APLICAÇÃO DE CADEIAS DE MARKOV E REGRESSÃO MÚLTIPLA LINEAR PARA ANÁLISE DOS ÍNDICES IBOVESPA, NASDAQ E NYSE

APPLICATION OF MARKOV CHAINS AND MULTIPLE LINEAR REGRESSION FOR ANALYSIS OF IBOVESPA, NASDAQ AND NYSE INDICES

Daniel Henrique Salvador, Leandro Corso

Universidade de Caxias do Sul, Campus Sede, R. Francisco Getúlio Vargas, 1130 - Petrópolis, Caxias do Sul - RS, 95070-560. E-mails: dhsalvador@ucs.br; llcorso@ucs.br

RESUMO

A bolsa de valores, apesar de dinâmica e sendo considerado um local de investimento de risco, é um meio capaz de render grandes ganhos financeiros a seus investidores. Nesse sentido, este artigo apresenta métodos desenvolvidos através dos modelos matemáticos de Cadeias de Markov Multivariadas (CMM) e de Regressão Linear Múltipla (RLM), para prever oscilações no índice IBOVESPA em função dos índices das bolsas americanas NYSE e NASDAQ. A fim de maximizar os ganhos e diminuir perdas na compra e venda de ações em carteira na bolsa de valores brasileira, utilizou-se uma base de dados de variação diária dos índices contemplando os anos de 2019 e 2020, que posteriormente foram aplicados para desenvolver as matrizes de transição de CMM e a equação linear de RLM. Então, através dos modelos propostos, foi avaliada a precisão de cada método, assim como uma simulação de retorno financeiro obtido com base nos resultados previstos. Constatou-se que o método de Cadeia de Markov Multivariada obteve maior precisão e maior lucro simulado, com uma média de acertos superior a 50% e 19% de retorno financeiro em todo o período delimitado.
Palavras-chave: Cadeia de Markov Multivariada. Regressão Linear Múltipla. Pesquisa Operacional. Bolsa de Valores. Investimento.

ABSTRACT

The stock exchange, despite being dynamic and considered to be a risky asset, is capable way of yielding great financial gains to its investors. In this sense, this article presents methods developed through the mathematical models of Multivariate Markov Chains (CMM) and Multiple Linear Regression (MRL), to predict oscillations in the IBOVESPA index as a function of the indices of the American stock exchanges NYSE and NASDAQ. In order to maximize gains and reduce losses in the purchase and sale of shares held on the Brazilian stock exchange, a database of daily variation of the indices was used, covering the years 2019 and 2020, which were later applied to develop the transition matrices of CMM and the linear equation of RLM. Then, through the proposed models, the precision of each method was evaluated, as well as a simulation of financial return obtained based on the predicted results. It was found that the Multivariate Markov Chain method obtained greater precision and greater simulated profit, with an average of hits greater than 50% and 19% of financial return throughout the delimited period.
Keywords: Multivariate Markov Chain. Multiple Linear Regression. Operational Research. Stock Exchange. Investment.

1. INTRODUÇÃO

À medida que a tecnologia e o fácil acesso à informação vêm se expandindo a nível global, o setor financeiro encontra-se em ascensão e assume uma posição fundamental dentro da economia e da sociedade (BANCO CENTRAL DO BRASIL, 2016). Entretanto, investir corretamente pode parecer um desafio em razão da volatilidade do mercado de ações, já que esse possui um perfil com aspectos que aparentam ser aleatórios ou influenciados por variáveis de diferentes formas, sejam elas econômicas, políticas, geográficas ou sociais.

Atualmente, a bolsa de valores brasileira possui cerca de 50% da sua participação concentrada na mão de estrangeiros (IBOVESPA, 2021), portanto ainda se encontra em uma situação em que a movimentação da economia externa exerce forte influência sobre seus índices. A busca pela formulação de um modelo capaz de prever os impactos causados pelas oscilações externas no mercado de ações, move investidores a criarem ferramentas e métodos matemáticos com o objetivo de diminuírem as incertezas e riscos em seus investimentos (GIACOMEL, 2016). Os autores Franco *et al.* (2021) descreveram, em sua publicação, a análise da relação entre as ações da Petrobrás, cotação do dólar e petróleo WTI, por meio da utilização de Cadeias de Markov Multivariadas, obtendo dessa forma os estados probabilísticos a serem utilizados para a previsão acurada desses dados.

As Cadeias de Markov podem ser definidas como um processo estocástico, que tem como premissa encontrar a distribuição de probabilidade para eventos futuros baseados apenas no seu estado presente, sem que sejam considerados os eventos anteriores do processo analisado (RODRIGUES, 2017). Os modelos Markovianos possuem aplicações em diversas áreas, como em ciências sociais, biológicas e administrativas (GRIGOLETTI, 2015).

O método da Regressão Linear Múltipla, no entanto, tem como objetivo formular uma equação baseada no histórico de uma variável dependente em função de uma ou mais variáveis independentes (MONTGOMERY; RUNGER, 2016). A utilização da Regressão Linear Múltipla se mostra adequada em situações em que se deseja compreender os efeitos de uma ou mais variáveis independentes, sobre uma variável dependente, onde o número de variáveis utilizadas tem influência positiva sobre a acuracidade da previsão (HORNGREN; FOSTER; DATAR, 1997).

