

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL
FACULDADE DE ENGENHARIAS, ARQUITETURA E URBANISMO E
GEOGRAFIA
CURSO DE ENGENHARIA DE PRODUÇÃO

Felipe Barcelos Blini Duarte

Análise da volatilidade do preço da saca de soja utilizando cadeias de Markov

Novembro de 2025

Felipe Barcelos Blini Duarte

Análise da volatilidade do preço da saca de soja utilizando cadeias de Markov

Trabalho de Conclusão de Curso
apresentado como requisito parcial à
obtenção do título de Bacharel em
Engenharia de Produção da Universidade
Federal de Mato Grosso do Sul.
Orientador(a): Profa. Dra. Kassia Tonheiro
Rodrigues

Novembro de 2025

RESUMO

Este trabalho analisa a dinâmica de preços da soja no Brasil por meio da aplicação de Cadeias de Markov. A soja é a principal mercadoria agrícola brasileira, e a volatilidade de seus preços representa um desafio persistente para produtores e gestores, influenciando decisões de estocagem, hedge e planejamento de safras. Em resposta, a metodologia de Cadeias de Markov, que modela transições probabilísticas entre estados de variação de preço, é empregada para compreender a variabilidade. A pesquisa utiliza dados diários do CEPEA, buscando um equilíbrio entre o detalhe da informação e o rigor metodológico para a tomada de decisão. O estudo justifica-se pela relevância de abordar um problema central para o agronegócio e pela aplicabilidade de um modelo probabilístico para subsidiar estratégias de comercialização e gestão de riscos menor tempo de retorno, indicando que o sistema retorna rapidamente a este estado. Em contraste, os estados de maior variação apresentam os maiores tempos de recorrência, reforçando que os períodos de estabilidade são mais frequentes. Essa análise do Tempo Médio de Recorrência possibilita a quantificação do risco de grandes flutuações de preço e direciona a gestão operacional para a otimização das oportunidades. A contribuição do estudo reside em oferecer instrumentos que permitem identificar janelas mais vantajosas de venda, contribuindo para a maximização do lucro, a redução de riscos e a estabilidade do setor.

Palavras-chave: Soja; Volatilidade de preços; Cadeias de Markov; Agronegócio; Gestão de riscos.

ABSTRACT

This study analyzes the soybean price dynamics in Brazil through the application of Markov Chains. Soybean is the main Brazilian agricultural commodity, and its price volatility represents a persistent challenge for producers and managers, influencing decisions regarding storage, hedging, and crop planning. In response, the Markov Chain methodology, which models probabilistic transitions between price variation states, is employed to understand this variability. The research uses daily data from CEPEA, seeking a balance between information detail and methodological rigor for decision-making. The study is justified by the relevance of addressing a central problem for agribusiness and the applicability of a probabilistic model to support commercialization strategies and risk management. The analysis demonstrated the suitability of Markov Chain modeling for the historical soybean price series. The main results revealed that the stability state has the shortest recurrence time, indicating that the system quickly returns to this state. In contrast, states of greater variation show the longest recurrence times, reinforcing that stability periods are more frequent. This Mean Recurrence Time analysis enables the quantification of the risk of large price fluctuations and directs operational management toward optimizing opportunities. The contribution of the study lies in offering instruments that allow the identification of more advantageous selling windows, contributing to profit maximization, risk reduction, and sector stability.

Keywords: Soybean; Price volatility; Markov Chains; Agribusiness; Risk management.

SUMÁRIO

1. Introdução	5
2. Referencial teórico.....	7
2.1 mercadorias agrícolas e volatilidade de preços	8
2.2 Processos estocásticos	9
2.3 Cadeias de Markov.....	11
2.4 Aplicações de Cadeias de Markov em Preços de <i>Commodities</i>	12
2.5 Ferramentas Computacionais em Modelagem Estocástica.....	15
3. Procedimentos metodológicos.....	16
3.1 Coleta de Dados.....	16
3.2 Cálculo das Variações Percentuais e Definição de Estados	17
3.3 Construção da Matriz de Transição.....	18
3.4 Validação da Hipótese Markoviana e Análise de Ordem.....	19
3.5 Análise de Convergência e Determinação da Distribuição Estacionária	20
3.6 Discussão Metodológica e Premissas	21
4. Resultados	23
4.1 Análise Preliminar dos Dados Coletados	23
4.2 Discretização dos Estados	25
4.3 Verificação da Propriedade Markoviana de Primeira Ordem.....	27
4.4 Matrizes de Contagem e Probabilidade de Transição	28
4.5 Distribuição Estacionária e Análise de Longo Prazo	31
4.6 Discussão e Síntese dos Resultados	34
5. Conclusão	35
Referências	37

1. INTRODUÇÃO

O Brasil produziu mais de 150 milhões de toneladas de soja, na safra 2022/23, mantendo o País na liderança mundial da produção desse grão. O sucesso da sojicultura brasileira é resultado de amplos investimentos em pesquisa, o que permitiu ao Brasil deixar de ser importador na década de 1970 para se tornar o maior produtor e exportador de soja no mundo, respaldado em ciência e inovação" (EMBRAPA, 2023). Essa relevância transcende a produção primária, pois a cadeia da soja e do biodiesel deve alcançar um Produto Interno Bruto (PIB) de R\$ 779,2 bilhões em 2025, com um crescimento projetado de 11,29% (ABIOVE, 2025; CEPEA, 2025). Além do impacto financeiro, o setor é um motor de desenvolvimento social, sendo responsável pela geração de 2,05 milhões de ocupações em toda a sua cadeia produtiva (ABIOVE; CEPEA, 2023, p. 4). Nesse contexto de extrema relevância econômica e social, a volatilidade dos preços da soja emerge como um fator crítico, afetando diretamente a rentabilidade dos produtores, as decisões de investimento e a estabilidade do mercado de *commodities* (IISD, 2024).

Para tanto, compreender a variabilidade dos preços da soja é fundamental para orientar decisões estratégicas de comercialização, especialmente diante da elevada participação da commodity na renda agrícola e no comércio exterior (CONAB, 2025). A magnitude da produção brasileira e sua relevância global tornam o mercado altamente exposto à volatilidade de preços, o que afeta diretamente produtores, exportadores e gestores de agronegócio (IISD, 2024). Estudos de previsão de preços de *commodities* agrícolas são essenciais para reduzir incertezas, orientar estratégias de estocagem e comercialização e apoiar a formulação de políticas de gestão de risco (Porto, 2022). Dessa forma, a análise da variabilidade de preços da soja não se justifica apenas pelo peso econômico, mas também pela necessidade de oferecer instrumentos que permitam identificar janelas mais vantajosas de venda, contribuindo para a maximização do lucro e para a estabilidade do setor (Xu *et al.*, 2025, p. 3). A volatilidade de preços da soja desencadeia uma série de consequências em cadeia, afetando desde o planejamento de safras e a capacidade de investimento dos produtores até a segurança alimentar e a balança comercial do país.

A volatilidade dos preços da soja representa um desafio persistente, impactando diretamente decisões relacionadas à estocagem, ao *hedge*, que consiste em operações

para proteção contra variações de preços, e ao planejamento de safras. Sykora (2013), ao examinar a relação entre o mercado à vista, no qual as transações são liquidadas com a entrega imediata do produto pelo preço corrente, e o mercado futuro, evidenciou a influência de benefícios da disponibilidade imediata do produto na formação de preços futuros, ilustrando a complexidade estocástica da commodity.

Em resposta a essa demanda por abordagens mais sofisticadas, a literatura internacional reforça a pertinência de metodologias estocásticas em contextos agrícolas. A complexidade inerente ao setor, marcada pela dependência espaço-temporal e pela influência de eventos climáticos extremos, exige modelos capazes de fornecer uma quantificação robusta da incerteza, especialmente em previsões, como apontado por Michaelides *et al.* (2025). Essa necessidade é corroborada pela afirmação de que modelos de previsão espaço-temporal têm sido amplamente aplicados no setor agrícola, especialmente para a estimativa da produtividade de culturas e a previsão de preços de commodities, em que a variabilidade sazonal e espacial afeta significativamente os resultados, de acordo com Mello *et al.* (2024). Dentro desse contexto de volatilidade, modelos estocásticos, como as Cadeias de Markov, são amplamente utilizados para capturar a dinâmica de preços e auxiliar na tomada de decisão.

Nessa mesma linha, a aplicação de Cadeias de Markov na análise de preços da soja já foi explorada em estudos anteriores, como o de Cechin e Corso (2017), que utilizou dados diários do Centro de Estudos Avançados em Economia Aplicada (CEPEA), demonstrando a viabilidade do método. Diante do exposto, o objetivo deste trabalho é analisar a dinâmica de preços da soja no Brasil, visando subsidiar a tomada de decisão estratégica e a gestão de riscos na sua comercialização, por meio da aplicação de Cadeias de Markov. A pesquisa utiliza uma frequência diária de dados, buscando um equilíbrio entre o detalhe da informação e a robustez estatística para a tomada de decisão.

Dessa forma, este trabalho justifica-se por sua relevância, aplicabilidade e viabilidade. Sua relevância é confirmada ao abordar um problema central para o agronegócio brasileiro, a volatilidade dos preços da soja, utilizando dados atualizados que refletem os cenários de mercado mais recentes. A aplicabilidade do estudo manifesta-se na proposta de um modelo probabilístico que pode ser diretamente utilizado por produtores e gestores como subsídio para estratégias de comercialização e mitigação de riscos, alinhando-se à engenharia de produção ao fornecer ferramentas de tomada de decisão

operacional e estratégica no agronegócio por meio da modelagem de Cadeias de Markov e da análise de volatilidade. Por fim, sua viabilidade é assegurada pelo uso de uma metodologia consolidada e de dados públicos do (CEPEA), o que garante a replicabilidade do modelo e o posiciona como uma ferramenta prática e acessível para o setor.

1.1 OBJETIVOS

O objetivo geral deste trabalho é analisar a dinâmica de preços da soja no Brasil, visando subsidiar a tomada de decisão estratégica e a gestão de riscos na sua comercialização, por meio da aplicação de Cadeias de Markov.

1.1.2 OBJETIVOS ESPECÍFICOS

- a) Caracterizar a volatilidade histórica dos preços da soja no mercado brasileiro, utilizando dados diários.
- b) Modelar a série temporal de preços da soja por meio da aplicação de Cadeias de Markov, identificando os estados de transição de preços.
- c) Estimar as probabilidades de transição entre os diferentes estados de preço (alta, baixa e estabilidade).
- d) Fornecer um modelo probabilístico que sirva de subsídio para a tomada de decisão estratégica e a gestão de riscos na comercialização da soja.

