

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE CIÊNCIAS ECONÔMICAS
CURSO DE CIÊNCIAS ECONÔMICAS

JOÃO VICTOR RUIZ RIBEIRO

**PROJEÇÃO DO ÍNDICE S&P 500: UMA APLICAÇÃO DO MÉTODO DE REGRESSÃO
DINÂMICA**

Juiz de Fora

2026

JOÃO VICTOR RUIZ RIBEIRO

**PROJEÇÃO DO ÍNDICE S&P 500: UMA APLICAÇÃO DO MÉTODO DE REGRESSÃO
DINÂMICA**

Monografia apresentada ao Curso de Ciências
Econômicas da Universidade Federal de Juiz de
Fora como requisito parcial à obtenção do título de
Bacharel em Ciências Econômicas.
Orientador: Prof. Dr. Alexandre Zanini

Juiz de Fora

2026

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Ruiz Ribeiro, João Victor.

PROJEÇÃO DO ÍNDICE S&P 500: UMA APLICAÇÃO DO MÉTODO DE REGRESSÃO DINÂMICA / João Victor Ruiz Ribeiro. -- 2026.

37 p.

Orientador: Alexandre Zanini

Coorientador: Douglas Sad Silveira,

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Economia, 2026.

1. S&P 500. 2. Regressão Dinâmica. 3. Séries de tempo. I. Zanini, Alexandre, orient. II. Sad Silveira,, Douglas , coorient. III. Título.



UNIVERSIDADE FEDERAL DE JUIZ DE FORA
REITORIA - FACECON - Depto. de Economia

FACULDADE DE ECONOMIA / UFJF

ATA DE APROVAÇÃO DE MONOGRAFIA II (MONO B)

Na data de 02/07/2026, a Banca Examinadora, composta pelos professores

1 - Alexandre Zanini - orientador; e

2 - Douglas Sad Silveira,

reuniu-se para avaliar a monografia do acadêmico **JOÃO VICTOR RUIZ RIBEIRO**, intitulada: **PROJEÇÃO DO ÍNDICE S&P500: UMA APLICAÇÃO DO MÉTODO DE REGRESSÃO DINÂMICA**.

Após primeira avaliação, resolveu a Banca sugerir alterações ao texto apresentado, conforme relatório sintetizado pelo orientador. A Banca, delegando ao orientador a observância das alterações propostas, resolveu **APROVAR** a referida monografia.

ASSINATURA ELETRÔNICA DOS PROFESSORES AVALIADORES



Documento assinado eletronicamente por **Alexandre Zanini, Professor(a)**, em 02/07/2026, às 15:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Douglas Sad Silveira, Professor(a)**, em 02/07/2026, às 16:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **3063158** e o código CRC **DF36F77A**.

Referência: Processo nº 23071.923719/2026-20

SEI nº 3063158

RESUMO

O presente trabalho tem como objetivo analisar a evolução histórica do S&P 500 e projetar o comportamento futuro do índice, no período de janeiro de 2010 a maio de 2026. O S&P 500 é um dos principais índices do mercado acionário mundial, reunindo empresas de grande capitalização de maior relevância global. Para a análise, foi utilizado o Método de Regressão Dinâmica (MRD), que aplica um conjunto de testes diagnósticos para chegar ao modelo final. Os dados foram obtidos junto ao Stooq, para a série do S&P 500, e ao Federal Reserve Economic Data (FRED), para as variáveis Federal Funds Rate (FEDFUNDS), Consumer Price Index (CPI) e taxa do título do Tesouro americano de dez anos (GS10). A estimação identificou um componente dinâmico na evolução do índice, expresso pela relação com o seu próprio valor no mês anterior, além de uma relação positiva com o CPI. As variáveis FEDFUNDS e GS10 não apresentaram significância estatística e, por isso, não integraram a equação final. O modelo apresentou bom ajuste aos dados, com R^2 ajustado de 99,37% e erro médio absoluto percentual (MAPE) de 3,17%. A partir do modelo estimado, projetou-se o índice para os doze meses seguintes, mediante a autoprojção do CPI, obtendo-se um cenário de alta para o S&P 500.

Palavras-chave: Regressão Dinâmica; S&P 500; Séries de Tempo; Econometria Financeira.

ABSTRACT

This study aims to analyze the historical evolution of the S&P 500 and to forecast the future behavior of the index, covering the period from January 2010 to May 2026. The S&P 500 is one of the leading indices of the global stock market, comprising large-capitalization companies of major worldwide relevance. The analysis was carried out using the Dynamic Regression method (DR), which applies a set of diagnostic tests to reach the final model. The data were obtained from Stooq, for the S&P 500 series, and from the Federal Reserve Economic Data (FRED), for the variables Federal Funds Rate (FEDFUNDS), Consumer Price Index (CPI), and the ten-year U.S. Treasury yield (GS10). The estimation identified a dynamic component in the evolution of the index, expressed by its relationship with its own value in the previous month, as well as a positive relationship with the CPI. The variables FEDFUNDS and GS10 were not statistically significant and therefore were not included in the final equation. The model showed a good fit to the data, with an adjusted R^2 of 99.33% and a mean absolute percentage error (MAPE) of 3.15%. Based on the estimated model, the index was projected for the following twelve months through the auto-projection of the CPI, resulting in an upward scenario for the S&P 500.

Keywords: Dynamic Regression; S&P 500; Time Series; Financial Econometrics.

