

ESTIMANDO ALÍQUOTAS DE IMPOSTO DE IMPORTAÇÃO DADO O NÍVEL DE COMPETITIVIDADE RELATIVA DO BRASIL

Eduardo Dornelas Munhoz¹

Gesner José Oliveira Filho²

José Matheus Andrade³

Rafael Pereira Oliveira⁴

Resumo: O objetivo deste artigo é propor uma metodologia baseada em modelos econométricos e de *machine learning* para estimar alíquotas de imposto de importação que seriam esperadas dado o nível de competitividade relativo do Brasil. Busca-se, assim, contribuir de maneira quantitativa com o debate em torno de qual deve ser o grau e velocidade adequados de processos de abertura comercial, considerando que existem gargalos estruturais e institucionais (ou ‘Custo Brasil’) que prejudicam o ambiente de negócios e limitam a competitividade das indústrias brasileiras frente à concorrência externa. A metodologia proposta, a qual pode ser aplicada para diferentes setores e produtos de interesse, utiliza indicadores do *Global Competitiveness Index* (GCI) como *proxies* para o grau de competitividade dos países. Juntamente com os fluxos de comércio do país para o produto de interesse, tais dados compõem o conjunto de covariadas que explicam a tarifa de importação. Calibra-se tal função a partir dos dados disponíveis entre 2017 e 2019 para os demais países e, então, estima-se qual seria a tarifa para o caso do Brasil. Apenas para fins de ilustração, este artigo realiza

¹ Mestre em Economia pela Universidade de Brasília. Coordenador de Infraestrutura na GO Associados.

² Ph.D. em Economia pela Universidade da Califórnia em Berkeley. Sócio executivo da GO Associados.

³ Mestre em Economia na Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. Consultor Sênior na GO Associados.

⁴ Mestre em Economia Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. Gerente de Projetos e Macrossetorial da GO Associados.

um estudo de caso para a indústria química, mais especificamente, para lixívia de soda cáustica.

Palavras-Chave: Custo Brasil. competitividade. proteção comercial. imposto de importação. machine learning.

Abstract: The objective of this article is to propose a methodology based on econometric and machine learning models to estimate import tax rates that would be expected given Brazil's relative competitiveness level. Thus, it aims to contribute quantitatively to the debate about the appropriate degree and speed of trade openness processes, considering that there are structural and institutional bottlenecks (or 'Custo Brasil') that hinder the business environment and limit the competitiveness of Brazilian industries against foreign competition. The proposed methodology, which can be applied to different sectors and products of interest, uses indicators from the Global Competitiveness Index (GCI) as proxies for the competitiveness level of countries. Together with the country's trade flows for the product of interest, these data constitute the set of covariates that explain the import tariff. This function is calibrated based on the available data between 2017 and 2019 for other countries, and then, it estimates what the tariff would be for Brazil. For illustrative purposes, this article conducts a case study for the chemical industry, specifically for caustic soda bleach.

Keywords: Custo Brasil. competitiveness. trade protection. import tax. machine learning.

1. Introdução

O objetivo deste artigo é propor uma metodologia baseada em modelos econométricos e de *machine learning* para estimar alíquotas de imposto de importação que seriam esperadas dado o nível de competitividade relativo do Brasil. Busca-se, assim, contribuir de maneira quantitativa com o debate político e acadêmico em torno de qual deve ser o grau e velocidade adequados de processos de abertura comercial, considerando que existem gargalos estruturais e institucionais (ou 'Custo Brasil') que prejudicam o ambiente de negócios e limitam a competitividade das indústrias brasileiras frente à concorrência externa.

A literatura acadêmica é extensa no que diz respeito aos benefícios econômicos de longo prazo da abertura comercial e da maior inserção dos países nas cadeias globais de valor. No entanto, o grau de

desenvolvimento de cada país, suas condições estruturais e qualidade das instituições, bem como dinâmicas setoriais tornam o processo de abertura complexo. A depender do grau de competitividade do país, reduções de barreiras tarifárias e aumento da concorrência externa podem gerar prejuízos significativos para a economia doméstica, como aumento do desemprego.

A interrupção das cadeias globais de valor causada pela pandemia de Covid-19 a partir de 2020, juntamente com a guerra entre Rússia e Ucrânia iniciada em 2022 e as mudanças climáticas, adicionaram ainda mais camadas de complexidade ao tema e colocaram em xeque a necessidade da liberalização comercial nos últimos anos pós-pandemia da Covid-19. Diversos governos têm se preocupado com suas relações comerciais e grau de dependência externa em termos de acesso a matérias-primas e insumos industriais⁵.

Tem-se observado cada vez mais uma tendência de desglobalização, marcada por países menos conectados em termos de comércio e uma maior valorização da produção e fontes renováveis domésticas. No Brasil, por exemplo, ganhou força em 2023 o discurso sobre a chamada ‘neointustrialização’⁶, a qual também possui forte componente ‘verde’ de aproveitamento do enorme potencial brasileiro na dimensão ambiental.

Assim, o debate político e acadêmico sobre o grau e velocidade adequados de processos de abertura comercial, que já era complexo por conta de uma série de aspectos estruturais de cada país, tem se tornado cada vez mais complexo devido a aspectos também conjunturais. Nesse sentido, a utilização de ferramentas quantitativas e criação de critérios objetivos sobre qual deve ser o posicionamento do Brasil no que tange à sua política de defesa comercial tornaram-se ainda mais relevantes.