2. REFERENCIAL TEÓRICO

Esta seção tem como objetivo estabelecer as bases conceituais e metodológicas para a análise da dinâmica de preços da soja por meio de Cadeias de Markov. A estrutura argumentativa parte da contextualização do problema, abordando as características das *commodities* agrícolas e a natureza de sua volatilidade de preços (seção 2.1). Em seguida, a revisão avança para o arcabouço teórico necessário para modelar tal fenômeno, introduzindo os processos estocásticos (seção 2.2) como a classe de modelos adequada para sistemas com incerteza. Dentre eles, aprofunda-se nas Cadeias de Markov (seção 2.3), detalhando suas propriedades e métricas fundamentais que serão empregadas na

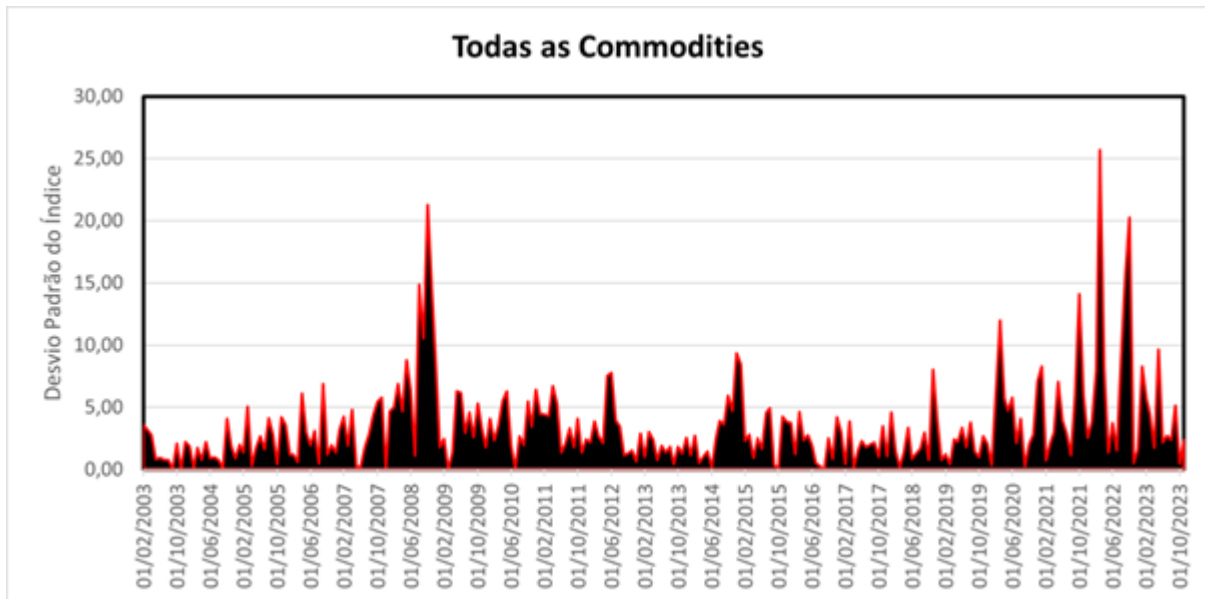
análise. Para validar a pertinência do método, a seção 2.4 explora aplicações práticas dessa ferramenta na precificação de *commodities*, consolidando sua relevância para o tema. Por fim, a seção 2.5 discute as ferramentas computacionais que viabilizam a implementação do modelo proposto. Coletivamente, este capítulo fundamenta a escolha metodológica e contextualiza a contribuição do trabalho frente à literatura existente.

2.1 MERCADORIAS AGRÍCOLAS E VOLATILIDADE DE PREÇOS

As *commodities* agrícolas são produtos primários, padronizados, produzidos em grande escala, comercializados globalmente e com preços influenciados por fatores climáticos, econômicos e especulativos (Agroadvance, 2024). Esses ativos, incluindo grãos como a soja, são caracterizados por uma volatilidade inerente, decorrente de incertezas climáticas, variações na produção e intervenções políticas, que amplificam as oscilações de preços e impactam a estabilidade econômica de produtores e consumidores (Embrapa, 2017). A literatura econômica enfatiza que os preços de diferentes *commodities* frequentemente apresentam uma tendência de se moverem em conjunto de forma excessiva, para além do que fundamentos macroeconômicos justificariam, atribuindo isso a fatores como expectativas de inflação ou comportamentos de rebanho entre investidores (Pindyck; Rotemberg, 1990). Esse padrão de movimentação conjunta é particularmente evidente em mercados agrícolas, onde choques comuns, como variações no petróleo ou na moeda, propagam efeitos cruzados entre *commodities* não relacionadas diretamente, gerando padrões de volatilidade que desafiam modelos tradicionais de previsão (Bini *et al.*, 2015).

A magnitude dessa volatilidade no mercado de commodities é ilustrada na Figura 1, que apresenta o desvio padrão do índice de preços de todas as commodities ao longo de um período de duas décadas.

Figura 1 - Desvio Padrão do Índice de Commodities (2003-2023)



Fonte: FMI (2023)

A volatilidade de preços em mercadorias agrícolas, definida como a dispersão das variações nos retornos ao longo do tempo, constitui um risco central para o setor, influenciando decisões de plantio, estocagem e comercialização. Estudos empíricos demonstram que essa volatilidade é exacerbada em mercados de grãos, como o de soja, devido à sazonalidade das safras e à dependência de condições meteorológicas, com evidências de repercussões de volatilidade de mercados de índices para contratos futuros agrícolas (Balg; Metcalf, 2010). No contexto da soja, que se destaca por sua relevância como fonte proteica e por seu papel em cadeias de valor globais, a volatilidade reflete não apenas dinâmicas locais, mas também interações internacionais, onde flutuações em mercados líderes propagam impactos sobre preços spot e futuros (Sykora, 2013 p. 6). Essa instabilidade demanda análises probabilísticas para capturar transições entre estados de preço, permitindo identificar padrões que suportem estratégias de mitigação de risco em ambientes de alta incerteza.

2.2 PROCESSOS ESTOCÁSTICOS

Conforme Karlin e Taylor (1975) os processos estocásticos configuram-se como sequências de variáveis aleatórias indexadas pelo tempo, representando sistemas

evolutivos onde o estado futuro é determinado por probabilidades condicionais, em contraste com modelos determinísticos, nos quais o comportamento é inteiramente previsível a partir de condições iniciais. Em um processo determinístico, o trajeto é fixo e reproduzível, ao passo que o estocástico incorpora elementos de aleatoriedade, permitindo modelar fenômenos com incertezas inerentes, como flutuações econômicas ou variações climáticas (Ross, 1995). Essa distinção é crucial para a análise de séries temporais, onde processos estocásticos capturam dependências probabilísticas que refletem a imprevisibilidade real de sistemas complexos, fornecendo ferramentas para inferir distribuições de probabilidade sobre trajetórias futuras (Box; Jenkins; Reinsel, 2008).

A formalização de processos estocásticos enfatiza a estrutura probabilística subjacente, com definições que abrangem espaços de estados discretos ou contínuos e funções de transição que governam a evolução temporal. De acordo com abordagens clássicas, um processo estocástico é caracterizado por uma família de variáveis aleatórias $\{X_t, t \in T\}$, onde T denota o conjunto de índices temporais, e as propriedades de independência ou dependência condicional definem subclasses específicas, como os processos markovianos (Karlin e Taylor, (1975). Essa estrutura conceitual sustenta aplicações em diversos domínios, permitindo a modelagem de dinâmicas aleatórias com rigor matemático e facilitando a derivação de propriedades como estacionariedade e ergodicidade, essenciais para previsões de longo prazo em contextos voláteis.

Na prática, a análise de séries temporais por métodos estocásticos recorre a diferentes classes de modelos paramétricos. Os modelos autorregressivos integrados de médias móveis (ARIMA) descrevem a dependência linear entre observações sucessivas e são sistematizados na abordagem Box-Jenkins para identificação, estimação e diagnóstico de modelos (Box; Jenkins; Reinsel, 2008). Para representar a dinâmica da variância condicional, sobretudo em séries financeiras e de preços de commodities, modelos da família ARCH e GARCH permitem que a volatilidade dependa de choques passados, capturando o fenômeno de aglomeração de volatilidade característico desses dados (Engle, 1982; Bollerslev, 1986). Adicionalmente, modelos de mudança de regime, como o Markov *Switching*, assumem que os parâmetros da série seguem um processo de Markov oculto, o que possibilita descrever transições entre estados, por exemplo, de alta e baixa volatilidade ou de expansão e contração da atividade econômica (Hamilton, 1989).

2.3 CADEIAS DE MARKOV

As cadeias de Markov representam uma classe específica de processos estocásticos discretos em tempo e espaço de estados, nos quais a evolução do sistema é governada pela propriedade de ausência de memória, segundo a qual o próximo estado depende apenas do estado atual e não dos anteriores (Norris, 1998), significando que a probabilidade de transição para o próximo estado depende exclusivamente do estado atual, independentemente da história pregressa do processo. Formalmente, para uma sequência de variáveis aleatórias $\{x_t, t = 0, 1, 2, \dots\}$ assumindo valores em um conjunto finito de estados, a propriedade de Markov estabelece que, $P(X_{t+1} = j \mid X_t = i, X_{t-1} = k_{t-1}, \dots, X_0 = k_0) = P(X_{t+1} = j \mid X_t = i)$ para todos os estados $i, j, k_{t-1}, \dots, k_0$ e todo $t \geq 0$ (Ross, 1995). A principal vantagem dessa definição reside na simplificação computacional e analítica da independência condicional. Tal propriedade torna eficiente a modelagem de fenômenos com dependências locais, superando o desafio imposto por modelos de memória infinita.

A matriz de transição constitui o elemento central para descrever o comportamento de uma cadeia de Markov, sendo denotada por $P = [p_{ij}]$, onde p_{ij} representa a probabilidade de transição do estado i para o estado j em um único passo. As entradas dessa matriz satisfazem as condições de não-negatividade, ou seja, $p_{ij} \geq 0$ para todos i e j , e a soma das probabilidades em cada linha iguala a unidade, $\sum_j p_{ij} = 1$, garantindo que a matriz seja estocástica e preserve a massa probabilística ao longo das transições (Kemeny e Snell, 1976, p. 14). Essa estrutura matricial facilita a computação de probabilidades de transição em múltiplos passos, por meio de potências da matriz, como $P^{(n)} = P^n$, que captura a dinâmica evolutiva do processo.

A distribuição estacionária, ou de equilíbrio, emerge como uma propriedade de longo prazo em cadeias de Markov irredutíveis e aperiódicas, representando um vetor de probabilidades $\pi = (\pi_1, \pi_2, \dots, \pi_m)$ que satisfaz a equação $\pi = \pi P$, sujeito à restrição de normalização $\sum_{i=1}^m \pi_i = 1$. Essa solução corresponde ao autovetor esquerdo associado ao autovalor unitário da matriz de transição e indica as probabilidades limite de ocupação de cada estado à medida que o número de passos tende ao infinito, independentemente da distribuição inicial (Hillier e Lieberman, 2005). A importância da distribuição estacionária

reside em sua capacidade de fornecer insights sobre o comportamento assintótico do sistema, permitindo inferir a fração de tempo gasto em cada estado em regimes de equilíbrio, o que é crucial para análises de estabilidade em modelagens de longo prazo (Norris, 1998).

O tempo médio de recorrência para um estado i , denotado por $m_i = \frac{1}{\pi_i}$, quantifica o número esperado de passos até o retorno ao estado i a partir dele mesmo, assumindo que a cadeia seja ergódica. Essa métrica deriva diretamente da distribuição estacionária e reflete a frequência relativa de visitas a um estado específico, com valores menores indicando estados mais recorrentes e, conseqüentemente, mais centrais na dinâmica do processo (Taha, 2007). Sua relevância estende-se à avaliação de persistência em sistemas estocásticos, auxiliando na identificação de padrões cíclicos e na previsão de intervalos entre eventos de interesse.

As cadeias de Markov possuem ampla aplicação em diversas áreas, como engenharia, física, biologia e finanças, e incluem aplicações consolidadas em teoria da confiabilidade em engenharia (Privault, 2018). A capacidade de quantificar incertezas por meio de probabilidades de transição torna este método uma ferramenta poderosa para a tomada de decisão em contextos dinâmicos (Alagoz *et al.*, 2010). Nesse escopo, sua aplicação na análise de preços de commodities emerge como um campo de especial interesse, que será detalhado na próxima seção.

2.4 APLICAÇÕES DE CADEIAS DE MARKOV EM PREÇOS DE COMMODITIES

A aplicação de Cadeias de Markov tem se estabelecido como uma metodologia robusta para a análise da dinâmica de preços em mercados de *commodities*, permitindo a modelagem de transições probabilísticas entre distintos regimes de volatilidade e variações econômicas. Essa abordagem é particularmente pertinente para mercados agrícolas, os quais são intrinsecamente suscetíveis a incertezas e flutuações (Porto, 2022).