LISTA DE FIGURAS

Figura 1: Construção de um modelo de regressão dinâmica.....	24
Figura 2: Evolução histórica das variáveis – Janeiro de 2010 a Maio de 2026.....	31
Figura 3: Função de Autocorrelação dos erros do modelo.....	32
Figura 4: Previsão do S&P 500	36

LISTA DE TABELAS

Tabela 1 – Principais Bolsas De Valores Do Mundo	11
Tabela 2 – Principais setores do S&P 500.....	14
Tabela 3 – Estatísticas de Desempenho do Modelo (%)	33
Tabela 4 – Previsão do S&P 500 – Intervalo de Confiança de 95%	35

SUMÁRIO

1 INTRODUÇÃO.....	7
2 REFERENCIAL TEÓRICO.....	10
2.1 O Mercado Acionário Norte-Americano.....	10
2.2 O índice S&P 500.....	11
2.2.1 Histórico do S&P 500.....	11
2.2.2 Metodologia de composição.....	12
2.2.3 Fórmula de cálculo	13
2.2.4 Composição setorial atual.....	13
2.3 Variáveis Macroeconômicas	15
2.3.1 Fed Funds Rates	15
2.3.2 CPI.....	15
2.3.3 Treasury 10y	16
2.4 Revisão Empírica.....	16
3 MÉTODO DE REGRESSÃO DINÂMICA	18
3.1 - Conceituação	18
3.2 - Estrutura dos Modelos de Regressão Dinâmica.....	19
3.2.1 - Modelos de Regressão Cochrane-Orcutt Generalizados.....	21
3.3 - Construção de modelos de regressão dinâmica.....	22
3.4 - Testes usados nos modelos de regressão dinâmica	25
3.4.1 - Testes de verificação da “dinâmica” do modelo	25
3.4.2 - Testes para a especificação das variáveis causais	27
3.4.3 - Testes baseados na autocorrelação dos resíduos.....	28
4 ANÁLISE DE DADOS E RESULTADOS	30
4.1 As Variáveis	30

4.2 Equação de Previsão.....	31
4.2 Previsões.....	34
5 CONCLUSÃO.....	37
6 REFERÊNCIAS BIBLIOGRÁFICAS	38

1 INTRODUÇÃO

Um índice de ações é uma das formas mais acessíveis de acompanhar o comportamento de determinado mercado. No Brasil, é comum observar notícias sobre a valorização ou desvalorização diária do índice Bovespa, principal indicador do mercado acionário nacional, formado a partir do desempenho das maiores empresas brasileiras segundo critérios previamente definidos (B3, 2025).

O S&P 500 é um índice de ações que mede o desempenho do segmento de grande capitalização do mercado acionário dos Estados Unidos. O índice é composto por 500 companhias líderes e cobre aproximadamente 80% da capitalização de mercado disponível do mercado acionário norte-americano (S&P Global, 2026).

O S&P 500 é mantido pela S&P Dow Jones Indices e segue metodologia oficial para definição de elegibilidade, manutenção e rebalanceamento. Os índices da família S&P U.S. Indices são ponderados pela capitalização de mercado ajustada ao free float, o que faz com que empresas de maior valor de mercado tenham maior peso relativo no índice (S&P Global, 2026).

O mercado acionário dos Estados Unidos permanece como o maior do mundo. Dados da World Federation of Exchanges mostram que, em 2025, a NYSE e a Nasdaq somavam mais de US\$ 62 trilhões em capitalização doméstica, evidenciando a centralidade do mercado norte americano no sistema financeiro global.

Nesse contexto, um índice de ações funciona como referência para acompanhar e monitorar o desempenho de um mercado. Assim, o S&P 500 pode ser considerado um dos principais índices do mundo, pois reúne empresas de grande capitalização do mercado acionário mais relevante em termos globais (S&P Global, 2026).

O mercado brasileiro de capitais cresceu significativamente nos últimos anos. Segundo a B3, no primeiro trimestre de 2026 foram registrados 5,6 milhões de CPFs com posição em renda variável, crescimento de 12% desde 2022. Além disso o número de investidores em ETFs chegou a 823 mil, em comparação com 608,7 mil no primeiro trimestre de 2025 e ainda o número de investidores em BDR, instrumento que permite investidores brasileiros investirem em empresas listadas em bolsas estrangeiras, chegou a cerca de 1 milhão. Isso mostra que ainda há um movimento expressivo de pessoas diversificando em investimentos no exterior, o que torna natural a busca por entender melhor o mercado americano (B3, 2026).

Realizar a previsão de séries financeiras é uma tarefa complexa devido à quantidade de fatores que afetam os preços dos ativos. Chen, Roll e Ross (1986) identificam produção industrial, inflação, spread da taxa de juros como variáveis macroeconômicas significativas que afetam o preço das ações. Fama (1981) argumenta que existe uma correlação negativa entre a inflação e o retorno sobre as ações, fato que mostra uma das dificuldades nos métodos de previsão, uma vez que se esperaria que ações funcionassem como proteção contra a inflação. O autor explica esse fenômeno por meio da “proxy hypothesis”: a atividade econômica real se correlaciona positivamente com retornos acionários e negativamente com inflação, de modo que a correlação observada entre inflação e bolsa seria, na verdade, mediada pela atividade real. Isso justifica a inclusão das variáveis inflação e atividade econômica real nos modelos de previsão.

Considerando esse cenário, esta monografia tem como objetivo projetar o índice S&P 500 através do Método de Regressão Dinâmica (MRD)¹, analisando o impacto de variáveis macroeconômicas dos Estados Unidos sobre o desempenho do índice entre janeiro de 2010 e maio de 2026. Mais claramente pode-se dizer que o problema de pesquisa deste trabalho monográfico concerne em verificar em que medida um modelo de regressão dinâmica, com variáveis macroeconômicas selecionadas, é capaz de projetar o comportamento do S&P 500 no curto prazo.

