

Significância de Efeitos Técnicos na Eficiência de Produção da Pesquisa Agropecuária*

Geraldo da Silva e Souza[†]

Sumário: 1. Introdução; 2. O Modelo Embrapa de Produção de Pesquisa; 3. O Modelo de Fronteira de Produção Estocástica; 4. Análise DEA; 5. Análise de Resultados; 6. Conclusões;

Palavras-chave: fronteira estocástica; modelos Tobit; DEA; eficiência da pesquisa agropecuária.

Códigos JEL: C1; C2; C5; C6.

Neste artigo utilizam-se modelos estatísticos censurados e truncados na avaliação de efeitos técnicos que potencialmente afetam a fronteira de produção dos centros de pesquisa da Embrapa (Empresa Brasileira de Pesquisa Agropecuária). Especificamente estuda-se a significância dos fatores nível da geração fundos para pesquisa externamente ao Tesouro Nacional, ações de parceria, qualidade técnica dos projetos de pesquisa, satisfação dos clientes, mudança administrativa, tipo e tamanho, na eficiência técnica de produção dos centros de pesquisa. As medidas de eficiência técnica são calculadas com base na Análise de Envelopamento de Dados (DEA) e na Análise de Fronteiras Estocásticas de Produção. Conclui-se que as medidas de eficiência DEA são mais informativas para a avaliação dos efeitos técnicos e que a geração de renda, as ações de parceria, a qualidade técnica dos projetos e o tamanho são fatores significantes. A análise estatística que conduz a esses resultados ajusta, às medidas de eficiência DEA, um modelo do tipo Tobit definido por uma distribuição na família gama. A análise Tobit é consistente com o ajuste de uma distribuição normal truncada aos resíduos obtidos de uma função de produção DEA.

This article uses censored and truncated statistical models to assess technical effects potentially affecting the production frontier of Embrapa's (Empresa Brasileira de Pesquisa Agropecuária) research centers. Specifically the article assesses the significance of the factors level of research funds generated externally

*O autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico suporte financeiro.

[†]Empresa Brasileira de Pesquisa Agropecuária - SGE. Universidade de Brasília - Departamento de Estatística. geraldo.souza@embrapa.br



to the National Treasury, actions of partnership, technical quality of research projects, client satisfaction, change in administration, type and size on the measurement of technical efficiency of the research centers. Efficiency measures are computed both using Data Envelopment Analysis (DEA) and Stochastic Frontier methods. It is concluded that the DEA efficiency measurements provide more information regarding the assessment of technical effects and that revenue generation, actions of partnership, technical quality of projects, and size are significant factors. The statistical analysis leading to these results fits a Tobit model, defined by a distribution in the gamma family, to DEA technical efficiency measurements. This analysis is consistent with the fit of a truncated normal distribution to residuals obtained from the fit of a DEA production function.

1. INTRODUÇÃO

Neste artigo exploram-se as propriedades estatísticas dos métodos de estimação de fronteiras de produção. Utilizam-se modelos de fronteiras de produção determinísticos e estocásticos na avaliação da importância de efeitos técnicos de interesse. Os dois objetivos principais do artigo são a apresentação das classes de modelos estatísticos que podem ser usados em geral na análise de dados de fronteira de produção e a avaliação, numa situação particular, da significância de certas covariáveis de interesse administrativo (efeitos técnicos) na especificação de componentes de eficiência técnica. No contexto desta aplicação, a população de medidas de eficiência de interesse é definida em associação ao processo de produção dos centros de pesquisa da Embrapa.

Para a aplicação em apreço as medidas de eficiência técnica geradas pelos modelos de fronteiras de produção determinístico e estocástico são comparadas, quanto ao seu conteúdo informativo, relativamente a significância dos fatores técnicos, seguindo-se a abordagem de Eseinbeis et alii (1999). A discussão sobre modelagem DEA representa uma generalização dos métodos de análise propostos por Banker (1993), McCarthy e Yaisawarnng (1993), Coelli et alii (1997) e Souza (2001, 2003).

O modelo de fronteira estocástica utilizado neste artigo é o mesmo considerado em Battese e Coelli (1993). Esse modelo ajusta uma superfície resposta do tipo Translog a dados de produção para os quais a ineficiência técnica é especificada com o uso de uma distribuição normal truncada cuja média e variância são funções dos efeitos técnicos.

A consideração de modelos censurados no ambiente DEA tem por objetivo tornar viável uma análise estatística de efeitos técnicos semelhante a que se utiliza na análise de fronteiras estocásticas. A técnica consiste em calcular, em um primeiro estágio, uma medida de eficiência e, posteriormente, postular um modelo estatístico flexível, dependente dos efeitos técnicos de interesse, para aproximar a distribuição da medida de eficiência.

Embora os procedimentos de avaliação em dois estágios tenham sido criticados recentemente na literatura de produtividade, com os modelos DEA, veja-se Simar e Wilson (2002) e Wilson (2003), a abordagem é válida quando o modelo de produção é do tipo determinístico e tem respostas univariadas, i.e, possui um único produto como resposta, como demonstrado em Banker (1993) e Souza (2001, 2003). A aplicação levada a efeito neste artigo é desse tipo.

A discussão apresentada no artigo prossegue como segue. Na Seção 2 introduz-se o modelo de produção de pesquisa da Embrapa compreendendo as variáveis de produção (insumos e produtos) e os efeitos técnicos considerados de importância gerencial. A Seção 3 introduz o modelo de fronteira de produção estocástico tal como desenvolvido em Battese e Coelli (1993). A Seção 4 trata dos problemas associados a abordagem DEA e de como generalizar os modelos estatísticos sugeridos em Banker (1993) e em Souza (2001, 2003) para modelos do tipo Tobit não normais. Na Seção 5 discutem-se aspectos descritivos das medidas de eficiência calculadas para os centros de pesquisa da Embrapa. As

fronteiras estocástica e determinística (DEA) são apresentadas nas Subseções 5.1 e 5.2 respectivamente. Finalmente compila-se um resumo e apresentam-se conclusões na Seção 6.

2. O MODELO EMBRAPA DE PRODUÇÃO DE PESQUISA

O sistema Embrapa de pesquisa compõe-se de 37 centros de pesquisa. Estes centros estão distribuídos pelo país e são classificados segundo suas missões e objetivos de pesquisa como centros ecorregionais (13), centros de produto (15) e centros temáticos (9). A empresa monitora 28 variáveis de produção e 3 variáveis de insumo em seu sistema de produção de pesquisa. Não se pretende descrever em detalhes a natureza destas variáveis aqui. Apenas um breve resumo será apresentado. Uma descrição completa é encontrada em Souza et alii (1999). As variáveis de produção podem ser classificadas em quatro categorias. Produção Científica (3), Produção de Publicações Técnicas (6), Desenvolvimento de Tecnologias, Produtos e Processos (8) e Difusão de Tecnologias e Imagem (11). A categoria de Produção Científica inclui a publicação de capítulos de livro, de artigos em periódicos com referato e de artigos e resumos em *proceedings* de congressos e reuniões técnicas. A categoria de Produção de Publicações Técnicas agrupa as publicações produzidas nos centros de pesquisa visando principalmente o negócio e a produção agrícola. Típicos desta categoria são as instruções e recomendações técnicas que são publicações escritas em linguagem simplificada e dirigidas a extensionists e fazendeiros. Tais publicações contêm recomendações técnicas com respeito a sistemas de produção agropecuária. Na categoria de Desenvolvimento de Tecnologias, Produtos, e Processos se agrupam variáveis relacionadas ao esforço feito por uma unidade de pesquisa para tornar sua produção disponível para a sociedade na forma de um produto final. Típico desta categoria é a produção de novos cultivares (variedades) de planta. Finalmente, a categoria de Difusão de Tecnologias e Imagem inclui variáveis relacionadas ao esforço de uma unidade em tornar seus produtos conhecidos pelo público e ao *marketing* de sua imagem. Típicos nesta categoria são a organização de dias de campo e de unidades de demonstração. O objetivo da primeira é a difusão de conhecimento, tecnologias, e inovações aos fazendeiros e extensionists e da última é o de demonstrar resultados de pesquisa já na forma de um produto acabado.