A pertinência do método é reforçada por sua capacidade de capturar o comportamento de sistemas em crises: Kristoufek e Vosvrda (2014), por exemplo, incorporaram métricas markovianas para investigar a eficiência de mercados futuros de *commodities*, avaliando a memória de longo prazo e detectando padrões persistentes de volatilidade. Em uma linha similar, Mehrdoust, Noorani e Kannianen (2024) utilizaram o

modelo de *Markov-switching* para identificar e modelar saltos súbitos nos preços em diferentes *commodities*, quantificando o risco associado às flutuações de oferta e demanda. A capacidade de identificar e quantificar esses regimes de volatilidade é crucial para a gestão de riscos e para a formulação de estratégias de mercado, especialmente no contexto de mercadorias agrícolas.

No cenário internacional, diversas aplicações demonstram a eficácia das Cadeias de Markov na compreensão da variabilidade dos preços de *commodities*. Mehrdoust *et al.* (2024), por exemplo, propõem um modelo de *Markov-switching* para avaliar a dinâmica de futuros e preços *spot* de *commodities*, como o petróleo, incorporando cadeias de Markov ocultas para modelar saltos súbitos nos preços e capturar diferentes estados econômicos. Embora focado no petróleo, a metodologia ressalta a relevância de considerar mudanças de regime para entender a formação de preços e a precificação de derivativos em mercados voláteis. Adicionalmente, Goyal *et al.* (2023) empregam métodos de *Markov Chain Monte Carlo* para decompor erros em previsões de estoques finais do USDA (Departamento de Agricultura dos Estados Unidos). para mercadorias como a soja, revelando ineficiências que afetam a precisão das projeções e demonstrando a utilidade da metodologia na quantificação de riscos associados a flutuações de oferta e demanda. Kristoufek e Vosvrda (2014) investigam a eficiência de mercados de futuros de *commodities*, incorporando métricas baseadas em Cadeias de Markov para avaliar a memória de longo prazo. Seus achados revelam padrões de volatilidade persistente em preços de bens agrícolas e energéticos, auxiliando na detecção de bolhas e na previsão de regimes instáveis. Tais estudos sublinham a importância de modelos que se ajustem às condições de mercado em constante evolução, um aspecto fundamental para a análise da soja.

A aplicação de Cadeias de Markov tem se mostrado valiosa para analisar a transmissão de risco e a interdependência entre mercados, bem como a variabilidade de rendimentos. Hu *et al.* (2024) investigam o transbordamento de risco extremo entre os mercados de futuros de soja dos EUA e da China, utilizando um modelo de cópula com mudança de regime de Markov. Os resultados indicam que a dependência entre esses mercados é assimétrica e intensifica-se em períodos de crise, o que é de grande relevância para a compreensão da dinâmica de preços da soja globalmente. Adicionalmente, Stokes *et al.* (2023) aplicam Cadeias de Markov para modelar a variabilidade intra-anual de

rendimentos de soja, integrando dados ambientais para estimar transições entre estados de produtividade, o que ilustra a capacidade da abordagem em interpretar os impactos climáticos sobre as variações de preços que derivam de colheitas variáveis.

De forma complementar, Xu, Cai e Zhang (2025) investigam como o risco transborda entre os mercados de futuros de soja dos EUA e da China, um fenômeno conhecido como *spillover*. Para isso, empregam um modelo de Vetor Autorregressivo Bayesiano (BVAR), uma técnica estatística que analisa a influência mútua entre múltiplas séries temporais combinado com mudanças de regime de Markov. Seus achados reforçam que esses *spillovers* são assimétricos e dependentes do regime de mercado, destacando a capacidade dos modelos markovianos em elucidar a transmissão de choques. A relevância dessa abordagem para a soja reside justamente em mapear essa interdependência, um fator crítico para produtores e investidores.

Além da análise de volatilidade, as Cadeias de Markov também contribuem para a compreensão do comportamento de agentes de mercado e para a previsão de preços. Mohamad (2024) aplica um modelo de regressão com parâmetros variáveis no tempo (TVP) e amostragem de Monte Carlo via Cadeias de Markov (MCMC) para analisar o comportamento de manada em mercados de *commodities* durante períodos de crise, como a pandemia de COVID-19 e a guerra Russo-Ucraniana. Embora o foco seja o comportamento de manada, a metodologia markoviana foi essencial para capturar a intensidade variável desse fenômeno ao longo do tempo, o que tem implicações diretas para a volatilidade dos preços da soja.

No âmbito nacional, o estudo de Cechin e Corso (2017) exemplifica a aplicação direta das Cadeias de Markov para prever variações diárias no preço da soja no Brasil, utilizando dados do CEPEA. Este trabalho calculou probabilidades de estados de flutuação e tempos de recorrência esperados, estabelecendo um precedente fundamental para análises subsequentes e demonstrando a utilidade prática da ferramenta para minimizar riscos e maximizar lucros por meio de decisões informadas de venda no setor agrícola brasileiro.

Em síntese, as aplicações das Cadeias de Markov em mercados de *commodities*, especialmente no mercado de soja, estendem-se desde a identificação de regimes de volatilidade e a avaliação da propagação de riscos até a análise do comportamento de mercado e a previsão de preços. A flexibilidade e a solidez desses modelos possibilitam

uma compreensão aprofundada das complexas interações que influenciam a variabilidade dos preços, oferecendo instrumentos cruciais para a gestão estratégica em um cenário de elevada incerteza. A habilidade de modelar a dependência sequencial e as transições entre os estados de preço confere às Cadeias de Markov um papel fundamental na otimização de estratégias de proteção (*hedging*) e na fundamentação de decisões comerciais no setor agrícola.

2.5 FERRAMENTAS COMPUTACIONAIS EM MODELAGEM ESTOCÁSTICA

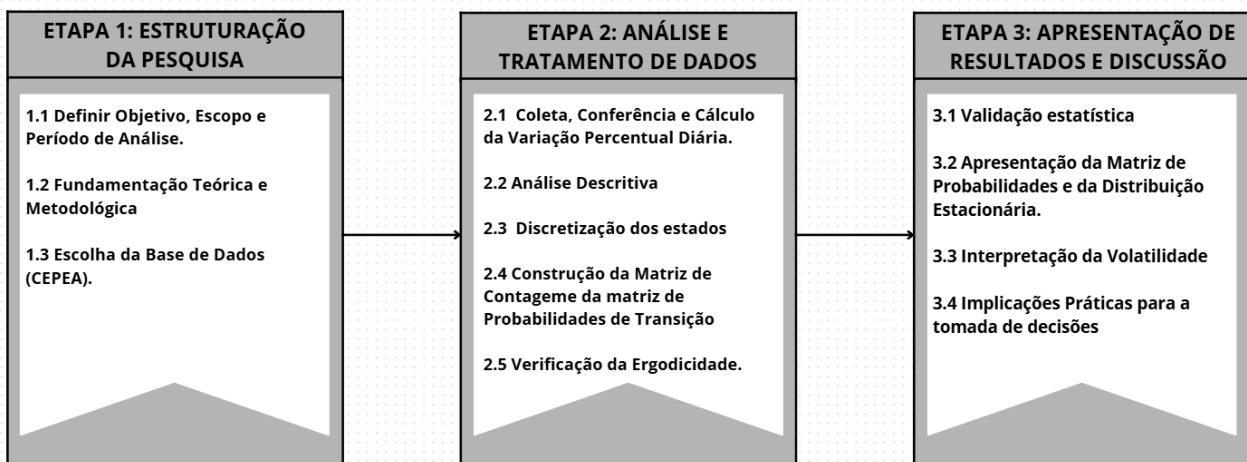
A implementação de modelagens estocásticas baseadas em Cadeias de Markov tem sido significativamente facilitada pelo avanço e pela popularização de ferramentas computacionais, como Excel, “R” e *Python*. Essas plataformas oferecem recursos que permitem a construção de matrizes de transição, a realização de simulações e a execução de análises estatísticas complexas, com um foco crescente na replicabilidade e na visualização dos resultados. A linguagem “R” utilizada junto com pacotes especializados como *markovchain*, oferece rotinas mais avançadas para estimação e plotagem de diagramas de estados, promovendo análises acessíveis tanto em contextos educacionais quanto aplicados (Lima, 2024).

Por sua vez, *Python* destaca-se pela sua flexibilidade e pelo vasto ecossistema de bibliotecas, que suportam desde a manipulação de dados até a construção de modelos estocásticos complexos (Maarleveld; Olivier; Bruggeman, 2013). Bibliotecas como *PyDTMC* são particularmente úteis para trabalhar com cadeias finitas, oferecendo funcionalidades para o cálculo de tempos de recorrência e a análise de ergodicidade. A integração do *Python* com ambientes de código aberto e sua capacidade de automação de simulações, como as de Monte Carlo, e a geração de gráficos interativos, são cruciais para validar modelos em séries temporais voláteis. Essa característica não apenas eleva o rigor da análise, mas também se alinha às demandas contemporâneas por transparência e reprodutibilidade científica em pesquisas quantitativas (Dudda *et al.*, 2025). A escolha da ferramenta computacional, portanto, deve considerar a complexidade do modelo, o volume de dados e a necessidade de replicabilidade e visualização, garantindo que a análise seja robusta e os resultados confiáveis.

3. PROCEDIMENTOS METODOLÓGICOS

Este capítulo descreve o percurso metodológico adotado neste trabalho, desde a coleta de dados até a aplicação das propriedades da Cadeia de Markov. A Figura 2 apresenta, de forma esquemática, as etapas realizadas. Esse fluxograma organiza o processo de modo sequencial, permitindo compreender a lógica do estudo e facilitar sua replicação.

Figura 2 - Fluxograma Metodológico
Fluxograma Metodológico: Etapas de Modelagem da Volatilidade
do Preço da Soja por Cadeias de Markov



Fonte: Autor (2025)

3.1 COLETA DE DADOS

Os dados utilizados neste estudo consistem em preços diários à vista (*spot*) da soja em grão no porto de Paranaguá (PR), coletados no Centro de Estudos Avançados em Economia Aplicada (CEPEA), vinculado à Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo (ESALQ-USP). A escolha dessa fonte deve-se à sua confiabilidade e disponibilidade pública, com dados atualizados e padronizados para o mercado brasileiro de mercadorias agrícolas.

O período analisado abrange os dias de negociação da soja de 02/01/2020 a 01/08/2025, totalizando 1393 observações. Esse intervalo foi selecionado por capturar diversos períodos, incluindo múltiplas safras e entressafras, permitindo uma análise

robusta das variações de preço. Os preços referem-se ao indicador da soja em grão no mercado de Paranaguá (PR), expressos em reais por saca de 60 kg.

A coleta dos dados foi realizada diretamente do site do CEPEA, com o *download* efetuado em formato (CSV). Posteriormente, os dados foram conferidos para a aplicação da metodologia em *Python*. Esta etapa incluiu uma verificação preliminar da série temporal de preços, a qual confirmou a ausência de erros e a integridade dos dados. As análises visuais e os resultados detalhados dessa verificação serão apresentados na seção 4.

3.2 CÁLCULO DAS VARIAÇÕES PERCENTUAIS E DEFINIÇÃO DE ESTADOS

A partir dos preços diários coletados, calculou-se a variação percentual diária para transformar a série temporal em uma sequência de mudanças relativas, adequada à modelagem probabilística. A equação (1) aplicada foi:

$$\Delta P_t = \frac{P_t - P_{t-1}}{P_{t-1}} \cdot 100 \quad \text{Equação (1)}$$

Onde P_t representa o preço registrado no dia t e P_{t-1} o preço no dia anterior. Essa operação resultou em 1392 variações percentuais, uma a menos que o número total de observações de preços (1393), devido à necessidade de pares consecutivos.

O tratamento e a análise da série temporal iniciaram-se com uma análise exploratória detalhada da variável contínua de interesse: a variação percentual diária do preço à vista. Esta etapa preliminar, de natureza fundamental, consistiu no cálculo de estatísticas descritivas essenciais, como média, desvio-padrão, valores mínimo e máximo.

O objetivo primário desta análise foi obter uma compreensão do comportamento estatístico e da distribuição da variação, servindo como subsídio empírico para a etapa de discretização subsequente. A visualização da série histórica e a distribuição completa das variações, que consolidam a base para a modelagem, serão apresentadas e discutidas em profundidade na Seção 4.

