CAIO DE ALMEIDA CAMILLI

MODELOS DE PREVISÃO DE PREÇOS SUCROALCOOLEIROS

São Paulo

2018

CAIO DE ALMEIDA CAMILLI

MODELOS DE PREVISÃO DE PREÇOS SUCROALCOOLEIROS

Trabalho de Formatura apresentado na Escola Politécnica da Universidade de São Paulo para a obtenção do diploma de Engenheiro de Produção

Orientadora: Profª. Drª. Celma de Oliveira Ribeiro

São Paulo

2018

Catalogação-na-publicação

Camilli, Caio de Almeida
Modelos de previsão de preços sucroalcooleiros / C. A. Camilli -- São Paulo, 2018.
120 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Análise de séries temporais 2.Cana-de-açúcar 3.ARIMA 4.Filtros de Kalman 5.Redes neurais I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

À minha família e amigos.

AGRADECIMENTOS

À professora Celma de Oliveira Ribeiro pela orientação e preciosos conselhos na concepção e realização desse trabalho.

A Victoria Morgado Mutran pelo acompanhamento no início do trabalho, pelo direcionamento dado na escolha do tema e pelas numerosas fontes bibliográficas indicadas.

À Escola Politécnica da USP pela inestimável formação fornecida, que me preparou para a vida profissional e acadêmica.

À École Centrale de Marseille por ter me recebido durante dois anos e contribuído no meu desenvolvimento intelectual e na expansão de meus horizontes. Ao governo francês e ao Campus France por terem me acompanhado e fornecido auxílio financeiro nessa jornada.

A meus pais, Alberto e Nilza Camilli, pelo suporte incondicional e encorajamentos que me dão forças para enfrentar a vida e que são indispensáveis para todas as minhas conquistas.

À minha companheira Solène, pelo apoio emocional e experiências compartilhadas que me tornam uma pessoa melhor.

À Hermès e a Romain Severini, que foi o melhor chefe que eu poderia ter tido na minha primeira experiência profissional e me ensinou muito sobre liderança e o mundo do trabalho.

À Criteo e a meus colegas de trabalho atuais, que criaram um ambiente de desafios e aprendizados no qual encontrei a minha vocação profissional.

A meus amigos, de infância e de faculdade, sem os quais a vida não tem graça.

"O presente é tão grande, não nos afastemos.

Não nos afastemos muito, vamos de mãos dadas"

Carlos Drummond de Andrade (1902-1987)

RESUMO

O setor sucroalcooleiro é de importância fundamental para a economia brasileira, contribuindo para mais de 2% do PIB em 2017 e gerando cerca de 800 mil empregos em território brasileiro. No entanto, dezenas de usinas fecharam as portas nos últimos anos devido a problemas financeiros, para os quais contribuem incertezas e flutuações nos preços de açúcar e etanol.

Este trabalho desenvolve modelos estatísticos para prever preços de derivados de cana-de-açúcar para um horizonte futuro de até três meses. São estudadas três classes de modelo: SARIMA/VAR, filtros de Kalman e redes neurais recorrentes do tipo LSTM. Esses modelos são fundamentados teoricamente, é retratado o cenário atual da indústria sucroalcooleira e são testadas as capacidades preditivas de uma série de variáveis externas, incluindo contratos futuros em bolsas de valores, preços de combustível no território nacional, índices internacionais de preços de petróleo bruto e taxas de câmbio.

Os modelos são formulados explicitamente e implementados em Python e R. Sua performance é avaliada e comparada utilizando validação cruzada e métricas clássicas da área de aprendizado estatístico, incluindo um modelo de persistência para referência (modelo de base). O melhor modelo obtém reduções de 25% a 40% nas taxas de erro para todos os horizontes e produtos estudados, mostrando potencial para ser utilizado em situações reais de decisões de *mix* de produção e gestão de estoques em uma usina de cana-de-açúcar.

Palavras-chave: cana-de-açúcar, SARIMA, VAR, filtro de Kalman, redes neurais, preços, previsão

ABSTRACT

The sugarcane sector is essential to the Brazilian economy, accounting for more than 2% of its GDP in 2017 and creating hundreds of thousands of jobs. However, dozens of plants were closed on recent years due to financial problems, to which contribute uncertainties and fluctuations on ethanol and sugar prices.

This work develops statistical models to forecast prices of sugarcane derivatives up to three months ahead. Three classes of models are studied: SARIMA/VAR, Kalman filters and LSTM recurrent neural networks. These models are grounded theoretically, a scenario of the sugarcane industry is depicted and several exogenous variables are tested for predictive power, among which are futures in the financial market, fuel prices in the Brazilian market, crude oil international indices and exchange rates.

The models are then explicitly formulated and implemented using Python and R. Their performance is assessed using cross-validation and classical metrics of statistical learning, including a persistence model for reference (baseline model). The best model achieves improvements between 25% and 40% on error rates for all products and forecast horizons, showing potential to be used in real-life decisions of production mix and stock management in sugarcane plants.

Keywords: sugarcane, SARIMA, VAR, Kalman filter, neural networks, prices, forecast

LISTA DE FIGURAS

Figura 1: Imagens das três principais funções neurais 33
Figura 2: Representação de um neurônio de uma rede neural 33
Figura 3: Rede neural *feedforward* 3-5-2-1 34
Figura 4: Representação de uma relação de recorrência por um diagrama de circuito 36
Figura 5: Representação da mesma relação de recorrência na forma de um grafo computacional aberto 36
Figura 6: Redes neurais recursivas com uma única saída. 37
Figura 7: Representação gráfica das equações de uma rede LSTM 40
Figura 8: Evolução histórica da produção de cana-de-açúcar no Brasil, por safra 46
Figura 9: *Mix* de produção das usinas sucroalcooleiras da Região Centro-Sul desde 2003/04 49
Figura 10: Preços mensais de etanol hidratado no estado de SP 56
Figura 11: Preços mensais de açúcar cristal no estado de SP 57
Figura 12: Preços de açúcar e etanol corrigidos com relação à inflação 58
Figura 13: Séries de etanol hidratado e de primeiros futuros da BM&F 61
Figura 14: CCF entre as séries diferenciadas de preços de etanol e contratos futuros de primeira expiração da BM&F 62
Figura 15: Séries de açúcar cristal e de primeiros futuros de açúcar da bolsa de Chicago 63
Figura 16: CCFs das séries de preços de açúcar cristal e futuros de açúcar de Chicago 64
Figura 17: CCFs de preços de gasolina comum com preços de etanol e de açúcar 66
Figura 18: CCFs de preços de óleo diesel com preços de etanol e de açúcar 66
Figura 19: CCFs de preços de GNV com preços de etanol e de açúcar 67
Figura 20: CCFs de preços de GLP com preços de etanol e de açúcar 67
Figura 21: CCFs de preços de petróleo Brent com preços de etanol e de açúcar 68
Figura 22: CCFs de preços de gasolina WTI com preços de etanol e de açúcar 69
Figura 23: CCFs de taxas de câmbio (dólar) com preços de etanol e de açúcar 69
Figura 24: ACF e PACF da série mensal de preços de açúcar sem diferenciação 73
Figura 25: ACF e PACF da série de preços de açúcar após diferenciação 74
Figura 26: Resíduos SARIMA normalizados para a série de preços de açúcar 76
Figura 27: ACF dos resíduos SARIMA para a série de preços de açúcar 77
Figura 28: Gráfico QQ para os resíduos SARIMA (açúcar) 77

Figura 29: Previsões SARIMA e valores reais da série de preços de açúcar 79
Figura 30: ACF e PACF da série de preços de etanol sem diferenciação 80
Figura 31: ACF e PACF da série de preços de etanol após diferenciação 81
Figura 32: Resíduos SARIMA normalizados para a série de preços de etanol 83
Figura 33: Gráfico QQ para os resíduos SARIMA (etanol) 83
Figura 34: ACF dos resíduos SARIMA para a série de preços de açúcar 84
Figura 35: Previsões SARIMA e valores reais da série de preços de etanol 85
Figura 36: Resíduos VAR normalizados para as séries de etanol e açúcar 88
Figura 37: Gráficos QQ dos resíduos VAR para as séries de etanol e açúcar 88
Figura 38: ACF dos resíduos VAR(2) para as séries de etanol e açúcar 89
Figura 39: Previsões do modelo VAR para as séries de preços de etanol e açúcar 90
Figura 40: Resíduos normalizados do modelo de espaço de estados 95
Figura 41: Gráfico QQ para os resíduos do modelo de espaço de estados 95
Figura 42: ACF dos resíduos do modelo de espaço de estados 96
Figura 43: Previsões realizadas pelo modelo de espaço de estados e valores reais 97
Figura 44: Evolução dos erros de treino e de validação com o treinamento da RNN 99
Figura 45: Previsões do modelo ótimo de redes neurais LSTM 101
Figura 46: Gráfico QQ dos resíduos do modelo LSTM. 101
Figura 47: ACF dos resíduos do modelo LSTM 102

LISTA DE TABELAS

Tabela 1: Maiores produtores mundiais de cana-de-açúcar em 2016 15
Tabela 2: Evolução da produção de cana-de-açúcar no Brasil, por região. 47
Tabela 3: Produção de cana-de-açúcar, açúcar e etanol por estado da Região Centro-Sul 47
Tabela 4: Estados brasileiros com o maior número de usinas de cana-de-açúcar. 48
Tabela 5: Evolução do número de usinas sucroalcooleiras na região Centro-Sul 48
Tabela 6: Estatísticas descritivas para as séries de açúcar e etanol 58
Tabela 7: Coeficientes estimados e valores de AIC e BIC para o Modelo 1 75
Tabela 8: Coeficientes estimados e valores de AIC e BIC para o Modelo 2 75
Tabela 9: Coeficientes estimados e valores de AIC e BIC para o Modelo 3 75
Tabela 10: Análise de resíduos para o modelo escolhido (Modelo 3) 77
Tabela 11: Previsões de preços de açúcar realizadas pelo modelo SARIMA 78
Tabela 12: Coeficientes estimados e valores de AIC e BIC para o Modelo 1 82
Tabela 13: Coeficientes estimados e valores de AIC e BIC para o Modelo 2 82
Tabela 14: Coeficientes estimados e valores de AIC e BIC para o Modelo 3 82
Tabela 15: Análise de resíduos para o modelo SARIMA$(0, 1, 0) \times (1, 0, 1)_6$ 84
Tabela 16: Previsões de preços de etanol realizadas pelo modelo SARIMA 85
Tabela 17: Valores de AIC e BIC para diferentes ordens de VAR 86
Tabela 18: Coeficientes estimados para o modelo VAR(2) 87
Tabela 19: Resultado das previsões VAR para os preços de açúcar e etanol 90
Tabela 20: Resultado da estimativa dos parâmetros do filtro de Kalman 94
Tabela 21: Resultado do modelo de espaço de estados para previsões um mês no futuro 96
Tabela 22: Resultados da rede neural LSTM na previsão dos preços de etanol e açúcar. 100
Tabela 23: Resultados consolidados dos modelos para previsões um mês no futuro 104
Tabela 24: Resultados consolidados dos modelos para previsões dois meses no futuro 104
Tabela 25: Resultados consolidados dos modelos para previsões três meses no futuro 104

LISTA DE SIGLAS E ABREVIATURAS

ACF	*Autocorrelation Function* (Função de Autocorrelação)
AIC	*Akaike Information Criterion* (Critério de Informação de Akaike)
ANN	*Artificial Neural Network* (Rede Neural Artificial)
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
BIC	*Bayes Information Criterion* (Critério de Informação de Bayes)
Cepea	Centro de Estudos Avançados em Economia Aplicada
CCF	*Cross Correlation Function* (Função de Correlação Cruzada)
FAO	*Food and Agriculture Organization*
IBGE	Instituto Brasileiro de Geografia e Estatística
LSTM	*Long Short-Term Memory*
MAE	*Mean Absolute Error* (Erro Médio Absoluto)
MAPE	*Mean Absolute Percentage Error* (Erro Médio Percentual Absoluto)
PACF	*Partial Autocorrelation Function* (Função de Autocorrelação Parcial)
PIB	Produto Interno Bruto
RMSE	*Root Mean Squared Error* (Raíz Quadrada do Erro Médio Quadrado)
RNN	*Recurrent Neural Network* (Rede Neural Recorrente)
SARIMA	*Seasonal Autoregressive Integrated Moving Average*
SQR	Soma dos Quadrados dos Resíduos
SQT	Soma dos Quadrados Total
UNICA	União da Indústria da Cana-de-Açúcar
VAR	*Vector Autoregression*

SUMÁRIO

1. INTRODUÇÃO 15
1.1. Importância do setor sucroalcooleiro para o Brasil 15
1.2. O problema do mix de produção 16
1.3. Objetivo do trabalho 16
1.4. Estrutura do trabalho 17
2. REVISÃO BIBLIOGRÁFICA 19
2.1. Regressão linear 19
2.1.1. Validação do modelo 20
2.2. Modelos de séries temporais 21
2.2.1. Definição e conceitos básicos 21
2.2.2. Modelos AR, MA e séries integradas (I) 22
2.2.3. Modelos ARIMA (p,d,q): autorregressivos, integrados e de média móvel 23
2.2.4. Modelos SARIMA (p,d,q)x(P,D,Q)$_S$: adição de sazonalidade 24
2.2.5. Construção de modelos ARIMA e SARIMA 25
2.2.6. Modelos VAR(p): Extensão multivariada dos modelos ARIMA 28
2.2.7. Teste de causalidade de Granger 28
2.3. Modelos de espaço de estado e filtros de Kalman 29
2.3.1. Filtro de Kalman 30
2.3.2. Estimação de parâmetros do modelo de espaço de estado 31
2.4. Redes neurais artificiais 31
2.4.1. Redes neurais *feedforward* 33
2.4.2. Redes neurais recorrentes 35
2.4.3. Metodologia prática 41
2.5. Modelagem de preços na literatura 42

3. INDÚSTRIA SUCROALCOOLEIRA NO BRASIL 45
3.1. História 45
3.2. Cenário atual 46
3.2.1. Regiões produtoras e volumes produzidos 46
3.2.2. Número de usinas e distribuição geográfica 47
3.2.3. *Mix* de produção e participação em exportações 49
3.3. Produção e distribuição 49
3.3.1. Plantio e colheita de cana-de-açúcar 49
3.3.2. Processamento da cana nas usinas 51
3.3.3. Produção de açúcar e etanol 51
3.3.4. Cogeração e excedente de energia 52
3.3.5. Cadeias de distribuição 52
4. DADOS E METODOLOGIA 55
4.1. Séries de preços de açúcar e etanol 55
4.1.1. Tratamento e exploração dos dados 57
4.2. Mercados futuros 59
4.2.1. Motivação 59
4.2.2. Contratos futuros de etanol hidratado 60
4.2.3. Contratos futuros de açúcar 63
4.3. Outras variáveis exógenas 65
4.3.1. Preços de combustíveis em São Paulo 65
4.3.2. Preços de petróleo internacionais 68
4.3.3. Dólar 69
4.4. Modelos a serem desenvolvidos no trabalho 70
4.4.1. Métricas de avaliação dos modelos 70
4.4.2. Separação de dados de treino e de teste 71
4.4.3. Ambiente de desenvolvimento 72

4.4.4. Correção com relação à inflação e transformação logarítmica 72
5. CONSTRUÇÃO DOS MODELOS E PREVISÕES 73
5.1. Modelos de séries temporais 73
5.1.1. SARIMA - Série de preços de açúcar 73
5.1.2. SARIMA - Série de preços de etanol 80
5.1.3 VAR - Modelagem conjunta de preços de etanol e de açúcar 86
5.2. Modelo de espaço de estados 91
5.2.1. Descrição 91
5.2.2. Cálculo do índice de sazonalidade para o etanol 93
5.2.3. Estimativa dos parâmetros e análise de resíduos 94
5.2.4. Previsões 96
5.3. Modelo de redes neurais LSTM 98
5.3.1. Arquitetura do modelo e variáveis de decisão 98
5.3.2. Previsões 100
5.3.3. Resíduos 101
5.4. Horizontes futuros e consolidação 103
5.4.1. Previsões até três meses no futuro 103
5.4.2. Discussão dos resultados 103
6. CONCLUSÃO 107
REFERÊNCIAS BIBLIOGRÁFICAS 109
ANEXO: CÓDIGO FONTE DO MODELO ESCOLHIDO 115

1. INTRODUÇÃO

1.1. Importância do setor sucroalcooleiro para o Brasil

O setor sucroalcooleiro ocupa um papel central na economia brasileira. Em 2017, a cadeia produtiva de cana-de-açúcar contribuiu com 156 bilhões de reais ao PIB brasileiro (2.4% do PIB total), alimentando uma produção de mais de 37 milhões de toneladas de açúcar e 26 milhões de metros cúbicos de etanol, segundo estimativas do Cepea/USP. Por ser o maior produtor mundial de cana-de-açúcar (Tabela 1), o Brasil é também o maior produtor e exportador de açúcar do mundo e o segundo maior produtor de etanol, atrás apenas dos Estados Unidos, que utilizam o milho como matéria-prima.

País	Produção (mil toneladas)
Brasil	666 823
Índia	348 448
China*	122 664
Tailândia*	87 468
Paquistão	65 451
México	56 447
Colômbia*	36 951
Austrália	34 403
Guatemala	33 533
Estados Unidos	29 926

Tabela 1: Maiores produtores mundiais de cana-de-açúcar em 2016 (* = produção estimada). Fonte: FAO/ONU

Existem atualmente 411 usinas de cana-de-açúcar em atividade no Brasil, empregando cerca de 800 mil pessoas ao longo da cadeia produtiva (ÚNICA, 2018).

1.2. O problema do mix de produção

É papel dos gestores das usinas sucroalcooleiras decidir o que será produzido após a colheita. A cana-de-açúcar pode ser transformada em três produtos principais:

1. açúcar, produzido tanto para exportação quanto para o mercado interno;
2. álcool hidratado, que possui em sua composição de 95,1% a 96% de etanol e o resto em água. É vendido em postos de gasolina como combustível;
3. álcool anidro, que possui no mínimo 99,6% de álcool em sua composição. É misturado à gasolina em proporções de cerca de 20% e também pode ser utilizado na fabricação de tintas, solventes e outros produtos químicos

Essa decisão é difícil, pois condições climáticas, instabilidades de políticas governamentais e oscilações nos preços criam incertezas que impactam as usinas. Nesse cenário, é importante desenvolver ferramentas para reduzir o risco financeiro ligado a essas incertezas. Duas abordagens são possíveis:

1. aproveitar a flexibilidade no mix de produção para reduzir o risco, utilizando teorias financeiras de gestão de portfólio, como fazem Rockafellar e Uryasev (2000);
2. utilizar o mercado de futuros e derivativos como cobertura, estocando uma parte da produção para ser vendida em momentos subsequentes, como fazem Ribeiro e Oliveira (2018)

Entre 2008 e 2014, 83 usinas sucroalcooleiras fecharam as suas portas em razão de problemas financeiros, segundo a consultoria Datagro. O fechamento de usinas resulta na perda de milhares de empregos e na diminuição da capacidade de processamento de cana no país em milhões de toneladas.

1.3. Objetivo do trabalho

As duas estratégias citadas necessitam de previsões de preços para o futuro próximo. A partir delas, é possível definir um *mix* de produção que forneça um retorno financeiro esperado, ao mesmo tempo em que se minimiza o risco, entendido como a variância dos retornos possíveis.

O objetivo desse trabalho é modelar matematicamente as séries de preços de açúcar e etanol, identificando padrões históricos e variáveis correlacionadas de forma a realizar

previsões de preços em um horizonte de até três meses no futuro. Esses modelos poderão ser usados para dar suporte a decisões de *mix* de produção e gestão de estoques em uma usina de cana-de-açúcar.

Serão testadas e comparadas três classes de modelos:

1. Modelos SARIMA/VAR, que utilizam os valores passados das próprias séries para prever a sua evolução futura;
2. Modelos de espaço de estado, que permitem a inclusão de outras variáveis, ditas exógenas, e dão maior controle ao analista;
3. Modelos de redes neurais recorrentes do tipo LSTM (*Long Short-Term Memory*), que oferecem uma abordagem alternativa aos modelos estatísticos clássicos pois se baseiam em técnicas de aprendizado de máquina (*machine learning*)

Para isso, será fundamental a experiência acumulada pelo autor em seu estágio como analista de *big data* em uma grande empresa de tecnologia (Criteo). A análise e modelagem de séries temporais utilizando linguagens de programação e conceitos estatísticos é uma das suas principais tarefas e esse trabalho pode ser considerado uma extensão de suas atividades profissionais para um ambiente acadêmico.

1.4. Estrutura do trabalho

O segundo capítulo contém a revisão bibliográfica, na qual são estudadas as bases teóricas para a compreensão dos modelos e é realizada uma exploração da literatura existente sobre a modelagem de preços no geral e mais especificamente no domínio sucroenergético.

O terceiro capítulo apresenta uma breve história da implantação da indústria sucroalcooleira no Brasil e traça o cenário atual da distribuição geográfica das usinas, técnicas de produção, cadeias de distribuição e mercados atendidos.

O capítulo quatro detalha o escopo do trabalho, as fontes dos dados utilizados e a metodologia de treinamento e avaliação dos modelos. Nele é também feito um estudo para definir a inclusão de outras variáveis na modelagem, entre as quais estão contratos futuros nas bolsas de Chicago e BM&F Bovespa, preços de combustíveis e derivados de petróleo no mercado nacional, índices internacionais de preços de petróleo bruto e taxas de câmbio.

No capítulo cinco são aplicados os modelos propostos, segundo as metodologias definidas no capítulo de revisão bibliográfica. Os resíduos dos modelos são analisados e realizam-se previsões para horizontes de até três meses no futuro. Os resultados são comparados para escolher o melhor modelo.

O sexto e último capítulo expõe a conclusão e considerações finais do autor.

2. REVISÃO BIBLIOGRÁFICA

2.1. Regressão linear

Das ferramentas de modelagem estatística, a regressão linear é a mais antiga, tendo sido desenvolvida por Sir Francis Galton no final do século XIX em um artigo que estudava a transmissão genética de características entre flores (Stanton, 2001). O objetivo da regressão linear é quantificar uma relação de causalidade entre variáveis independentes (entradas) e uma variável dependente (saída).

Segundo Hastie et al. (2009), o modelo de regressão pode ser escrito da seguinte forma:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j + \varepsilon ,$$

onde $X_1, X_2, \dots, X_p \in \mathbb{R}$ são as variáveis independentes, $Y \in \mathbb{R}$ representa a variável dependente e ε é uma variável aleatória que representa os erros cometidos na aproximação, chamados de resíduos. As seguintes hipóteses são feitas:

1. Existe uma relação linear entre as entradas e a saída;
2. As variáveis X_j são distribuídas normalmente, com pouca ou nenhuma multicolinearidade;
3. As observações (X, Y) são independentes entre si (não há autocorrelação);
4. Os resíduos ε possuem variância σ^2 constante, independente da magnitude de Y (homocedasticidade)

Os coeficientes β_0 e β_j devem ser estimados a partir de um conjunto de dados empíricos (x_1, y_1) ... (x_N, y_N), onde x_i é um vetor de p variáveis independentes que pode ser escrito como $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$. O método mais comum para estimar β é o método dos mínimos quadrados, que busca minimizar a soma dos quadrados dos resíduos (SQR) sobre o conjunto de observações empíricas. A função SQR é definida como:

$$SQR(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$$

Ainda segundo Hastie et al. (2009), para minimizar essa função, é possível passá-la para a forma matricial, em que X é uma matriz $N\ x\ (p+1)$ que contém as x_i observações empíricas acrescidas de uma coluna unitária, y = $(y_1, y_2, \dots, y_n)^T$ é um vetor coluna com as observações empíricas da variável dependente e $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$ é um vetor coluna com os coeficientes a serem estimados. Nessa notação, podemos reescrever a soma dos quadrados dos resíduos como:

$$SQR(\beta) = (y - X\beta)^T\ (y - X\beta)$$

Assumindo que a matriz X é invertível, para minimizar a SQR devemos anular sua primeira derivada com relação a β. A solução é:

$$\hat{\beta}\ =\ (X^T X)^{-1} X^T y,$$

e obtemos as seguintes estimativas para a variável dependente:

$$\hat{y}\ =\ X\hat{\beta}\ =\ X(X^T X)^{-1} X^T y$$

A variância σ^2 dos resíduos pode ser estimada pela fórmula:

$$\widehat{\sigma}^2 = \frac{SQR(\widehat{\beta})}{N-p-1} = \frac{1}{N-p-1}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2$$

2.1.1. Validação do modelo

Segundo James et al. (2013), uma questão importante no contexto de regressão linear é saber se existe realmente uma relação entre as variáveis independentes e a variável dependente; em outras palavras, determinar se algum dos coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ é diferente de zero. Para isso, é possível realizar um teste de hipóteses em que as hipóteses nula e alternativa são:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

H_1: pelo menos um dos β_j é diferente de zero

A estatística de teste para esse caso é o valor F dado pela equação:

$$F = \frac{(SQT - SQR)/p}{SQR/(N-p-1)}$$

na qual SQR é a soma dos quadrados dos resíduos $\sum_{i=1}^{N}(y_i - \widehat{y_i})^2$ e SQT é a soma dos quadrados totais na população observada $\sum_{i=1}^{N}(y_i - \overline{y_i})^2$. Se H_0 for verdadeira e os resíduos da regressão linear forem normalmente distribuídos, é possível mostrar que o valor F segue a distribuição F de Fisher-Snedecor, e a hipótese nula será rejeitada a um nível de significância α se $F > F_{p,N-p-1,\alpha}$.

2.2. Modelos de séries temporais

2.2.1. Definição e conceitos básicos

Segundo Shumway e Stoffer (2011), a análise de observações experimentais coletadas em momentos distintos introduz novos problemas de modelagem estatística, pois surgem entre pontos adjacentes dinâmicas de correlação que violam as hipóteses de independência amostral adotadas em métodos clássicos, como regressão linear. Modelos que levam em conta o ordenamento dos dados e a sua correlação temporal são chamados de modelos de séries temporais.

Dados de séries temporais são indexados com relação ao tempo, e normalmente são representados pela notação $x_1, x_2, \dots, x_T$, onde dado $t \in \mathbb{N}^*,\ t < T$, temos que a observação x_t sucedeu temporalmente a observação x_{t-1}, e assim por diante.

Um conceito importante no contexto de análise de séries temporais é o de estacionariedade. Uma série **estritamente estacionária** é aquela para qual a distribuição de probabilidades de uma coleção de valores $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$ é idêntica à da coleção deslocada no tempo $\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$ para todos os instantes $t_1, t_2, \dots, t_k$ e para todos os valores $h = 0, \pm1, \pm2, \dots$.

Uma definição menos restritiva é a de série **fracamente estacionária**, que é aquela para qual a esperança de seus elementos $\mu_t = E(x_t)$ é constante e não depende do tempo, e para qual o valor da função de autocovariância $\gamma(s,t) = E[(x_s - \mu_s)(x_t - \mu_t)]$ depende apenas da diferença $h = |s - t|$. Nesse caso, escreve-se $\mu_t = \mu$ e $\gamma(s,t) = \gamma(h)$.

Nesse trabalho, será adotada a convenção utilizada na maior parte dos textos em estatística, segundo a qual séries **fracamente estacionárias** são chamadas de **estacionárias**.

2.2.2. Modelos AR, MA e séries integradas (I)

2.2.2.1. Modelos autorregressivos (AR)

De acordo com Shumway e Stoffer (2011), os modelos autorregressivos tentam prever o valor da série em um instante t (x_t) em função de seus p valores anteriores $x_{t-1}, x_{t-2}, \dots, x_{t-p}$. O modelo de base AR($p$) assume a forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t,$$

onde x_t é uma série estacionária de média zero, $\phi_1, \phi_2, \dots, \phi_p$ são constantes, $\phi_p \neq 0$ e $w_t \sim N(0, {\sigma_w}^2)$, com ${\sigma_w}^2 > 0$. Se a média μ de x_t for não-nula, teremos:

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \cdots + \phi_p(x_{t-p} - \mu) + w_t$$

É conveniente no contexto de modelos ARIMA introduzir o operador de defasagem B, que satisfaz $x_{t-i} = B^i x_t$. Com a utilização desse operador, o modelo pode ser escrito como:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) x_t = w_t,$$

ou ainda:

$$\phi(B) x_t = w_t,$$

onde $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ é denominado operador de autorregressão de ordem *p*.

2.2.2.2. Modelos de média móvel (MA)

A ideia dos modelos de média móvel (MA) é similar à dos modelos AR, mas ao invés de expressar x_t como função de valores anteriores da série, a série é descrita como uma função dos q valores anteriores de w_t, que recebem o nome de inovações. O modelo MA de ordem q, MA(q), é escrito como:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q},$$

onde x_t é novamente uma série estacionária de média zero, $\theta_1, \theta_2, \dots, \theta_q$ são constantes, $\theta_q \neq 0$ e w_t uma série de variáveis normais identicamente distribuídas com média zero e variância

σ^2 (w_t é também chamado de ruído branco). Assim como no caso anterior, é possível escrever o modelo em termos do operador B:

$$x_t = \theta(B)w_t,$$

onde $\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$ é denominado operador de média móvel de ordem q.

2.2.2.3. Séries integradas (I)

Ainda segundo Shumway e Stoffer (2011), normalmente séries temporais não-estacionárias podem ser decompostas em dois componentes: um primeiro não-estacionário (tendência temporal) e um segundo estacionário de média zero. Para eliminar o primeiro componente e tornar a série estacionária, pode-se utilizar a operação de diferenciação, denotada por ∇, de forma que o resultado y_t, descrito pela equação abaixo, seja estacionário:

$$y_t = \nabla x_t = x_t - x_{t-1}$$

Se o resultado da operação ainda não for estacionário, o operador de diferenciação pode ser aplicado novamente, e assim por diante. Se uma série temporal torna-se estacionária após ter sido diferenciada d vezes, diz-se que ela é integrada de ordem d, ou I(d). Nesse caso, escreve-se $y_t = \nabla^{\mathrm{d}} x_t$, onde x_t é a série original e y_t é estacionária, e o operador ∇^{d} recebe o nome de operador de diferenciação de ordem *d*.

2.2.3. Modelos ARIMA (p,d,q): autorregressivos, integrados e de média móvel

A utilização de modelos ARIMA foi popularizada por Box e Jenkins (1970). Esses modelos são uma combinação dos três efeitos descritos anteriormente: a série é inicialmente diferenciada d vezes para se tornar estacionária e em seguida são aplicados o operador de autoregressão de ordem p sobre os valores da série x_t e o operador de média móvel de ordem q à série de inovações w_t. A descrição final do modelo, utilizando os operadores introduzidos nas seções anteriores, é:

$$\phi(B)\nabla^{\mathrm{d}} x_t = \theta(B)w_t,$$

em que x_t é uma série temporal qualquer (não necessariamente estacionária) , w_t é ruído branco de variância ${\sigma_w}^2 > 0$, $\nabla^{\mathrm{d}} = (1-B)^{\mathrm{d}}$, $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$ com $\phi_p \neq 0$, e $\theta(B) = \left(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q\right)$ com $\theta_q \neq 0$.

Por exemplo, um modelo $ARIMA(2,1,2)$ assume a forma explícita:

$$x_t = x_{t-1} + \phi_1(x_{t-1} - x_{t-2}) + \phi_2(x_{t-2} - x_{t-3}) + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2}$$

2.2.4. Modelos SARIMA (p,d,q)x(P,D,Q)$_S$: adição de sazonalidade

Shumway e Stoffer (2011) afirmam que modificações podem ser efetuadas sobre os modelos ARIMA para levar em conta comportamentos sazonais e não estacionários. Em algumas séries temporais observa-se uma forte dependência temporal em relação a um fator de sazonalidade *s*, que não pode ser modelado por um modelo ARIMA básico. Um exemplo é em dados mensais de séries macroeconômicas: os padrões temporais tendem a se repetir em ciclos de $s = 12$ meses, que correspondem a um ano do calendário. O mesmo fenômeno pode ocorrer com dados meteorológicos ou referentes a processos físicos ou biológicos.

Para levar em conta esse fator de sazonalidade, adicionam-se ao modelo ARIMA termos autorregressivos, de média móvel e de integração ligados diretamente ao período de sazonalidade *s*. Esses termos são expressos pelos seguintes operadores:

$$\phi_P(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \ldots - \phi_P B^{Ps}$$

$$\theta_Q(B^s) = 1 + \theta_1 B^s + \theta_2 B^{2s} + \ldots + \theta_Q B^{Qs}$$

$$\nabla_{\mathrm{s}}^{\mathrm{D}} x_t = (1 - B^s)^D x_t,$$

chamados respectivamente de **operador sazonal de autorregressão de ordem *P*** ($\phi_P(B^s)$), **operador sazonal de média móvel de ordem *Q*** ($\theta_Q(B^s)$) e **operador sazonal de integração de ordem *D*** ($\nabla_{\mathrm{s}}^{\mathrm{D}}$).

Incorporados ao modelo ARIMA, esses três operadores dão origem aos modelos SARIMA$(p,d,q)\ X\ (P,D,Q)_S$, regidos pela seguinte equação:

$$\phi(B)\phi_P(B^s)\nabla^{\mathrm{d}}\nabla_{\mathrm{s}}^{\mathrm{D}} x_t = \theta(B)\theta_Q(B^s)w_t,$$

onde $s \in \mathbb{N}^*, s > 1$ é o período de sazonalidade; $\phi(B), \phi_P(B^s), \theta(B), \theta_Q(B^s), \nabla^{\mathrm{d}}$ e $\nabla_{\mathrm{s}}^{\mathrm{D}}$ são os operadores introduzidos nas seções anteriores; w_t é uma série de ruído branco (também chamada de série de inovações) e x_t é a série modelada.

Por exemplo, um modelo SARIMA $(1,0,1)\ X\ (1,0,1)_{12}$ assume a forma explícita:

$$(x_t - \phi_{12}x_{t-12}) - \phi_1(x_{t-12} - \phi_{12}x_{t-13}) = (w_t + \theta_{12}w_{t-12}) + \theta_1(w_{t-1} + \theta_{12}w_{t-13})$$

Como o universo de modelos ARIMA/SARIMA inclui os modelos AR, MA e I descritos anteriormente, a seção seguinte fará referência exclusivamente a esses dois primeiros, embora se aplique a todos os outros.

2.2.5. Construção de modelos ARIMA e SARIMA

Box e Jenkins (1976) detalham as etapas básicas necessárias para construir modelos ARIMA e SARIMA em dados de séries temporais:

1. Inspeção visual dos dados para determinar se transformações matemáticas devem ser aplicadas para estabilizar a variância e definição da ordem *d* referente ao número de diferenciações necessárias para tornar a série estacionária;
2. Análise dos gráficos amostrais de autocorrelação (ACF) e de autocorrelação parcial (PACF) para determinar valores preliminares para as ordens *p* e *q* de autorregressão e de média móvel;
3. Dados os valores *p*, *q* e *d*, estimar os parâmetros ϕ_i e θ_j do modelo utilizando o método da máxima verossimilhança;
4. Escolha do modelo, utilizando AIC (Critério de Informação de Akaike) ou BIC (Critério de Informação de Bayes) para evitar sobreaprendizado;
5. Teste das inovações (resíduos) do modelo para homocedasticidade, normalidade e ausência de autocorrelação em série

Para compreender as etapas 2 e 4 são necessárias algumas definições suplementares, que serão realizadas a seguir.