O objetivo do presente estudo consiste em aplicar os métodos de Cadeias de Markov Multivariadas (CMM) e de Regressão Linear Múltipla (RLM), sobre os índices das bolsas de valores brasileira (IBOVESPA) e norte-americanas (NASDAQ e NYSE), durante os anos de 2019 e 2020, de forma que seja possível analisar e determinar a correlação entre os indicadores por meio dos valores estipulados pelos modelos de previsão, bem como compará-los entre si, visando buscar a técnica mais precisa a ser utilizada para amenizar perdas de capital ao se investir no mercado financeiro.

2. REFERENCIAL TEÓRICO

2.1. Mercados Financeiros

As movimentações dentro do mercado financeiro podem ser divididas em curto ou longo prazo, em função do objetivo do investidor. A avaliação das ações de longo prazo consiste na utilização de conhecimentos macroeconômicos, geopolíticos, índices financeiros e de balanços contábeis para determinar a taxa de retorno dos investimentos através de dividendos e no preço das ações projetadas no futuro; esta é chamada de análise fundamentalista.

Conforme Lemos (2018), a análise técnica, baseada em gráficos e modelos matemáticos, está melhor apta a atender as transações de curto prazo, tendo em vista que seu objetivo é fazer com que o investidor preveja as oscilações de preços das ações para determinar o melhor momento de comprá-las ou vende-las, dessa forma obtendo lucro na operação.

Vasanthi *et al.* (2011) demonstraram a eficiência da utilização de cadeias de Markov como modelo matemático aplicado na bolsa de valores da Índia. Dentro da mesma linha de estudo, Guruaj *et al.* (2019) apresentaram através de seu artigo modelos de previsão para variação de preços nas ações da Coca-Cola em bolsa de valores baseados no método de Regressão Linear Múltipla.

2.2. Processos Estocásticos e Cadeias de Markov (CM)

De acordo com Hillier (2013), processos estocásticos podem ser definidos como um conjunto de variáveis capazes de assumirem diferentes valores mensuráveis, dado um determinado índice correspondente ao tempo. Ao contrário de um processo determinístico, o processo estocástico é capaz de assumir valores aleatórios dependentes apenas de seu estado atual, mesmo que sua condição inicial seja conhecida. Essa condição está relacionada às áreas onde existe maior grau de incerteza sobre os parâmetros estudados para definição estatística das variáveis.

Os processos de Markov são processos estocásticos e podem ser definidos por estados X, ao decorrer do tempo *t*, onde $P(X_t)$ indica a probabilidade de um estado determinado em um intervalo de tempo. Essa probabilidade é representada através da Equação 1.

$$P\{X_{t+1} = j \,|X_0 = k_0, X_1 = k_1, \dots, X_{t-1} = k_{t-1} = i\} = P\{X_{t+1} = j|X_t = i\} \tag{1}$$

Taha (2011) explica que as probabilidades em um processo Markoviano possuem transição estacionária, o que significa que não sofrem alterações no decorrer do tempo, considerando os estados mudando em um intervalo de tempo de uma etapa ou de *n* etapas, como descrito nas Equações 2 e 3.

$$p_{ij} = P\{X_{t+1} = j|X_t = i\} \tag{2}$$
$$p_{ij}^{n} = P\{X_{t+n} = j|X_t = i\} \tag{3}$$

As probabilidades de transição entre estados podem ser descritas na forma matricial, onde as linhas representam vetores de probabilidade e seus elementos assumem valores de 0 até 1 ($0 \leq p_{ij} \leq 1$) e $\sum p_{ij} = 1$, onde $i = 1, 2, 3,4, \dots, n$. Conforme pode ser visualizado na Equação 4.

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1j} \\ \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ij} \end{bmatrix} \tag{4}$$

A Cadeia de Markov chega em seu estado irredutível quando os vetores de probabilidades entre estados convergem ao longo do tempo, tornando-se aperiódica ou também conhecida como ergódica (BARBOSA, 2009). Para *n* passos calculados, a matriz é considerada estável se respeitar às Equações 5 e 6, demonstradas a seguir.

$$\pi_j = \lim_{n\to\infty} a_j^{(n)} \tag{5}$$
$$\pi_j = \sum_{i=0}^{M} \pi_i p_{ij} \tag{6}$$

Onde $j = 1, 2, 3, \dots, M$

O estado de equilíbrio da matriz é então definido através da Equação 7, onde não há alterações nas probabilidades de transição.

$$\sum_{j=0}^{M} \pi_j = 1 \tag{7}$$

Através da matriz estável de transição é possível determinar o tempo médio de recorrência para o sistema voltar a um estado *j* anterior, obtendo o número de passos para chegar ao estado inicial, conforme descrito na Equação 8.

$$\mu_{ii} = \frac{1}{\pi_i} \tag{8}$$

Onde $i = 1, 2, 3, 4, \dots, M$

2.3. Cadeias de Markov Multivariáveis (CMM)

O processo multivariado de Cadeias de Markov é capaz de modelar as probabilidades do estado de um sistema baseado em múltiplos segmentos de dados históricos. Para Ching *et al.* (2008), para um único intervalo de tempo, existem s sequências categóricas e cada uma dessas possui m estados diferentes, onde a distribuição de probabilidade para j no intervalo $t = r + 1$ depende do estado atual $t = r$ de todas as sequências, incluindo a si. Esta relação pode ser definida através da Equação 9.

$$X_{r+1}^{(j)} = \sum_{k=1}^{s} \sum_{h=1}^{n} \lambda_{jk}^{(h)} P_h^{(jk)} X_{r-h+1}^{(k)}, j = 1, 2, 3, \ldots, s; r = n-1, n, \ldots \quad (9)$$

Onde X_r representa a probabilidade de transição na sequência j, λ_{jk} representa números reais não-negativos de soma igual a 1, P_h é o passo h da matriz de transição.

