

MODELO DE PRODUÇÃO PARA A PECUÁRIA DE CORTE BRASILEIRA COM BASE EM DADOS CENSITÁRIOS: ABORDAGENS DETERMINÍSTICA E ESTOCÁSTICA COM HIPÓTESE DE ENDOGENIDADE¹

Geraldo da Silva e Souza²
Eliane Gonçalves Gomes³
Urbano Gomes Pinto de Abreu⁴

RESUMO: Ajustou-se um modelo de produção para a pecuária de corte brasileira a partir de informações do Censo Agropecuário de 2006 no nível municipal. Optou-se por um modelo de produção que considera a renda bruta da atividade como função de gastos com terra, com mão de obra, com insumos e taxa de lotação. Covariáveis de interesse são indicadoras dos biomas brasileiros, uma variável ambiental e presença da assistência técnica. Consideraram-se duas especificações: uma função de produção na família Cobb-Douglas e uma fronteira DEA em dois estágios com retornos variáveis. Encontrou-se evidência estatística em favor do modelo de fronteira estocástica com componentes endógenas. A influência da variável ambiental é neutra nesse modelo e a componente de assistência técnica implica, *ceteris paribus*, em efeito negativo para a produção, resultado provável da presença de imperfeições de mercado. Notaram-se diferenças significativas na resposta de produção esperada nos diferentes biomas. A relação é dominada pelo bioma Pantanal na função de produção.

Palavras-chave: pecuária de corte, censo agropecuário, fronteira estocástica, DEA em dois estágios, endogeneidade.

A CENSUS-DATA BASED BRAZILIAN BEEF CATTLE PRODUCTION MODEL: DETERMINISTIC AND STOCHASTIC APPROACHES WITH ENDOGENEITY HYPOTHESIS

ABSTRACT: We have adjusted production model for the Brazilian beef cattle production, based on the 2006 Agricultural Census data at the municipal level. We have chosen a production model that considers the gross income of the activity as a function of the expenditure on land, labor, inputs and stocking rate. Covariables of interest are the functions indicating the Brazilian biomes, an environmental variable and the presence of technical assistance. Two specifications have been considered: a Cobb-Douglas production function and a two-stage DEA model with variable returns to scale. We found statistical evidence in favor of the endogenous stochastic frontier approach. The influence of the environmental variable was neutral in this model and the technical assistance component, *ceteris paribus*, implied a negative effect for the production, probably due to the presence of market imperfections. We have observed significant differences in the expected production response for the different biomes. The relationship in the production function was dominated by the Pantanal biome.

¹Registrado em CCTC, REA-07/2017.

²Matemático, Economista, PHD em Estatística, EMBRAPA, Secretaria de Gestão e Desenvolvimento Institucional (SGI), Brasília, Distrito Federal, Brasil (e-mail: geraldo.souza@embrapa.br).

³Engenheira Química, Doutora, EMBRAPA, Secretaria de Gestão e Desenvolvimento Institucional (SGI), Brasília, Distrito Federal, Brasil (e-mail: eliane.gomes@embrapa.br).

⁴Médico Veterinário, Doutor, EMBRAPA Pantanal, Corumbá, Mato Grosso do Sul, Brasil (e-mail: urbano.abreu@embrapa.br).

Key-words: *beef cattle, agricultural censos, stochastic frontier, two-stage dea approach, endogeneity.*

JEL Classification: *Q1, C52, C60.*

1 - INTRODUÇÃO

A atividade pecuária bovina é uma das principais atividades do meio rural brasileiro. O ramo pecuário representou, em 2015, 32% do PIB do agronegócio e 7% do PIB total do país, segundo dados do Centro de Estudos Avançados em Economia Aplicada (CEPEA, 2016).

Para atender à crescente demanda por carne bovina, os diferentes sistemas de produção pecuários têm incorporado ciência e tecnologia no processo produtivo, com consequente aumento dos índices de desempenho da produção. Conforme destacam Abreu et al. (2016), a atividade sustenta grande parte da estrutura da cadeia de produção de carne e os investimentos tecnológicos que nela se fizerem, com aumento da eficiência de produção, trarão benefícios para toda cadeia produtiva. Por outro lado, em um futuro próximo, a produção animal será afetada por restrições advindas de políticas de mitigação de emissões de gases do efeito estufa e de legislações ambiental e de bem-estar animal (THORNTON, 2010). Em paralelo, os sistemas de pastejo tropicais são complexos e requerem abordagem multidisciplinar para serem manejados com eficiência, sendo direcionados para intensificação (BERNDT; TOMKINS, 2013).

Nesse contexto, a taxa de lotação torna-se uma importante variável dos sistemas de produção pecuários. A taxa de lotação reflete a capacidade de suporte de pastejo, ou seja, é a pressão de utilização que permite equilíbrio entre o ganho por animal por unidade de área, maximizando o rendimento. A disponibilidade de forragem controla simultaneamente a qualidade e a quantidade das pastagens e, conseqüentemente, possibilita que as plantas se mantenham produtivas. Ao mesmo tempo, determina a taxa de lotação e define o sistema de produção animal. Daí a importância de considerar a interação da

disponibilidade de forragem e da produção animal quando da tomada de decisão em relação ao ajuste da taxa de lotação (EUCLIDES FILHO, 2001). O desempenho animal é diretamente afetado pela taxa de lotação, ou seja, a taxa de lotação é uma variável controle, enquanto a produtividade por animal constitui-se em uma variável resposta. A magnitude do efeito é modulada por diferentes variáveis que interagem dinamicamente (BARIONI et al., 2008). Abreu et al. (2013), trabalhando com taxa de lotação com objetivo de construir tipologias em escala municipal, verificaram grande diferença entre os sistemas de produção de gado de corte nos diferentes biomas do Brasil.

O objetivo deste artigo é estimar um modelo de produção para a pecuária de corte brasileira, utilizando as informações disponíveis no Censo Agropecuário de 2006 no nível municipal. Optou-se por um modelo de fronteira de produção que considera a renda bruta da atividade como função de gastos com terra, gastos com mão de obra, gastos com insumos e taxa de lotação (*proxy* para capital financeiro). Como variáveis contextuais, consideraram-se variáveis indicadoras dos biomas brasileiros, uma componente ambiental e assistência técnica. Neste estudo, foram consideradas duas classes de modelos: modelos de fronteira estocástica com componentes endógenas e modelos de análise de envoltória de dados em dois estágios. Nesse contexto, as abordagens metodológicas propostas visam contribuir para a formulação de estratégias para o setor pecuário.

2 - MODELO DE PRODUÇÃO

Conforme já apresentado, o objetivo da análise é estimar um modelo de produção para a pecuária de corte brasileira municipal, tomando por base informações disponíveis no Censo Agropecuário de

2006 (IBGE, 2009). Os dados aqui usados estão disponíveis na base censitária do IBGE e foram gerados a partir dos microdados do Censo Agropecuário de 2006, conforme estratégia amostral descrita em Souza et al. (2013).

