

REGIÃO DE VIABILIDADE E COMPARAÇÃO DE GRUPOS PARA A PERFORMANCE DEA NÃO ARQUIMEDIANA VIA REGRESSÃO FRACIONADA

Geraldo da Silva e Souza

Eliane Gonçalves Gomes

Empresa Brasileira de Pesquisa Agropecuária – Embrapa
Parque Estação Bilógica, Av. W3 Norte final, 70770-901, Brasília, DF, Brasil
geraldouso@embrapa.br
eliane.gomes@embrapa.br

RESUMO

Considera-se a utilização da constante não arquimediana ε em modelos DEA-CCR. A aplicação de interesse é definida pela medida de performance dos centros de pesquisa da Empresa Brasileira de Pesquisa Agropecuária. Caracteriza-se uma região de viabilidade e sugere-se um valor para a constante ε . Tipos de DMUs são, então, comparados por meio do uso de regressão fracionada e do método da máxima quase-verossimilhança. Conclui-se pela dominância dos centros de pesquisa do tipo produto. As medidas de performance clássica DEA-CCR e com restrições de folga não nula pela constante não arquimediana têm correlação de Spearman superior a 90%.

Palavras-chave: DEA não arquimediano; Pesos não nulos; Regressão fracionada.

Área principal: DEA - Análise Envoltória de Dados

ABSTRACT

In this paper we consider the use of the non-Archimedean constant ε in DEA-CCR models. The application of interest is defined as the performance measure of the Brazilian Agricultural Research Corporation research centers. We characterized an assurance region and suggest a value for the constant ε . Types of DMUs are then compared by using fractional regression models and the quasi-maximum likelihood approach. We conclude that the research centers whose mission is to research specific products performed better. The classic DEA-CCR performance measures and the solution provided by non-Archimedean model have Spearman correlation higher than 90%.

Key-words: Non-Archimedean DEA models; Non-zero weights; Fractional regression.

Main area: DEA - Data Envelopment Analysis

1. Introdução

Desde 1996, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) monitora a produção de seus centros de pesquisa com a utilização de um sistema de avaliação com base em um produto e três insumos. Os insumos do processo de produção são capital, custeio e mão de obra. O produto (*output*), unidimensional, é uma média ponderada de 28 atributos de produção, classificados em quatro categorias de produção: (a) produção técnico-científica; (b) produção de publicações técnicas; (c) transferência de tecnologias e promoção da imagem; (d) desenvolvimento de tecnologias, produtos e processos. Uma medida de eficiência é, então, calculada via modelos de análise de envoltória de dados (DEA). Veja Souza et al. (1999) para mais detalhes, notadamente no contexto DEA. O sistema de pesos utilizado para tornar os centros de pesquisa (DMUs) comparáveis é complexo. Toma por base a Lei dos Julgamentos Categóricos de Thurstone (Torgerson, 1958; Souza, 2002) e é dependente de um sistema de normalização que torna as grandezas de produção adimensionais. A subjetividade dos pesos utilizados e a redução dimensional drástica têm sido questionadas no âmbito da organização. Recentemente sugeriu-se o uso de um vetor de produtos de dimensão três, reagrupando os atributos em três categorias – produção técnico-científica, produção de publicações técnicas e outras atividades. O sistema de pesos utilizado nesses agrupamentos é objetivo e baseia-se na Análise Fatorial. Toda a classificação de performance baseia-se nos *ranks* das variáveis de produção, medidas de forma *per capita* e motivam o uso da solução DEA-CCR com três insumos e três produtos. Tipicamente, as soluções de desempenho unitário obtidas apresentam folgas e, portanto, não satisfazem ao conceito de eficiência de Pareto-Koopmans (Cooper et al., 2007).

A existência de produtos ou insumos com pesos nulos na solução do problema dos multiplicadores do modelo DEA-CCR clássico sugere o uso da constante não arquimediana no processo de avaliação. Em muitas aplicações DEA têm-se usado valores pequenos (10^{-6} e menores) para esta constante. Veja-se, por exemplo, Charnes et al. (1996). A escolha arbitrária desta constante não é aconselhável, pois pode conduzir a soluções errôneas (problemas numéricos), como enfatizado em Ali e Seiford (1993a, 1993b) e Cooper et al. (2007). Mehrabian et al. (2000) apresentam um método para o cálculo de uma região de viabilidade para a constante não arquimediana ϵ . Com base na região de viabilidade e na discussão de Cooper et al. (2007) sugere-se a utilização do extremo superior da região como valor adequado para a constante não arquimediana.

O estudo da significância estatística de parâmetros em modelos em que uma medida do tipo DEA é tomada como resposta apresenta dificuldades técnicas em dois contextos. Primeiramente, o método de cálculo envolvido na determinação da medida de performance induz correlação entre as unidades (DMUs) avaliadas. Mais problemático em aplicações mais complexas é a associação de variáveis contextuais com as medidas de erro dos modelos de regressão. Simar e Wilson (2007) discutem esses problemas em detalhes. Ramalho et al. (2010) sugerem o uso da abordagem de Papke e Wooldridge (1996) para o tratamento simultâneo dos problemas de respostas truncadas (valores em (0,1]) e correlação entre DMUs. Utiliza-se esta abordagem neste artigo para a comparação de tipos de DMUs.

Nossa exposição prossegue como segue. Na Seção 2 apresenta-se o modelo de avaliação atualmente em discussão na Embrapa. Na Seção 3 apresenta-se a técnica utilizada nos modelos DEA para a caracterização de uma região de viabilidade para a constante não arquimediana ϵ . A Seção 4 descreve os modelos de regressão fracionada. Na Seção 5 apresentam-se os resultados estatísticos e de programação obtidos. Finalmente, na Seção 6 apresentam-se um resumo e as conclusões do trabalho, seguida da lista de referências.