Em um movimento metodológico crucial para a aplicação da Cadeia de Markov de tempo discreto, a variável contínua de variação percentual foi submetida a um processo de discretização manual, global e estática. Esta transformação resultou em um conjunto finito e mutuamente exclusivo de sete estados (E1 a E7). A escolha por uma abordagem de

discretização classificada como não supervisionada e orientada ao uso foi estrategicamente definida para maximizar a interpretabilidade gerencial e garantir a aderência direta aos processos de tomada de decisão no contexto do agronegócio.

O critério central para a definição dos limiares de corte dos intervalos foi a garantia de uma contagem mínima de instâncias em cada estado. Este rigor visa mitigar o risco de categorias com fragmentação de dados, promovendo a robustez estatística das probabilidades de transição da Cadeia de Markov. Tal procedimento é justificado pela literatura, que estabelece a discretização de atributos contínuos como um pré-processamento essencial para algoritmos que demandam um espaço de estados discreto, sendo um fator determinante para a acurácia e a confiabilidade do modelo (DOUGHERTY; KOHAVI; SAHAMI, 1995, p. 7).

3.3 CONSTRUÇÃO DA MATRIZ DE TRANSIÇÃO

Com os estados devidamente definidos a partir das variações percentuais, procedeu-se à construção da matriz de transição. Esta matriz é essencial para capturar as probabilidades de mudança entre estados consecutivos, sendo a essência da aplicação da Cadeia de Markov. Inicialmente, percorreu-se a sequência temporal de estados, contando o número de transições de cada estado i para j . Para cada par de dias consecutivos, registrou-se se o estado no dia j era i e no dia $t + 1$ era j , registrando a ocorrência. O processo foi implementado em *Python*, utilizando a função “*pd.crosstab*” da biblioteca *pandas* para gerar a matriz de contagens.

Em seguida, as probabilidades de transição foram calculadas como frequências relativas, normalizando a matriz de contagens por linha, de acordo com a equação (2):

$$P_{ij} = \frac{\text{número de transições de } i \rightarrow j}{\text{total de transições saindo de } i} \quad \text{Equação (2)}$$

Isso garante que cada linha da matriz some 1, atendendo à propriedade estocástica. A normalização foi obtida com o método “.div(axis=0)” em *Python*. A Matriz de Contagem de Transições e a Matriz de Probabilidades de Transição resultantes deste procedimento serão apresentadas e analisadas detalhadamente na subseção 4.4.

3.4 VALIDAÇÃO DA HIPÓTESE MARKOVIANA E ANÁLISE DE ORDEM

A aplicação da Cadeia de Markov de tempo discreto é fundamentada na hipótese Markoviana de primeira ordem, que postula que a probabilidade de transição para o estado futuro é determinada exclusivamente pelo estado presente, sendo independente da história pregressa. Dada a complexidade dos dados de mercado e a potencial existência de dependência de longo prazo, a validade desta suposição foi submetida a um rigoroso escrutínio estatístico, empregando a metodologia proposta por Anderson e Goodman (1957).

Para tal, foram mobilizados testes estatísticos formais com o objetivo de comparar a adequação de um modelo de primeira ordem (M_1) em relação a um modelo de segunda ordem (M_2) (que incorpora a dependência dos dois estados anteriores). A metodologia de validação compreendeu:

1. Teste Qui-quadrado de Tripletas (χ^2): Utilizado para determinar se a inclusão do estado de anteontem (memória de dois dias) contribui significativamente para a melhoria da previsão do estado de amanhã, em comparação com a previsão baseada unicamente no estado de ontem.
2. Teste da Razão de Verossimilhança (LRT - Likelihood Ratio Test): Aplicado para contrastar a verossimilhança do modelo M_1 com a do modelo M_2 , estabelecendo qual ordem de dependência oferece um ajuste estatisticamente superior aos dados observados.
3. Teste de Chapman-Kolmogorov (CK): Empregado como um mecanismo de checagem global para avaliar a consistência interna do modelo de primeira ordem ao longo de múltiplos passos. Este teste compara a matriz de transição de dois passos prevista pelo modelo (P^2) com a matriz de transição de dois passos observada empiricamente ($P_{\{emp\}}^2$), verificando a coerência global da hipótese de primeira ordem.

A inclusão destes testes confere a validade à escolha do modelo de primeira ordem, ou, alternativamente, permite a identificação e quantificação de eventuais limitações do modelo.

3.5 ANÁLISE DE CONVERGÊNCIA E DETERMINAÇÃO DA DISTRIBUIÇÃO ESTACIONÁRIA

Uma vez estabelecida a validade da hipótese Markoviana de primeira ordem (Seção 3.4), o foco metodológico desloca-se para a análise das propriedades de longo prazo da cadeia. O primeiro passo consiste na verificação da ergodicidade da cadeia, uma propriedade fundamental que assegura a convergência do sistema para um estado de equilíbrio estável.

A verificação da ergodicidade compreendeu duas etapas: a avaliação da irredutibilidade, que exige que todos os estados sejam acessíveis entre si por meio de transições possíveis (diretas ou indiretas); e a confirmação da aperiodicidade, que é garantida pela presença de probabilidades positivas na diagonal principal da matriz, permitindo retornos a um estado em intervalos de tempo variados. A confirmação destas propriedades assegura a existência e a unicidade de uma distribuição estacionária (π).

O cálculo da distribuição estacionária (π) representa o estado de equilíbrio de longo prazo da Cadeia de Markov e é obtido pela solução do sistema de equações definido pela Equação (3):

$$\pi P = \pi \quad \text{Equação (3)}$$

A solução do sistema deve satisfazer as restrições de normalização e não negatividade apresentadas na Equação (4).

$$\sum_{i=1}^k \pi_i = 1, \pi_i \geq 0; (i = 1, \dots, k) \quad \text{Equação (4)}$$

Onde k é o número de estados. A matriz de transição k é assumida estocástica por linhas, conforme a Equação (5), o que implica que a soma das probabilidades em cada linha é unitária e todas as probabilidades são não negativas:

$$\sum_{j=1}^k p_{\{ij\}} = 1, \quad p_{\{ij\}} \geq 0; (i = 1, \dots, k) \quad \text{Equação (5)}$$

Adicionalmente, a distribuição estacionária π permite o cálculo do tempo médio de recorrência (T_i) para cada estado i , que representa o número esperado de passos para que a cadeia retorne ao estado i . Este tempo é determinado pela Equação (6):

$$T_i = \frac{1}{\pi_i} \tag{Equação (6)}$$

A implementação computacional dessas etapas foi realizada em linguagem de programação *Python*, com apoio de bibliotecas especializadas. A biblioteca *pandas* foi empregada para a manipulação da base de dados de preços, o cálculo das variações diárias e a construção da matriz de transição por meio da função *pd.crosstab*. A biblioteca *NumPy* foi utilizada para o cálculo de autovalores e autovetores da matriz de transição e para a resolução do sistema linear associado à Equação (3). Por fim, a biblioteca *PyDTMC* foi empregada na análise de Cadeias de Markov finitas, incluindo o cálculo dos tempos médios de recorrência e a verificação numérica da ergodicidade da cadeia. A síntese das bibliotecas de *Python* utilizadas para realização desta trabalho está representada na Figura 3.

Figura 3 - Principais bibliotecas em *Python* utilizadas na manipulação de dados e na análise da Cadeia de Markov

Pandas	NumPy	PyDTMC
• Manipulação de dados; • Cálculo de variações; • Construção de matrizes com (<i>pd.crosstab</i>). • Geração de gráficos descritivos a partir dos <i>DataFrames</i>	• Cálculo de autovetores e autovalores; • Resolução de sistemas de equações.	• Análise de Cadeias de Markov finitas; • Cálculo de tempos de recorrência e ergodicidade

Fonte: Autor 2025

A conclusão desta etapa de análise estabelece o perfil de equilíbrio da cadeia, fornecendo os parâmetros essenciais para a discussão metodológica e a análise de premissas que se segue na Seção 3.6.

3.6 DISCUSSÃO METODOLÓGICA E PREMISSAS

O arcabouço metodológico para a modelagem da volatilidade dos preços da soja por meio de Cadeias de Markov estabeleceu uma progressão lógica de procedimentos. Iniciou-se com o Processamento de Dados e Discretização da Variável de Variação (Seção 3.2), onde as variações percentuais diárias foram calculadas (Equação 1) e discretizadas em sete estados. A etapa subsequente consistiu na Estimação da Matriz de Probabilidades de Transição (Seção 3.3), quantificando a probabilidade de mudança entre estados (Equação 2). A Validação da Hipótese Markoviana e Análise de Ordem (Seção 3.4) foi crucial para o escrutínio estatístico da premissa de primeira ordem. Por fim, a Análise de Convergência e Determinação da Distribuição Estacionária (Seção 3.5) permitiu a obtenção da distribuição de equilíbrio pela solução do sistema e o cálculo do tempo médio de recorrência (Equação 6).

Essa abordagem é adequada para o problema de volatilidade nos preços da soja, pois transforma séries temporais contínuas em processos discretos probabilísticos através de estruturas de mudança de regime Markov, permitindo capturar quebras estruturais e padrões de volatilidade não observáveis (Luo *et al.*, 2019). Isso facilita a captura de dependências de curto prazo em mercados agrícolas sujeitos a influências sazonais e externas (Richter; Sorensen, 2002). A metodologia prepara a análise empírica ao fornecer uma estrutura para prever probabilidades de curto e médio prazo e tempos de retorno a estados específicos, permitindo avaliar riscos e oportunidades de comercialização sem demandar premissas paramétricas rígidas, comuns em modelos tradicionais.

O método baseia-se em premissas essenciais: a cadeia é de primeira ordem, dependendo apenas do estado anterior, e a matriz P assume-se constante ao longo do período analisado. Condições de irreducibilidade garantem comunicação entre todos os estados, enquanto aperiodicidade permite ciclos variados. Esses procedimentos alinham os objetivos do estudo, compreender e prever variações de preços da soja, com uma aplicação prática de Cadeias de Markov.

4. RESULTADOS

Este capítulo destina-se à apresentação e à análise crítica dos resultados obtidos a partir da aplicação dos conceitos de Cadeias de Markov, conforme detalhado no Capítulo 3. A estrutura deste capítulo segue o fluxo lógico do processamento de dados, iniciando-se pela Análise Preliminar dos Dados Coletados (Seção 4.1), passando pela Discretização dos Estados (Seção 4.2). Na sequência, aborda-se a Verificação da Propriedade Markoviana de Primeira Ordem (Seção 4.3) e a construção das Matrizes de Contagem e Probabilidade de Transição (Seção 4.4). Posteriormente, apresenta-se a Distribuição Estacionária e Análise de Longo Prazo (Seção 4.5), que fornece a base para a compreensão do comportamento de longo prazo do preço da *commodity*. Por fim, a Seção 4.6 apresenta a Discussão e Síntese dos Resultados, conectando os achados empíricos com o referencial teórico e os objetivos propostos por este trabalho.

4.1 ANÁLISE PRELIMINAR DOS DADOS COLETADOS

A análise preliminar dos dados de preços da soja, coletados do CEPEA, revelou a consistência da série temporal ao longo do período estudado. Conforme ilustrado na Figura 4, a série temporal dos preços diários da soja não apresentou anomalias ou erros, validando a base para as análises subsequentes.