2.2.5.1. Funções de autocorrelação (ACF) e autocorrelação parcial (PACF)

É possível definir para séries estacionárias a **função de autocorrelação** $\rho(h)$**,** dada pela seguinte fórmula:

$$\rho(h) = \frac{E[(x_{t+h} - \mu)(x_t - \mu)]}{\sqrt{E[(x_{t+h} - \mu)^2] * E[(x_t - \mu)^2]}}$$

Essa função, frequentemente abreviada como ACF (*Autocorrelation Function*), mede a dependência linear entre dois pontos de uma série observados em instantes diferentes, cuja defasagem é controlada pela variável h. Ela possui as seguintes propriedades:

1. $-1 \leq \rho(h) \leq 1$, para todo $h \in \mathbb{Z}$;
2. $\rho(0) = 1$, para toda série estacionária x_t;
3. $\rho(h) = \rho(-h)$, para todo $h \in \mathbb{Z}$

É possível provar que para uma série MA(q) a função de autocorrelação vale $\rho(h) = 0$ para $h > q$. Por isso, com a análise de ACF é possível determinar facilmente a ordem de um processo de média móvel. No entanto, o mesmo não vale para processos AR(p), cuja ACF decai lentamente e pode possuir valores estatisticamente significantes para $h > p$.

O desejo de encontrar uma função que permita facilmente determinar a ordem de um processo puramente autorregressivo leva à definição da função de autocorrelação parcial (PACF). A PACF busca eliminar a interferência de termos situados entre x_{t+h} e x_t nas suas correlações. Para isso, realizam-se as duas regressões lineares abaixo:

$$\hat{x}_{t+h} = \beta_1 x_{t-h-1} + \beta_2 x_{t-h-2} + \cdots + \beta_{h-1} x_{t+1}$$

$$\hat{x}_t = \beta_1 x_{t+1} + \beta_2 x_{t+2} + \cdots + \beta_{h-1} x_{t+h-1}$$

de forma que a definição da função de correlação parcial $\phi(h)$ para um processo estacionário é dada por:

$$\phi(h) = \begin{cases} \rho(1), se\ h = 1 \\ \dfrac{E[(x_{t+h} - \hat{x}_{t+h} - \mu)(x_t - \hat{x}_t - \mu)]}{\sqrt{E[(x_{t+h} - \hat{x}_{t+h} - \mu)^2] * E[(x_t - \hat{x}_t - \mu)^2]}}, se\ h > 1 \end{cases}$$

Para um processo AR(p), é possível mostrar que $\phi(h) = 0$ para $h > p$, como desejado. Já para processos MA(q), a PACF decai lentamente e pode possuir valores estatisticamente significantes mesmo para $h > q$.

Através da análise conjunta da ACF e da PACF, segundo Box e Jenkins (1976), é possível determinar se a série estudada se trata de um processo puramente AR, MA ou ARMA (o último ocorre se for observado um decaimento lento tanto da ACF quanto da PACF), e estimar as suas ordens *p* e *q*. Para encontrar sazonalidades de periodicidade *S*, basta analisar o comportamento da ACF e da PACF para valores de *h* múltiplos de *S*.

2.2.5.2. Critérios de informação de Akaike e Bayes

Em contextos de modelagem estatística, normalmente deseja-se criar um modelo que se adapte o melhor possível aos dados empíricos. Em um contexto de regressão linear, por exemplo, isso se traduz na minimização da soma dos quadrados dos resíduos SQR (ou, analogamente, da variância dos resíduos σ^2). No entanto, é possível mostrar que essas quantidades decrescem monotonicamente com o aumento do número de parâmetros *k* do modelo, de forma que um analista inexperiente poderia ficar tentado a criar modelos inverossímeis com um número enorme de parâmetros.

Os critérios de informação de Akaike (AIC) e de Bayes (BIC) evitam que isso aconteça ao adicionar um termo de penalidade que cresce com a adição de novos parâmetros. Dessa forma, constituem critérios objetivos a serem minimizados, levando em conta tanto a adequação aos dados empíricos quanto a complexidade de um modelo. Sendo *n* o número de observações disponíveis, *k* o número de parâmetros do modelo e $\hat{\sigma}_k^2$ a variância estimada dos resíduos do modelo com *k* parâmetros, dada pela fórmula $\hat{\sigma}_k^2 = SQR_k/n$, o AIC e BIC são definidos conforme as equações abaixo:

$$AIC = \log \hat{\sigma}_k^2 + \frac{n + 2k}{n}$$

$$BIC = \log \hat{\sigma}_k^2 + \frac{k \log n}{n}$$

Para modelos genéricos (não necessariamente de regressão linear multivariada), substitui-se $\hat{\sigma}_k^2$ pela função de verossimilhança L_k, definida conforme as hipóteses adotadas.

2.2.6. Modelos VAR(p): Extensão multivariada dos modelos ARIMA

Os modelos expostos até aqui foram criados para modelar séries temporais de uma única variável. No entanto, em muitos casos (incluindo no presente trabalho), pode ser de interesse do analista modelar simultaneamente duas ou mais séries temporais cujos valores são medidos nos mesmos instantes. Os modelos VAR (Vector Autoregression) são uma adaptação dos modelos ARIMA para estudar séries temporais multivariadas.

Nesse contexto, denota-se $\boldsymbol{x_t} = (x_{t,1}, x_{t,2}, \ldots, x_{t,k})^T$ o vetor coluna composto por *k* séries univariadas. São utilizados apenas termos de autorregressão porque, segundo Shumway e Stoffer (2011), a construção de um modelo multidimensional com termos de média móvel é um problema muito mais complexo.

Sendo assim, os modelos VAR de ordem *p*, VAR(p), são escritos na forma:

$$\boldsymbol{x_t} = \boldsymbol{\alpha} + \phi_1 \boldsymbol{x_{t-1}} + \phi_2 \boldsymbol{x_{t-2}} + \ldots + \phi_p \boldsymbol{x_{t-p}} + \boldsymbol{w_t},$$

onde ϕ_i é uma matriz de dimensões *k x k* que expressa a dependência de $\boldsymbol{x_t}$ com relação a $\boldsymbol{x_{t-i}}$, $\boldsymbol{\alpha}$ é um vetor de constantes e $\boldsymbol{w_t}$ é um vetor de variáveis normalmente distribuídas de média zero e matriz de covariância $\boldsymbol{\Sigma_w}$.

A ordem *p* do modelo é tipicamente determinada pela minimização dos critérios de informação de Akaike ou de Bayes (AIC ou BIC).

2.2.7. Teste de causalidade de Granger

Granger (1969) propôs um método estatístico para determinar relações de causalidade entre duas séries temporais. A noção de causalidade de Granger é centrada em torno de poder preditivo: diz-se que um sinal x_1 causa x_2 se os valores passados de x_1 ajudam na previsão de x_2, mesmo se forem incluídos em um modelo que leva em conta também os valores passados de x_2. Segundo Eichler (2011), essa é a definição mais utilizada no contexto de séries temporais.

O teste de causalidade de Granger consiste em um teste de hipótese, cuja hipótese nula é a de que x_1 não causa x_2. Inicialmente, é realizada uma autorregressão de x_2 até a ordem *m*:

$$x_{t,2} = a_o + \sum_{i=1}^{m} a_i \, x_{t-i,2} + \varepsilon_t,$$

onde ε_t é a série de resíduos do modelo. Em seguida, são incluídos passados de x_1 até a ordem *p*:

$$x_{t,2} = a_o + \sum_{i=1}^{m} a_i \, x_{t-i,2} + \sum_{j=1}^{p} b_i \, x_{t-i,1} + \eta_t,$$

com η_t representando a nova série de resíduos. Se os coeficientes b_i são significativos estatisticamente e a adição dos termos $x_{t-i,1}$ reduz a variância dos resíduos (o que é determinado respectivamente por testes *t* e *F*), rejeita-se a hipótese nula e diz-se que a série x_1 causa x_2 no sentido de Granger.

2.3. Modelos de espaço de estado e filtros de Kalman

Segundo Shumway e Stoffer (2011), os modelos de espaço de estado, também chamados de modelos dinâmicos lineares, são ferramentas muito poderosas que generalizam outras classes de modelos de séries temporais. Inicialmente introduzidos na área de pesquisa aeroespacial, foram estudados por Kalman (1960) e estendidos para o domínio de modelagem de séries estatísticas.

Esses modelos são constituídos de duas equações:

1. Uma equação de mudança de estado, na qual um vetor a ser determinado $\boldsymbol{x_t}$ é expresso como uma função linear de seu valor anterior $\boldsymbol{x_{t-1}}$, com a adição de um ruído gaussiano $\boldsymbol{w_t}$;
2. Uma equação de medição: assume-se que não podemos medir diretamente o vetor $\boldsymbol{x_t}$, mas sim uma versão linearmente transformada e com adição de ruído gaussiano que é denominada $\boldsymbol{y_t}$

Como relatam Grewal e Andrews (2010), esses modelos surgem naturalmente no problema de determinação da posição de um foguete: deseja-se conhecer suas coordenadas $\boldsymbol{x_t}$ em um instante *t*, mas seus sensores retornam um valor com ruído $\boldsymbol{y_t}$. Para diminuir a incerteza da medição, pode-se usar o fato de que a verdadeira posição do foguete evolui em função de sua posição anterior $\boldsymbol{x_{t-1}}$ e de sua velocidade.

Às equações de medição e transição podem ser adicionadas variáveis exógenas $\boldsymbol{u_t}$, de forma que os modelos de espaço de estado são escritos na seguinte forma geral:

$$\boldsymbol{y_t} = \mathrm{A_t}\mathbf{x_t} + \Gamma\boldsymbol{u_t} + \boldsymbol{v_t}$$

$$\mathbf{x_t} = \Phi\mathbf{x_{t-1}} + \Upsilon\boldsymbol{u_t} + \boldsymbol{w_t}$$

onde:

1. $\boldsymbol{x_t} = (x_{t,1}, x_{t,2}, \ldots, x_{t,p})^T$ é o vetor de *p* variáveis a serem determinadas;
2. $\boldsymbol{y_t} = (y_{t,1}, y_{t,2}, \ldots, y_{t,q})^T$ é o vetor de *q* medições;
3. $\boldsymbol{u_t} = (u_{t,1}, u_{t,2}, \ldots, u_{t,r})^T$ é um vetor de *r* variáveis exógenas;
4. Φ é uma matriz $p\ x\ p$, chamada de matriz de transição;
5. $\mathrm{A_t}$ é uma matriz $p\ x\ p$, chamada de matriz de medição;
6. Γ e Υ são matrizes $q\ x\ r$ e $p\ x\ r$, respectivamente;
7. $\boldsymbol{v_t}$ e $\boldsymbol{w_t}$ são séries de ruído gaussiano de média nula, sem correlação entre si e com matrizes de covariância R e Q, respectivamente

Embora as matrizes Φ, Γ e Υ também possam variar com o tempo, esse normalmente não é o caso; por isso o índice *t* foi omitido, de forma a aumentar a clareza.

Posto o sistema de equações do modelo de espaço de estado, o problema é estimar os valores de $\mathbf{x_t}$ em um instante *t*, dado um valor inicial $\mathbf{x_0}$ que obedece a uma distribuição normal $N(\boldsymbol{\mu_0}, \Sigma_0)$, o conjunto de medições $\boldsymbol{\psi_s} = \{\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_s}\}$ até o instante *s* e o valor das variáveis exógenas até o instante *s*. O algoritmo que resolve esse problema quando *s* = *t* é chamado de filtro de Kalman, e será descrito na seção a seguir.

2.3.1. Filtro de Kalman

Introduzindo a notação $\boldsymbol{x_t^s} = E(\boldsymbol{x_t} \mid \boldsymbol{\psi_s})$ e $\boldsymbol{P_t^s} = E\{(\boldsymbol{x_t} - \boldsymbol{x_t^s})(\boldsymbol{x_t} - \boldsymbol{x_t^s})^T \mid \boldsymbol{\psi_s}\}$, é possível provar os resultados abaixo utilizando teoria de probabilidades:

$$\boldsymbol{x_t^{t-1}} = \Phi\boldsymbol{x_{t-1}^{t-1}} + \Upsilon\boldsymbol{u_t}$$

$$\boldsymbol{P_t^{t-1}} = \Phi\boldsymbol{P_{t-1}^{t-1}}\Phi^{\mathrm{T}} + \mathrm{Q}$$

Utilizando os resultados acima, Kalman (1960) deduziu as seguintes relações:

$$\boldsymbol{x_t^t} = \boldsymbol{x_t^{t-1}} + K_t(\boldsymbol{y_t} - A_t\boldsymbol{x_t^{t-1}} - \Gamma\boldsymbol{u_t})$$

$$\boldsymbol{P_t^t} = [I - K_tA_t]\boldsymbol{P_t^{t-1}},$$

onde $K_t = \boldsymbol{P_t^{t-1}} A_t^T [A_t \boldsymbol{P_t^{t-1}} A_t^T + R]^{-1}$ recebe o nome de ganho de Kalman. Com essas relações, é possível chegar às estimativas $\boldsymbol{x_t^t}$ e $\boldsymbol{P_t^t}$ de forma recursiva, partindo de $\mathbf{x_0} = \boldsymbol{\mu_0}$ e $\boldsymbol{P_0} = \Sigma_0$ para obter $\boldsymbol{x_1^1}, \boldsymbol{x_2^2}, \ldots, \boldsymbol{x_t^t}$ e $\boldsymbol{P_1^1}, \boldsymbol{P_2^2}, \ldots, \boldsymbol{P_t^t}$.

Para realizar previsões, basta calcular os valores de $\boldsymbol{x_t^t}$ e $\boldsymbol{P_t^t}$ e servir-se das duas primeiras equações para obter a previsão ($\boldsymbol{x_{t+1}^t}$) e sua matriz de covariância ($\boldsymbol{P_{t+1}^t}$).

2.3.2. Estimação de parâmetros do modelo de espaço de estado

No caso em que alguns dos elementos das matrizes A_t, Φ, Γ, Υ, R ou Q são desconhecidos, é possível estimá-los utilizando o algoritmo a seguir, dado por Shumway e Stoffer (2011):

i. Escolher valores iniciais para os parâmetros desconhecidos, que serão representados pelo vetor $\Theta^{(0)}$
ii. Aplicar o filtro de Kalman aos dados experimentais usando os parâmetros $\Theta^{(0)}$ para obter uma série de inovações $\epsilon_t = (\boldsymbol{y_t} - A_t \boldsymbol{x_t^{t-1}} - \Gamma \boldsymbol{u_t})$ e de matrizes $\Sigma_t = (A_t \boldsymbol{P_t^{t-1}} A_t^T + R)$
iii. Calcular a função de verossimilhança $-ln\, L_Y(\Theta) = \frac{1}{2}\sum_{t=1}^{n} log\, |\Sigma_t| + \frac{1}{2}\sum_{t=1}^{n} \epsilon_t{}^T\, \Sigma_t{}^T \epsilon_t$
iv. Usar algum método de otimização numérica, como o método de Newton-Raphson, utilizando $-ln\, L_Y(\Theta)$ como função objetivo a ser minimizada para obter um novo conjunto de estimativas $\Theta^{(1)}$
v. Na iteração j, repetir a etapa 2 com $\Theta^{(j)}$ no lugar de $\Theta^{(j-1)}$ para obter novos valores de ϵ_t e de Σ_t. Utilizar esses valores nas etapas 3 e 4 para obter $\Theta^{(j+1)}$ e parar quando as estimativas de $-ln\, L_Y(\Theta)$ convergirem

2.4. Redes neurais artificiais

Segundo Goodfellow et al. (2016), redes neurais artificiais (ANNs) são sistemas computacionais inspirados em neurociência, cujo objetivo é aproximar uma função desconhecida $\boldsymbol{y} = f^*(\boldsymbol{x})$, encontrando o conjunto de parâmetros $\boldsymbol{\theta}$ que resulte na melhor aproximação $\hat{\boldsymbol{y}} = f(\boldsymbol{x}; \boldsymbol{\theta})$ para uma classe de funções f conhecidas.

Redes neurais artificiais foram desenvolvidas historicamente sobre o trabalho original de McCulloch e Pitts (1943), com contribuições significativas de Rosenblatt (1958) na invenção do *perceptron* (seu precursor), e de Ivakhnennko e Lapa (1967) que propuseram os primeiros modelos funcionais com múltiplas camadas.

Esses sistemas são compostos por uma rede de nós interconectados que recebem entradas de outros nós e computam um valor de saída. Esses nós são chamados de neurônios, e implementam funções não lineares, calculadas sobre uma soma ponderada de suas entradas. As funções mais comuns utilizadas em redes neurais são (sem ordem específica):

1. Tangente hiperbólica: $\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$, cuja imagem varia de -1 a 1;
2. Sigmoide: $S(x) = \frac{1}{1 + e^{-x}}$, cuja imagem varia de 0 a 1;
3. ReLU (Rectified Linear Unit): $ReLU(x) = \begin{cases} 0, se\ x < 0 \\ x, se\ x \geq 0 \end{cases}$, cuja imagem varia de 0 a $+\infty$

A Figura 1 mostra uma comparação entre as três principais funções neurais (também chamadas de funções de ativação), para valores de x variando de -5 a 5.

Graficamente, uma rede neural é representada por um grafo, no qual os nós correspondem aos neurônios e as arestas às conexões entre eles. À cada aresta é atribuído um valor chamado **peso**, de forma que os pesos determinam a ponderação realizada sobre a soma dos valores de entrada nos nós. Os nós são organizados em camadas, com conexões entre nós de camadas vizinhas mas não entre nós da mesma camada.

A Figura 2 esquematiza um neurônio genérico em uma rede neural. Esse neurônio recebe em sua entrada três valores, que são denotados por $e_1, e_2\ e\ e_3$, e calcula um valor de saída s. Sendo ϕ a função neural e w_i o peso associado à aresta que fornece a entrada e_i, o valor de saída é determinado pela expressão $s = \phi(\sum_{i=1}^{3} w_i e_i + b)$, onde b é chamado de viés.

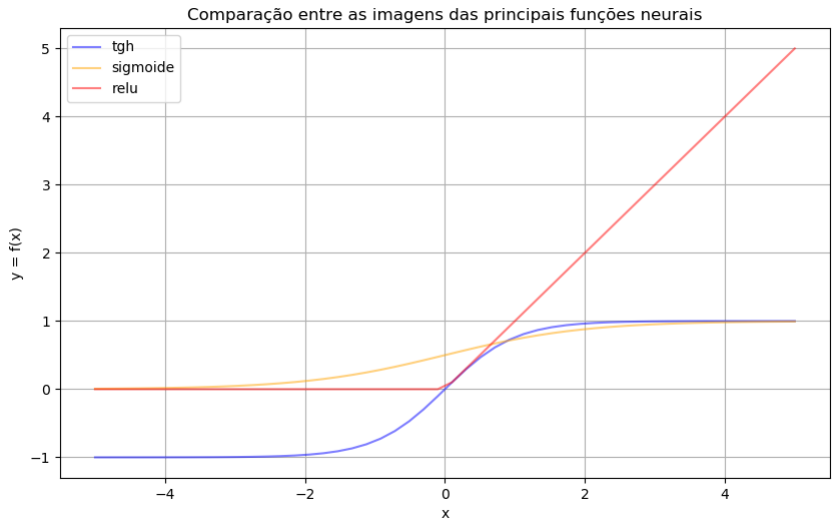

Figura 1: Imagens das três principais funções neurais. Fonte: elaborado pelo autor

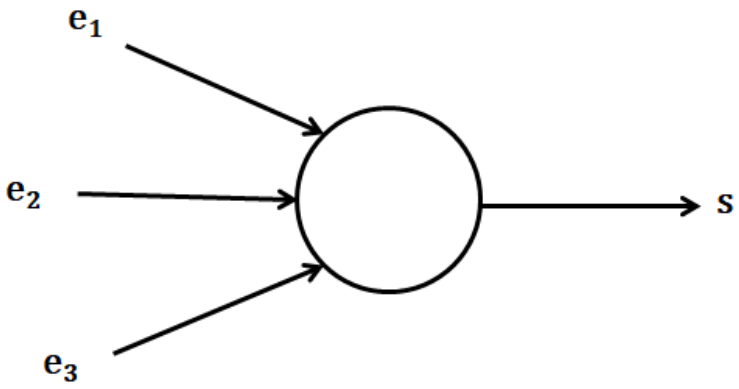

Figura 2: Representação de um neurônio de uma rede neural. Esse neurônio recebe três valores na entrada e calcula um valor de saída s. Fonte: elaborado pelo autor

2.4.1. Redes neurais *feedforward*

Redes neurais são classificadas em função de sua topologia. Segundo Goodfellow et al. (2016), a topologia mais simples é a de redes neurais chamadas *feedforward*, em que a informação flui diretamente da camada de entrada, que recebe os valores de $\boldsymbol{x}$, com um nó para cada dimensão, até a camada de saída, que fornece os valores de $\boldsymbol{y}$ calculados pela rede. Entre as camadas de saída e de entrada são colocadas camadas intermediárias, que são constituídas por neurônios como o da Figura 2. O número de camadas e o número de neurônios por camada devem ser decididos pelo analista.

A Figura 3 mostra o esquema de uma rede *feedforward* 3-5-2-1, com três neurônios na camada de entrada, cinco na primeira camada intermediária, dois na segunda camada intermediária e um neurônio de saída. Note que há conexões entre todos os neurônios de camadas consecutivas, o que é um padrão para essa topologia de rede. Os nós de entrada são representados por quadrados, os outros neurônios por círculos e os triângulos representam os vieses adicionados às camadas intermediárias e à camada de saída. A informação flui da esquerda para a direita.

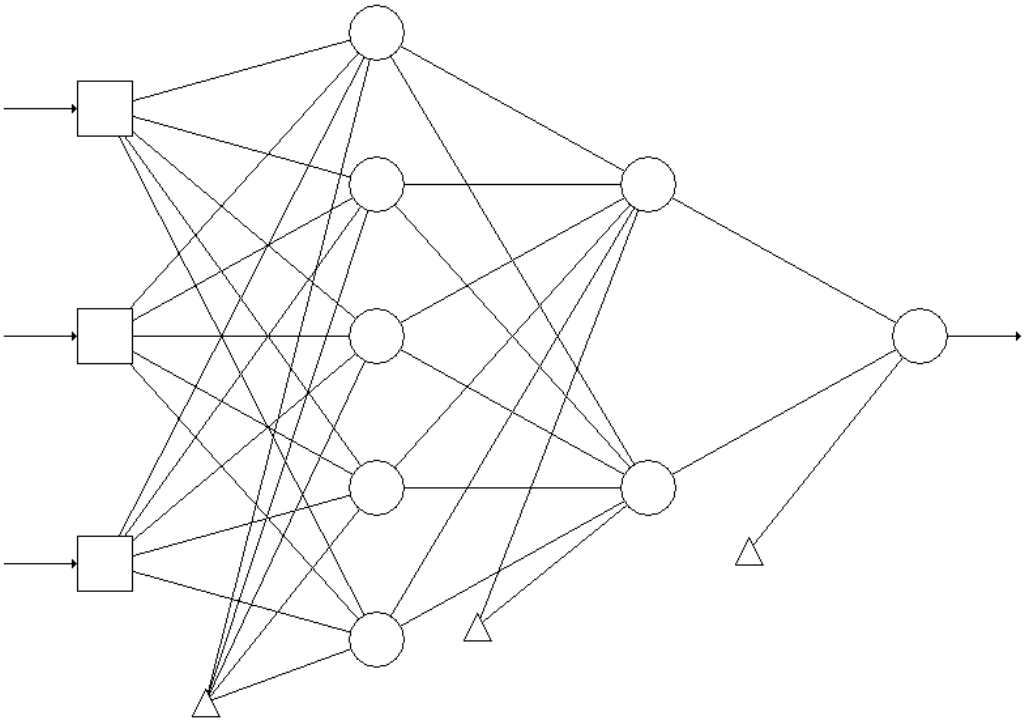

Figura 3: Rede neural *feedforward* 3-5-2-1. Elaborado pelo autor com ajuda do software MBP (Multiple Back-Propagation) v.2.2.5

Uma rede neural *feedforward* recebe *n* pares de valores $(\boldsymbol{x}^{(i)}, y^{(i)})$, e seu objetivo é fazer com que as saídas $\hat{y}^{(i)}$ que processa a partir das entradas $\boldsymbol{x}^{(i)}$ sejam as mais próximas possíveis de $y^{(i)}$. Para isso, é computada uma função de perda, que deve ser minimizada. As funções de perda mais utilizadas são:

1. Soma média dos quadrados: $f(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)}(\theta))^2$, para problemas de regressão;

2. Entropia cruzada: $f(\theta) = -\sum_{i=1}^{n} y^{(i)} \log \hat{y}^{(i)}(\theta) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}(\theta))$, para problemas de classificação

Uma vez decidida a topologia da rede (número de camadas e número de neurônios por camada) e as funções neurais a serem utilizadas em cada camada, o problema torna-se encontrar os valores dos pesos $\boldsymbol{w}$ para cada aresta e dos vieses $\boldsymbol{b}$ para cada nó que minimizam a função de perda $f(\boldsymbol{\theta})$, sendo $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ o vetor de *p* parâmetros da rede, que nesse caso são compostos por elementos de $\boldsymbol{w}$ e $\boldsymbol{b}$.

A minimização de $f(\boldsymbol{\theta})$ é realizada por um algoritmo de gradiente descendente estocástico, cujas etapas são as seguintes:

i. Inicializar o vetor $\boldsymbol{\theta}$ com valores aleatórios próximos de zero. O vetor inicial de parâmetros será denominado por $\boldsymbol{\theta}^{(\mathbf{0})}$

ii. Calcular numericamente $\overrightarrow{\nabla f}\left(\boldsymbol{\theta}^{(i)}\right) = \left(\frac{\partial f}{\partial \theta_1^{(i)}}, \frac{\partial f}{\partial \theta_2^{(i)}}, \ldots, \frac{\partial f}{\partial \theta_p^{(i)}}\right)$, o gradiente da função objetivo avaliado em $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}$. Essa operação é realizada por um algoritmo chamado *backpropagation*, cuja ideia é a de aplicar a regra da cadeia para todos os caminhos da rede neural que passam por $\theta_j^{(i)}$ para calcular as derivadas parciais

iii. Atualizar o valor de $\boldsymbol{\theta}^{(i)}$, segundo a relação $\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \gamma \overrightarrow{\nabla f}\left(\boldsymbol{\theta}^{(i)}\right)$. O parâmetro γ é chamado de taxa de aprendizado e pode ser predefinido pelo analista ou ser controlado por um outro algoritmo de otimização

iv. Repetir as etapas ii e iii, atualizando $\boldsymbol{\theta}^{(i)}$ até haver convergência

Um dos resultados teóricos mais importantes sobre ANNs é o teorema da aproximação universal de Cybenko (1989), que mostrou que uma rede neural *feedforward* com somente uma camada intermediária e um número finito de neurônios é capaz de aproximar qualquer função contínua em um subconjunto compacto de $\mathbb{R}^n$ com um nível de precisão arbitrário ε, utilizando uma função de ativação φ não-constante, limitada, contínua e que cresce monotonicamente.

2.4.2. Redes neurais recorrentes

Segundo Goodfellow et al. (2016), redes neurais recorrentes (RNNs) são especializadas no processamento de sequências de valores, algo que não é possível em redes

neurais *feedforward*, já que para estas em nenhum momento é introduzido conceitos de dependência temporal. Essa família de redes neurais foi proposta por Rumelhart et al. (1986).

Devido a essa característica, RNNs são usadas extensivamente na literatura para resolver problemas de processamento de linguagem natural, como no reconhecimento de frases escritas (Graves et al., 2009) ou de voz humana (Li e Wu, 2014), e também relacionados a séries temporais (Rout et al, 2017).

Redes neurais recursivas operam em sequências de vetores $\boldsymbol{x_t}$, com t variando de 1 a T. Em cada etapa temporal, seu estado oculto $\boldsymbol{h_t}$ evolui segundo uma relação de recorrência da forma $\boldsymbol{h_t = f(h_{t-1}, x_t; \theta)}$, onde $\boldsymbol{\theta}$ representa, como nas seções anteriores, o vetor de parâmetros a serem aprendidos pela rede. Essa relação pode também ser escrita explicitamente através de substituições sucessivas, como por exemplo $\boldsymbol{h_3 = f(f(f(h_1, x_1; \theta), x_2; \theta), x_3; \theta)}$.

Seguindo a mesma lógica, uma rede neural recorrente pode ser representada de duas formas: por um diagrama de circuito, como na Figura 4, ou na forma de um grafo computacional aberto, como na Figura 5.

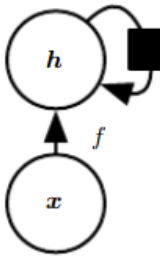

Figura 4: Representação de uma relação de recorrência por um diagrama de circuito. Fonte: Goodfellow et al. (2016)

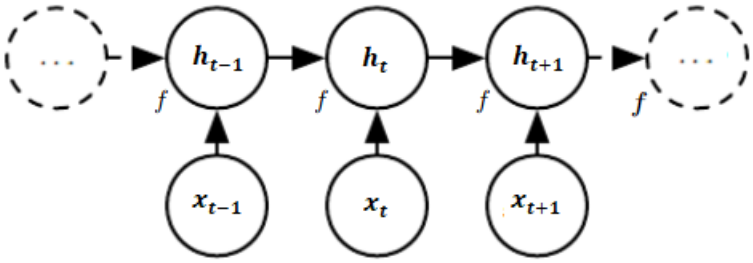

Figura 5: Representação da mesma relação de recorrência na forma de um grafo computacional aberto. Fonte: adaptado de Goodfellow et al. (2016)

Para entender melhor o papel do estado oculto $\boldsymbol{h_t}$, é importante conhecer a arquitetura de uma RNN. Em função do problema a ser resolvido, podem ser implementadas arquiteturas diferentes, sendo as principais:

1. RNNs que produzem uma saída $\boldsymbol{o_t}$ a cada etapa temporal e possuem relações de recorrência entre seus estados ocultos $\boldsymbol{h_t}$;
2. RNNs que produzem uma saída $\boldsymbol{o_t}$ a cada etapa temporal e introduzem relações de recorrência apenas entre a saída $\boldsymbol{o_t}$ de uma etapa e o estado oculto da etapa seguinte $\boldsymbol{h_{t+1}}$;
3. RNNs que possuem relações de recorrência entre seus estados ocultos $\boldsymbol{h_t}$ e que produzem uma única saída $\boldsymbol{o_t}$. Essas são as redes utilizadas na previsão de séries temporais.

As redes neurais de terceiro tipo, que serão usadas nesse trabalho, estão esquematizadas na Figura 6, na forma de grafo computacional aberto.

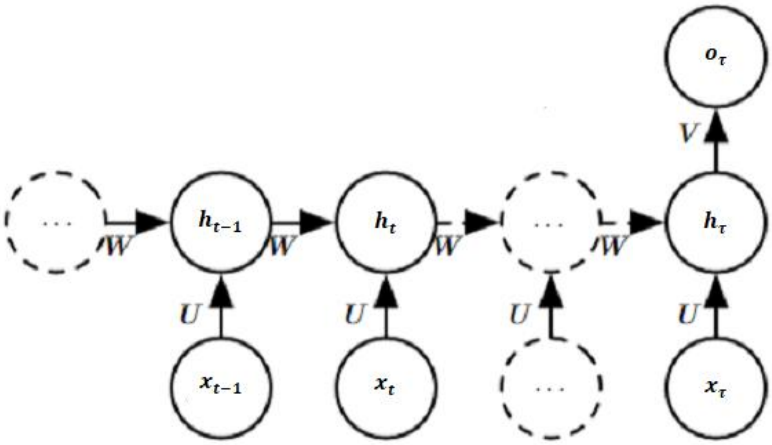

Figura 6: Redes neurais recursivas com uma única saída. Fonte: adaptado de Goodfellow et al. (2016)

Dada a arquitetura dessa rede neural, é possível especificar sua saída única $\boldsymbol{o_\tau}$ em função da sequência de entrada $\boldsymbol{x_1, x_2, \ldots, x_T}$.

$$\boldsymbol{h_t} = \sigma_h(\boldsymbol{b} + \boldsymbol{W h_{t-1}} + \boldsymbol{U x_t})$$

$$\boldsymbol{o_\tau} = \boldsymbol{V h_\tau}$$

$$\hat{y} = \sigma_o(\boldsymbol{o_\tau})$$

onde:

1. σ_h e σ_o são funções neurais previamente determinadas;
2. $\boldsymbol{h_t}$ é um vetor de dimensões *m* x 1, chamado de vetor de estados ocultos. A dimensão *m* é chamada de número de unidades, e deve ser escolhida pelo analista;
3. $\boldsymbol{x_t}$ é o vetor de entradas, de dimensão *p* x *1*;
4. $\boldsymbol{W}$ é uma matriz *m* x *m*, $\boldsymbol{U}$ é uma matriz *m* x *p* e $\boldsymbol{V}$ é uma matriz *1* x *m*. A matriz $\boldsymbol{V}$ é importante porque pondera o resultado de cada uma das unidades para produzir a estimativa final de $\hat{y}$;
5. $\boldsymbol{b}$ é um vetor de viés de dimensão *m* x *1*

De forma análoga ao caso das redes *feedforward*, uma rede neural recursiva recebe *n* pares de valores $(\boldsymbol{x_t^{(i)}}, y^{(i)})$, com *i* variando de 1 a *n*, e produz estimativas $\hat{y}^{(i)}$, em função de $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{W}$ e $\boldsymbol{b}$, que são representados pelo vetor de parâmetros $\boldsymbol{\theta}$**.** O objetivo é achar o valor de $\boldsymbol{\theta}$ que minimiza $f(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - \hat{y}^{(i)}(\boldsymbol{\theta}))^2$ ou outra função de perda a ser escolhida pelo analista.

Novamente esse problema é resolvido através do método de gradiente descendente, com as mesmas etapas descritas na seção anterior. A única diferença é no cálculo do gradiente de $f(\boldsymbol{\theta})$, que é calculado por um algoritmo ligeiramente diferente devido à maior complexidade do grafo. Esse algoritmo recebe o nome de *backpropagation through time* (BPTT), e detalhes de sua implementação podem ser vistos em Werbos (1990).

