



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E MATEMÁTICA APLICADA
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM E MÉTODOS
QUANTITATIVOS
MESTRADO ACADÊMICO EM MODELAGEM E MÉTODOS QUANTITATIVOS

YASMIN MARA DOS SANTOS VIEIRA

MODELOS GAMLSS PARA FLUXOS COMERCIAIS:
UMA ABORDAGEM ALTERNATIVA À GRAVITAÇÃO CLÁSSICA

FORTALEZA

2025

YASMIN MARA DOS SANTOS VIEIRA

MODELOS GAMLSS PARA FLUXOS COMERCIAIS:
UMA ABORDAGEM ALTERNATIVA À GRAVITAÇÃO CLÁSSICA

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Modelagem e Análise Quantitativa.

Orientador: Prof. Dr. Rafael Bráz Azevedo Farias.

Coorientador: Prof. Dr. Manoel Ferreira dos Santos Neto.

FORTALEZA

2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

V719m Vieira, Yasmin Mara dos Santos.

Modelos GAMLSS para fluxos comerciais : uma abordagem alternativa à gravitação clássica / Yasmin Mara dos Santos Vieira. – 2025.
101 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Modelagem e Métodos Quantitativos, Fortaleza, 2025.

Orientação: Prof. Dr. Rafael Bráz Azevedo Farias.

Coorientação: Prof. Dr. Manoel Ferreira dos Santos Neto.

1. Modelo gravitacional do comércio. 2. GAMLSS - ZAGA. 3. Comércio internacional - Modelos matemáticos. 4. Modelos lineares generalizados (MLG). 5. Acordos regionais. I. Título.

CDD 510

YASMIN MARA DOS SANTOS VIEIRA

MODELOS GAMLSS PARA FLUXOS COMERCIAIS:
UMA ABORDAGEM ALTERNATIVA À GRAVITAÇÃO CLÁSSICA

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem e Métodos Quantitativos do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Modelagem e Métodos Quantitativos. Área de Concentração: Modelagem e Análise Quantitativa.

Aprovada em: 05/11/2025.

BANCA EXAMINADORA

Prof. Dr. Rafael Bráz Azevedo Farias (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Manoel Ferreira dos Santos Neto (Coorientador)
Universidade Federal do Ceará (UFC)

Profª. Dra. Sílvia Maria de Freitas
Universidade Federal do Ceará (UFC)

Prof. Dr. Helton Saulo Bezerra dos Santos
Universidade de Brasília (UnB)

RESUMO

O modelo gravitacional é frequentemente usado para descrever os fluxos do comércio internacional, mas enfrenta dificuldades para lidar com dados caracterizados por heterocedasticidade, assimetria e excesso de zeros. Este trabalho analisa a eficácia dos Modelos Aditivos Generalizados para Localização, Escala e Forma (GAMLSS) na modelagem desses fluxos, em relação aos estimadores convencionais, como o Pseudo-Máxima Verossimilhança de Poisson (PPML) e o Pseudo-Máxima Verossimilhança de Gama (GPML). A aplicação empírica utiliza informações sobre as exportações brasileiras de cera de carnaúba, um produto de relevância econômica e estratégica, considerando fatores econômicos e institucionais que influenciam o comércio. Foram estimados modelos com distribuições contínuas e inflacionadas em zero, destacando-se a distribuição *Zero Adjusted Gamma* (ZAGA), que apresentou o melhor desempenho segundo critérios de informação e diagnóstico de resíduos. Os resultados mostram que, apesar de o PPML e o GPML serem consistentes em certos contextos, ambos exibiram limitações quando enfrentaram alta variabilidade e presença de zeros. Por outro lado, o modelo GAMLSS (ZAGA) alcança um desempenho melhor, possibilitando a modelagem simultânea da média, dispersão e probabilidade de fluxos nulos. Do ponto de vista substancial, os principais efeitos previstos pela teoria gravitacional foram confirmados, incluindo o impacto positivo do PIB e da população dos países importadores, bem como os efeitos negativos da distância, do câmbio e da ausência de litoral. Verificou-se ainda que os acordos regionais, embora não tenham elevado o volume médio das exportações, contribuíram para reduzir sua variabilidade e a frequência de períodos sem comércio, atuando como instrumentos de estabilização das relações comerciais. A principal contribuição desta pesquisa é demonstrar como os modelos GAMLSS são robustos e flexíveis para analisar fluxos comerciais, ampliando as opções de ferramentas disponíveis para estudos em economia internacional.

Palavras-chave: modelo gravitacional do comércio; GAMLSS - ZAGA; comércio internacional - modelos matemáticos; Modelos lineares generalizados (MLG); acordos regionais.

ABSTRACT

The gravity model is frequently used to describe international trade flows, but it faces challenges when dealing with data characterized by heteroskedasticity, asymmetry, and excess zeros. This study analyzes the effectiveness of Generalized Additive Models for Location, Scale, and Shape (GAMLSS) in modeling such flows, compared to conventional estimators such as the Poisson Pseudo-Maximum Likelihood (PPML) and the Gamma Pseudo-Maximum Likelihood (GPML). The empirical application uses data on Brazilian exports of carnauba wax—a product of economic and strategic relevance—considering both economic and institutional factors that influence trade. Models with continuous and zero-inflated distributions were estimated, highlighting the Zero Adjusted Gamma (ZAGA) distribution, which showed the best performance according to information criteria and residual diagnostics. The results indicate that although PPML and GPML are consistent in certain contexts, both displayed limitations when facing high variability and the presence of zeros. In contrast, the GAMLSS (ZAGA) model achieved superior performance, allowing for the simultaneous modeling of the mean, dispersion, and probability of zero trade flows. From a substantive standpoint, the main effects predicted by gravity theory were confirmed, including the positive impact of the importers' GDP and population, as well as the negative effects of distance, exchange rate, and lack of coastline. It was also found that regional agreements, although not increasing the average export volume, contributed to reducing its variability and the frequency of zero-trade periods, acting as stabilizing instruments for trade relations. The main contribution of this research is to demonstrate how GAMLSS models are robust and flexible tools for analyzing trade flows, expanding the range of methodologies available for studies in international economics.

Keywords: gravitational model of trade; GAMLSS - ZAGA; international trade - mathematical models; Generalized linear models (GLMs); regional trade agreements.

LISTA DE FIGURAS

Figura 1 – Ocorrência da extração da Carnaúba (2017).	52
Figura 2 – Evolução da produção nacional de cera de carnaúba (1997–2024).	53
Figura 3 – Participação dos principais países importadores no volume das exportações brasileiras de cera de carnaúba (2024).	55
Figura 4 – Volume e valor das exportações brasileira de cera de carnaúba (1997-2024).	55
Figura 5 – Distribuição das exportações antes e após transformação logarítmica.	61
Figura 6 – Gráficos de diagnóstico do ajuste do modelo com distribuição PPML.	66
Figura 7 – Gráficos de diagnóstico do ajuste do modelo com distribuição GPML.	67
Figura 8 – <i>Half-Normal Plots</i> com envelopes simulados para os resíduos dos modelos.	68
Figura 9 – <i>Half-Normal Plot</i> para os resíduos do modelo ZAGA.	75
Figura 10 – <i>Worm Plot</i> para os resíduos do modelo ZAGA.	75
Figura 11 – Gráficos de diagnóstico do ajuste do modelo com distribuição ZAGA.	76
Figura 12 – Densidade dos resíduos quantílicos dos modelos GPML, PPML e ZAGA.	78
Figura 13 – Exportações observadas em relação aos preditores lineares dos parâmetros do modelo MLG.	79
Figura 14 – Exportações observadas em relação aos preditores lineares dos parâmetros do modelo GAMLSS.	80
Figura 15 – Efeitos estimados das variáveis explicativas sobre os parâmetros do modelo ZAGA.	81
Figura 16 – Distribuição dos fluxos de exportação conforme a presença ($rta = 1$) ou ausência ($rta = 0$) de acordos regionais.	84
Figura 17 – Diferentes tipos de falhas no modelo indicados pelo gráfico de minhoca.	95

LISTA DE QUADROS

Quadro 1 – Nomenclatura e especificação das variáveis apresentadas.	58
-----------------------------------------------------------------------------	----

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas das variáveis contempladas no estudo.	60
Tabela 2 – Estimativas do modelo gravitacional para os modelos MLG.	65
Tabela 3 – Testes estatísticos aplicados aos resíduos do modelo PPML.	69
Tabela 4 – Testes estatísticos aplicados aos resíduos do modelo GPML.	70
Tabela 5 – Valores do GAIC para diferentes distribuições (resumo).	71
Tabela 6 – Valores do critério GAIC para seleção de variáveis nos ajustes dos parâmetros.	72
Tabela 7 – Estimativas do modelo ZAGA.	73
Tabela 8 – Testes estatísticos aplicados aos resíduos do modelo ZAGA.	77
Tabela 9 – Momentos estatísticos dos resíduos dos modelos estimados.	77
Tabela 10 – Efeitos de +10% (variáveis em log) e de +1/+10 pt (nível) sobre μ	99
Tabela 11 – Efeitos de +10% (variáveis em log) e de dummy sobre σ	99
Tabela 12 – Mudança nas <i>odds</i> de zeros (ν) para +10% (log) e +1 pt/dummy (nível).	99

LISTA DE ABREVIATURAS E SIGLAS

AIC	<i>Akaike Information Criterion</i>
AICc	Critério de Informação de Akaike Corrigido
ANOVA	Análise de Variância
ARMA	Modelo Autorregressivo de Médias Móveis
BIC	<i>Bayesian Information Criterion</i>
BNDES	Banco Nacional de Desenvolvimento Econômico e Social
CEPAL	Comissão Econômica para a América Latina e o Caribe
CEPII	<i>Centre d'Études Prospectives et d'Informations Internationales</i>
CES	Elasticidade de Substituição Constante
Comex stat	Sistema de Estatísticas do Comércio Exterior do Brasil
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EUA	Estados Unidos
FMI	<i>International Monetary Fund</i>
GAIC	<i>Generalized Akaike Information Criterion</i>
GAM	Modelos Aditivos Generalizados
GAMLSS	<i>Generalized Additive Models for Location, Scale and Shape</i>
GPML	<i>Gamma Pseudo Maximum Likelihood</i>
HNP	<i>Half Normal Plot</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
ILC	Índice de Liberdade Comercial
MDIC	Ministério do Desenvolvimento, Indústria e Comércio Exterior
MERCOSUL	Mercado Comum do Sul
MLG	Modelos Lineares Generalizados
MQO	Mínimos Quadrados Ordinários
NAFTA	Acordo de Livre Comércio da América do Norte
NCM	Nomenclatura Comum do Mercosul
NLS	Mínimos Quadrados Não Lineares
OMC	Organização Mundial do Comércio
PIB	Produto Interno Bruto
PML	<i>Pseudo Maximum Likelihood</i>

PPML	<i>Poisson Pseudo Maximum Likelihood</i>
rta	<i>Regional Trade Agreements</i>
SISCOMEX	Sistema Integrado de Comércio Exterior
ZAGA	<i>Zero-Adjusted Gamma</i>
ZAIG	<i>Zero Adjusted Inverse Gaussian</i>
ZANBI	<i>Zero Inflated Negative Binomial</i>
ZAP	<i>Zero Adjusted Poisson</i>
ZIP	<i>Zero-Inflated Poisson</i>
ZIP2	<i>Zero Inflated Poisson type 2</i>

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Considerações Gerais	12
1.2	O problema de pesquisa e sua relevância	13
1.3	Justificativa	14
1.4	Objetivos	14
1.4.1	<i>Objetivo Geral</i>	14
1.4.2	<i>Objetivos Específicos</i>	15
1.5	Estrutura do Trabalho	15
2	BASES TEÓRICAS E METODOLÓGICAS	17
2.1	O Comércio Internacional Brasileiro	17
2.2	Determinantes do Comércio Internacional	21
2.2.1	<i>Métodos de Coleta e Análise: Da Seleção à Modelagem</i>	21
2.2.2	<i>Variáveis Mais Impactantes nos Fluxos Comerciais</i>	24
2.2.2.1	<i>Revisão e Evidências Empíricas sobre o Comércio</i>	28
2.3	O Modelo Gravitacional no Comércio Internacional	29
3	MODELOS GRAVITACIONAIS E MÉTODOS DE ESTIMAÇÃO . . .	34
3.1	Modelo de Gravidade Tradicional	34
3.2	Modelo CES	35
3.3	Modelo PPML	37
3.3.1	<i>Estimadores não lineares</i>	38
3.3.2	<i>Técnicas de diagnóstico</i>	40
3.4	<i>Generalized Additive Model for Location, Scale and Shape (GAMLSS) . .</i>	40
3.4.1	<i>Estimação dos parâmetros do modelo</i>	42
3.4.2	<i>Modelos de Regressão para Dados Inflacionados e Ajustados</i>	43
3.4.2.1	<i>Distribuição Poisson Inflacionado em Zero (ZIP)</i>	43
3.4.2.2	<i>Distribuição Gama Ajustada a Zero (ZAGA)</i>	44
3.5	Técnicas de Diagnóstico	45
3.5.1	<i>Análise de Resíduos</i>	46
3.5.2	<i>Método Gráfico</i>	47
3.5.3	<i>Half Normal Plot</i>	47

3.5.4	<i>Worm Plot</i>	48
3.6	Seleção de Modelos	49
4	CARACTERIZAÇÃO DO MERCADO INTERNACIONAL DA CERA DE CARNAÚBA	51
4.1	Produção e Base Geográfica	51
4.1.1	<i>Trajetória da Produção</i>	53
4.2	Evolução Histórica das Exportações e Principais Destinos	54
5	APLICAÇÃO	57
5.1	Base de dados	57
5.1.1	<i>Dados e Variáveis do modelo</i>	58
5.2	Análise Descritiva	59
5.3	Equação gravitacional utilizada	63
6	ANÁLISE INFERENCIAL	65
6.1	Modelos Lineares Generalizados: PPML e GPML	65
6.2	Modelagem GAMLSS: seleção de distribuições contínuas e inflacionadas em zero	70
6.3	Comparação e Diagnóstico dos Modelos Estimados	76
6.4	Interpretação dos Coeficientes do Modelo Final	79
6.4.1	<i>Efeito dos acordos regionais (rta)</i>	83
7	CONCLUSÃO E TRABALHOS FUTUROS	86
	REFERÊNCIAS	88
	APÊNDICE A – LISTA DE PAÍSES NA AMOSTRA (1997 -2024) . . .	94
	APÊNDICE B – POSSÍVEIS INADEQUAÇÕES EM MODELOS VIA WORM PLOT	95
	APÊNDICE C – INTERPRETAÇÃO DAS ELASTICIDADES	96

1 INTRODUÇÃO

1.1 Considerações Gerais

O comércio internacional desempenha um papel estratégico na dinâmica econômica global, influenciando o crescimento das economias, a alocação eficiente de recursos, a difusão tecnológica e a especialização produtiva. A intensificação das trocas comerciais ao longo das últimas décadas revelou padrões complexos de interdependência entre países, exigindo instrumentos teóricos, analíticos e empíricos capazes de capturar com precisão a complexidade dos determinantes desses fluxos entre nações.

Nesse contexto, o modelo gravitacional de comércio, introduzido por Tinbergen (1962), consolidou-se como uma das abordagens mais influentes na literatura econômica. Inspirado na Lei da Gravitação Universal de Newton, esse modelo postula que os fluxos comerciais entre dois países são diretamente proporcionais ao nível de produto (geralmente representado pelo Produto Interno Bruto (PIB)) e inversamente proporcionais à distância entre eles. Posteriormente, contribuições como as de Anderson (1979), Bergstrand (1985) e Anderson e Wincoop (2003) forneceram fundamentos microeconômicos mais robustos ao modelo, incorporando elementos como resistências multilaterais ao comércio, políticas comerciais e estruturas de preferências.

Apesar de sua ampla aplicação, a estimação empírica tradicional do modelo enfrenta limitações importantes. A prática comum de aplicar a transformação logarítmica à variável dependente implica na exclusão de observações de valor zero que representam pares de países que, por razões políticas, econômicas, institucionais ou geográficas, não realizaram comércio em determinados períodos. Como destacam Linders e Groot (2006), essa exclusão pode gerar vieses de seleção e subestimar o impacto de barreiras comerciais ou de determinantes estruturais que inibem o comércio. Trata-se não só de uma perda de dados, mas da omissão de informação substantiva.

A heterocedasticidade, ou seja, a variação não constante dos resíduos, constitui outro desafio frequente na modelagem de dados de comércio. Quando ignorada, pode distorcer estimativas e comprometer a validade das inferências estatísticas. Para lidar com esse problema, Silva e Tenreiro (2006) propuseram o uso do estimador *Poisson Pseudo Maximum Likelihood* (PPML), que dispensa a log-linearização da equação de comércio, permite a inclusão de fluxos nulos e é robusto à presença de heterocedasticidade. Essa abordagem rapidamente se consolidou como uma das ferramentas metodológicas centrais nos estudos empíricos sobre comércio

internacional.

No entanto, mesmo o PPML não resolve integralmente as limitações associadas à distribuição dos dados. Os fluxos comerciais frequentemente apresentam assimetrias, excessos de curtose e heterogeneidade na dispersão: características que desafiam os pressupostos dos modelos tradicionais, como o PPML. Além disso, produtos e setores distintos apresentam dinâmicas comerciais próprias, como demonstrado por Corrêa (2021), ao analisar as exportações de açúcar, ou por Gopinath *et al.* (2020), que aplicaram técnicas de aprendizado de máquina à previsão de fluxos agrícolas.

Diante desse cenário, os *Generalized Additive Models for Location, Scale and Shape* (GAMLSS), desenvolvidos por Rigby e Stasinopoulos (2005), emergem como alternativa metodológica promissora. Esses modelos permitem especificar, de forma flexível, distribuições para a variável dependente, incluindo distribuições assimétricas e ajustadas para zero, como a *Zero-Adjusted Gamma* (ZAGA) e *Zero-Inflated Poisson* (ZIP). Tal flexibilidade é especialmente relevante para dados positivos com presença de zeros e variância não constante: características típicas dos fluxos de comércio internacional. Portanto, investigar o desempenho dos GAMLSS no contexto do comércio internacional é não apenas oportuno, mas necessário.

No contexto brasileiro, a relevância dessas discussões adquire contornos específicos quando aplicada a produtos de base extrativista, como a cera de carnaúba. Líder mundial na produção e exportação desse insumo, o Brasil enfrenta desafios relacionados à volatilidade dos preços, à concentração de mercados e às barreiras institucionais ao comércio. Assim, compreender os determinantes econômicos e estruturais dos fluxos de exportação da cera vegetal torna-se fundamental tanto para a formulação de políticas comerciais quanto para o fortalecimento sustentável dessa cadeia produtiva.

1.2 O problema de pesquisa e sua relevância

A aplicação dos modelos gravitacionais tradicionais, mesmo com o uso de técnicas mais robustas como o estimador PPML, tem-se mostrado insuficiente para lidar simultaneamente com a presença de assimetria, valores nulos e heterocedasticidade nos dados. A exclusão de observações com fluxos comerciais nulos ou a adoção de transformações logarítmicas *ad hoc* pode comprometer a validade empírica das estimativas. Ademais, grande parte das análises empíricas negligencia a possibilidade de modelar múltiplos parâmetros da distribuição, como variância, assimetria e curtose, limitando-se à modelagem da média. Essa limitação reduz

a capacidade dos modelos em capturar a influência de variáveis explicativas que afetam não somente o valor esperado dos fluxos comerciais, mas também sua dispersão e forma distributiva.

1.3 Justificativa

A centralidade do tema reside na necessidade premente de aprimorar os métodos de estimativa ligados ao modelo gravitacional, visando representar, com maior precisão, os padrões observados no comércio internacional. Os fluxos comerciais não se distribuem de maneira uniforme entre as nações, com diversos entraves comerciais, institucionais e estruturais influenciando a intensidade e a consistência das trocas. Frequentemente, o comércio bilateral mostra-se irregular ou mesmo ausente, gerando séries com grande número de valores nulos ou próximos de zero.

Ignorar essas complexidades pode comprometer a qualidade das análises empíricas, levando a estimativas enviesadas e inapropriadas, com potenciais consequências ruins para a criação de políticas comerciais. Como apontado por Ribeiro (2023), o modelo gravitacional é um instrumento fundamental não só para compreender os padrões comerciais, mas também para conceber estratégias de inserção internacional.

Diante desse cenário, a utilização dos modelos GAMLSS surge como uma alternativa metodológica promissora, ao permitir a modelagem conjunta de múltiplos parâmetros da distribuição dos dados. A estrutura oferece adaptabilidade, lidando com assimetria, heterocedasticidade e excessos de zero, frequentes em dados de comércio. Além disso, a presente proposta está alinhada à crescente incorporação de métodos avançados na análise econômica, como aprendizado de máquina (Gopinath *et al.*, 2020) e outras técnicas de regressão não linear.

1.4 Objetivos

1.4.1 *Objetivo Geral*

Avaliar o desempenho dos modelos GAMLSS na modelagem das exportações brasileiras de cera vegetal, em comparação com modelos gravitacionais tradicionais baseados em Modelos Lineares Generalizados (MLG)s, oferecendo uma análise mais detalhada dos fluxos comerciais ao modelar não somente a média, mas também a variabilidade e a assimetria dos dados.

1.4.2 Objetivos Específicos

- (i) Apresentar os principais avanços metodológicos, desde os modelos tradicionais de Tinbergen até a incorporação de técnicas mais recentes, como o PPML e a integração com modelos flexíveis como o GAMLSS;
- (ii) Verificar a adequação dos modelos GAMLSS em contextos com fluxos comerciais nulos e distribuições assimétricas;
- (iii) Comparar os resultados obtidos com os modelos GAMLSS e MLG;
- (iv) Analisar a capacidade dos modelos GAMLSS de capturar efeitos relacionados à localização, escala e forma na distribuição dos dados de comércio;
- (v) Discutir as implicações empíricas dos resultados para subsidiar a formulação de políticas comerciais mais precisas e direcionadas.

1.5 Estrutura do Trabalho

Além da introdução, este trabalho está organizado em cinco capítulos. O Capítulo 2 apresenta as bases teóricas e metodológicas do comércio internacional. Inicia com uma breve contextualização da história do comércio exterior brasileiro e, em seguida, discute os fatores que influenciam o comércio internacional. Na sequência, são abordadas as principais formas de analisar os fluxos comerciais na prática, com foco na seleção de variáveis e na formulação do modelo gravitacional. O Capítulo 3 trata dos modelos gravitacionais e de sua estimação. São apresentados o modelo tradicional, o com Elasticidade de Substituição Constante (CES), o estimador PPML e os modelos GAMLSS, com destaque para a estrutura dos modelos aditivos generalizados para localização, escala e forma. Também são explicadas as técnicas de estimação, os métodos de diagnóstico, os critérios de escolha de modelos e a análise dos resíduos. O Capítulo 4 analisa o mercado internacional da cera de carnaúba. Essa seção aborda a geografia da produção, a trajetória das exportações, os principais mercados de destino e os obstáculos da cadeia produtiva. O Capítulo 5 apresenta as bases de dados empregadas, as variáveis consideradas e a equação gravitacional estimada, incluindo uma análise descritiva dos dados que subsidia a modelagem posterior. Por fim, o Capítulo 6 apresenta e discute as análises inferenciais, comparando os Modelos Lineares Generalizados (Quasi-Poisson e Quasi-Gama) com os GAMLSS, incluindo distribuições inflacionadas e ajustadas para zero. Nesse capítulo, discute-se ainda a seleção do modelo final, os critérios de ajuste utilizados e a interpretação dos

coeficientes estimados.

2 BASES TEÓRICAS E METODOLÓGICAS

Compreender o comércio internacional exige analisar os vínculos que unem economias distintas e, muitas vezes, desiguais. Não se trata apenas de quantificar trocas entre nações, mas de captar as razões pelas quais os países se aproximam ou se afastam no cenário econômico global. Neste capítulo, parte-se de uma abordagem histórica e institucional para, em seguida, adentrar os fundamentos teóricos e metodológicos que explicam os fluxos comerciais. A intenção é mostrar que o comércio exterior do Brasil, em especial no setor agrícola e extrativista, decorre de decisões políticas, interesses geopolíticos e estratégias econômicas construídas ao longo do tempo (Krugman *et al.*, 2015).

A cera de carnaúba, ainda que modesta em termos de volume comparado a outras commodities, revela esse entrelaçamento entre território, política e mercado internacional. Por isso, retomar os marcos teóricos e metodológicos do comércio é essencial não só para justificar a escolha do modelo gravitacional neste trabalho, mas também para reconhecer que, por trás das variáveis e dos modelos, existem trajetórias, conflitos distributivos e decisões que moldam a realidade econômica.

2.1 O Comércio Internacional Brasileiro

O comércio internacional brasileiro de produtos agrícolas apresenta uma trajetória marcada por transformações, refletindo tanto mudanças internas na estrutura econômica e política do país quanto alterações no cenário econômico global. Desde o período colonial, o Brasil tem sido um importante *player* no mercado agrícola mundial, inicialmente com o comércio de cana-de-açúcar e, posteriormente, café, que se tornou o principal produto de exportação no século XIX e início do século XX (Moreira, 2023).

No decorrer do século XX, a diversificação dos produtos agrícolas exportados pelo Brasil ampliou-se consideravelmente. A soja, introduzida como uma cultura de rotação para recuperar solos desgastados pelo cultivo contínuo de café, ganhou escala com ganhos de produtividade e transformou-se em um dos principais produtos de exportação a partir da segunda metade do século. Paralelamente, o Brasil expandiu sua pauta exportadora para incluir produtos como carne bovina, suína, aves e uma variedade de frutas tropicais, posicionando-se como um fornecedor de commodities agrícolas no mercado internacional (Duarte, 2019).

O processo de expansão foi impulsionado por uma série de políticas governamentais

voltadas para o fortalecimento do agronegócio, incluindo incentivos fiscais¹, investimentos em infraestrutura logística e fomento à pesquisa agrícola por meio de instituições como a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Essas medidas ampliaram a capacidade produtiva do setor, melhoraram a qualidade dos produtos agrícolas e aumentaram a competitividade do Brasil no mercado internacional (Gonçalves, 2021).

A entrada do Brasil em blocos comerciais como o Mercado Comum do Sul (MERCOSUL) e a negociação de acordos de livre comércio com outras regiões também auxiliou na expansão do comércio agrícola. Ao mesmo tempo, reduziram barreiras tarifárias e não tarifárias, facilitando um maior fluxo de produtos agrícolas brasileiros para mercados internacionais (Nascimento e Aguiar, 2024).

No entanto, o comércio agrícola brasileiro também enfrentou desafios, incluindo disputas comerciais internacionais, como as barreiras impostas por questões sanitárias e fitossanitárias, e críticas relativas ao impacto ambiental da expansão agrícola, especialmente no que tange ao desmatamento da Amazônia. Essas questões destacam a complexidade das relações comerciais internacionais e a necessidade de uma abordagem equilibrada que considere tanto o desenvolvimento econômico quanto a sustentabilidade (Gonçalves, 2021).

A região Nordeste do Brasil, com sua diversidade climática e geográfica, destaca-se como um importante centro de produção agrícola, contribuindo para a pauta de exportação do país com uma variedade de produtos distintos. A cera de carnaúba, exclusiva do Nordeste brasileiro, é amplamente utilizada em uma variedade de indústrias internacionais, incluindo as de cosméticos, automotivas e de alimentos, devido às suas propriedades únicas de resistência ao calor e à água. A exportação de cera de carnaúba contribui para a economia local e coloca o Brasil como líder global neste segmento de mercado (Carvalho e Gomes, 2009).

A estrutura econômica dos países parceiros comerciais do Brasil exerce influência sobre os padrões de comércio internacional. Estes países, variando amplamente em termos de desenvolvimento econômico, capacidade industrial e estabilidade política, interagem com o Brasil por meio de um complexo tecido de relações comerciais que reflete suas respectivas necessidades e ofertas econômicas (Gonçalves, 2021).

Nos países desenvolvidos, como os Estados Unidos e os membros da União Europeia, a estrutura econômica é caracterizada por economias altamente diversificadas, com setores de

¹ Os incentivos fiscais incluíram, por exemplo, isenções e reduções tributárias para exportação de commodities, além de linhas de crédito subsidiadas por instituições como o Banco Nacional de Desenvolvimento Econômico e Social (BNDES).

serviços dominantes e indústrias avançadas. Esses países tendem a importar do Brasil produtos agrícolas e matérias-primas importantes para suas cadeias de produção, ao mesmo tempo que exportam para o Brasil bens de alto valor agregado, como tecnologia e equipamentos industriais. A demanda nesses países por produtos agrícolas e naturais, frequentemente direcionada por padrões de consumo que valorizam a sustentabilidade e a qualidade, molda as exportações brasileiras.

Por outro lado, em países em desenvolvimento, especialmente aqueles localizados na África e na América Latina, a estrutura econômica pode ser menos diversificada, frequentemente centrada em torno de poucos setores, como agricultura, mineração e manufatura leve. Nestes mercados, o Brasil serve como um fornecedor de produtos similares, competindo com produtores locais e como um importante exportador de bens manufaturados e tecnologia intermediária. Esses países podem depender mais de importações de alimentos básicos e bens de consumo, áreas nas quais o Brasil possui vantagens competitivas (Oliveira, 2021).

A estrutura econômica dos países parceiros também é moldada por fatores como políticas governamentais, acordos comerciais e a participação em blocos econômicos. Países com políticas abertas ao comércio tendem a facilitar maiores fluxos comerciais com o Brasil, enquanto aqueles com regimes mais protecionistas podem impor barreiras que alteram os termos e o volume de comércio. Os acordos comerciais, como os negociados através do MERCOSUL ou entre o Brasil e a União Europeia, procuram diminuir estas barreiras, criando um ambiente mais favorável ao intercâmbio de bens e serviços (Duarte, 2019).

Países com economias estáveis e previsíveis são parceiros comerciais mais confiáveis, proporcionando oportunidades de longo prazo para o comércio. Inversamente, países com instabilidades frequentes podem apresentar riscos para acordos comerciais, afetando a continuidade e a segurança das trocas comerciais (Moreira, 2023).

As políticas comerciais e os acordos bilaterais, estrategicamente desenvolvidos e implementados, visam promover o comércio exterior e proteger setores internos sensíveis, equilibrando interesses econômicos, políticos e sociais (Gonçalves, 2021). No contexto brasileiro, essas políticas foram moldadas por uma série de reformas e iniciativas que buscam expandir o acesso a mercados externos para produtos brasileiros, enquanto gerenciam as importações de maneira a não prejudicar as indústrias locais. Essas políticas incluem tarifas alfandegárias², quotas

² O Brasil aplica tarifas *ad valorem*, sendo taxas cobradas como uma porcentagem do valor aduaneiro do produto importado (Sistema Integrado de Comércio Exterior (Siscomex), 2025).

de importação, subsídios para exportação e medidas de defesa comercial, como *antidumping* e compensatórias, empregadas para combater práticas comerciais desleais.

