



UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
Faculdade de Ciências e Tecnologia  
Câmpus de Presidente Prudente

# Estudo de Séries Temporais dos Preços da Gasolina, Etanol e Açúcar no Estado de São Paulo através da Transfer Entropy

Yuliana Apaza Flores

Orientadora: Profa. Dra. Marilaine Colnago

Programa: Matemática Aplicada e Computacional



**UNIVERSIDADE ESTADUAL PAULISTA**

Faculdade de Ciências e Tecnologia de Presidente Prudente

Programa de Pós-Graduação em Matemática Aplicada e Computacional

# **Estudo de Séries Temporais dos Preços da Gasolina, Etanol e Açúcar no Estado de São Paulo através da Transfer Entropy**

Yuliana Apaza Flores

Orientadora: Profa. Dra. Marilaine Colnago

Dissertação apresentada ao Programa de Pós-Graduação em Matemática Aplicada e Computacional da Faculdade de Ciências e Tecnologia da UNESP para a obtenção do título de Mestre em Matemática Aplicada e Computacional.

Presidente Prudente, abril de 2025

F634e Flores, Yuliana Apaza  
Estudo de séries temporais dos preços da gasolina, etanol e açúcar  
no estado de São Paulo através da Transfer Entropy / Yuliana Apaza  
Flores. -- Presidente Prudente, 2025  
71 p.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP),  
Faculdade de Ciências e Tecnologia, Presidente Prudente  
Orientadora: Marilaine Colnago

1. Séries Temporais. 2. Transfer Entropy. 3. Preços de  
Commodities. 4. Transferência de informação. I. Título.

**CERTIFICADO DE APROVAÇÃO**

**TÍTULO DA DISSERTAÇÃO:** Estudo de séries temporais dos preços da gasolina, etanol e açúcar no estado de São Paulo através da Transfer Entropy

**AUTORA:** YULIANA APAZA FLORES

**ORIENTADORA:** MARILAINE COLNAGO

Aprovada como parte das exigências para obtenção do Título de Mestra em Matemática Aplicada e Computacional, pela Comissão Examinadora:

Profa. Dra. MARILAINE COLNAGO (Participação Virtual)

Departamento de Ciências de Computação e Estatística / Instituto de Biociências, Letras e Ciências Exatas - UNESP

Prof. Dr. MESSIAS MENEGUETTE JUNIOR (Participação Virtual)

Departamento de Matemática e Computação / Faculdade de Ciências e Tecnologia de Presidente Prudente - FCT/Unesp

Profa. Dra. MIRELA VANINA DE MELLO (Participação Virtual)

Universidade Estadual de Santa Cruz

Presidente Prudente, 27 de fevereiro de 2025

*Dedico este trabalho ao meu pai, Alberto, e à  
minha mãe, Magdalena, com muito amor e  
gratidão, pelo apoio incondicional e por sem-  
pre acreditarem em mim.*

---

# Agradecimentos

Agradeço em primeiro lugar a Deus por minha vida, a saúde, por me ouvir, abençoar, guardar e cuidar nesta caminhada e enquanto permaneço longe de minha família de estar sempre cuidando deles, eu devo todas as graças.

Agradeço às pessoas mais importantes da minha vida: meu pai, Alberto Apaza Cartagena, minha mãe, Magdalena Flores Geronimo, e meu irmão, Albert Dayiro Apaza Flores, pelo exemplo de vida e caráter, bem como pelo apoio incondicional que sempre me deram nos meus estudos, na minha formação intelectual e profissional. Desde o dia em que decidi prosseguir meus estudos de pós-graduação, mesmo sabendo que isso significaria estar longe deles, nunca deixaram de me incentivar e apoiar nesse caminho.

Agradeço imensamente à minha orientadora, Professora Doutora Marilaine Colnago, por todo o apoio e orientação ao longo da elaboração desta dissertação. Sua constante disponibilidade, a empatia com que acolheu minhas ideias, bem como seu estímulo e suporte científico, foram fundamentais para a construção deste trabalho.

Agradeço aos professores da banca examinadora pela disponibilidade e colaboração na análise e contribuição para a melhoria desta dissertação.

Gratidão aos professores do PósMAC que fizeram parte da minha formação acadêmica nesta etapa, por me ajudar tirando minhas dúvidas nas disciplinas. Os meus agradecimentos aos amigos, especialmente ao Nelson, Susan e aos colegas do PósMAC, por sua disposição em me ajudar, tirando minhas dúvidas e me dando conselhos sobre outros aspectos da vida cotidiana.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*"Nossos problemas na vida são cálculos de matemática. Basta subtrair, dividir, somar ou multiplicá-los. O resultado final é igual à capacidade de resolvê-los."*

**Angelita Loturco**

# Resumo

---

As *commodities* são produtos de alta comercialização e desempenham um papel crucial na economia mundial, com seus preços determinados no mercado internacional pela dinâmica de oferta e demanda. Neste trabalho, foi analisada a dinâmica das séries temporais dos preços das seguintes *commodities*: gasolina, etanol e açúcar, no estado de São Paulo, no período de 9 de maio de 2004 a 19 de novembro de 2023. Para isso, foi utilizada a medida de Transfer Entropy, um parâmetro assimétrico que avalia a quantidade de informação transferida de uma fonte para um destino, identificando mudanças estruturais associadas a fatores econômicos e de mercado. A análise foi aplicada às séries temporais de preços, retornos e volatilidade dessas *commodities*, considerando os períodos pré-pandemia, durante a pandemia e pós-pandemia de Covid-19. Esses três períodos apresentam dinâmicas claramente distintas. Antes da pandemia, os mercados eram estáveis e previsíveis, com interações moderadas entre as *commodities*. Durante a pandemia, a volatilidade e a interdependência aumentaram significativamente, refletindo as disrupções globais. No período pós-pandemia, os mercados começaram a se estabilizar, mas com um novo equilíbrio e dinâmicas influenciadas por fatores externos, como tensões geopolíticas e políticas energéticas. Os resultados obtidos mostram que a Covid-19 afetou a relação entre os preços das *commodities* açúcar, etanol e gasolina. A Transfer Entropy se mostrou eficiente para quantificar o fluxo de informações entre as séries temporais de preços dessas *commodities* e elucidar a direcionalidade da transferência.

Palavras-Chave: *Séries Temporais, Transfer Entropy, Preços de Commodities, Transferência de informação.*



# Abstract

---

*Commodities* are highly traded products and play a crucial role in the global economy, with their prices determined in the international market by the dynamics of supply and demand. In this study, the dynamics of the time series of the prices of the following *commodities* were analyzed: gasoline, ethanol, and sugar, in the state of São Paulo, from 9 May 2004 to 19 November 2023. For this purpose, the Transfer Entropy measure was used, an asymmetric parameter that evaluates the amount of information transferred from a source to a destination, identifying structural changes associated with economic and market factors. The analysis was applied to the time series of prices, returns, and volatility of these *commodities*, considering the pre-pandemic, pandemic, and post-pandemic periods of Covid-19. These three periods exhibit clearly distinct dynamics. Before the pandemic, the markets were stable and predictable, with moderate interactions among the *commodities*. During the pandemic, volatility and interdependence increased significantly, reflecting global disruptions. In the post-pandemic period, the markets began to stabilize, but with a new equilibrium and dynamics influenced by external factors, such as geopolitical tensions and energy policies. The results obtained show that Covid-19 affected the relationship between the prices of the *commodities* sugar, ethanol, and gasoline. Transfer Entropy proved to be an efficient method for quantifying the information flow between the time series of these *commodities* prices and elucidating the directionality of the transfer.

Keywords: *Time Series, Transfer Entropy, Commodity Prices, Information Transfer.*



---

## Lista de Figuras

|                                                                                                                    |    |
|--------------------------------------------------------------------------------------------------------------------|----|
| 3.1.1 Decomposição de uma escolha entre três possibilidades. . . . .                                               | 29 |
| 3.2.1 A relação entre a entropia de variáveis aleatórias, entropias de margem e entropia de cópula. . . . .        | 35 |
| 3.3.1 Uma representação do $k$ -vizinhos mais próximo e da $\epsilon$ -bola. . . . .                               | 37 |
| 5.0.1 Séries de preço, retorno e volatilidade do açúcar. . . . .                                                   | 46 |
| 5.0.2 Séries de preço, retorno e volatilidade da gasolina. . . . .                                                 | 47 |
| 5.0.3 Séries de preço, retorno e volatilidade do etanol. . . . .                                                   | 48 |
| 5.0.4 Transfer Entropy entre as séries da volatilidade de preços entre as <i>commodities</i> estudadas. . . . .    | 49 |
| 5.0.5 Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol. . . . .                       | 50 |
| 5.0.6 Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina. . . . .                     | 50 |
| 5.0.7 Transfer Entropy entre as séries da volatilidade de preços de etanol e gasolina. . . . .                     | 51 |
| 5.1.1 Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (pré-pandemia). . . . .        | 53 |
| 5.1.2 Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (pré-pandemia). . . . .      | 54 |
| 5.1.3 Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (pré-pandemia). . . . .      | 55 |
| 5.1.4 Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (período pandêmico). . . . .   | 57 |
| 5.1.5 Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (período pandêmico). . . . . | 58 |
| 5.1.6 Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (período pandêmico). . . . . | 59 |
| 5.1.7 Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (pós-pandemia). . . . .        | 61 |
| 5.1.8 Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (pós-pandemia). . . . .      | 62 |
| 5.1.9 Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (pós-pandemia). . . . .      | 63 |



---

## Lista de Tabelas

|                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------|----|
| 4.1.1 Moeda e unidade de medida das <i>commodities</i> . . . . .                                                       | 41 |
| 5.0.1 Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para<br>o período total estudado. . . . . | 49 |
| 5.1.1 Períodos estudados e tamanho das séries. . . . .                                                                 | 51 |
| 5.1.2 Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para<br>o período pré-pandemia. . . . .   | 52 |
| 5.1.3 Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para<br>o período pandêmico. . . . .      | 56 |
| 5.1.4 Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para<br>o período pós-pandemia. . . . .   | 60 |
| 5.2.1 Quadro comparativo dos períodos das <i>commodities</i> . . . . .                                                 | 64 |



---

# Índice de Notações

- $\mathbb{R}$  - denota respectivamente o conjunto dos números reais.
- Conab - Companhia Nacional de Abastecimento.
- PROÁLCOOL - Programa Nacional do Alcool.
- TE - Transfer Entropy.
- IM - Informação Mútua.
- GC - Causalidade de Granger.
- VAR - Vetor Autoregressivo.
- FGV - Fundação Getúlio Vargas.
- INPM - Instituto Nacional de Pesos e Medidas.
- VHP - Polarização Muito Alta.
- ARIMA - Média Móvel Integrada Autoregressiva..
- ARCH - Heterocedasticidade Condicional Autorregressiva.
- RNAs - Redes Neurais Artificiais.
- MAPE - Erro Percentual Absoluto Médio.
- V.A - Variável Aleatória.
- CDF - Função de Distribuição Cumulativa.
- CE - Entropia de Copula.
- ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis.
- CEPEA - Centro de Estudos Avançados em Economia Aplicada.
- USP - Universidade de São Paulo.
- CSV - Valores separados por vírgula.
- KDD - Descoberta de Conhecimento em Bases de Dados.
- KSG - Kraskov-Stögbauer-Grassberger.
- kNN -  $k$ -vizinhos mais próximos.
- Dom - Domínio de uma função.
- Ran - Rango de uma função.



# Sumário

---

|                                                                                                                        |           |
|------------------------------------------------------------------------------------------------------------------------|-----------|
| Resumo                                                                                                                 | 5         |
| Abstract                                                                                                               | 7         |
| Lista de Figuras                                                                                                       | 9         |
| Lista de Tabelas                                                                                                       | 11        |
| Índice de Notações                                                                                                     | 13        |
| Sumário                                                                                                                | 15        |
| <b>1 Introdução</b>                                                                                                    | <b>17</b> |
| 1.1 Objetivos . . . . .                                                                                                | 19        |
| 1.2 Estruturação dos capítulos . . . . .                                                                               | 19        |
| <b>2 Revisão de literatura</b>                                                                                         | <b>21</b> |
| 2.1 <i>Commodities</i> . . . . .                                                                                       | 21        |
| 2.1.1 Açúcar . . . . .                                                                                                 | 22        |
| 2.1.2 Etanol . . . . .                                                                                                 | 22        |
| 2.1.3 Gasolina . . . . .                                                                                               | 23        |
| 2.2 Estudos e Aplicações de Modelos de Séries Temporais em Mercados Finan-<br>ceiros e de <i>Commodities</i> . . . . . | 24        |
| <b>3 Conceitos Preliminares</b>                                                                                        | <b>27</b> |
| 3.1 Entropia e Medidas de Dependência . . . . .                                                                        | 27        |
| 3.1.1 Introdução ao Conceito de Transfer Entropy . . . . .                                                             | 27        |
| 3.1.2 Entropia de Shannon . . . . .                                                                                    | 28        |
| 3.1.3 Séries Temporais Simbólicas . . . . .                                                                            | 30        |
| 3.2 Cópulas: Modelagem da Dependência Multivariada . . . . .                                                           | 30        |
| 3.2.1 Teorema de Sklar . . . . .                                                                                       | 33        |
| 3.2.2 Construindo a Densidade de Cópula Empírica . . . . .                                                             | 33        |
| 3.2.3 Informação Mútua e Entropia de Cópula . . . . .                                                                  | 34        |
| 3.2.4 Estimando informações mútuas via Copula . . . . .                                                                | 35        |
| 3.2.5 Transfer Entropy via Entropia de Copula . . . . .                                                                | 36        |
| 3.3 Estimativas de Entropia . . . . .                                                                                  | 36        |
| 3.3.1 Estimando entropia com o método KNN . . . . .                                                                    | 36        |
| 3.4 Testes Estatísticos Baseados em Entropia . . . . .                                                                 | 38        |
| 3.4.1 Teste de independência condicional . . . . .                                                                     | 38        |
| 3.4.2 Teste de Normalidade Multivariada . . . . .                                                                      | 39        |
| 3.4.3 Teste de duas amostras com CE . . . . .                                                                          | 39        |

|          |                                                                                                           |           |
|----------|-----------------------------------------------------------------------------------------------------------|-----------|
| <b>4</b> | <b>Metodologia</b>                                                                                        | <b>41</b> |
| 4.1      | Base de Dados . . . . .                                                                                   | 41        |
| 4.2      | Etapas da Pesquisa . . . . .                                                                              | 41        |
| 4.3      | Ferramentas Utilizadas . . . . .                                                                          | 42        |
| <b>5</b> | <b>Resultados e discussões</b>                                                                            | <b>45</b> |
| 5.1      | Transfer Entropy nos períodos de pré-pandemia, durante a pandemia e<br>pós-pandemia de Covid-19 . . . . . | 51        |
| 5.1.1    | Pré Pandemia (2018 - 2019) . . . . .                                                                      | 52        |
| 5.1.2    | Durante a Pandemia (2020 - 2021) . . . . .                                                                | 55        |
| 5.1.3    | Pós Pandemia (2022 - 2023) . . . . .                                                                      | 59        |
| 5.2      | Discussão sobre o Comparativo dos Períodos . . . . .                                                      | 63        |
| <b>6</b> | <b>Conclusões</b>                                                                                         | <b>65</b> |
|          | <b>Referências Bibliográficas</b>                                                                         | <b>67</b> |

# Introdução

O cultivo da cana-de-açúcar no Brasil remonta ao ano de 1532, quando Martim Afonso de Souza introduziu a planta na Capitania Hereditária de São Vicente, marcando o início de uma trajetória histórica profundamente ligada ao desenvolvimento econômico e social do país [1]. A cana-de-açúcar, sendo uma planta perene, apresenta a capacidade de produzir durante um período de quatro a seis anos, desde que foi manejada adequadamente. Com seu cultivo disseminado por quase todo o território brasileiro, as regiões Centro-Sul e Norte-Nordeste destacam-se como as principais áreas de produção, com o estado de São Paulo liderando tanto em área plantada quanto na produção de seus derivados.

Historicamente, a cana-de-açúcar foi o principal pilar da economia colonial brasileira, sendo o açúcar bruto o único produto derivado exportado durante o período colonial, com destino principal aos mercados europeus, como a Holanda [72]. Com o passar dos séculos, o papel da cana-de-açúcar evoluiu significativamente. Atualmente, além de ser essencial para a produção de açúcar, ela também sustenta a pujante indústria de biocombustíveis, consolidando o Brasil como o maior produtor global de açúcar e o segundo maior exportador de etanol.

Os dados recentes da Companhia Nacional de Abastecimento (Conab) mostram que cerca de 93% da produção de cana-de-açúcar está concentrada na região Centro-Sul, enquanto apenas 7% ocorre no Norte-Nordeste, evidenciando a predominância da produção em áreas de clima mais favorável e infraestrutura mais desenvolvida. O estado de São Paulo, por exemplo, destaca-se não apenas pela extensão de suas áreas cultivadas, mas também pelo alto grau de mecanização e eficiência em suas lavouras.

Além de seu impacto econômico direto, a cultura da cana-de-açúcar tem desempenhado um papel central no comércio internacional brasileiro desde o século XVI. Inicialmente associada à exportação de açúcar bruto, ela evoluiu para atender a uma cadeia de produção diversificada que inclui açúcar e etanol. O aumento da produção de etanol iniciou-se na década de 1970, em resposta à crise do petróleo, quando o Brasil tomou a iniciativa de produzir etanol anidro e hidratado como combustíveis para veículos, visando reduzir a dependência do país do petróleo importado [44]. Em 1975, o governo brasileiro lançou o Programa Nacional do Alcool (PROÁLCOOL), com o objetivo de aumentar a produção de etanol e incentivar a fabricação de veículos *flex-fuel*, que podem ser movidos tanto a etanol hidratado quanto a gasolina, ou ainda, a uma mistura dos dois combustíveis.

O setor sucroenergético, que integra a produção de açúcar e etanol a partir da cana-de-açúcar, desempenha um papel crucial tanto na economia paulista quanto na nacional. O mercado de combustíveis, composto principalmente pela gasolina e pelo etanol, está intimamente ligado ao setor agrícola e é influenciado por uma variedade de fatores, como políticas de preços, oferta e demanda, condições climáticas e eventos internacionais. Dessa

forma, o comportamento dos preços de *commodities*, como gasolina, etanol e açúcar, é de interesse contínuo, especialmente em economias fortemente dependentes de produtos agrícolas e energéticos, como São Paulo [63].

A relação entre os preços do etanol, açúcar e gasolina é bastante complexa e envolve diversos fatores econômicos e produtivos. A usina, ao lidar com a cana-de-açúcar, tem a flexibilidade de direcionar sua produção para o etanol ou para o açúcar, dependendo das condições do mercado. Quando o preço do açúcar está em alta no mercado internacional, as usinas tendem a priorizar a produção desse subproduto para maximizar seus lucros [61]. Por outro lado, quando os preços do etanol são mais atraentes, há uma mudança de foco para a produção do biocombustível. Essa interdependência torna o mercado do etanol e do açúcar altamente sensível às variações de preço e demanda de ambos os produtos.

Além disso, o etanol está intrinsecamente ligado ao mercado da gasolina, já que é utilizado na mistura do combustível vendido nos postos de abastecimento, especialmente no Brasil, onde o etanol anidro compõe parte da gasolina [56]. Assim, variações no preço da gasolina também influenciam diretamente o preço do etanol, criando uma correlação entre as três *commodities*.

A grande correlação existente entre os preços do etanol, do açúcar e da gasolina ressalta a importância de monitorar as transferências de informação entre esses mercados, uma vez que flutuações em um podem rapidamente impactar os outros, gerando um efeito em cadeia na precificação e na alocação de recursos das usinas. Ao estudar a transferência de informação, avalia-se a quantidade de conhecimento que uma variável aleatória contém sobre outra, ou seja, a redução de incerteza em uma variável a partir da informação fornecida por outra [55].

Na análise de séries temporais multivariadas, é comum estudar o acoplamento entre as variáveis, então uma medida eficaz para esse tipo de análise é a Transfer Entropy [70], que quantifica o fluxo de informações de uma variável para outra. Esse método é não-paramétrico e capaz de capturar interações não lineares entre as variáveis. Além disso, Transfer Entropy (TE) é uma medida assimétrica, que diferente da Informação Mútua (IM) e Causalidade de Granger (GC), que apenas mede a quantidade de informação compartilhada entre duas variáveis, mensura a direção e o fluxo de informações entre elas, oferecendo uma análise mais detalhada das inter-relações [14, 52].

Mais ainda, TE tem capacidade de capturar relações não lineares, isso a torna mais eficaz em cenários onde a GC falha. Em relação à Informação Mútua (IM), que também é uma ferramenta importante na análise de séries temporais, ela mede a quantidade total de informação compartilhada entre duas variáveis, mas não identifica a direção do fluxo de informação, tornando-a inadequada para estudos de causalidade.

Sendo assim, a presente dissertação objetiva a análise dos padrões de Transfer Entropy dos preços de *commodities* importantes para o mercado nacional. Este estudo tem grande importância para a formulação de políticas no setor e para o desempenho econômico tanto do Estado quanto do País. Essas análises fornecem subsídios valiosos para o desenvolvimento de estratégias eficazes voltadas ao gerenciamento dos riscos associados às flutuações nos preços das *commodities*. Dada a importância do agronegócio para a sociedade, o uso de modelos matemáticos e econométrico-estatísticos para análise e previsão de preços no setor agropecuário torna-se essencial devido aos riscos financeiros associados às decisões estratégicas e às operações com derivativos realizadas pelas organizações produtivas [71].

Por fim, a relevância dessa pesquisa se intensifica diante da escassez de estudos na literatura nacional que explorem essas abordagens no contexto das *commodities* negociadas no mercado, ressaltando a importância de investigações direcionadas a essa temática.

## 1.1 Objetivos

O objetivo geral desse trabalho é analisar a dinâmica das séries temporais dos preços da gasolina, etanol e açúcar no estado de São Paulo, desde de 9 de maio de 2004 a 19 de novembro de 2023, utilizando o método de Transfer Entropy, a fim de investigar a transferência de informação entre essas *commodities* e identificar mudanças estruturais associadas a fatores econômicos e de mercado.

Como objetivos específicos, tem-se:

- Investigar a direção e intensidade do acoplamento entre os preços da gasolina, etanol e açúcar ao longo do tempo.
- Identificar possíveis mudanças estruturais nas séries temporais dos preços dessas *commodities* que possam estar relacionadas a eventos econômicos e de mercado.
- Analisar a influência do mercado internacional nos preços locais de etanol, açúcar e gasolina no estado de São Paulo.
- Aplicar o método de Transfer Entropy via cópula para detectar mudanças na intensidade do acoplamento entre as séries temporais de preços.
- Examinar a eficácia do método Transfer Entropy na identificação de mudanças no comportamento dos preços dessas *commodities* ao longo do período analisado.