2.4.2.1. Redes neurais LSTM

Hochreiter (1991) identificou o desafio de aprender dependências temporais longas em redes neurais recorrentes, também conhecido como problema de desaparecimento do gradiente. Segundo Goodfellow et al. (2016), a ideia básica é a de que o cálculo do gradiente em uma RNN com muitas etapas envolve a aplicação da regra da cadeia múltiplas vezes, o que tende a fazer com que termos próximos do início do grafo desapareçam ou cresçam exponencialmente. Para entender o porquê, basta imaginar uma rede neural simples em que a equação de atualização dos estados ocultos é dada por:

$$\boldsymbol{h_t = W^T h_{t-1}}$$

Essa relação de recorrência pode ser simplificada para:

$$\boldsymbol{h_t = (W^t)^T h_0}$$

onde $\boldsymbol{W^t}$ denota a exponenciação da matriz $\boldsymbol{W}$. Assumindo que $\boldsymbol{W}$ possa ser decomposta em autovetores **Q** e autovalores **Λ**, sendo **Q** uma matriz ortogonal, é possível escrever

$$\boldsymbol{h_t = Q^T \Lambda^t Q h_0.}$$

Os autovalores **Λ** estão elevados a *t*, de forma que para *t* suficientemente grande, quantidades menores do que um (o que normalmente ocorre) tenderão a zero e quantidades maiores do que um crescerão exponencialmente. Um fenômeno similar ocorre no algoritmo de *backpropagation through time*, de forma que a contribuição de valores iniciais da sequência torna-se pequena para o aprendizado dos parâmetros $\boldsymbol{\theta}$ da RNN.

Para solucionar esse problema, Hochreiter e Schmidhuber (1997) propuseram uma nova classe de redes neurais recursivas, batizadas de LSTM (Long Short-Term Memory). A grande inovação desses modelos é a introdução de "portões" que controlam a retenção de informação na rede. Através desses portões, cujos parâmetros devem ser aprendidos pelo algoritmo, é possível priorizar o aprendizado de padrões de curto ou longo prazo na sequência de entrada. Existem três portões: um portão de entrada $\boldsymbol{i_t}$, um portão de saída $\boldsymbol{q_t}$ e um portão de esquecimento $\boldsymbol{f_t}$. As equações de uma rede LSTM são:

$$\boldsymbol{f_t} = \sigma_g(\boldsymbol{W_f x_t} + \boldsymbol{U_f h_{t-1}} + \boldsymbol{b_f})$$

$$\boldsymbol{i_t} = \sigma_g(\boldsymbol{W_i x_t} + \boldsymbol{U_i h_{t-1}} + \boldsymbol{b_i})$$

$$\boldsymbol{q_t} = \sigma_g(\boldsymbol{W_o x_t} + \boldsymbol{U_o h_{t-1}} + \boldsymbol{b_o})$$

$$\boldsymbol{c_t} = \boldsymbol{f_t} \circ \boldsymbol{c_{t-1}} + \boldsymbol{i_t} \circ \sigma_h(\boldsymbol{W_c x_t} + \boldsymbol{U_c h_{t-1}} + \boldsymbol{b_c})$$

$$\boldsymbol{h_t} = \boldsymbol{q_t} \circ \sigma_h(\boldsymbol{c_t})$$

onde:

1. $\boldsymbol{i_t}$, $\boldsymbol{q_t}$ e $\boldsymbol{f_t}$ são vetores *m* x 1 que representam os portões descritos anteriormente. Seus valores vão de 0 a 1;

2. $\boldsymbol{c_t}$ é um vetor *m* x 1 chamado de vetor de estado de célula. O portão de esquecimento $\boldsymbol{f_t}$ controla a sua inércia ao longo da recursão, enquanto que o portão de entrada $\boldsymbol{i_t}$ controla a incorporação de novas informações ao estado da célula;
3. $\boldsymbol{h_t}$ é mais uma vez o vetor *m* x 1 de estado oculto. Note que nesse caso esse vetor é calculado a partir do estado de célula e controlado pelo portão de saída $\boldsymbol{q_t}$;
4. $\boldsymbol{W}$ e $\boldsymbol{U}$ são matrizes *m* x *p* e *m* x *m* como para RNNs convencionais, mas nesse caso há quatro matrizes diferentes, uma para calcular o valor de cada portão e outra para atualizar o estado de célula. O mesmo se aplica para os vieses $\boldsymbol{b}$;
5. σ_h é a função tangente hiperbólica e σ_g é a função sigmoide;
6. O símbolo ∘ denota o produto matricial de Hadamard: se $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$, então $c_{ij} = a_{ij} * b_{ij}$ (multiplicação elemento por elemento)

A Figura 7 mostra uma representação gráfica dessas equações.

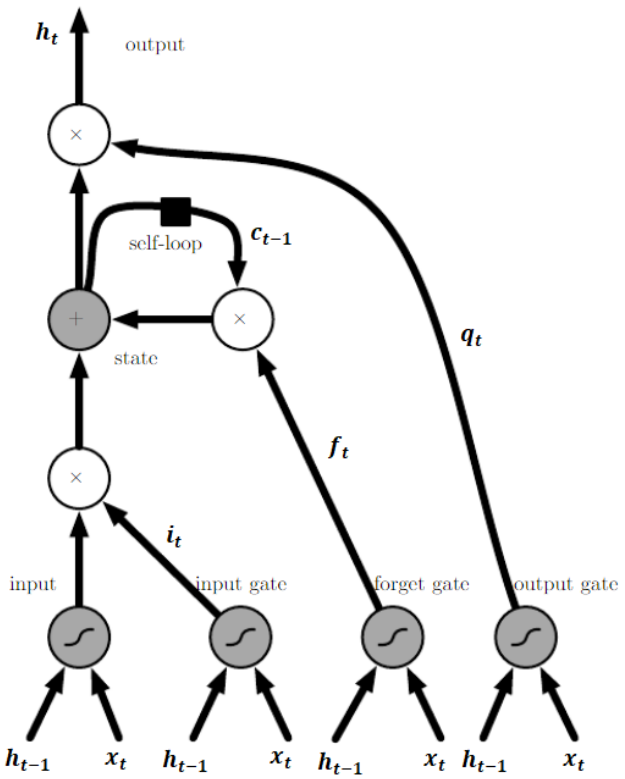

Figura 7: Representação gráfica das equações de uma rede LSTM. Fonte: adaptado de Goodfellow et. al (2016)

2.4.3. Metodologia prática

Goodfellow et al. (2016) argumentam que não basta conhecer quais algoritmos existem e saber seus princípios básicos de funcionamento para aplicar um modelo de redes neurais com sucesso. Existem decisões práticas que devem ser tomadas por um analista que afetam diretamente o resultado final de um modelo, como:

1. Saber se é necessário coletar mais dados para treinar o algoritmo;
2. Aumentar ou diminuir a capacidade de um modelo - no caso de redes neurais feedforward, isso se reflete pelo número de camadas e número de neurônios por camada; no caso de redes neurais recursivas, a complexidade do modelo é expressa pelo número de unidades internas (dimensão do estado oculto);
3. Escolher o número de variáveis preditivas (dimensão das entradas);
4. Implementar corretamente o modelo e saber corrigir erros de código

Tendo isso em vista, os autores recomendam o seguinte processo de concepção de um modelo de redes neurais:

i. Determinar os objetivos do modelo: baseado no problema que deve ser resolvido, isso envolve escolher uma função de erro apropriada e colocar uma meta para o erro do algoritmo, sabendo que na grande maioria dos casos não existe erro zero. Uma boa forma de estabelecer a meta é usar a performance de um algoritmo mais simples
ii. Definir a sequência de tratamento de dados: escolha de variáveis, normalização, separação de conjuntos de teste/treino, aplicação do algoritmo e medição do erro
iii. Avaliar o sistema para gargalos de performance: identificar problemas nos dados de entrada, erros de código e diagnosticar se os problemas são devidos a sobre aprendizado (*overfitting*) ou sub aprendizado (*underfitting*)
iv. Utilizar as conclusões da etapa anterior para realizar mudanças incrementais no modelo, como ajustes de parâmetros ou coleta de mais dados. Se a performance no conjunto de treino está abaixo do esperado, deve-se aumentar a capacidade do modelo; se o problema está no conjunto de teste, pode-se coletar mais dados ou utilizar técnicas como o *dropout* (abandono), que consiste em excluir aleatoriamente células ou unidades internas de uma rede neural durante a fase de treino para reduzir *overfitting*

2.5. Modelagem de preços na literatura

A literatura de modelagem de preços é muito extensa, e a alta volatilidade e comportamento errático de séries financeiras dividem analistas quanto à sua verdadeira natureza, mesmo após mais de meio século de pesquisa.

Kendall (1953) analisou séries financeiras agregadas semanalmente e concluiu que as variações de preços eram predominantemente aleatórias, sendo difícil determinar estatisticamente se os fracos padrões observados eram significativos ou puramente devidos ao acaso. Seu trabalho influenciou o desenvolvimento da teoria dos passeios aleatórios (*random walks*), que segundo Fama (1965) afirma que os movimentos passados do preço de uma ação não podem ser usados para prever sua variação futura. A justificativa para essa teoria é a *hipótese dos mercados eficientes*, que acredita que novas informações são incorporadas muito rapidamente ao preço das ações devido ao grande numero de investidores e à alta liquidez, e por isso o melhor indicador do seu valor intrínseco é o preço atual.

Outros autores como Bessembinder et al. (1995) defendem que, em certos mercados como o de *commodities*, os preços eventualmente retornam à média ou à tendência a longo prazo. Como afirmam Wets e Rios (2012), essa noção é fundamentada por teoria básica microeconômica: quando os preços estão altos, produtores de maior custo entram no mercado, o que aumenta a oferta e gera pressão deflacionária; por outro lado, quando os preços abaixam, a oferta diminui porque alguns produtores não entram no mercado, o que gera pressão inflacionária.

Embora uma resposta definitiva não possa ser dada à questão da verdadeira natureza de séries financeiras, é importante conhecer essas teorias e testar modelos diferentes para ver qual se aplica melhor ao mercado e produto em questão. Essa abordagem é presente no trabalho de Reichsfeld e Roache (2011), que estudam a performance de modelos ARIMA, *random walk* e de contratos futuros na previsão do preço de dez *commodities* diferentes.

Schwartz (1997) compara três modelos estocásticos na previsão de preços de *commodities* como petróleo, ouro e cobre, e encontra fortes indícios de reversão à média. Os parâmetros desses três modelos são estimados utilizando filtros de Kalman, e as performances são avaliadas pela raiz quadrada do erro quadrático médio (RMSE) de estimativas dentro e fora da amostra de treino.

Wets e Rios (2012) propõem um modelo para a modelagem de preços de *commodities* que separa componentes de curto e longo prazo, utilizando movimentos brownianos para a estimação de ambos e incorporando um componente de reversão à média ao fator de longo prazo. Nas estimações de volatilidade e de arrasto são utilizados contratos futuros, que segundo os autores são mais eficientes do que projeções de estoques, ofertas e demandas porque todas essas informações já são levadas em conta por *traders* na sua precificação.

Mais recentemente, na área de redes neurais, Lago et al. (2018) constroem quatro modelos de redes neurais para prever preços de eletricidade incluindo RNNs e LSTM, e concluem que essas classes de modelos têm melhor performance comparadas a métodos estatísticos tradicionais. Já Baek e Kim (2018) usam redes neurais LSTM para prever índices dos mercados de ações americano e coreano.

Na área de *commodities* sucroalcoleeiras, Ribeiro e Oliveira (2011) utilizam um modelo híbrido na previsão de preços de açúcar, baseado em filtros de Kalman e em redes neurais artificiais, utilizando valores de contratos futuros e variáveis exógenas como preço do petróleo, vendas de veículos *flex* e taxas de câmbio.

Zafeiriou et al. (2014) utilizam um modelo de correção de erro (VECM) para estudar o impacto de preços de petróleo bruto, preços de gasolina e emissões de gases de efeito estufa na volatilidade de preços de etanol. Oliveira et al. (2018) estudam o problema da determinação do mix de produção utilizando técnicas financeiras de gestão de portfólio e estimativas de preços baseadas em modelos de espaço de estado e filtros de Kalman.

3. INDÚSTRIA SUCROALCOOLEIRA NO BRASIL

3.1. História

A história da cana-de-açúcar no Brasil se confunde com a do país. O primeiro engenho de açúcar brasileiro foi construído na Feitoria de Itamaracá (posteriormente chamada de capitania de Pernambuco) em 1516 e administrado por Pero Capico; na região Sudeste, a primeira unidade produtora foi instalada por Martim Afonso de Souza em 1533 na capitania de São Vicente. As mudas de cana foram trazidas da Ilha da Madeira pelos portugueses, e a produção era destinada a atender a demanda europeia.

De meados do século XVI até o fim do século XVII, a cultura do açúcar foi um verdadeiro motor para o povoamento e desenvolvimento do território brasileiro, consolidando a colonização portuguesa e constituindo a atividade econômica mais importante na época. Esse período da história brasileira é conhecido como ciclo do açúcar, e terminou devido à concorrência dos holandeses nas Antilhas e à descoberta do ouro em Minas Gerais.

O cultivo de açúcar continuou a perder importância até o final do século XIX, quando começaram a ser instaladas as primeiras usinas de cana-de-açúcar no Nordeste, substituindo os antigos engenhos. Na região Sudeste, o cultivo de açúcar ganhou força na década de 1930, com a crise do café.

Na década de 1970, o cultivo de cana-de-açúcar ganhou um incentivo com a criação do Programa Nacional do Álcool (Pró-Álcool), que buscava desenvolver motores a combustão interna movidos à álcool para reduzir a dependência do país com relação a derivados do petróleo. Para isso, eram dados subsídios governamentais a produtores de cana-de-açúcar e créditos para a construção de usinas sucroalcooleiras. O programa foi um sucesso, e viu a produção subir de 123 milhões de toneladas por ano na safra 1981/82 para 220 milhões de toneladas por ano na safra de 1986/87, como mostra a Figura 8.

No final da década de 1980, a crise econômica e a superinflação enfrentados pelo governo forçaram a diminuição dos subsídios, o que levou a indústria de álcool a uma estagnação que durou até a introdução dos veículos *flex* em 2003. Desde então, a produção de cana-de-açúcar está em plena expansão.

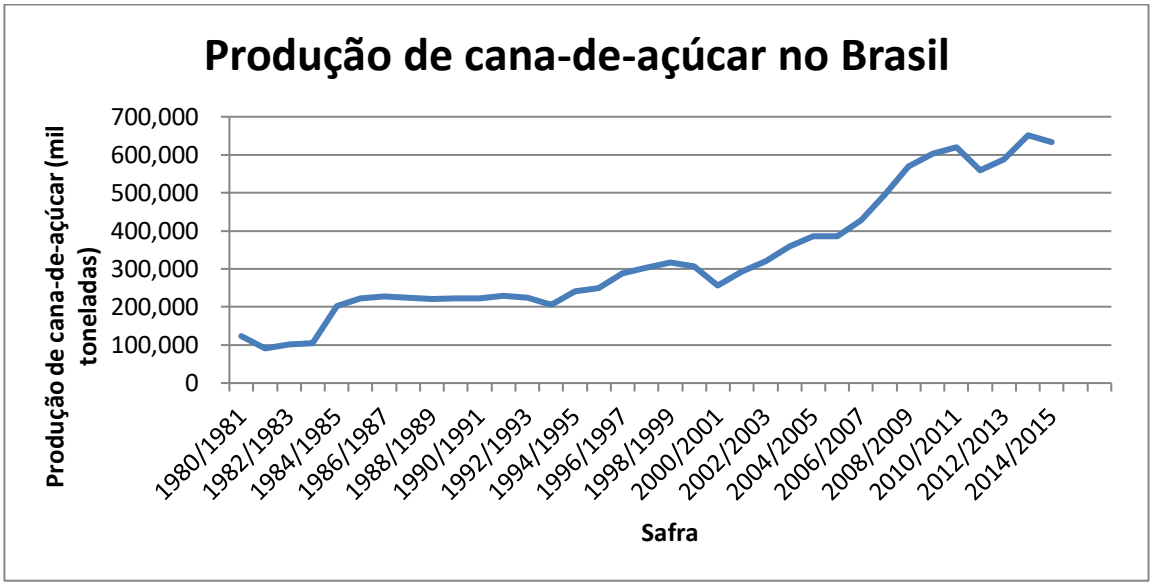

Figura 8: Evolução histórica da produção de cana-de-açúcar no Brasil, por safra. Fonte: UNICA (2016)

3.2. Cenário atual

3.2.1. Regiões produtoras e volumes produzidos

A indústria de cana-de-açúcar é dividida em duas regiões de atuação: Nordeste e Centro-Sul, que são os dois núcleos históricos de desenvolvimento da cultura desse produto. A região Centro-Sul é responsável por mais de 90% da produção total no Brasil, e esse número vem crescendo devido à diminuição de produção na Região Nordeste nas últimas safras, como podemos constatar na Tabela 2.

A Região Centro-Sul é constituída pelos estados do Espírito Santo (ES), Goiás (GO), Mato Grosso (MT), Mato Grosso do Sul (MS), Minas Gerais (MG), Paraná (PR), Rio de Janeiro (RJ), Rio Grande do Sul (RS) e São Paulo (SP). A Tabela 3 mostra a produção de cana-de-açúcar, açúcar e etanol total por estado, na safra 2017/2018. É possível ver que o estado de São Paulo responde por cerca de 61% da produção de cana-de-açúcar, 68% da produção de açúcar e 51% da produção de etanol da Região Centro-Sul. Por conta de sua importância, esse estado será escolhido como referência para esse estudo: os preços sucroalcooleiros levantados serão os preços sentidos pelos produtores do estado de SP.

Safra	Região		Total
	Centro-Sul	Nordeste	
2011/2012	493 159	66 056	**559 215**
2012/2013	532 758	55 720	**588 478**
2013/2014	597 061	54 233	**651 294**
2014/2015	573 145	60 782	**633 927**
2015/2016	617 709	49 114	**666 823**
2016/2017	607 136	44 703	**651 839**
2017/2018	596 330	44 738	**641 058**

Tabela 2: Evolução da produção de cana-de-açúcar (em mil toneladas) no Brasil, por região. Fonte: UNICA (2018)

Estado	Produto		
	Cana-de-açúcar	Açúcar	Etanol
ES	2 380 657	126 842	90 471
GO	70 621 968	2 241 828	4 618 388
MT	16 134 127	410 524	1 498 675
MS	46 940 207	1 491 650	2 631 783
MG	64 956 358	4 241 214	2 709 559
PR	37 047 410	2 920 682	1 268 948
RJ	1 061 788	35 374	46 432
RS	44 822	0	2 485
SP	365 989 639	24 591 393	13 222 807
TOTAL	**596 329 679**	**36 059 507**	**26 089 458**

Tabela 3: Produção de cana-de-açúcar (em toneladas), açúcar (em toneladas) e etanol (em milhares de litros) por estado da Região Centro-Sul, na safra 2017/2018. Fonte: UNICA (2018)

3.2.2. Número de usinas e distribuição geográfica

O site especializado *Nova Cana* faz o levantamento de todas as usinas sucroalcooleiras no Brasil e sua localização. Existem atualmente 411 usinas em atividade no Brasil, das quais 333 estão localizadas na região Centro-Sul e 172 no estado de São Paulo. A Tabela 4 mostra os sete estados brasileiros com maior número de usinas de cana-de-açúcar: é possível ver que São Paulo possui quatro vezes mais usinas que o segundo colocado, Minas Gerais.

Estado	Número de usinas
São Paulo (SP)	172
Minas Gerais (MG)	42
Goías (GO)	39
Paraná (PR)	30
Alagoas (AL)	25
Mato Grosso do Sul (MS)	24
Pernambuco (PE)	17

Tabela 4: Estados brasileiros com o maior número de usinas de cana-de-açúcar. Fonte: Nova Cana (2018)

Safra	Número de novas unidades	Número de unidades fechadas	Número total de usinas
2007/2008	22	3	337
2008/2009	29	3	363
2009/2010	21	3	381
2010/2011	10	4	387
2011/2012	4	16	375
2012/2013	2	13	364
2013/2014	2	10	356
2014/2015	0	9	347
2015/2016	2	8	341
2016/2017	1	6	336
2017/2018	1	4	333

Tabela 5: Evolução do número de usinas sucroalcooleiras na região Centro-Sul nas últimas onze safras. Fonte: UNICA (2018)

Apesar da tendência de aumento de produção de cana-de-açúcar no Brasil nos últimos anos, especialmente na Região Centro-Sul, o número de usinas está diminuindo anualmente desde a safra 2011/2012, como podemos ver na Tabela 5. Nos últimos sete anos, 68 unidades foram fechadas na região.

3.2.3. *Mix* de produção e participação em exportações

A maior parte da plantação de cana na região Centro-Sul é utilizada na produção de etanol, como mostra a Figura 9. Após atingir um máximo de 60.3% na safra 2008/2009, atualmente 53.5% da cana-de-açúcar colhida é empregada na produção de etanol.

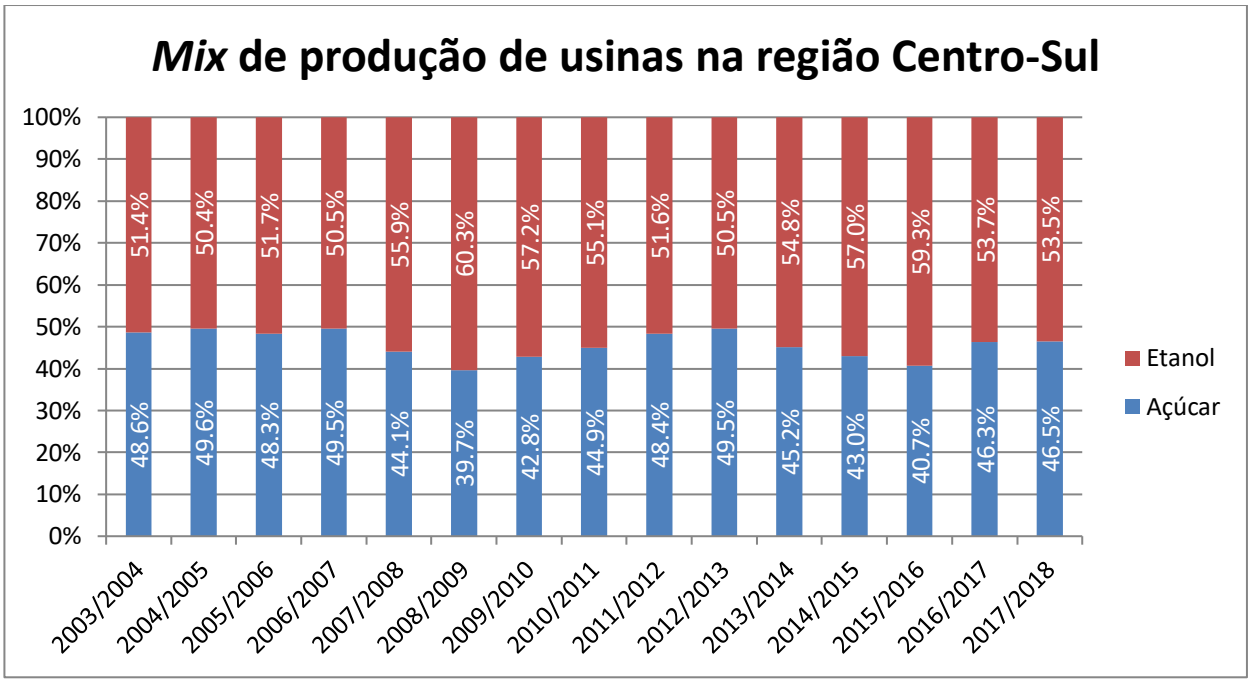

Figura 9: *Mix* de produção das usinas sucroalcooleiras da Região Centro-Sul desde 2003/2004. Fonte: ÚNICA (2018)

A produção de etanol é destinada principalmente para alimentar o mercado interno brasileiro, enquanto que boa parte da produção de açúcar é exportada. Segundo dados da ÚNICA (2018), da safra 2017/2018 foram exportados 916.944 dos 26.089.458 milhares de litros de etanol produzidos (cerca de 3,5%) e 26.484.000 das 36.509.057 toneladas de açúcar produzidas (cerca de 72,5%).

3.3. Produção e distribuição

3.3.1. Plantio e colheita de cana-de-açúcar

A cana-de-açúcar necessita de irrigação abundante, altas temperaturas e elevada incidência solar para se desenvolver plenamente. É preciso preparar o terreno antes do plantio, realizando limpeza, calagem, aração e cavação de sulcos. Uma vez o terreno

preparado, é aplicado o adubo e em seguida as mudas de cana são colocadas nos sulcos e cobertas de terra. Ainda é muito comum a colocação das mudas ser realizada manualmente, apesar do avanço da mecanização, que tem sido aplicada à maior parte das outras etapas.

É possível realizar em média cinco cortes em uma cultura de cana-de-açúcar, antes de replantá-la. O primeiro corte é o mais produtivo, chamado de corte da cana-planta; os cortes subsequentes são chamados de socas e vão decrescendo de produtividade até que seja necessário o plantio de novas mudas.

Existem duas possibilidades para a época de realização do plantio: na primeira (**cana de 18 meses**), espera-se seis meses entre a última colheita e o plantio de novas mudas, período no qual podem ser aplicadas outras culturas de rotação (soja, amendoim, entre outros) para aumentar a nitrogenação do solo; após esse período realiza-se o plantio e espera-se 18 meses para cortar a cana-planta. Essa técnica é aplicada em 80% dos casos, pois maximiza a produtividade do primeiro corte. A segunda opção (**cana de 12 meses**) é plantar a cana logo após a última soca e colhê-la após um ano: apesar de mais rápida, essa técnica é menos produtiva, e por isso é aplicada em apenas 20% dos casos.

A colheita da cana-de-açúcar deve ser realizada quando as mudas possuírem o maior teor possível de açúcar. O pico de maturação das culturas ocorre entre novembro e abril na região Nordeste, e entre abril e novembro na Região Centro-Sul, segundo a Ageitec (Agência Embrapa e Informação Tecnológica). Das técnicas de colheita existentes, é possível citar o método tradicional com colheita manual e a colheita mecanizada. Essa última é predominante nos estados de São Paulo e Mato Grosso, mas ainda há desafios técnicos para reduzir o seu impacto na produtividade das safras futuras, devido a danos nas mudas e perdas devido à contaminação da cana com terra e impurezas minerais (Braunbeck e Magalhães, 2006).

A prática de queimar parte da plantação para facilitar a colheita e a limpeza do canavial, conhecida como queimada, têm sido desencorajada e rigorosamente controlada pela legislação, devido aos prejuízos à fertilidade do solo e aos riscos de incêndios em regiões vizinhas. Uma discussão completa dos impactos ambientais e socioeconômicos das queimadas pode ser encontrada em Ronquim (2010).

3.3.2. Processamento da cana nas usinas

As etapas iniciais de processamento da cana-de-açúcar colhida são as mesmas, tanto para a produção de álcool quando para a produção de açúcar, até a extração do caldo de açúcar. Começa-se por lavar a cana para reduzir suas impurezas: se o corte foi realizado manualmente, realiza-se a lavagem com água; se o corte foi mecânico, não se pode utilizar água pois isso acarretaria em uma perda elevada de sacarose, o que motivou o desenvolvimento de sistemas de limpeza com jatos de ar.

Em seguida, a cana-de-açúcar é picada e passa por uma máquina chamada desfibrador, que retira as fibras do produto de modo a expor as células ricas em sacarose. Depois de passar pelo desfibrador, a cana é comprimida entre dois rolos para a extração do seu caldo, em um processo conhecido como moagem.

Após a moagem, separa-se o bagaço do caldo da cana. O bagaço representa de 25% a 30% do total em massa da cana moída, e parte dele será utilizado nas caldeiras como combustível. Já o caldo é filtrado e tratado quimicamente para eliminação de impurezas e correção do pH. Após o tratamento, poderá ser destinado para a produção de açúcar ou de etanol.

3.3.3. Produção de açúcar e etanol

As usinas de cana-de-açúcar possuem duas seções diferentes que podem receber o caldo tratado: uma fábrica de açúcar e uma destilaria de etanol. O gestor de produção da usina decide qual será a fração do caldo alocada para cada uma das duas seções, de forma a obter uma certa proporção entre a produção final de açúcar e de etanol.

Na fábrica de açúcar, realiza-se a evaporação do caldo, que perde cerca de três quartos da sua água e se transforma em um xarope concentrado. Esse xarope passa por um processo de cozimento, durante o qual a sacarose se cristaliza, resultando em uma massa que será centrifugada e lavada. O açúcar centrifugado passa em seguida por um processo de secagem e é levado enfim para ser estocado e ensacado.

Nas destilarias de etanol, o caldo de açúcar passa por um processo de fermentação em tanques, com o uso de leveduras (*Saccharomyces cerevisiae*), que se alimentam do açúcar produzindo etanol e gás carbônico. O etanol produzido por esses microrganismos é destilado,

chegando a um teor alcoólico de 96%, que é o do álcool hidratado, utilizado como combustível. Para produzir o álcool anidro, é necessário desidratar esse líquido, chegando a 99.5% de teor alcoólico. O álcool anidro é vendido para ser misturado à gasolina.

3.3.4. Cogeração e excedente de energia

Toda a energia necessária para o funcionamento de uma usina de cana-de-açúcar pode ser produzida pela queima do bagaço de cana, separado na etapa de moagem. Essa queima é realizada em caldeiras, gerando vapor e energia elétrica que são utilizados no acionamento e alimentação das máquinas. Esse processo de conversão da energia térmica do bagaço da cana-de-açúcar em energia mecânica e elétrica é denominado cogeração.

A cogeração nas usinas de cana-de-açúcar produz ainda um excedente de energia que pode ser vendido a concessionárias, constituindo uma fonte de renda adicional. Boa parte das máquinas e equipamentos utilizados nas usinas são fabricados nacionalmente e otimizados para aproveitar a energia da queima do bagaço de cana-de-açúcar.

Segundo Dantas (2009), a eficiência do processo de cogeração aumenta com a elevação da pressão e da temperatura nas caldeiras e nos ciclos de vapor. Atualmente, os sistemas mais eficientes operam com vapor a 65 bar e uma temperatura de 480ºC. Existem caldeiras no mercado brasileiro que conseguem operar a temperaturas próximas de 520ºC, mas há uma limitação tecnológica para passar desse valor, já que os aços fabricados no Brasil não resistem a temperaturas maiores.

Simulações mostram que é possível chegar a um excedente de energia entre 90kWh e 120kWh por tonelada de cana-de-açúcar em um sistema otimizado de cogeração (Nova Cana, 2018). Sabendo que a produção de cana-de-açúcar no Brasil em 2016 foi de cerca de 691 milhões de toneladas, e que nesse ano a energia total produzida no país esteve na ordem de 900.000GWh (EME, 2017), é possível estimar que a energia gerada pelo bagaço de cana-de-açúcar tem potencial para representar até 9% da matriz brasileira.

3.3.5. Cadeias de distribuição

Santos et al. (2013) e Milanez et al. (2010) estudaram as cadeias de distribuição do açúcar e do etanol no brasil. Os principais atores identificados foram:

1. Fornecedores de cana-de-açúcar: embora muitas usinas possuam suas próprias fazendas de plantação de cana, existem também produtores independentes que possuem contratos com as usinas;
2. Produtores (usinas): construídas para produzir tanto etanol quanto açúcar, são frequentemente possuídas por grupos sucroalcooleiros, entre os quais se destacam o grupo Raízen Energia S/A (24 usinas localizadas na região Centro-Sul) e Biosev (12 usinas localizadas tanto na região Centro-Sul como na região Nordeste);
3. Compradores (exportação): nas cadeias logísticas de etanol e açúcar para exportação, existem empresas que fazem a intermediação entre as usinas e as importadoras. Essas empresas são chamadas de *tradings*, e entre elas se destaca a Copersucar, a maior exportadora de açúcar do mundo e uma das maiores exportadoras de etanol (Neves et al., 2016);
4. Compradores (mercado interno): existem atacadistas e indústrias alimentícias que compram o açúcar diretamente das usinas. Já para o etanol, é proibida a venda direta das usinas para os postos de combustível: é necessário passar por intermediários, conhecidos como distribuidoras

Os modais de transporte utilizados para o açúcar são o rodoviário e o ferroviário, sendo predominante o primeiro. Para o etanol, é utilizado majoritariamente o modal rodoviário (90% do volume transportado). Segundo o site especializado Nova Cana, esse domínio do modal rodoviário é uma limitação para o escoamento de etanol para exportação. Janotti et al. (2012) apontam a falta de vagões-tanque especializados como um fator impeditivo para a utilização do modal ferroviário (embora existam projetos para contornar essa dificuldade) e afirmam que a melhor opção para o transporte do etanol é a utilização de dutos.

O acondicionamento de açúcar pode ser feito a granel (no interior de silos), em sacos de 50 kg ou em *big-bags*, que são sacos de polipropileno de 1200 kg. A sua estocagem pode ser realizada na própria usina, em armazéns intermediários ou em terminais portuários. Os armazéns intermediários são utilizados principalmente pelos *traders* para controlar o momento de venda, e se situam próximos às usinas. Alguns grupos sucroalcooleiros recorrem menos a armazéns intermediários porque possuem tanto usinas como unidades compradoras

da cadeia de produção, e com isso conseguem alinhar melhor produção e demanda e gerenciar seus níveis de estoque de acordo com sua capacidade.

O etanol é armazenado em tanques especializados, nas usinas ou nos terminais portuários. Segundo Janotti et al. (2012), existem poucos armazéns intermediários para esse produto, o que é mais um fator que dificulta a utilização dos modais ferroviário e hidroviário, já que eles são necessários para a consolidação da carga nesse contexto.

4. DADOS E METODOLOGIA

4.1. Séries de preços de açúcar e etanol

A fonte dos dados desse estudo é o Centro de Estudos Avançados em Economia Aplicada (Cepea). Esse órgão faz parte da Escola Superior de Agricultura Luiz de Queiroz (Esalq/USP), localizada em Piracicaba, no estado de São Paulo, e é constituído por professores e pesquisadores de mestrado e doutorado com conhecimentos sobre agronomia, economia, estatística, entre outros. Esse grupo está em contato com cerca de 6 mil profissionais do agronegócio, que compartilham informações detalhadas sobre sua área de atuação, constituindo uma fonte confiável e valiosa de dados.

No site do Cepea é possível consultar séries históricas de vários tipos de produtos agropecuários, como açúcar, arroz, café, etanol, leite, soja, entre outros. Dentro de cada categoria há diferentes escolhas de séries de preços e de indicadores. No caso desse estudo, estamos interessados nos preços dos principais derivados da cana-de-açúcar: álcool e etanol. Para o etanol, é possível escolher:

1. Natureza do produto: etanol hidratado, anidro ou outros fins;
2. Localidade geográfica: estados de SP, AL, PE, PB, MT, PE ou GO;
3. Nível de agregação: semanal, mensal ou anual

Como o alvo do estudo são os produtores do estado de SP e o seu objetivo é utilizar dinâmicas históricas de flutuação a médio prazo, optou-se por escolher um **nível de agregação mensal, com dados do estado paulista**. O produto escolhido foi o etanol hidratado, que é produzido em grandes volumes para ser utilizado como combustível em veículos a álcool ou *flex*.

Os preços estão expressos em R$/L e correspondem aos valores sem frete, a retirar na unidade de produção. Isso é importante porque significa que esse indicador reflete os preços percebidos pelo produtor, e não em outro ponto da cadeia logística. Os levantamentos são feitos diariamente junto a unidades produtoras, intermediários de vendas e distribuidoras. Para o cálculo do indicador mensal, os valores são ponderados pelo volume de comercialização.