Enquanto isso, o Brasil tem se engajado ativamente na negociação de acordos bilaterais e multilaterais que facilitam o comércio e a integração econômica. Um exemplo é o MERCOSUL, o qual é um bloco econômico que inclui Argentina, Brasil, Paraguai e Uruguai, com Venezuela suspensa desde 2016. Este bloco eliminou a maioria das tarifas entre os estados-membros e estabeleceu uma tarifa externa comum para regular as importações de países externos ao bloco (Oliveira, 2021).

O Brasil tem explorado acordos de livre comércio com outras regiões e países, buscando diversificar seus mercados de exportação e reduzir sua dependência de mercados tradicionais. Acordos recentes incluem negociações com a União Europeia, que prometem abrir mercados para produtos agrícolas e manufaturados, sujeitos à ratificação e implementação das políticas acordadas (Moreira, 2023). Essas políticas e acordos são formulados e negociados com base em considerações econômicas e considerando fatores políticos e diplomáticos, eles necessitam de um equilíbrio delicado entre abrir mercados externos para empresas brasileiras e proteger a economia local de competições externas potencialmente devastadoras. Esta dualidade é particularmente evidente em setores onde o Brasil é simultaneamente um grande produtor e consumidor, como na agricultura e na manufatura (Duarte, 2019).

Portanto, o dinamismo deste contexto exige uma constante reavaliação das políticas para assegurar que permaneçam relevantes e eficazes no fomento ao desenvolvimento econômico sustentável do Brasil. As barreiras tarifárias e não tarifárias são instrumentos regulatórios que os países utilizam para controlar o volume e a natureza das mercadorias que atravessam suas fronteiras. Essas barreiras têm um impacto substancial no comércio internacional, influenciando tanto a competitividade das exportações quanto a disponibilidade de importações no mercado interno (Nascimento e Aguiar, 2024).

Barreiras tarifárias, como tarifas de importação e exportação, são taxas impostas pelo governo sobre bens que entram ou saem do país. Essas tarifas visam proteger os produtores nacionais de competição externa, aumentando o custo de produtos importados para torná-los menos atraentes em comparação com os produtos locais. Contudo, no longo prazo, tais práticas podem estagnar a inovação e eficiência no país, uma vez que a falta de competição externa reduz os incentivos para melhorias.

Além disso, barreiras comerciais podem provocar tensões entre países, levando a

disputas comerciais que são frequentemente mediadas por organizações internacionais como a Organização Mundial do Comércio (OMC). A resolução de tais disputas vai encontrar um equilíbrio justo entre a proteção da indústria doméstica e a promoção de um ambiente de comércio livre e justo em 2021.

2.2 Determinantes do Comércio Internacional

2.2.1 Métodos de Coleta e Análise: Da Seleção à Modelagem

A seleção de dados e fontes, em função de sua qualidade, relevância e confiabilidade, determina a precisão dos resultados obtidos e a validade das conclusões derivadas. Neste contexto, deve-se exercer rigor na escolha de bases de dados e metodologias de coleta de informações para garantir que suas análises reflitam adequadamente os fenômenos estudados (Castelo e Marinho, 2021).

A seleção de dados para estudos do comércio internacional envolve identificar conjuntos de dados que proporcionem informações abrangentes sobre fluxos comerciais, tarifas, barreiras não tarifárias e outros indicadores econômicos relevantes. Dados primários podem ser coletados diretamente por meio de *surveys* e entrevistas com *stakeholders*-chave, como empresas exportadoras, agências governamentais e organizações internacionais. Essas informações primárias são vitais para entender nuances e detalhes que dados secundários não podem fornecer (Sebok, 2021).

Contudo, a maioria dos estudos depende de dados secundários, acessíveis mediante bases de dados públicas e privadas. Organizações como o Instituto Brasileiro de Geografia e Estatística (IBGE), o Banco Mundial, a OMC e a Comissão Econômica para a América Latina e o Caribe (CEPAL) fornecem extensos repositórios de dados econômicos e comerciais que visam a análise de tendências de comércio e avaliação de políticas econômicas. Estudos longitudinais, que analisam dados ao longo de um período extenso, requerem acesso a séries históricas consistentes e confiáveis para permitir uma análise de tendências e a identificação de padrões ao longo do tempo. A consistência temporal assegura que mudanças nos padrões comerciais possam ser atribuídas a eventos econômicos ou políticas específicas, em vez de serem artefatos de alterações na metodologia de coleta de dados. Assim, os dados precisam ser verificados quanto à origem e à precisão, já que erros de coleta, vieses na seleção de amostras e metodologias de reporte podem distorcer a análise e levar a conclusões errôneas. Nesse sentido, a validação cruzada de

informações de várias fontes é recomendada para assegurar a integridade dos dados (Pereira *et al.*, 1999).

Os métodos estatísticos para análise de dados representam uma faceta da pesquisa econômica, proporcionando ferramentas necessárias para a interpretação e compreensão de complexos conjuntos de dados. No contexto do comércio internacional e da análise econômica, a aplicação desses métodos permite aos pesquisadores extrair padrões, testar hipóteses e realizar inferências sobre as relações entre variáveis econômicas. A estatística descritiva é frequentemente o ponto de partida nas análises, oferecendo um panorama inicial através de médias, modas, medianas, variâncias e desvios padrão. Essas medidas fornecem uma compreensão elementar das características dos dados, como a distribuição central e a dispersão. Além disso, gráficos como histogramas, gráficos de caixa e diagramas de dispersão são utilizados para visualizar a distribuição dos dados e identificar possíveis *outliers* ou anomalias que poderiam influenciar os resultados das análises mais complexas (Sebok, 2021).

Para além da descrição básica, a inferência estatística é utilizada para testar a robustez das relações observadas entre as variáveis. Testes de hipóteses, intervalos de confiança e Análise de Variância (ANOVA) são métodos comuns que ajudam a determinar se os padrões observados nos dados da amostra podem ser generalizados para uma população maior.

Os modelos de regressão representam outra ferramenta estatística que permite estimar as relações quantitativas entre variáveis dependentes e independentes. Modelos de regressão linear, regressão logística e modelos de séries temporais são amplamente utilizados para modelar e prever comportamentos econômicos, ajustando-se para fatores confundidores e medindo a força e a significância das relações entre as variáveis (Silva e Hernández, 2020). Além disso, a econometria avança sobre os métodos estatísticos tradicionais, integrando teoria econômica com técnicas estatísticas para modelar economicamente relações complexas. Métodos como modelos de correção de erro, cointegração e Modelo Autorregressivo de Médias Móveis (ARMA) são particularmente úteis em análises de dados econômicos, onde as relações entre as variáveis são frequentemente influenciadas por fatores de *lag* temporal e outras dinâmicas de mercado.

A construção de um modelo gravitacional especificamente adaptado para a análise do comércio de um produto requer uma abordagem meticulosa que considere as características únicas deste produto e as dinâmicas específicas de seu mercado. Tradicionalmente utilizado para prever os fluxos comerciais com base no tamanho econômico dos países e na distância entre eles, deve ser ajustado para incorporar as particularidades do comércio de cera do produto, que não

são totalmente capturadas pelos determinantes econômicos gerais (Portes *et al.*, 2019).

O PIB dos países importadores e exportadores continua sendo uma variável, refletindo a capacidade econômica geral e a demanda potencial, utilizada em diversos produtos industriais e de consumo. A distância geográfica, um *proxy* para os custos de transporte e barreiras culturais e linguísticas, também é fundamental. Contudo, variáveis adicionais relacionadas à especificidade do produto, como a existência de regulamentos ambientais e padrões de qualidade nos países importadores, que podem afetar o comércio deste produto, podem ser incorporadas ao modelo (Pereira, 2024).

Os acordos comerciais e tarifas preferenciais que podem existir entre o Brasil e os países parceiros podem reduzir os custos de comércio e criar um ambiente mais favorável para as exportações. As barreiras não tarifárias, como certificações ecológicas e restrições de uso baseadas em considerações ambientais ou de saúde, também podem ser meticulosamente modeladas, dado que estas podem influenciar fortemente a demanda internacional pelo produto (Castelo e Marinho, 2021).

A metodologia para estimar o modelo envolve a aplicação de técnicas econométricas robustas para garantir que as inferências sobre as relações causais sejam válidas e confiáveis. A análise de sensibilidade e a validação dos modelos através de dados históricos são práticas padrão para testar a robustez das previsões do modelo (Sebok, 2021).

Os testes de sensibilidade e robustez são importantes para assegurar que as conclusões derivadas do modelo sejam consistentes sob diferentes condições e resilientes nas suposições e parâmetros do modelo. Para o modelo gravitacional aplicado ao comércio internacional, tais testes são indispensáveis para garantir a confiabilidade das previsões e das análises de políticas baseadas nesses resultados (Pereira, 2024).

A sensibilidade do modelo é testada alterando-se sistematicamente os valores das variáveis de entrada e observando as mudanças resultantes nas previsões de saída, ajudando a identificar quais variáveis têm o maior impacto nas conclusões do modelo e quão sensíveis são as previsões a pequenas alterações nessas variáveis. Por exemplo, pode-se ajustar o PIB dos países importadores ou as tarifas aplicadas sobre a cera vegetal para examinar como essas modificações afetam o volume estimado de comércio e essa abordagem destaca as variáveis mais críticas e ajuda a identificar possíveis pontos de fragilidade no modelo que podem requerer atenção adicional (Portes *et al.*, 2019).

Os testes de robustez, por outro lado, envolvem utilizar diferentes conjuntos de dados

ou métodos alternativos para estimar o mesmo modelo, para verificar se as conclusões são consistentes independentemente das fontes de dados ou das técnicas econométricas empregadas. Por exemplo, a utilização de dados de diferentes anos ou de fontes alternativas para validar as estimativas do modelo pode revelar a estabilidade das relações modeladas ao longo do tempo e sob diferentes condições econômicas (Silva e Hernández, 2020).

Os resultados fornecem uma base sólida para recomendações de políticas e estratégias comerciais e orientam melhorias contínuas no modelo. Identificar áreas onde o modelo é particularmente sensível a certas variáveis pode levar ao aprimoramento das técnicas de coleta de dados ou à revisão das metodologias analíticas utilizadas (Pereira, 2024). A análise comparativa com dados históricos é uma abordagem metodológica na avaliação da eficácia e precisão de modelos econômicos, incluindo aqueles desenvolvidos para o comércio de produtos específicos como a cera vegetal.

Na construção de um modelo gravitacional para a cera vegetal, a análise comparativa envolve o confronto das previsões do modelo com os fluxos comerciais reais registrados em anos anteriores. Este procedimento começa com a compilação de dados comerciais históricos sobre a exportação de cera vegetal, incluindo volumes de comércio, preços, mercados principais, e barreiras tarifárias e não tarifárias enfrentadas. Esses dados são então comparados com os resultados gerados pelo modelo para verificar a precisão das estimativas (Pereira *et al.*, 1999).

Utilizar dados históricos também ajuda a identificar padrões e anomalias nos fluxos comerciais que podem não ser imediatamente evidentes. Uma análise pode revelar ciclos econômicos ou efeitos sazonais que influenciam o comércio de cera vegetal, como variações na demanda devido a tendências de moda ou tecnologia em indústrias consumidoras como a automotiva e a de cosméticos. Alterações nas políticas comerciais, acordos internacionais ou crises econômicas globais podem ter efeitos sobre o comércio de produtos como a cera de carnaúba. Ao comparar os períodos antes e depois desses eventos nos dados históricos, os pesquisadores podem avaliar como tais eventos alteram as condições de mercado e ajustar o modelo para antecipar ou responder a essas mudanças (Portes *et al.*, 2019).

2.2.2 Variáveis Mais Impactantes nos Fluxos Comerciais

O PIB, representando a magnitude econômica de um país, quando elevado, usualmente sugere uma economia robusta com recursos substanciais para produzir bens e serviços. Para as exportações, isso significa que o país consegue produzir quantidades importantes de bens

para a exportação e possui um mercado interno potencialmente grande, que pode servir como um amortecedor contra as flutuações do mercado externo. Além disso, países com economias maiores tendem a ter mais infraestrutura desenvolvida e tecnologia, facilitando a produção eficiente e a logística de exportação (Rinaldi *et al.*, 2021).

A premissa básica do modelo gravitacional é que quanto maior a distância entre dois países, maiores serão os custos de transporte, o que pode desencorajar o comércio bilateral e esses custos adicionais podem tornar os bens exportados menos competitivos devido ao aumento dos preços finais para os consumidores estrangeiros. A distância também pode refletir barreiras menos tangíveis, como diferenças culturais e linguísticas, que podem complicar as transações comerciais e as operações de mercado (Viegas *et al.*, 2023).

Entretanto, os efeitos da distância são mitigados ao longo do tempo por meio de avanços tecnológicos em transporte e comunicação, facilitando assim uma maior interação econômica global. A globalização e a integração econômica, como observado em blocos comerciais regionais e acordos de livre comércio, também ajudam a reduzir o impacto da distância física, promovendo um ambiente onde os produtos podem fluir mais livremente entre as fronteiras nacionais (Viana *et al.*, 2022).

A análise da interação entre o PIB e a distância geográfica em modelos gravitacionais proporciona uma visão detalhada das forças que moldam o comércio internacional. A compreensão dessas dinâmicas ajuda os formuladores de políticas que procuram promover o comércio externo ou proteger indústrias domésticas. Por exemplo, entender como a distância e a escala econômica influenciam o comércio e podem auxiliar na formulação de políticas que visam melhorar a infraestrutura logística, negociar acordos comerciais estratégicos ou implementar medidas que promovam a competitividade externa de indústrias específicas (Filho e Santana, 2020).

A cera vegetal, conhecida por suas propriedades duráveis e seu brilho natural, é amplamente utilizada em várias indústrias, incluindo automotiva, cosmética, alimentícia e farmacêutica. Como resultado, os padrões de comércio dessa *commodity* são influenciados por uma gama diversificada de fatores econômicos e regulatórios nos países importadores (Ataíde *et al.*, 2019).

Os Estados Unidos e os países da União Europeia, como Alemanha e França, estão entre os maiores importadores de cera vegetal. Esses mercados são caracterizados por suas economias avançadas com setores industriais sofisticados que demandam insumos de alta qualidade para a fabricação de uma variedade de produtos. Além disso, a presença de

regulamentos estritos relacionados à qualidade, segurança e sustentabilidade em tais regiões impulsiona a demanda por cera vegetal, valorizada por ser um produto natural e ecologicamente sustentável (Rinaldi *et al.*, 2021).

Na América do Norte, especialmente nos Estados Unidos, a indústria automotiva e de cuidados pessoais são os principais consumidores de cera vegetal. A cera é utilizada em polidores de carro e em vários cosméticos, onde suas propriedades de formação de filme e brilho são altamente valorizadas. E, a tendência crescente de produtos *eco-friendly* e orgânicos nos Estados Unidos (EUA) tem aumentado ainda mais a demanda por cera vegetal, percebida como uma alternativa sustentável às ceras sintéticas (GLOBAL GROWTH INSIGHTS, 2023).

Na Europa, a demanda é impulsionada tanto pelas indústrias automotiva e cosmética quanto pela indústria de alimentos, onde a cera vegetal é usada como agente de revestimento para confeitos e frutas, aproveitando sua característica comestível e não tóxica. Os regulamentos rigorosos da União Europeia sobre materiais seguros e naturais em produtos de consumo direcionam uma preferência clara por ingredientes como a cera vegetal. Na Ásia, especialmente países como Japão e Coreia do Sul, têm aumentado suas importações devido ao crescimento das indústrias que utilizam cera vegetal em aplicações variadas, desde componentes eletrônicos até produtos de beleza. A valorização de produtos naturais e a ascensão de normas ambientais mais estritas também contribuem para o crescimento da demanda na região asiática (Viana *et al.*, 2022).

Acordos comerciais, sejam eles bilaterais ou multilaterais, têm o potencial de moldar os padrões de comércio entre as nações ao modificar as condições sob as quais os bens e serviços cruzam as fronteiras. Estes geralmente visam reduzir ou eliminar tarifas, simplificar procedimentos alfandegários, limitar barreiras não tarifárias e, de modo mais amplo, facilitar o fluxo de comércio (Ataíde *et al.*, 2019). Ao remover ou reduzir as tarifas, os acordos comerciais diminuem diretamente o custo de exportar bens e serviços para os mercados parceiros, tornando os produtos de um país mais competitivos no exterior. Este efeito de redução de custos pode levar a um aumento nos volumes de exportação. Por exemplo, a eliminação de tarifas no âmbito do Acordo de Livre Comércio da América do Norte (NAFTA) impulsionou o comércio entre os EUA, Canadá e México, facilitando um ambiente no qual as indústrias poderiam exportar com custos mais baixos e previsibilidade aumentada (Rangel *et al.*, 2019).

Além de reduzir as tarifas, muitos acordos comerciais abordam barreiras não tarifárias, que podem incluir regulamentos sanitários e fitossanitários, restrições quantitativas e

normas técnicas. Ao alinhar regulamentos e padronizar procedimentos, os acordos reduzem a complexidade e o custo da conformidade para os exportadores, o que é especialmente relevante para os produtos que requerem conformidade regulatória rigorosa antes de poderem ser vendidos em mercados estrangeiros (Moreira, 2022).

Os acordos comerciais também promovem a integração econômica mais ampla, que pode incluir o comércio de bens e de serviços e capital. Esta integração pode estimular o investimento entre os países membros do acordo, à medida que as empresas buscam aproveitar as novas oportunidades de mercado abertas pela redução das barreiras comerciais. O investimento resultante pode, por sua vez, fortalecer os laços econômicos entre os países e promover ainda mais o comércio (Rinaldi *et al.*, 2021).

A relevância econômica das exportações de cera de carnaúba para o Brasil é substancial, refletindo-se tanto no plano econômico quanto no social e ambiental. Este produto natural, extraído das folhas da palmeira de carnaúba, predominante no Nordeste brasileiro, é altamente valorizado em mercados internacionais devido às suas propriedades únicas, que incluem resistência ao calor, impermeabilidade e um brilho distinto. Estas características tornam a cera de carnaúba um componente essencial em diversas indústrias, incluindo a automotiva, alimentícia, cosmética e farmacêutica (Moreira, 2022).

O país é o maior produtor mundial deste produto, dominando quase a totalidade do mercado global. A demanda constante por cera de carnaúba nos mercados internacionais contribui para a estabilidade econômica das regiões produtoras, especialmente no Nordeste, onde as opções de emprego podem ser limitadas e a economia depende frequentemente de produtos agrícolas e naturais (Rangel *et al.*, 2019).

Além de seu impacto direto na geração de receita, a exportação de cera de carnaúba também auxilia no desenvolvimento socioeconômico das comunidades locais. As operações de extração e processamento oferecem empregos para muitas famílias, muitas vezes em áreas rurais onde a pobreza é mais prevalente. A renda gerada por essas atividades pode melhorar o padrão de vida, aumentar o consumo e incentivar a economia local através do efeito multiplicador (Carvalho e Gomes, 2009).

Do ponto de vista ambiental, a cera de carnaúba é promovida como um produto sustentável, derivado de fontes renováveis. A extração da cera não requer o abate das árvores de carnaúba, permitindo a sustentabilidade da produção. Além disso, a crescente demanda por produtos naturais e ecologicamente corretos nos mercados internacionais reforça a imagem do

Brasil como um fornecedor de produtos sustentáveis, o que pode abrir portas para outras iniciativas ecológicas e aumentar a conscientização ambiental (Moreira, 2022). Entretanto, é preciso que as práticas de produção e exportação sejam continuamente monitoradas e regulamentadas para assegurar que os benefícios econômicos não ocorram às custas do meio ambiente ou do bem-estar das comunidades locais. A gestão responsável dessas exportações é vital para garantir que a cera de carnaúba continue sendo um recurso renovável e benéfico para o Brasil no longo prazo.

2.2.2.1 *Revisão e Evidências Empíricas sobre o Comércio*

Ao realizar essa comparação, os pesquisadores começam por revisar meticulosamente a literatura acadêmica para identificar estudos similares que tenham examinado os fatores influenciando o comércio internacional de produtos comparáveis ou que tenham utilizado abordagens metodológicas análogas. Essa revisão abrange artigos publicados em periódicos de renome, trabalhos apresentados em conferências acadêmicas, teses doutorais e relatórios de instituições econômicas internacionais. Entre os estudos fundamentais, Anderson (1979, 2011) fornece a base teórica para o modelo gravitacional. Outros estudos, como Bergstrand (1985, 1989), aprofundam as bases microeconômicas do modelo. No contexto da cera de carnaúba, pesquisas como Rangel *et al.* (2019) examinam a demanda externa do produto, enquanto Moreira (2022) discute as condições econômicas e regulamentações que afetam seu comércio.

Os coeficientes do modelo gravitacional, que quantificam a influência de variáveis como o PIB, a distância geográfica, as tarifas e as barreiras não tarifárias sobre os fluxos comerciais, são então comparados com os resultados desses estudos. Se a literatura sugere que um aumento nos países importadores geralmente resulta em um aumento nas importações de cera de carnaúba, espera-se que os coeficientes no modelo apresentem um sinal positivo e sejam estatisticamente importantes. A magnitude desses coeficientes também é analisada para assegurar que estejam em faixas razoáveis, conforme reportado por estudos prévios. Estudos anteriores frequentemente encontram que distâncias maiores suprimem o comércio devido aos custos de transporte mais altos. Se os coeficientes relacionados à distância no modelo atual contradizerem essa tendência geral, isso pode indicar a necessidade de revisão das especificações do modelo ou das técnicas de estimativa (Viana *et al.*, 2022).

Alterações na inclusão de variáveis de controle, na definição das variáveis dependentes e independentes, ou na metodologia de estimação podem afetar os coeficientes. Comparar essas diferenças ajuda a entender a robustez dos resultados e a identificar quais aspectos do modelo

são mais influenciados por decisões metodológicas (Rangel *et al.*, 2019). Na discussão sobre as variáveis que mais impactam o comércio, é preciso identificar e analisar aqueles fatores que exercem influência sobre os fluxos de comércio internacional. Essas variáveis definem a intensidade e a direção dos fluxos comerciais e oferecem determinantes para a formulação de políticas e estratégias econômicas. Dentre as variáveis mais impactantes, o PIB, a distância geográfica, as políticas comerciais, as condições econômicas e as regulamentações sanitárias e fitossanitárias são frequentemente destacadas (Moreira, 2022).

2.3 O Modelo Gravitacional no Comércio Internacional

Diante de um contexto tão plural, há inúmeras formas de compreender e analisar o comércio internacional, entre elas o modelo gravitacional. Esse modelo, cuja concepção se inspira na analogia com a lei da gravidade de Newton, é amplamente utilizado para explicar e prever os fluxos de comércio entre países. A metodologia estima o volume de comércio com base na massa econômica dos países envolvidos, geralmente medida pelo PIB, e na distância entre eles. A premissa básica é que, assim como grandes massas se atraem com maior força em um contexto físico, economias maiores tendem a ter um volume maior de comércio, enquanto distâncias maiores tendem a reduzi-lo (Sandambi, 2021).

Desde sua introdução, o modelo passou por várias reformulações e melhorias para abordar tanto as críticas quanto as limitações inerentes ao conceito original. Essas revisões incluem incorporar variáveis como tarifas alfandegárias, acordos comerciais, barreiras não tarifárias, e diferenças linguísticas e culturais, que também podem afetar o comércio bilateral (Bobato *et al.*, 2020). Tinbergen propôs o modelo em seu estudo econométrico sobre o padrão das relações entre países na ausência de barreiras comerciais, afirmando que “*os fluxos de comércio estão diretamente relacionados à dimensão econômica dos países envolvidos e inversamente proporcionais à distância entre eles*” (Tinbergen, 1962).

De acordo com Deardorff (1998), embora inicialmente o modelo gravitacional fosse somente intuitivo, ele ganhou amplo uso devido à fundamentação teórica rigorosa que recebeu posteriormente, bem como ao seu sucesso empírico na previsão dos fluxos de comércio bilateral de diversas commodities em diferentes situações. Além disso, destaca-se a flexibilidade do método para estimativas envolvendo uma ampla variedade de políticas comerciais, que entram como variáveis influenciando os custos de comércio.

A aplicação do modelo evoluiu de um mecanismo simples para um instrumento

sofisticado de análise econômica, abrangendo não apenas o comércio de bens, mas também o de serviços e os fluxos de capital. Essa expansão natural reflete sua crescente utilidade à medida que o ambiente econômico global se torna mais complexo e interconectado. A capacidade do modelo de se adaptar para incorporar novas dimensões e variáveis demonstra sua robustez e relevância contínua na análise dos padrões de comércio global (Mendes *et al.*, 2020).

O modelo ganhou maior qualidade estatística com a inclusão de efeitos fixos para controle dos termos de resistência multilateral, conforme indicado por Anderson e Wincoop (2003). Esses termos representam o efeito que a posição dos países exportadores e importadores no mercado global e sua conjuntura econômica têm sobre seu comércio bilateral. Em outras palavras, capturam os efeitos fixos específicos de fatores inerentes a cada país em cada período, permitindo controlar a heterogeneidade entre os países, além de mitigar o efeito de variáveis omitidas, os quais são não observáveis ou difíceis de mensurar, como os índices de resistência multilateral.

Essencialmente, o modelo serve como uma ferramenta preditiva, permitindo que formuladores de políticas antecipem os efeitos de mudanças nas políticas econômicas e nos acordos comerciais. Por exemplo, ele pode prever como um acordo de livre comércio pode intensificar o comércio entre seus membros ou como barreiras econômicas adicionais podem impactar negativamente o fluxo de bens e serviços. O contínuo refinamento e adaptação sublinham sua importância como um pilar da análise econômica contemporânea, importante para navegar na complexa rede de interações econômicas que define o cenário global atual (Corrêa, 2021).

O modelo gravitacional, uma ferramenta estimada na economia internacional, tem sido aplicado extensivamente para elucidar o volume e a dinâmica dos fluxos comerciais entre nações. As aplicações deste modelo abrangem desde o estudo de tratados de livre comércio até o impacto de políticas protecionistas, fornecendo uma base sólida para a compreensão dos determinantes do comércio global (Maraschin, 2020).

A flexibilidade também permite sua adaptação para explorar fluxos de comércio em setores específicos, proporcionando uma visão detalhada sobre como diferentes setores respondem a alterações nas políticas comerciais e econômicas. Por exemplo, estudos que utilizam o modelo têm demonstrado como barreiras não tarifárias, como normas sanitárias e fitossanitárias, afetam desproporcionalmente setores onde a qualidade e a conformidade são essenciais (Sandambi, 2021).

Por meio de sua capacidade de simular cenários futuros, o modelo auxilia na avaliação

de como mudanças propostas em políticas comerciais podem influenciar os fluxos comerciais, o que é especialmente relevante em negociações comerciais, onde entender as consequências de alterações tarifárias ou a introdução de novos acordos comerciais pode determinar o curso das discussões (Mendes *et al.*, 2020).

Sendo amplamente empregado para analisar o comércio internacional, distingue-se de outros modelos econômicos pelo seu foco específico e pela metodologia empregada. Comparativamente, enquanto emprega a analogia física da lei de Newton sobre a gravidade, outros modelos econômicos podem utilizar abordagens distintas para elucidar as dinâmicas comerciais entre países (Maraschin, 2020).

Um desses modelos é o modelo de Heckscher-Ohlin, que se fundamenta na teoria das vantagens comparativas baseadas nas diferenças de fatores de produção entre os países. Diferente do modelo que considera principalmente a massa econômica e a distância, o modelo de Heckscher-Ohlin analisa como a abundância relativa de fatores como capital e trabalho pode influenciar o comércio. Este modelo é particularmente útil para entender os padrões de comércio baseados na especialização, mas não aborda diretamente a influência de fatores como distância e barreiras tarifárias (Corrêa, 2021).

Outro modelo relevante é o modelo de Linder, que sugere que países com demandas internas similares tendem a comercializar mais entre si, presumindo que consumidores com rendimentos parecidos possuem preferências similares. Este complementa o modelo gravitacional ao focar na similaridade das preferências dos consumidores como um fator de impulso ao comércio, uma dimensão que não explora profundamente (Mendes *et al.*, 2020).

O modelo de Ricardo foca em como as diferenças na produtividade do trabalho entre países impulsionam o comércio internacional. A premissa é que mesmo um país com uma desvantagem absoluta em todos os campos pode se beneficiar do comércio se se especializar nos setores onde possui a menor desvantagem relativa. Enquanto o modelo de Ricardo fornece uma base robusta para entender o comércio sob uma perspectiva de produtividade, o modelo gravitacional oferece uma explicação mais direta baseada em tamanho e distância econômica (Ribeiro, 2023).

Cada modelo traz luz sobre diferentes aspectos do comércio, como a importância de fatores econômicos, geográficos, produtivos e preferenciais. No âmbito, diversas variáveis visam explicar os fluxos de comércio entre os países. Essas, selecionadas com base em sua influência teórica e empírica sobre o comércio, são cruciais para a elaboração de previsões precisas e

análises detalhadas dos padrões de comércio internacional (Corrêa, 2021). O PIB serve como um indicador da massa econômica de um país, pressupondo que economias maiores tenham uma capacidade maior de gerar e receber fluxos de comércio. Uma economia robusta produz uma gama maior de bens e serviços e possui uma demanda interna elevada, estimulando o comércio tanto de importações quanto de exportações (Maraschin, 2020).

Historicamente, a distância tem sido vista como uma barreira ao comércio, onde maiores distâncias implicam em custos logísticos e de transporte mais altos. No entanto, a distância no modelo gravitacional pode ser interpretada de maneira mais ampla, incluindo barreiras culturais e linguísticas, que também podem impedir o comércio. Por exemplo, países que compartilham uma língua comum ou têm laços históricos frequentemente experimentam volumes de comércio mais relevantes devido à menor “distância” cultural e comunicativa (Mendes *et al.*, 2020). Outras variáveis incorporadas incluem a população dos países, que indica o tamanho do mercado potencial, e variáveis *dummy* para acordos comerciais, como zonas de livre comércio ou uniões econômicas. Estas variáveis sugerem que países participantes de blocos comerciais tendem a ter um volume de comércio maior entre si devido à redução ou eliminação de tarifas e à facilitação de normas regulatórias.

Embora o modelo gravitacional seja uma ferramenta amplamente utilizada na análise do comércio internacional, ele não está isento de limitações e críticas que questionam sua aplicabilidade em certos contextos econômicos e sua capacidade de capturar todas as nuances do comércio global (Corrêa, 2021).

Uma das principais limitações do modelo gravitacional tradicional é a suposição de isotropicidade, que implica que o comércio entre dois países depende exclusivamente do tamanho de suas economias e da distância entre eles, assumindo que esses efeitos são homogêneos em todas as direções. Essa simplificação desconsidera a possibilidade de que os fluxos comerciais possam ser influenciados por direcionalidade específica, relações geopolíticas e características estruturais de cada país. Na prática, o comércio internacional é moldado por uma variedade de fatores dinâmicos, como barreiras tarifárias, acordos regionais, políticas econômicas e choques externos, incluindo pandemias e crises financeiras (Bobato *et al.*, 2020).

