

ECOEFIÊNCIA DA AGROPECUÁRIA NO BIOMA AMAZÔNIA: ÍNDICES ROBUSTOS E DETERMINANTES COM O MODELO DEA ESTOCÁSTICO

1 Introdução

A agropecuária bioma Amazônia enfrenta dois grandes desafios. O primeiro enxerga a necessidade de o setor tornar-se mais competitivo, produtivo e eficiente. Isso significa ofertar mais produtos com maior qualidade e menor preço para satisfazer as crescentes necessidades humanas. O segundo desafio foca no imperativo de transformar o setor em uma atividade mais amigável com a natureza. Sugere que não se pode comprometer os serviços ambientais prestados pelas florestas nem expor a biodiversidade ao risco de degradação. Logo, deve-se reduzir a emissões de gases de efeito estufa (GEE) e a intensidade de uso dos recursos a um nível pelo menos em linha com a capacidade de resiliência do Planeta.

Para superar esses desafios, os tomadores de decisão e a sociedade como um todo precisam de convincentes respostas às seguintes questões: é possível aumentar a produção e, simultaneamente, reduzir os impactos ambientais e a utilização de insumos? Em que medida a agropecuária amazônica pode ser mais ecoeficiente? Quais são os fatores determinantes da ecoeficiência? Em torno destas questões está o problema central de pesquisa deste trabalho.

Neste contexto, o presente estudo tem como objetivo estimar escores robustos de ecoeficiência DEA com *bootstrap* em dois estágios para a agropecuária amazônica, em base municipal, considerando-se os dados do último Censo Agropecuário (2017). Para isso, utilizou-se no primeiro estágio, como na maioria dos casos estudados por Suzigan et al. (2020), os insumos e produtos clássicos do setor. Porém, aqui se adicionam como *outputs* três externalidades ambientais, sendo uma positiva e duas negativas. Neste estágio utilizou-se técnicas *bootstrap* para detectar *outliers*, testar o tipo de retornos de escala, estimar o viés e corrigir os escores. No segundo estágio, variáveis contextuais relacionadas a indicadores de desenvolvimento social, regional e econômico são usadas para explicar a ecoeficiência com o uso de modelos robustos de regressão truncada.

2 Conceitos e medida de eficiência e ecoeficiência

A análise de eficiência e ecoeficiência decorre do conceito conjunto de possibilidade de produção (CPP) – *technology set* (T). O CPP define um espaço no negativo em \mathbb{R}^{p+q} , formado pelos p insumos e os q produtos (incorporando as externalidades negativas e positivas) das n unidades produtivas integrantes do CPP. O limite superior desse espaço determina a fronteira do CPP, que envolve as coordenadas das unidades produtivas observadas (doravante denominada Decision Making Unit – DMUs). Para Farrell (1957), essa fronteira pode ser construída a partir de combinações lineares das DMUs ecoeficientes. As DMUs que operam em pontos abaixo da fronteira, interior do CPP, são apontados de ecoineficientes, já que não estão maximizando a produção nem minimizando os insumos.

Diante do exposto, surge o conceito da eficiência e ecoeficiência. Segundo Koopmans (1951), uma unidade produtiva é eficiente se, e somente se, é tecnologicamente impossível aumentar qualquer *output* (e/ou reduzir qualquer *input*) sem simultaneamente reduzir outro *output* (e/ou aumentar qualquer outro *input*). Assim, incluindo neste conceito as pressões ambientais, obtemos a definição de ecoeficiência. Esta se revela quando uma DMU obtém o maior nível de produção desejado possível com um determinado nível de insumo e impacto ambiental ou quando emprega a menor

quantidade possível de insumos e minimiza o impacto ambiental para produzir um número dado de produtos desejados.

A mensuração da eficiência e ecoeficiência resulta do cálculo da distância euclidiana que separa cada DMU da fronteira. Essa medida foi proposta por Farrell (1957) e usualmente chamada de “eficiência de Farrell”. Esta possibilita definir duas medidas radiais de eficiência técnica (θ): i) a orientada ao insumo, que foca na minimização equiproporcional dos insumos, mantendo a produção constante, θ_{IO} ; ii) a orientada ao produto que foca na maximização equiproporcional dos produtos, mantendo os insumos constantes, θ_{OO} . Esta forma de mensurar a eficiência é a inversa da eficiência radial de Shephard (1970) e, desconsiderando as possíveis folgas nos insumos e produtos, é um caso particular do conceito de Koopmans (1951).

2.1 Modelos DEA

Foi baseado no trabalho de Farrell (1957) que Charnes, Cooper e Rhodes (1978) introduziram o modelo *Data Envelopment Analysis* (DEA-CCR ou CRS) considerando a propriedade mais ampla do CPP, a tecnologia de retornos constantes de escala – CRS. Partindo de um conjunto de n observações do CPP, o modelo identifica as melhores práticas, constrói com elas uma fronteira eficiente e mede o nível de eficiência relativa de unidades que não pertencem à fronteira, identificando os *benchmarks* com os quais essas DMUs ineficientes podem ser comparadas.

O estimador DEA da ecoeficiência orientado aos *inputs* para cada DMU _{i} com $(\mathbf{x}_i, \mathbf{y}_i)$ e uma tecnologia \hat{T}_{CCR} é definido resolvendo o seguinte problema de programação linear (PPL):

$$\hat{\theta}_i(\mathbf{x}_i, \mathbf{y}_i)_{CCR-IO} = \min\{\hat{\theta}_i | \mathbf{y}_i \leq Y\lambda, \hat{\theta}_i \mathbf{x}_i \geq X\lambda, \lambda \in \mathcal{R}_+^n\} \quad (1)$$

onde Y e X representam, respectivamente, as matrizes dos *outputs* (desejáveis e indesejáveis) e *inputs* das n observações no CPP; \mathbf{x}_i e \mathbf{y}_i — os vetores de insumos e produtos, respectivamente, da DMU _{i} avaliada; $\lambda = [\lambda_1, \dots, \lambda_n]_{1 \times n}$ — o vetor que determina a combinação linear das melhores práticas na construção da fronteira. $\hat{\theta}_i$ é o índice de ecoeficiência orientado ao *input* de DMU _{i} que deve ser menor ou igual a 1. Se $\hat{\theta}_i = 1$, a DMU _{i} considera-se ecoeficiente. Caso contrário, é ecoineficiente. $\hat{\theta}_i \mathbf{x}_i$ computa a projeção de $(\mathbf{x}_i, \mathbf{y}_i)$ na fronteira, o nível mínimo de *inputs* possíveis para o nível dado de *output*.

De forma idêntica, pode-se definir um novo PPL e encontrar o estimador da ecoeficiência orientado ao *output* $\hat{\theta}_i$ para \hat{T}_{CCR} para cada DMU _{i} , como segue:

$$\hat{\theta}_i(\mathbf{x}_i, \mathbf{y}_i)_{CCR-OO} = \max\{\hat{\theta}_i | \hat{\theta}_i \mathbf{y}_i \leq Y\lambda, \mathbf{x}_i \geq X\lambda, \lambda \in \mathcal{R}_+^n\} \quad (2)$$

onde $\hat{\theta}_i \geq 1$. Se $\hat{\theta}_i = 1$, a DMU _{i} é considerada ecoeficiente e se $\hat{\theta}_i > 1$ será ecoineficiente. $\hat{\theta}_i \mathbf{y}_i$ determina a projeção de $(\mathbf{x}_i, \mathbf{y}_i)$ na fronteira ecoeficiente e indica em quanto é possível aumentar a produção com nível dado de insumo. É necessário registrar que, com a tecnologia com retornos constantes, $1/\hat{\theta}_i_{CCR-OO} = \hat{\theta}_i_{CCR-IO}$.

