



Fundação Educacional do Município de Assis
Instituto Municipal de Ensino Superior de Assis
Campus "José Santilli Sobrinho"

ANDRÉ BARBOSA ALVES DOS SANTOS

**MODELAGEM E PREVISÃO DO DIFERENCIAL DE BASE DA SOJA DO
MUNICÍPIO DE CÂNDIDO MOTA: UMA APLICAÇÃO DO MODELO SARIMA**

**Assis/SP
2024**



Fundação Educacional do Município de Assis
Instituto Municipal de Ensino Superior de Assis
Campus "José Santilli Sobrinho"

ANDRÉ BARBOSA ALVES DOS SANTOS

**MODELAGEM E PREVISÃO DO DIFERENCIAL DE BASE DA SOJA DO
MUNICÍPIO DE CÂNDIDO MOTA: UMA APLICAÇÃO DO MODELO SARIMA**

Projeto de pesquisa apresentado ao curso de Análise e Desenvolvimento de Sistemas, do Instituto Municipal de Ensino Superior de Assis – IMESA e à Fundação Educacional do Município de Assis – FEMA, como requisito para a conclusão do projeto de iniciação científica.

Orientado: André Barbosa Alves dos Santos

Orientador: Dr. Almir Rogério Camolesi

Linha de pesquisa: Ciências Exatas e da Terra

**Assis/SP
2024**

ABSTRACT

Knowing the importance of agribusiness for society and also the importance of marketing for agribusiness, this study aims to measure the efficiency of SARIMA¹ as a predictive model in the historical series of the soybean base differential in the municipality of Cândido Mota - SP. As a result, it was found that, although the series presented satisfactory statistical characteristics for the application of the model, it was not sufficiently capable of accurately predicting the behavior of the base differential, suggesting that a more robust model is needed for this purpose.

Keywords: agribusiness, basis, forecasting, SARIMA.

¹ Seasonal Autoregressive Integrated Moving Average.

RESUMO

Sabendo da importância do agronegócio para a sociedade e também da importância da comercialização para o agronegócio, o presente trabalho objetiva mensurar a eficiência do SARIMA² como modelo preditivo diante da série histórica do diferencial de base da soja do município de Cândido Mota – SP. Como resultado, verificou-se que, apesar da série apresentar características estatísticas satisfatórias para a aplicação do modelo, este não foi suficientemente capaz de prever com precisão o comportamento do diferencial de base, dando a entender que a série demanda um modelo mais robusto para este fim.

Palavras-chave: agronegócio, diferencial de base, análise preditiva, SARIMA.

² Modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade.

SUMÁRIO

INTRODUÇÃO.....	6
1. REVISÃO DA LITERATURA.....	8
1.1. DIFERENCIAL DE BASE	8
1.2. ANÁLISE PREDITIVA DE SÉRIES TEMPORAIS	9
1.3. PROCESSO ESTOCÁSTICO ESTACIONÁRIO	10
1.4. PROCESSO AUTORREGRESSIVO	11
1.5. PROCESSO DE MÉDIA MÓVEL	11
1.6. PROCESSO AUTORREGRESSIVO DE MÉDIAS MÓVEIS	12
1.7. PROCESSO AUTORREGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS.....	12
2. HIPÓTESE.....	13
3. OBJETIVOS	13
3.1. OBJETIVO GERAL.....	13
3.2. OBJETIVOS ESPECÍFICOS	13
4. METODOLOGIA.....	13
4.1. COLETA DOS DADOS.....	13
4.2. TRANSFORMAÇÃO DOS DADOS	13
4.3. ANÁLISE DOS DADOS E RESULTADOS	13
5. CONCLUSÃO.....	19
BIBLIOGRAFIA	19

INTRODUÇÃO

Segundo Mazoyer e Roudart (2010), desde sua origem, a agricultura se tornou o principal fator de transformação da *ecosfera*³, condicionando o aumento da população e o desenvolvimento social. Ao longo dos anos, as revoluções agrícolas proporcionaram o desenvolvimento de outras atividades humanas, tais como o comércio, a logística e a indústria, e estes, por sua vez, forneceram à agricultura novos meios de produção, criando uma complexa cadeia produtiva por trás dessa atividade, denominada hoje como agronegócio (ARAÚJO, 2017).

O termo agronegócio⁴ surgiu em 1957 no trabalho intitulado *A Concept Of Agribusiness*, escrito por John H. Davis e Ray A. Goldberg. Nele, os autores dissertam sobre a transformação ocorrida na cadeia produtiva da agricultura ao longo do tempo e sobre a necessidade de olhá-la através de uma nova ótica que não dissocie a agricultura da indústria e do setor de serviços, mas sim integre-os.

É a partir dessa ótica sugerida pelos autores que é possível entender a grandeza desse setor e sua importância como motor econômico gerador de renda, haja vista que parte da cadeia produtiva de alimentos, da indústria mecânica, da indústria química e do setor de serviços, de alguma forma, estão conectados a ele, principalmente em países exportadores de commodities agrícolas, como é o caso do Brasil, que em 2019 teve 21,4% de seu PIB gerado pelo agronegócio (CEPEA, 2020).

Outro importante aspecto do agronegócio para o país é sua participação no acúmulo de *divisas*⁵ e no *equilíbrio cambial*⁶. Como mostra a Figura 1, o superávit do setor na balança comercial entre 1997 e 2019 foi crescente, salvo alguns poucos anos. Portanto, é possível afirmar que nesse período o agronegócio foi responsável por parte significativa das divisas que entraram no país, tendo, conseqüentemente, um papel fundamental no controle cambial deste.