Este artigo busca preencher essa lacuna, ao menos no que tange aos aspectos estruturais relevantes para o debate, e apresenta abordagem baseada em modelos econométricos e de *machine learning* capaz de

⁵ WEF. Global Competitiveness Report – 2018. World Economic Forum, 2018. Disponível em: <https://www3.weforum.org/docs/GCR2018/05FullReport/The-GlobalCompetitivenessReport2018.pdf>. Acesso: 30/07/2023.

⁶ SILVA, L. I. L.; ALCKMIN, G. Neointustrialização para o Brasil que queremos. **Governo Federal**, 2023. Disponível em: <https://www.gov.br/planalto/pt-br/vice-presidencia/central-de-conteudo/artigos/neointustrializacao-para-o-brasil-que-queremos>. Acesso em: 30/07/2023.

estimar as alíquotas esperadas de imposto de importação após levar em consideração a qualidade das instituições e o grau de competitividade relativo do Brasil frente a outros países.

Para tanto, indicadores do *Global Competitiveness Index* (GCI)⁷, disponíveis para 152 países, são utilizados como *proxies* e, juntamente com os fluxos de comércio do país para um dado produto de interesse, compõem o conjunto de covariadas que explicam a tarifa de importação. Calibra-se tal função a partir dos dados disponíveis entre 2017 e 2019 para os demais países e, então, estima-se qual seria a tarifa para o produto de interesse no caso do Brasil. De modo a ilustrar a metodologia, realiza-se estudo de caso para a indústria química, mais especificamente, para lixívia de soda cáustica.

Este artigo está organizado da seguinte forma, para além desta Introdução. A Seção 2 resume o debate a respeito dos benefícios e prejuízos do processo de abertura comercial. A Seção 3 detalha a metodologia quantitativa para estimar alíquotas de imposto de importação que seriam esperadas dado o nível de competitividade relativo do Brasil. A Seção 4 apresenta um estudo de caso para fins de ilustração, com aplicação da metodologia proposta para um produto escolhido de forma aleatória. Por fim, a Seção 5 apresenta as considerações finais.

2. Breve revisão de literatura

A literatura acadêmica é extensa no que diz respeito aos benefícios econômicos de longo prazo da abertura comercial e da maior inserção dos países nas cadeias globais de valor. Há diversos estudos que encontram evidências de efeitos positivos de reduções tarifárias no crescimento econômico e no bem-estar social⁸. Há também trabalhos que apontam que a redução de tarifas está ligada à incentivos

⁷ WEF. *Global Competitiveness Report – 2018*. World Economic Forum, 2018. Disponível em: <https://www3.weforum.org/docs/GCR2018/05FullReport/The-GlobalCompetitivenessReport2018.pdf>. Acesso: 30/07/2023.

⁸ FURCERI, D; HANNAN, S; OSTRY, J; ROSE, A. Are tariffs bad for growth? Yes, say five decades of data from 150 countries. *Journal of Policy Modeling*, v.42, p.850-859, 2020.

à exportação, notoriamente em setores em que a qualidade do produto é relevante⁹.

Tais resultados, de certa forma, já eram aguardados pela Teoria Econômica, visto que a imposição de impostos (ou outras formas de barreiras, tarifárias ou não) tendem a criar distorções na economia as quais geram uma alocação subótima dos recursos existentes em um país¹⁰. De fato, evidências sugerem que, após a eliminação dessas distorções via abertura comercial, há queda nos investimentos em indústrias de baixa produtividade e em setores mais expostos à competição com as importações, indicando realocação de recursos para agentes mais produtivos da economia¹¹.

Por outro lado, há estudos que mostram que a imposição de tarifas de importação em alguns setores pode gerar crescimento da renda nacional, por exemplo, em atividades que tenham ganhos claros via *learning by doing* e pesquisa e desenvolvimento¹². Não à toa existe extensa literatura (como em DIXIT, 1985¹³; BOND, 1990¹⁴; COSTINOT et al, 2015¹⁵) que analisa qual seria o patamar de uma *tarifa ótima de importação* e que, em geral, considera modelos de equilíbrio geral ou parcial para encontrar seus resultados.

Muitos modelos teóricos parecem não captar fatores institucionais e estruturais importantes que afetam a alocação dos recursos em

⁹ HAYAKAWA, K.; ISHIKAWA, J.; TARUI, N. What goes around comes around: Export-enhancing effects of import-tariff reductions. *Journal of International Economics*, v.126, 2020.

¹⁰ SAMUELSON, P. Ohlin was right. *Swedish Journal of Economics*, p.365–384, 1971.

¹¹ PIERCE, J.; SCHOTT, P. Investment responses to trade liberalization: Evidence from U.S. industries and establishments. *Journal of International Economics*, v.115, p.203-222, 2018.

¹² NUNN, N.; TREFLER, D. The structure of tariffs and long-term growth. *American Economic Journal: Macroeconomics*, v.2, n.4, p.158-194, 2010.

¹³ DIXIT, A. Chapter 6: Tax policy on open economies in *Handbook of Public Economics* vol.1. *Elsevier Science Publishers B.V.*, 1985.

¹⁴ BOND, E. The optimal tariff structure in higher dimensions. *International Economic Review*, v.31, p.103-116, 1990.

¹⁵ COSTINOT, A. DONALDSON, D.; VOGEL, J.; WERNING, I. Comparative advantage and optimal trade policy. *The Quarterly Journal of Economics*, p.659-702, 2015.

cada país. Embora seja discussão relativamente recente na Teoria Econômica, tem surgido cada vez mais trabalhos empíricos sobre os efeitos da qualidade das instituições e até de aspectos culturais e históricos sobre crescimento econômico¹⁶. No caso específico de fluxos de comércio e abertura comercial, existem evidências de que tais fatores institucionais e culturais podem, inclusive, ser acrescentados às vantagens comparativas na troca entre dois países¹⁷.