Para atingir o objetivo, esta monografia busca: (i) mostrar o mercado acionário norte americano e a metodologia do índice S&P 500; (ii) analisar as variáveis macroeconômicas que afetam o desempenho de índices acionários; (iii) estimar, por meio de Regressão Dinâmica, a relação entre o S&P 500 e as variáveis Fed Funds Rate, CPI e Treasury 10y. (iv) avaliar o modelo estimado por meio de testes diagnósticos e indicadores de desempenho; (v) projetar o S&P 500 para o horizonte de doze meses.

A contribuição esperada deste trabalho é a difusão do conhecimento do mercado acionário americano ampliando a base de análises para o investidor brasileiro, através do Método de Regressão Dinâmica, a princípio pouco utilizado para previsão do S&P 500. Esta monografia dialoga com o trabalho de Faria (2023), que analisa pela mesma ótica o mercado de ações brasileiro, e deixa como sugestão de continuação do trabalho a análise do S&P 500.

Esta monografia está organizada em cinco capítulos. Além desta Introdução, o capítulo 2 apresenta o referencial teórico, onde é contextualizado o mercado acionário norte-americano, o índice S&P 500, as variáveis macroeconômicas que podem afetar seu desempenho e uma revisão

¹ Goodrich (1989) e Zanini (2000).

da literatura empírica sobre previsão de índices acionários. O capítulo 3 mostra a metodologia, com a apresentação do método de Regressão Dinâmica. O capítulo 4 discute os resultados da estimação e da projeção do modelo. Por fim, o capítulo 5 traz as considerações finais.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta a fundamentação teórica do trabalho, contextualizando o mercado acionário norte-americano, o índice S&P 500 e sua metodologia de cálculo, as variáveis macroeconômicas selecionadas como explicativas e uma revisão da literatura empírica sobre previsão de índices acionários.

2.1 O MERCADO ACIONÁRIO NORTE-AMERICANO

O mercado norte-americano é o maior do mundo em capitalização, com a NYSE e a Nasdaq como principais bolsas de valores, e com um total de aproximadamente 4.642 empresas listadas de diferentes setores, como tecnologia, finanças, energia, saúde, indústria e varejo (WFE, 2025). Em 2024, o volume médio diário de negociação ultrapassou 12 bilhões de ações nas bolsas de valores americanas, demonstrando a maturidade e a liquidez presentes nesse mercado (SIFMA, 2025). Essa profundidade torna o mercado norte-americano uma referência para o mercado acionário global.

A NYSE foi fundada em 1792 e é a bolsa mais antiga dos Estados Unidos, onde estão listadas as empresas mais tradicionais e de setores mais maduros. (NYSE, 2026). Já a Nasdaq foi criada em 1971, sendo a primeira bolsa totalmente eletrônica do mundo, e é onde estão as empresas de alta tecnologia e crescimento (Nasdaq, 2021). Todas as empresas presentes no índice S&P 500 estão listadas em uma dessas duas bolsas de valores americanas. A Tabela 1 apresenta o valor de mercado das dez maiores bolsas do mundo em 2024.

Tabela 1 – Principais Bolsas De Valores Do Mundo

País	Valor de mercado
Estados Unidos	US\$ 69,0 trilhões
China	US\$ 14,75 trilhões
Japão	US\$ 8,08 trilhões
Hong Kong	US\$ 6,49 trilhões
Índia	US\$ 5,13 trilhões
Reino Unido	US\$ 4,17 trilhões
Canadá	US\$ 3,85 trilhões
Coreia do Sul	US\$ 3,76 trilhões
França	US\$ 3,69 trilhões
Taiwan	US\$ 3,32 trilhões

Fonte: Elaboração própria a partir de World Bank (2024).

A Tabela 1 evidencia o tamanho do mercado norte-americano, mais de quatro vezes maior que o do segundo colocado, a China. Isso reforça a importância desse mercado e a sua influência global. Cabe destacar, ainda, que o S&P 500 reúne as 500 maiores empresas desse mercado em capitalização, sendo o principal índice utilizado como referência para acompanhar seu desempenho.

2.2 O ÍNDICE S&P 500

2.2.1 HISTÓRICO DO S&P 500

O índice S&P 500 foi criado pela Standard & Poor's em 1957, como sucessor de um índice anterior, mas com apenas 90 empresas. Foi o primeiro indicador de mercado gerado em tempo real por um computador (S&P Global, 2026). Atualmente, o índice é operado pela S&P Dow Jones Índices.

2.2.2 METODOLOGIA DE COMPOSIÇÃO

O índice S&P 500 segue a metodologia oficial publicada pela S&P Dow Jones Indices, que define os critérios de elegibilidade, seleção, ponderação e manutenção das empresas que o compõem. A metodologia é pública e garante transparência sobre as regras aplicadas para construir o índice, que reúne 500 empresas de grande capitalização do mercado norte-americano (S&P Dow Jones Indices, 2026).

Dentre os principais critérios para inclusão no índice, destacam-se: a empresa precisa ser dos Estados Unidos e estar listada em uma das seguintes bolsas: NYSE, NYSE Arca, NYSE American, Nasdaq, Cboe BZX, Cboe BYX, Cboe EDGA ou Cboe EDGX. Além disso, precisa ter capitalização de mercado mínima de US\$ 22,7 bilhões e capitalização ajustada ao free float correspondente a pelo menos 50% do total. O volume de negociações deve ser de no mínimo 250.000 ações em cada um dos seis meses anteriores à data de avaliação, e o setor da empresa é considerado para preservar o balanceamento setorial do índice (S&P Dow Jones Indices, 2026).