Além disso, críticos apontam que o modelo frequentemente falha em incorporar variáveis não diretamente quantificáveis, como a qualidade das instituições, transparência governamental e estabilidade política, que podem facilitar ou restringir o comércio. A ausência desses elementos pode resultar em estimativas enviesadas e na subestimação de barreiras não

tarifárias, levando a previsões imprecisas (Sandambi, 2021).

O modelo pode sofrer de problemas metodológicos relacionados à multicolinearidade entre variáveis independentes, particularmente quando variáveis como PIB e distância são fortemente correlacionadas com outras variáveis explicativas. Essa correlação pode distorcer os resultados de regressões e obscurecer o entendimento preciso dos fatores que verdadeiramente influenciam o comércio (Bobato *et al.*, 2020).

Embora o modelo gravitacional tradicional seja eficaz para descrever padrões comerciais em momentos específicos, ele pode não capturar adequadamente a evolução desses padrões ao longo do tempo. O comércio internacional é um fenômeno dinâmico, influenciado por inovações tecnológicas, variações nos custos de transporte e mudanças nas políticas comerciais, fatores que podem alterar significativamente as relações entre os países (Maraschin, 2020).

3 MODELOS GRAVITACIONAIS E MÉTODOS DE ESTIMAÇÃO

O modelo gravitacional, ao longo das últimas décadas, firmou-se como um dos principais instrumentos teóricos e empíricos para captar tais dinâmicas. Sua força reside tanto na simplicidade intuitiva quanto na flexibilidade para incorporar diferentes especificações econométricas. No entanto, como qualquer representação da realidade, sua eficácia depende das escolhas metodológicas que fundamentam a estimação. Este capítulo apresenta os principais desdobramentos do modelo gravitacional: do tradicional ao estrutural com base CES, passando pelos avanços em estimação robusta via PPML e em adição ao aqui proposto GAMLSS. Ao longo da exposição, buscou-se não somente apresentar as fórmulas, mas refletir sobre as decisões técnicas que traduzem a complexidade das relações comerciais em evidência empírica confiável.

3.1 Modelo de Gravidade Tradicional

Supondo uma massa de bens, trabalho ou outros fatores de produção fornecidos na origem i , Y_i , é atraída por uma massa de demanda por bens no destino j , Y_j , mas o fluxo potencial é reduzido pela distância entre eles, d_{ij} . Aplicando estritamente essa analogia, a forma elementar do modelo gravitacional é dada por:

$$T_{ij} = \frac{Y_i Y_j}{d_{ij}^\beta},$$

em que T_{ij} representa o fluxo comercial entre os países i e j , e β é um parâmetro que capta o efeito da distância sobre o comércio.

Tinbergen (1962) foi o primeiro a aplicar essa abordagem para explicar os fluxos comerciais. Na formulação original, assume-se que os expoentes das variáveis de massa são iguais a 1 e o expoente da distância bilateral é igual a -2 . Na prática, estimações empíricas frequentemente resultam em coeficientes próximos de 1 para as variáveis de massa e em torno de -1 para a distância (Anderson, 2011).

A precisão do modelo gravitacional pode ser aprimorada com a inclusão de variáveis adicionais que capturam fricções comerciais, como fronteiras políticas e idioma comum (Bobato *et al.*, 2020). Entretanto, apenas essas fricções bilaterais podem não ser suficientes para explicar completamente as barreiras ao comércio. Os fluxos comerciais entre i e j também são influenciados pela acessibilidade de i a outros mercados e pela concorrência enfrentada por j de outros fornecedores. Para capturar esse efeito, a literatura introduziu índices de afastamento, que

medem a distância média efetiva de um país em relação aos seus parceiros comerciais (Portes *et al.*, 2019).

3.2 Modelo CES

A consolidação teórica do modelo gravitacional ocorreu com Anderson (1979), que fundamentou o modelo a partir de preferências do tipo CES e bens diferenciados por origem. Extensões importantes, como as de Bergstrand (1989) e Deardorff (1998), incorporaram elementos como competição monopolística e especialização no estilo Heckscher-Ohlin, sem abandonar a estrutura CES.

Com base nesses desenvolvimentos, Anderson e Wincoop (2003) propuseram uma formulação mais avançada do modelo, decompondo a resistência comercial em três componentes intuitivos: (i) barreira comercial bilateral entre i e j ; (ii) resistência ao comércio com todas as regiões; e (iii) resistência de j ao comércio com todas as regiões. Assume-se que cada região é especializada na produção de um único bem, cuja oferta é fixa. Além disso, os consumidores possuem preferências idênticas e homotéticas, representadas por uma função CES:

$$U_j = \left(\sum_i \beta_i^{\frac{(1-\sigma)}{\sigma}} c_{ij}^{\frac{(\sigma-1)}{\sigma}} \right)^{\frac{\sigma}{(\sigma-1)}}, \quad (3.1)$$

sujeita à restrição orçamentária:

$$\sum_i p_{ij} c_{ij} = y_j, \quad (3.2)$$

em que σ é a elasticidade de substituição entre os bens, β_i é um parâmetro de distribuição positivo, c_{ij} representa a quantidade do bem produzido na região i e consumido em j , y_j é a renda nominal dos residentes da região j e p_{ij} é o preço pago por j por bens oriundos de i , sendo $p_{ij} = p_i t_{ij}$, com t_{ij} representando o fator de custo de comércio. Assume-se que os custos comerciais são suportados pelo exportador. O valor das exportações de i para j é $x_{ij} = p_{ij} c_{ij}$, e a renda total da região i é, portanto, $y_i = \sum_j x_{ij}$.

A demanda por bens da região i por consumidores da região j , que maximizam a utilidade em (3.1) sujeita à restrição (3.2), é dada por:

$$x_{ij} = \left(\frac{\beta_i p_i t_{ij}}{P_j} \right)^{1-\sigma} y_j, \quad (3.3)$$

em que P_j é o índice de preços da região j , dado por:

$$P_j = \left[\sum_i (\beta_i p_i t_{ij})^{1-\sigma} \right]^{\frac{1}{1-\sigma}}. \quad (3.4)$$

A estrutura de equilíbrio geral do modelo impõe a condição de livre mercado, que implica em:

$$y_i = \sum_j x_{ij} = \sum_j \left(\frac{\beta_i p_i t_{ij}}{P_j} \right)^{1-\sigma} y_j = (\beta_i p_i)^{1-\sigma} \sum_j \left(\frac{t_{ij}}{P_j} \right)^{1-\sigma} y_j, \quad \forall i. \quad (3.5)$$

Definindo a renda mundial por $y^W \equiv \sum_j y_j$ e as parcelas de renda por $\theta_j \equiv y_j/y^W$, obtém-se:

$$x_{ij} = \frac{y_i y_j}{y^W} \left(\frac{t_{ij}}{\Pi_i P_j} \right)^{1-\sigma},$$

onde,

$$\Pi_i \equiv \left(\sum_j \left(\frac{t_{ij}}{P_j} \right)^{1-\sigma} \theta_j \right)^{\frac{1}{1-\sigma}}. \quad (3.6)$$

Substituindo os preços escalonados de equilíbrio em (3.4), obtém-se

$$P_j = \left(\sum_i \left(\frac{t_{ij}}{\Pi_i} \right)^{1-\sigma} \theta_i \right)^{\frac{1}{(1-\sigma)}}. \quad (3.7)$$

Juntos, (3.6) e (3.7) podem ser resolvidas para todos os Π_i e P_j em termos de participação na renda θ_i , barreiras comerciais bilaterais t_{ij} e de σ . Assumindo custos de comércio simétricos ($t_{ij} = t_{ji}$), tornando facilmente verificado que uma solução para (3.6) e (3.7) é ($\Pi_i = P_i$) com:

$$P_j^{1-\sigma} = \sum_i P_i^{r-1} \theta_{ij} t_j^{1-\sigma}, \quad \forall j.$$

Os autores chegam a uma solução implícita¹ em que a equação de gravidade:

$$x_{ij} = \frac{y_i y_j}{y^W} \left(\frac{t_{ij}}{P_i P_j} \right)^{1-\sigma}. \quad (3.8)$$

A Equação 3.8 sintetiza os fundamentos microeconômicos do modelo gravitacional a partir de uma estrutura de preferências do tipo CES, conforme proposta por Anderson (1979). A teoria implica que:

$$\ln x_{ij} = \kappa + \ln y_i + \ln y_j - \ln y^W + (1 - \sigma) \ln d_{ij} - (1 - \sigma) \ln P_i - (1 - \sigma) \ln P_j, \quad (3.9)$$

¹ Sendo a solução para os índices de preços como uma função de todas as barreiras comerciais bilaterais e participações de renda.

onde κ é uma constante.

Essa formulação confere uma base teórica sólida para o modelo empírico de gravidade, ao incorporar a especialização produtiva, os custos de comércio e as preferências dos consumidores consistentemente. Assim, o uso do modelo gravitacional como ferramenta empírica se justifica não apenas por sua adequação estatística, mas também por sua coerência com os princípios da teoria do comércio internacional. Essa conexão entre teoria e prática reforça sua ampla adoção nos estudos aplicados contemporâneos.

3.3 Modelo PPML

Na literatura de comércio internacional, o estimador Mínimos Quadrados Ordinários (MQO) log-linearizado geralmente fornece estimativas inconsistentes para os coeficientes β . Para corrigir problemas de vieses e heterocedasticidade em modelos log-lineares, o estimador de PPML tem se mostrado uma alternativa robusta ao modelo log-linear padrão (Silva e Tenreyro, 2006).

Em sua forma não estocástica, a relação entre o modelo multiplicativo de elasticidade constante e sua formulação aditiva logarítmica linear é trivial. No entanto, as relações econômicas não seguem leis físicas exatas, e tudo o que se pode esperar é que se sustentem, em média. Essa questão é discutida em detalhes por Silva e Tenreyro (2006).

A teoria econômica sugere que y e x estão ligados por um modelo de elasticidade constante da forma:

$$y_i = \exp(x_i\beta), \quad (3.10)$$

em que $\exp(x_i\beta)$ é interpretada como a expectativa condicional de y_i dado x , ou seja, $\mathbb{E}[y_i | x]$. No entanto, a transformação logarítmica da variável dependente pode introduzir um viés na estimativa de $\mathbb{E}[y_i | X = x]$, caso o termo residual seja heterocedástico e não normalmente distribuído.

Como discutido por Silva e Tenreyro (2006), a estimação o modelo log-linear:

$$\ln y_i = x_i\beta + \varepsilon_i, \quad (3.11)$$

via MQO pode ser inadequada por várias razões. Primeiramente, a transformação logarítmica não é viável para $y_i = 0$, pois $\ln(0) = -\infty$. Além disso, mesmo quando todas as observações de y_i são positivas, o valor esperado do resíduo no modelo log-linear dependerá das covariáveis.

Portanto, se reescrevermos o modelo original na forma multiplicativa:

$$y_i = \exp(x_i\beta)\eta_i, \quad (3.12)$$

onde $\eta_i = 1 + \varepsilon_i/\exp(x_i\beta)$ e $\mathbb{E}[\eta_i | x] = 1$. Supondo que y_i é positivo, o modelo pode ser tornado linear nos parâmetros tomando logaritmos de ambos os lados da equação, levando a

$$\ln y_i = x_i\beta + \ln \eta_i. \quad (3.13)$$

Para obter um estimador consistente dos parâmetros de declive estimando em (3.13) por MQO, é necessário que $\mathbb{E}[\ln \eta_i | x]$ não dependa de x_i . Além disso, a estimativa consistente do intercepto também requer que $\mathbb{E}[\ln \eta_i | x] = 0$. Como $\eta_i = 1 + \varepsilon_i/\exp(x_i\beta)$, essa condição é atendida somente se ε_i puder ser escrito como $\varepsilon_i = \exp(x_i\beta)v_i$, onde v_i é uma variável aleatória estatisticamente independente de x_i . Nesse caso, $\eta_i = 1 + v_i$ e, portanto, implica que $\mathbb{E}[\ln \eta_i | x]$ é constante e estatisticamente independente de x_i . Dessa forma, somente sob condições muito específicas no termo de erro, a representação log-linear do modelo de elasticidade constante é útil como um dispositivo para estimar os parâmetros de interesse.

Além disso, como y_i é não negativo, quando $\mathbb{E}[y_i | x]$ se aproxima de 0, a probabilidade de y_i ser positivo também deve se aproximar de 0. Isso implica que $V[y_i | x]$, a variância condicional, tende a desaparecer à medida que $\mathbb{E}[y_i | x]$ se aproxima de zero. No entanto, é possível observar grandes desvios da média condicional (levando assim a uma maior dispersão) quando o valor esperado de y_i está muito longe de seu limite inferior. O termo residual ε_i é provavelmente heteroscedástico e sua variância dependerá de (3.10).

Como resultado, a regressão de $\ln y_i$ em x_i por MQO levará a estimativas inconsistentes de β . A principal razão para a heteroscedasticidade afetar a consistência de um estimador é que a transformação não linear da variável dependente altera as propriedades do termo residual. A menos que fortes suposições sejam impostas à forma de distribuição, recuperar informações sobre a expectativa de y_i da média condicional de $\ln y_i$ pode não ser possível, pois o logaritmo do termo residual é correlacionado com os regressores. Em geral, mesmo que todas as observações em y_i sejam positivas, estimar β do modelo log linear produzirá estimadores inconsistentes e heteroscedasticidade entre os regressores.

3.3.1 Estimadores não lineares

Tradicionalmente, esse modelo é log-linearizado para fins de estimação, produzindo (3.11), a qual é estimada com o MQO. Contudo, essa abordagem apresenta duas limitações

importantes: (i) a impossibilidade de lidar com observações de comércio nulo (*zero trade flows*), uma vez que o logaritmo de zero é indefinido, e (ii) a possibilidade de viés nos estimadores quando a variância do termo de erro é correlacionada com os regressores, o que frequentemente ocorre quando a variância do comércio cresce com seu nível médio, uma situação típica nas amostras de comércio internacional.

Em Silva e Tenreyro (2006), a equação do modelo multiplicativo pode ser escrita como a seguinte função exponencial, (3.13), em que η_{ij} é uma variável aleatória log-normal com média 1 e variância σ_i^2 .

A estimação do modelo gravitacional também pode ser conduzida via Mínimos Quadrados Não Lineares (NLS), com a função:

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n [y_i - \exp(x_i^{\top} \beta)]^2,$$

cuja condição de primeira ordem é:

$$\sum_{i=1}^n [y_i - \exp(x_i^{\top} \hat{\beta})] \exp(x_i^{\top} \hat{\beta}) x_i = 0.$$

Embora o NLS produza estimadores consistentes, sua eficiência depende da homocedasticidade dos erros. Dada a frequente heterocedasticidade nos dados de comércio, o NLS tende a sobreponderar observações de grande magnitude, tornando-o ineficiente.

Silva e Tenreyro (2006) argumentam ser possível obter estimadores mais eficientes para modelos de gravidade com variáveis dependentes não negativas e heterocedasticidade presente nos resíduos, ao se utilizar o estimador de PPML. Além de ser eficiente sob a condição de variância condicional proporcional à média, os autores demonstram que o estimador PPML permanece consistente em uma variedade de cenários de variância.

Nesse contexto, os autores mostram que o estimador PPML pode ser obtido como uma forma reponderada do estimador de NLS, ao se dividir a condição de primeira ordem do NLS pela variância condicional dos resíduos, resultando:

$$\sum_{i=1}^n [y_i - \exp(x_i^{\top} \tilde{\beta})] x_i = 0. \quad (3.14)$$

O estimador definido em (3.14) é numericamente equivalente ao estimador de *Pseudo Maximum Likelihood* (PML) de Poisson, amplamente empregado na modelagem de dados de contagem. No entanto, essa forma da equação implica que é suficiente a especificação correta da média condicional, ou seja, $\mathbb{E}[y_i | x_i] = \exp(x_i^{\top} \beta)$, independentemente de y_i ser inteiro ou seguir

uma distribuição de Poisson (Gourieroux *et al.*, 1984). Dessa forma, a regressão de Poisson pode ser empregada mesmo quando a variável dependente assume valores contínuos e positivos.

3.3.2 Técnicas de diagnóstico

As técnicas de diagnóstico gráfico aplicadas aos MLG consistem, em sua essência, na análise visual de resíduos e valores ajustados, além de outras métricas derivadas do modelo estimado. Tais abordagens são fundamentais para avaliar a adequação do modelo especificado, verificar pressupostos sobre a distribuição dos erros e detectar possíveis observações influentes ou discrepantes.

Conforme sugerido por Paula (2013), essas técnicas são particularmente relevantes quando se trabalha com dados assimétricos ou com distribuições pertencentes à família exponencial, onde os pressupostos dos modelos clássicos nem sempre são válidos.

As principais ferramentas gráficas recomendadas para a avaliação dos MLG incluem:

- Gráficos de resíduos versus valores ajustados: utilizados para verificar a presença de padrões sistemáticos, como heterocedasticidade ou não linearidade, indicando possíveis violações da especificação do modelo.
- Gráfico normal de probabilidades com envelope (*QQ-plot*): permite avaliar se os resíduos seguem uma distribuição aproximadamente normal (ou outra, conforme o modelo). A presença de pontos fora dos envelopes pode indicar inadequações na escolha da distribuição da variável resposta.
- Gráfico de \hat{z}_i contra $\hat{\eta}_i$: utilizado para avaliar a adequação da função de ligação. A presença de um padrão aproximadamente linear sugere que a função de ligação especificada é apropriada para os dados.

Assim, a combinação dessas ferramentas gráficas fornece uma base sólida para verificar a consistência e a confiabilidade do modelo estimado, além de orientar possíveis reespecificações.

3.4 Generalized Additive Model for Location, Scale and Shape (GAMLSS)

Especificamente, o modelo GAMLSS assume que, para $i = 1, 2, \dots, n$, as observações independentes Y_i , condicionadas ao vetor de parâmetros θ^i , possuem função de probabilidade (ou densidade) $f(y_i | \theta^i)$, em que $\theta^i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{ip})$ é um vetor de p parâmetros relacionados

às variáveis explicativas e aos possíveis efeitos aleatórios.

Os GAMLSS são modelos de regressão semiparamétricos que requerem a especificação de uma distribuição paramétrica para a variável resposta. Eles foram propostos por Rigby e Stasinopoulos (2005) como uma extensão das classes de MLG (Nelder e Wedderburn, 1972) e Modelos Aditivos Generalizados (GAM) (Hastie, 2017).

Considerando $\mathbf{y}^\top = (y_1, y_2, \dots, y_n)$ o vetor de observações. Para $k = 1, 2, \dots, p$, seja $g_k(\cdot)$ uma função de ligação monotônica conhecida, que relaciona $\boldsymbol{\theta}_k = (\theta_{1k}, \theta_{2k}, \dots, \theta_{nk})^\top$ ao preditor linear $\boldsymbol{\eta}_k = (\eta_{k1}, \eta_{k2}, \dots, \eta_{kn})^\top$. Assim, define-se o modelo geral:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\gamma}_{jk}, \quad (3.15)$$

onde $\boldsymbol{\beta}_k^\top = (\beta_{1k}, \dots, \beta_{J'_k k})$ é o vetor de efeitos fixos, \mathbf{X}_k é a matriz de delineamento de dimensão $n \times J'_k$, \mathbf{Z}_{jk} é a matriz de delineamento associada aos termos aditivos ($n \times q_{jk}$) e $\boldsymbol{\gamma}_{jk}$ é um vetor aleatório q_{jk} -dimensional.

A formulação em (3.15) pode ser escrito na forma de efeitos aleatórios, considerando que $g_k(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável, de modo que:

$$\begin{aligned} g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \boldsymbol{\gamma}_{j1}, & g_3(\boldsymbol{\nu}) &= \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \boldsymbol{\gamma}_{j3}, \\ g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \boldsymbol{\gamma}_{j2}, & g_4(\boldsymbol{\tau}) &= \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \boldsymbol{\gamma}_{j4}. \end{aligned} \quad (3.16)$$

Assume-se que $\boldsymbol{\gamma}_{jk} \sim \mathcal{N}_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^{-1})$, em que \mathbf{G}_{jk} é uma matriz de penalização que pode depender de hiperparâmetros λ_{jk} . Quando \mathbf{G}_{jk} é singular, a densidade associada é interpretada como imprópria, sendo proporcional a $\exp\left(-\frac{1}{2} \boldsymbol{\gamma}_{jk}^\top \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk}\right)$.

O modelo (3.15) é bastante flexível, pois permite incluir termos paramétricos e não paramétricos das variáveis independentes, além de componentes que capturam efeitos aleatórios. Os parâmetros nos termos aditivos podem ser modelados de forma não linear por meio de técnicas de suavização, como suavização por *kernel*, *splines* e polinômios fracionários, entre outras. Os vetores $\boldsymbol{\gamma}_{jk}$, para $j = 1, 2, \dots, J_k$, podem ser combinados em um único vetor $\boldsymbol{\gamma}_k$ com uma matriz de design \mathbf{Z}_k unificada. No entanto, a formulação (3.15) é preferida, pois é adequada ao algoritmo de *backfitting* e facilita a incorporação de diferentes tipos de termos aditivos e/ou efeitos aleatórios (Rigby e Stasinopoulos, 2005).

Quando $J_k = 0$, não há termos aditivos associados aos parâmetros da distribuição, e o modelo (3.15) reduz-se a um modelo totalmente paramétrico:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k. \quad (3.17)$$

Caso $\mathbf{Z}_{jk} = \mathbf{I}_n$ e $\boldsymbol{\gamma}_{jk} = h_{jk}(\mathbf{x}_{jk})$, a formulação torna-se aditiva semiparamétrica:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \quad (3.18)$$

em que $h_{jk}(\cdot)$ é uma função desconhecida da variável explicativa x_{jk} , e $\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ é o vetor que avalia a função h_{jk} em \mathbf{x}_{jk} . Assume-se que os vetores explicativos \mathbf{x}_{jk} sejam conhecidos.

3.4.1 Estimação dos parâmetros do modelo

No caso do modelo GAMLSS paramétrico, mostrado em (3.17), a estimação é realizada através do método da máxima verossimilhança. De acordo com Rigby e Stasinopoulos (2005), a função é dada por:

$$\ell = \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}^i), \quad (3.19)$$

em que $f_Y(\cdot)$ representa a função densidade de probabilidade da variável resposta. A função de máxima verossimilhança penalizada ℓ_p para o modelo (3.17) é dada por:

$$\ell_p = \ell - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \boldsymbol{\gamma}_{jk}^\top \mathbf{G}_{jk} \boldsymbol{\gamma}_{jk}, \quad (3.20)$$

onde $\ell = \sum_{i=1}^n \log f(y_i | \boldsymbol{\theta}^i)$ é a função de máxima verossimilhança dos dados condicionais a $\boldsymbol{\theta}^i$, $i = 1, \dots, n$.

A classe GAMLSS possui características bastante atrativas para o uso na modelagem de fluxos de comércio internacional no contexto de modelos gravitacionais. Por exemplo, ela abrange uma ampla gama de distribuições de probabilidade, permite a modelagem conjunta de quatro parâmetros da distribuição (como média, variância, assimetria e curtose), permite a inclusão de componentes não paramétricos e também de efeitos aleatórios.

Como a abordagem deste trabalho abrange apenas as funções paramétricas do GAMLSS, (3.20) será reduzida para ℓ , e os vetores $\boldsymbol{\beta}_k$ serão estimados pela maximização desta função. Dessa maneira, o processo de estimação será equivalente ao apresentado em (3.19).

No trabalho de Rigby e Stasinopoulos (2005) são apresentados dois algoritmos para o ajuste dos modelos GAMLSS, a saber: CG e RS. O primeiro é uma generalização do algoritmo

proposto por Cole e Green (1992); este usa a primeira derivada e o valor esperado ou aproximado das derivadas de segunda ordem e das derivadas cruzadas da função de log-verossimilhança em relação aos parâmetros da distribuição $\theta = (\mu, \sigma, \nu, \tau)$, para uma distribuição com quatro parâmetros.

Entretanto, para muitas funções de densidade de probabilidade, $f(y | \theta)$, os parâmetros θ são ortogonais, ou seja, os valores esperados das derivadas cruzadas da função de log-verossimilhança são iguais a 0. Neste caso, é utilizado o algoritmo RS, sendo uma generalização do algoritmo usado por Rigby e Stasinopoulos (1996) no ajuste da média e da dispersão de modelos aditivos e que, ao contrário do CG, não necessita das derivadas cruzadas da função de log-verossimilhança. Ambos os métodos utilizam os procedimentos de Newton-Raphson/Escore de Fisher.

3.4.2 Modelos de Regressão para Dados Inflacionados e Ajustados

Os modelos de regressão para dados inflacionados ou ajustados são essenciais para lidar com variáveis de resposta que apresentam uma frequência elevada de zeros ou que incluem zeros de forma ajustada. A seguir, são apresentadas algumas distribuições comumente utilizadas nesses contextos.

3.4.2.1 Distribuição Poisson Inflacionado em Zero (ZIP)

A distribuição Poisson inflacionada em zero é apropriada para modelar dados de contagem com excesso de zeros. Existem duas parametrizações comuns para essa distribuição:

(i) Primeira parametrização

Seja $Y = 0$ com probabilidade ν e $Y \sim \text{Po}(\mu)$ com probabilidade $(1 - \nu)$, então Y tem uma distribuição de Poisson inflada em zero, denotada por $\text{ZIP}(\mu, \nu)$, dada por

$$p_Y(y | \mu, \nu) = \begin{cases} \nu + (1 - \nu)e^{-\mu}, & \text{se } y = 0, \\ (1 - \nu)\frac{\mu^y}{y!}e^{-\mu}, & \text{se } y = 1, 2, 3, \dots \end{cases}$$

A média de Y nesta parametrização é $\mathbb{E}(Y) = (1 - \nu)\mu$ e sua variância por $V(Y) = \mu(1 - \nu)[1 + \mu\nu]$.

(ii) Segunda parametrização Thomas *et al.* (2018) apresenta uma parametrização diferente da distribuição de Poisson inflacionada em zero, $\text{ZIP}(\lambda, \nu)$, dada por

$$p_Y(y \mid \mu, \nu) = \begin{cases} \nu + (1 - \nu)e^{-\left(\frac{\mu}{1-\nu}\right)}, & \text{se } y = 0, \\ \frac{(1 - \nu)\mu^y}{y!(1 - \nu)^y} e^{-\left(\frac{\mu}{1-\nu}\right)}, & \text{se } y = 1, 2, 3, \dots \end{cases}$$

A média de Y é dada por $\mathbb{E}(Y) = \mu$ e a variância por $V(Y) = \mu + \mu^2 \left(\frac{\nu}{1-\nu}\right)$.

3.4.2.2 Distribuição Gama Ajustada a Zero (ZAGA)

A distribuição ZAGA é uma distribuição de probabilidade com três parâmetros, que permite a ocorrência do valor zero com probabilidade positiva.

A função de probabilidade contínua-discreta de $Y \sim \text{ZAGA}(\mu, \sigma, \nu)$ é dada por

$$f_Y(y) = \begin{cases} p_0, & \text{se } y = 0, \\ (1 - p_0) f_W(y), & \text{se } 0 < y < \infty, \end{cases}$$

para $0 \leq y < \infty$, onde $W \sim \text{GA}(\mu, \sigma)$ tem uma distribuição gama com $0 < \mu < \infty$ e $0 < \sigma < \infty$ e $0 < \nu < 1$, onde $\nu = p_0$. Para $\text{ZAGA}(\mu, \sigma, \nu)$ as funções de ligação padrão para μ e σ são a função de ligação log e para ν a função de ligação *logit*, fornecendo preditores $\eta_1 = \log \mu$, $\eta_2 = \log \sigma$ e $\eta_3 = \log[\nu/(1 - \nu)]$. Obtém-se uma parametrização (Stasinopoulos *et al.*, 2017) para $f_W(y)$ dada por

$$f_W(y \mid \mu, \sigma) = \frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2}-1} e^{-y/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)},$$

em que $y > 0$, onde $\mu > 0$ e $\sigma > 0$ e $\mathbb{E}(W) = \mu$ e $V(W) = \sigma^2 \mu^2$.

A função de verossimilhança assume uma forma particular quando se trabalha com distribuições mistas discreta-contínuas, como a distribuição ZAGA utilizada neste estudo. Conforme Mills (2013), a função de probabilidade pode ser fatorada em duas partes: uma associada à probabilidade de ocorrência de valores nulos, e outra referente à densidade contínua condicional aos valores positivos da variável resposta.

A função de probabilidade mista é expressa como:

$$f(y) = (1 - \pi)I_{y=0} + (\pi g(y))I_{y>0}, \quad (3.21)$$

em que $I_{y=0}$ e $I_{y>0}$ são indicadores que assumem valor 1 quando suas respectivas condições são

verdadeiras e 0 caso contrário. Assim, a função de verossimilhança total é escrita como:

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^n f(y_i) = \prod_{i=1}^n [(1 - \pi)I_{y_i=0} + \pi g(y_i)I_{y_i>0}] \\ &= \left[\prod_{y_i=0} (1 - \pi) \prod_{y_i>0} \pi \right] \left[\prod_{y_i>0} g(y_i) \right]. \end{aligned} \quad (3.22)$$

Consequentemente, a função de log-verossimilhança correspondente é:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \log \left[\prod_{y_i=0} (1 - \pi) \prod_{y_i>0} \pi \right] + \log \left[\prod_{y_i>0} g(y_i) \right] \\ &= \sum_{y_i=0} \log(1 - \pi) + \sum_{y_i>0} [\log(\pi) + \log(g(y_i))]. \end{aligned} \quad (3.23)$$

A forma fatorada da função de verossimilhança permite a estimação separada dos componentes de probabilidade (parte discreta) e de média (parte contínua), independentemente das covariáveis incluídas em cada submodelo. Dessa forma, os coeficientes $\boldsymbol{\beta}^\top$ associados a cada parâmetro da distribuição são estimados de maneira independente, o que simplifica o processo computacional e a interpretação econométrica dos resultados.

3.5 Técnicas de Diagnóstico

No contexto dos modelos GAMLSS, o diagnóstico é realizado com base na análise dos resíduos quantílicos normalizados (Dunn e Smyth, 1996), os quais são definidos a partir da inversa da função de distribuição acumulada, apresentando distribuição assintoticamente normal padrão quando os parâmetros do modelo são adequadamente estimados. Essa abordagem é mais robusta do que as utilizadas em modelos clássicos, uma vez que considera a complexidade dos modelos com múltiplos parâmetros.

