



CENTRO UNIVERSITÁRIO SENAI CIMATEC

**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL**

Mestrado em Modelagem Computacional e Tecnologia Industrial

Dissertação de Mestrado

Etanol no Brasil: uma abordagem preditiva de preços.

Apresentada por: Fabrício Bisset Silva de Brito
Orientador: Prof. Dr. Roberto Luiz de Souza Monteiro

02 de maio de 2019

Fabício Bisset Silva de Brito

Etanol no Brasil: uma abordagem preditiva de preços.

Dissertação de Mestrado apresentado ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Mestrado em Modelagem Computacional e Tecnologia Industrial do Centro Universitário SENAI CIMATEC, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Sistemas Complexos

Orientador: Prof. Dr. Roberto Luiz de Souza Monteiro
Centro Universitário SENAI CIMATEC

Salvador
Centro Universitário SENAI CIMATEC
2019

Ficha catalográfica elaborada pela Biblioteca do Centro Universitário SENAI CIMATEC

B862e Brito, Fabrício Bisset Silva de

Etanol no Brasil: uma abordagem preditiva de preços / Fabrício Bisset Silva de Brito. – Salvador, 2019.

64 f. : il. color.

Orientador: Prof. Dr. Roberto Luiz de Souza Monteiro.

Dissertação (Mestrado em Modelagem Computacional e Tecnologia Industrial) – Programa de Pós-Graduação, Centro Universitário SENAI CIMATEC, Salvador, 2019.
Inclui referências.

1. Commodities energéticas. 2. Etanol. 3. Séries temporais. 4. Previsão de séries temporais. I. Centro Universitário SENAI CIMATEC. II. Monteiro, Roberto Luiz de Souza. III. Título.

CDD: 338.4

Centro Universitário SENAI CIMATEC

Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial

Mestrado em Modelagem Computacional e Tecnologia Industrial

A Banca Examinadora, constituída pelos professores abaixo listados, leram e recomendam a aprovação do Dissertação de Mestrado, intitulada "Etanol no Brasil: uma abordagem preditiva de preços.", apresentada no dia 23 de abril de 2019, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Orientador:

Prof.º. Prof. Dr. Roberto Luiz de Souza Monteiro
Centro Universitário SENAI CIMATEC

Membro externo da Banca:

Prof. Dr. José Roberto de Araújo Fontoura
Universidade do Estado da Bahia - UNEB

Membro interno da Banca:

Prof.^a. Dr. Ingrid Winkler
Centro Universitário SENAI CIMATEC

Agradecimentos

Agradeço pelo teto na infância, pela família e pela educação recebida, tudo muito humildade, porém com bastante dignidade - a Yalorixá Valdete Maria Ferreira de Brito (*in memoriam*) e a Manoel Gregório de Brito, meus avós paternos. Pai e mãe, de acordo com a sabedoria popular, são aqueles que criam.

Camila: minha esposa e companheira. Muito obrigado pelo apoio e presença em momentos importantes da minha vida.

Meu respeito e obrigado ao meu orientador, o Professor Dr. Roberto Monteiro, sempre tranquilo e paciente com seus orientandos. Eu espero continuar meus trabalhos acadêmicos com o Senhor e sua equipe.

Meu igualmente sincero agradecimento, óbvio, aos funcionários da Secretaria Acadêmica e da Coordenação do programa de Pós-graduação e Modelagem Computacional e Tecnologia Industrial do SENAI / CIMATEC, em especial a Julie e Urânia, pela presteza, educação e gentileza.

Salvador, Brasil
dia de 02 de maio de 2019

Fabício Bisset Silva de Brito

Resumo

No contexto econômico contemporâneo - de transações comerciais eletrônicas e inúmeras bolsas de valores e de mercadorias conectadas ao redor do mundo - o Brasil em 2017 posicionou-se como o maior produtor de etanol combustível do mundo.

Diante do cenário - dinâmico - da economia global das últimas décadas e da relevância do Brasil na produção mundial da citada commodity, o presente trabalho se propõe a responder qual modelo matemático apresenta melhor acurácia na previsão de preço da commodity etanol combustível no Brasil para o dia seguinte (próximo ponto) em uma série temporal.

Pesquisas relacionadas a modelos preditivos de produção e de preços de commodities, o que inclui o etanol combustível no Brasil, possuem relevância: (i) para os agricultores, que necessitam tomar decisões mercadológicas e de produção; (ii) para a indústria, que necessita tomar decisões de estoque pertinente a cadeia produtiva; e (iii) para os governos de diversos países, que necessitam tomar decisões visando proteger a economia doméstica e proporcionar a segurança alimentar.

Pesquisas realizadas, até o presente momento, trabalharam com o preço de outras commodities que não o etanol combustível no Brasil ou, quando trabalham com os preços do etanol combustível, trabalham com outro intervalo de tempo que não seja o diário, como é o caso da pesquisa realizada por (DAVID *et al.*, 2018). Dessa forma, não foram observadas na revisão da literatura – até o momento - estudos científicos relacionados a previsão de preços da commodity etanol combustível em série temporal, para o dia seguinte.

Em termos gerais, o presente trabalho - através de experimento de aplicação – compara as medidas de acurácia dos modelos matemáticos ARIMA, ARFIMA e Exponencial Suavizado, em série temporal de preços da commodity etanol combustível no Brasil, em uma previsão para o dia seguinte (próximo ponto); com o objetivo de selecionar o modelo mais preciso. No experimento proposto, o modelo ARIMA (1,1,1) apresenta vantagem comparativa em relação aos demais modelos, quanto a acurácia preditiva de preço para o dia seguinte da commodity etanol combustível, no período de 25 de janeiro de 2010 a 04 de fevereiro de 2014.

A vantagem comparativa do modelo ARIMA (1,1,1) resultada do experimento realizado, portanto, é insuficiente para esgotar a questão proposta para a pesquisa, sendo necessário a introdução de novos modelos matemáticos - em novos experimentos - na comparação da acurácia preditiva, considerando o amplo universo de modelos matemáticos preditivos existentes e observados na revisão da literatura deste trabalho.

Palavras-chave: Commodities; Etanol; Séries Temporais; Previsão.

Abstract

In the modern economic context - electronic trading commodities and various connected stock exchanges and commodities around the world - Brazil in 2017 positioned itself as the world's largest producer of fuel ethanol.

Given the scenario - dynamic - global economy of the last decades and the relevance of Brazil in the world production of commodities cited, the present work applies to a mathematical model that better responds to the forecast price of fuel ethanol commodities in Brazil for the following day (next point) in a time series.

Research related to predictive models of production and commodity prices, including fuel ethanol in Brazil, including: (i) for farmers, who make marketing and production decisions; (ii) for industry, to make inventory decisions pertinent to the production chain; and (iii) to the governments of various countries, which adopt control measures to protect the domestic economy and provide food security.

Researches conducted so far work on the price of other non-fuel ethanol commodities in Brazil or, when using with fuel ethanol prices, use a time interval other than daily, such as research by name david2018fractional. Thus, no scientific studies related to the forecast of commodity ethanol fuel prices in time series for the following day were observed in the literature review.

In general terms, the present work - through application experimentation - compares with the precision measures of the mathematical models ARIMA, ARFIMA and Smoothed Exponential, in the time series of the fuel ethanol commodity prices in Brazil, in a forecast for the day. next (hereinafter point); in order to select the most accurate model. No proposed experiment, or ARIMA model (1,1,1), has a comparative advantage over the other models, as it costs the premeditated price for the next day fuel ethanol fuel, from January 25, 2010 to February 4, 2014.

The comparative advantage of the ARIMA model (1,1,1) resulting from the experiment performed, therefore, is insufficient to exhaust the proposed research question, requiring the introduction of new mathematical models - in new experiments - in the comparison of predictive accuracy. considering the wide universe of existing predictive mathematical models and observed in the literature review of this work.

Keywords: Commodities; Ethanol; Temporal Series; Forecast.

Sumário

| | | |
|----------|--|-----------|
| 1 | Introdução | 1 |
| 1.1 | Definição do problema | 4 |
| 1.2 | Objetivo | 4 |
| 1.3 | Importância da pesquisa | 4 |
| 1.4 | Motivação | 4 |
| 1.5 | Limites e limitações | 5 |
| 1.6 | Organização da Dissertação de Mestrado | 5 |
| 2 | Fundamentação Teórica - Commodities | 7 |
| 2.1 | Commodities: conceito | 7 |
| 2.2 | Commodities no contexto econômico contemporâneo | 7 |
| 2.3 | Contextualização da <i>commodity</i> etanol no Brasil contemporâneo | 9 |
| 2.4 | Commodities e Complexidade | 13 |
| 3 | Fundamentação Teórica - Séries Temporais | 17 |
| 3.1 | Séries Temporais: conceitos fundamentais | 17 |
| 3.2 | Modelos para séries temporais | 26 |
| 3.2.1 | Modelo ARIMA | 28 |
| 3.2.2 | Modelo SARIMA | 30 |
| 3.2.3 | Modelo ARFIMA | 31 |
| 3.2.4 | Modelos de Suavização Exponencial | 31 |
| 3.2.4.1 | Modelo de Suavização Exponencial Simples | 31 |
| 3.2.4.2 | Modelo de Suavização Exponencial Dupla (Método de Holt) | 32 |
| 3.2.4.3 | Modelo Holt-Winters (Suavização Exponencial Tripla) | 33 |
| 3.3 | Medidas de acurácia aplicadas a série temporal dos preços do etanol brasileiro | 36 |
| 4 | Base de Dados e Metodologia | 39 |
| 4.1 | Base de Dados | 39 |
| 4.2 | Metodologia | 39 |
| 4.2.1 | Aplicação do Modelo ARIMA | 39 |
| 4.2.2 | Aplicação do Modelo ARFIMA | 40 |
| 4.2.3 | Aplicação do Modelo Exponencial Suavizado | 40 |
| 5 | Resultados | 42 |
| 5.1 | Resultado das Previsões por Modelo Matemático | 42 |
| 5.2 | Acurácia Preditiva dos Modelos Aplicados no Experimento. | 42 |
| 6 | Conclusões e Perspectivas | 43 |
| | Referências | 49 |

Lista de Tabelas

- 5.1 Resultado das previsões do preço do etanol para ao dia seguinte por modelo matemático em Dólar Americano (US\$). Fonte: Autor. 42
- 5.2 Acurácia Preditiva dos Modelos Aplicados no Experimento. Fonte: Autor. 42
- 6.1 Tabela Comparativa das Medidas de Acurácia por Modelo Preditivo. Fonte: Autor. 43

Lista de Figuras

| | | |
|-----|--|----|
| 2.1 | Preços nominais indexados e previsões anuais de commodities. Fonte:BM . | 7 |
| 2.2 | Produção mundial do etanol no ano 2017. Fonte: RFA | 11 |
| 2.3 | Licenciamento total de automóveis e comerciais leves por combustível no ano 2017. Fonte: ANFAVEA | 12 |
| 3.1 | Preço diário da commodity etanol no Brasil. Fonte: CEPEA | 18 |
| 3.2 | Consumo anual de energia elétrica no Brasil. Fonte: MME | 18 |
| 3.3 | Preços diários das ações da Petrobrás (Petr4) na bolsa de valores B3. Fonte: B3 | 19 |
| 3.4 | Representação de um sistema dinâmico: Fonte: Morettin; Tolo (2004). . . | 19 |
| 3.5 | Série não-estacionária quanto ao nível e inclinação. Fonte: Morettin; Tolo (2004) | 21 |
| 3.6 | Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: Morettin; Tolo, (2004) | 23 |
| 3.7 | Um processo estocástico interpretado como uma família de trajetórias.Fonte: Morettin; Tolo (2004). | 24 |
| 3.8 | Processo estocástico e Série temporal. Fonte: Souza; Camargo (2004) . . . | 25 |
| 3.9 | Observações de uma série temporal com previsões de origem t e horizonte h . Fonte: Morettin; Tolo (2004). | 27 |
| 4.1 | Metodologia aplicada ao experimento. Fonte: Autor. | 41 |
| 6.1 | Comparaçãõ da medida de acurácia ME entre os modelos matemáticos. Fonte: Autor. | 44 |
| 6.2 | Comparaçãõ da medida de acurácia RMSE entre os modelos matemáticos. Fonte: Autor. | 44 |
| 6.3 | Comparaçãõ da medida de acurácia MAE entre os modelos matemáticos. Fonte: Autor. | 44 |
| 6.4 | Comparaçãõ da medida de acurácia MPE entre os modelos matemáticos. Fonte: Autor. | 45 |
| 6.5 | Comparaçãõ da medida de acurácia MAPE entre os modelos matemáticos. Fonte: Autor. | 45 |
| 6.6 | Comparaçãõ da medida de acurácia MASE entre os modelos matemáticos. Fonte: Autor. | 45 |

Lista de Siglas

| | |
|-------------|--|
| ADCC | Correlação Condicional Dinâmica Assimétrica |
| ANFAVEA .. | Associação Nacional dos Fabricantes de Veículos Automotores |
| AFRIMA ... | Média Móvel Integrada Fracionada |
| AR | Auto-regressivo |
| ARDL | Atraso Distribuído Auto-regressivo |
| ARIMA | Auto-regressivo Integrado de Média Móvel |
| ARMA | Auto-regressivo de Média Móvel |
| AST | Análise de Série Temporal |
| B3 | Brasil Bolsa Balcão |
| BM | Banco Mundial |
| BOVESPA .. | Bolsa de Valores de São Paulo |
| CEPEA C... | Centro de Estudos Avançados em Economia Aplicada |
| DCCA | Análise de Correlação Cruzada Destendenciada |
| DFA | Análise de Flutuação Destendenciada |
| ECM | Correção de Erros |
| EU | União Européia |
| EUA | Estados Unidos da América |
| GTAP | Global Trade Analysis Project Framework |
| IDE | Ambiente de Desenvolvimento Integrado |
| KF | Filtro de Kalman |
| MA | Média Móvel |
| ML | Machine Learning |
| MME | Ministério de Minas e Energia - República Federativa do Brasil |
| MSM-POT .. | Comutação de Limite de Pico Multifractal Markoviana |
| MSVR | Regressão Vetorial de Suporte Multi-saída |
| PETRA | Ação da PETROBRAS S/A - Tipo Preferencial (PN) |
| RFA | Renewable Fuels Associations |
| RNA | Redes Neurais Artificiais |
| SARIMA ... | Auto-regressivo de Média Móvel Integrado Sazonal |
| ST | Série Temporal |
| TOP | Caminho Ótico Térmico |
| USP | Universidade de São Paulo |
| VECM | Modelo Vetorial de Correção de Erros |

Introdução

A partir da década de 70, no cenário econômico global, ocorreram mudanças significativas de caráter negocial (i.e. abertura de novas bolsas de valores, de mercadorias e de contratos e futuros no mundo) e de caráter tecnológico (i.e. automação das transações comerciais) nos contratos de futuros e a termo das commodities (i.e. cobre, soja, tribo, ouro, dentre outras), dinamizando as suas trocas - de forma substancial - principalmente através da mediação feita pela interface eletrônica ([JARROW; CHATTERJEA, 2013](#)).

Dentro do contexto econômico contemporâneo - de transações comerciais eletrônicas e inúmeras bolsas de valores e de mercadorias conectadas ao redor do mundo - o Brasil em 2017 posicionou-se como o 2º maior produtor de etanol combustível do mundo ([ASSOCIATION *et al.*, 2018](#)). Diante desse cenário - dinâmico - da economia global das últimas décadas e da relevância do Brasil na produção mundial da citada commodity, o presente trabalho se propõe a responder qual modelo matemático apresenta melhor acurácia na previsão de preço da commodity etanol combustível no Brasil para o dia seguinte (próximo ponto) em uma série temporal.

Em experimento de aplicação, ao tempo em que compara a acurácia preditiva entre os modelos matemáticos ARIMA e ARFIMA em uma série temporal de preços do etanol combustível no Brasil, em uma previsão para o período de 365 dias (previsão de um ano a frente), ([DAVID *et al.*, 2018](#)) obteve vantagem quantitativa para o modelo ARFIMA no que diz respeito a acurácia preditiva.

O experimento feito por ([DAVID *et al.*, 2018](#)) é único na literatura específica - até o presente momento - no que se refere a previsão de preços da commodity etanol combustível no Brasil, conforme revisão bibliográfica realizada. Entretanto, o citado trabalho não atende a necessidade da comparação da acurácia dos modelos matemáticos na previsão de preços para o dia seguinte, considerando a distinção entre o intervalos de tempo da série utilizada (anual) e a necessidade de se descobrir o melhor modelo matemático para uma previsão de preços para o dia seguinte (diária).

O aspecto temporal (passo de um ano a frente na previsão) da série temporal de preços do etanol combustível no Brasil, utilizada por ([DAVID *et al.*, 2018](#)), não atende ao interesse

de parte dos agentes econômicos envolvidos nas transações diárias de compra e venda de commodities, no que diz respeito a previsão de preços: as negociações das commodities em preço à vista são realizadas - em regra - em pregões diários e não em pregões anuais.