Os 3 insumos considerados são x_1 -custos de pessoal, x_2 -custos operacionais e x_3 -custos de capital.

As variáveis de produção (produto e insumos) da Embrapa constam da tabela 1. Uma única medida de produto (y) é ali apresentada como resultado de uma média ponderada dos 28 índices de produção definidos pelo conjunto de variáveis consideradas como produtos. Cada componente dessa média é um índice que mede desempenho relativamente a um ponto de referência. Dessa forma a produção combinada dá informação sobre quanto cada centro de pesquisa desvia-se, em média, de um padrão básico especificado para todo o conjunto de variáveis de produto. Os insumos também estão medidos como índices relativos a um ponto de referência. Para os dados (y , x_1 , x_2 e x_3) da tabela 1 o ponto de referência para cada variável, quer se trate de insumo (x_i) ou produto (y), foi escolhido como a média da empresa no ano de 2000.

O uso de uma medida agregada de produto é motivada como segue. O ajuste de uma fronteira de produção quer no contexto determinístico, quer no estocástico, pressupõe homogeneidade das unidades de produção envolvidas quanto a natureza do processo de produção subjacente. Esta hipótese não se verifica estritamente na Embrapa. Embora todas as unidades produzam alguma quantidade de todas as variáveis de produção consideradas, e se utilizem dos mesmos tipos de insumo, têm percepções distintas sobre a importância relativa de cada categoria de produção. Um centro de biotecnologia, por exemplo, considera como categoria mais importante a produção científica enquanto um centro de produto, tipicamente, considera como mais importante a categoria de Desenvolvimento de Tecnologias, Produtos e Processos. A Embrapa procurou resolver este problema agregando o produto com o uso de um sistema de pesos variável por unidade de pesquisa. Tal sistema de pesos não pode ser determinado de modo automático por uma técnica como DEA, por exemplo, devido a falta de homogeneidade das unidades envolvidas na avaliação. Outro problema técnico pertinente ao método DEA é que o uso

**Tabela 1 – Variáveis de Produção**

Unidade	y	x_1	x_2	x_3
CPAA	0,524	1,224	2,442	1,414
CPATU	1,290	2,724	1,537	2,111
CNPAF	1,110	1,474	0,979	1,230
CNPGC	1,041	1,081	1,122	1,202
CNPSA	0,885	0,834	0,962	0,982
CPPSE	0,841	0,650	0,621	0,815
CNPMA	1,091	0,868	0,922	1,185
CTAA	1,929	0,754	0,846	1,158
CPAF.AP	0,955	0,317	0,333	0,304
CPAF.RO	0,439	0,598	0,640	0,442
CPAF.RR	0,837	0,344	0,468	0,345
CPAO	0,734	0,549	0,576	0,498
CPAP	0,354	0,515	0,401	0,564
CNPC	0,298	0,634	0,476	0,555
CNPDIA	0,889	0,341	0,476	0,617
CNPTIA	1,860	0,428	0,484	0,525
CPAC	1,503	2,033	1,469	1,862
CPACT	1,656	1,799	1,207	1,916
CNPGL	1,447	1,447	0,895	1,825
CNPMS	1,268	1,622	1,517	1,740
CNPSO	1,855	1,514	1,553	1,399
CENARGEN	2,234	1,987	3,973	2,798
CPAMN	0,890	1,117	0,967	0,878
CPATC	0,516	0,897	0,778	0,565
CPATSA	1,224	1,430	1,079	0,820
CNPA	0,968	0,873	0,670	0,693
CNPF	0,808	1,014	0,791	0,662
CNPH	1,381	1,030	0,956	1,084
CNPMF	1,114	1,122	1,411	0,844
CNPT	1,819	1,230	0,887	0,963
CNPUV	0,756	0,674	0,759	1,036
CNPAB	1,772	0,688	0,760	0,752
CNPAT	0,534	1,036	1,513	1,189
CNPS	0,703	0,955	0,738	0,758
CPAF.AC	0,681	0,549	0,630	0,578
CPPSUL	0,452	0,468	0,465	0,512
NMA	0,318	0,180	0,694	0,175

excessivo de variáveis de produção, sem algum tipo de agregação a priori, torna todos os centros de pesquisa eficientes.

A definição de um sistema de pesos a priori que sirva ao propósito da obtenção de uma medida univariada representativa do *ouput* e que permita comparações entre unidades é uma tarefa complexa. No caso da Embrapa os pesos foram definidos como resultado de um estudo que envolveu cerca de 500 pesquisadores e todos os administradores da empresa. A cada participante da pesquisa associada ao estudo pediu-se que manifestasse na escala 1 (menos importante) a 5 (mais importante), sua percepção sobre a importância de cada categoria de produção e de cada variável de produção em sua categoria respectiva. O modelo postulado para a análise desses dados é conhecido como Lei dos Julgamentos Categóricos e é derivado da Lei dos Julgamentos Comparativos proposta por Thurstone (1927). Esta lei teve seu uso popularizado em economia por McFadden (1974) que a utiliza para modelar preferências. A Lei dos Julgamentos Categóricos tem por objetivo transformar a escala de percepção ordinal e individual dos avaliadores numa escala intervalar representativa da população da qual a amostra de avaliadores foi extraída. A importância relativa de cada variável pode então ser determinada tendo por base as diferenças estimáveis na escala contínua. Para mais detalhes sobre este processo veja Souza et alii (2000) e Souza (2002). Os pesos resultantes da técnica de Thurstone tem estrutura semelhante aos gerados pelo método AHP de Saaty (1994).

No contexto acima descrito a Embrapa tem-se utilizado de um modelo DEA para avaliar a eficiência e a produtividade de seus centros de pesquisa. Algumas questões de interesse gerencial surgiram como consequência do processo de avaliação. Todas associadas a fatores exógenos afetando o processo de avaliação. Argumenta-se, por exemplo, que o processo de avaliação aumenta a competição interna entre os centros e inibe parcerias úteis para a pesquisa. Outras variáveis contextuais, medidas para cada centro de pesquisa, que potencialmente podem estar associadas ao nível de eficiência técnica de um centro de pesquisa, tem a ver com a ocorrência de mudanças administrativas, com o nível de satisfação dos clientes, com a qualidade técnica de projetos, com a capacidade das lideranças das unidades em obter recursos orçamentários para a pesquisa externos ao Tesouro e com o tipo e o tamanho de uma unidade.

Com o intuito de avaliar a importância das variáveis contextuais na medida de eficiência técnica este artigo propõe modelos estatísticos que incorporam efeitos técnicos associados a essas variáveis na especificação da distribuição da componente de ineficiência do modelo de fronteira. A especificação desses efeitos para a aplicação particular da Embrapa consta da Seção 5.

3. O MODELO DE FRONTEIRA DE PRODUÇÃO ESTOCÁSTICA

Seguindo Battese e Coelli (1993) e Kumbhakar e Lovell (2000), considere para um conjunto de firmas, o modelo de fronteira de produção estocástica definido por

$$y_j = \exp \{ x_j' \beta + v_j - u_j \} \quad (1)$$

onde y_j representa a produção da firma j , x_j é um vetor de dimensão k de funções conhecidas dos insumos, β é um vetor de parâmetros desconhecidos de dimensão k , os v_j são erros não correlacionados, normalmente distribuídos, com média zero e variância constante σ_v^2 e os u_j são variáveis aleatórias independentes, não negativas, relacionadas a ineficiência técnica. Supõem-se adicionalmente que u_j e v_j são independentes.