## 1.2 Estruturação dos capítulos

O presente trabalho está estruturado da seguinte maneira: no Capítulo 2, é realizada uma revisão da literatura, abordando de forma detalhada as principais características das *commodities* analisadas neste estudo, como gasolina, etanol e açúcar, além de explorar os métodos mais relevantes utilizados para a análise da dinâmica de séries temporais. Essa revisão estabelece o embasamento teórico necessário para a compreensão da relação entre essas *commodities* no mercado.

O Capítulo 3, é dedicado à apresentação das definições e conceitos fundamentais relacionados ao método de Transfer Entropy, que constitui a principal ferramenta metodológica deste trabalho. São abordadas as propriedades teóricas desse método e sua aplicabilidade na detecção de transferências de informação entre séries temporais de diferentes naturezas. No Capítulo 4, apresenta a parte metodológica, onde mostra os períodos estudados e os dados das *commodities*, também as etapas da pesquisa juntamente com as ferramentas utilizadas.

No Capítulo 5, são expostos os resultados obtidos a partir da aplicação dos métodos descritos. Essa seção busca evidenciar as tendências identificadas e oferece uma análise inicial sobre a intensidade e a direção do acoplamento entre os preços das *commodities* ao longo do período estudado.

Finalmente, no Capítulo 6, são discutidos os principais desafios enfrentados durante o desenvolvimento da pesquisa.



## Revisão de literatura

### 2.1 *Commodities*

As *commodities* são matérias-primas ou produtos agrícolas primários que servem como base para processos industriais e de manufatura. Esses bens, geralmente padronizados em qualidade e negociados em mercados globais, desempenham um papel central no funcionamento da economia mundial [16]. As *commodities* representam um alicerce do desenvolvimento econômico, contribuindo significativamente para a economia de diversos países, incluindo o Brasil. O país se destaca como um importante exportador e produtor de *commodities* agrícolas e minerais, graças à sua abundância de recursos naturais e condições climáticas favoráveis, consolidando sua posição de destaque no mercado global.

Historicamente, o Brasil tem sido um dos maiores exportadores mundiais de *commodities*, com uma economia fortemente baseada na produção de matérias primas e bens agropecuários. Segundo dados da Fundação Getúlio Vargas (FGV), cerca de 65% das exportações brasileiras são compostas por *commodities*, demonstrando a relevância desse setor para o crescimento e desenvolvimento econômico do país [72].

O termo *commodity* vem do inglês, e significa mercadoria. Ele é utilizado para designar transações comerciais nas bolsas de valores de produtos de origem primária ou com pouca industrialização, produzidos em larga escala por diferentes produtores e que possuem alta demanda no mercado global. Esses bens, como, petróleo e açúcar, são produzidos em grandes quantidades com pouca variação de qualidade, independentemente de sua origem. Ou seja, são produtos com durabilidade suficiente para serem armazenados por um certo período de tempo sem perder sua qualidade.

De acordo com Teweles e Bradley [68], os principais tipos de *commodities* e alguns exemplos delas são:

- **Agrícolas:** café, trigo, soja, milho, açúcar, farelo de soja.
- **Minerais:** ouro, petróleo, ferro, alumínio.
- **Financeiras:** dólar, euro, real, índices futuros.
- **Ambientais:** créditos de carbono, água, energia.

A seguir serão apresentadas as *commodities* que farão parte do tema central do presente trabalho.

### 2.1.1 Açúcar

A palavra “açúcar” é derivada de *shakkar*, palavra em sânscrito, que significava “grão” ou “areia” e era usada para descrever os cristais de açúcar extraídos da cana. A cultura da cana-de-açúcar teve origem na Polinésia, sendo disseminada pelos árabes ao redor do mundo. Esta cultura se adaptou muito bem nas Américas, devido ao clima favorável para seu desenvolvimento. A cana-de-açúcar é uma planta perene, podendo produzir por um período médio de quatro a seis anos, e seu manejo é relativamente fácil. Oficialmente, foi Martim Affonso de Souza quem trouxe a primeira muda de cana-de-açúcar ao Brasil, em 1532, e iniciou seu cultivo na Capitania de São Vicente, em São Paulo. Nessa mesma Capitania foi construído o primeiro engenho de açúcar do Brasil [1].

Atualmente, o cultivo da cana-de-açúcar no país se concentra na região Centro-Sul, que é a maior produtora e possui a maior área plantada dessa cultura. Consequentemente, o estado de São Paulo é o maior produtor brasileiro de açúcar e etanol etílico [60].

O açúcar é um produto que faz parte da cesta básica de alimentos dos brasileiros e também se constitui em uma *commodity* comercializada tanto no mercado físico, quanto no mercado futuro<sup>1</sup>, em todo o mundo. Seu processo de fabricação consiste em extraí-lo através do caldo proveniente do esmagamento da cana-de-açúcar, seguido de branqueamento, decantação, evaporação, flotação e cristalização. Existem diversos tipos de açúcar para diferentes destinos, seja para o consumidor final (refinarias) ou para os diversos segmentos do mercado industrial de alimentos e bebidas, cada qual com seu valor no mercado, além das *tradings*, principal canal de exportação.

O preço doméstico do açúcar não depende apenas de variáveis regionais ou nacionais, mas é influenciado também por variações do preço do açúcar no mercado internacional, pela taxa de câmbio interna, entre outras variáveis [30]. Quando o real se valoriza em relação ao dólar, o açúcar brasileiro se torna mais caro no mercado internacional, o que, de acordo com a lei da oferta e demanda, reduz o volume de exportações. Com essa diminuição nas exportações, há um aumento na oferta de açúcar no mercado interno, forçando os produtores a reduzir os preços para escoar a produção localmente ou, em alguns casos, redirecionar parte dessa produção para o etanol, caso este mercado se mostre mais atrativo. No entanto, o preço do açúcar no mercado internacional não depende apenas da taxa de câmbio, mas também de fatores como a liquidez global, o índice de preços de *commodities* CRB-Reuters<sup>2</sup> e o preço do petróleo [30].

### 2.1.2 Etanol

O álcool etílico, também chamado de etanol, é encontrado em bebidas como cerveja, vinho, aguardente, na indústria de perfumaria e em produtos de limpeza. Na produção do etanol é possível obter dois compostos, o etanol anidro e o etanol hidratado [18]. O etanol anidro, que é misturado à gasolina é caracterizado pelo teor alcoólico mínimo de 99,3%, determinado pelo Instituto Nacional de Pesos e Medidas (INPM). O etanol hidratado, utilizado como combustível, é uma mistura de etanol e água com teor alcoólico mínimo de 92,6% (INPM).

Na década de 1970, em meio à crise do petróleo, o Brasil adotou uma estratégia voltada para a redução de sua dependência do petróleo importado. Essa medida visava não apenas

---

<sup>1</sup>O mercado físico refere-se ao mercado onde as transações são realizadas com o propósito de entrega imediata (ou em curto prazo) do produto físico negociado. Já no mercado futuro, as negociações envolvem contratos futuros, ou seja, acordos para comprar ou vender um ativo ou *commodity* em uma data específica no futuro, por um preço previamente definido.

<sup>2</sup>*Commodity Research Bureau* - amplamente reconhecido como um dos principais indicadores de preços globais de *commodities*.

a diversificação da matriz energética nacional, mas também a segurança energética e o estímulo ao desenvolvimento de uma indústria renovável e sustentável [44]. Em 1975, o governo federal lançou o Programa Nacional do Álcool (PROÁLCOOL), um marco para o setor energético brasileiro. O PROÁLCOOL teve como objetivo central ampliar a produção de etanol hidratado a partir da cana-de-açúcar, promovendo a substituição gradual da gasolina por esse combustível renovável, além de incentivar a criação e a produção de veículos *flex fuel*, que são projetados para funcionar com etanol hidratado, gasolina ou uma mistura de ambos, oferecendo flexibilidade ao consumidor e contribuindo para a redução da dependência do petróleo.

Como o valor do petróleo oscila muito, dependendo da conjuntura mundial de produção e consumo, logo após a criação do PROÁLCOOL o preço do barril de petróleo diminuiu muito, de modo que a importância do programa reduziu significativamente, voltando a ter uma razoável importância apenas em 1979.

Apesar dos altos e baixos do programa PROÁLCOOL, o cultivo da cana-de-açúcar vem contribuindo para a melhoria do cenário social e industrial brasileiro, propiciando o desenvolvimento do agronegócio do país e creditando ao etanol um importante espaço no setor automobilístico nacional pelo fato de ser um combustível ambientalmente correto e plenamente sustentável [25].

Em relação ao mercado de etanol, as variações no valor do biocombustível ocorrem devido à concorrência com o açúcar, especialmente em momentos de alta de sua cotação no mercado internacional, ao patamar dos preços e da demanda de gasolina automotiva, seu principal concorrente, e às condições de financiamento para a expansão da área plantada de cana-de-açúcar [30].

### 2.1.3 Gasolina

O petróleo, atualmente, é um dos recursos mais relevantes da nossa sociedade. Mesmo com todos os adventos tecnológicos e pesquisas em alternativas para sua substituição, o petróleo é ainda a principal fonte de energia mundial, e seus derivados são utilizados principalmente no abastecimento dos diferentes meios de transporte e na geração de energia elétrica, sendo de vital importância para a economia mundial [31].

A locomoção é uma das necessidades fundamentais do homem moderno, e o petróleo, como fonte energética, desempenha um papel essencial para atender a essa demanda. Isso se deve principalmente à gasolina, um dos principais derivados do refino do petróleo, que continua sendo o combustível mais utilizado no Brasil. Atualmente, as pessoas percorrem distâncias que, no passado, levavam dias para serem cobertas a cavalo, e agora são feitas em minutos ou horas, graças à mobilidade proporcionada pelos veículos movidos a gasolina.

Dessa forma, a gasolina, que é um produto derivado do petróleo, é tão importante que os impactos na sua demanda e preço causam reflexos em toda sociedade. Por exemplo, aumento do seu preço reflete nos índices de inflação, na taxa básica de juros e no produto interno bruto, já que é um insumo para o funcionamento de quase todas as atividades econômicas [28].

Além dos fatores já mencionados, o preço da gasolina também pode ser impactada pelas oscilações nos preços do etanol. Primeiro, em virtude da adição obrigatória de etanol anidro à gasolina, em proporção legalmente estabelecida e, segundo, pois existe a possibilidade de migração imediata da demanda entre gasolina e etanol hidratado, devido ao uso de motores *flex-fuel* nos veículos desde 2003, permitindo uma rápida troca de combustível conforme a variação de preços [30].

## 2.2 Estudos e Aplicações de Modelos de Séries Temporais em Mercados Financeiros e de *Commodities*

Esta seção oferece uma revisão detalhada dos principais estudos que exploram a dinâmica de séries temporais aplicadas a mercados financeiros e de *commodities*. A análise dessas séries tem se tornado essencial para entender a volatilidade, prever tendências e tomar decisões estratégicas, tanto no setor público quanto no privado. Diversos modelos e métodos, como ARCH, ARIMA, SARIMA e Redes Neurais, têm sido amplamente utilizados para captar a complexidade desses mercados. A seguir, serão discutidos trabalhos relevantes que ilustram a aplicação desses modelos, com foco em mercados relacionados ao açúcar, etanol e gasolina no Brasil, assim como serão apresentados estudos utilizando Transfer Entropy (TE), que é a metodologia principal utilizada neste trabalho.

Em [59], os autores conduziram estudos sobre a análise de séries temporais econômicas com volatilidade variável ao longo do tempo, introduzindo o conceito de ARCH (sigla, em inglês, para Heterocedasticidade Condicional Autorregressiva). Esse modelo é fundamental para descrever a volatilidade, que é a variação no preço dos instrumentos financeiros em função do risco. A volatilidade geralmente apresenta padrões caracterizados por períodos de alta turbulência seguidos por momentos de maior estabilidade. Assim, se o objetivo for modelar exclusivamente a volatilidade de uma única série temporal, o modelo ARCH é uma escolha adequada. No entanto, quando o interesse está em analisar interações e fluxos de informação entre diferentes séries temporais, a Transfer Entropy se mostra um método mais poderoso. Isso ocorre porque a TE não impõe restrições paramétricas rígidas e permite capturar relações causais dinâmicas, sendo mais adequada para identificar dependências não lineares entre variáveis [19].

Já no estudo apresentado em [17], que tem como tema central a previsão de produção de cana-de-açúcar e seus derivados na macrorregião Centro-Sul do Brasil, foi aplicada a metodologia ARIMA (sigla, em inglês, para Média Móvel Integrada Autorregressiva) para prever a produção de açúcar, etanol anidro e etanol hidratado no período de 2020 a 2025. O uso de séries temporais entre 1980 e 2020 permitiu a construção de modelos preditivos aplicados à produção anual dessas *commodities*. Apesar dos resultados fornecerem informações valiosas para o planejamento governamental e agrícola no Brasil, eles apresentaram uma margem de erro relativamente alta, com um MAPE (sigla, em inglês, para Erro Percentual Absoluto Médio) de 13,6%. Isso se deu pelo fato de que a metodologia ARIMA tem maior eficiência em previsões de curto prazo [9].

Ainda sobre o método ARIMA, em [50], o trabalho apresentou um estudo sobre a transmissão espacial de preços entre os mercados de cana-de-açúcar de São Paulo e Paraná, porém utilizando o método Box-Jenkins para modelos ARIMA. Seus resultados indicaram uma forte relação de cointegração entre os dois mercados, com choques de preços em São Paulo sendo transmitidos para o Paraná de forma significativa, tanto no curto quanto no longo prazo.

No trabalho apresentado em [11], foram estudados os modelos ARMA, ARIMA e SARIMA para a previsão de taxas de juros, com ênfase nas funções de autocorrelação e autocorrelação parcial. Além disso, modelos de Redes Neurais foram combinados com o SARIMA gerando previsões mais precisas. O modelo combinado foi capaz de superar tanto o SARIMA isolado quanto as Redes Neurais, mostrando que a junção dessas abordagens pode oferecer resultados mais robustos em problemas complexos de previsão financeira. As Redes Neurais Artificiais (RNAs) também foram estudadas em [7] para prever os preços do açúcar no Brasil, analisando a influência de fatores como o preço do petróleo e a taxa de câmbio. A capacidade das RNAs de estabelecer relações complexas entre diversas

variáveis demonstrou ser uma ferramenta eficaz para auxiliar na tomada de decisões no setor de *commodities*.

Já em [48], o autor investigou a formação de preços no mercado brasileiro de *commodities* utilizando a Teoria da Microestrutura e a co-integração de Johansen. Seus resultados indicam que, embora o Brasil influencie as cotações internacionais, o país ainda é tomador de preços no mercado global. O estudo destacou a vulnerabilidade do Brasil devido à dependência de um número limitado de produtos de exportação, criando incertezas tanto para o setor público quanto para o privado, devido à alta volatilidade dos preços.

Em [44], o artigo abordou a importância histórica da cana-de-açúcar na economia brasileira e o impacto do Programa Nacional do Alcool (PROÁLCOOL). Criado em resposta à crise do petróleo na década de 1970, o programa ajudou a consolidar a produção de etanol como uma alternativa energética no Brasil, com altos e baixos ao longo das décadas.

Por fim, em [10], foi analisada a relação entre os preços da gasolina e o setor sucroalcooleiro no Brasil, utilizando um modelo VAR (Autorregressivo Vetorial). O estudo destacou como as políticas de controle de preços da gasolina impactam diretamente os preços do etanol, reforçando a necessidade de o governo considerar os efeitos dessas políticas no setor sucroalcooleiro.

Já no que diz respeito a aplicação da Transfer Entropy em séries temporais, diversos estudos têm explorado essa metodologia, fornecendo uma visão mais clara sobre a transferência de informações e sua direcionalidade entre diferentes eventos. Em [51], por exemplo, TE foi utilizada para analisar as séries temporais de preços de *commodities* agrícolas como açúcar, boi, café e soja entre 1997 e 2019, mostrando o impacto direto dessas *commodities* na economia mundial. Outro trabalho, [40], propôs a construção de redes de entropia de transferência para medir a influência mútua entre eventos representados por séries temporais, permitindo identificar os nós mais influentes usando métricas de centralidade de rede.

No estudo [21], foi proposto um estimador não paramétrico baseado em  $k$ -vizinhos mais próximos para melhorar a precisão da TE em cenários com alta dimensionalidade ou correlações fortes entre variáveis. Já [47] focou em medir a complexidade de sinais biológicos por meio da entropia multiescala, destacando a importância de capturar múltiplas escalas temporais em dados reais, como em variações no batimento cardíaco humano.

Em [22], a TE foi aplicada a processos autorregressivos acoplados, investigando como diferentes parâmetros de ruído e acoplamento influenciam a transferência de informações entre processos. O estudo demonstrou que a entropia de transferência aumenta com a força do acoplamento entre processos e varia conforme as variâncias de ruído.

Por outro lado, [45] utilizou TE para estudar a transferência de informações entre séries de preço de *commodities* como etanol, petróleo e açúcar durante a crise financeira mundial, revelando a interdependência entre essas séries e a direcionalidade do fluxo de informações, especialmente em momentos de instabilidade econômica.

A Transfer Entropy (TE) tem sido amplamente comparada a outros métodos de análise de relações causais em séries temporais, como a Informação Mútua (IM) e a Causalidade de Granger (GC), que também são utilizados para esse fim. Em [70], os autores introduziram a TE como uma medida não paramétrica capaz de detectar a transferência de informação em sistemas dinâmicos não lineares. Diferentemente da IM, que é simétrica e não distingue entre causa e efeito, a TE é assimétrica, permitindo identificar a direção do fluxo de informação entre variáveis.

Em [57], os autores destacam a TE como uma abordagem superior para a análise da direcionalidade em interações entre variáveis. Já o estudo de [20] resalta que a IM não diferencia entre quem influencia e quem é influenciado, além de não considerar expli-

tamente a dependência temporal entre as variáveis. Essa limitação torna a IM menos adequada para a análise de séries temporais, onde a direção e os atrasos são essenciais. De forma semelhante, [32] reforça que a IM apresenta restrições na análise de sistemas em que o fator tempo desempenha um papel crítico.

Com relação à Causalidade de Granger, os trabalhos de [35] e [66] apontam que esse método pode falhar na detecção de causalidade em sistemas dinâmicos não lineares. Além disso, erros na escolha do modelo VAR podem comprometer significativamente os resultados da GC.

Esses estudos ressaltam a versatilidade e a eficácia da Transfer Entropy na análise de interações complexas entre variáveis temporais, destacando-a como uma ferramenta valiosa para a investigação de relações causais em diferentes contextos.

## Conceitos Preliminares

Neste capítulo, será apresentada uma série de definições e resultados essenciais para facilitar a leitura e promover uma melhor compreensão das ideias desenvolvidas ao longo deste trabalho. As referências utilizadas para a elaboração deste capítulo incluem os seguintes estudos: [70, 45, 8, 54, 58, 62, 69].

### 3.1 Entropia e Medidas de Dependência

O conceito de complexidade é amplamente utilizado nos dias de hoje, abrangendo uma vasta gama de abordagens científicas e tecnológicas aplicadas a sistemas naturais, artificiais e sociais. No entanto, sua definição quantitativa pode variar dependendo do contexto. Entre as abordagens mais relevantes e eficazes para analisar a complexidade de sistemas não lineares estão os métodos derivados da mecânica estatística e da teoria da informação [45]. Esses métodos, especialmente aqueles baseados em medidas de entropia, se destacam como ferramentas poderosas para a análise de séries temporais complexas, ajudando a identificar padrões e características de sistemas que não podem ser descritos adequadamente por métodos lineares tradicionais.

#### 3.1.1 Introdução ao Conceito de Transfer Entropy

O conceito de Transfer Entropy foi introduzido no início dos anos 2000, a partir da crescente necessidade de ferramentas mais robustas para a análise de causalidade em sistemas dinâmicos e não lineares. Thomas Schreiber [70] foi o pioneiro ao formalizar a TE, propondo-a como uma extensão e generalização da causalidade de Granger. Sua formulação é baseada em conceitos de teoria da informação, mais especificamente na Informação Mútua Condicional, permitindo a análise da transferência de informação entre duas séries temporais, condicionada ao passado de uma das séries.

Enquanto a entropia de Shannon mede a incerteza de um sistema, a Transfer Entropy vai além, medindo a direção e magnitude da transferência de informação entre duas variáveis ao longo do tempo.

Desde sua criação, o conceito de TE foi amplamente expandido e aplicado em diferentes áreas. Novas variações foram propostas, como a Transfer Entropy Condicional (que elimina a influência de variáveis de confusão) e métodos de estimativa não-paramétricos mais eficientes. Atualmente, a TE é amplamente usada em campos como neurociência para estudar a conectividade funcional do cérebro, finanças para analisar fluxos de informação no mercado, e física para investigar a dinâmica de sistemas complexos.

### 3.1.2 Entropia de Shannon

A entropia de Shannon pode ser derivada de maneira axiomática, como discutido em [58], baseando-se em algumas propriedades intuitivas de uma “medida de informação”. Essa abordagem nos ajuda a entender o conceito de entropia a partir de princípios básicos.

Suponha que tenhamos uma série de eventos  $E_k$ , cada um com uma probabilidade de ocorrência  $p_k$ . A informação associada ao evento  $E_k$  pode ser descrita como<sup>1</sup>:

$$I(E_k) = \log(1/p_k). \quad (3.1.1)$$

Esta fórmula reflete o quanto de “surpresa” ou “incerteza” está associada a cada evento, com base nas seguintes observações:

- Se  $p_k = 1$ , isso significa que não há surpresa na ocorrência do evento  $E_k$ .
- Se  $p_k = 0$ , então há surpresa infinita, afinal o evento  $E_k$  é impossível.
- $I(E_k) = -\log p(E_k)$  é a auto-informação do evento ou mensagem  $E_k$ .

Para uma variável aleatória (v.a.)  $X$ , denotamos a probabilidade de  $X = x$  como  $p(x)$ . O valor esperado de  $X$ , que mede o valor médio que  $X$  assume, é dado por:

$$X : E[X] = EX = \sum_x xp(x). \quad (3.1.2)$$

Agora, se considerarmos uma função  $g(x)$  aplicada a essa variável aleatória, o valor esperado de  $g(X)$  será:

$$Eg(X) = \sum_x g(x)p(x). \quad (3.1.3)$$

Se escolhermos a função  $g(x) = \log\left(\frac{1}{p(x)}\right)$ , então  $g(x)$  mede a surpresa de observar o evento  $X = x$ . Assim, o valor esperado dessa surpresa é dado por:

$$E \left[ \log \left( \frac{1}{p(x)} \right) \right] = \sum_x p(x) \log \left( \frac{1}{p(x)} \right). \quad (3.1.4)$$

Este valor é conhecido como a entropia de  $X$  e mede a incerteza média associada à variável aleatória.