Nesta subseção, são apresentadas duas das principais ferramentas gráficas de diagnóstico: os gráficos padrão gerados por meio da função `plot()`, que incluem visualizações clássicas de resíduos, e o *Worm Plot*, um gráfico diagnóstico baseado em resíduos quantílicos normalizados, proposto por Buuren e Fredriks (2001). Ambas as técnicas são utilizadas para verificar a conformidade dos resíduos com a normalidade teórica esperada e para identificar possíveis inadequações na especificação do modelo.

3.5.1 Análise de Resíduos

A análise de resíduos é uma das ferramentas mais aplicadas para avaliar a qualidade de ajuste dos modelos. Os modelos GAMLSS usam os resíduos quantílicos normalizados (Dunn e Smyth, 1996), enquanto os MLGs empregam os resíduos componente do desvio (deviance) e de Pearson (McCullagh e Nelder, 1989).

A padronização permite uma melhor comparação dos resíduos entre observações com diferentes influências no ajuste do modelo. Devido à diversidade de distribuições dos modelos GAMLSS, os resíduos brutos, definidos pela diferença entre o valor observado e o valor ajustado e representados por $r_i = y_i - \hat{y}_i$ para $i = 1, 2, \dots, n$, se tornam inadequados² para estes modelos. Como alternativa, as análises de diagnóstico desta classe de modelos se baseiam nos resíduos quantílicos aleatorizados propostos por Dunn e Smyth (1996).

Os resíduos são definidos da seguinte forma:

$$r_{q,i} = \Phi^{-1} \left(F(y_i; \hat{\theta}^i) \right), \quad i = 1, 2, \dots, n,$$

em que, y_i corresponde à i -ésima observação da variável resposta; $F(y_i; \hat{\theta}^i)$ é a distribuição acumulada da função de probabilidade assumida no modelo; $\hat{\theta}^i = (\hat{\mu}_i, \hat{\sigma}_i, \hat{\nu}_i, \hat{\tau}_i)$ são as estimativas dos parâmetros do modelo; $\Phi^{-1}(\cdot)$ é a função de distribuição acumulada normal padrão.

Para distribuições discretas ou mistas, torna-se necessária a definição de uma nova variável aleatória, u_i , de distribuição uniforme no intervalo (a_i, b_i) , para a_i e b_i obtidos por:

$$a_i = F(y_i - 1 | \hat{\theta}^i) \quad \text{e} \quad b_i = F(y_i | \hat{\theta}^i).$$

Sua introdução nos resíduos $r_{q,i}$, possibilita a geração de resíduos aleatórios contínuos, definidos por

$$r_{q,i} = \Phi^{-1}(u_i).$$

Este procedimento tem por finalidade evitar massas de pontos na distribuição dos resíduos, e retoma os resíduos quantílicos aleatorizados (chamados apenas de resíduos, subsequentemente). Considerando que o modelo foi especificado corretamente, estes resíduos terão distribuição normal padrão. Isto posto, diversas técnicas de diagnóstico usuais podem ser realizadas para verificação do modelo.

² Isso ocorre pois o comportamento esperado destes resíduos seria condizente com uma variável de distribuição normal.

3.5.2 Método Gráfico

A saída padrão de diagnóstico gerada pelo comando `plot()` no software R (pacote `gamlss`) é composta pelos quatro gráficos definidos abaixo, acrescidas das medidas descritivas do resíduo:

- Resíduos versus valores ajustados;
- Resíduos versus um índice ou uma variável explicativa;
- Densidade estimada de Kernel dos resíduos;
- Gráfico quantil-quantil normal (QQ-plot) dos resíduos.

Espera-se que a média desses resíduos seja aproximadamente zero, com desvio padrão próximo de um e as medidas de assimetria e curtose em torno de zero e três, respectivamente, características correspondentes à uma variável com distribuição normal padrão. Para os gráficos, esperam-se comportamentos típicos de resíduos com distribuição normal, conforme apresentado por Montgomery *et al.* (1992).

3.5.3 Half Normal Plot

O *Half Normal Plot* (HNP) é uma técnica gráfica utilizada para avaliar a adequação do ajuste de um modelo estatístico por meio da comparação entre os resíduos observados e os valores esperados sob o modelo ajustado. Esse método, originalmente proposto por Atkinson (1985), baseia-se na ordenação dos resíduos (frequentemente o deviance) em valores absolutos e na sua comparação com os quantis teóricos de uma distribuição *half-normal* teórica. A presença de pontos que se afastam significativamente do envelope indica possíveis *outliers* ou valores com alta influência no ajuste.

No contexto dos modelos GAMLSS, o HNP pode ser empregado para verificar a presença de observações influentes, heterocedasticidade ou desvios de normalidade nos resíduos. O gráfico no entanto requer a definição de funções auxiliares específicas para sua aplicação. Para viabilizar a construção do hnp, foram implementadas as funções `diagfun`, `simfun` e `fitfun`, responsáveis, respectivamente, pela extração dos resíduos, simulação de amostras e reestimação do modelo. Essa adaptação metodológica segue as diretrizes propostas por Moral *et al.* (2017), garantindo que os envelopes simulados sejam comparáveis, permitindo uma avaliação visual mais rigorosa: (i) se os pontos estiverem próximos da linha de referência (dentro das bandas de confiança), indica-se um bom ajuste; e (ii) observações fora das bandas podem sugerir possíveis

outliers, má especificação da distribuição ou necessidade de termos adicionais no modelo.

3.5.4 *Worm Plot*

Dentre as técnicas diagnósticas baseadas em resíduos quantílicos normalizados, o *Worm Plot*, proposto por Buuren e Fredriks (2001), é uma ferramenta gráfica utilizada para avaliar a adequação de modelos. Essa abordagem baseia-se na análise dos resíduos quantílicos normalizados para identificar discrepâncias sistemáticas entre os dados observados e a distribuição assumida pelo modelo.

A construção do *Worm Plot* envolve a transformação dos resíduos em quantis da distribuição normal padrão, dividindo-os em grupos de tamanhos semelhantes ao longo da variável preditora. Cada grupo de resíduos é então representado em um gráfico de dispersão contra os quantis esperados de uma distribuição normal padrão. O nome deriva da aparência curvada que os pontos assumem quando o modelo apresenta especificações inadequadas, semelhante ao formato de uma minhoca.

O bom ajuste dos dados ao modelo é caracterizado pela disposição dos pontos aleatoriamente ao redor do eixo horizontal, sem padrões sistemáticos. Para um dado quantil z com probabilidade associada p e tamanho amostral n , o intervalo de confiança de 95% é calculado como:

$$\pm 1,96 \times \left[f(z)^{-1} \sqrt{\frac{p(1-p)}{n}} \right],$$

onde $f(z)$ representa a densidade da distribuição normal padrão na posição z ³.

O formato da minhoca revela padrões específicos de inadequação do modelo e sugere possíveis ajustes. Segundo Buuren e Fredriks (2001), a interpretação do gráfico segue diretrizes bem definidas. Na Figura 17 (Apêndice B), um modelo bem ajustado resulta em pontos distribuídos aleatoriamente ao redor da linha horizontal, sem padrões visíveis. Padrões sistemáticos fornecem informações sobre a inadequação do modelo: (i) um formato em “U” ou invertido indica problemas na curtose, sugerindo que as caudas da distribuição estão mais pesadas ou leves do que o esperado; (ii) uma forma em “S” sugere assimetria, indicando que a distribuição dos resíduos não é simétrica em torno de zero e (iii) desvios inclinados para um dos lados podem indicar a necessidade de transformar a variável resposta ou ajustar melhor os

³ O intervalo de confiança de 95% dá uma impressão da variação da amostragem e delinea a região onde a minhoca deve estar localizada na maioria das vezes se as distribuições empírica e teórica concordarem.

parâmetros do modelo.

3.6 Seleção de Modelos

A seleção adequada da distribuição e da estrutura paramétrica é uma etapa fundamental no processo de modelagem. Quando dois modelos não são aninhados, o teste de razão de verossimilhança não pode ser aplicado, tornando necessária a utilização de outros critérios para a seleção do modelo mais adequado. Para esse fim, diversos critérios de informação foram propostos na literatura.

Akaike (1974) introduziu o *Akaike Information Criterion* (AIC) para a seleção de modelos, que é expresso por:

$$AIC = -2\ell(\beta) + 2p,$$

onde $\ell(\beta)$ representa o logaritmo da função de verossimilhança do modelo e p é o número de parâmetros estimados. O AIC busca encontrar um equilíbrio entre qualidade de ajuste e parcimônia, favorecendo modelos com menor valor de AIC.

Outro critério amplamente utilizado é o *Bayesian Information Criterion* (BIC), proposto por Schwarz (1978), dado por:

$$BIC = -2\ell(\beta) + p \ln(n),$$

assim, além dos termos já definidos, n é o número de observações. O BIC impõe uma penalização maior para modelos mais complexos em comparação ao AIC, favorecendo modelos mais parcimoniosos em amostras grandes.

Adicionalmente, quando o tamanho da amostra n é pequeno e o número de parâmetros p é relativamente grande, recomenda-se o uso do Critério de Informação de Akaike Corrigido (AICc) para evitar a seleção tendenciosa de modelos superparametrizados.

Os modelos GAMLSS apresentam diferentes funções para a seleção de variáveis explicativas. A função considerada é construída com base no *Generalized Akaike Information Criterion* (GAIC), definido por

$$GAIC(k) = GD + k \cdot df,$$

em que GD é a deviance global dada por -2ℓ , df corresponde aos graus de liberdade do modelo e k é um fator de penalização ajustável. O GAIC é uma generalização do AIC, sendo que para

$k = 2$ recuperamos o critério tradicional. Dessa forma, o AIC é considerado um caso especial do GAIC.

Dentre os métodos de seleção de variáveis, Paula (2013) descreve quatro abordagens principais: métodos *forward*, *backward* e *stepwise* e a seleção baseada no critério de Akaike. Assim, utiliza-se o método *backward*, no qual as variáveis são gradativamente removidas do modelo GAMLSS, buscando a minimização do GAIC. Esse processo mantém a propriedade de parcimônia do AIC, pois modelos com menos parâmetros tendem a apresentar penalidades menores e, conseqüentemente, valores reduzidos de GAIC.

Além disso, a seleção de variáveis pode ser realizada separadamente para cada parâmetro da distribuição adotada no GAMLSS. Primeiramente, selecionam-se as variáveis para o modelo da média (μ), seguido pelo modelo do parâmetro de dispersão (σ) e, quando aplicável, também para parâmetros de inflação de zeros ou curtose, como ν e τ . Esse processo é realizado sequencialmente, garantindo uma modelagem flexível e adaptada às especificidades dos dados.

4 CARACTERIZAÇÃO DO MERCADO INTERNACIONAL DA CERA DE CARNAÚBA

A cera de carnaúba, embora tecnicamente um insumo vegetal, possui relevância econômica, social e simbólica que transcende sua materialidade. Endêmica do semiárido nordestino, sua cadeia produtiva é, ao mesmo tempo, expressão da riqueza da biodiversidade brasileira e reflexo das desigualdades históricas que atravessam o interior do país. No cenário internacional, a carnaúba ocupa posição estratégica, abastecendo indústrias de alto valor agregado ao redor do mundo, enquanto, em sua origem, segue atrelada a práticas extrativistas tradicionais. Este capítulo propõe uma análise da estrutura e da dinâmica do mercado internacional da cera de carnaúba, com especial atenção aos paradoxos que marcam sua trajetória: o protagonismo exportador do Brasil; a valorização do produto no exterior frente à instabilidade dos preços internos; e a crescente demanda global diante dos desafios ambientais e sociais locais. Trata-se, portanto, de compreender não apenas fluxos de produção e comércio, mas também as relações institucionais e territoriais que sustentam esse mercado.

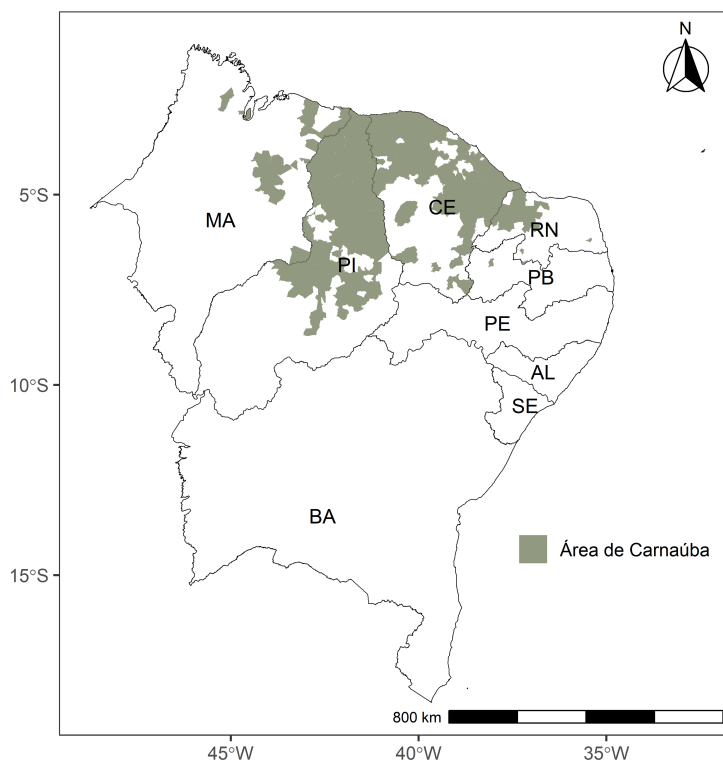
4.1 Produção e Base Geográfica

A carnaúba, *Copernícia prunifera*, pertencente ao gênero *Copernicia*, é uma palmeira nativa e endêmica do Nordeste brasileiro. É reconhecida como a “planta extrativa por excelência” devido ao alto valor agregado da cera extraída de suas folhas. Esse produto nobre é obtido a partir de um pó branco, fino, que reveste as células epidérmicas das folhas da planta. Desde o século XIX, a busca por métodos mais eficientes de extração e beneficiamento da cera tem mobilizado pesquisadores, técnicos e produtores (Moreira, 2022)

Embora algumas espécies do gênero *Copernícia* estejam presentes em outras regiões do Brasil e em alguns países, a *C. prunifera* se concentra no semiárido nordestino, principalmente em margens de rios intermitentes e várzeas. A Figura 1 ilustra as principais áreas de ocorrência dos carnaubais, abrangendo os estados do Maranhão, Piauí, Ceará, Rio Grande do Norte e Paraíba.

A cadeia produtiva da carnaúba estrutura-se em um modelo tradicionalmente extrativista, no qual o pó cerífero é colhido manualmente e processado em diversas etapas até se transformar na cera de carnaúba, amplamente utilizada pelas indústrias de cosméticos, farmacêutica, alimentícia e automobilística. Outros subprodutos também são aproveitados, como a fibra da folha, empregada em diversas finalidades artesanais e industriais (Moreira, 2022).

Figura 1 – Ocorrência da extração da Carnaúba (2017).



Fonte: Elaboração própria a partir de IBGE (2019).

Em 2024, a produção nordestina alcançou 556 toneladas de cera, 15.042 toneladas de pó e 1.507 toneladas de fibra de carnaúba, movimentando R\$ 220 milhões (IBGE, 2025). O Brasil permanece como o principal exportador mundial de cera de carnaúba, com receitas que somaram US\$ 108 milhões, consolidando o país como líder absoluto no comércio internacional desse insumo (MDIC, 2025).

Apesar do protagonismo econômico, o setor enfrenta sérios desafios estruturais e sociais. Muitas comunidades envolvidas com a extração da carnaúba são vulneráveis, sujeitas a trabalho informal, condições precárias de segurança e saúde ocupacional e, em casos extremos, trabalho forçado ou infantil (Filho e Santana, 2020). A ausência de inovações tecnológicas limita a produtividade, enquanto o desinteresse das novas gerações pela atividade ameaça a continuidade da prática extrativista.

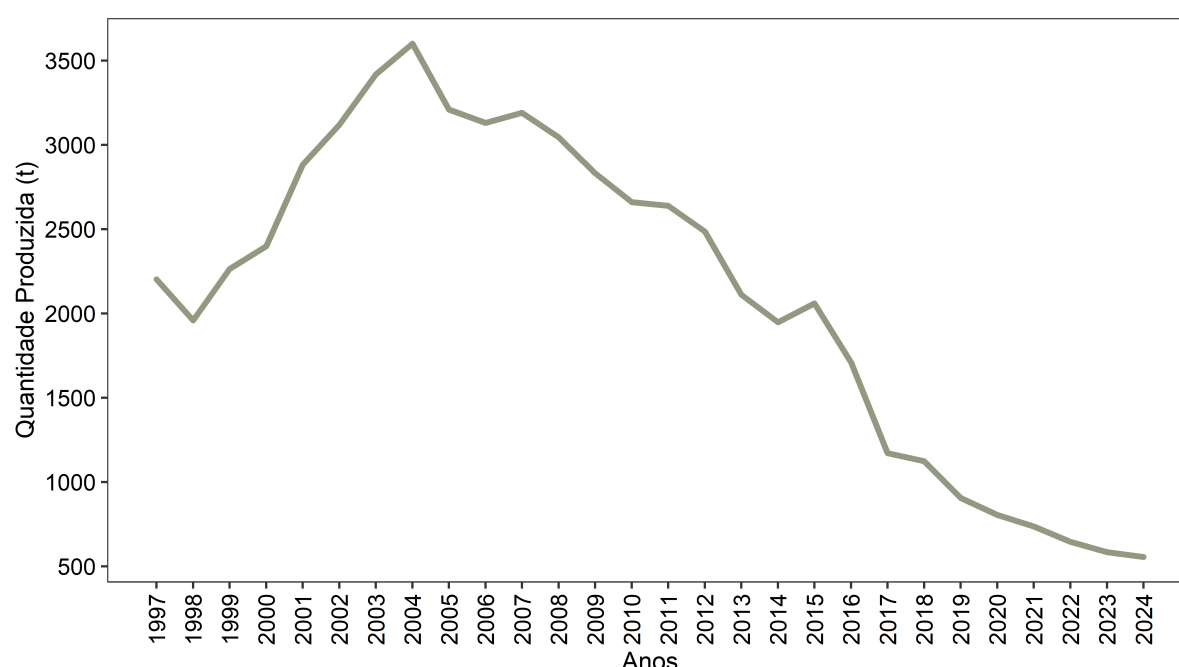
Adicionalmente, a biodiversidade regional está sob risco, comprometendo a regeneração natural dos carnaubais e, por conseguinte, a sustentabilidade do setor. No entanto, há importantes oportunidades: a manutenção dos carnaubais contribui para a preservação das florestas em pé, proteção das bacias hidrográficas e valorização econômica da biodiversidade nativa. O uso sustentável da carnaúba, por meio de práticas de manejo responsáveis, também desestimula a conversão de terras e limita o avanço de espécies invasoras, fortalecendo o elo

entre conservação ambiental e geração de renda (Câmara Setorial da Carnaúba, 2009).

4.1.1 *Trajetoória da Produção*

A produção brasileira de cera de carnaúba apresentou significativa retração entre 1997 e 2024, conforme ilustrado na Figura 2. Entre 1998 e 2004, experimentou-se um período de expansão que culminou no pico produtivo de aproximadamente 3.600 toneladas em 2004; observa-se uma trajetória de declínio quase contínuo, intensificada a partir de 2013. Em 2024, o volume produzido caiu para menos de 600 toneladas, o menor patamar da série histórica (IBGE, 2025).

Figura 2 – Evolução da produção nacional de cera de carnaúba (1997–2024).



Fonte: IBGE (2025).

Historicamente, a produção nacional é fortemente orientada para o mercado externo, que ainda hoje constitui a principal fonte de demanda e de sustentação econômica da cadeia produtiva. No entanto, por se tratar de uma matéria-prima de características singulares e amplamente valorizada no mercado internacional, a cera de carnaúba tornou-se objeto de estratégias comerciais assimétricas. Grupos empresariais estrangeiros, com elevado poder de barganha, têm exercido controle indireto sobre as cotações internacionais, apropriando-se de parcela significativa do excedente econômico gerado, em detrimento da renda doméstica (Moreira, 2023).

Um mecanismo recorrente adotado por esses compradores internacionais consiste na antecipação de contratos de aquisição, muitas vezes com preços substancialmente abaixo dos praticados em períodos de pico. Essa prática limita a capacidade de negociação dos produtores nacionais, reduzindo suas margens de lucro e tornando-os mais vulneráveis à volatilidade do mercado.

4.2 Evolução Histórica das Exportações e Principais Destinos

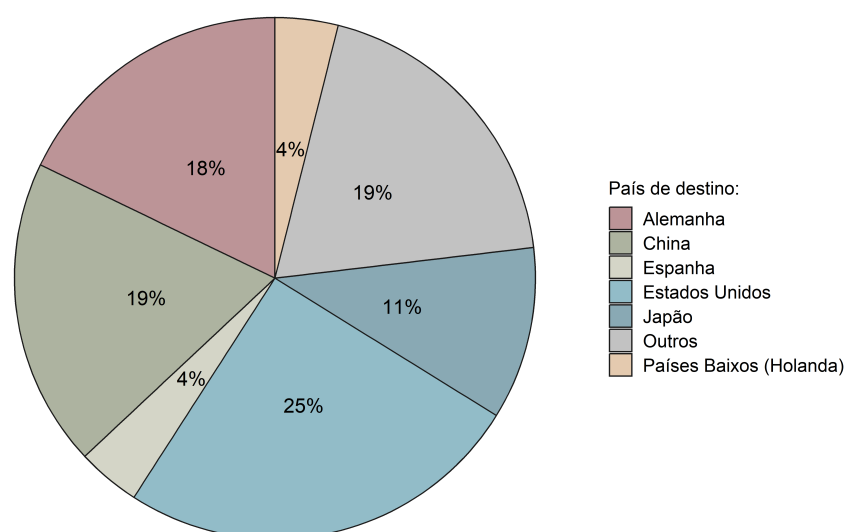
Com efeito, do confronto entre produção e exportação para o período de 1997 a 2024, desprende-se que o mercado internacional vem absorvendo o volume produzido anualmente. Ainda que se considere a possibilidade de superestimação nos dados de exportação e/ou subnotificação da produção, é plausível que a produção seja, de fato, superior à registrada. Assim, os dados disponíveis indicam forte dependência da demanda internacional.

As exportações brasileiras de cera de carnaúba, para o resto do mundo, defrontam-se com um mercado bastante diversificado, haja vista o número de países que importam o produto. Neste sentido, as estatísticas do Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC) mostram que nada menos que 44 países demandaram cera de carnaúba em 2024. Apesar da ampla base geográfica, a concentração das vendas externas é significativa: apenas quatro países — Estados Unidos, China, Alemanha e Japão — responderam, conjuntamente, por 71% do volume exportado naquele ano (Figura 3). Tal configuração, ainda que demonstre inserção global, evidencia uma significativa dependência de poucos mercados estratégicos.

A trajetória das exportações entre 1997 e 2024 (Figura 4) revela padrões distintos ao longo do período. No que diz respeito ao volume exportado, observa-se uma trajetória relativamente estável até meados dos anos 2000, com oscilações em torno de 14 mil toneladas anuais. A partir de 2010, observa-se um movimento consistente de crescimento nas exportações, com os volumes exportados se aproximando de 17 mil toneladas em determinados anos.

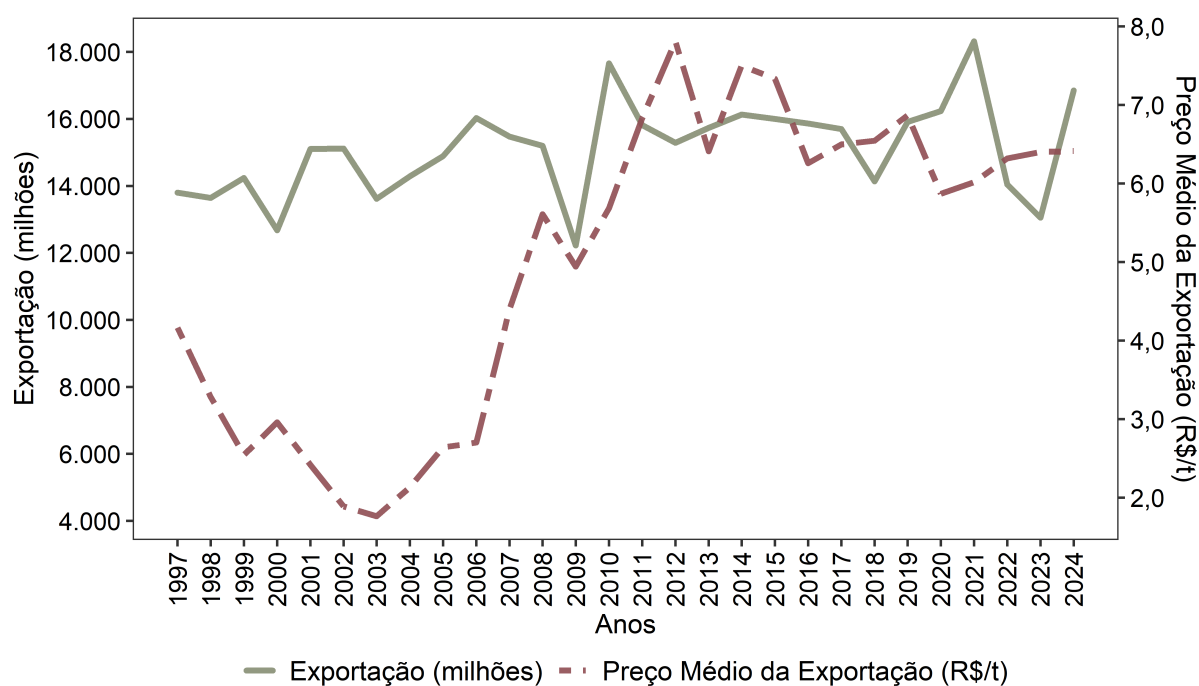
Dada a forte dependência do mercado externo evidenciada pelo elevado grau de concentração das exportações em poucos países e pela sensibilidade às oscilações da demanda global, o setor produtivo enfrentou uma queda significativa na demanda internacional, restrições de crédito e volatilidade nos mercados financeiros globais. Após esse período de contração, as exportações se recuperaram gradualmente, impulsionadas pelo aumento da demanda pela produção e de matérias-primas, bem como pela valorização das commodities no mercado internacional.

Figura 3 – Participação dos principais países importadores no volume das exportações brasileiras de cera de carnaúba (2024).



Fonte: MDIC (2025).

Figura 4 – Volume e valor das exportações brasileira de cera de carnaúba (1997-2024).



Fonte: MDIC (2025).

Apesar de variações pontuais, o patamar médio de exportações manteve-se elevado nos últimos anos, com leve recuo em anos mais recentes. Todavia, sob uma ótica abrangente, parece sensato afirmar que as quantidades exportadas têm permanecido relativamente estáveis ao longo da série analisada, com uma média próxima de 15.000 toneladas anuais.

O preço médio da exportação apresentou tendência declinante entre 1997 e 2003,

quando recuou de aproximadamente US\$ 9,00/t para menos de US\$ 2,00/t, configurando um processo de aviltamento dos preços. Esse movimento refletiu o enfraquecimento do comércio global, resultado da reversão do boom das ações da nova economia nos Estados Unidos, da recessão no Japão e da estagnação na União Europeia (Costa *et al.*, 2013).

A partir de 2004, nota-se uma recuperação expressiva, com elevação contínua até 2011. Esse período se caracteriza por uma valorização significativa do produto exportado, atingindo um pico próximo a US\$ 8,00/t. O preço das exportações durante o “super ciclo” (2003–2012) aumentou, oscilando em seguida, com nova queda após a recessão de 2014–16. Observa-se que as exportações cresceram significativamente até 2010, sofreram uma redução durante a recessão e, após seu término em 2016, voltaram a crescer, atingindo um ponto alto em 2018. Com a chegada da pandemia, houve um novo crescimento em 2021, alcançando o ponto mais alto da série histórica.

Ao nível de mercado interno, destaca-se a importância dos seguintes fatores: a variabilidade na qualidade do produto, associada às práticas tradicionais de extração e beneficiamento; a elevada relação de preços entre a cera de carnaúba e seus substitutos (naturais e sintéticos); a oferta inelástica, dada a limitação territorial da carnaubeira; e a ineficiência dos canais de comercialização. Quanto ao mercado externo, ressalta-se que as restrições à expansão da produção brasileira têm levado indústrias internacionais a reformular suas formulações, diante da impossibilidade de garantir oferta compatível com a crescente demanda. Soma-se a isso a desconfiança quanto à uniformidade e à pureza do produto exportado, além de denúncias pontuais envolvendo condições de trabalho análogas à escravidão — fatores que comprometem a imagem da cadeia produtiva brasileira e estimulam o desenvolvimento de substitutos sintéticos (Filho e Santana, 2020).

5 APLICAÇÃO

Após a construção teórica e a contextualização histórica que fundamentam este estudo, este capítulo tem por objetivo aplicar os métodos propostos aos dados observados. A escolha pelo uso de modelos gravitacionais, especialmente sob a estrutura flexível dos GAMLSS, justifica-se não apenas por sua robustez metodológica, mas sobretudo por sua capacidade de lidar com características recorrentes em dados de comércio internacional, como assimetria, heterocedasticidade e, em especial, a presença significativa de valores zero.

5.1 Base de dados

No que concerne à verificação dos determinantes das exportações de cera vegetal, tem-se que os dados para as exportações da cera foram coletados no sistema de consultas e extração de dados do comércio exterior brasileiro, MDIC¹, bem como no banco de dados UN Comtrade. Os indicadores referentes ao PIB brasileiro e dos países importadores foram obtidos por meio do World Bank (2025).

As distâncias entre os países foram coletadas do *Centre d'Études Prospectives et d'Informations Internationales* (CEPII). Cada distância foi mensurada em termos de quilometragem, pelas coordenadas geográficas de suas principais cidades, ou seja, latitude e longitude. O Índice de Liberdade Comercial (ILC), seguindo Arevalo *et al.* (2016) e Aguiar (2018), de maneira que esse indicador possa variar de 0 a 100, expressando: liberdade restrita (0–49,9), predominantemente restrita (50–59,9), parcialmente livre (60–69,9), predominantemente livre (70–79,9) e completamente livre (80–100)². A taxa de câmbio nominal foi coletada no BANCO CENTRAL DO BRASIL (2025). Além disso, foram construídas variáveis binárias indicando se o país importador tem acesso ao mar e se há acordos comerciais vigentes.

O período analisado abrange os anos de 1997 a 2024, escolhido em função da disponibilidade de dados e por representar de forma adequada a evolução das exportações ao longo de diferentes ciclos econômicos. Salienta-se que foram considerados 44 (quarenta e quatro) países (Apêndice A), que representam 98% do destino das exportações brasileiras de cera vegetal, tendo sido considerado o período de 1997 a 2024, totalizando 1.232 observações.

¹ Sendo realizada com base na Tabela da Nomenclatura Comum do Mercosul (NCM): 15211000 — Ceras vegetais.

² Os dados foram obtidos da *Heritage Foundation*.