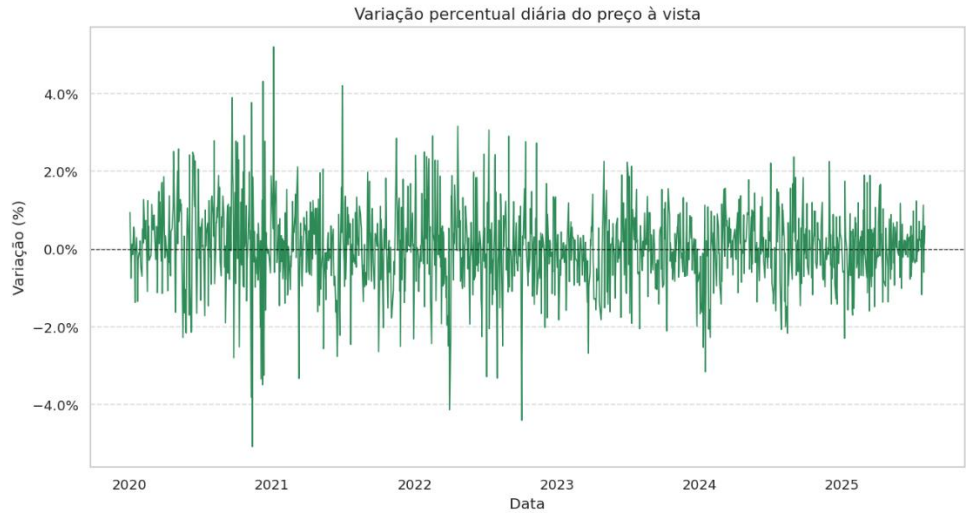
Figura 4 - Evolução do Preço da Saca de Soja (R\$) — Indicador CEPEA/B3



Fonte: Autor (2025)

Observou-se, a partir da série temporal, a presença de variações diárias nos preços. Para uma análise detalhada dessas flutuações, utilizou-se a equação (1), resultando na construção do gráfico de variação percentual diária. A Figura 5, ilustra a variação percentual diária do preço à vista, evidenciando a dinâmica de oscilação dos valores ao longo do tempo.

Figura 5 - Variação percentual diária do preço à vista



Fonte: Autor (2025)

Para uma compreensão mais aprofundada do comportamento dessas variações, a Tabela 1 sumariza as estatísticas descritivas básicas das variações percentuais diárias, fornecendo uma visão quantitativa de seu comportamento.

Tabela 1 - Estatísticas Descritivas das Variações Percentuais Diárias dos Preços da Soja

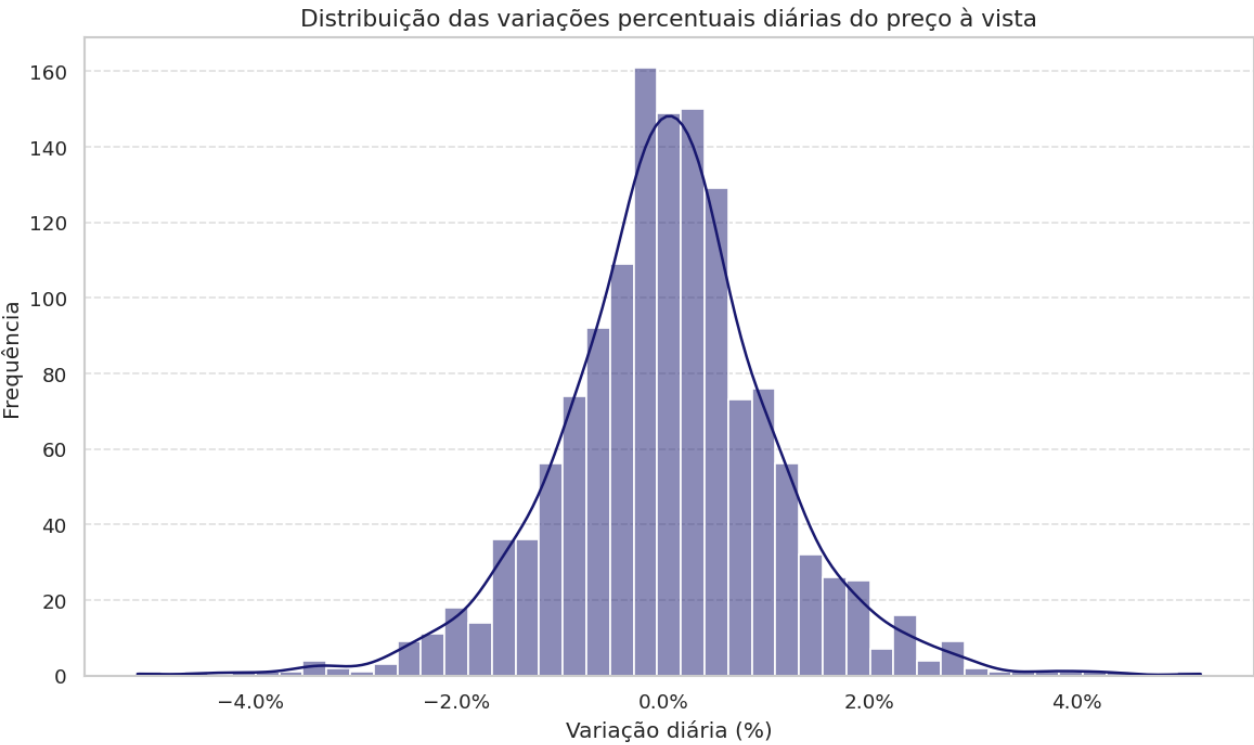
Métrica	Resultado
Média Variação diária	0.04%
Desvio-padrão diário	1.06%
Mínimo diário	-5.09%
Máximo diário	5.20%
Número de dias	1.392

Fonte: Autor (2025)

A análise dessas métricas revelou uma distribuição concentrada em torno de zero, com caudas pesadas. Para uma visualização mais detalhada, o Histograma das Variações Diárias é apresentado na Figura 6, confirmando uma forma leptocúrtica com pico central e

declínio simétrico nas extremidades. Este perfil de distribuição orientou a escolha dos estados, priorizando a densidade em torno de zero.

Figura 6 - Histograma das variações diárias com curva de densidade sobreposta.



4.2 DISCRETIZAÇÃO DOS ESTADOS DE VARIAÇÃO DE PREÇO

Conforme delineado na seção 3.2, a discretização das variações percentuais diárias constitui uma etapa fundamental para a aplicação da metodologia de Cadeias de Markov de tempo discreto. Este procedimento visa transformar a série contínua de variações em um conjunto finito de estados, permitindo a análise probabilística das transições entre diferentes níveis de variação. A definição desses estados foi realizada empiricamente, com base na distribuição observada das variações, adotando sete intervalos assimétricos para otimizar a representação da densidade de frequência e garantir a robustez estatística, conforme descrito na metodologia.

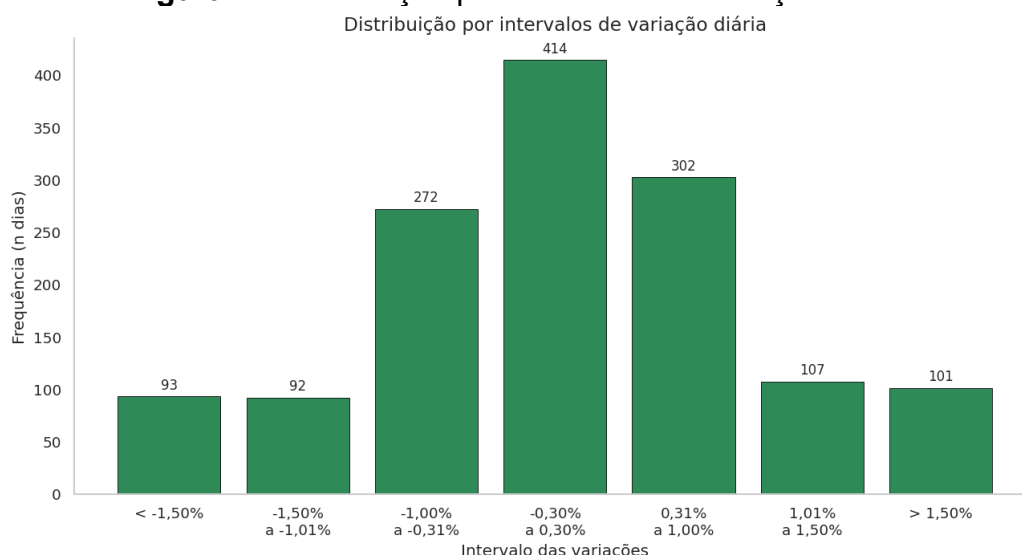
A Tabela 2 detalha a definição de cada um dos sete estados utilizados na discretização das variações percentuais diárias, incluindo suas respectivas frequências e proporções, que foram cruciais para a construção da matriz de transição.

Tabela 2 - Definição e Frequência dos Estados de Variação Percentual Diária.

Estado	Intervalo de variação diária (%)	Frequência	Proporção (%)	Frequência acumulada	Proporção acumulada (%)
E1	< -1,50%	93	6,68%	93	6,68%
E2	-1,50% a -1,01%	92	6,61%	185	13,29%
E3	-1,00% a -0,31%	272	19,54%	457	32,83%
E4	-0,30% a 0,30%	414	29,74%	871	62,57%
E5	0,31% a 1,00%	302	21,70%	1 173	84,27%
E6	1,01% a 1,50%	107	7,69%	1 280	91,96%
E7	> 1,50%	101	7,26%	1 392	100,00%

Fonte: Autor (2025)

A Figura 7 ilustra a distribuição das frequências por esses intervalos, demonstrando visualmente a adequação da discretização em acomodar a maior frequência no centro e garantir a representação das caudas. Observa-se uma concentração significativa das variações em torno de zero, com a frequência diminuindo à medida que os intervalos se afastam do ponto central, o que corrobora a natureza leptocúrtica da distribuição identificada na análise preliminar. Essa distribuição assimétrica e a concentração central justificam a escolha dos sete estados para a modelagem.

Figura 7 - Distribuição por intervalos de variação diária.

Fonte: Autor (2025)

Com os estados devidamente definidos e caracterizados, o próximo passo metodológico é a verificação da adequação do modelo de Cadeia de Markov. Esta análise é crucial para validar a suposição de que o comportamento futuro das variações de preço depende apenas do estado atual.

4.3 VERIFICAÇÃO DA PROPRIEDADE MARKOVIANA DE PRIMEIRA ORDEM

A validade da suposição de primeira ordem da Cadeia de Markov foi testada por meio de três análises estatísticas, seguindo a abordagem estatística de Anderson e Goodman (1957). Os resultados obtidos são sintetizados na Tabela 3.

Tabela 3 - Resultados dos Testes de Propriedade Markoviana.

Teste	Estatística	gl	p
χ^2 de Tripletas	305,893	252	0,011385
LRT Ordem 1 vs 2	325,285	187	$1,526 \times 10^{-9}$
Chapman–Kolmogorov (CK)	35,015	42	0,7689

Fonte: Autor (2025)

Os testes χ^2 e *LRT* indicaram que a inclusão de um segundo dia (ordem 2) melhora o ajuste do modelo em alguns padrões de transição de preço, sugerindo a presença de uma memória de mercado que se estende além de um dia. Contudo, o teste de Chapman-

Kolmogorov (CK) demonstrou que, no agregado, o modelo de primeira ordem não apresenta inconsistência significativa ao longo do tempo.

A escolha pela primeira ordem é particularmente relevante no contexto da tomada de decisão gerencial diária, como em operações de *trading* ou estratégias de *hedge*. Nesses cenários, a simplicidade do modelo se traduz em maior interpretabilidade gerencial, pois o estado do sistema no dia anterior é o dado mais imediato, prático e relevante para a projeção do estado futuro. Embora a evidência estatística aponte para uma memória de segunda ordem em padrões específicos, a complexidade adicional de um modelo de ordem superior não se justificaria pela sua aplicação prática imediata, onde a informação de "ontem" é o principal insumo para a ação de "hoje". Portanto, a evidência de memória de segunda ordem será considerada uma limitação do modelo e um ponto de discussão na última seção.

Com os estados devidamente definidos e caracterizados, procedeu-se à construção da matriz de transição, que será o foco da próxima subseção, permitindo a modelagem do comportamento futuro das variações.