Banker, Charnes e Cooper (1984) estenderam o trabalho de Charnes et al. (1978) e apresentaram o segundo modelo clássico. Este novo DEA é chamado de modelo DEA-BCC ou VRS, já que considera uma tecnologia com retornos variáveis de escala (VRS). Assim, a propriedade da proporcionalidade entre *inputs* e *outputs*, que é a base do modelo DEA-CRS, é substituída pela propriedade da convexidade. A DEA-BCC permite que DMUs que operam com baixos valores de *inputs* tenham retornos crescentes de escala, e as que operam com altos valores tenham retornos decrescentes de escala. O PPL, para o estimador da ecoeficiência orientada ao *input* para \hat{T}_{BCC} , é definido para cada DMU _{i} por:

$$\hat{\theta}_i(\mathbf{x}_i, \mathbf{y}_i)_{BCC-IO} = \min\{\hat{\theta}_i | \mathbf{y}_i \leq Y\lambda, \hat{\theta}_i \mathbf{x}_i \geq X\lambda, \mathbf{1}\lambda = 1, \lambda \in \mathcal{R}_+^n\} \quad (3)$$

Em contrapartida, o PPL do modelo DEA-BCC orientado ao produto é dado por:

$$\hat{\theta}_i(x_i, y_i)_{BCC-OO} = \max\{\hat{\theta}_i \mid \hat{\theta}_i y_i \leq Y\lambda, x_i \geq X\lambda, \mathbf{1}\lambda = 1, \lambda \in \mathcal{R}_+^n\} \quad (4)$$

Assim, para uma dada unidade i , $\hat{\theta}_{BCC}$ é estimada adicionando-se, no problema de programação linear DEA-CCR, a restrição $\mathbf{1}\lambda = 1$, ou seja, $\sum_{i=1}^n \lambda_i = 1$. Isso é o que permite incorporar a propriedade de retornos variáveis e assegurar que as DMUs só sejam comparadas com DMUs de tamanho equivalente, desconsiderando a ineficiência de escala.

Neste modelo, $\hat{\theta}_{BCC-IO}$ e $\hat{\theta}_{BCC-OO}$ têm a mesma interpretação que no DEA-CCR no que se refere ao nível mínimo possível de *inputs* e ao nível máximo de *outputs* possíveis respectivamente. Porém, a orientação do modelo torna $\hat{\theta}_{BCC-IO} \neq 1/\hat{\theta}_{BCC-OO}$. Além disso, esses escores de ecoeficiência tendem a ser menos restritivos que os obtidos com DEA-CCR, ou seja, a ecoeficiência obtida com DEA-BCC será sempre superior ou igual à alcançada com DEA-CCR. Deste modo, o DEA-BCC tem um menor poder de discriminação entre as DMUs e o número de DMUs ecoeficientes geralmente é maior.

2.2 Inferência estatística dos estimadores DEA com *bootstrap*

Um dos dilemas dos modelos DEA tradicionais decorre de seu carácter determinístico e não-paramétrico, ou seja, em estimar valores pontuais de eficiência relativos sem discutir a incerteza em torno dessa estimativa. Porém, é importante para uso empírico, a complementação da análise DEA com procedimentos de inferência estatística que permitam estimar intervalos de confiança, fazer testes de hipóteses e indagar se outras variáveis explicam a eficiência. Isso possibilita detectar *outliers*, corroborar a confiabilidade e a aceitação dos modelos DEA, comparar a eficiência entre DMUs e estudar o efeito de variáveis exógenas (ambientais ou contextuais). Para superar essas limitações e obter indicadores mais robustos, foi desenvolvido a DEA estocástica (ou semi-paramétrica) alicerçada em técnicas de *bootstrap*.

A abordagem DEA estocástica parte do fato de que o cálculo da eficiência é realizado com base nos dados de uma amostra do CPP que não é exaustivamente conhecido. Essas DMUs são apenas um subconjunto de práticas, não necessariamente as melhores possíveis. Além disso, os valores dos *inputs* e *outputs* podem incluir erros de medição e choques aleatórios que a organização não pode controlar, como por exemplo a queda na produção devido a uma estiagem. Portanto, deve-se supor que tanto os *inputs* e *outputs* quanto os estimadores dos índices de eficiência são variáveis aleatórias, resultantes de um oculto processo gerador de dados (PGD) e de uma subjacente função de densidade da qual outras observações e amostras podem ser geradas. Consequentemente, assume-se a existência de uma diferença (um viés) entre a verdadeira tecnologia T (desconhecida e não observável) e a tecnologia estimada \hat{T} com base em uma amostra finita dada de DMUs.

Quando a amostra da variável aleatória é pequena e/ou se desconhece sua distribuição, o *bootstrap* é o método mais apropriado para a inferência estatística. O *bootstrap* foi introduzido por Bradley Efron (1979), mas só mais tarde foi utilizado na DEA (Simar, 1992; Simar & Wilson, 1998). É um método computacional amplamente utilizado em inferência estatística para obter estimativas com maior precisão, calcular os erros padrões, intervalos de confiança e realizar testes de hipóteses sobre parâmetros de interesse. A ideia básica do *bootstrap* é obter um “novo” conjunto de dados aleatórios, reamostrados com substituição sobre o conjunto de dados original. Este processo é repetido várias vezes para criar amostras de réplicas (pseudoamostras) do mesmo tamanho que o original. Destas amostras obtemos a distribuição das estimativas do parâmetro de interesse, que deve ser semelhante ao verdadeiro, e podemos tirar conclusões robustas sobre as estatísticas que nos interessam. Em outras palavras, a essência do *bootstrap* é gerar um mundo fictício tão semelhante ao real ou ao verdadeiro

que os resultados nele obtidos possam ser considerados semelhantes aos que seriam obtidos com dados populacionais, sem erros de medição e omissões.

Neste trabalho, o *bootstrap* será utilizado para detectar os casos atípicos, estimar intervalos de confiança da ecoeficiência e testar duas hipóteses: existência de retorno de escala e dependência entre a ecoeficiência e fatores externos.

2.3 Detecção das observações influentes nos cálculos de ecoeficiência

Para Wilson (1993), o método DEA requer um procedimento prévio de limpeza dos dados. Como qualquer desvio em relação à fronteira do CPP significa ineficiência, os valores pontuais de eficiência DEA são fortemente sensíveis tanto aos erros de medição quanto à existência de *outliers*. Os *outliers* de alto desempenho enviesam para baixo os índices de eficiência das demais DMUs. Assim, Bogetoft e Otto (2011) sugerem a utilização do método de nuvem de dados multivariados em uma primeira etapa, como uma análise exploratória de dados, antes de usar qualquer estimativa de fronteira. Porém, esse método exige a execução de várias etapas e uma capacidade computacional muito intensa que pode ser não factível para amostras muito grandes e no nível computacional disponível.

Uma forma mais eficiente para lidar com os *outliers* em grandes amostras é o método *Jackstrap*, proposto por Sousa e Stosic (2005), que combina método de reamostragem *bootstrap* e *jackknife* para a detecção de *outliers*. Esta abordagem busca reduzir estocasticamente o impacto das DMUs atípicas nas medidas de eficiência usando o conceito de alavancagem e partindo do princípio de que as DMUs mais influentes exibem alavancagem muito acima da média. A alavancagem indica o impacto da remoção de uma DMU sobre os escores de eficiência das $n-1$ DMUs restantes, ou seja, mensura o deslocamento da fronteira eficiente quando uma determinada DMU está ausente. Esse procedimento está disponível na biblioteca R em <https://CRAN.R-project.org/package=jackstrap> (K. M. Sousa, M. D. C. S. Sousa, & Monte, 2020).