2. O Modelo de Desempenho da Embrapa

O modelo de performance atualmente em discussão na Embrapa baseia-se em uma medida DEA-CCR que se utiliza de três insumos e três produtos e é calculada aqui para 37 centros de pesquisa componentes da instituição. Atualmente, a empresa conta com 42 unidades de pesquisa classificadas em três tipos – centros de pesquisa de produto, ecorregionais e

temáticos. Os centros de pesquisa de produto têm como missão viabilizar soluções de pesquisa, desenvolvimento e inovação em produtos agrícolas específicos (ex.: soja, milho e sorgo, algodão, gado de leite etc.). Os centros ecorregionais e temáticos dedicam-se, respectivamente, à pesquisa para sustentabilidade da agricultura nas diferentes ecorregiões brasileiras (ex.: pantanal, tabuleiros costeiros, cerrado etc.) e em temas básicos de interesse (ex.: monitoramento por satélite, solos, informática e instrumentação agropecuárias etc.). Cinco das 42 unidades foram recentemente criadas e não constam de nossa análise. As componentes de insumos do modelo de performance são custos com pessoal, custeio e capital. As componentes de produção são as categorias de produção técnico-científica – PTC, produção de publicações técnicas – PPT, e outras atividades de produção – POUT. A categoria PTC inclui artigos em periódico indexado, capítulo de livro, artigo e resumo em anais de congressos e orientação de teses. A componente PPT inclui as variáveis circular técnica, comunicado técnico/recomendações técnicas, boletim de pesquisa e desenvolvimento e documentos. A categoria POUT é definida pelas variáveis dias de campo, organização de eventos, unidade demonstrativa e de observação, estágios de pós-graduação, metodologia científica e monitoramento. Todas as variáveis são normalizadas pelo quantitativo de pessoal e, posteriormente, transformadas em *ranks*. A normalização reduz variabilidade e facilita comparações, reduzindo diferenças de escala. A transformação em *ranks* elimina a influência de observações atípicas, define respostas adimensionais e empresta propriedades não paramétricas aos métodos com base em normalidade utilizados na determinação dos pesos dos atributos dentro das dimensões. Esses métodos se utilizam de Análise Fatorial e podem ser vistos em detalhes em Souza e Gomes (2013); não serão discutidos aqui. A Tabela 1 apresenta os valores dos seis atributos de produção para o ano de 2009.

3. DEA e Modelos Não Arquimedianos

3.1. Conceitos gerais

Dado um conjunto de unidades produtivas (*Decision Making Units* – DMUs), para as quais são conhecidos os níveis de recursos empregados (*inputs*) e de resultados obtidos (*outputs*), os modelos DEA têm como finalidade calcular uma medida de eficiência para cada DMU. Em contraste às funções de produção clássicas, DEA destaca cada observação individual com o objetivo de calcular uma fronteira de eficiência, determinada pelas unidades Pareto-eficientes. Essas unidades servem de *benchmark* para aquelas ineficientes.

Genericamente, a medida de eficiência é dada pela razão entre a soma ponderada de produtos e a soma ponderada de insumos necessários para gerá-los. Os pesos usados nessas ponderações são obtidos, no contexto de DEA, por problemas de programação matemática, que atribuem a cada DMU os pesos que maximizam a sua eficiência.

Os dois modelos DEA mais referenciados na literatura são o CCR e o BCC (Cooper et al., 2007), os quais diferem quanto às hipóteses de convexidade da fronteira de eficiência (convexa no caso BCC) e retornos de escala (constantes – CCR – ou variáveis – BCC). Na busca da fronteira de eficiência são possíveis duas orientações radiais para esses modelos: a *inputs* (minimização dos *inputs*, sem alteração do nível de *outputs*) e a *outputs* (aumento dos *outputs*, sem mudança nos *inputs*). Sob a hipótese de retornos constantes (modelo DEA-CRR), as duas orientações são equivalentes no sentido de que induzem à mesma medida de eficiência.

Em (1) e em (2) apresentam-se, respectivamente, o modelo DEA CCR dos Multiplicadores e do Envelope (duais), com orientação a *inputs*. Considera-se que cada DMU k , $k=1..n$, é uma unidade de produção que utiliza r *inputs* x_i^k , $i=1..r$, para produzir s *outputs* y_j^k , $j=1..s$; x_i^o e y_j^o são os *inputs* e *outputs* da DMU o ; v_i e u_j são os pesos calculados pelo modelo para *inputs* e *outputs*, respectivamente. Em (5), θ_o é a eficiência da DMU o em análise e λ_k representa a contribuição da DMU k na formação do alvo da DMU o . As formulações primal e dual do modelo DEA BCC são obtidas adicionando-se um fator de escala à função objetivo e à desigualdade de (1), e a restrição $\sum_{k=1}^n \lambda_k = 1$ ao PPL (2). Uma interpretação do modelo do envelope é a que segue: se X representa a matriz de utilização de insumos e Y a matriz de

produtos, então o par $(X\lambda, Y\lambda)$ representa o vetor de produção de um pseudoprodutor, sendo λ o vetor com componentes λ_k . Deste modo, uma unidade é considerada eficiente se não existe pseudoprodutor que seja capaz de produzir pelo menos o que a unidade produziu, gastando menos. O conjunto dos pseudoprodutores inclui todas as unidades em avaliação.

Tabela 1: Variáveis de produção do modelo DEA e tipos de centros de pesquisa.