Os dados foram coletados de abril de 2006 a abril de 2018, totalizando 145 observações mensais. A Figura 10 mostra a série original disponível no site do Cepea. É possível observar uma tendência de crescimento com o passar do tempo, e oscilações periódicas que pressupõem a existência de sazonalidade.

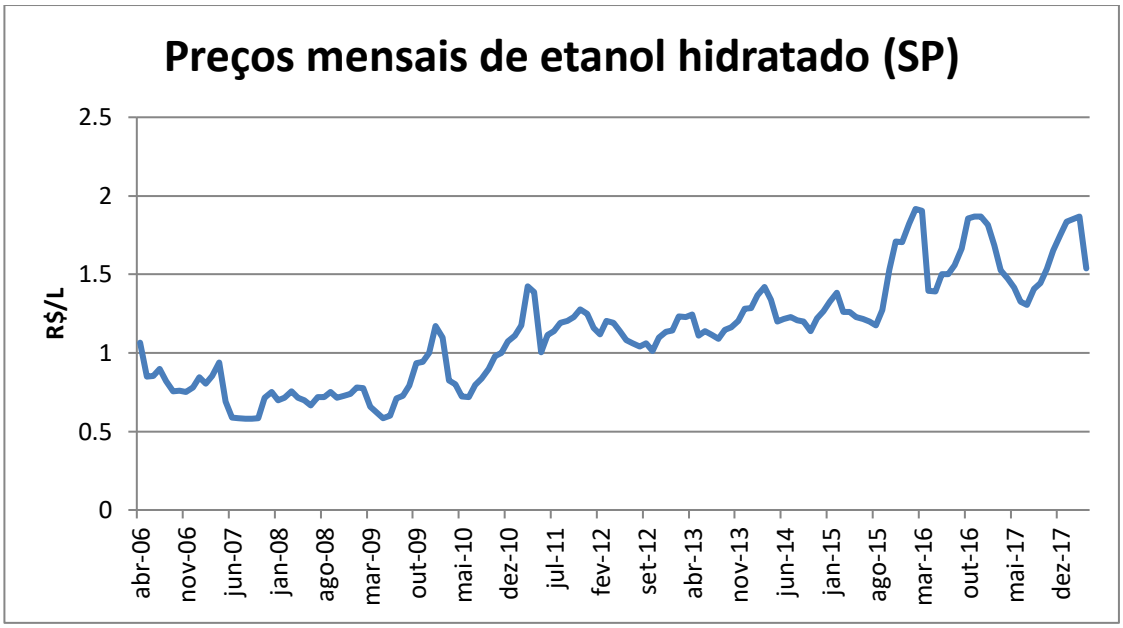

Figura 10: Preços mensais de etanol hidratado no estado de SP, de abril de 2006 a abril de 2018. Fonte: Cepea (2018)

Já para o açúcar, no site da Cepea, é possível escolher a natureza do produto (açúcar cristal ou refinado amorfo), a localidade geográfica (SP, AL, PB ou PE) e o nível de agregação dos dados (diária, semanal, mensal ou anual). Será utilizada a série de preços de **açúcar cristal**, pois trata-se de um produto produzido em grandes volumes para alimentar o mercado interno brasileiro (o açúcar refinado amorfo é produzido principalmente para exportação). Serão utilizados **dados agregados mensalmente**, ponderados com relação ao volume de comercialização, e referentes ao estado de SP, de forma a estares alinhados com os dados de etanol.

Os preços de açúcar cristal estão em R$/saca de 50kg. Os valores estão indicados sem frete, com retirada na unidade de produção, e são referentes a um produto de alta qualidade (polarização mínima de 99.7 graus, máximo de 0.1% de umidade e 0.07% de cinzas) ensacado em sacas de polipropileno. A Figura 11 mostra a série de preços de abril de 2006 a abril de 2018 (145 valores). É possível observar um crescimento nos preços com o passar do tempo, e as sazonalidades são menos presentes do que na série de etanol.

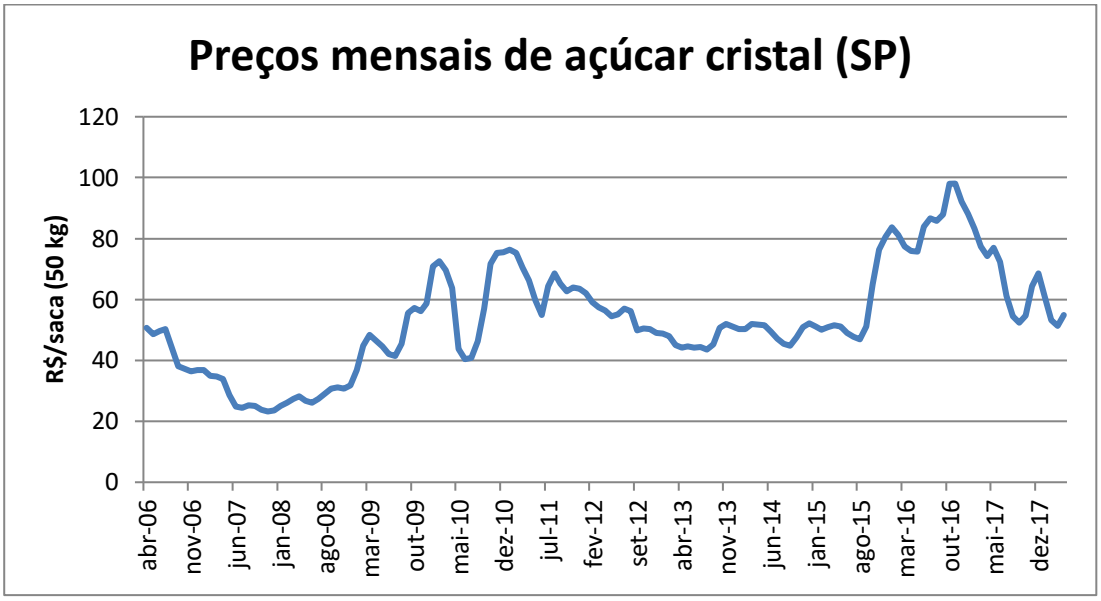

Figura 11: Preços mensais de açúcar cristal no estado de SP, de abril de 2006 a abril de 2018. Fonte: Cepea (2018)

4.1.1. Tratamento e exploração dos dados

Embora os dados mostrem uma tendência crescimento a longo prazo, é importante diferenciar entre crescimento nominal e crescimento real. Como a série escolhida está distribuída em um período de 12 anos, os dados serão corrigidos segundo o IPCA (Índice Nacional de Preços ao Consumidor Amplo), calculado pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O IPCA é o índice oficial do governo brasileiro e do Banco Central para medição da inflação e sua série histórica pode ser consultada no site do IBGE.

Segundo a série histórica do IPCA, os preços no Brasil subiram em média 5,6% ao ano de abril de 2006 a abril de 2018. Isso representa um crescimento acumulado de 92,1% no período, e mostra que a inflação é, pelo menos parcialmente, responsável pela tendência crescente nos dados originais.

A Figura 12 mostra as séries de etanol hidratado e açúcar cristal corrigidas com relação à inflação (o mês de referência para a correção foi janeiro de 2001). É possível observar que, para as duas séries, a tendência de crescimento foi eliminada. Isso é importante porque retira uma fonte conhecida de variação dos dados e elimina a necessidade de utilizar um termo de crescimento nos modelos, reduzindo o número de variáveis a serem estimadas e aumentando a precisão das previsões.

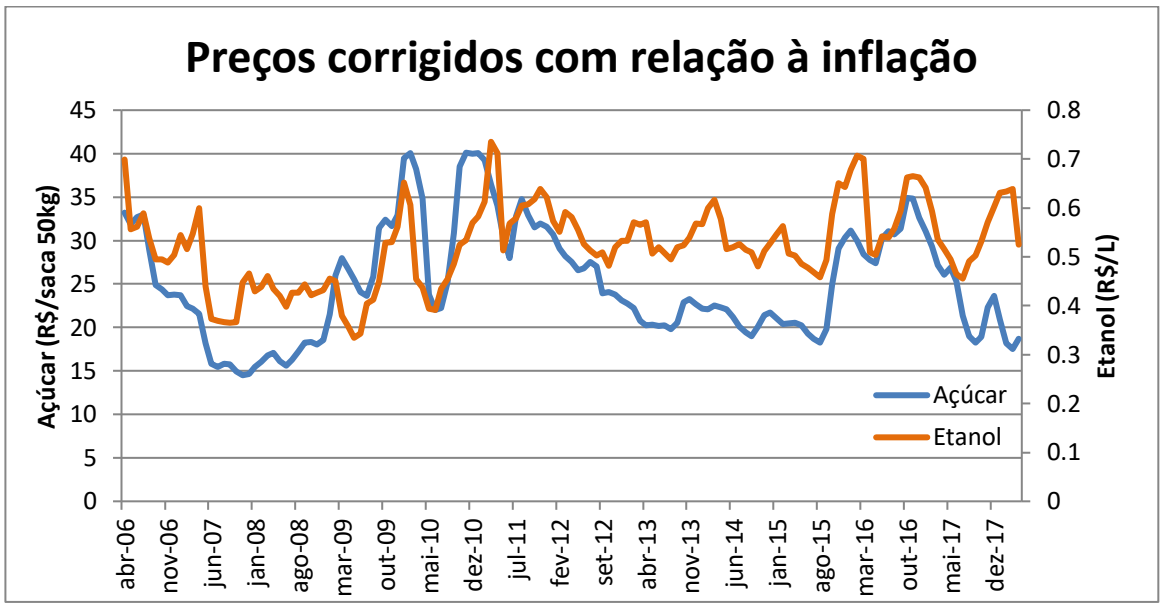

Figura 12: Preços de açúcar (azul) e etanol (laranja) corrigidos com relação à inflação, de abril de 2006 a abril de 2018. O mês de referência é janeiro de 2001. Fonte: elaborado pelo autor

A Tabela 6 mostra os valores de algumas estatísticas descritivas básicas (média, desvio padrão, mediana, mínimo e máximo) para as duas séries. Para o etanol, os valores de média e mediana são muito próximos, indicando que a distribuição de preços é bem centralizada, com oscilações acima ou abaixo da média igualmente prováveis. Já para o açúcar, a mediana é inferior à média, indicando a existência de uma assimetria na distribuição de preço, gerada pelos altos valores observados entre o final de 2009 e o começo de 2011.

	Açúcar (R$/saca 50kg)	Etanol (R$/L)
Média	25.04	0.523
Desvio Padrão	6.47	0.083
Mediana	23.73	0.520
Mínimo	14.50	0.334
Máximo	40.11	0.735

Tabela 6: Estatísticas descritivas para as séries de açúcar e etanol corrigidas com relação à inflação

Como açúcar e etanol são derivados da mesma matéria-prima, é natural suspeitar que as variações de seus preços estejam de certa forma correlacionadas. O exame atento da Figura 12 reforça essas suspeitas: em muitas situações observa-se variações na mesma direção das duas séries de preços, como durante os anos de 2010 e 2016. Uma forma de quantificar essa impressão é calcular o coeficiente de correlação de Pearson *r*, definido segundo a equação:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

onde x_i e y_i são as séries consideradas (no nosso caso, açúcar cristal e etanol hidratado) e $\bar{x}$, $\bar{y}$ denotam os seus valores médios. O coeficiente *r* varia de -1 (correlação linear negativa perfeita) a +1 (correlação linear positiva perfeita), sendo nulo quando não há correlação linear entre as variáveis consideradas.

O valor de *r* medido entre as séries de açúcar cristal e etanol hidratado é de 0.54, indicando uma correlação linear positiva moderada. Os modelos desenvolvidos nesse trabalho tentarão tirar vantagem da correlação entre as duas variáveis, utilizando os valores de uma para tentar prever a outra.

4.2. Mercados futuros

4.2.1. Motivação

Muitos autores, como Bowman e Husein (2014) ou ainda Zhang e Liu (2018), argumentam que os contratos futuros ajudam a prever as variações de preços de algumas *commodities*. Um contrato futuro é uma obrigação legal de comprar ou vender um certo produto a um preço predeterminado em um momento especificado no futuro. Esses contratos são padronizados e negociados em bolsas de valores, que garantem a qualidade dos produtos e a execução das obrigações.

Um dos modelos clássicos de precificação de contratos futuros, utilizado por exemplo em Reichsfeld e Roache (2011), assume que o preço no instante τ do contrato futuro de uma *commodity* com expiração em $t = T$, que denotaremos $F(T, \tau)$, é igual ao preço esperado dessa *commodity* em $t = T$ (denotado por $\mathrm{E}(\mathrm{S}(T))$) corrigido pela taxa interna de retorno ρ, que representa o custo de oportunidade de investir em outro ativo. Essa relação é expressa pela seguinte equação:

$$F(T, \tau) = e^{\rho(T-\tau)}\mathrm{E}(\mathrm{S}(T))$$

Segundo essa equação, é possível utilizar os preços dos contratos futuros com expiração em T para prever os preços de uma *commodity* no instante T. Além disso, à medida

que o tempo passa e que nos aproximamos de T, os preços dos contratos futuros tendem a convergir para o valor real de $S(T)$.

É comum construir séries de preços de valores futuros para um determinado ativo usando os contratos de mais próxima expiração. Por exemplo, para a série de primeiros futuros, são utilizados os contratos futuros de data de expiração T_1 mais próxima; para a série de segundos futuros, é usada a segunda data de expiração T_2 mais próxima, e assim por diante.

4.2.2. Contratos futuros de etanol hidratado

A referência de mercado futuro para os produtores de etanol hidratado no Brasil é a Bolsa de Valores, Mercadorias e Futuros de São Paulo (BM&F Bovespa, chamada de B3 desde 2017). A BM&F Bovespa é a bolsa de valores oficial do Brasil, com origens que datam de 1890 e uma capitalização de mercado superior a $1 trilhão, que a coloca entre as 20 maiores do mundo (World Federation of Exchanges, 2018).

A BM&F Bovespa possui contratos futuros de etanol hidratado com expiração para todos os meses do ano. Sendo assim, os primeiros futuros são sempre os contratos futuros de expiração no mês seguinte, e iremos verificar se a sua série histórica tem o potencial para prever os preços de etanol hidratado com um mês de antecedência.

A Figura 13 mostra as séries de preços de etanol hidratado (Cepea) e primeiros futuros de etanol hidratado (BM&F), de junho de 2010 a abril de 2018. É possível ver que as duas séries são muito correlacionadas: em alguns momentos os preços futuros são superiores aos preços de etanol e em outros o comportamento é invertido, mas ao longo de toda a série histórica os valores se mantiveram muito próximos. Os valores foram corrigidos com relação à inflação.

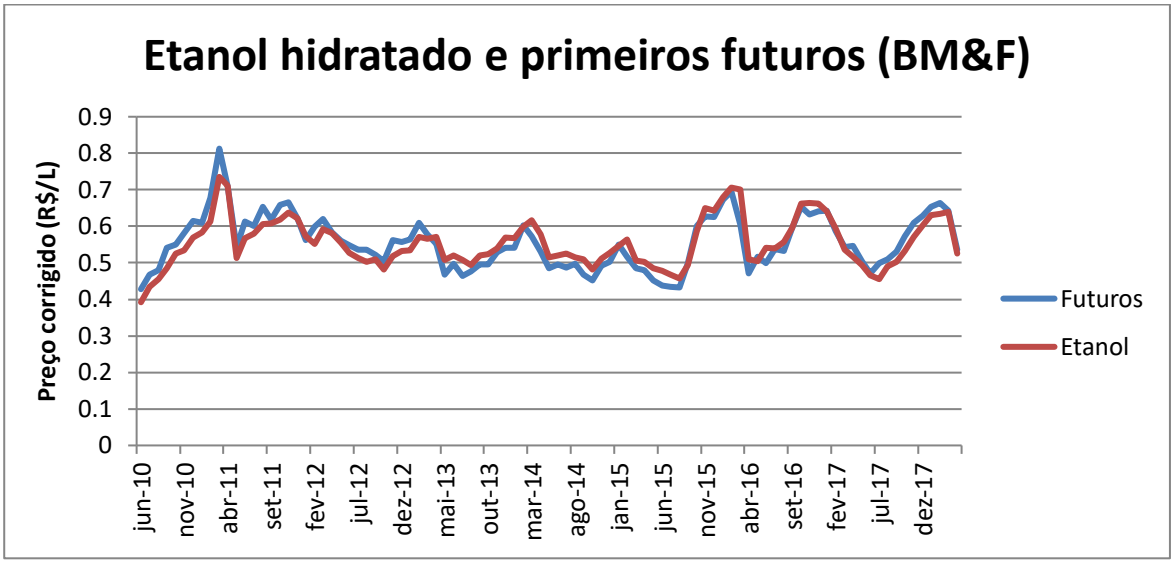

Figura 13: Séries de etanol hidratado (vermelho) e de primeiros futuros da BM&F (azul), de junho de 2010 a abril de 2018. Valores corrigidos com relação à inflação. Fonte: elaborado pelo autor

No entanto, para efeitos de poder preditivo, não basta haver uma alta correlação entre duas séries: mesmo que isso seja verdade, seria preciso conhecer os valores futuros de uma para prever a outra, o que implica em modelar e prever as duas! Por isso, para estudar a inclusão de variáveis exógenas no modelo, será utilizada a função de correlação cruzada (CCF), cuja definição é inspirada no coeficiente de correlação de Pearson *r*. A função de correlação cruzada $\rho_{xy}(h)$ entre duas séries temporais *x* e *y* é definida segundo a expressão:

$$\rho_{xy}(h) = \begin{cases} \dfrac{\sum_{i=1}^{n-h}(x_{i+h} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1+h}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n-h}(y_i - \bar{y})^2}}, se\ h \geq 0 \\ \dfrac{\sum_{i=1-h}^{n}(x_{i+h} - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n+h}(x_i - \bar{x})^2}\sqrt{\sum_{i=1-h}^{n}(y_i - \bar{y})^2}}, se\ h < 0 \end{cases}$$

O valor de h corresponde ao deslocamento temporal (defasagem) da série x_t. Para $h = 0$, a função retorna o valor do coeficiente de correlação de Pearson *r* para as duas séries. Para $h < 0$, a função retorna a correlação entre valores passados de x_t e valores presentes de y_t. Para $h > 0$, encontramos o valor da correlação entre valores passados de y_t e valores presentes de x_t.

É importante destacar que, contrariamente às funções de autocorrelação (ACF) e autocorrelação parcial (PACF), a CCF não é simétrica com relação a h ($\rho_{xy}(h) \neq \rho_{xy}(-h)$). Além disso, a ordem de x_t e y_t é importante: $\rho_{xy}(h) = \rho_{yx}(-h)$.

Para esse estudo, consideraremos sempre x_t como a variável endógena e y_t como a variável exógena. Sendo assim, estamos interessados nos valores de $\rho_{xy}(h)$ quando $h > 0$: se a CCF apresentar um pico nessa região, isso significa que valores passados da variável exógena estão correlacionados com valores presentes da variável endógena, indicando poder preditivo.

Para aplicar a função de correlação cruzada, é necessário que as duas séries sejam estacionárias. Para eliminar tendências de curto prazo e a influência de fatores autorregressivos no resultado, será sempre realizada uma operação prévia de diferenciação temporal nas duas séries: $\nabla y_t = y_t - y_{t-1}$ e $\nabla x_t = x_t - x_{t-1}$.

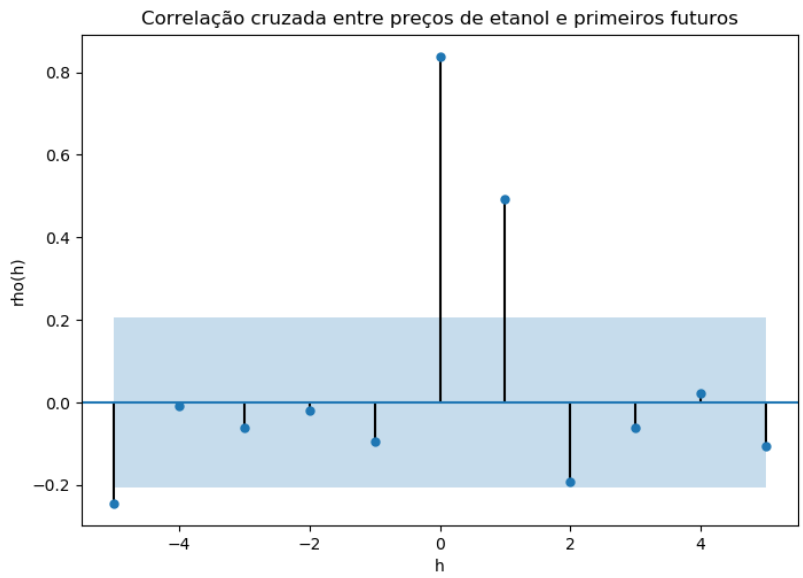

Figura 14: CCF entre as séries diferenciadas de preços de etanol e contratos futuros de primeira expiração da BM&F. Fonte: elaborado pelo autor

A Figura 14 mostra o gráfico da CCF entre a série de etanol hidratado e a série de primeiros futuros da Bovespa. É possível ver um pico em $h = 0$, que confirma a alta correlação entre as duas séries observada na Figura 13, e outro pico em $h = 1$ que sugere que os valores passados dos primeiros futuros têm poder preditivo sobre a série de preços de etanol hidratado.

4.2.3. Contratos futuros de açúcar

Para o açúcar cristal, a BM&F Bovespa possui em teoria contratos futuros, mas não há liquidez pois pouquíssimos acordos são fechados mensalmente. Por isso, utilizaremos os dados de contratos futuros de açúcar da Bolsa de Chicago (CME - Chicago Mercantile Exchange), o maior mercado de opções e contratos futuros do mundo, incluindo muitas *commodities*.

O contrato de futuros escolhido foi o do açúcar nº 11, considerado referência mundial para preços de açúcar ao redor do mundo. Esses contratos possuem expirações nos meses de março, maio, julho e outubro, e cada contrato representa 112 mil libras de açúcar bruto (aproximadamente 50,8 toneladas). Os valores, originalmente em dólares, serão convertidos para reais segundo a série de taxas de câmbio mensais emitida pelo Cepea e corrigidos com relação à inflação.

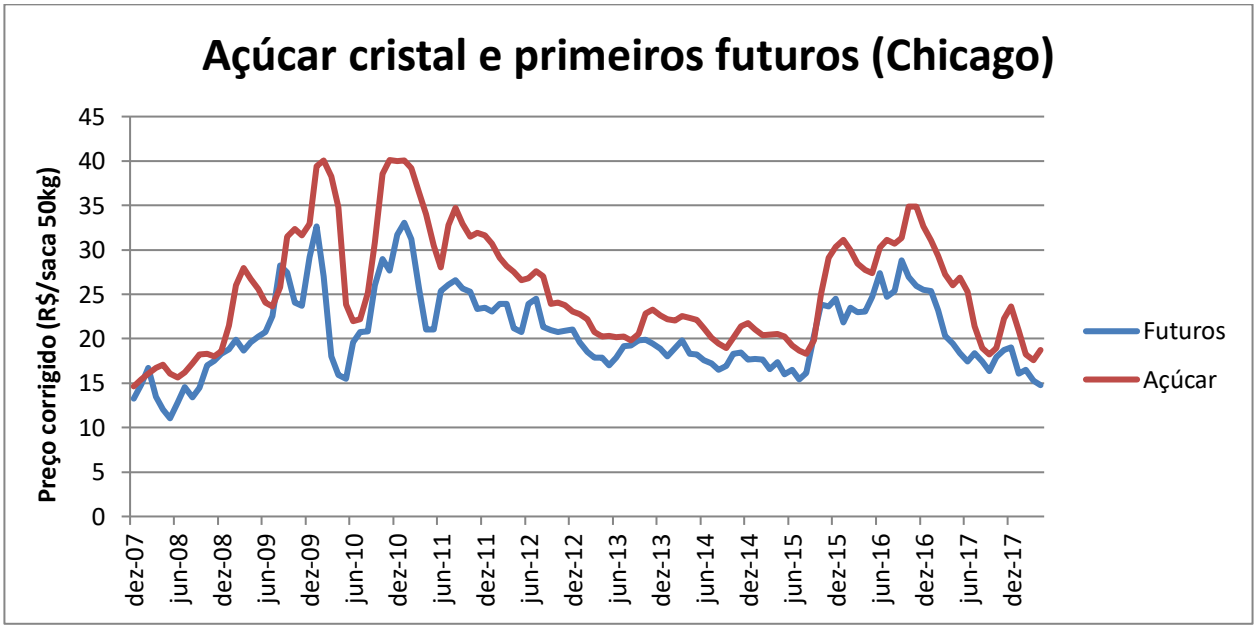

Figura 15: Séries de açúcar cristal (vermelho) e de primeiros futuros de açúcar nº 11 da bolsa de Chicago (azul), de dezembro de 2007 a abril de 2018. Valores corrigidos com relação à inflação. Fonte: elaborado pelo autor

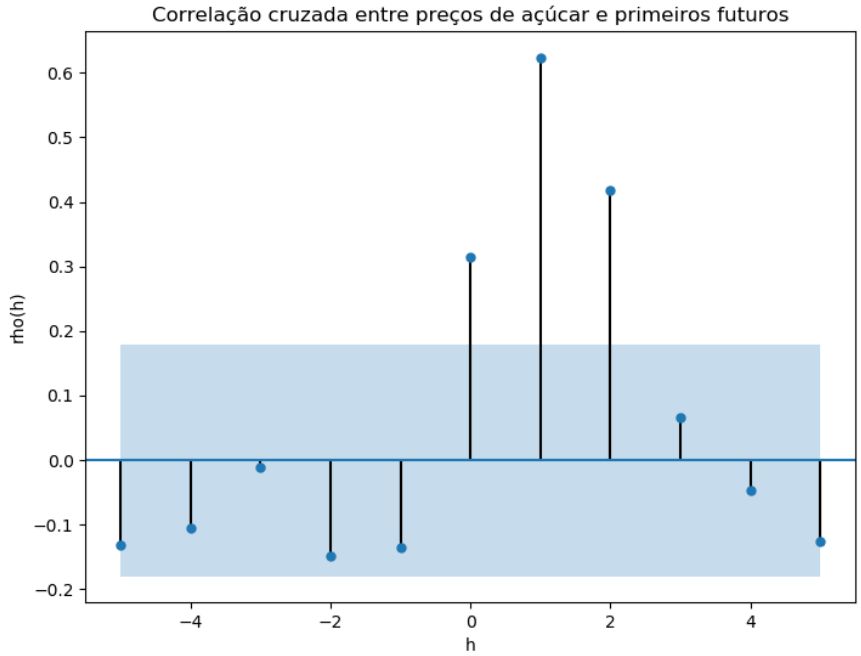

Figura 16: CCF das séries diferenciadas de preços de açúcar cristal e contratos futuros de açúcar nº 11 da Bolsa de Chicago. Fonte: elaborado pelo autor

A análise das Figuras 16 e 17 mostra que existe uma alta correlação entre os valores passados dos primeiros futuros de Chicago e valores presentes dos preços de açúcar cristal para produtores de SP. É interessante notar que a relação temporal ($h = 1$ e $h = 2$) é muito mais forte do que a relação presente ($h = 0$), confirmando o forte caráter influenciador dos futuros de Chicago nos preços de açúcar ao redor do mundo. Usaremos, então, valores passados dos primeiros futuros de Chicago como variáveis exógenas nos modelos a serem desenvolvidos.

4.3. Outras variáveis exógenas

4.3.1. Preços de combustíveis em São Paulo

Sendo o etanol hidratado um dos combustíveis mais utilizados em solo brasileiro, serão estudadas as suas correlações com os preços de outros combustíveis e derivados de petróleo utilizados em solo nacional. Para isso, a referência utilizada serão as séries históricas da ANP (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis), que contém preços de distribuição e revenda, por estado brasileiro, dos seguintes produtos:

1. Gasolina comum;
2. Óleo diesel não aditivado;
3. Gás natural veicular (GNV);
4. Gás liquefeito de petróleo (GLP)

Foram escolhidos os preços desses produtos no estado de São Paulo, que é a referência para esse trabalho, agregados mensalmente desde 2001. Foi realizada a análise de correlação cruzada (CCF), com os seguintes resultados:

1. Com relação ao etanol hidratado, as CCFs de óleo diesel, GNV e GNP não apresentam nenhum pico de correlação estatisticamente significante. Entre etanol e gasolina comum há altos valores de correlação para $h = 0$ e $h = -1$, o que significa que são na verdade os preços de gasolina que reagem a variações no preço de etanol, e não o contrário. A ausência de picos de correlação na região $h > 0$ implica que a gasolina comum não poderá ser utilizada como variável exógena nos modelos;
2. Com relação ao açúcar cristal, nenhuma das quatro variáveis apresentou CCFs com picos estatisticamente significantes.

Os resultados da análise podem ser vistos nas Figuras 17 a 20. Atenção para as escalas dos eixos *y*, que variam de gráfico para gráfico.

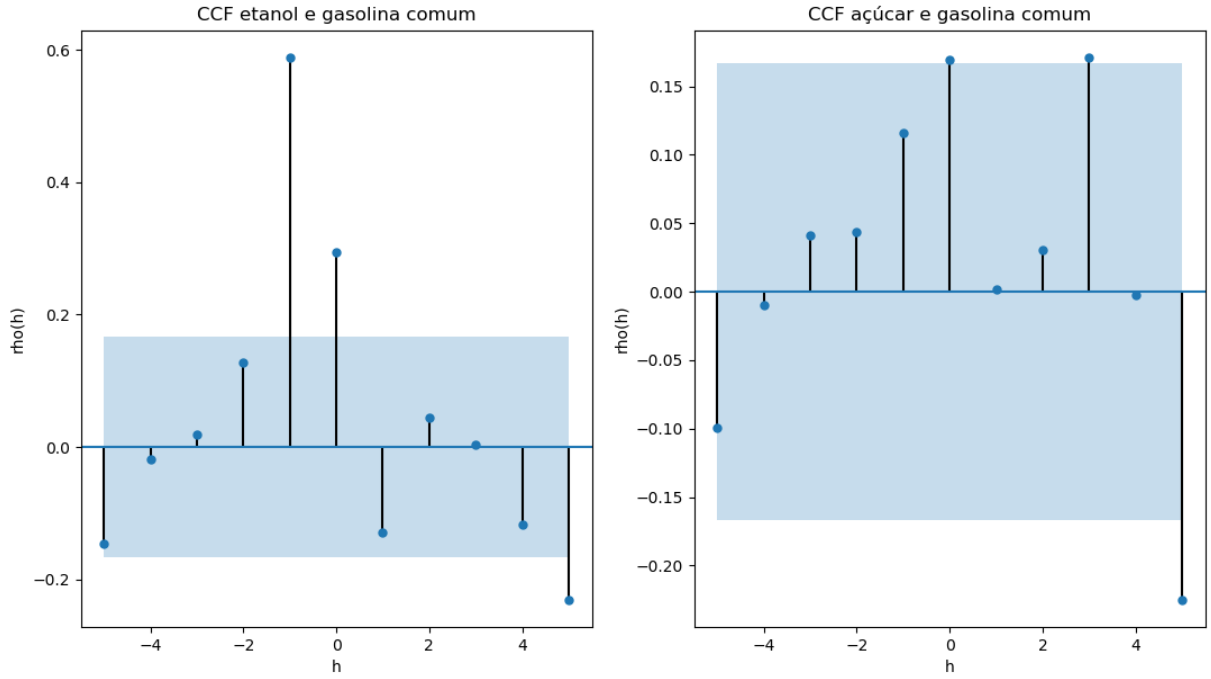

Figura 17: Análise de correlação cruzada entre preços de etanol hidratado e gasolina comum (esquerda), e entre preços de açúcar cristal e gasolina comum (direita). Fonte: elaborado pelo autor

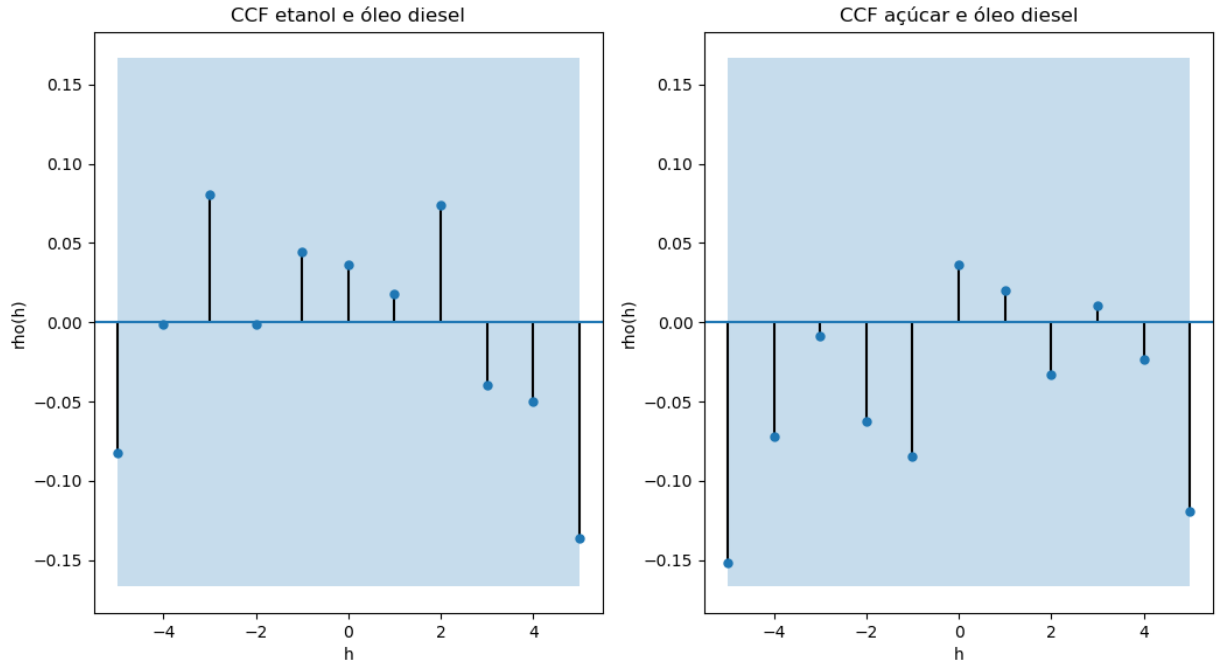

Figura 18: Análise de correlação cruzada entre preços de etanol hidratado e óleo diesel (esquerda), e entre preços de açúcar cristal e óleo diesel (direita). Fonte: elaborado pelo autor