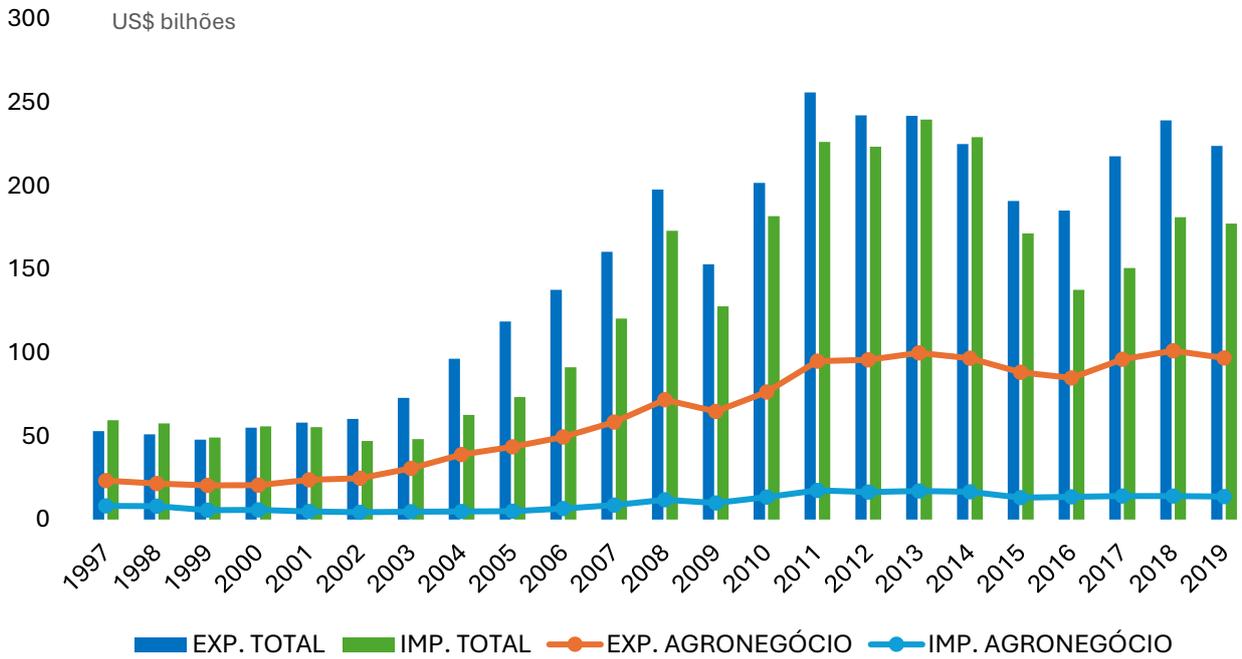
³ No contexto da ecologia, a *ecosfera* é definida como o conjunto de todos os ecossistemas da terra.

⁴ Do inglês *agrobusiness*, que segundo Davis e Goldberg (1957), por definição, significa a soma total de todas as operações envolvidas na fabricação e distribuição de suprimentos agrícolas, operações de produção na fazenda, armazenagem, processamento e distribuição de commodities agrícolas e itens feitos a partir destes.

⁵ Divisas são moedas que estão presentes em um país diferente de sua origem, como por exemplo, o dólar presente no Brasil.

⁶ O Brasil opta por fazer uso de um regime cambial chamado câmbio flutuante, onde a taxa de uma determinada moeda é definida pela oferta e demanda da moeda estrangeira no mercado, sem que haja controle por parte da autoridade monetária. Ou seja, nesse sistema, quanto maior a oferta de uma moeda estrangeira em um determinado país, menor será seu valor em relação à moeda local, e o contrário também é verdade.

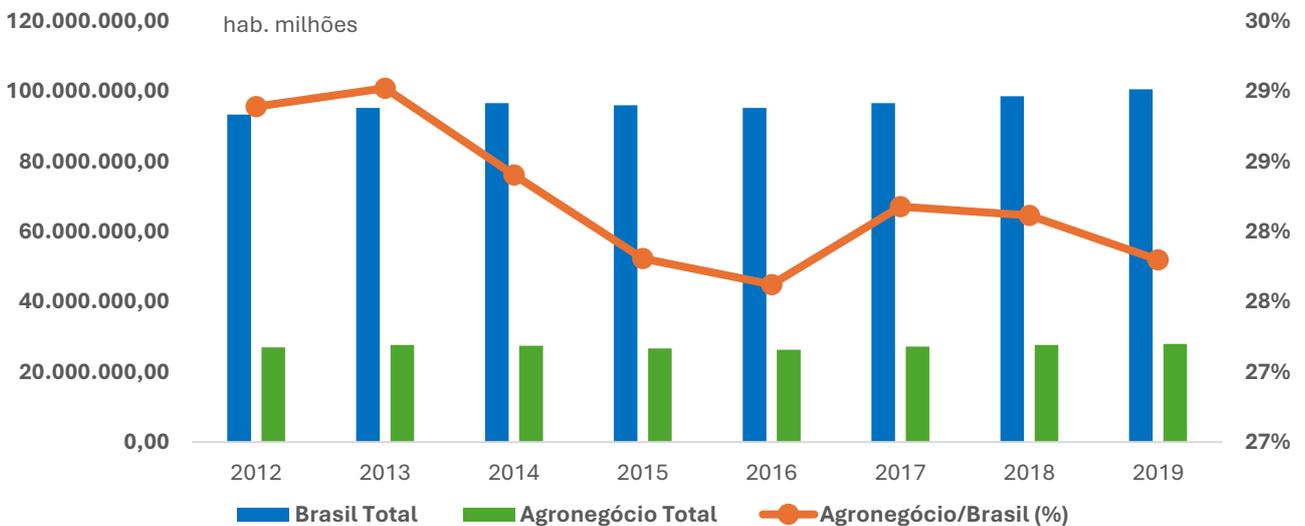
Figura 1 – Evolução anual da balança comercial brasileira e do agronegócio – 1997 a 2019 (em US\$ bi)



Fonte: Elaboração DCNC / SCRI / MAPA - Agrostat Brasil, a partir de dados da SECEX/MDIC

A representatividade do agronegócio no mercado de trabalho também é algo notório. Entre os anos de 2012 e 2019, um pouco mais de ¼ (um quarto) do total de ocupados no país, em média, estavam empregados em atividades relacionadas ao agronegócio, como mostra a Figura 2.