Quando se consideram tais fatores em conjunto com as tarifas de importação, essas perdem seu efeito para explicar o volume fluxo de comércio – ainda que os efeitos da interação entre tarifas e instituições mostre que os efeitos positivos da redução de tarifas são intensificados quanto melhores forem as instituições –¹⁸. Tal resultado sugere que, em caso de instituições fracas, os efeitos sobre o fluxo de comércio após reduções de barreiras comerciais podem ser reduzidos. Na mesma linha, há evidências de que o grau de abertura comercial é menos relevante para o crescimento econômico do que a qualidade das instituições e as políticas comerciais em si (tarifações e subsídios)¹⁹.

Considerando que a retirada ou diminuição de barreiras à importação pode causar efeitos negativos, ao menos no curto prazo, à economia doméstica devido ao fechamento de empresas e destruição de postos de trabalho²⁰, a literatura parece convergir para a conclusão de que a solução ótima é de sincronismo entre a redução de barreiras comerciais e aprimoramento do ambiente de negócios, melhora das instituições e eliminação de gargalos estruturais.

¹⁶ ACEMOGLU, D.; JOHNSON, S.; ROBINSON, A. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, v.91, n.5, p.1369-1401, 2001.

¹⁷ BELLOC, M.; BOWLES, S. Persistence and Change in Culture and Institutions under Autarchy, Trade, and Factor Mobility. *American Economic Journal: Microeconomics*, v.9, p.245-276, 2017.

¹⁸ JANSEN, M.; NORDAS, H. K. Institution, trade policy and trade flows. Staff Working Paper ERSD-2004-02, *World Trade Organization*, Economic Research and Statistics Division, 2004.

¹⁹ MAMOON, D.; MURSHED, S. M. Trade policy, openness, institutions. *The Pakistan Development Review*, v.45, n.1, p.99-119, 2006.

²⁰ Tais efeitos nocivos independem se o nível de produção era ou não eficiente, isto é, se a alocação dos recursos era ótima ou se o setor era artificialmente beneficiado pelo maior grau de proteção às importações.

No Brasil, popularizou-se o termo “Custo Brasil” para descrever o conjunto de gargalos estruturais, burocráticos e econômicos que encarecem o investimento e prejudicam o ambiente de negócios no país. Dentre tais gargalos estão fatores diretamente ligados às instituições, como elevada burocracia e corrupção, tamanho da carga tributária, bem como aspectos estruturais como alto custo logístico (derivado de uma gestão ineficiente pelo Estado) e ambiente de negócios pouco competitivo. A lógica do sincronismo é que uma redução gradual do Custo Brasil criaria condições para que as indústrias e empresas domésticas concorressem com o mercado externo a partir da redução de barreiras comerciais. Tal processo gradual de abertura permitiria o país absorver ganhos de bem estar e crescimento econômico ao longo do tempo, mas sem observar os potenciais impactos negativos de curto prazo.

Embora intuitiva, a ideia do sincronismo entre a redução de barreiras comerciais e redução de Custo Brasil não é trivial de ser implementada. Isso porque, para além dos aspectos conjunturais discutidos anteriormente, inexistem atualmente ferramentas quantitativas e critérios objetivos que indiquem qual é o patamar de proteção comercial esperado de uma dada indústria dado o nível de competitividade do Brasil frente ao resto do mundo. A metodologia quantitativa proposta neste artigo e apresentada na próxima seção busca suprir essa lacuna.

3. Metodologia quantitativa para estimação da alíquota esperada de imposto de importação

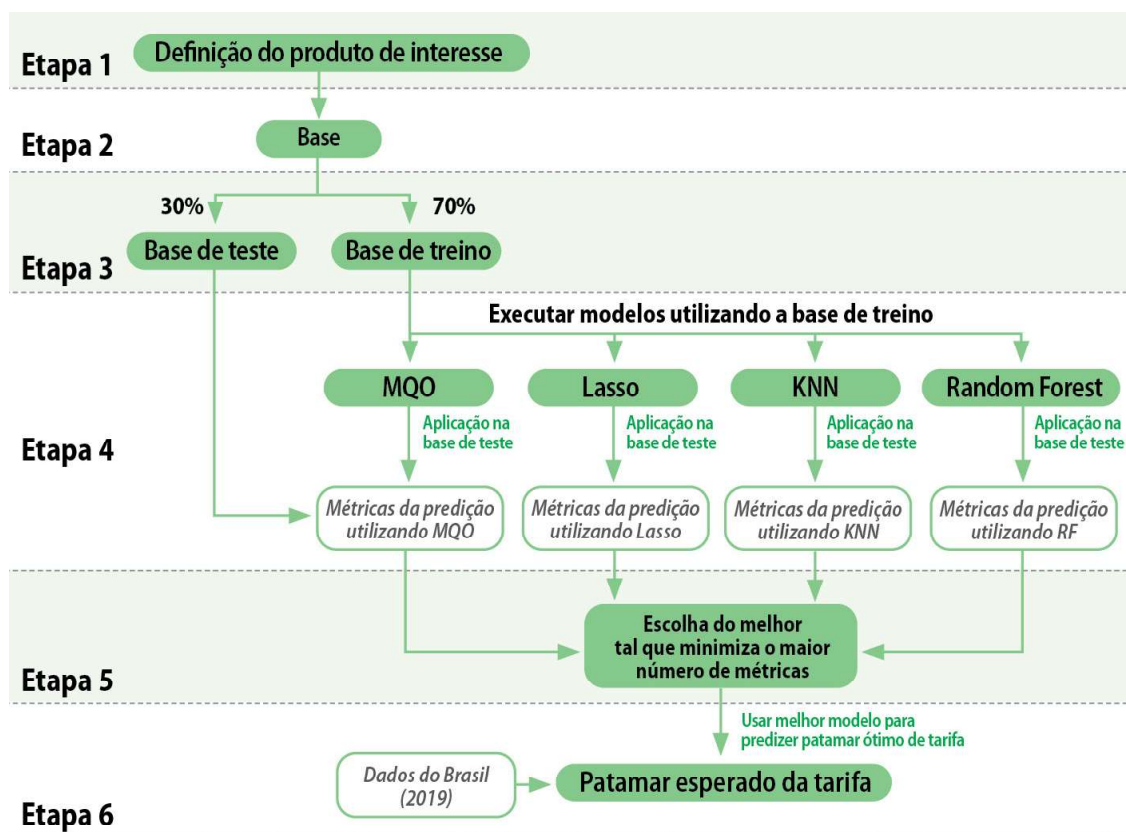
A metodologia de identificação da alíquota de imposto de importação esperada dado o nível de competitividade relativo do Brasil frente a outros países pode ser explicada em seis etapas, descritas a seguir e ilustradas na Figura 1:

- (i) A primeira etapa é a definição do produto de interesse, isto é, aquele cuja alíquota de imposto de importação esperada no Brasil se deseja estimar;
- (ii) Uma vez definido o produto de interesse, a segunda etapa é a coleta dos dados disponíveis sobre alíquotas de imposto de importação para todos os países, inclusive o Brasil. Constrói-se, assim, uma base de dados que possui

informações sobre todos os países no que tange às alíquotas do produto de interesse e um conjunto de variáveis explicativas – isto é, variáveis relevantes para explicar o patamar vigente da alíquota daquele produto;

- (iii) A terceira etapa é a segregação dos conjuntos de dados para a “base de treino” e para a “base de teste”. A base de treino será utilizada para que os modelos econométricos e de *machine learning* possam “aprender” sobre a relação existente entre as alíquotas de imposto de importação dos países e seu nível de competitividade, dado pela performance do país nas diferentes variáveis explicativas. Uma vez aprendida tal relação, a base de teste será utilizada para testar os modelos comparando suas estimativas de alíquota esperada com as alíquotas efetivamente observadas;
- (iv) A quarta etapa é a estimação dos diferentes modelos econométricos e de *machine learning* comumente utilizados na literatura acadêmica: Mínimos Quadrados Ordinários (MQO), Regressão de Lasso (Lasso), *K-Nearest Neighbors* (KNN) e *Random Forest*;
- (v) A quinta etapa é a de avaliação da performance dos modelos estimados na etapa anterior a partir de diferentes critérios comumente utilizados na literatura acadêmica para avaliação de modelos de previsão: Erro médio absoluto (*Mean Absolute Error* - MAE, na sigla em inglês), Erro médio absoluto escalado (*Mean Absolute Scaled Error* - MASE, na sigla em inglês), Erro quadrático médio (*Mean Squared Error* - MSE, na sigla em inglês) e Raiz quadrada do erro quadrático médio (*Root Mean Square Error* - RMSE, na sigla em inglês);
- (vi) Uma vez identificado o melhor modelo de previsão da alíquota de imposto de importação esperada dado o nível de competitividade relativo, a sexta etapa é a aplicação deste modelo para o produto de interesse no Brasil.

Figura 1: etapas da metodologia quantitativa de estimação da alíquota de imposto de importação esperada²¹



Fonte: Elaboração própria

3.1. Etapa 1: definição do produto de interesse

A primeira etapa da metodologia é a definição do produto de interesse, isto é, aquele cuja alíquota de imposto de importação esperada no Brasil se deseja estimar. Idealmente, o produto escolhido deve ser tal que seu NCM seja suficientemente desagregado para permitir a comparação entre os países e a interpretação sobre os patamares vigentes e esperados das alíquotas de imposto de importação. No caso em questão, conforme será discutido na Seção 4, apenas para demonstrar a metodologia, optou-se aleatoriamente pela utilização do produto *lixívia de soda cáustica*, cujo código de seis dígitos do sistema harmonizado é 2815.12.

²¹ Elaboração própria.













3.2. *Etapa 2: construção da base de dados*

A base de dados deve conter tanto as alíquotas de imposto de importação de diversos países quanto variáveis que reflitam o nível de competitividade de cada economia e que sejam capazes de, em alguma medida, explicar a alíquota vigente do produto de interesse. A seguir é apresentado um conjunto de variáveis explicativas utilizadas neste artigo para a estimação da alíquota esperada do produto de interesse.

3.2.1. *Dados do GCI 4.0*

De forma a captar o impacto da competitividade dos países na determinação de suas alíquotas de imposto de importação para o produto de interesse, foram coletados dados do GCI 4.0²² para todos os países disponíveis e referentes aos anos de 2017 a 2019, último ano disponível²³. Consideraram-se os 12 pilares do GCI apresentados na Figura 2 como variáveis explicativas elegíveis para comporem os modelos quantitativos, além do próprio índice agregado.