**Definição 3.1.1** (Entropia). *A entropia de uma variável aleatória  $X$ , com distribuição de probabilidade  $p(x)$ , é formalmente definida [62] como :*

$$H(X) = E \left[ \log \left( \frac{1}{p(X)} \right) \right] = - \sum_x p(x) \log p(x), \quad (3.1.5)$$

*essa expressão representa a quantidade média de incerteza ou informação que está associada ao valor de  $X$ . Note que usamos logaritmo na base 2, então a entropia será medida em bits.*

---

<sup>1</sup>Com o surgimento dos computadores, a base 2 é de longe a mais usada na tecnologia atual. Consequentemente, a entropia é frequentemente expressa com um logaritmo para a base 2. A unidade correspondente é o bit, uma contração de dígito binário. Assim,  $M$  valores possíveis correspondem a  $\log_2 M$  bits. Foi o artigo de Shannon [62] de 1948 que introduziu a palavra "bit" como unidade fundamental de informação. Sendo assim, no presente trabalho, sempre utilizaremos a base 2 para o cálculo do logaritmo, desta forma,  $\log \equiv \log_2$ , a menos que seja especificado o contrário.

### 3.1.2.1 Propriedades da Entropia

Supondo um conjunto de eventos cujas probabilidades de ocorrências sejam dadas por  $p_1, p_2, \dots, p_n$ , é possível encontrar uma medida de quanta “escolha” está envolvida na seleção de um evento. Para tal medida  $H(p_1, p_2, \dots, p_n)$ , é razoável requerermos as seguintes propriedades:

- 1) **Continuidade:**  $H$  deve ser contínuo em  $p_i$  ;
- 2) **Crescimento Monótono:** Se todos os  $p_i$  são iguais,  $p_i = \frac{1}{n}$ , então  $H$  deve ser uma função monotonicamente crescente de  $n$  (quando temos eventos equiprováveis, teremos mais incerteza quão maior for o número de eventos possíveis);
- 3) **Adição de Escolhas:** Se uma escolha pode ser dividida em várias etapas independentes, a entropia total deve ser a soma ponderada das entropias de cada etapa. Isso reflete a ideia de que o processo de escolha pode ser quebrado em partes menores, sem perder a incerteza total do sistema. A Figura (3.1.1) ilustra essa ideia: inicialmente, escolhemos entre três possibilidades com probabilidades  $p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{6}$ . Na segunda etapa, uma escolha adicional é feita entre dois sub-eventos, mantendo as probabilidades finais inalteradas. De acordo com essa propriedade, a entropia da escolha original é a soma ponderada das entropias das etapas. Ou seja:

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H\left(\frac{2}{3}, \frac{1}{3}\right)$$

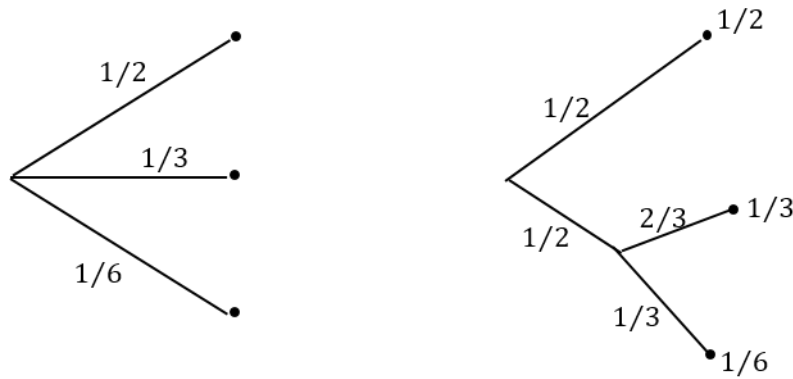


Figura 3.1.1: Decomposição de uma escolha entre três possibilidades.

A única função que satisfaz todas essas propriedades é a entropia de Shannon, dada pela fórmula:

$$H = -K \sum_{i=1}^n (p_i) \log(p_i); \quad (3.1.6)$$

onde  $K$  é uma constante positiva,  $p_i$  é a probabilidade de ocorrência de cada evento  $i$  e  $\sum_{i=1}^n p_i = 1$ .

### 3.1.3 Séries Temporais Simbólicas

Séries temporais simbólicas são representações de séries temporais onde os valores contínuos ou discretos das variáveis são mapeados para um conjunto finito de símbolos. Uma das grandes vantagens das séries temporais simbólicas é que elas preservam as características fundamentais da dinâmica do sistema original, enquanto tornam a análise mais acessível. Essa técnica é amplamente utilizada em análise de sistemas não lineares, permitindo o uso de ferramentas como a Transfer Entropy para estudar o fluxo de informação entre diferentes partes do sistema.

Uma série temporal simbólica é gerada a partir de um alfabeto de duas letras, por exemplo,  $\lambda = 2$  com os símbolos  $(0, 1)$ . Assim, a sequência simbólica pode ser representada como 0110100110010110.... Essa sequência pode ser lida em blocos consecutivos de comprimento  $n$ . Utilizando  $n = 2$ , por exemplo, obtemos os seguintes blocos: 01/10/10/01/10/01/01/10 [45].

O número de todos os tipos possíveis de palavras é  $\lambda^n = 2^2 = 4$ , ou seja, 00, 01, 10, 11. Sendo assim, as probabilidades necessárias para a estimativa de uma entropia,  $p_{00}$ ,  $p_{01}$ ,  $p_{10}$ ,  $p_{11}$  são as frações dos diferentes blocos (palavras), na série temporal simbólica, ou seja,  $p_{00} = 0$ ;  $p_{01} = \frac{4}{8}$ ;  $p_{10} = \frac{4}{8}$ ;  $p_{11} = 0$  [45]. Com base nessas probabilidades, podemos estimar a medida de entropia probabilística,  $H_s$ , introduzida por Shannon [62]:

$$H_s = - \sum_{i=1}^{\lambda^n} (p_i) \log_2(p_i),$$

em que  $p_i$  são as probabilidades associadas às configurações microscópicas. O logaritmo de base 2 é tomado aqui por conveniência no caso de uma simbolização binária (dando origem a uma interpretação da entropia de Shannon como a média negativa da quantidade de informação, medida em bits), mas pode ser substituída por outra base, por exemplo, usando  $\log_\lambda$ , sem perda de generalidade.

Nesse sentido, a entropia, introduzida por Claude Shannon em 1948, é um conceito fundamental da teoria da informação que quantifica a incerteza associada a uma variável aleatória. Em termos mais simples, a entropia mede o grau de imprevisibilidade ou desordem de um sistema.

## 3.2 Cópulas: Modelagem da Dependência Multivariada

Cópulas são funções matemáticas que permitem descrever e modelar a dependência entre variáveis aleatórias, separando essa dependência das distribuições marginais das variáveis. Essa abordagem é especialmente útil quando as distribuições marginais das variáveis são complexas ou desconhecidas, mas queremos focar na estrutura de dependência entre elas. Em particular, cópulas têm sido amplamente utilizadas em estatísticas, finanças, meteorologia, e análise de séries temporais, onde as relações entre variáveis não são necessariamente lineares.

Antes de introduzir o conceito de função de cópula, são lembrados alguns conceitos e resultados relacionados ao estudo de funções de distribuição. Estes resultados abordam a extensão do conceito de função crescente univariada para o caso multivariado.

**Definição 3.2.1.** *Sejam  $S_1$  e  $S_2$  dois subconjuntos não vazios de  $\mathbb{R} = \langle -\infty, +\infty \rangle$ , e seja  $H$  uma função de duas variáveis reais tal que seu domínio é  $S_1 \times S_2$ . Seja  $B =$*

$[x_1, x_2] \times [y_1, y_2]$  um retângulo tal que  $x_1, x_2, y_1, y_2 \in \text{Dom } H$ . Então, define-se o  $H$ -volume de  $B$  como:

$$V_H(B) := H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \quad (3.2.1)$$

*Observação 1.1.* Note que se definirmos as diferenças de primeira ordem de  $H$  no retângulo  $B$  como:

$$\Delta_{x_1}^{x_2} H(x, y) := H(x_2, y) - H(x_1, y) \text{ e } \Delta_{y_1}^{y_2} H(x, y) := H(x, y_2) - H(x, y_1), \quad (3.2.2)$$

então o  $H$ -volume do retângulo  $B$  é a diferença de segunda ordem de  $H$  em  $B$ ,

$$V_H(B) = \Delta_{x_1}^{x_2} \Delta_{y_1}^{y_2} H(x, y). \quad (3.2.3)$$

Esta interpretação nos dá uma possível generalização do conceito de função crescente para funções de  $n$  variáveis, tomando as diferenças de ordem  $n$  - ésimo para medir o crescimento da função em relação a dois pontos.

**Definição 3.2.2.** Uma função de duas variáveis reais  $H$  é bi-crescente se  $V_H(B) \geq 0$  para todos os retângulos  $B$  cujos vértices pertencem ao  $\text{Dom } H$ .

Esta extensão tem algumas peculiaridades e não tem relação alguma (não implica nem é implicada) com o conceito de função não decrescente<sup>2</sup> em cada um de seus argumentos.

**Exemplo 3.2.1.** Seja  $H : I^2 \rightarrow \mathbb{R}$  definida como:

$$H(x, y) := (2x - 1)(2y - 1),$$

e seja um retângulo qualquer  $B \in I^2$  definido por  $B := [x_1, x_2] \times [y_1, y_2]$ . Então

$$V_H(B) = 4(x_2 - x_1)(y_2 - y_1) \geq 0,$$

pelo que  $H$  é bi-crescente. No entanto, para um valor fixo  $y_0$ , temos que

$$F(x) := H(x, y_0) = [2(2y_0 - 1)]x - (2y_0 - 1),$$

isto é, uma reta com inclinação  $2(2y_0 - 1)$  que é negativa para  $y_0 < 1/2$ , e nesse caso teríamos que  $H$  é decrescente em seu primeiro argumento, apesar de ser bi-crescente.

O lema a seguir pode ser encontrado em [8].

**Lema 3.2.1.** Sejam  $S_1$  e  $S_2 \subseteq \mathbb{R}$  e  $H$  uma função bi-crescente, ou seja, funções que são crescentes em ambos os argumentos, com domínio  $S_1 \times S_2$ . Sejam  $x_1, x_2 \in S_1$  com  $x_1 \leq x_2$  e  $y_1, y_2 \in S_2$  com  $y_1 \leq y_2$ . Então a função  $t \in S_1 \rightarrow \varphi_1(t) := H(t, y_2) - H(t, y_1)$  não é decrescente em  $S_1$ , e a função  $t \in S_1 \rightarrow \varphi_2(t) := H(x_2, t) - H(x_1, t)$  não é decrescente em  $S_2$ .

**Demonstração.** Se verifica apenas para  $\varphi_1$ , pois é análogo para  $\varphi_2$ . Deve-se observar que  $\varphi_1$  não é decrescente, ou seja,

$$\begin{aligned} \forall t_1, t_2 \in S_1 : t_1 \leq t_2 &\implies \varphi_1(t_1) \leq \varphi_1(t_2) \Leftrightarrow \varphi_1(t_2) - \varphi_1(t_1) \geq 0. \\ \varphi_1(t_2) - \varphi_1(t_1) &= H(t_2, y_2) - H(t_2, y_1) - (H(t_1, y_2) - H(t_1, y_1)) \\ &= H(t_2, y_2) - H(t_2, y_1) - H(t_1, y_2) + H(t_1, y_1) \\ &= V_H([t_1, t_2] \times [y_1, y_2]) \geq 0. \end{aligned}$$

Então  $\varphi_1$  não é decrescente.

---

<sup>2</sup>Função crescente sem sê-lo no sentido estrito.

**Definição 3.2.3.** *Sejam  $S_1, S_2 \subset \mathbb{R}$  não vazios, diremos que a função  $H : S_1 \times S_2 \rightarrow \mathbb{R}$  é uma função grounded se, tomando  $a_1 = \inf S_1$  e  $a_2 = \inf S_2$ , verifica-se que*

$$H(x, a_2) = H(a_1, y) = 0, \forall (x, y) \in S_1 \times S_2.$$

**Definição 3.2.4.** *Dada a função  $H : S_1 \times S_2 \rightarrow \mathbb{R}$ , com  $S_1, S_2 \subset \mathbb{R}$ , definem-se as marginais de  $H$ , como as funções  $F$  e  $G$  dadas por:*

$$\begin{aligned} F(x) &= H(x, b_2), \quad \forall x \in \text{Dom } F = S_1, \text{ com } b_1 = \sup S_1 \text{ e} \\ G(y) &= H(b_1, y), \quad \forall y \in \text{Dom } G = S_2, \text{ com } b_2 = \sup S_2. \end{aligned}$$

De acordo com [54], primeiro definimos subcópulas como uma certa classe de funções bi-crescentes com margens; então definimos cópulas como subcópulas com domínio  $I^2$ .

**Definição 3.2.5.** *Uma subcópula bidimensional (ou 2-subcópula, ou abreviadamente, uma subcópula) é uma função  $C'$  com as seguintes propriedades:*

1.  $\text{Dom } C' = S_1 \times S_2$ , onde  $S_1$  e  $S_2$  são subconjuntos de  $I = [0, 1]$  contendo 0 e 1;
2.  $C'$  é grounded e bi-crescente;
3. Para cada  $u$  em  $S_1$  e cada  $v$  em  $S_2$ ,

$$C'(u, 1) = u \text{ e } C'(1, v) = v. \quad (3.2.4)$$

Observe que para todo  $(u, v)$  em  $\text{Dom } C'$ ,  $0 \leq C'(u, v) \leq 1$ , de modo que  $\text{Ran } C'$  também é um subconjunto de  $I$ .

**Definição 3.2.6.** *Uma cópula bidimensional (ou 2-cópula, ou brevemente, uma cópula) é a 2-subcópula  $C$  cujo domínio é  $I^2$ .*

*Equivalentemente, uma cópula é uma função  $C$  de  $I^2$  a  $I$  com as seguintes propriedades:*

1. Para cada  $u, v$  em  $I$ ,

$$C(u, 0) = 0 = C(0, v) \quad (3.2.5)$$

e

$$C(u, 1) = u \text{ e } C(1, v) = v; \quad (3.2.6)$$

2. Para cada  $u_1, u_2, v_1, v_2$  em  $I$  de tal modo que  $u_1 \leq u_2$  e  $v_1 \leq v_2$ ,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0. \quad (3.2.7)$$

Sendo assim, uma cópula é uma função que junta distribuições marginais univariadas para formar uma distribuição conjunta multivariada. Logo, dada  $F_X(x)$  uma função de distribuição cumulativa (CDF) de uma variável aleatória  $X$  e  $F_Y(y)$  a CDF de uma variável aleatória  $Y$ , a cópula  $C$  das variáveis  $X$  e  $Y$  é uma função que satisfaz:

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)), \quad (3.2.8)$$

onde  $F_{X,Y}(x, y)$  é a CDF conjunta das variáveis  $X$  e  $Y$ . A cópula  $C$  captura toda a dependência entre  $X$  e  $Y$ , independentemente das distribuições marginais  $F_X(x)$  e  $F_Y(y)$ .

### 3.2.1 Teorema de Sklar

O Teorema de Sklar é um dos resultados centrais na teoria das cópulas, afirmando que qualquer distribuição de probabilidade conjunta multivariada pode ser decomposta em duas partes: uma cópula que descreve a dependência entre as variáveis e as distribuições marginais que capturam o comportamento individual de cada variável.

Assim, começamos esta seção com uma breve discussão sobre funções de distribuição.

**Definição 3.2.7.** *Uma função de distribuição é uma função  $F$  com domínio  $\mathbb{R}$  tal que*

1.  $F$  não é decrescente,
2.  $F(-\infty) = 0$  e  $F(\infty) = 1$ .

**Definição 3.2.8.** *Uma função de distribuição conjunta (bivariada) é uma função  $H$  com domínio  $\mathbb{R}^2$  tal que*

1.  $H$  é bi-crescente,
2.  $H$  é contínuo à direita em cada variável,
3.  $H(x, -\infty) = H(-\infty, y) = 0$ ,  $\forall x, y \in \mathbb{R}$  e  $H(\infty, \infty) = 1$ .

Deste modo  $H$  é limitada inferiormente, e porque  $\text{Dom } H = \mathbb{R}^2$ ,  $H$  tem marginais  $F$  e  $G$ , dadas por  $F(x) = H(x, \infty)$  e  $G(y) = H(\infty, y)$ .

**Teorema 3.2.1.** *Seja  $H$  uma função de distribuição conjunta com distribuições marginais  $F$  e  $G$ . Então existe uma cópula  $C$  para todo  $x, y$  em  $\mathbb{R}$ ,*

$$H(x, y) = C(F(x), G(y)), \quad \forall x, y \in \mathbb{R}. \quad (3.2.9)$$

*Se  $F$  e  $G$  são contínuas, então  $C$  é única; caso contrário,  $C$  é exclusivamente determinado em  $\text{Ran } F \times \text{Ran } G$ . Por outro lado, se  $C$  é uma cópula e  $F$  e  $G$  são funções de distribuição, então a função  $H$  definida por (3.2.9) é uma função de distribuição conjunta com distribuições marginais  $F$  e  $G$ .*

### 3.2.2 Construindo a Densidade de Cópula Empírica

Tomando em conta o artigo de Shannon [58], a teoria da informação tem sido amplamente utilizada em muitas ramas da ciência. Na teoria da informação, entropia e informação mútua (IM) são dois conceitos fundamentais diferentes [20]. Entropia é definida como a medida de incerteza associada a variáveis aleatórias e a IM multivariada como uma medida da dependência mútua de variáveis aleatórias definidas como segue:

$$I(\mathbf{x}) = \sum_{i=1}^N H(x_i) - H(\mathbf{x}), \quad (3.2.10)$$

onde  $\mathbf{x} = [x_1, \dots, x_N] \in \mathbb{R}^N$  é uma variável aleatória,  $H$  é a entropia e  $I$  é a IM. Ao contrário das medidas estatísticas tradicionais de dependência, como correlação (ou seja, estatísticas de segunda ordem) e estatísticas de ordem superior [41], a IM é uma medida de todas as ordens de dependência.

Teoricamente, a cópula é uma função que representa a dependência entre variáveis aleatórias [39]. Tal representação é garantida pelo teorema de Sklar [65]. Seja  $P$  uma função de distribuição conjunta com margens  $\{F_i, i = 1, \dots, N\}$ . O teorema de Sklar

afirma que  $P$  pode ser representado como uma função de cópula  $C$  com  $F$  como seus argumentos:

$$P(\mathbf{x}) = C(F_1(x_1), \dots, F_N(x_N)). \quad (3.2.11)$$

Se os valores aleatórios  $F$  forem separados com margens da função de distribuição conjunta, a cópula tem todas as informações de dependência das variáveis aleatórias.

### 3.2.3 Informação Mútua e Entropia de Cópula

**Definição 3.2.9. (Entropia da cópula)** Seja  $\mathbf{x} \in \mathbb{R}^N$  variável aleatória com funções marginais  $\mathbf{u} = [F_1, \dots, F_N]$  e densidade de cópula  $c(\mathbf{u})$ . A entropia da cópula de  $\mathbf{x}$  é definida como:

$$H_c(\mathbf{x}) = - \int_{\mathbf{u}} c(\mathbf{u}) \log c(\mathbf{u}) d\mathbf{u}, \quad (3.2.12)$$

onde  $c(\mathbf{u}) = \frac{d^N C(\mathbf{u})}{du_1 du_2 \dots du_N}$ .

**Definição 3.2.10. (Entropia de Margem)** Seja  $\mathbf{x} \in \mathbb{R}^N$  variável aleatória com funções marginais  $[p_1, \dots, p_N]$ . A entropia de margem é definida como

$$H(x_i) = - \int_{x_i} p_i(x_i) \log p_i(x_i) dx_i, \quad (3.2.13)$$

onde  $i = 1, \dots, N$ .

**Teorema 3.2.2.** A informação mútua de variáveis aleatórias é equivalente à sua entropia de cópula negativa:

$$I(\mathbf{x}) = -H_c(\mathbf{x}). \quad (3.2.14)$$

**Demonstração.**

$$\begin{aligned} I(\mathbf{x}) &= \int_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{\prod_i p_i(x_i)} d\mathbf{x} \\ I(\mathbf{x}) &= \int_{\mathbf{x}} c(\mathbf{u}_x) \prod_i p_i(x_i) \log c(\mathbf{u}_x) d\mathbf{x} \\ I(\mathbf{x}) &= \int_{\mathbf{x}} c(\mathbf{u}_x) \log c(\mathbf{u}_x) d\mathbf{u}_x \\ I(\mathbf{x}) &= -H_c(\mathbf{x}). \end{aligned}$$

■

O teorema 3.2.2 pode ser interpretado como IM medindo exatamente a incerteza de dependência de variáveis aleatórias.

**Corolário 3.2.1.** A entropia de variáveis aleatórias conjuntas é composta de entropias de margem e a entropia de cópula.

$$H(\mathbf{x}) = \sum_{i=1}^N H(x_i) + H_c(\mathbf{x}). \quad (3.2.15)$$

**Demonstração.**

$$\begin{aligned} I(\mathbf{x}) = \sum_i H(x_i) - H(\mathbf{x}) &\Leftrightarrow H(\mathbf{x}) = \sum_i H(x_i) - I(\mathbf{x}) \\ &= \sum_i H(x_i) + H_c(\mathbf{x}). \end{aligned}$$

■

O corolário 3.2.1 indica que a entropia de variáveis aleatórias é composta de  $N$  entropias de margem e a entropia da cópula, ou seja, entropia de suas inter-relações.

Esse relacionamento é ilustrado na Figura 3.2.1, que dá uma visão na relação interna entre IM e cópula como uma conexão entre teoria da informação e cópula.

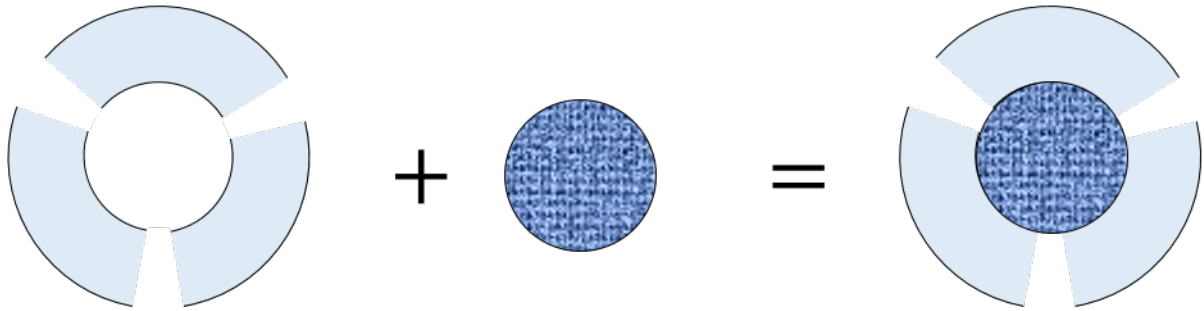


Figura 3.2.1: A relação entre a entropia de variáveis aleatórias, entropias de margem e entropia de cópula.

### 3.2.4 Estimando informações mútuas via Copula

As estimativas de IM são um problema comum em muitos campos científicos. Todos os métodos tradicionais são baseados na equação (3.2.10) que leva a análises tanto da entropia de margem quanto da entropia de todas as variáveis [6]. O Teorema 3.2.2 mostra que a estimativa de entropia e a IM serão os mesmos se a cópula puder ser estimada. Tais estimativas podem ser obtidas com cópula empírica.