Figura 2 – Total de ocupados no agronegócio em relação ao total do Brasil – 2012 a 2019 (em hab. mi)



Fonte: Elaboração própria / SANTOS, 2024 – a partir de dados do CEPEA (2023)

Entre todas as atividades que compõem a cadeia produtiva do agronegócio, uma em particular interessa a esse trabalho: a comercialização. Ao longo dos anos, a comercialização agrícola sofreu incontáveis transformações até chegar ao modelo atual, onde grandes negócios são articulados por *tradings companies*⁷ e onde os preços das commodities são referenciados por bolsas de mercadorias. Essas transformações trouxeram à comercialização agrícola e também à indústria desse setor maior segurança jurídica e financeira, além de padronizar a qualidade do produto agrícola.

Mais recentemente, com o surgimento das TICs (Tecnologias de Informação e Comunicação), grande parte dos processos da cadeia produtiva do agronegócio passaram a depender dessas tecnologias em certo grau, principalmente aqueles que lidam com a geração, armazenagem e análise de dados, como é o caso da comercialização. Para estes processos, uma área das TICs em particular se tornou fundamental: a ciência de dados.

Segundo Morettin e Singer (2023), a ciência de dados consiste principalmente na aplicação de técnicas estatísticas à problemas que exigem grande capacidade computacional. Assim, tratar, analisar e interpretar dados significa desenvolver tecnologias que suportem as tomadas de decisões dos agentes e que auxiliem o crescimento do respectivo setor.

Nesse sentido, este trabalho pretende contribuir com a aplicação da ciência de dados no setor do agronegócio e com a informatização deste, em particular a comercialização, através da modelagem da série histórica do diferencial de base da soja do município de Cândido Mota – SP e da aplicação na série temporal do modelo autorregressivo integrado de médias móveis com sazonalidade (SARIMA), com o objetivo de prever o comportamento deste indicador.

1. REVISÃO DA LITERATURA

1.1. DIFERENCIAL DE BASE

O diferencial de base, ou simplesmente base, é um tema estudado já há algum tempo pela ciência econômica. Um dos primeiros autores que se dedicou a estudar o comportamento desse indicador foi John Maynard Keynes, em seu trabalho *Treatise on Money* (1930), onde este desenvolve sua teoria sobre o que ele chama de *Normal Backwardation*⁸. Outra teoria

⁷ Empresas comerciais exportadoras.

⁸ Em síntese, essa teoria ressalta o risco financeiro oferecido pela necessidade de se armazenar estoques de um período a outro. Para o autor, o preço normal de fornecimento à vista inclui uma remuneração pelo risco de flutuação dos preços durante o período de produção, enquanto o preço a prazo exclui essa remuneração (SILVEIRA, 2017).

bastante conhecida e amplamente usada em trabalhos acadêmicos como referência é a *Teoria da Estocagem*⁹, proposta inicialmente por Kaldor (1939) e Working (1930).

Como o intuito do presente trabalho não é verificar se o comportamento do diferencial de base da soja do município de Cândido Mota – SP está em conformidade com as teorias econômicas, e sim encontrar o modelo autorregressivo que preveja com maior precisão o comportamento deste indicador, será adotada uma definição mais prática para este, como a exposta no trabalho de Tonin e Tonin (2013).

Segundo os autores, o diferencial de base é a diferença entre o preço *spot*¹⁰ de uma commodity em uma determinada praça de comercialização e o preço dessa mesma commodity negociado no mercado futuro (TONIN e TONIN, 2013 *apud* LEUTHOLD, JUNKUS e CORIDIER, 1989).

Ainda neste trabalho, Tonin e Tonin (2013) *apud* Purcell e Koontz (1999) diz que o comportamento do diferencial de base está associado aos custos presentes na cadeia produtiva da commodity, como logística, capacidade de estocagem, oscilações na oferta e demanda do produto e de seus substitutos nas respectivas praças de comercialização.

1.2. ANÁLISE PREDITIVA DE SÉRIES TEMPORAIS

Segundo Dias (2023) *apud* Marquesone (2018), a análise preditiva é aquela cujo o objetivo é visualizar eventos futuros a partir da compreensão do passado. Um modelo de análise preditiva nada mais é que uma função matemática que relaciona as variáveis observadas a uma variável não observada, e pode ser usado para prever o valor da variável não observada em novos dados.

Variáveis observadas em análises preditivas que tem como objetivo investigar dados financeiros, comumente advêm de uma *série temporal*¹¹, e estas, por sua vez, são estimadas majoritariamente por processos autorregressivos. A explicação do processo autorregressivo utilizado neste trabalho será apresentada nas seções seguintes, e foi fundamentada na obra de Domodar N. Gujarati e Down C. Porter intitulada *Econometria*

⁹ Esta teoria defende que o diferencial de base pode ser explicado através dos custos de manutenção intertemporal do produto estocável, como por exemplo, o custo de oportunidade do capital empregado no processo, custos de estocagem e o rendimento de conveniência (SILVEIRA, 2017).

¹⁰ Preço de uma determinada commodity no mercado à vista.

¹¹ É um conjunto de observações de uma determinada variável feita em períodos sucessivos de tempo ao longo de um determinado intervalo (FARIA et al., 2008).

Básica, 5ª edição, publicada em 2011. Porém, antes é preciso entender o conceito de processo estocástico estacionário ou simplesmente estacionariedade.

1.3. PROCESSO ESTOCÁSTICO ESTACIONÁRIO

Segundo Gujarati e Porter (2011), um processo estocástico¹² será chamado de estacionário se sua média e variância forem constantes ao longo do tempo, e se o valor da covariância entre dois períodos de tempo dependa da distância, do intervalo ou da defasagem entre os dois períodos, e não do tempo ao qual a covariância é computada.