Essas decisões são tomadas pelo Comitê de Índices da S&P Dow Jones Indices, que tem discricionariedade na escolha, diferentemente de outros índices, nos quais a seleção é feita de forma automática por ranking de capitalização. Junto com os critérios citados acima, o comitê analisa o balanceamento setorial do índice, sua estabilidade e a representatividade da empresa no mercado, e suas decisões são comunicadas publicamente (S&P Dow Jones Indices, 2026).

O S&P 500 é ponderado pela capitalização de mercado ajustada ao free float, de modo que empresas maiores têm peso proporcionalmente maior no índice. Apenas as ações disponíveis para negociação pública entram no cálculo: ações detidas por insiders, governos ou em bloqueio são excluídas. A cada três meses ocorre o rebalanceamento, com ajustes na quantidade de ações em circulação das empresas componentes. Já as inclusões e exclusões de empresas podem ocorrer a qualquer momento, conforme decisão do Comitê de Índices, geralmente motivadas por aquisições, falências ou perda de elegibilidade (S&P Dow Jones Indices, 2026).

2.2.3 FÓRMULA DE CÁLCULO

A equação 1 a seguir é utilizada para o cálculo do índice:

$$\text{índice S\&P 500} = \frac{\left\{ \sum_{i=1}^{500} (P_i \times Q_i \times IWF_i) \right\}}{\{D\}}$$

Equação 1 — Fórmula de cálculo do S&P 500

Fonte: S&P Dow Jones Indices (2026).

No numerador, é apresentado a soma da capitalização ajustada ao *free float* para todas as empresas do índice. Sendo: P é o preço da ação, Q a quantidade de ações em circulação da empresa e IWF o fator entre 0 e 1 que indica a proporção de ações em *free float*. Já no denominador, temos D , uma constante ajustada periodicamente para manter a continuidade do índice (S&P Dow Jones Indices, 2026).

Cabe destacar a função do divisor D : sem ele, qualquer evento corporativo iria gerar uma alteração artificial no índice, por exemplo, se houver a troca de uma empresa por outra. A capitalização da nova empresa não é igual à da antiga; por isso, se for feita uma troca simples, o índice subirá ou descera de forma artificial. Portanto, sempre que há um evento que afeta o índice, é necessário ajustar o D para refletir o valor real do índice. (S&P Dow Jones Indices, 2026).

2.2.4 COMPOSIÇÃO SETORIAL ATUAL

Uma carteira de ações diversificada tende a apresentar menor risco do que a soma ponderada dos riscos individuais de cada ativo. Este argumento foi desenvolvido por Markowitz (1952) em sua teoria de seleção de portfólio, referência fundamental para estudos sobre diversificação e alocação de ativos. Esse princípio aplica-se de forma direta a índices acionários compostos por múltiplas empresas, como o S&P 500.

O S&P 500 é composto por empresas em 11 setores da economia norte-americana, o que reforça seu papel como indicador representativo e diversificado do mercado acionário dos Estados Unidos. A Tabela 2 apresenta a composição setorial do índice.

Tabela 2 – Principais setores do S&P 500

Setor	Participação
Tecnologia	29%
Finanças	13%
Saúde	12%
Serviços não essenciais	10%
Comunicação	9%
Indústrias	8%
Serviços essenciais	6%
Energia	4%
Materiais	3%
Utilitários	3%
Negócios imobiliários	3%

Fonte: Elaboração própria a partir de S&P Global (2026).

A Tabela 2 evidencia a força das empresas de tecnologia na economia atual, com 29% de participação no índice. Em seguida, estão os setores mais tradicionais, finanças e saúde, com 13% e 12%, respectivamente. Os outros setores apresentam pesos diversificados, variando entre 3% e 10%, o que demonstra que, apesar da predominância de empresas tecnológicas, o índice ainda possui equilíbrio setorial.

Segundo Vieira (2024), o S&P 500 constitui um exemplo relevante de diversificação, pois sua estrutura setorial tende a mitigar parte dos riscos específicos das empresas componentes, inclusive em contextos de maior volatilidade.

2.3 VARIÁVEIS MACROECONÔMICAS

Esta seção apresenta as variáveis macroeconômicas selecionadas como explicativas do modelo de previsão do S&P 500, e que afetam sistematicamente o retorno das ações segundo Chen, Roll e Ross (1986) e Fama (1981): dentre elas estão: *Fed Funds Rate*, *CPI* e *Treasury 10y*. Todas serão detalhadas nas próximas subseções.

2.3.1 FED FUNDS RATE

A Fed Funds Rate é a taxa básica de juros dos Estados Unidos, cuja meta é definida pelo FOMC (Federal Open Market Committee), que se reúne em oito reuniões por ano (Federal Reserve, 2026). Espera-se uma relação negativa com o S&P 500: quando a taxa de juros sobe, a renda fixa fica mais atraente e há uma migração de capital da renda variável para a renda fixa, resultando na queda do índice. Chen, Roll e Ross (1986) identificam o spread de juros como variável precificada pelo mercado de ações americano e Faria (2023) encontrou esse efeito negativo no caso brasileiro, utilizando a SELIC e o IBOV.

2.3.2 CPI

O Consumer Price Index (CPI) é o principal indicador de inflação ao consumidor dos Estados Unidos, calculado pelo Bureau of Labor Statistics (BLS, 2026). A visão tradicional é de que ações funcionam como hedge contra a inflação, uma vez que empresas tendem a repassar a alta de preços ao consumidor final. No entanto, evidências empíricas mostram uma relação negativa entre inflação e retornos acionários reais.

Fama (1981) propôs a *proxy hypothesis* para explicar esse fenômeno: a atividade econômica real correlaciona-se positivamente com retornos acionários e negativamente com inflação, de modo que a correlação observada entre inflação e bolsa seria mediada pela atividade real, e não direta.