2.4 O bootstrap para estimar o viés e o intervalo de confiança dos índices DEA

Após identificar e excluir os *outliers*, pode-se passar a estimativa do viés e intervalo de confiança. Simar e Wilson (1998) propuseram um tipo especial de *bootstrap* chamado de “alisado (*smooth*) homogêneo”. Nesta proposta, a partir dos índices de eficiência iniciais, estima-se novos índices *bootstrap* $\hat{\theta}_i^*$ para cada DMU_{*i*}, utilizando-se de uma estimativa de densidade por Kernel com métodos de suavização e reflexão. Esses índices são utilizados para gerar uma nova amostra S_n^* representada por $Y^* = [y_1^*, \dots, y_i^*, \dots, y_n^*]$ e $X^* = [x_1^*, \dots, x_i^*, \dots, x_n^*]$. Essa amostra é estruturada (respeitando a natureza radial da eficiência) da seguinte forma: para estimar $\hat{\theta}_i^*(x_i^*, y_i^*)_{OO}$ usa-se $y_i^* = y_i \frac{\hat{\theta}_i^*}{\hat{\theta}_i}$ e $x_i^* = x_i$; para estimar $\hat{\theta}_i^*(x_i^*, y_i^*)_{IO}$ é dada por $x_i^* = x_i \frac{\hat{\theta}_i^*}{\hat{\theta}_i}$ e $y_i^* = y_i$. Dessa nova amostra de *inputs* e *outputs* resulta um novo índice de eficiência $\hat{\theta}_i^*$. Repetindo este processo B vezes, para cada DMU teremos um grande número de medições de eficiência *bootstrap*, um conjunto $\{\hat{\theta}_i^b\}_{b=1}^B$ e uma função de densidade conjunta $\hat{f}(x_i^*, y_i^*)$ para cada DMU_{*i*}. Essa distribuição fornece uma estimativa confiável do índice de eficiência verdadeiro de DMU_{*i*}, sendo essa confiabilidade maior quando $B \rightarrow \infty$ e $n \rightarrow \infty$. Simar e Wilson (1998) recomendam que pelo menos 2000 amostras são necessárias para que os resultados assintóticos sejam considerados bons, embora números maiores representem índices de precisão melhores.

Como ressaltam os autores supracitados, quando o *bootstrap* é consistente, observa-se que a diferença entre a eficiência inicial ($\hat{\theta}_i$) e a eficiência *bootstrap* ($\hat{\theta}_i^*$) deve

ser semelhante à diferença entre a eficiência inicial e a eficiência verdadeira (θ_i) para dada DMU_i. Dessa forma, o viés para cada DMU_i pode ser expresso por:

$$\widehat{viés}_i = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_i^* - \hat{\theta}_i = \bar{\theta}_i^* - \hat{\theta}_i \quad (5)$$

onde $\bar{\theta}_i^*$ é a média das B estimativas de eficiência *bootstrap* da DMU_i. Daí, o estimador de $\hat{\theta}_i$ com correção do viés é:

$$\tilde{\theta}_i = \hat{\theta}_i - \widehat{viés}_i = \hat{\theta}_i - \bar{\theta}_i^* + \hat{\theta}_i = 2\hat{\theta}_i - \bar{\theta}_i^* \quad (6)$$

No entanto, para Simar e Wilson (1998), a correção de viés não deve ser usada, exceto quando:

$$\hat{\sigma}^2 < \frac{1}{3} [\widehat{viés}_i]^2 \quad (7)$$

onde $\hat{\sigma}^2$ é estimado com a variância dos estimadores *bootstrap* de tal forma que:

$$\hat{\sigma}^2 = \frac{1}{B} \sum_{b=1}^B \left[\hat{\theta}_i^b - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_i^b \right]^2 \quad (8)$$

Ressalta-se também que, conhecendo a distribuição empírica da eficiência *bootstrap* e do viés, é possível encontrar intervalos de confiança com um nível de significância (1- α), calculando os valores críticos a_α e b_α , de forma que:

$$Pr[-b_\alpha \leq \hat{\theta}_i^* - \hat{\theta}_i \leq -a_\alpha] \approx Pr[-b_\alpha \leq \hat{\theta}_i - \theta_i \leq -a_\alpha] \approx 1 - \alpha \quad (9)$$

Para estimar a_α e b_α , Simar e Wilson (1998, 2000) sugerem o uso do procedimento do percentil. Este procedimento baseia-se na ordenação dos vieses do maior para o menor e na eliminação, em ambas as extremidades dessa ordenação, da percentagem ($\alpha/2$) dos valores. Logo, o intervalo estimado é:

$$\hat{\theta}_i + a_\alpha \leq \theta_i \leq \hat{\theta}_i + b_\alpha \quad (10)$$

Este procedimento, formulado para a DMU_i, é o mesmo que deve ser seguido para cada uma das demais DMUs que compõem a amostra original. Pode ser operacionalizado usando-se a função *R boot.sw98* do *package FEAR* (Wilson, 2008) descrito ainda por Bogetoft e Otto (2011).

2.5 Teste do tipo de retornos de escala

A discricionariedade na escolha de um modelo (CCR ou BCC) ou a adoção de vários modelos pode não ser adequada. Impor *a priori* um suposto sobre os retornos de escala da fronteira pode distorcer seriamente as medidas de eficiência se a verdadeira tecnologia for outra (Simar & Wilson, 2002). Portanto, no desenvolvimento e na validação do método, deve-se testar se a tecnologia T, do qual nossas observações são amostradas, exhibe retornos constantes de escala ou não. Os autores acima citados formalizam o teste com um par de hipóteses: a hipótese nula (H_0): T é CCR, ou a hipótese alternativa (H_a): T é BCC.

Para fazer o teste, Simar e Wilson (2002) propõem utilizar o estimador:

$$S = \frac{\sum_{i=1}^n \hat{\theta}_{i,CCR}}{\sum_{i=1}^n \hat{\theta}_{i,BCC}} \quad (11)$$

onde $S \leq 1$, já que $\sum_{i=1}^n \hat{\theta}_{i,CCR} \geq \sum_{i=1}^n \hat{\theta}_{i,BCC}$. Se S é próximo de 1 então H_0 não deve ser rejeitada, e se S for significativamente menor que 1, H_a deve ser considerada verdadeira. Assim, devemos buscar um valor crítico c_α que determinará se devemos rejeitar H_0 quando $S < c_\alpha$ de acordo com um nível de significância. Normalmente o nível de significância (α) é de 5% e indica a probabilidade de rejeitar a hipótese nula, embora ela seja verdadeira (erro do tipo I).

Contudo, a distribuição do parâmetro populacional de S é desconhecida, logo, não é possível determinar c_α diretamente. Para estimar a distribuição desse parâmetro, Simar

e Wilson (2002) recomendam usar o método *bootstrap*. Com ajuda da função *R boot.sw98*, o algoritmo obtém o conjunto de eficiência *bootstrap* para o modelo CCR, o conjunto *bootstrap* para o modelo BCC e a distribuição *bootstrap* de S. Isso permite comparar o valor estimado de S em (11) com o valor crítico a 5% para tomar a decisão sobre H_0 .

2.6 Teste de dependência da ecoeficiência de fatores exógenos

Após testar as hipóteses para as tecnologias e estimar os intervalos de confiança da ecoeficiência, com frequência surgem as seguintes questões: por que algumas DMUs são mais ecoeficientes do que outras? A ecoeficiência está relacionada à disponibilidade de créditos e assistência técnica, à escolaridade dos gestores e servidores? À existência ou não de cooperativas e associações de produtores ou ao ambiente de negócios em diferentes regiões e / ou outros fatores?