- Média: $E(Y_t) = \mu$
- Variância: $var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$
- Covariância: $\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$

onde γ_k a covariância (ou autocovariância) na defasagem k , é a covariância entre os valores de Y_t e Y_{t+k} , isto é, entre dois valores de Y separados por k . Se $k = 0$, obtemos γ_0 , que é simplesmente a variância de Y (σ^2); se $k = 1$, γ_1 é a covariância entre os dois valores adjacentes de Y .

Por exemplo, suponha que haja uma mudança na origem de Y de Y_t para Y_{t+m} . Se Y_t for estacionário, a média, a variância e autocovariância de Y_{t+m} deverão ser iguais àsquelas de Y_t , ou seja, se uma série temporal for estacionária, essas características estatísticas - em variadas defasagens - permanecerão as mesmas, não importando em que ponto esta seja mensurada: isto é, invariantes no tempo.

De forma prática, tal série temporal tenderá a retornar para a sua média, e flutuações em torno desta média – mensurada por sua variância – terão, de modo geral, uma amplitude constante. Em outras palavras, um processo estacionário não se desviará muito de seu valor médio em virtude da variância finita.

Já uma série temporal não estacionária terá uma média que varia com o tempo, ou uma variância que varia com o tempo, ou, ainda, ambas. Desta forma, é possível estudar seu comportamento apenas pelo período de tempo em consideração, sendo cada conjunto de

¹² Um processo estocástico é uma coleção de variáveis aleatórias ordenadas no tempo. O valor do diferencial de base, por exemplo, é a realização de um processo estocástico do diferencial de base, assim como uma amostra é um corte transversal de uma população. Portanto, da mesma maneira que se utiliza amostras de dados para extrair inferência sobre a população, se utiliza realizações para extrair inferências sobre o processo estocástico.

dados específicos para cada episódio, não sendo possível, como consequência, generalizá-lo para outros períodos, tendo pouco valor prático para o propósito de previsão.

1.4. PROCESSO AUTORREGRESSIVO (AR)

Ao contrário dos modelos de regressão, onde o Y_t é explicado pelos regressores $k X_1, X_2, \dots, X_k$, modelos autorregressivos permitem que Y_t seja explicado pelos valores passados, ou defasados, do próprio Y e dos termos de erro estocástico. Se Y_t for modelado como

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t$$

em que δ é a média de Y e u_t é um erro aleatório não correlacionado, com média zero e variância constante σ^2 (trata-se de um *ruído branco*¹³), então é dito que Y_t segue um processo **autorregressivo estocástico de primeira ordem** ou **AR(1)**.

Aqui o valor de Y no período t depende do seu valor no período anterior e de um termo aleatório; os valores de Y são expressos como desvios baseados em um valor médio. Mas, se o modelo considerado for

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t$$

então é dito que Y_t segue um processo autorregressivo de segunda ordem, ou AR(2). O valor de Y no período t depende do seu valor nos dois períodos prévios, sendo os valores de Y expressos em torno dos seus valores médios δ . Em geral, pode-se ter

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t$$

no caso Y_t é um processo autorregressivo de ordem p -ésima, ou AR(p). Perceba que nos exemplos anteriores não há outros regressores $k X_1, X_2, \dots, X_k$, apenas o valor atual e os valores anteriores, indicando que *os dados falam por si mesmos*.

1.5. PROCESSO DE MÉDIA MÓVEL (MA)

Suponha que Y seja modelado como se segue:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1}$$

em que μ é uma constante e u , como antes, é um termo de erro estocástico de ruído branco. Aqui Y no período t é igual a uma constante mais uma média móvel dos termos de erro

¹³ É chamado de **processo estocástico puramente aleatório** ou **ruído branco** se o processo tem média zero, variância constante σ^2 e é serialmente não correlacionado.

atuais e passados. Neste caso, é dito que Y segue um processo de média móvel de primeira ordem, ou um MA(1). Mas, se Y segue a expressão

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}$$

então é um processo MA(2). De forma mais geral,

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q}$$

é um processo **MA(q)**. Ou seja, um processo de média móvel é uma combinação linear de termos de erro de ruído branco.

1.6. PROCESSO AUTORREGRESSIVO DE MÉDIAS MÓVEIS (ARMA)

É muito provável que Y possua características tanto de AR quanto de MA e seja, portanto, **ARMA**. Então, Y_t segue um processo ARMA(1,1) se pode ser escrito como

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1}$$

porque há um termo autorregressivo e um termo de média móvel. Na equação acima, θ representa um termo constante. Em geral, em um processo **ARMA (p, q)**, haverá termos autorregressivos p e termos de média móvel q .

1.7. PROCESSO AUTORREGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS (ARIMA)

Sabe-se que, majoritariamente, séries temporais são não estacionárias, e sabe-se também que quando a série temporal é não estacionária, é necessário torna-la estacionária. Uma das técnicas para tornar uma série não estacionária em uma série estacionária é a diferenciação. É dado o nome de *integrada* à série que passa por um processo de diferenciação a fim de tornar-se estacionária. Em geral, se uma série temporal é $I(d)$, depois de diferenciá-la d vezes, obtêm-se uma série $I(0)$.

Se uma série temporal precisa passar por um processo de diferenciação d vezes para torna-se estacionária e aplicar o modelo ARMA (p, q), é dito que a série temporal original é **ARIMA (p, d, q)**, ou seja, ela é uma série temporal **autorregressiva integrada de médias móveis**, em que p denota os números dos termos autorregressivos, d o número de vezes que a série deve ser diferenciada antes de tornar-se estacionária e q o número de termos de média móvel. Portanto, dados os valores de p, d e q , é possível dizer qual processo está sendo modelado.

2. HIPÓTESE

O presente trabalho considera a possibilidade de gerar dados preditivos sobre o comportamento do diferencial de base da soja do município de Cândido Mota – SP, com baixa discrepância em relação aos dados reais, através de modelos autorregressivos univariados¹⁴.

