charnes, Cooper e Rhodes) e BCC (abreviatura dos nomes dos criadores do modelo, sendo eles: Banker, Charnes e Cooper) da DEA para analisar a eficiência dos 23 melhores jogadores do *ranking* FIFA (FIFA Ballon d'Or, 2015). Além disso, utilizamos o método de eficiência cruzada para a formação de um *ranking*, aumentando assim a capacidade discriminatória da Análise Envoltória de Dados.

## 2. Análise Envoltória de Dados (DEA)

A análise envoltória de dados (DEA) é classificada como uma técnica não paramétrica, que possibilita a determinação de fronteiras de eficiência entre diversas unidades tomadoras de decisão (DMUs), em um cenário que contempla diversos recursos (inputs) e diversos resultados (outputs). Este modelo tem sido utilizado como base para avaliações de desempenho e benchmarking em diversos setores como escolas, hospitais, agências bancárias, fábricas de produção, etc. [Charnes *et al.*, 2013].

A Análise Envoltória de Dados surgiu com o primeiro modelo proposto por Charnes *et al.* [1978], denominado CCR e caracterizado por um retorno constante de escala, isto é, se há uma variação nas variáveis de entrada (inputs) há também uma variação proporcional nas variáveis de saída (outputs). Segundo Biondi [2001] este modelo tem a vantagem de reduzir os múltiplos produtos e múltiplos insumos de cada DMU para um cenário 'virtual' de um único produto e um único insumo. A próxima seção apresenta, de forma sintética, o modelo CCR.

O primeiro modelo proposto por Charnes *et al.* [1978], foi o CCR, caracterizado por um retorno constante de escala, isto é, se há uma variação nas variáveis de entrada (inputs) há também uma variação proporcional nas variáveis de saída (outputs). O principal ponto deste modelo, segundo Biondi [2001] é a redução de múltiplos produtos e múltiplos insumos para cada DMU para um cenário 'virtual' de um único produto e um único insumo.



## 2.1. Modelo CCR-DEA

Em alguns casos, a comparação entre DMUs com o retorno constante de escala não é um método justo de análise, principalmente quando há entre as unidades cenários diferentes de análise. Se as unidades tomadoras de decisão apresentam cenários de investimentos e retornos próximos, a utilização do CCR-DEA passa a ser recomendável, mas se o cenário for discrepante, se recomenda a utilização do modelo VRS ou BCC-DEA (retorno variável de escala).

Para entendimento dos modelos CCR será apresentado nesta seção as equações matemáticas presentes no mesmo.

Considere  $n$  DMUs, cuja eficiência deve ser analisada com relação a  $m$  variáveis de entradas, denotadas por  $x_{ik}$  e  $s$  variáveis de saída, denotadas por  $y_{rk}$ . A medida de eficiência para a  $k$ -ésima DMU é definida por:

$$h_k = \max \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}},$$

onde os pesos  $v_i \geq 0$  e  $u_i \geq 0$  devem ser escolhidos para cada DMU respeitando a restrição

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \text{ for } j = 1, \dots, n.$$

Onde

$u_r$  = pesos de outputs  $r, r = 1, \dots, s$ ;

$v_i$  = pesos de inputs  $i, i = 1, \dots, m$ ;

$y_{rk}$  = outputs  $r, k = 1, \dots, n$ ;

$x_{ik}$  = inputs  $i, i = 1, \dots, n$ .

A taxa de eficiência  $h_k$  é um número que varia entre 0 e 1 e  $k$ -ésima DMU é considerada eficiente se for possível encontrar um conjunto de pesos  $\{v_1, \dots, v_m\}$  e  $\{u_1, \dots, u_s\}$  tal que  $h_k = 1$  e  $h_j \leq 1$  para todos  $j \neq k$ . Assim, cada unidade pode formular um problema de otimização para encontrar os pesos que mais lhe favorecem na análise comparada.

O CCR-DEA pode ser formulado como o seguinte problema programação linear (PL):

$$\max h_k = \sum_{r=1}^s u_r y_{rk}$$

Sujeito à:

$$\begin{aligned} \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0 \\ \sum_{i=1}^m v_i x_{ik} &= 1 \\ u_r &\geq \varepsilon > 0, v_i &\geq \varepsilon > 0 \end{aligned}$$

O PL acima deve ser resolvido para cada unidade e o resultado final é a determinação de hiperplanos que definem uma fronteira de eficiência. As DMUs que atingem  $h_k=1$  estão sob a fronteira, e as demais, que ficam “envelopadas”, são declaradas comparativamente ineficientes.