4.4 MATRIZES DE CONTAGEM E PROBABILIDADE DE TRANSIÇÃO

A aplicação do procedimento metodológico descrito na seção 3.3 resultou na Matriz de Contagem de Transições e, subsequentemente, na Matriz de Probabilidades de Transição. A Matriz representada na tabela (4), detalha as contagens de transições observadas entre os sete estados consecutivos. Esta matriz é fundamental por representar a frequência absoluta de todas as mudanças de estado identificadas na série temporal analisada, sendo a base para o cálculo das probabilidades.

Tabela 4 - Matriz de Contagem de Transições

DE/PARA	E1	E2	E3	E4	E5	E6	E7
E1	20	12	17	14	16	7	7
E2	14	7	23	24	9	12	3
E3	12	10	60	92	70	12	16
E4	24	30	73	148	84	28	27
E5	11	18	58	86	75	27	26
E6	5	9	19	26	27	12	9
E7	7	6	22	24	20	9	13

Fonte: Autor (2025)

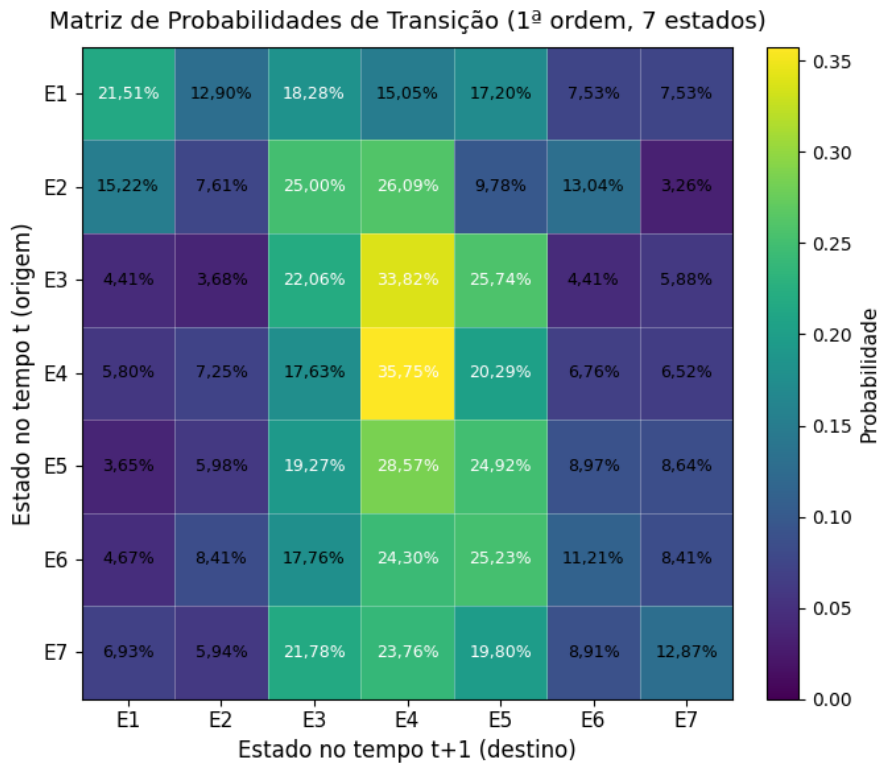
A partir da normalização da Matriz da tabela (4) obteve-se a Matriz presente na tabela (5), que representa a Matriz de Probabilidades de Transição. Cada elemento P_{ij} desta matriz indica a probabilidade do preço da soja transitar do estado i para o estado j no dia seguinte. A análise desta matriz exposta na tabela (5) é crucial para a compreensão da dinâmica do mercado e para a modelagem preditiva, pois revela a persistência de cada estado (elementos da diagonal principal) e a probabilidade de mudanças significativas.

Tabela 5 - Matriz de Probabilidades de Transição

De/Para	E1	E2	E3	E4	E5	E6	E7
E1	21,51%	12,90%	18,28%	15,05%	17,20%	7,53%	7,53%
E2	15,22%	7,61%	25,00%	26,09%	9,78%	13,04%	3,26%
E3	4,41%	3,68%	22,06%	33,82%	25,74%	4,41%	5,88%
E4	5,80%	7,25%	17,63%	35,75%	20,29%	6,76%	6,52%
E5	3,65%	5,98%	19,27%	28,57%	24,92%	8,97%	8,64%
E6	4,67%	8,41%	17,76%	24,30%	25,23%	11,22%	8,41%
E7	6,93%	5,94%	21,78%	23,76%	19,80%	8,91%	12,87%

Fonte: Autor (2025)

Para facilitar a interpretação visual dos dados apresentados na tabela (5), a Figura 6 ilustra a mesma matriz em formato de mapa de calor. Nesta representação, as cores mais quentes (amarelo) indicam as transições com maior probabilidade de ocorrência, enquanto as cores mais frias (roxo) apontam as menos prováveis.

Figura 8 - Mapa de Calor da Matriz de Probabilidades de Transição

Fonte: Autor (2025)

A Matriz de Probabilidades de Transição, visualizada no Mapa de Calor da Figura 8, constitui a base empírica para a compreensão da dinâmica de curto prazo do mercado. A concentração de probabilidades elevadas na diagonal principal, notadamente no estado E_4 (variação diária entre $-0,30\%$ e $0,30\%$), com $35,75\%$, confirma a inércia do mercado, indicando que a permanência no estado de estabilidade é o movimento mais provável. Para o produtor, a análise deve ir além da estabilidade e focar nas transições que representam oportunidades de otimização de lucro e minimização de risco. As transições fora da diagonal principal sinalizam mudanças de regime e são cruciais para a tomada de decisão de venda.

A análise das transições a partir dos estados de variação positiva para os estados de variação negativa é fundamental para a decisão de venda. Observa-se que o estado E_6 (variação entre $1,01\%$ e $1,50\%$) apresenta a maior probabilidade de reversão significativa para um estado de variação negativa, com $17,76\%$ de chance de transicionar para E_3 (variação entre $-1,00\%$ e $-0,31\%$). Embora a probabilidade de o mercado continuar em alta seja maior, esta transição representa o risco de correção mais acentuado após um dia

de forte valorização. Para o produtor, este dado sugere que um dia de variação entre 1,01% e 1,50% deve ser um ponto de atenção máxima, podendo justificar a antecipação da venda de parte da produção para garantir a realização do lucro e mitigar o risco de uma desvalorização moderada no dia seguinte.

Em contraste, o estado E_2 (variação entre $-1,50\%$ e $-1,01\%$) demonstra ser o menos estável em termos de permanência, com apenas 7,61% de probabilidade de se manter. A alta probabilidade de transição de E_2 para E_4 (Estabilidade, 26,09%) e E_3 (variação entre $-1,00\%$ e $-0,31\%$, com 25,00%) indica que, após um dia de variação negativa entre $-1,50\%$ e $-1,01\%$, o mercado tende a moderar a desvalorização ou buscar a estabilidade. Esta dinâmica é relevante para a minimização de perdas e a otimização de lucros futuros, pois sugere que o choque de uma variação negativa é rapidamente absorvido, e a probabilidade de uma variação ainda mais extrema ($E_1 < -1,50\%$) é baixa (15,22%).

A análise da Matriz de Transição revela a dinâmica de curto prazo do mercado, essencial para a definição de estratégias de comercialização táticas. Contudo, a perspectiva de longo prazo é igualmente vital para a formulação de uma política de comercialização sólida. A partir desta matriz, a próxima etapa da análise consiste na determinação da distribuição estacionária e do tempo médio de recorrência para cada estado. Estes elementos, detalhados na Seção 4.5, complementam a análise de curto prazo, fornecendo uma base preditiva mais consistente para a decisão de comercialização e permitindo a avaliação de outras estratégias que consideram o comportamento estrutural do mercado.

4.5 DISTRIBUIÇÃO ESTACIONÁRIA E ANÁLISE DE LONGO PRAZO

Prossegue-se a análise iniciada na Seção 3, na qual foram definidos os estados de variação diária da série temporal e estabelecida a ergodicidade da cadeia de Markov. Conforme descrito anteriormente, primeiro discretizaram-se as variações percentuais diárias em sete estados equilibrados e, em seguida, verificou-se a irredutibilidade e aperiodicidade da matriz de transição (P) o que assegurou a existência de uma única distribuição estacionária $\pi...$ Aplicando a equação (3), juntamente com a restrição de que

a soma das probabilidades deve ser 1, obtém-se o sistema de equações lineares que define o comportamento do sistema no longo prazo, expresso por:

$$\begin{aligned}
 \pi_1 &= 0.215\pi_1 + 0.152\pi_2 + 0.044\pi_3 + 0.058\pi_4 + 0.037\pi_5 + 0.047\pi_6 + 0.069\pi_7 \\
 \pi_2 &= 0.129\pi_1 + 0.076\pi_2 + 0.037\pi_3 + 0.072\pi_4 + 0.060\pi_5 + 0.084\pi_6 + 0.059\pi_7 \\
 \pi_3 &= 0.183\pi_1 + 0.250\pi_2 + 0.221\pi_3 + 0.176\pi_4 + 0.193\pi_5 + 0.178\pi_6 + 0.218\pi_7 \\
 \pi_4 &= 0.151\pi_1 + 0.261\pi_2 + 0.338\pi_3 + 0.357\pi_4 + 0.286\pi_5 + 0.243\pi_6 + 0.238\pi_7 \\
 \pi_5 &= 0.172\pi_1 + 0.098\pi_2 + 0.257\pi_3 + 0.203\pi_4 + 0.249\pi_5 + 0.252\pi_6 + 0.198\pi_7 \\
 \pi_6 &= 0.075\pi_1 + 0.130\pi_2 + 0.044\pi_3 + 0.068\pi_4 + 0.090\pi_5 + 0.112\pi_6 + 0.089\pi_7 \\
 \pi_7 &= 0.075\pi_1 + 0.033\pi_2 + 0.059\pi_3 + 0.065\pi_4 + 0.086\pi_5 + 0.084\pi_6 + 0.129\pi_7 \\
 \pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_5 + \pi_6 + \pi_7 &= 1
 \end{aligned}$$

A Tabela 7 apresenta os resultados da distribuição estacionária π , obtidos a partir da resolução do sistema de equações. Os valores calculados indicam a probabilidade de, no longo prazo, o preço da soja se encontrar em cada um dos sete estados de variação diária definidos.

Tabela 7 - Distribuição estacionária das variações diárias do preço da soja

Estado	Faixa de variação diária (%)	Frequência (%)	Tempo médio de retorno T (dias)
E1	< -1,50%	6,74%	0.067392
E2	-1,50% a -1,01%	6,67%	0.066667
E3	-1,00% a -0,31%	19,71%	0.197101
E4	-0,30% a 0,30%	30,00%	0.300000
E5	0,31% a 1,00%	21,81%	0.218116
E6	1,01% a 1,50%	7,75%	0.077536
E7	> 1,50%	7,32%	0.073188

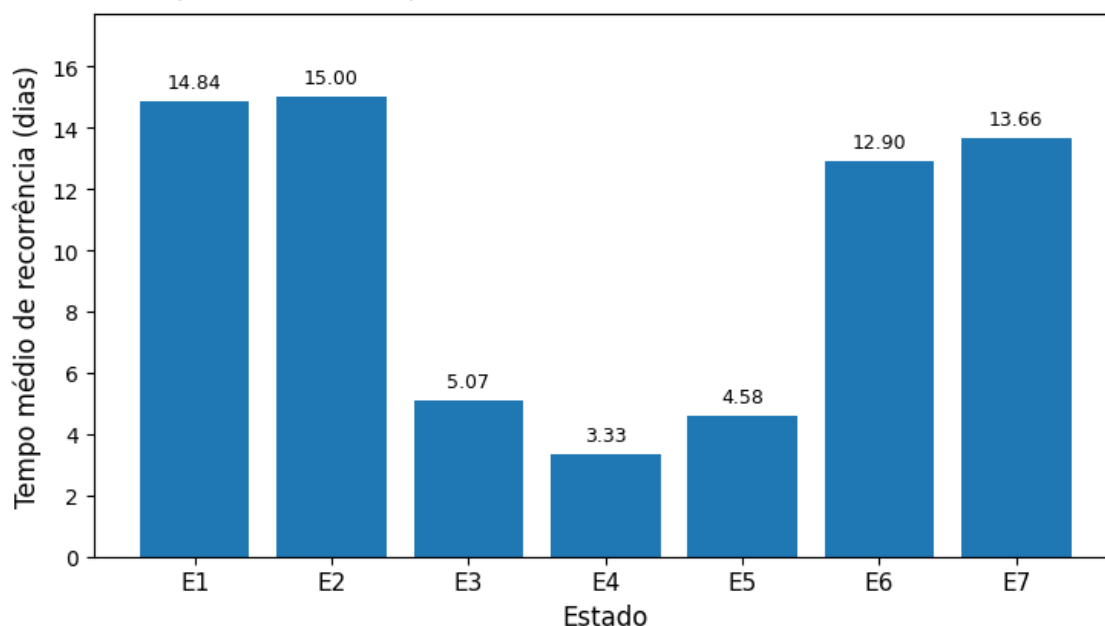
Fonte: Autor (2025)

Observa-se que o estado E4, que representa uma variação diária mínima (entre - 0,30% e 0,30%), é o mais provável no longo prazo, com uma probabilidade de ocorrência de 30,00%. Isso sugere uma forte tendência de estabilidade nos preços da soja, onde o sistema passa a maior parte do tempo. Os estados E3 (19,71%) e E5 (21,81%), que indicam variações moderadas, também apresentam probabilidades significativas. Em contrapartida, os estados de variação extrema, como E1 e E7, são os menos frequentes.