3. OBJETIVOS

3.1. OBJETIVO GERAL

O objetivo geral deste trabalho é verificar a eficiência do modelo autorregressivo integrado de médias móveis com sazonalidade diante da série histórica do diferencial de base da soja do município de Cândido Mota – SP.

3.2. OBJETIVOS ESPECÍFICOS

Os objetivos específicos deste trabalho são:

- Coletar, tratar e modelar a série histórica do diferencial de base da soja do município de Cândido Mota – SP.
- Encontrar os melhores parâmetros para o modelo autorregressivo integrado de médias móveis com sazonalidade que, após testado e validado, possa servir como ferramenta de análise de dados da mesma natureza, ou então que colabore com o desenvolvimento de outros estudos que tenham como tema de investigação o mesmo deste.

4. METODOLOGIA

A metodologia utilizada na pesquisa fundamentou-se naquela comumente utilizada em análises de séries temporais, e foi dividida entre as etapas que serão apresentadas nas seções seguintes.

4.1. COLETA DOS DADOS

Os valores que formam a amostra utilizada nessa pesquisa foram coletados em diferentes repositórios, e o dia da semana definido como padrão para a coleta foi a quarta-feira. Nas quartas-feiras que não houveram publicações de valores, estes foram supridos por aqueles válidos com data mais próxima, no entanto, quando isto também não foi possível, o método

¹⁴ Esse tipo de modelo pressupõe que as características da série de interesse são explicadas exclusivamente a partir do comportamento da própria série, ou seja, *os dados falam por si mesmos*.

escolhido para o preenchimento dos valores ausentes foi a média aritmética simples entre o valor da data anterior e o valor da data posterior à data com valor ausente¹⁵.

- Preço regional da soja: estes dados se referem ao preço da soja no município de Cândido Mota – SP, e parte deles foram coletados na *Cooperativa Agroindustrial de Cândido Mota (COOPERMOTA)* e parte no periódico *O Diário do Vale*.
- Preço internacional da soja: estes se referem ao preço da soja na *Chicago Board of Trade*, e foram coletados na plataforma *Investing*.
- Preço do dólar em relação ao real: como o preço internacional da soja é cotado em dólar, foi necessário também coletar do *Instituto de Pesquisa Econômica Aplicada (IPEA)*, o preço do dólar em relação ao real a fim de parear os valores regionais aos internacionais.

4.2. TRANSFORMAÇÃO DOS DADOS

Após a coleta, os dados foram armazenados em um arquivo *Excel*, onde passaram por um processo de transformação que basicamente calculou o diferencial de base da soja para todas as quartas-feiras desde janeiro de 2000 até a segunda semana de dezembro de 2024. Para a realização do cálculo, foi usada a seguinte fórmula:

$$\text{Diferencial de Base}^{16} = ((\text{Preço Internacional}) - (\text{Preço Regional} \div \text{US\$} \div 2,2046)) \times 100$$

onde 2,2046 representa o fator de conversão¹⁷ entre um bushel e uma saca de soja, e o US\$ representa o valor do dólar frente ao real. Em seguida, essa base de dados foi importada para o *Google Colaboratory*, onde através da linguagem *Python* foi possível realizar o tratamento e a análise dos dados.

4.3. ANÁLISE DOS DADOS E RESULTADOS

Antes de aplicar o modelo autorregressivo integrado de médias móveis na série, verificou se esta possui características de uma série estacionária. Para isso, foi aplicado o teste Dickey-Fuller Aumentado (ADF), que identifica a presença ou não de raiz unitária na série. De acordo com o teste, o valor-p foi de 0.00000000000002%, consideravelmente abaixo do

¹⁵ De maneira geral, qualquer método escolhido teria, na prática, o mesmo efeito sobre o resultado da pesquisa. Isso porque foram coletados dados de 1301 dias, sendo que destes, apenas 10 valores, e não dias, estavam ausentes.

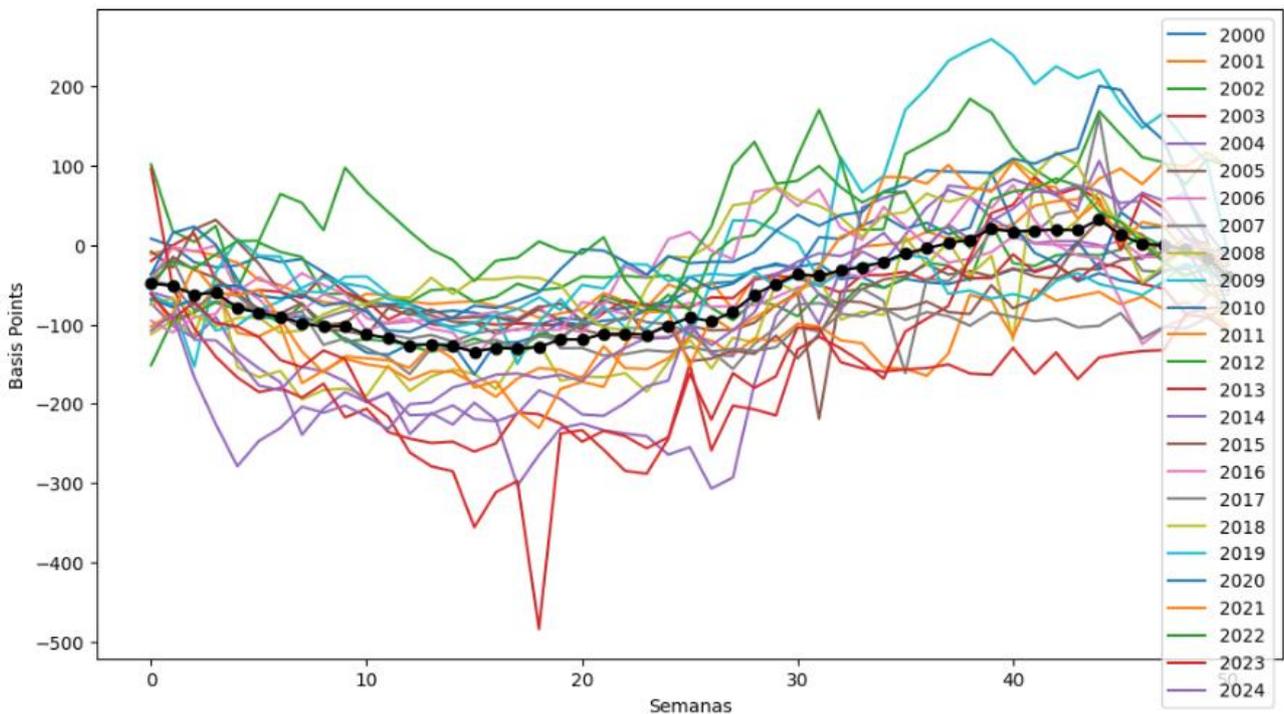
¹⁶ O cálculo resultará em valores em pontos por bushel, também chamado de pontos base (do inglês *basis points*), que nada mais é que o valor em centavos por bushel multiplicado por 100.