Battese e Coelli (1993) sugerem o uso da família de distribuições normais truncadas (em zero) $N(\mu_j, \sigma_u^2)$ para modelar a distribuição das componentes u_j . A inclusão de efeitos técnicos z_j no modelo é feita postulando que $\mu_j = z_j \delta$ sendo δ um parâmetro desconhecido.

A média e a variância da distribuição $N^+(\mu_j, \sigma_u^2)$ são dadas por $\mu_j + \sigma_u \lambda_j$ e



$$\sigma_u^2 \left[1 - \lambda_j \left(\frac{\mu_j}{\sigma_u} + \lambda_j \right) \right], \quad (2)$$

respectivamente, com

$$\lambda_j = \frac{\phi(\mu_j/\sigma_u)}{\Phi(\mu_j/\sigma_u)}. \quad (3)$$

As funções $\phi(\cdot)$ e $\Phi(\cdot)$ representam as funções densidade de probabilidades e de distribuição da normal padrão, respectivamente. Como a média e a variância da normal truncada são funções monotônicas de μ_j , segue que o modelo de fronteira de produção estocástica incorpora heteroscedasticidade monotônica. Pode-se modelar também as variâncias de modo a se permitir heteroscedasticidade não monotônica. Por exemplo, pode-se postular que $\sigma_{u(j)}^2 = \exp\{z_j'\gamma\}$. Esta abordagem aparece em Wang (2002) e não será utilizada aqui.

Coelli et alii (1997) sugerem que a distribuição normal truncada é suficientemente flexível na maioria das aplicações. Competitiva da formulação normal-normal truncada é a especificação do tipo normal-gama. Neste caso pode-se supor que μ_j tem distribuição $\Gamma(p, \zeta_j)$, onde o parâmetro de forma p é fixo e $\zeta_j = \exp\{-\mu_j\}$. Este modelo também assume heteroscedasticidade monotônica e generaliza o caso exponencial ($p = 1$). Contudo a formulação normal-gama não é analiticamente tratável, particularmente para a caracterização da função de verossimilhança e da distribuição condicional de dado u_j dado ϵ_j . Veja-se Greene (2003).

A especificação adequada das funções dos insumos dotam de flexibilidade o modelo de fronteira estocástica. O uso de logs, por exemplo, conduz às funções de Cobb-Douglas e ao Translog. Com normalização adicional dos dados e acrescentando-se termos em senos e cosenos ao Tranlog, obtém-se a forma flexível de Fourier (Gallant (1982). Neste artigo consideraremos a superfície resposta do modelo de fronteira estocástica na família Translog com a especificação de erros normal-normal truncada.

A eficiência técnica no modelo de fronteira de produção estocástica é orientada para produto e é definida para a j -ésima firma sob estudo por

$$\frac{E(y_j | u_j, x_j)}{E(y_j | u_j = 0, x_j)}. \quad (4)$$

Esta quantidade é precisamente $\exp\{-u_j\}$ e é estimada pela esperança condicional

$$E(\exp\{-u_j\} | \epsilon_j) \quad (5)$$

onde $\epsilon_j = v_j - u_j$.

Battese e Coelli (1993) mostram que para a especificação de erros normal-normal truncada

$$E(\exp\{-u_j\} | \epsilon_j) = \exp\{-\mu_j^* + \frac{1}{2}\sigma_*^2\} \left\{ \frac{\Phi[(\mu_j^*/\sigma_*) - \sigma_*]}{\Phi[\mu_j^*/\sigma_*]} \right\} \quad (6)$$

onde

$$\mu_j^* = \frac{\sigma_v^2 \mu_j - \sigma_u^2 \epsilon_j}{\sigma_v^2 + \sigma_u^2}$$

e

$$\sigma_*^2 = \frac{\sigma_u^2 \sigma_v^2}{\sigma_u^2 + \sigma_v^2}. \quad (7)$$

A função *log* verossimilhança a ser maximizada como função dos parâmetros β , δ , σ_u^2 e σ_v^2 , vem dada por

$$-\frac{n}{2}(\ln 2\pi + \ln \sigma^2) + \sum_{i=1}^n \left\{ \ln \left(\Phi \left(\frac{-\lambda + \mu_j/\lambda}{\sigma} \right) \right) - \ln \left(\Phi \left(\frac{\mu_j \sqrt{1 + \lambda^2}}{\sigma \lambda} \right) \right) - \frac{(\epsilon_j + \mu_j)^2}{2\sigma^2} \right\} \quad (8)$$

onde $\lambda = \sigma_u/\sigma_v$ e $\sigma^2 = \sigma_u^2 + \sigma_v^2$. Usando a reparametrização $\gamma = \sigma_u/\sigma_v$ obtém-se

$$-\frac{n}{2}(\ln 2\pi + \ln \sigma^2) + \sum_{i=1}^n \left\{ \ln \left(\Phi \left(\frac{-\gamma \epsilon_j + (1 - \gamma) \mu_j}{\sqrt{\gamma(1 - \gamma)} \sigma} \right) \right) - \ln \left(\Phi \left(\frac{\mu_j}{\sqrt{\gamma} \sigma} \right) \right) - \frac{(\epsilon_j + \mu_j)^2}{2\sigma^2} \right\}. \quad (9)$$

O ajuste do modelo de fronteira de produção estocástica pode ser levado a efeito com o uso do procedimento PROC NL MIXED do SAS ou com o uso do programa Frontier (Coelli et alii (1997)). Este último faz uso da função *log* verossimilhança reparametrizada na forma dependente do parâmetro γ .

Coelli et alii (1997) descrevem em detalhes como escolher valores iniciais para os parâmetros no processo de maximização da função de verossimilhança. Primeiramente toma-se $\delta = 0$. O estimador de mínimos quadrados ordinários para o modelo resultante fornece estimadores consistentes com exceção do intercepto β_0 e de σ_2 . Num segundo passo a função *log* verossimilhança é calculada para um grid de valores de γ no intervalo (0,1). Nesses cálculos, os estimadores de mínimos quadrados e são ajustados por

$$\beta_0 = \beta_{0(ls)} \sqrt{\frac{2\gamma\sigma^2}{\pi}} \quad (10)$$

sendo

$$\sigma^2 = \sigma_{ls}^2 \frac{(n - k)\pi}{n(\pi - 2\gamma)}. \quad (11)$$

Os melhores valores do grid são utilizados para inicializar o algoritmo de maximização da função *log* verossimilhança.

O teste estatístico da hipótese $H : \gamma = 0$ contra a alternativa unicaudal $A : 0 < \gamma < 1$ pode ser levado a efeito com uso da razão de verossimilhança. A hipótese H é equivalente a especificação do modelo de produção sem a componente de ineficiência u_j . A estatística teste é obtida calculando-se $LR = -2\ln(L(H)) - \ln(L(A))$ onde L é a função de verossimilhança. Para o teste com nível de significância α tal que $2\alpha \in (0,1)$, utiliza-se o valor crítico a para o qual $P(\chi_1^2 > a) = 2\alpha$.

Coelli et alii (1997) sugerem que valores de γ próximos a um são indicativos de que as observações suportam um modelo de fronteira determinística, i.e, um modelo sem o termo u_j . No meu nível de conhecimento um teste formal desta hipótese não está disponível na literatura no momento.