O interesse despertado por essa questão conduziu ao surgimento do procedimento comumente denominado “análise de dois estágios” (Emrouznejad & Yang, 2018). No primeiro são calculados os índices de ecoeficiência e, em seguida, na segunda etapa, busca-se estabelecer a influência de outras variáveis (Z) na ecoeficiência das DMUs. Aqui, Lovell (1993) sugere o uso de variáveis sob o controle do gerente no primeiro estágio e variáveis sobre as quais o produtor não controla no segundo estágio. Portanto, as variáveis (Z) são chamadas de variáveis exógenas, ambientais ou não discricionárias e podem ser internas ou externas ao processo de produção. Além disso, recomenda-se que as variáveis Z não sejam correlacionadas com os *inputs* e *outputs* do primeiro estágio.

Com frequência para indagar se um vetor de Z explica a eficiência ($\tilde{\theta}$), usa-se a regressão *tobit*. Esse método econométrico, estimado por máxima verossimilhança, é utilizado quando a variável dependente está limitada entre faixas de valores. No entanto, como os índices DEA mostram autocorrelação e as variáveis Z são frequentemente correlacionadas com os resíduos, as premissas básicas do modelo são violadas e, portanto, podem-se gerar resultados enviesados.

Para superar essas desvantagens, Simar e Wilson (2007) propõem a aplicação de *bootstrap* a um modelo de regressão truncado (*tobit* ou equivalente) em que $\tilde{\theta}_i = Z_i\beta + \varepsilon_i$. Aqui, $\tilde{\theta}_i$ representa a eficiência corrigida usando-se do procedimento de *bootstrap* homogêneo a partir das estimativas no primeiro estágio; Z_i corresponde ao vetor de variáveis explicativas que se espera que influenciem os índices de eficiência por meio do vetor de parâmetros β a ser estimado. O vetor ε_i é o ruído ou erro aleatório associado a cada DMU, indicando a parte da ineficiência não explicada por Z_i . Assume-se que ε_i possui distribuição normal com média 0 e variância σ_ε^2 censurada entre 0 e 1. A regressão truncada com *bootstrap* testa de forma consistente quais são as variáveis influentes e significativas, seu sinal (positivo ou negativo) e permite ainda corrigir os índices de eficiência com base nos parâmetros estimados.

Simar e Wilson (2007) sugerem dois algoritmos que incorporam o procedimento de *bootstrap* em um modelo de regressão truncado. Para amostras com menos de 400 unidades e até três saídas e três entradas, o Algoritmo # 1 proposto se ajusta melhor do que o Algoritmo # 2, que é mais eficiente a partir de amostras que excedem 800 unidades. Como o tamanho de nossa amostra está mais próximo a 400, aplicamos o Algoritmo # 1, que é descrito por Boueri, R., Rocha, F. e Rodopoulos, F. (2015).

3 Variáveis

Conforme o Censo Agropecuário de 2017 do IBGE (2019), existem nos municípios do bioma Amazônia 677.596 estabelecimentos agropecuários, em que 90% são de pequenos agricultores. Esses estabelecimentos geram uma receita anual de 68.417

bilhões de reais e ocupam uma área de 97,655,962.00 ha. Grande parte desta área ainda é coberta por vegetação nativa dada a exigência de 80% de Reserva Legal, prevista no Código Florestal Brasileiro para esta região.

Para estimar a ecoeficiência dos 552 municípios (DMUs) do bioma, no primeiro estágio, foram usados os insumos e produto cotejados na literatura, adicionando-se como *outputs* três externalidades ambientais. Como insumos incluíram-se: x_1 – área dos estabelecimentos em hectares, x_2 – máquinas, medidas pelo *proxy* gasto anual com combustíveis e lubrificantes em mil reais, x_3 – despesa anual com insumos para produção vegetal e animal em mil reais, x_4 – mão de obra ocupada nos estabelecimentos (assalariada e familiar), x_5 – outras despesas das propriedades em mil reais; e como *output*, y_1 — receita bruta anual em 10 milhão de reais. Entre as externalidades, foram eleitas uma positiva y_2 e duas negativas y_3 e y_4 , sendo os valores destes dois últimos invertidos para que, ao maximizar o vetor produção, essas externalidades negativas sejam minimizadas. Elas são: y_2 — área de matas e florestas naturais e plantadas nos estabelecimentos em 10 mil hectares que representa um *proxy* dos serviços ambientais prestados pela flora e pela fauna nessa área, y_3 – índice de impacto da agricultura na biodiversidade, calculado pelo Índice de Shannon (SHDI), descrito por Beltrán-Estevé, Reig-Martínez e Estruch-Guitart, (2017), y_4 — emissões anuais de gases de efeito estufa pelo sector agropecuário em milhões de toneladas GWP – AR5 (potencial de aquecimento global segundo os fatores de conversão estabelecidos no 5º relatório do IPCC). Essas informações por municípios foram extraídas do Censo Agropecuário 2017, exceto y_4 que foram baixadas do site do Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG) (<https://seeg.eco.br/o-que-e-o-seeg>).

Para investigar a associação de fatores exógenos com a ecoeficiência no segundo estágio, foram selecionadas 12 variáveis. A escolha considerou as variáveis identificadas na literatura sobre o tema (Bravo-Ureta & Pinheiro, 1993) disponíveis no Censo, porém, foram incluídas outras não abarcadas na literatura, o que dá à análise realizada um caráter exploratório. As variáveis são: z_1 – percentagem (%) de propriedades classificadas como agricultura familiar nos municípios, z_2 – % de propriedades que receberam assistência técnica, z_3 — % de propriedades associadas à cooperativa, z_4 – % dos estabelecimentos em que o produtor possui pelo menos ensino médio, z_5 – % da mão de obra familiar em relação ao total, z_6 — % dos estabelecimentos em que o produtor é proprietário da terra, z_7 — % dos estabelecimentos em que o gestor é mulher, z_8 – % de propriedade que ocupam mão de obra infantil (menos de 14 anos), z_9 – % dos estabelecimentos que obtiveram financiamento, z_{10} – densidade populacional do município (hab./km²), *proxy* do tamanho do mercado, z_{11} – Latitude e z_{12} – Longitude. Estas duas últimas variáveis representam as coordenadas geográficas da sede do município e permitirão identificar o impacto da localização espacial na ecoeficiência.

4 Análises dos resultados

Na sequência reportam-se os resultados encontrados. Inicialmente analisa-se a estatística descritiva das variáveis utilizadas e a detecção de *outliers*. Em seguida, debate-se os resultados da inferência estatística da ecoeficiência: o teste de hipóteses de retorno de escala e o intervalo de confiança dos índices DEA. Por fim, indaga-se o impacto das variáveis contextuais nos índices de ecoeficiência calculados.

4.1 Análise descritiva, detecção de *outlier* e amostra final

Para ter uma visão global da amostra inicial (552 municípios), calculou-se as estatísticas descritivas dos insumos, dos produtos e das variáveis ambientais. Essa análise mostra um grande espalhamento dos dados e evidencia que, na maioria das variáveis, os

valores mínimos estão bastante abaixo do 1º quartil e os máximos estão muito acima do 3º quartil. Isso indica a existência de significativa heterogeneidade nos dados e de indícios de potenciais observações extremas (*outliers*).

A presença de *outliers* e de heterogeneidade dos dados podem explicar-se pela adoção de boas práticas, pelo grande número e pela diversidade tecnológica das propriedades rurais, mas também devido a possíveis problemas de erros e omissões, inclusive aqueles decorrentes do sistema auto declarativo do Censo. Esses problemas devem afetar os índices de ecoeficiência, subestimando consideravelmente a estimativa. Portanto, para assegurar a confiabilidade dos índices encontrados, é crucial a correção desses problemas. Somente assim poder-se-á obter escores que sigam a tendência geral e sejam úteis para a avaliação do desempenho da região.