5.1.1 Dados e Variáveis do modelo

Os dados utilizados para a estimação do modelo são provenientes de fontes secundárias, sendo obtidos nas seguintes bases de dados: Sistema de Estatísticas do Comércio Exterior do Brasil (Comex stat), Banco Mundial (*World Bank*), CEPII, Sistema Integrado de Comércio Exterior (SISCOMEX), Banco Central do Brasil e *International Monetary Fund* (FMI).

A amostra utilizada compreende 44 países, os quais foram selecionados conforme a intensidade de exportação desse produto ao longo do período analisado, sendo este de 1997 a 2024, conforme o Quadro 1.

Quadro 1 – Nomenclatura e especificação das variáveis apresentadas.

Variável	Descrição	Fonte	Sinal Esperado
X_{ij}	Fluxo de comércio. Valor atribuído pelos valores das exportações.	ComexStat UN Comtrade	★
Y_i e Y_j	PIB do país exportador e do país importador, respectivamente.	Banco Mundial FMI	(+)
D_{ij}	Distâncias entre o país exportador e o país importador.	CEPII	(-)
Pop_i e Pop_j	População do país exportador e do país importador, respectivamente.	Banco Mundial FMI	(+)
tC_j	Taxa de Câmbio.	Banco Central	(-)
ILC_j	Índice de liberdade comercial.	<i>Heritage Foundation</i>	(+)
rta^*	Acordos Comerciais Vigentes.	CEPII	(+)
$landlocked_j^*$	Países sem acesso ao Mar.	CEPII	(-)

Fonte: Elaboração própria com base nos dados da pesquisa.

Nota: ★: Variável Dependente. *: Variável *dummy*.

A escolha das variáveis foi fundamentada no arcabouço teórico do modelo gravitacional, tal como em: Corrêa (2021), Anderson e Wincoop (2003), Ferreira (2021), bem como em abordagens empíricas mais recentes, a fim de proporcionar uma análise abrangente dos determinantes dos fluxos de exportações.

É importante destacar que o tamanho das economias é crucial para explicar o volume de comércio entre duas nações, uma vez que o PIB do país exportador reflete seu potencial produtivo, enquanto o PIB do país importador reflete seu potencial de consumo e a condição financeira do país para adquirir e consumir produtos. Da mesma forma como o PIB per capita dos países exportadores e importadores constitui um dos principais determinantes no modelo, essa variável, obtida pela razão entre o PIB e a população, reflete a renda média dos habitantes e, consequentemente, a capacidade produtiva e de consumo dos países.

A taxa de câmbio é outro fator relevante para a análise do comércio internacional.

Espera-se um coeficiente positivo para essa variável, pois uma depreciação cambial (ou seja, um aumento na taxa de câmbio) torna os produtos brasileiros mais baratos para os países importadores, aumentando a competitividade das exportações. Estudos como o de Eichengreen (2007) corroboram a ideia de que taxas de câmbio favoráveis incentivam exportações, justificando um coeficiente positivo esperado.

A distância entre as nações torna o comércio mais difícil, uma vez que, quanto maior a distância entre elas, maiores os custos de transporte e os preços dos produtos importados, levando os países a terem uma tendência natural de comércio com as nações mais próximas, uma vez que a distância pode inviabilizar certas importações. Disdier e Head (2008) destacam que a elasticidade da distância é consistentemente negativa, indicando que os fluxos comerciais são inversamente proporcionais à distância. Com isso, espera-se um coeficiente negativo para a variável, uma vez que o comércio tende a se reduzir à medida que a distância aumenta.

O índice de liberdade comercial mede o ambiente regulatório e a facilidade de operação empresarial em um país. Economias com maior liberdade econômica oferecem condições mais favoráveis para o comércio, como menor burocracia e maior atratividade para investimentos estrangeiros. Conforme Gwartney *et al.* (2023), maior liberdade econômica está associada a volumes comerciais mais elevados, indicando um coeficiente positivo esperado.

Quanto aos países, importador e exportador, possuem acordos comerciais, reflete-se na redução de barreiras tarifárias e não tarifárias, assim há uma tendência de comércio entre ambos, sendo acrescentada ao modelo apresentado uma *dummy* para esta variável, a qual assumirá valor 1 caso ambos os países apresentem acordo e 0 caso contrário. Baier e Bergstrand (2007) demonstram que tais acordos podem aumentar significativamente os fluxos comerciais ao estimular a integração econômica. Assim, espera-se um coeficiente positivo para essa variável.

Por fim, a variável *landlocked_j*, que indica se um país possui ou não saída para o mar, é essencial para capturar os efeitos da geografia sobre o comércio. Países sem acesso marítimo enfrentam custos de transporte significativamente mais altos, o que prejudica sua competitividade internacional.

5.2 Análise Descritiva

A tabela a seguir apresenta as estatísticas descritivas das variáveis contempladas neste estudo. As variáveis analisadas incluem o volume de exportações de cera vegetal, distância geográfica entre o Brasil e seus parceiros comerciais, PIB per capita do Brasil e dos países

importadores, taxa de câmbio, índice de liberdade comercial e população dos países importadores. A seleção dessas variáveis fundamenta-se em sua relevância para a análise dos fluxos comerciais de cera vegetal, conforme o modelo gravitacional adaptado.

Tabela 1 – Estatísticas descritivas das variáveis contempladas no estudo.

Variável	Mínimo	Média	Máximo	Desvio Padrão	Coefficiente de Variação (%)
Exportações de Cera Vegetal (milhões US\$)	0	1,71	39,29	4,49	261,32
Distância geográfica (km)	1.457	9.028	18.837	4.981,88	53,35
PIB per capita do Brasil (mil US\$)	2,83	7,69	13,20	0,31	40,36
PIB per capita dos países importadores (mil US\$)	0,40	17,19	88,43	1,78	103,49
Taxa de câmbio (US\$/R\$)	1,07	2,80	5,39	1,28	44,45
Índice de liberdade comercial (ILC)	24,70	64,59	95,00	10,09	15,62
População dos Países importadores (milhões)	0,74	114,80	1.450,93	266,19	231,87

Fonte: Elaboração própria com base nos dados da pesquisa.

Nota: As exportações e populações foram ajustadas para milhões para maior clareza.

O PIB per capita do Brasil tem valores entre US\$ 2,83 mil e US\$ 13,20 mil, com uma média de US\$ 7,69 mil e desvio padrão de 0,31 mil US\$. O coeficiente de variação de 40,36% revela que há uma variação moderada no PIB per capita do Brasil ao longo dos anos. A variável PIB per capita dos países importadores apresenta uma maior dispersão, com valores variando entre US\$ 0,40 mil e US\$ 88,43 mil. A média de US\$ 17,19 mil e o coeficiente de variação de 103,49% indicam uma grande disparidade econômica entre os países importadores. Como o PIB per capita é uma variável econômica contínua e com comportamento próximo do log-normal, é apropriada a aplicação de uma função logarítmica, de modo a capturar elasticidades e reduzir o impacto de assimetrias.

A distância geográfica entre o Brasil e os países importadores varia entre 1.457 km e 18.837 km, com média de 9.337 km e desvio-padrão de 4.982 km. Dada sua distribuição assimétrica e a relação inversa esperada com os fluxos comerciais, aplicou-se a transformação logarítmica, conforme recomendado na literatura de modelos gravitacionais.

A Taxa de Câmbio (US\$) tem valores entre 1,07 e 5,39, com uma média de 2,80 e desvio padrão de 1,22. Por ser uma variável contínua e positiva, com comportamento potencialmente não linear, a função de ligação logarítmica também é indicada. A taxa de câmbio afeta os preços relativos entre exportadores e importadores, influenciando diretamente a competitividade do produto exportado brasileiro no mercado internacional.

O ILC varia de 24,70 a 95,00, com uma média de 64,59 e desvio padrão de 10,09. O coeficiente de variação de 15,62% sugere uma menor dispersão em comparação com outras

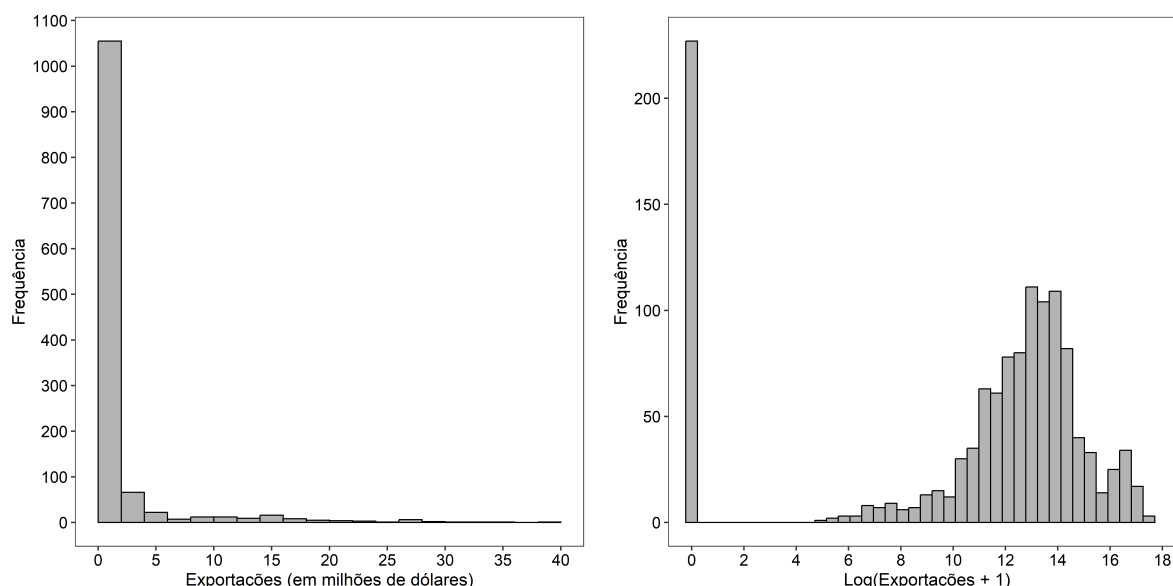
variáveis. Como se trata de um índice composto com escala delimitada, é comum tratá-lo com uma relação linear, especialmente quando se busca avaliar seu efeito direto sobre os fluxos comerciais, assumindo que maior liberdade econômica facilita o comércio exterior.

A população dos países importadores (milhões de pessoas) tem valores entre 0,74 e 1.450,93 milhões, com uma média de 114,80 milhões e desvio padrão de 266,19 milhões. O coeficiente de variação de 231,87% evidencia altíssima disparidade no tamanho dos mercados consumidores. Considerando a forte assimetria e a hipótese de que países mais populosos demandam mais produto, é comum aplicar uma função logarítmica, o que permite captar relações não lineares e controlar o impacto de países com populações muito superiores à média.

A literatura sobre modelos gravitacionais reconhece que variáveis como distância, PIB per capita, população e taxa de câmbio apresentam, frequentemente, distribuição assimétrica e elevada variabilidade. Trabalhos clássicos, como os de Silva e Tenreiro (2006) e Head e Mayer (2014), abordam o uso da transformação logarítmica para estimar elasticidades e reduzir a distorção causada por *outliers*. No contexto da cera de carnaúba, cuja comercialização envolve parceiros com perfis econômicos e demográficos altamente heterogêneos, a adoção de funções logarítmicas revela-se particularmente apropriada.

Figura 5 – Distribuição das exportações antes e após transformação logarítmica.

(a) Variável Exportações em Milhões de Dólares (X_{ij}). (b) Variável Exportações Log-transformada.



Fonte: Elaboração própria.

A variável dependente, X_{ij} , apresenta forte assimetria positiva, conforme evidenciado pelo histograma apresentado na Figura 5a. A maioria das observações está concentrada em

valores muito baixos de exportações, com diversos registros iguais a zero e poucos valores extremos. Essa característica é corroborada pelas estatísticas descritivas da Tabela 1, em que se observa um coeficiente de variação elevado (261,32%), além de uma média (1,71) muito inferior ao valor máximo observado (US\$ 39,29 milhões). Esse padrão indica elevada dispersão e cauda longa à direita, comprometendo a adequação de modelos que assumem normalidade e homocedasticidade dos resíduos.

Para lidar com essa distribuição altamente assimétrica, foi aplicada a transformação logarítmica natural $\log(X_{ij} + 1)$, cujo resultado é ilustrado na Figura 5b. O acréscimo da constante 1 garante a definição da transformação em observações nulas, evitando a exclusão de fluxos comerciais iguais a zero. O histograma indica que essa transformação reduziu consideravelmente a assimetria, aproximando a distribuição de uma forma mais simétrica. No entanto, ainda se observa um pico pronunciado em zero, evidenciando a persistência de fluxos comerciais nulos. A presença de uma cauda longa à direita sugere que, mesmo após a transformação, subsistem valores extremos substanciais.

Esse padrão reforça a necessidade de modelagem estatística que acomode adequadamente a frequência de valores nulos e a dispersão dos dados. Para modelar adequadamente essa estrutura, podem-se utilizar abordagens estatísticas que considerem tanto a presença de zeros excessivos quanto a variabilidade dos fluxos comerciais. Nesse contexto, os GAMLSS destacam-se por sua flexibilidade em acomodar diferentes distribuições de dados, incluindo aquelas ajustadas a zeros.

No caso brasileiro, a ausência de fluxos comerciais em algumas rotas de exportação pode decorrer tanto de zeros estruturais quanto de zeros de censura, a depender do contexto. Os zeros estruturais ocorrem quando não há demanda efetiva pelo produto em determinados mercados, seja por disponibilidade de substitutos, barreiras comerciais, normas sanitárias ou restrições regulatórias. Por exemplo, a exploração de trabalho análogo ao escravo na cadeia produtiva da cera de carnaúba pode afastar compradores estrangeiros, resultando em queda nas exportações para certos mercados (Filho e Santana, 2020).

Por outro lado, zeros de censura surgem quando os volumes exportados são muito pequenos ou inferiores ao limiar mínimo de registro³, ou quando transações são realizadas por intermediários, ocultando o destino nos dados brasileiros. Portanto, no Brasil, ambos os tipos de

³ É comum que sistemas de comércio internacional adotem tais práticas para simplificar a coleta de dados e reduzir certos custos, resultando na omissão de pequenos fluxos de exportação.

zeros podem coexistir, embora os de censura tendam a ser mais frequentes devido à forma como os dados de comércio exterior são coletados e reportados.

5.3 Equação gravitacional utilizada

Nesta equação, o comércio (X_{ij}) representa as exportações do produto analisado entre os países i (exportador) e j (importador), enquanto as variáveis explicativas refletem características econômicas, geográficas e institucionais dos países participantes.

Nos MLGs será avaliada a adequabilidade das distribuições sugeridas para modelagem do parâmetro da média, apresentado na Seção 6.1. Assim, todos os modelos irão contar com as variáveis explicativas de (5.1).

$$\begin{aligned} \mathbb{E}[X_{ij} | Z] = \mu_{ij} = \exp \left\{ \beta_0 + \beta_1 \log \left(\frac{Y_i}{\text{Pop}_i} \right) + \beta_2 \log \left(\frac{Y_j}{\text{Pop}_j} \right) + \beta_3 \log (D_{ij}) + \beta_4 \log (\text{Pop}_j) \right. \\ \left. + \beta_5 \text{ILC}_j + \beta_6 \log (\text{tC}_j) + \beta_7 \text{landlocked}_j + \beta_8 \text{rta}_{ij} \right\}. \end{aligned} \quad (5.1)$$

Na equação acima, μ_{ij} representa a média condicional do fluxo de comércio entre os países i (exportador) e j (importador), estimada com base nas variáveis explicativas incluídas no modelo.

Dada a possibilidade de modelar o parâmetro de dispersão e proporção de zeros, a próxima abordagem será com modelos GAMLSS. Na Seção 6.2 têm-se os ajustes dos modelos. As equações deste modelo são dadas por:

$$\begin{aligned} \log (\mu_{ij}) &= \beta_0 + \beta_1 \log \left(\frac{Y_i}{\text{Pop}_i} \right) + \beta_2 \log \left(\frac{Y_j}{\text{Pop}_j} \right) + \beta_3 \log (D_{ij}) + \beta_4 \log (\text{Pop}_j) + \beta_5 \text{ILC}_j \\ &\quad + \beta_6 \log (\text{tC}_j) + \beta_7 \text{landlocked}_j + \beta_8 \text{rta}_{ij}, \\ \log (\sigma_{ij}) &= \gamma_0 + \gamma_1 \log \left(\frac{Y_i}{\text{Pop}_i} \right) + \gamma_2 \log \left(\frac{Y_j}{\text{Pop}_j} \right) + \gamma_3 \log (D_{ij}) + \gamma_4 \log (\text{Pop}_j) + \gamma_5 \text{ILC}_j \\ &\quad + \gamma_6 \log (\text{tC}_j) + \gamma_7 \text{landlocked}_j + \gamma_8 \text{rta}_{ij}, \\ \log \left(\frac{v_{ij}}{1 - v_{ij}} \right) &= \alpha_0 + \alpha_1 \log \left(\frac{Y_i}{\text{Pop}_i} \right) + \alpha_2 \log \left(\frac{Y_j}{\text{Pop}_j} \right) + \alpha_3 \log (D_{ij}) + \alpha_4 \log (\text{Pop}_j) \\ &\quad + \alpha_5 \text{ILC}_j + \alpha_6 \log (\text{tC}_j) + \alpha_7 \text{landlocked}_j + \alpha_8 \text{rta}_{ij}. \end{aligned} \quad (5.2)$$

Essas três equações compõem o modelo GAMLSS, no qual são especificadas funções de regressão separadas para os três parâmetros da distribuição da variável resposta X_{ij} , que

representa o fluxo de comércio, medido em termos dos valores das exportações de cera vegetal para o país j . μ_{ij} , é média condicional do fluxo comercial entre i e j ; σ_{ij} é parâmetro de dispersão, que permite modelar a variabilidade em torno da média; ν_{ij} é parâmetro relacionado à inflação de zeros, que captura a probabilidade de observações nulas (quando a distribuição utilizada possui esse componente); β_0 , γ_0 e α_0 representam a constante do modelo de regressão; Y_i e Y_j referem-se ao PIB da região exportadora i e do país importador j , respectivamente; N_i e N_j referem-se à população da região exportadora i e do país importador j , respectivamente; $\frac{Y}{N}$, é a relação PIB/População, produzindo o PIB per capita tanto do país i quanto do país j ; D_{ij} , é a distância geográfica entre o país i e j ; ILC_j é o índice de liberdade comercial do país j ; tC é a taxa de câmbio (real (R\$) / dólar americano (US\$)); $landlocked_j$ é uma variável *dummy* que representa se o país j tem saída ao mar; rta é a *dummy* refere-se os países (i e j) têm acordos comerciais vigentes.

Certamente, há outras variáveis que poderiam ser incluídas no modelo, porém, para este estudo, esta equação apresenta-se de satisfatória abrangência. Aplicadas as notações fundamentais ao modelo, parte-se para explicar o comportamento das variáveis estudadas.

6 ANÁLISE INFERENCIAL

Ao longo dos capítulos anteriores foram apresentados os fundamentos teóricos e metodológicos que sustentam a modelagem dos fluxos comerciais com base na estrutura gravitacional. Essa fundamentação permitiu delinear os principais determinantes econômicos e institucionais do comércio internacional e discutir os desafios empíricos associados à presença de zeros, à heterocedasticidade e à assimetria nos dados. A partir dessas considerações, este capítulo tem como objetivo aplicar diferentes estratégias de estimação, incluindo abordagens clássicas e métodos mais flexíveis, a fim de avaliar o modelo proposto e explorar suas implicações.

6.1 Modelos Lineares Generalizados: PPML e GPML

Os modelos lineares generalizados PPML e *Gamma Pseudo Maximum Likelihood* (GPML) são frequentemente empregados na literatura para lidar com a natureza heterocedástica dos dados de fluxos comerciais. Para garantir a comparabilidade entre as abordagens neste trabalho, todos os modelos seguem a especificação apresentada em (5.1).

A Tabela 2 exibe os resultados da estimação do modelo gravitacional utilizando duas abordagens: o PPML e o GPML. Tais especificações foram escolhidas por sua capacidade de acomodar características específicas dos dados de comércio internacional, como a alta assimetria da variável resposta e a presença de fluxos nulos.

Tabela 2 – Estimativas do modelo gravitacional para os modelos MLG.

Variáveis	PPML		GPML	
	Estimativa	(Erro padrão)	Estimativa	(Erro padrão)
Constante	-20,468***	(0,97)	-9,394***	(0,74)
$\log PIB_{pci}$	0,334***	(0,06)	0,312***	(0,06)
$\log PIB_{pcj}$	1,143***	(0,05)	0,687***	(0,03)
$\log Dist_{ij}$	0,128*	(0,06)	-0,490***	(0,05)
$\log tC$	-0,298***	(0,05)	-0,384***	(0,06)
$\log Pop_j$	1,100***	(0,02)	1,004***	(0,03)
ILC_j	0,008*	(0,00)	0,025***	(0,00)
$landlocked_j$	-1,055***	(0,06)	-0,623***	(0,07)
rta	0,086	(0,09)	-0,670***	(0,07)
Num. obs.	1232		1005	

Fonte: Elaboração própria a partir de estimações.

Nota: Erros-padrão entre parênteses. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

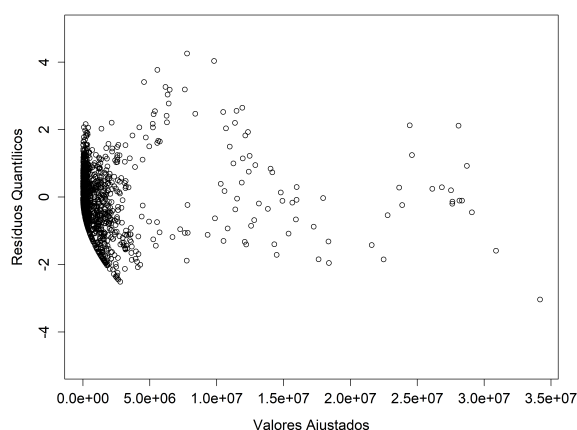
De modo geral, os sinais e significâncias estimados estão alinhados com a teoria do modelo gravitacional e com a literatura, embora algumas diferenças relevantes mereçam destaque

entre as duas especificações. As diferenças refletem as formas com que cada modelo lida com os fluxos comerciais nulos e positivos. O PPML incorpora as observações nulas e tende a suavizar os efeitos de penalização associados à distância, enquanto o GPML, por lidar apenas com valores positivos, evidencia mais fortemente o impacto negativo da distância sobre o volume exportado. Essa divergência pode ser atribuída à maior sensibilidade do GPML a variações na dispersão dos dados.

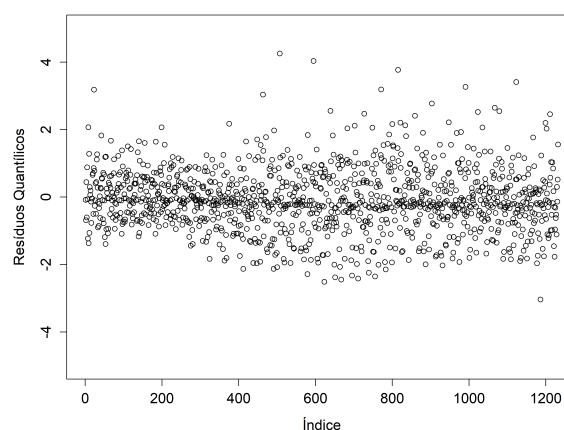
Os gráficos de diagnóstico do modelo com distribuição PPML (Figura 6) apresentam sinais evidentes de sobredispersão nos resíduos, indicando que a variância dos dados excede sistematicamente a média.

Figura 6 – Gráficos de diagnóstico do ajuste do modelo com distribuição PPML.

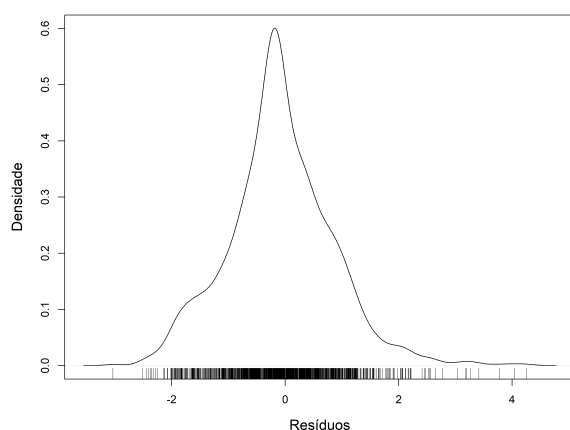
(a) Resíduos vs Valores Ajustados.



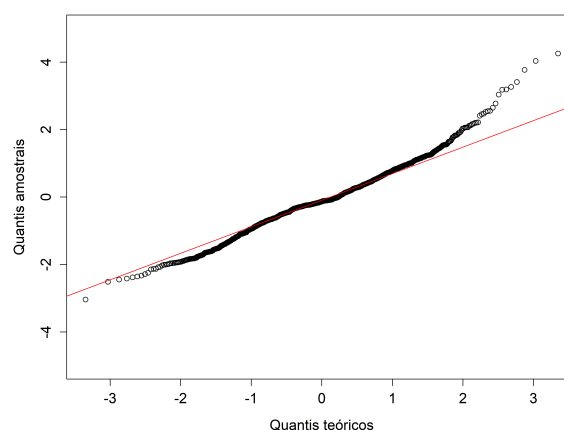
(b) Resíduos vs Índice



(c) Estimativa da Densidade.



(d) Normal QQ-Plot.



Fonte: Elaboração própria.

Nota: Diagnóstico baseado em resíduos quantílicos.

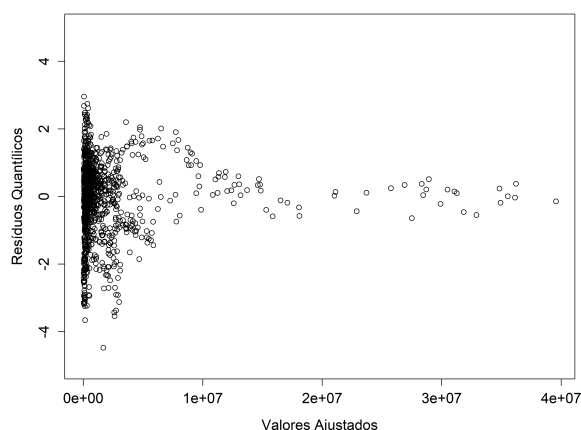
No gráfico de resíduos versus valores ajustados (Figura 6a), nota-se a presença de heterocedasticidade, com maior dispersão dos resíduos à medida que os valores ajustados

aumentam. Além disso, há indícios de assimetria, com resíduos concentrados próximos de zero e cauda longa à direita. O gráfico de resíduos versus índice (Figura 6b) sugere relativa aleatoriedade em torno da linha central, embora seja possível identificar pontos extremos. A estimativa da densidade dos resíduos (Figura 6c) reforça a presença de assimetria, com pico elevado em torno de zero e cauda mais pronunciada à direita, o que se distância da distribuição normal. Esse diagnóstico é confirmado pelo QQ-Plot (Figura 6d), em que os resíduos acompanham a linha de referência apenas na região central, mas apresentam desvios consideráveis nas caudas, sobretudo na superior.

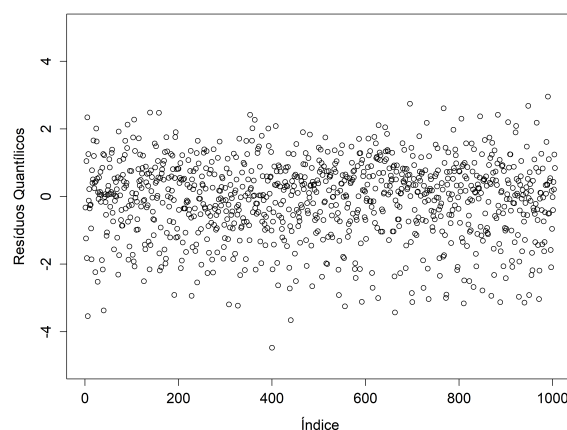
A Figura 7, por sua vez, apresenta os gráficos de diagnóstico para o modelo ajustado sob a distribuição GPML. Comparativamente, verifica-se uma melhora substancial na qualidade do ajuste.

Figura 7 – Gráficos de diagnóstico do ajuste do modelo com distribuição GPML.

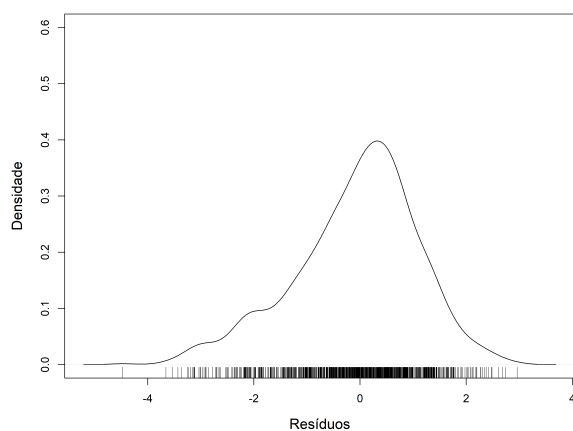
(a) Resíduos vs Valores Ajustados.



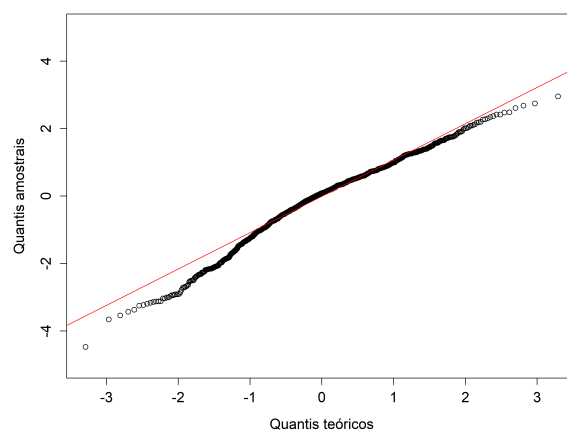
(b) Resíduos vs Índice



(c) Estimativa da Densidade.



(d) Normal QQ-Plot.



Fonte: Elaboração própria.

Nota: Diagnóstico baseados em resíduos quantílicos.

O gráfico de resíduos versus valores ajustados (Figura 7a) também aponta para a presença de heterocedasticidade, embora em menor intensidade do que no PPML. A dispersão é mais homogênea, com menor concentração extrema de resíduos. O gráfico de resíduos versus índice (Figura 7b) não revela padrões sistemáticos importantes, indicando que os resíduos estão mais bem distribuídos em torno da média zero.

A estimativa da densidade (Figura 7c) mostra uma distribuição mais simétrica, com caudas menos acentuadas, aproximando-se mais do comportamento esperado. No QQ-Plot (Figura 7d), observa-se maior alinhamento dos resíduos à reta de referência em comparação ao PPML, embora persistam desvios nas caudas. Tal desempenho é compatível com o uso dessa distribuição para modelar variáveis contínuas e estritamente positivas. De forma geral, o modelo GPML apresenta desempenho superior na adequação dos resíduos.