¹⁷ Os contratos de soja negociados na *Chicago Board of Trade* possuem outra unidade de medida, o bushel, que é uma unidade de medida de volume, e no caso da soja, equivale à 60 libras-peso, ou 27,2152 kg, ou seja, em uma saca de 60 quilos de soja cabem aproximadamente 2,2046 bushels.

limite estipulado pelo nível de significância de 5%, sugerindo que a série possui características de uma série estacionária.

Outra importante característica da série que também foi identificada foi a presença de sazonalidade. Para isso, foram plotados em um gráfico dados sobre o comportamento do diferencial de base ao longo do ano, de todos os anos contemplados na série, como mostra a Figura 3.

Figura 3 – Comportamento do diferencial de base ao longo do ano (2000 até 11-12-2024)



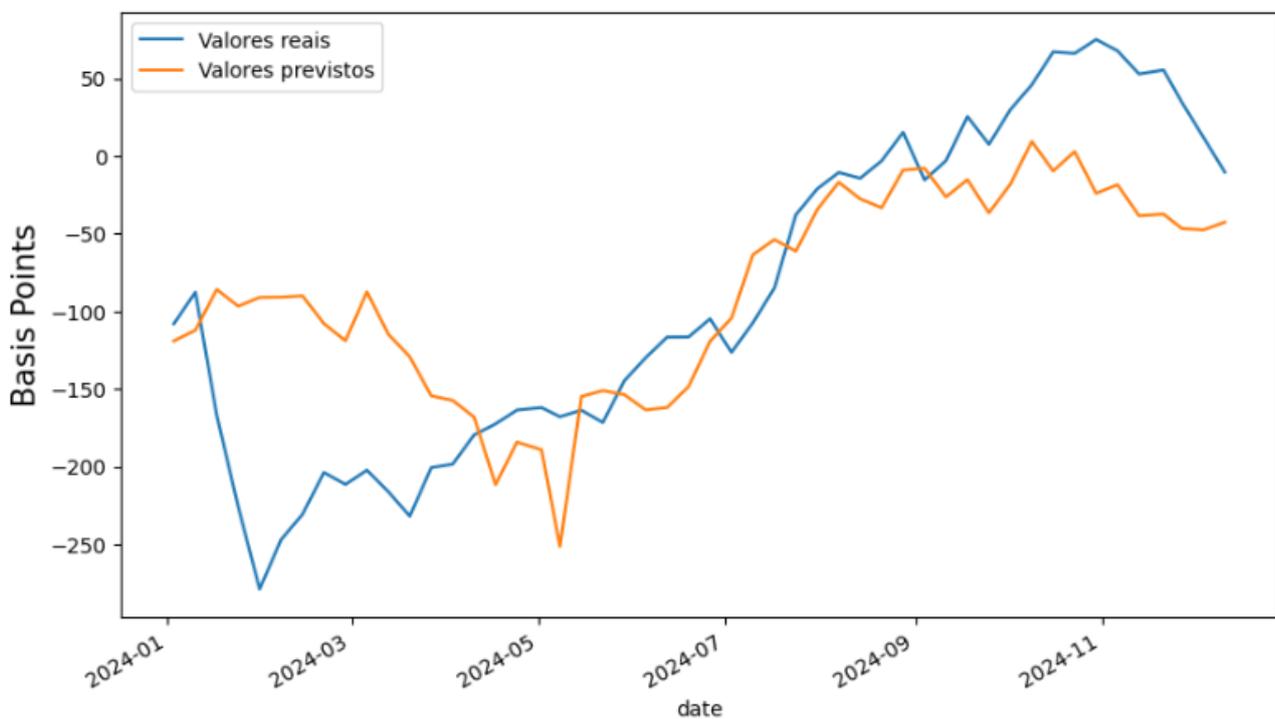
Fonte: Elaboração própria / SANTOS, 2024 – a partir de dados da COOPERMOTA, CBOT e IPEA

De forma intuitiva, é possível observar o comportamento sazonal mostrado pela linha pontilhada preta, que representa a média entre os valores de todos os anos em cada período ao longo do ano. Percebe-se através dessa linha que, na média, o diferencial de base se mantém enfraquecido durante o primeiro semestre do ano, quando ocorre a colheita da soja no país, e passa a se fortalecer a partir do segundo semestre, durante o período chamado de entressafra.

Dito isso, o próximo passo foi submeter a série à função *autoARIMA* da biblioteca *pmdarima* da linguagem *Python*. Essa função conduz testes de diferenciação para determinar a ordem d e então realiza ajustes dentro dos intervalos p e q definidos. Pelo fato de a série ter apresentado presença de sazonalidade, optou-se por habilitar parte da função que busca

identificar os parâmetros ótimos P e Q e a diferenciação ótima D , ou seja, os parâmetros sazonais do modelo SARIMA¹⁸. Utilizando o critério de informação de Akaike¹⁹ (AIC), a função determinou que o modelo mais adequado para a série seria o SARIMA(2, 0, 0)(2, 1, 0)[52], onde 52 (semanas) é o número de observações por ciclo sazonal. A série então foi dividida em duas partes, amostra de treino e amostra de teste, onde na primeira foram imputados valores anteriores à 2024, enquanto na segunda foram imputados valores de 2024, gerando os resultados apresentados no gráfico da Figura 4.