Optou-se por modelar a renda bruta da atividade pecuária de corte como função dos gastos com terra estimados como fluxo, gastos com mão de obra na atividade, gastos com insumos específicos (sal mineral e medicamentos) e taxa de lotação como *proxy* do uso de capital. Os gastos com mão de obra na atividade incluem salários pagos a família e a empregados, e contratação de serviços de terceiros. Os gastos com mão de obra para a atividade pecuária não estão disponíveis nas bases de dados do IBGE. Nesse contexto, como *proxy* para gastos com mão de obra na pecuária, considerou-se o produto dos gastos totais com mão de obra pela proporção da área de pastagem relativamente à área total. Entende-se que a pecuária usa menos mão de obra por unidade de área do que outras atividades e, portanto, a *proxy* utilizada pode superestimar a elasticidade da mão de obra. Mesmo assim, a evidência empírica encontrada é que esse insumo, embora significativa estatisticamente, é o que menos afeta a produção, e o viés de alta não altera a conclusão do estudo. O fluxo de gastos com terra foi determinado como em Souza et al. (2013) e corresponde a 4% do valor da terra. A taxa de lotação foi definida pelo quociente entre o número de animais e a área de pastagem em hectares.

Uma função de produção define o máximo de produto y possível, dados os níveis de insumo definidos pelo vetor x . As condições tecnológicas são caracterizadas por uma função real não negativa $f(\cdot)$ – função de produção, monótona não decrescente em x e côncava e, desse modo, $y = f(x)$. A forma funcional mais utilizada na especificação de uma função de produção é a forma funcional Cobb-Douglas, na qual o log do produto é a uma função linear, com coeficientes positivos, dos logs dos insumos. Tipicamente, utilizam-se como insumos as componentes de capital, trabalho e terra em aplicações na agropecuária. O termo constante dessa relação é interpretado como nível tecnológico prevalente. Outras variáveis

podem afetar a especificação, provocando deslocamentos na relação de produção, e o nível de eficiência de um determinado produtor. Tais variáveis são consideradas contextuais. Expectativas quanto ao sinal dessas componentes (efeitos crescentes ou decrescentes) podem ser estabelecidas para essas variáveis, mas não se impõem condições teóricas *a priori* sobre os sinais. Variáveis contextuais consideradas neste artigo são as variáveis indicadoras dos seis biomas brasileiros (Amazônia, Caatinga, Cerrados, Mata Atlântica, Pampas e Pantanal), uma componente ambiental, dada pela razão área de pastagem degradada por área total de pastagem, e assistência técnica (intensidade da assistência técnica municipal), definida pela proporção de estabelecimentos no município que receberam assistência técnica. A degradação das pastagens compromete a rentabilidade da pecuária brasileira e a recuperação das pastagens é uma estratégia promissora para a intensificação sustentável da pecuária (SILVA et al., 2017). No estudo de Alves et al. (2012), foi observado impacto positivo e significativo da assistência técnica para maior eficiência da atividade agropecuária como um todo. No entanto, ao categorizar as atividades por tipo de produção principal, a componente de assistência técnica deixou de ser significativa para o grupo pecuária.

Na análise aqui levada a efeito forma considerados 2.584 municípios, para os quais há observações completas válidas para todos os biomas, os quais representam 46% da população total de municípios brasileiros em 2006.

3. - ABORDAGEM METODOLÓGICA

3.1 - Fronteira Estocástica

Referências básicas para nossa discussão nesta seção são Khumbhakar e Lovell (2000), Coelli et al. (2005), Greene (2011) e Stata (2015). Os modelos de fronteira de produção apareceram primeiramente na literatura no artigo de Aigner, Lovell e Schmidt (1977). Segue-se o mesmo enfoque de Souza et al. (2013).

As ideias básicas envolvidas na análise econômica de fronteiras de produção estocásticas passam inicialmente pela especificação de uma função de produção real $f(x, z, \theta)$ dependente do vetor de insumos x de dimensão k , do vetor de efeitos contextuais z de dimensão g e de um vetor paramétrico θ de dimensão finita d . Sem erros aleatórios e ineficiência, o máximo de produção y_j que pode ser obtido pelo estabelecimento j com o uso de x_j , na presença das covariáveis z_j , é dado por $y_j = f(x_j, z_j, \theta)$.

A possibilidade de ineficiência no processo de produção pressupõe a existência de uma componente estocástica $\eta_j \in (0,1)$, tal que a produção na realidade vem dada por $y_j = f(x_j, z_j, \theta)\eta_j$. Se η_j se aproxima de 1, isto significa que o estabelecimento tem sua produção próxima do ótimo, definido pela função de produção $f(x, z, \theta)$. Quando $\eta_j < 1$, o estabelecimento não está produzindo o máximo possível em face da tecnologia disponível para o conjunto de produtores e incorporada na função de produção $f(x, z, \theta)$.

Tipicamente, as observações de produção também estão sujeitas a variações aleatórias resultantes de efeitos de per se desprezáveis, mas que apresentam deslocamentos na produção. Desse modo, é comum postular também a presença de choques estocásticos na função de produção e presumir a existência de variáveis aleatórias reais v_j , tais que $y_j = f(x_j, z_j, \theta)\eta_j \exp(v_j)$.

A especificação acima é equivalente ao modelo estatístico $\ln y_j = \ln f(x_j, z_j, \theta) + v_j - u_j$, onde u_j é uma variável aleatória não negativa representando a componente de ineficiência do modelo, isto é, $u_j = -\ln(\eta_j)$.

Uma função de produção de uso corrente em Teoria de Produção vem dada pela especificação Cobb-Douglas $f(x, z, \theta) = C \prod_{v=1}^k x_v^{\beta_v} \exp(z'w)$. Nesta representação, $\theta = (\beta, w)$, sendo $\beta_v > 0$, a elasticidade do insumo x_v . Portanto, tipicamente, tomando logs, obtém-se a representação

$$\ln(y_j) = \ln(C) + \sum_{v=1}^k \beta_v \ln(x_{v,j}) + \sum_{l=1}^g w_l z_{l,j} + v_j - u_j.$$

Especificações estocásticas distintas para as

componentes de erro levam a modelos de fronteira alternativos. Tipicamente, assume-se que os v_j são distribuídos independentemente da componente de ineficiência u_j . Representam uma amostra aleatória da distribuição normal com média zero e variância σ^2 . Para u_j , assumem-se observações independentes provenientes da distribuição exponencial com variância σ_u^2 , da distribuição meia normal, proveniente do truncamento positivo da normal com média zero e variância σ_u^2 , ou da distribuição normal truncada resultante do truncamento positivo da distribuição normal com média μ e variância σ_u^2 . Ineficiências esperadas são dadas por σ_u^2 para a distribuição exponencial, $\sqrt{2/\pi}\sigma_u$ para a distribuição meia normal, e $\mu + \sigma_u \lambda$ com $\lambda = \phi(\mu/\sigma_u)/\Phi(\mu/\sigma_u)$ para a distribuição normal truncada, onde $\phi(\cdot)$ e $\Phi(\cdot)$ são as funções densidade de probabilidades e de distribuição de probabilidades da normal padrão.