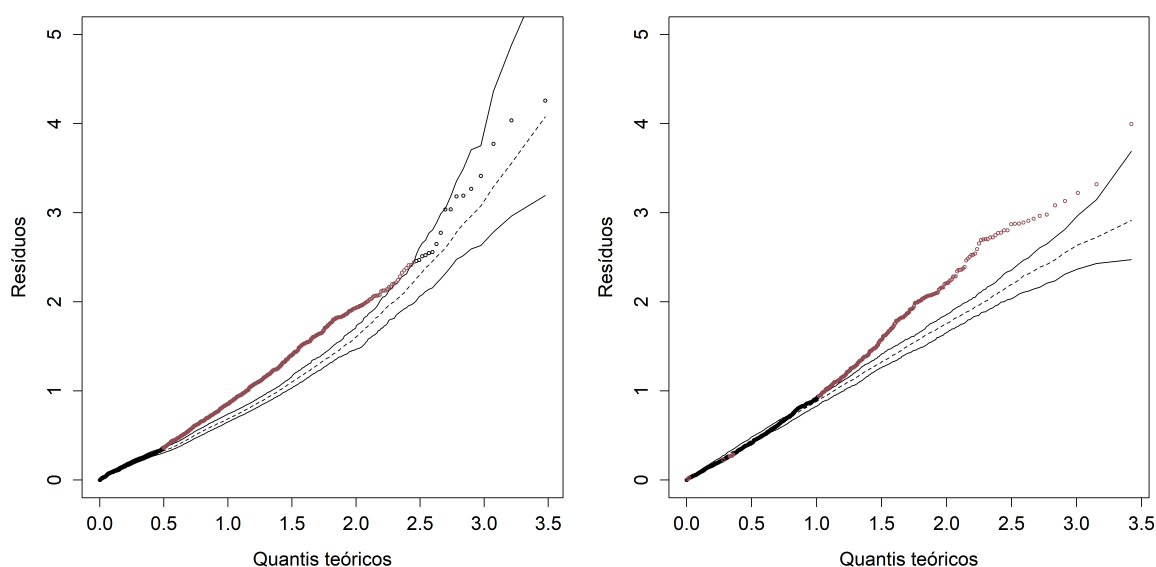
As análises gráficas dos resíduos, ilustradas nas Figuras 6 e 7, revelam que o modelo com distribuição PPML apresenta sinais evidentes de sobredispersão e padrões estruturais nos resíduos, indicando inadequação na representação da variabilidade dos dados. Em contrapartida, o modelo GPML apresenta melhor desempenho diagnóstico, com menor estrutura nos resíduos e maior aderência à distribuição teórica esperada, especialmente nos quantis centrais, embora ainda haja desvios nos extremos.

A Figura 8 apresenta os *Half-Normal Plots* com envelopes simulados para os resíduos dos modelos PPML e GPML.

Figura 8 – *Half-Normal Plots* com envelopes simulados para os resíduos dos modelos.

(a) PPML

(b) GPML



Fonte: Elaboração própria.

No modelo PPML (Figura 8a), cerca de 60,63% (747) dos pontos ficaram fora do envelope, indicando desempenho insatisfatório e confirmando o desvio substancial dos resíduos em relação ao comportamento esperado sob a hipótese do modelo. Já o GPML (Figura 8b) apresentou melhora relativa, com 30,73% (339) dos pontos fora do envelope, o que representa, de certa forma, maior aderência em comparação ao PPML. No entanto, embora o ajuste seja superior, ainda não atinge um nível de adequação considerado satisfatório, evidenciando limitações mesmo sob a especificação da distribuição gama.

Cabe destacar que o modelo GPML não acomoda observações com valor zero na variável dependente, levando à sua exclusão do processo de estimação. Essa limitação metodológica pode influenciar significativamente os resultados dos testes aplicados aos resíduos. Assim, parte do desempenho distinto do GPML pode refletir um viés de seleção, e não necessariamente uma adequação estrutural superior ao modelo PPML.

Além da análise visual, foram aplicados testes estatísticos aos resíduos, a fim de verificar a presença de heterocedasticidade (teste de Breusch-Pagan), autocorrelação serial (Breusch-Godfrey e Durbin-Watson) e desvios da normalidade (Shapiro-Wilk, Jarque-Bera e Kolmogorov-Smirnov). O uso desses procedimentos segue o referencial metodológico adotado por Silva (2015), Bobato *et al.* (2020) e Corrêa (2021), que destacam a importância desses testes para assegurar a robustez das inferências.

O modelo PPML (Tabela 3) apresentou violações significativas de algumas das suposições clássicas da econometria. Os testes de Breusch-Pagan indicaram forte presença de heterocedasticidade nos resíduos (Figura 6a). Embora os testes de Breusch-Godfrey (ordens 1 e 2) e de Durbin-Watson não tenham apontado evidências de autocorrelação serial, os testes de normalidade (Shapiro-Wilk, Jarque-Bera e Kolmogorov-Smirnov) rejeitaram a hipótese nula de normalidade, evidenciando distorções estruturais na distribuição dos resíduos (Figura 6c).

Tabela 3 – Testes estatísticos aplicados aos resíduos do modelo PPML.

Teste	Estatística	<i>p</i> -valor	Conclusão (nível 5%)
Breusch-Pagan (Heterocedasticidade)	87,329	< 0,001	rejeita H_0
Breusch-Godfrey (ordem 1)	0,735	0,3912	não rejeita H_0
Breusch-Godfrey (ordem 2)	3,984	0,1364	não rejeita H_0
Durbin-Watson (Autocorrelação)	1,939	0,1344	não rejeita H_0
Shapiro-Wilk (Normalidade)	0,982	< 0,001	rejeita H_0
Jarque-Bera (Normalidade)	119,400	< 0,001	rejeita H_0
Kolmogorov-Smirnov (Normalidade)	0,057	0,0005	rejeita H_0

Fonte: Elaboração própria.

Esses resultados sugerem que, apesar do PPML ser amplamente recomendado na literatura, principalmente por lidar adequadamente com fluxos comerciais nulos e preservar a consistência na presença de heterocedasticidade, sua eficiência pode ser afetada em contextos de elevada variabilidade e distribuição assimétrica dos fluxos comerciais positivos, como no presente conjunto de dados.

Em contraste, o modelo GPML (Tabela 4), embora também apresente rejeição à hipótese de normalidade dos resíduos, não demonstrou desempenho superior ao PPML no que se refere às propriedades dos resíduos. Os testes de Breusch-Godfrey indicaram ausência de autocorrelação de ordem 1, mas evidenciaram dependência serial de ordem 2, resultado corroborado pelo teste de Durbin-Watson (Figura 7b). Ademais, a presença de heterocedasticidade foi confirmada pelo teste de Breusch-Pagan.

Tabela 4 – Testes estatísticos aplicados aos resíduos do modelo GPML.

Teste	Estatística	<i>p</i> -valor	Conclusão (nível 5%)
Breusch-Pagan (Heterocedasticidade)	15,574	0,0001	rejeita H_0
Breusch-Godfrey (ordem 1)	0,043	0,8344	Não rejeita H_0
Breusch-Godfrey (ordem 2)	22,117	< 0,001	rejeita H_0
Durbin-Watson (Autocorrelação)	1,815	0,0014	rejeita H_0
Shapiro-Wilk (Normalidade)	0,979	< 0,001	rejeita H_0
Jarque-Bera (Normalidade)	47,097	< 0,001	rejeita H_0
Kolmogorov-Smirnov (Normalidade)	0,064	0,0004	rejeita H_0

Fonte: Elaboração própria.

O modelo GPML, embora represente uma alternativa flexível, não está isento de violações relevantes, especialmente relacionadas à heterocedasticidade e autocorrelação serial de maior ordem, características recorrentes em dados comerciais e macroeconômicos em painel. Diante desses resultados, torna-se evidente a necessidade de recorrer a alternativas metodológicas mais flexíveis, capazes de lidar explicitamente com padrões de dispersão assimétrica, excesso de observações nulas e variância não constante. Nesse contexto, os modelos da família GAMLSS configuram-se como uma abordagem promissora. A próxima seção é dedicada à sua especificação, estimativa e avaliação empírica no âmbito do presente estudo.

6.2 Modelagem GAMLSS: seleção de distribuições contínuas e inflacionadas em zero

Inicialmente, estimou-se um modelo de referência (5.2) utilizando a função `gamlss()`, adotando-se a distribuição Normal padrão (`family = N0`). Esse procedimento teve caráter

preliminar, sendo empregado unicamente como ponto de comparação. Embora inadequado para fins de inferência final, tal modelo constitui um ponto de partida para a seleção automatizada da distribuição mais apropriada.

Na sequência, empregou-se a função `chooseDist()`¹, que possibilita a escolha automatizada da distribuição mais adequada com base em critérios de informação, como o GAIC (*Generalized Akaike Information Criterion*). Para essa etapa, definiu-se o argumento `type = "realplus"`, de modo a restringir a análise a distribuições contínuas com suporte positivo (a escolha coerente com a natureza da variável resposta, que assume apenas valores positivos). Adicionalmente, incluiu-se o argumento `extra = c()`², permitindo a avaliação de distribuições inflacionadas em zero, como ZAGA, *Zero Inflated Negative Binomial* (ZANBI), *Zero Adjusted Inverse Gaussian* (ZAIG), *Zero Adjusted Poisson* (ZAP) e ZIP, pois permitem modelar separadamente a inflação de zeros e a distribuição dos valores positivos. Também foram consideradas distribuições da família Pareto (PARETO2 e PARETO2o), que se mostram úteis em contextos com forte assimetria e presença de valores extremos, embora não tratem especificamente do problema dos zeros.

A Tabela 5 apresenta os valores do Critério GAIC, calculado sob três penalizações distintas ($k = 2$, $k = 3,84$ e $k = 7,08$). Os resultados indicam que a distribuição ZAGA apresenta o melhor desempenho, com os menores valores de GAIC em todas as penalizações consideradas. A distribuição mostra-se particularmente adequada para modelar variáveis contínuas não negativas com presença significativa de zeros, característica verificada nos dados de exportação analisados.

Tabela 5 – Valores do GAIC para diferentes distribuições (resumo).

Distribuição	GAIC ($k = 2$)	GAIC ($k = 3,84$)	GAIC ($k = 7,08$)
ZAGA	29.685,91	29.735,59	29.824,15
ZAIG	30.692,50	30.742,18	30.830,74
ZANBI	80.638,31	80.687,99	80.776,55
PARETO2	35.486,28	35.519,40	35.578,44
PARETO2o	31.260,08	31.293,20	31.352,24
ZIP / ZAP*	1.922.248.000	1.922.248.000	1.922.248.000
ZIP2*	2.075.074.000	2.075.074.000	2.075.074.000

Fonte: Elaboração própria.

Nota: *: Valor incompatível com a escala dos dados, indicando falha no ajuste do modelo.

¹ <https://rdrr.io/cran/gamlss/man/fitDist.html>

² Pode-se listar as distribuições extras que devem ser tentadas, que não estão na lista de `type = "realplus"`.

Em comparação, a distribuição ZAIG apresentou desempenho inferior à ZAGA, mas ainda melhor que as demais alternativas contínuas, enquanto a PARETO2o teve ajuste intermediário, refletindo alguma capacidade de representar caudas pesadas e assimetria nos dados, embora sem superar a ZAGA. A distribuição PARETO2 apresentou GAIC ainda mais elevado, indicando ajuste menos adequado. Por outro lado, o modelo ZANBI apresentou GAIC significativamente elevados, revelando baixa qualidade de ajuste.

As distribuições inflacionadas de zero, como ZIP, *Zero Inflated Poisson type 2* (ZIP2) e ZAP, não se mostraram adequadas: em alguns casos, retornaram valores extremamente elevados de GAIC e, em outros, não puderam sequer ser estimadas, sugerindo incompatibilidade estrutural com os dados analisados. De forma semelhante, outras distribuições contínuas (como Weibull, Lognormal, Inversa Gaussiana, entre outras) não produziram resultados válidos para o GAIC, retornando NA (*Not Available*), possivelmente em função de dificuldades de convergência nos algoritmos de estimação.

Seguindo o procedimento sugerido por Rigby e Stasinopoulos (2005), a seleção de variáveis foi realizada via critério GAIC, utilizando o método *backward*³ para os parâmetros μ , σ e ν . A seleção de variáveis por eliminação, embora empírica, é uma estratégia comum para evitar sobreparametrização, conforme orientam Stasinopoulos *et al.* (2017). Para μ , mantiveram-se todas as variáveis originalmente especificadas (5.2), dado que todas se mostraram significativas. Para o parâmetro σ , foram removidas as variáveis *landlocked_j* e *ILC_j*, enquanto ν , retiraram-se *landlocked_j* e *log PIBpc_i*.

A Tabela 6 apresenta os valores do GAIC para cada etapa do processo de seleção de variáveis. Verifica-se que os modelos finais estimados para os parâmetros σ e ν apresentam menores valores de GAIC, indicando ajustes mais adequados em relação aos avaliados.

Tabela 6 – Valores do critério GAIC para seleção de variáveis nos ajustes dos parâmetros.

	μ	σ	ν
Modelo Completo	29.685,91	29.685,91	29.685,91
- <i>landlocked_j</i>	-	29.684,53	29.682,11
- <i>ILC_j</i>	-	29.682,97	-
- <i>log PIBpc_i</i>	-	-	29.684,00

Fonte: Elaboração própria.

Ressalta-se que o modelo completo indicado para cada parâmetro refere-se ao

³ O processo inicia com todas as variáveis candidatas incluídas, removendo-se uma a uma aquelas cuja exclusão reduz o AIC, até que nenhuma remoção traga melhoria.

conjunto inicial de variáveis explicativas especificado exclusivamente para aquele parâmetro, conforme a estrutura dos modelos GAMLSS. Assim, o modelo completo de μ pode conter variáveis diferentes daquelas presentes nos modelos completos de σ ou ν .

Dessa forma, as equações finais para os parâmetros são apresentadas a seguir:

$$\begin{aligned}
 \log(\mu_i) &= \beta_0 + \beta_1 \log\left(\frac{Y_i}{Pop_i}\right) + \beta_2 \log\left(\frac{Y_j}{Pop_j}\right) + \beta_3 \log(D_{ij}) + \beta_4 \log(Pop_j) + \beta_5 ILC_j \\
 &\quad + \beta_6 \log(tC_j) + \beta_7 landlocked_j + \beta_8 rta_{ij}, \\
 \log(\sigma_i) &= \gamma_0 + \gamma_1 \log\left(\frac{Y_i}{Pop_i}\right) + \gamma_2 \log\left(\frac{Y_j}{Pop_j}\right) + \gamma_3 \log(D_{ij}) + \gamma_4 \log(Pop_j) \\
 &\quad + \gamma_5 \log(tC_j) + \gamma_6 rta_{ij}, \\
 \log\left(\frac{\nu_i}{1 - \nu_i}\right) &= \alpha_0 + \alpha_1 \log\left(\frac{Y_j}{Pop_j}\right) + \alpha_2 \log(D_{ij}) + \alpha_3 \log(Pop_j) + \alpha_4 ILC_j \\
 &\quad + \alpha_5 \log(tC_j) + \alpha_6 rta_{ij}.
 \end{aligned}
 \tag{6.1}$$

A Tabela 7 apresenta os coeficientes estimados, erros-padrão e níveis de significância para os coeficientes dos preditores lineares para os três parâmetros da distribuição ZAGA: a média condicional (μ), a dispersão (σ) e a probabilidade de ocorrência de valores zero (ν).

Tabela 7 – Estimativas do modelo ZAGA.

Variável	μ		σ		ν	
	Estimativa	(Erro)	Estimativa	(Erro)	Estimativa	(Erro)
Constante	-9,123***	(0,80)	3,534***	(0,55)	13,588***	(1,67)
$\log PIB_{pci}$	0,332***	(0,06)	0,104*	(0,05)	—	—
$\log PIB_{pcj}$	0,699***	(0,03)	-0,175***	(0,02)	-0,322***	(0,09)
$\log Dist_{ij}$	-0,598***	(0,05)	-0,091*	(0,04)	0,955***	(0,20)
$\log tC$	-0,333***	(0,06)	0,127**	(0,05)	-0,484*	(0,21)
$\log Pop_j$	1,035***	(0,03)	-0,115***	(0,02)	-0,924***	(0,08)
ILC_j	0,025***	(0,00)	—	—	-0,062***	(0,01)
$landlocked_j$	-0,801***	(0,08)	—	—	—	—
rta	-0,652***	(0,07)	-0,407***	(0,06)	-1,573***	(0,21)

Fonte: Elaboração própria com base nas estimativas do modelo.

Nota: Erros-padrão entre parênteses. *** $p < 0,01$; ** $p < 0,05$; * $p < 0,1$.

No parâmetro μ , que representa o valor esperado condicional à existência de fluxo positivo de exportação, observa-se que todas as variáveis explicativas são estatisticamente significativas ao nível de 1%. O sinal dos coeficientes segue, em grande medida, a estrutura teórica do modelo gravitacional: variáveis como $\log PIB_i$, $\log PIB_j$ e $\log Pop_j$ apresentam efeito positivo sobre o volume exportado, enquanto variáveis associadas a barreiras ao comércio, como

$\log Dist_{ij}$, a ausência de acesso ao mar ($landlocked_j$) e a taxa de câmbio ($\log tC$), impactam negativamente o fluxo.

O parâmetro σ , responsável por modelar a variância condicional do modelo, revela que fatores como PIB, distância, população e acordos comerciais influenciam significativamente a dispersão dos fluxos comerciais positivos. Em particular, o sinal negativo de $\log PIB_j$, $\log Pop_j$ e rta sugere que esses elementos estão associados a menor variabilidade nos fluxos, ou seja, relações comerciais mais estáveis e previsíveis entre esses países. Já o PIB do exportador e a taxa de câmbio tendem a aumentar a variabilidade, embora esta última não seja estatisticamente significativa ao nível de 1%.

Por fim, o parâmetro ν , modelado com função de ligação logit, reflete a probabilidade de ausência de comércio entre dois países. Os resultados indicam que variáveis como liberdade econômica, população e acordos comerciais estão negativamente associadas à probabilidade de fluxo nulo, ou seja, favorecem a existência de comércio entre os pares. O PIB per capita do exportador ($\log PIB_i$) não se mostrou estatisticamente significativo nesse componente, indicando que, embora influencie o volume exportado, não necessariamente determina a existência ou não de um fluxo comercial.

A avaliação do modelo GAMLSS com distribuição ZAGA incluiu gráficos de resíduos quantílicos e do *Worm Plot* (envelope simulado). A Figura 9 apresenta os melhores resultados, com apenas 4,63% (57) dos pontos fora do envelope. A clara concentração dos resíduos dentro dos limites simulados indica um ajuste residual satisfatório, corroborando a capacidade do modelo em capturar a estrutura dos dados de forma consistente.

A Figura 10 apresenta o envelope normal simulado do modelo ajustado, gerado com o comando `wp()`. No gráfico, a maioria das observações encontra-se dentro dos limites do envelope, e a linha de tendência dos resíduos (em vermelho) permanece próxima à linha de referência (zero), com pequenas flutuações aleatórias. Isso confirma que não há evidências sistemáticas de má especificação na escolha da distribuição, nem violações severas de normalidade nos resíduos padronizados.

Os gráficos de diagnóstico do modelo com distribuição ZAGA (Figura 11) sugerem um ajuste globalmente satisfatório. No gráfico de resíduos versus valores ajustados (Figura 11a), observa-se maior concentração de resíduos próximos de zero para valores ajustados baixos, mas com dispersão que tende a diminuir à medida que os valores aumentam, indicando leve heterocedasticidade. Apesar disso, não se identificam padrões sistemáticos de tendência, o que é

Figura 9 – *Half-Normal Plot* para os resíduos do modelo ZAGA.

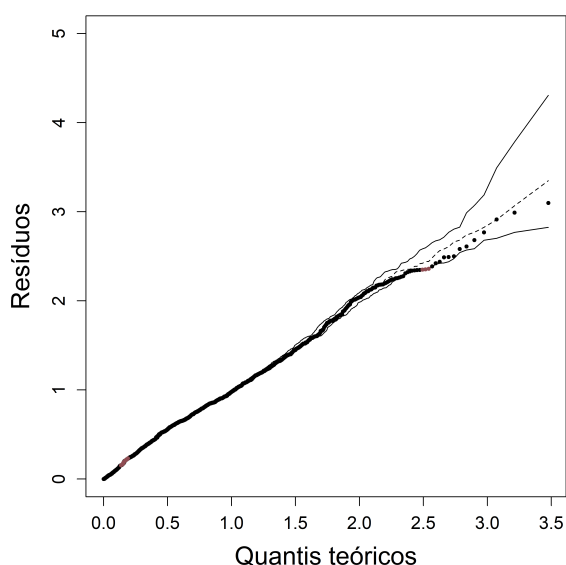
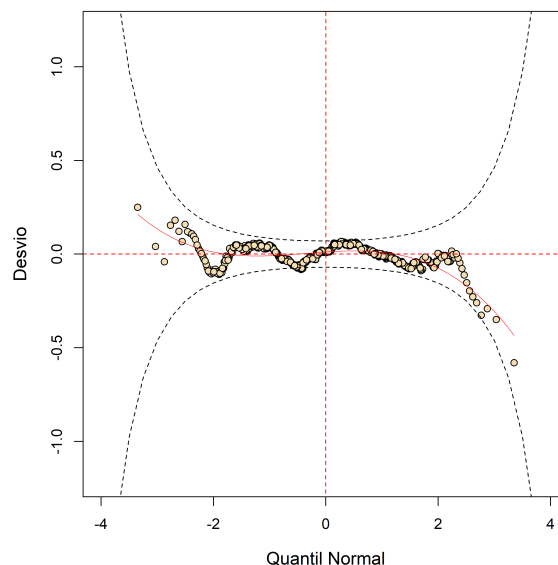


Figura 10 – *Worm Plot* para os resíduos do modelo ZAGA.



Fonte: Elaboração própria.

um índice positivo de ajuste adequado.

O gráfico de resíduos versus índice (Figura 11b) evidencia distribuição aleatória dos resíduos em torno da linha central, sem indicar autocorrelação ou estrutura remanescente nos dados. A estimativa da densidade dos resíduos quantílicos (Figura 11c) mostra distribuição aproximadamente simétrica, sem caudas muito acentuadas, o que representa um avanço em relação aos modelos PPML e GPML. Por fim, o gráfico Q-Q Plot (Figura 11d) confirma esse comportamento, com resíduos aderindo bem à linha de referência ao longo de quase toda a distribuição, apresentando apenas pequenos desvios nas caudas.

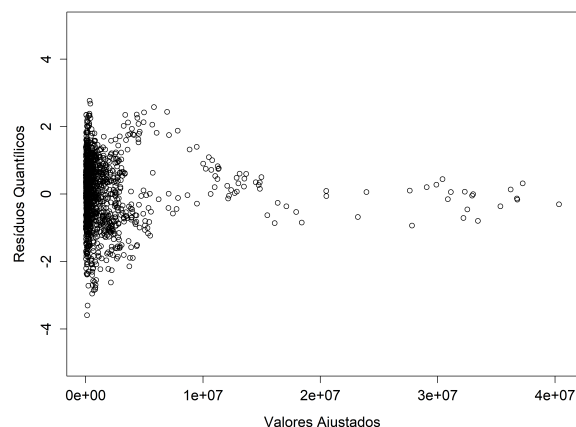
Complementando a análise visual, a Tabela 8 apresenta os resultados dos principais testes estatísticos aplicados aos resíduos do modelo. O teste de Breusch-Pagan indicou a presença de heterocedasticidade nos resíduos. Para verificar autocorrelação, utilizaram-se os testes de Breusch-Godfrey de ordem 1 e 2, além do teste de Durbin-Watson. Todos apontaram a ausência de autocorrelação significativa nos resíduos.

Quanto à normalidade, os testes de Shapiro-Wilk, Jarque-Bera e Kolmogorov-Smirnov não rejeitaram a hipótese nula, confirmam que os resíduos não diferem significativamente da normalidade teórica, conforme demonstrado tanto pelos testes estatísticos quanto pela aderência no gráfico Q-Q Plot (Figura 11d).

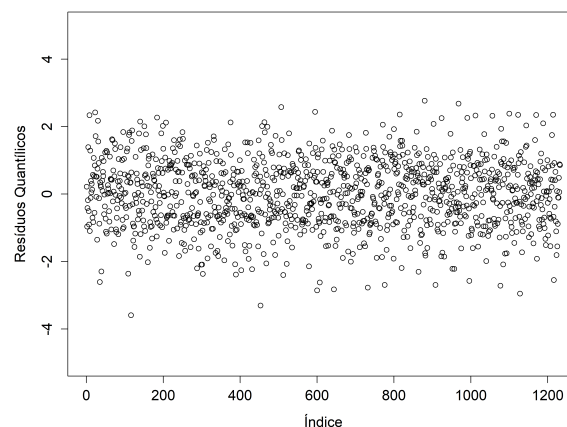
Os resultados apresentados confirmam que a especificação do modelo GAMLSS com distribuição ZAGA é estatisticamente consistente com os dados analisados. Enquanto os modelos

Figura 11 – Gráficos de diagnóstico do ajuste do modelo com distribuição ZAGA.

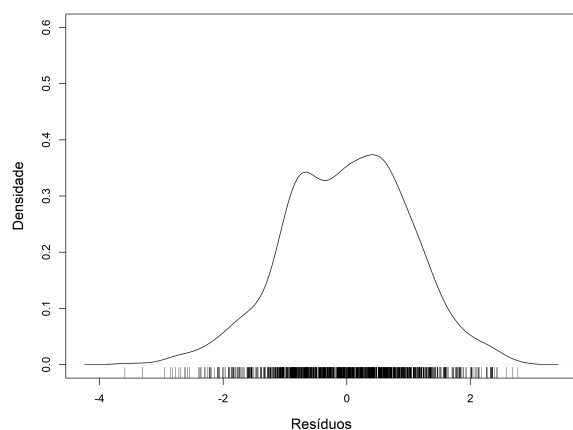
(a) Resíduos vs Valores Ajustados.



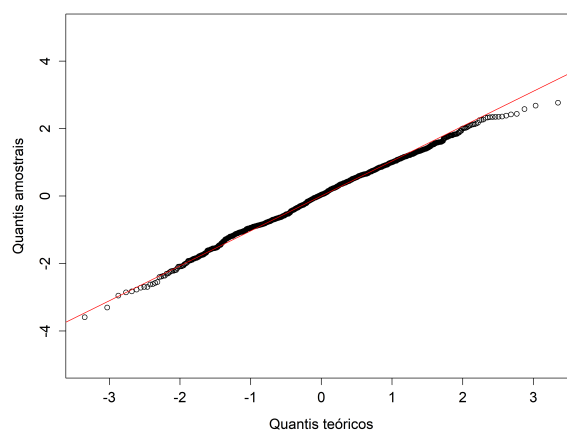
(b) Resíduos vs Índice



(c) Estimativa da Densidade.



(d) Normal QQ-Plot.



Fonte: Elaboração própria.

Nota: Diagnóstico baseado em resíduos quantílicos.

estimados por GPML e PPML apresentaram falhas importantes, como heterocedasticidade combinada com violações significativas da hipótese de normalidade dos resíduos e, no caso do GPML, também autocorrelação serial, o modelo ZAGA destacou-se por gerar resíduos aproximadamente normais e sem evidências de autocorrelação. Além disso, ainda que nenhum dos modelos tenha eliminado totalmente a heterocedasticidade, sua intensidade foi substancialmente menor no ZAGA em comparação aos demais.

6.3 Comparação e Diagnóstico dos Modelos Estimados

Nesta seção são comparadas as especificações PPML, GPML e GAMLSS (ZAGA) aplicadas à modelagem dos fluxos comerciais. Combinamos momentos dos resíduos, densidades e dispersões observadas vs preditores para comparar os ajustes. O objetivo é avaliar a adequação

Tabela 8 – Testes estatísticos aplicados aos resíduos do modelo ZAGA.

Teste	Estatística	p-valor	Conclusão (nível 5%)
Breusch-Pagan (Heterocedasticidade)	8,656	0,0032	rejeita H_0
Breusch-Godfrey (ordem 1)	2,529	0,1117	Não rejeita H_0
Breusch-Godfrey (ordem 2)	2,840	0,2460	Não rejeita H_0
Durbin-Watson (Autocorrelação)	1,939	0,1344	Não rejeita H_0
Shapiro-Wilk (Normalidade)	0,996	0,0091	Não rejeita H_0
Jarque-Bera (Normalidade)	5,603	0,0607	Não rejeita H_0
Kolmogorov-Smirnov (Normalidade)	0,027	0,3036	Não rejeita H_0

Fonte: Elaboração própria a partir dos resultados dos testes.

Nota: H_0 representa a hipótese nula de ausência de problemas (homocedasticidade, independência ou normalidade).

estatística e empírica de cada modelo frente às particularidades dos dados: presença de zeros, heterocedasticidade e assimetria.

A avaliação dos resíduos é fundamental para aferir a qualidade dos ajustes. A Tabela 9 resume os quatro primeiros momentos estatísticos dos resíduos quantílicos de cada modelo. De acordo com a literatura, espera-se que resíduos bem comportados apresentem média próxima de zero, variância igual a um, assimetria nula e curtose em torno de três, valores que caracterizam uma distribuição normal padrão e, portanto, indicam ajuste adequado do modelo.

Tabela 9 – Momentos estatísticos dos resíduos dos modelos estimados.