Unidade	Tipo	Inputs			Outputs		
		Pessoal	Capital	Custeio	PTC	PPT	POUT
DMU1	Temático	37	35	36	31,53	22,96	13,69
DMU2	Produto	22	6	7	22,22	16,91	8,47
DMU3	Temático	16	29	25	34,27	30,88	21,57
DMU4	Produto	4	9	3	11,88	17,99	12,62
DMU5	Temático	32	30	33	33,88	18,23	13,20
DMU6	Produto	11	15	30	10,89	18,96	18,10
DMU7	Temático	34	37	35	36,67	32,14	23,45
DMU8	Produto	30	11	16	27,26	21,55	27,69
DMU9	Produto	9	34	19	23,80	15,33	11,88
DMU10	Produto	14	23	1	20,14	8,46	26,86
DMU11	Produto	5	16	8	14,89	17,95	11,81
DMU12	Temático	33	36	34	10,75	14,78	14,89
DMU13	Temático	29	28	14	23,09	13,13	27,68
DMU14	Produto	23	10	24	30,59	23,15	25,91
DMU15	Produto	13	19	23	13,48	21,20	14,88
DMU16	Temático	35	24	15	18,88	23,72	27,31
DMU17	Produto	12	22	13	11,13	7,76	22,92
DMU18	Produto	15	25	32	27,24	8,22	25,25
DMU19	Produto	10	12	9	18,70	26,47	23,69
DMU20	Temático	36	32	27	24,40	22,35	21,01
DMU21	Produto	7	31	26	20,91	10,85	19,84
DMU22	Ecorregional	17	3	10	3,40	13,39	11,16
DMU23	Ecorregional	21	13	20	23,21	23,20	14,44
DMU24	Ecorregional	1	17	6	14,44	27,11	25,95
DMU25	Ecorregional	28	20	37	19,11	5,72	11,34
DMU26	Ecorregional	31	8	28	7,55	7,56	9,46
DMU27	Ecorregional	8	1	2	3,64	25,03	17,24
DMU28	Ecorregional	26	27	22	21,95	33,71	21,78
DMU29	Ecorregional	3	2	5	6,01	15,60	13,62
DMU30	Ecorregional	19	14	29	9,01	18,51	16,71
DMU31	Ecorregional	27	18	11	16,61	24,82	18,69
DMU32	Ecorregional	24	5	21	24,65	28,63	24,93
DMU33	Ecorregional	2	7	4	12,15	7,89	18,22
DMU34	Ecorregional	20	4	12	5,73	6,98	19,88
DMU35	Produto	18	26	18	27,28	25,73	23,73
DMU36	Produto	6	21	17	12,00	25,09	33,39
DMU37	Temático	25	33	31	29,69	21,03	9,77

$$\begin{aligned} & \text{Max } \sum_{j=1}^s u_j y_j^o \\ & \text{sujeito a} \\ & \sum_{i=1}^r v_i x_i^o = 1 \\ & - \sum_{i=1}^r v_i x_i^k + \sum_{j=1}^s u_j y_j^k \leq 0, \forall k \\ & u_j, v_i \geq 0, \forall j, i \end{aligned} \quad (1)$$

$$\begin{aligned} & \text{Min } \theta_o \\ & \text{sujeito a} \\ & \theta_o x_i^o - \sum_{k=1}^n x_i^k \lambda_k \geq 0, \forall i \\ & - y_j^o + \sum_{k=1}^n y_j^k \lambda_k \geq 0, \forall j \\ & \lambda_k \geq 0, \forall k \end{aligned} \quad (2)$$

3.2. Modelos DEA não arquimedianos e região de viabilidade para ε

Os modelos clássicos de DEA podem atribuir eficiência a unidades não verdadeiramente eficientes (eficiência fraca devido à presença de folgas não nulas; ver Cooper et al., 2007). Para contornar essa característica estão propostos na literatura os chamados modelos não arquimedianos, os quais tentam eliminar a possibilidade de o modelo do envelope ser ilimitado e, conseqüentemente, a formulação dos multiplicadores ser inviável. Nestes modelos são incorporadas folgas à função objetivo da formulação do envelope ou restrições aos pesos no modelo dos multiplicadores. Neste último caso, os multiplicadores devem ser maiores que ε , um número muito pequeno ou infinitesimal (constante não arquimediana). Note-se que ao incorporar restrições aos pesos, elimina-se o problema de pesos zero em DEA (para discussões sobre este tema ver, por exemplo, Angulo Meza e Lins, 2002; Thanassoulis et al., 2004; Portela e Thanassoulis, 2006) Os equivalentes não arquimedianos dos modelos (1) e (2) são apresentados em (3) e (4), onde s_i^- e s_r^+ representam as folgas para *inputs* e *outputs*, respectivamente.

$$\begin{aligned} & \text{Max } \sum_{j=1}^s u_j y_j^o \\ & \text{sujeito a} \\ & \sum_{i=1}^r v_i x_i^o = 1 \\ & - \sum_{i=1}^r v_i x_i^k + \sum_{j=1}^s u_j y_j^k \leq 0, \forall k \\ & u_j, v_i \geq \varepsilon > 0, \forall j, i \end{aligned} \quad (3)$$

$$\begin{aligned} & \text{Min } \theta_o - \varepsilon \left(\sum_{i=1}^r s_i^- + \sum_{j=1}^s s_r^+ \right) \\ & \text{sujeito a} \\ & \theta_o x_i^o - \sum_{k=1}^n x_i^k \lambda_k - s_i^- = 0, \forall i \\ & - y_j^o + \sum_{k=1}^n y_j^k \lambda_k - s_r^+ = 0, \forall j \\ & \lambda_k, s_i^-, s_r^+ \geq 0, \forall k, i, j \end{aligned} \quad (4)$$

Conforme discutido em Ali e Seiford (1993b) e Cooper et al. (2007) é bastante tentador representar $\varepsilon > 0$ por um número valor numérico explícito e muito pequeno, como 10^{-5} ou 10^{-6} . No entanto, podem ser obtidos resultados errôneos, que podem ser piores com a imposição de valores ainda menores. Ali (1990) faz uma análise de sensibilidade para diferentes valores de $\varepsilon > 0$ e mostra suas implicações no cálculo das medidas de eficiência DEA. Gomes Junior (2006) usou como critério de escolha de $\varepsilon > 0$ o valor que não tornou os problemas de programação linear (PPLs) de DEA inviáveis, nem impossibilitou a distinção das variáveis de *output* em termos dos pesos a elas atribuídos.