Dadas amostras de distribuição idênticas e independentes  $\mathbf{X}_T = \{X_1, \dots, X_T\}$  geradas a partir de um  $N$  dimensional  $\mathbf{x} \in \mathbb{R}^N$ , a IM,  $I(\mathbf{x})$ , pode ser estimado como:

---

---

**Input:**  $\mathbf{X}_T = \{X_1, \dots, X_T\}$

**Output:**  $I(\mathbf{x})$

- 1 **Step** Determine a densidade de cópula empírica  $\hat{\mathbf{U}}_T$  de  $\mathbf{X}_T$ .
  - 2 **Step** Estime a entropia de  $\hat{\mathbf{U}}_T$  como  $I(\mathbf{x})$ .
- 

A densidade de cópulas empírica,  $\hat{\mathbf{U}}_T$ , pode ser determinada usando as funções empíricas:

$$U_t^i = \frac{1}{T} \sum_{t=1}^T 1_{\{x_t^i \leq x_t^i\}}, \quad (3.2.16)$$

onde  $i = 1, \dots, N$  e  $1$  é uma função indicadora. Em Step 1 pode-se usar outros métodos, como métodos analíticos ou kernel [64]. O método empírico é a opção mais fácil

porque é intuitivo e facilmente calculado, pois pode ser resolvido como um problema de classificação. A seleção de um método de estimativa depende de muitos fatores, como informações básicas, requisitos de precisão ou preferências pessoais, etc. Muitos métodos foram desenvolvidos para estimativas de entropia, então a entropia de  $\hat{\mathbf{U}}_T$  a Step 2 pode ser estimada usando muitos métodos bem estabelecidos [15], [46]. O método atual não é paramétrico, então todos os métodos anteriores para estimar a IM podem ser transformados neste método que precisa apenas da estimativa de entropia. Assim, as estimativas de IM podem ser obtidas simplesmente estimando uma entropia em vez de muitas como na equação (3.2.10). Essa melhoria é muito significativa, pois as estimativas de IM são amplamente utilizadas e computacionalmente onerosas.

### 3.2.5 Transfer Entropy via Entropia de Copula

Nesta seção, propomos um método para estimar TE via CE [37]. Primeiro, definimos a TE que é inspirada pela noção de Informação Mútua (IM), como segue:

**Definição 3.2.11.** (*Entropia de Transferência*). TE entre as séries temporais  $X_i, Y_i, i = 1, \dots, T$ , é definida como:

$$TE = \sum p(Y_{i+1}, Y^i, X_i) \log \left( \frac{p(Y_{i+1}|Y^i, X_i)}{p(Y_{i+1}|Y^i)} \right), \quad (3.2.17)$$

onde  $Y^i = (Y_1, \dots, Y_i)$ .

Daí, derivamos uma representação de TE com CE.

**Proposição 3.2.1.** TE pode ser representado apenas com CE.

**Demonstração.** A demonstração começa a partir da definição de TE (3.2.17). Após expandir a equação de definição, a definição pode ser facilmente transformada na equação composta de quatro termos CE diferentes.

$$TE = \sum p(Y_{i+1}, Y^i, X_i) \log \frac{p(Y_{i+1}|Y^i, X_i)}{p(Y_{i+1}|Y^i)} \quad (3.2.18)$$

$$= \sum p(Y_{i+1}, Y^i, X_i) \log \frac{p(Y_{i+1}, Y^i, X_i)p(Y^i)}{p(Y_{i+1}, Y^i)p(Y^i, X_i)} \quad (3.2.19)$$

$$= I(Y_{i+1}; Y^i, X_i) - I(Y_{i+1}; Y^i) - I(Y^i; X_i) \quad (3.2.20)$$

$$= -H_c(Y_{i+1}; Y^i, X_i) + H_c(Y_{i+1}; Y^i) + H_c(Y^i; X_i) \quad (3.2.21)$$

$$= -H_c(Y_{i+1}, Y^i, X_i) + H_c(Y_{i+1}, Y^i) + H_c(Y^i; X_i) - H_c(Y^i). \quad (3.2.22)$$

O último termo  $H_c(Y^i)$  em (3.2.22) é igual a 0 se  $Y^i = Y_i$ .

## 3.3 Estimativas de Entropia

### 3.3.1 Estimando entropia com o método KNN

Seja a variável aleatória em consideração  $\mathbf{X} \in \mathbb{R}^d$  e sua densidade de probabilidade denotada por  $p_{\mathbf{X}}(\mathbf{x})$ . Sua entropia é definida como:

$$H(\mathbf{X}) = \int_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \log \left( \frac{1}{p_{\mathbf{X}}(\mathbf{x})} \right) d\mathbf{x}, \quad (3.3.1)$$

onde  $x$  é o vetor no suporte da densidade  $p_{\mathbf{X}}(\mathbf{x})$ . O objetivo é estimar  $H(\mathbf{X})$  a partir de  $N$  amostras finitas,  $\mathbf{x}_i$   $i = 1 \dots N$ , a partir da distribuição  $p_{\mathbf{X}}(\mathbf{x})$ . Uma estimativa de Monte-Carlo da entropia pode ser escrita como

$$\hat{H}(\mathbf{X}) = \frac{1}{N} \sum_{i=1}^N \log \left( \frac{1}{p_{\mathbf{X}}(\mathbf{x}_i)} \right). \quad (3.3.2)$$

No entanto, como  $p_{\mathbf{X}}(\mathbf{x}_i)$  é desconhecida, uma estimativa  $\hat{p}_{\mathbf{X}}(\mathbf{x}_i)$  deve ser substituída na equação (3.3.2) para obter  $\hat{H}(\mathbf{X})$ .

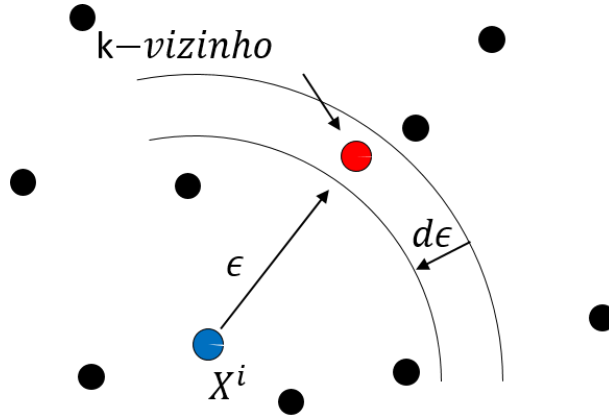


Figura 3.3.1: Uma representação do  $k$ -vizinhos mais próximo e da  $\epsilon$ -bola.

A ideia-chave é estimar  $\hat{p}_{\mathbf{X}}(\mathbf{x}_i)$  através de  $k$ -vizinhos mais próximos (kNN) de  $\mathbf{x}_i$ . Considere a densidade de probabilidade  $p_k(\epsilon)$  de  $\epsilon$ , a distância de  $\mathbf{x}_i$  para seu kNN (veja a Figura (3.3.1)). A probabilidade  $p_k(\epsilon)d\epsilon$  é a probabilidade de que exatamente um ponto esteja em  $[\epsilon, \epsilon + d\epsilon]$ , exatamente  $k - 1$  pontos estejam a distâncias menores que o kNN, e os pontos restantes estejam mais distantes que o kNN. Então, segue que

$$p_k(\epsilon)d\epsilon = \binom{N-1}{1} \frac{dP_i(\epsilon)}{d\epsilon} d\epsilon \binom{N-2}{k-1} (P_i(\epsilon))^{k-1} (1 - P_i(\epsilon))^{N-k-1} \quad (3.3.3)$$

onde,  $P_i(\epsilon)$  é a massa de probabilidade de uma bola  $\epsilon$  centrada em um ponto de amostra  $\mathbf{x}_i$ . A região dentro da bola  $\epsilon$  é  $\|\mathbf{x} - \mathbf{x}_i\| < \epsilon$  e é denotada por  $B(\epsilon, \mathbf{x}_i)$ . A massa de probabilidade em  $B(\epsilon, \mathbf{x}_i)$  é

$$P_i(\epsilon) = \int_{B(\epsilon, \mathbf{x}_i)} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (3.3.4)$$

O valor esperado de  $\log(P_i)$  pode ser obtido a partir das equações (3.3.3) e (3.3.4)

$$E(\log P_i) = \int_0^\infty \log P_i(\epsilon) p_k(\epsilon) d\epsilon = \Psi(k) - \Psi(N), \quad (3.3.5)$$

onde  $\Psi$  é a função digama.

Se a massa de probabilidade em  $B(\epsilon, \mathbf{x}_i)$  pode ser escrita na seguinte forma

$$P_i \approx \eta_i p_{\mathbf{X}}(\mathbf{x}_i) \quad (3.3.6)$$

então, considerando o logaritmo e tomando expectativas em ambos os lados da equação (3.3.6) e usando as equações (3.3.5) e (3.3.2), a estimativa de entropia pode ser escrita como

$$\hat{H}(\mathbf{X}) = \Psi(N) - \Psi(k) + \frac{1}{N} \sum \log \eta_i. \quad (3.3.7)$$

### A. Estimadores clássicos

As estimativas clássicas de Kozachenko e Leonenko [42], e similarmente de Singh [34], assumem que a densidade probabilidade  $p_{\mathbf{X}}(\mathbf{x})$  é constante dentro de  $B(\epsilon, \mathbf{x}_i)$ . Por exemplo, Kozachenko e Leonenko [6], [42] assumem que

$$P_i \approx c_d \epsilon^d p_{\mathbf{X}}(\mathbf{x}_i), \quad (3.3.8)$$

onde  $c_d$  é o volume da bola unitária  $d$ -dimensional ( $B(\epsilon, \mathbf{x}_i)$  com  $\epsilon = 1$ ). A expressão para  $c_d$  depende do tipo de norma usada para calcular as distancias; por exemplo, para a norma máxima ( $L_\infty$ ),  $c_d = 2^d$  e para a norma euclidiana ( $L_2$ ),  $c_d = \pi^{d/2}/\Gamma(1 + d/2)$ , onde  $\Gamma$  é a função Gama. Usando a equação (3.3.8) nas equações (3.3.6) e (3.3.7), a estimativa de entropia pode ser escrita como

$$\hat{H}(\mathbf{X}) = \Psi(N) - \Psi(k) + \log(c_d) + \frac{d}{N} \sum_{i=1}^N \log(\epsilon(i)), \quad (3.3.9)$$

onde  $\epsilon(i)$  é a distância da amostra  $i$ -ésima até seu vizinho mais próximo  $k$ -ésimo. Este estimador é chamado de estimador KL.

## 3.4 Testes Estatísticos Baseados em Entropia

### 3.4.1 Teste de independência condicional

Granger [33] desenvolveu a noção de causalidade entre duas séries temporais estacionárias com base na filosofia de que a causa deve melhorar a previsão de efeitos. Dadas duas variáveis aleatórias  $X_t, Y_t, t = 1, \dots, T$ , Causalidade de Granger (GC) é definida como segue:

**Definição 3.4.1.** (*Causalidade de Granger*)

$$\delta^2(Y_{t+1}|Y_t, X_t) < \delta^2(Y_{t+1}|Y_t), \quad (3.4.1)$$

se  $X$  causa  $Y$ , onde  $\delta^2$  denota variância.

Entropia de Transferência é outra medida teórica da informação de causalidade definida para séries temporais estacionárias por Schreiber [70].

A definição 3.2.11 pode ser escrita como a Informação Mutua Condicional da seguinte forma:

$$TE = I(Y_{i+1}; X_i | Y^i). \quad (3.4.2)$$

De acordo com essa definição, a TE é uma medida independente de modelo e pode ser interpretada como o fluxo de informação de causa para efeito. A TE assume que as séries temporais são estacionárias. Adicionalmente, como probabilidades condicionais são usadas na definição, isso implica que a TE também assume a propriedade de Markov. Embora tenha grande potencial de aplicação em muitos campos, a TE tem sido considerada notoriamente difícil de ser estimada [43]. Tanto a Causalidade de Granger quanto a Entropia de Transferência são definidas com base na mesma filosofia de causalidade. Barnett, mostra que a GC e a TE são equivalentes sob a suposição de Gaussianidade [3]. Além disso, a TE pode ser aplicada a casos mais amplos que a GC. A Informação Direcionada é uma noção intimamente relacionada à TE, sendo definida como a soma de um grupo de TE e uma IM [36].

### 3.4.2 Teste de Normalidade Multivariada

Dado o vetor aleatório  $\mathbf{X} \in \mathbb{R}^d$ , a hipótese para o teste de normalidade multivariada é

$$H_0 : p(\mathbf{x}) \in N_d(V_x), \quad (3.4.3)$$

onde  $N_d$  denota distribuição normal, e  $V_x$  são as covariâncias de  $\mathbf{X}$ .

A estatística de teste para  $H_0$  é definida como segue:

$$T_{ce} = H_c(\mathbf{x}) - H_c(\mathbf{x}_n), \quad (3.4.4)$$

onde  $\mathbf{x}_n$  é o vetor aleatório gaussiano com as mesmas covariâncias. O primeiro termo em (3.4.4) pode ser estimado com o estimador CE não paramétrico. O segundo termo em (3.4.4) pode ser derivado analiticamente como

$$H_c(\mathbf{x}_n) = \frac{1}{2} \log(|V_x|). \quad (3.4.5)$$

É fácil saber  $T_{ce} = 0$  para distribuições normais.

### 3.4.3 Teste de duas amostras com CE

Dadas duas amostras  $\mathbf{X}_1 = \{X_{11}, \dots, X_{1m}\} \sim P_1$ ,  $\mathbf{X}_2 = \{X_{21}, \dots, X_{2n}\} \sim P_2$ , a hipótese nula para o teste de duas amostras é

$$H_0 : P_1 = P_2, \quad (3.4.6)$$

e a alternativa é

$$H_1 : P_1 \neq P_2. \quad (3.4.7)$$

Onde  $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^d$  e  $P_1, P_2$  são as funções de distribuição de probabilidade correspondentes.

Sejam  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$  e  $Y_0, Y_1$  duas variáveis de rotulagem para as duas hipóteses, respectivamente  $Y_1 = (0_1, \dots, 0_m, 1_1, \dots, 1_n)$  e  $Y_0 = (1_1, \dots, 1_{m+n})$ . Então a CE entre  $\mathbf{X}$  e  $Y_i$  pode ser calculada como

$$H_c(\mathbf{X}; Y_i) = H_c(\mathbf{X}, Y_i) - H_c(\mathbf{X}). \quad (3.4.8)$$

Então a estatística de teste para  $H_0$  é definida como a diferença entre as CEs das duas hipóteses, como segue:

$$T_{ce}(X_1, X_2) = H_c(\mathbf{X}, Y_0) - H_c(\mathbf{X}, Y_1). \quad (3.4.9)$$

É fácil perceber que  $T_{ce}$  terá um valor pequeno se  $H_0$  for verdadeiro e um valor grande se  $H_1$  for verdadeiro. A estatística de teste em (3.4.9) pode ser facilmente estimada a partir dos dados, estimando os dois termos nela presentes com o estimador não paramétrico da CE. Como o estimador de CE é não paramétrico, ele pode ser aplicado em qualquer caso sem a necessidade de assumir hipóteses específicas. Outra vantagem desse estimador da estatística de teste é que ele não requer hiperparâmetros.



# Metodologia

## 4.1 Base de Dados

Os dados utilizados são os preços semanais da gasolina e do etanol, disponibilizados pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP)<sup>1</sup>, enquanto o preço da saca do açúcar é disponibilizado pelo Centro de Estudos Avançados em Economia Aplicada (Cepea), da Esalq/USP<sup>2</sup>. Analisamos as *commodities*: açúcar, etanol e gasolina, durante o período de 9 de maio de 2004 a 19 de novembro de 2023.

A tabela (4.1.1), apresenta as unidades de medida e o tipo de moeda utilizada das *commodities* analisadas.

| Nome     | Moeda/Unidade de medida |
|----------|-------------------------|
| Açúcar   | R\$ / Saca de 50kg      |
| Etanol   | R\$ / Litros            |
| Gasolina | R\$ / Litros            |

Tabela 4.1.1: Moeda e unidade de medida das *commodities*.

Os dados também foram divididos em três amostras. Uma antes da pandemia de Covid-19 (07 de janeiro de 2018 a 29 de dezembro de 2019), uma durante a pandemia (05 de janeiro de 2020 a 26 de dezembro de 2021) e uma depois da pandemia (02 de janeiro de 2022 a 26 de novembro de 2023), com o objetivo de avaliar se a pandemia afetou a direção de fluxo de informação entre *commodities*.

## 4.2 Etapas da Pesquisa

- **Coleta e tratamento dos dados:** A primeira etapa do pré-processamento consiste em identificar e definir fontes confiáveis de dados, realizar a limpeza (*data cleaning*), organizar e rotular os dados, eliminando inconsistências, redundâncias e erros que possam afetar os resultados da análise. Esse processo é fundamental para modelar os dados de forma que se tornem úteis e manejáveis para algoritmos específicos de análise.

<sup>1</sup>Disponível em: <https://www.gov.br/anp/pt-br/assuntos/precos-e-defesa-da-concorrencia/precos>.

<sup>2</sup>Disponível em: <https://www.cepea.esalq.usp.br/br/indicador/acucar.aspx>.

Essa etapa faz parte do processo de Descoberta de Conhecimento em Dados (do inglês *Knowledge Discovery in Databases - KDD*) [23], uma área multidisciplinar que combina Estatística, Banco de Dados e Visualização Gráfica. O objetivo é entender mais sobre os dados, utilizando técnicas de exploração e visualização a fim de identificar inconsistências e/ou padrões.

- **Procedimento Metodológico:** Neste trabalho, foram empregadas técnicas de visualização para a geração de representações gráficas das séries temporais de *commodities*, com o objetivo de identificar padrões, tendências e possíveis anomalias nos dados. A visualização dessas séries auxilia na interpretação dos comportamentos subjacentes e na extração de percepções relevantes.

Além disso, foi aplicada uma metodologia baseada em análise de informação para caracterizar a dinâmica temporal das séries. Essa abordagem permite quantificar a transferência de informação entre variáveis ao longo do tempo, possibilitando uma compreensão mais profunda das relações causais e das interdependências existentes no conjunto de dados.

Os métodos desenvolvidos serão implementados sobre a base de dados fornecida pela fonte, garantindo uma análise consistente e fundamentada para a extração de conhecimento a partir das séries temporais estudadas.

### 4.3 Ferramentas Utilizadas

- **Microsoft Excel (Microsoft Corporation):** A criação da base com os dados já mencionados, foi realizado utilizando um arquivo do tipo CSV (do inglês, *comma-separated values*) através do Excel. Esse tipo de arquivo proporciona maior estabilidade para o tratamento de dados, ajudando a evitar tempo excessivo de processamento ou problemas de armazenamento.
- **Python:** Para a aplicação do método Transfer Entropy (TE), é essencial que os dados estejam sincronizados. Esse requisito já é atendido, uma vez que as séries temporais de açúcar, etanol e gasolina possuem periodicidade semanal. As análises foram realizadas utilizando o pacote *Transfer Entropy* do software Python, especificamente por meio da biblioteca *Copent*.

A estimativa da entropia de cópula segue um processo em duas etapas: primeiro, a cópula empírica é estimada por meio da estatística de ordem; em seguida, a entropia de cópula é calculada utilizando o método KSG. Esse conceito matemático permite medir e testar a independência estatística multivariada e, conforme demonstrado, é equivalente à informação mútua. Diferentemente do coeficiente de correlação de Pearson, a entropia de cópula é aplicável a relações não lineares, de ordem superior e multivariadas, tornando-se uma ferramenta versátil para a análise de dependências complexas entre variáveis.

A Transfer Entropy foi calculada para três diferentes representações das séries temporais:

- Preços ( $P_t$ ) das *commodities*;
- Retornos ( $R_t$ ), definidos como  $R_t = \ln P_t - \ln P_{t-1}$ ;
- Volatilidade ( $V_t$ ), definida como  $V_t = |R_t|$ .

Para cada par de *commodities*, a Transfer Entropy foi estimada em ambas as direções:

$$TE(Y \longrightarrow X) \text{ e } TE(X \longrightarrow Y),$$

em que  $X$  e  $Y$  representam as séries temporais analisadas.

Se  $TE(Y \longrightarrow X) > 0$ , considera-se que há um fluxo de informação de  $Y$  para  $X$ ; caso contrário, considera-se que a direção do fluxo de informação ocorre de  $X$  para  $Y$  [38].

A média da Transfer Entropy é o valor médio TE calculada ao longo de múltiplas janelas temporais ou entre diferentes pares de séries. É usada para:

- Suavizar variações: Reduzir o impacto de ruídos ou flutuações temporais.
- Generalizar relações causais: Identificar padrões estáveis de influência entre variáveis.

Pelo tanto a média da TE é simplesmente a média aritmética dos valores de TE obtidos:

$$\text{Média}T_{Y \rightarrow X} = \frac{1}{N} \sum_{i=1}^N T_{Y \rightarrow X}^{(i)},$$

onde:

- $N$  : Número de janelas, pares ou ensaios.
- $T_{Y \rightarrow X}^{(i)}$  : TE calculada para a  $i$ -ésima janela, par ou ensaio.



## Resultados e discussões

Neste capítulo, buscamos realizar uma análise das séries temporais, a partir do método de Transfer Entropy, como uma forma de interpretar o comportamento dos preços da gasolina, etanol e açúcar, no estado de São Paulo.

Conforme já mencionado, os valores estão expressos em reais e abrangem o período de 9 de maio de 2004 a 19 de novembro de 2023. Para a aplicação da Transfer Entropy, é fundamental que os dados estejam sincronizados temporalmente. Como os preços de açúcar, gasolina e etanol são reportados semanalmente, essa regularidade temporal foi mantida nas análises, que foram conduzidas por meio do pacote Transfer Entropy, disponível na linguagem Python, através do Google Colaboratory.

As Figuras 5.0.1, 5.0.2 e 5.0.3 apresentam as séries temporais de preço, retorno e volatilidade dos preços do açúcar, gasolina e etanol, respectivamente. Observamos que os preços variam ao longo do tempo: em alguns anos, os preços são mais baixos, enquanto em outros há um aumento brusco.

A Figura 5.0.1 ilustra a evolução dos preços do açúcar ao longo do tempo, indicando um crescimento marcado por picos e vales notáveis. O preço mínimo registrado foi de R\$ 21,3480 por saca, observado nos primeiros meses da análise, enquanto o preço máximo, alcançado nos últimos meses de 2023, chegou a R\$ 158,225 por saca, refletindo um aumento expressivo. A média dos preços foi de R\$ 65,5632 por saca, sugerindo uma tendência de alta a longo prazo no mercado de açúcar, impulsionada por fatores como a demanda global e a produção.