Para contornar esses problemas, utilizou-se o método *Jackstrap* de detecção de *outliers*. A aplicação desse método selecionou bolhas que continham 10% do número total de DMUs e 2000 replicações *bootstrap* e calculou a alavancagem média de cada DMU, assim como o valor de limiar com os dois modelos básico: DEA-CRS e DEA-VRS. O método *Jackstrap* para a DEA-CRS encontrou 25 e 13 *outliers* para os critérios da função *Heaviside* e do teste *Kolmogorov-Smirnov* (K-S) respectivamente; para a DEA-VRS, detectaram-se 20 e 15 *outliers* da função *Heaviside* e do teste K-S na mesma ordem. Dessas quatro opções, decidiu-se pela mais estrita: o critério da função *Heaviside* da DEA-CRS.

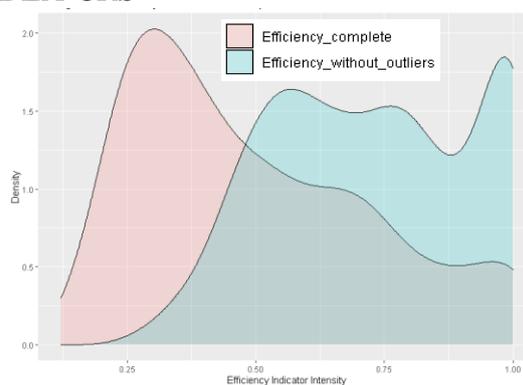
O Quadro 1 apresenta as 25 DMUs a serem eliminadas, ordenadas pelas alavancagens. Analisando-se cuidadosamente os registros do quadro, constatou que essas unidades contavam com valores baixos para determinados insumos ou elevados para certos produtos. Nota-se que algumas DMUs são capazes de acrescentar seus insumos várias vezes e seguir sendo ecoeficientes. Isso é muito improvável quando se supõe que o nível tecnológico desses municípios não está muito distante da média encontrada na região, a menos que essas diferenças representem erros nos dados. Por exemplo, temos o caso de Anajás (PA), que reportou um índice de supereficiência (Andersen & Petersen, 1993) com CRS de 5,55, que indica que pode aumentar seus recursos em mais de 400% e ainda assim permanecerem na fronteira de ecoeficiência.

Quadro 1 - As DMUs (25) consideradas *outliers*, ordenadas pelas alavancagens

Municípios
Xapuri (AC); Caapiranga (AM); Carauari (AM); Uarini (AM); Anajás (PA); Bagre (PA); Benevides (PA); Bujaru (PA); Currealinho (PA); Marituba (PA); Salinópolis (PA); Santa Izabel do Pará (PA); São Francisco do Pará (PA); Terra Alta (PA); Ulianópolis (PA); Vigia (PA); Amapá (AP); Cutias (AP); Itaúbal (AP); Tartarugalzinho (AP); Carmolândia (TO); Conquista D'Oeste (MT); Lambari D'Oeste (MT); Nortelândia (MT); Nova Olímpia (MT).

Fonte: elaborado pelos autores.

Figura 1 - Densidade das eficiências de DEA-CRS



Fonte: elaborado pelos autores.

A Figura 1 mostra as curvas de densidades da ecoeficiência obtidas por meio do modelo DEA-CRS aplicado antes e depois da retirada das 25 DMUs de maior alavancagem. Nota-se que a remoção dos *outliers* gera um grande impacto sobre a distribuição dos escores de ecoeficiência computados. A nova distribuição indica uma média e mediana maior e um deslocamento para a região de maior ecoeficiência,

indicando que os escores de eficiência originais são enviesados para baixo devido à existência de *outliers*.

Salienta-se também que o número de DMUs removidos representa menos de 4.5% da amostra original e 48% delas encontra-se no Estado do Pará. Portanto, a partir de agora se trabalhará com uma amostra de 527 DMUs.

4.2 Inferência estatística dos índices DEA

Definida a nova amostra, realizou-se o teste de significância de retorno de escala, uma vez que a escolha arbitrária da tecnologia pode resultar em conclusões inadequadas. Seguindo o método descrito, esse teste foi realizado usando a distribuição dos escores de ecoeficiência CCR e BCC obtidas do *bootstrap* com 2000 amostragens e tendo a hipótese nula (H_0): T é CCR e a hipótese alternativa (H_a): T é BCC. Devemos rejeitar H_0 se S (Equação 11) é significativamente menor que o valor crítico para $\alpha = 0.05$.

Já que o valor estimado de $S = 0.95$ é maior que o valor crítico $c_\alpha = 0.92$, não podemos rejeitar a hipótese nula. Os resultados obtidos mostram ainda um erro do tipo I igual a 0.4480 (superior à α), apontando que estaremos errando com uma probabilidade de 44.8% se rejeitarmos a hipótese nula — com base nesta estimativa. Consequentemente, podemos afirmar que a tecnologia T, do qual nossas observações são amostradas, exhibe retornos constantes de escala, ou seja, o pequeno, médio e grande município podem ser igualmente ecoeficientes. Este resultado está em dissonância com a evidência empírica obtida por Souza e Gomes (2019), que rejeita a hipótese de retornos constantes de escala na análise da eficiência técnica para 4.964 municípios brasileiros. Porém, este confirma os resultados de Freitas, Teixeira e Braga (2019), que encontraram a possibilidade de eficiência em propriedades de todos os tamanhos. Os dados usados em ambos os trabalhos são os do censo agrícola de 2006.

Aceitando a tecnologia CRS, passou-se a estimar o escore de ecoeficiência de Shaphard orientado aos *outputs* e seu intervalo de confiança, considerando o viés aleatório inerente aos dados. Usou-se o procedimento de Simar & Wilson (1998) disponível na biblioteca FEAR em: <https://pww.people.clemson.edu/Software/FEAR/fear.html>. O resultado com 2000 amostragens permitiu estabelecer uma fronteira estocástica de produção que deve ser muito próxima da real. Mas, devido ao tamanho da base de dados, são mostrados apenas os cinco municípios mais ecoeficientes e os cinco mais ecoineficientes (Tabela 1). Para comparar esses resultados, acrescentam-se na Tabela 1 os escores determinísticos estimados anteriormente.

A Tabela 1 mostra que, na região estudada, existem comportamentos ecoineficientes relevantes. É possível aumentar os produtos desejados e reduzir, simultaneamente, o dano ambiental com os mesmos insumos. Por exemplo, o município Parauapebas (PA) foi o mais bem-posicionado no *ranking* de ecoeficiência corrigida. Mesmo assim, a divisão do vetor produto pelo escalar, que representa o índice de ecoeficiência ($y/0.936$), mostra as possíveis melhorias e dará a projeção desse município na fronteira estocástica, o nível ótimo de *outputs* sem modificar seus insumos. O mesmo deve ser realizado para os outros municípios. Os valores agregados indicam que é possível aumentar, em toda a região estudada, a receita anual em 50,17%, as áreas preservadas em 54,06% e reduzir, concomitantemente, o impacto no índice de biodiversidade em 21,4% e as emissões de GEE em 37,15%.

Os resultados da Tabela 1 indicaram ainda que a média global foi de 0.615, mas as médias de ecoeficiências entre os municípios diferem entre os estados. O estado com a maior média foi o Amapá com um escore de 0.72, seguido pelo Acre (0.70), Roraima (0.699), Mato Grosso (0.697), Amazonas (0.665), Pará (0.63), Tocantins (0.566), Maranhão (0.528) e Rondônia (0.513).