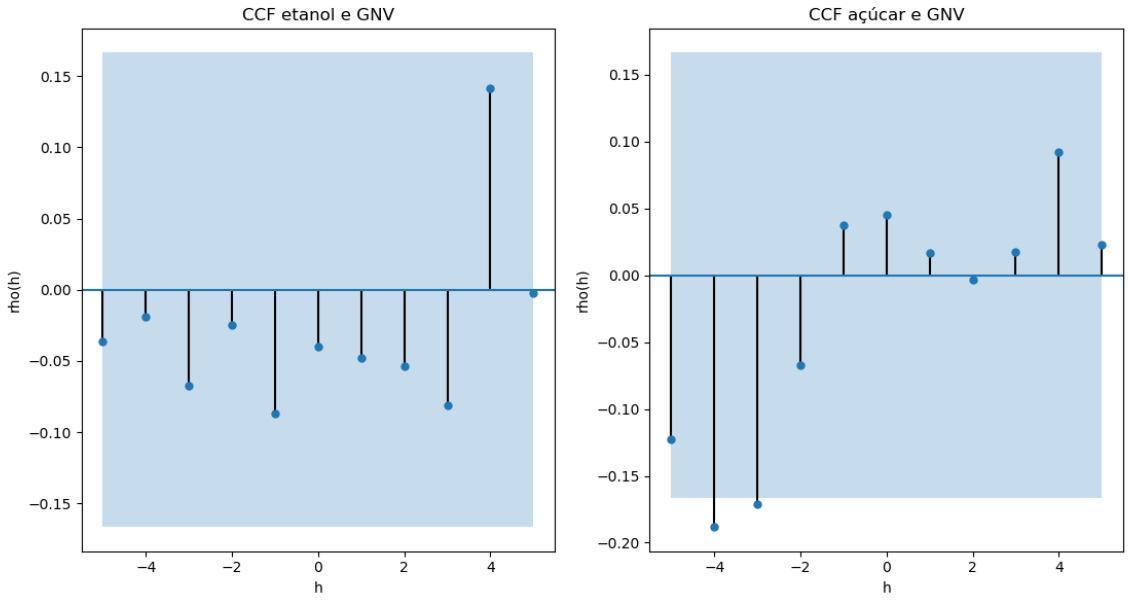

Figura 19: Análise de correlação cruzada entre preços de etanol hidratado e GNV (esquerda), e entre preços de açúcar cristal e GNV (direita). Fonte: elaborado pelo autor

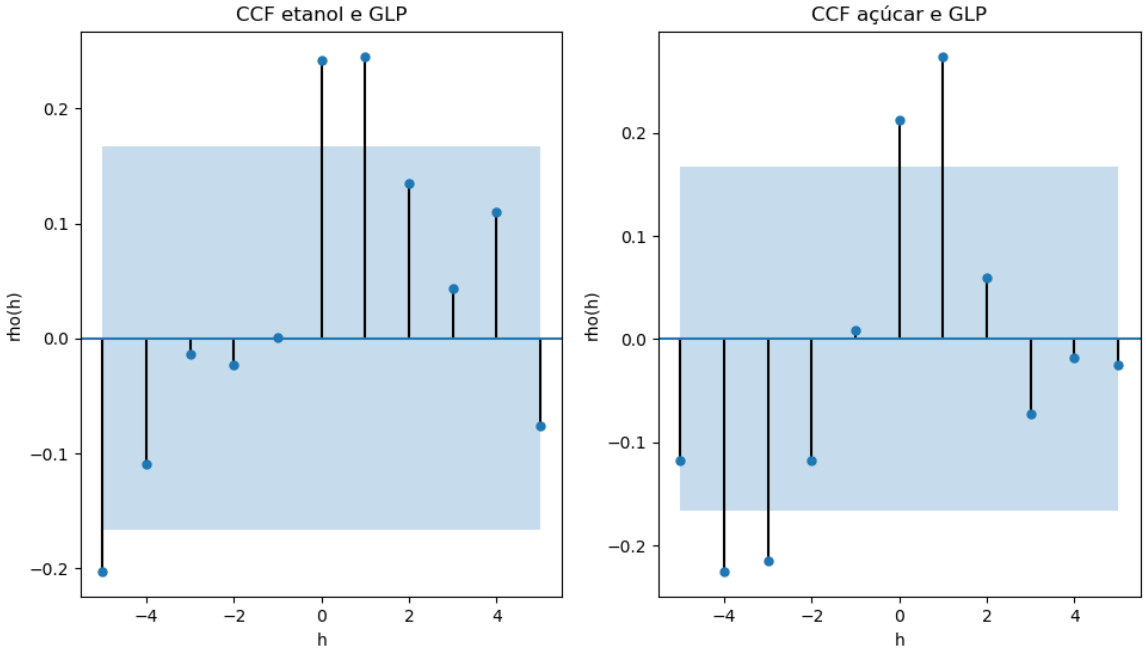

Figura 20: Análise de correlação cruzada entre preços de etanol hidratado e GLP (esquerda), e entre preços de açúcar cristal e GLP (direita). Fonte: elaborado pelo autor

4.3.2. Preços de petróleo internacionais

Outros candidatos a variáveis exógenas são os preços de petróleo bruto no mercado internacional. As duas principais séries utilizadas mundialmente por investidores como referência são:

1. Petróleo Brent, extraído de uma dúzia de plataformas localizadas no mar do Norte, na Europa setentrional. É utilizado na precificação da produção de petróleo da Europa, África e Oriente Médio;
2. Petróleo WTI (West Texas Intermediate), extraído na cidade de Cushing, Oklahoma (Estados Unidos)

As séries de petróleo bruto Brent e WTI foram extraídas do site da U.S. Energy Information Administration (EIA). As Figuras 21 e 22 mostram que não foi encontrado nenhum pico de correlação estatisticamente significante com as séries de etanol hidratado ou açúcar cristal.

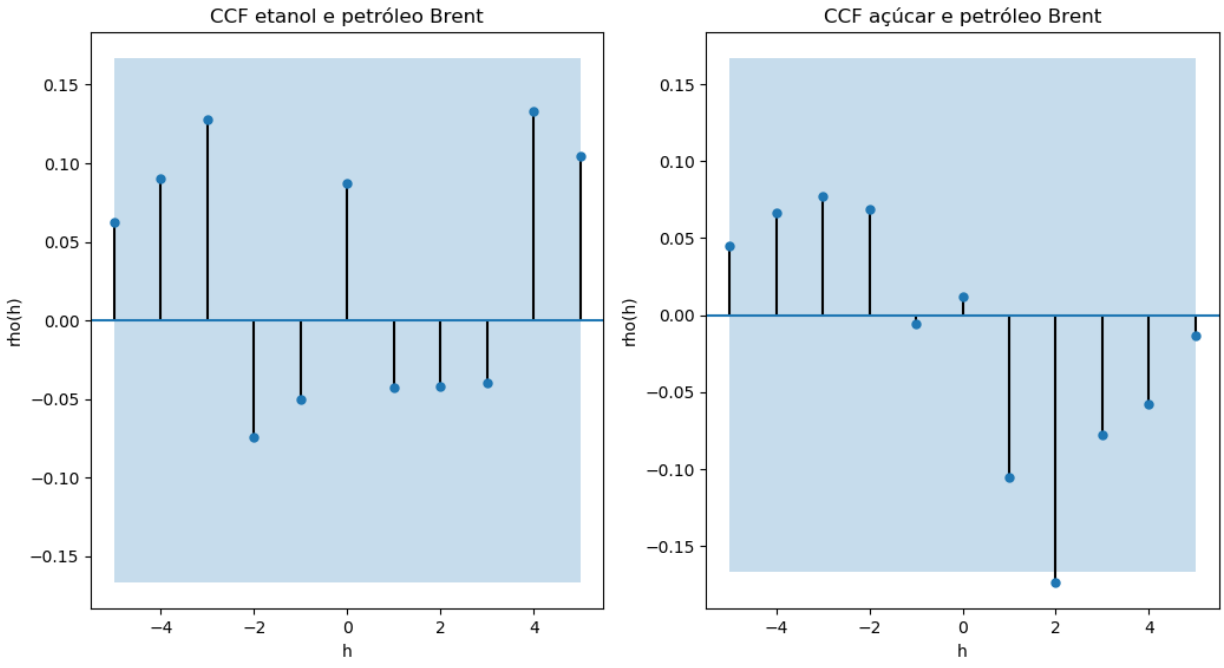

Figura 21: Análise de correlação cruzada entre preços de etanol hidratado e petróleo Brent (esquerda), e entre preços de açúcar cristal e petróleo Brent (direita). Fonte: elaborado pelo autor

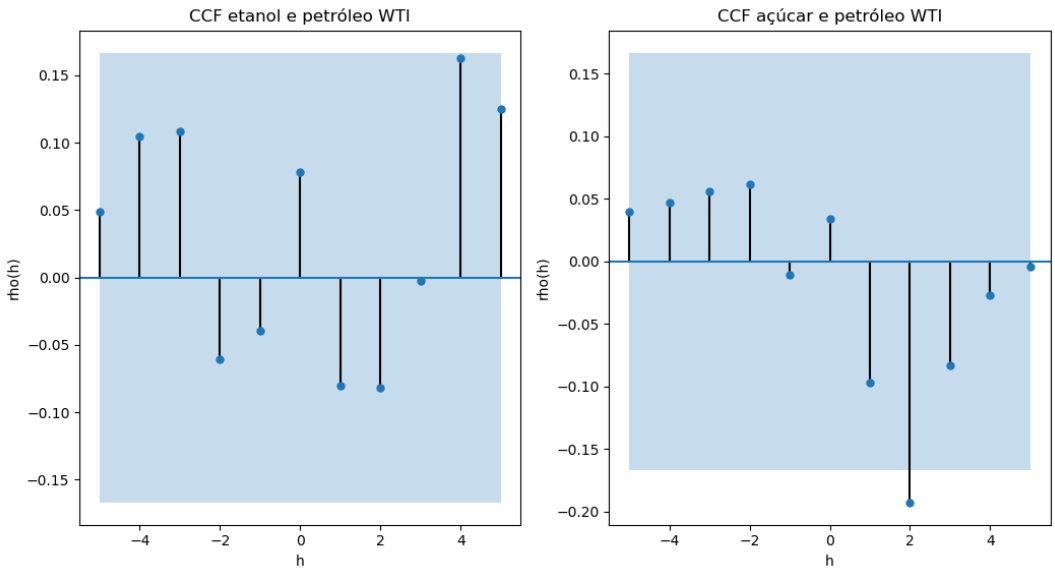

Figura 22: Análise de correlação cruzada entre preços de etanol hidratado e petróleo WTI (esquerda), e entre preços de açúcar cristal e petróleo WTI (direita). Fonte: elaborado pelo autor

4.3.3. Dólar

Finalmente, foram estudadas as correlações com a série de taxas de câmbio dólar/real do Cepea. A Figura 23 mostra a ausência de picos de correlação estatisticamente significantes, que atestam o fraco poder preditivo dessa variável para os preços de açúcar e etanol.

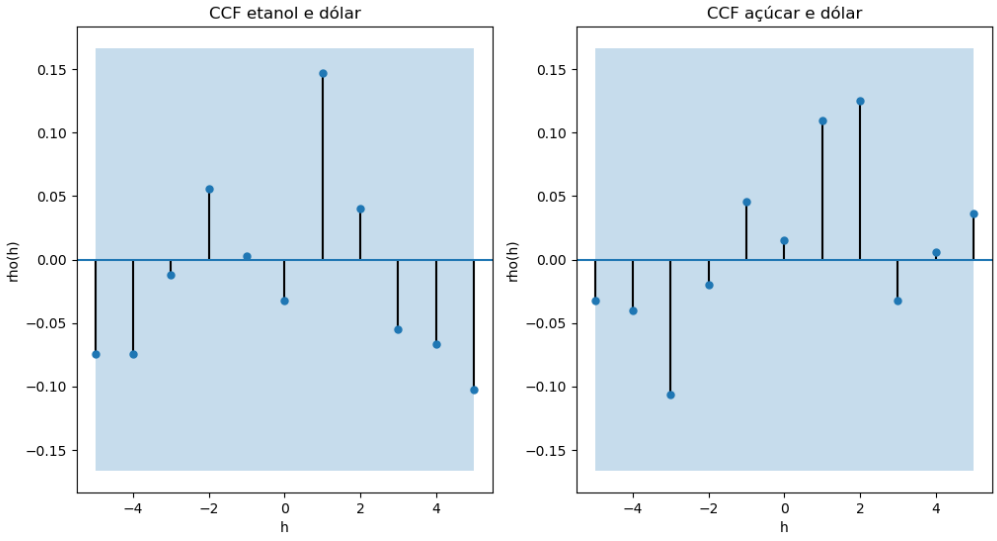

Figura 23: Análise de correlação cruzada entre preços de etanol hidratado e valores do dólar (esquerda), e entre preços de açúcar cristal e valores do dólar (direita). Fonte: elaborado pelo autor

4.4. Modelos a serem desenvolvidos no trabalho

Nesse trabalho, serão desenvolvidas três classes de modelos para os preços de açúcar e etanol. A primeira classe de modelos baseia-se unicamente em valores históricos das séries temporais, utilizando termos autorregressivos, sazonais e de média móvel: são os modelos SARIMA, que serão estendidos também para a sua versão multidimensional (VAR), permitindo interações entre as duas séries mas sem utilizar variáveis exógenas. Os resultados e aprendizados obtidos com a aplicação desses modelos serão utilizados no desenvolvimento dos modelos subsequentes.

A segunda classe é a de modelos de espaço de estado, que dão ao analista a possibilidade de incluir variáveis exógenas e maior controle sobre a formulação matemática e hipóteses do modelo. Por meio da equação de transição, será possível também chegar a uma expressão explícita da matriz de covariância dos preços de açúcar e etanol, de forma a realizar simulações de Monte Carlo para a resolução do problema de gestão de portfólio em usinas de cana-de-açúcar.

Por último, será experimentada uma abordagem não paramétrica e não linear com o uso de redes neurais recorrentes do tipo LSTM. Essa terceira classe de modelos é otimizada para aprender seletivamente dependências de curto e longo prazo em sequências, e será interessante comparar o resultado dessa abordagem mais recente com técnicas clássicas de análise de séries temporais, empregadas nas duas primeiras classes de modelos.

4.4.1. Métricas de avaliação dos modelos

Para avaliar a eficácia dos modelos, serão utilizadas três métricas: raiz quadrada do erro quadrado médio (RMSE), erro médio absoluto (MAE) e erro médio percentual absoluto (MAPE). Sendo y_i os valores reais e $\hat{y}_i$ os valores previstos por um modelo, essas métricas podem ser calculadas da seguinte forma:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{N}}$$

$$MAE = \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{N}$$

$$MPAE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

Devido à presença do termo quadrático, o RMSE penaliza mais severamente erros de previsão de grande magnitude do que o MAE: é possível provar matematicamente que o valor do primeiro será sempre maior ou igual que o valor do segundo. O MPAE é uma versão modificada do MAE que leva em conta o erro relativo de previsão, expresso em termos percentuais.

Essas métricas são também chamadas de *funções de perda*. Quanto mais próximas da realidade forem as previsões, menores são os seus valores, de forma que elas dão ao analista critérios objetivos para julgar os resultados obtidos e comparar diferentes modelos.

4.4.2. Separação de dados de treino e de teste

Segundo Hastie et al. (2009), na elaboração de modelos estatísticos para previsão, é importante separar os dados disponíveis em conjuntos de treino e de teste. O conjunto de treino serve para calibrar os parâmetros do modelo, enquanto que o conjunto de teste serve para avaliar a sua performance. É importante que o modelo não tenha contato com os dados do conjunto de teste durante a sua calibração, pois deseja-se estimar a sua precisão quando for aplicado em uma situação real de previsão, na qual os dados reais são desconhecidos.

No contexto desse trabalho, estão disponíveis dados mensais de preços de açúcar e etanol de abril de 2006 a abril de 2018, totalizando 145 valores. Desses, 75% serão usados para treino e 25% para teste: a performance dos modelos será avaliada sobre os 36 valores mais recentes das séries, medidos entre maio de 2015 e abril de 2018.

Para modelos de redes neurais, existe ainda uma terceira subdivisão, chamada de conjunto de validação. Essa adição é necessária porque redes neurais possuem um alto número de hiperparâmetros a serem otimizados: número de camadas (ou de estados internos), número de épocas de treinamento, porcentagem de *dropout*, entre outros. Contrariamente às duas primeiras classes de modelos, as redes neurais são aproximadores universais (Cybenko, 1989), de forma que, se o modelo tiver uma capacidade suficiente, a taxa de erro no conjunto de treino tenderá a diminuir para níveis arbitrariamente próximos a zero quanto mais complexa for a rede neural e quanto mais extensivo for o seu treinamento.

O conjunto de validação, assim como o conjunto de teste, é formado por dados com os quais a rede neural não teve contato durante a fase de treino, e é utilizado para medir a performance do modelo. A diferença é que serão tomadas decisões de otimização da rede neural sobre esse conjunto, com a escolha dos hiperparâmetros de rede. Como essas escolhas têm alto potencial de influenciar os resultados finais do modelo, a verdadeira performance do modelo é medida sobre o conjunto de teste, que não foi utilizado para treinar o modelo e nem para otimizar os seus hiperparâmetros.

4.4.3. Ambiente de desenvolvimento

Existem muitos *softwares* de modelagem estatística e aprendizado de máquina disponíveis. O autor optou por ferramentas gratuitas e abertas ao público, de forma que os modelos aplicados nesse trabalho possam ser reproduzidos por leitores interessados e que o próprio código fonte possa ser lido e interpretado como parte do trabalho. Para não interromper a leitura, os principais programas escritos pelo autor serão disponibilizados em anexo.

As duas linguagens de programação utilizadas nesse trabalho foram Python e R. Os modelos SARIMA e VAR foram desenvolvidos em Python utilizando a biblioteca *statsmodels* (Perktold et al., 2009). Os modelos de espaço de estado foram escritos em R, utilizando a biblioteca *FKF* (Luethi et al., 2010). Finalmente, as redes neurais LSTM foram implementadas em Python utilizando a biblioteca *Keras* (Chollet et al., 2015).

4.4.4. Correção com relação à inflação e transformação logarítmica

Todas as séries serão convertidas para reais (caso necessário) e corrigidas com relação a inflação, tomando como base o Índice de Preços ao Consumidor Amplo (IPCA) publicado pelo IBGE e o ano de 2001 como referência.

Além disso, será aplicado o logaritmo natural sobre todas as quantidades estudadas. Como explica Feng et al. (2014), essa é uma transformação importante para casos em que a variabilidade dos dados é proporcional à sua magnitude, como ocorre com séries de preços, nas quais mudanças costumam ser percebidas em termos de porcentagens do valor total. Os modelos serão treinados sobre as séries logarítmicas e as suas previsões serão corrigidas com uma transformação exponencial.

5. CONSTRUÇÃO DOS MODELOS E PREVISÕES

5.1. Modelos de séries temporais

5.1.1. SARIMA - Série de preços de açúcar

Utilizando a metodologia de Box e Jenkins (1970), o primeiro passo a ser realizado é a análise das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série de preços para determinar os parâmetros do modelo e investigar a existência de sazonalidade.

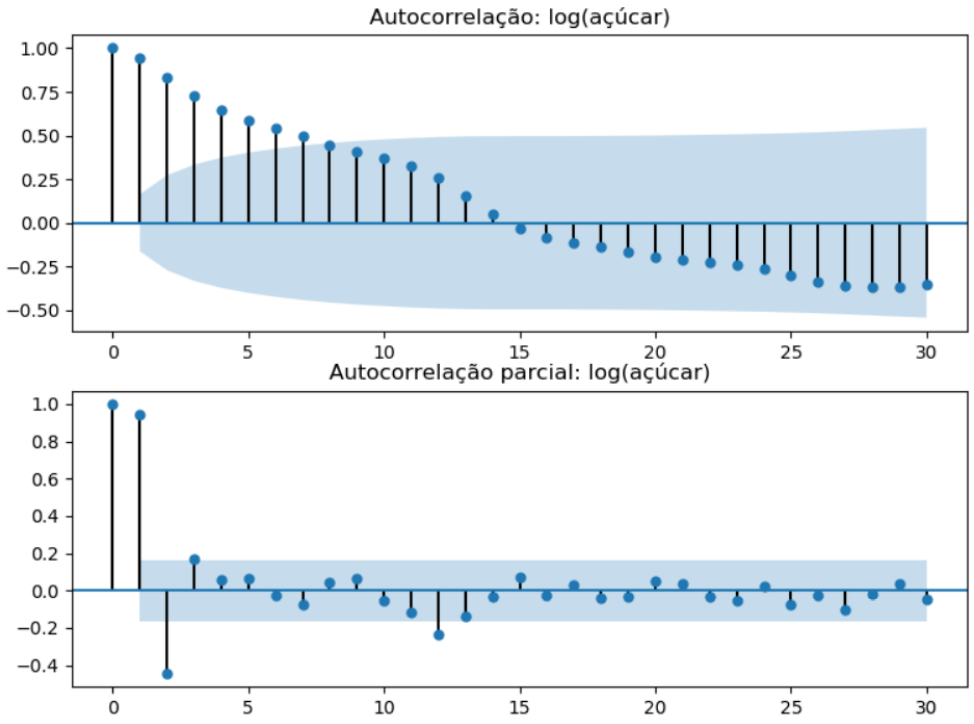

Figura 24: ACF e PACF da série mensal de preços de açúcar sem diferenciação

Analisando a Figura 24, é possível observar que a função de autocorrelação decai lentamente, enquanto que a função de autocorrelação parcial possui picos nos pontos h=1, h=2 e h=12. Esse comportamento sugere a existência de sazonalidade anual (S=12) e um comportamento autorregressivo de segunda ordem (p=2). Não há elementos que indicam a existência de termos de média móvel, de forma que o primeiro candidato a modelo (Modelo 1) é SARIMA(2, 0, 0) x $(1, 0, 0)_{12}$.

No entanto, a alta magnitude do pico em h=1 da função de autocorrelação parcial e o decaimento lento da função de autocorrelação sugerem que a série pode conter um elemento de integração; dessa forma, podemos realizar uma operação de diferenciação $\nabla y_t = y_t - y_{t-1}$ para estabilizar suas ACF e PACF. O resultado dessa operação pode ser visto na Figura 25.

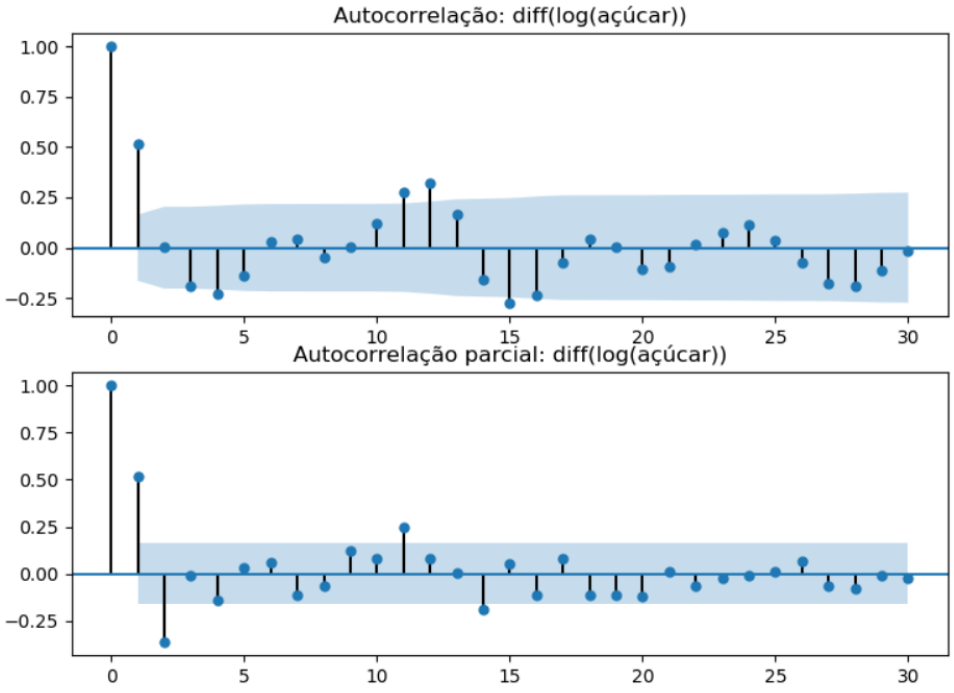

Figura 25: ACF e PACF da série de preços de açúcar após diferenciação

A análise das funções diferenciadas revela picos em h=1 e h=12 na função de autocorrelação, enquanto que persistem os dois picos em h=1 e h=2 na função de autocorrelação parcial. Sendo assim, foram encontrados mais dois modelos candidatos (Modelo 2 e Modelo 3): SARIMA (2, 1, 1) x $(1, 0, 0)_{12}$ e SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$, respectivamente.

Em seguida, os parâmetros dos três modelos foram estimados utilizando o método da máxima verossimilhança. Os resultados estão listados abaixo, nas Tabelas 7 a 9:

Modelo 1: SARIMA (2, 0, 0) x $(1, 0, 0)_{12}$				
AIC = -357.1			BIC = -345.2	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR(1)	1.483	0.064	23.145	0.000
AR(2)	-0.485	0.064	-7.579	0.000
AR.S(12)	0.2807	0.092	3.049	0.002
Sigma2	0.004	0.000	13.139	0.000

Tabela 7: Coeficientes estimados e valores de AIC e BIC para o Modelo 1

Modelo 2: SARIMA (2, 1, 1) x $(1, 0, 0)_{12}$				
AIC = -373.6			BIC = -358.7	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR(1)	0.827	0.253	3.267	0.001
AR(2)	-0.398	0.150	-2.652	0.008
MA(1)	-0.194	0.266	-0.728	0.467
AR.S(12)	0.168	0.101	1.670	0.095
Sigma2	0.004	0.000	14.307	0.000

Tabela 8: Coeficientes estimados e valores de AIC e BIC para o Modelo 2

Modelo 3: SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$				
AIC = -375.4			BIC = -363.5	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR(1)	0.656	0.064	10.185	0.000
AR(2)	-0.316	0.098	-3.236	0.001
AR.S(12)	0.163	0.100	1.636	0.102
Sigma2	0.004	0.000	15.227	0.000

Tabela 9: Coeficientes estimados e valores de AIC e BIC para o Modelo 3

Pelo critério da minimização dos critérios de informação de Akaike e Bayes (AIC e BIC), o modelo escolhido foi o Modelo 3: SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$. Podemos destacar a alta significância estatística de seus termos autorregressivos AR(1) e AR(2) e da variância das inovações (Sigma2). A sua equação final explícita é:

$$\nabla \log a_t = 0.66 * \nabla \log a_{t-1} - 0.32 * \nabla \log a_{t-2} + 0.16 * \nabla \log a_{t-12} -$$

$$- 0.66 * 0.16 * \nabla \log a_{t-13} + 0.32 * 0.16 * \nabla \log a_{t-14} + \varepsilon_t$$

com $\varepsilon_t \sim N(0, 0.004)$ representando a série de inovações (resíduos) e a_t representando a série de preços de açúcar corrigida com relação à inflação. Para entender de onde aparecem os últimos dois termos que envolvem a_{t-13} e a_{t-14} é necessário aplicar os coeficientes calculados na Tabela 9 à definição de processo SARIMA da seção 2.2.4. e realizar a multiplicação dos operadores autorregressivos normais e sazonais.

O gráfico dos resíduos do modelo pode ser visto na Figura 26. Apesar da presença de um *outlier* muito marcante entre 2010 e 2011, não é visível nenhuma heterocedasticidade nos dados. Essa impressão é confirmada pelo teste de heterocedasticidade, que com um valor *p* de 0.58 não permite rejeitar a hipótese nula de homocedasticidade dos resíduos.

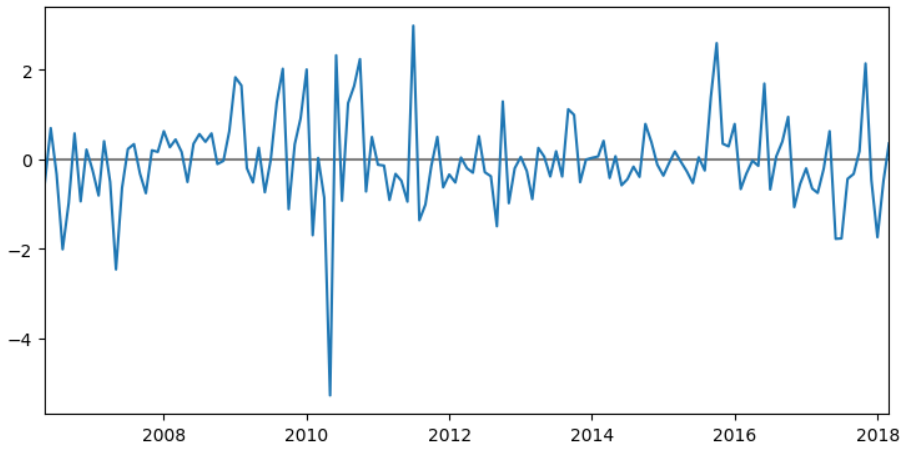

Figura 26: Resíduos normalizados do modelo SARIMA(2, 1, 0) x (1, 0, 0)$_{12}$ para a série de preços de açúcar

O teste de Ljung-Box procura autocorrelação em uma série temporal e a sua hipótese nula é a de resíduos não correlacionados. Pelo valor de p=0.40, não podemos rejeitar a hipótese nula a um nível de significância de 5%, o que também podemos observar na Figura 27, que mostra o gráfico da ACF para os resíduos do modelo.

Finalmente, o teste de Jarque-Bera sugere que os resíduos não seguem uma distribuição normal, já que sua hipótese nula é rejeitada a níveis de significância menores do que 1% (valor p<0.01). Isso também pode ser constatado no gráfico de quantis empíricos versus quantis teóricos (Figura 28): vemos que os pontos experimentais desviam

significativamente da linha vermelha, que é a referência para dados distribuídos normalmente.

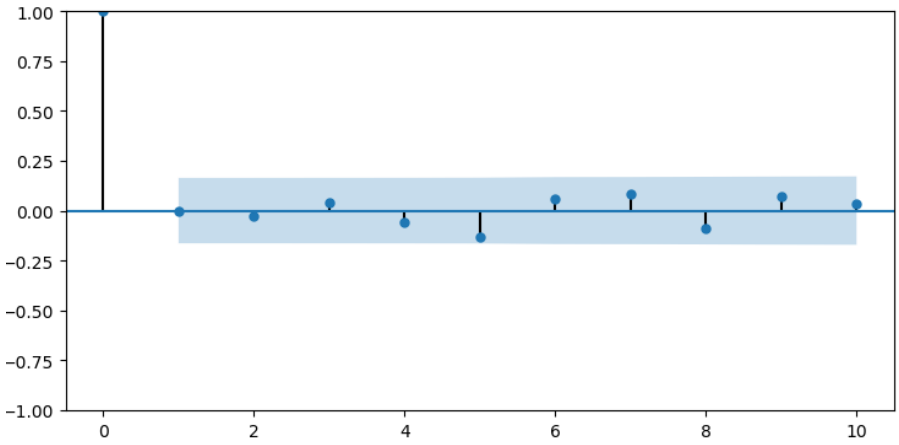

Figura 27: Função de autocorrelação dos resíduos do modelo SARIMA(2, 1, 0) x (1, 0, 0)$_{12}$ para a série de preços de açúcar

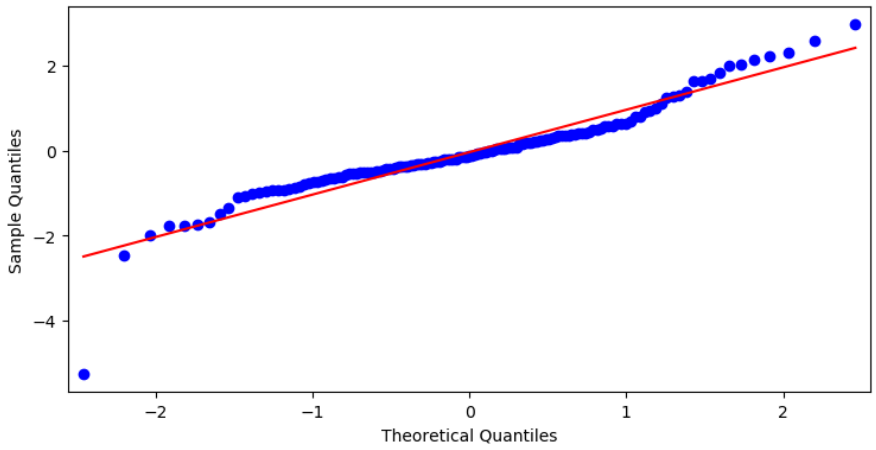

Figura 28: Quantis empíricos x quantis teóricos para os resíduos do modelo SARIMA(2, 1, 0) x (1, 0, 0)$_{12}$ para a série de preços de açúcar

Modelo 3: SARIMA (2, 1, 0) x (1, 0, 0)$_{12}$				
Teste	Hipótese nula (H0)	Estatística de teste	Valor p	Decisão
Ljung-Box	Ausência de autocorrelação dos resíduos	41.67	0.40	Aceitar H0
Heterocedasticidade	Resíduos homocedásticos	0.85	0.58	Aceitar H0
Jarque-Bera	Resíduos normalmente distribuídos	159.58	0.00	Rejeitar H0

Tabela 10: Análise de resíduos para o modelo escolhido (Modelo 3)

5.1.1.1. Previsões

O modelo selecionado foi utilizado para realizar previsões dos preços de açúcar nos últimos três anos, de maio de 2015 até abril de 2018. A técnica aplicada foi a de treinar um modelo SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$ em um período imediatamente anterior ao valor que se deseja prever, de forma que o modelo aprende o comportamento dos *m* meses anteriores e realiza a previsão sobre um valor que nunca viu antes. Por exemplo, para *m*=72 e para prever o valor do preço de açúcar no mês abril de 2016, o modelo é treinado nas observações dos seis anos anteriores, de abril de 2010 até março de 2016; para prever o mês de maio de 2016 o modelo é treinado no período de maio de 2010 ate abril de 2016, e assim por diante. Essa técnica é discutida em Bergmeir e Benítez (2012).

Os valores previstos foram então comparados com os valores reais utilizando três métricas: raiz quadrada do erro quadrado médio (RMSE), erro médio absoluto (MAE) e erro médio percentual absoluto (MAPE). Foram realizados testes sobre 3 valores diferentes de *m*: 60, 72 e 84, de forma que comparou-se o uso de valores históricos dos últimos 5, 6 e 7 anos na previsão dos preços. Os resultados foram comparados aos de um modelo de base chamado de *modelo de persistência*, que prevê simplesmente que o preço em um instante *t* será o mesmo que no instante *t*-1.