Figura 2: Pilares do GCI 4.0 (WEF, 2018)

Enabling Environment	Markets	Human Capital
 Pillar 1 Institutions	 Pillar 7 Product market	 Pillar 5 Health
 Pillar 2 Infrastructure	 Pillar 8 Labour market	 Pillar 4 Skills
 Pillar 3 ICT adoption	 Pillar 9 Financial system	Innovation Ecosystem
 Pillar 4 Macroeconomic stability	 Pillar 10 Market size	 Pillar 11 Business dynamism
		 Pillar 12 Innovation capability

²² WEF. Global Competitiveness Report – 2018. World Economic Forum, 2018. Disponível em: <https://www3.weforum.org/docs/GCR2018/05FullReport/The-GlobalCompetitivenessReport2018.pdf>. Acesso: 30/07/2023

²³ Em 2020, o relatório do Global Competitiveness Index foi publicado, mas apenas com os rankings de países entre os pilares, e não foi possível encontrar os dados abertos, como nos casos de 2017 a 2019. Desde aquele ano, não há novas publicações disponíveis.

3.2.2. *Dados exportações líquidas*

Além dos dados do GCI 4.0, considera-se também como variável explicativa as exportações líquidas do produto de interesse de cada país, calculadas como a diferença entre o valor exportado e o valor importado, em dólares, conforme os dados da base UN Comtrade. Com tal variável, espera-se captar efeitos sobre o volume de comércio exterior do produto de interesse em cada país na determinação de sua alíquota de imposto importação.

3.2.3. *Alíquota de imposto de importação*

Enquanto as variáveis explicativas descritas anteriormente são *regressores* – isto é, fatores que explicam a variável de interesse a ser regredida –, a alíquota de imposto de importação é a variável de interesse, ou seja, a variável regredida. Tais dados são coletados junto à OMC e se referem às alíquotas máximas do código NCM do produto de interesse. Em caso de eventuais dados faltantes na base, assumiu-se sempre a alíquota do último ano disponível. Ao final da construção da base de todos, todos os países que continham informações incompletas foram expurgados da base.

Para evitar qualquer viés no aprendizado dos modelos a serem testados, os dados referentes ao Brasil foram retirados da base de treino e de teste. Além disso, foram utilizados algoritmos com métodos de seleção automática de variáveis explicativas, incluindo o Mínimos Quadrados Ordinários com método *stepwise*. Isto é, o próprio procedimento econométrico automaticamente escolhe quais covariadas, dentre as disponibilizadas, são mais relevantes para estimar as alíquotas de imposto de importação. Tal seleção automática de variáveis explicativas é especialmente relevante para a seleção dos pilares do GCI com maior capacidade preditiva, além de reduzir drasticamente qualquer interferência do pesquisador sobre os resultados.

3.3. *Etapa 3: segregação dos conjuntos de treino e teste*

A terceira etapa da abordagem quantitativa é a preparação para estimação dos modelos econométricos e de *machine learning*. Para aplicação dos modelos, a base de dados construída foi segregada aleatoriamente em duas: uma base para teste (correspondente a 30%

das observações); e uma base para treino (70% das observações). Os modelos são treinados utilizando a base de treino; e sua capacidade preditiva é, posteriormente, mensurada a partir da comparação entre a previsão das alíquotas de imposto de importação por cada um dos modelos baseando-se nas variáveis explicativas presentes na base de teste e as próprias alíquotas observadas na base de teste.

3.4. Etapa 4: estimação dos modelos

Na quarta etapa, estimam-se quatro modelos, escolhidos pelo seu popular uso em estudos regressionais e preditivos na literatura acadêmica (SHAPI *et al*, 2018²⁴; IDICULA-THOMAS *et al*, 2021²⁵; DUDEK, 2022²⁶; KIM, 2017²⁷; BREINLICH *et al*, 2021²⁸), assim como – especialmente para os métodos de *machine learning* – sua robustez estatística. Em todos os casos, o método de seleção de regressores (variáveis utilizadas para explicar a alíquota de imposto de importação) é automatizado, de forma que se elimina quaisquer vieses de seleção por parte do pesquisador. Os quatro modelos considerados são descritos a seguir:

- Mínimos Quadrados Ordinários (MQO): método econométrico linear clássico. Neste artigo foi considerado em conjunto com o método *stepwise*, que escolhe automaticamente o melhor conjunto de regressores a partir do modelo que apresenta

²⁴ SHAPI, M.; RAMLI, N.; AWALIN, L. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in Built Environment*, v.5, mar. 2018.

²⁵ IDICULA-THOMAS, S.; GAWDE, U.; JHA, P. *Comparison of machine learning algorithms applied to symptoms to determine infectious causes of death in children: national survey of 18,000 verbal autopsies in the Million Death Study in India*. BMC Public Health, out. 2021. DOI: 10.1186/s12889-021-11829-y.

²⁶ DUDEK, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies*, v. 15, n. 20, p. 7547, 2022.

²⁷ KIM, I. S. Political cleavages within industry: Firm-level lobbying for trade liberalization. *American Political Science Review*, v. 111, n. 1, p. 1-20, 2017.

²⁸ BREINLICH, H.; CORRADI, V.; ROCHA, N.; RUTA, M.; SILVA, J. M. C. S.; ZYLKIN, T. Machine learning in international trade research: Evaluating the impact of trade agreements. Policy Research Working Paper 9629, *World Bank Group*. 2021.

a menor estatística AIC (“*Akaike Information Criterion*”) dentre os conjuntos testados;

- Regressão de Lasso (Lasso): método linear de *machine learning* que também realiza seleção de regressores, com a característica adicional de penalizar modelos com muitos regressores e / ou com regressores altamente correlacionados entre si, de forma a evitar o problema de “*overfitting*”;
- *K-Nearest Neighbors* (KNN): algoritmo de *machine learning* capaz de agrupar observações (nesse caso, países) em grupos (“vizinhos”) conforme as características apresentadas pelos regressores. Quando requisitado que estime as observações da base de teste, o algoritmo analisa cada observação de teste e, matematicamente, encontra o vizinho mais próximo; e
- *Random Forest*: método paramétrico de *machine learning* bastante utilizado em análises quantitativas. É baseado no método “*ensemble*”, em que diversos modelos diferentes são gerados (neste caso, “árvores de decisão”) e o resultado é gerado a partir da agregação dos resultados individuais.