Outro aspecto a se considerar, a partir do experimento de (DAVID *et al.*, 2018), é a possibilidade de incremento de modelos matemáticos na verificação da precisão, observando que uma das propostas do trabalho elaborado pelo citado autor é comparar qual o modelo matemático apresenta a melhor acurácia preditiva e que o mesmo trabalhou, somente, com dois modelos matemáticos (o ARIMA e o ARFIMA) para tal objetivo. Assim, o universo de modelos matemáticos aplicados no trabalho realizado por (DAVID *et al.*, 2018) pode ser substancialmente ampliado com o desenvolvimento de novos experimentos ao longo do tempo.

Além do trabalho desenvolvido por (DAVID *et al.*, 2018), a etapa da revisão da literatura desta pesquisa não encontrou outra produção científica que abordasse a previsão de preços da commodity etanol combustível no Brasil em séries temporais. Entretanto, a revisão da literatura coletou uma farta produção científica - em estado da arte - que aborda séries temporais de preços de outras commodities, enquanto objeto científico, numa perspectiva dos Sistemas Complexos.

Pesquisas relacionadas a modelos preditivos de produção e de preços de commodities, o que inclui o etanol combustível no Brasil, possui relevância:

- (i) para os agricultores, que necessitam tomar decisões mercadológicas e de produção;
- (ii) para a indústria, que necessita tomar decisões de estoque pertinente a cadeia produtiva; e
- (iii) para os governos de diversos países, que necessitam tomar decisões visando proteger a economia doméstica e proporcionar a segurança alimentar.

Considerando a importância das pesquisas relacionadas a acurácia dos modelos preditivos de preços de commodities, este trabalho - através de experimento de aplicação - compara a acurácia preditiva dos modelos matemáticos ARIMA, ARFIMA e Exponencial Suavi-

zado, em uma série temporal de preços da commodity etanol combustível no Brasil, em uma previsão para o dia seguinte (próximo ponto). Para tanto, utiliza-se do experimento feito por (DAVID *et al.*, 2018) adaptando-o a previsão diária, ao tempo em que incluiu-se mais um modelo matemático ao estudo comparativo de referência.

Cabe ressaltar, em tempo, que a escolha do modelo Exponencial Suavizado, para ampliar a quantidade de modelos matemáticos além da utilizada no experimento feito por (DAVID *et al.*, 2018), se deu pelo fato do citado modelo já ter sido utilizado por (LAGO; RIDDER; SCHUTTER, 2018) em trabalho científico na previsão de preços de energia elétrica em séries temporais ao tempo em que compara a acurácia preditiva em meio a 27 (vinte e sete) modelos preditivos.

O estudo desenvolvido por (LAGO; RIDDER; SCHUTTER, 2018), portanto, apresenta-se em meio a literatura científica específica, inclusive, como referencial para trabalhos futuros na âmbito da comparação da acurácia entre modelos preditivos em séries temporais de preços da commodity etanol combustível no Brasil na medida em que compara modelos matemáticos tradicionais (como o ARIMA) com modelos preditivos mais recentes (como os baseados em ML ou com modelos híbridos, que agregam características tradicionais e de ML ao mesmo tempo, dentre outros). Assim, observa-se que ainda há um longo caminho pela frente, no que diz respeito a comparação das ferramentas metodológicas relacionadas ao objeto da pesquisa em lide.

Para a realização do experimento proposto, convém esclarecer, este trabalho utilizou-se das bibliotecas e funções da linguagem e ambiente de programação R (PROJECT.ORG, 2018; HORNIK, 2012) em todas as iterações matemáticas dos modelos matemáticos elencados no Capítulo 3 de forma automática, sendo desnecessária qualquer iteração manual das fórmulas apresentadas no referido capítulo. A linguagem de e ambiente de programação R é amplamente utilizada na realização de cálculos matemáticos, econômicos e estatísticos pra fins acadêmicos e científicos.

Por fim, para solucionar o problema ao qual se propõe resolver, o presente estudo científico utiliza-se dos dados disponibilizados por (CEPEA, 2018), entidade vinculada a USP responsável no país pela coleta, tabulação e divulgação de dados históricos dos preços das commodities brasileiras à sociedade, no endereço eletrônico *www.cepea.esalq.usp.br* na *internet*.

1.1 Definição do problema

Qual modelo matemático apresenta melhor acurácia na previsão de preço da commodity etanol combustível no Brasil para o dia seguinte (próximo ponto) em uma série temporal?

1.2 Objetivo

O presente trabalho - através de experimento de aplicação - compara a acurácia preditiva dos modelos matemáticos ARIMA, ARFIMA e Exponencial Suavizado, em uma série temporal de preços da commodity etanol combustível no Brasil, em uma previsão para o dia seguinte (próximo ponto).

1.3 Importância da pesquisa

Pesquisas relacionadas a acurácia de modelos preditivos de produção e de preços de commodities, o que inclui o etanol combustível no Brasil, possui relevância:

- (i) para os agricultores, que necessitam tomar decisões mercadológicas e de produção;

- (ii) para a indústria, que necessita tomar decisões de estoque pertinente a cadeia produtiva; e

- (iii) para os governos de diversos países, que necessitam tomar decisões visando proteger a economia doméstica e proporcionar a segurança alimentar.

1.4 Motivação

Como motivação a presente pesquisa apresenta-se o fato deste estudo possuir um objeto de escopo macroeconômico, considerando que os preços das commodities brasileiras, enquanto variáveis, possuem grande relevância ao país, levando em consideração a depen-

dência econômica das commodities nas exportações nacionais. Dependência econômica essa que remonta a época do Brasil Colonial, com exploração da cana-de-açúcar, do ouro e do pau-brasil.

Outra fato que motivou a realização da presente pesquisa é a inexistência de estudos relacionados a previsão de preços da commodity etanol, para o dia seguinte, na literatura científica econômica mundial.

1.5 Limites e limitações

A série temporal trabalhada na pesquisa aqui dissertada é, especificamente, referente aos preços da commodity etanol combustível no mercado brasileiro, limitada ao período compreendido entre 25 de janeiro de 2010 e 04 de fevereiro de 2014 (1000 pontos), obtidos em pregões diários em preços à vista.

Os modelos matemáticos utilizados no experimento são os mesmos utilizados por (DAVID *et al.*, 2018) (o ARIMA e o ARFIMA), acrescido de mais um modelo: o Exponencial Suavizado.

Um outro fator limitante ao experimento é a janela de tempo para previsão: que é o dia seguinte (próximo ponto ou um passo a frente).

1.6 Organização da Dissertação de Mestrado

Este documento apresenta 7 capítulos e está estruturado da seguinte forma:

- **Capítulo 1 - Introdução:**
- **Capítulo 2 - Fundamentação Teórica - Commodities:**
- **Capítulo 3 - Fundamentação Teórica - Séries Temporais:**
- **Capítulo 4 - Base de Dados e Metodologia:**
- **Capítulo 5 - Resultados:**

- **Capítulo 6 - Conclusões:**

Fundamentação Teórica - Commodities

2.1 Commodities: conceito

Commodities – no âmbito da pesquisa realizada – são produtos intensivos em recursos naturais em estado bruto (de caráter primário) ou com mínimo teor de industrialização, que são produzidos de forma padronizada e em grandes quantidades; cujos preços são formados em bolsa de valores dentro do país ou no exterior (VERÍSSIMO; XAVIER, 2014).

De acordo com (BRESSER-PEREIRA, 2010) as commodities são categorizadas – sejam elas em estado bruto ou com reduzido grau de industrialização – como produtos agrícolas, minerais ou energia. Pode-se verificar a categorização das commodities citada anteriormente de forma semelhante no Sumário Executivo Commodity Markets Outlook (BANK, 2017), publicado mensalmente pelo Banco Mundial, conforme ilustrado seguir(Figura 2.1):

Figura 2.1: Preços nominais indexados e previsões anuais de commodities. Fonte:BM

| | Price Indexes (2010=100) | | | | | | Change (%) | | Revision ² | |
|------------------------------|--------------------------|-------|-------|-------|--------------------|--------------------|------------|---------|-----------------------|--------------------|
| | 2013 | 2014 | 2015 | 2016 | 2017f ¹ | 2018f ¹ | 2016-17 | 2017-18 | 2017f ¹ | 2018f ¹ |
| Energy | 127 | 118 | 65 | 55 | 68 | 71 | 23.7 | 4.0 | -1.1 | -4.0 |
| Non-Energy ³ | 102 | 97 | 82 | 80 | 84 | 85 | 4.9 | 0.6 | 0.8 | 0.7 |
| Agriculture | 106 | 103 | 89 | 89 | 89 | 90 | -0.6 | 1.2 | -0.7 | -0.7 |
| Beverages | 83 | 102 | 94 | 91 | 83 | 84 | -8.4 | 0.7 | -1.7 | -1.7 |
| Food | 116 | 107 | 91 | 92 | 92 | 93 | -0.1 | 1.2 | -0.2 | -0.2 |
| Oils and meals | 116 | 109 | 85 | 90 | 89 | 91 | -0.5 | 1.7 | -2.7 | -2.5 |
| Grains | 128 | 104 | 89 | 82 | 82 | 83 | -0.2 | 1.9 | 2.4 | 2.3 |
| Other food | 104 | 108 | 100 | 105 | 106 | 106 | 0.4 | 0.1 | 0.6 | 0.5 |
| Raw Materials | 95 | 92 | 83 | 80 | 82 | 83 | 2.4 | 1.6 | -1.2 | -1.1 |
| Fertilizers | 114 | 100 | 95 | 75 | 72 | 72 | -4.1 | -0.2 | -3.7 | -5.5 |
| Metals and Minerals | 91 | 85 | 67 | 63 | 77 | 76 | 22.4 | -0.7 | 4.3 | 4.3 |
| Precious Metals ³ | 115 | 101 | 91 | 97 | 97 | 97 | -0.2 | -0.8 | 1.3 | 1.7 |
| Memorandum items | | | | | | | | | | |
| Crude oil (\$/bbl) | 104 | 96 | 51 | 43 | 53 | 56 | 23.8 | 5.7 | -2.0 | -4.0 |
| Gold (\$/toz) | 1,411 | 1,266 | 1,161 | 1,249 | 1,250 | 1,238 | 0.1 | -1.0 | 25.0 | 31.7 |

2.2 Commodities no contexto econômico contemporâneo

No mercado global contemporâneo, segundo (JARROW; CHATTERJEA, 2013), os contratos de negócios futuros e a termo são utilizados em transações de commodities (são exemplos de commodities: cobre, ouro, trigo, soja, dentre outras) e em contratos de ativos financeiros (como, por exemplo, ações, moedas e índices financeiros); tais contratos

sofreram mudanças significativas a partir da década de 70 que podem ser destacadas em quatro categorias:

- (i) a introdução de novas espécies contratos de futuros no mercado;
- (ii) a abertura de novas bolsas no mundo;
- (iii) a consolidação e maior vínculo institucional entre as bolsas; e
- (iv) a automação das negociações em função das novas tecnologias.

Ainda, segundo ([JARROW; CHATTERJEA, 2013](#)), uma variedade de novas espécies de contratos foi concebida nas últimas três décadas do século XX e as trocas envolvendo commodities tornaram-se cada vez mais conectadas e dinâmicas, inclusive, o comércio se tornou cada vez mais automatizado, muitas vezes eliminando corretores e intermediários do processo para permitir que os comerciantes interagissem em uma interface eletrônica.

No contexto econômico contemporâneo, o Brasil, conforme a publicação técnica da ([UNCTAD, 2016](#)), é descrito como um país em desenvolvimento que depende de maneira significativa das commodities em suas exportações, no momento em que as mesmas representam 63 % de todos os produtos exportados nos biênios 2009 – 2010 e 2014 – 2015.

No Brasil, nos dias atuais, a instituição Brasil Bolsa Balcão (B3) – nome em vigor da antiga bolsa de valores Bovespa – é a empresa brasileira de infraestrutura de mercado financeiro de relevância mundial, responsável por transações de bolsa e de balcão – inclusive – por negociações de contratos futuros de commodities como açúcar, boi gordo, café, etanol, milho, ouro, petróleo e soja.

Em paralelo as mudanças significativas ocorridas nas formas de negociação das commodities no cenário econômico contemporâneo no mundo, organismos internacionais, assim como a comunidade científica, apresentam interesse investigativo nos preços das commodities, enquanto variável, para análise econômica das nações.

Exemplo do interesse científico pelos preços das commodities na atualidade é o trabalho de (KOHZADI *et al.*, 1996), na medida em que comparam a acurácia preditiva dos modelos Média Móvel Auto-Regressiva (ARIMA) e RNA na previsão dos preços mensal à vista do trigo e da carne bovina americana.

Consequente ao trabalho de (KOHZADI *et al.*, 1996) listado anteriormente, (FAO, 2017) acompanha a evolução dos preços das commodities no contexto econômico global contemporâneo, no documento *Commodities and Development Report 2017: Commodity Markets, Economic Growth and Development*, na medida em que correlaciona o grau de desenvolvimento econômico de um país com a dependência das commodities, bem como retrata a relevância da flutuação dos seus preços dentro de um escopo macroeconômico.

Já (MOREIRA, 2014), ainda dentro do contexto econômico global contemporâneo, assume que a mudança nos preços das commodities funciona como uma importante variável na determinação da inflação de consumo em países em desenvolvimento, bem como em economias emergentes ; e (TADASSE *et al.*, 2016) investigam a volatilidade nos preços trigo, milho e soja, e mostram como esses fatores desencadeiam crises e mudanças extremas nos preços do alimentos, numa abordagem relacionada a segurança alimentar nos países.

2.3 Contextualização da *commodity* etanol no Brasil contemporâneo

A produção de biocombustíveis líquidos emergiu como alternativa de uso de combustíveis fósseis para fins de transporte no mundo (TRIANA, 2011). Assim, a mitigação das emissões de carbono, a segurança energética e o desenvolvimento agrícola global são os principais impulsionadores deste tipo de projetos de bioenergia (TRIANA, 2011).

O etanol, enquanto alternativa energética, portanto, é descrito como um combustível de energia renovável que ajuda a mitigar as mudanças climáticas, destacando os avanços tecnológicos que poderiam permitir a produção não só de etanol tradicional, mas também do etanol celulósico de segunda geração a partir do bagaço de cana (BENITES-LAZARO; MELLO-THÉRY; LAHSEN, 2017).

O Brasil em 2017 produziu 26 % do etanol no mundo ([ASSOCIATION *et al.*, 2018](#)), ocupando, dessa forma, a segunda posição entre os países produtores (Figura 2.2).

No Brasil, atualmente, o etanol combustível é derivado da cana-de-açúcar e é usado puro ou misturado à gasolina em uma mistura chamada gasohol (25% de etanol, 75% de gasolina) ([OLIVEIRA; VAUGHAN; RYKIEL, 2005](#)).

Veículos dedicados ao etanol foram introduzidos no Brasil a partir do final da década de 70, pelo programa do governo brasileiro chamado Proálcool – como política governamental de enfrentamento às duas crises mundiais do petróleo ([CAVALCANTI; SZKLO; MACHADO, 2012](#)).

Consequente a introdução dos veículos a álcool no final da década de 70, ocorreria a evolução no mercado brasileiro dos veículos comerciais leves e de passageiros com a inovação da tecnologia Flex Fuel – que utiliza simultaneamente Gasolina tipo C ou Etanol hidratado – a partir de março 2003, impactaram significativamente na demanda pelo etanol no mercado consumidor brasileiro e quantitativamente na variação (a maior) de seu preço frente a Gasolina tipo C ([GOMEZ; LEGEY, 2015](#)).

Por consequência da introdução no Brasil da tecnologia Flex Fuel, acompanhado do aumento da demanda pelo açúcar no mercado internacional, impulsionou um crescimento significativo na indústria de cana-de-açúcar nas últimas décadas ([MONTASSER *et al.*, 2015](#)).

Em 2017 a quantidade de emplacamentos de veículos automóveis e comerciais leves, com motor de tecnologia de combustível Flex Fuel, correspondeu a 88,60 % do total de veículos emplacados (Figura 2.3) ([ANFAVEA, 2018](#)).