Complementarmente, a Figura 9 ilustra o tempo médio de recorrência para cada estado, ou seja, o número médio de dias esperado até que o sistema retorne a um determinado estado.

Figura 9 - Tempo de recorrência por estado.

Tempo de recorrência por estado – Cadeia de Markov (1ª ordem, 7 estados)

**Fonte:** Autor (2025)

Os resultados do tempo de recorrência corroboram a análise da distribuição estacionária. O estado E4 possui o menor tempo de retorno, de aproximadamente 3,33 dias, indicando que o sistema volta rapidamente a este estado de estabilidade. Por outro lado, os estados de maior variação, E1 e E2, apresentam os maiores tempos de recorrência, com 14,84 e 15,00 dias, respectivamente. Isso reforça a conclusão de que os períodos de estabilidade são mais frequentes e resilientes, enquanto as grandes variações são eventos mais raros e demoram mais a se repetir.

A correlação deste Tempo Médio de Recorrência com o horizonte de planejamento do produtor é um ponto de inflexão para a gestão da comercialização. O longo período de recorrência de aproximadamente 15 dias para os estados de maior variação (E1 e E2) estabelece um limite temporal para a expectativa de ocorrência de eventos extremos. Em um contexto de tomada de decisão que se estende por dias ou semanas, essa baixa frequência de retorno sugere que a estratégia de espera por picos de preço é estatisticamente ineficiente. A análise de probabilidade de transição já indica que, mesmo ao atingir um estado de alta variação, a probabilidade de transição para um estado de menor variação no dia seguinte é alta.

O Tempo Médio de Recorrência de 15 dias complementa esta visão, fornecendo uma métrica de escassez temporal que direciona a decisão para a otimização das oportunidades nos regimes mais frequentes e de menor recorrência, como o E4, onde a ação proativa de venda se torna mais vantajosa do que a reativa. Esta integração de métricas de longo e curto prazo é crucial para a formulação de uma política de comercialização sólida e operacional.

4.6 DISCUSSÃO E SÍNTESE DOS RESULTADOS

A análise dos resultados obtidos, que abrange desde a caracterização das variações diárias até a determinação da distribuição estacionária, proporciona uma compreensão robusta da dinâmica de preços da soja. A constatação inicial de que as variações seguem uma distribuição leptocúrtica está em plena consonância com os achados da literatura especializada em mercados financeiros e de *commodities*. Tais mercados são frequentemente caracterizados por distribuições com caudas pesadas e alta concentração em torno da média, o que, conforme discutido por Li *et al.* 1 (2016), indica que grandes oscilações de preço são eventos de baixa frequência, mas de alta magnitude potencial.

A aplicação da Cadeia de Markov, por sua vez, revelou padrões de transição bem definidos. A alta probabilidade de permanência no estado de baixa volatilidade (E4), confirmada pela maior frequência na diagonal da matriz de transição e pelo peso de 30% na distribuição estacionária, sugere que o preço da soja tende a se manter próximo do equilíbrio na maior parte do tempo. Este resultado é fundamental, o que confirma a premissa de que a modelagem probabilística de primeira ordem é adequada para capturar a memória de curto prazo do processo, onde o estado atual influencia diretamente o estado futuro.

A discussão dos resultados é particularmente relevante para os diversos atores envolvidos na comercialização da soja. Enquanto a análise diária da volatilidade é crucial para atores que comercializam nesse mercado que operam em horizontes de tempo mais curtos, já para o produtor rural, a variável mensal ou semanal de preço pode ser mais contributiva e operacionalmente útil. Isso se deve ao fato de que as decisões estratégicas do produtor, como o momento ideal de venda da safra ou a contratação de hedge, geralmente se alinham a um horizonte temporal mais amplo, como o ciclo mensal de

mercado. Portanto, a adaptação da análise para uma frequência mensal ou semanal, em trabalhos futuros, poderia otimizar a aplicabilidade direta do modelo no planejamento da comercialização da safra.

5. CONCLUSÃO

Essa pesquisa teve como objetivo analisar a dinâmica de preços da saca de soja no Brasil, visando subsidiar a tomada de decisão estratégica e a gestão de riscos na sua comercialização, empregando o modelo de Cadeias de Markov para modelar a volatilidade dessa *commodity* agrícola. Para tanto, utilizou-se como recorte temporal o período compreendido entre 02 de janeiro de 2020 e 01 de agosto de 2025 para aplicação do estudo.

Para alcançar este fim, foi aplicado o modelo de Cadeias de Markov, os procedimentos adotados basearam-se na coleta e processamento de dados diários do preço da saca de soja. O processo envolveu a transformação das variações percentuais diárias em um espaço de estados discreto, composto por sete estados (E1 a E7), que representam diferentes níveis de volatilidade. A partir dessa discretização, foi construída a Matriz de Probabilidade de Transição, que quantifica a probabilidade do preço transitar de um estado de volatilidade para outro em um dia subsequente. A análise da Matriz de Transição e a determinação da Distribuição Estacionária foram os pilares para a compreensão do comportamento de longo prazo do preço.

Os principais resultados da aplicação do modelo de Cadeias de Markov revelaram padrões de persistência e atração na volatilidade dos preços da soja. A análise da Matriz de Transição demonstrou que o estado de "Estabilidade" (E4), caracterizado por variações percentuais diárias entre -0,30% e 0,30%, possui a maior probabilidade de permanência, indicando uma tendência do preço se manter em um patamar de baixa volatilidade após atingi-lo. A Distribuição Estacionária confirmou essa tendência, ao indicar que o preço da soja tende a passar a maior parte do tempo no estado E4 no longo prazo. Além disso, os resultados forneceram *insights* valiosos sobre a probabilidade de ocorrência de eventos extremos (E1 e E7), permitindo aos gestores quantificar o risco de grandes quedas ou altas de preço em um horizonte temporal definido.

A contribuição do estudo reside em oferecer uma ferramenta analítica robusta e acessível, baseada em Cadeias de Markov, para a gestão de riscos no agronegócio brasileiro. Ao quantificar a probabilidade de transição entre diferentes regimes de volatilidade, o modelo provê um subsídio direto para a tomada de decisão estratégica, como a definição de janelas mais vantajosas para a comercialização, a otimização de estratégias de *hedge* e o planejamento de safras. O trabalho alinha-se, assim, à Engenharia de Produção ao aplicar um modelo estocástico para aprimorar a gestão operacional e estratégica em um setor de fundamental importância para a economia nacional.

Ainda, este estudo apresenta limitações inerentes à modelagem. A principal delas reside na assunção de que a Matriz de Transição é constante ao longo do período analisado, o que implica desconsiderar a possível influência de fatores exógenos, como a sazonalidade da safra ou grandes choques macroeconômicos, que poderiam alterar as probabilidades de transição ao longo do tempo. Adicionalmente, a análise estatística de primeira ordem, incluindo testes Qui-Quadrado e LRT (Razão de Verossimilhança), que apresentaram valores superiores a 0,5, sugere que a propriedade de Markov de primeira ordem pode não capturar toda a dependência temporal da série.

Com base nas limitações identificadas, sugere-se para pesquisas futuras a exploração de modelos de Cadeias de Markov de segunda ordem ou superior, a fim de verificar se a inclusão de uma memória mais longa na série temporal melhora a aderência do modelo. Outra linha de pesquisa promissora seria a incorporação de variáveis exógenas ao modelo, como o câmbio (dólar/real), o preço do petróleo ou índices de produção global, por meio de modelos de Cadeias de Markov com covariáveis. Por fim, a aplicação da metodologia em outras *commodities* agrícolas relevantes para o Brasil, como o milho ou o café, permitiria validar a generalidade do modelo proposto e expandir o escopo de ferramentas de gestão de risco para o agronegócio.

REFERÊNCIAS

- ANDERSON, T. W.; GOODMAN, L. A. Statistical inference about Markov chains. ***The Annals of Mathematical Statistics***, v. 28, n. 1, p. 89-110, 1957. Disponível em: <https://scirp.org/reference/ReferencesPapers.aspx?ReferenceID=1524370>. Acesso em: 28 out. 2025.
- BALG, Basher A.; METCALF, Hugh. Modeling exchange rate volatility. ***Review of International Economics***, v. 18, n. 1, p. 109-120, 2010. DOI: 10.1111/j.1467-9396.2009.00872.x. Acesso em: 28 out. 2025.
- BINI, D. A.; CANEVER, M. D.; DENARDIM, A. A. Correlação e causalidade entre os preços de *commodities* e energia. ***Nova Economia***, v. 25, n. 1, p. 143-160, 2015. DOI: 10.1590/0103-6351/1985.
- BOSCHIERO, B. N. Retrospectiva de 2024 e Perspectivas para 2025 no Agronegócio Brasileiro. **Blog Agroadvance**, 2024. Disponível em: <https://agroadvance.com.br/blog-retrospectiva-de-2024-e-perspectivas-para-2025-no-agronegocio-brasileiro/>. Acesso em: 11 nov. 2025.
- BOX, George E. P.; JENKINS, Gwilym M.; REINSEL, Gregory C. **Time Series Analysis: Forecasting and Control**. 4. ed. Hoboken, NJ: John Wiley & Sons, 2008. 746 p.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento; EMBRAPA. **Projeções do Agronegócio: Brasil 2016/2017 a 2026/2027**. Brasília, DF: Ministério da Agricultura, Pecuária e Abastecimento, 2017. 94 p. Disponível em: Acesso em: 28 out. 2025.
- CARVALHO, João Carlos de; PAVAN, Lucca Simeoni; HASEGAWA, Marcos Minoru. Transmissões de volatilidade de preços entre *Commodities* agrícolas brasileiras. ***Revista de Economia e Sociologia Rural***, v. 58, n. 3, e193763, 2020. Disponível em: <https://doi.org/10.1590/1806-9479.2020.193763>. Acesso em: 28 out. 2025.
- CEBALLOS, F. *et al.* Grain price and volatility transmission from international to domestic markets. ***World Development***, v. 96, p. 1-13, ago. 2017. DOI:

10.1016/j.worlddev.2017.03.001. Disponível em:
<https://doi.org/10.1016/j.worlddev.2017.03.001>. Acesso em: 28 out. 2025.