Na Tabela 1 também é possível observar os intervalos de confiança. Estes indicam que múltiplos índices corrigidos podem ser estatisticamente iguais, já que, caso dois intervalos se cruzem, seria possível afirmar que as unidades não possuem diferentes níveis de ecoeficiência. Este é o caso dos cinco primeiros municípios do *ranking*. Isso indica que deve haver grande cautela ao se realizar análises comparativas entre DMUs.

Tabela 1 - Ecoeficiências CRS de Shaphard orientado ao *output*

DMU	Sem os <i>outliers</i> e com correção	Intervalo de confiança (95%)		Sem os <i>outliers</i> e sem correção	Com a amostra inicial
		Limite inferior	Limite superior		
Parauapebas (PA)	0.936	0.910	0.983	0.998	0.450
Novo Airão (AM)	0.897	0.861	0.973	0.985	0.847
Medicilândia (PA)	0.891	0.869	0.938	0.953	0.506
Presidente Figueiredo (AM)	0.885	0.856	0.971	0.977	0.852
Rio Preto da Eva (AM)	0.880	0.858	0.985	1.000	0.849
.
.
.
Santa L. do Paruá (MA)	0.274	0.259	0.304	0.308	0.178
Garrafão do Norte (PA)	0.268	0.255	0.330	0.334	0.215
Anajatuba (MA)	0.267	0.261	0.299	0.303	0.121
Governador Nunes F. (MA)	0.255	0.244	0.288	0.291	0.128
Itapecuru Mirim (MA)	0.247	0.238	0.275	0.278	0.137
Média	0.615	0.601	0.720	0.730	0.519

Fonte: elaborada pelos autores.

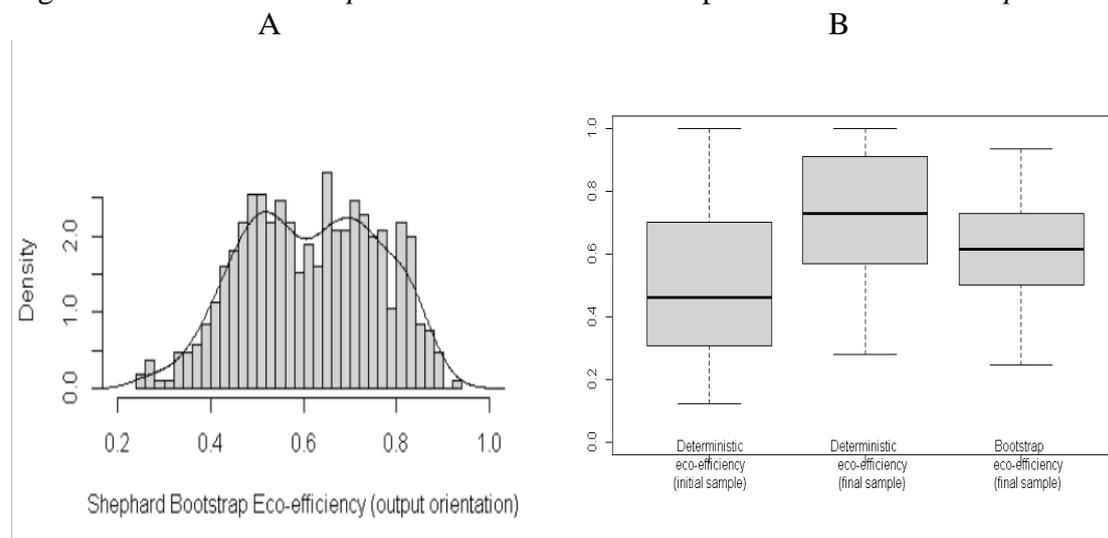
Além disso, nota-se que os índices de ecoeficiência corrigidos pelo viés em geral são superiores aos índices da amostra original e inferiores aos índices com a amostra final. Isso sugere que os escores determinísticos da amostra final devem estar superestimados e os da amostra inicial subestimados.

Esse fato é ainda corroborado pela Figura 2. A Figura 2A exibe a curva de densidade dos índices de ecoeficiência corrigidos, indicando que é muito diferente das curvas dos escores de ecoeficiência determinísticos antes e depois de retirar os *outliers* representadas na Figura 1. Portanto, as probabilidades de cada faixa de valores de ecoeficiência diferem em cada modelo. A Figura 2B exibe os diagramas de caixas das três distribuições indicando que as medianas, médias e as dispersões dos índices são diferentes. A média e mediana dos índices corrigidos situou-se entre a média e mediana das outras distribuições, tendo sua amplitude interquartílica menor. Isso mostra que os índices gerados pelo procedimento de *bootstrap*, com seus respectivos intervalos de confiança, são mais eficientes e robustos.

Este fato sugeriu testes de comparação de distribuições. Os testes de ANOVA e de Kruskal-Wallis (com p-value=2e-16 e p-value = 2.2e-16 respectivamente) indicaram que existem diferencia significativas entre as distribuições. Os vários testes Post-hot confirmam também que há diferença significativas entre os grupos de distribuição par a par. Esta constatação evidencia a relevância de se realizar a análise de *bootstrap* ao se estimar os índices de eficiências de DMUs.

Contudo, os índices de ecoeficiência obtidos nesta etapa só podem se considerar convincentes se todos os municípios estiverem operando no mesmo contexto. Isso não é evidente quando observamos os valores das variáveis contextuais. Logo, é necessário incorporar essas variáveis na análise da ecoeficiência.

Figura 2. Densidade e *boxplots* das eficiências de Shephard orientado aos *outputs*



Fonte: elaborada pelos autores.

4.3 Impacto das variáveis contextuais nos índices de ecoeficiência

A influência dos fatores ambientais nos índices de ecoeficiência CRS orientado aos *outputs* corrigidos foi estimada seguindo o algoritmo de Simar e Wilson (2007) de segundo estágio usando a biblioteca do R FEAR. A Tabela 2 apresenta os resultados da regressão *tobit* com 2000 reamostragens *bootstrap*. Como os coeficientes (betas) estimados das variáveis não têm uma interpretação direta igual aos modelos lineares, incluímos os efeitos marginais médios.

Tabela 2 - Resultados das estimativas do modelo aplicado na segunda etapa

Variáveis ambientais	Coeficiente Beta	Intervalo de confiança a 95%		Efeito Marginal Médio
		Limite inferior	Limite superior	
Intercepto	2.2958	2.2871	2.3045	
z1 - agricultura familiar	0.2168	0.2081	0.2254	0.216756
z2 - assistência técnica	-0.4731	-0.4807	-0.4656	-0.4731345
z3 - cooperativa	-0.0957	-0.1079	-0.0835	-0.09570354
z4 - ensino médio	-0.6816	-0.6912	-0.6720	-0.6815853
z5 - mão de obra familiar	-0.1277	-0.1333	-0.1221	-0.1276654
z6 - proprietário da terra	-0.0390	-0.0426	-0.0354	-0.03900976
z7 - mulher	-0.2225	-0.2337	-0.2112	-0.222456
z8 - mão de obra infantil	0.0188	0.0037	0.0338	0.01877187
z9 - financiamento	1.1642	1.1536	1.1749	1.164225
z10 - densidade	0.0001	0.0001	0.0001	0.00005767993
z11 - Latitude	0.0019	0.0017	0.0021	0.001917629
z12 - Longitude	0.0078	0.0077	0.0079	0.007816003
Desvio do erro	0.3966	0.3962	0.3971	

Fonte: elaborada pelos autores.

De acordo com esses resultados, todas as variáveis contextuais são diferentes de zero e significativas a 5%. Porém, quatro delas, tendo valores muito próximos a zero, afetam irrelevantemente a ecoineficiência. Trata-se das variáveis associadas a: z₈ — mão de obra infantil, z₁₀ — densidade populacional do município, *proxy* do tamanho do mercado, z₁₁ — Latitude e z₁₂ — Longitude. A baixa significância destas duas últimas variáveis indica, inesperadamente, que a localização espacial tem baixa importância na ecoeficiência.