Surco (2004), na concepção de seu *software* DEA-SAED, usou o teorema de Ali e Seiford (1993a), que estabelece um limite superior para o elemento não arquimediano ε . Segundo esses autores, os PPLs de DEA têm soluções ótimas finitas quando $\varepsilon < 1 / \min_{k=1 \dots n} \left(\sum_{i=1 \dots r} x_{ik} \right)$.

Mehranian et al. (2000) provaram, com um contraexemplo, que o limite proposto por Ali e Seiford (1993a) é inválido e propuseram uma região de viabilidade (*assurance region*) para o valor de ε . Jahanshahloo e Khodabakhshi (2004), Alirezae (2005) e MirHassani e Alirezae (2005) apresentaram abordagens alternativas para o cálculo de intervalos de segurança para ε .

Neste artigo optou-se pela abordagem proposta por Mehrabian et al. (2000). De fato, a pesquisa bibliográfica feita sugere que esta é a abordagem base para as demais que tratam de

regiões de viabilidade para ε . Segundo essa proposta, busca-se o maior valor de ε que garanta a viabilidade do modelo dos multiplicadores, o que implica em que o modelo do envelope não seja ilimitado. A formulação (5) é representativa desta proposta e a formulação (6) é o seu dual.

$$\begin{array}{ll}
 \text{Max } \varepsilon_o & \text{Min } \beta_o \\
 \text{sujeito a} & \text{sujeito a} \\
 \sum_{i=1}^r v_i x_i^o = 1 & \sum_{i=1}^r s_i^- + \sum_{j=1}^s s_j^+ = 1 \\
 -\sum_{i=1}^r v_i x_i^k + \sum_{j=1}^s u_j y_j^k \leq 0, \forall k & \theta_o x_i^o - \sum_{k=1}^n x_i^k \lambda_k - s_i^- = 0, \forall i \\
 u_j \geq \varepsilon & \sum_{k=1}^n y_j^k \lambda_k - s_r^+ = 0, \forall j \\
 v_i \geq \varepsilon & \lambda_k, s_i^-, s_r^+ \geq 0, \forall k, i, j
 \end{array} \quad (5) \qquad (6)$$

O intervalo de viabilidade é definido por $[0, \varepsilon_o^*]$, onde ε_o^* é o valor ótimo do modelo (5). O intervalo de viabilidade global é dado por $[0, \varepsilon^*]$, onde $\varepsilon^* = \min\{\varepsilon_1^*, \dots, \varepsilon_n^*\}$. Cada elemento no intervalo global de viabilidade define um valor de viabilidade para ε . O Teorema 1 de Mehrabian et al. (2000) afirma que se $[0, \varepsilon^*]$ é o maior intervalo de viabilidade para ε , então $\varepsilon^* = \varepsilon_o^*$.

4. Modelos de Regressão Fracionada

A comparação de tratamentos (grupos) em problemas de análise de variância com respostas DEA pode ser influenciada pela correlação induzida entre as DMUs pelo cálculo da medida de performance. Ramalho et al. (2010) discutem a utilização de modelos de regressão fracionada em contextos mais gerais onde um vetor x de variáveis contextuais afeta linearmente a resposta DEA. Ramalho et al. (2010) consideram duas alternativas de modelagem. Considera-se aqui aquela derivada do modelo de Papke e Wooldridge (1996). Se y representa a resposta DEA, x o vetor de atributos contextuais (na aplicação aqui em apreço definindo um conjunto de variáveis indicadoras), e $G(\cdot)$ uma função não linear com valores em $[0,1]$ postula-se (7).

$$E(y | x) = G(x\theta) \quad (7)$$

Escolhas usuais para a função $G(\cdot)$ são definidas pela representação logística (8) e por $G(\cdot) = \Phi(\cdot)$, sendo $\Phi(\cdot)$ a inversa da função de distribuição da normal padrão. Papke e Wooldridge (1996), na realidade, sugerem como especificações possíveis qualquer função de distribuição adequada para o estudo de dados binários.

$$G(x\theta) = \frac{e^{x\theta}}{1 + e^{x\theta}} \quad (8)$$

O modelo de regressão fracionada postula, portanto, que o valor esperado da medida de performance é uma função monótona do construto $\mu = x\theta$. Para estimar θ a partir das observações $(x_i, y_i) \ i = 1, \dots, n$, procura-se pelo vetor $\hat{\theta}$ que maximiza a função de quase-verossimilhança (9).

$$\sum_{i=1}^n (y_i \log(G(x_i\theta)) + (1 - y_i) \log(1 - G(x_i\theta))) \quad (9)$$

Demonstra-se que se a função da média for especificada corretamente, então $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V)$, onde V é estimada por (10). Este é o resultado que permite inferência estatística válida envolvendo o parâmetro θ . A hipótese de observações não correlacionadas não é necessária.

$$\begin{aligned} \hat{V} &= (\hat{A})^{-1} \hat{B} \hat{A} \\ \hat{A} &= \frac{1}{n} \sum_{i=1}^n \frac{\hat{g}_i^2}{\hat{G}_i(1-\hat{G}_i)} x_i' x_i \\ \hat{B} &= \frac{1}{n} \sum_{i=1}^n \frac{\hat{u}_i^2 \hat{g}_i^2}{(\hat{G}_i(1-\hat{G}_i))^2} x_i' x_i \\ \hat{G}_i &= G(x_i, \hat{\theta}), \hat{g}_i = G'(x_i, \hat{\theta}), \hat{u}_i = y_i - \hat{G}_i \end{aligned} \quad (10)$$

5. Resultados Estatísticos e de Programação

A Tabela 2 apresenta os resultados das medidas de desempenho DEA padrão (DEA-CCR) e os valores dos multiplicadores. Para tal usou-se o *software* SIAD (Angulo Meza et al., 2005). A escolha do modelo DEA-CCR deve-se ao fato da não existência de problemas de escala na resposta (variáveis expressas como *ranks*) e do uso de DEA como uma medida de performance ao invés de característica de uma fronteira de produção empírica. Neste contexto favorece-se uma abordagem instrumentalista e menos benevolente. Testes formais de retorno à escala não são necessários.