Previsões de preço de açúcar utilizando um modelo SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$			
Resultados das previsões de maio de 2015 a abril de 2018			
Janela de treinamento (*m*)	RMSE	MAE	MAPE
Últimos 5 anos (*m* = 60)	0.063	0.045	1.4%
Últimos 6 anos (*m* = 72)	0.063	0.045	1.4%
Últimos 7 anos (*m* = 84)	**0.062**	**0.044**	**1.4%**
Modelo de persistência	0.084	0.066	2.1%

Tabela 11: Previsões de preços de açúcar para um mês no futuro realizadas pelo modelo SARIMA

A Tabela 11 resume os resultados das previsões para a série de açúcar. É possível ver que o modelo SARIMA (2, 1, 0) x $(1, 0, 0)_{12}$ teve uma melhor performance que o modelo de persistência para todos os indicadores e todos os valores da janela de treinamento *m*. Entre todos os valores de *m*, o melhor resultado foi encontrado para *m* = 84, **com uma melhoria de 26.8% no RMSE, 33.3% no MAE e 34.7% no MAPE em relação ao modelo de base**. Os

valores previstos pelo modelo SARIMA e os valores reais da série de preços de açúcar podem ser vistos na Figura 29:

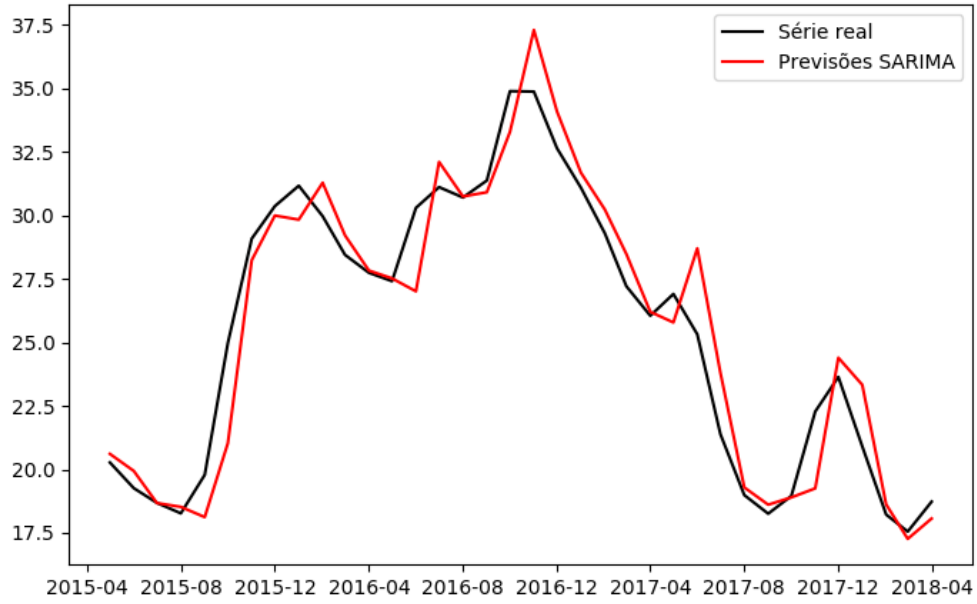

Figura 29: Previsões SARIMA e valores reais da série de preços de açúcar de maio de 2015 a abril de 2018. Fonte: elaborado pelo autor

5.1.2. SARIMA - Série de preços de etanol

Novamente seguindo a metodologia de Box e Jenkins (1970), começamos por analisar as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série de preços logarítmicos de etanol. A ACF da série original apresenta um decaimento lento nos valores iniciais até o valor de h=6 e em seguida uma nova subida até h=12. Está claro que existe uma sazonalidade, mas é difícil detectá-la sem diferenciação. Já a PACF possui alguns picos em h=1, h=2 e h=4 e um valor no limite da significância estatística em h=7, como podemos ver na Figura 30:

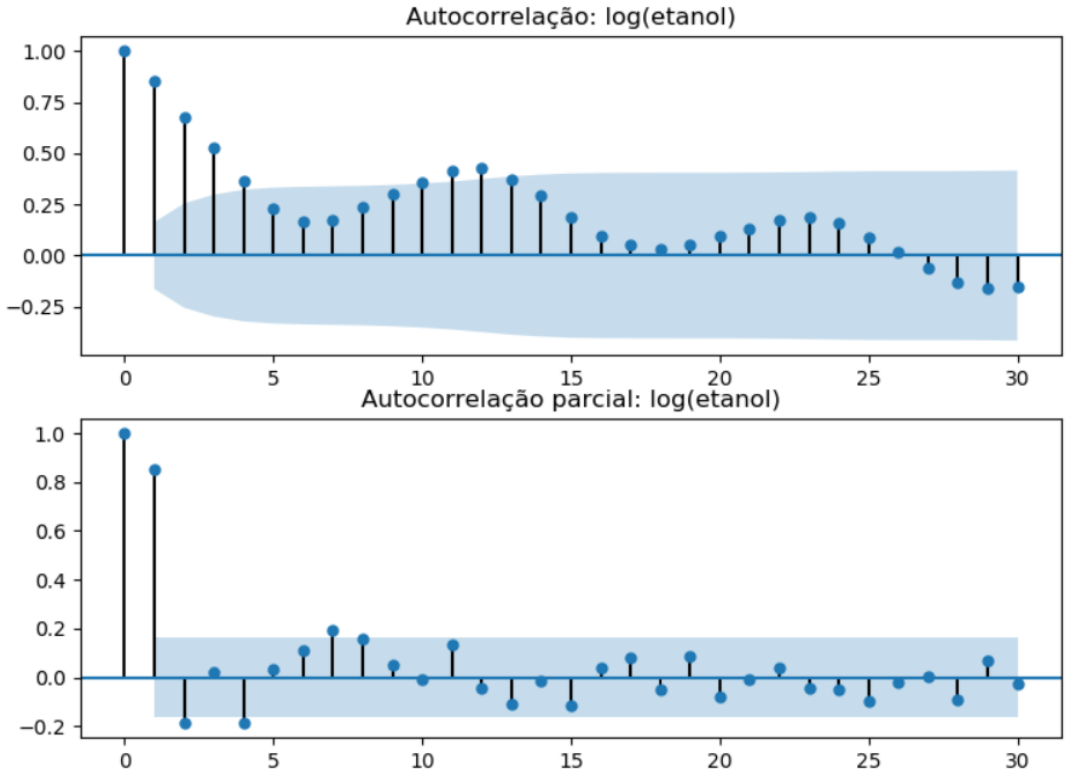

Figura 30: ACF e PACF da série logarítmica de preços de etanol sem diferenciação. Fonte: elaborado pelo autor

Após a diferenciação $\nabla y_t = y_t - y_{t-1}$ da série, é possível ver na Figura 31 picos significantes estatisticamente para os pontos h=5, h=6 e h=12 na ACF e h=6 na PACF. Isso sugere que a sazonalidade da série de etanol hidratado é semestral, e que essa dinâmica predomina sobre os elementos de autorregressão e média móvel. Vemos também que os picos em h=6 são negativos, enquanto os picos em h=12 são positivos.

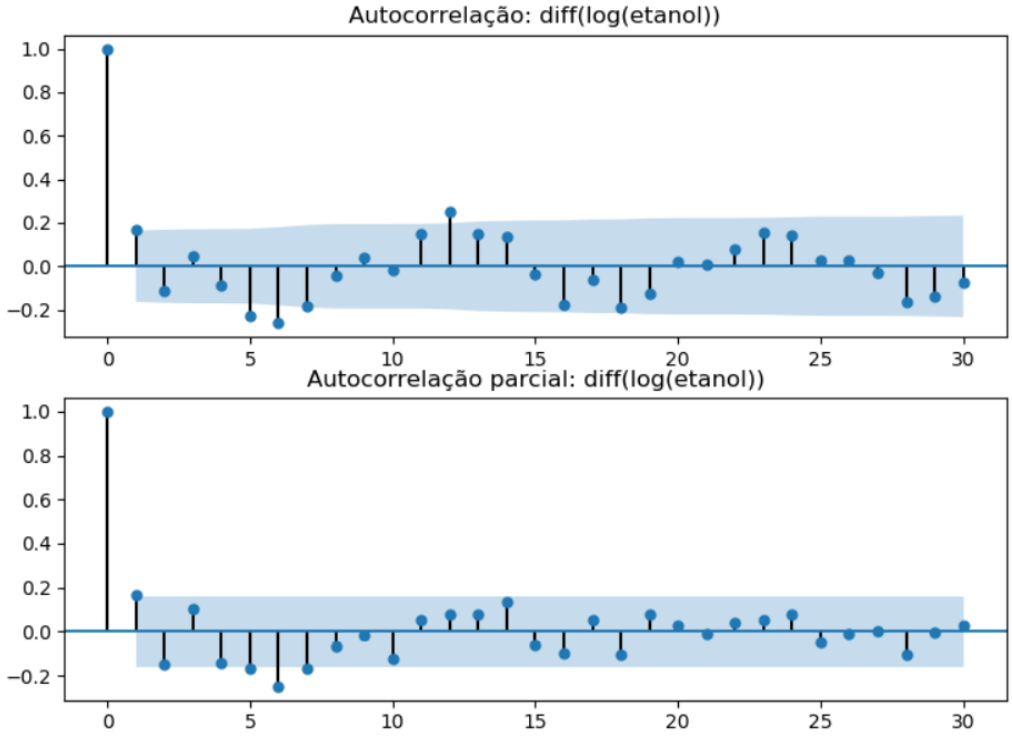

Figura 31: ACF e PACF da série logarítmica de preços de etanol após diferenciação. Fonte: elaborado pelo autor

Sendo assim, os modelos candidatos serão compostos exclusivamente de termos sazonais. Serão testados três modelos:

1. Modelo 1: SARIMA (0, 1, 0) x $(1, 0, 0)_6$;
2. Modelo 2: SARIMA (0, 1, 0) x $(1, 0, 1)_6$;
3. Modelo 3: SARIMA (0, 1, 0) x $(2, 0, 1)_6$

Os resultados da estimativa dos parâmetros dos modelos pelo método da máxima verossimilhança, além dos valores dos critérios de informação de Bayes e Akaike (BIC e AIC) podem ser vistos nas Tabelas 12 a 14. O modelo escolhido foi o Modelo 2: SARIMA (0, 1, 0) x $(1, 0, 1)_6$, pois é o que minimiza o AIC e BIC. A sua forma explícita é:

$$\nabla \log e_t = -0.981 * \nabla \log e_{t-6} + \varepsilon_t + 0.889 * \varepsilon_{t-6}$$

onde $\varepsilon_t \sim N(0, 0.006)$ é a série de inovações (resíduos) e e_t representa a série de preços de etanol corrigidos com relação à inflação.

Modelo 1: SARIMA (0, 1, 0) x $(1, 0, 0)_6$				
AIC = -303.0			BIC = -297.1	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR.S(6)	-0.207	0.071	-2.909	0.004
Sigma2	0.007	0.000	14.418	0.000

Tabela 12: Coeficientes estimados e valores de AIC e BIC para o Modelo 1

Modelo 2: SARIMA (0, 1, 0) x $(1, 0, 1)_6$				
AIC = -315.6			BIC = -306.6	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR.S(6)	-0.981	0.028	-34.588	0.000
MA.S(6)	0.889	0.084	10.578	0.000
Sigma2	0.006	0.001	11.335	0.000

Tabela 13: Coeficientes estimados e valores de AIC e BIC para o Modelo 2

Modelo 3: SARIMA (0, 1, 0) x $(2, 0, 1)_6$				
AIC = -314.2			BIC = -302.4	
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR.S(6)	-1.072	0.116	-9.229	0.000
AR.S(12)	-0.083	0.094	-0.886	0.375
MA.S(6)	0.924	0.121	7.643	0.000
Sigma2	0.006	0.001	9.852	0.000

Tabela 14: Coeficientes estimados e valores de AIC e BIC para o Modelo 3

Dessa vez, a análise dos resíduos do modelo (Figura 32) sugere a existência de heterocedasticidade: a variância dos dados no período de 2006 a 2012 parece ser maior que a variância de 2012 a 2018. Essa suspeita é confirmada pelo teste de heterocedasticidade, que com um valor *p* de 0.02 rejeita a hipótese nula de homocedasticidade dos resíduos a um nível de confiança de 5%.

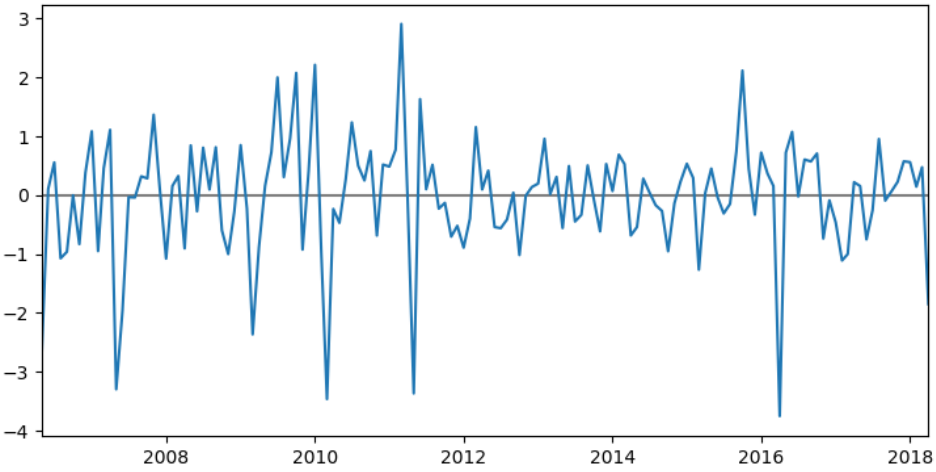

Figura 32: Resíduos normalizados do modelo SARIMA(0, 1, 0) x (1, 0, 1)$_6$ para a série de preços de etanol

A hipótese de normalidade dos resíduos também é rejeitada, tanto pelo teste de Jarque-Bera ($p < 0.01$) quanto pela análise gráfica de quantis empíricos versus quantis teóricos (Figura 33). Isso não é surpreendente, dado que os resíduos apresentam heterocedasticidade. Já o teste de Ljung-Box ($p = 0.35$) não sugere a existência de correlação em série nos resíduos. O resumo da análise de resíduos para o Modelo 2 pode ser visto na Tabela 15.

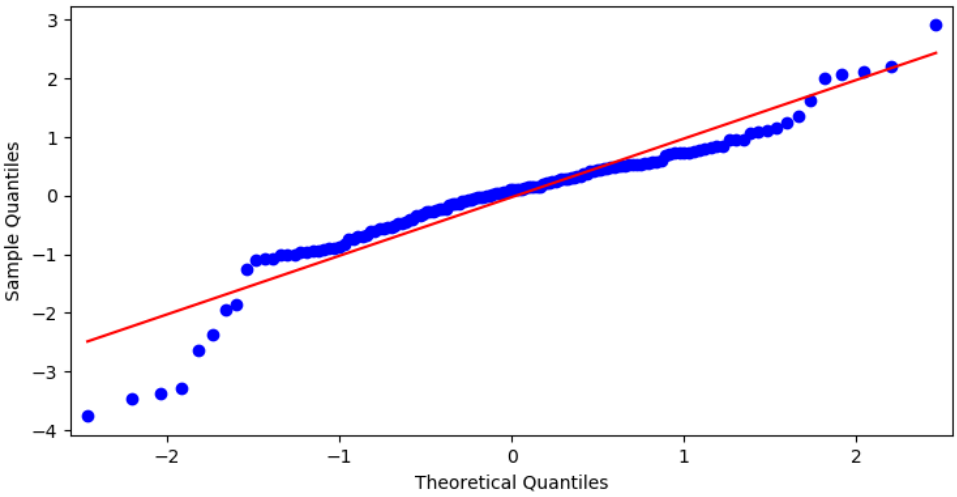

Figura 33: Gráfico QQ para os resíduos do modelo SARIMA(0, 1, 0) x (1, 0, 1)$_6$ para a série de preços de etanol

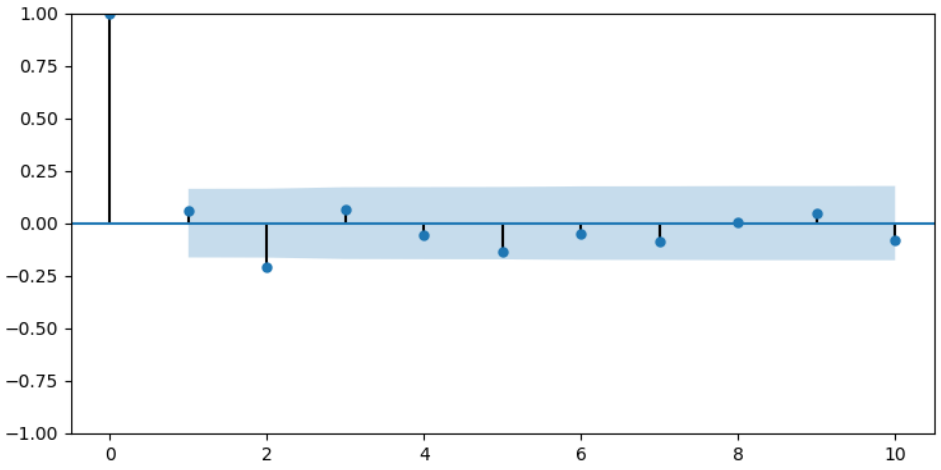

Figura 34: Função de autocorrelação dos resíduos do modelo SARIMA(0, 1, 0) x (1, 0, 1)$_6$ para a série de preços de açúcar

Modelo 2: SARIMA (0, 1, 0) x (1, 0, 1)$_6$				
Teste	Hipótese nula (H0)	Estatística de teste	Valor p	Decisão
Ljung-Box	Ausência de autocorrelação dos resíduos	42.8	0.35	Aceitar H0
Heterocedasticidade	Resíduos homocedásticos	0.50	0.02	Rejeitar H0
Jarque-Bera	Resíduos normalmente distribuídos	78.23	0.00	Rejeitar H0

Tabela 15: Análise de resíduos para o modelo SARIMA(0, 1, 0) x (1, 0, 1)$_6$

5.1.2.1. Previsões

De forma análoga ao item 5.1.1.1, o modelo escolhido foi utilizado para realizar previsões de preços de etanol entre maio de 2015 e abril de 2018. As mesmas métricas e janelas de previsão foram utilizadas, e os resultados estão resumidos na Tabela 16. Os melhores resultados foram obtidos para uma janela de treinamento de 6 anos, **com uma melhoria de 12.5% no RMSE, 13.4% no MAE e 14.8% no MAPE com relação ao modelo de persistência**. A Figura 35 mostra graficamente os resultados das previsões para o melhor modelo.

Previsões de preço de etanol utilizando um modelo SARIMA (0, 1, 0) x $(1, 0, 1)_6$			
Resultados das previsões de maio de 2013 a abril de 2018			
Janela de treinamento (*m*)	RMSE	MAE	MAPE
Últimos 5 anos (*m* = 60)	0.075	0.051	9.0%
Últimos 6 anos (*m* = 72)	**0.074**	**0.050**	**8.7%**
Últimos 7 anos (*m* = 84)	0.076	0.052	9.2%
Modelo de persistência	0.085	0.058	10.2%

Tabela 16: Previsões de preços de etanol para um mês no futuro realizadas pelo modelo SARIMA

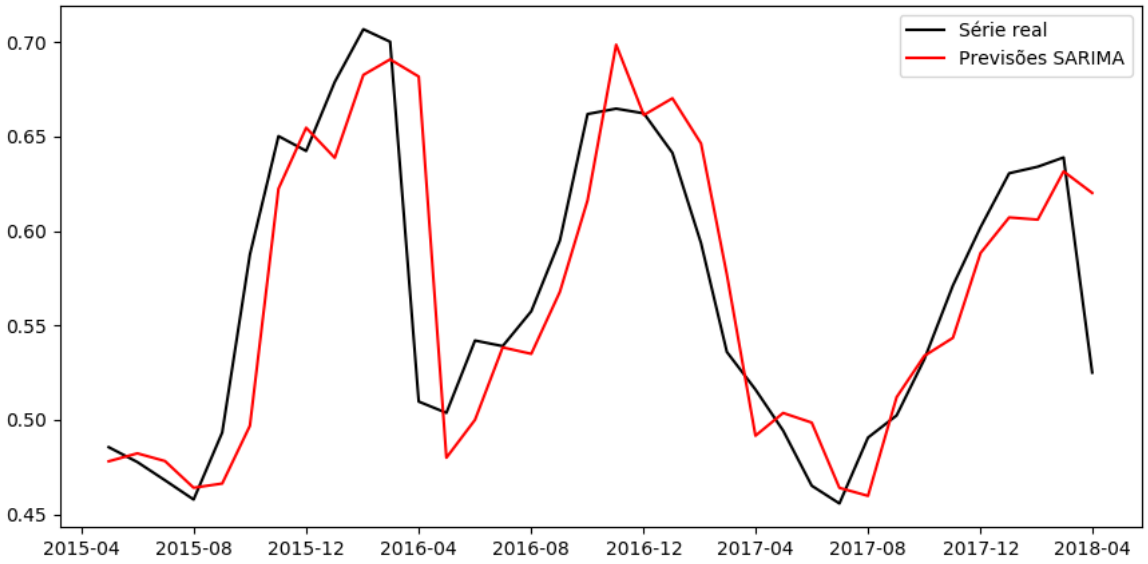

Figura 35: Previsões SARIMA e valores reais da série de preços de etanol de maio de 2015 a abril de 2018. Fonte: elaborado pelo autor

5.1.3 VAR - Modelagem conjunta de preços de etanol e de açúcar

Álcool e açúcar são ambos derivados da cana-de-açúcar. Por conta disso, é natural suspeitar que os seus preços estejam de alguma forma correlacionados. A Figura 12, que mostra as séries históricas de preços de açúcar e de etanol no mesmo gráfico, reforça essas suspeitas.

Para avaliar a interação entre as duas séries temporais, foi utilizado um modelo VAR (*Vector Autoregression*). Esse modelo é mais simples do que o SARIMA porque utiliza apenas elementos autorregressivos, mas permite estimar interações entre séries temporais: como os valores passados de etanol influenciam os valores presentes de açúcar, e vice-versa.

Enquanto em um modelo SARIMA é necessário escolher sete hiperparâmetros (*p, d, q, P, D, Q* e *S*) antes de realizar as estimativas, para o modelo VAR basta escolher o valor de *p*, referente à ordem de autorregressão. Como descrito na seção 2.2.6, isso é feito empiricamente: são aplicados *n* modelos diferentes e escolhe-se o valor de *p* que minimiza os critérios de informação de Bayes e de Akaike (BIC e AIC).

Utilizando os aprendizados dos modelos SARIMA aplicados nas seções anteriores, decidiu-se por diferenciar as séries de açúcar e de etanol antes de escolher a ordem do modelo: dessa forma a autorregressão será realizada sobre $\nabla x_t = x_t - x_{t-1}$.

A Tabela 17 mostra os valores de AIC e BIC para diferentes ordens *p* de modelo. Foram considerados valores até *p*=5, pois para valores muito maiores o modelo terá muitas variáveis e se distanciará muito da realidade, dado que o número *k* de variáveis estimadas pelo modelo se relaciona com a sua ordem *p* pela expressão $k = 2(p + 1)$. Foi escolhido o valor *p*=2, que minimizou tanto o AIC quanto o BIC.

Ordem do modelo (*p*)	AIC	BIC
p=1	-10.40	-10.27
p=2	**-10.52**	**-10.31**
p=3	-10.50	-10.20
p=4	-10.51	-10.13
p=5	-10.50	-10.04

Tabela 17: Valores de AIC e BIC para diferentes ordens de VAR

Os parâmetros estimados pelo modelo de ordem 2 podem ser vistos na Tabela 18. Como existem duas séries de preços diferentes, existem dois conjuntos de parâmetros: os que estimam os preços de etanol e os que estimam os preços de açúcar. Na coluna **coeficiente,** o termo AR(*p*).*etanol* faz referência ao p-ésimo elemento regressivo da série de preços de etanol, por exemplo.

Sendo $\boldsymbol{a_t}$ e $\boldsymbol{e_t}$ as séries de açúcar e de etanol corrigidas pela inflação, respectivamente, as equações finais do modelo são as seguintes:

$$\boldsymbol{\nabla \log a_t} = 0.654 * \boldsymbol{\nabla \log a_{t-1}} + 0.121 * \boldsymbol{\nabla \log e_{t-1}} - 0.378 * \boldsymbol{\nabla \log a_{t-2}} + 0.029 * \boldsymbol{\nabla \log e_{t-2}} + \boldsymbol{\varepsilon_t^a}$$

$$\boldsymbol{\nabla \log e_t} = 0.223 * \boldsymbol{\nabla \log a_{t-1}} + 0.147 * \boldsymbol{\nabla \log e_{t-1}} - 0.097 * \boldsymbol{\nabla \log a_{t-2}} - 0.176 * \boldsymbol{\nabla \log e_{t-2}} + \boldsymbol{\varepsilon_t^e}$$

com $\boldsymbol{\varepsilon_t^a} \sim N(0, 0.004)$ e $\boldsymbol{\varepsilon_t^e} \sim N(0, 0.006)$ representando as séries de resíduos.

Modelo VAR(2)				
AIC = -10.52		BIC = -10.31		
Resultados para a equação de preços de etanol				
Coeficiente	Valor	Desvio padrão	Valor de z	P > |z|
AR(1).etanol	0.147	0.091	1.612	0.109
AR(1).açúcar	0.223	0.108	2.063	0.041
AR(2).etanol	-0.176	0.090	-1.961	0.052
AR(2).açúcar	-0.097	0.105	-0.922	0.358
Sigma2.etanol	0.006	0.001	7.253	0.000
Resultados para a equação de preços de açúcar				
AR(1).etanol	0.121	0.072	1.680	0.095
AR(1).açúcar	0.654	0.085	7.684	0.000
AR(2).etanol	0.029	0.071	0.408	0.684
AR(2).açúcar	-0.378	0.082	-4.586	0.000
Sigma2.açúcar	0.004	0.001	5.690	0.000

Tabela 18: Coeficientes estimados para o modelo VAR(2)

Para a equação de preços de açúcar, os valores passados da própria série são estatisticamente significantes, com valores de *p* menores do que 0.01. Isso é coerente com o que foi visto na análise SARIMA, dado que o modelo ótimo utilizava dois termos autorregressivos. Os valores passados da série de etanol não parecem ser relevantes na previsão de preços de açúcar, haja vista os altos valores de *p* calculados.

Já para a equação de preços de etanol, nenhum dos termos é estatisticamente significante a um nível de confiança de 1%. Os valores do mês anterior da série de açúcar são os que chegam mais perto de serem relevantes, com um valor de *p* de 0.04. Para os valores passados de etanol, era esperado que eles não fossem significantes, visto que o modelo ótimo SARIMA não possuía nenhum termo de autorregressão.

A inspeção dos resíduos do modelo VAR nos permite chegar a conclusões semelhantes às realizadas com os modelos SARIMA: há uma certa heterocedasticidade na série de etanol (Figura 36), e nenhuma das séries de resíduos parece ser normalmente distribuída (Figura 37).

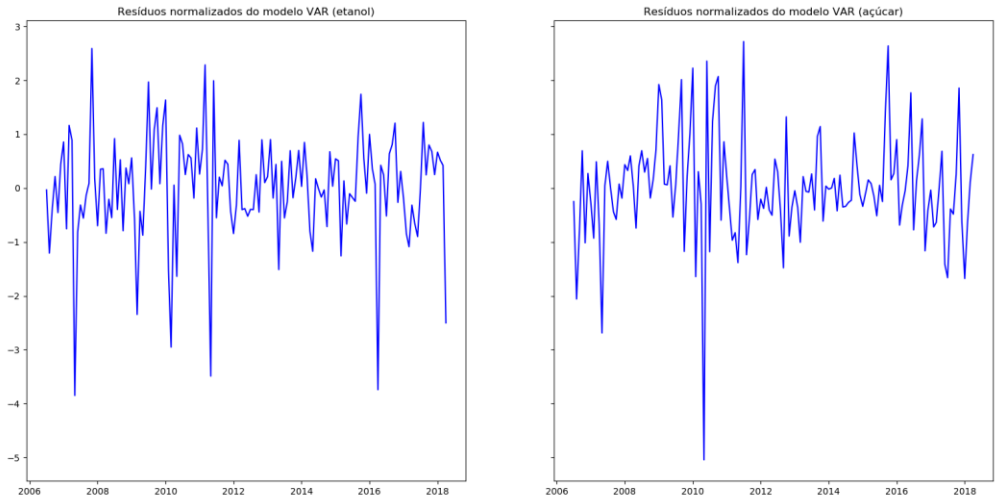

Figura 36: Resíduos normalizados do modelo VAR(2) para as séries de etanol (esquerda) e açúcar (direita). Fonte: elaborado pelo autor

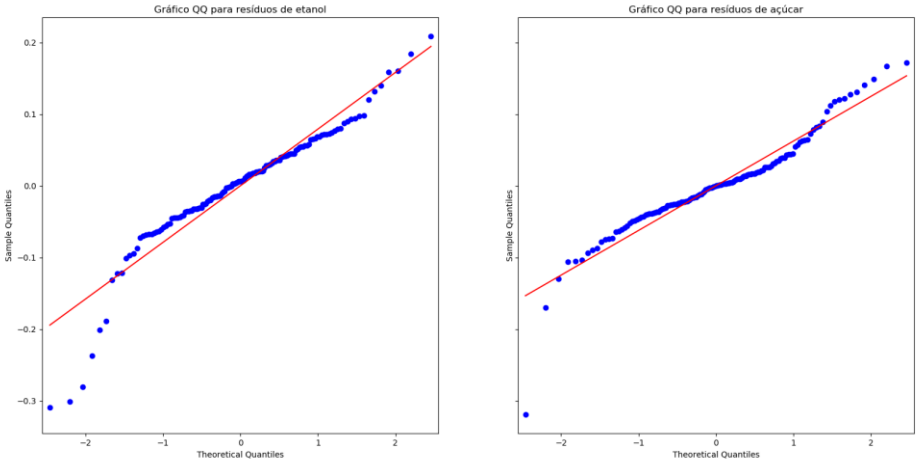

Figura 37: Gráficos QQ dos resíduos do modelo VAR(2) para as séries de etanol (esquerda) e açúcar (direita). Fonte: elaborado pelo autor

Note que a Figura 38 mostra a existência de autocorrelação nos resíduos de etanol: isso ocorre porque não foi possível introduzir termos sazonais no modelo VAR, e eles são importantes para prever os preços de etanol (como visto em 5.1.2). Isso será remediado na próxima seção com a utilização de modelos de espaço de estado.

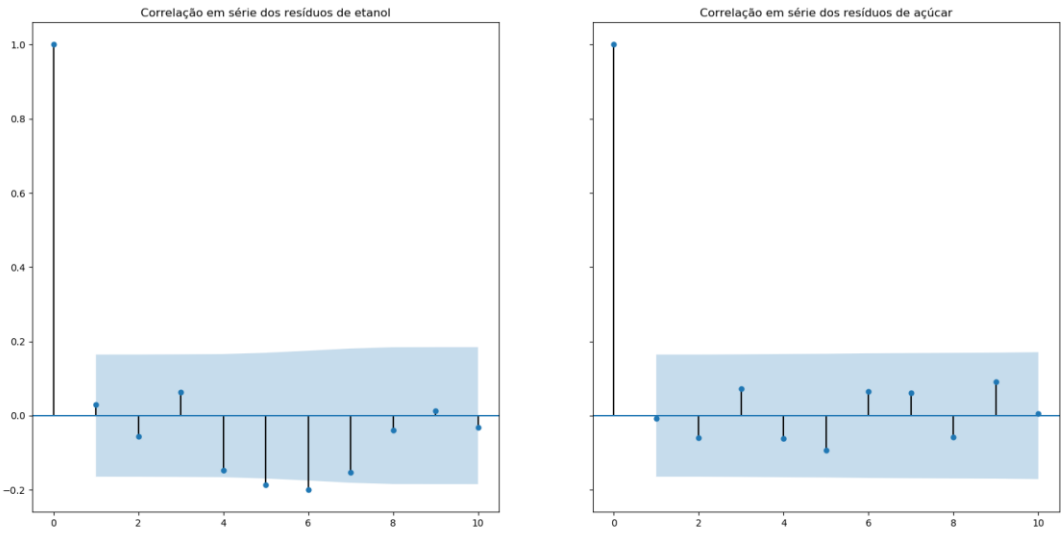

Figura 38: ACF dos resíduos do modelo VAR(2) para as séries de etanol (esquerda) e açúcar (direita). Elaborado pelo autor

5.1.3.1. Previsões

Assim como nos casos anteriores, o modelo VAR(2) foi utilizado para realizar previsões de preços de açúcar e de etanol de maio de 2015 a abril de 2018, utilizando as métricas definidas e janelas passadas de 5, 6 e 7 anos. Os resultados das previsões estão expressos na Tabela 19. É possível observar que o modelo VAR teve uma performance pior que a do modelo SARIMA para os preços de etanol, o que pode ser explicado pela ausência de termos sazonais, que são os termos com maior poder preditivo para essa série. Já para os preços de açúcar, os resultados são bons, equivalentes aos do modelo SARIMA.

Previsões de preços de etanol e açúcar utilizando um modelo VAR(2)						
	Etanol			Açúcar		
Janela de treinamento	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Últimos 5 anos	0.084	0.059	10.3%	0.066	0.052	1.6%
Últimos 6 anos	**0.083**	**0.059**	**10.2%**	**0.060**	**0.044**	**1.4%**
Últimos 7 anos	0.086	0.063	11.0%	0.064	0.048	1.5%
Modelo de persistência	0.085	0.058	10.2%	0.084	0.066	2.1%

Tabela 19: Resultado das previsões para os preços de açúcar e etanol para um mês no futuro realizadas pelo modelo VAR

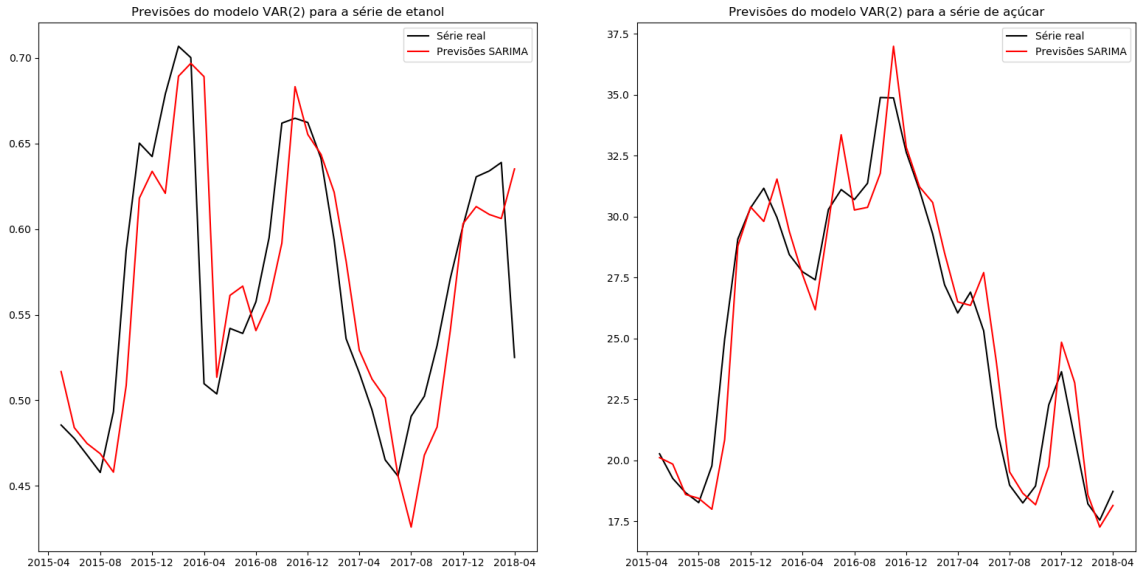

Figura 39: Previsões VAR(2) para a série de preços de etanol (esquerda) e açúcar (direita). Elaborado pelo autor

5.2. Modelo de espaço de estados

5.2.1. Descrição

Os modelos clássicos de séries temporais da seção anterior trouxeram resultados aceitáveis, especialmente no caso das modelagens individuais utilizando SARIMA. No entanto, eles apresentam a limitação de não poder incluir variáveis exógenas. No capítulo 4 identificamos que os contratos futuros de etanol hidratado da Bovespa e de açúcar bruto número 11 de Chicago possuem potencial preditivo para os preços sucroalcooleiros, e nesta seção criaremos um modelo completo usando termos autorregressivos, sazonais e variáveis exógenas para melhorar as previsões.