A probabilidade dos estados de um sistema da sequência $X_{r+1}^{(j)}$ no intervalo de tempo $t = r + 1$ é calculado a partir da média ponderada da probabilidade P^{jk} e $x_r^{(k)}$, onde a k-ésima sequência representa um passo de tempo t, para o próximo passo na j-ésima sequência e $x_r^{(k)}$, conforme Ching *et al.* (2008). Assim apresentado na forma matricial pela Equação 10.

$$X_{r+1} = \begin{pmatrix} X_{r+1}^{(1)} \\ X_{r+1}^{(2)} \\ \vdots \\ X_{r+1}^{(s)} \end{pmatrix} = \begin{pmatrix} \lambda_{11}P^{(11)} & \lambda_{12}P^{(12)} & \ldots & \lambda_{1s}P^{(1s)} \\ \lambda_{21}P^{(21)} & \lambda_{22}P^{(22)} & \ldots & \lambda_{2s}P^{(2s)} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{s1}P^{(s1)} & \lambda_{s2}P^{(s2)} & \cdots & \lambda_{ss}P^{(ss)} \end{pmatrix} \begin{pmatrix} X_r^{(1)} \\ X_r^{(2)} \\ \vdots \\ X_r^{(s)} \end{pmatrix} \quad (10)$$

2.4. Regressão Linear Múltipla

A Regressão Linear Múltipla (RLM) é um modelo matemático probabilístico que consiste em sua formulação a utilização de duas ou mais variáveis independentes $x_1, x_2, \ldots, x_k$ que resultam em uma variável resposta y. Segundo Devore (2018), a análise de regressão descreve a relação entre as variáveis independentes, correlacionando-as com a variável dependente, como demonstrado na Equação 11.

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_k X_{kj} + \epsilon_j \quad (11)$$

Onde Y_j representa a variável resposta, β_i os coeficientes parciais de regressão estimados, X_{ij} as variáveis explicativas ou independentes e ϵ_i são os erros aleatórios estimados calculados a partir da variância σ^2 da amostra. Para Montgomery e Runger (2016), é esperado uma variação na variável resposta Y_j para qualquer variação unitária em X_j quando o restante das variáveis regressoras $X_i (i \neq j)$ se mantiverem constantes.

3. METODOLOGIA

Com o objetivo de mensurar a acuracidade dos modelos matemáticos de previsões para os índices das bolsas de valores apresentadas no estudo, a metodologia foi dividida na seguinte sequência, apresentada na Figura 1. A seguir, cada tópico é aportado, detalhando os passos seguidos.

Figura 1 – Etapas da metodologia
Fonte: Elaborado pelos autores (2021)

3.1. Coleta de dados

No presente artigo, foram utilizados os dados históricos das bolsas de valores de São Paulo (IBOVESPA) e de Nova Iorque (NASDAQ e NYSE), durante os anos de 2019 e 2020, considerando a variação dos índices no momento do fechamento diário do mercado de ações.

Os dados foram obtidos no endereço eletrônico *investing.com* (2021) e organizados de forma que contemplem a data do fechamento das bolsas e a variação diária dos índices. O Quadro 1 apresenta uma demonstração dos dados coletados.

Quadro 1 – Demonstração da base de dados

	NASDAQ	**IBOVESPA**	**NYSE**
DATA	**VARIAÇÃO DIÁRIA (%)**		
01/04/2019	1,35%	-1,34%	1,10%
02/04/2019	0,28%	-3,10%	-0,13%
03/04/2019	0,60%	-0,08%	0,11%
01/04/2020	-4,19%	3,67%	-4,44%
02/04/2020	2,00%	-5,51%	2,21%
03/04/2020	-1,41%	1,65%	-1,81%

Fonte: Elaborado pelos autores (2021)

3.2. Organização dos dados

Em função do funcionamento dos mercados de valores ocorrerem apenas em dias úteis, os dados foram manipulados para que fossem desconsiderados da base os feriados existentes no Brasil e Estados Unidos, onde não existe registro em comum para a variação diária dos mercados entre os dois países. No total, foram contabilizados 244 dias para o ano de 2020 e 242 dias no ano de 2019. A organização desses dados foi realizada utilizando o software *Excel*, assim como os gráficos apresentados a seguir.

A Figura 2 apresenta um gráfico onde estão descritos a variação das três bolsas de valores durante o ano de 2019.