Modelo	Média	Variância	Assimetria	Curtose
PPML	-0,084	0,863	0,395	4,303
GPML	-0,075	1,356	-0,537	3,251
ZAGA	0,009	0,999	-0,165	2,990

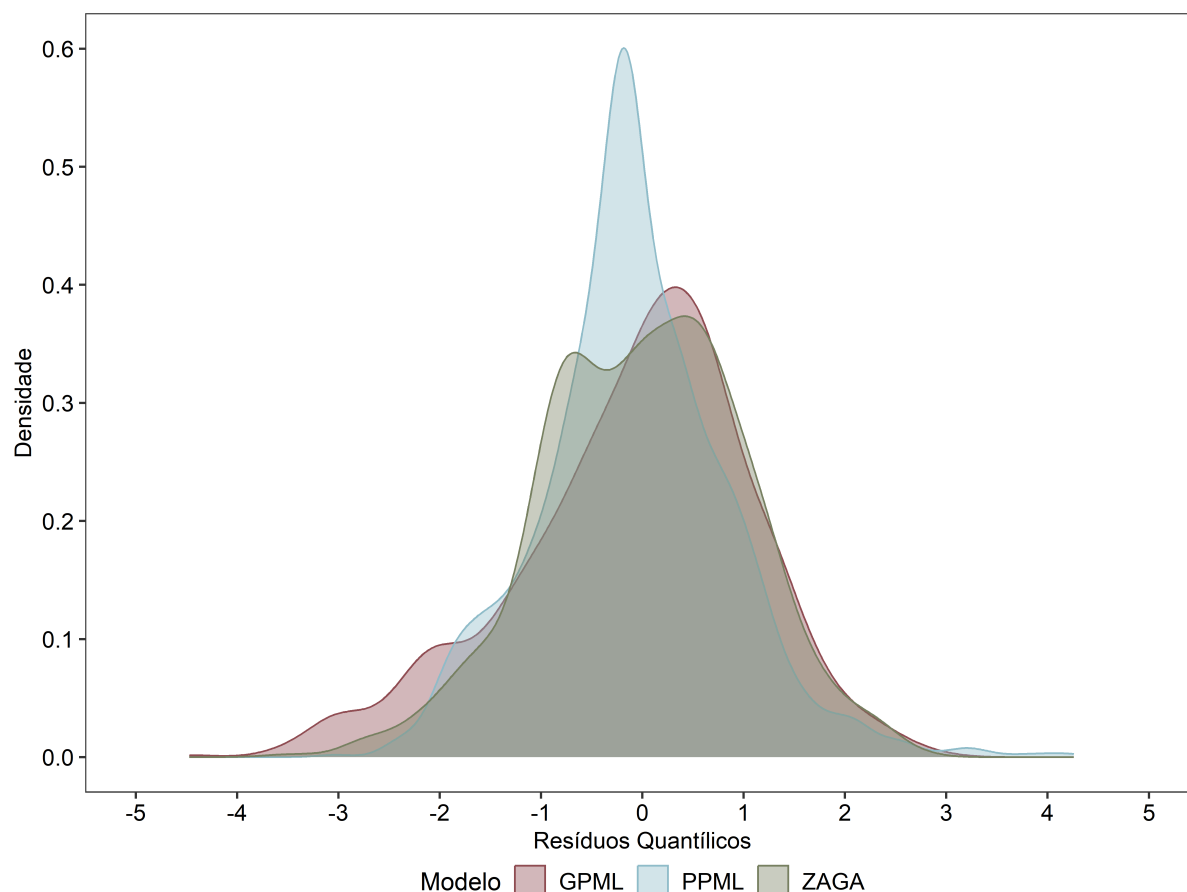
Nota: Diagnóstico baseado em resíduos quantílicos padronizados.

Os resíduos do modelo ZAGA se aproximam mais da normalidade padrão: média próxima de zero, variância aproximadamente um, assimetria próxima de zero e curtose em torno de três. Em contraste, o PPML apresenta curtose elevada e assimetria positiva, indicando caudas pesadas à direita e possível subestimação da variabilidade. No GPML, a assimetria negativa e variância inflacionada sugerem dispersão excessiva e leve deslocamento da distribuição.

Essas observações quantitativas são reforçadas pela Figura 12, que apresenta as densidades empíricas estimadas dos resíduos quantílicos. A análise da densidade dos resíduos quantílicos (Figura 12) reforça as evidências obtidas pelos testes estatísticos. A curva associada ao modelo ZAGA apresenta formato mais simétrico e concentrado em torno de zero, indicando maior proximidade com a hipótese de normalidade e, portanto, melhor adequação do ajuste. Em contraste, a densidade do modelo PPML evidencia uma cauda mais alongada à direita, coerente

com a presença de assimetria positiva. Já o modelo GPML apresenta a distribuição de resíduos levemente deslocada para a esquerda, o que reflete a assimetria negativa identificada. Esses resultados confirmam que, entre os três modelos, o ZAGA produz resíduos com distribuição mais bem comportada, sendo estatisticamente mais consistente com os dados analisados.

Figura 12 – Densidade dos resíduos quantílicos dos modelos GPML, PPML e ZAGA.



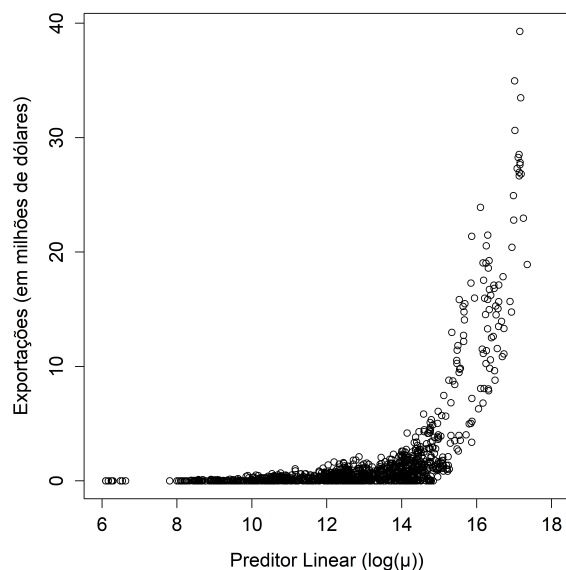
Fonte: Elaboração própria.

A Figura 13, por sua vez, mostra os resultados obtidos a partir dos ajustes via MLG, os gráficos de dispersão entre os valores observados de exportações e os preditores lineares do parâmetro μ . Em todos os casos, observa-se uma relação positiva entre o preditor linear de μ e as exportações, refletindo a adequação dos modelos na captura da tendência central dos dados.

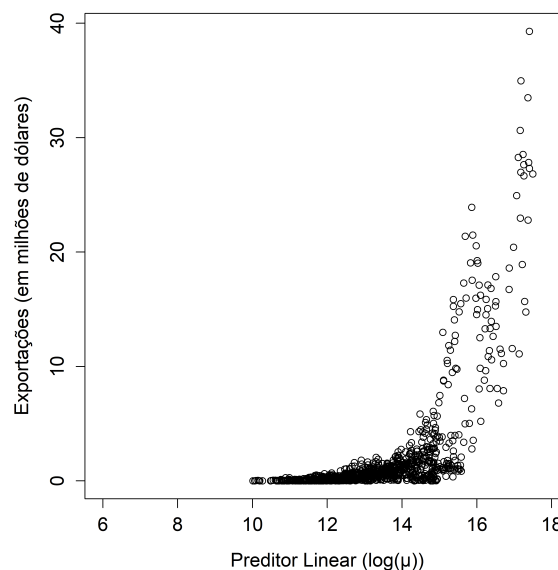
O modelo PPML (Figura 13a) também revela aderência entre o preditor μ e os valores observados, mas apresenta maior densidade de pontos nos extremos inferiores e maior heterogeneidade à medida que o preditor cresce. Já o modelo GPML (Figura 13b) exibe comportamento semelhante, porém com maior concentração de pontos entre $\log(\mu) \in [13, 16]$, sugerindo menor flexibilidade para capturar variações nos extremos da distribuição.

Figura 13 – Exportações observadas em relação aos preditores lineares dos parâmetros do modelo MLG.

(a) Exportações vs Preditor Linear - PPML.



(b) Exportações vs Preditor Linear - GPML.



Fonte: Elaboração própria.

A Figura 14 ilustra a relação entre os valores observados de exportações e os preditores lineares dos parâmetros do modelo ZAGA. O componente μ apresenta relação positiva clara com o nível de exportação (Figura 14a).

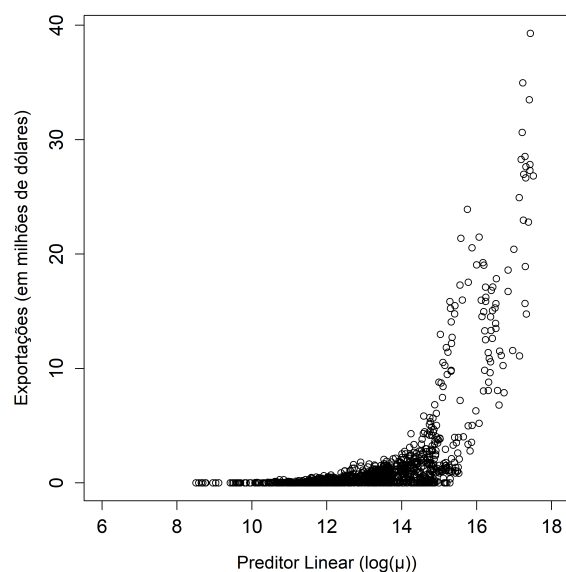
Já o parâmetro σ (Figura 14b) evidencia uma concentração de exportações em baixos níveis do preditor linear, compatível com a heterogeneidade dos fluxos de menor magnitude, enquanto valores mais elevados de $\log(\sigma)$ estão associados a menor dispersão, sugerindo que o modelo captura de forma adequada a redução relativa da variabilidade em fluxos maiores. No caso do parâmetro ν (Figura 14c), observa-se uma relação inversa com os valores de exportações: níveis mais altos de $\text{logit}(\nu)$ concentram observações nulas, enquanto valores negativos estão relacionados a fluxos positivos, o que evidencia a capacidade do modelo em incorporar explicitamente a ocorrência de zeros.

6.4 Interpretação dos Coeficientes do Modelo Final

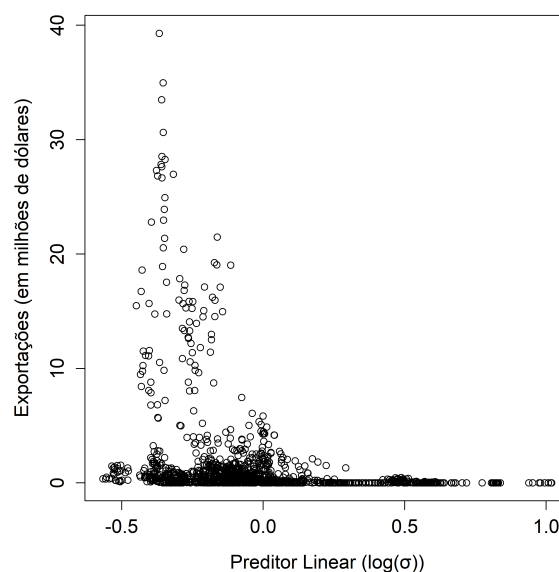
A Tabela 7 (anteriormente apresentada na seção 6.2) é retomada aqui para facilitar a interpretação detalhada dos coeficientes estimados no modelo final. A modelagem foi conduzida com três equações distintas: a primeira representa o componente de localização (μ), a segunda modela a dispersão (σ) dos fluxos positivos e a terceira incorpora o parâmetro de inflacionamento de zeros (ν), todos estimados simultaneamente. A escolha das funções de ligação seguiu a

Figura 14 – Exportações observadas em relação aos preditores lineares dos parâmetros do modelo GAMLSS.

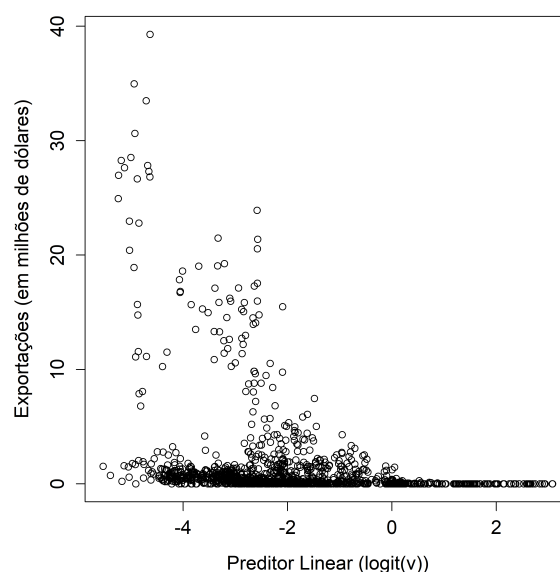
(a) Preditor Linear (μ).



(b) Preditor Linear (σ).



(c) Preditor Linear (ν).

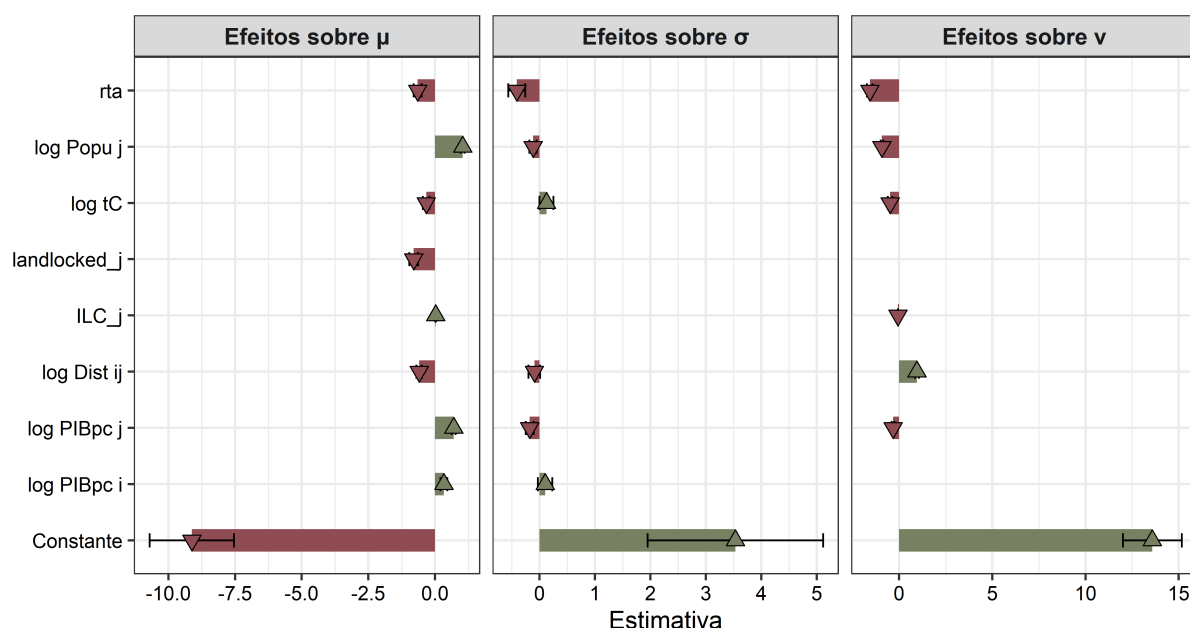


configuração mais adequada à natureza de cada parâmetro: logarítmica para μ e σ , de modo a capturar efeitos multiplicativos sobre os fluxos e sua dispersão, e logística para ν , apropriada para modelar probabilidades.

A Figura 15 sintetiza os efeitos estimados para cada variável explicativa nos três parâmetros do modelo. A representação gráfica facilita a visualização da direção e magnitude dos impactos de cada variável explicativa, permitindo identificar quais determinantes exercem influência ampliadora ou redutora sobre os fluxos médios, sobre sua variabilidade e sobre a

chance de não haver comércio.

Figura 15 – Efeitos estimados das variáveis explicativas sobre os parâmetros do modelo ZAGA.



Nota: i) μ , associado à média condicional dos fluxos comerciais positivos; (ii) σ , que representa a dispersão relativa em torno da média; e (iii) ν , referente à probabilidade de ocorrência de zeros estruturais. Os valores apresentados correspondem aos coeficientes estimados com intervalos de confiança, transformados segundo as funções de ligação.

Os resultados para o componente μ corroboram os efeitos tradicionais previstos pela teoria gravitacional. Tanto o PIB per capita do exportador quanto o do importador exibem coeficientes positivos e estatisticamente significativos, reforçando a ideia de que o tamanho da economia é um fator central para explicar os fluxos de comércio (Anderson e Wincoop, 2003). Um aumento de 10% no PIB per capita brasileiro está associado a um incremento médio de 3,21% nas exportações, enquanto o crescimento de igual magnitude no PIB per capita dos países importadores aumenta 6,89% nos fluxos.

A população do país importador também apresenta impacto positivo, sinalizando que mercados mais amplos, além de serem mais ricos, constituem maior potencial de demanda. Onde, aumento de 10% da população eleva as exportações em 10,37%, reforçando o papel do tamanho de mercado na atração do fluxo comercial. Por outro lado, a distância entre os países atua como barreira ao comércio: um acréscimo de 10% na distância aparece com o sinal esperado, refletindo os custos de transporte e de informação, em consonância com Arevalo *et al.* (2016). Assim, sugerindo que um aumento de 10% na distância reduz em aproximadamente 5,54% o comércio bilateral.

Entre as variáveis dummies, o índice de liberdade econômica apresenta efeito positivo,

com um acréscimo de uma unidade resultando em aumento médio de 2,5% nas exportações, sugerindo que ambientes mais abertos favorecem a integração comercial. Em contraste, a condição de país importador sem saída ao mar reduz o comércio em 55,11%, evidenciando a desvantagem estrutural de não possuir acesso marítimo, aspecto exposto por Silva (2015). A existência de saída para o oceano contribui positivamente para os fluxos comerciais entre o Brasil e seus parceiros. Surpreendentemente, a existência de *Regional Trade Agreements* (rta) se mostrou estatisticamente significativa, porém com efeito negativo, estando associada a uma queda média de 47,9%. A literatura gravitacional frequentemente encontra efeito positivo para a integração regional, uma vez que a redução de barreiras tarifárias e não tarifárias deveria estimular o comércio intra-bloco. No entanto, neste estudo, a estimativa sugere que a mera existência de acordos não se traduziu em ganhos efetivos de comércio.

No componente de dispersão, responsável por modelar a variabilidade dos fluxos de comércios positivos, observa-se que o PIB per capita do país exportador apresenta coeficiente positivo e significativo. Em termos percentuais, tal resultado indica que um aumento no nível de renda do exportador está associado a uma maior dispersão nos valores exportados. Enquanto o PIB per capita do país importador quanto à distância exibem coeficientes negativos e estatisticamente significativos, sugerindo que fluxos destinados a países de maior renda ou a parceiros mais distantes tendem a reduzir a variabilidade. O resultado apresenta-se consistente com a própria base de dados e com os resultados da seção 4.2, enquanto os principais destinos do Brasil: Alemanha, China, Estados Unidos e Japão, constituem mercados ricos e geograficamente distantes, nos quais o padrão de comércio tende a ser mais previsível e concentrado.

Adicionalmente, verifica-se que os acordos regionais reduzem a dispersão em aproximadamente 33,5%. Esse resultado contrasta com o impacto negativo sobre a média, mas sugere que, embora os blocos não tenham aumentado o nível de comércio, eles podem ter contribuído para tornar os fluxos mais estáveis, algo já mencionado por Bobato *et al.* (2020) no contexto de acordos bilaterais brasileiros. Ou seja, esses arranjos podem funcionar mais como garantidores de regularidade de fornecimento do que como motores de crescimento dos volumes médios.

A equação associada ao parâmetro ν modela a probabilidade de ocorrência de zeros nos fluxos comerciais, ou seja, a chance de inexistência de relações comerciais entre pares de países. Como a função de ligação é logística, os coeficientes refletem mudanças nas odds (razão de chances) de comércio nulo.

Os resultados indicam que tanto o PIB per capita quanto a população do país importador reduzem significativamente a chance de fluxos nulos. Especificamente, um aumento de 10% no PIB per capita do importador reduz em torno de 3% as odds de zeros, ao passo que o crescimento da população contribui para reduzir em 8,4% as chances de comércio nulo, refletindo maior tamanho de mercado. Conforme Helpman *et al.* (2008), países maiores e mais ricos são mais propensos a participar do comércio internacional, reduzindo a ausência de trocas comerciais.

Em contrapartida, a distância apresentou coeficiente positivo, indicando que, quanto maior for a separação geográfica, maior a chance de não haver comércio entre dois países. Em termos percentuais, um aumento de 10% na distância aumenta as odds em 9,5%, reforçando o peso dos custos de transporte. Em outras palavras, o aumento na distância entre os mercados tende a diminuir o comércio, ou seja, quanto maior for a distância, mais desfavorecido será o mercado (Arevalo *et al.*, 2016). Adicionalmente, o índice de liberdade econômica do parceiro comercial apresenta coeficiente negativo e significativo, ou seja, exerce influência favorável ao comércio, onde o aumento de uma unidade reduz em 6% as chances de ausência de fluxos, enquanto um incremento de 10% na taxa de câmbio diminui as odds em 4,5%.

Por fim, a variável associada à participação em acordos regionais mostra coeficiente negativo e altamente significativo, efeito mais expressivo, diminuindo as odds em cerca de 79,3%, mostrando que, mesmo não elevando o volume médio, eles aumentam significativamente as chances de que haja pelo menos algum fluxo comercial. Assim, corroborando com a expectativa teórica de que acordos regionais podem funcionar como “porta de entrada” que garante algum nível de comércio, ainda que não consiga impulsionar volumes médios expressivos.

6.4.1 Efeito dos acordos regionais (rta)

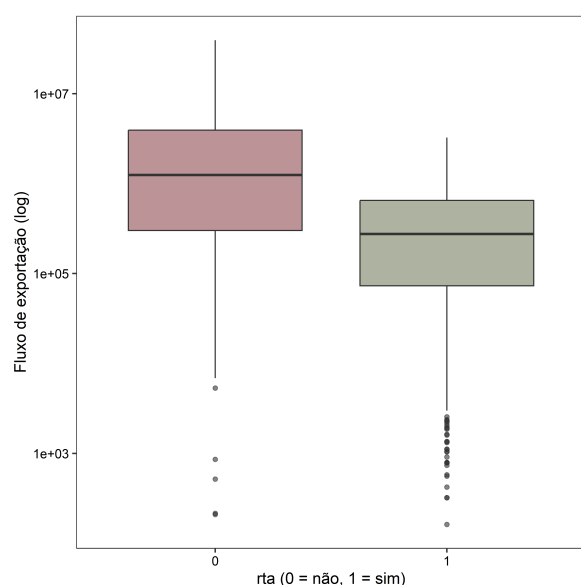
O impacto negativo da *dummy* de acordos comerciais (rta) no componente de média, em relação ao produto analisado, indica que a mera presença de acordos não resultou em um aumento nos volumes exportados. Assim, esse resultado diverge da expectativa convencional, que geralmente associa efeitos positivos à integração regional devido à diminuição de barreiras tarifárias e não tarifárias. Uma interpretação plausível é que os acordos regionais firmados pelo Brasil mostram que o produto em questão enfrenta barreiras não tarifárias (sanitárias, técnicas ou ambientais) que atenuam os benefícios tarifários.

A Figura 16 ilustra a distribuição dos fluxos de exportação conforme a presença ou

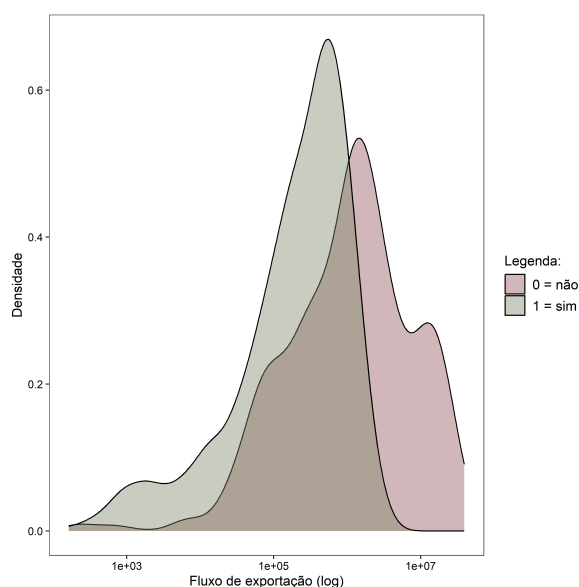
não de acordos regionais. Observa-se que, apesar de as medianas e médias se apresentarem mais elevadas na falta de acordos, os fluxos em sua presença exibem menor dispersão, corroborando a análise de que os rtas atuam como dispositivos de estabilização, diminuindo a frequência de zeros e de valores extremos negativos, ao invés de funcionarem como impulsionadores da elevação dos volumes médios.

Figura 16 – Distribuição dos fluxos de exportação conforme a presença (rta = 1) ou ausência (rta = 0) de acordos regionais.

(a) Boxplot por acordo comercial.



(b) Densidade (escala log) por acordo comercial.



Fonte: Elaboração própria.

O boxplot da Figura 16a mostra que, embora os fluxos de exportação sejam mais elevados, em média, para pares de países sem acordos, os fluxos entre países signatários de rtas apresentam menor dispersão, sinalizando maior estabilidade. Essa evidência é coerente com o efeito estimado para o componente de dispersão, no qual os acordos regionais reduzem significativamente a variabilidade. Em outras palavras, o acordo não garante que os volumes sejam maiores, mas parece reduzir a chance de grandes oscilações, conferindo mais previsibilidade ao comércio.

Essa interpretação é corroborada pela curva de densidade em escala log. Nos países que possuem acordo, a distribuição é mais concentrada, apresentando um pico acentuado e caudas menos extensas. Isso implica que casos de exportações muito pequenas ou inexistentes são mais raros, assim como episódios de valores extremos. No caso dos pares sem acordos, a curva se alarga, apresentando uma cauda esquerda mais pesada (com maior ocorrência de fluxos muito baixos) e uma cauda direita mais longa (com alguns fluxos muito altos). Esse gráfico

mostra uma maior heterogeneidade, o que contribui para entender por que a média pode parecer mais alta, com alguns episódios de exportações muito grandes, mas coexistindo com muitos casos de exportações quase inexistentes.

As estatísticas descritivas corroboram o padrão visual. Em geral, os pares com acordos regionais exibem fluxos de exportação mais altos (11,1 contra 9,88), porém as medianas são bastante semelhantes (12,3 e 12,9). Isso sugere que a diferença média é afetada por alguns valores extremos no grupo sem acordo. A principal diferença está na dispersão: o desvio-padrão reduz-se quase pela metade entre os países signatários (3,66 em comparação com 6,49), indicando que os rtas estão ligados a uma distribuição mais estável e menos diversa. Dessa forma, o efeito dos acordos (rta) não se limita apenas ao “quanto” os países exportam, mas também ao “como” esses fluxos se comportam. Os acordos parecem estabelecer um ambiente onde o comércio é menos instável, o que possibilita que as trocas se tornem mais constantes.

7 CONCLUSÃO E TRABALHOS FUTUROS

A proposta inicial deste estudo partiu da necessidade de aprimorar os métodos de estimação aplicados ao modelo gravitacional, tendo em vista as limitações dos modelos tradicionais diante de dados de comércio internacional caracterizados por assimetria, heterocedasticidade e excesso de zeros. Como destacado na justificativa, a inadequação em lidar com essas propriedades estatísticas pode conduzir a interpretações equivocadas e, conseqüentemente, a políticas comerciais mal calibradas. Nesse sentido, a avaliação comparativa entre os modelos PPML, GPML e GAMLSS demonstrou-se crucial para testar a consistência das diferentes abordagens.

Os resultados empíricos confirmaram que, embora o PPML continue sendo uma ferramenta amplamente utilizada e consistente ao incorporar fluxos nulos, seu desempenho foi limitado por forte heterocedasticidade e resíduos não normais. O GPML, por outro lado, ajustou-se de forma mais adequada aos fluxos positivos, mas sua exclusão das observações nulas resultou em viés de seleção e em problemas de autocorrelação serial. Ambas as limitações reafirmam que os modelos lineares generalizados, ainda que úteis, não são suficientes para representar a complexidade dos fluxos comerciais atuais.

Em contraste, o modelo GAMLSS com distribuição ZAGA mostrou-se mais robusto. Ao possibilitar a modelagem simultânea de múltiplos parâmetros, ofereceu uma análise abrangente que não se restringe ao volume médio de comércio, mas incorpora também sua variabilidade e a chance de inexistência de trocas comerciais. Essa flexibilidade resultou em resíduos com comportamento próximo da normalidade, ausência de autocorrelação significativa e redução da intensidade da heterocedasticidade, consolidando sua superioridade metodológica em relação aos modelos tradicionais.

Do ponto de vista substantivo, os resultados reforçam pilares da teoria gravitacional: PIB e população dos parceiros importadores emergiram como determinantes centrais e positivos dos fluxos comerciais, enquanto distância, taxa de câmbio e ausência de saída para o mar se confirmaram como barreiras relevantes. Entretanto, um resultado inovador deste trabalho diz respeito ao papel dos acordos regionais. Contrariando a expectativa tradicional, esses acordos não se associaram a aumentos significativos nos volumes médios exportados; ao contrário, apresentaram efeito negativo sobre a média. Contudo, reduziram a dispersão e a probabilidade de fluxos nulos, sugerindo que sua contribuição não se dá no estímulo direto ao crescimento do comércio, mas na estabilização e continuidade das trocas.

Essas evidências dialogam diretamente com os objetivos propostos: demonstram a capacidade dos GAMLSS de lidar com distribuições assimétricas e inflacionadas em zero, ampliam a compreensão da estrutura dos fluxos comerciais e apontam para implicações de política econômica. Em particular, mostram que acordos regionais podem ser mais eficazes como instrumentos de previsibilidade e segurança comercial do que como catalisadores imediatos de expansão de volumes. Assim, do ponto de vista metodológico e empírico, este estudo contribui para a literatura ao reforçar a relevância dos modelos flexíveis na análise econômica aplicada, em linha com a tendência contemporânea de incorporar técnicas mais sofisticadas na modelagem de fenômenos complexos.

Entretanto, algumas limitações devem ser reconhecidas. Primeiramente, embora o modelo GAMLSS represente um avanço metodológico significativo, ainda enfrenta desafios relacionados à complexidade computacional. Por fim, a análise concentrou-se em variáveis clássicas da equação gravitacional, não incorporando dimensões recentes como qualidade institucional, barreiras não tarifárias ou indicadores ambientais, que poderiam enriquecer a compreensão dos fluxos.

Essas limitações abrem espaço para pesquisas futuras. Uma primeira direção seria expandir a análise para diferentes períodos, setores e produtos, de modo a testar a robustez dos resultados em contextos mais amplos. Outra possibilidade seria integrar os GAMLSS às técnicas de aprendizado de máquina. Ademais, recomenda-se avaliar comparativamente o desempenho dos modelos GAMLSS em relação a abordagens gravitacionais tradicionais, como o PPML e o GPML, de modo a verificar em que medida a maior flexibilidade do GAMLSS representa ganhos efetivos em diferentes contextos empíricos. Por fim, análises que incluam dimensões institucionais, regulatórias e ambientais podem contribuir para uma visão mais abrangente dos determinantes contemporâneos do comércio internacional.

Em síntese, a principal conclusão é que o GAMLSS, especialmente com distribuições inflacionadas em zero como a ZAGA, constitui um avanço significativo frente às abordagens convencionais, não somente por produzir estimativas estatisticamente mais confiáveis, mas também por oferecer uma visão mais rica sobre o funcionamento do comércio internacional. Essa contribuição é particularmente relevante para a formulação de políticas públicas, ao ampliar a capacidade de identificar não somente os fatores que ampliam o comércio, mas também aqueles que tornam as relações mais estáveis e previsíveis ao longo do tempo.

