

Determinantes da Eficiência em Empresas de Comércio: uma abordagem bayesiana

March 16, 2024

Abstract

Este artigo busca investigar quais fatores operacionais e financeiros são mais determinantes para explicar o nível de eficiência das companhias brasileiras que atuam nos setores de “Comércio” e “Comércio e Distribuição”. Para isso, faz uso de um modelo de fronteira estocástica para estimar os termos de eficiência, que são posteriormente analisados através de um modelo de árvores de regressão bayesianas aditivas (*bayesian additive regression trees*, ou BART na sigla em inglês). Parte dos resultados converge para o sugerido pela literatura - corroborando com a hipótese de que empresas maiores são mais ineficientes. Por outro lado, identifica-se que empresas mais alavancadas são menos eficientes. Também é notado um efeito da pandemia da COVID-19 sobre a performance das companhias, bem como a ausência de evidências que sugiram relação entre governança corporativa e eficiência. Por fim, ressalta-se o amplo potencial de futuras pesquisas sobre o tema, utilizando metodologias não-paramétricas como BART.

1 Introdução

Desde a eclosão da pandemia da COVID-19, diversas atividades econômicas passaram a enfrentar novos desafios em todo o mundo. O cenário crítico que emergiu com a crise sanitária - incluindo aumento do desemprego, das taxas de juros, contração da demanda internacional por *commodities* e colapso das cadeias de suprimentos afetou, em particular, as economias emergentes (Hevia, Neumeyer, et al. (2020)), que precisaram simultaneamente enfrentar os efeitos humanitários, financeiros e econômicos de uma crise sem precedentes. No Brasil, este fenômeno não foi diferente.

Em particular, companhias que atuam em atividades relacionadas ao comércio atravessaram um contexto ainda mais inóspito, aliando-se à crise global uma série de mudanças estruturais do setor, com a forte penetração de *players* estrangeiros no *e-commerce* e a tendência de maior integração entre os diversos canais de vendas *online* e *offline* (Costa, Sâmia da Silva Fôro, and Lima Vieira (2020); Cruz (2021); Delardas et al. (2022)). Destacam-se também as fortes mudanças no padrão de consumo, trazendo novas características como a redução do número de visitas a estabelecimentos físicos de comércio, o aumento da demanda por alimentos e forte predileção pelos canais digitais de venda (Gupta and Mukherjee (2022)). Apesar de relacionadas ao período pandêmico, muitas dessas mudanças notadamente tornaram-se um novo padrão, com efeitos perceptíveis de médio e longo prazo.

Simultaneamente, indicadores financeiros das principais companhias brasileiras passaram e causar preocupação no mercado. Já em 2023, após o pedido de Recuperação Judicial de uma das empresas mais tradicionais do varejo nacional (americanas s.a., fruto da fusão entre o conglomerado de *e-commerce* B2W Digital e a tradicional Lojas Americanas S.A.), elevou-se ainda mais o grau de incerteza e insegurança de investidores, bancos e consumidores. Dado esse cenário, mostra-se relevante analisar a performance dessas companhias em termos de eficiência técnica - em outras palavras, o quão bem essas empresas estão utilizando os fatores de produção que demandam.

Para isso, uma das metodologias econométricas mais relevantes envolve os modelos de Análise de Fronteira Estocástica (SFA, do inglês *Stochastic Frontier Analysis*), cujos primeiros desenvolvimentos remontam à década de 1970. Essencialmente, essa abordagem se propõe a estimar uma fronteira de produção que representa, a partir dos dados disponíveis e da definição *a priori* de uma função de produção, o maior nível de produto possível para cada combinação de fatores/insumos. A diferença entre os valores estimados da fronteira e os valores observados é, então, explicada por dois componentes estocásticos: um erro aleatório ξ e uma medida de eficiência τ .

Entretanto, são recorrentes na literatura as discussões a respeito de como analisar os fatores que explicam os níveis de eficiência. Tim Coelli, Perelman, and Romano (1999) apresentam duas abordagens:

a primeira inclui no próprio modelo de fronteira estocástica os componentes exógenos que ajudariam a explicar os desvios de eficiência - ou seja - deslocam a tecnologia da firma e geram medidas de eficiência "líquidas", controladas pelas idiosincrasias de cada observação; a segunda sugere uma análise em duas etapas, estimando um novo modelo econométrico que busca relacionar a medida de eficiência "bruta" e um vetor de co-variadas que ajudariam a explicá-la.

Apesar de, tipicamente, ambas as abordagens sugerirem resultados (e interpretações) convergentes, seus significados são distintos. No presente estudo, a segunda estratégia será aplicada, fazendo uso da metodologia de *bayesian additive regression trees* (BART) - ou árvores de regressão aditivas bayesianas.

Esta ferramenta se caracteriza numa soma de árvores de regressão cujos hiper-parâmetros são amostras de funções densidade de probabilidade definidas *a priori* e que induzem ao encolhimento das árvores. Com isso, a estratégia se baseia em construir árvores que sejam *weak learners*, ou seja, que pouco expliquem da variável dependente, evitando assim o excesso de ajuste aos dados de treino (*overfitting*).

Dessa maneira, esta metodologia permite enfrentar alguns problemas primordiais que surgem neste tipo de modelagem, não se restringindo a uma forma funcional pré-definida, além de não ser penalizada pelas interações entre as co-variadas. Ademais, fornece métodos intuitivos para mensurar a importância de cada variável explicativa dentro do modelo - aspecto particularmente útil para o tópico que será investigado neste trabalho.

Desse modo, este artigo se propõe a investigar quais indicadores financeiros e operacionais mais ajudam a explicar os níveis de eficiência das companhias brasileiras listadas na B3 nos setores de "Comércio" e "Comércio e Distribuição". Metodologicamente, se divide em duas etapas: primeiramente, é realizada a estimação da Fronteira Estocástica e dos respectivos termos de eficiência; a partir desses resultados, o modelo BART é empregado. Na seção seguinte são apresentados maiores detalhes metodológicos, incluindo uma apresentação da base de dados utilizada. Posteriormente, seguem os resultados estimados, que são discutidos e desdobrados na seção final.

2 Metodologia

Nesta seção serão detalhados os procedimentos metodológicos de ambas as etapas da pesquisa. Inicialmente, é apresentado o modelo de fronteira estocástica, estimado pela abordagem frequentista tradicional. Posteriormente, o modelo BART é introduzido, explicitando o funcionamento do algoritmo e como foi implementado. Por fim, traz-se uma breve descrição das bases de dados utilizadas.

2.1 Análise de Fronteira Estocástica

Os modelos de Análise de Fronteira Estocástica (SFA, do inglês *Stochastic Frontier Analysis*) foram inicialmente desenvolvidos por Aigner, Lovell, and Schmidt (1977), numa tentativa de encurtar as distâncias entre a teoria econômica convencional e os trabalhos econométricos.