De acordo com os estudos de ([DAVID *et al.*, 2018](#)) a volatilidade dos preços do etanol no Brasil em anos consecutivos está associada principalmente aos seguintes fatores:

- (i) quantidade de produção de cana-de-açúcar;

Figura 2.2: Produção mundial do etanol no ano 2017. Fonte: RFA



(ii) percentual de cana para a produção de etanol, ou seja, para o mix de produção;

(iii) renda do consumidor;

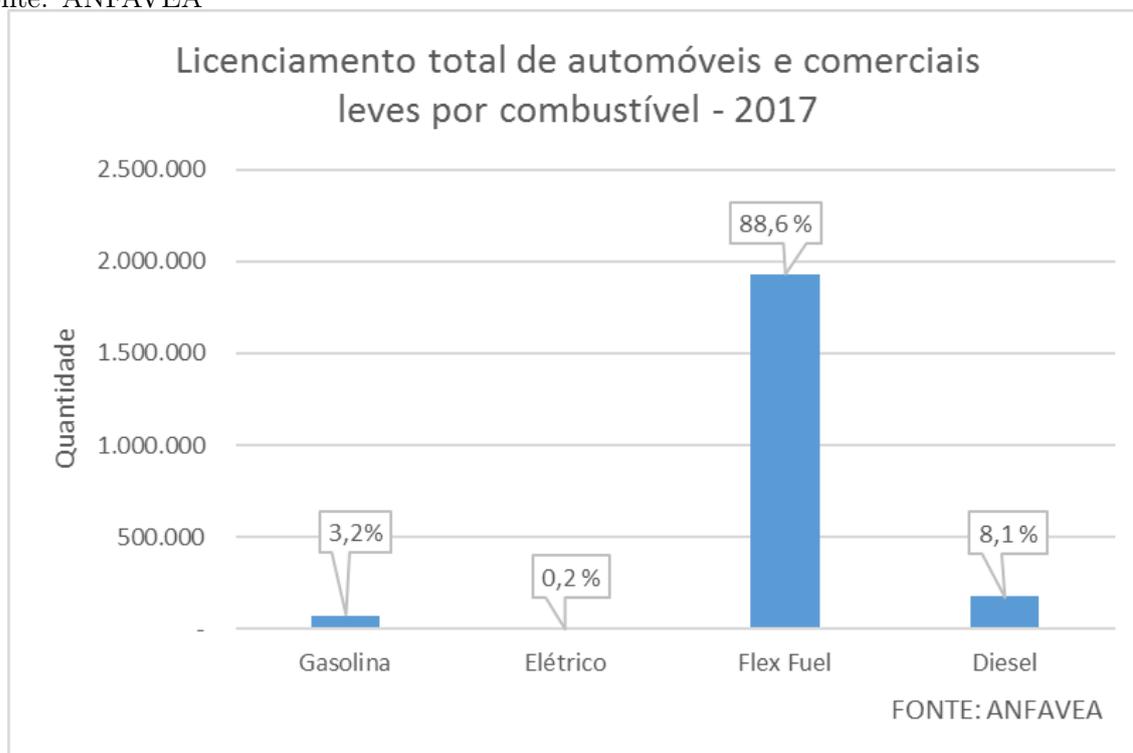
(iv) número de veículos da frota comercial leve; e

(v) preço da gasolina tendo em vista a mistura compulsória de tal tipo de etanol na venda de gasolina.

Na expectativa de compreender cientificamente a problemática da volatilidade nos preços do etanol brasileiro na contemporaneidade, no âmbito econômico, a comunidade científica efetuaram estudos recentes.

(FREITAS; KANEKO, 2011) formularam análise empírica utilizando uma abordagem de cointegração com teste de limite de atraso distribuído (ARDL) – com o objetivo de oferecer uma estimativa de demanda do etanol no Brasil após 2003 – utilizando se como elementos no modelo: o preço do etanol, a renda média dos consumidores, o crescimento da frota de veículos e o efeito cruzado com o gasohol.

Figura 2.3: Licenciamento total de automóveis e comerciais leves por combustível no ano 2017. Fonte: ANFAVEA



Já ([AJANOVIC, 2011](#)) investigaram se a produção de biodiesel no mundo – com abordagem da produção do etanol brasileiro – tem a capacidade de incrementar os preços dos alimentos, utilizando-se como método a relação fundamental entre a quantidade produzida, os custos da produção e os preços do mercado resultantes.

([HUANG *et al.*, 2012](#)), por outro lado, avaliaram impactos futuros da produção dos biocombustíveis nos setores agrícolas regionais e afins, especificamente nas nações em desenvolvimento, dessa forma, utilizando-se da modelagem GTAP para explicar as interações globais dos mercados regionais de biocombustíveis e alimentos, concluem que o aumento da produção de biocombustíveis nos Estados Unidos (EUA), Brasil e União Europeia (UE) leva a aumento de commodities agrícolas (alimentares ou não) e que a magnitude exata e a fonte do impacto dos biocombustíveis no mundo em desenvolvimento dependem do preço internacional do petróleo e do grau de substituição entre o petróleo e os biocombustíveis.

Colaboraram, também, quanto a produção científica recente sobre os preços da commodity combustível etanol no Brasil ([CAVALCANTI; SZKLO; MACHADO, 2012](#)), quando testam o nível de correlação, através da análise econométrica direta dos logaritmos, entre

os preços e o petróleo brent.

Já (BARROS; GIL-ALANA; WANKE, 2014) examinaram o grau de persistência no consumo de etanol e de gasolina no Brasil usando integração fracionária em dados mensais de janeiro de 2000 até dezembro de 2012, juntamente com as respectivas séries de preços

(MONTASSER *et al.*, 2015), também dentro do escopo da commodity etanol combustível no Brasil contemporâneo, testaram a existência de “bolhas” na relação entre os preços do etanol e os da gasolina, no período entre 2000 e 2012, através de um modelo recursivo generalizado do Dickey-Fuller Aumentado (DFA).

(GOMEZ; LEGEY, 2015) contribuíram com seus estudos ao combinar os modelos de Correção de Erros (ECM) com o Filtro de Kalman (KF), estudam a influência dos preços relativos do etanol hidratado e Gasolina C no nível consumo destes combustíveis entre 2001 e 2011.

Em outro trabalho científico relacionado a commodities, (DAVID *et al.*, 2018) aplicaram vários modelos matemáticos numa série temporal de preços do etanol brasileiro, dentre eles, a análise de Flutuação retirando a tendência (DFA), os expoentes de Hurst e Lyapunov, examinam a acurácia preditiva entre os modelos Média Móvel Integrada Fracionada (AFRIMA) e Média Móvel Auto-Regressiva (ARIMA) em uma previsão de 365 dias dos preços do etanol numa série temporal apresentando, como resultado, uma vantagem do modelo ARFIMA sobre o ARIMA.

2.4 Commodities e Complexidade

Enquanto os preços da maioria das commodities e alimentos estavam no mesmo nível entre 1980 e 2000, aumentaram dramaticamente desde o início dos anos 2000. Os preços atingiram o pico em 2008, despencaram durante a crise financeira global e iniciaram forte recuperação no início de 2009 (BELKE; BORDON; VOLZ, 2013).

Dez anos após o início da crise financeira global, tornou-se evidente que a expansão do setor financeiro, a sofisticação fenomenal dos produtos financeiros e a velocidade sem precedentes de transações financeiras juntas alteraram profundamente a relação entre fi-

nanças, economia e sociedade (LAGOARDE-SEGOT, 2017). Nesse contexto de expansão e críticas ao Teorema do Mercado Eficiente, (KARYOTIS; ALIJANI, 2016; ULUSOY; ONBIRLER, 2017) argumentam que as flutuações nos preços das ações refletem assimetrias de informação que resultam de movimentos especulativos e levam a efeitos de transbordamento nos mercados financeiros e de commodities.

No cenário de alta volatilidade dos preços das commodities agrícolas apresentado nas últimas décadas, medir e prever com precisão a dinâmica dos preços futuros de commodities agrícolas é, inevitavelmente, um componente importante, não apenas na gestão de riscos comerciais, mas também na especulação de preços. A previsão do preço dos futuros de commodities agrícolas é considerada uma tarefa desafiadora, devido ao fato de que os preços são altamente voláteis, complexos e dinâmicos, sendo, portanto, de grande interesse para pesquisadores de finanças, profissionais do mercado e formuladores de políticas (XIONG *et al.*, 2015).

Compreender a dinâmica dos preços das commodities nas últimas décadas é de interesse não apenas dos negociantes de commodities, mas também dos formuladores de políticas nos países exportadores e importadores de commodities (LOF; NYBERG, 2017). Infelizmente, os preços das commodities são notoriamente difíceis de prever. Quanto a muitos outros ativos negociados em mercados financeiros competitivos, os preços das commodities exibem um comportamento quase aleatório, com mudanças praticamente imprevisíveis por preços passados ou outros fatores observáveis (LOF; NYBERG, 2017).

Diante da dificuldade na manipulação dos preços das commodities enquanto objeto de estudo científico, inclusive no que diz respeito a previsão, surgiram estudos recentes sobre as flutuações desses preços embasados teoricamente na econofísica, disciplina esta que abriu novos horizontes de pesquisa e contribuição científica para a economia (PEREIRA; SILVA; PEREIRA, 2017), trazendo aplicações em Agentes Baseados em Modelos, em memória de longo prazo, em correlação cruzada e em Sistemas Complexos.

Assim, são exemplos atuais da literatura científica que abordam os preços das commodities num contexto dinâmico, aleatório, não linear e complexo:

i) (BENEDETTO *et al.*, 2014) utilizam método de processamento de sinais de máxima entropia para a análise de séries temporais de preços financeiros e de commodities agrícolas e não agrícolas;

- ii) ([ANTONAKAKIS *et al.*, 2018](#)) propuseram examinar a relação causal entre fundos de commodities e retornos usando dados mensais para o período de maio de 1997 a agosto de 2015, utilizando wavelets dada a forte evidência de não-linearidade e quebras estruturais;
- iii) ([WILES; ENKE, 2014](#))[10] propuseram contribuir para a literatura de modelos de previsão de preços agrícolas, utilizando inteligência artificial e modelagem não linear ;
- iv) ([LI; LU, 2011](#)) propuseram investigar as características multifractais das séries temporais de volatilidade dos mercados futuros de commodities agrícolas da China, usando Análise Multifractal de Flutuação Detectada e análise de espectro multifractal ;
- v) ([BEKIROU *et al.*, 2017](#)) propuseram investigar as ligações causais dinâmicas entre os mercados futuros de ações e commodities dos EUA através da utilização de teoria complexa de redes;
- vi) ([HE; CHEN, 2011](#)) utilizou o DCCA para investigar a correlação cruzada entre duas grandezas características, a saber, relações preço-volume nos mercados de commodities agrícolas;
- vii) ([HERRERA; RODRIGUEZ; PINO, 2017](#)) propuseram um modelo Markov Switching Multifractal Peaks-Over-Threshold (MSM-POT) para capturar o comportamento dinâmico das ocorrências aleatórias de eventos extremos que excedam um limiar alto em séries temporais de retornos de em preços diários de 06 commodities agrícolas e não agrícolas;
- viii) ([JIA *et al.*, 2016](#)) propuseram utilizar o método do caminho ótimo térmico (TOP) para estudar a relação dinâmica entre os mercados de commodities agrícolas chinês e americano, tanto nos retornos quanto na volatilidade;
- ix) ([KABLAN; FTITI; GUESMI, 2017](#)) propuseram analisar a interdependência entre os ciclos de preços de crédito e de commodities entre os exportadores africanos de commodities usando uma análise co-espectral evolucionária que estabelece uma medida de correlação dinâmica e variável no tempo;

- x) (LAHMIRI; UDDIN; BEKIROS, 2017) propuseram analisar os co-movimentos entre os principais mercados de ações financeiros internacionais, mercados de câmbio e mercados de commodities após a recente crise financeira global através das wavelets;
- xi) (LÜBBERS; POSCH, 2016), usando um modelo de fator dinâmico generalizado, investigam a volatilidade individual dos mercados de commodities correlacionando uma amostra ampla de trinta e um retornos futuros de commodities entre 1996 e 2015;
- xii) (LOF; NYBERG, 2017) propuseram investigar a relação entre os preços de commodities e as taxas de câmbio através de regressões preditivas lineares padrões e, também, através de regressões preditivas não casuais;
- xiii) (ORDU; ORAN; SOYTAS, 2018) propuseram investigar o incremento da correlação entre o mercado de ações e as commodities agrícolas a partir de 2008, em um fenômeno descrito pelo mesmo como “financeirização” das commodities, através do método Correlação Condicional Dinâmica Assimétrica (ADCC);
- xiv) (PAL; MITRA, 2017) propuseram avaliar a associação entre os preços do petróleo bruto e os índices mundiais de preços dos alimentos, primeiro no espaço e no tempo, e depois dentro da esfera combinada de tempo-frequência, utilizando a cointegração de Johansen e Wavelets;
- xv) (SHALINI; PRASANNA, 2016) propuseram estudar o impacto da crise financeira através da análise da presença de ruptura estrutural / mudança de regime na volatilidade nos preços das commodities, através do modelo de Markov e Wavelet durante a crise financeira;
- xvi) Xiong et al (XIONG *et al.*, 2015) propuseram investigar a previsão de preços de intervalos de séries temporais através do modelo vetorial de correção de erros (VECM) e regressão vetorial de suporte multi-saída (MSVR), capturando simultaneamente padrões lineares e não-lineares.

Fundamentação Teórica - Séries Temporais

3.1 Séries Temporais: conceitos fundamentais

De maneira geral, as investigações empíricas dependem de dados arranjados em forma cronológica e a análise de séries temporais (AST) objetiva descrever e analisar os comportamentos passados da série, tendo em vista à compreensão do comportamento da série e a conseqüente previsão de movimentos futuros (FONSECA; MARTINS; TOLEDO, 1985).

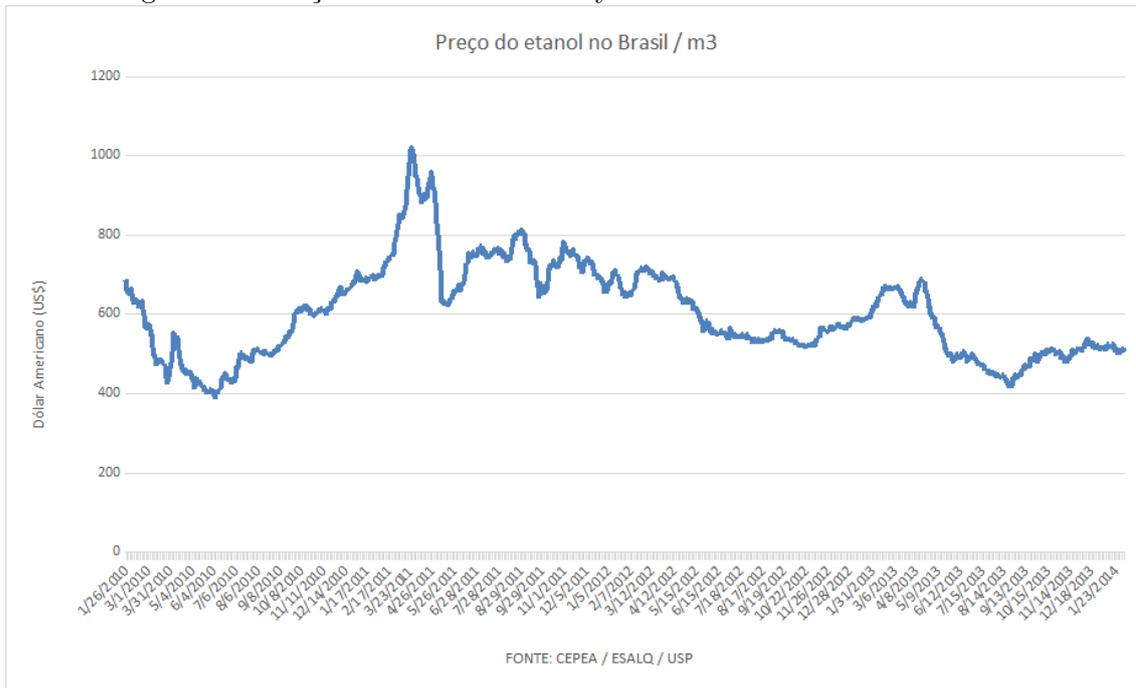
A análise de séries temporais tem sido de grande relevância em diversas áreas do conhecimento como, por exemplo, economia, ciências sociais, demografia, epidemiologia, entre outras. Série temporal é um conjunto de observações ordenadas no tempo, conforme o entendimento de Morettin e Tolo (MORETTIN; TOLOI, 2004). Seja Y a variável observada e t a variável tempo. Uma série temporal é definida pelos valores $Y_1, Y_2, Y_3 \dots Y_n$, nos tempos $t_1, t_2, t_3 \dots t_n$. Assim Y é uma função de t simbolizada por:

$$Y = f(t) \tag{3.1}$$

São exemplos de series temporais:

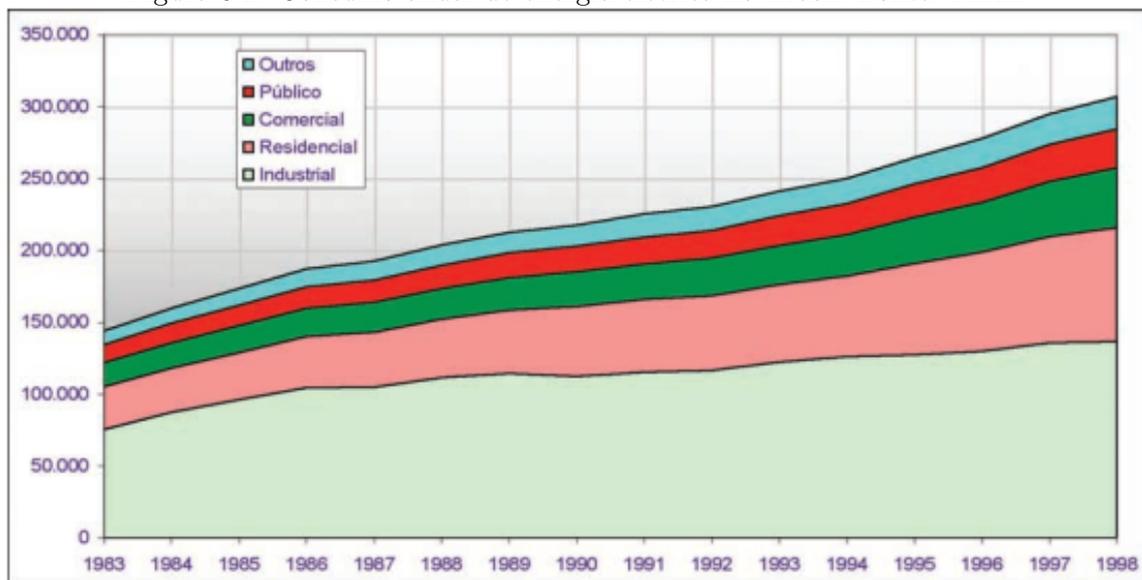
- (i) Preço diário do etanol no Brasil (Figura 3.1);

Figura 3.1: Preço diário da commodity etanol no Brasil. Fonte: CEPEA



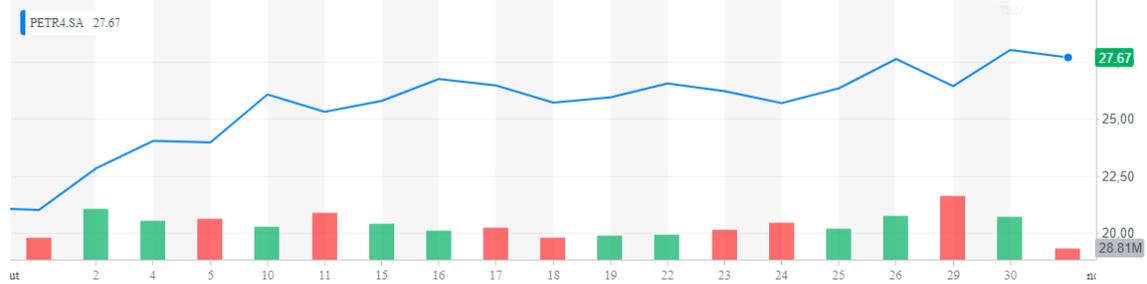
(ii) Consumo anual de energia elétrica no Brasil (Figura 3.2);

Figura 3.2: Consumo anual de energia elétrica no Brasil. Fonte: MME



(iii) Preço diário das ações da Petrobrás (PETR4) na bolsa de valores B3 (Figura 3.3).

Figura 3.3: Preços diários das ações da Petrobrás (Petr4) na bolsa de valores B3. Fonte: B3



A AST é a forma científica pela qual são identificados os segregados e fatores relacionados com o tempo que influenciam os valores observados na série. Uma vez identificados, podem ser usados para auxiliar na interpretação e na projeção de valores da série temporal. Segundo (MORETTIN; TOLOI, 2004) existem, de maneira geral, dois enfoques utilizados nos procedimentos de análise de séries temporais, ambos com objetivos determinados. No primeiro enfoque, a análise é feita no domínio do tempo, ou seja, é considerada a evolução temporal do processo, e os modelos propostos são modelos paramétricos (com um número finito de parâmetros). Já no segundo enfoque a análise é concebida no domínio de frequência, e os modelos propostos são modelos não paramétricos como, por exemplo, a análise espectral.

Em variadas situações no campo das ciências físicas, exatas, biológicas e humanas freqüente o conceito de sistema dinâmico, submetendo-se a uma série de entrada $X(t)$, uma série de saída $Z(t)$ e uma função de transferência v (Figura 3.4).

Figura 3.4: Representação de um sistema dinâmico: Fonte: Morettin; ToloI (2004).



Os sistemas lineares, segundo (VEIGA; SÁFADI, 1999) são de relevante importância, onde a saída está relacionada com entrada por meio de um funcional linear envolvendo $v(t)$. Segundo estes, um exemplo característico é a chamada função de transferência, representada pela seguinte expressão:

$$Z(t) = \sum_{r=0}^{\infty} v(\tau)X(t - \tau) \quad (3.2)$$

Daí surgem problemas de interesse que são:

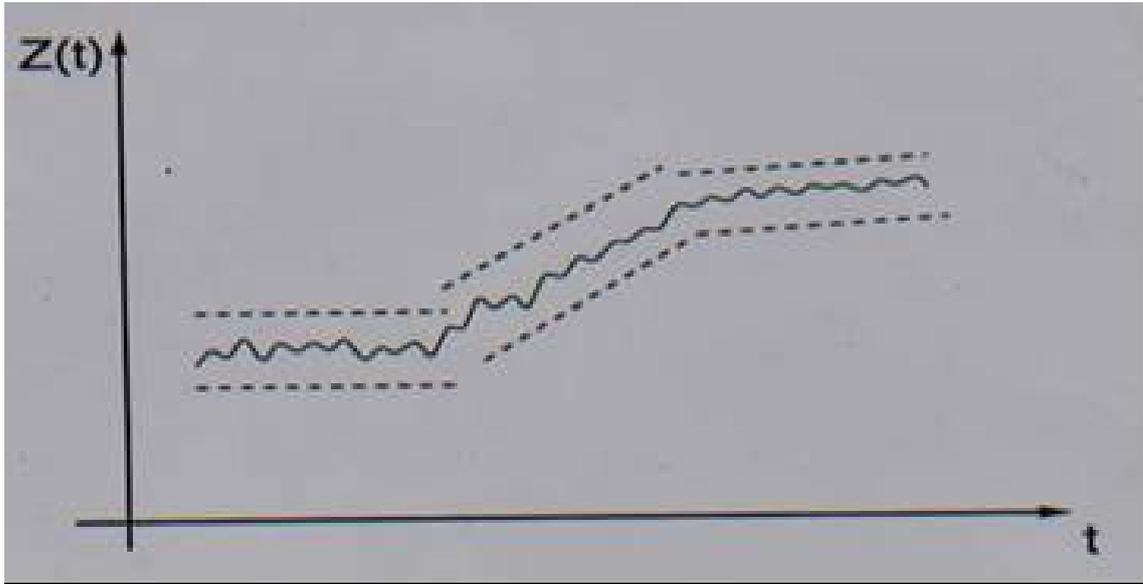
- a) estimar a função de transferência $v(t)$, conhecendo as séries de entrada e de saída;
- b) fazer previsões da série $Z(t)$, com o conhecimento de observações de observações da série de entrada $X(t)$;
- c) estudar o comportamento do sistema, simulando-se a série de saída;
- d) controlar a série de saída $Z(t)$, de modo a trazê-la o mais próximo possível de um valor almejado, ajustando-se de maneira conveniente a série de entrada $X(t)$; este controle é necessário devido a perturbações que normalmente afetam um sistema dinâmico.

Um caráter relevante da AST é identificar se a série é estacionária, em outras palavras, se a série desenvolve aleatoriamente no tempo em torno de uma média constante, caracterizando assim certo padrão de equilíbrio estável. Mas, na maioria das vezes, as séries temporais apresentam característica de não-estacionariedade. Outra característica relevante ao analisar uma série temporal é identificar componentes periódicos e separar componentes determinísticos, dos puramente estocásticos ([ANTENEODO, 2004](#)).

Os modelos utilizados em séries cronológicas fundamentam-se preponderantemente no conceito de estacionariedade. Sendo assim, em sua maioria, os modelos considerados são lineares estacionários fracos, como modelos autoregressivos (AR), de médias móveis (MA) e mistos (ARMA). Existem, também, modelos capazes de descrever estatisticamente séries estacionárias e não estacionárias, como por exemplo, o modelo ARIMA (auto-regressivos integrados médias móveis) desde que não apresentem comportamento explosivo ([MORETTIN; TOLOI, 2004](#)). Uma série temporal pode apresentar característica de estacionariedade durante um período longo ou apenas em períodos muito curtos, mudando de nível e/ou de inclinação (Figura 3.5).

Devido ao fato de grande parte dos modelos estatísticos de AST supor que as séries sejam estacionárias, é preciso transformar os dados de origem, caso estes não se caracterizam

Figura 3.5: Série não-estacionária quanto ao nível e inclinação. Fonte: Morettin; Toloí (2004)



como de uma série estacionária. O procedimento de transformação mais elementar é estabelecer diferenças sucessivas da série original, até conseguir a desejada estacionariedade (MORETTIN; TOLOI, 2004). Dessa forma a primeira diferença de $Z(t)$ é definida pela expressão a seguir:

$$\Delta Z(t) = Z(t) - Z(t - 1), \quad (3.3)$$

a segunda diferença

$$\Delta^2 Z(t) = \Delta[Z(t) - Z(t - 1)], \quad (3.4)$$

de forma geral, a n -ésima diferença de $Z(t)$ é

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]. \quad (3.5)$$

Ocorre com grande frequência em ST econômicas e financeiras, por exemplo, a presença de tendências, podendo gerar um incremento na variância ao longo do tempo. Neste caso, uma transformação não linear, como a logarítmica pode ser aplicada (MORETTIN; TOLOI, 2004).

Existem, basicamente, dois motivos para se transformar os dados originais de uma série temporal: estabilizar a variância e tornar o efeito sazonal aditivo. Uma outra razão de se transformar os dados, de acordo com (MORETTIN; TOLOI, 2004), é obter uma série mais próxima possível de uma distribuição simétrica aproximadamente normal, ou seja $\mu = 0$ (média igual a 0) e $\delta^2 = 1$ (variância igual a 1).

Caso consideremos Z_1, Z_2, \dots, Z_n como variáveis aleatórias, tendo uma distribuição conjunta $f(z_1, \dots, z_n)$ e distribuições marginais $f(z_1), \dots, f(z_n)$, a série será estacionária se, para qualquer momento de tempo t e n , tivermos $f(z_t) = f(z_{t+n})$, $n = \pm 1, \pm 2, \dots$, isto é, as distribuições de Z_1, Z_2, \dots, Z_n são invariantes por translações ao longo do tempo. Assim, Z_t e Z_{t+n} terão os mesmos parâmetros, por exemplo,

$$E(Z_t) = E(Z_{t+n}), \quad (3.6)$$

$$Var(Z_t) = Var(Z_{t+n}), \quad (3.7)$$

onde $E(Z_t)$ é o valor esperado da variável Z e $Var(Z_t)$ é a variância de Z .

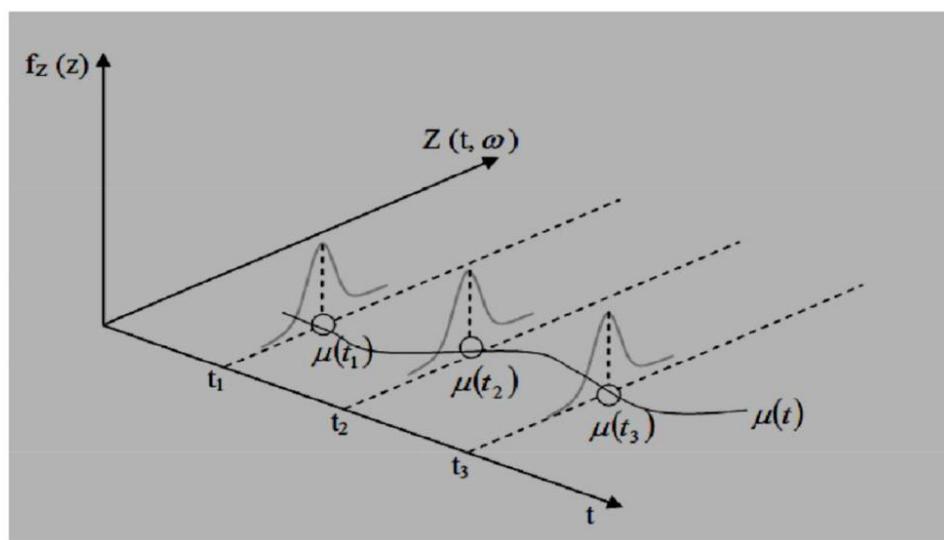
Ou seja, a média e a variância de uma série estacionária são constantes, independentemente do instante de tempo t , e pode-se representar da seguinte forma:

$$E(Z_t) = \mu, \forall t \quad (3.8)$$

$$Var(Z_t) = \delta^2, \forall t. \tag{3.9}$$

Algumas séries temporais são baseadas em processos estocásticos, isto é, processos controlados por leis probabilísticas. Se relacionarmos T como um conjunto arbitrário. Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória (MORETTIN; TOLOI, 2004; PEREIRA, 1984). Nestas condições um processo estocástico pode ser considerado uma família de variáveis aleatórias (v.a), e podemos supor que elas sejam definidas no mesmo espaço amostral de probabilidade. Na realidade $Z(t)$ é uma função de dois argumentos, $Z(t, \omega), t \in T, \omega \in \Omega$ (Figura 3.6).

Figura 3.6: Um processo estocástico interpretado como uma família de variáveis aleatórias. Fonte: Morettin; Toloi, (2004)



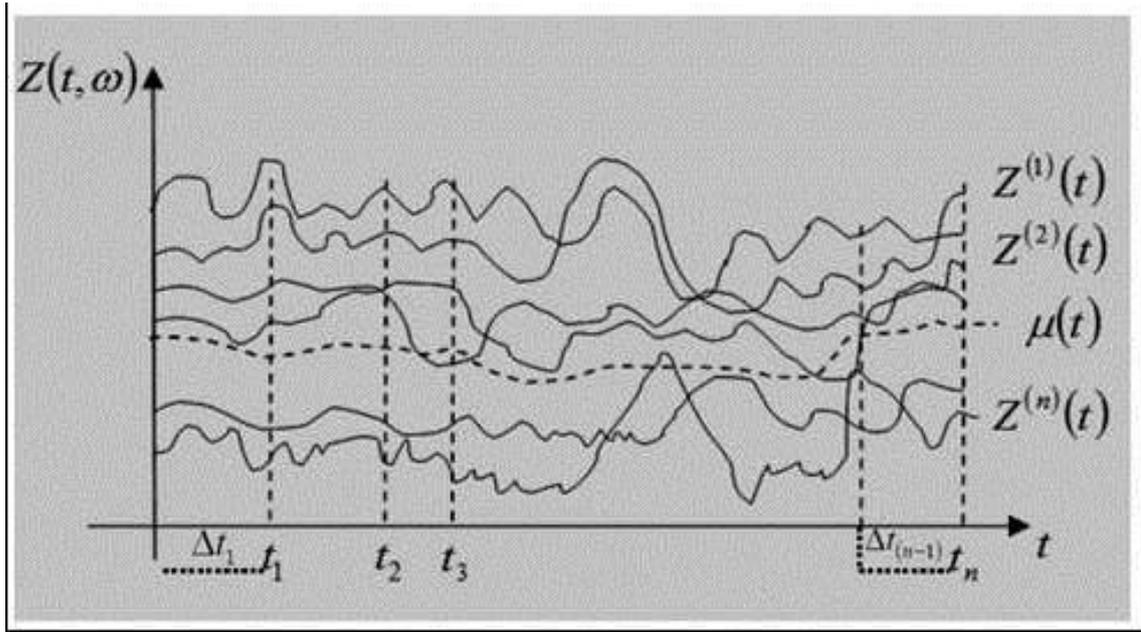
Ao analisar a Figura 3.6 verifica-se, que para cada $t \in T$, teremos uma v.a. $Z(t, \omega)$, como uma distribuição de probabilidade, ou seja é possível que a função densidade de probabilidade $f_z(Z)$ no momento t_1 difira da $f_z(Z)$ no momento t_2 .

Se designarmos as possíveis realizações de $Z(t, \omega)$, por $Z^{(1)}(t), Z^{(2)}(t), \dots, Z^{(n)}(t)$, poderemos chamar o conjunto de todas as possíveis trajetórias como “ensemble”. E, também, podemos observar que cada realização $Z^j(t)$ é uma função do tempo t não aleatória e, para cada t fixo, $Z^j(t)$ é um número real.