O vetor de parâmetros θ do modelo é estimado para n observações ou estabelecimentos pelo método de máxima verossimilhança. O processo de inferência estatística é válido assintoticamente. As seguintes funções são otimizadas na obtenção da estimativa $\hat{\theta}$ do vetor θ (STATA, 2015):

1) Modelo normal-exponencial:

$$L(\theta^*) = \sum_{j=1}^n \left\{ -\ln \sigma_u + \frac{\sigma^2}{2\sigma_u^2} + \ln \left(\frac{-\varepsilon_j - \frac{\sigma^2}{\sigma_u}}{\sigma} \right) + \frac{\varepsilon_j}{\sigma} \right\}$$

2) Modelo normal-meia normal:

$$L(\theta^*) = \sum_{j=1}^n \left\{ \frac{1}{2} \ln \left(\frac{2}{\pi} \right) - \ln(\sigma_s) + \ln \left(-\frac{\rho \varepsilon_j}{\sigma_s} \right) - \frac{\varepsilon_j^2}{2\sigma_s^2} \right\}$$

3) Modelo normal-normal truncada:

$$L(\theta^*) = \sum_{j=1}^n \left\{ \frac{1}{2} \ln(2\pi) - \ln(\sigma_s) - \ln \Phi \left(\frac{\mu}{\sigma_s \sqrt{\gamma}} \right) + \ln \Phi \left(\frac{(1-\gamma)\mu - \gamma \varepsilon_j}{\sigma_s \sqrt{\gamma(1-\gamma)}} \right) \right\} - \sum_{j=1}^n \left\{ \frac{1}{2} \left(\frac{\varepsilon_j + \mu}{\sigma_s} \right)^2 \right\} - \sum_{j=1}^n \left\{ \frac{1}{2} \left(\frac{\varepsilon_j + \mu}{\sigma_s} \right) \right\}$$

Nas expressões acima, $\varepsilon_j = v_j - u_j$ representa a diferença $\ln y_j - \ln f(x_j, z_j, \theta)$ entre a variável

resposta e a parte determinística do modelo, $\sigma_s^2 = \sigma^2 + \sigma_u^2$, $\rho = \sigma_u / \sigma$ e $\gamma = \sigma_u^2 / \sigma_s^2$. O parâmetro θ^* inclui θ e a parametrização adicional usada na componente de ineficiência.

Efeitos associados a variáveis contextuais afetando a eficiência técnica são modelados por meio dos parâmetros envolvidos nas especificações das distribuições associadas à ineficiência. Nos casos exponencial e meia normal, postula-se que $\sigma_{uj}^2 = \exp(m_j' b)$, onde m é um vetor de covariáveis e b o vetor de efeitos correspondentes. Para a distribuição normal truncada, postula-se $\mu_j = m_j' b$. O valor esperado da ineficiência em qualquer caso é uma função monótona do construto linear $m_j' b$. No caso da normal truncada, o valor esperado vem dado por $\mu_j + \sigma_u \lambda_j$, $\lambda_j = \phi(\mu_j / \sigma_u) / \Phi(\mu_j / \sigma_u)$, sendo $\phi(\cdot)$ a função densidade da distribuição normal padrão e $\Phi(\cdot)$ sua função de distribuição. Heteroscedasticidade nas componentes v e/ou u pode ser obtida impondo um tipo análogo de especificação para as respectivas componentes de variância. Tal opção é tipicamente utilizada nas especificações exponencial e meia normal, e não é necessária para a normal truncada.

Em nossa aplicação, a possibilidade da existência de variáveis contextuais endógenas exige a utilização de técnica de variáveis instrumentais. Nesse contexto, utilizou-se a técnica sugerida por Karakaplan e Kutlu (2017) e Karakaplan (2017) para a estimação do modelo de fronteira estocástica e o teste de endogeneidade associado. Neste artigo, optou-se pela especificação normal-meia normal, com as variáveis contextuais afetando diretamente a função de produção. A medida de eficiência técnica te_j é, assim, estimada por (KARAKAPLAN, 2017):

$$te_j = \left(\frac{1 - \Phi(\sigma_* - \mu_{*j} / \sigma_*)}{1 - \Phi(-\mu_{*j} / \sigma_*)} \right) \exp\left(-\mu_{*j} + \frac{1}{2} \sigma_*^2\right)$$

onde:

$$\mu_{*j} = -e_j \frac{\sigma_u^2}{\sigma_s^2}, \quad \sigma_s^2 = \sigma_u^2 + \sigma_w^2, \quad \text{e} \quad \sigma_*^2 = \sigma_w^2 \sigma_u^2 / \sigma_s^2,$$

$\sigma_w = \sigma \sqrt{1 - \xi' \xi}$, sendo ξ o vetor de correlações do

erro da regressão das variáveis endógenas nos instrumentos W_j , com o erro da variável resposta no modelo de produção:

$$e_j = y_j - f(x_j, z_j, \theta) - \xi'(z_j^* - W_j \delta)$$

Nesta última expressão, z_j^* representa o subvetor de z considerado endógeno e δ um vetor de parâmetros.

A função de verossimilhança a ser estimada é a que segue (KARAKAPLAN, 2017):

$$\log L(\theta) = \sum_{j=1}^n \left\{ \frac{\ln(2/\pi) - \ln(\sigma_s^2) - (e_j / \sigma_s^2)}{2} + \ln \Phi\left(\frac{\xi_j}{\sigma_s}\right) \right\} + \sum_{j=1}^n \left\{ \frac{-p \ln 2\pi - \ln(\det \Omega) - \tilde{\epsilon}_j' \Omega^{-1} \tilde{\epsilon}_j}{2} \right\}$$

com $\xi = \sigma_u / \sigma_w$, p é a dimensão de z_j^* , $\tilde{\epsilon}_j = z_j^* - W_j \delta$, Ω é a matriz de variância-covariância de $\tilde{\epsilon}_j$. Outras formulações podem ser testadas e ajustadas considerando o modelo de fronteira $y_j = f(x_j, z_j, \theta) + \xi'(z_j^* - W_j \delta) + w_j - u_j$, estimando o modelo de variáveis instrumentais por mínimos quadrados ordinários e considerando uma das formulações clássicas apresentadas acima. Nesta formulação, o erro idiossincrático w_j tem variância σ_w^2 .

3.2 - DEA em Dois Estágios

Considere-se um processo de produção com n unidades de produção. Cada unidade se utiliza de quantidades variáveis de S insumos para produzir um único produto y . Represente-se por $Y = (y_1, \dots, y_n)$ o vetor de produção $1 \times n$ e por $X = (x_1, \dots, x_n)$ a matriz $S \times n$ de uso de insumos no processo de produção.

Seja K um subconjunto compacto, convexo e com interior não vazio do ortante não negativo de R^S . O nível de produção máximo (fronteira) realizável com o uso de $x \in K$ é determinado pela função de produção $y = g(x)$. Supõe-se que $g(x)$ seja contínua e que, adicionalmente:

- Se $x \geq w$ são vetores de K então $g(x) \geq g(w)$.
- Se x e w são vetores de K e $t \in [0; 1]$, então $t g(x) + (1 - t) g(w) \leq g(tx + (1 - t)w)$.

- Para cada $j = 1, \dots, n, x_j \in K, g(x_j) \geq y_j$.

- Com respeito à escala de operação da tecnologia definida por $g(x)$, diz-se que:

a) $g(x)$ mostra retornos constantes à escala se $g(\eta x) = \eta g(x)$ para todo $\eta \geq 0$ e $x \in K$ tal que $\eta x \in K$.

b) $g(x)$ mostra retornos crescentes à escala se $g(\eta x) \geq \eta g(x)$ para todo $\eta \geq 1$ e $x \in K$ tal que $\eta x \in K$.

c) $g(x)$ mostra retornos decrescentes à escala se $g(\eta x) \leq \eta g(x)$ para todo $\eta \geq 1$ e $x \in K$ tal que $\eta x \in K$.

d) $g(x)$ mostra retornos variáveis à escala se nenhuma das propriedades (a), (b), ou (c) se verifica.