Figura 4 – Comportamento previsto do diferencial de base vs comportamento real do diferencial de base



Fonte: Elaboração própria / SANTOS, 2024 – a partir de dados da COOPERMOTA, CBOT e IPEA

De maneira intuitiva, percebe-se que a previsão (linha laranja) não acompanhou de maneira precisa o comportamento real do diferencial de base (linha azul) no período, tampouco teve um comportamento comumente observado em previsões, onde os dados previstos de curto prazo performam melhor que os previstos no médio e longo prazo. Aqui os dados previstos são mais assertivos conforme o horizonte preditivo aumenta, salvo os três últimos meses.

¹⁸ Modelos autorregressivos integrados de médias móveis com sazonalidade são processos que buscam prever valores de uma série temporal com base na mesma combinação de termos do modelo ARIMA, porém considerando a sazonalidade da série, ou seja, agora o modelo é composto por componentes autorregressivos, de integração e de média móvel sazonais (P , Q , D), além das ordens de defasagens dos componentes não sazonais (p , d , q) (SILVA, 2023).

¹⁹ O critério de informação de Akaike é uma métrica da inferência estatística que mensura a qualidade de um modelo priorizando a simplicidade deste.

Com o intuito de obter uma avaliação mais criteriosa, foi aplicado o cálculo do erro absoluto médio (MAE) para o horizonte preditivo de um, três, seis e 12 meses. Os resultados foram, respectivamente: 86.74, 98.79, 64.27 e 55.24 pontos por bushel.

Diante desses resultados e visto que a série apresentou características de uma série estacionária, viu-se necessidade de reamostrar os dados a fim de observar a performance do modelo a partir de recortes distintos da série temporal. O critério adotado para a reamostragem buscou dividir as novas amostras pensando na comparação entre antes e depois do início da pandemia da *covid 19*²⁰ devido ao fato desta ter causado um notório desequilíbrio na cadeia produtiva mundial que refletiu diretamente nos preços das commodities. A Tabela 1 deixa mais claro o critério adotado.

Tabela 1 – Critério de reamostragem

	Amostra de Treino	Amostra de Teste
Pré-pandemia:	2000 - 2015	2016
	2000 - 2016	2017
	2000 - 2017	2018
	2000 - 2018	2019
Pós-pandemia:	2000 - 2019	2020
	2000 - 2020	2021
	2000 - 2021	2022
	2000 - 2022	2023

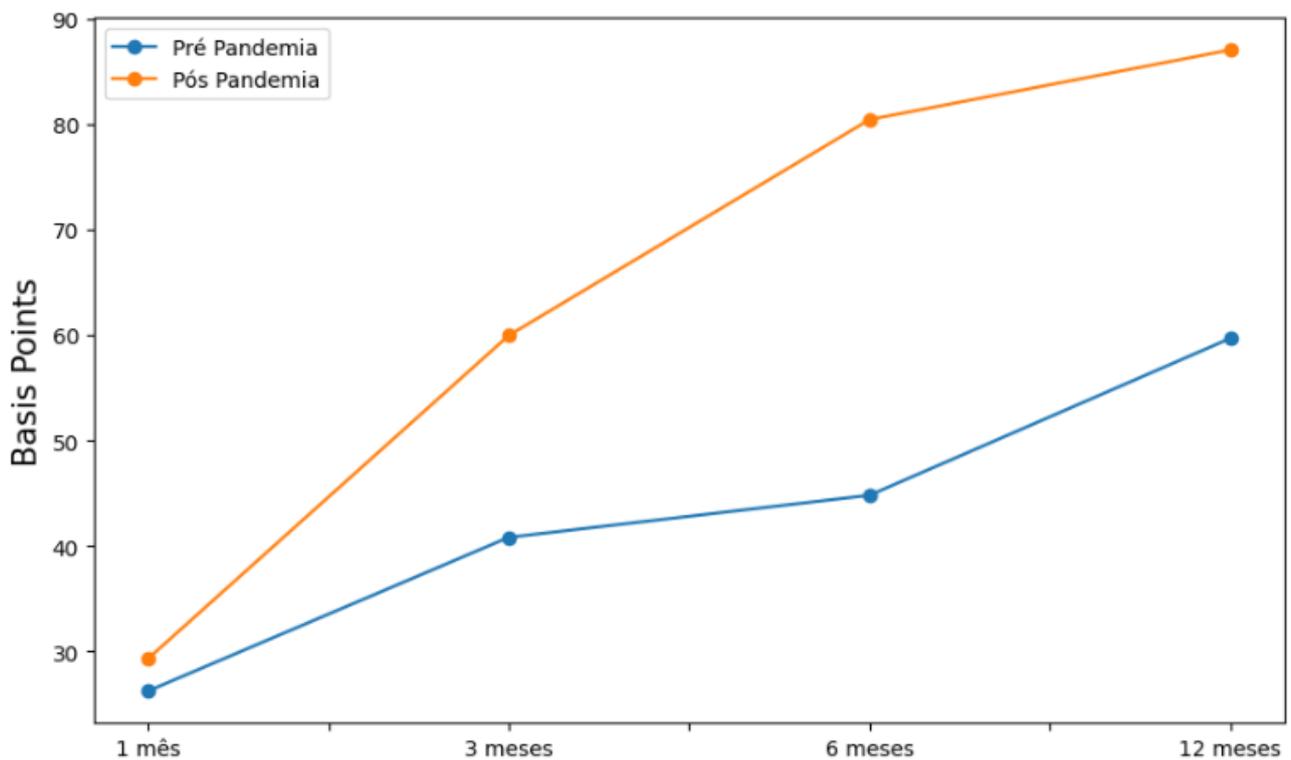
A partir do critério adotado, calculou-se o erro absoluto médio (MAE) para os horizontes preditivos de um, três, seis e doze meses de cada amostra definida na Tabela 1, e partir dos resultados, calculou-se a média entre os dados pré-pandêmicos e pós-pandêmicos. Os resultados podem ser vistos na Tabela 2 e também no gráfico da Figura 5.