Em alguns casos, a comparação entre DMUs com o retorno constante de escala não é um método justo de análise, principalmente quando há unidades de tamanhos desproporcionais a serem comparadas. Se as unidades tomadoras de decisão apresentam cenários de investimentos e retornos próximos, a utilização do CCR-DEA passa a ser recomendável, mas se o cenário for discrepante, se recomenda a utilização do modelo VRS ou BCC-DEA (retorno variável de escala), que foi proposto por Banker *et al.* [1984] e será descrito na próxima seção.

## 2.2. O modelo BCC-DEA

Como alternativa para o modelo CCR, Banker *et al.* [1984] propõem um modelo de retorno variável de escala, conhecido como BCC, em homenagem à seus autores. Para viabilização desta proposta é acrescentado uma variável,  $u_*$  que tem como função ser um fator de escala que auxilia a ajustar a escala de acordo com cada DMU.

O modelo BCC-DEA foi proposto por Banker *et al.* [1984], ele tem como características ser um modelo de retorno variável de escala, procurando evitar problemas existentes em situações de competição imperfeita.

Com isso, o problema de programação linear (PL) é formalizado da seguinte maneira:

$$\max h_k = \sum_{r=1}^s u_{r*} y_{rk} - u_*$$

Sujeito à:

$$\begin{aligned} \sum_{r=1}^s u_{r*} y_{rj} - \sum_{i=1}^m v_{i*} x_{ij} - u_* &\leq 0 \\ \sum_{i=1}^m v_{i*} x_{ik} &= 1 \\ u_r &\geq \varepsilon > 0, v_i \geq \varepsilon > 0 \end{aligned}$$

Onde

$u_r$  = pesos de outputs  $r, r = 1, \dots, s$ ;

$v_i$  = pesos de inputs  $i, i = 1, \dots, m$ ;

$y_{rk}$  = outputs  $r, k = 1, \dots, n$ ;

$x_{ik}$  = inputs  $i, i = 1, \dots, n$ ;

$u_*$  = fator de escala.

Assim como no modelo de retorno constante de escala (CCR), no modelo BCC se faz necessário a resolução de um PL para cada DMU estudada. As DMUs que atingem  $h_k = 1$  estão sob a fronteira, e as demais, que ficam “envelopadas”, são declaradas relativamente ineficientes.

## 2.3. Ordenamento por eficiência cruzada

Devido à sua característica de flexibilidade irrestrita dos pesos no DEA, os modelos citados acima não permitem a obtenção de um ordenamento total (*ranking*) das unidades. Com isso, segundo Talluri [2000], unidades que são eficientes em suas análises podem ser não eficientes em um contexto global. Diversos estudos propuseram métodos de ordenamento utilizando como base os cálculos realizados na DEA. Em seu estudo de revisão sobre classificação a partir da Análise Envolvória de Dados, Adler *et al.* [2002] apresentou 6 grupos para o ordenamento no contexto da DEA, sendo eles:

- Métodos de classificação de eficiência cruzada [Sexton *et al.*, 1986];
- Técnicas de classificação de super-eficiência [Andersen e Petersen, 1993];



- Método de classificação de referência [Torgersen *et al.*, 1996];
- Ranking com estatísticas multivariadas no contexto da DEA [Friedman e Sinuany-Stern, 1997];
- O ranking das unidades de tomada de decisão ineficientes [Bardhan *et al.*, 1996];
- DEA e métodos de decisão multicritérios [Troutt, 1995].

No presente estudo utiliza-se o método de classificação por eficiência cruzada [Sexton *et al.*, 1986]. Sobre as vantagens na utilização da eficiência cruzada Adler *et al.* [2002] aponta que este modelo apresenta como vantagem a capacidade de ranquear unidades eficientes, apresentando pesos em comum para as mesmas e assim, evitando problemas ocasionais de inviabilidade dos métodos.

A matriz de avaliação cruzada proposta por Sexton *et al.* [1986] apresenta como base a utilização dos pesos de uma unidade em sua própria análise e quando analisada com o peso das demais unidades.