2.3.3 TREASURY 10Y

A *Treasury 10y* é a taxa de juros de longo prazo, medida pelo rendimento dos títulos do Tesouro americano com vencimento em 10 anos, sendo uma referência global para a precificação de ativos (Federal Reserve, 2026). Ela é definida pelo mercado por meio da oferta e demanda nos leilões primários e do mercado secundário. A principal diferença do *Treasury 10y* para a *Fed Funds Rate* é que a primeira é de longo prazo e calculada pelo mercado, enquanto a segunda é definida por política monetária e é considerada de curto prazo. A relação esperada com o S&P 500 é inversa, pois essa taxa é referência para o desconto de fluxos futuros das empresas, portanto, se ela sobe, o *valuation* cai (Chen, Roll e Ross, 1986).

2.4 REVISÃO EMPÍRICA

A previsão de índices de ações é um tema amplamente discutido na literatura, apresentando desafios significativos. Diversos métodos são utilizados na vasta literatura empírica, como: séries temporais, modelos macro, *machine learning* e redes neurais. Além disso, as variáveis utilizadas também variam entre macroeconômicas, dados financeiros e até dados de redes sociais.

Chen, Roll e Ross (1986), utilizando dados mensais de 1953 até 1983, testaram se variáveis macroeconômicas são precificadas pelo mercado americano. A hipótese central dos autores era de que movimentos inesperados em variáveis macroeconômicas afetam o retorno acionário. As variáveis produção industrial, inflação não esperada, mudança na inflação esperada, prêmio de risco e spread entre taxas longas e curtas foram identificadas como significativas.

Em estudo sobre a relação entre retornos acionários, atividade real, inflação e oferta de moeda no período de 1953 a 1977, Fama (1981) documentou a correlação negativa entre inflação e retornos acionários reais, observação contraintuitiva uma vez que ações são frequentemente vistas como *hedge* contra a inflação. Para explicar esse fenômeno, o autor propôs a *proxy hypothesis*: a atividade real correlaciona-se positivamente com retornos acionários e negativamente com

inflação, de modo que a correlação negativa observada entre inflação e retornos seria, na verdade, mediada pela atividade real.

Jiao e Jakubowicz (2017) destacam a crescente importância da predição de movimentos de ações e mostram que informações recentes de outros mercados podem elevar a previsibilidade de modelos aplicados ao S&P 500. No campo das novas fontes de informação, Mao et al. (2012) analisam a relação entre o volume diário de publicações no Twitter sobre ações do S&P 500 e indicadores de mercado, encontrando evidências de correlação entre a atividade na rede social e alguns movimentos do mercado acionário.

A literatura sobre previsão do S&P 500 utiliza métodos como *machine learning* e redes neurais. Até onde se sabe, não foram identificados trabalhos que apliquem o Método de Regressão Dinâmica especificamente a esse índice. Este trabalho pretende contribuir para preencher essa lacuna.

Faria (2023) aplicou o Método de Regressão Dinâmica ao Ibovespa utilizando as variáveis: SELIC, IPCA, câmbio, IBC-Br e uma *dummy* para pandemia. Sugeriu como continuação de seu trabalho o uso do mesmo método para o S&P 500, o que esse trabalho busca desenvolver.

3 MÉTODO DE REGRESSÃO DINÂMICA²

3.1 - CONCEITUAÇÃO

Nos modelos de regressão linear estudados comumente na literatura, supõe-se que os erros “gerados” pelo modelo possuem algumas características como: média zero, variância constante, distribuição Normal e independência (o que implica na inexistência de correlação serial).

Entretanto, na prática, ao modelar séries econômicas, os resíduos tendem a apresentar correlações positivas, e erros positivos tendem a ser seguidos por outros também positivos (o mesmo comportamento é observado para resíduos negativos). O gráfico dos resíduos versus o índice dos tempos revela, nesta situação, que os resíduos tendem a se agrupar em “blocos” de resíduos com o mesmo sinal. Segundo Barros e Souza (1995), ao tentar modelar uma série temporal através de um modelo de regressão, a hipótese de independência dos ruídos não é realista, e os resultados e testes usados nos modelos de regressão não são válidos. Por exemplo, segundo os autores, algumas das consequências da autocorrelação dos resíduos são:

- 1) Os estimadores usuais por mínimos quadrados são ainda não tendenciosos, mas não têm variância mínima.
- 2) Os estimadores da variância e dos erros padrões dos coeficientes da regressão são subestimados, o que levaria à conclusão de que os estimadores são mais precisos do que na realidade.
- 3) Os intervalos de confiança para os parâmetros da regressão e os testes de hipóteses relacionados a estes intervalos perdem a validade, como uma consequência direta de 2).

Desta forma, estes três motivos implicam na necessidade de procurar procedimentos para tratar o problema de autocorrelação dos erros, pois ignorá-los leva, em geral, a inúmeras conclusões errôneas.

Em particular, dado que a hipótese de independência dos erros não é realista no contexto de séries temporais, os modelos de regressão dinâmica estendem os modelos usuais de regressão ao levantarem esta restrição.

² Este capítulo é baseado fundamentalmente em Zanini (2000).

Os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. Atenta-se que o termo “regressão dinâmica” não indica que os parâmetros do modelo evoluem no tempo³. Ao contrário, a palavra “dinâmica” significa aqui um modelo de regressão no qual incluímos a estrutura de dependência de uma série temporal.

Modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente (série a ser explicada) indicar que não podemos supor a independência dos erros.