Segundo (MORETTIN; TOLOI, 2004) uma forma de analisar uma distribuição de proba-

bilidade $Z(t, \omega)$, para um t fixo, é levar em consideração durante a análise a quantidade de trajetórias que passam por uma janela de amplitude Δt de modo que, tal quantidade de trajetórias será relacionada à $f_z(Z)\Delta t$. Onde $\Delta t = t_j - t_{(j-1)}$, j variando de $1, \dots, n$ pontos da série (Figura 3.7).

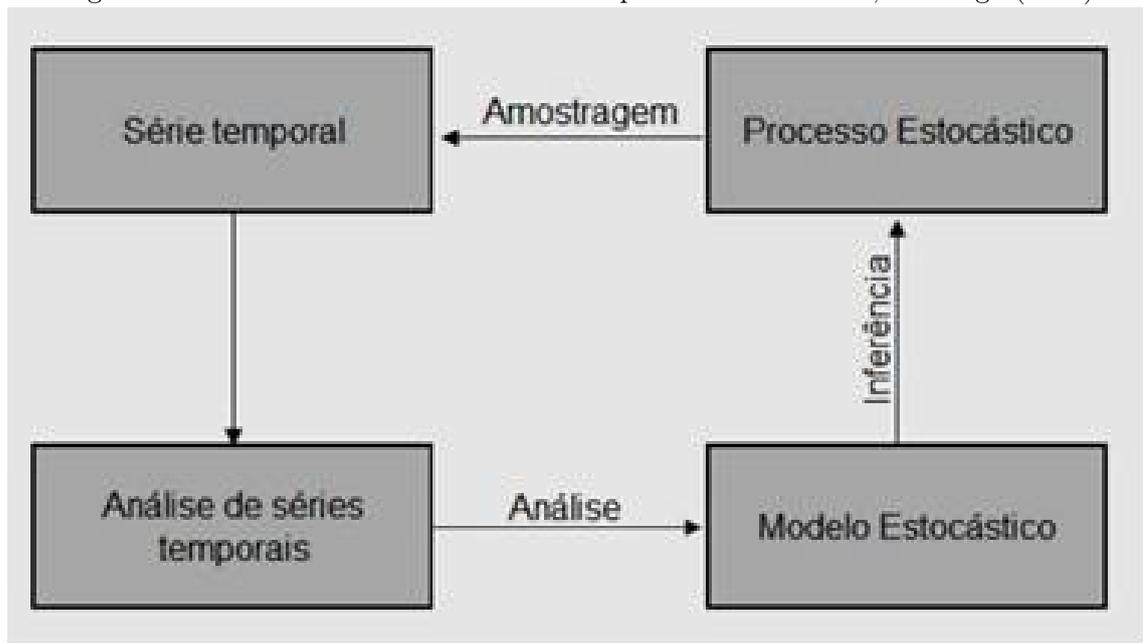
Figura 3.7: Um processo estocástico interpretado como uma família de trajetórias. Fonte: Morettin; Toloí (2004).



Um processo estocástico está estatisticamente determinado quando é conhecido suas funções de distribuição até a N -ésima ordem. Por exemplo, dado um processo estocástico $Z = Z(t), t \in T$, o mesmo estará especificado se para todo $n > 1$ conhecemos as distribuições finito-dimensionais. Isto denota que, para $n = 1$ conhecemos as distribuições de v.a $Z(t_1), t \in T$ para $n = 2$, conhecemos as distribuições bidimensionais da v.a $(Z(t_1), Z(t_2)), t_1, t_2 \in T$, e assim por diante (MORETTIN; TOLOI, 2004; PEREIRA, 1984). Porém, na prática, segundo Souza e Camargo (2004), dificilmente se conhece todas as funções de distribuição até a N -ésima ordem e frequentemente tem-se apenas uma realização do processo estocástico em análise. Com esta situação, é necessário realizar uma inferência do mecanismo gerador da série (Figura 4.8).

A Figura 3.8 ilustra a sequência e o objetivo do estudo de séries temporais. Em outras palavras, considerando um processo estocástico em análise retira-se uma amostra finita de observações (série temporal), e através do estudo dessa amostra, é possível definir um modelo com propósito de inferir sobre o comportamento da realidade (GUJARATI; PORTER, 2011). Uma relevante classe de modelos estocásticos corresponde ao processo estacionário, i.e., aquele que apresenta média constante com equilíbrio em seu desenvolvimento. Um processo estocástico $Z = \{Z(t), t \in T\}$ é considerado estacionário, se para

Figura 3.8: Processo estocástico e Série temporal. Fonte: Souza; Camargo (2004)



todo t tivermos:

$$E [Z(t)] = \mu, \quad (3.10)$$

$$V [Z(t)] = V [Z(t) - \mu]^2 = \delta^2; \quad (3.11)$$

$$E [(Z(t) - \mu) (Z(t - k) - \mu)] = f_z(k), k = 1, 2, 3... \quad (3.12)$$

onde μ é a média amostral esperada por $E \{Z(t)\}$ e δ^2 a variância esperada em $V[Z(t)]$, que não variam ao longo do tempo. A Equação 4.12 exprime que as autocovariâncias independem no tempo. Ou seja, elas variam apenas em função da distância ou ordem k que afasta as duas observações consideradas (ENDERS, 2003).

Além das propriedades citadas anteriormente, outra propriedade pertinente em uma ST

é a autocorrelação, que determina se a série possui memória de longa duração ou não, ou seja, se a mesma é persistente, antipersistente ou descorrelacionada. Uma série temporal pode ser caracterizada como persistente se observações adjacentes estão correlacionadas positivamente. A ST é denominada como antipersistente se a série apresentar correlação inversa, em outras palavras, se existir a tendência de valores grandes serem seguidos por valores pequenos como, por exemplo, o batimento cardíaco. Se uns dos casos acima mencionados não ocorrerem à série é dita descorrelacionada, como ocorre com o ruído branco (ANTENEODO, 2004).

3.2 Modelos para séries temporais

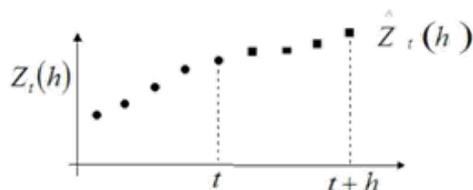
Explicitar por meio de modelos matemáticos sistemas e fenômenos observáveis, desde a antiguidade, sempre foi um desafio. Uma das significativas mudanças, ocorridas na metade dos anos 90, foi à crescente tendência no desenvolvimento de se obter modelos matemáticos a partir de dados observados e não mais exclusivamente com base nas equações que descrevem a física do processo. Uma das possíveis razões dessa mudança foi o acesso a computadores com baixo custo e bom desempenho (AGUIRRE, 2004).

A palavra modelo pode ser denotada como uma representação simplificada da realidade, estruturada de tal maneira que possibilite a compreensão do funcionamento total ou parcial de uma dada realidade ou fenômeno (MATOS, 2000). “Por um modelo matemático entende-se uma construção abstrata de um fenômeno real, que deveria comportar-se em concordância com a situação real em todos os seus aspectos relevantes” (SOUZA; CARMARGO, 2004).

Os modelos utilizados em ST são os processos estocásticos. Ao descrever determinada ST por um modelo, é necessário especificar uma função-perda como, por exemplo, o erro quadrático médio (EQM) (Equação 3.12), além do modelo proposto, para daí então obter um procedimento de previsão. Digamos que dispomos de observações de uma série temporal até o momento t e precisamos prever o valor da série no instante $t + h$ (Figura 3.9), neste caso t é a origem e $Z_t(h)$ denotará a previsão de $Z(t+h)$, de origem t e horizonte h , poderemos definir o erro quadrático médio pela seguinte expressão (MORETTIN; TOLOI, 2004):

$$EQM = [Z(t+h) - Z_t(h)]^2. \quad (3.2)$$

Figura 3.9: Observações de uma série temporal com previsões de origem t e horizonte h . Fonte: Morettin; Tolo (2004).



De posse de um dado modelo que possa descrever uma ST até o momento t e considerando que temos o objetivo de minimizar a expressão acima, obteremos uma expressão para $Z_t(h)$. A previsão de uma ST pode ser classificada como de curto, médio e longo prazo, isso dependendo do valor associado ao horizonte máximo de previsão. Mas, qualquer que seja o horizonte de previsão (curto, médio ou longo prazo), não há um consenso entre os estudiosos sobre o horizonte máximo de previsão. Este estará correlacionado com o grau de previsibilidade da ST e também com o erro de previsão. Além disso, a qualidade da previsão de um determinado modelo, não pode ser sustentada quando forem identificadas mudanças significativas geradas a partir da origem, em relação àquelas utilizadas para a construção do modelo.

Comumente os modelos de previsão de ST fundamentam-se na suposição de que observações passadas contêm propriedades sobre o padrão de comportamento da ST. O objetivo dos métodos, neste caso, é estabelecer diferença entre o padrão de qualquer ruído que possa estar contido nas observações para daí então usar esse padrão para prever valores ou movimentos futuros da série.

A classificação dos modelos de ST, de acordo com o número de parâmetros envolvidos, se dá através de duas classes: modelos paramétricos (aqueles para os quais este número de parâmetro é finito) e não-paramétricos (os que envolvem um número infinito de parâmetros). Na classe de modelos paramétricos, cuja análise é realizada no domínio do tempo, os mais frequentes na utilização estão os modelos de erro (ou de regressão), os modelos auto-regressivos e de médias móveis (ARMA), os modelos auto-regressivos integrados e de médias móveis (ARIMA), modelos de memória longa (ARFIMA), modelos estruturais e modelos não-lineares. E os modelos não-paramétricos comumente utilizados são a função de autocovariância (autocorrelação) e a transformada de Fourier.

Inclusive, pode-se classificar, também, os modelos de ST em duas categorias: modelos automáticos e modelos não automáticos. Os modelos automáticos correspondem aqueles

cuja aplicação se dá diretamente com o auxílio do computador. Já os não automáticos exigem a intervenção de pessoal especializado, para serem modelados, devido à dificuldade de automatização do modelo. Entre os modelos automáticos podemos citar os modelos de alisamento exponencial simples, os modelos auto-regressivo stepwise e a filtragem adaptativa. No rol de modelos não-automáticos pode-se mencionar o chamado método Box-Jenkins e o Bayesiano. A utilização frequente dos modelos automáticos de ST é devida à simplicidade, à eficiência computacional e à sua razoável previsão (TOLOI; MORETTIN, 1987; VEIGA; SÁFADI, 1999).

Grande parte dos métodos de ST estão fundamentados apenas na análise de uma série temporal visando especificar algum modelo que a descreva, porém existem alguns procedimentos de previsão que tentam modelar o comportamento de uma ST por meio de outras séries. Com essa premissa, (SOUZA, 1989) classifica os modelos de ST, dependendo do número de ST's envolvidas na análise, em modelos univariados e multivariados. A modelagem de uma única série temporal denomina-se análise univariada enquanto que análise de mais de uma série distinta trata-se de uma análise multivariada (MADDALA; WELLER, 2003).

Existe na literatura uma significativa variedade de modelos de séries temporais, cada um com suas peculiaridades, em outras palavras, cada uma com suas capacidades e limitações (WHEELWRIGHT; MAKRIDAKIS *et al.*, 1985). Diante disso, fica evidente que qualquer que seja a classificação do modelo de ST, é possível modelar estatisticamente um número significativo de diferentes modelos para analisar o comportamento de uma série particular.

A construção e seleção de modelos de ST dependem de diversos fatores como, por exemplo, comportamento do fenômeno ou conhecimento a priori que temos de sua natureza e do objetivo do estudo e, também, da acessibilidade a software adequados (MORETTIN; TOLOI, 2004).

3.2.1 Modelo ARIMA

O modelo Autorregressivo Integrado de Média Móvel – Autoregressive Integrated Moving Average (ARIMA) é um procedimento popular entre os modelos estatísticos de análise de séries temporais (LEE; KO, 2011). Esse modelo foi proposto por (BOX; JENKINS, 1976) e tem origem nos modelos autorregressivo (AR), médias móveis (MA) e da combinação dos modelos AR e MA (ARMA). Além de incluir modelos não estacionários (ARIMA) e

sazonais (SARIMA). Cada um destes modelos pode modelar uma série isolada ou combinadamente.

No modelo AR(p), o valor atual da série é expresso como um agregado linear de p valores anteriores e um ruído aleatório. Segundo (BOX; JENKINS, 1976), (MORETTIN; TOLOI, 2004), (EHLERS, 2009) e (LEE; KO, 2011), um modelo AR(p) é dado pela representação matemática (3.2.1):

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t, \quad (3.2.1)$$

em que p é a ordem de AR, Z_t é a série temporal observada no momento t, a_t é o ruído aleatório em uma distribuição normal com média 0 e variância igual a 1 e ϕ_1, \dots, ϕ_p são os parâmetros de AR(p). Ao introduzir o operador de defasagem B, que define $Z_{t-1} = BZ_t$; e, conseqüentemente, $Z_{t-p} = B^p Z_t$, a Equação (3.2.1) pode ser reescrita por (3.2.2)

$$\phi_P(B)Z_t = a_t, \quad (3.2.2)$$

em que $\phi_P(B)$ é o operador de AR(p), definido por

$$\phi_P(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p. \quad (3.2.3)$$

O modelo MA (q), que explora a estrutura de autocorrelação dos resíduos de previsão do período atual com aqueles ocorridos em períodos anteriores é descrito pela equação (BOX; JENKINS, 1976; MORETTIN; TOLOI, 2004; EHLERS, 2009; LEE; KO, 2011):

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (3.2.4)$$

e pode ser reescrito pela seguinte equação (3.2.5)

$$Z_t = \theta_q(B)a_t, \quad (3.2.5)$$

em que q é a ordem de MA (q), $1, \dots, q$ são os parâmetros do modelo MA(q), e $q(B)$ o qual é definido por $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. Para aumentar a flexibilidade na adaptação de séries temporais as equações (3.2.2) e (3.2.5) são combinadas para obter o modelo ARMA (p, q) que apresenta processos mistos AR(p) e (q), dado por

$$\theta_p(B)Z_t = \theta_q(B)a_t, \quad (3.2.6)$$

em que $\Delta^d Z_t = (1 - B)^d Z_t$ e d é a ordem de diferenciação.

3.2.2 Modelo SARIMA

Muitas séries temporais apresentam padrões repetitivos que aparecem regularmente a cada intervalo de tempo. Como é o caso das indústrias, em que vendas e produção seguem uma sazonalidade forte em determinados períodos do ano. Para lidar com as séries que apresentam autocorrelação sazonal, (BOX; JENKINS, 1976) generalizaram o modelo ARIMA e definiram o modelo ARIMA sazonal multiplicativo, conhecido como SARIMA(p, d, q)(P, D, Q) e representado por

$$\phi_p(B)\Phi_p(B^S)\Delta^d\Delta_s^D Z_t = \theta_q(B)\Theta_Q(B^S)a_t, \quad (3.2.2)$$

em que $\Delta_s^D Z_t = (1 - B^S)^D Z_t$, D é a ordem de diferenciação sazonal, $\Phi_p(B^S)$ é o operador sazonal AR (p) definido por $\Phi_p(B^S) = 1 - \Phi_1(B^S) - \Phi_2(B^{2S}) - \dots - \Phi_p(B^{pS})$; $\Theta_Q(B^S)$ é o operador sazonal MA (q) por $\Theta_Q(B^S) = 1 - \Theta_1(B^S) - \Theta_2(B^{2S}) - \dots - \Theta_Q(B^{QS})$ e Φ_1, \dots, Φ_p são parâmetros do modelo sazonal AR (p) e $\Theta_1, \dots, \Theta_Q$ são parâmetros do modelo sazonal MA (q).

3.2.3 Modelo ARFIMA

Segundo (MORETTIN, 2008), os modelos ARFIMA são classes de modelos que são capazes de descrever de maneira simultânea as dinâmicas de memórias de curta e longa de processos estacionários. Assim, dizemos que $\{X_t\}$ é um processo auto regressivo fracionário integrado de média móvel, denominado ARFIMA (p, d, q) com $d \in (-\frac{1}{2}, \frac{1}{2})$ se $\{X_t\}$ for estacionário e satisfizer a equação

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t, \quad (3.2.3)$$

onde $a_t \sim RB(0, \sigma_a^2)$ e $\phi(B)$ e $\theta(B)$ são polinômios em B de graus p e q , respectivamente.

3.2.4 Modelos de Suavização Exponencial

A suavização exponencial é um modelo estatístico que começa de uma equação de médias móveis, com ponderações exponenciais, que tem como objetivo produzir ajustes nas variações aleatórias dos dados de uma série temporal específica. Esse procedimento de suavização utiliza uma ponderação distinta para cada valor observado na série temporal, de modo que valores mais recentes recebam pesos maiores. Dessa forma, os pesos formam um conjunto que decai exponencialmente a partir de valores mais recentes (JR, 2006).