Das restantes, chama atenção a variável z9 – financiamento. Com o maior nível de influência, essa variável apresentou um sinal inesperado, segundo a teoria, e pode indicar uma séria falha de mercado. Os efeitos marginais médios da Tabela 2 expõem que um incremento de 1% no número dos estabelecimentos com financiamento aumenta o valor predito da ecoineficiência dos municípios em 1.164225 pontos. Já na década de 1980, Taylor, Drummond, & Gomes (1986), estudando agricultores de Minas Gerais, encontraram que o financiamento teve um efeito ligeiramente negativo na eficiência alocativa e na eficiência técnica. Souza & Gomes (2019) também obtiveram um coeficiente beta imprevisto para essa variável, seguindo o algoritmo #1 e #2 de Simar e Wilson (2007) na análise da eficiência técnica de 4.964 municípios brasileiros com os dados do censo agrícola de 2006. Certamente isso pode ser explicado pelas políticas de crédito subsidiado, que estimula a compra de insumos e máquinas com preço ineficiente, diminuindo a eficiência alocativa (Fernandes & Pascual, 2015).

A segunda variável de maior impacto é o nível educacional dos gestores. Os efeitos marginais médio da Tabela 2 mostram que o aumento em 1% no número de estabelecimentos em que o gestor possui pelo menos o ensino médio deve reduzir o índice de ecoineficiência em 0.6816 pontos. Ademais, indica que os gestores rurais com melhor formação podem ter conhecimentos que lhes permitam atribuir mais importância à questão econômica e ambiental, bem como na gestão das atividades e na escolha das novas tecnologias. Isso corrobora os resultados de outras investigações (Gasson, 1998; Ondersteijn, Giesen, & Huirne, 2006). Para Fernandes & Pascual (2015), agricultores, como quaisquer outros gestores, tomam melhores decisões se estão melhor preparados.

Outro fator relevante para explicar as diferenças observadas nos índices de ecoeficiência é a assistência técnica. A Tabela 2 indica que o efeito marginal médio é de -0.4731345. Tal fato pode ser explicado pelas evidências de que, ao receberem essa assistência, os produtores tendem a corrigir problemas de uso inadequado de insumos e também aqueles relacionados a o que, quanto e como produzir com menor impacto ambiental. Isto é corroborado no trabalho de Gomes et al. (2018) que constata a existência de relações positivas entre o tempo em que o produtor recebe a assistência técnica e a eficiência.

Vale salientar ainda que, apesar das mulheres serem minoria na gestão das propriedades rurais, a presença delas na administração tem um efeito positivo e significativo na ecoeficiência, como observado na Tabela 2. O Censo Agropecuário de 2017 identificou que o número de mulheres responsáveis pela gestão de propriedades rurais no país cresceu e chegou ao milhão de um universo de 5.07 milhões de gestores. Isso, sem dúvida, é um importante indicador na equidade de gênero em um meio ainda muito conservador. É um progresso (ainda não suficiente) na reparação histórica da discriminação contra a mulher a fim de estabelecer a igualdade com base no reconhecimento das características próprias de cada gênero e das capacidades das mulheres serem tão competentes quanto os homens.

No que se refere ao impacto da agricultura familiar, nota-se, na Tabela 2, que um acréscimo de 1% no número de propriedades assim enquadrada aumenta a ecoineficiência em 0.216756 pontos. Seguramente que esse fato está relacionado ao baixo nível de escolaridade dos produtores dessa classe, o que impede a incorporação de novas tecnologias, a alocação eficiente dos fatores de produção postos a sua disposição e a determinação do mix ótimo de produção. Esse resultado é consistente com os achados em parte da literatura empírica pertinente. Porém, não contradiz o dito anteriormente sobre a possibilidade da pequena propriedade ser tão competitiva como a grande, já que como mostra Del Grossi, Florido, Rodrigues e Oliveira, (2019) a agricultura familiar é pequena, mas a pequena propriedade não é necessariamente familiar. Assim, as políticas

públicas precisam reunir em seu arsenal iniciativas que deem à agricultura familiar chances de competirem em igualdade com as grandes propriedades, principalmente quanto aos preços que pagam pelos insumos e aos preços de venda dos produtos (Souza et al. 2018).

Os resultados da regressão indicaram também que a utilização de mão de obra familiar é outro fator importante na elevação da ecoeficiência. Como mostram os dados do Censo, a proporção de familiares em relação ao total de mão de obra utilizada na propriedade é alta, o que está relacionado à elevada frequência da agricultura familiar. Porém, o impacto na ecoeficiência dessas duas variáveis é divergente, o que exige um estudo mais detalhado.

De forma similar, como mostrado na Tabela 2, é significativa a relação entre a ecoeficiência e o associativismo a cooperativas, que é muito baixo na região. Isso abre grande possibilidade para a melhoria da ecoeficiência, já que o associacionismo está relacionado ao melhor acesso às informações, sendo útil para comparar os resultados e reproduzir as boas práticas sustentáveis. O associacionismo permite ao agricultor potencializar os serviços de extensão e assistência técnica, elevar seu poder de barganha na comercialização dos produtos e aquisições dos insumos e ameniza a indivisibilidade dos meios de produção de alto custo (por exemplo as colheitadeiras e os silos).

Por fim, verifica-se que a propriedade da terra é outro fator importante na redução da ecoineficiência apesar do valor do impacto ser pequeno (-0.039). Esse resultado é importante diante da recorrente apropriação ilegal de terras públicas na região, denominada grilagem. Nestas áreas irregulares, o desenvolvimento de atividades agropastoris ou florestais viola normas ambientais, agrárias, civis e tributárias, bem como cria incertezas sobre a situação fundiária, causando conflitos sociais (principalmente com comunidades indígenas) e dificultando a implantação de projetos de conservação e de desenvolvimento econômico na região (Araújo, Benatti, & Pena, 2006).

Considerações finais

Dos resultados, derivam-se resumidamente as seguintes inferências:

1) A pesquisa mostra que a curva de densidade dos índices de ecoeficiência corrigidos é estatisticamente diferente das curvas dos escores de ecoeficiência determinísticos antes e depois de retirar os *outliers*, indicando que estes escores devem estar enviesados. Ademais, advertem que múltiplos índices corrigidos podem ser estatisticamente iguais e que é possível afirmar que DMUs não possuem diferentes níveis de ecoeficiência. Essas constatações recomendam que deve haver grande cautela ao se realizar análises comparativas entre DMUs e evidenciam a relevância de se realizar a análise de *bootstrap* e aferir intervalos de confiança ao se estimar os índices de eficiências de DMUs;

2) O teste de significância de retorno de escala indica que a tecnologia da qual nossas observações são amostradas exhibe retornos constantes, indicando que o pequeno, médio ou grande municípios podem ser igualmente ecoeficientes. Esse resultado aflora o antigo debate sobre a sustentabilidade dos diversos tamanhos de propriedade a longo prazo. Se consideramos o município como uma unidade produtiva, nossa pesquisa corrobora as teses de que o tamanho não é condição para que uma empresa seja competitiva, alinhado com a intuição de Caves & Porter (1977) e Porter (1979) que afirmam que os pequenos podem ser competitivos se encontrarem eficientemente seu nicho de mercado e aproveitarem sua flexibilidade e a capacidade para inovar; por exemplo, na agricultura agroecológica e no abastecimento dos mercados locais. Este fato é de fundamental importância para demonstrar aos tomadores de decisão e à opinião

pública que na definição das políticas públicas deve ser dado peso igual tanto para a pequena agricultura quanto para a grande;