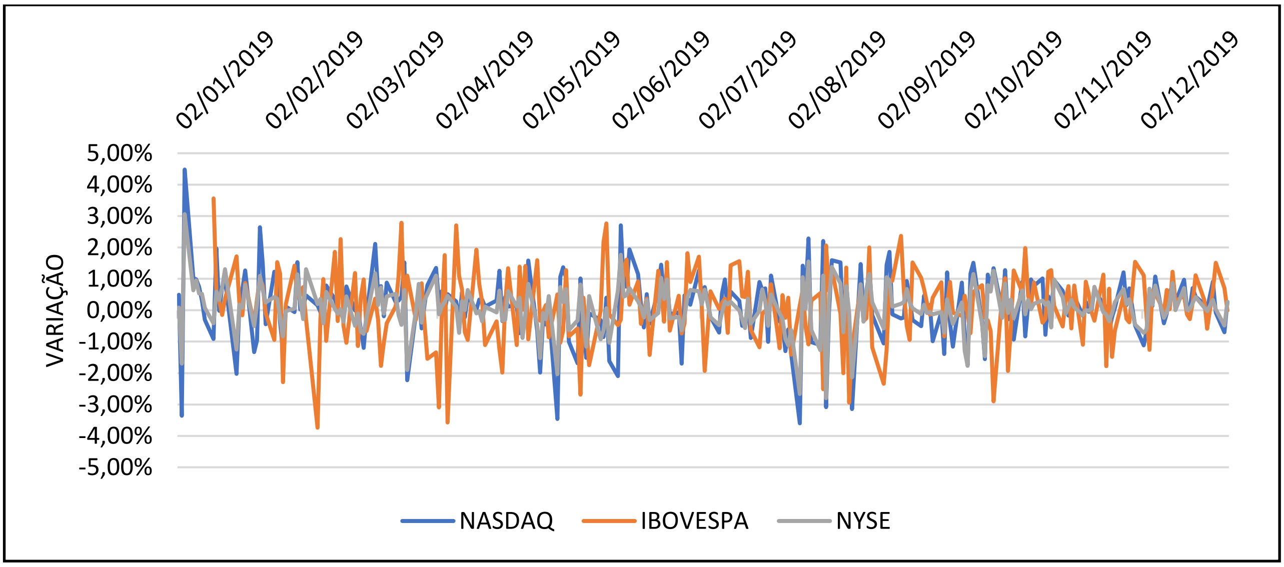

Figura 2 – Variação Percentual dos índices em 2019
Fonte: Elaborado pelos autores (2021)

A Figura 3 mostra as variações percentuais dos índices das bolsas de valores durante o ano de 2020.

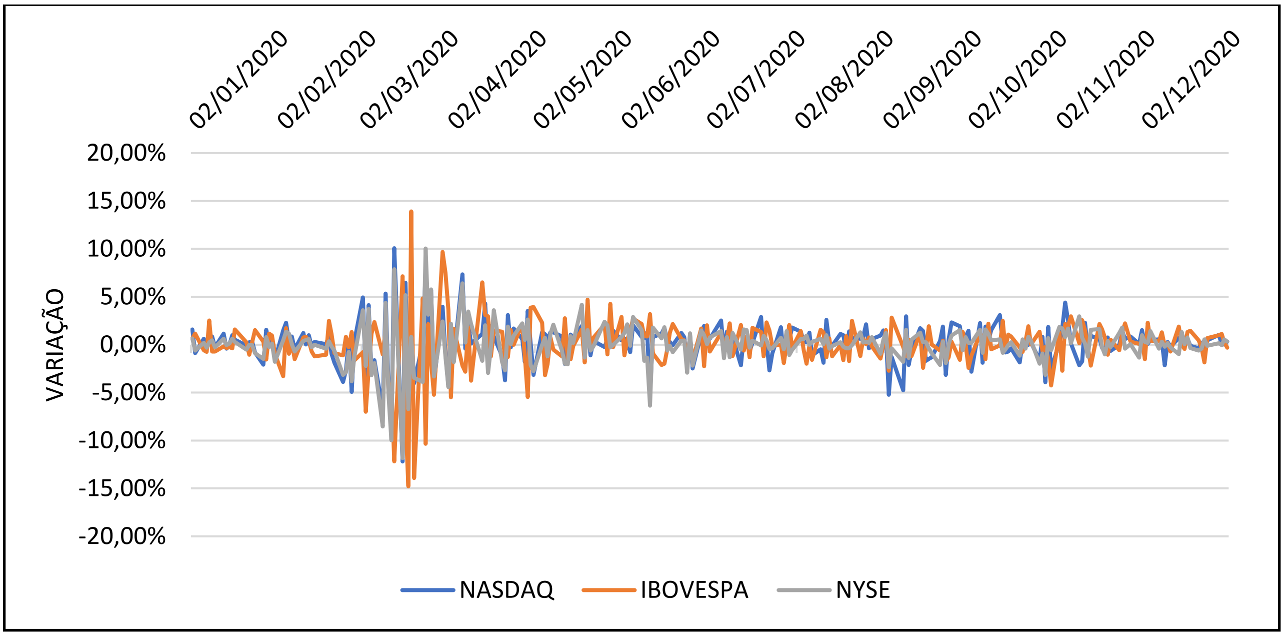

Figura 3 – Variação Percentual dos índices em 2020
Fonte: Elaborado pelos autores (2021)

3.3. Aplicação de Cadeias de Markov Multivariadas

Visando prever as variações do índice IBOVESPA, em decorrência de seu histórico e dos índices NYSE e NASQAD, a aplicação de Cadeias de Markov Multivariadas se dá nas seguintes etapas:

1. Definição dos intervalos de variações para a composição da matriz de frequência, de maneira que se mantenha uma alta dispersão de dados para cada intervalo, não limitando as informações a serem utilizadas posteriormente. Os intervalos definidos para este estudo estão apresentados no Quadro 2.

Quadro 2 – Intervalo de variações

A	< -1,5%
B	de -1,5% a -1,01%
C	de -1% a -0,01%
D	de 0% a 1%
E	de 1,01% a 1,5%
F	> 1,5%

Fonte: Elaborado pelos autores (2021)

2. A partir dos intervalos definidos, a matriz de frequência é aplicada sobre os dados de variação de índices coletados. A matriz demonstra a periodicidade em que ocorre cada transição da variação um índice para outro. No caso do estudo abordado, foram criadas as matrizes De NASDAQ/Para IBOVESPA e De NYSE/Para IBOVESPA, contabilizando o somatório da quantidade de vezes em que os índices transitavam de uma faixa estabelecida no intervalo para outra em um passo de tempo de um dia ($n = 1$).
3. A partir da matriz de frequência, é obtida a matriz de transição, ao ser efetuada a divisão da quantidade de vezes da mudança de estado pelo somatório da respectiva linha dos elementos dentro da matriz, desse modo resultando em uma matriz contemplando as probabilidades de ocorrência de cada transição de estado.