Estes resultados das pontuações são expressados na matriz de eficiência cruzada  $H = h_{kj}$ , demonstrada na seguinte equação:

$$h_{kj} = \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ij}}, k = 1, \dots, n, j = 1, \dots, n.$$

Onde

$h_{kj}$  = eficiência atribuída a unidade  $j$  com os pesos da unidade  $k$ ;

$u_r$  = pesos de outputs  $r, r = 1, \dots, s$ ;

$v_i$  = pesos de inputs  $i, i = 1, \dots, m$ ;

$y_{rk}$  = outputs  $r, k = 1, \dots, n$ ;

$x_{ik}$  = inputs  $i, i = 1, \dots, n$ .

Tendo esta formulação, nota-se que a diagonal da matriz  $H$  representa a eficiência obtida para a própria DMU, ou seja,  $h_{11}$  é o pontuação na unidade 1, com o peso ótimo para si mesmo. Se observado a matriz em seu todo, observa-se que,  $0 \leq h_{kj} \leq 1$ . Esta restrição citada implica na escolha de apenas um modelo do DEA, o CCR. Ao adicionar o fator de escala, o modelo BCC permite que com os pesos das demais DMUs, a unidade apresente índices inferiores a 0.

Neste estudo foi escolhido a aplicação da média pois, desta forma, o resultado final da classificação representa o comportamento médio das unidades em todas as condições obtidas através da determinação dos pesos, incluindo os que melhor representam a unidade analisada. Esta média, é representada por:

$$h_k = \frac{\sum_{j=1}^n h_{kj}}{n}$$

### 3. Proposta do Estudo



Este estudo tem dois objetivos principais, o primeiro é a aplicação dos modelos CCR e BCC tendo em vista analisar a eficiência dos 23 melhores jogadores do mundo em 2015, assim como a influência da escolha dos modelos no resultado. O segundo objetivo o ornamento das DMUs de cada caso através de aplicação do modelo de eficiência cruzada proposto por Sexton *et al.* [1986].

Como dados de *input* foram utilizados o salário que cada atleta recebe de seu clube e o gasto do clube para realização de sua transferência dividido pelo número de temporadas que o jogador se encontra neste clube. Como *outputs* foram considerados a soma do tempo em campo em cada jogo da temporada, porcentagem de pontos obtidos e saldo de gols das partidas, que é composto pelo número de gols marcado pela equipe menos os gols sofridos nos momentos em que o jogador esteve em campo, todos referentes aos jogos realizados em 2015, além da diferença entre o valor do jogador no mercado no momento de sua chegada ao clube e o final do ano de 2015.

#### 4. Resultados

Realizado a aplicação de ambos os modelos, CCR e BCC, obtêm-se os resultados demonstrados na Tabela 1. Sendo assim, se observa que apenas 3 unidades são eficientes no modelo CCR e 7 são eficientes no modelo BCC. Esta diferença entre a aplicação do modelo pode ser notada, pois 21 unidades apresentaram aumento de eficiência quando foram analisadas com BCC em comparação a análise com CCR, e 2 unidades mantiveram seus índices.