Salienta-se que a estimação de parâmetros num modelo de regressão dinâmica é feita através de mínimos quadrados ordinários⁴, a exemplo dos modelos de regressão usuais. Entretanto, a estimação em modelos de regressão dinâmica é mais complicada, e envolve um procedimento iterativo com vários estágios.

Nos modelos de regressão dinâmica, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas. Atenta-se, neste momento, para uma outra distinção entre os modelos de regressão dinâmica e modelos de espaço de estados. Nos modelos de regressão dinâmica, as variáveis exógenas são tratadas como “números fixos” e não como variáveis aleatórias.

Já nos modelos de espaço de estados, as variáveis exógenas são tratadas como séries temporais, ou seja, realizações de processos estocásticos. Logo, no contexto da modelagem em espaço de estados, a estrutura de autocovariâncias e autocorrelações das séries de variáveis exógenas é uma informação de interesse, enquanto este aspecto é ignorado nos modelos de regressão dinâmica.

3.2 - ESTRUTURA DOS MODELOS DE REGRESSÃO DINÂMICA

Os modelos de regressão dinâmica podem ser descritos pela equação:

$$\varphi(B)Y_t = \beta x_t + \varepsilon_t$$

³ Caso dos modelos de espaço de estados que usam o Filtro de Kalman (Hamilton, 1994).

⁴ Dudewicz & Mishra (1988).

onde:

Y_t = variável dependente (endógena) no instante t

β = vetor de coeficientes das variáveis causais, que será estimado por mínimos quadrados

x_t = vetor de variáveis causais (exógenas)⁵ no instante t

ε_t = ruído aleatório associado ao modelo, onde supomos que os ε_t são independentes e identicamente distribuídos com densidade $N(0, \sigma^2)$

$\varphi(B)$ = polinômio autoregressivo de ordem p , isto é:

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ sendo B o operador de atraso

A estrutura do modelo de regressão dinâmica permite considerar como elementos x e t variáveis causais e também suas defasagens.

A presença do polinômio $\varphi(B)$ no modelo traz uma grande flexibilidade desta classe de modelos mas, ao mesmo tempo, dificulta a procura por um modelo adequado.

Observa-se que, se $\varphi(B) = 1$, não existem defasagens da variável dependente, e a interpretação do modelo é muito simples, pois as variáveis causais influenciam diretamente a variável endógena.

Ao contrário, quando $\varphi(B) \neq 1$, o modelo pode ser usado para representar relações bastante complicadas.

Por fim, salienta-se que uma grande diferença entre os modelos de Regressão Dinâmica e os modelos ARIMA consiste no fato dos modelos de regressão dinâmica incluírem efeitos de variáveis causais através do termo βx_t . Os modelos ARIMA univariados de Box e Jenkins, por sua vez, não incluem tais efeitos, e apenas o passado da série y_t e os valores defasados da série de erros são usados na modelagem e previsão da série Y_t .

Procura-se, portanto, dentro de metodologia definida, evoluir na modelagem, buscando em outras causalidades, que não somente a própria série de demanda, um melhor modelo para prever esta demanda. Este modelo, como já mencionado, foi o modelo de regressão dinâmica que pode

⁵ As variáveis exógenas representadas por x_t têm seus valores determinados “fora” do modelo de regressão, enquanto a variável endógena Y_t é determinada a partir das variáveis exógenas e do ruído ε_t .

ser considerado como um caso particular do que é conhecido na literatura como modelos de Cochrane e Orcutt generalizados.

3.2.1 - MODELOS DE REGRESSÃO COCHRANE-ORCUTT GENERALIZADOS⁶

O modelo de regressão generalizado de Cochrane e Orcutt (1949) é dado por:

$$\varphi(B)Y_t = \beta x_t + w_t, (3.1)$$

$$R(B)w_t = \varepsilon_t. (3.2)$$

Onde $R(B)$ = polinômio autoregressivo

Pode-se notar que a equação (3.2) tem a mesma forma da equação (3.1) da regressão dinâmica, entretanto os erros w_t apresentam uma estrutura AR dada pela equação (3.3)⁷.

Este modelo dado pelas equações (3.2) e (3.3) ainda pode ser descrito em termos de uma única equação como a seguir. Note que da equação (3.2):

$$w_t = \varphi(B)Y_t - \beta x_t$$

Substituindo esta última expressão na equação (3.3) tem-se que:

$$R(B) \cdot [\varphi(B)Y_t - \beta X_t] = \varepsilon_t \quad (3.4)$$

Desta última expressão nota-se que o modelo de regressão generalizado de Cochrane e Orcutt introduz defasagens tanto na variável dependente (Y_t) quanto nas causais. A expressão (3.4) indica também que a relação de causalidade entre Y_t e x_t não é afetada pela introdução do polinômio autoregressivo $R(B)$.

⁶ Barros e Souza (1995).

⁷ A formulação original de Cochrane e Orcutt supõe que os ruídos w_t apresentam apenas uma estrutura AR(1), ou seja, a equação (3.3) é dada por: $w_t = \alpha w_{t-1} + \varepsilon_t$ onde os ε_t são i.i.d e apresentam distribuição $N(0, \sigma^2)$.

Esta última equação pode ainda ser escrita em termos de novas variáveis Y_t^* e x_t^* dadas por:

$$\begin{aligned} Y_t^* &= R(B)Y_t \\ x_t^* &= R(B)x_t \end{aligned}$$

onde $R(B)$ é chamado de “fator comum” e representa a estrutura de correlação presente no erro w_t .