4. A matriz de probabilidade é então utilizada como modelo preditivo para as oscilações dos índices da bolsa de valores, e esses valores são comparados ao histórico real durante os anos de 2019 e 2020, possibilitando que seja verificada a acuracidade da previsão para o método CMM.

3.4. Aplicação de Regressão Linear Múltipla

O método de Regressão Linear Múltipla tem como objetivo determinar o valor de uma variável resposta por meio de uma função linear constituída por variáveis independentes. A variável dependente do método consiste no índice IBOVESPA no período $t+1$, a partir das variáveis independentes IBOVESPA, NASDAQ e NYSE no período t.

A partir dos coeficientes de regressão definidos para a função, é possível determinar a previsão para o índice IBOVESPA um dia a frente do tempo, baseado no histórico de variação dos índices da própria IBOVESPA e das bolsas de valores americanas.

Para esse processo, foi utilizada a ferramenta de Análise de Dados do software *Excel*. Por meio da disponibilização dos dados coletados, o programa analisa as variáveis resposta em função das variáveis independentes regressoras do período anterior, retornando à equação linear utilizada para prever os índices no dia seguinte, com coeficiente de correção, bem como a série de dados previstos já calculada a partir da função.

3.5. Aplicabilidade dos métodos

Para cada método proposto, foram realizadas simulações nas operações na IBOVESPA, considerando o investimento ou venda de ações de acordo com a previsão para os índices. Dessa maneira, foi possível verificar o retorno financeiro para cada método, analisando a eficácia de cada um ao ser comparado parcialmente com os dados coletados.

Nas técnicas aplicadas, foi considerado o período atual t para a variação dos índices das bolsas de valores, buscando o ganho previsto no período $t+1$, comprando ou vendendo a ação nas condições propostas nos fluxos apresentados pela Figura 4.

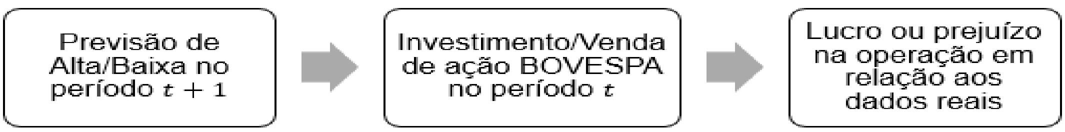

Figura 4 – Fluxo da simulação para previsão do índice
Fonte: Elaborado pelos autores (2021)

Desta forma, foi possível avaliar e comparar os rendimentos totais simulados para cada método aplicado no fim do período de 2019 e no 2020, obtendo a técnica mais apropriada para o caso abordado no estudo.

4. RESULTADOS

Com base nos métodos apresentados nos capítulos anteriores, foram determinados os resultados das previsões dos modelos de Cadeias de Markov Multivariadas e de Regressão Linear Múltipla, a fim de comparar a precisão e aplicabilidade de cada método. Os resultados obtidos são apresentados a seguir.

4.1. Cadeias de Markov Multivariadas – resultados

A seguir estão demonstradas as matrizes de transição para o índice IBOVESPA em relação aos índices NASDAQ e NYSE, durante os anos de 2019 e 2020.

A Figura 5 apresenta a frequência das ocorrências entre as transições definidas nos intervalos A até F, onde a coluna "De" representa a variação do índice NASDAQ em 2019 dentro do intervalo estipulado no tempo t, e a linha "Para" está associada à variação do índice IBOVESPA em 2019 no tempo $t + 1$.

De \ Para	IBOVESPA					
NASDAQ	**A**	**B**	**C**	**D**	**E**	**F**
A	0,000	0,000	0,500	0,286	0,143	0,071
B	0,154	0,231	0,000	0,385	0,154	0,077
C	0,053	0,040	0,333	0,333	0,107	0,133
D	0,059	0,119	0,257	0,396	0,089	0,079
E	0,083	0,042	0,417	0,250	0,167	0,042
F	0,200	0,000	0,400	0,200	0,067	0,133

Figura 5 – Matriz de transição NASDAQ x IBOVESPA 2019
Fonte: Elaborado pelos autores (2022)

A Figura 6 apresenta a frequência das ocorrências entre as transições definidas nos intervalos A até F, onde a coluna "De" representa a variação do índice NYSE em 2019 dentro do intervalo estipulado no tempo t, e a linha "Para" está associada à variação do índice IBOVESPA em 2019 no tempo $t + 1$.

De \ Para	IBOVESPA					
NYSE	**A**	**B**	**C**	**D**	**E**	**F**
A	0,000	0,000	0,429	0,429	0,143	0,000
B	0,000	0,000	0,714	0,143	0,000	0,143
C	0,090	0,079	0,292	0,360	0,090	0,090
D	0,049	0,089	0,293	0,341	0,130	0,098
E	0,214	0,071	0,286	0,286	0,071	0,071
F	0,000	0,000	0,000	0,500	0,000	0,500

Figura 6 – Matriz de transição NYSE x IBOVESPA 2019
Fonte: Elaborado pelos autores (2022)

A Figura 7 apresenta a frequência das ocorrências entre as transições definidas nos intervalos A até F, onde a coluna "De" representa a variação do índice NASDAQ em 2020 dentro do intervalo estipulado no tempo t, e a linha "Para" está associada à variação do índice IBOVESPA em 2020 no tempo $t + 1$.