A variabilidade dos preços também é notável, com um desvio padrão de R\$ 33,6582 por saca, o maior entre as três *commodities* analisadas, o que indica que o mercado de açúcar é particularmente suscetível a flutuações abruptas. Esse comportamento cíclico, com períodos de alta e baixa volatilidade, está fortemente relacionado à sazonalidade agrícola e à demanda internacional [53]. Além disso, os retornos sobre o preço do açúcar exibem uma alta frequência de variações bruscas, características comuns em mercados de *commodities* agrícolas. Em suma, a volatilidade e a dispersão nos preços refletem a sensibilidade do açúcar a condições climáticas, variações na oferta e fatores políticos internacionais [2].

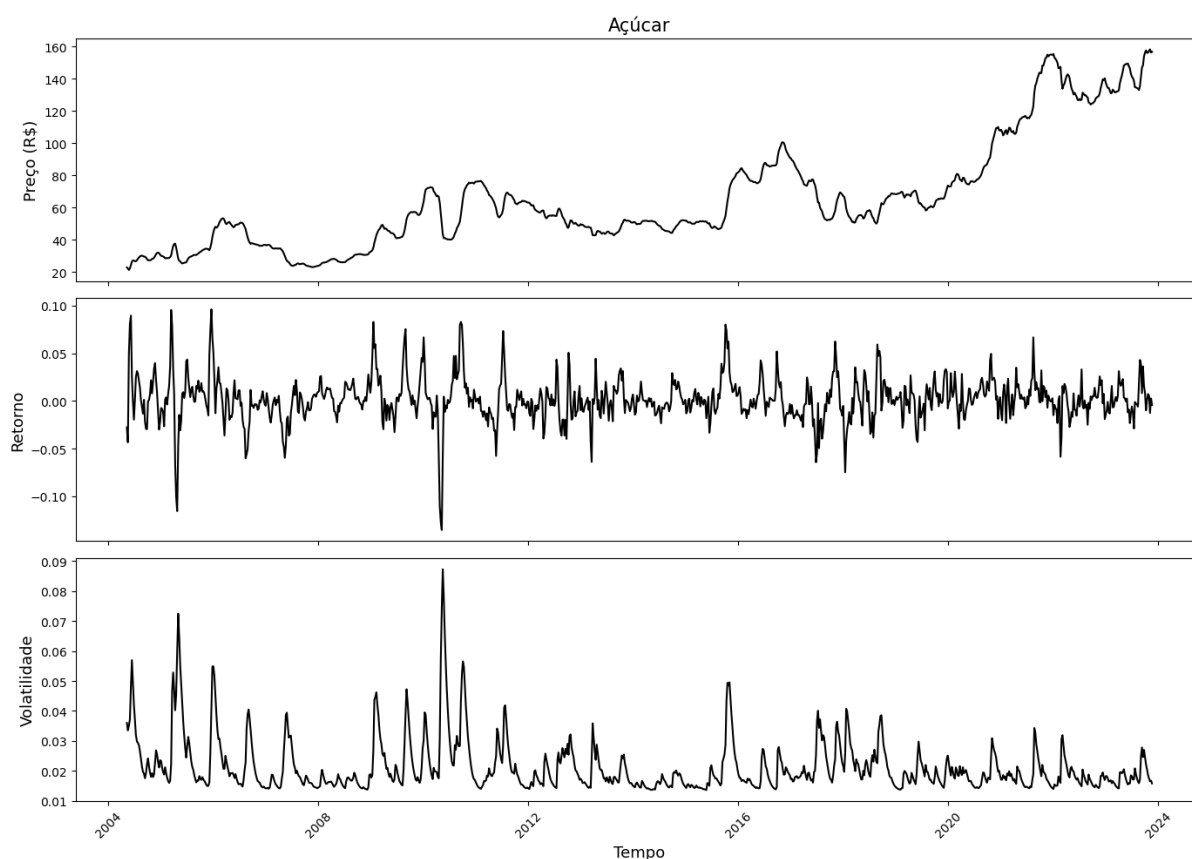


Figura 5.0.1: Séries de preço, retorno e volatilidade do açúcar.

A Figura 5.0.2 ilustra a evolução temporal dos preços da gasolina, destacando um crescimento gradual e sustentado ao longo dos anos, intercalado por períodos de relativa estabilidade e volatilidade moderada. Entre 2020 e 2022, período mais crítico da pandemia de Covid-19, observamos um aumento acelerado nos preços, seguido por uma correção. Durante todo o período de crescimento, o preço mínimo registrado foi de R\$ 1,8880 por litro, reflexo de uma fase de estabilidade. Já o preço máximo alcançou R\$ 6,992 por litro, em 2022. A média do preço da gasolina ao longo do período foi de R\$ 3,3458 por litro, enquanto o desvio padrão foi de R\$ 1,1551 por litro, sugerindo que, embora o preço tenha sofrido variações, essas oscilações foram relativamente moderadas.

Além disso, o retorno representa variações percentuais semanais do preço da gasolina. Observa-se um comportamento oscilatório, com momentos de retornos positivos e negativos, indicando períodos de ganhos e perdas. A volatilidade mede a intensidade das variações no retorno. Nota-se que há períodos de maior volatilidade, caracterizados por picos, sugerindo momentos de maior incerteza e instabilidade nos preços da gasolina.

Apesar das flutuações, a gasolina se manteve como um produto com comportamento estável em termos de preços, apresentando uma tendência de alta consistente ao longo do tempo.

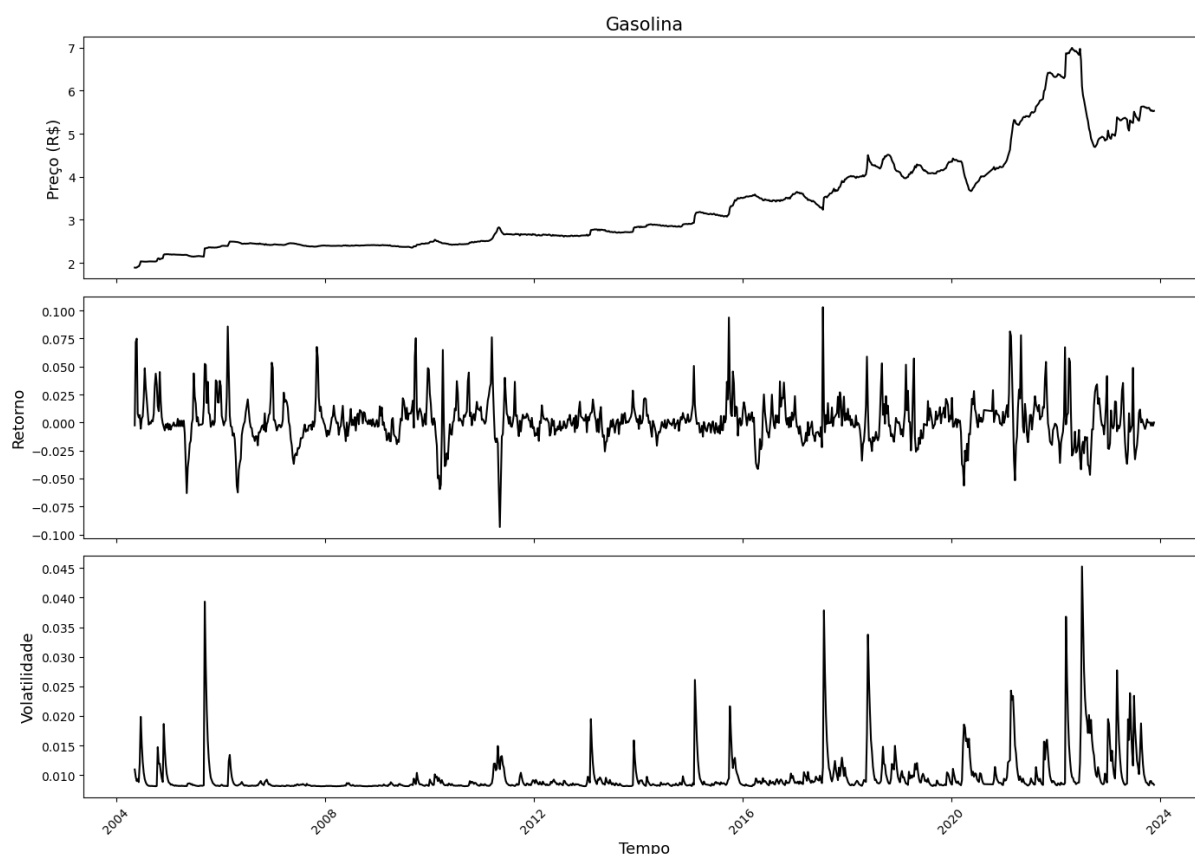


Figura 5.0.2: Séries de preço, retorno e volatilidade da gasolina.

A Figura 5.0.3 retrata a evolução temporal dos preços do etanol, evidenciando um aumento progressivo com maior variabilidade em comparação aos preços da gasolina. A série temporal do etanol revela mudanças bruscas em diversos períodos, especialmente em torno de 2010, período de chuvas excessivas que afetaram a produção de cana-de-açúcar, reduzindo a oferta e elevando os preços do etanol e, assim como a gasolina, entre 2020 e 2022, devido a Covid-19. O preço mínimo registrado foi de R\$ 0,766 por litro, ocorrido nas primeiras semanas da análise, enquanto o preço máximo atingiu R\$ 5,348 por litro, no período pandêmico. A média dos preços foi de R\$ 2,2003 por litro, sugerindo que o etanol é uma alternativa de custo mais baixo em comparação à gasolina. Com um desvio padrão de R\$ 0,9564 por litro, fica evidente que o preço do etanol apresenta maior volatilidade e está sujeito a mais flutuações do que gasolina.

O retorno, apresenta oscilações constantes ao longo do tempo, com valores positivos e negativos. Há momentos em que os retornos são mais acentuados, indicando períodos de maior variação nos preços. A dispersão dos retornos sugere que, em determinados períodos, os preços do etanol experimentaram mudanças bruscas, refletindo possíveis impactos de fatores externos. Também na volatilidade, evidencia a intensidade das variações no retorno ao longo do tempo. Os picos de volatilidade indicam momentos de maior instabilidade nos preços do etanol, especialmente em períodos de crise ou choques no mercado. Nota-se que há períodos de baixa volatilidade intercalados com momentos de alta volatilidade, o que sugere que a incerteza no mercado do etanol ocorre de forma não uniforme.

Pelo tanto, o retorno reflete as flutuações dos preços do etanol, enquanto a volatilidade destaca os momentos de maior ou menor estabilidade no mercado. Esses comportamentos podem estar relacionados a fatores econômicos, condições agrícolas, e políticas relacionadas aos biocombustíveis que influenciam a oferta e a demanda dessa *commodity*.

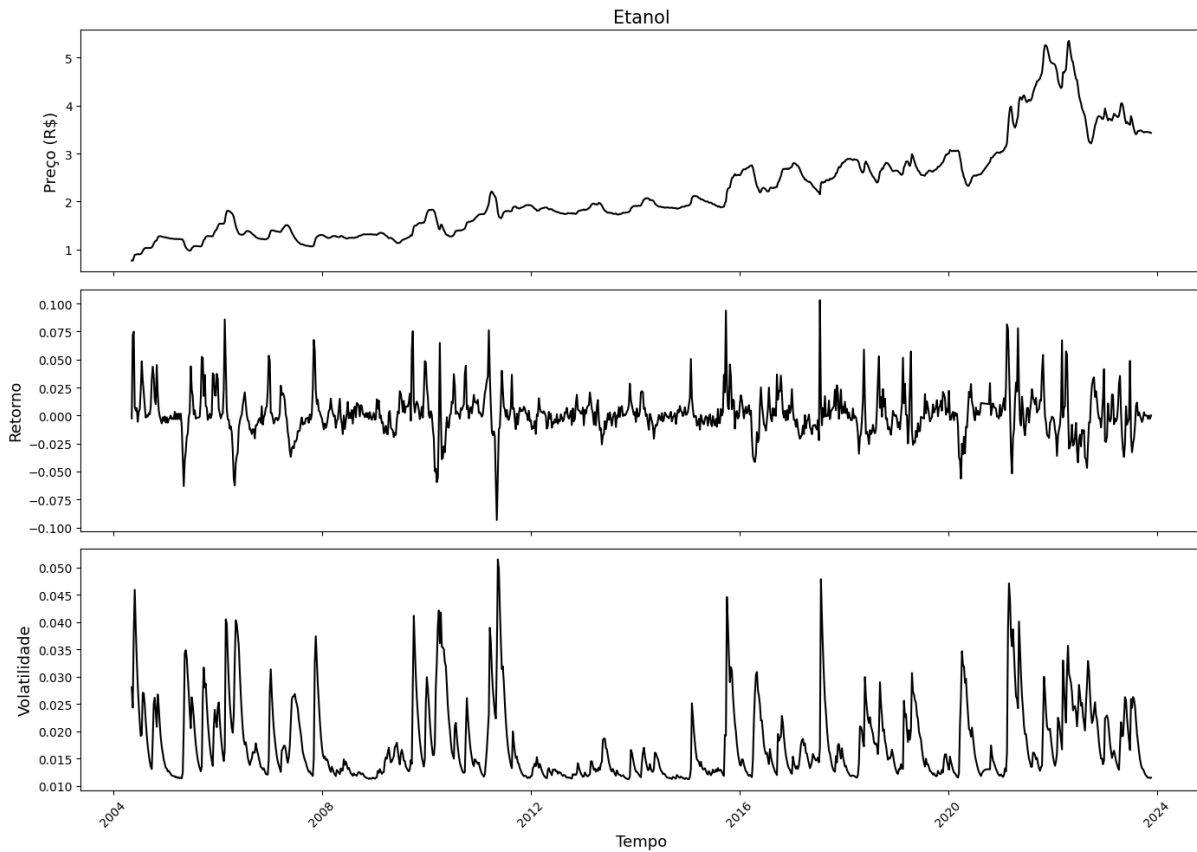


Figura 5.0.3: Séries de preço, retorno e volatilidade do etanol.

A Tabela 5.0.1 e a Figura 5.0.4 apresentam os valores de Transfer Entropy entre cada par de séries, refletindo a quantidade de informação que a série na linha transfere para a série na coluna. Observa-se uma relação bidirecional entre os preços do etanol e da gasolina, sendo que a influência da gasolina sobre o etanol é ligeiramente maior do que a influência inversa. Esse resultado é coerente, uma vez que o etanol, como biocombustível, compete diretamente com a gasolina em certos mercados, especialmente em regiões que incentivam a mistura de biocombustíveis com combustíveis fósseis. Quando o preço da gasolina aumenta, a demanda por etanol como alternativa também pode crescer, pressionando seus preços para cima.

No que diz respeito à relação entre o etanol e o açúcar, a influência do preço do etanol sobre o açúcar (0.505381) é bastante significativa, enquanto a influência do açúcar sobre o etanol (0.442021) é um pouco menor, mas ainda relevante. Isso é consistente com o fato de que o etanol é produzido a partir da cana-de-açúcar, estabelecendo uma interdependência entre os dois mercados.

Já na interação entre os preços da gasolina e do açúcar, o primeiro exerce uma influência moderada sobre o segundo (0.458101), enquanto a transferência inversa é mais fraca (0.363305). Essa maior influência da gasolina sobre o açúcar pode ser entendida como uma consequência da conexão entre o mercado de açúcar e a produção de biocombustíveis, onde variações nos custos de energia afetam diretamente os preços das *commodities* agrícolas ligadas à produção de biocombustíveis.

| →        | Açúcar   | Etanol   | Gasolina |
|----------|----------|----------|----------|
| Açúcar   | -        | 0.442021 | 0.363305 |
| Etanol   | 0.505381 | -        | 0.287834 |
| Gasolina | 0.458101 | 0.351449 | -        |

Tabela 5.0.1: Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para o período total estudado.

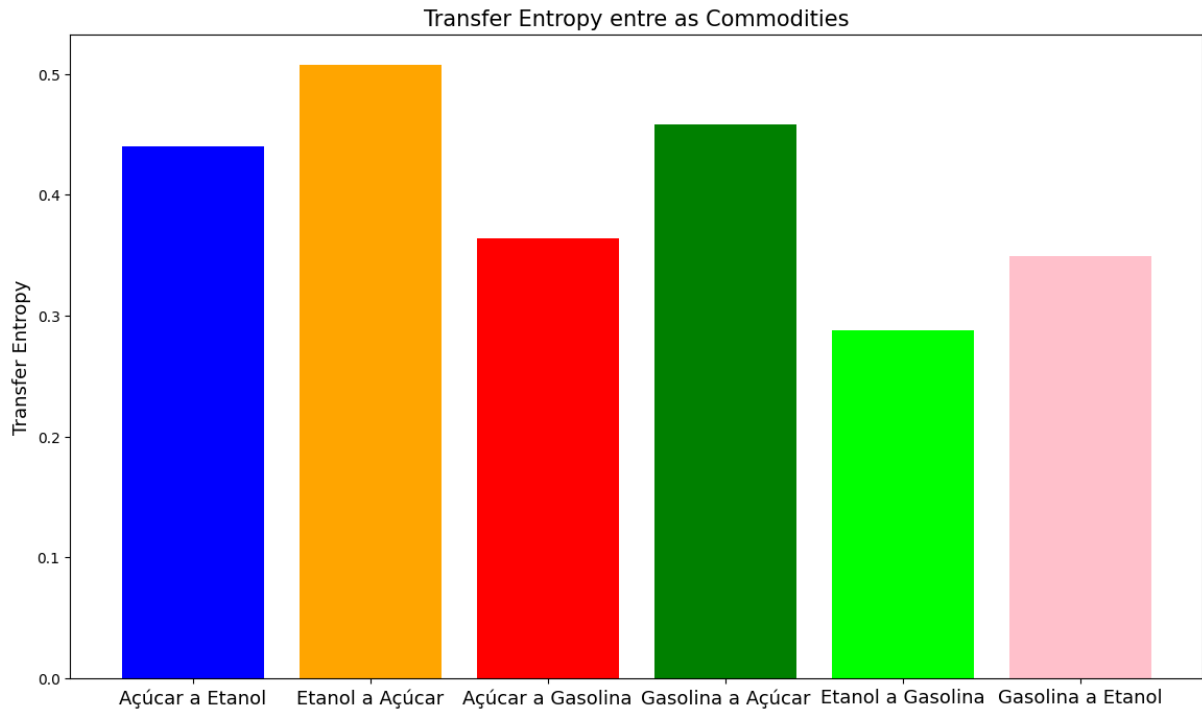


Figura 5.0.4: Transfer Entropy entre as séries da volatilidade de preços entre as *commodities* estudadas.

Ao examinar a Figura 5.0.5, observa-se que, na maioria das escalas de tempo, o fluxo de informação no sentido etanol  $\rightarrow$  açúcar apresenta valores de Transfer Entropy mais elevados do que no sentido contrário, açúcar  $\rightarrow$  etanol. Isso indica que o etanol transfere mais informação para o preço do açúcar do que o açúcar transfere para o etanol, em diferentes escalas temporais, conforme mencionado anteriormente.

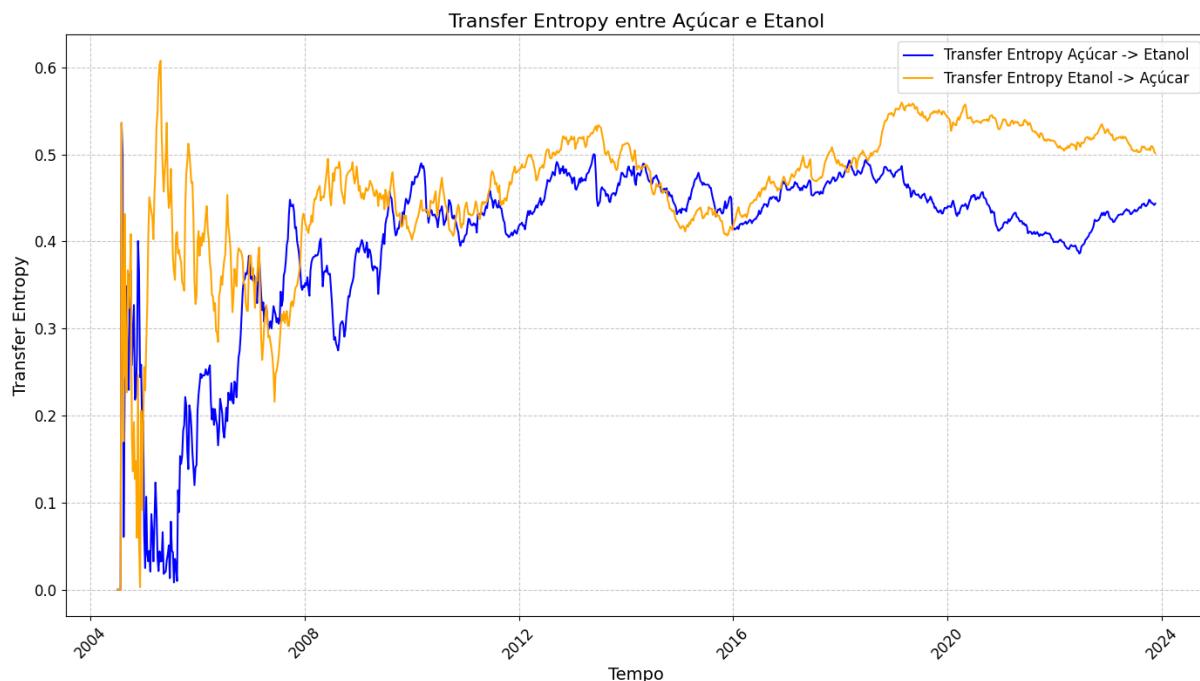


Figura 5.0.5: Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol.

Já a Figura 5.0.6 revela que, na maioria das escalas temporais, o fluxo de informação no sentido gasolina  $\rightarrow$  açúcar apresenta valores de Transfer Entropy superiores ao sentido inverso, açúcar  $\rightarrow$  gasolina. Isso indica que a gasolina transfere mais informação para o preço do açúcar do que o açúcar transfere para o preço da gasolina nas diferentes escalas de tempo.