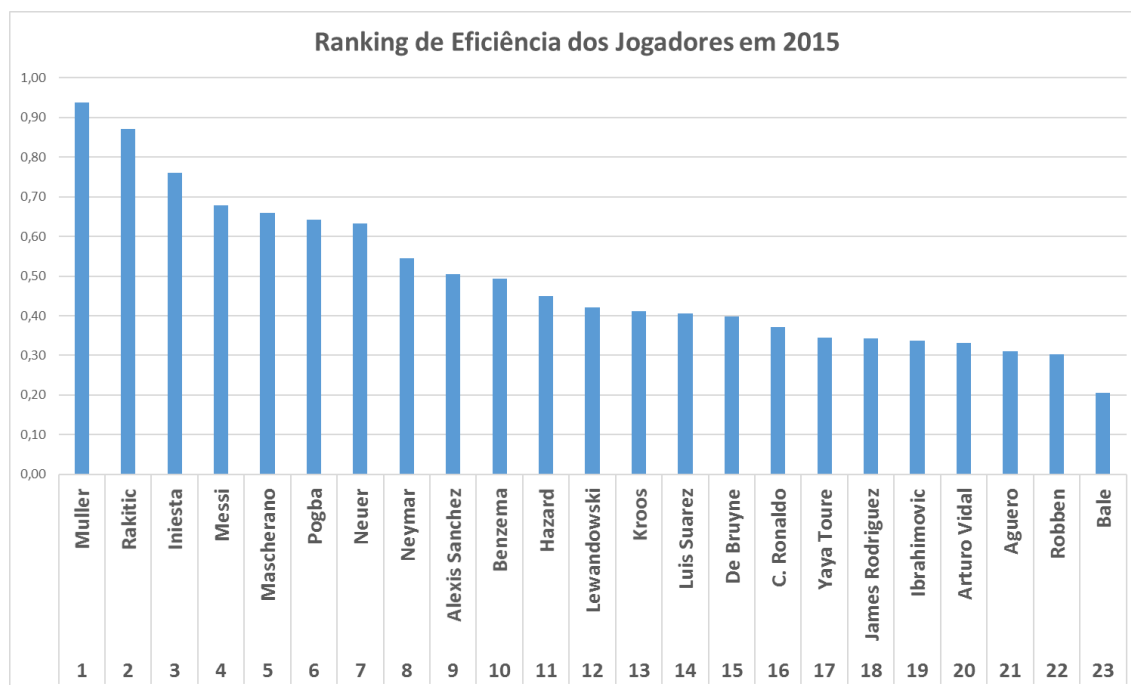
**Tabela 1-** Eficiência dos 23 melhores jogadores do mundo em 2015 nos modelos CCR e BCC.

#	JOGADOR	CCR	BCC
1	Messi	0,86	1,00
2	C. Ronaldo	0,51	0,98
3	Neymar	0,87	1,00
4	Aguero	0,40	0,55
5	Benzema	0,71	0,83
6	Hazard	0,63	0,69
7	Iniesta	1,00	1,00
8	Lewandowski	0,57	0,66
9	Muller	1,00	1,00
10	Pogba	0,86	0,86
11	Robben	0,73	0,75
12	Alexis Sanchez	0,62	0,64
13	Yaya Toure	0,53	0,62
14	Bale	0,25	0,30
15	De Bruyne	0,69	0,88
16	Ibrahimovic	0,52	0,53
17	Kroos	0,52	0,66
18	Mascherano	0,99	1,00
19	Neuer	0,91	1,00
20	Rakitic	1,00	1,00
21	James Rodriguez	0,68	0,72
22	Luis Suarez	0,54	0,97



23                      Arturo Vidal                      0,50                      0,50

A partir dos resultados obtidos no modelo CCR, aplicou-se o modelo de ordenamento proposto por (Sexton *et al.*, 1986). Este modelo de eficiências cruzadas aponta que a melhor unidade analisada é o jogador Muller, com média de 0,94. A unidade com pior desempenho médio neste estudo foi Bale, com 0,20.



**Figura 1** – Ranking de Eficiência dos 23 melhores jogadores do mundo em 2015.

Quando observado que a base para a seleção dos jogadores foi a eleição de melhor jogador do mundo de 2015, se faz necessário um destaque para os 3 finalistas. O terceiro colocado, Neymar apresenta índice de 0,54, já Cristiano Ronaldo, segundo colocado, e Messi, o primeiro colocado, apresentaram respectivamente 0,37 e 0,68.

## 5. Conclusão

Tendo em vista os resultados apresentados e as características de ambos os modelos da DEA, podemos concluir que o número de DMUs eficientes é baixo, 30% do total de unidades alcançaram eficiência no modelo BCC, sendo o mesmo o mais benevolente. Se observado os dados do modelo CCR este número cai para 13% das unidades.

Outra conclusão que se pode chegar ao se analisar as unidades presentes neste estudo é que, por ser composto pela seleção de 23 melhores jogadores do mundo, os jogadores tende a apresentar índices de desempenho próximos, sendo que o mesmo padrão não se é notado quando observa se os investimentos realizados. Quando observado este cenário, entende-se que os clubes tendem a superestimar seus investimentos, ou seja, investem altas quantias de dinheiro e em contrapartida, recebem retornos menores que os necessários para alcançar altas índices de eficiência. Isto reflete, analisando os dados, que os altos investimentos exigem retornos que por vezes podem ser inalcançáveis.

## Referências



ADLER, N.; FRIEDMAN, L.; SINUANY-STERN, Z. Review of ranking methods in the data envelopment analysis context. **European journal of operational research**, v. 140, n. 2, p. 249-265, 2002. ISSN 0377-2217.