Segundo (MORETTIN; TOLOI, 2004), o tipo de classificação de um modelo de suavização exponencial além da componente tendência pode também estar relacionada com o fator sazonalidade e pela introdução do erro, tanto na forma aditiva como na multiplicativa. A seguir são elencadas as três formas básicas de suavização exponencial muito utilizadas na prática.

3.2.4.1 Modelo de Suavização Exponencial Simples

O tipo mais trivial de série temporal é aquele em que os valores da série flutuam aleatoriamente ao redor de um valor fixo, sem manifestação de qualquer tendência. Se a série temporal se mantém constante em relação a um nível médio, uma suavização exponencial simples preconizada por (BROWN, 1959; BROWN, 1963) pode ser aplicada

para a previsão de valores futuros da série. Sua representação matemática é ilustrada por (MAKRIDAKIS; HYNDMAN; WHEELWRIGHT, 1998).

$$\widehat{Z}_{t+1} = \alpha Z_t + (1 + \alpha)\widehat{Z}_t, 0 \leq \alpha \leq 1, \quad (3.2.4.1)$$

onde \widehat{Z}_{t+1} é o valor estimado para o intervalo de tempo $t + 1$, α é o parâmetro constante de suavização do método; que pode assumir valores entre 0 e 1 ($0 \leq \alpha \leq 1$); Z_t é o valor real atual para o período de tempo t e, \widehat{Z}_t é o valor previsto para o mesmo período de tempo t . De acordo com a equação acima, o valor previsto para o próximo período \widehat{Z}_{t+1} é uma junção do valor previsto atual \widehat{Z}_t e o valor da observação atual Z_t . Assim, seleciona-se de forma aleatória um valor inicial para a constante, a partir do qual previsões são geradas. Comparam-se os valores estimados com os reais, e calcula-se a média do quadrado das diferenças entre os mesmos; o parâmetro que minimiza essa média é utilizado no modelo resultante (PELLEGRINI; FOGLIATTO, 2001).

3.2.4.2 Modelo de Suavização Exponencial Dupla (Método de Holt)

A suavização exponencial dupla é frequentemente um modelo eficaz de previsão, introduzido por Holt (1957), que ampliou a suavização exponencial simples para dados de série temporal que exibam uma tendência linear. Este modelo incorpora elementos adicionais, à medida que introduz uma constante de suavização que interfere na tendência da série. A função de previsão do modelo de Holt é representada por (4.2.4.2):

$$\widehat{Z}_{t+n} = L_t + nT_t, \quad (4.2.4.2)$$

onde a previsão para o intervalo de tempo $t + n$ (\widehat{Z}_{t+n}) é igual estimativa do nível esperado da série temporal no período de tempo t (L_t) mais a influência esperada da tendência (taxa de aumento ou redução) durante os próximos n períodos (nT). A função de previsão na equação (3.2.4.2) pode ser utilizada para obter períodos de tempo n para previsões futuras, onde $n = 1, 2, 3$, e assim sucessivamente.

O modelo preconizado por Holt faz o uso dos parâmetros α e β denominados de constantes de suavização. Nesse modelo, além da função acima que efetua a previsão duas outras

funções são utilizadas para estimar o nível e a tendência da série temporal, de acordo com as equações (3.2.4.3) e (4.2.4.4), a seguir:

$$L_t = \alpha Z_t + (1 - \alpha)(L_{t-1} + T_{t-1}), 0 \leq \alpha \leq 1, \quad (3.2.4.3)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, 0 \leq \beta \leq 1, \quad (3.2.4.4)$$

onde L_t é a estimativa para o nível da série no período t e T_t é a estimativa da inclinação (tendência) da série no mesmo período t . A equação (3.2.4.2) ajusta diretamente L_t para a tendência do período que antecede T_{t-1} , acrescentando, conseguinte, o último valor suavizado L_{t-1} .

3.2.4.3 Modelo Holt-Winters (Suavização Exponencial Tripla)

Os modelos estatísticos de séries temporais são técnicas quantitativas comumente aplicadas para se obter prognósticos de variáveis, dentre os quais se encontram os modelos de suavização exponencial, particularmente o modelo Holt-Winters que é uma ampliação do modelo (HOLT, 1957) aprofundado por (WINTERS, 1960) para aplicação em séries temporais que possuem características de tendência e sazonalidade.

Levando em consideração que algumas séries específicas possuem um fator sazonal, além do nível e tendência que capta características da série que se repetem a intervalos regulares de tempo, Winters propõe métodos de projeção para essas séries, abordando dois tipos de efeitos sazonais: multiplicativo e aditivo.

O método de Holt-Winters para efeitos sazonais multiplicativos é largamente utilizado na modelagem de dados onde a amplitude do ciclo sazonal varia na mesma proporção ao nível da série ao longo do tempo. Seja um modelo de série sazonal, de período s , fator sazonal multiplicativo e tendência aditiva, isto é,

$$Z_t = L_t S_t + T_t + \epsilon_t, t = 1, 2, \dots \quad (3.2.4.3)$$

que representa o comportamento estrutural da série. As projeções dos valores futuros da série são realizadas por meio da função de previsão do modelo representada por

$$\widehat{Z}_{t+n} = (L_t + nT_t)S_{t-s+n}, n = 1, 2, \dots \quad (3.2.4.4)$$

onde \widehat{Z}_{t+n} é a previsão para n períodos à frente ($t + n$). Nesse modelo, além da equação (4.2.4.4), que efetua a previsão, três outras equações são utilizadas para estimar o nível, a tendência da série no período atual e os valores do fator sazonal respectivo ao último período de sazonalidade, de acordo com as equações (3.2.4.5), (3.2.4.6) e (3.2.4.7), a seguir:

$$L_t = \alpha \frac{Z_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1}), 0 \leq \alpha \leq 1, \quad (3.2.4.5)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}; 0 \leq \beta \leq 1, \quad (3.2.4.6)$$

$$S_t = \gamma \frac{Z_t}{L_t} + (1 - \gamma)S_{t-s}, 0 \leq \gamma \leq 1, \quad (3.2.4.7)$$

onde α , β e γ são as constantes de suavização que regulam o peso em relação ao nível (L_t), a tendência (T_t) e a sazonalidade (S_t), respectivamente. As equações (3.2.4.5) e (3.2.4.7) afirmam que, no período de tempo (t), há uma estimativa do índice sazonal no período de tempo ($t - s$) ou que há um valor (S_{t-s}). Dessa forma, precisa-se estimar os valores para (S_1, S_2, \dots, S_p). Uma maneira trivial de realizar as referidas estimativas sazonais preliminares requeridas para a utilização das equações de suavização é permitir que

$$S_t = \frac{Z_t}{\sum_{n=1}^s \frac{Z_n}{S}}, t=1, 2, \dots, s. \quad (3.2.4.8)$$

O modelo de Holt-Winter para efeitos sazonais aditivos é usado na modelagem de dados onde a amplitude do ciclo sazonal não depende do nível local da série, ou em outras palavras, permanece constante ao longo do tempo. Seja um modelo cuja série sazonal, de período s , é estabelecida pela soma do nível, tendência, um fator sazonal e um erro aleatório, dado por

$$Z_t = L_t + T_t + S_t + \epsilon_t, t = 1, 2, \dots \quad (3.2.4.9)$$

As projeções dos valores futuros da série são realizadas através da função de previsão do modelo descrita por

$$\hat{Z}_{t+n} = L_t + nT_t + S_{t-s+n}, n = 1, 2, \dots \quad (3.2.4.10)$$

onde \hat{Z}_{t+n} é a previsão para n períodos à frente ($t+n$). No modelo em questão, além da equação (3.2.4.10), que efetua a previsão, três outras equações são trabalhadas para estimar o nível, a tendência da série no período atual e, os valores do fator sazonal correspondente ao último período de sazonalidade, de acordo com as equações (3.2.4.11), (3.2.4.12) e (3.2.4.13), a seguir:

$$L_t = \alpha(Z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}); 0 < \alpha < 1, \quad (3.2.4.11)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}; 0 < \beta < 1, \quad (3.2.4.12)$$

$$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-s}, 0 < \gamma < 1, \quad (3.2.4.13)$$

onde α , β e γ são as constantes de suavização que regulam o peso em relação ao nível (L_t), a tendência (T_t) e a sazonalidade (S_t), respectivamente. As equações (3.2.4.11) e (3.2.4.13) afirmam que, no período de tempo (t), há uma estimativa do índice sazonal no período de tempo ($t - s$) ou que há um valor (S_{t-s}). Dessa forma, precisa-se estimar os valores para (S_1, S_2, \dots, S_p).

3.3 Medidas de acurária aplicadas a série temporal dos preços do etanol brasileiro

O termo acuracidade, esclarece ([MAKRIDAKIS; HYNDMAN; WHEELWRIGHT, 1998](#)), implica em qual medida um modelo preditivo se adequa aos dados de uma série temporal já conhecida. Segundo esses autores, portanto, a acurácia representa a eficiência de um modelo em relação à capacidade de reduzir erros de uma previsão de dados que ainda não são conhecidos.

No experimento de aplicação dos modelos matemáticos ARIMA e ARFIMA em série temporal de preços do etanol no Brasil, em uma previsão para o período de 365 dias (previsão de um ano a frente), ([DAVID *et al.*, 2018](#)) utiliza-se das medidas de acurácia U de Theils, MAE, ACF1 e MASE, obtendo vantagem quantitativa em todas as medidas de acurácia citadas para o modelo AFRIMA.

Para que seja possível identificar a acurácia, para fins comparativos, dos modelos em séries temporais de preços elencados no trabalho aqui proposto, torna-se necessário a aplicação de medidas estatísticas de erros nesses mesmos modelos.

As medidas de acurácia Erro Médio (ME) e MAE (do inglês Mean Absolute Error), foram detalhadas matematicamente, quanto a sua utilização, por ([MONTGOMERY; JENNINGS; KULAHCI, 2015](#)).

A medida mais trivial é o Erro Médio (ME), que consiste simplesmente na média da série de erros de previsão, conforme fórmula a seguir:

$$ME = \frac{\sum_t^T \varepsilon_t}{T} \quad (3.3)$$

O Erro Médio Absoluto, mede o valor médio de erro entre as séries observadas e ajustadas. Nesta análise esta variável será representada por MAE (do inglês Mean Absolute Error), cuja representação matemática é dada por:

$$MAE = \frac{\sum_t^T |\varepsilon_t|}{T} \quad (3.4)$$

A Raiz do Erro Quadrático Médio (simbolizado por RMSE, sendo esta abreviatura do inglês Root Mean Squared Error) representa as diferenças individuais quadráticas entre as séries temporais observadas e ajustadas:

$$RMSE = \sqrt{\frac{\sum_t^T |\varepsilon_t|^2}{T}} \quad (3.5)$$

Outra forma de medida de erro é a Média do Erro Absoluto Percentual (simbolizado por MAPE, do inglês Mean Absolute Percentage Error). A grande vantagem de utilizar esta expressão está na sua representação em termos percentuais que fornece um rápido entendimento, e uma desvantagem que deve ser considerada está no sentido de que se, por acaso, o valor do observado for muito pequeno, qualquer discrepância faz o MAPE “explodir”. A expressão utilizada com esta variável é representada por:

$$MAPE = \frac{\sum_t^T |P_t|}{T} \quad (3.6)$$

O Erro de Percentual Médio (MPE) é a média calculada dos erros percentuais, pelos quais as previsões de um modelo diferem dos valores reais:

$$MPE = \frac{\sum_t^T P_t}{T} \quad (3.7)$$

Maiores detalhes sobre estatística de erros, em especial as Eqs. (7), (8), e (9) que serão usadas podem ser obtidos em (Cochran, 1977; Montgomery et al., 2008).

Além das medidas descritas anteriormente, (HYNDMAN; KOEHLER, 2006) sugerem o uso de uma nova medida, escalonando o erro, baseado no erro absoluto médio da amostra treinameto em relação ao passeio aleatório do erro a um passo de previsão. A medida é chamada de ASE (Absolute Scale Error), dada pela seguinte equação:

$$q_t = \frac{|e_t|}{\frac{1}{n-m} \sum_{i=m+1}^n |y_i - y_{i-m}|} \quad (3.8)$$

Desta forma, portanto, deriva-se a medida Mean Absolute Scaled Error (MASE) da média(qt).

No que diz respeito a medidas de acurária, temos também o U de Theil (U). A citada medida pertence a categoria de medidas relativas e, neste caso específico, trata-se da razão entre a raiz quadrada da média do erro percentual quadrático e a raiz quadrada da média da taxa de variação percentual quadrática da série em análise. Desta forma, o U de Theil é a simples comparação entre o modelo em questão e um simples passeio aleatório. Dessa forma, quanto menor o índice, melhor será o modelo (em comparação com um passeio aleatório). Além disso, como se trata de valores ao quadrado, pesos maiores são relacionados a erros mais discrepantes:

$$U = \sqrt{\frac{\sum_t^T p_t^2}{\sum_t^T \left(\frac{Y_{t+1}-Y_t}{Y_t}\right)^2}} \quad (3.9)$$

Base de Dados e Metodologia

4.1 Base de Dados

A base de dados utilizada no experimento é a série temporal dos preços diários do etanol em Dólar Americano US\$ referente ao período de 25/01/2010 até 04/02/2014 (mil pontos). A referida base de dados, disponibilizada por ([CEPEA, 2018](#)), possui como representação gráfica a Figura 4.1.

4.2 Metodologia

A metodologia se refere a aplicação dos modelos matemáticos elencados no Capítulo 3 na base de dados indicada no subitem 4.1, através da linguagem de programação R e ambiente de desenvolvimento integrado (IDE) R-Studio ([HORNİK, 2012](#)), dos modelos matemáticos ARIMA, AFRIMA e Exponencial Suavizado, de acordo com as seguintes etapas (Figura 4.1):

- i) Coleta de dados;
- ii) Aplicação dos modelos matemáticos;
- iii) Resultado das previsões por modelo matemático aplicado;
- iv) Comparação da acurácia preditiva dos modelos aplicados na etapa (ii);

4.2.1 Aplicação do Modelo ARIMA

Para aplicar o modelo matemático ARIMA na série temporal objeto de estudo utilizamos a função `auto.arima()` disponível na biblioteca `forecast` do linguagem R. Segundo a documentação da linguagem em questão, a função `auto.arima()` retorna o melhor modelo ARIMA (p,d,q). A função conduz uma pesquisa sobre o modelo possível dentro das restrições de ordem fornecidas ([HORNİK, 2012](#)).

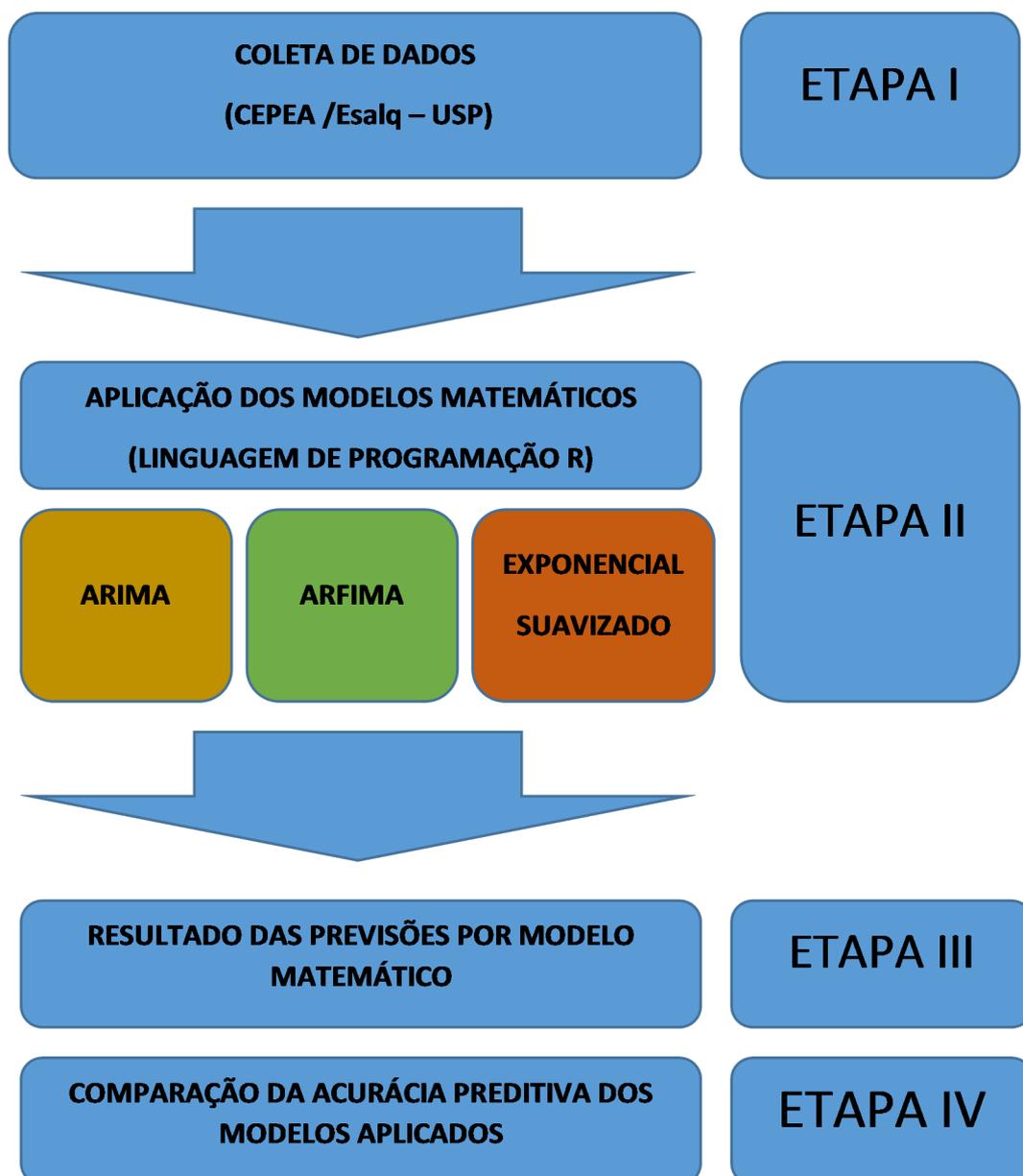
4.2.2 *Aplicação do Modelo ARFIMA*

Para aplicar o modelo matemático ARFIMA na série temporal objeto de estudo foi utilizado a função `arfima()` disponível na biblioteca `forecast` da linguagem R. Segundo a documentação da linguagem em questão, um modelo ARFIMA (p, d, q) ótimo é selecionado e estimado automaticamente ([HORNIK, 2012](#)).