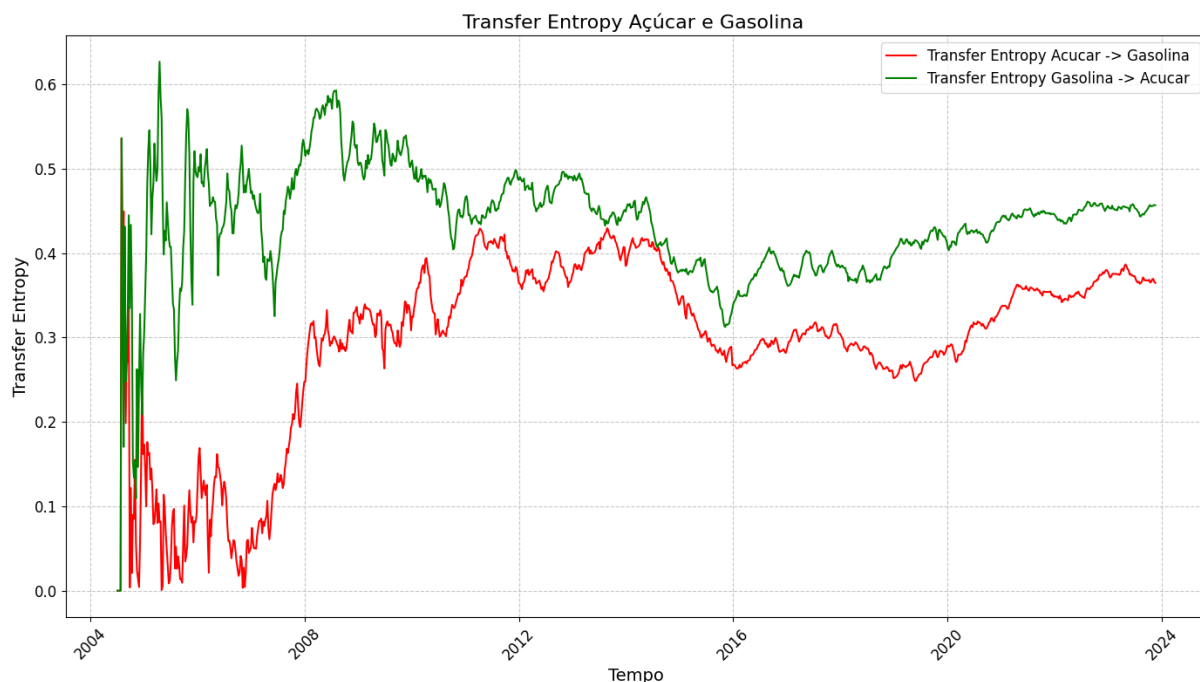


Figura 5.0.6: Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina.

Por fim, a Figura 5.0.7 mostra que, na maioria das escalas temporais, o sentido gasolina  $\rightarrow$  etanol apresenta valores de Transfer Entropy superiores ao sentido inverso, etanol  $\rightarrow$  gasolina. Isso indica uma maior influência da gasolina sobre o etanol, refletindo o fato de que os preços do petróleo e dos combustíveis fósseis continuam a ser fatores decisivos no mercado de biocombustíveis, como o etanol. Dessa forma, qualquer alteração significativa nos preços do petróleo bruto ou nas políticas energéticas que impactem o preço da gasolina também terá repercussões no mercado de etanol.

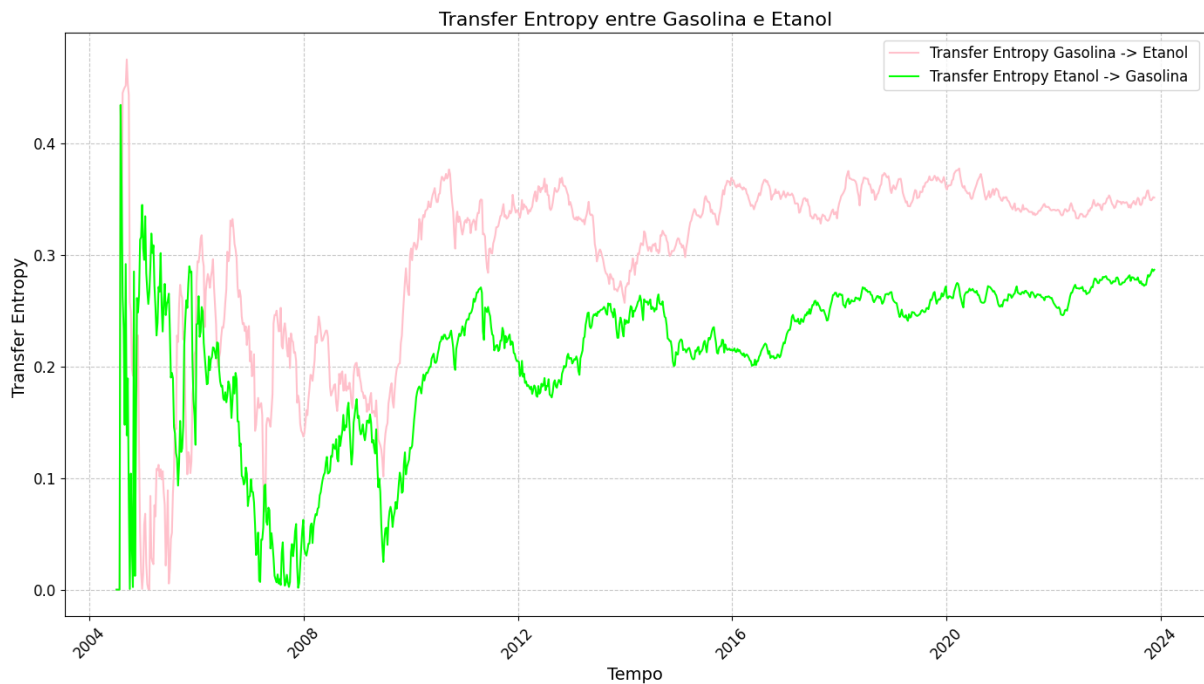


Figura 5.0.7: Transfer Entropy entre as séries da volatilidade de preços de etanol e gasolina.

## 5.1 Transfer Entropy nos períodos de pré-pandemia, durante a pandemia e pós-pandemia de Covid-19

Neste capítulo serão apresentados os resultados da Transfer Entropy em períodos determinados, abordando o impacto da pandemia de Covid-19, que teve um efeito significativo na economia brasileira e, em particular, no mercado de *commodities* como o açúcar, o etanol e a gasolina.

A Tabela 5.1.1 apresenta os períodos utilizados para o estudo: antes, durante e após a pandemia, assim como os tamanhos das séries estudadas: 104, 104 e 100 observações, respectivamente.

| Períodos | Datas                   | Tamanho das séries |
|----------|-------------------------|--------------------|
| Antes    | 07/01/2018 a 29/12/2019 | 104                |
| Durante  | 05/01/2020 a 26/12/2021 | 104                |
| Após     | 02/01/2022 a 26/11/2023 | 100                |

Tabela 5.1.1: Períodos estudados e tamanho das séries.

A seguir, serão apresentadas as análises de cada período mencionado.

### 5.1.1 Pré Pandemia (2018 - 2019)

No período anterior à pandemia, a economia brasileira estava em um processo de lenta recuperação após a recessão de 2015 – 2016 [29]. A demanda interna por gasolina e etanol manteve-se estável, enquanto o mercado de exportação de açúcar brasileiro apresentou uma recuperação gradual devido à demanda global. Durante esse período, os preços globais do açúcar não registraram grandes variações, sendo influenciados pela oferta e demanda mundiais, com o Brasil como um dos principais produtores e exportadores.

No caso do etanol, a demanda impulsionada pelo programa de biocombustíveis RenovaBio [5], manteve os preços estáveis no mercado interno. A mistura obrigatória de etanol na gasolina também contribuiu para estabilizar a produção e a demanda.

Por fim, no caso da gasolina, os preços neste período acompanharam as variações dos preços internacionais do petróleo e a taxa de câmbio do dólar. Embora a Petrobras [13], empresa estatal, influenciasse nos preços, o contexto pré-pandemia apresentou estabilidade na oferta e nos preços dos combustíveis.

A Tabela 5.1.2 apresenta os valores de Transfer Entropy entre cada par de séries, refletindo a quantidade de informação que a série na linha transfere para a série na coluna, no período mencionado.

| →        | Açúcar   | Etanol   | Gasolina |
|----------|----------|----------|----------|
| Açúcar   | -        | 0.213937 | 0.128837 |
| Etanol   | 0.324835 | -        | 0.119801 |
| Gasolina | 0.351413 | 0.030508 | -        |

Tabela 5.1.2: Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para o período pré-pandemia.

Na Figura 5.1.1, mostra a análise de Transfer Entropy entre o açúcar e o etanol durante 2018 e 2019, onde revela flutuações significativas na transferência de informação entre esses mercados, refletindo mudanças em suas dinâmicas de influência mútua. No início de 2018, observa-se um forte fluxo de informação do açúcar para o etanol, atingindo seu pico em maio. No entanto, na segunda metade do ano, essa interação diminui consideravelmente, sugerindo um período de estabilização em ambos os mercados.

Em 2019, a dinâmica muda, os preços do etanol começam a influenciar mais os do açúcar, especialmente nos primeiros meses do ano, devido a fatores como mudanças na demanda por biocombustíveis, políticas energéticas ou variações na produção. No final de 2019, o fluxo de informação se estabiliza novamente, indicando um equilíbrio entre os dois mercados.

Em síntese, a análise mostra como as relações entre açúcar e etanol variam de acordo com o contexto econômico e as políticas energéticas, destacando a importância da Transfer Entropy para identificar influências direcionais e não lineares em sistemas econômicos complexos.

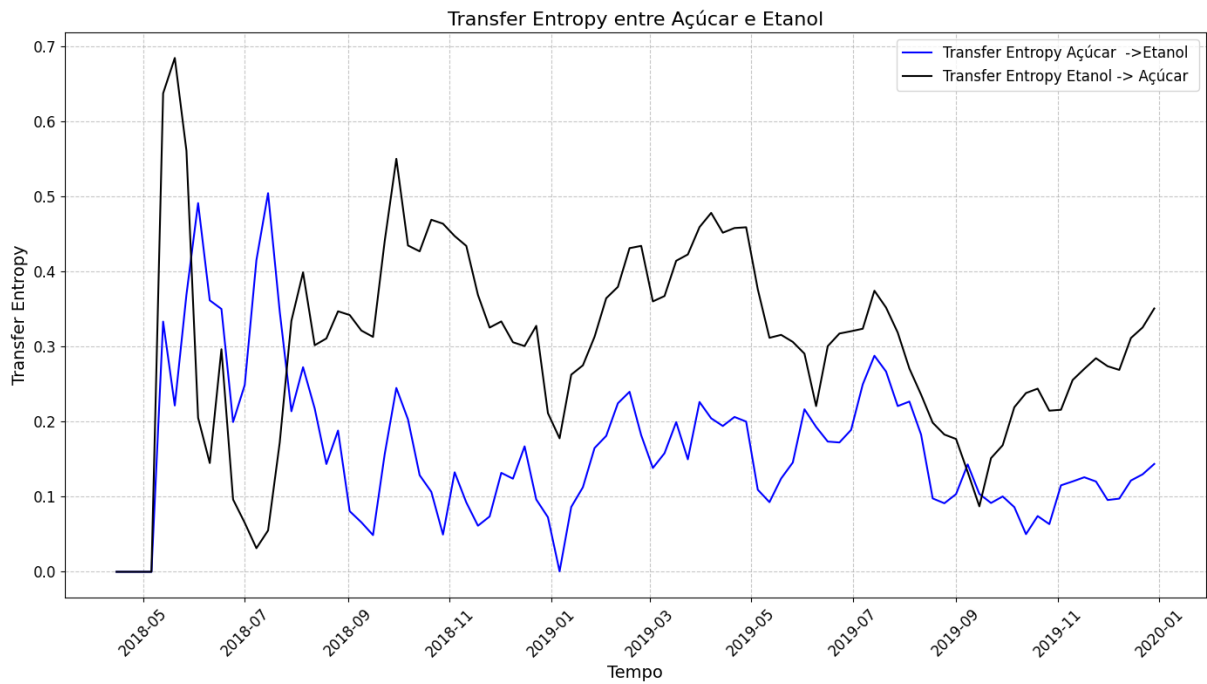


Figura 5.1.1: Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (pré-pandemia).

Na Figura 5.1.2, no início de 2018, observa-se uma transferência significativa de informação da gasolina para o açúcar, indicando que, nesse momento, o preço da gasolina influenciava fortemente o preço do açúcar. Posteriormente, os valores flutuam, mas a Transfer Entropy de gasolina para açúcar se mantém consistentemente maior do que a açúcar para gasolina, sugerindo que o preço da gasolina exerce uma influência mais significativa no preço do açúcar ao longo do período.

A Transfer Entropy de açúcar para gasolina apresenta valores relativamente baixos e menos variáveis, indicando que a influência do açúcar sobre a gasolina é limitada e ocorre em menor intensidade. Entre 01/2019 e 03/2019, nota-se uma redução geral na transferência de informação em ambas as direções, seguida de uma recuperação gradual.

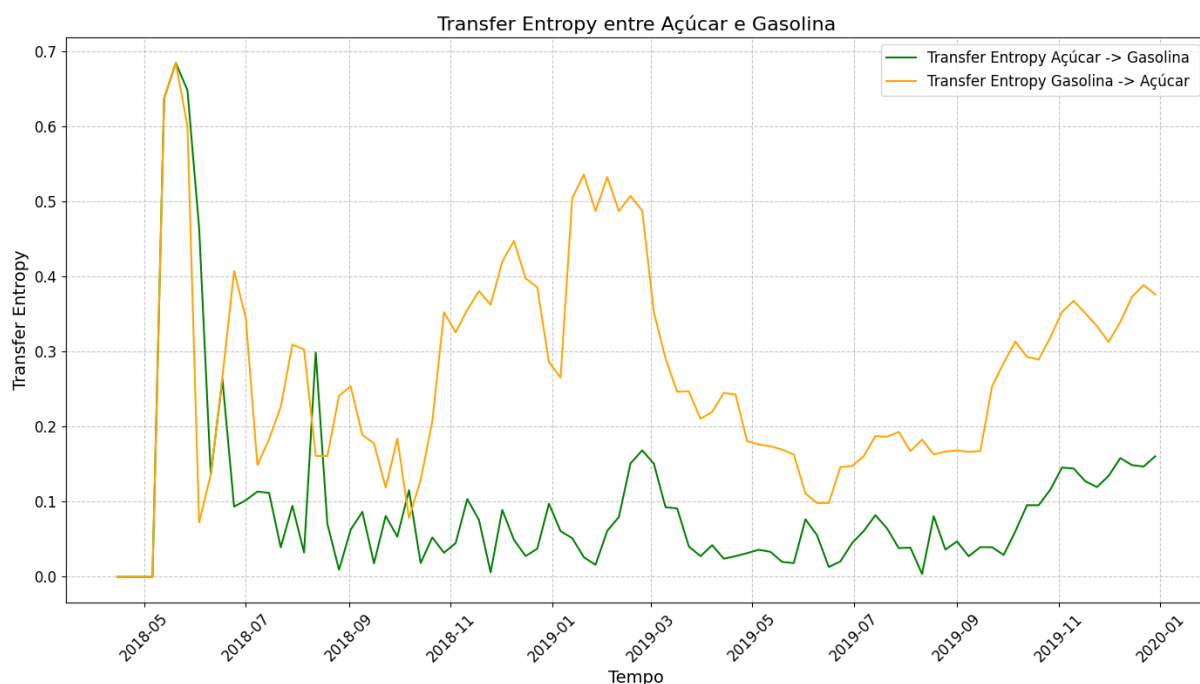


Figura 5.1.2: Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (pré-pandemia).

No início de 2018, o etanol exerce uma forte influência sobre a gasolina (ver Figura 5.1.3), atingindo picos significativos de transferência de informação. Isso reflete a dependência do mercado de gasolina em relação ao etanol como uma alternativa energética, especialmente em regiões com mandatos de mistura de biocombustíveis.

A influência da gasolina sobre o etanol é menos pronunciada, sugerindo que o preço do etanol é mais determinado por outros fatores, como a oferta de açúcar e políticas relacionadas.

Em 2019, a transferência de informação do etanol para a gasolina continua dominante, mas sua intensidade diminui gradualmente, indicando uma menor dependência da gasolina em relação ao mercado de etanol. A direção oposta (gasolina para etanol) apresenta uma leve recuperação, o que pode refletir ajustes na relação de preços devido a fatores macroeconômicos e energéticos globais.

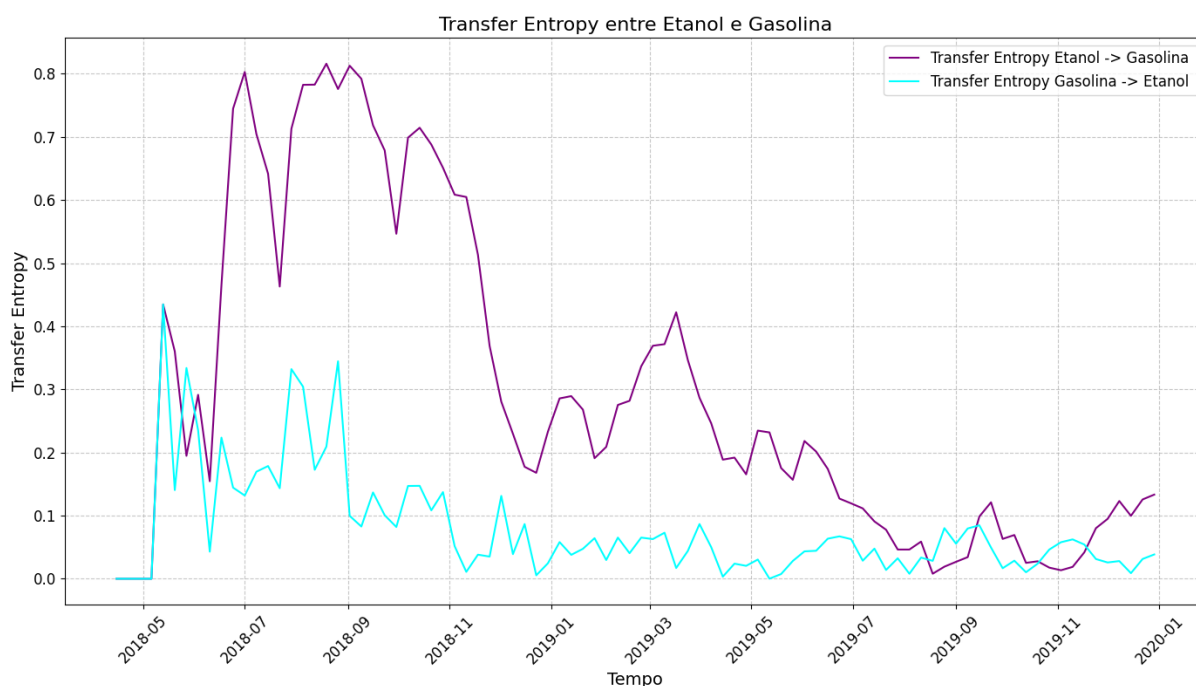


Figura 5.1.3: Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (pré-pandemia).

### 5.1.2 Durante a Pandemia (2020 - 2021)

A pandemia da COVID-19 gerou disruptions significativas nas cadeias de suprimento e reduziu a demanda global de diversas *commodities* devido às restrições sanitárias e ao impacto na mobilidade e na produção [53]. Para o Brasil, os efeitos foram variados entre os mercados de etanol, açúcar e gasolina, afetando diretamente o setor sucroenergético.

No país, a produção de etanol alcançou quase 33 bilhões de litros em 2020, utilizados como combustível único nos veículos *flex fuel* (etanol hidratado) e em mistura com a gasolina (etanol anidro). No entanto, as restrições de mobilidade provocaram uma queda acentuada na demanda de combustíveis, incluindo o etanol. Entre 2 de março e 3 de abril de 2020, o preço do etanol hidratado ao produtor, livre de impostos, caiu de R\$ 2,16 para R\$ 1,28 por litro. Durante o mesmo período, o consumo de etanol hidratado nos postos de revenda caiu 58,4 %, enquanto o consumo de gasolina, que contém 27 % de etanol anidro, caiu 53,8 % [49].

Além disso, a desvalorização do real brasileiro elevou os custos de produção, pressionando ainda mais o setor. Apesar da queda na demanda, parte da cana-de-açúcar foi redirecionada para a produção de etanol, devido à instabilidade nos preços do açúcar no mercado internacional [4]. Com a chegada da pandemia, houve uma redução significativa na demanda global por açúcar, assim, parte da produção de cana foi redirecionada para o etanol, equilibrando parcialmente os impactos. A colheita de cana-de-açúcar, iniciada na safra 2020/2021, enfrentou desafios logísticos, já que o produto não pode ser armazenado nem transportado por longas distâncias e deve ser processado em até, no máximo, 48 horas [12].

Essa característica do setor sucroenergético, que opera em sistema de fluxo, contrastou com a flexibilidade da produção de combustíveis fósseis, como petróleo e gás, que podem ser estocados. Além disso, durante a pandemia, a demanda por gasolina caiu drasticamente, refletindo a redução na mobilidade e no consumo global de combustíveis.

A Petrobras ajustou os preços internos da gasolina para acompanhar a queda nos preços internacionais do petróleo, especialmente no início de 2020 [13]. Esses ajustes, somados à desvalorização do real, provocaram volatilidade nos preços internos da gasolina.

Pelo tanto em um curto período, a crise da Covid-19 gerou uma grande instabilidade a um setor estratégico para o Brasil. A combinação de queda na demanda, flutuações nos preços e desafios logísticos afetou profundamente o setor sucroenergético, exigindo medidas urgentes para evitar uma desestruturação ainda maior desse setor vital para a economia nacional.

A Tabela 5.1.3 apresenta os valores de Transfer Entropy entre cada par de séries, no período mencionado.

| →        | Açúcar   | Etanol   | Gasolina |
|----------|----------|----------|----------|
| Açúcar   | -        | 0.170471 | 0.145073 |
| Etanol   | 0.204847 | -        | 0.119801 |
| Gasolina | 0.196155 | 0.025094 | -        |

Tabela 5.1.3: Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para o período pandêmico.

A Figura 5.1.4 mostra a análise de Transfer Entropy entre o açúcar e o etanol durante 2020 e 2021, revelando flutuações significativas na transferência de informação entre esses mercados, refletindo mudanças em suas dinâmicas de influência mútua. No início de 2020, observa-se um forte fluxo de informação do etanol para o açúcar, isso devido a Covid-19. No entanto, na segunda metade do ano, essa interação diminui consideravelmente, sugerindo um período de estabilização em ambos os mercados.

Em 2021, a dinâmica muda por um curto período de tempo: os preços do açúcar começam a influenciar mais os do etanol, especialmente nos primeiros meses do ano, devido a mudanças na demanda por biocombustíveis. No final de 2021, o fluxo de informação se estabiliza novamente, indicando um equilíbrio entre os dois mercados.