Uma medida de bondade de ajuste (*goodness of fit*) para o modelo de fronteira estocástica se obtém calculando um coeficiente de correlação entre valores observados e preditos. Para o modelo em forma logarítmica o valor predito do *log* de uma observação é dado por $\ln(x)' \hat{\beta} - (\hat{\mu} + \hat{\sigma}_u \hat{\lambda})$ onde $\ln(x)$ é o vetor com *logs* dos insumos.

4. ANÁLISE DEA

Considere um processo de produção com n firmas. Cada firma se utiliza de quantidades variáveis de s insumos na produção do produto r dimensional y . Denote por $Y = (y_1, \dots, y_n)$ a matriz de produção $r \times n$ e por $X = (x_1, \dots, x_n)$ a matriz $s \times n$ de utilização de insumos. Note-se que o elemento y_j é o



vetor produzido pela firma j e que x_j é o vetor de insumos utilizados pela firma j na produção de y_j . Supõe-se que pelo menos uma componente de cada vetor de insumos e de cada vetor de produtos seja estritamente positiva. Tipicamente, nas discussões de modelos DEA, supõe-se também que as variáveis de insumo e de produto sejam medidas em quantidades físicas. Isto não é estritamente necessário e proxies podem ser utilizadas refletindo a intensidade de uso de insumos e da produção.

É possível definir medidas DEA de eficiência técnica em vários contextos. Veja Färe et alii (1994), Coelli et alii (1997), Cooper et alii (2000) e Thanassoulis (2001). Aqui lidar-se-á com o modelo mais próximo da discussão levada a efeito sobre fronteiras estocásticas na seção precedente.

Definição 4.1. A medida de eficiência técnica (DEA), da firma j , sob a hipótese de retornos variáveis à escala e orientada para a produção, é dada pela solução do problema de programação linear $\max_{\phi, v} \phi$ sujeito a restrições

1. $v = (v_1, \dots, v_n)' \geq 0$ e $\sum_{\nu=1}^n v_\nu = 1$.
2. $Yv \geq \phi y_j$.
3. $Xv \leq x_j$.

A solução ótima $\phi^* \geq 1$, do problema da Definição 4.1, é a eficiência técnica de produção da firma j . As firmas para as quais os valores correspondentes das ineficiências técnicas são unitários são eficientes. Um valor $\phi^* > 1$ significa que a firma é ineficiente e que, no contexto do modelo de produção DEA, deve aumentar radialmente sua produção por este fator para tornar-se eficiente.

Quando $r = 1$ é possível delinear um modelo estatístico de produção relativamente ao qual as medidas de eficiência DEA têm propriedades não paramétricas ótimas. De fato, neste caso, suponha que os pares (x'_j, y_j) satisfaçam o modelo de fronteira determinística $y_j = g(x_j) - \epsilon_j$, onde $g(x)$ é uma função de produção não conhecida. Supõe-se que o espaço de insumos K é compacto e convexo em \mathcal{R}^s . A função $g(x)$ é monotônica, côncava e para todo j , $g(x) \geq y_j$. As variáveis ϵ_j são não negativas e independentemente distribuídas.

A técnica DEA pode ser utilizada para estimar $g(x)$ quando x for um elemento do conjunto

$$K^* = \left\{ x \in K; x \geq \sum_{j=1}^n v_j x_j, \quad v_j \geq 0, \quad \sum_{j=1}^n v_j = 1 \right\}. \quad (12)$$

Para cada $x \in K^*$ o estimador DEA de $g(x)$ é dado por

$$g_n^*(x) = \sup \left\{ \sum_{j=1}^n v_j y_j; \quad \sum_{j=1}^n v_j x_j \leq x \right\}. \quad (13)$$

Pode-se demonstrar que para cada j , $g_n^*(x) = \phi_j^* y_j$, onde ϕ_j^* é a eficiência técnica calculada como na Definição 4.1.

A função $g_n^*(x)$ é uma função de produção definida no conjunto compacto e convexo, K^* , i.e, é monótona, côncava e satisfaz $g_n^*(x_j) \geq y_j$. Tem também a propriedade da extrapolação máxima, i.e, para qualquer função de produção $f(x)$, $f(x) \geq g_n^*$ para $x \in K^*$. Mostra-se em Banker (1993) e em Souza (2001), que g_n^* é um estimador de máxima verossimilhança não paramétrico de $g(x)$. Souza (2001) mostra também que $g_n^*(x)$ é fortemente consistente e que os resíduos $\epsilon_{nj}^* = (\phi_j^* - 1)y_j$, quando n é grande, têm o mesmo comportamento distribucional dos resíduos ϵ_j .

Extendendo resultados devidos a Banker (1993), Souza (2001) discute duas famílias de distribuições para os ϵ_j consistentes com os resultados descritos acima. As famílias exponencial e normal truncada com parâmetros variáveis. A consideração dessas distribuições, na especificação do modelo estatístico

de fronteira determinística, permite o uso, no ambiente DEA, de métodos de análise típicos de fronteiras estocásticas de produção. Especificamente, pode-se postular heteroscedasticidade monotônica e, simultaneamente, modelar a presença de efeitos técnicos, através do estabelecimento de uma dependência dos parâmetros de localização e escala dessas duas famílias de distribuições, em construtos lineares das covariáveis de interesse.

Os resultados obtidos em Souza (2001) sugerem que os modelos censurados como o de Tobit possam ser também utilizados no ajuste das ineficiências ϵ_j . Modelos deste tipo são particularmente convenientes pois, em qualquer aplicação, alguns dos ϵ_{nj}^* serão nulos.

Uma vantagem adicional dos modelos censurados é que estes modelos podem ser utilizados no ajuste das próprias medidas de eficiência em qualquer contexto em que estas estão definidas, desde que se tenham razões para acreditar que as firmas tomem suas decisões de produção independentemente. Nestas condições, acredita-se que as correlações observadas entre as medidas de eficiência não serão fortes o suficiente para invalidar a análise. Os resultados de Banker (1993) e Souza (2001), faz-se mister observar aqui, confirmam esta conjectura no caso univariado, para amostras grandes. Os modelos de Tobit tem sido usados neste contexto na literatura. Veja-se Mccarthy e Yaisawarng (1993).

Neste artigo, além das distribuições sugeridas em Souza (2001), propõe-se, para o caso de produção unidimensional, extensões do tipo Tobit supondo heteroscedasticidade e não normalidade. Com este propósito exploram-se duas extensões para o modelo Tobit. O Tobit-gama e o Tobit-normal truncada.

Para os modelos Tobit começa-se a discussão com o Tobit clássico (normal) prosseguindo-se com as generalizações para o Tobit-normal heteroscedástico, Tobit-normal truncada e Tobit-gama. A notação utilizada para os construtos lineares dos efeitos técnicos é a mesma da seção anterior.

Considere o modelo de Tobit clássico com ponto de censura a onde a resposta w_j (o resíduo ϵ_{nj}^* do modelo de fronteira determinística, ou qualquer medida de eficiência técnica DEA) é definida por

$$w_j = \begin{cases} w_j^* & \text{se } w_j^* > a \\ a & \text{se } w_j^* \leq a \end{cases} \quad (14)$$

com $w_j^* = \mu_j + o_j$. Os erros aleatórios o_j têm distribuição $N(0, \sigma^2)$. Nas aplicações de interesse para este artigo $a = 1$ ou $a = 0$. Os parâmetros δ_j dos efeitos técnicos presentes em μ_j podem ser estimados pelo método de máxima verossimilhança. Esta abordagem exige a maximização da função de verossimilhança

$$L(\delta, \sigma) = \prod_{j; w_j = a} \Phi\left(\frac{a - \mu_j}{\sigma}\right) \prod_{j; w_j > a} \frac{1}{\sigma} \phi\left(\frac{w_j - \mu_j}{\sigma}\right) \quad (15)$$

relativamente a δ e σ .