A equação obtida então usando-se estas novas variáveis é:

$$\phi(B)Y_t^* = \beta x_t^* + \epsilon_t$$

Logo, o modelo de regressão generalizado de Cochrane e Orcutt reduz-se ao modelo de regressão dinâmica usual ao se considerar as novas variáveis Y_t^* e x_t^* . O modelo original de Cochrane e Orcutt tem como fator comum $R(B) = 1 - \alpha B$ e então a equação (3.5) reduz-se a:

$$\phi(B) \cdot [Y_t - \alpha Y_{t-1}] = \beta [x_t - \alpha x_{t-1}] + \epsilon_t \quad (3.6)$$

Salienta-se que, neste modelo, o procedimento de estimação é sequencial. A estimativa inicial de α é 0, e a partir dela estimamos β e $\phi(B)$ por mínimos quadrados ordinários. A partir destas estimativas encontra-se um estimador de $R(B)$ através da equação $R(B)w_t = \epsilon_t$. O polinômio estimado $R(B)$ é então usado para transformar y_t e x_t e reestimar β e $\phi(B)$. O processo é repetido até que se alcance a convergência dos parâmetros.

3.3 - CONSTRUÇÃO DE MODELOS DE REGRESSÃO DINÂMICA

Geralmente os modelos econométricos têm uma estrutura conhecida, baseada em considerações teóricas e o problema reduz-se ao problema de estimação dos parâmetros do modelo já conhecido. Entretanto, este é raramente o caso no contexto de séries temporais, onde a estratégia é construir modelos a partir dos dados.

A estratégia usualmente empregada para construir um modelo de regressão dinâmica é uma estratégia bottom-up, isto é, partimos de um modelo simples e o refinamos e incluímos novas variáveis até encontrar um modelo apropriado. A elaboração de um modelo de regressão dinâmica é muitas vezes um procedimento difícil, pois é preciso não apenas escolher as variáveis a serem incluídas no modelo, mas também os lags (defasagens) destas variáveis.

Na definição do modelo adequado, é necessário levar em conta não só a significância dos parâmetros mas também uma certa estrutura “lógica” do modelo. Por exemplo, vendas (ou demanda) de um produto são geralmente afetadas por seu preço. O aumento do preço, por sua vez, tende a diminuir as vendas (ou demanda), e vice-versa.

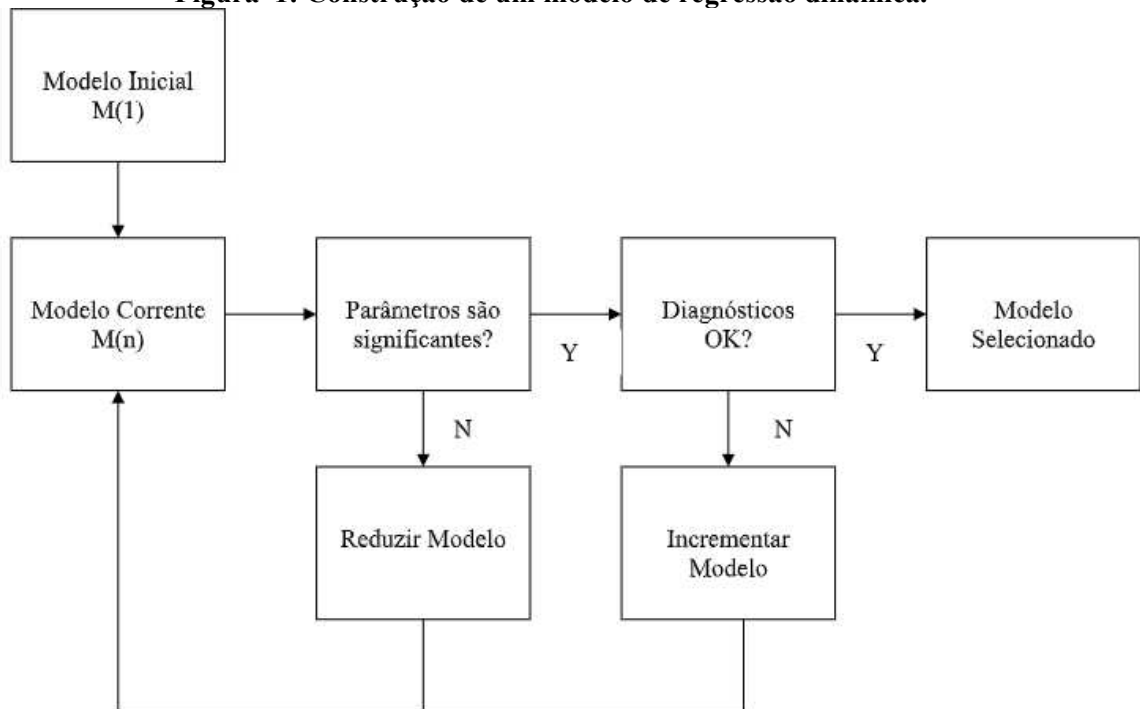
Logo, se o modelo de regressão encontrado para explicar venda pelo preço apresenta um coeficiente positivo para a variável preço, é bom desconfiar por mais bem ajustado que esteja o modelo, pois a relação apontada pelo modelo não é, em geral, verdadeira⁸. Em síntese, na escolha de um modelo de regressão, não é necessário apenas encontrar um ajuste de parâmetros adequado, mas fundamentalmente faz-se mister verificar se os coeficientes estimados são coerentes.

As previsões geradas por um modelo de regressão dinâmica dependem não só de valores passados da série mas também dos valores previstos para as variáveis causais. Logo, para obtermos as previsões da série Y_t para $T+1$, $T+2$, $T+3$, etc, é necessário fornecer ao modelo os valores futuros do vetor de variáveis causais x_t . Se as previsões destas variáveis exógenas não forem apropriadas, o modelo de regressão dinâmica irá também gerar previsões inadequadas.