Pode-se utilizar o conjunto de observações (x_j, y_j) e o modelo de programação matemática conhecido como análise de envoltória de dados (DEA) para estimar $g(x)$ somente para x em um dos conjuntos seguintes:

$$K_1^* = \left\{ x \in K; x \geq \sum_{j=1}^n \lambda_j x_j, \lambda_j \geq 0 \right\},$$

$$K_2^* = \left\{ x \in K; x \geq \sum_{j=1}^n \lambda_j x_j, \lambda_j \geq 0, \sum_{j=1}^n \lambda_j \geq 1 \right\},$$

$$K_3^* = \left\{ x \in K; x \geq \sum_{j=1}^n \lambda_j x_j, \lambda_j \geq 0, \sum_{j=1}^n \lambda_j \leq 1 \right\},$$

$$K_4^* = \left\{ x \in K; x \geq \sum_{j=1}^n \lambda_j x_j, \lambda_j \geq 0, \sum_{j=1}^n \lambda_j = 1 \right\}.$$

Cada um desses conjuntos pressupõe uma escala de operação distinta para a função de produção DEA: K_1^* impõe retornos constantes, K_2^* retornos crescentes, K_3^* retornos decrescentes e K_4^* retornos variáveis. Entende-se que a hipótese de trabalho mais adequada aos dados em questão é a de retornos variáveis.

Para $x \in K_4^*$, define-se a função $g_n^*(x) = \sup \left\{ \sum_j \lambda_j y_j; \sum_j \lambda_j x_j \leq x \right\}$, onde o sup é restrito aos valores λ satisfazendo as restrições K_4^* .

Para cada r , $g_n^*(x_r) = \phi_r^* y_r$, onde ϕ_r^* é o valor ótimo do problema de programação linear $\max_{\phi, \lambda} \phi$, sujeito a $\sum_j \lambda_j y_j \geq \phi y_r, \sum_j \lambda_j x_j \leq x_r,$

$\lambda \geq 0$ e $\sum_j \lambda_j = 1$. Tipicamente em análises envolvendo DEA modela-se o inverso de ϕ_r^* - uma medida de eficiência no intervalo (0,1]. Para maiores detalhes sobre modelos DEA, sugerem-se Souza (2003), Coelli et al. (2005) e Cooper, Seiford e Zhu (2011).

Seja $\varphi_r^* = 1 / \phi_r^*$ na presença de um vetor de fatores contextuais z , Ramalho et al. (2010) assumem a dependência funcional $E(\varphi_r^* | z_j) = F(z_j \beta)$ e estimam β por quasi-verossimilhança otimizando $l(\beta) = \eta \left(\sum_{\varphi_j^* < 1} \log F(z_j \beta) \right) + (1 + \eta) \left(\sum_{\varphi_j^* > 1} \log (1 - F(z_j \beta)) \right)$, sendo η a função indicadora de uma unidade eficiente. As propriedades estatísticas desse estimador são descritas em Papke e Wooldridge (1996) e Ramalho, Ramalho e Henriques (2010). Note-se que no caso em estudo $z=(1, \text{tipos de bioma, pastagem degradada, assistência técnica})$, sendo tipos definidos por um conjunto de variáveis indicadoras. A análise está disponível no Stata v. 14 (STATA, 2015). A possibilidade de endogeneidade de algum componente contextual pode invalidar a análise. Nesse caso, pode-se considerar métodos com a utilização de variáveis instrumentais (SOUZA; GOMES, 2015; RAMALHO; RAMALHO, 2017). Neste artigo, a possibilidade de endogeneidade das variáveis pastagem degradada e assistência técnica levou a estimativa do valor esperado da eficiência por meio de mínimos quadrados não lineares em dois estágios.

3.3 - Considerações

A abordagem mais indicada para o uso de um modelo de fronteira de produção diz respeito à existência ou não de uma componente de erro idiosincrática representando choques aleatórios não controláveis pelas firmas na definição da fronteira. A abordagem DEA é indicada na ausência dessa componente de erro. Nesse contexto, desvios do máximo de produção devem-se somente à presença de ineficiência de produção.

As duas especificações foram utilizadas na presente análise com resultados mais satisfatórios e intuitivos do ponto de vista estatístico para a formu-

lação da fronteira estocástica. A correlação de valores preditos com observados para o modelo estocástico é superior a 87%, e a quantidade similar no modelo de fronteira determinística é de 41%. Além disso, a formulação de fronteira estocástica rejeita a hipótese $\sigma_u = 0$ contra a alternativa $\sigma_u > 0$ significativamente, esta última hipótese equivalente à formulação com as duas componentes de erro. É interessante observar adicionalmente que o teste de aleatoriedade da amostra – *runs test* (BRADLEY, 1968) – indica correlação serial negativa para eficiências DEA retornos variáveis, invalidando a teoria assintótica relevante para os modelos fracionários. Portanto, é necessário utilizar técnicas de reamostragem no cálculo de desvios padrão e intervalos de confiança, o que cria dificuldades adicionais de análise.

4 - RESULTADOS ESTATÍSTICOS

A análise inicia-se com a formulação DEA com retornos variáveis e orientada a produto. As variáveis de produção são: produto – renda com a atividade pecuária de corte, insumos – gastos com mão de obra, gastos com terra, gastos com insumos específicos e taxas de lotação. Como fatores contextuais que afetam a medida DEA, tomaram-se as variáveis indicadoras dos biomas, a variável ambiental e a intensidade da assistência técnica municipal. Como já salientado, observações completas válidas para todos os biomas são em número de 2.584 municípios. O processo inferencial está restrito a esses casos. Dados perdidos são resultados de restrições à liberação de informação pelo IBGE e à ausência do valor de alguma variável no nível municipal.