Como comentado anteriormente, tais modelos são executados levando-se em conta a base de treino. Posteriormente, são avaliados por sua capacidade de prever as alíquotas de imposto de importação da base de teste a partir das variáveis explicativas contidas nesta base.

3.5. *Etapa 5: avaliação da performance dos modelos estimados e Etapa 6: aplicação*

Realizado o treinamento e o teste de todos os modelos, avalia-se sua performance a partir de diferentes métricas, o que garante robustez à abordagem quantitativa. Quatro métricas avaliaram o desempenho dos modelos²⁹, todas elas comuns na avaliação de modelos de previsão:

- i. Erro médio absoluto (*Mean Absolute Error* - MAE, na sigla em inglês): calcula a média da diferença absoluta entre os valores reais e os previstos, isto é, o módulo do erro observado;

²⁹ Todas as métricas utilizadas foram computadas pelo pacote *Metrics*, (HAMMER, FRASCO E LeDELL, 2022) da linguagem de programação R. Disponível em: <https://cran.r-project.org/web/packages/Metrics/Metrics.pdf>. Acesso: 30/07/2023.

- ii. Erro médio absoluto escalado (*Mean Absolute Scaled Error* - MASE, na sigla em inglês): tem a mesma lógica que o MAE, porém os erros são escalados por uma estimativa “naive” do valor da série histórica. Isto é, compara-se o erro observado ao erro que seria observado fazendo uma estimativa naive de usar o próprio valor da amostra do período anterior para predizer o próximo período. Caso o MASE seja superior a um, a estimativa naive é melhor do que a do modelo avaliado;
- iii. Erro quadrático médio (Mean Squared Error - MSE, na sigla em inglês): calcula a média dos erros de cada predição e os eleva ao quadrado. Uma vez que cada erro é elevado ao quadrado, este método dá um peso maior a erros mais distantes dos valores atuais. Note que a unidade dessa métrica é o quadrado da unidade das observações (exemplo: se as observações estão na casa de dezenas, então o MSE mostrará uma medida de erro na casa de centenas);
- iv. Raiz quadrada do erro quadrático médio (Root Mean Square Error - RMSE, na sigla em inglês): é a raiz quadrada do MSE, procedimento que serve para que a unidade da métrica seja igual à unidade das observações. Além disso, tal procedimento também faz com que o RMSE seja menos sensível à presença de outliers.

Especificamente para os modelos Lasso, KNN e Random Forest, que possuem algum grau de aleatoriedade em seus métodos, os modelos testados foram executados a partir de diversos *seeds*³⁰ de modo a manter a replicabilidade da análise e, ao mesmo tempo, também realizar análises de sensibilidade. Os resultados das previsões de cada um desses modelos consideram a média dos critérios entre os vários *seeds* aplicados.

³⁰ Em métodos que possuem algum grau de aleatoriedade em sua construção, executá-lo mais de uma vez pode gerar resultados ligeiramente diferentes entre si, mesmo utilizando a mesma base para treino. Ainda que o resultado, de forma geral, não se altere, tal característica pode atrapalhar a replicabilidade do exercício. Portanto, para os três modelos supracitados, utilizam-se diversos *seeds*, isto é, funções que retiram a aleatoriedade dos geradores de números aleatórios presentes em *softwares* e linguagens de programação, de forma que diversas execuções do mesmo modelo sob o mesmo conjunto de treino sempre apresentarão o mesmo resultado sob o mesmo *seed*.

Os modelos foram estimados utilizando-se o *software* econométrico R. Em especial, os seguintes pacotes foram utilizados: (i) Caret: (KUNH, 2008³¹) utilizado, por exemplo, para treinar os modelos *Random Forest* e KNN; (ii) Stats:³² utilizado, por exemplo, para aplicar o método MQO junto com *stepwise*; (iii) Metrics: (HAMNER E FRASCO, 2018³³) utilizado, por exemplo, para calcular as métricas de erro para escolher o melhor modelo; e (iv) glmnet (FRIEDMAN *et al*, 2010³⁴): utilizado, por exemplo, para aplicar o modelo Lasso.

Uma vez identificado o melhor modelo de previsão da alíquota de imposto de importação esperada dado o nível de competitividade relativo, a sexta etapa é a aplicação deste modelo para o produto de interesse no Brasil. Para tanto, estima-se, a partir do melhor modelo, a alíquota que seria esperada para o produto de interesse dada a performance do Brasil nas variáveis explicativas.

4. Estudo de caso: aplicação da metodologia na indústria química

A metodologia de identificação da alíquota de imposto de importação esperada dado o nível de competitividade relativo do Brasil frente a outros países foi aplicada, de maneira aleatória, para a indústria química, mais especificamente, para a lixívia de soda cáustica. Os resultados são apresentados na Tabela 1.

Os modelos de MQO, Lasso, KNN e Random Forest foram estimados a partir das variáveis explicativas presentes na base de teste para prever a alíquota de imposto de importação da lixívia de soda cáustica. Os resultados na Tabela 1 mostram que o KNN foi o modelo

³¹ KUNH, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v.28, n.5, pág.1-26, 2008.