O modelo será escrito na forma de espaço de estados, que permite grande controle das variáveis e parâmetros a serem estimados. Será realizada uma regressão multivariada com erros ARIMA, como descrito em Durbin e Koopman (2012). As variáveis utilizadas serão:

1. $\boldsymbol{a_t}$ se refere à série temporal de preços de açúcar cristal levantada pelo Cepea;
2. $\boldsymbol{e_t}$ se refere à série temporal de preços de etanol hidratado levantada pelo Cepea;
3. $\boldsymbol{f_t^a}$ se refere à série de preços futuros de açúcar bruto número 11 da bolsa de Chicago;
4. $\boldsymbol{f_t^e}$ se refere à série de preços futuros de etanol hidratado da BM&F Bovespa;
5. $\boldsymbol{s_t^e}$ se refere ao índice sazonal de preços de etanol, calculado pelo autor baseado nos coeficientes AR e MA estimados no modelo da seção 5.1.2. Esse cálculo será detalhado na seção 5.2.2.

As equações individuais para as séries de preços de açúcar e de etanol são dadas por:

$$\boldsymbol{\nabla \log a_t} = \varphi_1^{\mathrm{a}} \boldsymbol{\nabla \log a_{t-1}} + \varphi_2^{\mathrm{a}} \boldsymbol{\nabla \log a_{t-2}} + \beta_{\mathrm{fat}} \boldsymbol{\nabla \log f_t^a} + \boldsymbol{\varepsilon_a^t}$$

$$\boldsymbol{\nabla \log e_t} = \beta_{\mathrm{set}} \nabla \boldsymbol{\log f_t^e} + \beta_{\mathrm{fet}} \boldsymbol{s_t^e} + \boldsymbol{\varepsilon_e^t}$$

com $\boldsymbol{\varepsilon_a^t} \sim N(0, \sigma_{\mathrm{a}}^2)$ e $\boldsymbol{\varepsilon_e^t} \sim N(0, \sigma_{\mathrm{e}}^2)$ e $\boldsymbol{\nabla x_t} = \boldsymbol{x_t} - \boldsymbol{x_{t-1}}$. Note que os modelos serão aplicados sobre as séries logarítmicas diferenciadas uma vez, da mesma forma que na seção 5.1. (por isso a presença do operador de diferenciação $\boldsymbol{\nabla}$).

Relembrando o que foi visto na revisão bibliográfica, os modelos de espaço de estado podem ser escritos na seguinte forma geral:

$$\boldsymbol{y_t} = \mathrm{A_t}\mathbf{x_t} + \Gamma\boldsymbol{u_t} + \boldsymbol{v_t}$$

$$\mathbf{x_t} = \Phi\mathbf{x_{t-1}} + \Upsilon\boldsymbol{u_t} + \boldsymbol{w_t}$$

Nesse caso específico, teremos:

1. $\boldsymbol{y_t} = \begin{bmatrix} \boldsymbol{\nabla \log a_t} \\ \boldsymbol{\nabla \log e_t} \end{bmatrix}$ é o vetor coluna com as medições das variáveis de interesse;
2. $\mathbf{x_t} = \begin{bmatrix} \boldsymbol{\nabla \log a_t} \\ \boldsymbol{\nabla \log e_t} \\ \boldsymbol{\nabla \log a_{t-1}} \\ \boldsymbol{\nabla \log e_{t-1}} \end{bmatrix}$ é o vetor de estado;
3. $\boldsymbol{u_t} = \begin{bmatrix} \boldsymbol{s_t^e} \\ \nabla \boldsymbol{\log f_t^e} \\ \nabla \boldsymbol{\log f_t^a} \end{bmatrix}$ é o vetor de variáveis exógenas;
4. $\boldsymbol{v_t} = 0$;
5. $\boldsymbol{w_t} = N(0, \Sigma)$

As matrizes do modelo são dadas por:

1. $A_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$
2. $\Gamma = \begin{bmatrix} 0 & 0 & \beta_{\mathrm{fat}} \\ \beta_{\mathrm{set}} & \beta_{\mathrm{fet}} & 0 \end{bmatrix}$
3. $\Phi = \begin{bmatrix} \varphi_1^{\mathrm{a}} & 0 & \varphi_2^{\mathrm{a}} & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$
4. $\Upsilon = 0$
5. $\Sigma = \begin{bmatrix} \sigma_{\mathrm{a}}^2 & \sigma_{\mathrm{ae}} & 0 & 0 \\ \sigma_{\mathrm{ae}} & \sigma_{\mathrm{e}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Existem oito parâmetros a serem estimados para o modelo:

1. φ_1^{a} e φ_2^{a} são os coeficientes de autorregressão do açúcar, como os calculados para o modelo do item 5.1.1;
2. β_{set} é um coeficiente que representa os termos sazonais (média móvel e autorregressão) do etanol, como calculados para o item 5.1.2. Esses termos foram unificados em um único coeficiente e colocados como variável exógena para simplificar a dimensão do problema: seria necessário um vetor de estados de 12

dimensões para incorporar a sazonalidade na formulação da equação de estado, o que tornaria o problema insolúvel computacionalmente;

3. β_{fat} e β_{fet} são os coeficientes que representam as contribuições dos preços futuros de açúcar e de etanol na formação dos seus respectivos preços;
4. σ_a^2, σ_e^2 representam as variâncias dos preços de açúcar e de etanol na equação de estado;
5. σ_{ae} representa a covariância dos preços de açúcar e etanol na equação de estado. Note que a modelagem bidimensional permite a introdução desse termo, que conecta os dois termos de erro $\boldsymbol{\varepsilon_a^t}$ e $\boldsymbol{\varepsilon_e^t}$

5.2.2. Cálculo do índice de sazonalidade para o etanol

No item 5.1.2. vimos que a série de etanol podia ser modelada na forma SARIMA $(0,1,0) \times (1,0,1)_6$**,** com coeficientes MA.S(6) = 0.889 e AR.S(6) = -0.981. É possível provar (ver Shumway e Stoffer, 2011) que esse modelo admite uma representação autorregressiva infinita, dada pela equação:

$$\nabla \mathbf{e_t} = \sum_{\mathbf{i=1}}^{\infty} (-0.889)^{\mathbf{i-1}} * \ (0.889 - 0.981) * \nabla \mathbf{log}\, \boldsymbol{e}_{\mathbf{t-(6*i)}}$$

Para aproveitar o potencial dos termos sazonais sem aumentar demais a dimensão do problema (como já foi mencionado, a incorporação de um termo sazonal de ordem 6 na equação de estado aumentaria a sua dimensão para 12), calcula-se previamente essa somatória infinita, que é incluída como a variável exógena $\mathbf{s_t}$ (índice de sazonalidade):

$$\mathbf{s_t} = \sum_{\mathbf{i=1}}^{\infty} (-0.889)^{\mathbf{i-1}} * \ (0.889 - 0.981) * \nabla \mathbf{log}\, \boldsymbol{e}_{\mathbf{t-(6*i)}}$$

É importante notar que na prática calcula-se a somatória até o último termo disponível (os dados de etanol foram coletados de abril de 2006 a abril de 2018). Por exemplo, para o índice $\mathbf{s_t}$ de junho de 2012, a somatória é calculada com valores $\nabla \mathbf{log}\, \boldsymbol{e}_{\mathbf{t-(6*i)}}$ de dezembro de 2011, junho de 2011, ..., até junho de 2006.

5.2.3. Estimativa dos parâmetros e análise de resíduos

Os oito parâmetros do modelo foram estimados utilizando o algoritmo do Filtro de Kalman, implementado na biblioteca FKF do software R. Os resultados estão expressos na Tabela 20.

O modelo final é aplicado sobre observações de setembro de 2010 até abril de 2018 (são perdidos valores anteriores porque a série de futuros da BM&F Bovespa começa em 2010), totalizando 92 observações. Como são estimados 8 parâmetros, a distribuição *t* associada possui 84 graus de liberdade. O seu valor crítico para um nível de confiança de 1% é de 2.635; portanto pode-se concluir que todos os parâmetros são estatisticamente significantes a 1%, com exceção de σ_{ae}.

Parâmetro	Estimativa	Desvio Padrão	Valor de *t*
φ_1^a	0.7012	0.0999	7.0144
φ_2^a	-0.3724	0.0962	-3.8695
σ_a^2	0.0022	0.0001	20.351
σ_e^2	0.0040	0.0005	8.1325
σ_{ae}	0.0000	0.0002	0.3256
β_{fet}	0.3613	0.0917	3.9368
β_{fat}	0.2685	0.0572	4.6875
β_{set}	0.9575	0.2478	3.8643

Tabela 20: Resultado da estimativa dos parâmetros utilizando o algoritmo do filtro de Kalman

A análise de resíduos das Figuras 40 a 42 mostra que o modelo retrata bem a série de preços de açúcar: os resíduos são normalmente distribuídos, homocedásticos e não apresentam autocorrelação em série. Já os resíduos do modelo para a série de preços de etanol não são normalmente distribuídos, apesar de serem homocedásticos e de não apresentarem autocorrelação em série.

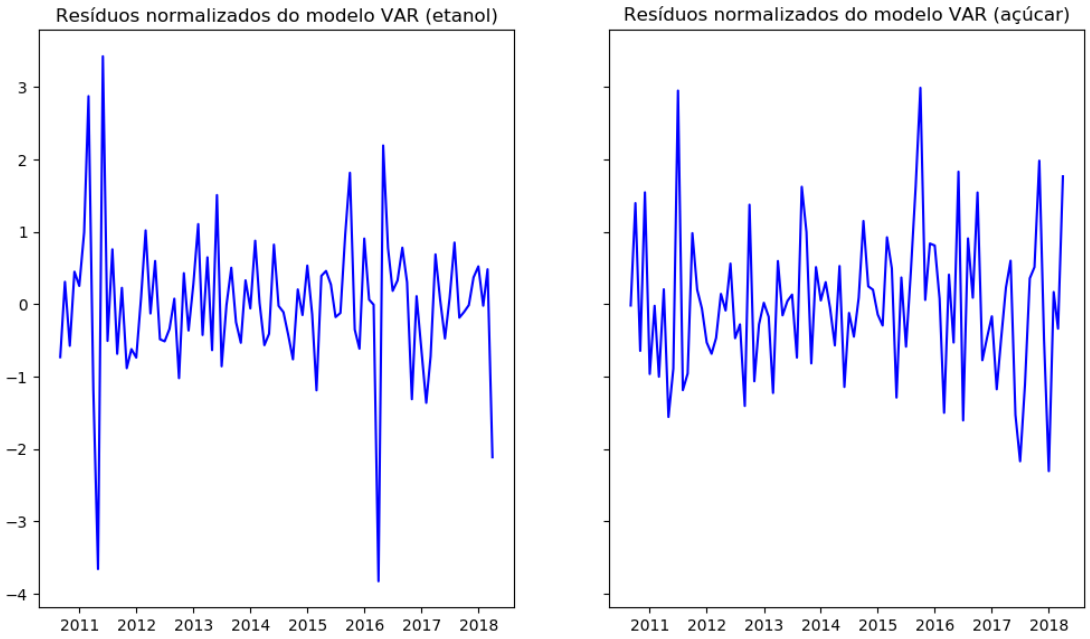

Figura 40: Gráfico dos resíduos normalizados para o modelo de espaço de estados. Fonte: elaborado pelo autor

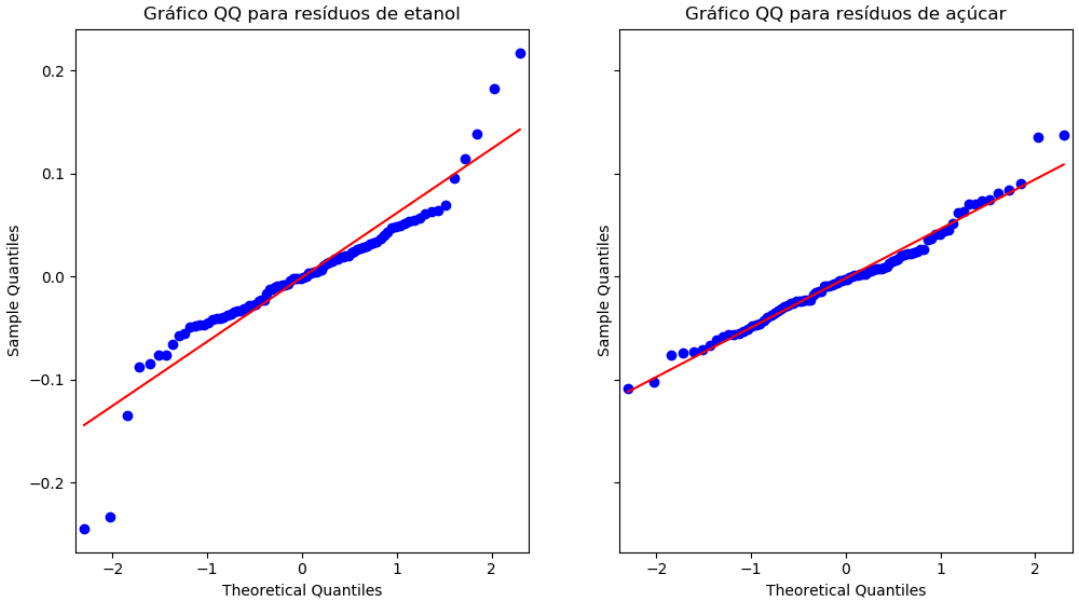

Figura 41: Gráfico QQ para os resíduos do modelo de espaço de estados. Fonte: elaborado pelo autor

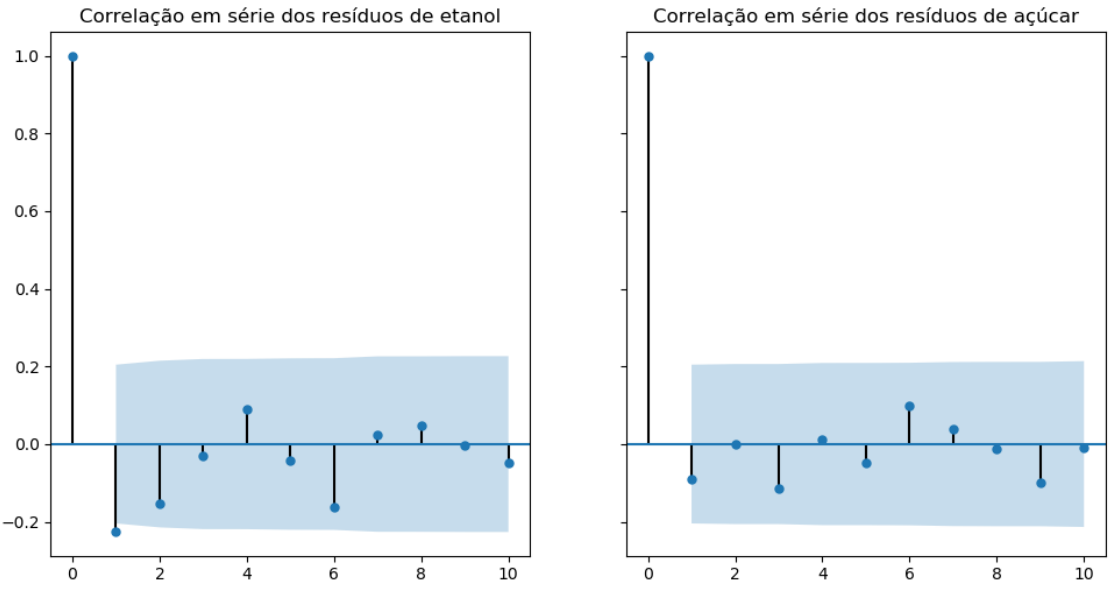

Figura 42: Função de autocorrelação dos resíduos do modelo de espaço de estados. Fonte: elaborado pelo autor

5.2.4. Previsões

O modelo de espaço de estados foi utilizado para realizar previsões dos preços de açúcar e etanol entre maio de 2015 e abril de 2018, utilizando como métricas as funções RMSE, MAE e MAPE. Foram testadas janelas de 24 meses, 36 meses e 48 meses para o treinamento, utilizando um procedimento semelhante aos realizados anteriormente. Os resultados estão expressos na Tabela 21.

Previsões de preços de etanol e açúcar utilizando o modelo de espaço de estado						
Resultados das previsões de maio de 2015 a abril de 2018						
	Etanol			Açúcar		
Janela de treinamento (*m*)	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Últimos 2 anos (*m* = 24)	0.067	0.043	7.4%	0.061	0.049	1.5%
Últimos 3 anos (*m* = 36)	0.068	0.042	7.4%	0.063	0.050	1.5%
Últimos 4 anos (*m* = 48)	**0.067**	**0.044**	**7.7%**	**0.059**	**0.046**	**1.4%**
Modelo de persistência	0.085	0.058	10.2%	0.084	0.066	2.0%

Tabela 21: Resultado do modelo de espaço de estados para previsões um mês no futuro

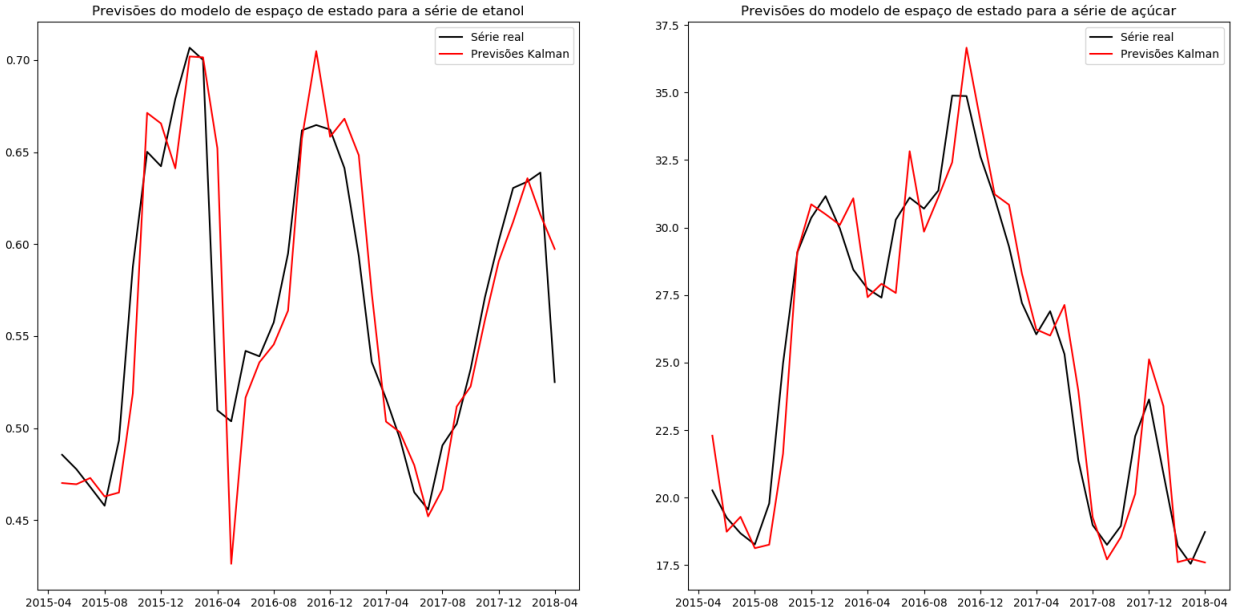

Figura 43: Previsões realizadas pelo modelo de espaço de estado (em vermelho) e valores reais (em preto), para as séries de etanol (esquerda) e açúcar (direita). Fonte: elaborado pelo autor

5.3. Modelo de redes neurais LSTM

5.3.1. Arquitetura do modelo e variáveis de decisão

Por fim, será testado o modelo de redes neurais LSTM, que como vimos no capítulo 2 possui uma grande capacidade de aprender dependências temporais a curto e longo prazo. Os modelos de redes neurais não são paramétricos, pois são controlados por uma série de pesos e variáveis internos à rede que introduzem dinâmicas complexas e não lineares. Dessa forma, os modelos são descritos pela arquitetura e parâmetros da rede em si, e não por equações explícitas com as variáveis de entrada.

As principais decisões a serem tomadas para otimizar a rede são:

1. Além das séries $\boldsymbol{a_t}$ e $\boldsymbol{e_t}$ de preços de açúcar e etanol, quais outras variáveis serão alimentadas à rede. As possibilidades são as duas séries de primeiros futuros $\boldsymbol{f_t^a}$ e $\boldsymbol{f_t^e}$. (lembrando que antes do treinamento do modelo são aplicadas as diferenças logarítmicas $\boldsymbol{\nabla \log x_t = \log x_t - \log x_{t-1}}$ sobre todas as séries);
2. Quantos valores passados serão utilizados para cada série. Essa variável influencia muito no resultado: embora redes LSTM sejam desenhadas para aprender dependências temporais de longo prazo, a introdução de muitos valores pode gerar ruído que atrapalhará nas previsões, além de reduzir o tamanho do conjunto de treino (se *L* valores passados forem utilizados, serão perdidos *L* elementos). Como o espaço de observações já é restrito, serão consideradas apenas duas possibilidades: utilizar os valores dos **últimos três meses** ou dos **últimos seis meses**;
3. O número de estados internos da camada LSTM. Analogamente ao número de nós em uma rede neural *feedforward*, esse parâmetro determina quantas células LSTM funcionarão em paralelo: quanto maior esse número, mais facilmente a rede se adaptará aos dados de treino, mas maior será o risco de a rede possuir dificuldade para ser generalizada para dados nunca vistos. Esse fenômeno é conhecido como sobreaprendizado, e é discutido extensivamente em Hastie et al. (2009);
4. O número de iterações (épocas) de treinamento da rede neural. Para isso é utilizado o conjunto de validação: a taxa de erro nesse conjunto é medida durante o treino, e decide-se parar quando esse valor começa a aumentar, como explica Goodfellow et al. (2016). A Figura 44 mostra um exemplo típico de evolução das curvas de treino e teste durante o treinamento de uma rede neural

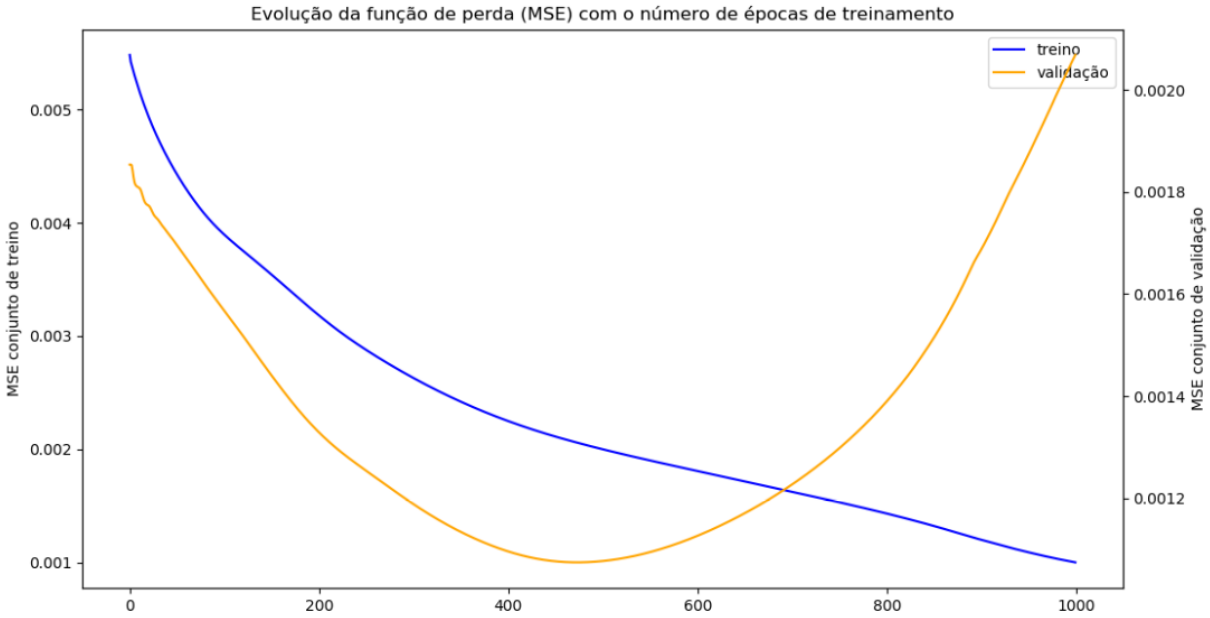

Figura 44: Evolução dos erros de treino (azul) e de validação (laranja) com o aumento do número de épocas de treinamento. É possível observar que a curva de treino decresce monotonicamente, enquanto que o erro de validação começa a crescer após cerca de 500 épocas de treinamento. Fonte: elaborado pelo autor

Com a utilização da série de primeiros futuros de etanol hidratado da BM&F Bovespa, o conjunto de dados disponível consiste em observações realizadas entre julho de 2010 e abril de 2018. O conjunto de teste será formado por 36 observações entre maio de 2015 e abril de 2018, o conjunto de validação por 18 observações entre novembro de 2013 e abril de 2015, e o conjunto de treino pelas observações restantes. O tamanho do conjunto de treino depende do número de valores passados a serem utilizados no treinamento: se forem três, o conjunto de treino será formado por 37 observações entre outubro de 2010 e outubro de 2013; caso sejam utilizados os últimos seis meses de observações para treinamento, esse valor cai para 34.

A divisão dos dados experimentais em 55 (ou 52) observações de treino + validação e 36 observações de teste corresponde à uma proporção de aproximadamente 60% dos dados utilizados para otimização da rede neural e 40% dos dados utilizados para teste, que é uma proporção razoável segundo Goodfellow et al. (2016).

A arquitetura da rede utilizada é simples: uma camada de neurônios LSTM que receberá P séries de N valores passados das variáveis de entrada e uma camada de saída com dois neurônios, um para prever os valores da série de etanol hidratado e outro para prever os

valores da série de açúcar cristal. A função objetivo a ser minimizada é o erro médio quadrado e o algoritmo de otimização para as rotinas de gradiente descendente estocástico é o Adam, proposto por Kingma e Ba (2014).

5.3.2. Previsões

Dada a arquitetura da rede, as diferentes combinações de variáveis de decisão, número de valores passados e número de estados ocultos da camada LSTM foram testados para escolher o melhor modelo. Como há muitas combinações, mostrar todos os testes atrapalharia a fluidez do trabalho e não agregaria nenhum valor: será mostrada somente a otimização do número de estados internos do melhor modelo, que utiliza $\boldsymbol{a_t}$, $\boldsymbol{e_t}$ e $\boldsymbol{f_t^e}$ como variáveis preditivas (os futuros de açúcar de Chicago são portanto deixados de fora) e **três valores passados** na entrada da rede.

A Tabela 22 mostra os resultados do modelo de redes neurais LSTM para diferentes números de estados internos da camada recorrente: é possível ver que a performance se estabiliza com 70 estados internos. Para o etanol, há uma melhoria de 22.4% no RMSE, 12.1% no MAE e 12.7% no MAPE; para o açúcar, as melhorias são de 26.2% no RMSE, 30.3% no MAE e 25.0% no MAPE. As previsões estão representadas graficamente na Figura 45.

	Etanol			Açúcar		
Nº de estados internos	RMSE	MAE	MAPE	RMSE	MAE	MAPE
10	0.075	0.055	9.5%	0.065	0.049	1.6%
20	0.074	0.061	10.5%	0.066	0.052	1.7%
30	0.068	0.054	9.3%	0.061	0.047	1.5%
40	0.068	0.053	9.2%	0.062	0.048	1.5%
50	0.067	0.053	9.2%	0.062	0.046	1.5%
60	0.067	0.052	9.0%	0.061	0.046	1.5%
70	**0.066**	**0.051**	**8.9%**	**0.062**	**0.046**	**1.5%**
80	0.066	0.051	8.8%	0.062	0.047	1.5%
Modelo de persistência	0.085	0.058	10.2%	0.084	0.066	2.0%

Tabela 22: Resultados da rede neural LSTM na previsão dos preços de etanol e açúcar, utilizando a série de primeiros futuros de açúcar como variável auxiliar e três valores passados na alimentação da rede

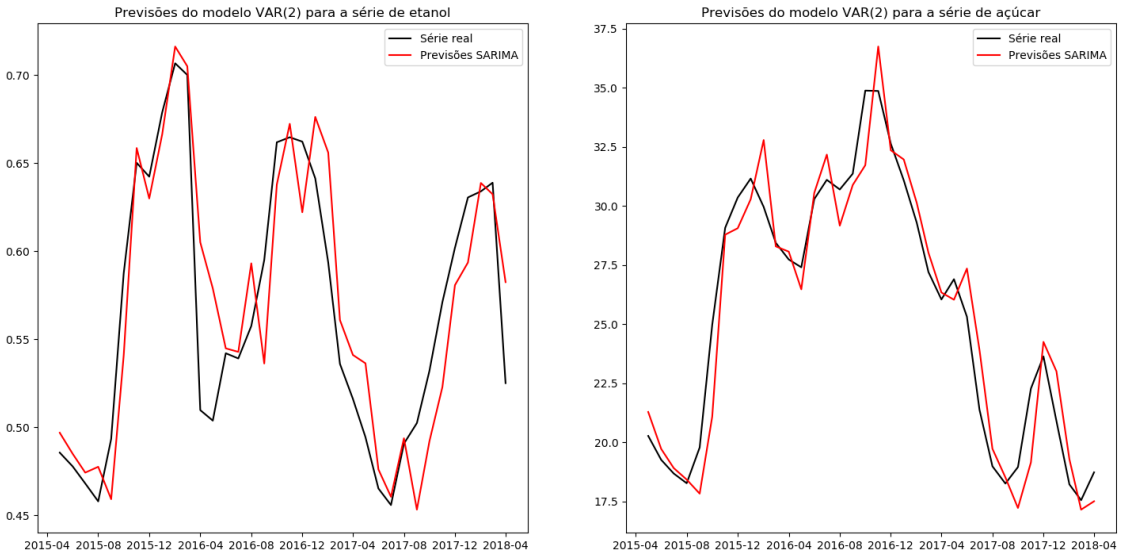

Figura 45: Gráficos das previsões do modelo ótimo de redes neurais LSTM para a série de preços de etanol (esquerda) e açúcar (direita). Fonte: elaborado pelo autor

5.3.3. Resíduos

Embora o modelo de redes neurais não faça nenhuma hipótese explícita sobre as distribuições de probabilidade ou outras propriedades das séries e de seus resíduos, ainda é interessante analisar os resíduos produzidos. É possível observar que, tanto para o etanol quanto para o açúcar, os resíduos não são normalmente distribuídos (Figura 46), e existe um resquício de correlação em série no limite da significância estatística para o etanol (Figura 47).

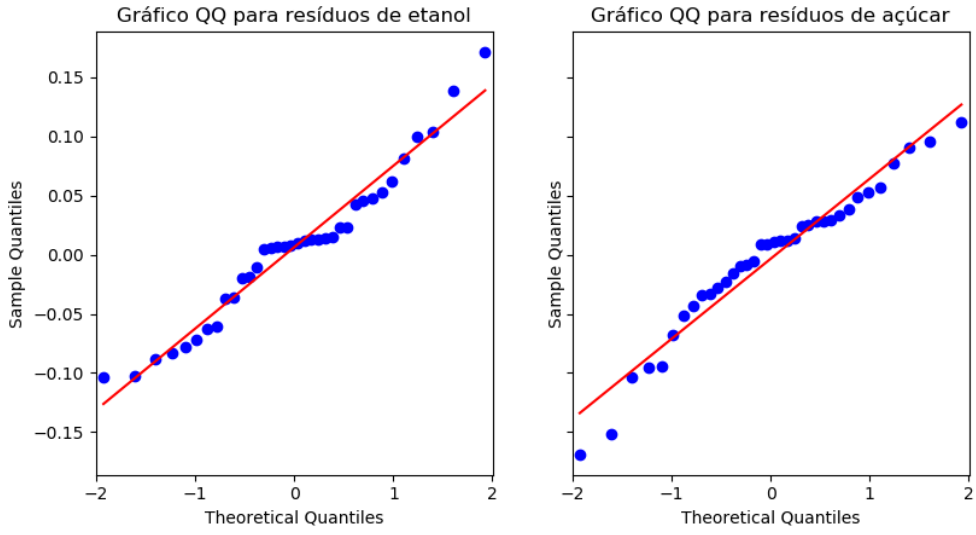

Figura 46: Gráfico QQ dos resíduos produzidos pelo modelo LSTM selecionado no item anterior. Fonte: elaborado pelo autor

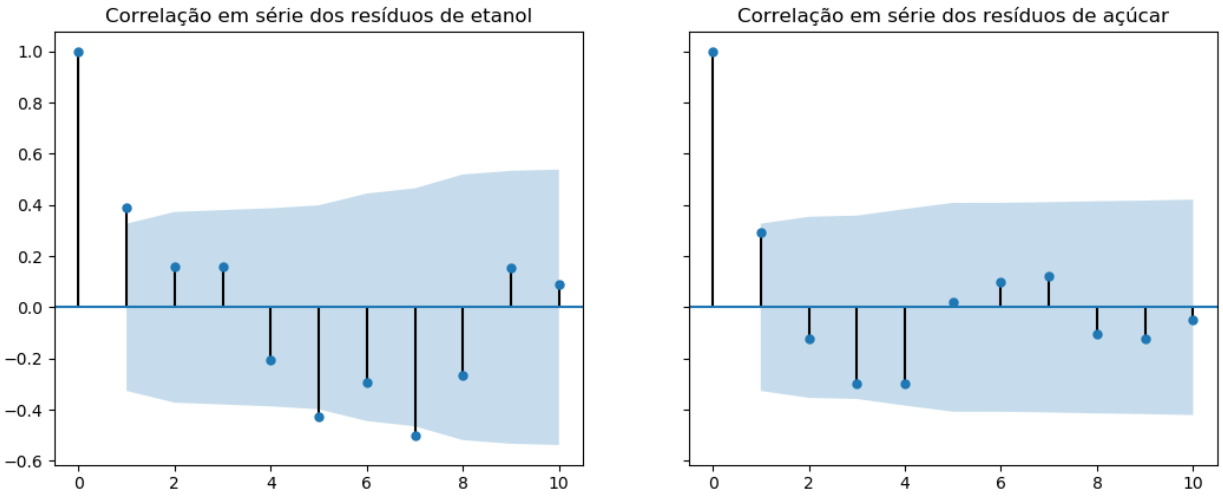

Figura 47: Função de autocorrelação dos resíduos produzidos pelo modelo LSTM. Fonte: elaborado pelo autor

5.4. Horizontes futuros e consolidação

5.4.1. Previsões até três meses no futuro

Nas três seções anteriores, foram construídos cinco modelos diferentes para a previsão de preços de açúcar e etanol, e esses modelos foram testados para realizar previsões um mês no futuro. Nessa seção, as previsões realizadas serão estendidas para até três meses no futuro e os modelos serão comparados para selecionar o melhor.