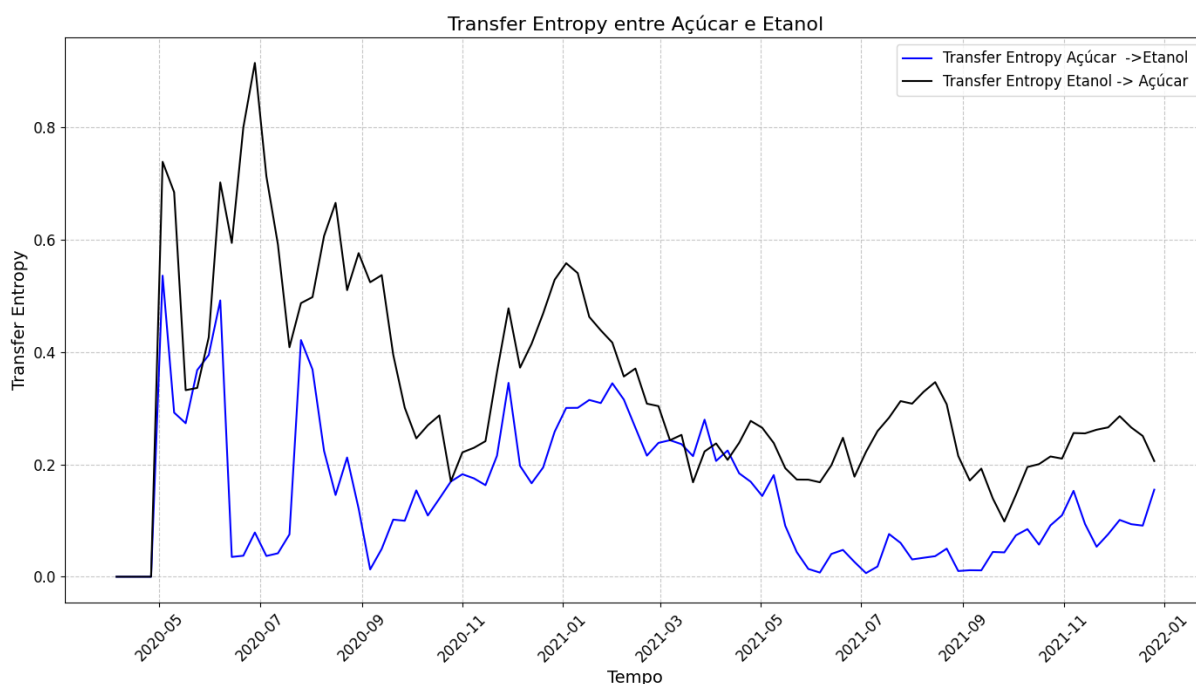


Figura 5.1.4: Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (período pandêmico).

Já na Figura 5.1.5, no início de 2020, apresenta uma transferência significativa de informação da gasolina para o açúcar, com picos notáveis nos primeiros meses. Esse comportamento pode ser atribuído a variações nos preços internacionais do petróleo devido a pandemia, que afetam indiretamente a produção de açúcar também ao uso na produção de etanol.

A final do ano 2020, os preços do açúcar começam a influenciar mais sobre a gasolina, e a dinâmica parece se inverter parcialmente, com maior estabilidade no fluxo de informação do açúcar para a gasolina em certos períodos e nos últimos meses de 2021, os fluxos de informação mostram certa convergência, sugerindo um possível equilíbrio na interação entre esses mercados.

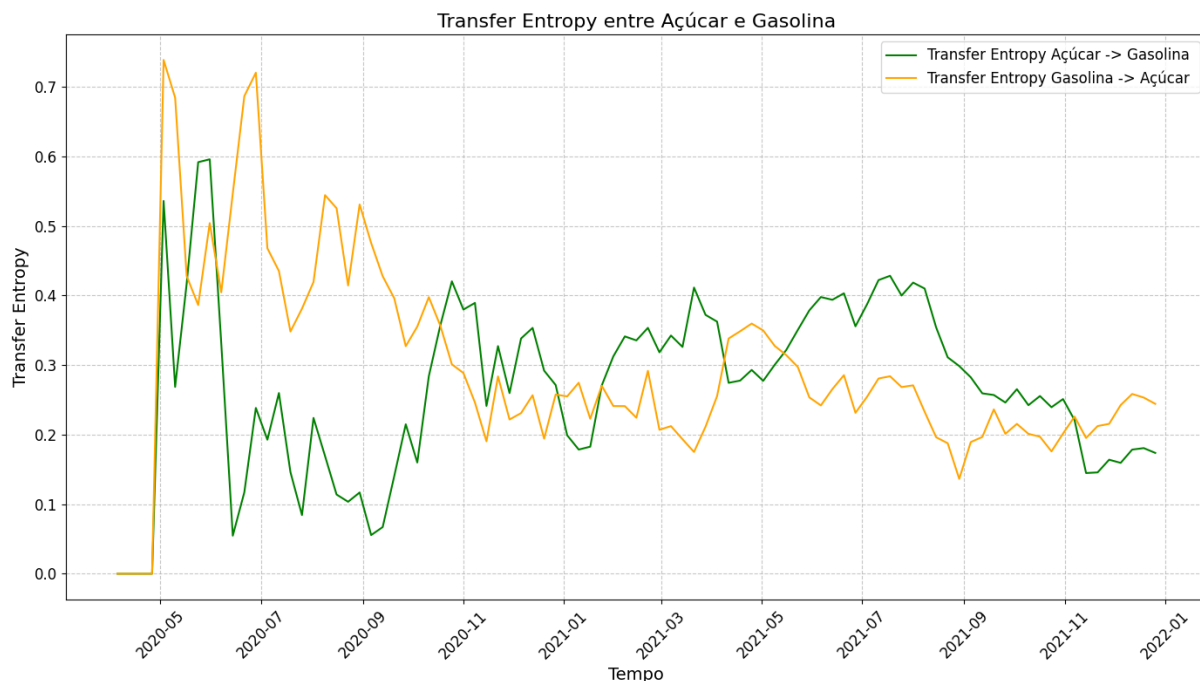


Figura 5.1.5: Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (período pandêmico).

Por fim, nota-se pela Figura 5.1.6, no início de 2020, o etanol exerce uma forte influência sobre a gasolina, atingindo picos significativos de transferência de informação. Isso reflete a dependência do mercado de gasolina em relação ao etanol como uma alternativa energética, especialmente em regiões com mandatos de mistura de biocombustíveis.

A influência da gasolina sobre o etanol é menos pronunciada, sugerindo que o preço do etanol é mais determinado por outros fatores, como a oferta de açúcar e políticas relacionadas.

Em 2021, a transferência de informação do etanol para a gasolina continua dominante, mas sua intensidade diminui gradualmente, indicando uma menor dependência da gasolina em relação ao mercado de etanol. A direção oposta (gasolina para etanol) apresenta uma leve recuperação, o que pode refletir ajustes na relação de preços devido a fatores macroeconômicos e energéticos globais.

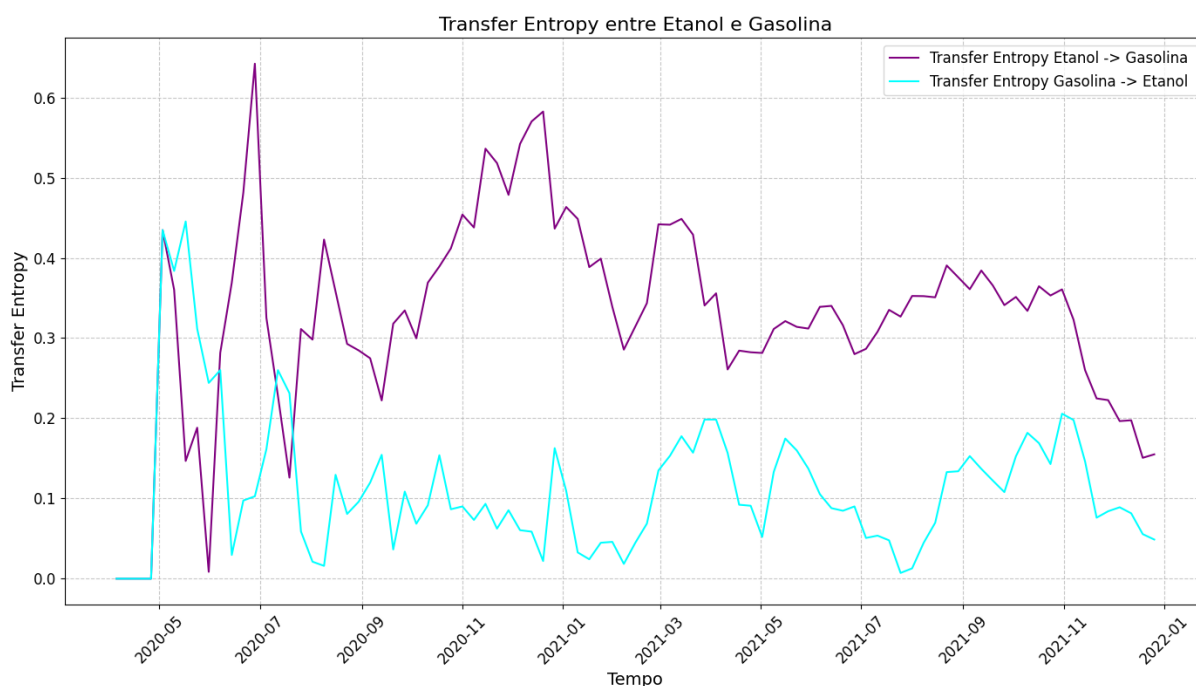


Figura 5.1.6: Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (período pandêmico).

### 5.1.3 Pós Pandemia (2022 - 2023)

Com a recuperação gradual da economia após a pandemia da Covid-19, a demanda por combustíveis e produtos agrícolas começou a se normalizar, apesar de novos desafios emergirem durante este período. Em 2022, o conflito entre Rússia e Ucrânia gerou impactos significativos nos mercados globais, incluindo as *commodities* brasileiras, ao mesmo tempo em que as flutuações no câmbio continuaram afetando a economia nacional [13].

Os preços do açúcar no mercado internacional começaram a se recuperar em 2022, impulsionados pelo aumento da demanda global [67]. A maior produção de açúcar no Brasil foi favorecida pela retomada da demanda externa, à medida que os mercados internacionais voltaram à normalidade. O país manteve sua posição de liderança como um dos maiores exportadores globais de açúcar, beneficiando-se das condições favoráveis do mercado externo e do aumento da produção interna [26].

A reativação econômica e o aumento da mobilidade impulsionaram também a demanda por etanol. O programa RenovaBio [27], que incentiva o uso de biocombustíveis, continuou desempenhando um papel importante na promoção da produção nacional de etanol. Adicionalmente, o Brasil se destacou no cenário internacional com um aumento expressivo nas exportações de etanol. Segundo a Conab, na safra 2022 e 2023, as exportações alcançaram cerca de 2,84 bilhões de litros, o que representou um crescimento de 58% em comparação à safra anterior. Esse aumento foi favorecido por diversos fatores, como:

- **Valorização do dólar**, que tornou as exportações mais competitivas.
- **Questões tributárias** no mercado interno brasileiro, que incentivaram a busca por mercados externos.
- **Redução da oferta internacional de petróleo**, que aumentou a demanda por combustíveis alternativos, especialmente na Europa.

Os preços da gasolina experimentaram aumentos em 2022, principalmente devido aos impactos da guerra na Ucrânia, que elevaram os preços globais do petróleo [24]. No Brasil, a volatilidade dos preços internos foi intensificada pela desvalorização do real frente ao dólar, mas a recuperação da demanda por gasolina foi notável conforme a economia brasileira retomava seu ritmo e a mobilidade voltava aos níveis pré-pandêmicos.

Pelo tanto, os anos de 2022 e 2023 marcaram a recuperação dos mercados de açúcar, etanol e gasolina no Brasil. Apesar das dificuldades impostas por fatores globais, como o conflito na Ucrânia e as flutuações cambiais, o setor agrícola e energético brasileiro mostrou resiliência e adaptabilidade. A combinação de políticas internas, como o programa RenovaBio, e condições externas, como a demanda crescente por combustíveis alternativos, foi essencial para fortalecer o papel do Brasil como líder global nesses mercados.

A Tabela 5.1.4 apresenta os valores de Transfer Entropy entre cada par de séries, no período pós-pandemia.

| →        | Açúcar   | Etanol   | Gasolina |
|----------|----------|----------|----------|
| Açúcar   | -        | 0.190874 | 0.172560 |
| Etanol   | 0.194233 | -        | 0.348543 |
| Gasolina | 0.048862 | 0.318385 | -        |

Tabela 5.1.4: Transfer Entropy para as séries de preços de açúcar, etanol e gasolina para o período pós-pandemia.

Na figura 5.1.7, durante o ano de 2022, observa-se uma notável influência do etanol sobre o açúcar, especialmente no início do ano, quando a Transfer Entropy do etanol para o açúcar atinge picos significativos. Esse comportamento reflete a alta dependência do mercado do açúcar em relação aos preços do etanol, provavelmente impulsionada pelo aumento dos preços internacionais do petróleo após a recuperação econômica pós-pandemia. Esse aumento no preço do petróleo fortaleceu a demanda por etanol como alternativa energética, no contexto de políticas que favoreceram a mistura de biocombustíveis com gasolina. No entanto, na metade e no final do ano 2022, a intensidade dessa relação diminui gradualmente, indicando uma estabilização tanto nos preços do petróleo quanto na oferta agrícola de cana-de-açúcar, após as flutuações iniciais causadas pela pandemia.

Ao longo de 2023, a relação entre os dois mercados muda significativamente, mostrando uma maior influência do açúcar sobre o etanol. A Transfer Entropy do açúcar para o etanol mantém-se mais estável e com maior predominância do que no ano anterior. No meio e no final de 2023, os fluxos de informação entre açúcar e etanol mostram uma convergência, sugerindo um equilíbrio mais consistente na interação entre esses dois mercados.

A variação de preços após a pandemia também é evidente. Durante 2022, as disrupções nas cadeias de suprimento globais e o aumento na demanda por biocombustíveis impulsionaram os preços do etanol. Isso teve um impacto direto no açúcar, já que parte da produção de cana foi direcionada para a fabricação de etanol. Em contraste, 2023 marca uma etapa de estabilização e ajuste, favorecida por uma recuperação completa das cadeias de suprimento e um melhor desempenho nas condições agrícolas, que permitiram uma interação mais equilibrada entre o açúcar e o etanol. Esse comportamento destaca a estreita interdependência entre ambos os mercados, especialmente no Brasil, onde a cana-de-açúcar desempenha uma função dupla como alimento e como fonte de energia renovável.

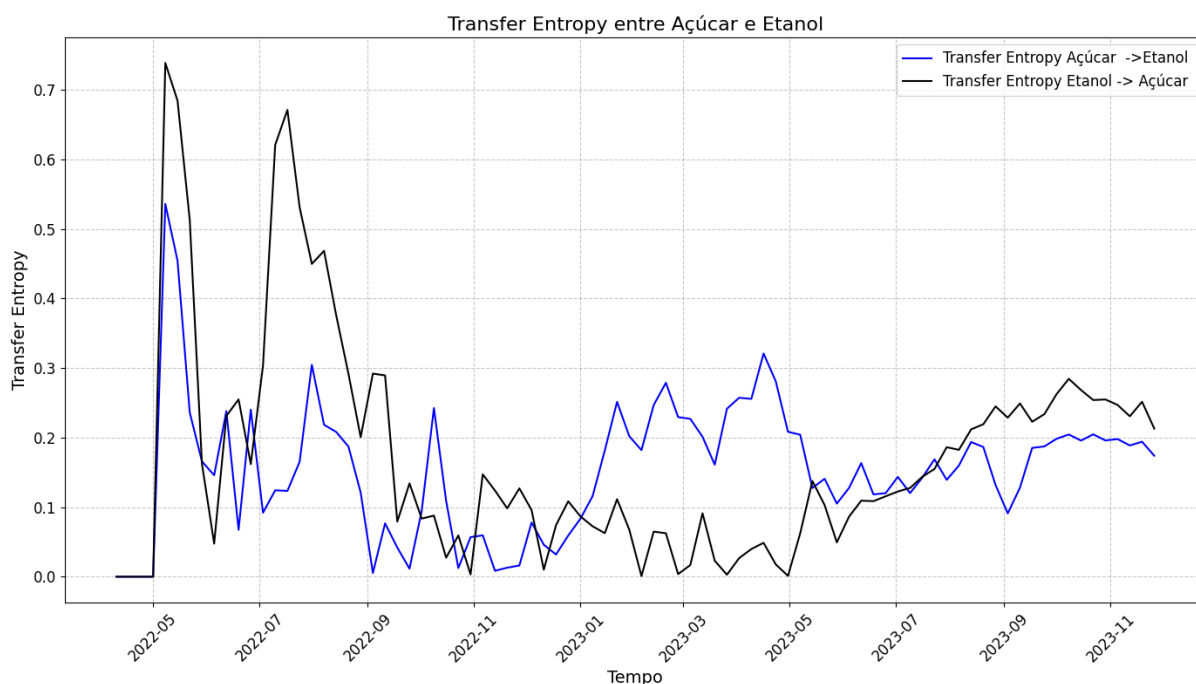


Figura 5.1.7: Transfer Entropy entre as séries da volatilidade de preços do açúcar e etanol (pós-pandemia).

O gráfico 5.1.8, mostra a interação entre os preços do açúcar e da gasolina, através da Transfer Entropy, durante os anos de 2022 e 2023. Essa análise é especialmente relevante no contexto pós-pandemia, onde se observam flutuações significativas em ambas as direções de influência.

Em 2022, a Transfer Entropy da gasolina para o açúcar apresenta picos consideráveis no início do ano. Isso pode ser explicado pelo forte impacto que os preços internacionais do petróleo tiveram nos mercados globais, incluindo o mercado de açúcar, dado que grande parte da produção de cana-de-açúcar é destinada à produção de etanol, um biocombustível alternativo à gasolina. A guerra na Ucrânia e a instabilidade geopolítica afetaram os preços do petróleo, gerando um aumento nos custos energéticos e, consequentemente, uma pressão sobre a produção agrícola e os preços do açúcar. Por volta de meados de 2022, a influência da gasolina diminuiu consideravelmente, o que pode ser atribuído à estabilização dos preços do petróleo e a ajustes nos mercados energéticos e agrícolas.

Por outro lado, a Transfer Entropy do açúcar para a gasolina mostra um comportamento mais consistente ao longo do tempo, com uma influência crescente no final de 2022 e durante 2023. Essa mudança reflete a importância da produção de açúcar na determinação dos preços do etanol, que, por sua vez, afeta os mercados de gasolina. Em 2023, a tendência se inverte parcialmente, com o açúcar exercendo uma influência mais dominante, devido a uma boa safra agrícola no Brasil e à recuperação completa das cadeias de suprimento, o que estabilizou a oferta de açúcar e permitiu um impacto mais direto sobre os biocombustíveis.

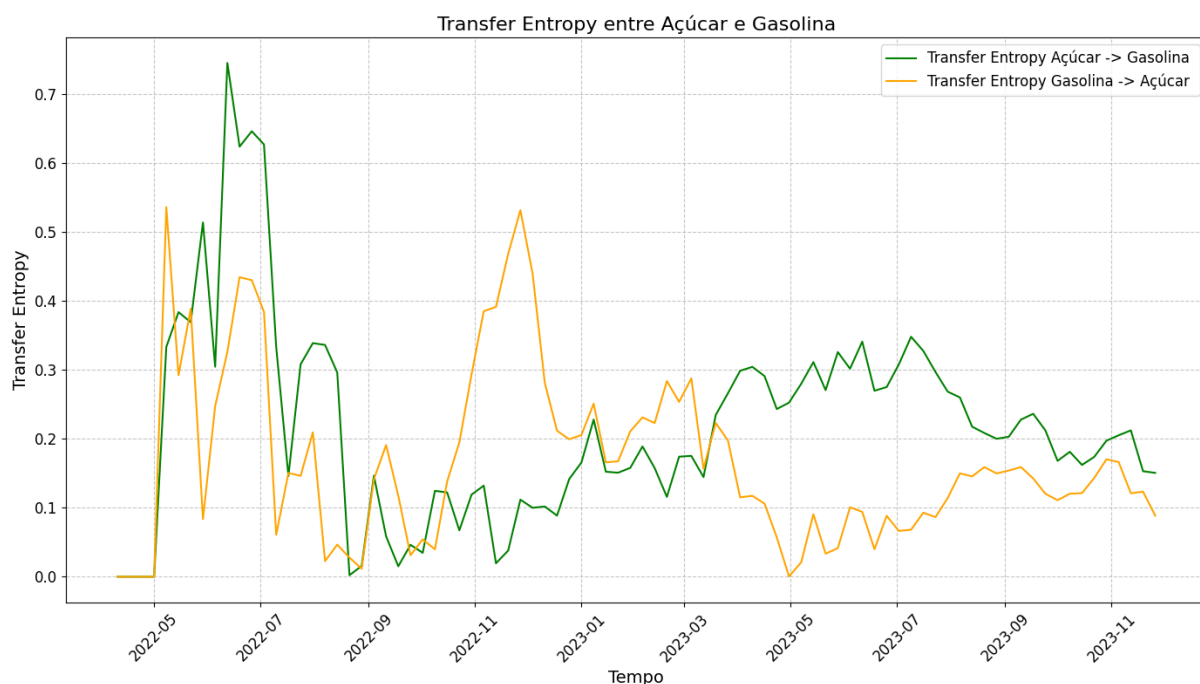


Figura 5.1.8: Transfer Entropy entre as séries da volatilidade de preços do açúcar e gasolina (pós-pandemia).

Na figura 5.1.9, no início do ano de 2022, a influência da gasolina sobre o etanol apresenta picos consideráveis. Entretanto, em meados de 2022, a tendência se inverte parcialmente, com o etanol exercendo uma influência mais dominante. Isso se deve à recuperação pós-pandemia e à normalização total das cadeias de suprimento, o que estabilizou a oferta de açúcar e permitiu um impacto mais direto sobre os biocombustíveis.

Em contraste, 2023 marca uma fase de estabilização e ajuste, favorecida por uma recuperação completa das cadeias de suprimento, o que permitiu uma interação mais equilibrada entre o etanol e a gasolina.

Portanto, o gráfico reflete como os mercados de etanol e gasolina estão intrinsecamente conectados, com flutuações significativas em suas interações durante 2022 e 2023. Esses movimentos foram impulsionados por fatores econômicos globais, políticas energéticas e mudanças na oferta agrícola, tudo dentro do contexto de uma economia global que busca se recuperar e se ajustar após os efeitos da pandemia.