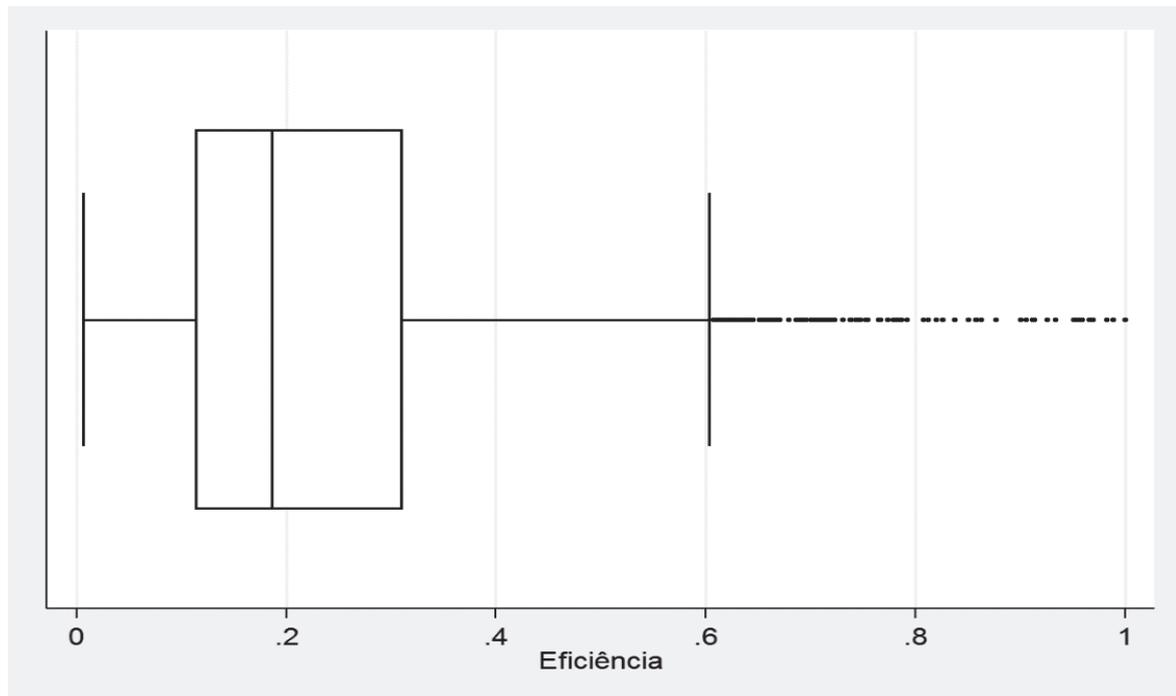
O resumo de cinco números da medida DEA de eficiência vem dado por Min=0,006, Q1=0,160, Mediana=0,186, Q3=0,310 e Max=1. A distribuição é acentuatadamente assimétrica à direita, com valores típicos muito pequenos, indicando um nível alto de ineficiência. Observações com nível de eficiência maior do que 0,6 são raras e representam apenas cerca de 5% da população. O gráfico de caixa (*box-plot*) e o histograma da figura 1 ilustram essas observações.

A tabela 1 apresenta os resultados estatísticos do modelo de regressão fracionária para a medida DEA levado a efeito com mínimos quadrados não lineares em dois estágios. As variáveis contextuais são as *dummies* de bioma, o indicador ambiental e a assistência técnica. As duas últimas foram consideradas endógenas. O modelo passa no teste de especificação de Hansen, com p-valor de 8,3%. Somente as variáveis indicadoras dos biomas são significantes. Não é necessário utilizar desvios robustos via *bootstrap*, uma vez que esses desvios dominam a estimativa usual e, portanto, não alteram a conclusão quanto à significância.

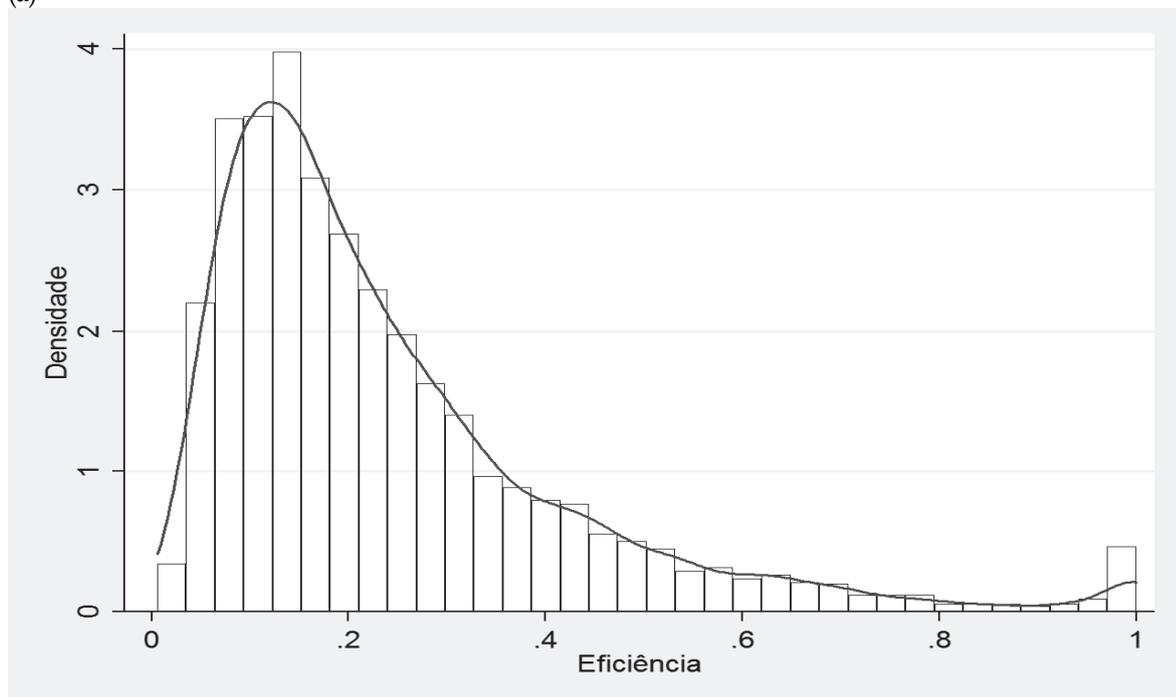
A variável ambiental tem o sinal correto, apesar de não significativa. O aumento da área degradada implica em redução monotônica do nível de eficiência técnica. Analogamente, o sinal negativo da assistência técnica é compatível com outros estudos (SOUZA; GOMES; ALVES, 2017) e, provavelmente, é resultante de imperfeições de mercado. Imperfeições de mercado são as principais causas que inibem o acesso dos agricultores à tecnologia e, conseqüentemente, à inclusão produtiva. Esse conceito é discutido em Alves e Souza (2015). São o resultado de assimetria em crédito, infraestrutura de produção, disponibilidade de informação, extensão rural e assistência técnica, entre outros. As imperfeições de mercado são tipicamente desfavoráveis à pequena produção, que é dominante na agricultura brasileira.

Notam-se diferenças significativas entre biomas, com dominância para o bioma Pantanal, seguido por Amazônia, Pampa, Cerrados, Caatinga e Mata Atlântica. Esses resultados são aparentes na tabela 2, obtidos via modelos de regressão fracionária com correção de viés e 2.000 repetições *bootstrap*, desconsiderando as variáveis potencialmente endógenas, dado que não há evidência de que afetem significativamente a medida de eficiência técnica (Tabela 1).

Relativamente à fronteira estocástica, foram consideradas três formulações distintas: um modelo normal-normal truncado, um modelo normal-meia normal e um modelo normal-exponencial, com a especificação da variável ambiental e de assistência téc-



(a)



(b)

Figura 1 - Box-Plot (a) e Histograma (b) da Medida de Eficiência DEA, Brasília, Distrito Federal, Janeiro de 2017.
Fonte: Elaborada pelos autores a partir do STATA (2015).

Tabela 1 - Ajuste do Modelo de Regressão Fracionária via Mínimos Quadrados Não Lineares em Dois Estágios¹, Brasil, Janeiro de 2017

Variável	Coefficiente	Desvio padrão	Estatística t	Pr > t
Constante	0,7065	0,2956	2,39	0,0169
Variável ambiental	-5,4723	3,3490	-1,63	0,1024
Assistência técnica	-0,1628	0,1437	-1,13	0,2576
Amazônia	-0,3771	0,2087	-1,81	0,0708
Caatinga	-1,0328	0,2004	-5,15	<0,0001
Cerrados	-0,9050	0,1989	-4,55	<0,0001
Mata Atlântica	-1,1902	0,1964	-6,06	<0,0001
Pampa	-1,0476	0,2268	-4,62	<0,0001

¹Instrumentos: insumos do modelo DEA, *dummies* de biomas, escore demográfico, escore de infraestrutura, índice de desempenho da educação básica, índice de desempenho do Sistema Único de Saúde, taxa de alfabetização.