Assuma a seguinte função de produção

$$Y_{it} = f(\mathbf{X}_{it}; \beta) \xi_{it} \tau_{it} \quad (1)$$

onde Y_{it} representa a produção obtida através da tecnologia descrita em $f(\cdot)$, do vetor de insumos X_{it} e do vetor de parâmetros desconhecidos β para a firma i , no período t . Além disso, os termos estocásticos ξ_{it} , que representa um choque aleatório, e $0 < \tau_{it} < 1$, que representa uma medida de eficiência técnica da firma no período, também afetam o produto.

Assumindo uma função de produção do tipo Cobb-Douglas, a partir da Equação 1 pode-se obter

$$y_{it} = \sum_{j=1}^K (x_{itj} \cdot \beta_j) + v_{it} - u_{it} \quad (2)$$

onde $y_{it} = \ln Y_{it}$; $x_{itj} = \ln X_{itj}$; $v_{it} = \ln \xi_{it}$; $u_{it} = \ln \tau_{it}$; e $j = 1, \dots, K$ denota o número de insumos utilizados no processo produtivo. Neste estudo, será adotada a formulação mais recorrente na literatura, onde $v_{it} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ e $u_{it} \stackrel{iid}{\sim} N^+(\mu, \sigma_u^2)$ e, como especificado por Battese and T.J. Coelli (1992), o estimador para a medida de eficiência técnica da firma i no período t é obtido por:

$$E\{\exp(u_{it}) | \epsilon_{it}\} = \left[\frac{1 - \phi\{\eta_{it}\tilde{\sigma}_i - (\tilde{\mu}_i/\tilde{\sigma}_i)\}}{1 - \phi(-\tilde{\mu}_i/\tilde{\sigma}_i)} \right] \exp\left(\eta_{it}\tilde{\mu}_i + \frac{1}{2}\eta_{it}^2\tilde{\sigma}_i^2\right) \quad (3)$$

onde

$$\tilde{\mu}_i = \frac{\mu\sigma_v^2 - \sum_{t=1}^{T_i} \eta_{it}\epsilon_{it}\sigma_u^2}{\sigma_v^2 + \sum_{t=1}^{T_i} \eta_{it}^2\sigma_u^2} \quad (4)$$

$$\tilde{\sigma}_i^2 = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sum_{t=1}^{T_i} \eta_{it}^2\sigma_u^2} \quad (5)$$

$$\epsilon_{it} = y_{it} - \sum_{j=1}^K (x_{itj} \cdot \beta_j) \quad (6)$$

Além disso,

$$\eta_{it} = \exp[-\eta(t - T)] \quad (7)$$

onde η é um parâmetro desconhecido e $t \in \mathcal{T}_i$, o conjunto de T_i períodos para os quais existem observações da firma i . Ou seja, o modelo possibilita que o termo de eficiência varie ao longo do tempo.

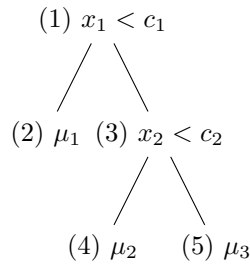
Com isso, o modelo pode ser estimado por máxima verossimilhança a partir de uma base de dados em painel - balanceado ou desbalanceado. Para modelos que não variam no tempo, tem-se que $\eta = 0$.

2.2 Bayesian Additive Regression Trees (BART)

O modelo de árvores de regressão aditivas bayesianas (BART) foi introduzido por Chipman, George, and McCulloch (2010), propondo uma nova abordagem para lidar com problemas típicos de modelos de regressão e classificação baseados em árvores de decisão.

Os primeiros modelos desse tipo foram desenvolvidos por Breiman et al. (1984), propondo uma abordagem não-paramétrica para problemas de regressão e classificação. A ideia central reside em segmentar a amostra a partir dos valores observados de \mathbf{X}_i - o vetor de variáveis explicativas. Para cada observação dentro destes segmentos, é estimado um mesmo valor para a variável dependente Y_i (tipicamente é utilizada a média, mas algumas abordagens sugerem utilizar a mediana ou mesmo realizar um ajuste linear). Essa segmentação é definida com base em *splits*, decisões binárias tomadas ao se comparar o valor observado de x_{ik} (observação da variável k para o indivíduo i) com uma regra de corte como, por exemplo, a média \bar{x}_k .

Figure 1: Exemplo de uma árvore de regressão



A Figura 1 ilustra o comportamento decisório para estimação de uma árvore de regressão. O primeiro *split* é realizado com base na co-variada x_1 , comparando os valores observados em cada observação x_{1i} com uma constante c_1 . Se $x_{1i} < c_1$, estima-se que a variável dependente seja μ_1 (usualmente, a média da variável dependente na partição onde atende-se ao critério). Caso contrário, o modelo analisa a co-variada x_2 , de modo análogo ao primeiro.

Entretanto, a literatura aponta um problema recorrente neste tipo de modelagem: há uma forte tendência ao excesso de ajuste do modelo aos dados da amostra de treino, efeito chamado de *overfitting*. Com isso, os estimadores perdem sua capacidade de generalização, tornando-se “viciados” à forma com que as variáveis interagem na amostra.

Com isso, algumas estratégias foram desenvolvidas para a mitigação deste comportamento indesejável. Algoritmos como *boosting*, florestas aleatórias (*random forests*) e *bagging* (junção dos termos *bootstrap* e *adding*) se baseiam na premissa de que a concatenação de diversas árvores pequenas e pouco explicativas (*weak learners*) é mais eficiente do que a adoção de árvores individuais e grandes. Enquanto os dois últimos geram um grande número de árvores aleatórias e tomam a média das estimativas,

o primeiro realiza a soma de uma sequência de árvores, onde cada árvore é ajustada para explicar o resíduo deixado pelas árvores anteriores.

Modelos BART seguem premissa similar (daí o termo *additive*). Entretanto, como ferramenta para garantir que cada uma das m árvores seja uma *weak learner*, faz uso de hiper-parâmetros não-determinísticos. Ou seja, define-se que estes serão tratados como variáveis aleatórias, que seguem funções densidade de probabilidade definidas *a priori* e que tendem a gerar árvores pequenas. Nos termos dos autores, “[...] BART pode ser visualizado como uma abordagem Bayesiana não-paramétrica que ajusta um modelo rico em parâmetros utilizando uma distribuição *a priori* de forte influência.” (Chipman, George, and McCulloch 2010).