A Tabela 3 mostra os escores de desempenho do modelo DEA não arquimediano (Desemp), assim como os valores dos multiplicadores. Esses resultados foram obtidos por procedimentos de programação via SAS, com adaptação da macro ORDEA apresentada em Emrouznejad e Ho (2012) e desenvolvimento de macros para cálculo de medidas DEA com restrições aos pesos.

A medida de desempenho Desemp tem correlação de posição de 0,929 com a medida DEA-CCR. O valor máximo da constante não arquimediana que pode ser utilizada é de 0,002652 levando a região de viabilidade $[0; 2,652 \times 10^{-2}]$ (coluna Epsil da Tabela 3). Para esse valor da constante obtêm-se as medidas Desemp. Na solução não arquimediana os pesos relativos de pessoal, custeio e capital são, respectivamente, 19,4%, 24,6% e 56,0%. Pessoal é a componente relativamente mais barata para o sistema. Em relação às componentes de produto obtêm-se 66,9%, 20,3% e 12,8% para a produção técnico-científica, publicações técnicas e de outras atividades. Os preços sombra da produção técnico-científica dominam o cálculo da medida de desempenho. *Ceteris paribus*, DMUs com boa produção técnico-científica e baixos valores de gastos com capital têm melhor performance. Note-se que não há pesos nulos nessa solução, haja vista a inclusão das restrições aos pesos dos multiplicadores no modelo não arquimediano. Os valores mais baixos da medida Desemp são esperados na solução com restrições aos pesos, pois forçam as DMUs a operarem componentes de produção que não conduzem à sua eficiência máxima, obtida com pesos nulos na solução DEA original.

A distribuição da medida de desempenho é mostrada na Figura 1, onde se apresenta um estimador não-paramétrico da função densidade de probabilidades. O gráfico sugere a superposição de populações, com um grupo com desempenho nitidamente superior.

A Figura 2 apresenta os gráficos de caixa (desenhos esquemáticos) correspondentes à distribuição do desempenho por tipo de DMU. As distribuições por tipo diferem e DMUS classificadas como do tipo produto têm performance superior, seguidas das ecorregionais e temáticas. As medianas de performance são, respectivamente, 0,610, 0,421 e 0,260. Não se observam atipicidades nas distribuições.

Tabela 2: Resultados da análise de desempenho DEA usual sob a hipótese de retornos constantes de escala (DEA-CCR) e multiplicadores.

Unidade	DEA-CCR	Multiplicadores					
		Pessoal	Capital	Custeio	PTC	PPT	POUT
DMU1	0,3934	0,0065	0,0174	0,0041	0,0125	0,0000	0,0000
DMU2	1,0000	0,0208	0,0525	0,0325	0,0409	0,0054	0,0000
DMU3	0,5997	0,0092	0,0245	0,0058	0,0175	0,0000	0,0000
DMU4	1,0000	0,0268	0,0640	0,1054	0,0699	0,0094	0,0000
DMU5	0,4871	0,0075	0,0201	0,0047	0,0144	0,0000	0,0000
DMU6	0,3382	0,0585	0,0238	0,0000	0,0187	0,0071	0,0000
DMU7	0,4525	0,0065	0,0172	0,0041	0,0123	0,0000	0,0000
DMU8	0,7617	0,0096	0,0386	0,0179	0,0242	0,0000	0,0037
DMU9	0,4233	0,0694	0,0110	0,0000	0,0178	0,0000	0,0000
DMU10	1,0000	0,0061	0,0359	0,0881	0,0410	0,0000	0,0065
DMU11	0,6555	0,0498	0,0304	0,0332	0,0280	0,0133	0,0000
DMU12	0,1411	0,0056	0,0143	0,0088	0,0111	0,0015	0,0000
DMU13	0,4661	0,0025	0,0149	0,0365	0,0170	0,0000	0,0027
DMU14	0,9118	0,0138	0,0517	0,0069	0,0298	0,0000	0,0000
DMU15	0,3320	0,0483	0,0196	0,0000	0,0155	0,0058	0,0000
DMU16	0,3968	0,0025	0,0150	0,0368	0,0171	0,0000	0,0027
DMU17	0,3476	0,0190	0,0039	0,0529	0,0000	0,0000	0,0152
DMU18	0,5278	0,0203	0,0278	0,0000	0,0194	0,0000	0,0000
DMU19	0,8025	0,0184	0,0465	0,0288	0,0362	0,0048	0,0000
DMU20	0,3382	0,0063	0,0159	0,0098	0,0124	0,0016	0,0000
DMU21	0,4492	0,0839	0,0133	0,0000	0,0215	0,0000	0,0000
DMU22	0,3053	0,0237	0,1990	0,0000	0,0517	0,0000	0,0116
DMU23	0,6492	0,0147	0,0391	0,0092	0,0280	0,0000	0,0000
DMU24	1,0000	0,0570	0,0377	0,0505	0,0313	0,0166	0,0037
DMU25	0,3487	0,0096	0,0255	0,0060	0,0182	0,0000	0,0000
DMU26	0,2191	0,0078	0,0573	0,0107	0,0255	0,0000	0,0028
DMU27	1,0000	0,1059	0,1529	0,0000	0,0000	0,0400	0,0000
DMU28	0,4037	0,0078	0,0196	0,0122	0,0153	0,0020	0,0000
DMU29	1,0000	0,1750	0,0680	0,0678	0,0000	0,0641	0,0000
DMU30	0,2306	0,0134	0,0358	0,0084	0,0256	0,0000	0,0000
DMU31	0,4625	0,0051	0,0181	0,0487	0,0234	0,0015	0,0020
DMU32	1,0000	0,0112	0,0820	0,0153	0,0365	0,0000	0,0040
DMU33	1,0000	0,0267	0,1069	0,0496	0,0671	0,0000	0,0102
DMU34	0,4153	0,0187	0,1567	0,0000	0,0407	0,0000	0,0092
DMU35	0,5395	0,0089	0,0226	0,0140	0,0176	0,0023	0,0000
DMU36	0,6415	0,0749	0,0262	0,0000	0,0000	0,0029	0,0170
DMU37	0,4277	0,0075	0,0201	0,0047	0,0144	0,0000	0,0000

Tabela 3: Resultados da análise de desempenho não arquiimediana e multiplicadores. Desemp é a medida não arquiimediana de performance, Epsil é o multiplicador ótimo da folga.