Fonte: Elaborada pelos autores a partir dos dados do IBGE (2009) e STATA (2015).

Tabela 2 - Estimativa Final do Modelo de Regressão Fracionária, Brasil, Janeiro de 2017

Variável	Coefficiente	Desvio padrão <i>bootstrap</i>	z	P> z	[Intervalo de confiança 95%]
Amazônia	-0,5448	0,3188	-1,71	0,087	-1,1698 0,0801
Caatinga	-1,0885	0,3191	-3,41	0,001	-1,7138 -0,4631
Cerrados	-1,0016	0,3174	-3,16	0,002	-1,6237 -0,3794
Mata Atlântica	-1,2584	0,3165	-3,98	0,000	-1,8787 -0,6382
Pampa	-0,8939	0,3186	-2,81	0,005	-1,5183 -0,2695
Constante (Pantanal)	0,3753	0,3161	1,19	0,235	-0,2443 0,9949

Fonte: Elaborada pelos autores a partir dos dados do IBGE (2009) e STATA (2015).

nica na função de produção como componentes endógenas. O modelo normal-normal truncado não convergiu e o melhor ajuste (maior valor da função log-verossimilhança) foi da combinação normal-meia normal. Embora conjuntamente endógenas, a variável ambiental tem o sinal correto, mas não se mostrou significativa. Desse modo, essa variável foi retirada do modelo, já que a correção de viés por endogeneidade dessa variável é desprezável.

A tabela 3 apresenta os resultados obtidos com a abordagem da fronteira estocástica com hipótese de endogeneidade. As variáveis de produção (produto - renda com a atividade pecuária de corte; insumos - gastos com mão de obra, gastos com terra, gastos com insumos específicos e taxas de lotação) são medidas em

logs e representam observações de um modelo na família Cobb-Douglas para a função de produção. A função de produção assume coeficientes técnicos distintos para os biomas. A componente de ineficiência é modelada pela distribuição meia normal. O modelo de produção sem componente de ineficiência, estimado por mínimos quadrados em dois estágios (MQL2), tem $R^2=0,682$ e coeficiente de assimetria residual de $-0,318$, significativamente negativa, resultado favorável à inclusão da componente de ineficiência. O modelo de fronteira estocástica tem correlação de $0,791$ entre valores preditos e observados, e de $0,911$ com os valores preditos do modelo estimado com MQL2 que produz estimativas consistentes dos parâmetros sob endogeneidade. O intervalo de confiança para a variância da componente de ineficên-

Tabela 3 - Ajuste da Fronteira Estocástica Normal-Meia Normal com Componente Endógena¹, Brasil, Janeiro de 2017

Item	Coefficiente	Desvio padrão	z	P> z	[Intervalo de confiança 95%]	
Renda						
Mão de obra	0,0390	0,0144	2,70	0,007	0,0107	0,0673
Terra	0,3743	0,0231	16,21	0,000	0,3290	0,4195
Insumos	0,7466	0,0158	47,26	0,000	0,7156	0,7775
Lotação	0,4082	0,0225	18,11	0,000	0,3641	0,4524
Amazônia	-0,3432	0,2606	-1,32	0,188	-0,8540	0,1676
Caatinga	-0,2391	0,2630	-0,91	0,363	-0,7546	0,2765
Cerrados	-0,5068	0,2585	-1,96	0,050	-1,0135	-0,0001
Mata Atlântica	-0,5579	0,2602	-2,14	0,032	-1,0678	-0,0479
Pampa	-0,0274	0,2684	-0,10	0,919	-0,5534	0,4986
Assistência técnica	-2,7126	0,1478	-18,35	0,000	-3,0023	-2,4229
Constante	2,6960	0,3488	7,73	0,000	2,0124	3,3796
Instrumentalização de assistência técnica						
Social	0,5991	0,0313	19,16	0,000	0,5378	0,6604
Demo	-0,0110	0,0340	-0,32	0,746	-0,0775	0,0556
Educação básica	0,0019	0,0175	0,11	0,911	-0,0323	0,0362
Mão de obra	-0,0244	0,0033	-7,41	0,000	-0,0309	-0,0180
Terra	0,0632	0,0042	15,02	0,000	0,0550	0,0715
Insumos	0,0168	0,0036	4,65	0,000	0,0097	0,0239
Lotação	0,0125	0,0056	2,26	0,024	0,0017	0,0234
Amazônia	0,0526	0,0638	0,82	0,410	-0,0725	0,1776
Caatinga	0,1084	0,0641	1,69	0,091	-0,0174	0,2341
Cerrados	0,0721	0,0631	1,14	0,254	-0,0517	0,1958
Mata Atlântica	0,1035	0,0632	1,64	0,102	-0,0204	0,2273
Pampa	0,1103	0,0653	1,69	0,091	-0,0176	0,2382
Constante	-0,3448	0,0825	-4,18	0,000	-0,5066	-0,1830
ξ _Assistência técnica						
Constante	2,3425	0,1592	14,71	0,000	2,0304	2,6546
$\ln \sigma_u$						
Constante	-0,7314	0,0727	-10,06	0,000	-0,8738	-0,5889
$\ln \sigma_w$						
Constante	-1,9150	0,0704	-27,20	0,000	-2,0530	-1,7770

¹Teste de endogeneidade: Ho - correção para endogeneidade não é necessária; Ha - há endogeneidade no modelo e a correção é necessária.

ξ _Assistência técnica = 0; $\chi^2(1) = 216,44$; Prob> $\chi^2 = 0,0000$; Resultado: rejeita Ho (nível de 0.1%).

Fonte: Elaborada pelos autores a partir dos dados do IBGE (2009) e STATA (2015).

cia leva à rejeição da hipótese de $\sigma_u = 0$, evidenciando, mais uma vez, a presença da componente de ineficiência.

O teste de Hausman, realizado comparando o modelo de fronteira estocástica com componente de ineficiência exponencial sem hipótese de endogeneidade com o modelo consistente MQL2, indica significância para os dois parâmetros (das variáveis ambiental e de assistência técnica), favorecendo a hipótese de endogeneidade (p-valor < 0,0001). Esse resultado confirma-se no modelo de eficiência estocástica de Karakaplan e Kutlu (2017). A significância do teste deve-se, principalmente, à variável assistência técnica. A variável ambiental não apresentou significância na função de produção e, por isso, foi eliminada do modelo. A tabela 3 apresenta o modelo final, incluindo o teste de endogeneidade da componente assistência técnica (teste de $\xi = 0$). Nesse modelo, consideraram-se como instrumentos as variáveis exógenas da função de produção, as *dummies* de biomas, um escore de infraestrutura (social), um escore demográfico (demo) e um escore de qualidade da educação básica. Mais detalhes sobre os três últimos instrumentos podem ser vistos em Souza, Gomes e Alves (2017).