Formalizando os conceitos apresentados, considere que há interesse em estimar a variável aleatória Y dado uma matriz de co-variadas x , tal que

$$Y = f(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \quad (8)$$

Deseja-se aproximar $f(x) = E(Y|x)$ utilizando um modelo de soma de árvores $h(x) = \sum_{j=1}^m g_j(x)$ onde cada g_j representa uma árvore.

Seja T uma árvore binária constituída por um conjunto de nós de decisão, com regras de *split* (por exemplo, os nós 1 e 2 da Figura 1) e b nós terminais, com estimativas para Y (nós 2, 4 e 5 da Figura 1). Além disso, $M = \{\mu_1, \mu_2, \dots, \mu_b\}$ representa o conjunto de parâmetros associado com cada nó terminal. Por “binário”, entende-se que, nesta árvore, cada nó de decisão origina apenas dois novos nós, com um critério de *split* no formato $x_k \leq c_k$ vs $x_k > c_k$ para variáveis contínuas. Para variáveis *dummy*, $c = 0$.

Nessa estrutura, cada observação i da matriz de co-variadas x está associada com um, e apenas um nó terminal de T , a partir da cadeia de decisões binárias que os originam. Assim, $g(x; T; M)$ denota a função que associa cada observação de co-variadas x_i a um $\mu_i \in M$. Desse modo, constrói-se o modelo de soma de árvores

$$Y = \sum_{j=1}^m g(x; T_j; M_j) + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \quad (9)$$

onde $E(Y|x)$ corresponde à soma de cada nó terminal μ .

Finalmente, resta especificar a *priori* de encolhimento que permite ao modelo gerar árvores *weak learners*. Para fins de simplificação, os autores concentram atenção apenas aos parâmetros $(T_1, M_1) \dots (T_m, M_m)$ e σ que são independentes entre si.

Para $p(T_j)$, define-se que a probabilidade de um nó de profundidade $d = 0, 1, 2, \dots$ não ser terminal (i.e., a probabilidade de realizar o *split* num nó de profundidade d) é dada por

$$\alpha(1 + d)^{-\beta}, \quad \alpha \in (0, 1), \beta \in [0, \inf) \quad (10)$$

Ou seja, por α e β é possível limitar probabilisticamente o tamanho das árvores de modo que, a cada novo nível de profundidade, é menor a probabilidade de ser realizado um novo *split*. É recorrente na literatura o uso de validação cruzada para testar diferentes configurações para esses hiper-parâmetros.

Além disso, é utilizada uma distribuição Uniforme para selecionar qual co-variada será utilizada em um nó, assim como para definir o critério de seleção adotado em cada nó.

Com relação a $p(\mu_i|T_j)$, é utilizada a distribuição Normal conjugada $N(\mu_\mu, \sigma_\mu^2)$, recorrente em diversas formulações da abordagem bayesiana graças aos ganhos computacionais obtidos com sua utilização. Os autores argumentam que, em aplicações empíricas do modelo, é esperada alta probabilidade de os valores *a priori* de μ_{ij} estarem localizados no intervalo entre os pontos mínimo e máximo da variável dependente, y_{min} e y_{max} . Logo, é coerente optar por especificações de μ_μ e σ_μ^2 que concentrem a massa de probabilidade de $p(\mu_i|T_j)$ dentro desse intervalo. Os autores fornecem essas especificações, mas ressalta-se que, em termos de aplicabilidade, basta centralizar Y entre -0.5 e 0.5 , adotando a versão transformada de Y como a variável dependente do modelo e definir um valor de σ_μ^2 que respeite a identidade $h\sqrt{m}\sigma_\mu = 0.5$ para um dado valor da constante h . Ou seja,

$$\mu_i|T_j \sim N(0, \sigma_\mu^2) \quad \sigma_\mu^2 = \frac{0.5}{h\sqrt{m}} \quad (11)$$

Vale ressaltar que esta definição de $p(\mu_i|T_j)$ induz o modelo a gerar árvores pouco explicativas ao encolher os parâmetros μ_{ij} a zero e, conseqüentemente, reduzindo o peso da árvore T_j em $E(Y|x)$.

Por fim, $p(\sigma)$ é definida como uma distribuição Qui-Quadrado Invertida conjugada, tal que

$$\sigma^2 \sim \frac{\nu\lambda}{\chi_\nu^2} \quad (12)$$

Novamente, os hiper-parâmetros podem ser escolhidos a partir de um procedimento de validação cruzada. Porém, são mapeados pelos autores combinações que geram estimativas mais ou menos conservadoras.

Por fim, o número de árvores que compõem o modelo, m , também deve ser definido previamente. Assim como os demais, este hiper-parâmetro que pode ser ajustado por validação cruzada, mas os autores sugerem adotar $m = 200$ e, posteriormente, testar se pequenas variações neste valor geram alterações substanciais na performance do modelo. Via de regra, valores acima de 50 são suficientes para gerar métricas de performance satisfatórias.

Os procedimentos de amostragem e estimação por Cadeias de Markov Monte-Carlo (MCMC) seguem o procedimento detalhado por Chipman, George, and McCulloch (2010) e Kapelner and Bleich (2016).

Já a importância de cada co-variada é mensurada através da proporção em que aparecem nos *splits* das m árvores que compõem o modelo, calculando-se a média entre as iterações por MCMC. Como ressaltado pelo autores, essa maneira intuitiva de identificar as variáveis com maior poder explicativo decorre da “competição” travada entre elas durante o ajuste das árvores e do processo iterativo de poda e cultivo. Além disso, para variáveis selecionadas, são construídos gráficos de dependência parcial (*partial dependence plots*), ferramenta sugerida por Friedman (2001) que consiste em marginalizar o efeito da co-variada x calculando o impacto médio que há sobre as estimativas de Y quando forçam-se todas as observações de x a um dos valores (ou intervalo de valores) em seu domínio.

2.3 Base de dados e variáveis utilizadas

No presente estudo, são analisadas as companhias brasileiras listadas em bolsa classificadas pela B3 nos setores de “Comércio” e “Comércio e Distribuição”. A ampla maioria dos indicadores utilizados, tanto para o modelo de fronteira estocástica (SFA), quanto para a análise que utiliza o modelo BART, foram obtidos de fonte secundária - a plataforma *Economática*. Esta ferramenta consolida, organiza e disponibiliza centenas de indicadores relacionados a ativos financeiros negociados no Brasil e demais países da América Latina e EUA.

Para deflacionar os indicadores mensurados em valores monetários, foi utilizado o Índice de Preços ao Consumidor Amplo (IPCA), medida oficial da inflação brasileira. Os dados relacionados a comércio exterior foram extraídos via API do Instituto de Pesquisa Econômica Aplicada (IPEA).