REFERÊNCIAS

- AGUIAR, G. M. B. **Condicionantes da exportação de carne bovina do Brasil: uma abordagem gravitacional**. 47 p. Dissertação (Dissertação Mestrado) – Centro de Ciências Agrárias, Universidade Federal do Ceará, Fortaleza, 2018.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, dez. 1974. Disponível em: <http://dx.doi.org/10.1109/TAC.1974.1100705>. Acesso em: 02 out. 2024.
- ANDERSON, J. E. A theoretical foundation for the gravity equation. **The American economic review**, JSTOR, v. 69, n. 1, p. 106–116, 1979.
- ANDERSON, J. E. The gravity model. **Annual Review of Economics**, v. 3, n. 1, p. 133–160, 2011. Disponível em: <https://doi.org/10.1146/annurev-economics-111809-125114>. Acesso em: set. 2024.
- ANDERSON, J. E.; WINCOOP, E. V. Gravity with gravitas: A solution to the border puzzle. **American economic review**, American Economic Association, v. 93, n. 1, p. 170–192, 2003.
- AREVALO, J. L. S.; ANDRADE, Á. M. F. d.; SILVA, G. A. B. Uma nota sobre modelos gravitacionais aplicados à exportação de café de Brasil, Colômbia e Peru. **Revista Brasileira de Economia**, SciELO Brasil, v. 70, p. 271–280, 2016.
- ATAÍDE, E. M.; LISBÔA, A. M. G.; BASTOS, D.; SOUZA, J.; COSTA, R. Cera de carnaúba e extrato de própolis na conservação pós-colheita de laranja Natal CNPMF 112 em condição refrigerada. In: ATAÍDE, E. M.; LISBÔA, A. M. G.; BASTOS, D.; SOUZA, J.; COSTA, R. **Anais do Congresso Brasileiro de Fruticultura**, 26., Petrolina: Embrapa Semiárido, UNIVASF, SBF, 2019. Fruticultura de precisão: desafios e oportunidades.
- ATKINSON, A. C. Plots, transformations, and regression an introduction to graphical methods of diagnostic regression analysis. **Clarendon Press**, Oxford, England Clarendon, 1985.
- BAIER, S. L.; BERGSTRAND, J. H. Do free trade agreements actually increase members' international trade? **Journal of international Economics**, Elsevier, v. 71, n. 1, p. 72–95, 2007.
- BANCO CENTRAL DO BRASIL. **Taxa de câmbio comercial para compra: real (R\$) / dólar americano (US\$) - média (GM366_ERC366)**. 2025. Disponível em: <https://api.bcb.gov.br/>. Acesso em: 08 ago. 2025. Frequência: diária, de 2 jan. 1985 até 07 ago. 2025.
- BERGSTRAND, J. H. The gravity equation in international trade: some microeconomic foundations and empirical evidence. **The review of economics and statistics**, JSTOR, p. 474–481, 1985.
- BERGSTRAND, J. H. The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. **The review of economics and statistics**, JSTOR, p. 143–153, 1989.
- BOBATO, A. M.; CORONEL, D. A.; FEISTEL, P. R. Os determinantes do comércio intraindústria entre o Brasil, a União Europeia e a China, no período de 2006 a 2017: uma aplicação do modelo gravitacional. **Geosul**, v. 35, n. 77, p. 695–731, 2020.

BUUREN, S. v.; FREDRIKS, M. Worm plot: a simple diagnostic device for modelling growth reference curves. **Statistics in Medicine**, Wiley, v. 20, n. 8, p. 1259–1277, abr. 2001. ISSN 1097-0258. Disponível em: <http://dx.doi.org/10.1002/sim.746>. Acesso em: set. 2024.

Câmara Setorial da Carnaúba. A carnaúba: preservação e sustentabilidade. **Fortaleza: Câmara Setorial da Carnaúba**, p. 01–40, 2009. Disponível em: https://www.adece.ce.gov.br/wp-content/uploads/sites/98/2012/10/cartilha_carnauba.pdf. Acesso em: jun. 2025.

CARVALHO, J. N. F. de; GOMES, J. M. A. Pobreza, emprego e renda na economia da carnaúba. **Revista Econômica do Nordeste**, v. 40, n. 2, p. 361–378, 2009. Disponível em: <https://www.bnb.gov.br/revista/ren/article/view/357/306>. Acesso em: 12 jul. 2024.

CASTELO, C. E. F.; MARINHO, R. de S. O comércio internacional do acre com o mercado andino: Identificando problemas. **UÁQUIRI-Revista do Programa de Pós Graduação em Geografia da Universidade Federal do Acre**, v. 3, n. 2, 2021.

CEPII. **Base de Dados do CEPII**. 2024. Disponível em: <http://www.cepii.en>. Acesso em: 08 ago. 2025.

COLE, T. J.; GREEN, P. J. Smoothing reference centile curves: The lms method and penalized likelihood. **Statistics in Medicine**, Wiley, v. 11, n. 10, p. 1305–1319, jan. 1992. ISSN 1097-0258. Disponível em: <http://dx.doi.org/10.1002/sim.4780111005>. Acesso em: 10 jun. 2025.

CORRÊA, M. A. K. **Modelo gravitacional aplicado a bens agrícolas: o caso das exportações brasileiras de açúcar**. 138 p. Tese (Tese de Doutorado em Desenvolvimento Regional e Agronegócio) – Universidade Estadual do Oeste do Paraná, Toledo, 2021.

COSTA, V. L. dos S.; SOARES, T. F. O.; GOMES, J. M. A. Política de garantia de preços mínimos para a cera de carnaúba: comparação entre preços mínimos e preços de mercado das safras de 2003/2004-2011/2012. **INFORME ECONÔMICO (UFPI)**, v. 30, n. 2, 2013.

DEARDORFF, A. Determinants of bilateral trade: Does gravity work in a neoclassical world? In: FRANKEL, J. A. (Ed.). **The Regionalization of the World Economy**. Chicago: University of Chicago Press, 1998. v. 1, p. 7–32. ISBN 0-226-25995-1.

DISDIER, A.-C.; HEAD, K. The puzzling persistence of the distance effect on bilateral trade. **The Review of Economics and statistics**, The MIT Press, v. 90, n. 1, p. 37–48, 2008.

DUARTE, M. E. **Evolução econômica do município de Mossoró/RN**. 119 p. Dissertação (Dissertação Mestrado) – Universidade Federal do Rio Grande do Norte, Centro de Ciências Sociais Aplicadas, Programa de Pós-graduação em Economia, 2019.

DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Informa UK Limited, v. 5, n. 3, p. 236–244, set. 1996. ISSN 1537-2715. Disponível em: <http://dx.doi.org/10.1080/10618600.1996.10474708>. Acesso em: 10 jul. 2024.

EICHENGREEN, B. The real exchange rate and economic growth. **Social and Economic Studies**, JSTOR, p. 7–20, 2007.

FERREIRA, C. R. C. Comércio internacional e o modelo de gravidade: evidências recentes em análise teórica e empírica. **A Economia em Revista - AERE**, v. 29, n. 1, 2021.

FILHO, B. D. L. E. S.; SANTANA, S. C. d. O trabalho escravo e a subordinação estrutural na cadeia produtiva da cera da carnaúba. **Revista da Escola Nacional da Inspeção do Trabalho**, 2020.

GLOBAL GROWTH INSIGHTS. Carnauba wax used in cosmetics market size, growth analysis, forecast 2023-2032. **Global Growth Insights**, 2023. Disponível em: <https://www.globalgrowthinsights.com/pt/market-reports/carnauba-wax-used-in-cosmetics-market-103402>. Acesso em: 13 fev 2025.

GONÇALVES, T. F. C. **A participação da China na expansão do comércio exterior da região nordeste do Brasil**. Dissertação (Trabalho de Conclusão de Curso (Graduação em Relações Internacionais) – Departamento de Relações Internacionais, Centro de Ciências Sociais Aplicadas) – Universidade Federal de Sergipe, São Cristovão, SE, 2021.

GOPINATH, M.; BATARSEH, F.; BECKMAN, J. **Machine Learning in Gravity Models: An Application to Agricultural Trade**. [S. n.], 2020. Disponível em: <http://dx.doi.org/10.3386/w27151>. Acesso em: nov. 2024.

GOURIEROUX, C.; MONFORT, A.; TROGNON, A. Pseudo maximum likelihood methods: Theory. **Econometrica**, JSTOR, v. 52, n. 3, p. 681, maio 1984. ISSN 0012-9682. Disponível em: <http://dx.doi.org/10.2307/1913471>. Acesso em: mai. 2025.

GWARTNEY, J.; LAWSON, R.; MURPHY, R. **Economic Freedom of the World: 2023 Annual Report**. Fraser Institute, 2023. Disponível em: <http://dx.doi.org/10.53095/88975012>.

HASTIE, T. J. Generalized additive models. In: HASTIE, T. J. **Statistical Models in S**. Routledge, 2017. p. 249–307. ISBN 9780203738535. Disponível em: <http://dx.doi.org/10.1201/9780203738535-7>. Acesso em: set. 2024.

HEAD, K.; MAYER, T. Gravity equations: Workhorse, toolkit, and cookbook. In: HEAD, K.; MAYER, T. **Handbook of international economics**. [S. l.]: Elsevier, 2014. v. 4, p. 131–195.

HELPMAN, E.; MELITZ, M.; RUBINSTEIN, Y. Estimating trade flows: Trading partners and trading volumes. **The quarterly journal of economics**, MIT Press, v. 123, n. 2, p. 441–487, 2008.

IBGE. **Atlas Nacional Digital do Brasil 2019: Carnaúba - Existência de extração em quilogramas (2017)**. 2019. Disponível em: <https://l1nk.dev/SA2Qb>. Acesso em: 26 mai. 2025. Formato: Shapefile.

IBGE. **Produção da Extração Vegetal e da Silvicultura**. 2025. Disponível em: <https://sidra.ibge.gov.br/tabela/289>. Acesso em: 26 mai. 2025.

KRUGMAN, P. R.; OBSTFELD, M.; MELITZ, M. J. **Economia Internacional**. São Paulo: Pearson Education do Brasil, 2015.

LINDERS, G.-J.; GROOT, H. L. D. **Estimation of the gravity equation in the presence of zero flows**. [S. l.]: Tinbergen institute discussion paper, 2006.

MARASCHIN, R. V. Os determinantes institucionais do comércio mundial bilateral e por classificação de intensidade tecnológica de produtos, nos anos 2000 a 2016: uma aplicação do modelo gravitacional. **Pontifícia Universidade Católica do Rio Grande do Sul**, 2020.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. Springer US, 1989. ISBN 9781489932426. Disponível em: <http://dx.doi.org/10.1007/978-1-4899-3242-6>. Acesso em: jul. 24.

MDIC. **ComexStat**. 2025. Disponível em: <http://comexstat.mdic.gov.br/pt/geral>. Acesso em: set. 2025.

MENDES, K.; HIDALGO, A. B.; LUCHINE, A. A. O comércio internacional dos municípios brasileiros entre 2000 e 2010: uma aplicação do modelo gravitacional. In: MENDES, K.; HIDALGO, A. B.; LUCHINE, A. A. **Anais do 48º Encontro Nacional de Economia**, [S. l.]: ANPEC, 2020. Evento realizado em formato virtual.

MILLS, E. D. **Adjusting for covariates in zero-inflated gamma and zero-inflated log-normal models for semicontinuous data**. [S. l.]: The University of Iowa, 2013.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. Introduction to linear regression analysis. John Wiley and sons. **Inc. Canada**, 1992.

MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. Half-normal plots and overdispersed models in r: the hnp package. **Journal of Statistical Software**, v. 81, p. 1–23, 2017.

MOREIRA, P. I. Entre extrativismo e ciência: a história da exploração da cera de carnaúba no nordeste do Brasil na primeira metade do século XX. **Historia Ambiental Latinoamericana y Caribeña (HALAC) revista de la Solcha**, v. 12, n. 2, p. 107–139, 2022.

MOREIRA, P. I. **Brasil, terra da carnaúba: ciência, política e indústria no processo de comoditização da cera de carnaúba no Nordeste do Brasil, 1900-1970**. Tese (Tese de Doutorado em História das Ciências e da Saúde) – Casa de Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, 2023. Acesso em: 22 dez. 2024.

NASCIMENTO, M. V. V. do; AGUIAR, C. W. dos S. A geohistória no centro da capital cearense: o papel da casa boris-frères na economia e na sociedade de fortaleza na belle époque. **Revista Caribeña de Ciencias Sociales**, v. 13, n. 8, p. e42226–e42226, 2024.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society. Series A (General)**, JSTOR, v. 135, n. 3, p. 370, 1972. ISSN 0035-9238. Disponível em: <http://dx.doi.org/10.2307/2344614>. Acesso em: jul. 2024.

OLIVEIRA, E. J. de. A centralidade de Mossoró (RN) na rede urbana-regional (1858-1915). **Revista da Casa da Geografia de Sobral (RCGS)**, v. 23, n. 2, p. 192–220, 2021. Disponível em: <https://rcgs.uvanet.br/index.php/RCGS/article/view/769/597>. Acesso em: 12 jan. 2025.

PAULA, G. A. **Modelos de regressão: com apoio computacional**. [S. l.]: IME-USP São Paulo, 2013.

PEREIRA, J. P. A. A desindustrialização do Brasil e seu impacto nas relações de comércio exterior: Causas, consequências e estratégias de recuperação. **Journal of Technology & Information (JTnI)**, v. 4, n. 2, 2024.

PEREIRA, M. d. N. F.; RIBEIRO, C. J. S.; TRACTENBERG, L.; MEDEIROS, P. L. Bases de dados na economia do conhecimento: a questão da qualidade. **Ciência da Informação**, SciELO Brasil, v. 28, p. 215–223, 1999.

PORTES, L. H.; MACHADO, C. V.; TURCI, S. R. B. Coordenação governamental da política de controle do tabaco no brasil. **Ciência & Saúde Coletiva**, SciELO Brasil, v. 24, p. 2701–2714, 2019.

RANGEL, S. B.; NETO, J. D. S.; AMIN, M. Demanda externa de cera de carnaúba: uma análise econométrica. **Revista de Economia e Sociologia Rural**, v. 24, n. 3, p. 311–330, 2019.

RIBEIRO, S. **A importância do modelo gravitacional como instrumento do comércio internacional**. OBSERVARE. Universidade Autónoma de Lisboa, 2023. ISBN 9789899002289. Disponível em: <http://dx.doi.org/10.26619/978-989-9002-28-9.30>. Acesso em: 22 dez. 2024.

RIGBY, R. A.; STASINOPOULOS, D. M. A semi-parametric additive model for variance heterogeneity. **Statistics and Computing**, Springer Science and Business Media LLC, v. 6, n. 1, p. 57–65, 1996. ISSN 1573-1375. Disponível em: <http://dx.doi.org/10.1007/bf00161574>. Acesso em: 14 jan. 2025.

RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. **Journal of the Royal Statistical Society Series C: Applied Statistics**, Oxford University Press (OUP), v. 54, n. 3, p. 507–554, abr. 2005. ISSN 1467-9876. Disponível em: <http://dx.doi.org/10.1111/j.1467-9876.2005.00510.x>. Acesso em: jul. 2024.

RINALDI, M. M.; DIANESE, A.; COSTA, A. M. Avaliação do uso de cera de carnaúba na conservação pós-colheita de frutos de *Passiflora cincinnata* cv. BRS Sertão Forte. **Agrotrópica (Itabuna)**, Agrotopica, v. 33, n. 1, p. 29–38, abr. 2021. ISSN 0103-3816.

SANDAMBI, N. L. A. **Comércio externo entre Portugal-Palop “Aplicação dos modelos gravitacional e VAR em painel”**. Dissertação (Mestrado) – Universidade de Évora, 2021.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, mar. 1978. ISSN 0090-5364.

SEBOK, N. d. F. A. Coronavírus e o comércio exterior: Estudo bibliográfico. **Research, Society and Development**, v. 10, n. 7, p. e59310717008–e59310717008, 2021. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/17008/14825>. Acesso em: 22 dez. 2024.

SILVA, D. C. d.; HERNÁNDEZ, L. G. Aplicação metodológica da análise de conteúdo em pesquisas de análise de política externa. **Revista Brasileira de Ciência Política**, SciELO Brasil, p. e218584, 2020.

SILVA, J. S.; TENREYRO, S. The log of gravity. **The Review of Economics and statistics**, JSTOR, p. 641–658, 2006.

SILVA, T. J. J. d. **Desempenho exportador da manga e da uva brasileira no comércio internacional: uma análise entre o período de 2003 a 2013**. Dissertação (Dissertação de Mestrado - Economia / CAA) – Universidade Federal de Pernambuco (UFPE), Recife, 2015.

Sistema Integrado de Comércio Exterior (Siscomex). **Barreiras Comerciais**. 2025. Disponível em: <https://www.gov.br/siscomex/pt-br/servicos/aprendendo-a-exportar/2-identificacao-do-mercado-alvo/Barreiras%20Comerciais>. Acesso em: set. 2025.

STASINOPOULOS, M.; ENEA, M.; RIGBY, R. A. **Zero adjusted distributions on the positive real line**. 2017. Disponível em: <https://www.gamlss.com/wp-content/uploads/2023/06/ZeroAdjustedDistributions.pdf>. Acesso em: nov. 2024.

THOMAS, G.; NAKAMURA, L. R.; MORAL, R. A.; DEMÉTRIO, C. G. Modeling data with zero inflation and overdispersion using gamlss. **arXiv preprint arXiv:1810.02618**, 2018.

TINBERGEN, J. Shaping the world economy: Suggestions for an international economic policy. **The Twentieth Century Fund**, 1962.

United Nations Statistics Division. **UN Comtrade TradeFlow**. 2025. Disponível em: <https://comtradeplus.un.org/TradeFlow>. Acesso em: 14 out. 2025.

VIANA, A. C.; AIRES, T. L. S.; CLEMINTINO, L. S. Cera de carnaúba incorporada com a argila caulim para conservação de manga (mangifera indica l.). **Revista Semiárido De Visu**, v. 10, n. 2, 2022.

VIEGAS, E.; LEVY, O.; HAVLIN, S.; TAKAYASU, H.; TAKAYASU, M. A complexity perspective on the geographical location of companies: How distance reduce trade between firms. **arXiv preprint arXiv:2311.15760**, 2023.

WOOLDRIDGE, J. M. **Introductory econometrics a modern approach**. [S. l.]: South-Western cengage learning, 2016.

World Bank. **GDP (current US\$)**. 2025. Disponível em: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?most_recent_year_desc=false. Acesso em: 10 set. 2025.

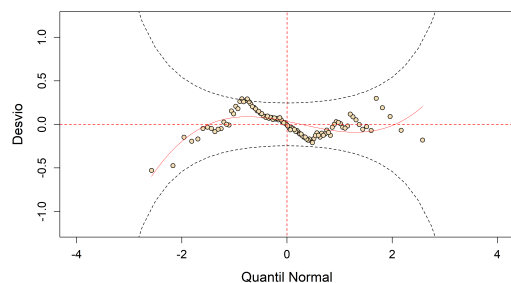
APÊNDICE A – LISTA DE PAÍSES NA AMOSTRA (1997 -2024)

Lista de Países			
ISO	Descrição	ISO	Descrição
ALB	Albânia	JPN	Japão
ARE	Emirados Árabes Unidos	KEN	Quênia
ARG	Argentina	KOR	Coreia do Sul
AUS	Austrália	LVA	Letônia
BEL	Bélgica	MEX	México
BOL	Bolívia	NGA	Nigéria
CAN	Canadá	NLD	Países Baixos (Holanda)
CHL	Chile	PAK	Paquistão
CHN	China	PER	Peru
COL	Colômbia	PHL	Filipinas
DEU	Alemanha	POL	Polônia
ECU	Equador	PRY	Paraguai
EGY	Egito	RUS	Rússia
ESP	Espanha	SGP	Singapura
FRA	França	SLV	El Salvador
GBR	Reino Unido	THA	Tailândia
GTM	Guatemala	TUR	Turquia
GUY	Guiana Francesa	TWN	Taiwan (Formosa)
IDN	Indonésia	URY	Uruguai
IND	Índia	USA	Estados Unidos
ISR	Israel	VEN	Venezuela
ITA	Itália	ZAF	África do Sul

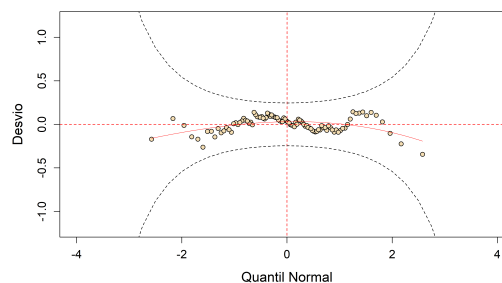
APÊNDICE B – POSSÍVEIS INADEQUAÇÕES EM MODELOS VIA *WORM PLOT*

Figura 17 – Diferentes tipos de falhas no modelo indicados pelo gráfico de minhoca.

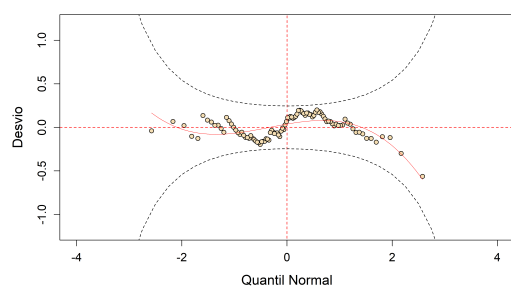
(a) resíduo médio muito alto.



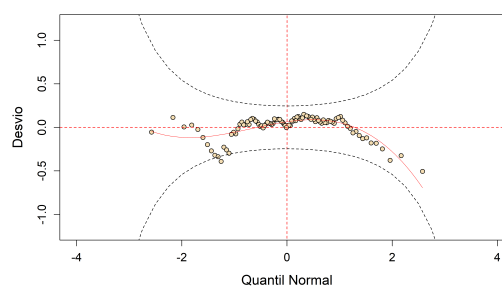
(b) resíduo médio muito baixo.



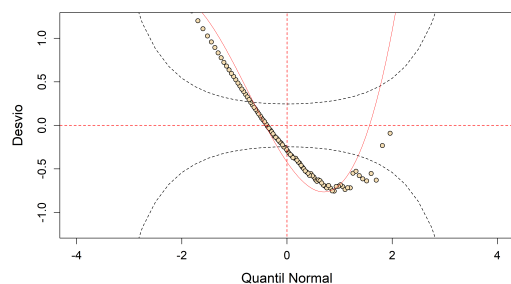
(c) variância residual muito alta.



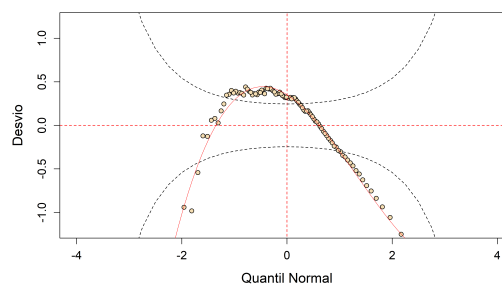
(d) variância residual muito baixa.



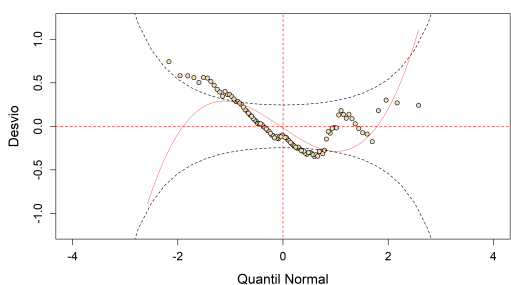
(e) assimetria positiva residual.



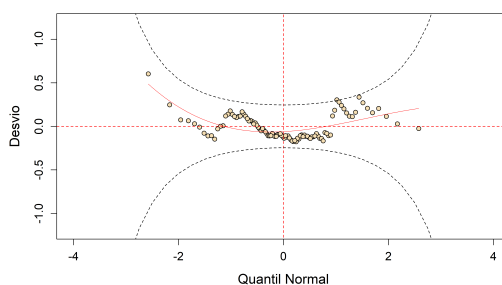
(f) assimetria negativa residual.



(g) resíduo leptocúrtico.



(h) resíduo platicúrtico.



Fonte: Elaboração própria a partir das especificações de Buuren e Fredriks (2001).

Nota: (a) e (b) indicam falha no ajuste correto da localização; (c) e (d) indicam falha no ajuste correto da escala; (e) e (f) indicam falha no ajuste correto da assimetria; e (g) e (h) indicam falha no ajuste correto da curtose.

APÊNDICE C – INTERPRETAÇÃO DAS ELASTICIDADES

A interpretação dos coeficientes estimados no modelo ZAGA deve considerar a forma funcional dos links adotados em cada parâmetro: média condicional das exportações positivas (μ), dispersão relativa (σ) e probabilidade de fluxos nulos (ν) (Rigby e Stasinopoulos, 2005). A seguir, são apresentadas as principais regras de interpretação (Wooldridge, 2016).

Efeitos sobre μ (média, link log)

- **Variáveis em log:** o coeficiente (β) é uma elasticidade:

$$\frac{\partial \log \mu}{\partial \log x} = \beta.$$

Assim, uma variação percentual de g em x gera impacto aproximado de $\beta \cdot g \cdot 100\%$ em μ .

- **Variáveis em nível:** o coeficiente (β) é uma semi-elasticidade:

$$\frac{\partial \log \mu}{\partial x} = \beta,$$

e a variação percentual em μ é dada por $100 \cdot [\exp(\beta \Delta x) - 1]$.

- **Dummies:** o coeficiente em log indica efeito multiplicativo:

$$\frac{\mu_1}{\mu_0} = e^\beta,$$

de modo que o impacto percentual é $100 \cdot (e^\beta - 1)$.

Efeitos sobre σ (dispersão relativa, link log)

- A interpretação segue as mesmas regras de μ : coeficientes em log indicam elasticidades e em nível indicam semi-elasticidades.
- Dummies indicam efeito multiplicativo sobre a dispersão.

Efeitos sobre ν (probabilidade de fluxos nulos, link logit)

- Os coeficientes de ν afetam o logit da probabilidade de zeros, e seus efeitos devem ser interpretados em termos de variações nas *odds*.
- Para variáveis em nível e *dummy*:

$$\frac{\partial \log(odds)}{\partial x} = \beta, \quad \text{em que } odds = \frac{\nu}{1 - \nu}$$

de modo que uma variação Δx altera as *odds* em $\exp(\beta \Delta x)$ e a variação percentual nas *odds* é dada por $100 \cdot [\exp(\beta \Delta x) - 1]$.

– Para variáveis em log:

$$\frac{\partial \log(odds)}{\partial \log x} = \beta,$$

Assim, um aumento percentual de g em x altera as *odds* em $OR = (1 + g)^\beta$, e a variação percentual correspondente é $100 \cdot [(1 + g)^\beta - 1]$.

Elasticidades e Semi-Elasticidades dos Parâmetros do Modelo ZAGA

Ressalta-se que o modelo completo indicado para cada parâmetro refere-se ao conjunto inicial de variáveis explicativas especificado exclusivamente para aquele parâmetro, conforme a estrutura dos modelos GAMLSS. Assim, o modelo completo de μ pode conter variáveis diferentes daquelas presentes nos modelos completos de σ ou ν .

As fórmulas utilizadas foram:

– **Variáveis em log (link log):**

$$\% \Delta = 100 \left[(1 + g)^{\beta} - 1 \right],$$

onde $g = 0,10$ representa um aumento de 10%.

– **Variáveis em nível (link log):**

$$\% \Delta = 100 [e^{\beta \Delta x} - 1],$$

sendo $\Delta x = 1$ ponto (ou 10 pontos quando indicado).

– **Dummies (link log):**

$$\% \Delta = 100 (e^{\beta} - 1).$$

– **Variáveis em log no logit:**

$$OR = (1 + g)^{\alpha}, \quad \% \Delta \text{ odds} = 100 (OR - 1).$$

– **Variáveis em nível ou dummies no logit:**

$$OR = e^{\alpha \Delta x}, \quad \% \Delta \text{ odds} = 100 (OR - 1).$$

Tabela 10 – Efeitos de +10% (variáveis em log) e de +1/+10 pt (nível) sobre μ .

Variável	Fórmula utilizada	Efeito
$\log PIB_{pci}$ ($\beta = 0,332$)	$100[(1 + 0,10)^{0,332} - 1]$	(3,21%)
$\log PIB_{pcj}$ ($\beta = 0,699$)	$100[(1 + 0,10)^{0,699} - 1]$	(6,89%)
$\log Dist_{ij}$ ($\beta = -0,598$)	$100[(1 + 0,10)^{-0,598} - 1]$	(-5,54%)
$\log tC$ ($\beta = -0,333$)	$100[(1 + 0,10)^{-0,333} - 1]$	(-3,13%)
$\log Pop_j$ ($\beta = 1,035$)	$100[(1 + 0,10)^{1,035} - 1]$	(10,37%)
ILC_j ($\beta = 0,025$)	$100[e^{0,025 \cdot 1} - 1]$	(2,50%)
	$100[e^{0,025 \cdot 10} - 1]$	(25,00%)
$landlocked_j$ dummy ($\beta = -0,801$)	$100(e^{-0,801} - 1)$	(-55,11%)
rta dummy ($\beta = -0,652$)	$100(e^{-0,652} - 1)$	(-47,88%)

Tabela 11 – Efeitos de +10% (variáveis em log) e de dummy sobre σ .

Variável	Fórmula utilizada	Efeito
$\log PIB_{pci}$ ($\beta = 0,104$)	$100[(1 + 0,10)^{0,104} - 1]$	(0,99%)
$\log PIB_{pcj}$ ($\beta = -0,175$)	$100[(1 + 0,10)^{-0,175} - 1]$	(-1,66%)
$\log Dist_{ij}$ ($\beta = -0,091$)	$100[(1 + 0,10)^{-0,091} - 1]$	(-0,87%)
$\log tC$ ($\beta = 0,127$)	$100[(1 + 0,10)^{0,127} - 1]$	(1,22%)
$\log Pop_j$ ($\beta = -0,115$)	$100[(1 + 0,10)^{-0,115} - 1]$	(-1,09%)
rta dummy ($\beta = -0,407$)	$100(e^{-0,407} - 1)$	(-33,45%)

Tabela 12 – Mudança nas odds de zeros (ν) para +10% (log) e +1 pt/dummy (nível).

Variável	Fórmula (odds)	OR	% nas odds
$\log PIB_{pcj}$ (-0,322)	$e^{(-0,322 \cdot \log(1+0,10))}$	0,970	(-3,00%)
$\log Dist_{ij}$ (0,955)	$e^{(0,955 \cdot \log(1+0,10))}$	1,095	(9,50%)
$\log tC$ (-0,484)	$e^{(-0,484 \cdot \log(1+0,10))}$	0,955	(-4,50%)
$\log Pop_j$ (-0,923)	$e^{(-0,923 \cdot \log(1+0,10))}$	0,916	(-8,40%)
ILC_j (-0,062)	$e^{-0,062 \cdot 1}$	0,939	(-6,00%)
rta dummy (-1,572)	$e^{-1,572 \cdot 1}$	0,207	(-79,30%)

Nota: Em ν , os coeficientes afetam o *logit*; por isso os resultados são expressos em *odds* (e não diretamente em probabilidade). Converter odds em probabilidade requer o nível base de ν .