De \ Para	IBOVESPA					
NASDAQ	**A**	**B**	**C**	**D**	**E**	**F**
A	0,293	0,024	0,171	0,195	0,098	0,220
B	0,167	0,167	0,167	0,083	0,250	0,167
C	0,200	0,125	0,200	0,175	0,125	0,175
D	0,145	0,096	0,205	0,241	0,072	0,241
E	0,074	0,222	0,148	0,185	0,111	0,259
F	0,167	0,104	0,208	0,146	0,083	0,292

Figura 7 – Matriz de transição NASDAQ x IBOVESPA 2020
Fonte: Elaborado pelos autores (2022)

A Figura 8 apresenta a frequência das ocorrências entre as transições definidas nos intervalos A até F, onde a coluna “De” representa a variação do índice NYSE em 2020 dentro do intervalo estipulado no tempo t, e a linha “Para” está associada à variação do índice IBOVESPA em 2020 no tempo $t + 1$.

De \ Para	**IBOVESPA**					
NYSE	**A**	**B**	**C**	**D**	**E**	**F**
A	0,371	0,057	0,143	0,057	0,086	0,286
B	0,214	0,000	0,214	0,143	0,214	0,214
C	0,141	0,141	0,234	0,219	0,094	0,172
D	0,113	0,138	0,213	0,263	0,113	0,163
E	0,059	0,059	0,235	0,176	0,176	0,294
F	0,220	0,098	0,098	0,146	0,024	0,415

Figura 8 – Matriz de transição NYSE x IBOVESPA 2020
Fonte: Elaborado pelos autores (2022)

4.2. Regressão Linear Múltipla – resultados

Por meio da RLM, foram calculados os coeficientes de regressão e intersecção das equações, para que fossem determinadas as previsões das variações futuras dos índices das bolsas de valores nos anos de 2019 e 2020.

O Quadro 3 apresentado a seguir descreve os valores dos coeficientes encontrados em cada ano.

Quadro 3 – Coeficientes de Regressão

COEFICIENTES	**2019**	**2020**
Interseção	0,076	-0,070
NYSE	-0,427	-0,004
NASDAQ	0,338	0,073

Fonte: Elaborado pelos autores (2022)

Com base nos coeficientes calculados, foi possível criar as equações para determinar a variação no índice IBOVESPA em função das bolsas americanas, dessa forma criando um cenário simulado de oscilação na IBOVESPA durante os anos de 2019 e 2020, como apresentado nas Figuras 9 e 10.

A simulação apresenta os dados históricos do índice IBOVESPA no tempo $t + 1$ comparado aos dados calculados pelo método de Regressão Linear para o tempo t em função das variações dos índices NASDAQ e NYSE.

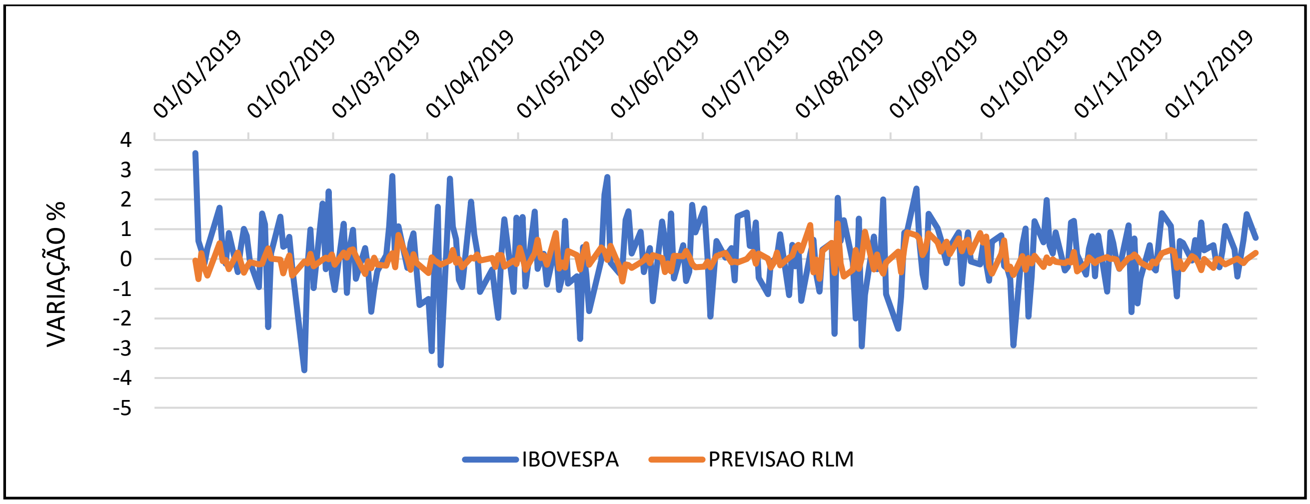

Figura 9 – Simulação RLM em 2019
Fonte: Elaborado pelos autores (2022)