O modelo de fronteira estocástica considerou observações trimestrais, entre 2011Q1 e 2023Q3 e inclui as seguintes variáveis *proxy* para os fatores de produção: o Ativo Imobilizado, que é utilizado para aproximar o montante de capital empregado, mensura todos os bens tangíveis de propriedade da companhia que são utilizados para geração de renda; os gastos com Obrigações Sociais e Trabalhistas aproximam o uso de fator trabalho, mensurando dispêndios como salários, benefícios e direitos trabalhistas. A variável de *output* é a Receita. Estatística descritiva para o logaritmo natural das variáveis do modelo são apresentadas na Tabela 1.

Table 1: Estatística Descritiva das Variáveis - modelo SFA

Variável	Média	Desvio-Padrão	Máximo	Mínimo	Observações
ln_receita	14.7	1.6	18.5	10.8	879
ln_obrig_soc_trab	11.4	1.6	14.4	6.4	879
ln_ativo_imobilizado	13.5	1.8	17.3	9.4	879

O segundo modelo, que busca identificar os determinantes da eficiência técnica das companhias, possui co-variadas organizadas em 4 blocos. Estatística descritiva é apresentada na Tabela 2.

O primeiro bloco considera informações cadastrais das companhias na B3: “segmento_b3” identifica o segmento de atuação - “Eletrodomésticos”, “Material de Transporte”, “Medicamentos e outros produtos”, “Produtos diversos”, “Tecidos, vestuário e calçados” e “Alimentos”. Foram geradas variáveis *dummy* para as 5 primeiras categorias; “subsetor_b3” gerou uma variável *dummy* igual a 1 caso a companhia esteja listada no sub-setor “Comércio e Distribuição” e igual a 0 caso esteja enquadrada como “Comércio”.

Em seguida, constam os indicadores financeiros selecionadas para caracterizar a estrutura de capital, *market share* e performance no mercado financeiro: “ec_div_ativo” traz a proporção entre Dívida Bruta e Ativo Total, um indicador tipicamente utilizado para avaliar o nível de alavancagem das firmas; “ec_ebitda_despfin” traz a proporção entre EBITDA e as despesas financeiras, identificando a capacidade da firma em arcar com seus débitos; “ms_valor_mercado” mensura o valor de mercado da companhia; “mk_market_share” mensura a proporção exercida pelas receitas da companhia dentro do total de receitas do seu respectivo sub-setor.

Neste bloco também está incluído o Índice de Negociabilidade, identificado pela variável “mk_indnegoc”. Esse indicador busca mensurar a relevância da companhia nas transações realizadas na B3 e é calculado como

$$IN = \frac{1}{P} \cdot \sum_{i=1}^P \sqrt[3]{\frac{n_{ai}}{N_i} \cdot \left(\frac{v_{ai}}{V_i}\right)^2} \quad (13)$$

onde P corresponde ao número de pregões do período, n_{ai} e v_{ai} correspondem ao montante de negociações e ao valor negociado do ativo a no pregão i , respectivamente; N_i e V_i correspondem ao total de negociações e valores negociados no pregão i .

Além disso, a Taxa Interna de Retorno (TIR) do ativo também é considerada, através da variável “mk_tir”.

O terceiro bloco traz indicadores operacionais das companhias: “ope_pm_estoque” mensura o Prazo Médio de Estoque, que indica o período médio que os itens permanecem armazenados - valores mais baixos apontam para um giro mais ágil, eficiente e com menores custos de *handling*; “ope_pm_fornec” traz o Prazo Médio de Pagamento aos Fornecedores, que engloba o período entre a compra de novas mercadorias e a efetuação do pagamento; “ope_pm_receb”, por sua vez, mensura o tempo médio entre a efetuação de uma venda e o recebimento dos valores. Concatenando essas informações, o ciclo financeiro também é incorporado no modelo, pela variável “ope_cf” e mensura o período entre a venda de bens ou serviços e o pagamento dos respectivos fornecedores - para atividades de varejo, este indicador é particularmente relevante, uma vez que ciclos financeiros mais ágeis propiciam fluxos de caixa mais saudáveis. Por fim, o ciclo operacional (“ope_co”) também é considerando, mensurando todo o período entre a aquisição de bens ou serviços de fornecedores e o recebimento das vendas.

Table 2: Estatística Descritiva das Variáveis - modelo BART

Variável	Média	Desvio-Padrão	Máximo	Mínimo	Observações
vl_import	5065.1	583.8	6162.3	3915.6	339
ec_div_ativo	25.6	20.4	100.8	0	339
ec_ebitda_despfin	6.2	11.4	102.9	-20.5	339
ms_valor_mercado	7370486.5	8952243.4	47864350.4	60243	339
ms_market_share	0.1	0.1	0.7	0	339
mk_indnegoc	0.2	0.3	1.5	0	339
mk_tir	11.6	50.1	233.2	-88.8	339
ope_pm_estoque	81.1	48.9	415.7	0	339
ope_pm_fornec	65.4	37.3	242.9	1.1	339
ope_pm_receb	72.7	46	325.6	0	339
ope_cf	88.4	57.9	284.8	-47.2	339
ope_co	153.8	76.9	469.8	20.9	339
ano	2017.6	3.7	2023	2011	339
segmento_eletrdomesticos	0.1	0.2	1	0	339
segmento_material_de_transporte	0.2	0.4	1	0	339
segmento_medicamentos_e_outros	0.2	0.4	1	0	339
segmento_produtos_diversos	0	0.2	1	0	339
segmento_tecidos_vest_e_calçados	0.5	0.5	1	0	339
d_comercio_dist	0.2	0.4	1	0	339
d_novo_mercado	0.7	0.4	1	0	339
eficiencia	0.3	0.1	0.9	0.1	339

Por fim, o quarto bloco inclui o valor das importações brasileiras de bens de consumo semi-duráveis e não-duráveis voltados ao consumidor final, através da variável “vl_import”. Argumenta-se que a concorrência com plataformas estrangeiras de *e-commerce*, cuja participação no mercado brasileiro ainda é pouco explorada, tenha trazido novas dinâmicas ao varejo nacional e impactado a performance das companhias. Além disso, também foi gerada uma variável *dummy* para identificar as observações que compreendem o período após a eclosão da pandemia da COVID-19, a partir do primeiro trimestre de 2020 - “d_pandemia”.

A base de dados cobre o mesmo período da anterior, entre 2011Q1 e 2023Q3. Entretanto, a disponibilidade de indicadores operacionais e financeiros para as companhias analisadas é escassa para períodos mais antigos, prévios a 2016. Desse modo, o número de observações cai consideravelmente.

3 Resultados

Nesta seção serão apresentados os resultados de ambos os modelos.