A versão heteroscedástica do modelo de Tobit que será considerada aqui supõe que $\sigma_j^2 = \exp\{l_j' \xi\}$, sendo l_j um vetor de observações em variáveis que afetam a variância e ξ o vetor de parâmetros correspondente. Com modificações óbvias na função de verossimilhança procura-se agora pelo máximo relativamente a δ e ξ .

Apresentam-se agora outras versões do modelo de Tobit que permitem censura num valor $a > 0$. Primeiramente suponha que a distribuição de w_j^* seja a distribuição normal truncada com média μ_j e variância σ^2 . A função de verossimilhança para este modelo vem dada por

$$L(\delta, \sigma) = \prod_{j; w_j = a} \frac{\Phi\left(\frac{a - \mu_j}{\sigma}\right) - \Phi\left(-\frac{\mu_j}{\sigma}\right)}{\Phi(\mu_j/\sigma)} \prod_{j; w_j > a} \frac{\phi\left(\frac{w_j - \mu_j}{\sigma}\right)}{\sigma \Phi(\mu_j/\sigma)}. \quad (16)$$

Finalmente suponha que a distribuição de w_j^* seja gama com parâmetro de forma $p > 0$ e parâmetro de escala $\zeta_j = \exp\{-\mu_j\}$, onde, como anteriormente, $\mu_j = z_j' \delta$. A distribuição exponencial é um caso



particular com $p = 1$. Denote por $\Gamma(\cdot)$ a função gama e por $G_p(\cdot)$ a função de distribuição da distribuição gama com parâmetro de forma p e parâmetro de escala unitário. A função de verossimilhança para a versão Tobit-gama vem dada por

$$L(p, \delta) = \prod_{j; w_j = a} G_p(\zeta_j) \prod_{j; w_j > a} \frac{\zeta_j^p w_j^{p-1} \exp\{-\zeta_j w_j\}}{\Gamma(p)} \quad (17)$$

No contexto de modelos estatísticos do tipo Tobit, para avaliar o ajuste, bem como a direção de influência dos efeitos marginais na variável resposta, é necessário calcular o valor esperado da variável resposta, dada as observações nos efeitos técnicos e valores para os parâmetros. Para o modelo Tobit clássico tem-se

$$E(w_j | z_j, \delta) = \Phi\left(\frac{a - \mu_j}{\sigma}\right) + \mu_j \left(1 - \Phi\left(\frac{a - \mu_j}{\sigma}\right)\right) + \sigma \phi\left(\frac{a - \mu_j}{\sigma}\right) \quad (18)$$

com modificações óbvias para acomodar o caso heteroscedástico. Para o modelo Tobit-normal truncada, tem-se

$$E(w_j | z_j, \delta) = \frac{\Phi\left(\frac{a - \mu_j}{\sigma}\right) - \Phi\left(-\frac{\mu_j}{\sigma}\right) + \phi\left(\frac{a - \mu_j}{\sigma}\right) + \mu_j \left(1 - \Phi\left(\frac{a - \mu_j}{\sigma}\right)\right)}{\Phi\left(\frac{\mu_j}{\sigma}\right)}. \quad (19)$$

Finalmente, para a especificação Tobit-gama,

$$E(w_j | z_j, \delta) = G_p(\zeta_j) + \frac{\Gamma(p+1)}{\zeta_j \Gamma(p)} (1 - G_{p+1}(\zeta_j)). \quad (20)$$

As três esperanças matemáticas mostradas acima são funções monotônicas do construto μ_j .

5. ANÁLISE DE RESULTADOS

A Tabela 2 mostra valores das variáveis contextuais e das eficiências obtidas com as abordagens DEA (Rvs) e de fronteira estocástica (Sf) para o ano 2000. As fronteiras são tratadas em detalhes em 5.1 e 5.2. Note-se que, com a notação da Seção 4, $Rvs = 1/\phi^*$. A variável Sf é calculada com o uso da especificação completa do Translog, incluindo os efeitos técnicos na componente de ineficiência.

As variáveis contextuais de interesse são a intensidade da geração de fundos para a pesquisa (Rp), ações de parceria (Ap), satisfação de clientes (Sc), qualidade técnica dos projetos de pesquisa (Qt), tipo, tamanho e mudança administrativa (Vd) ocorrida durante os seis primeiros meses do período de avaliação ou durante os seis meses precedentes a este período. As variáveis tipo, tamanho e Vd são categóricas. Tipo tem níveis E (ecorregional), T (temático) e P (produto). Tamanho tem níveis G (grande), M (Médio) e P (pequeno). A variável Vd assume o valor 1 se houve mudança administrativa na unidade e o valor 0 de outro modo. A classificação em tamanho foi levada a efeito com a utilização de análise de conglomerados, com o método de Ward, aplicada aos insumos. Rp, Ap, Sc e Qt são escores quantitativos normalizados para variarem entre 0 e 1. A escala, em qualquer caso, é monotônica com o valor 1 representando o maior valor. Rp e Ap são índices relativos. A variável Rp é o quociente dos fundos obtidos pelo total dos recursos liberados pelo Tesouro para a unidade, excetuando-se despesas com pessoal. Ap é definida como o número médio de parcerias por pesquisador. Sc e Qt derivam de escores de avaliação calculados para a empresa por consultores externos.

A correlação de postos entre Rvs e Sf é de 32% indicando uma associação de fraca intensidade entre as abordagens DEA e de fronteira estocástica. Isto sugere que as classificações das firmas segundo os dois critérios diferem. Os valores medianos (Tabela 3) das medidas de eficiência técnica Rvs e Sf apontam na mesma direção quanto a tipo. Os centros temáticos tem performance superior aos centros ecoregionais e de produto. Para Rvs os centros maiores tem melhor performance e não se nota diferença

Tabela 2 – Variáveis Contextuais e Medidas de Eficiência

Unidade	Tipo	Tamanho	Vd	Rp	Ap	Sc	Qt	Rvs	Sf
CPAA	E	L	0	0,465	0,126	0,951	0,743	0,261	0,808
CPATU	E	L	0	0,427	0,064	0,947	0,843	0,646	0,793
CNPAF	P	M	0	0,251	0,132	0,960	0,866	0,572	0,875
CNPGC	P	M	0	0,228	0,180	0,953	0,791	0,534	0,844
CNPSA	P	M	0	0,513	0,303	0,984	0,980	0,461	0,863
CPPSE	P	M	0	1,000	1,000	0,969	0,882	0,446	0,827
CNPMA	T	M	0	0,339	0,236	0,931	0,745	0,564	0,967
CTAA	T	M	0	0,261	0,365	0,929	0,865	1,000	0,971
CPAFAP	E	S	0	0,111	0,266	0,967	0,799	1,000	0,424
CPAFRO	E	S	0	0,157	0,160	0,942	0,749	0,289	0,425
CPAFRR	E	S	0	0,186	0,234	0,909	0,810	0,746	0,468
CPAO	E	S	0	0,285	0,198	0,964	0,727	0,419	0,500
CPAP	E	S	0	0,217	0,429	0,889	0,751	0,261	0,370
CNPC	P	S	0	0,080	0,000	0,987	0,819	0,165	0,319
CNPDIA	T	S	0	0,172	0,283	0,984	0,859	0,712	0,843
CNPTIA	T	S	0	0,062	0,047	0,973	0,663	1,000	0,909
CPAC	E	L	1	0,295	0,100	0,989	0,716	0,755	0,873
CPACT	E	L	1	0,339	0,248	0,964	0,878	0,843	0,887
CNPGL	P	L	1	0,197	0,172	0,976	0,837	0,748	0,814
CNPMS	P	L	1	0,818	0,600	1,000	0,766	0,636	0,759
CNPSO	P	L	1	0,758	0,095	0,980	0,911	0,933	0,896
CENARGEN	T	L	1	0,035	0,168	0,956	0,842	1,000	0,962
CPAMN	E	M	1	0,176	0,235	0,911	0,817	0,465	0,936
CPATC	E	M	1	0,098	0,165	0,927	0,632	0,277	0,917
CPATSA	E	M	1	0,358	0,122	0,900	0,797	0,641	0,956
CNPA	P	M	1	0,489	0,177	1,000	0,860	0,514	0,931
CNPF	P	M	1	0,085	0,137	0,922	0,744	0,429	0,905
CNPH	P	M	1	0,257	0,318	0,964	0,770	0,716	0,918
CNPMF	P	M	1	0,172	0,239	0,956	0,857	0,583	0,919
CNPT	P	M	1	0,326	0,059	0,964	0,863	0,949	0,948
CNPUV	P	M	1	0,358	0,143	0,938	0,752	0,395	0,905
CNPAB	T	M	1	0,392	0,220	0,958	1,000	0,936	0,978
CNPAT	T	M	1	0,243	0,031	0,979	0,765	0,271	0,969
CNPS	T	M	1	0,698	0,138	0,958	0,716	0,372	0,975
CPAFAC	E	S	1	0,377	0,394	0,927	0,827	0,364	0,589
CPPSUL	P	S	1	0,263	0,128	0,947	0,800	0,260	0,419
NMA	T	S	1	0,165	0,958	0,967	0,804	1,000	0,867