³²R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. 2013. Disponível em: <http://www.R-project.org/>. Acesso em: 30/07/2023

³³ HAMNER, B; FRASCO, M. Metrics: Evaluation Metrics for Machine Learning. *R Package*. 2018. Disponível em: <<https://CRAN.R-project.org/package=Metrics>>. Acesso em: 30/07/2023.

³⁴ FRIEDMAN, J; HASTIE, T; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, v. 33, n. 1, p. 1, 2010.

com melhor performance na previsão das alíquotas de imposto de importação da lixívia de soda cáustica da base teste em três das quatro métricas e, portanto, foi considerado o melhor modelo e aquele a ser aplicado para o caso brasileiro.

Tabela 1: Avaliação da performance dos modelos estimados

Modelo*	Média de RMSE	Média de MAE	Média de MSE	Média de MASE	Predição da tarifa
KNN	3,22	2,09	10,38	0,57	12,75
MQO	4,06	3,03	16,44	0,60	5,28
Lasso	4,00	2,92	16,03	1,01	5,65
Random Forest	3,27	2,30	10,68	0,46	11,58

Fonte: Elaboração própria. *Apresentou-se a média dos resultados de cada modelo para as diferentes *seeds* consideradas.

Considerando os dados do Brasil em 2019, bem como as diversas variáveis explicativas para o país, estimou-se a alíquota de imposto de importação esperada da lixívia de soda cáustica dado o nível de competitividade relativa do país a partir do modelo KNN. Observa-se que a alíquota esperada para o produto de interesse é de 12,75%, enquanto a alíquota vigente é de 7,2%. Isto é, a análise quantitativa proposta neste artigo sugere que uma alíquota de 12,75% seria mais adequada para o produto de interesse tendo em vista o nível de competitividade relativo do Brasil.

Em suma, considerando a abordagem quantitativa proposta, percebe-se que, em relação ao nível de competitividade do Brasil frente ao resto do mundo – refletido por um conjunto de variáveis explicativas, em especial, do GCI –, a alíquota estimada pelo modelo para o país é de 12,75%. Isto significa um aumento de 5,55 pontos percentuais em relação à alíquota vigente.

Deve-se ressaltar algumas limitações da metodologia proposta. Por exemplo, o modelo não capta potenciais efeitos socioeconômicos

ou ambientais de se produzir localmente determinados insumos. Tampouco considera questões estratégicas de segurança em termos de fornecimento de matérias-primas nem a existência de setores estratégicos ou preocupações relacionadas à soberania nacional. Ademais, os modelos levam em consideração *proxies* agregadas do grau de competitividade do país e sua relação com as tarifas de importação ao redor do mundo, mas não capta diretamente a diferença de custos de capital e outras variáveis financeiras que são importantes para determinar a rentabilidade de um projeto de investimento.

Entre os possíveis aprimoramentos da metodologia proposta, pode-se citar: a inclusão de mais variáveis explicativas, que podem variar entre os diferentes setores da economia; inclusão de uma variável *dummy* para captar se o país em questão fabrica aquele produto; utilização de uma base de dados mais atualizada (a depender de divulgações mais recentes do GCI ou outras fontes alternativas); inclusão de fator dinâmico no modelo (o qual considera as variáveis em *cross section*).

5. Considerações finais

Este artigo propõe uma metodologia baseada em modelos econométricos e de *machine learning* para estimar alíquotas de imposto de importação que seriam esperadas dado o nível de competitividade relativo do Brasil. Busca-se, assim, suprir a ausência de ferramentas quantitativas e critérios objetivos que permitam a implementação da ideia do sincronismo entre a redução de barreiras comerciais e redução de Custo Brasil.

A metodologia proposta utiliza a relação entre os pilares do *Global Competitiveness Index* (GCI), mensurado pelo Fórum Econômico Mundial, como uma *proxy* para o grau de competitividade dos países. Juntamente com os fluxos de comércio do país (exportações líquidas) para um dado produto de interesse, tais dados compõem o conjunto de covariadas que explicam a tarifa de importação. São utilizados dados de 2017 a 2019.

Calibra-se tal função para os demais países e, então, estima-se qual seria a tarifa para o caso do Brasil considerando seu grau de competitividade. A lógica subjacente da metodologia proposta é: qual seria a alíquota de imposto de importação esperada para o Brasil se a

mesma relação entre os pilares do GCI e alíquotas aplicadas no conjunto dos demais países vigorasse para o caso brasileiro.

Tal metodologia é passível de ser aprimorada por meio, por exemplo, da inclusão de novas variáveis explicativas ou uso de outras bases de dados. Mas o que chama atenção é sua flexibilidade: pode ser aplicada para qualquer setor ou produto de interesse. De modo a ilustrar seu funcionamento, este artigo aplicou a metodologia para o caso da “lixívia de soda cáustica”, produto escolhido aleatoriamente e cujo código no sistema harmonizado é “2815.12”.

Após testar os modelos Mínimos Quadrados Ordinários (MQO) com algoritmo *stepwise*, Lasso, KNN, Random Forest, verificou-se que, para o exemplo em questão, o modelo com melhor desempenho dentre as métricas de erro consideradas foi o KNN. Seu resultado foi o de que a alíquota de imposto de importação que é esperada para o Brasil dado o seu nível de competitividade relativa seria de 12,75%. Isto significa um aumento de 5,55 pontos percentuais em relação à alíquota vigente, de 7,2%.