3) Foi observado que, na região estudada, existem comportamentos ecoineficientes acentuados. É possível, a nível agregado, aumentar em toda a região estudada a receita anual em 50,17%, as áreas preservadas em 54,06% e reduzir, concomitantemente, o impacto no índice de biodiversidade em 21,4% e as emissões de GEE em 37,15% com os mesmos insumos. Assim, acreditamos que esses resultados fornecem importantes contribuições práticas para subsidiar a harmonização do crescimento econômico e da preservação ambiental, amparando a sustentabilidade da região;

4) Os resultados da regressão indicaram que os escores de ecoeficiência são também afetados significativamente por fatores exógenos. Esse impacto pode ser classificado como de baixa ou alta importância. As de baixa importância apresentaram impactos com valores próximos a zero, sendo a mão de obra infantil, a densidade populacional do município e a localização geográfica. As de alta importância foram a agricultura familiar, a assistência técnica, a associação à cooperativa, a escolaridade, a mão de obra familiar, a propriedade da terra, as mulheres gestoras e o financiamento. No entanto, esta última variável apresentou um sinal inesperado segundo a teoria e pode indicar uma séria falha na política pública para o setor.

Para finalizar, cabe frisar que esta análise deve ser complementada pelo exame da dinâmica dos indicadores de ecoeficiência no tempo, identificando as mudanças tecnológicas e de eficiência do setor. Portanto, a extensão natural da investigação seria incluir dados de painel na análise. Isto será objeto de pesquisas futuras, já que é outra lacuna encontrada na literatura.

REFERÊNCIAS

- Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science*, 39(10), 1261-1264.
- Alves, E. (2004). Retornos à escala e mercado competitivo: teoria e evidências empíricas. *Revista de Economia e Agronegócio*, 2(3).
- Araújo, R., Benatti, J. E., & Pena, S. (2006). *Grilagem de terras públicas na Amazônia Brasileira*. Brasília: MMA.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30,1078–1092.
- Beltrán-Esteve, M., Reig-Martínez, E., & Estruch-Guitart, V. (2017). Assessing eco-efficiency: A metafrontier directional distance function approach using life cycle analysis. *Environmental Impact Assessment Review*, 63, 116–127.
- Bogetoft, P., & Otto, L. (2010). *Benchmarking with DEA, SFA, and R* (vol. 157). New York, USA: Springer Science & Business Media.
- Boueri, R., Rocha, F., & Rodopoulos, F. (Eds.). (2015). Avaliação da qualidade do gasto público e mensuração da eficiência. Ministério da Fazenda, Secretaria do Tesouro Nacional.
- Bravo-Ureta B. E., & Pinheiro A. E. (1993). Efficiency Analysis of Developing Country Agriculture: A Review of Frontier Function. *Agricultural and Resource Economics Review*, 22, 88-101.

- Caves, R. E., & Porter, M. E. (1977). From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition. *The quarterly journal of economics*, 241-261.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444.
- Del Grossi, M., Florido, A. C. S., Rodrigues, L. F. P., & Oliveira, M. S. (2019). Comunicação de Pesquisa: Delimitando a Agricultura Familiar nos Censos Agropecuários Brasileiros. *Revista NECAT-Revista do Núcleo de Estudos de Economia Catarinense*, 8(16), 40-45.
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist*, 7(1), 1-26.
- Emrouznejad, A., & Yang, G. L. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-economic planning sciences*, 61, 4-8.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society: Series A (General)*, 120(3), 253-281.
- Fernandes de Oliveira, L. A., & Pascual, U. (2015). Análise da eficiência da agricultura familiar agroecologista. *Revibec: revista de la Red Iberoamericana de Economía Ecológica*, 24, 0221-233.
- Freitas, C. O., Teixeira, E. C., Braga, M. J., & de Souza Schuntzemberger, A. M. (2019). Technical efficiency and farm size: an analysis based on the Brazilian agriculture and livestock census. *Italian Review of Agricultural Economics*, 74(1), 33-48. doi: <https://doi.org/10.13128/REA-25478>.
- Gasson R. (1998). Educational qualifications of UK farmers: A review. *J Rural Stud*. 14, 487-498.
- Gomes, A. P., Ervilha, G. T., Freitas, L. F. D., & Nascif, C. (2018). Assistência técnica, eficiência e rentabilidade na produção de leite. *Revista de Política Agrícola*, 27(2), 79.
- Instituto Brasileiro de Geografia e Estatística. (2019). *Censo Agropecuário 2017: Resultados definitivos*. Rio de Janeiro: IBGE.
- Koopmans, T. C. (1951). An analysis of production as an efficient combination of activities. In: T.C. Koopmans (Ed.). *Activity Analysis of Production and Allocation, Cowles Commission for Research in Economics*. New York: Mongraph 13.
- Lovell, C. K. (1993). Production frontiers and productive efficiency. *The measurement of productive efficiency: techniques and applications*, 3, 67.
- Ondersteijn C. J. M., Giesen G. W. J., & Huirne, R. B. M. (2006). Perceived environmental uncertainty in Dutch dairy farming: The effect of external farm context on strategic choice. *Agr. Syst*, 88, 205-226.
- Porter, M. E. (1979). The structure within industries and companies' performance. *The review of economics and statistics*, 214-227.
- Shephard, R. W. (1970). *Theory of cost and production function*. Princeton, NJ: Princeton University.
- Simar, L. (1992). Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric Non-Parametric and Semi-Parametric Methods with Bootstrapping, *Journal of Productivity Analysis*, 3, 167-203.
- Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management science*, 44(1), 49-61.

- Simar, L., & Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: the state of the art. *Journal of Productivity Analysis*, 13(1), 49-78.
- Simar, L., & Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139(1), 115-132.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, 136(1), 31-64.
- Sousa, K. M., Sousa, M. D. C. S., & Monte, P. A. D. (2020). *Package jackstrap*: <https://CRAN.R-project.org/package=jackstrap>.
- Sousa, M. D. C. S., & Stošić, B. (2005). Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. *Journal of Productivity analysis*, 24(2), 157-181.
- Souza, G. S., Gomes, E. G., & Alves, E. R. A. (2018). Imperfeições de mercado e concentração de renda na produção agrícola. *Revista de Política Agrícola*, XXVII (2), 31-38.
- Sousa, M. D. C. S. D., & Souza, J. C. F. (2014). Escores robustos de eficiência e seus determinantes: o caso das agências do banco do Brasil. *Pesquisa e Planejamento Econômico*, 44(1), 34.
- Souza, G. S., & Gomes, E. G. (2019). Métodos alternativos para análise de regressão em dois estágios com resposta DEA: uma aplicação para a agricultura brasileira. *Simpósio de Pesquisa Operacional e Logística da Marinha*, 19 (pp. 1644-1657), Rio de Janeiro.
- Suzigan, L. H., Peña, C. R., & Guarnieri, P. (2020). Eco-efficiency Assessment in Agriculture: A Literature Review Focused on Methods and Indicators. *J. Agric. Sci*, 12, 118.
- Taylor, T. G., Drummond, H. E., & Gomes, A. T. (1986). Agricultural credit programs and production efficiency: an analysis of traditional farming in southeastern Minas Gerais, Brazil. *American Journal of Agricultural Economics*, 68(1), 110-119.
- Wilson, P. W. (1993). Detecting outliers in deterministic nonparametric frontier models with multiple outputs. *Journal of Business and Economics Statistics*, 11(3), 319-323. doi: 10.2307/1391956
- Wilson, P. W. (2008). FEAR 1.0: A Software Package for Frontier Efficiency Analysis with R. *Socio-Economic Planning Sciences*, 42, 247-254.