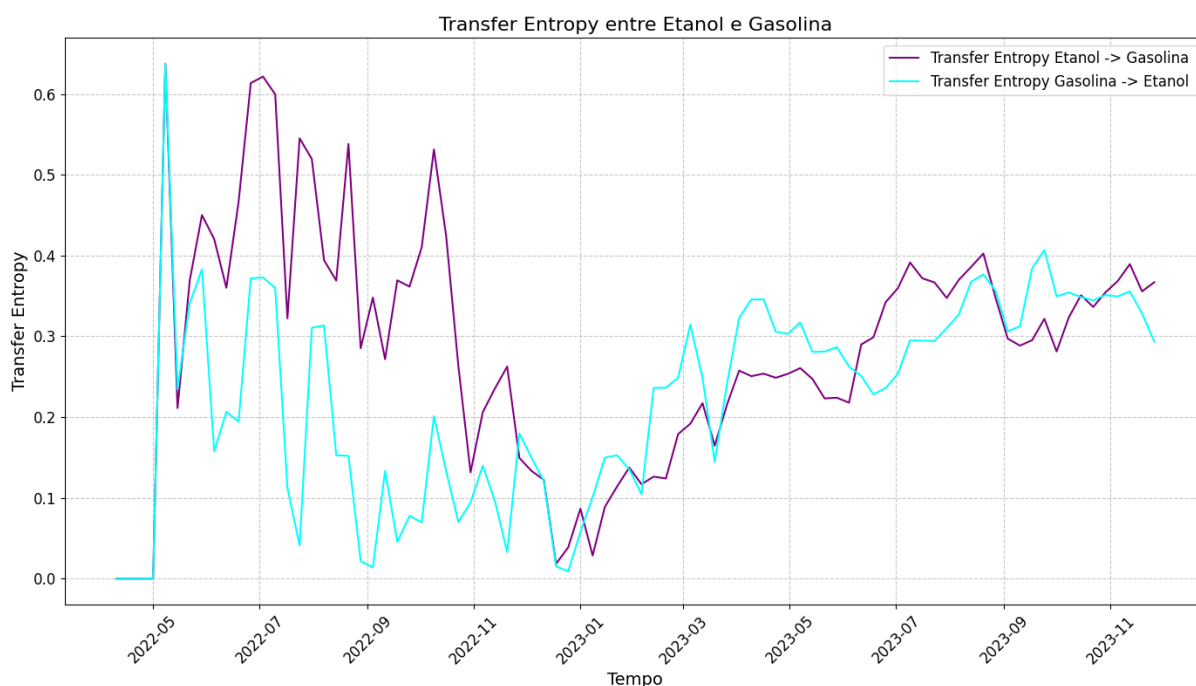


Figura 5.1.9: Transfer Entropy entre as séries da volatilidade de preços do etanol e gasolina (pós-pandemia).

## 5.2 Discussão sobre o Comparativo dos Períodos

O comportamento dos mercados de açúcar, gasolina e etanol durante os períodos antes, durante e depois da pandemia revela padrões distintos influenciados por fatores econômicos, geopolíticos e climáticos. Cada período foi marcado por dinâmicas particulares que destacam a interação entre essas *commodities* e os choques externos que as afetaram.

**Pré Pandemia (2018 - 2019)** Nos anos anteriores à pandemia, os mercados de açúcar, gasolina e etanol apresentaram estabilidade e uma interação moderada. Os gráficos refletem um fluxo de informação constante e previsível entre essas *commodities*. A gasolina influenciava moderadamente os preços do açúcar, especialmente nos momentos em que o preço do petróleo gerava impactos indiretos na produção agrícola destinada aos biocombustíveis. Por outro lado, a influência do açúcar sobre a gasolina era menos expressiva, indicando que as mudanças no mercado agrícola não impactavam significativamente os combustíveis fósseis.

No caso do etanol, as interações com o açúcar foram estáveis, impulsionadas por um equilíbrio na produção de cana-de-açúcar entre seu uso para açúcar refinado e para etanol. A relação com a gasolina era forte, mas sem grandes disrupções, devido à estabilidade nas políticas de mistura de biocombustíveis e à ausência de eventos externos que alterassem significativamente a demanda ou a oferta.

**Durante a pandemia (2020 - 2021)** A pandemia trouxe uma fase de alta volatilidade e disrupções significativas nos mercados. Os gráficos mostram um aumento considerável na Transfer Entropy, refletindo uma maior interdependência e flutuações intensas entre as *commodities*.

Nesse período, os preços da gasolina despencaram inicialmente devido às restrições de mobilidade e à queda na demanda por combustíveis. Esse efeito foi transferido para o mercado de etanol, que depende diretamente do consumo de gasolina, resultando em uma redução da demanda e pressionando os produtores de cana-de-açúcar. A influência

do etanol sobre o açúcar intensificou-se, devido ao desvio de parte da produção de cana para o mercado de biocombustíveis como forma de manter margens de lucro.

Por outro lado, os preços do açúcar enfrentaram alta volatilidade, influenciados por disrupções nas cadeias de suprimento globais e pela desvalorização do real brasileiro, que amplificou os efeitos das mudanças nos preços internacionais. Esse período se destacou pela incerteza e pela forte conexão entre os mercados agrícolas e energéticos, onde a volatilidade de um impactava diretamente os outros.

**Pós Pandemia (2022-2023)** Nos anos pós-pandemia, os gráficos mostram uma estabilização progressiva nas interações entre as *commodities*.

A influência da gasolina sobre o açúcar diminuiu significativamente em comparação aos anos de pandemia, refletindo a recuperação das cadeias de suprimento globais e a estabilização dos preços do petróleo. Em contraste, a influência do açúcar sobre a gasolina aumentou gradualmente, destacando o papel da produção de cana-de-açúcar na determinação dos preços do etanol, que, por sua vez, afeta os mercados de combustíveis fósseis.

No mercado de etanol, os preços recuperaram níveis semelhantes aos anteriores à pandemia, impulsionados pela retomada da mobilidade e pelas políticas que promoveram misturas obrigatórias com gasolina. Os fluxos de informação entre o açúcar e o etanol se estabilizaram, com uma relação mais previsível e um impacto equilibrado entre ambos os mercados.

Além disso, fatores externos, como a guerra na Ucrânia, introduziram novas dinâmicas no mercado energético, provocando aumentos nos preços do petróleo e impactando indiretamente a produção e os custos das *commodities* agrícolas. Apesar disso, uma boa safra agrícola no Brasil e a recuperação econômica global permitiram um ajuste e equilíbrio nesses mercados.

Agora vamos a mostrar as comparações de Transfer Entropy entre os preços de açúcar, gasolina e etanol nos períodos antes, durante e após a pandemia.

| Períodos→                    | Pré Pandemia | Durante a Pandemia | Pós Pandemia |
|------------------------------|--------------|--------------------|--------------|
| <b>TE(Açúcar → Etanol)</b>   | 0.213937     | 0.170471           | 0.190874     |
| <b>TE(Etanol → Açúcar)</b>   | 0.324835     | 0.204847           | 0.194233     |
| <b>TE(Gasolina → Etanol)</b> | 0.351413     | 0.025094           | 0.318385     |
| <b>TE(Etanol → Gasolina)</b> | 0.119801     | 0.119801           | 0.348543     |
| <b>TE(Gasolina → Açúcar)</b> | 0.351413     | 0.196155           | 0.048862     |
| <b>TE(Açúcar → Gasolina)</b> | 0.128837     | 0.145073           | 0.172560     |

Tabela 5.2.1: Quadro comparativo dos períodos das *commodities*.

Pelo tanto, os três períodos apresentam dinâmicas claramente diferenciadas. Antes da pandemia, os mercados eram estáveis e previsíveis, com interações moderadas. Durante a pandemia, a volatilidade e a interdependência entre as *commodities* aumentaram significativamente devido às disrupções globais. Após a pandemia, os mercados começaram a se estabilizar, mas com um equilíbrio renovado e novas dinâmicas impulsionadas por fatores externos, como tensões geopolíticas e políticas energéticas. Essa análise destaca como eventos globais podem transformar profundamente as relações entre os mercados agrícolas e energéticos.

## Conclusões

Este trabalho apresentou um estudo a partir da metodologia de Transfer Entropy, para análise dos preços das seguintes *commodities*: açúcar, etanol e gasolina. A escolha dessas variáveis para a realização das análises se deve ao fato de que o açúcar ocupa uma posição significativa nas exportações brasileiras, enquanto o etanol e a gasolina são amplamente utilizados como combustível no mercado interno, além do etanol ser adicionado à gasolina e ser fruto da mesma matéria-prima do açúcar, a cana-de-açúcar.

A partir dos resultados apresentados, foi possível uma análise aprofundada do comportamento das séries temporais dos preços das *commodities* citadas, utilizando a metodologia de Transfer Entropy no período de maio de 2004 a novembro de 2023, no estado de São Paulo. Essa medida, caracterizada como um parâmetro assimétrico, quantifica a quantidade de informação transferida de uma fonte para um destino, permitindo identificar a direção e intensidade dos fluxos de informação entre as variáveis analisadas. A TE foi aplicada nas séries de preços, retornos (taxas de variação logarítmica dos preços) e volatilidade (magnitude das variações de preços) dessas *commodities*, proporcionando uma visão detalhada das relações dinâmicas entre os mercados agrícolas e energéticos.

Além disso, foi realizada uma análise específica da Transfer Entropy em três períodos distintos: pré-pandemia (2018 – 2019), durante a pandemia (2020 – 2021) e pós-pandemia (2022 – 2023). Esses períodos apresentam dinâmicas claramente diferenciadas, evidenciando como os eventos globais influenciaram as interações entre as *commodities*. No período pré-pandemia, observou-se estabilidade e previsibilidade nos mercados, com interações moderadas entre os preços. Durante a pandemia, as disrupções globais causaram alta volatilidade e um aumento na interdependência entre os mercados, destacando a importância dos choques externos nas cadeias de suprimento e nos preços. Já no período pós-pandemia, os mercados começaram a se estabilizar, mas com um equilíbrio renovado, influenciado por fatores como tensões geopolíticas, políticas energéticas e a recuperação econômica global.

A Transfer Entropy demonstrou-se uma ferramenta eficaz para quantificar o fluxo de informações entre as séries temporais dos preços das *commodities* analisadas, elucidando com clareza a direcionalidade das transferências de informação. Os resultados evidenciaram uma interdependência entre os preços do etanol e da gasolina, com a gasolina exercendo uma influência ligeiramente maior sobre o etanol. Isso é consistente com o papel do etanol como biocombustível alternativo, especialmente em mercados que promovem a mistura de combustíveis fósseis. Além disso, observou-se uma relação significativa entre os preços do etanol e do açúcar, refletindo a estreita conexão entre a produção de etanol e a cana-de-açúcar. A influência moderada da gasolina sobre o açúcar sugere uma

ligação indireta por meio do mercado de biocombustíveis, reforçando a complexidade das interações entre essas *commodities*.

Essa abordagem possibilitou compreender, de forma detalhada, as mudanças estruturais associadas a fatores econômicos, geopolíticos e de mercado, destacando a relevância do uso da Transfer Entropy como ferramenta analítica para estudar as relações dinâmicas entre *commodities* em diferentes contextos temporais e econômicos.

---

## Referências Bibliográficas

- [1] A história da cana-de-açúcar - da antiguidade aos dias atuais. <https://www.udop.com.br/noticia/2003/01/01/a-historia-da-cana-de-acucar-da-antiguidade-aos-dias-atuais.html>, 2003. Acessado em: 30 jan. 2025.
- [2] Volatilidade do mercado e safra recorde no Brasil impactam preços globais de açúcar, 2024. <https://pensaragro.com.br/volatilidade-do-mercado-e-safra-recorde-no-brasil-impactam-precos-globais-de-acucar/>. Acessado em: 30 jan. 2025.
- [3] B. B. Adam and A. K. Anil. Granger causality and transfer entropy are equivalent for gaussian variables. *Phys. Rev. Lett.*, 103:238701, 2009.
- [4] Agência Brasileira de Desenvolvimento Industrial (ABDI). *Resiliência em Cadeias de Valor*. Governo Federal do Brasil, 2024. <https://www.gov.br/fazenda/pt-br/central-de-conteudo/publicacoes/conjuntura-economica/estudos-economicos/2024/relatorio-resiliencia-em-cadeias-de-valor.pdf>. Acessado em: 17 nov. 2024.
- [5] Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). *Renova-Bio: Política Nacional de Biocombustíveis*, 2023. <https://www.gov.br/anp/pt-br/assuntos/renovabio>. Acesso em: 16 jan. 2025.
- [6] K. Alexander, S. Harald, and G. Peter. Estimating mutual information. *Phys. Rev. E*, 69:066138, 2004.
- [7] S. R. Alexandre. Previsões de séries temporais combinando modelos arma e redes neurais artificiais. Dissertação de mestre, 2010.
- [8] M. M. K. Alvaro. Comparação de modelos estatísticos para estimação do intervalo de tempos entre ultrapasses de um limiar de temperatura na cidade de P. Prudente - SP. Tesis de maestría, 2019.
- [9] C. C. André, C. Ademir, and L. C. Antônio. Aplicação do modelo arima para previsão do preço do frango inteiro resfriado no grande atacado do estado de São Paulo. *Anais do Congresso Brasileiro de Custos*, 2006.
- [10] S. M. Andre and B. S. Yony. Uma nota sobre o impacto do preço do açúcar, do etanol e da gasolina na produção do setor sucroalcooleiro. *Revista Brasileira de Economia*, 70, 2016.
- [11] P. M. Andres. Series temporales aplicadas. Universidad Complutense de Madrid, 2022.
- [12] P. R. Antonio. Safra 20/21 termina com crescimento na moagem de cana. União da Indústria da Cana-de-Açúcar, 2021.

- [13] L. M. Arthur, C. O. Brena, M. S. Juan, and T. D. Marcos. Petrobrás, preço, paridade de importação e tributação: Análise de aspectos atuais no mercado brasileiro, 2022.
- [14] A. Q. Barbi. A informação mútua como medida de dependência não linear na estrutura de rede do mercado brasileiro de ações. Master's thesis, Universidade de São Paulo, 2017.
- [15] J. Beirlant, E. J. Dudewicz, L. Györfi, and E.C. Meulen. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 2001.
- [16] R. M. Carneiro. *Commodities, choques externos e crescimento: reflexões sobre a América Latina*. Number 117 in Serie Macroeconomía del Desarrollo. Comisión Económica para América Latina y el Caribe (CEPAL), Santiago de Chile, 2012. Símbolo de la ONU: LC/L.3435.
- [17] C. S. Celso, B. A. João, A. C. Marcia, F. C. Raul, M. F. Daniel, and G. V. Hevellyne. Previsão de produção de cana-de-açúcar e seus derivados, através de arima, no período de 2020 a 2025, na macrorregião centro-sul do Brasil. *Research, Society and Development*, 11:e6211427048, 2022.
- [18] C. S. Celso, B. C. João, F. C. Márcia, A. C. Raul, and M. F. Daniel. Aplicação de inteligência artificial na previsão de produções de açúcar e etanol anidro, no período de 2020 até 2025, nas macrorregiões centro-sul e norte nordeste do Brasil. *Revista Brasileira de Geografia Física*, 15(5), 2022.
- [19] A. Chainarong, Z. Elena, and B. Tanya. Variable-lag granger causality and transfer entropy for time series analysis. *ACM Transactions on Knowledge Discovery from Data*, 15:1–30, 05 2021.
- [20] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 2012.
- [21] L. Damiano and P. Sanjay. A non-parametric k-nearest neighbour entropy estimator. *Physical Review E*, 93(1), 2016.
- [22] W. H. Daniel and D. P. Shawn. Transfer entropy for coupled autoregressive processes. *Entropy [electronic only]*, 3, 2013.
- [23] M. Dayana Carla and S. N. Macedo Matos. Extração de conhecimento através da mineração de dados. *Revista de Engenharia e Tecnológica*, 2010.
- [24] ICEX España Exportación e Inversiones. Informe económico y comercial de Brasil 2023. Technical report, 2023. <https://www.icex.es/content/dam/es/icex/oficinas/022/documentos/2023/03/anexos/informe-economico-comercial-brasil-2023.pdf>.
- [25] F. N. Elis. Cana-de-açúcar: a produção de etanol e seus benefícios. *Monografia/Ciências e Tecnologia de São Paulo – Campus Barretos*, 2017.
- [26] Empresa de Pesquisa Energética. Análise de conjuntura dos biocombustíveis - ano 2022. Technical report, Empresa de Pesquisa Energética, 2023. [https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-756/NT-EPE-DPG-SDB-2023-01\\_Analise\\_de\\_Conjuntura\\_dos\\_Biocombustiveis\\_Ano2022.pdf](https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-756/NT-EPE-DPG-SDB-2023-01_Analise_de_Conjuntura_dos_Biocombustiveis_Ano2022.pdf).

- [27] Brasil Energia. Reflexos do renovabio na produção de biocombustíveis e seus impactos na carteira de gasolina e diesel no Brasil no decênio 2020-2029, 2024. <https://brasilenergia.com.br/cenariospetroleo/reflexos-do-renovabio-na-producao-de-biocombustiveis-e-seus-impactos-na-carreira-de-gasolina-e-diesel-no-brasil-no-decenio-2020-2029>. Acesso em: 12 dez. 2024.
- [28] C. M. Fernando and D. R. Santos. Fatores determinantes para consumo de gasolina no estado de São Paulo: o estudo da elasticidade preço da gasolina. *Revista de Gestão e Secretariado*, 13:2337–2352, 2022.
- [29] F. Filipe. Fundo monetario internacional. <https://www.imf.org/es/News/Articles/2019/07/25/NA072519-Six-Charts-on-Boosting-Growth-in-Brazil>, 2019.
- [30] F. S. Gabrielly. Examinando a influência do mercado da gasolina a sobre os preços do etanol hidratado e açúcar ao produtor no Brasil, 2020.
- [31] M. A. Gauto, M. A. Daniela, C. A. Messias, and C. A. Paulo. *Petróleo e Gás: Princípios de Exploração, Produção e Refino*. Tekne. Bookman Editora, 2016.
- [32] G. H. German, W. Wei, R. Kalle, C. S. Miguel, P. Gordon, and V. Raul. Assessing coupling dynamics from an ensemble of time series. *Entropy*, 17, 2010.
- [33] C. W. J. Granger. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37(3):424–438, 1969.
- [34] S. Harshinder, V. Hnizdo, F. D. Adam, and N. Misra. Nearest neighbor estimates of entropy. *American Journal of Mathematical and Management Sciences*, 23, 2003.
- [35] L. Helmut. *The New Introduction to Multiple Time Series Analysis*. 2005.
- [36] L. M. James. Causality, feedback and directed information. Computer Science, 1990.
- [37] M. A. Jian. Estimating transfer entropy via copula entropy. Computer Science: Machine Learning, 2021.
- [38] H. Jiayi and S. Pengjian. Comparison of transfer entropy methods for financial time series. *Physica A: Statistical Mechanics and its Applications*, 482:772–785, 2017.
- [39] H. Joe. *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1997.
- [40] P. N. José. Transfer entropy networks to rank influence over time with application to the brazilian stock market. Tese de Doutorado, Universidade Federal do Rio de Janeiro, 2023.
- [41] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume 1. Hafner Publishing Company, New York, 1969.
- [42] L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. *Probl. Peredachi Inf.*, 23(2):9–16, 1987.
- [43] Z. Kun, P. Jonas, J. Dominik, and S. Bernhard. Kernel-based conditional independence test and application in causal discovery, 2012.

- [44] C. C. Leidiane, C. O. Regiane, M. C. Marina, F. Ana Laura, and F. G. Ana. Cana-de-açúcar e Álcool combustível: Histórico, sustentabilidade e segurança energética. *Enciclopédia Biosfera*, 9(16), 2013.
- [45] M. R. Leila. *Estudo de séries temporais de preços de petróleo, etanol e açúcar através da Transfer Entropy*. PhD thesis, Universidade Federal Rural de Pernambuco, 2018.
- [46] P. Liam. Estimation of entropy and mutual information. *Neural Computation*, 15:1191–1253, 2003.
- [47] C. Madalena, G. Ary, and P. Chung-Kang. Multiscale entropy of biological signals. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 71:021906, 2005.
- [48] F. G. Marcos. Formação de preços de commodities no Brasil. Dissertação apresentada ao Curso de Pós-Graduação da FGV /EAESP, 2002.
- [49] F. V. Maria. Produção e uso de biocombustíveis no brasil. Mestrado em Economia Rural, 2021.
- [50] C. T. Mariza, A. M. Mario, and A. S. Pery. Transmissão de preços no mercado de cana-de-açúcar entre os estados de são paulo e paraná. *Revista Brasileira de Estudos Regionais e Urbanos*, 17:19–37, 2013.
- [51] P. P. Matheus. Estudo de series temporais de preços das commodities agrícolas através de transfer entropy. *Dissertação de Mestrado, Universidade Federal Rural de Pernambuco*, 2021.
- [52] P. P. Matheus, M. S. Joelma, S. A. Lidiane, and S. Tatijana. Fluxo de informações no mercado brasileiro de alimentos: A influência da crise financeira de 2008. *Research, Society and Development*, 10(7), 2021.
- [53] N. V. Mauricio. Análise de volatilidade e previsão dos preços das principais commodities brasileiras (café, milho e soja) durante a pandemia de covid-19 através de modelos de séries temporais. Dissertação de mestrado, Universidade de Brasília, Brasília, DF, 2023.
- [54] R. B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, 2007.
- [55] M. F. Nice. O processo de transferência da informação. *Ciência da Informação*, 8(2):119–138, 1979.
- [56] F. T. Raeder. Assimetria na transmissão de preços da gasolina no brasil: os efeitos do mandato de etanol e o papel da nova estratégia de precificação da petrobrás. Master’s thesis, Universidade Federal Fluminense, 2020.
- [57] V. Raul, W. Michael, L. Michael, and P. Gordon. Transfer entropy a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, 30(1):45–67, 2011.
- [58] O. Rioul. *This is IT: A Primer on Shannon’s Entropy and Information*, pages 49–86. Springer International Publishing, Cham, 2021.

- [59] F. E. Robert. Modelización de series univariantes: Garch. [https://bookdown.org/victor\\_morales/SeriesdeTiempo/modelizaci%C3%B3n-de-series-univariantes-garch.html](https://bookdown.org/victor_morales/SeriesdeTiempo/modelizaci%C3%B3n-de-series-univariantes-garch.html), 2003. Acesso em: 01 de out. de 2024.
- [60] C. C. Ronquim and M. F. Fonseca. Avanço das áreas de cana-de-açúcar e alterações em áreas de agropecuária no interior paulista. Campinas: Embrapa Territorial, 2018.
- [61] G. R. Santos. *Quarenta anos de etanol em larga escala no Brasil: desafios, crises e perspectivas*. Instituto de Pesquisa Econômica Aplicada (Ipea), Brasil, 2016.
- [62] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [63] G. R. Silva. Cana-de-açúcar na produção de etanol. Trabalho de conclusão de curso, Centro Universitário Internacional Uninter, 2022.
- [64] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.
- [65] A. M. Sklar. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231, 1959.
- [66] L. B. Steven and K. S. Anil. Wiener–granger causality: A well established methodology. *NeuroImage*, 58(2):323–329, 2011.
- [67] SuperVarejo. Em contramão aos demais itens da cesta básica, açúcar tem aumento de preço, dezembro 2024. <https://www.supervarejo.com.br/economia/em-contramao-aos-demais-itens-da-cesta-basica-acucar-tem-aumento-de-preco>. Acesso em: 12 dez. 2024.
- [68] R. J. Teweles and E. S. Bradley. *The Stock Market*. Wiley Investment. Wiley, 1998.
- [69] M. C. Thomas and A. T. Joy. *Entropy, Relative Entropy, and Mutual Information*, chapter 2, pages 13–55. John Wiley & Sons, Inc, 2005.
- [70] S. Thomas. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, 2000.
- [71] F. Vieira, E. R. Jose, and F. Albert. Agricultura e indústria no Brasil : inovação e competitividade, 2017. <https://repositorio.ipea.gov.br/handle/11058/7682>.
- [72] Redação Warren. Commodities. Online; acesso em 5 out. 2024. <https://warren.com.br/magazine/commodities/>.