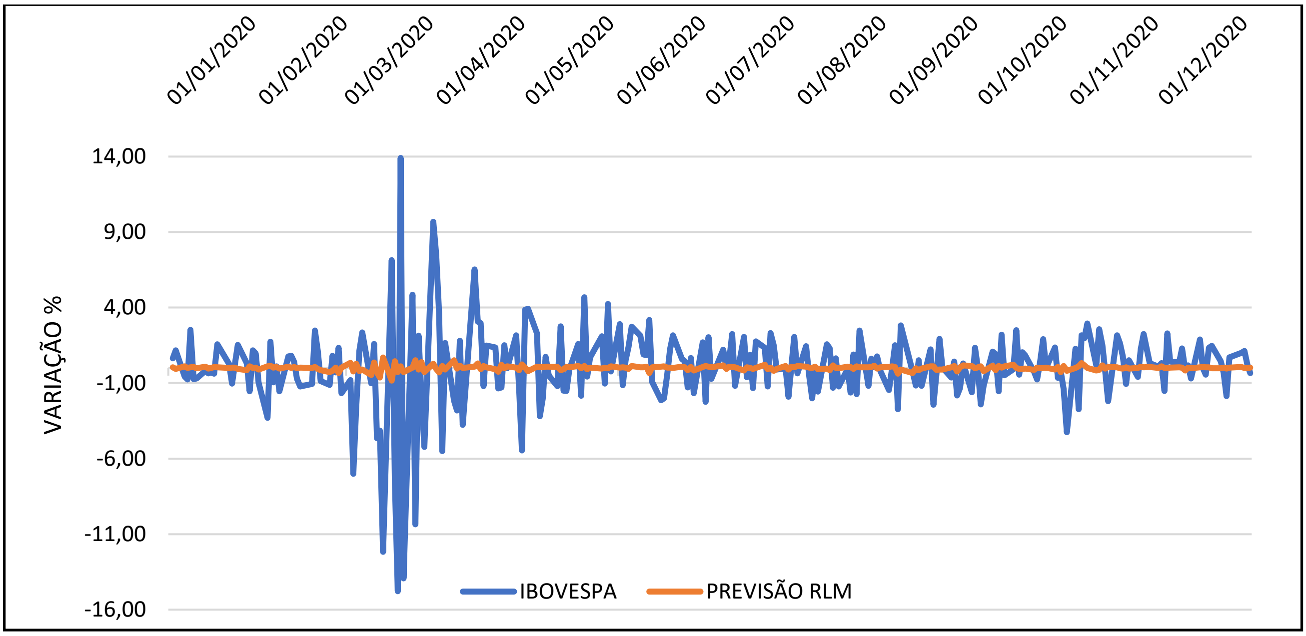

Figura 10 – Simulação RLM em 2020
Fonte: Elaborado pelos autores (2022)

Aqui se pode verificar que o comportamento dos dados apresentou diferenças importantes a serem consideradas entre os 2 períodos analisados. Este fato fica evidenciado uma vez que os coeficientes apresentados pelos modelos de regressão para NYSE e NASDAQ para os períodos de 2019 e 2020 sejam bastante diferentes.

4.3. Retornos financeiros nos cenários simulados

Em um contexto financeiro, a partir das simulações dos modelos apresentadas, foi possível medir avaliar a precisão de acertos de cada método levando em consideração como correta a previsão no caso de elevação ou queda do índice IBOVESPA no dia futuro. Foi simulada a venda de uma ação em carteira no tempo presente t em uma queda prevista no dia futuro $t + 1$; ou a compra de uma ação no tempo presente t no caso de uma variação positiva prevista para o dia futuro $t + 1$. Ou seja, apenas considerando o sentido da variação do índice.

Esta análise foi realizada durante os meses de outubro, novembro e dezembro nos anos de 2019 e 2020. O período de tempo foi escolhido de forma arbitrária, sem qualquer aspecto lógico intrínseco, apenas para possibilitar a análise dos métodos neste cenário. O Quadro 4 apresenta a precisão dos modelos de previsão considerando apenas o aspecto financeiro do estudo por meio do método de CMM.

Quadro 4 – Acertos no sentido da variação pelo método CMM em 2019 e 2020

	2019		**2020**	
	NASDAQ	**NYSE**	**NASDAQ**	**NYSE**
QTD. DADOS	64	64	64	64
QTD. ACERTOS	38	36	37	36
% ACERTO	59,38%	56,25%	57,81%	56,25%

Fonte: Elaborado pelos autores (2022)

O Quadro 5 apresenta a precisão na compra e venda de ações em carteira pertencentes ao índice IBOVESPA previstos por meio do modelo de RLM.

Quadro 5 – Acertos no sentido da variação pelo método de RLM em 2019 e 2020

RLM	2019	2020
QTD. DADOS	64	64
QTD. ACERTOS	25	34
% ACERTO	39,06%	53,13%

Fonte: Elaborado pelos autores (2022)

O Quadro 6 apresenta os ganhos financeiros simulados de ambos os métodos nos anos de 2019 e 2020, baseado em um investimento inicial, com compra e vendas de ações em carteira baseada nas previsões e alterada pelas variações dos índices reais nos meses de outubro, novembro e dezembro.

Quadro 6 – Ganhos financeiros simulados por método em out/nov/dez de 2019 e 2020

GANHOS FINANCEIROS		
MÉTODO	2019	2020
CMM/NASDAQ	12,74%	19,06%
CMM/NYSE	8,08%	15,65%
RLM	-10,92%	5,68%

Fonte: Elaborado pelos autores (2022)

No fim do exercício, foi medido um lucro positivo simulado na operação de compra e venda de ações, que resultou em até 19% de retorno financeiro utilizando o modelo de CMM.

5. CONCLUSÃO

Este estudo teve como principal objetivo demonstrar a aplicabilidade de métodos de previsão matemáticos no ambiente financeiro, que mesmo inserido em um cenário de incertezas e especulações, possui padrões a serem traçados por meio de ferramentas de forma que maximizem os lucros de investidores voltados às operações de renda variável.