3.1 Análise de Fronteira Estocástica

Os resultados da análise de fronteira estocástica são apresentados na Tabela 3. Ambos os insumos tiveram coeficientes estatisticamente significantes a 99%. Pela estruturação do modelo, sabe-se que os coeficientes estimados são idênticos às elasticidades de cada insumo em relação ao *output*. Logo, mostra-se pertinente avaliar como o setor combina os recursos utilizados. Para tal, foi realizado um teste de Wald χ^2 para avaliar se a soma dos coeficientes é igual a 1, o que caracterizaria uma tecnologia com retornos constantes de escala. A hipótese nula ($ln_ativo_imobilizado + ln_obrig_soc_trab = 1$) foi rejeitada e os resultados apontam para um tecnologia intensiva em fator trabalho.

Table 3: Resultados do Modelo de Fronteira Estocástica

	Coef	StdError	z	P z	95% ConfInterval
ln_ativo_imobilizado*	0.1956	0.0346	5.64	0.000	0.1276 0.2635
ln_obrig_soc_trab*	0.6207	0.0539	11.52	0.000	0.5151 0.7263
const.*	5.9889	0.5813	10.30	0.000	4.8495 7.1281
μ^*	0.9963	0.2975	3.35	0.001	0.4132 1.5795
η^*	-0.0052	0.0017	-3.03	0.002	-0.0086 -0.0018
σ_u^2	0.3965	.2146			0.0535 0.7395
σ_v^2	0.3492	.0076			0.3159 0.3825

* estatisticamente significativa a 99%

Ambos os parâmetros relacionados à eficiência das firmas, μ e η , foram estatisticamente significantes. Apesar de relativamente próximo de zero, $\eta < 0$ sugere uma tendência de decréscimo da performance das firmas ao longo do tempo. Além disso, a proporção $\sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$, tipicamente denotada por γ , foi de aproximadamente 53,17%. Ou seja, da variância total do termo de erro $\epsilon_{it} = v_{it} + u_{it}$, pouco mais da metade é explicada pela variância do termo de eficiência técnica.

A partir do estimador definido na Equação 3 foram estimados os termos de eficiência técnica de cada firma, ao longo do período observado. A Tabela 4 sintetiza esses resultados.

Chama atenção a elevada presença de companhias dos segmentos de Alimentos e Eletrodomésticos no topo do *ranking* - grupos essencialmente caracterizados por empresas que atuam nos ramos de supermercados (varejo, atacado ou ainda no modelo de negócio recentemente apelidado de “atacarejo”) e lojas de departamentos, respectivamente. Na metade inferior do *ranking* predominam as empresas classificadas nos ramos têxtil e de diversidades. As empresas que atuam no ramo de medicamentos aparecem distribuídas ao longo de toda a classificação - embora concentradas num nível intermediário de eficiência técnica.

As evidências do modelo SFA sugerem alguns novos questionamentos com relação a eficiência das companhias brasileiras de comércio e distribuição: o valor estimado negativo para o parâmetro η está, de alguma forma, relacionado à pandemia da COVID-19? A aparente segregação dos níveis de eficiência

Table 4: Ranking de Eficiência

Companhia	Média Efic. Téc.	Segmento B3
Allied	0.8523	Eletrodomésticos
Carrefour BR	0.8289	Alimentos
Viveo	0.7308	Medicamentos e outros produtos
P.Acucar-Cbd	0.7147	Alimentos
Assai	0.6993	Alimentos
Magaz Luiza	0.6523	Eletrodomésticos
Casas Bahia	0.6479	Eletrodomésticos
Grupo Mateus	0.6382	Alimentos
Pague Menos	0.6117	Medicamentos e outros produtos
RaiaDrogasil	0.5449	Medicamentos e outros produtos
Minasmaquina	0.4992	Material de transporte
Dimed	0.4511	Medicamentos e outros produtos
Wlm Ind Com	0.3694	Material de transporte
Lojas Marisa	0.3304	Tecidos vestuário e calçados
Arezzo Co	0.3222	Tecidos vestuário e calçados
Emmpar S/A	0.3217	Material de transporte
Lojas Renner	0.2919	Tecidos vestuário e calçados
Grupo Sbf	0.2877	Produtos diversos
Grazziotin	0.2855	Tecidos vestuário e calçados
Cea Modas	0.2708	Tecidos vestuário e calçados
Petz	0.2596	Produtos diversos
Blau	0.2555	Medicamentos e outros produtos
Grupo Soma	0.2542	Tecidos vestuário e calçados
Quero-Quero	0.2489	Produtos diversos
Guararapes	0.2401	Tecidos vestuário e calçados
Hypera	0.2229	Medicamentos e outros produtos
Le Biscuit	0.2137	Produtos diversos
Espacolaser	0.1528	Produtos diversos
Veste	0.1499	Tecidos vestuário e calçados

entre os segmentos de atuação das companhias é, de fato, relevante? Ou suas idiossincrasias financeiras e operacionais explicam melhor o comportamento observado? Espera-se que os resultados do modelo BART sejam capazes de trazer luz a essas questões.

3.2 Determinantes da eficiência: modelo BART

O modelo BART foi ajustado utilizando as parametrizações *standard* sugeridas pelos autores: foram estimadas 50 árvores, em um total de 1250 iterações de MCMC (250 de *burn-in*); para as distribuições a priori, adotou-se $\alpha = 0.95$, $\beta = 2$, $h = 2$ e $\nu = 3$; a variável dependente (as medidas de eficiência técnica estimadas através da Equação 3) foi centralizada entre -0.5 e 0.5. Em relação às métricas de erro, foram gerados resultados com Média dos Erros Percentuais Absolutos (MAPE) igual a 1.15% e Raiz do Erro Quadrático Médio (RMSE) igual a 0.008 (ambos *in sample*, não foi realizada a divisão da amostra entre dados de treino e teste).

A média da proporção de inclusão das co-variadas entre os *splits* do modelo é apresentada na Tabela 5. Ressalta-se o protagonismo das variáveis “ano” e “d_pandemia” - ambas figuram entre as 5 mais relevantes. A variável *proxy* relacionada ao *market share* também mostrou-se de significativa relevância, bem como algumas das variáveis *dummies* geradas a partir dos segmentos de atuação das companhias.