4.2.3 *Aplicação do Modelo Exponencial Suavizado*

Para aplicar um modelo matemático Exponencial Suavizado na série temporal objeto de estudo foi utilizado a função `ets()` disponível na biblioteca `forecast` da linguagem R. Segundo a documentação da linguagem em questão, a aplicação da metodologia é totalmente automática e acordo com o valor AIC, AICc ou BIC de acordo com o melhor modelo. O único argumento requerido para a função `ets()` é a série temporal. Um modelo Exponencial Suavizado ótimo é escolhido automaticamente se não for especificado ([HORNIK, 2012](#)).

Figura 4.1: Metodologia aplicada ao experimento. Fonte: Autor.



Resultados

5.1 Resultado das Previsões por Modelo Matemático

A tabela 5.1, a seguir, apresenta a previsão do 1001º ponto por modelo matemático otimizado, gerados automaticamente pela linguagem R - modelo ARIMA do tipo (1, 1, 1), modelo ARFIMA do tipo (2, 0, 2) e o modelo exponencial suavizado do tipo (M, Ad, N) - com intervalos de confiança de 80% e 95%:

Tabela 5.1: Resultado das previsões do preço do etanol para ao dia seguinte por modelo matemático em Dólar Americano (US\$). Fonte: Autor.

| MODELO | PONTO | PREVISÃO | BAIXA 80 | ALTA 80 | BAIXA 95 | ALTA 95 |
|---------------------------|-------|----------|----------|----------|----------|----------|
| ARIMA (1,1,1) | 1001º | 520.9558 | 511.8058 | 530.1059 | 506.9621 | 534.9496 |
| EXP. SUAVIZADO (M, Ad, N) | 1001º | 520.4249 | 512.579 | 528.2708 | 508.4257 | 532.4241 |
| ARFIMA (2, 0, 2) | 1001º | 521.1896 | 512.0597 | 530.3196 | 507.2266 | 535.1527 |

5.2 Acurácia Preditiva dos Modelos Aplicados no Experimento.

A partir da aplicação dos modelos matemáticos ARFIMA, ARIMA e Exponencial suavizado na base de dados objeto deste trabalho, obtivemos as medidas de acurácia de cada modelo, a saber: Erro Médio (ME), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE), Erro Percentual Médio (MPE), Erro de Porcentual Médio Absoluto (MAPE) e Erro Escalonado Médio Absoluto (MASE), conforme a Tabela 5.2 a seguir:

Tabela 5.2: Acurácia Preditiva dos Modelos Aplicados no Experimento. Fonte: Autor.

| MODELO | ME | RMSE | MAE | MPE | MAPE | MASE |
|---------------------------|---------|--------|--------|---------|--------|--------|
| ARFIMA(2, 0, 2) | -0.0573 | 7.1099 | 5.1481 | -0.0245 | 0.8551 | 0.9156 |
| ARIMA(1,1,1) | -0.0618 | 7.1291 | 5.1357 | -0.0082 | 0.8534 | 0.9134 |
| EXP. SUAVIZADO (M, Ad, N) | -0.0483 | 7.1669 | 5.1533 | -0.0042 | 0.8562 | 0.9165 |

Conclusões e Perspectivas

De acordo com os dados elencados na Tabela 5.1, pode-se inferir que todos os modelos selecionados para o experimento em questão realizaram previsões de alta para o 1.001º ponto (previsões para o dia seguinte) da série temporal de preços da commodity etanol objeto do experimento, em relação ao 1.000 ponto, conforme a seguir: ARIMA (1,1,1) com preço previsto de US\$ 520,9558; Exponencial Suavizado (M, Ad, N) com preço previsto de com preço US\$ 520,4249 e ARFIMA (2,0,2) com preço previsto de US\$ 521,1896.

A aplicação dos modelos matemáticos ARIMA, ARFIMA e Exponencial Suavizado no experimento proposto pela presente pesquisa demonstra vantagem do modelo ARIMA do tipo (1,1,1) em relação comparativa aos demais modelos, quanto a melhor acurácia preditiva de preço para o dia seguinte da commodity etanol combustível, de 25 de janeiro de 2010 a 04 de fevereiro de 2014.

A vantagem do modelo ARIMA em relação aos demais modelos é verificada nos Resultados desta pesquisa (Capítulo 5, Subitem 5.2), onde pode-se observar menores erros de acurácia para o modelo ARIMA (1,1,1) em 04 (quatro) das 06 (seis) medidas disponíveis para comparação - ME, MAE, MPE e MASE -, elencadas de forma ilustrativa na tabela 6.1 e Figuras 6.1, 6.2, 6.3, 6.4, 6.5 e 6.6 a seguir:

Tabela 6.1: Tabela Comparativa das Medidas de Acurácia por Modelo Preditivo. Fonte: Autor.

| MODELO | ME | RMSE | MAE | MPE | MAPE | MASE |
|------------------------------|----|------|-----|-----|------|------|
| ARFIMA(2, 0, 2) | - | X | - | -X | - | - |
| ARIMA(1,1,1) | X | - | X | - | X | X |
| EXP. SUAVIZADO (M, Ad, N) | - | - | - | - | - | - |

Figura 6.1: Comparação da medida de acurácia ME entre os modelos matemáticos. Fonte: Autor.

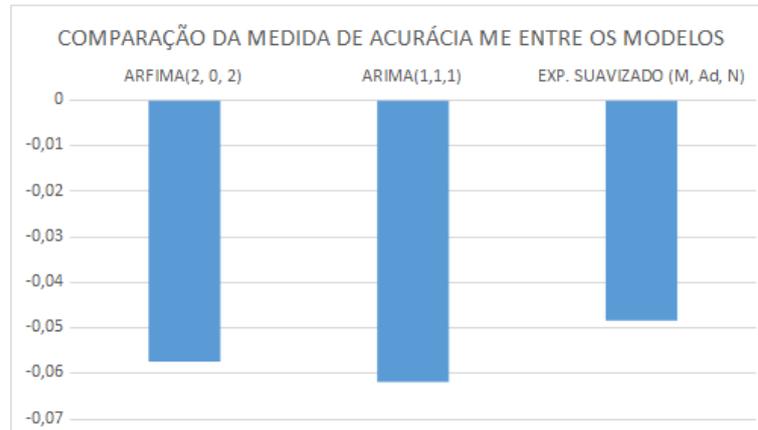


Figura 6.2: Comparação da medida de acurácia RMSE entre os modelos matemáticos. Fonte: Autor.

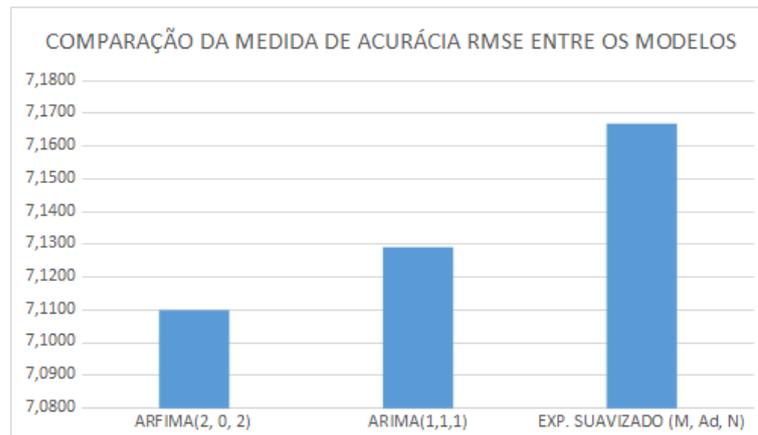


Figura 6.3: Comparação da medida de acurácia MAE entre os modelos matemáticos. Fonte: Autor.

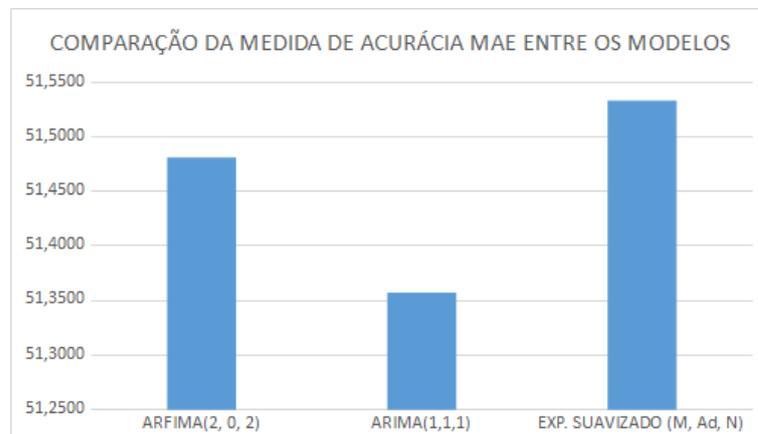


Figura 6.4: Comparação da medida de acurácia MPE entre os modelos matemáticos. Fonte: Autor.

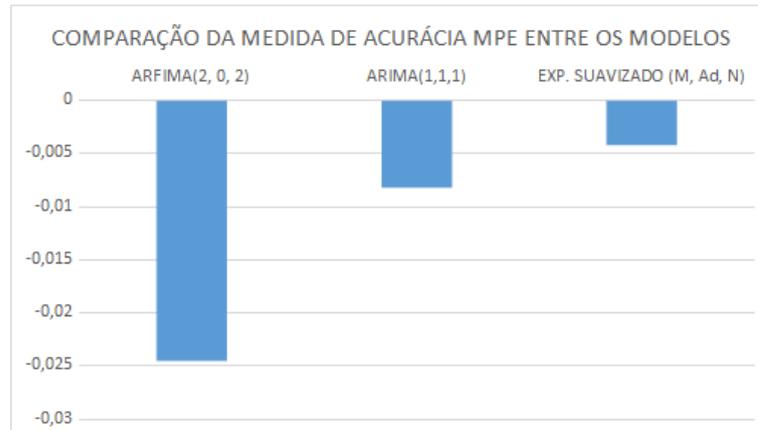


Figura 6.5: Comparação da medida de acurácia MAPE entre os modelos matemáticos. Fonte: Autor.

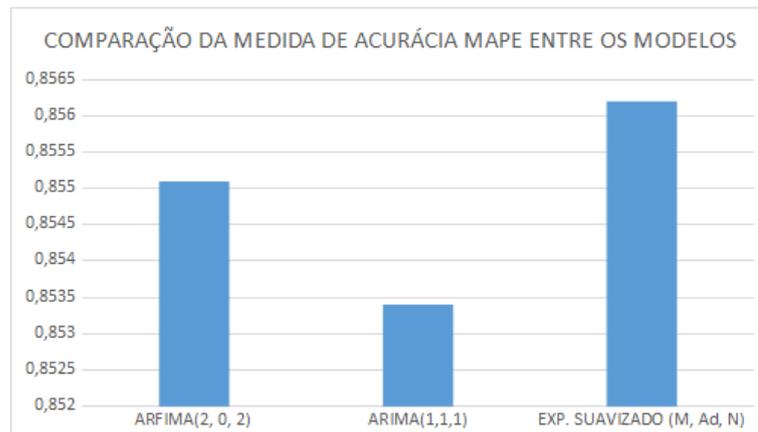
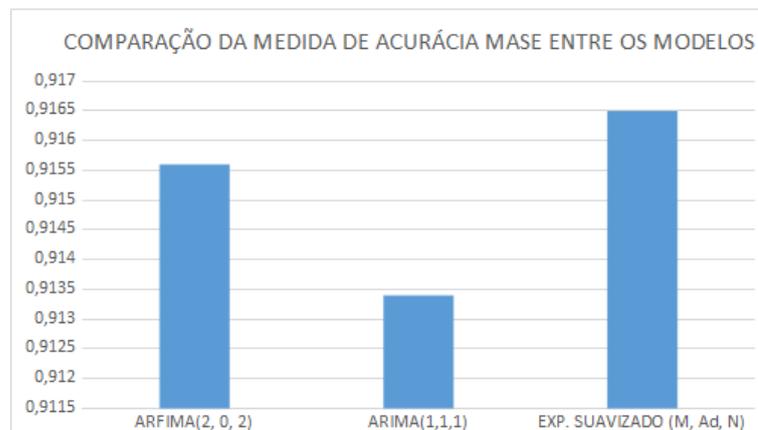


Figura 6.6: Comparação da medida de acurácia MASE entre os modelos matemáticos. Fonte: Autor.



Após a verificação do modelo preditivo mais acurado, percebe-se que o presente trabalho responde parcialmente a questão inicial da pesquisa, pois limita-se aos três modelos

introduzidos no experimento. Em outras palavras, ainda que o modelo ARIMA (1, 1, 1) tenha prevalecido em comparação aos modelos ARFIMA (2, 0, 2) e ao Exponencial Suavizado (M, Ad, N), tal resultado é insuficiente para esgotar a questão de qual modelo matemático apresenta melhor acurácia preditiva, frente a significativa quantidade de modelos disponíveis para aplicação e que não foram utilizados nesta pesquisa.

Convém salienta, neste ponto, que o levantamento e o registro da produção científica efetuados na revisão bibliográfica contidos no presente discorrer possui relevância na continuidade deste trabalho, pois sinaliza outras possibilidades metodológicas na previsão de preços da commodity etanol combustível no Brasil em trabalhos futuros.

Alguns modelos matemáticos relacionados aos Sistemas Complexos e a ML - que já foram aplicados em séries temporais de preços de outras commodities, modelos esses brevemente elencados na Fundamentação Teórica (Capítulo 3, Subitem 3.2.2) - carecem de aprofundamento específico quanto a previsão de preços em série temporal da commodity etanol combustível no Brasil, com o objetivo de buscar maior esclarecimento quanto a questão da presente pesquisa.

Um exemplo de possibilidade de trabalho futuro - que se encontra na seara dos Sistemas Complexos - é a aplicação da inteligência artificial em conjunto com a modelagem não-linear na previsão de preço da commodity etanol combustível no Brasil utilizando os mesmos caminhos percorridos por (WILES; ENKE, 2014).

Um outro assunto bastante em voga na abordagem commodities e Sistemas Complexos, conforme visto na revisão da literatura, é a correlação entre os preços de commodities em séries temporais distintas, bem como a correlação dos preços de uma determinada commodity e a inflação de um país. As correlações aqui citadas foram percebidas, por exemplo, nos trabalhos de (BEKIROUS *et al.*, 2017) e (PAL; MITRA, 2017) e abrem novas perspectivas de pesquisa - tanto no que diz respeito a aplicação de métodos existentes, quanto no âmbito da inovação métodos - para a commodity etanol combustível no Brasil, no que se refere a série temporal dos preços.

Fato intrigante nos resultados da pesquisa é o intervallo de preços da commodity etanol previstos no experimento - quando analisados independente do modelo de maior acurácia, em outras palavras, todos os valores previstos para o 1001º ponto independente do modelo matemático utilizado - que apresenta-se como uma possível ferramenta de análise

na seara da previsão de alta (ou de baixa) de preços de commodities. Tal constatação, detalhada a seguir, não faz parte do escopo do objeto do presente trabalho e carece de maior aprofundamento no que diz respeito a trabalhos futuros.