Utilizando o método Markoviano, foi obtida uma precisão de acerto de 59% e 57% para NASDAQ em relação a IBOVESPA no último trimestre de 2019 e 2020 respectivamente. Para a relação entre NYSE e IBOVESPA, manteve-se em 56% de precisão nos dois anos. Utilizando a Regressão Linear, a precisão foi de 39% no último trimestre de 2019 e 53% em 2020. Também se pode constatar que o período de pandemia afetou o comportamento dos dados, o que corrobora com essa diferença. Fica destacado no gráfico o comportamento relativamente diferente no mês 03/2020. Este fato sugere que os modelos de regressão possam apresentar resultados melhores se forem para períodos menores, para isso, novos testes podem ser sugeridos neste tópico.

Foi possível observar que, apesar das oscilações nas variações dos índices que diminuem a precisão dos cálculos, a taxa de acerto das previsões manteve-se positiva na maior parte das aplicações. As cadeias de Markov demonstram ser uma ferramenta mais precisa em relação ao método de Regressão Linear no contexto aplicado, tendo em vista que nas CMM existe a relação distinta entre cada uma das bolsas de valores americanas com a IBOVESPA, sem que fossem correlacionadas na mesma equação, como foi feito no método de regressão, o que tornou a previsão com uma razão menor que 50% de acertos no ano de 2019.

É importante levar em consideração que modelos matemáticos estão limitados apenas à dados quantitativos para suas análises, além do presente estudo estar reservado à apenas três bolsas de valores no mundo, entre milhares de índices existentes que podem vir a influenciar uns aos outros. Para estudos futuros, podem ser feitas análises mais minuciosas por meio de uma maior base de dados, abrangendo outros ativos ao redor do mundo e em maiores intervalos de tempo, considerando também uma relação de mercado pré-existente para utilização da metodologia de Regressão Linear, como o dólar e o petróleo, por exemplo. Dessa forma, seria possível determinar com maior precisão dados futuros baseados em informações do passado e presente.

Por fim, o objetivo proposto do estudo foi atingido de forma que comprovou a eficácia do modelo de previsão matemático de Cadeias de Markov para direcionamento e tomada de decisão no mercado financeiro, podendo ser utilizada como parte de um planejamento em conjunto com outras ferramentas e do olhar analítico de um investidor sobre o cenário global econômico para maximizar seus lucros.

6. REFERÊNCIAS BIBLIOGRÁFICAS

BANCO CENTRAL DO BRASIL, 2016. **Evolução do Sistema Financeiro Nacional**. Disponível em: <http://www.bcb.gov.br/htms/deorf/r199812/texto.asp?idpai=revsfn199812>.

BARBOSA, Helenice Lopes. **Métodos estatísticos em Cadeias de Markov**. 2009. 49 f. Dissertação (Mestrado em Probabilidade e Estatística; Modelagem Matemática) – Universidade Federal do Rio Grande do Norte, Natal, 2009.

CHING, Wai-Ki; NG, Michael K.; FUNG, Eric S. Higher-order multivariate Markov chains and their applications. **Linear Algebra and its Applications**, v. 428, n. 2-3, p. 492-507, 2008.

DEVORE, Jay L. **Probabilidade e estatística para engenharia e ciências**; tradução da 9ª edição norte-americana. São Paulo: Cengage Learning Brasil, 2018.

FRANCO, Mateus M.; CECHIN, Rafaela B.; PASOLINI, Mônica; MONEGAT, Amanda D. R.; CORSO, Leandro L. **Cadeia de Markov Multivariada aplicada na bolsa de valores utilizando dados da Petrobrás, dólar e petróleo WTI**. ConTexto, Porto Alegre, v. 21, n. 47, p. 79-94, jan./abr. 2021.

GIACOMEL, Felipe dos Santos. **Um método algorítmico para operações na bolsa de valores baseado em ensembles de redes neurais para modelar e prever os movimentos dos mercados de ações**. 2016. 92 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

GRIGOLETTI, Pablo. **Cadeias de Markov**. 2015. Artigo para Escola de Informática – Universidade Católica de Pelotas (UCPel).

GURURAJ, Vaishnavi; SHRIYA, V. R.; ASHWINI, K. **Stock market prediction using linear regression and support vector machines**. International Journal of Applied Engineering Research, v. 14, n. 8, p. 1931-1934, 2019.

HILLIER, Frederick S.; LIEBERMAN, Gerald J. **Introdução à pesquisa operacional**. 9. ed. Porto Alegre: Amgh, 2013.

HORNGREN, Charles T., FOSTER, George, DATAR, Srikant M. **Contabilidade de Custos**; tradução José Luiz Paravato. 9.ed. Rio de Janeiro: LTC, 1997.

IBOVESPA, 2021. **Participação dos Investidores**. Disponível em: <https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/participacao-dos-investidores/volume-total-acumulado/>.

INVESTING, 2021. **Site institucional**. Disponível em: <https://br.investing.com/?ref=www>.

LEMOS, Flávio. **Análise técnica dos mercados financeiros**. São Paulo; Saraiva, 2018.

MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. 6. ed. Rio de Janeiro: LTC, 2016.

RODRIGUES, R. **Pesquisa Operacional**. Porto Alegre: Grupo A, 2017.

VASANTHI, S.; SUBHA, M. V.; NAMBI, S. Thirupparkadal. An empirical study on stock index trend prediction using markov chain analysis. **Journal of Banking Financial Services and Insurance Research**, v. 1, n. 1, p. 72-91, 2011.

TAHA, Hamdy A. **Operations research**: an introduction. Upper Saddle River: Pearson/Prentice Hall, 2011.