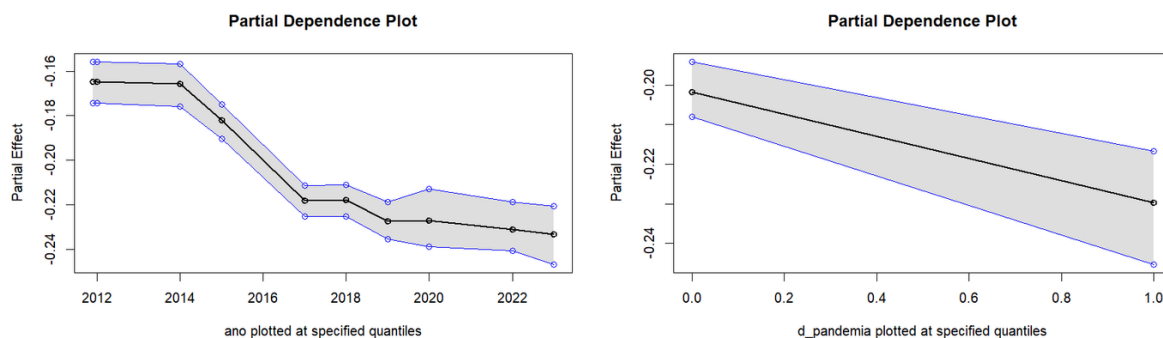
Dentre os indicadores financeiros selecionados para o modelo, a proporção entre Dívidas e Ativos, utilizada para mensurar o nível de alavancagem das firmas, esteve entre as mais relevantes para determinar a eficiência técnica. Curiosamente, a proporção EBITDA/Despesas Financeiras - relacionada à capacidade de arcar com as dívidas - esteve entre as menos relevantes. Valor de mercado e o Índice de Negociabilidade surgem na metade superior deste *ranking* - ratificando a hipótese de que o tamanho das firmas está intrinsecamente relacionado ao nível de eficiência técnica.

Table 5: Proporção de Inclusão das Co-Variadas do Modelo BART

Variável	Média % de Inclus.	Desvio-Padrão
ano	0.1175	0.0206
ms_market_share	0.0757	0.0217
segmento__eletrodomesticos	0.0687	0.0117
d_pandemia	0.0659	0.0188
ec_div__ativo	0.0629	0.0169
segmento__medicamentos_e_outros_produtos	0.0596	0.0167
segmento__tecidos_vestuario_e_calçados	0.0552	0.0242
ms_valor_mercado	0.0545	0.0317
d_comercio_dist	0.0533	0.0241
mk_indnegoc	0.0523	0.0218
ope_pm_estoque	0.0418	0.0206
ope_co	0.0391	0.0106
ope_cf	0.039	0.0155
segmento__produtos_diversos	0.0367	0.0152
ope_pm_fornec	0.0338	0.0178
segmento__material_de_transporte	0.0331	0.0243
ope_pm_receb	0.0255	0.0179
d_novo_mercado	0.0249	0.0098
ec_ebitda__despfn	0.0235	0.0158
vl_import	0.0223	0.012
mk_tir	0.0147	0.0119

Para variáveis selecionadas, foram construídas visualizações da dependência parcial em relação à variável endógena, permitindo visualizar os efeitos marginais gerados. As linhas pretas representam a média, enquanto a área hachurada entre as linhas azuis demarca os valores dentro do intervalo de credibilidade de 95%.

Figure 2: Dependências Parciais: ano e pandemia

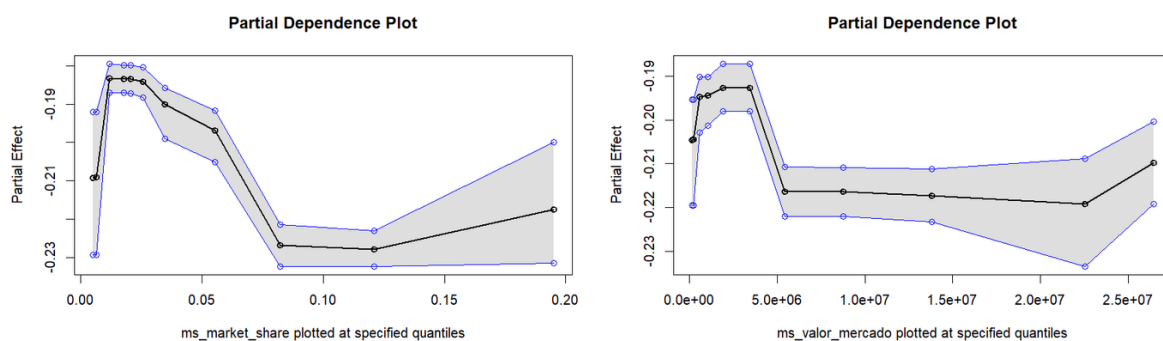


Em consonância com os resultados do modelo SFA, os efeitos marginais da variável “ano” são negativos e decrescentes. O período mais acentuado de queda, porém, está localizado entre 2014 e 2016, e não no período da pandemia da COVID-19 - embora a variável *dummy* relacionada à pandemia tenha sido relevante e marginalmente negativa.

Deve-se ressaltar, todavia, a sensibilidade da função de dependência parcial à quantidade de dados observados em cada quantil de x . Como este estudo utiliza uma base de dados em painel desbalanceado, este é um indicador que tende a sofrer maior peso conforme a quantidade de observações aumenta ao longo do tempo, dado a maior disponibilidade de dados nos anos mais recentes. Por isso a opção de avaliar o efeito da pandemia através de uma variável isolada - “d_pandemia”. Esta co-variada teve efeito marginal negativo, ratificando a hipótese de que a pandemia da COVID-19 impactou (e segue impactando) negativamente a performance operacional das firmas.

Alinhado com evidências da literatura, identifica-se uma relação inversamente proporcional entre o

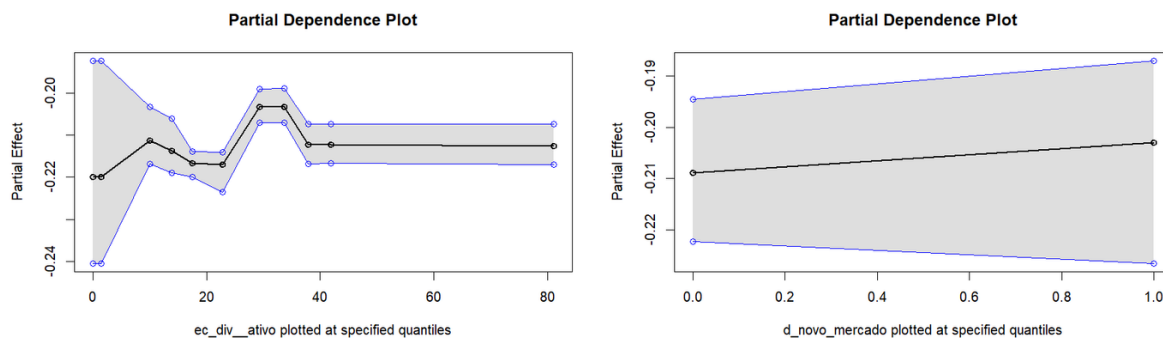
Figure 3: Dependências Parciais: *market share* e valor de mercado



tamanho das firmas e os níveis de eficiência. M. Angeles Diaz (2008) apontam resultados similares ao analisar o mercado espanhol, ressaltando a hipótese de *market selection*: para empresas relativamente menores, torna-se imperativo a adoção de modelos mais eficientes, uma vez que precisam enfrentar empresas maiores e com elevado poder de mercado. Além disso, questões organizacionais, logísticas e gerenciais podem gerar gargalos e propiciar ambientes mais ineficientes em empresas de grande porte. Fenn et al. (2008) alcançam conclusões semelhantes.

