



CONEPA
CONGRESSO NACIONAL DE ESTUDANTES
E PROFISSIONAIS DE ADMINISTRAÇÃO

10ª Edição 2024 | 22 e 23 de novembro de 2024

João Pessoa, Paraíba (Região Nordeste)

PREVISÃO DO ÍNDICE BOVESPA: APLICAÇÃO DO MODELO AUTO-ARIMA NA ANÁLISE DE SÉRIES TEMPORAIS DE CURTO PRAZO

Diogo Daniel Bandeira de Albuquerque¹
Doutorando em Economia
Universidade Estadual de Montes Claros
diogo.albuquerque@unimontes.br

¹ Agradeço o apoio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) BPG-00068-23.

Resumo

Este estudo apresenta uma análise preditiva do Índice Bovespa (IBOV), utilizando o modelo ARIMA e sua variante automatizada, o Auto-ARIMA, para prever preços de fechamento no curto prazo. A pesquisa abrange dados coletados de janeiro de 2010 a maio de 2024, aplicando um fluxo automatizado para seleção de parâmetros do modelo. Os resultados mostraram que o Auto-ARIMA conseguiu capturar a direção geral das tendências, embora tenha apresentado limitações em prever mudanças abruptas e alta volatilidade. O Erro Quadrático Médio (RMSE) foi utilizado como métrica de avaliação do modelo, indicando a necessidade de ajustes adicionais ou de incorporação de variáveis exógenas para melhorar a precisão em cenários mais voláteis.

Palavras-chave: Índice Bovespa, ARIMA, Auto-ARIMA, previsão de séries temporais, mercado financeiro.

ABSTRACT

This study presents a predictive analysis of the Bovespa Index (IBOV), using the ARIMA model and its automated variant, Auto-ARIMA, to forecast short-term closing prices. The research covers data collected from January 2010 to May 2024, applying an automated flow for model parameter selection. The results showed that Auto-ARIMA could capture the general direction of trends, although it exhibited limitations in predicting abrupt changes and high volatility. The Root Mean Square Error (RMSE) was used as a model evaluation metric, indicating the need for additional adjustments or the incorporation of exogenous variables to enhance accuracy in more volatile scenarios.

Keywords: Bovespa Index, ARIMA, Auto-ARIMA, time series forecasting, financial market.

1. INTRODUÇÃO

O mercado financeiro brasileiro é caracterizado por sua volatilidade e complexidade, refletindo eventos econômicos, políticos e sociais que afetam diretamente o comportamento de seus índices. O Índice Bovespa (IBOV), principal indicador do desempenho das ações negociadas na B3, é amplamente utilizado como referência para medir a saúde econômica do país. Entender e prever o comportamento desse índice é de grande interesse para investidores, analistas financeiros e gestores de portfólio, pois possibilita a tomada de decisões mais embasadas e estratégicas, tanto no curto quanto no longo prazo. Nesse contexto, a modelagem de séries temporais surge como uma ferramenta crucial para identificar padrões e tendências nos preços das ações.

Entre os métodos mais conhecidos para previsão de séries temporais está o modelo ARIMA (Autoregressive Integrated Moving Average), amplamente utilizado devido à sua capacidade de capturar componentes autoregressivos e de médias móveis em séries estacionárias e não estacionárias. Este estudo foca na aplicação do modelo ARIMA, com a versão automatizada Auto-ARIMA, para prever o comportamento do Índice Bovespa no curto prazo, utilizando dados históricos coletados do Yahoo Finance. A pesquisa visa avaliar a precisão do modelo e explorar seu potencial para melhorar a estratégia de previsão em um mercado marcado por flutuações abruptas e complexas.

2. REFERENCIAL TEÓRICO E CONTEXTUALIZAÇÃO

A previsão de séries temporais é uma área fundamental da estatística e econometria, amplamente utilizada em estudos econômicos, financeiros e de planejamento. Desde os primeiros trabalhos de Yule (1926), que introduziu a técnica autoregressiva, até as contribuições de Box e Jenkins (1970), o campo evoluiu significativamente, introduzindo métodos que aprimoram a precisão e parcimônia na modelagem de séries temporais (RAMOS, 2010).