²⁰ Segundo a *Organização das Nações Unidas*, a pandemia tem início no dia 11 de março de 2020.

Tabela 2 – Média do erro absoluto médio: pré-pandemia e pós-pandemia

	Pré Pandemia	Pós Pandemia
Um mês:	26.24	29.30
Três meses:	40.80	59.94
Seis meses:	44.79	80.40
Doze meses:	59.67	87.00

Figura 5 – Média do erro absoluto médio: pré-pandemia e pós-pandemia



Fonte: Elaboração própria / SANTOS, 2024 – a partir de dados da COOPERMOTA, CBOT e IPEA

A partir do gráfico da Figura 5, é possível observar que as médias dos erros absolutos médios das previsões pré-pandemia satisfazem melhor os parâmetros aceitos pelo modelo de cálculo utilizado, ou seja, quanto mais próximo de 0 estiver a média, melhor. É possível também observar que o comportamento tanto da média pré-pandemia quanto da média pós-pandemia segue um comportamento comum em previsões, onde os resultados de curto prazo são melhores que os de médio e longo prazo.

5. CONCLUSÃO

Posto tudo isso, entende-se que o modelo sazonal autorregressivo de médias móveis (SARIMA) não foi capaz de prever com tanta precisão o comportamento do diferencial de base da soja do município de Cândido Mota – SP. Portanto, podemos considerar a hipótese de que modelos autorregressivos univariados não são capazes de prever com precisão séries temporais como esta, porém, para que isso seja verdade, é preciso testar outros modelos como estes, como por exemplo, os da família ARCH²¹.

Segundo Gujarati e Porter (2001), séries temporais financeiras frequentemente apresentam o fenômeno da aglomeração por volatilidade, isto é, períodos com grandes oscilações no comportamento seguidos por períodos nos quais há relativa calma. Para a previsão de séries com este tipo de comportamento, modelos autorregressivos de heterocedasticidade²² condicional (ARCH) são os comumente usados, pois são modelos que consideram a existência de autocorrelação entre a heterocedasticidade observada ao longo de diferentes períodos da série.

Portanto, é razoável pensar na viabilidade de uma nova pesquisa de iniciação científica que proponha a aplicação deste modelo para a série histórica do diferencial de base da soja do município de Cândido Mota – SP, visto que o comportamento deste, por se tratar de um indicador formado por dados financeiros, possa sofrer influência da volatilidade.

BIBLIOGRAFIA

ARAUJO, Marcos. **O Segredo do Grão: o comércio de *commodities* agrícolas**. 1º ed. Curitiba: Do Autor, 2017.

CEPEA, 2020. Disponível em: [http://www.cepea.esalq.usp.br/upload/kceditor/files/Cepea_PIB_CNA_2019\(1\).pdf](http://www.cepea.esalq.usp.br/upload/kceditor/files/Cepea_PIB_CNA_2019(1).pdf). Acesso em: 8 jan. 2024.

CEPEA, 2023. **Mercado de Trabalho do Agronegócio**. Disponível em: [https://cepea.esalq.usp.br/upload/kceditor/files/Planilha_Mercado_Trabalho_SITE\(14\).xlsx](https://cepea.esalq.usp.br/upload/kceditor/files/Planilha_Mercado_Trabalho_SITE(14).xlsx). Acesso em: 8 jan. 2024.

DAVIS, J. H.; GOLDBERG, R. A. **A Concept Of Agribusiness**. 1º ed. Boston: Harvard University, 1957.

²¹ *Autoregressive Conditional Heteroskedasticity*.

²² O termo refere-se ao comportamento inconstante da variância ao longo da série.

DIAS, J. C.; DIAS J. C. **Análise Preditivas de Dados: uma abordagem de conceitos e algoritmos**. Revista Processando o Saber, Praia Grande, v.15, p. 63-74, jun. 2023.

FARIA, E. L. *et. al.* **Previsão de séries temporais utilizando métodos estatísticos**. 2008. Centro Brasileiro de Pesquisas Físicas - CBPF, Pontifícia Universidade Católica – PUC. Rio de Janeiro, p. 63, 2008.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5º ed. Porto Alegre: AMGH, 2011.

MAPA, 2020. **Estatística do Agronegócio**. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/relacoes-internacionais/documentos/estatisticas-do-agronegocio/SERIEHISTORICABCARESUMIDA19972019.xls/@@download/file>. Acesso em 8 jan. 2024.

MAZOYER, Marcel; ROUDART, Laurence. **História das Agriculturas no mundo: do neolítico à crise contemporânea**. Universidade Estadual Paulista - UNESP. São Paulo: Editora UNESP, 2010.

MORETTIN, P. A.; SINGER, J. M. **Estatística e Ciência de Dados**. 1º ed. Rio de Janeiro: LTC, 2023.

SILVA, Fernando. **Previsão Econômica com Modelos Arima. Análise Macro**, 2023. Disponível em: < <https://analisemacro.com.br/econometria-e-machine-learning/previsao-economica-com-modelos-arima/>>. Acesso em: 23 de set. de 2024.

SILVEIRA, G. A. **Análise da Teoria da Estocagem sobre a Base dos Contratos Futuros de Soja no Brasil**. 2017. Dissertação (Mestrado) – Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, USP, Ribeirão Preto, 2017.

TONIN, J. M.; TONIN, J. R. **Análise do Comportamento e Previsão da Base para a Soja em Paranaguá com Utilização de Contratos Futuros na CBOT: uma aplicação do modelo SARIMA**. ResearchGate, 2013. Disponível em: <https://researchgate.net/publication/283117119>. Acesso em: 5 jan. 2024.