apreciável entre os centros médios e pequenos. Para Sf os centros médios tem a melhor performance seguido dos grandes e pequenos. Para ambas as variáveis os centros pequenos são os que mostram a pior performance. A mudança administrativa conduziu a valores maiores de Rvs e Sf. As análises de covariância (Ancova) nos postos de Rvs e Sf (Tabelas 4 e 5), como esperado, diferem. A Ancova é marginalmente significativa para Rvs e altamente significativa para Sf. Para a variável Rvs, Rp e Qt são significantes com tipo, tamanho e Ap seguindo em importância. Não se nota significância das variáveis Vd e Sc. Para a análise com a resposta Sf, tipo, tamanho e Vd são altamente significantes e Ap é significativa. As variáveis Rp, Sc e Qt não são significantes. Sabe-se que as análises de covariância, com sabor não-paramétrico, como consideradas aqui, são robustas relativamente a presença de correlação moderada nas observações da variável resposta, atipicidades, heteroscedasticidade e não normalidade. A consistência dessas Ancovas com as abordagens paramétricas é um dos critérios utilizados na avaliação dos modelos paramétricos.

5.1. Análise da Fronteira Estocástica

Para os dados da Embrapa, a especificação na forma Translog leva ao modelo estatístico

**Tabela 3 – Eficiências Medianas**

Classificação	Nível	Rvs	Sf
Tipo	E	0,465	0,793
	P	0,534	0,876
	T	0,936	0,967
Tamanho	P	0,419	0,467
	M	0,524	0,924
	G	0,752	0,823
Vd	0	0,549	0,817
	1	0,636	0,917

Tabela 4 – Análise de Covariância para os postos de Rvs

Fonte	gl	Soma de Quadrados	Quadrado Médio	F	Valor p
Modelo	9	1748,742	194,305	2,13	0,062
Rp	1	404,420	404,420	4,44	0,045
Ap	1	200,335	200,335	2,20	0,150
Sc	1	55,912	55,912	0,61	0,440
Qt	1	402,748	402,748	4,42	0,045
Vd	1	2,879	2,879	0,03	0,860
Tamanho	2	408,806	204,403	2,24	0,125
Tipo	2	362,566	181,283	1,99	0,156
Erro	27	2458,758	91,065		
Total	36	4207,500			

Tabela 5 – Análise de Covariância para os postos de Sf

Fonte	gl	Soma de Quadrados	Quadrado Médio	F	Valor p
Modelo	9	3913,243	434,805	38,52	< 0,001
Rp	1	0,878	0,878	0,08	0,782
Ap	1	58,039	58,039	5,14	0,032
Sc	1	0,121	0,121	0,01	0,918
Qt	1	8,733	8,733	0,77	0,387
Vd	1	261,175	261,175	23,14	< 0,001
Tamanho	2	1229,615	614,807	54,47	< 0,001
Tipo	2	1187,292	593,646	52,59	< 0,001
Erro	27	304,757	11,287		
Total	36	4218			

$$\ln y_j = \beta_0 + \beta_1 \ln x_{j1} + \beta_2 \ln x_{j2} + \beta_3 \ln x_{j3} + \beta_4 (\ln x_{j1})^2 + \beta_5 (\ln x_{j2})^2 + \beta_6 (\ln x_{j3})^2 + \beta_7 \ln x_{j1} \ln x_{j2} + \beta_8 \ln x_{j1} \ln x_{j3} + \beta_9 \ln x_{j2} \ln x_{j3} + v_j - u_j \quad (21)$$

com

$$\mu_j = \delta_0 + \delta_1 \text{Rp}_j + \delta_2 \text{Ap}_j + \delta_3 \text{Sc}_j + \delta_4 \text{Qt}_j + \delta_5 \text{Vd}_j + \delta_6 \text{M}_j + \delta_7 \text{G}_j + \delta_8 \text{T}_j + \delta_9 \text{P}_j \quad (22)$$

sendo M e G variáveis indicadoras para os centros médios e grandes e T e P variáveis indicadoras de centros temáticos e de produto, respectivamente. As variáveis x_i representam os insumos, y a medida de produto e Rp, Ap, Sc, Qt e Vd as demais variáveis contextuais.

Na análise preliminar de fronteira estocástica compararam-se as especificações Translog (modelo completo) e Cobb-Douglas (modelo reduzido). O modelo ajustado em cada caso superpõe à estrutura de erros

normais $N(0, \sigma_v^2)$, ineficiências $N^+(\mu_j, \sigma_u^2)$, onde μ_j é o construto linear definido pelos efeitos técnicos. O processo de estimação, via máxima verossimilhança, foi levado a efeito com o software Frontier. Os critérios de informação de Akaike e de Shwarz¹ (tabela 6) não apontam para o mesmo melhor modelo. O critério de Akaike indica a formulação Translog e o de Shwarz a formulação Cobb-Douglas. A estatística do teste da razão de verossimilhança é significativa no nível de 5% e aponta na direção do Translog.

Tabela 6 – Estatísticas de ajuste do modelo de fronteira estocástica. ll é o valor do log da verossimilhança e k é o número de parâmetros. LR é a estatística teste da hipótese nula Cobb-Douglas

Modelo	$-2ll$	k	LR	AIC	BIC	Valor p
Cobb-Douglas	30,28	16	12,8	62,3	88,1	0,0462
Translog	17,48	22		61,5	96,9	

No contexto do ajuste do Translog (Tabela 7) a estatística do teste da razão de verossimilhança, para a presença de efeitos técnicos, tem por valor 13,4 com 9 graus de liberdade. O valor p correspondente é 0,145 e não é significativo. Este resultado difere marcadamente da análise de covariância da medida de eficiência técnica gerada por este modelo.