Os modelos ARIMA (Autoregressive Integrated Moving Average), desenvolvidos por Box e Jenkins, são amplamente adotados para análise e previsão de séries temporais, particularmente em séries econômicas e financeiras que apresentam estacionaridade após diferenciação (BOX; JENKINS, 1976). A metodologia de Box-Jenkins envolve as etapas de identificação, estimação e diagnóstico, visando modelar processos geradores de séries temporais com variância constante (GOMES, 1989). Entretanto, para séries que apresentam

variância não constante, técnicas de transformação, como o uso de logaritmos, são frequentemente aplicadas para estabilizar a variância antes da aplicação dos modelos ARIMA (LIMA SILVA, 2023).

Os modelos ARIMA são complementados por modelos de heterocedasticidade condicional, como os modelos ARCH e GARCH, propostos por Engle (1982) e Bollerslev (1986), para lidar com a volatilidade persistente em séries financeiras. Esses modelos permitem a variação da variância condicional ao longo do tempo, capturando a heteroscedasticidade característica das séries de retornos financeiros (RAMOS, 2010). Além disso, a cointegração, introduzida por Engle e Granger (1987), é utilizada para modelar relações de longo prazo entre séries não estacionárias, enquanto o modelo VEC (Vector Error Correction) é empregado quando há cointegração (JOHANSEN; JUSELIUS, 1990).

A aplicação prática dos modelos ARIMA em contextos brasileiros pode ser exemplificada em estudos de previsão de preços de commodities e índices financeiros, como no trabalho de previsão do preço do frango inteiro resfriado no atacado em São Paulo (CONGRESSO BRASILEIRO DE CUSTOS, 2006), e na análise do índice Bovespa a curto prazo (GOMES, 1989). Esses estudos mostram a robustez dos modelos ARIMA para capturar as dinâmicas subjacentes de preços e índices em contextos econômicos reais.

Recentemente, abordagens híbridas, combinando modelos estatísticos e técnicas de inteligência artificial, como redes neurais artificiais, têm sido exploradas para melhorar o desempenho da previsão de séries temporais financeiras (RAMOS, 2010). Essas abordagens buscam compensar as limitações dos modelos lineares, aproveitando a capacidade das redes neurais de capturar padrões não lineares e complexos nos dados, o que é essencial para prever séries com características de não-linearidade e alta volatilidade (WEIGEND; GERSHENFELD, 1994).

2.1 AMOSTRA

Os dados do índice Bovespa foram coletados no Yahoo Finance, abrangendo o período de 02/01/2010 a 02/05/2024. A análise estatística descritiva revela informações fundamentais, como contagem total de observações (3.548), média (78.888,43), desvio padrão (25.979,81), valores mínimos (37.497) e máximo (134.194).

O gráfico 1 apresenta o histórico de preços ajustados de fechamento do índice Bovespa (IBOV) de janeiro de 2010 a maio de 2024, mostrando uma tendência geral de alta, partindo de cerca de 70.000 pontos e atingindo vales abaixo dos 40.000 (em 2016) picos acima de 130.000 (em 2021 e 2024). A série é marcada por alta volatilidade, especialmente em momentos como a crise que culminou com o impeachment de Dilma Rousseff, em 2016, e a pandemia de 2020.

Gráfico 1 – Índice Bovespa entre janeiro 2010 e maio de 2024



Fonte Yahoo Finance. Elaboração Própria.

Para o processo de previsão, foram utilizados os últimos 60 dias de dados, dos quais 30 são usados para treinamento e 30 para teste. O modelo ARIMA foi ajustado iterativamente a cada observação do conjunto de teste, usando o `auto_arima` do pacote `pmdarima` do Python para determinar as melhores ordens (p , d , q). Durante as iterações, foram escolhidos modelos ARIMA variados para cada dado da série, como $ARIMA(0,0,1)$, $ARIMA(1,0,0)$, e $ARIMA(0,1,1)$, refletindo a complexidade e variabilidade dos dados.