A função de produção, como esperado, tem todas as elasticidades de insumos positivas e significantes estatisticamente. Os insumos sal mineral e medicamentos dominam, seguidos pela taxa de lotação, terra e trabalho, representando 47,61% ($\pm 1,09$), 26,04% ($\pm 1,20$), 23,87% ($\pm 1,31$) e 2,49% ($\pm 0,93$) da elasticidade total, respectivamente. Insumos tecnológicos (sal mineral e medicamentos mais capital) são responsáveis, portanto, por 73,65% da elasticidade total. Esses números são mais expressivos para a pecuária de corte do que para a agricultura geral, como reportado em Souza et al. (2013). Esses autores calcularam em 68,1% a participação dos insumos tecnológicos.

O coeficiente linear do Pantanal domina na função de produção. A ordem dos biomas induzida pelos coeficientes é Pantanal, Pampa, Caatinga, Amazônia, Cerrados e Mata Atlântica. Diferenças não significantes são Amazônia e Caatinga, Amazônia e Pantanal, Caatinga e Pantanal, Pampa e Pantanal, Cerrado e Mata Atlântica. Esses testes não são transitivos. Desvios da

fronteira estocástica, associados às componentes de eficiência, não diferem significativamente. A distribuição correspondente é apresentada por bioma na figura 2. As respostas medianas são próximas para todos os biomas.

No Pantanal ocorre a concentração dos produtores na atividade de cria, havendo recria apenas das novilhas de reposição. Segundo Abreu, McManus e Santos (2010), em função das características ambientais peculiares, a região produtora do Pantanal apresenta sistema extensivo de produção, quase na totalidade baseado em pastagens nativas. Dada a economia de escala (tamanho médio da propriedade em torno de 4.000 hectares), apresenta baixo volume de custos, embora seja o bioma que concentra o maior número de matrizes de cria por área em relação aos outros biomas. Ou seja, o uso da terra no bioma é basicamente ocupado pela pecuária extensiva.

As variáveis contextuais de ambiente (não estatisticamente significante) e de assistência técnica agem no sentido de redução do valor esperado da resposta. À parte a significância estatística, esses resultados coincidem com a abordagem determinística e substanciam a hipótese de imperfeições de mercado. Como discutido em Souza, Gomes e Alves (2017), a extensão rural provavelmente não alcança a maioria dos estabelecimentos rurais devido à falta de infraestrutura apropriada e aos desequilíbrios de mercado gerados pelos preços recebidos e pagos pelos produtores, diferenciados pelo porte do produtor.

5 - CONSIDERAÇÕES FINAIS

Ajustou-se um modelo de produção para a pecuária de corte brasileira. O modelo envolve uma especificação da função de produção na família Cobb-Douglas, com insumos definidos por gastos com mão de obra, terra e insumos específicos para a pecuária. Como *proxy* do capital financeiro da pecuária, considerou-se a taxa de lotação. Alternativamente, determinou-se uma fronteira DEA com retornos variáveis na ótica do produto. As duas abordagens são marcadamente distintas. Encontraram-se evidências estatísticas significativas em

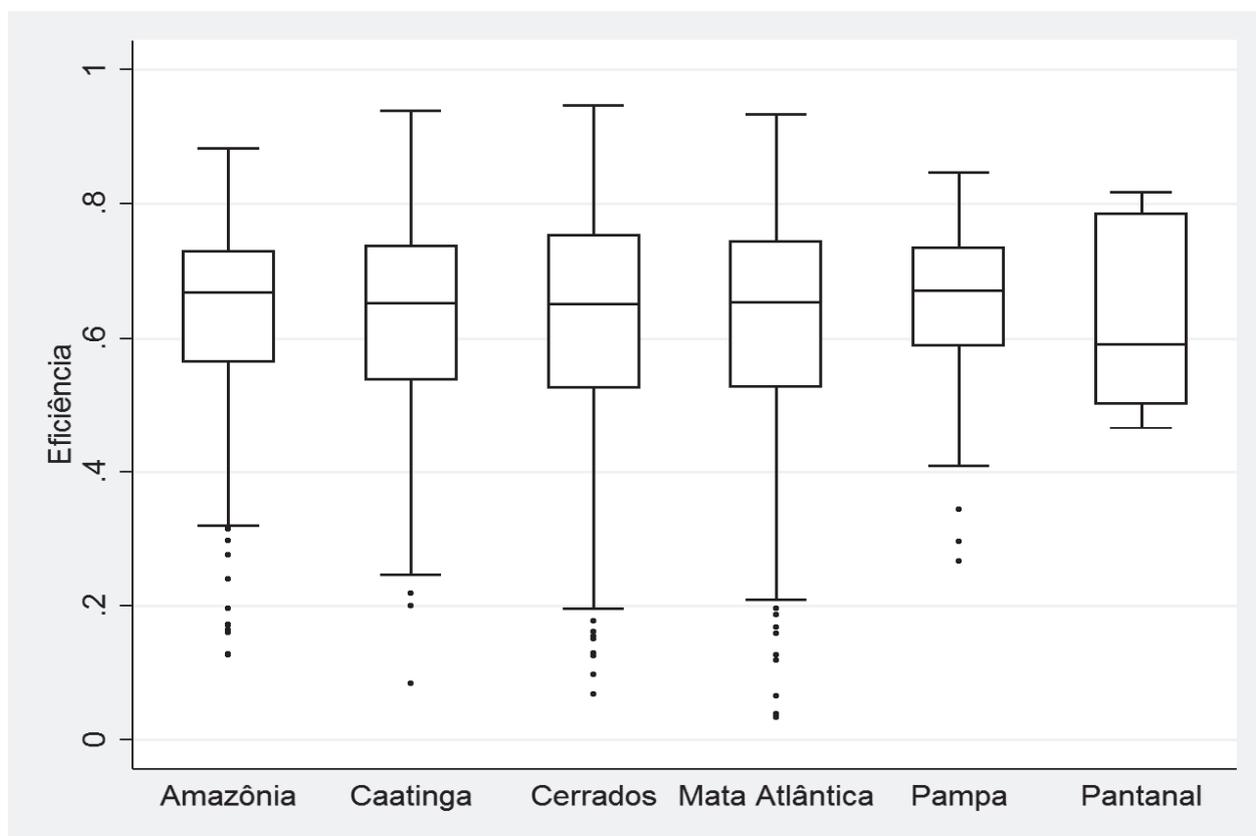


Figura 2 - Box-Plot da Medida de Eficiência Técnica Estocástica por Bioma, Brasil, Janeiro de 2017.
Fonte: Elaborada pelos autores a partir do Stata (2015).

favor de um modelo de fronteira estocástica com componentes endógenas.

Segundo o modelo de fronteira estocástica, a influência do fator ambiental, medido pela proporção de pastagens degradadas, é neutra, i.e., não há evidência de que seu coeficiente na função de produção difira de zero. Qualquer que seja a hipótese relativamente à natureza da fronteira, a componente de assistência técnica tem sinal negativo, mas só é significativo no modelo de fronteira estocástica. Esses resultados da assistência técnica são consoantes com estudos anteriores, e resultam muito provavelmente de imperfeições de mercado que impossibilitam a efetividade da extensão rural. Notam-se similaridades entre os biomas quanto à eficiência técnica no modelo de fronteira estocástica. A função de

produção do bioma Pantanal tem coeficiente técnico dominante seguido de Pampa, Caatinga, Amazônia, Cerrados e Mata Atlântica.