CECHIN, Rafaela Boeira; CORSO, Leandro Luís. Previsão da variação do preço da soja, utilizando Cadeia de Markov. In: CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO, 2017, Foz do Iguaçu. **Anais eletrônicos** [...]. Foz do Iguaçu: APREPRO, 2017. Disponível em:
<https://anteriores.aprepro.org.br/conbrepro/2017/down.php?id=2981&q=1>. Acesso em: 28 out. 2025.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA; ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. Cadeia da soja e biodiesel deve crescer 11,29% em 2025. São Paulo: Agência Safras, 16 out. 2025. Disponível em: [URL da Agência Safras]. Acesso em: 29 out. 2025.

CENTRO DE ESTUDOS AVANÇADOS EM ECONOMIA APLICADA; ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. Cadeia da soja e do biodiesel no Brasil: empregos e PIB. Piracicaba: Cepea/Abiove, 2023. Disponível em:
<https://digital.agrishow.com.br/artigos/cadeia-da-soja-e-do-biodiesel-no-brasil-gera-empregos-e-contribui-com-o-pib-do-agro/>. Acesso em: 29 out. 2025.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Boletim Logístico da Conab**. Ano VIII, jul. 2025. Brasília, DF: Conab, 2025. Disponível em: <https://www.gov.br/conab/pt-br/atuacao/logistica/boletim-logistico/boletim-logistico-julho-2025.pdf>. Acesso em: 30 out. 2025.

DOUGHERTY, James; KOHAVI, Ron; SAHAMI, Mehran. Supervised and Unsupervised Discretization of Continuous Features. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 12., 1995, Tahoe City. **Proceedings of the Twelfth International Conference on Machine Learning**. San Francisco: Morgan Kaufmann, 1995. p. 130-139.

DUDDA, Lukas *et al.* Open science interventions to improve reproducibility and replicability. **Royal Society Open Science**, v. 12, n. 2, p. 2304257, 2025. Disponível em: <https://royalsocietypublishing.org/doi/10.1098/rsos.242057>. Acesso em: 08 out. 2025.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (Embrapa Soja). **Soja em números (safra 2024/25)**. Brasília, DF: Embrapa. Disponível em: [URL da fonte de informação do portal, não fornecida]. Acesso em: 29 out. 2025.

GOYAL, Raghav; ADJEMIAN, Michael K.; GLAUBER, Joseph W.; MEYER, Seth. Decomposing USDA ending stocks forecast errors. **Journal of Agricultural and Resource Economics**, v. 48, n. 1, p. 1-19, abr. 2023. Disponível em: <https://ageconsearch.umn.edu/record/320674/files/JARE320674.pdf>. Acesso em: 23 out. 2025.

HILLIER, Frederick S.; LIEBERMAN, Gerald J. **Introduction to Operations Research**. 8. ed. Nova York: McGraw-Hill, 2005.

HU, X.; ZHU, B.; ZHANG, B.; ZENG, L. Extreme risk *spillovers* between US and Chinese agricultural futures markets in crises: a dependence-switching copula-CoVaR model. **PLoS ONE**, v. 19, n. 3, p. e0299237, 2024. Disponível em: [URL não fornecida]. Acesso em: 28 out. 2025.

INTERNATIONAL INSTITUTE FOR SUSTAINABLE DEVELOPMENT. **Global Market Report: Soybean prices and sustainability**. Winnipeg: International Institute for Sustainable Development, fev. 2024. Disponível em: <https://www.iisd.org/system/files/2024-02/global-market-report-soybean-prices-sustainability.pdf>. Acesso em: 29 out. 2025.

KARLIN, Samuel; TAYLOR, Howard M. **A First Course in Stochastic Process**. Nova York: Academic Press, 1975.

KEMENY, John G.; SNELL, J. Laurie. **Finite Markov Chains**. New York: Springer-Verlag, 1976.

KRISTOUFEK, Ladislav; VOSVRDA, Miloslav. Measuring capital market efficiency: Long-term memory, fractal dimension and approximate entropy. ***The European Physical Journal B: Condensed Matter and Complex Systems***, v. 87, n. 7, p. 1-9, 2014. DOI: 10.1140/epjb/e2014-50113-x. Disponível em:

<https://link.springer.com/article/10.1140/epjb/e2014-50113-x>. Acesso em: 28 out. 2025.

KUMAR, R. *et al.* Spatio-temporal predictive modeling techniques for different domains: a survey. ***ACM Computing Surveys***, v. 57, n. 2, p. 1-42, 2024. DOI: 10.1145/3696661. Disponível em: <https://dl.acm.org/doi/10.1145/3696661>. Acesso em: 17 out. 2025.

LI, Na; KER, Alan P.; SAM, Abdoul G.; ARADHYULA, Satheesh. **Modelling regime-dependent agricultural commodity price volatilities**. Guelph: University of Guelph, Department of Food, Agricultural and Resource Economics, Institute for the Advanced Study of Food and Agricultural Policy, 2016. (Institute White Paper, 2016.2). Disponível em: https://aede.osu.edu/sites/aede/files/publication_files/Volatility%20working%20paper.pdf. Acesso em: 11 nov. 2025.

LIMA, Jéssica Valéria Gomes Vilar. **Ferramentas computacionais de apoio às aplicações de Cadeias de Markov na Engenharia de Produção**. 2024. 58 f. Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Produção) – Universidade Federal de Alagoas, Campus do Sertão, Delmiro Gouveia, AL, 2024.

LUO, Jiawen; KLEIN, Tony; JI, Qiang; HOU, Chenghan. Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. ***QMS Research Paper***, n. 2019/10. Belfast: Queen's University Belfast, Queen's Management School, 2019. DOI: 10.2139/ssrn.3435054. Disponível em: <https://doi.org/10.2139/ssrn.3435054>. Acesso em: 15 out. 2025.

MAARLEVELD, T. R.; OLIVIER, B. G.; BRUGGEMAN, F. J. StochPy: a comprehensive, user-friendly tool for simulating stochastic biological processes. **PLoS ONE**, v. 8, n. 11, p. e79345, 2013. DOI: 10.1371/journal.pone.0079345. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0079345>. Acesso em: 19 out. 2025.

MEHRDOUST, F.; NOORANI, I.; KANNIAINEN, J. Valuation of option price in commodity markets described by a Markov-switching model: a case study of WTI crude oil market. **Mathematics and Computers in Simulation**, v. 215, p. 228–269, 2024. Disponível em: [URL não fornecida]. Acesso em: 28 out. 2025.

MELLO, Fernando Dupin da Cunha; KUMAR, Prashant; NASCIMENTO, Erick G. Sperandio. Advancements in Soybean Price Forecasting: Impact of AI and Critical Research Gaps in Global Markets. **Economies**, v. 12, n. 11, p. 310, 2024. DOI: <https://doi.org/10.3390/economies12110310>. Acesso em: 28 out. 2025.

MICHAELIDES, Marie; MAILHOT, Mélina; LI, Yongkun. Probabilistic crop yields forecasts with spatio-temporal conditional copula using extreme weather covariates. Montréal: Concordia University, 2025. Disponível em: <https://arxiv.org/html/2503.22807v1>. Acesso em: 24 out. 2025.

MOHAMAD, A. A. A. Herding behaviour surrounding the Russo–Ukraine war and COVID-19 pandemic: evidence from energy, metal, livestock and grain *commodities*. **Review of Behavioral Finance**, v. 16, n. 5, p. 925–957, 2024. DOI: 10.1108/RBF-12-2023-0339. Disponível em: <https://doi.org/10.1108/RBF-12-2023-0339>. Acesso em: 28 out. 2025.

NEUFELD, Ariel; SESTER, Julian; ŠIKIĆ, Mario. Markov decision processes under model uncertainty. **Mathematical Finance**, [v., n., p.], 2023. DOI: 10.1111/mafi.12381. Disponível em: <https://doi.org/10.1111/mafi.12381>. Acesso em: 28 out. 2025.

NORRIS, J. R. **Markov Chains**. Cambridge, Reino Unido: Cambridge University Press, 1998. 237 p.

PINDYCK, Robert S.; ROTEMBERG, Julio J. The Excess Co-Movement of Commodity Prices. *The Economic Journal*, v. 100, n. 403, p. 1173-1189, dez. 1990. DOI: 10.2307/2233966. Acesso em: 28 out. 2025.

PORTO, Bernardo M. Previsão de preços das commodities agrícolas: uma revisão bibliométrica sobre modelos. *Revista GeSec*, São Paulo, v. 13, n. 3, p. 881-912, set./dez. 2022. DOI: <http://dx.doi.org/10.7769/gesec.v13i3.1380>. Acesso em: 28 out. 2025.

PRIVAUULT, Nicolas. **Understanding Markov chains: examples and applications**. 2. ed. Singapore: Springer, 2018.em: 11 out. 2025.

PUTERMAN, Martin L. **Markov Decision Processes: Discrete Stochastic Dynamic Programming**. Hoboken, NJ: John Wiley & Sons, 2005.

RICHTER, Martin Christian; SØRENSEN, Carsten. **Stochastic volatility and seasonality in commodity futures and options: the case of soybeans**. Copenhagen: Copenhagen Business School, Department of Finance, 2002. 45 p. (Working Papers, Department of Finance, Copenhagen Business School, No. 2002-4). Disponível em: <https://research-api.cbs.dk/pure/en/publications/stochastic-volatility-and-seasonality-in-commodity-futures-markets>. Acesso em: 1 out. 2025.

ROSS, Sheldon M. **Introduction to Probability Models**. Upper Saddle River, NJ: Prentice Hall, 1995.

STOKES, B. J. *et al.* A Markov chain model of crop conditions and intrayear crop yield forecasting. *Forecasting*, v. 5, n. 4, p. 747-763, 2023. DOI: 10.1002/for.3052. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/for.3052>. Acesso em: 28 out. 2025.

SYKORA, Nelson Danilo. **Preços de commodities agrícolas e o comportamento de mercado invertido (backwardation): o caso da soja**. 2013. Dissertação (Mestrado em Finanças e Economia Empresarial) – Escola de Pós-Graduação em Economia, Fundação

Getulio Vargas, Rio de Janeiro, 2013. Disponível em: <https://repositorio.fgv.br>. Acesso em: 28 out. 2025.

TAHA, Hamdy A. **Operations research: an introduction**. 8. ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2007. 811 p.

XU, C. The *asymptotic tails of limit distributions of continuous-time Markov chains*. **Advances in Applied Probability**, v. 56, n. 3, p. 707-738, 2024. Disponível em: <https://www.cambridge.org/core/journals/advances-in-applied-probability/article/asymptotic-tails-of-limit-distributions-of-continuous-time-markov-chains/A2126D5E4A476E5DFF1E94DE0D65ADAB>. Acesso em: 05 out. 2025.

XU, X.; CAI, S.; ZHANG, L. Dynamic influence of international price fluctuation on soybean market price in China: based on Bayesian-VAR model. **Frontiers in Sustainable Food Systems**, v. 9, n. 1594210, 2025. DOI: 10.3389/fsufs.2025.1594210. Disponível em: <https://doi.org/10.3389/fsufs.2025.1594210>. Acesso em: 28 out. 2025.

LI, N.; KER, A.; SAM, A.; ARADHYULA, S. **Modelling regime-dependent agricultural commodity price volatilities**. Guelph: Institute for the Advanced Study of Food and Agricultural Policy, Department of Food, Agricultural and Resource Economics, University of Guelph, 2016. (Institute White Paper, 2016.2).

ENGLE, Robert F. **Autoregressive conditional heteroscedasticity with estimates of the variance** of United Kingdom inflation. *Econometrica*, v. 50, n. 4, p. 987-1008, 1982. DOI: 10.2307/1912773. Disponível em: <https://www.jstor.org/stable/1912773>. Acesso em: 23 nov. 2025.

HAMILTON, James D. **A new approach to the economic analysis of nonstationary time series and the business cycle**. *Econometrica*, v. 57, n. 2, p. 357-384, 1989. DOI: 10.2307/1912559. Disponível em: <https://www.jstor.org/stable/1912559>. Acesso em: 23 nov. 2025.