Unidade	Desemp	Multiplicadores						Epsil
		Pessoal	Capital	Custeio	PTC	PPT	POUT	
DMU1	0,1808	0,012208	0,006754	0,008663	0,002653	0,002652	0,002652	0,002652
DMU2	1,0000	0,002652	0,034659	0,104815	0,041984	0,002652	0,002652	0,007173
DMU3	0,4795	0,013266	0,015820	0,013158	0,009934	0,002652	0,002652	0,004364
DMU4	1,0000	0,051966	0,002652	0,256089	0,035665	0,030167	0,002652	0,021248
DMU5	0,2446	0,012533	0,008736	0,010208	0,004759	0,002652	0,002652	0,003017
DMU6	0,3212	0,053722	0,021966	0,002652	0,014425	0,006121	0,002652	0,006061
DMU7	0,2654	0,012302	0,007051	0,009167	0,003218	0,002652	0,002652	0,002739
DMU8	0,6100	0,014349	0,026294	0,017519	0,017584	0,002652	0,002652	0,004579
DMU9	0,3550	0,050960	0,014440	0,002652	0,011886	0,002652	0,002652	0,005261
DMU10	1,0000	0,002652	0,002652	0,901876	0,045008	0,002652	0,002652	0,018031
DMU11	0,6288	0,044632	0,029263	0,038580	0,024551	0,012919	0,002652	0,011124
DMU12	0,1188	0,012388	0,007333	0,009624	0,003734	0,002652	0,002652	0,002820
DMU13	0,3773	0,010405	0,009484	0,030908	0,011657	0,002652	0,002652	0,004107
DMU14	0,6545	0,014287	0,025694	0,017269	0,017146	0,002652	0,002652	0,004768
DMU15	0,3049	0,045189	0,018502	0,002652	0,011694	0,005089	0,002652	0,005454
DMU16	0,3305	0,010493	0,008448	0,028665	0,010336	0,002652	0,002652	0,003714
DMU17	0,3214	0,024246	0,009620	0,038261	0,002652	0,002652	0,011839	0,006505
DMU18	0,3802	0,030752	0,018154	0,002652	0,010698	0,002652	0,002652	0,004340
DMU19	0,7738	0,016711	0,049135	0,027030	0,034269	0,002652	0,002652	0,009435
DMU20	0,2287	0,012519	0,008600	0,010152	0,004660	0,002652	0,002652	0,002981
DMU21	0,3868	0,061583	0,016128	0,002652	0,014607	0,002652	0,002652	0,005573
DMU22	0,2979	0,024096	0,187948	0,002652	0,053425	0,002652	0,007257	0,009032
DMU23	0,5127	0,014377	0,026569	0,017634	0,017786	0,002652	0,002652	0,005125
DMU24	1,0000	0,044055	0,002652	0,151810	0,019357	0,015614	0,011454	0,014816
DMU25	0,1714	0,012794	0,011259	0,011259	0,006602	0,002652	0,002652	0,003326
DMU26	0,1765	0,013186	0,064621	0,002652	0,017390	0,002652	0,002652	0,004290
DMU27	1,0000	0,002652	0,002652	0,488066	0,002652	0,037745	0,002652	0,021783
DMU28	0,3398	0,013102	0,014236	0,012499	0,008777	0,002652	0,002652	0,003842
DMU29	1,0000	0,114346	0,075514	0,101186	0,062761	0,033359	0,007497	0,028381
DMU30	0,2169	0,013800	0,020987	0,015310	0,013708	0,002652	0,002652	0,004591
DMU31	0,4207	0,009952	0,014760	0,042328	0,018384	0,002652	0,002652	0,004933
DMU32	1,0000	0,015005	0,110458	0,004171	0,034807	0,002652	0,002652	0,005894
DMU33	1,0000	0,019026	0,053171	0,147438	0,076629	0,002652	0,002652	0,024939
DMU34	0,3589	0,019140	0,146343	0,002652	0,042619	0,002652	0,004826	0,007436
DMU35	0,4659	0,013596	0,019018	0,014489	0,012269	0,002652	0,002652	0,004811
DMU36	0,6149	0,069804	0,025528	0,002652	0,002652	0,003714	0,014674	0,007634
DMU37	0,2596	0,012708	0,010425	0,010912	0,005993	0,002652	0,002652	0,003329

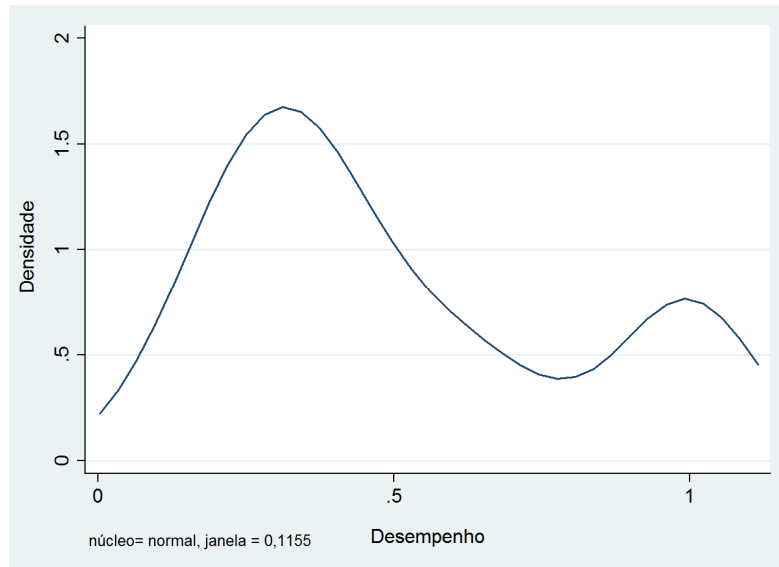


Figura 1: Função densidade de probabilidade estimada para o desempenho não arquimediano.