O resultado significativo e positivo da taxa de locação na função de produção corrobora com as estratégias de intensificação dos sistemas de pecuária de corte no Brasil. Essa estratégia é função do capital disponível para o investimento, do risco e da taxa de retorno em cada situação. O planejamento técnico, aliado ao financeiro, e o monitoramento constante, em cada bioma, são fundamentais para verificar a viabilidade operacional e econômica das estratégias assumidas dentro dos sistemas de produção e fornecer, com maior precisão, as informações necessárias para a tomada de decisão e manter a atividade competitiva no mercado (BARBOSA et al., 2012).

LITERATURA CITADA

- ABREU, U. G. P. et al. Greenhouse gas emissions intensity assessment in beef cattle production systems: a data envelopment analysis (DEA) approach with variable returns to scale. In: SECOND INTERNATIONAL SYMPOSIUM ON GREENHOUSE GASES IN AGRICULTURE - SIGEE, 2., 2016, Campo Grande. **Anais...** Brasília: Embrapa, 2016. p. 195-199.
- _____. et al. Beef cattle stocking rate, a key to the conservation of the pastoral system in Brazilian biomes. In: SIXTH INTERNATIONAL CONFERENCE ON AGRICULTURAL STATISTICAL - ICAS, 6., 2013, Rio de Janeiro. **Anais...** IBGE: Rio de Janeiro, 2013. p. 1-15.
- _____.; MCMANUS, C.; SANTOS, S. A. Cattle ranching, conservation and transhumance in the Brazilian Pantanal. **Pastoralism: research, policy and practice**, London, Vol. 1, Issue 1, pp. 99-114, Jan. 2010.
- AIGNER, D.; LOVELL, C. A. K.; SCHIMIDT, P. Formulation and estimation of stochastic frontier production function models. **Journal Of Econometrics**, Vol. 6, Issue 1, pp. 21-37, July. 1977.
- ALVES, E. R.; SOUZA, G. S. Pequenos estabelecimento em termos de área também enriquecem? Pedras e tropeços. **Revista de Política Agrícola**, Brasília, v. 24, n. 3, p. 7-21, jul./ago./set. 2015.
- _____. et al. Um modelo de produção para a agricultura brasileira e a importância da pesquisa da Embrapa. **Revista de Política Agrícola**, Brasília, v. 21, p. 35-59, out./nov./dez. 2012.
- BARBOSA, F. A. et al. Gerência e competitividade na bovinocultura de corte. In: SIMPÓSIO DE PRODUÇÃO DE GADO DE CORTE - SIMCORTE, 8., Viçosa, 2012. **Anais...** Viçosa: UFV, 2012. p. 159-182.
- BARIONI, L. G. et al. Visão sistêmica aplicada à otimização da produtividade do animal e da área. In: SIMPÓSIO DE PRODUÇÃO DE GADO DE CORTE - SIMCORTE, 6., Viçosa, 2008. **Anais...** Viçosa: UFV, 2008. p. 321-344.
- BERNDT, A.; TOMKINS, N. W. Measurement and mitigation of methane emissions from beef cattle in tropical grazing systems: a perspective from Australia and Brazil. **Animal**, Ireland, Vol. 7, pp. 363-372, June 2013.
- BRADLEY, J. V. **Distribution-Free statistical tests**. New Jersey: Prentice Hall, 1968. 388 p.
- CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA - CEPEA. PIB do agronegócio. Piracicaba: CEPEA/ESALQ/USP, 2016. Disponível em: <<http://www.cepea.esalq.usp.br/br/pib-do-agronegocio-brasileiro.aspx>>. Acesso em: 17 jan. 2017.
- COELLI, T. J. et al. **An introduction to efficiency and productivity analysis**. 2. ed. New York: Springer, 2005. 367 p.
- COOPER, W. W.; SEIFORD, L. M.; ZHU, J. (Eds.) **Handbook on data envelopment analysis**. 2. ed. New York: Springer, 2011. 493 p.
- EUCLIDES FILHO, K. Interação genótipo-ambiente-mercado na produção de carne bovina nos trópicos. In: SIMPÓSIO DE PRODUÇÃO DE GADO DE CORTE - SIMCORTE, 2., Viçosa, 2001. **Anais...** Viçosa: UFV, 2001. p. 93-115.
- GREENE, W. H. **Econometric Analysis**. 7. ed. New York: Prentice Hall, 2011. 1232 p.
- KARAKAPLAN, M. U. Fitting endogenous stochastic frontier models in Stata. **The Stata Journal**, Vol. 17, Issue 1, pp. 39-55, 2017.
- _____.; KUTLU, L. Handling endogeneity in stochastic frontier analysis. **Economics Bulletin**, Vol. 37, Issue 2, p. 889-901, May 2017. Disponível em: <<http://www.mukarakaplan.com/Karakaplan%20-%20EndoSFA.pdf>>. Acesso em: 27 jun. 2017.
- KHUMBHAKAR, S. C.; LOVELL, C. A. K. **Stochastic frontier analysis**. Cambridge: Cambridge University Press, 2000. 344 p.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Censo agropecuário 2006. Rio de Janeiro: IBGE, 2009.
- PAPKE, L. E.; WOOLDRIDGE, J. M. Econometric methods goes fractional response variables with an application to 401(k) plan participation rates. **Journal of Applied Economics**, Vol. 11, Issue 6, pp. 619-632, Nov. 1996.
- RAMALHO, E. A.; RAMALHO, J. J. S. Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses. **Econometric Reviews**, London, Vol. 36, Issue 4, p. 397-420, 2017.
- _____.; _____.; HENRIQUES, P. D. Fractional regression models for second stage DEA efficiency analyses. **Journal of Productivity Analysis**, United States, Vol. 34, pp. 239-255, Dec. 2010.
- SILVA, R. O. et al. Sustainable intensification of Brazilian livestock production through optimized pasture restoration. **Agricultural Systems**, Vol. 153, pp. 201-211, May 2017.
- SOUZA, G. S. **Funções de produção: uma abordagem estatística com o uso de modelos de encapsulamento de dados**. Brasília: Embrapa, 2003. 49 p. (Texto para Discussão n. 17).
- _____. et al. Um modelo de produção para a agricultura brasileira e a importância da pesquisa da Embrapa. In: ALVES, E. R. A.; SOUZA, G. S.; GOMES, E. G. (Eds.). **Contribuição da Embrapa para o desenvolvimento da agricultura no Brasil**. 1. ed. Brasília: Embrapa Informação Tecnológica, 2013. v. 1, p. 49-86.
- _____.; GOMES, E. G. Management of agricultural research centers in Brazil: a DEA application using a dynamic GMM approach. **European Journal of Operational Research**, Vol. 240, pp. 819-824, 2015.
- SOUZA, G. S.; GOMES, E. G.; ALVES, E. R. A. Conditional FDH

efficiency to assess performance factors for Brazilian agriculture. **Pesquisa Operacional**, Rio de Janeiro, v. 77, p. 93-106, jan./abr. 2017.

STATA. **Stata base reference manual release 14**. Frontier: stochastic frontier models. College Station: Stata Press, 2015. v. 1,

2556 p.

THORNTON, P. K. Livestock production: recent trends, future prospects. **Philosophical Transactions of the Royal Society B**, London, Vol. 365, pp. 2853-2867, Sept. 2010.

Recebido em 17/02/2017. Liberado para publicação em 08/01/2018.

REV. DE ECONOMIA AGRÍCOLA, SÃO PAULO, v. 63, n. 1, p. 21-34, jan.-jun. 2016