Estender as previsões para vários meses no futuro é muito natural para os modelos de séries temporais: basta aplicar recursivamente as equações sobre os resultados obtidos para obter uma sequência de previsões a um horizonte arbitrário. Para os modelos que utilizam variáveis exógenas, isso é mais complicado porque o modelo não prevê diretamente essas variáveis. Existem duas abordagens possíveis nesse caso: a primeira consiste em criar modelos paralelos para prever as variáveis exógenas e a segunda consiste em utilizá-las nas primeiras previsões e omiti-las quando não há mais valores disponíveis. O autor escolheu a segunda abordagem para evitar a propagação de incertezas decorrente do encadeamento de modelos de previsão.

As Tabelas 23 a 25 mostram os resultados obtidos pelos modelos na realização de previsões de um a três meses no futuro, comparando-os com o modelo de base que supõe que os preços se manterão constantes e iguais ao valor atual (modelo de persistência). Nos três casos, as melhores previsões foram realizadas pelo **modelo de espaço de estado**, que inclui os preços dos primeiros contratos futuros como variáveis exógenas.

5.4.2. Discussão dos resultados

A modelagem iniciou-se com a aplicação da metodologia de Box e Jenkins (1970) para analisar individualmente as séries de preços de açúcar cristal e etanol hidratado baseado unicamente em padrões observados historicamente, traduzidos em termos sazonais, autorregressivos e de média móvel. Foi detectada uma dinâmica predominantemente sazonal para o etanol e autorregressiva para o açúcar, e as previsões realizadas representaram melhorias significativas com relação ao modelo de base.

Previsões para o horizonte de **um mês** no futuro						
	Etanol			Açúcar		
Modelo	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SARIMA (açúcar)	-	-	-	0.062	0.044	1.4%
SARIMA (etanol)	0.074	0.050	8.7%	-	-	-
VAR	0.083	0.059	10.2%	0.060	0.044	1.4%
Espaço de estado	**0.067**	**0.044**	**7.7%**	**0.059**	**0.046**	**1.4%**
Redes neurais	0.066	0.051	8.9%	0.062	0.046	1.5%
Modelo de base	0.085	0.058	10.2%	0.084	0.066	2.0%

Tabela 23: Resultados consolidados dos modelos para previsões um mês no futuro

Previsões para o horizonte de **dois meses** no futuro						
	Etanol			Açúcar		
Modelo	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SARIMA (açúcar)	-	-	-	0.129	0.097	3.0%
SARIMA (etanol)	0.106	0.079	14.2%	-	-	-
VAR	0.137	0.112	20.2%	0.133	0.100	3.1%
Espaço de estado	**0.093**	**0.065**	**11.6%**	**0.105**	**0.084**	**2.7%**
Redes neurais	0.136	0.108	19.4%	0.144	0.110	3.5%
Modelo de base	0.132	0.105	18.9%	0.148	0.115	3.6%

Tabela 24: Resultados consolidados dos modelos para previsões dois meses no futuro

Previsões para o horizonte de **três meses** no futuro						
	Etanol			Açúcar		
Modelo	RMSE	MAE	MAPE	RMSE	MAE	MAPE
SARIMA (açúcar)	-	-	-	0.176	0.137	4.3%
SARIMA (etanol)	0.123	0.103	19.1%	-	-	-
VAR	0.177	0.152	28.3%	0.185	0.143	4.5%
Espaço de estado	**0.101**	**0.077**	**13.6%**	**0.128**	**0.097**	**3.1%**
Redes neurais	0.165	0.137	24.4%	0.202	0.158	5.0%
Modelo de base	0.168	0.146	27.0%	0.189	0.154	4.8%

Tabela 25: Resultados consolidados dos modelos para previsões três meses no futuro

Para a série de açúcar, é possível ver que os resultados do modelo SARIMA vão se aproximando do *benchmark* à medida que o horizonte de previsão se torna mais longo, refletindo a reversão rápida à média de um modelo puramente autorregressivo; o mesmo não ocorre para o modelo de etanol a curto prazo por conta da sua sazonalidade.

A modelagem individual é importante para a análise exploratória e compreensão inicial das séries, mas é de interesse limitado para técnicas de gestão de portfólio e minimização de risco que são de maior interesse para o planejamento da produção em uma usina de cana-de-açúcar. Por isso, o trabalho segue com a aplicação de três modelos bidimensionais, de forma a aproveitar as interações entre as séries de açúcar e etanol e, se possível, gerar estimativas conjuntas de distribuições de probabilidades para valores futuros.

O primeiro desses três modelos foi um modelo VAR (*Vector Autoregression*). Esse modelo possui limitações devido à incapacidade de estimar sazonalidades, e isso se traduz nas previsões ruins realizadas para a série de etanol. Outra limitação é a impossibilidade de omitir termos de defasagem menores do que a ordem *p* do modelo: no caso desse trabalho, utilizar um VAR(2) implicou a inclusão de termos autorregressivos e interações cruzadas de primeira e segunda ordem para as duas séries, independente de sua significância estatística. Isso gera ruídos que vão se amplificando à medida que as previsões vão se estendendo para o futuro, e resulta na deterioração rápida das previsões com o aumento do horizonte futuro.

Essas limitações, assim como a incapacidade de incorporar variáveis externas, são resolvidas pelo modelo de espaço de estados. A grande adaptabilidade da sua formulação geral e a facilidade de sua implementação e resolução computacional devido ao algoritmo do filtro de Kalman permitiram ao autor selecionar apenas os termos relevantes identificados pela análise SARIMA e incluir também as variáveis exógenas de grande poder preditivo identificadas pelo capítulo 4. Os resíduos desse modelo se mostraram mais regulares (inclusive com aceitação da hipótese de normalidade para o açúcar) e a melhoria com relação ao modelo de base foi mantida para todos os horizontes de previsão a curto prazo (de um a três meses), provando a superioridade desse modelo com relação aos demais.

Por último, foi testada uma abordagem alternativa não-paramétrica com o uso de redes neurais recorrentes do tipo LSTM (Long Short-Term Memory). Essas redes neurais obtiveram resultados similares aos do modelo de espaço de estado para previsões em um horizonte de um mês, o que é surpreendente dado o conjunto limitado de dados para treino à

sua disposição e atesta o seu potencial para a utilização em estruturas de dados sequenciais. No entanto, a ausência de uma formulação explícita de equações de recursão sobre as saídas do modelo dificulta a realização previsões a horizontes mais distantes, o que se refletiu na degradação rápida de sua qualidade com relação ao modelo de base.

6. CONCLUSÃO

Nesse trabalho, foram construídos e comparados quatro modelos diferentes de previsão de preços sucroalcooleiros. O modelo de melhor performance obteve melhorias superiores a 30% com relação ao modelo de base tanto para o etanol quanto para o açúcar, considerando diferentes métricas de avaliação e diferentes horizontes de previsão, ao mesmo tempo em que permitiu relacionar os resíduos das duas séries ao calcular a sua matriz de covariância. Essa formulação atende aos pré-requisitos para utilização na modelagem de Rockafellar e Uryasev (2000) e pode ser aplicada em casos reais por gestores de usinas sucroalcooleiras para decisões de *mix* de produção e gestão de estoques, reduzindo o risco financeiro a que estão submetidos, que tem penalizado severamente produtores de álcool e de açúcar no território brasileiro nos últimos anos.

Para chegar a esse resultado, foi importante compreender o funcionamento da indústria sucroalcooleira como um todo e estudar suas interações com outros mercados, como o de contratos futuros em bolsas de valores e o de combustíveis no território nacional. A metodologia de análise de inclusão de variáveis exógenas utilizando funções de correlação cruzada (CCF) foi fundamental para estabelecer um critério objetivo de seleção de variáveis relevantes, baseado na existência de correlação temporal e no conceito de causalidade no sentido de Granger (1969).

A ausência de normalidade nos resíduos para o etanol poderá ser corrigida em futuros trabalhos com a inclusão de variações estocásticas também para as inovações na equação de estado, como feito por Schwartz (1997). Outra possível melhoria é a utilização de contratos futuros de maturidades diversas, possibilitando a realização de previsões de preços para horizontes mais longos.

Embora esse trabalho seja focado em usinas sucroalcooleiras do estado de São Paulo, os modelos utilizados podem ser facilmente adaptados para séries temporais relativas a outros produtos e outras localidades. O autor espera que sua abordagem e metodologia possam ser aproveitadas por outros estudantes na área de modelagem estatística.

REFERÊNCIAS BIBLIOGRÁFICAS

ADEBIYI, A.; AYO, C.; ADEWUMI, A. Stock price prediction using the ARIMA model. **UKSim-AMSS 16th Conference on Computer Modelling and Simulation**, p. 105-111, 2014.

BAEK, Y.; KIM, H. Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. **Expert Systems with Applications**, v. 113, p. 457-480, 2018.

BERGMEIR, C.; BENÍTEZ, J. M. On the use of cross-validation for time-series predictor evaluation. **Information Sciences**, v. 191, p. 192-213, 2012.

BESSEMBINDER, H; COUGHENOUR, J. F; SEGUIN, P. J. Mean reversion in equilibrium asset prices: evidence from the futures term structure. **The Journal of Finance**, v. 50, n. 1, p. 361-375, 1995.

BOWMAN, C.; HUSAIN, A. M. Forecasting commodity prices: futures versus judgement. **International Monetary Fund Working Papers**, 2014.

BOX, G.; JENKINS, G. Time series analysis: forecasting and control. 1 ed. **Holden Day**, 1970.

BRAUNBECK, O.; MAGALHÃES, P. Colheita de cana-de-açúcar com auxílio mecânico. **Faculdade de Engenharia Agrícola, UNICAMP**, 2006.

CYBENKO, G. Approximations by superpositions of sigmoidal functions. **Mathematics of Control, Signals and Systems**, v. 2, n. 4, p. 303-314, 1989.

DURBIN, J.; KOOPMAN, S. J. Time series analysis by state space methods. 2 ed. **Oxford University Press**, 2012.

EICHLER, M. Causal inference in time-series analysis. In: BERZUINI, C.; DAWID, P.; BERNARDINELLI, L. Causality: statistical perspectives and applications, **Wiley Series in Probability and Statistics**, 2012.

EICHLER, M. Causal inference with multiple time series: principles and problems. **Philosophical Transactions of the Royal Society (series A)**, v. 371, n. 1997, p. 1-17, 2013.

FAMA, E. F. The behaviour of stock-market prices. **Journal of Business**, v. 38, n. 1, p. 34-105, 1965.

FAMA, E. F. Efficient capital markets: a review of theory and empirical work. **Journal of Finance**, v. 25, n. 2, p. 383-417, 1970.

FENG, E. *et al.* Log-transformation and its implications for data analysis. **Shanghai Archives of Psychiatry**, v. 26, n. 2, p. 105-109, 2014.

FREITAS, L.; KANEKO, S. Ethanol demand in Brazil: regional approach. **Energy Policy**, v. 39, p. 2289-2298, 2011.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. **MIT Press**, 2016.

GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. **Econometrica**, v. 37, n. 3, p. 424–438, 1969.

GRAVES, A. *et al*. A novel connectionist system for improved unconstrained handwriting recognition. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 5, p. 855-868, 2009.

GREWAL, M.; ANDREWS, A. Applications of Kalman filtering in aerospace 1960 to the present. **IEEE Control Systems Magazine**, v. 30, n. 3, p. 69-78, 2010.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The elements of statistical learning. 2 ed. **Springer Series in Statistics**, 2009.

HOCHREITER, S. Untersuchungen zu dynamischen neuronalen Netzen. **Tese de graduação - TU München**, Munique, 1991.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735-1780, 1997.

HORNIK, K. Approximation capabilities of multilayer feedforward networks. **Neural Networks**, v. 4, n. 2, p. 251-257, 1991.

JAMES, G *et al*. An introduction to statistical learning. 1 ed. **Springer Series in Statistics**, 2013.

JANOTTI, P. *et al*. A logística do açúcar e do etanol entre usinas paulistas e o porto de Santos: um estudo comparativo entre agentes comerciais. **Revista de Administração da UNIMEP**, v. 10, n. 2, p. 101-126, 2012.

KENDALL, M. G. The analysis of economic time series - part I: prices. **Journal of the Royal Statistical Sociaty (series A),** v. 116, n. 1, p. 11-34, 1953.

KINGMA, D. P.; BA, J. Adam: a method for stochastic optimization. **3rd International Conference for Learning Representations**, San Diego, 2015.

LAGO, J.; RIDDER, F.; SCHUTTER, B. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. **Applied Energy**, v. 221, p. 386-405, 2018.

LEE, D. Econometric assessment of bioenergy development. **International Journal of Hydrogen Energy**, v. 42, p. 27701-27717, 2017.

LI, X.; WU, X. Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition. **Peking University**, 2015.

MCCULLOCH, W.; PITTS, W. A. A logical calculus of ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115-133, 1943.

MILANEZ, A. *et al*. Logística para o etanol: situação atual e desafios futuros. **BNDES Setorial - Sucroenergético**, v. 31, p. 49-98, 2010.

NEVES, M.; GRAY, A.; BOURQUARD, B. Copersucar: a world leader in sugar and ethanol. **International Food and Agribusiness Management Review**, v. 19, n. 2, p. 207-240, 2016.

NOVACANA. Novacana, 2018. Tudo sobre etanol, cana, açúcar e cogeração. Disponível em <https://novacana.com/>. Acesso em: 8 de ago. de 2018.

OLIVEIRA, S.; RIBEIRO, C.; CICOGNA, M. P. Uncertainty effects on production mix and on hedging decisions: the case of Brazilian ethanol and sugar. **Energy Economics**, v. 70, p. 516-524, 2018.

REICHSFELD, D. A.; ROACHE, S. K. Do commodity futures help forecast spot prices? **International Monetary Fund Working Papers,** 2011.

RIBEIRO, C.; OLIVEIRA, S. A hybrid commodity price-forecasting model applied to the sugar-alcohol sector. **The Australian Journal of Agricultural and Resource Economics**, v. 55, p. 180-198, 2011.

ROCKAFELLAR, R.; URYASEV, S. Optimization of conditional value-at-risk. **Journal of Risk**, v. 2, p. 21-41, 2000.

RONQUIM, C. Queimada na colheita da cana-de-açúcar: impactos ambientais, sociais e econômicos. **Embrapa Monitoramento por Satélite. Documentos**, v. 77, p. 1-48, 2010.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386-408, 1958.

ROUT, A. *et al*. Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. **Journal of King Saud University - Computer and Information Sciences**, v. 29, n. 4, p. 536-552, 2017.

RUMELHART, D.; HINTON, G.; WILLIAMS, R. Learning internal representations by error propagation. In: RUMELHART, D.; MCCLELLAND, J. Parallel distributed processing: explorations in the microstructure of cognition, **MIT Press Cambridge**, v. 1, p. 318-362, 1986.

SANTOS, M. *et* al. Incerteza na cadeia de exportação de açúcar. **Pretexto**, v. 14, n. 3, p. 64-80, 2013.

SCHWARTZ, E, S. The stochstic behavior of commodity prices: implications for valuation and hedging. **The Journal of Finance**, v. 52, n. 3, p. 923-973, 1997.

SHUMWAY, R.; STOFFER, D. Time series analysis and its applications. 3 ed. **Springer Science**, 2011.

TUSELL, F. Kalman filtering in R. **Journal of Statistical Software**, v. 39, n. 2, 2011.

ÚNICA (União da Indústria da Cana-de-Açúcar). Unicadata, 2018. Relatórios sobre a colheita de cana-de-açúcar. Disponível em: <www.unicadata.com.br>. Acesso em: 9 de set. de 2018.

WERBOS, P. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550-1560, 1990.

WETS, R.; RIOS, I. Modeling and estimating commodity prices: copper. **Universities of California and Chile**, 2013.

ZAFEIRIOU, E. *et al*. The impact of energy prices on the volatility of ethanol prices and the role of gasoline emissions. **Renewable and Sustainable Energy Reviews**, v. 33, p. 87-95, 2014.

ZHANG, Y.; LIU, L. The lead-lag relationships between spot and futures prices of natural gas. **Physica A**, v. 490, p. 203-211, 2018.

ANEXO: CÓDIGO FONTE DO MODELO ESCOLHIDO

```
rm(list = ls())
require(FKF)
require(dplyr)
require(ggpubr)
require(tsoutliers)
require(stats)

#definir diretório de trabalho
setwd("C:\\Users\\c.camilli\\Desktop\\Faculdade\\Poli\\TF\\Dados")

#função para o cálculo do índice de sazonalidade
calculate_seasonal_et <- function(series, ma_coef, ar_coef){
  depth = length(series)%/%6 - 2
  lista = rep(0, depth)
  for (i in 1:depth){
    lista[i] = 6*i
  }
  vec = rep(0, length(series))
  ma_coef = 0.889
  ar_coef = -0.981
  for (i in 1:length(lista)){
    factor = ((-1)**(i+1))*(ar_coef*(ma_coef**(i-1)) + (ma_coef**i))
    col = c(rep(0, lista[i]), series[1:(length(series)-lista[i])])
    vec = vec + col*factor
  }
  return (vec)
}

#função que define explicitamente as equações do filtro de Kalman
kalman_f <- function(phi1_a, phi2_a, sigma2_a, sigma2_e, sigma2_ae, alpha_fet, alpha_fat,
alpha_set){
  if (sigma2_ae<0){
    sigma2_ae = -1*sigma2_ae
  }
  ct <- matrix(c(0, alpha_set, 0, alpha_fet, alpha_fat, 0), nrow=2)%*%matrix(t(dfFinal[begin:end,
c("Et_season", "Fet_1", "Fat_1")]), nrow=3)
  Tt <- matrix(c(phi1_a, 0, 1, 0, 0, 0, 0, 1, phi2_a, 0, 0, 0, 0, 0, 0, 0), nrow=4)
  HHt <- matrix(c(sigma2_a, sigma2_ae, 0, 0, sigma2_ae, sigma2_e, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
nrow=4)
  dt <- matrix(c(0,0,0,0), nrow=4)
  Zt <- matrix(c(1, 0, 0, 1, 0, 0, 0, 0), nrow=2)
  GGt <- matrix(c(0, 0, 0, 0), nrow=2)
  a0 <- c(dfFinal[begin, "At"], dfFinal[begin, "Et"], dfFinal[begin, "At_1"], dfFinal[begin, "Et_1"])
  P0 <- matrix(c(var(dfFinal[begin:end, "At"]), 0, 0, 0, 0, var(dfFinal[begin:end, "Et"]), 0, 0, 0, 0,
var(dfFinal[begin:end, "At_1"]), 0, 0, 0, 0, var(dfFinal[begin:end, "Et_1"])), nrow=4)
  return(list(a0=a0, P0=P0, ct=ct, dt=dt, Zt=Zt, Tt=Tt, GGt=GGt, HHt=HHt))
}

#função objetivo a ser minimizada: verossimilhança do filtro de Kalman
```

```
objective <- function(theta, yt){
  sp <- kalman_f(theta["phi1_a"], theta["phi2_a"], theta["sigma2_a"], theta["sigma2_e"],
theta["sigma2_ae"], theta["alpha_fet"], theta["alpha_fat"], theta["alpha_set"])
  ans <- fkf(a0=sp$a0, P0=sp$P0, dt=sp$dt, ct=sp$ct, Tt=sp$Tt, Zt=sp$Zt, HHt=sp$HHt,
GGt=sp$GGt, yt=yt)
  return(-ans$logLik)
}

#ler e formatar dados
dfInflacao = read.csv("ipca_brasil_desde_2001.csv", sep=';')
dfInflacao$Date = as.Date(dfInflacao$Date, format="%d/%m/%Y")

dfAcucar = read.csv("cepea_acucar_cristal_produtor_mensal_sp.csv", sep=";")
dfAcucar$Date = as.Date(dfAcucar$Date, format="%d/%m/%Y")
dfAcucar = dfAcucar[order(dfAcucar$Date),c("Date", "Preco_BRL")]
dfAcucar = merge(dfAcucar, dfInflacao, by="Date")
dfAcucar$Preco_BRL = log(dfAcucar$Preco_BRL/dfAcucar$Index)
dfAcucarDiff = data.frame(Date = dfAcucar[2:dim(dfAcucar)[1], 'Date'])
dfAcucarDiff$At = dfAcucar[2:dim(dfAcucar)[1], 'Preco_BRL'] - dfAcucar[1:(dim(dfAcucar)[1]-1),
'Preco_BRL']
dfAcucarDiff$Preco_BRL_Acucar = dfAcucar[2:dim(dfAcucar)[1],"Preco_BRL"]

dfEtanol = read.csv("cepea_etanol_hidratado_mensal_sp_comb.csv", sep=";")
dfEtanol$Date = as.Date(dfEtanol$Date, format="%d/%m/%Y")
dfEtanol = dfEtanol[order(dfEtanol$Date),c("Date", "Preco_BRL")]
dfEtanol = merge(dfEtanol, dfInflacao, by="Date")
dfEtanol$Preco_BRL = log(dfEtanol$Preco_BRL/dfEtanol$Index)
dfEtanolDiff = data.frame(Date = dfEtanol[2:dim(dfEtanol)[1], 'Date'])
dfEtanolDiff$Et = dfEtanol[2:dim(dfEtanol)[1], 'Preco_BRL'] - dfEtanol[1:(dim(dfEtanol)[1]-1),
'Preco_BRL']
dfEtanolDiff$Et_season = calculate_seasonal_et(dfEtanolDiff$Et, 0.889, -0.981)
dfEtanolDiff$Preco_BRL_Etanol = dfEtanol[2:dim(dfEtanol)[1],"Preco_BRL"]

dfBovespa = read.csv("bovespa_etanol_hidratado_futures.csv", sep=";")
dfBovespa$Date = as.Date(dfBovespa$Date, format="%d/%m/%Y")
colnames(dfBovespa)[2] = "Ethanol_Futures"
dfBovespa$Ethanol_Futures = dfBovespa$Ethanol_Futures/1000
dfBovespa = merge(dfBovespa[,c("Date", "Ethanol_Futures")], dfInflacao, by="Date")
dfBovespa$Fet = log(dfBovespa$Ethanol_Futures/dfBovespa$Index)
dfBovespaDiff = data.frame(Date = dfBovespa[2:dim(dfBovespa)[1], 'Date'])
dfBovespaDiff$Fet = dfBovespa[2:dim(dfBovespa)[1], 'Fet'] - dfBovespa[1:(dim(dfBovespa)[1]-1),
'Fet']

dfChicago = read.csv("chicago_sugar11_futures_final.csv", sep=";")
dfChicago$Date = as.Date(dfChicago$Date, format="%d/%m/%Y")
colnames(dfChicago)[2] = "Sugar_Futures"
dfChicago$Ethanol_Futures = dfChicago$Ethanol_Futures
dfChicago = merge(dfChicago[,c("Date", "Sugar_Futures")], dfInflacao, by="Date")
dfChicago$Fat = log(dfChicago$Sugar_Futures/dfChicago$Index)
dfChicagoDiff = data.frame(Date = dfChicago[2:dim(dfChicago)[1], 'Date'])
dfChicagoDiff$Fat = dfChicago[2:dim(dfChicago)[1], 'Fat'] - dfChicago[1:(dim(dfChicago)[1]-1),
'Fat']

dfTudo = merge(dfAcucarDiff, dfEtanolDiff, by="Date", suffixes=c("_Acucar", "_Etanol"))
dfTudo = merge(dfTudo, dfBovespaDiff, by="Date")
```

```
dfTudo = merge(dfTudo, dfChicagoDiff, by="Date")

dfFinal = data.frame(Date = dfTudo[3:dim(dfTudo)[1], 'Date'])
dfFinal$At = dfTudo[3:dim(dfTudo)[1], 'At']
dfFinal$Et = dfTudo[3:dim(dfTudo)[1], 'Et']
dfFinal$Et_season = dfTudo[3:dim(dfTudo)[1], 'Et_season']
dfFinal$At_1 = dfTudo[2:(dim(dfTudo)[1]-1), 'At']
dfFinal$At_2 = dfTudo[1:(dim(dfTudo)[1]-2), 'At']
dfFinal$Et_1 = dfTudo[2:(dim(dfTudo)[1]-1), 'Et']
dfFinal$Et_2 = dfTudo[1:(dim(dfTudo)[1]-2), 'Et']
dfFinal$Fet_1 = dfTudo[2:(dim(dfTudo)[1]-1), 'Fet']
dfFinal$Fat_1 = dfTudo[2:(dim(dfTudo)[1]-1), 'Fat']

### determinação dos parâmetros do modelo
begin = 1
end = dim(dfFinal)[1]
yt = matrix(t(dfFinal[begin:end, c("At", "Et")]), nrow=2)
theta <- c(phi1_a=0.66, phi2_a=-0.33, sigma2_a = var(dfFinal[begin:end, "At"]), sigma2_e =
var(dfFinal[begin:end, "Et"]),
        sigma2_ae = cov(dfFinal[begin:end, "Et"], dfFinal[begin:end, "At"]), alpha_fet=0.4,
alpha_fat=0.4, alpha_set=1)
fit <- optim(theta, objective, yt = rbind(yt), hessian=TRUE)
SE = sqrt(diag(solve(fit$hessian)))
u = round(cbind(estimate=fit$par, SE, t_value = fit$par/SE), 6)
rownames(u) = c("phi1_a", "phi2_a", "sigma2_a", "sigma2_e", "sigma2_ae", "alpha_fet", "alpha_fat",
"alpha_set")
print(u)


### cálculo e análise dos resíduos
sp <- kalman_f(fit$par["phi1_a"], fit$par["phi2_a"], fit$par["sigma2_a"], fit$par["sigma2_e"],
fit$par["sigma2_ae"], fit$par["alpha_fet"], fit$par["alpha_fat"], fit$par["alpha_set"])
ans <- fkf(a0=sp$a0, P0=sp$P0, dt=sp$dt, ct=sp$ct, Tt=sp$Tt, Zt=sp$Zt, HHt=sp$HHt,
GGt=sp$GGt, yt=rbind(yt))

c_mat = matrix(c(0, fit$par["alpha_set"], 0, fit$par["alpha_fet"], fit$par["alpha_fat"], 0),
nrow=2)%*%matrix(t(dfFinal[begin:end, c("Et_season", "Fet_1", "Fat_1")]), nrow=3)
Zt <- matrix(c(1, 0, 0, 1, 0, 0, 0, 0), nrow=2)
preds = c_mat + Zt%*%ans$at[,1:(end-begin+1)]

preds_et = preds[2,]
preds_ac = preds[1,]
real_et = dfFinal$Et
real_ac = dfFinal$At
resid_et = real_et - preds_et
resid_ac = real_ac - preds_ac

JarqueBera.test(resid_ac)
JarqueBera.test(resid_et)

Box.test(resid_ac, type="Ljung-Box")
Box.test(resid_et, type="Ljung-Box")

par(mfrow=c(2,1))
```

```
Time = dfFinal[begin:end, "Date"]
plot(Time, resid_et, main="Resíduos do modelo para a série de etanol")
plot(Time, resid_ac, main="Resíduos do modelo para a série de açúcar")

par(mfrow=c(2,1))
et_acf = acf(resid_et, plot=FALSE)
ac_acf = acf(resid_ac, title="ACF para os resíduos do modelo (açúcar)", plot=FALSE)
plot(et_acf, main="ACF para os resíduos do modelo (etanol)")
plot(ac_acf, main="ACF para os resíduos do modelo (açúcar)")

par(mfrow=c(2,1))
ggqqplot(resid_et, title="Gráfico QQ dos resíduos (etanol)")
ggqqplot(resid_ac, title="Gráfico QQ dos resíduos (açúcar)")

### validação cruzada para avaliação do modelos

#janela de treinamento
l = 48

#horizonte de previsão em meses
steps_ahead = 3

#inicialização das variáveis
previsoes_et = rep(0, 36)
previsoes_ac = rep(0, 36)
reais_ac = rep(0, 36)
reais_et = rep(0, 36)

it = 0
bench_RMSE_ac = 0
model_RMSE_ac = 0
bench_MAE_ac = 0
model_MAE_ac = 0
bench_MAPE_ac = 0
model_MAPE_ac = 0

bench_RMSE_et = 0
model_RMSE_et = 0
bench_MAE_et = 0
model_MAE_et = 0
bench_MAPE_et = 0
model_MAPE_et = 0

for (i in (dim(dfFinal)[1]-35):dim(dfFinal)[1])
{

it = it + 1
print(i-dim(dfFinal)[1]+36)

#definir janela de treinamento
begin = i-l-(steps_ahead-1)
end = i-steps_ahead
yt = matrix(t(dfFinal[begin:end, c("At", "Et")]), nrow=2)

#parâmetros iniciais do modelo
```

```
theta <- c(phi1_a=0.66, phi2_a=-0.33, sigma2_a = var(dfFinal[begin:end, "At"]), sigma2_e =
var(dfFinal[begin:end, "Et"]),
        sigma2_ae = cov(dfFinal[begin:end, "Et"], dfFinal[begin:end, "At"]), alpha_fet=0.4,
alpha_fat=0.4, alpha_set=1)

#rotina de otimização dos parâmetros
fit <- optim(theta, objective, yt = rbind(yt), hessian=TRUE)

#cálculo do filtro com parâmetros ótimos
sp <- kalman_f(fit$par["phi1_a"], fit$par["phi2_a"], fit$par["sigma2_a"], fit$par["sigma2_e"],
fit$par["sigma2_ae"], fit$par["alpha_fet"], fit$par["alpha_fat"], fit$par["alpha_set"])
ans <- fkf(a0=sp$a0, P0=sp$P0, dt=sp$dt, ct=sp$ct, Tt=sp$Tt, Zt=sp$Zt, HHt=sp$HHt,
GGt=sp$GGt, yt=rbind(yt))

#realizar previsões
pred_etanol=0
pred_acucar=0
int_state = ans$at[,l+1]
Tt <- matrix(c(fit$par["phi1_a"], 0, 1, 0, 0, 0, 0, 1, fit$par["phi2_a"], 0, 0, 0, 0, 0, 0, 0), nrow=4)

for (k in 1:steps_ahead){
  if (k==1){
  c_mat = matrix(c(0, fit$par["alpha_set"], 0, fit$par["alpha_fet"], fit$par["alpha_fat"], 0),
nrow=2)%*%matrix(t(dfFinal[end+k, c("Et_season", "Fet_1", "Fat_1")]), nrow=3)
  Zt <- matrix(c(1, 0, 0, 1, 0, 0, 0, 0), nrow=2)
  preds = c_mat + Zt%*%int_state
  pred_etanol = pred_etanol + preds[2]
  pred_acucar = pred_acucar + preds[1]
  }
  else{
    c_mat = matrix(c(0, fit$par["alpha_set"], 0, 0, 0, 0), nrow=2)%*%matrix(t(dfFinal[end+k,
c("Et_season", "Fet_1", "Fat_1")]), nrow=3)
    Zt <- matrix(c(1, 0, 0, 1, 0, 0, 0, 0), nrow=2)
    preds = c_mat + Zt%*%int_state
    pred_etanol = pred_etanol + preds[2]
    pred_acucar = pred_acucar + preds[1]
  }
  int_state = Tt%*%int_state
}

real_etanol = dfTudo[i+2, "Preco_BRL_Etanol"]
prev_etanol = pred_etanol + dfTudo[i+1, "Preco_BRL_Etanol"]
bench_etanol = dfTudo[i+(2-steps_ahead), "Preco_BRL_Etanol"]
previsoes_et[it] = prev_etanol
reais_et[it] = real_etanol

real_acucar = dfTudo[i+2, "Preco_BRL_Acucar"]
prev_acucar = pred_acucar + dfTudo[i+1, "Preco_BRL_Acucar"]
bench_acucar = dfTudo[i+(2-steps_ahead), "Preco_BRL_Acucar"]
previsoes_ac[it] = prev_acucar
reais_ac[it] = real_acucar

model_RMSE_ac = model_RMSE_ac + (real_acucar - prev_acucar)**2
bench_RMSE_ac = bench_RMSE_ac + (real_acucar - bench_acucar)**2
```

```
model_MAE_ac = model_MAE_ac + abs(real_acucar - prev_acucar)
bench_MAE_ac = bench_MAE_ac + abs(real_acucar - bench_acucar)
model_MAPE_ac = model_MAPE_ac + abs((real_acucar - prev_acucar)/real_acucar)
bench_MAPE_ac = bench_MAPE_ac + abs((real_acucar - bench_acucar)/real_acucar)

model_RMSE_et = model_RMSE_et + (real_etanol - prev_etanol)**2
bench_RMSE_et = bench_RMSE_et + (real_etanol - bench_etanol)**2
model_MAE_et = model_MAE_et + abs(real_etanol - prev_etanol)
bench_MAE_et = bench_MAE_et + abs(real_etanol - bench_etanol)
model_MAPE_et = model_MAPE_et + abs((real_etanol - prev_etanol)/real_etanol)
bench_MAPE_et = bench_MAPE_et + abs((real_etanol - bench_etanol)/real_etanol)
}


#### cálculo e impressão dos resultados finais
model_RMSE_ac = sqrt(model_RMSE_ac/(36))
bench_RMSE_ac = sqrt(bench_RMSE_ac/(36))
model_MAE_ac = model_MAE_ac/36
bench_MAE_ac = bench_MAE_ac/36
model_MAPE_ac = model_MAPE_ac/36
bench_MAPE_ac = bench_MAPE_ac/36

model_RMSE_et = sqrt(model_RMSE_et/(36))
bench_RMSE_et = sqrt(bench_RMSE_et/(36))
model_MAE_et = model_MAE_et/36
bench_MAE_et = bench_MAE_et/36
model_MAPE_et = model_MAPE_et/36
bench_MAPE_et = bench_MAPE_et/36


cat("RMSE of SARIMAX model for sugar:", model_RMSE_ac)
cat("RMSE of baseline model for sugar:", bench_RMSE_ac)

cat("\nMAE of SARIMAX model for sugar:", model_MAE_ac)
cat("MAE of baseline model for sugar:", bench_MAE_ac)

cat("\nMAPE of SARIMAX model for sugar:", model_MAPE_ac)
cat("MAPE of baseline model for sugar:", bench_MAPE_ac)

cat("\n\nRMSE of SARIMAX model for ethanol:", model_RMSE_et)
cat("RMSE of baseline model for ethanol:", bench_RMSE_et)

cat("\nMAE of SARIMAX model for ethanol:", model_MAE_et)
cat("MAE of baseline model for ethanol:", bench_MAE_et)

cat("\nMAPE of SARIMAX model for ethanol:", model_MAPE_et)
cat("MAPE of baseline model for ethanol:", bench_MAPE_et)
```