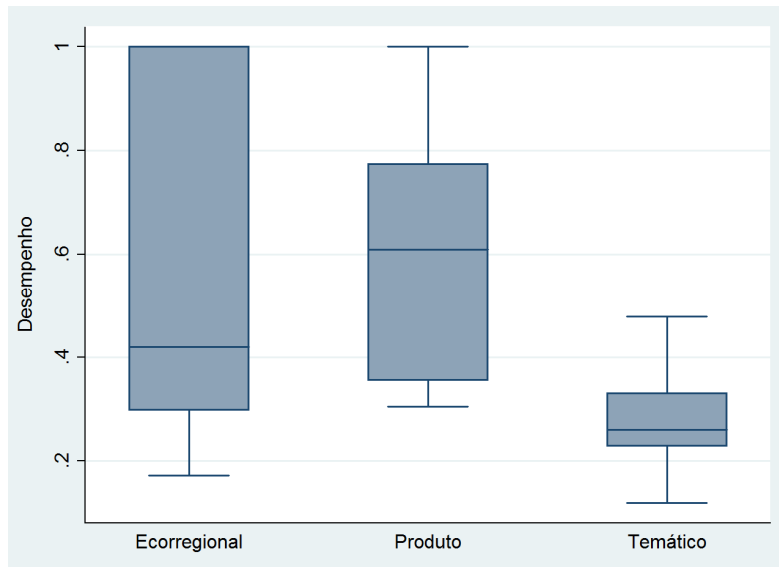


Figura 2: Desenhos esquemáticos das medidas de desempenho, segundo modelo não arquimediano, por tipo de DMU.

A análise estatística formal com o estimador de máxima quase-verossimilhança corrobora a impressão dos gráficos distribucionais. Ajustou-se o modelo com resposta logística, com resposta esperada dependente do construto linear $\mu_i = b_0 + b_1 d_i^1 + b_2 d_i^2$. Nesta expressão, as variáveis d_i^1 e d_i^2 são variáveis indicadoras dos tipos produto e temático, respectivamente, e os b_j são parâmetros desconhecidos. O vetor de estimativas de $b = (b_0, b_1, b_2)$ vem dado por $\hat{b} = (0,4811, 0,1324, -0,8395)$ evidenciando a baixa performance dos centros temáticos. A

matriz de covariâncias V é estimada por $\hat{V} = \begin{pmatrix} 0,1679529 & -0,167953 & -0,167953 \\ -0,167953 & 0,2520028 & 0,1679529 \\ -0,167953 & 0,1679529 & 0,221968 \end{pmatrix}$.

A significância conjunta dos parâmetros do modelo é avaliada através da estatística $l = \hat{b}(\hat{V})^{-1} \hat{b}'$, que se distribui como o qui-quadrado com 3 graus de liberdade sob a hipótese

$b = 0$. Na instância aqui em apreço, $l = 8,234$, com p-valor de 0,041. As diferenças entre centros temáticos e de produto, centros ecorregionais e temáticos e de temáticos e de produto têm com estatísticas de teste $z^1 = \frac{\hat{b}u'_1}{\sqrt{u_1\hat{V}u'_1}}$, $z^2 = \frac{\hat{b}u'_2}{\sqrt{u_2\hat{V}u'_2}}$, $z^3 = \frac{\hat{b}u'_3}{\sqrt{u_3\hat{V}u'_3}}$, $u_1 = (0,1,-1)$, $u_2 = (1,0,-1)$ e $u_3 = (0,1,-1)$, que se distribuem como a normal padrão sob as hipóteses de diferenças nulas. Os valores obtidos para essas estatísticas foram $z^1 = 2,616$, $z^2 = 1,550$ e $z^3 = 0,401$, com p-valores 0,004, 0,061 e 0,344, respectivamente. Vê-se que os centros de produto dominam. A diferença entre os centros ecorregionais e de produto é marginal.

6. Resumo e Considerações Finais

Neste artigo utilizou-se a constante não arquimediana ε em modelos DEA-CCR. Os modelos DEA não arquimedianos eliminam a possibilidade de o modelo do envelope ser ilimitado e o modelo dos multiplicadores ser inviável. Ao serem incorporadas folgas à função objetivo da formulação do envelope, ou restrições aos pesos no modelo dos multiplicadores, elimina-se tanto o problema de pesos zero em DEA quanto a possibilidade de DMUs falsamente eficientes.

Com base nas discussões de Mehrabian et al. (2000) foi proposta uma região de viabilidade para ε , sugeriu-se um valor para a constante e foram calculadas tanto as medidas de desempenho quanto os multiplicadores para *inputs* e *outputs*.

A aplicação de interesse foi definida pela medida de performance dos centros de pesquisa da Embrapa, considerando-se gastos com pessoal, custeio e capital como *inputs*, e produção técnico-científica, de publicações técnicas e outras atividades como *outputs*. As medidas de desempenho DEA-CCR e com restrições de folga não nula pela constante ε apresentaram correlação de Spearman de cerca de 93%. Na solução do modelo DEA não arquimediano, pessoal foi a componente relativamente mais barata para o sistema. Os preços sombra da produção técnico-científica dominaram o cálculo da medida de desempenho. Entendeu-se, assim, que DMUs com alta produção técnico-científica e gastos com capital reduzidos têm melhor desempenho.