Para ambas as variáveis, nota-se também um comportamento de estabilização dos efeitos marginais a partir de um determinado ponto.

Figure 4: Dependências Parciais: alavancagem e listagem na B3



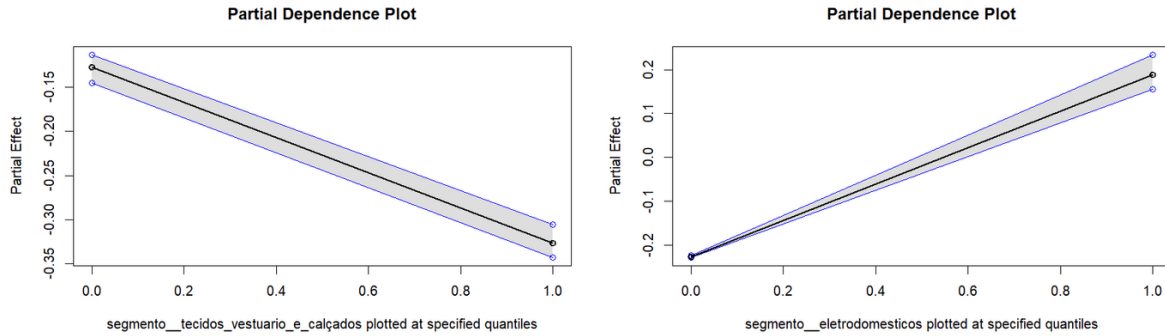
Adicionalmente, é identificado um efeito marginal negativo e estável do nível de alavancagem das firmas sobre a eficiência. No campo teórico, a relação de causa e efeito entre estes dois indicadores não é consenso, havendo hipóteses que sugerem ambas as direções. Em trabalhos empíricos como o de Margaritis and Psillaki (2007), também são identificadas evidências que corroboram ambas as possibilidades, mas os resultados sugerem uma relação diretamente proporcional: empresas mais alavancadas mostraram-se mais eficientes.

Sob a ótica na qual a ineficiência é causadora da alavancagem, pode-se argumentar que essas empresas necessitam de liquidez para viabilizar suas operações e, por isso, aumentam o endividamento. Seria sintoma desse cenário a existência de ciclos financeiros maiores, que causam fragilidade nos fluxos de caixa. Entretanto, o teste de correlação de Spearman entre as variáveis “ope_cf” (Ciclo Financeiro) e “ec_div_ativo” (alavancagem) gerou resultados fracos e estatisticamente insignificantes (coeficiente igual a 0.027 e p-valor igual a 0.62).

Em relação a governança corporativa, a variável “d_novo_mercado” figurou entre as menos relevantes no modelo (vide Tabela 5) e o efeito marginal estimado ao alterar os valores de 0 para 1 foi desprezível. Ou seja, os resultados sugerem que a adoção de boas práticas de governança, critério exigido pela B3 para incluir as companhias no mais célebre segmento de listagem - o Novo Mercado - não gera efeito sobre os níveis de eficiência.

Por fim, ratificando a tendência observada na Tabela 4, identificam-se efeitos marginais negativos para

Figure 5: Dependências Parciais: segmentos



companhias que atuam no segmento têxtil, enquanto firmas classificadas nos segmento de “Eletrodomésticos” de fato tendem a possuir níveis mais elevados de eficiência técnica.

Deve-se ressaltar que a visualização das dependências parciais se baseia na hipótese consideravelmente restritiva de que não há interação entre as co-variadas - premissa que não é adotada na estimação do modelo BART. De fato, há fortes motivos para crer que, pelo contrário, existam profundas interações entre as variáveis explicativas, justamente um dos motivos que induzem a adoção de modelos de estimação não-paramétricos. Além disso, a já mencionada sensibilidade das dependências parciais em relação à distribuição dos dados observados também surge como um potencial problema para a adoção da metodologia.

Entretanto, a consonância observada entre a proporção de inclusão das co-variadas e os efeitos marginais é fator que também deve ser levado em consideração, ratificando a relevância dos resultados e seus desdobramentos.

4 Considerações Finais

Este trabalho buscou analisar os determinantes da eficiência técnica das companhias brasileiras de capital aberto que atuam nos setores de “Comércio” e “Comércio e Distribuição”. Para isso, optou-se por aliar duas metodologias que partem de abordagens e contextos distintos: para estimação da eficiência técnica, utilizou-se o método consolidado de Análise de Fronteira Estocástica; para identificar os determinantes da eficiência, um modelo de árvores de regressão aditivas bayesianas (BART), situado na fronteira dos modelos de regressão não-paramétricos.

Como resultado, foi possível construir evidências que apontam à formação de um setor intensivo em fator capital, cujos níveis de eficiência não notadamente diferenciados entre os diversos segmentos que o compõe. Identificou-se também uma tendência de aumento da ineficiência ao longo dos últimos anos, ratificando a preocupação do mercado, de investidores e de consumidores em relação à sustentabilidade e viabilidade dessas companhias - em especial, dados os efeitos da pandemia da COVID-19.

Empresas que atuam no varejo supermercadista destacaram-se pelo elevado nível de eficiência. Algumas hipóteses que surgem para explicar esse fenômeno se relacionam ao período pandêmico - quando a demanda por alimentos cresceu e estes estabelecimentos mantiveram-se em plena operação, por se enquadrarem no rol de “atividades essenciais”. A rápida adaptação desses *players* a plataformas de venda *omnichannel*, que unem o espaço físico ao digital, também pode ajudar a justificar essa performance destacável.

Os mesmos argumentos se aplicam a algumas das companhias segmentadas na área de medicamentos e farmácias, como Dimed, Raia-Drogasil e Pague Menos: se caracterizam enquanto atividades essenciais, com elevada capilaridade e rápido crescimento no *e-commerce*, em especial após a popularização de serviços digitais de tele-entregas, como aplicativos de *delivery*. De modo similar, companhias tradicionais do varejo brasileiro também apresentaram um desempenho elevado, como Magazine Luiza e Casas Bahia.