As previsões resultaram em um erro quadrático médio (RMSE) de 1.006,92, indicando a precisão do modelo na estimativa dos preços de fechamento. Portanto, a modelagem ARIMA foi bem-sucedida, embora o RMSE indique uma margem de erro que pode ser melhorada com ajustes nos parâmetros do modelo ou inclusão de novas variáveis explicativas.

2.2 INSTRUMENTO DE AVALIAÇÃO

A metodologia utilizada para previsão de preços de ações baseou-se no modelo ARIMA (Autoregressive Integrated Moving Average), um dos métodos mais reconhecidos para análise e previsão de séries temporais. O ARIMA é especialmente eficaz em capturar padrões lineares em séries temporais estacionárias ou não estacionárias, ajustando componentes autoregressivos (AR), de média móvel (MA), e a ordem de diferenciação (I) para tornar a série estacionária antes da modelagem. No contexto deste estudo, o modelo foi aplicado aos dados de preços ajustados de fechamento das ações, extraídos de um banco de dados histórico do Yahoo Finance.

A aplicação do ARIMA aos dados seguiu um fluxo automatizado de seleção de parâmetros. O procedimento começou com a identificação de diferentes símbolos de ações (tickers) no banco de dados. Para cada ticker, os dados históricos de preços foram ordenados cronologicamente, e a função `auto_arima` foi utilizada para identificar automaticamente a melhor combinação de parâmetros (p, d, q), onde:

- p: representa a ordem do componente autoregressivo,
- d: é o número de diferenciações necessárias para tornar a série estacionária,
- q: indica a ordem do componente de média móvel.

A escolha automática dos parâmetros baseou-se no critério de informação AIC (Akaike Information Criterion), visando minimizar esse critério para selecionar o modelo mais parcimonioso. A modelagem foi realizada de forma não sazonal, já que a análise estava focada em prever o comportamento de curto prazo dos preços das ações.

Após a identificação do melhor modelo ARIMA, foram geradas previsões para os 30 dias subsequentes à última data disponível no conjunto de dados. Junto com as previsões, foram calculados intervalos de confiança para cada ponto, proporcionando uma estimativa da incerteza associada a cada previsão. Os resultados, incluindo os valores previstos e seus respectivos limites inferior e superior, foram armazenados no banco de dados. Cada entrada incluí o ticker correspondente, a data da previsão, o valor previsto do preço ajustado de fechamento, os limites de confiança, e a configuração específica do modelo ARIMA utilizado (expressa como ARIMA(p, d, q)).

A aplicação desta metodologia resultou em um conjunto de previsões pontuais para cada ação analisada, permitindo uma análise mais fundamentada do possível comportamento futuro

dos preços. O uso de ARIMA demonstrou-se eficaz na identificação de tendências e flutuações de curto prazo, oferecendo uma ferramenta quantitativa robusta para apoiar decisões financeiras no mercado de ações.

3. METODOLOGIA

O modelo ARIMA (Autoregressive Integrated Moving Average), ou Modelo Integrado Autorregressivo de Médias Móveis, é amplamente utilizado para análise e previsão de séries temporais. Ele combina três componentes: autorregressão (AR), integração (I) e médias móveis (MA), representados matematicamente pelas ordens p , d , e q , respectivamente.

O componente autorregressivo considera uma relação linear entre os valores atuais e os valores passados da série. Ele é representado pela ordem p , que define o número de lags incluídos no modelo.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

onde:

- X_t é o valor da série no tempo t .
- ϕ_l são os coeficientes dos lags, com l de 1 a p .
- ε_t é o termo de erro aleatório no tempo t , assumido como ruído branco com média zero e variância constante.

A integração lida com a estacionariedade da série temporal, ou seja, transforma uma série não estacionária em estacionária através da diferenciação. A ordem d indica o número de diferenças aplicadas para alcançar a estacionariedade.

$$X'_t = (1 - B)^d X_t$$

onde:

- B é o operador de defasagem (lag operator).
- ϕ_l é a ordem de diferenciação necessária para tornar a série estacionária.