Tabela 7 – Fronteira estocástica Translog. Listagem do Frontier

Variável	Estimativa	Desvio Padrão	z
Intercepto (Produção)	0,179	0,113	1,583
$\ln(x_1)$	0,125	0,354	0,354
$\ln(x_2)$	-0,058	0,366	-0,157
$\ln(x_3)$	0,426	0,434	0,981
$[\ln(x_1)]^2$	-1,422	1,117	-1,272
$[\ln(x_2)]^2$	-1,453	0,695	-2,090
$[\ln(x_3)]^2$	-0,096	0,776	-0,124
$\ln(x_1) \ln(x_2)$	2,837	0,962	2,951
$\ln(x_1) \ln(x_3)$	0,802	1,574	0,600
$\ln(x_2) \ln(x_3)$	0,185	1,029	0,179
Intercepto (Efeitos técnicos)	0,712	0,839	0,849
Rp	-0,459	0,887	-0,517
Ap	0,462	0,899	0,513
Sc	0,176	0,841	0,209
Qt	-0,057	0,864	-0,066
Vd	-0,287	0,341	-0,841
M (tamanho)	-1,088	0,545	-1,997
G (tamanho)	-0,579	0,570	-1,015
T (tipo)	-0,864	0,652	-1,327
P (tipo)	0,315	0,423	0,744
σ^2	0,132	0,053	2,511
γ	0,307	0,244	1,259

O valor obtido para o teste da razão de verossimilhança da hipótese é $LR = 13,41$ e este valor excede o nível crítico de 2,71 para um teste no nível de significância 5%, favorecendo o modelo estocástico. Para o Translog a correlação de posições entre valores observados e preditos é 0,578.

¹O critério de informação de Akaike é definido por $-2ll+2k$ onde ll é a função log verossimilhança e k o número de parâmetros ajustados. O critério bayesiano de Shwarz é definido por $-2ll+k\log(n)$ onde n é o número de firmas. O melhor modelo deve possuir o menor valor para ambos os critérios.



5.2. Análise das Respostas DEA

O modelo DEA ajustado para a Embrapa aos dados da Tabela 1 assume retornos variáveis à escala. É o mesmo da Definição 4.1. Os insumos são x_1 , x_2 e x_3 e o produto y .

Muitas publicações na área de análise DEA fazem uso das medidas de eficiência correspondentes como variáveis dependentes em problemas de regressão. Exemplos típicos de tal abordagem aparecem em McCarty e Yaisawarng (1993), Arnold, Bardan e Cooper (1994), Eisenbeis, Ferrier e Kwan (1999) e Sathie (2001). Estas aplicações implicitamente assumem que as medidas de eficiência DEA são não correlacionadas. Esta hipótese se justifica para o modelo de produção determinístico, com respostas univariadas, como demonstrado em Banker (1993) e Souza (2001). Em casos mais gerais (produto multidimensional) Banker e Natarajan (2004) sugerem que os mesmos resultados se verifiquem. É possível também que se possa estender os resultados de Souza (2001) para funções custo, mas esses resultados ainda não estão disponíveis na literatura. Como suporte adicional, para a aplicação presente, aos resultados assintóticos de Banker (1993) e Souza (2001), procuramos por evidência de correlações significantes entre as medidas de eficiência DEA (ϕ^*) dos centros de pesquisa da Embrapa, através da realização do teste de corridas (*runs test*). Veja Spiegel e Stephens (1999). Sob a hipótese de independência o número de corridas $R = 20$ (calculado depois de arranjar os dados dos centros de pesquisa em ordem alfabética pelo nome do centro) tem distribuição normal com média e variância

$$\mu_R = \frac{2N_1N_2}{N_1 + N_2} + 1 \quad \text{e} \quad \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1 + N_2)^2(N_1 + N_2 - 1)}, \quad (23)$$

respectivamente. Nestas expressões $N_1 = 18$ é o número de observações abaixo da mediana (1,74825) e $N_2 = 18$ é o número de observações acima da mediana. Segue que o teste estatístico da hipótese de independência vem dado por $z = (20 - 19)/2,9568 = 0,338$ e não é significativo.

Para o ajuste dos resíduos ϵ_{nj}^* consideram-se quatro modelos probabilísticos consoantes com a discussão levada a efeito na Seção 4. Estes são o Tobit-clássico, o Tobit heteroscedástico (com função de ligação exponencial para a variância dependente do tamanho de cada centro de pesquisa), a distribuição normal truncada e a distribuição exponencial. O ajuste estatístico desses modelos foi levado a efeito com o PROC NL MIXED do SAS (Tabelas 8 e 9). O melhor modelo segundo os critérios de informação é o obtido da distribuição normal truncada. A correlação de posições entre valores observados e preditos para esse modelo é 0,587. Rp, Ap, Qt, tamanho e tipo são fatores significantes. Aumentos em Ap e Qt, tudo o mais constante, implicam em aumento do escore de efeito técnico e portanto conduzem a um nível maior de eficiência pois os resíduos, na média, decrescem. A variável Rp aponta na direção contrária de Ap e Qt. A habilidade de um centro em gerar recursos de fomento à pesquisa não afeta a eficiência técnica positivamente.

Tabela 8 – Estatísticas de ajuste para os modelos definidos para os resíduos $\epsilon^* = (\phi^* - 1)y$. ll é o log da verossimilhança e k o número de parâmetros

Modelo	$-2ll$	k	AIC	BIC
Tobit	40,5	11	62,5	80,2
Tobit heteroscedástico	35,5	13	61,5	82,4
Exponencial	35,0	10	55,0	71,1
Normal truncada	17,2	11	39,2	56,9

Os modelos tendo como resposta (ϕ^*) foram ajustados, com o uso do PROC QLIM e do PROC NL MIXED do SAS (Tabelas 10 e 11). Estes são o Tobit, o Tobit heteroscedástico, o Tobit-normal truncada, o Tobit-exponencial e o Tobit-gama. Os critérios de informação e a função de verossimilhança dão indicação de que a melhor alternativa é o modelo Tobit-gama. Para este modelo a correlação de posições entre valores observados e preditos é 0,557. Vê-se aqui que os resultados estatísticos concordam em geral com a análise dos resíduos do modelo de produção determinístico. As variáveis importantes na explicação

Tabela 9 – Ajuste da normal truncada aos resíduos $\epsilon^* = (\phi^* - 1)y$. Método de máxima verossimilhança. Listagem SAS

Variável	Estimativa	Desvio Padrão	t	Valor p
Intercept	4.310	3.710	1.16	0.252
Rp	1,522	0,592	2,57	0,014
Ap	-1,139	0,568	-2,01	0,052
Sc	-0,534	3,851	-0,14	0,890
Qt	-3,856	1,369	-2,82	0,008
Vd	-0,167	0,183	-0,91	0,367
M (tamanho)	0,117	0,228	0,51	0,612
G (tamanho)	-0,587	0,316	-1,85	0,072
T (tipo)	-0,720	0,307	-2,34	0,025
P (tipo)	0,051	0,250	0,20	0,839
σ^2	0,168	0,052	3,24	0,003

do nível de ϕ^* são Rp, Ap, Qt, tamanho e, marginalmente, tipo. Como no caso dos resíduos $\epsilon_{n,j}^*$, Rp aponta em direção oposta a Ap e Qt. Esses resultados mostram muito mais concordâncias com a Ancova (não paramétrica) do que os obtidos com Sf.