Apenas em caráter ilustrativo, no que diz respeito a possibilidade de ferramenta de previsão de alta (ou de baixa) do preço da commodity etanol no Brasil para o dia seguinte, de acordo com o intervalo dos preços previstos exposto na Tabela 5.1, tal intervalo apreciado de maneira independente da melhor acurácia matemática do experimento, temos: para 80% de intervalo de confiança a maior baixa o preço US\$ 511,8058 no modelo ARIMA e maior alta o preço US\$ 530,3196 no modelo ARFIMA; já no intervalo de confiança de 95% temos a maior baixa o preço US\$ 507,2266 e a maior alta o preço US\$ 535,1527.

A análise da maior alta e da maior baixa dos preços previstos como ferramenta de previsão, independente do modelo com melhor acurácia preditiva, pode oferecer a um operador (pessoa do Estado, da indústria ou do mercado financeiro em geral, responsável pela decisão de compra ou de venda de uma commodity) um quantitativo maior de cenários preditivos do preço da commodity em questão para dia seguinte. Por exemplo: um operador que almeja minimizar suas perdas em um cenário de baixa de preços, para os valores contidos no intervalo de 95% confiança, poderá adotar em sua ação os valores contidos entre o menor preço na baixa - ARIMA (1,1,1), US\$ 506,9621 - e o menor preço na alta - Exponencial Suavizado (M, Ad, N), US\$ 532,4241; já um operador que objetiva ampliar seus lucros em um cenário de alta de preços, no mesmo intervalo de confiança, poderá adotar em sua ação os valores contidos entre o maior preço na baixa - Exponencial Suavizado (M,Ad,N), US\$ 508,4257 - e o maior preço na alta - ARFIMA (2,0,2), US\$ 535,1527. Como já dito, tal percepção carece de maior aprofundamento científico em trabalhos futuros.

Considerando que o presente trabalho não encerrou o problema, muito pelo contrário, apenas introduziu na literatura econômica a discussão relacionada a comparação de acurácia entre modelos matemáticos na previsão de preços da commodity etanol combustível para o dia seguinte (um passo a frente), este trabalho sinaliza continuidade para um debate de um tema com uma variedade significativa de modelos preditivos na literatura científica específica.

Como perspectivas de ação após a realização da presente etapa, inicial, partindo do pressuposto o amplo horizonte disponível para novas aplicações e o desenvolvimento de novos produtos relacionados a série temporal de preços da commodity etanol combustível no

Brasil - tanto através de modelos matemáticos tradicionais, quanto por meio dos modelos derivados dos Sistemas Complexos -, apresentam-se como possibilidades:

- i) Ampliar o experimento elaborado com outros modelos matemáticos, como o SARIMA, descrito no Capítulo 3, Subitem 3.2.2, porém não aplicado no experimento por limitação de tempo;
- ii) Ampliar o experimento elaborado com outros modelos baseados em Sistemas Complexos, conforme possibilidades referenciais elencadas no Capítulo 2, Subitem 2.4;
- iii) Ampliar o experimento elaborado com outros modelos baseados em ML, conforme possibilidades apontadas por (LAGO; RIDDER; SCHUTTER, 2018) e do modelo Learning Matrix introduzido por (MONTEIRO *et al.*, 2016);
- iv) Modelagem computacional em UML e desenvolvimento de nova biblioteca na Linguagem R para o modelo Learnig Matrix, proposto por (MONTEIRO *et al.*, 2016), para a previsão de preços da série temporal da commodity etanol combustível no Brasil.

Referências Bibliográficas

- AGUIRRE, Luis Antonio. *Introdução à identificação de sistemas—Técnicas lineares e não-lineares aplicadas a sistemas reais*. [S.l.]: Editora UFMG, 2004.
- AJANOVIC, Amela. Biofuels versus food production: does biofuels production increase food prices? *Energy*, Elsevier, v. 36, n. 4, p. 2070–2076, 2011.
- ANFAVEA, Estatísticas. *Associação Nacional de Fabricantes de Veículos Automotores*. 2018.
- ANTENEODO, Celia. Processos estocásticos. *V Escola do Centro Brasileiro de Pesquisas Físicas*, 2004.
- ANTONAKAKIS, Nikolaos *et al.* The relationship between commodity markets and commodity mutual funds: A wavelet-based analysis. *Finance Research Letters*, Elsevier, v. 24, p. 1–9, 2018.
- ASSOCIATION, Renewable Fuels *et al.* *Online at www. ethanolrfa.org/industry/statistics/# B, (as of March 10, 2018)*. 2018.
- BANK, World. Commodity markets outlook. In: WASHINGTON. [S.l.], 2017.
- BARROS, Carlos P; GIL-ALANA, Luis A; WANKE, Peter. Ethanol consumption in brazil: Empirical facts based on persistence, seasonality and breaks. *Biomass and Bioenergy*, Elsevier, v. 63, p. 313–320, 2014.
- BEKIROU, Stelios *et al.* Information diffusion, cluster formation and entropy-based network dynamics in equity and commodity markets. *European Journal of Operational Research*, Elsevier, v. 256, n. 3, p. 945–961, 2017.
- BELKE, Ansgar; BORDON, Ingo G; VOLZ, Ulrich. Effects of global liquidity on commodity and food prices. *World Development*, Elsevier, v. 44, p. 31–43, 2013.
- BENEDETTO, Francesco *et al.* *Maximum entropy estimator for the predictability of energy commodity market time series*. [S.l.], 2014.
- BENITES-LAZARO, LL; MELLO-THÉRY, NA; LAHSEN, M. Business storytelling about energy and climate change: The case of brazil’s ethanol industry. *Energy Research & Social Science*, Elsevier, v. 31, p. 77–85, 2017.
- BOX, George EP; JENKINS, Gwilym M. Time series analysis: forecasting and control, rev. edn. *Holder-Day, San Francisco, CA, USA*, 1976.
- BRESSER-PEREIRA, Luiz Carlos. *Doença holandesa e indústria*. [S.l.]: Editora FGV, 2010.
- BROWN, Robert Goodell. *Statistical forecasting for inventory control*. [S.l.]: McGraw/Hill, 1959.
- BROWN, Robert G. Smoothing. *Forecasting and Prediction of Discrete Time Series*, Prentice-Hall Englewood Cliffs, NJ, p. 238, 1963.

- CAVALCANTI, Marcelo; SZKLO, Alexandre; MACHADO, Giovani. Do ethanol prices in brazil follow brent price and international gasoline price parity? *Renewable Energy*, Elsevier, v. 43, p. 423–433, 2012.
- CEPEA, DATA. *Centro de Estudos Avançados em Economia Aplicada*. 2018.
- DAVID, Sergio Adriani *et al.* Fractional dynamic behavior in ethanol prices series. *Journal of Computational and Applied Mathematics*, Elsevier, v. 339, p. 85–93, 2018.
- EHLERS, RS. Apostila análise de séries temporais, 5a edição. *UFPR, Curitiba, 94p*, 2009.
- ENDERS, W. *Applied econometric time series*. New York (USA) Wiley, 2003.
- FAO, UN. *Commodities and development report 2017: Commodity markets, economic growth and development*. In: NEW YORK - GENEVA. [S.l.], 2017.
- FONSECA, Jairo Simon da; MARTINS, Gilberto de Andrade; TOLEDO, G S. *Estatística Aplicada*. [S.l.]: São Paulo: Atlas, 1985. v. 6.
- FREITAS, Luciano Charlita De; KANEKO, Shinji. Ethanol demand under the flex-fuel technology regime in brazil. *Energy Economics*, Elsevier, v. 33, n. 6, p. 1146–1154, 2011.
- GOMEZ, José MA; LEGEY, Luiz FL. An analysis of the impact of flex-fuel vehicles on fuel consumption in brazil, applying cointegration and the kalman filter. *Energy*, Elsevier, v. 81, p. 696–705, 2015.
- GUJARATI, Damodar N; PORTER, Dawn C. *Econometria Básica-5*. [S.l.]: Amgh Editora, 2011.
- HE, Ling-Yun; CHEN, Shu-Peng. Nonlinear bivariate dependency of price–volume relationships in agricultural commodity futures markets: a perspective from multifractal detrended cross-correlation analysis. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 390, n. 2, p. 297–308, 2011.
- HERRERA, Rodrigo; RODRIGUEZ, Alejandro; PINO, Gabriel. Modeling and forecasting extreme commodity prices: A markov-switching based extreme value model. *Energy Economics*, Elsevier, v. 63, p. 129–143, 2017.
- HOLT, CC. *Forecasting trends and seasonals by exponentially weighted averages*. *Carnegie Institute of Technology*. [S.l.], 1957.
- HORNIK, Kurt. The comprehensive r archive network. *Wiley Interdisciplinary Reviews: Computational Statistics*, Wiley Online Library, v. 4, n. 4, p. 394–398, 2012.
- HUANG, Jikun *et al.* Biofuels and the poor: Global impact pathways of biofuels on agricultural markets. *Food Policy*, Elsevier, v. 37, n. 4, p. 439–451, 2012.
- HYNDMAN, Rob J; KOEHLER, Anne B. Another look at measures of forecast accuracy. *International journal of forecasting*, Elsevier, v. 22, n. 4, p. 679–688, 2006.
- JARROW, Robert A; CHATTERJEA, Arkadev. *An introduction to derivative securities, financial markets, and risk management*. [S.l.]: WW Norton & Company, 2013.
- JIA, Rui-Lin *et al.* Correlation between agricultural markets in dynamic perspective—evidence from china and the us futures markets. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 464, p. 83–92, 2016.

- JR, Everette S Gardner. Exponential smoothing: The state of the art—part ii. *International journal of forecasting*, Elsevier, v. 22, n. 4, p. 637–666, 2006.
- KABLAN, Sandrine; FTITI, Zied; GUESMI, Khaled. Commodity price cycles and financial pressures in african commodities exporters. *Emerging Markets Review*, Elsevier, v. 30, p. 215–231, 2017.
- KARYOTIS, Catherine; ALIJANI, Sharam. Soft commodities and the global financial crisis: implications for the economy, resources and institutions. *Research in International Business and Finance*, Elsevier, v. 37, p. 350–359, 2016.
- KOHZADI, Nowrouz *et al.* A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, Elsevier, v. 10, n. 2, p. 169–181, 1996.
- LAGO, Jesus; RIDDER, Fjo De; SCHUTTER, Bart De. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, Elsevier, v. 221, p. 386–405, 2018.
- LAGOARDE-SEGOT, Thomas. Financialization: Towards a new research agenda. *International Review of Financial Analysis*, Elsevier, v. 51, p. 113–123, 2017.
- LAHMIRI, Salim; UDDIN, Gazi Salah; BEKIROS, Stelios. Clustering of short and long-term co-movements in international financial and commodity markets in wavelet domain. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 486, p. 947–955, 2017.
- LEE, Cheng-Ming; KO, Chia-Nan. Short-term load forecasting using lifting scheme and arima models. *Expert Systems with Applications*, Elsevier, v. 38, n. 5, p. 5902–5911, 2011.
- LI, Zhihui; LU, Xinsheng. Multifractal analysis of china’s agricultural commodity futures markets. *Energy Procedia*, Elsevier, v. 5, p. 1920–1926, 2011.
- LOF, Matthijs; NYBERG, Henri. Noncausality and the commodity currency hypothesis. *Energy Economics*, Elsevier, v. 65, p. 424–433, 2017.
- LÜBBERS, Johannes; POSCH, Peter N. Commodities’ common factor: An empirical assessment of the markets’ drivers. *Journal of Commodity Markets*, Elsevier, v. 4, n. 1, p. 28–40, 2016.
- MADDALA, Gangadharrao Soundalyarao; WELLER, Leonardo. *Introdução à econometria*. [S.l.]: LTC, 2003.
- MAKRIDAKIS, Spyros; HYNDMAN, Rob J; WHEELWRIGHT, Steven C. *Forecasting: Methods and Applications*. [S.l.]: John Wiley & Sons, Inc., 1998.
- MATOS, Orlando Carneiro de. *Econometria basica: teoria e aplicacoes*. [S.l.]: Atlas, 2000.
- MONTASSER, Ghassen El *et al.* Are there multiple bubbles in the ethanol–gasoline price ratio of brazil? *Renewable and Sustainable Energy Reviews*, Elsevier, v. 52, p. 19–23, 2015.
- MONTEIRO, Roberto LS *et al.* A model for improving the learning curves of artificial neural networks. *PloS one*, Public Library of Science, v. 11, n. 2, p. e0149874, 2016.

- MONTGOMERY, Douglas C; JENNINGS, Cheryl L; KULAHCI, Murat. *Introduction to time series analysis and forecasting*. [S.l.]: John Wiley & Sons, 2015.
- MOREIRA, Ricardo Ramallete. Commodities prices volatility, expected inflation and gdp levels: an application for a net-exporting economy. *Procedia Economics and Finance*, Elsevier, v. 14, p. 435–444, 2014.
- MORETTIN, Pedro Alberto. *Econometria financeira: um curso em séries temporais financeiras*. [S.l.]: Edgard Blücher, 2008.
- MORETTIN, Pedro A; TOLOI, Clélia. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2004.
- OLIVEIRA, Marcelo E Dias De; VAUGHAN, Burton E; RYKIEL, Edward J. Ethanol as fuel: energy, carbon dioxide balances, and ecological footprint. *AIBS Bulletin*, American Institute of Biological Sciences, v. 55, n. 7, p. 593–602, 2005.
- ORDU, Beyza Mina; ORAN, Adil; SOYTAS, Ugur. Is food financialized? yes, but only when liquidity is abundant. *Journal of Banking & Finance*, Elsevier, v. 95, p. 82–96, 2018.
- PAL, Debdatta; MITRA, Subrata K. Time-frequency contained co-movement of crude oil and world food prices: A wavelet-based analysis. *Energy Economics*, Elsevier, v. 62, p. 230–239, 2017.
- PELLEGRINI, Fernando R; FOGLIATTO, Flávio S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. *Production*, SciELO Brasil, v. 11, n. 1, p. 43–64, 2001.
- PEREIRA, BB. Séries temporais multivariadas. *VI Simpósio Nacional de Probabilidade e Estatística, Rio de Janeiro, 218p Google Scholar*, 1984.
- PEREIRA, Eder Johnson de Area Leão; SILVA, Marcus Fernandes da; PEREIRA, HBB. Econophysics: Past and present. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 473, p. 251–261, 2017.
- PROJECT.ORG [www.r](http://www.r-project.org). *Online at www.r-project.org*. 2018.
- SHALINI, Velappan; PRASANNA, Krishna. Impact of the financial crisis on indian commodity markets: Structural breaks and volatility dynamics. *Energy Economics*, Elsevier, v. 53, p. 40–57, 2016.
- SOUZA, RC; CAMARGO, ME. *Análise de Séries Temporais: os modelos ARIMA*. [S.l.]: Gráfica e Editora Regional, Rio de Janeiro, 2004.
- SOUZA, Reinaldo Castro. *Modelos estruturais para previsão de séries temporais: abordagens clássica e bayesiana*. [S.l.]: IMPA, 1989.
- TADASSE, Getaw *et al*. Drivers and triggers of international food price spikes and volatility. In: *Food Price Volatility and Its Implications for Food Security and Policy*. [S.l.]: Springer, 2016. p. 59–82.
- TOLOI, CMC; MORETTIN, PA. Previsão de séries temporais. *São Paulo: Atual Editora*, 1987.
- TRIANA, Carlos Ariel Ramírez. Energetics of brazilian ethanol: Comparison between assessment approaches. *Energy Policy*, Elsevier, v. 39, n. 8, p. 4605–4613, 2011.

ULUSOY, Veysel; ONBIRLER, Özgür Ünal. Marginal speculation and hedging in commodity markets. *Finance Research Letters*, Elsevier, v. 23, p. 269–282, 2017.

UNCTAD. The state of commodity dependence. In: UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT GENEVA. [S.l.], 2016.

VEIGA, RD; SÁFADI, T. Análise de regressão e séries temporais. *Lavras: Universidade Federal de Lavras*, 1999.

VERÍSSIMO, Michele Polline; XAVIER, Clésio Lourenço. Tipos de commodities, taxa de câmbio e crescimento econômico: evidências da maldição dos recursos naturais para o brasil. *Revista de Economia Contemporânea*, v. 18, n. 2, p. 267–295, 2014.

WHEELWRIGHT, Steven C; MAKRIDAKIS, Spyros G *et al.* Forecasting methods for management. Wiley, 1985.

WILES, Phoebe S; ENKE, David. Nonlinear modeling using neural networks for trading the soybean complex. *Procedia Computer Science*, Elsevier, v. 36, p. 234–239, 2014.

WINTERS, Peter R. Forecasting sales by exponentially weighted moving averages. *Management science*, INFORMS, v. 6, n. 3, p. 324–342, 1960.

XIONG, Tao *et al.* A combination method for interval forecasting of agricultural commodity futures prices. *Knowledge-Based Systems*, Elsevier, v. 77, p. 92–102, 2015.