Conforme destacado, os resultados da aplicação da metodologia proposta devem ser analisados com parcimônia e à luz de suas limitações. Dentre elas, pode-se destacar que o modelo não capta, do ponto de vista macroeconômico, potenciais efeitos socioeconômicos ou ambientais de se produzir localmente determinados insumos. Tampouco considera questões estratégicas de segurança em termos de fornecimento de matérias-primas nem a existência de setores estratégicos ou preocupações relacionadas à soberania nacional. Do ponto de vista microeconômico, o modelo não capta aspectos como o custo de capital e de instalação de maquinário, de modo que a alíquota esperada resultante não necessariamente é aquela que permitiria equiparar o retorno de se investir no Brasil para produzir determinado produto ao retorno que se pode obter em outros países.

Em que pesem tais limitações, os resultados para o caso da lixívia de soda cáustica ilustram a capacidade de a metodologia proposta adicionar uma dimensão quantitativa fundamental e atualmente inexistente ao processo de determinação de alíquotas de imposto de importação no Brasil. Na mesma linha, contribui com o debate político e acadêmico em torno de qual deve ser o grau e velocidade adequados de processos de abertura comercial.

Referências bibliográficas

- ACEMOGLU, D.; JOHNSON, S.; ROBINSON, A. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, v.91, n.5, p.1369-1401, 2001.
- BELLOC, M.; BOWLES, S. Persistence and Change in Culture and Institutions under Autarchy, Trade, and Factor Mobility. *American Economic Journal: Microeconomics*, v.9, p.245-276, 2017.
- BOND, E. The optimal tariff structure in higher dimensions. *International Economic Review*, v.31, p.103-116, 1990.
- BREINLICH, H.; CORRADI, V.; ROCHA, N.; RUTA, M.; SILVA, J. M. C. S.; ZYLKIN, T. Machine learning in international trade research: Evaluating the impact of trade agreements. Policy Research Working Paper 9629, *World Bank Group*. 2021.
- COSTINOT, A. DONALDSON, D.; VOGEL, J.; WERNING, I. Comparative advantage and optimal trade policy. *The Quarterly Journal of Economics*, p.659-702, 2015.
- DIXIT, A. Chapter 6: Tax policy on open economies in *Handbook of Public Economics vol.1. Elsevier Science Publishers B.V.*, 1985.
- DUDEK, G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies*, v. 15, n. 20, p. 7547, 2022.
- FAN, H.; LI, Y.; YEAPLE, S. Trade liberalization, quality and export prices. *Review of Economics and Statistics*, v.97, n.5, 2015.
- FRIEDMAN, J; HASTIE, T; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, v. 33, n. 1, p. 1, 2010.
- FURCERI, D; HANNAN, S; OSTRY, J; ROSE, A. Are tariffs bad for growth? Yes, say five decades of data from 150 countries. *Journal of Policy Modeling*, v.42, p.850-859, 2020.
- HAMNER, B; FRASCO, M. Metrics: Evaluation Metrics for Machine Learning. *R Package*. 2018. Disponível em: <<https://CRAN.R-project.org/package=Metrics>>. [Acesso em: 30/07/2023](#).

HAYAKAWA, K.; ISHIKAWA, J.; TARUI, N. What goes around comes around: Export-enhancing effects of import-tariff reductions. *Journal of International Economics*, v.126, 2020.

IDICULA-THOMAS, S.; GAWDE, U.; JHA, P. *Comparison of machine learning algorithms applied to symptoms to determine infectious causes of death in children: national survey of 18,000 verbal autopsies in the Million Death Study in India*. BMC Public Health, out. 2021. DOI: 10.1186/s12889-021-11829-y.

JANSEN, M.; NORDAS, H. K. Institution, trade policy and trade flows. Staff Working Paper ERSD-2004-02, *World Trade Organization*, Economic Research and Statistics Division, 2004.

KIM, I. S. Political cleavages within industry: Firm-level lobbying for trade liberalization. *American Political Science Review*, v. 111, n. 1, p. 1-20, 2017.

KUHN, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, v.28, n.5, pág.1-26, 2008.

MAMOON, D.; MURSHED, S. M. Trade policy, openness, institutions. *The Pakistan Development Review*, v.45, n.1, p.99-119, 2006.

NAIDIN, L. C.; VEIGA, P. da M.; RIOS, S. P. *Liberalização comercial sob Bolsonaro: O que foi feito e o que pode ser feito até o final do governo*. Centro de Estudos de Integração e Desenvolvimento (CINDES). 2021.

NUNN, N.; TREFLER, D. The structure of tariffs and long-term growth. *American Economic Journal: Macroeconomics*, v.2, n.4, p.158-194, 2010.

PIERCE, J.; SCHOTT, P. Investment responses to trade liberalization: Evidence from U.S. industries and establishments. *Journal of International Economics*, v.115, p.203-222, 2018.

R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. 2013. Disponível em: <http://www.R-project.org/>. Acesso em: 30/07/2023.

SAMUELSON, P. Ohlin was right. *Swedish Journal of Economics*, p.365–384, 1971.

SILVA, L. I. L.; ALCKMIN, G. Neoindustrialização para o Brasil que queremos. *Governo Federal*, 2023. Disponível em: <https://www.gov.br/planalto/pt-br/vice-presidencia/central-de-conteudo/artigos/neoindustrializacao-para-o-brasil-que-queremos>. Acesso em: 30/07/2023.

SHAPI, M.; RAMLI, N.; AWALIN, L. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in Built Environment*, v.5, mar. 2018.

WEF. Global Competitiveness Report – 2018. World Economic Forum, 2018. Disponível em: <https://www3.weforum.org/docs/GCR2018/05FullReport/TheGlobalCompetitivenessReport2018.pdf>. Acesso: 30/07/2023