Tabela 10 – Estatísticas de ajuste para os modelos definidos para ϕ^* . ll é a log verossimilhança e k o número de parâmetros. Listagem SAS

Modelo	$-2ll$	k	AIC	BIC
Tobit	99,32	11	121,3	139,0
Tobit heteroscedástico	85,35	13	111,4	132,3
Normal truncada	97,00	11	119,0	136,7
Exponencial	119,9	10	139,9	156,0
Gama	88,20	11	110,2	128,0

Tabela 11 – Tabela 9: Ajuste do modelo Tobit-Gama a ϕ^* . Método de máxima verossimilhança. Listagem SAS

Variável	Estimativa	Desvio Padrão	t	Valor p
p	5,792	1,468	3,94	< 0,001
Intercepto	1,580	3,533	0,45	0,657
Rp	1,040	0,468	2,22	0,032
Ap	-1,080	0,437	-2,47	0,018
Sc	-0,349	3,957	-0,09	0,930
Qt	-2,499	1,023	-2,44	0,020
Vd	-0,095	0,154	-0,62	0,539
M (tamanho)	-,239	0,197	-1,21	0,232
G (tamanho)	-0,712	0,260	-2,73	0,001
T (tipo)	-0,411	0,260	-1,58	0,122
P (tipo)	0,050	0,229	0,22	0,830

6. CONCLUSÕES

Investigou-se o conteúdo informativo das medidas de eficiência de produção estocástica e determinística da produção de pesquisa dos centros da Embrapa, em termos de associações com fatores técnicos de importância gerencial. A abordagem no cálculo das medidas de eficiência é voltada para a produção



e assume retornos variáveis à escala. No caso estocástico a especificação da fronteira é feita com o uso da forma flexível Translog. No caso não-paramétrico (determinístico) utilizou-se um modelo DEA com retornos variáveis. Esta abordagem considera, como pares, na avaliação de cada unidade, apenas unidades aproximadamente de mesmo tamanho que a unidade sendo avaliada.

Do ponto de vista do ajuste estatístico de modelos paramétricos, a medida de eficiência técnica com mais conteúdo informativo é fornecida pela abordagem DEA. Esta mostra também consistência com a análise de covariância não-paramétrica. Quanto a abordagem estocástica não se detecta significância dos efeitos técnicos nem consistência com a análise de covariância não paramétrica correspondente.

Para a abordagem DEA o ajuste da distribuição normal truncada aos resíduos da produção e o modelo Tobit-gama, com censura unitária, produzem essencialmente os mesmos resultados e representam os melhores ajustes.

Os efeitos técnicos, associados às variáveis contextuais, significantes são o esforço na captação de recursos para a pesquisa externos ao Tesouro, ações em parceria, qualidade técnica de projetos, tipo e tamanho. Observa-se que o esforço na captação de recursos tem efeito negativo sobre a medida de eficiência de produção. Por outro lado observam-se efeitos positivos associados às ações em parcerias e na qualidade técnica de projetos. As mudanças administrativas ocorridas na vizinhança do tempo de avaliação não parecem ter causado mudanças significantes na eficiência. Os centros maiores tendem a ser mais eficientes. O mesmo se aplica aos centros temáticos. Em parte como resultado dessas observações a Embrapa tem tornado mais justo seu sistema de avaliação. Tem procurado também facilitar a administração das suas unidades melhorando processos e envidando esforços para a liberação mais expedita de recursos do Tesouro. Outra ação gerencial importante para a melhora dos níveis de eficiência e produtividade foi o estabelecimento de metas de operação mais claras e métodos mais rígidos de acompanhamento da execução dos planos diretores e do desempenho institucional. Espera-se que tais medidas tendam a reverter, por exemplo, o efeito negativo de Rp. Faz-se mister ressaltar que a direção da Embrapa tem conseguido manter um nível mediano crescente de eficiência técnica ao longo do tempo. O desafio importante do momento é o de associar a esse crescimento uma redução significativa na variabilidade das medidas de eficiência dos centros de pesquisa. O efeito positivo das parcerias na avaliação da eficiência respondeu a uma crítica importante ao sistema de avaliação. As parcerias não inibem eficiência e devem ser estimuladas. Neste contexto duas decisões gerenciais importantes foram considerados. A manutenção do uso das medidas de eficiência como critério de avaliação institucional e o incentivo às parcerias internas e externas com a criação de um indicador de desempenho institucional adicional incluindo explicitamente o indicador de parcerias.

As medidas de eficiência técnica DEA sob a hipótese de retornos constantes apresentam correlação de posição apreciável (cerca de 69%) com as calculadas sob a hipótese de retornos variáveis. Isto sugere que os resultados obtidos aqui para as medidas calculadas com retornos variáveis tenham contrapartida semelhante quando as medidas são calculadas sob retornos constantes. Contudo, as duas medidas diferem bastante em magnitude. Este efeito escala leva a necessidade do uso das medidas de eficiência com retornos variáveis na modelagem para atenuar o efeito do tamanho nas avaliações.

Referências Bibliográficas

- Banker, R. D. (1993). Maximum likelihood, consistency and dea: a statistical foundation. *Management Science*, 39(10):1265–1273.
- Banker, R. D. & Natarajan, R. (2004). Statistical tests based on dea efficiency scores. In Cooper, W. W., Seiford, L. M., & Zhu, J., editors, *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Norwell.
- Battese, G. E. & Coelli, T. J. (1993). A stochastic frontier production function incorporating a model for

- technical inefficiency effects. Technical Report 69, Department of Econometrics, University of New England, Armidale.
- Coelli, T., Rao, D. P., & Battese, G. E. (1997). *An Introduction to Efficiency and Productivity Analysis*. Springer.
- Cooper, W. W., Tone, K., & Seiford, L. M. (2000). *Data Envelopment Analysis, A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer, Boston.
- Eseinbeis, R. A., Kwan, S. H., & Ferrier, G. D. (1999). The informativeness of stochastic frontier and programming frontier efficiency scores: cost efficiency and other measures of bank holding company performance. *Working Paper. Federal Reserve Bank of Atlanta*, pages 99–23.
- Färe, R., Grosskopf, S., & Lovell, C. A. K. (1994). *Production Frontiers*. Cambridge University Press, New York.
- Gallant, A. R. (1982). Unbiased determination of production technologies. *Journal of Econometrics*, 20:285–323.
- Greene, W. H. (2003). Simulated likelihood estimation of the normal-gama stochastic frontier function. *Journal of Productivity Analysis*, 19:179–190.
- Kumbhakar, S. & Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, New York.
- Mccarthy, T. A. & Yaisawarng, S. (1993). *The Measurement of Productive Efficiency*. Oxford University Press, New York.
- Sathie, M. (1990). X-efficiency in australian banking: an empirical investigation. *Journal of Banking and Financing*, 25:613–630.
- Simar, L. & Wilson, P. W. (2002). Estimation and inference in two-stage, semi-parametric models of production processes. Discussion Paper no 0307, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Bélgica.
- Souza, G. S. (2001). Statistical properties of data envelopment analysis estimators of production function. *Brazilian Review of Econometrics*, 21:291–322.
- Souza, G. S. (2002). The law of categorical judgement revisited. *Brazilian Journal of Probability and Statistics*, 16:123–140.
- Souza, G. S. (2003). Funções de produção, uma abordagem estatística com o uso de modelos de encapsulamento de dados. Texto para Discussão No. 17. Embrapa Informação Tecnológica Brasília. DF.
- Souza, G. S., Ávila, A. F. D., & Alves, E. (1997). Produtividade e eficiência relativa de produção em sistemas de produção de pesquisa agropecuária. *Revista Brasileira de Economia*, 51(3):281–307.
- Souza, G. S., Ávila, A. F. D., & Alves, E. (1999). Technical efficiency in agricultural research. *Scientometrics*, 46:141–16.
- Souza, G. S., Ávila, A. F. D., & Alves, E. (2000). Psicometria linear da escalagem ordinal: uma aplicação na caracterização da importância relativa de atividades de produção em ciência e tecnologia. *Cadernos de Ciência e Tecnologia*, 17(3):11–27.
- Spiegel, M. R. & Stephens, L. J. (1999). *Theory and Problems of Statistics*. Schawm's outline. McGraw-Hill, New York, 3 edition.



- Thanassoulis, E. (2001). *Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software*. Kluwer Academic Publishers, Boston.
- Wang, H. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis*, 18:241–253.
- Wilson, P. W. (2003). Testing independence in models of productive efficiency. *Journal of Productivity Analysis*, 20(3):361–390.