Por outro lado, companhias da área têxtil figuraram entre as menos eficientes. Especificidades do setor podem ajudar a explicar esse fenômeno, como a mudança do padrão de consumo e a entrada de *players* internacionais no mercado da moda. Embora a variável “vl_import” tenha tido pouca relevância no modelo, deve-se ressaltar a carência de dados específicos para o segmento, que permitiriam uma

análise mais precisa das mudanças que afetam o setor, bem como de pesquisas que aprofundem as discussões a respeito dos desdobramentos que foram (e ainda serão) observados. Os recorrentes debates públicos acerca da taxação das importações voltadas ao consumidor final carecem deste tipo de análise criteriosa.

Em relação aos indicadores financeiros, os resultados geraram evidência a favor da hipótese de *market selection*, apontando maiores níveis de eficiência em companhias menores e com poder de mercado mais restrito. Quanta a estrutura de capital, os resultados mostram relação inversa entre alavancagem e eficiência. Este comportamento, que contraria estudos prévios aplicados em outros países, carece de futuras investigações. A hipótese de custos de agência não pode ser descartada - seja por eventuais conflitos de interesses entre gestores e acionistas, seja por falhas gerenciais, de tomada de decisão e de posicionamento no mercado.

Destaca-se também a necessidade de aprofundamento com relação aos aspectos operacionais das companhias analisadas. Há notável carência de dados a respeito do tema, o que traz nebulosidade aos estudos que se propõem a tecer análises setoriais. A interação entre indicadores operacionais e financeiros também se mostra fonte promissora de estudos voltados ao setor, em especial com a crescente digitalização do varejo nacional, fenômeno que altera significativamente o modelo de negócio das empresas, os riscos que tomam e a necessidade de liquidez.

Por fim, ressalta-se a relevância das metodologias empregadas neste estudo - em especial, do modelo BART. Método relativamente recente na literatura, suas aplicações ainda são pouco exploradas, em especial nos trabalhos em língua portuguesa. A ferramenta abre a possibilidade de avaliar intuitivamente a relevância das variáveis explicativas, sendo, simultaneamente, recurso poderoso para estudos de inferência causal e preditiva. Desdobramentos do presente estudo podem incluir a utilização desta abordagem no ajuste da própria fronteira estocástica, abrindo mão de uma série de hipóteses restritivas do modelo tradicional e viabilizando novas formas de observar o comportamento das firmas.

References

- Aigner, Dennis, C.A.Knox Lovell, and Peter Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". In: *Journal of Econometrics* 6.1, pp. 21–37. ISSN: 0304-4076. DOI: [https://doi.org/10.1016/0304-4076\(77\)90052-5](https://doi.org/10.1016/0304-4076(77)90052-5). URL: <https://www.sciencedirect.com/science/article/pii/0304407677900525>.
- Battese, G.E. and T.J. Coelli (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India". In: *Journal of Productivity Analysis* 3.1/2, pp. 153–169. ISSN: 0895562X, 15730441. URL: <http://www.jstor.org/stable/41770578> (visited on 01/11/2024).
- Breiman, L. et al. (1984). *Classification and Regression Trees*. Taylor & Francis. ISBN: 9780412048418. URL: <https://books.google.com.br/books?id=JwQx-W0mSyQC>.
- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (2010). "BART: Bayesian additive regression trees". In: *The Annals of Applied Statistics* 4.1, pp. 266–298. DOI: 10.1214/09-A0AS285. URL: <https://doi.org/10.1214/09-A0AS285>.
- Coelli, Tim, Sergio Perelman, and Elliot Romano (1999). "Accounting for Environmental Influences in Stochastic Frontier Models: With Application to International Airlines". In: *Journal of Productivity Analysis* 11.3, pp. 251–273. ISSN: 0895562X, 15730441. URL: <http://www.jstor.org/stable/41769985> (visited on 01/09/2024).
- Costa, Agnes de Souza, Glinda Sâmia da Silva Fôro, and Jeferson de Lima Vieira (ago. 2020). "COVID-19 e as cadeias de suprimentos: uma revisão bibliográfica dos principais impactos no Brasil". In: *Revista Vianna Sapiens* 11.2, p. 28. DOI: 10.31994/rvs.v11i2.687. URL: <https://viannasapiens.emnuvens.com.br/revista/article/view/687>.
- Cruz, Wander Luis de Melo (July 2021). "Crescimento do e-commerce no Brasil: desenvolvimento, serviços logísticos e o impulso da pandemia de Covid-19". In: *GeoTextos* 17.1. DOI: 10.9771/geo.v17i1.44572. URL: <https://periodicos.ufba.br/index.php/geotextos/article/view/44572>.
- Delardas, Orestis et al. (2022). "Socio-Economic impacts and challenges of the coronavirus pandemic (COVID-19): an updated review". In: *Sustainability* 14.15, p. 9699.

- Fenn, Paul et al. (2008). “Market structure and the efficiency of European insurance companies: A stochastic frontier analysis”. In: *Journal of Banking Finance* 32.1. Dynamics of Insurance Markets: Structure, Conduct, and Performance in the 21st Century, pp. 86–100. ISSN: 0378-4266. DOI: <https://doi.org/10.1016/j.jbankfin.2007.09.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0378426607002610>.
- Friedman, Jerome H. (2001). “Greedy function approximation: A gradient boosting machine.” In: *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: [10.1214/aos/1013203451](https://doi.org/10.1214/aos/1013203451). URL: <https://doi.org/10.1214/aos/1013203451>.
- Gupta, Astha Sanjeev and Jaydeep Mukherjee (2022). “Long-term changes in consumers’ shopping behavior post-pandemic: an exploratory study”. In: *International Journal of Retail & Distribution Management* 50.12, pp. 1518–1534.
- Hevia, Constantino, Pablo Andrés Neumeyer, et al. (2020). “A perfect storm: COVID-19 in emerging economies”. In: *COVID-19 in developing economies* 1.1, pp. 25–37.
- Kapelner, Adam and Justin Bleich (2016). “bartMachine: Machine Learning with Bayesian Additive Regression Trees”. In: *Journal of Statistical Software* 70.4, pp. 1–40. DOI: [10.18637/jss.v070.i04](https://doi.org/10.18637/jss.v070.i04). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v070i04>.
- M. Angeles Diaz, Rosario Sanchez (2008). “Firm size and productivity in Spain: a stochastic frontier analysis”. In: *Small Business Economics* 30, pp. 315–323. DOI: <https://doi.org/10.1007/s11187-007-9058-x>. URL: <https://link.springer.com/article/10.1007/s11187-007-9058-x>.
- Margaritis, Dimitris and Maria Psillaki (2007). “Capital Structure and Firm Efficiency”. In: *Journal of Business Finance & Accounting* 34.9-10, pp. 1447–1469. DOI: <https://doi.org/10.1111/j.1468-5957.2007.02056.x>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-5957.2007.02056.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5957.2007.02056.x>.