Os tipos de centros de pesquisa foram comparados por meio de modelos de regressão fracionada e do método da máxima quase-verossimilhança. Verificou-se a dominância dos centros de pesquisa do tipo produto, com diferença marginal para os centros ecorregionais.

A utilização da constante não arquimediana no modelo DEA distribuiu melhor os pesos do problema primal (formulação dos multiplicadores) nas soluções ótimas, sem afastar-se demasiadamente da classificação original. Isso foi verdade notadamente no caso das variáveis de produto (*outputs*), onde os preços-sombra relativos passaram de 76,4%, 15,6% e 7,8% para produção técnico-científica, publicações técnicas e outras atividades, para 66,9%, 20,3% e 18,8%, respectivamente, amenizando a influência da componente técnico-científica e impondo performance em outras atividades. A distribuição para o vetor de insumos passou de 36,6%, 47,0% e 22,4% para pessoal, capital e custeio, para 19,4%, 56,4% e 24,6%, respectivamente. Fica bem mais evidente no caso não arquimediano a sugestão de controle nas componentes de custeio e capital. Os preços-sombra dos *outputs* servem também ao propósito de definir um sistema de avaliação de objetivos em que metas são comparadas com realizações de um ano base, ponderadas por pesos internos e, finalmente, agregadas via os multiplicadores relativos não arquimedianos para produzir um escore final de avaliação.

7. Referências

- Ali, A.I. Data Envelopment Analysis - Computational Issues. Computers, Environment and Urban Systems, v. 14, p. 157-165, 1990.
- Ali, A.I.; Seiford, L.M. Computational accuracy and infinitesimals in data envelopment analysis. INFOR, v. 31, n. 4, p. 290-297, 1993a.

- Ali, A.I.; Seiford, L.M. The mathematical programming approach to efficiency analysis. In: Fried, H.O.; Lovell, C.A.K.; Schmidt, S.S. (eds.) *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press: Oxford, pp. 120-159, 1993b.
- Alirezaee, M.R. The overall assurance interval for the non-archimedean epsilon in DEA models: A partition base algorithm. *Applied Mathematics and Computation*, v. 164, n. 3, p. 667-674, 2005.
- Angulo Meza, L.; Lins, M.P.E. Review of methods for increasing discrimination in data envelopment analysis. *Annals of Operations Research*, v. 116, p. 225-242, 2002.
- Angulo Meza, L.; Biondi Neto, L.; Soares de Mello, J.C.C.B.; Gomes, E.G. ISYDS - Integrated System for Decision Support (SIAD - Sistema Integrado de Apoio à Decisão): a software package for Data Envelopment Analysis model. *Pesquisa Operacional*, v. 25, n.3, p. 493-503, 2005.
- Charnes, A.; Cooper, W.W.; Lewin, A.Y.; Seiford, L.M. *Data Envelopment Analysis: Theory, Methodology and Application*. Kluwer Academic Publishers: Boston, 1996.
- Cooper, W.W.; Seiford, L.M.; Tone, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. 2nd Ed. Springer: New York, 2007.
- Emrouznejad, A.; Ho, W. *Applied Operational Research with SAS*. CRC Press: New York, 2012.
- Gomes Junior, S.F. Métodos não convencionais de restrições aos pesos em DEA aplicados ao Campeonato Mundial de Fórmula 1. *Dissertação (Mestrado em Engenharia de Produção)*, Universidade Federal Fluminense. 2006.
- Jahanshahloo, G.R.; Khodabakhshi, M. Determining assurance interval for non-archimedean element in the improving outputs model in DEA. *Applied Mathematics and Computation*, v. 151, n. 2, p. 501-506, 2004.
- Mehrabian, S.; Jahanshahloo, G.R.; Alirezaee, M.R.; Amin, G.R. An assurance interval for the non-Archimedean epsilon in DEA models. *Operations Research*, v. 48, n. 2, p. 344-347, 2000.
- MirHassani, S.A.; Alirezaee, M.R. An efficient approach for computing non-archimedean epsilon in DEA based on integrated models. *Applied Mathematics and Computation*, v. 166, n. 2, p. 449-456, 2005.
- Papke, L.E.; Wooldridge, J.M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Economics*, v. 11, n. 6, p. 619-632, 1996.
- Portela, M.C.A.S.; Thanassoulis, E. Zero weights and non-zero slacks: Different solutions to the same problem. *Annals of Operations Research*, v. 145, n. 1, p. 129-147, 2006.
- Ramalho, E.A.; Ramalho, J.J.S.; Henriques, P.D. Fractional regression models for second stage DEA efficiency analyses. *Journal of Productivity Analysis*, v. 34, p. 239-255, 2010.
- Simar, L.; Wilson, P.W. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, v. 136, n. 1, p. 31-64, 2007.
- Souza, G.S.; Avila, A.F.D.; Alves, E.R. Technical efficiency of production in agricultural research. *Scientometrics*, v. 46, n. 1, p. 141-160, 1999.
- Souza, G.S.; Gomes, E.G. Análise fatorial e análise de envoltória de dados na construção de indicadores e pesos em modelos de produção com respostas múltiplas. Embrapa/SGE: Brasília, 2013 (documento não publicado).
- Souza, G.S.; Moreira, T.B.S.; Pinto, M.B.P. Uma metodologia alternativa para mensuração de pressão sobre o mercado de câmbio. *Estudos Econômicos*. v. 34, n.1, p. 73-99, 2004.
- Surco, D.F. Desenvolvimento de uma ferramenta computacional para avaliação da eficiência técnica baseada em DEA *Dissertação (Mestrado em Métodos Numéricos em Engenharia)* Universidade Federal do Paraná. 2004.
- Thanassoulis, E.; Portela, M.C.A.S.; Allen, R. Incorporating value judgments in DEA. In: Cooper, W.W.; Seiford, L.W.; Zhu, J. (eds.) *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers: Boston, pp. 99-138, 2004.
- Torgerson, W.S. *Theory and Methods of Scaling*. New York: Wiley. 1958.