ANDERSEN, P.; PETERSEN, N. C. A procedure for ranking efficient units in data envelopment analysis. **Management science**, v. 39, n. 10, p. 1261-1264, 1993. ISSN 0025-1909.

BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. **Management science**, v. 30, n. 9, p. 1078-1092, 1984. ISSN 0025-1909.

BARDHAN, I. et al. Models for efficiency dominance in data envelopment analysis. **Part**, v. 1, p. 322-332, 1996.

BARROS, C. P.; GARCIA-DEL-BARRIO, P. Productivity drivers and market dynamics in the Spanish first division football league. **Journal of Productivity Analysis**, v. 35, n. 1, p. 5-13, 2011. ISSN 0895-562X.

BIONDI, L. Neuro-DEA: nova metodologia para determinação da eficiência relativa de Unidades Tomadoras de Decisão. **Neuro-DEA: nova metodologia para determinação da eficiência relativa de Unidades Tomadoras de Decisão**, 2001.

CHARNES, A. et al. **Data envelopment analysis: Theory, methodology, and applications**. Springer Science & Business Media, 2013. ISBN 9401106371.

CHARNES, A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. **European journal of operational research**, v. 2, n. 6, p. 429-444, 1978. ISSN 0377-2217.

FARRELL, M. J. The measurement of productive efficiency. **Journal of the Royal Statistical Society. Series A (General)**, v. 120, n. 3, p. 253-290, 1957. ISSN 0035-9238.

FRIEDMAN, L.; SINUANY-STERN, Z. Scaling units via the canonical correlation analysis in the DEA context. **European Journal of Operational Research**, v. 100, n. 3, p. 629-637, 1997. ISSN 0377-2217.

GARCÍA-SÁNCHEZ, I. Efficiency and effectiveness of Spanish football teams: a three-stage-DEA approach. **Central European Journal of Operations Research**, v. 15, n. 1, p. 21-45, 2007. ISSN 1435-246X.

GONZÁLEZ-GÓMEZ, F.; PICAZO-TADEO, A. J. Can we be satisfied with our football team? Evidence from Spanish professional football. **Journal of Sports Economics**, v. 11, n. 4, p. 418-442, 2010. ISSN 1527-0025.

GUZMÁN, I.; MORROW, S. Measuring efficiency and productivity in professional football teams: evidence from the English Premier League. **Central European Journal of Operations Research**, v. 15, n. 4, p. 309-328, 2007. ISSN 1435-246X.

HAAS, D.; KOCHER, M. G.; SUTTER, M. Measuring efficiency of German football teams by data envelopment analysis. **Central European Journal of Operations Research**, v. 12, n. 3, p. 251, 2004. ISSN 1435-246X.

HAAS, D. J. Productive efficiency of English football teams—a data envelopment analysis approach. **Managerial and Decision Economics**, v. 24, n. 5, p. 403-410, 2003. ISSN 1099-1468.

PESTANA BARROS, C.; ASSAF, A.; SÁ-EARP, F. Brazilian football league technical efficiency: a Simar and Wilson approach. **Journal of Sports Economics**, v. 11, n. 6, p. 641-651, 2010. ISSN 1527-0025.





SEXTON, T. R.; SILKMAN, R. H.; HOGAN, A. J. Data envelopment analysis: Critique and extensions. **New Directions for Evaluation**, v. 1986, n. 32, p. 73-105, 1986. ISSN 1551-2371.

TALLURI, S. Data envelopment analysis: models and extensions. **Decision Line**, v. 31, n. 3, p. 8-11, 2000.

TIEDEMANN, T.; FRANCKSEN, T.; LATACZ-LOHMANN, U. Assessing the performance of German Bundesliga football players: a non-parametric metafrontier approach. **Central European Journal of Operations Research**, v. 19, n. 4, p. 571-587, 2011. ISSN 1435-246X.

TORGERSEN, A. M.; FØRSUND, F. R.; KITTELSEN, S. A. Slack-adjusted efficiency measures and ranking of efficient units. **Journal of Productivity Analysis**, v. 7, n. 4, p. 379-398, 1996. ISSN 0895-562X.

TROUTT, M. D. A maximum decisional efficiency estimation principle. **Management Science**, v. 41, n. 1, p. 76-82, 1995. ISSN 0025-1909.