O componente de médias móveis modela a relação entre o valor atual e os erros passados (resíduos) da série. A ordem q define quantos termos de erros passados são considerados.

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

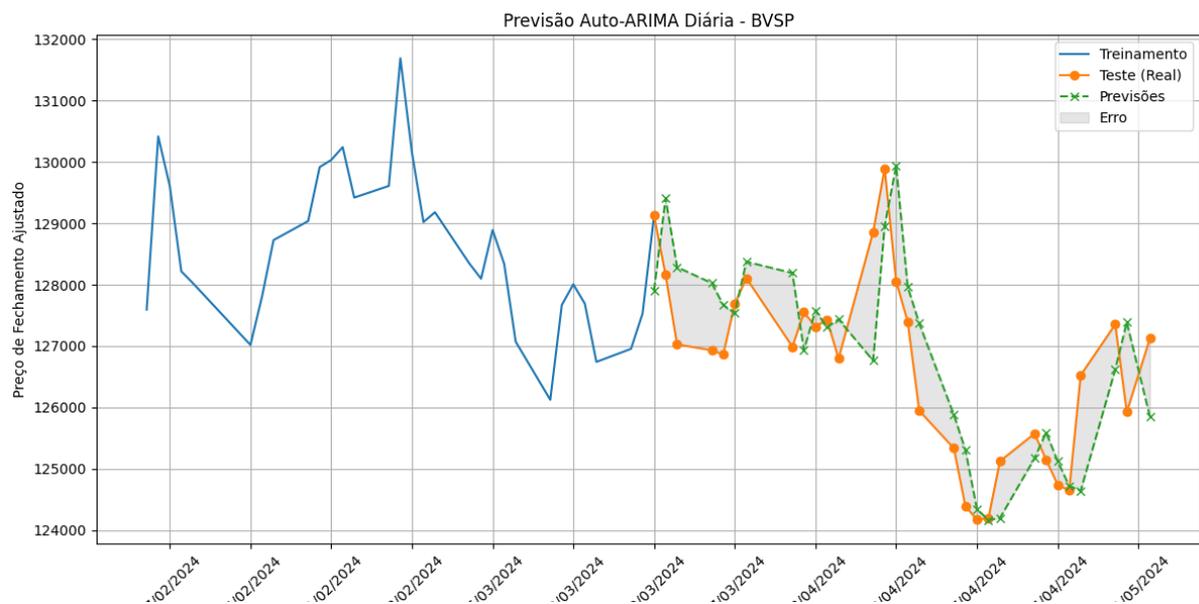
onde:

- ε_t é o termo de erro atual (ruído branco).
- θ_l são os coeficientes das médias móveis dos lags, com l de 1 a q .

4. RESULTADOS E DISCUSSÕES

Através dos dados foram realizadas previsões diárias do índice Bovespa utilizando o modelo Auto-ARIMA para um período de 30 dias, entre 20/03/2024 e 02/05/2024². O resultado está apresentado no gráfico 2, que apresenta as previsões e os valores reais do mercado, permitindo identificar tanto os acertos quanto os erros do modelo em cada iteração.

Gráfico 2 – Previsão do Índice Bovespa



Fonte: Yahoo Finance. Elaboração Própria

² O período se estende por mais de um mês devido ao fato da bolsa não negociar ativos nos fins de semana e feriados.

No gráfico 2 existem três linhas para representar as seguintes series:

1. **Linha Azul:** Representa o histórico de treinamento, com os dados anteriores usados para calibrar o modelo e gerar as previsões.
2. **Linha Laranja:** Indica os valores reais observados durante o período de teste.
3. **Linha Verde Pontilhada:** Corresponde às previsões feitas pelo modelo para o período de 30 dias.
4. **Área Cinza Sombreada:** Representa o erro entre os valores previstos e os valores reais, destacando as discrepâncias em cada ponto de previsão.

O gráfico mostra que, de maneira geral, o modelo conseguiu capturar a tendência geral dos preços de fechamento do índice Bovespa. No entanto, há momentos de maior discrepância, especialmente nos dias de alta volatilidade, quando as previsões se distanciaram consideravelmente dos valores reais. Esses desvios são evidenciados pelas áreas cinzas mais amplas, indicando maior erro de previsão nessas datas.

De modo geral o Erro Quadrático Médio (RMSE) das previsões: 1006,92. Podemos separar as previsões em:

- **Previsões com Superestimação:** Em algumas iterações, como em 10/04/2024 e 30/04/2024, o modelo superestimou os valores reais, resultando em diferenças positivas de 1.881,63 e 1.464,34, respectivamente.
- **Previsões com Subestimação:** Por outro lado, em 08/04/2024 e 26/04/2024, o modelo subestimou os valores reais, com diferenças de -2.095,52 e -1.883,83, respectivamente.

Deste modo, podemos verificar que o modelo Auto-ARIMA foi capaz de prever a direção geral do índice Bovespa, mas apresentou limitações em capturar mudanças bruscas ou movimentos de alta volatilidade. Esses resultados sugerem que, embora o modelo seja útil para prever tendências de curto prazo, ele pode se beneficiar da incorporação de variáveis exógenas ou de técnicas complementares para melhorar a precisão das previsões em ambientes financeiros voláteis.

5. CONSIDERAÇÕES FINAIS

Em resumo, a análise do índice Bovespa utilizando o modelo ARIMA e sua variante automatizada, o Auto-ARIMA, demonstrou-se eficaz na previsão de tendências de curto prazo, mas também expôs limitações na captura de movimentos abruptos e de alta volatilidade do mercado. O modelo conseguiu acompanhar a direção geral dos preços durante o período analisado, contudo, enfrentou dificuldades em prever de forma precisa os picos de volatilidade e as mudanças bruscas, resultando em discrepâncias significativas entre os valores previstos e os observados. Essas limitações foram evidenciadas tanto no cálculo do Erro Quadrático Médio (RMSE) quanto nas discrepâncias observadas no gráfico de previsão, especialmente em dias de maior instabilidade.

Portanto, a utilização do Auto-ARIMA mostrou-se uma ferramenta valiosa para análises financeiras e previsão de preços, especialmente em cenários de estabilidade relativa. No entanto, para ambientes financeiros mais voláteis, a precisão das previsões pode ser aprimorada com a incorporação de variáveis exógenas que considerem fatores econômicos e políticos externos, ou mesmo pela aplicação de modelos híbridos que combinem o ARIMA com outras técnicas, como redes neurais ou modelos de aprendizado de máquina mais complexos. Isso permitiria uma análise mais robusta, capaz de identificar não apenas a tendência, mas também as variações de curto prazo de forma mais assertiva.

REFERÊNCIAS

- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: Forecasting and control**. San Francisco: Holden-Day, 1976.
- ENGLE, R. F. **Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation**. *Econometrica*, v. 50, p. 987-1008, 1982.
- ENGLE, R. F.; GRANGER, C. W. J. **Co-integration and error correction: representation, estimation and testing**. *Econometrica*, v. 55, p. 251-276, 1987.
- GOMES, F. C. **Os modelos Arima e a abordagem de Box-Jenkins: uma aplicação na previsão do IBOVESPA a curtíssimo prazo**. 1989.

10ª Edição 2024 | 22 e 23 de novembro de 2024
João Pessoa, Paraíba (Região Nordeste)

JOHANSEN, S.; JUSELIUS, K. **Maximum likelihood estimation and inference on cointegration** – with applications to the demand for money. Oxford Bulletin of Economics and Statistics, v. 52(2), p. 169-210, 1990.

LIMA SILVA, R. G. F. **Avaliação da Precisão dos Modelos ARIMA com e sem Transformação Estabilizadora da Variância na Previsão de Séries Temporais Anuais**. Repositorio UFU. 2023.

RAMOS, A. S. **Previsões de séries temporais combinando modelos ARMA e Redes Neurais Artificiais**. Tese (Doutorado em Economia) – Universidade Federal de Pernambuco, 2010.

WEIGEND, A. S.; GERSHENFELD, N. A. **Time Series Prediction: Forecasting the Future and Understanding the Past**. Perseus Books, 1994.