Isto caracteriza um aspecto importante dos modelos de regressão dinâmica que consiste na possibilidade de consecução de cenários ao se chegar a um modelo relacional de variáveis dependentes em relação a variáveis explicativas. Isto é, surge a possibilidade de montagem de vários cenários para as variáveis causais o que enriquece, por exemplo, qualquer trabalho de planejamento de vendas de um determinado produto. Feitas estas considerações, o fluxograma a seguir indica, de maneira genérica, os passos usados na construção de um modelo de regressão dinâmica.

⁸ Na verdade, “casos estranhos” como este ocorrem na prática com certa frequência e uma possível saída é olhar para o coeficiente da variável preço com alguma defasagem, e verificar se é possível encontrar algum resultado coerente.

Figura 1: Construção de um modelo de regressão dinâmica.



Fonte: Zanini (2000).

Como dito anteriormente, a especificação correta de um modelo de regressão dinâmica envolve a precisa especificação da relação causal entre as variáveis e da estrutura dinâmica do modelo.

Nos modelos de regressão dinâmica podem ser usadas também variáveis de intervenção (ou variáveis dummy). O objetivo deste tipo de procedimento é considerar situações atípicas como, por exemplo, aumento das vendas de brinquedo no Natal e no Dia da Criança (neste caso é razoável incluir dummies para os meses de dezembro e outubro no modelo). O mesmo procedimento pode ser usado para levar em conta os efeitos de situações incomuns como greves e planos econômicos⁹.

Os modelos de regressão dinâmica incorporam ainda diretamente a sazonalidade da série ao modelo, ao invés de supor que a série será previamente dessazonalizada. Atenta-se que existem duas maneiras de tratar a sazonalidade: via dummies sazonais ou diretamente, através de defasagens na variável dependente ou nos erros estruturados.

⁹ Variáveis dummies são geralmente definidas como 1 (no período de ocorrência do fato relevante) e 0 (fora deste período).

3.4 - TESTES USADOS NOS MODELOS DE REGRESSÃO DINÂMICA

Como dito anteriormente, em regressão dinâmica a construção do modelo envolve vários passos até se chegar a um modelo “final”. Diversos testes da adequação de um modelo de regressão podem ser mencionados. Estes testes¹⁰ são aplicados em diversos estágios da modelagem da série. Temos, por exemplo:

- i) testes como o objetivo de definir a especificação do modelo explicativo;
- ii) testes visando encontrar a dinâmica do modelo, isto é, a inclusão ou não de variáveis defasadas,
- iii) testes para verificar o ajuste do modelo.

Para que se entenda um pouco melhor o modelo de regressão dinâmica que será apresentado para a demanda da gasolina automotiva no Brasil, será abordado de uma forma geral, como se processa o “algoritmo” dos testes na prática:

3.4.1 - TESTES DE VERIFICAÇÃO DA “DINÂMICA” DO MODELO

Como dito anteriormente, a dinâmica de um modelo acontece através dos lags da variável dependente e/ou através da presença de erros estruturados num modelo de Cochrane-Orcutt. A cada momento da elaboração do modelo, são realizados testes de hipóteses¹¹ sobre a “dinâmica” do modelo.

Em todos os casos a seguir, a hipótese nula afirma que a dinâmica do modelo está corretamente especificada, ou seja, a inclusão de outros lags da variável dependente ou outros erros estruturados não é necessária. A hipótese alternativa, em cada caso, representa a necessidade de inclusão de novos termos. Sendo assim tem-se :

¹⁰ A maioria dos testes empregados em regressão dinâmica é uma variante dos testes de Multiplicadores de Lagrange (testes LM) e são baseados na distribuição Qui-Quadrado. (Barros e Souza, 1995).

¹¹ Dudewicz e Mishra (1988).

1) Teste de defasagem da variável endógena

Suponha que a variável dependente Y_t e seus *lags* $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p-1}$ estão presentes no modelo atual. A hipótese alternativa consiste em adicionar a variável defasada Y_{t-p} ao modelo, isto é, adiciona-se o primeiro lag ainda não presente no modelo atual. Se esta variável for considerada significativa, a hipótese nula é rejeitada e deve-se adicionar a variável Y_{t-p} ao modelo.

2) Teste da defasagem sazonal da variável endógena

Este teste é semelhante ao anterior. A hipótese alternativa consiste em adicionar ao modelo atual a variável defasada até o primeiro *lag* sazonal Y_{t-ps} ainda não presente ao modelo. Se o coeficiente de Y_{t-ps} for significativo, esta variável deve ser incluída no modelo, e a hipótese nula deve ser rejeitada.

3) Teste da sequência de defasagens da variável endógena

A hipótese alternativa consiste em adicionar todos os lags da variável dependente que ainda não estão presentes no modelo.

4) Teste da defasagem dos resíduos

A hipótese alternativa consiste em adicionar ao modelo o primeiro termo defasado ε_{t-p} ainda não incluído no modelo atual.

5) Teste da defasagem sazonal dos resíduos

Na hipótese alternativa adiciona-se ao modelo atual o primeiro *lag* sazonal ε_{t-ps} ainda não presente no modelo.

6) Teste da sequência de defasagens dos resíduos

Na hipótese alternativa adiciona-se às variáveis do modelo atual uma seqüência de resíduos defasados ε_{t-1} , ε_{t-2} , ε_{t-S} onde S é o período sazonal. É importante ressaltar que, na hipótese alternativa, inclui-se apenas os resíduos ainda ausentes no modelo atual.

3.4.2 - TESTES PARA A ESPECIFICAÇÃO DAS VARIÁVEIS CAUSAIS

O objetivo de todos estes testes é verificar se a inclusão de uma ou mais variáveis ainda não contempladas no modelo resulta numa melhora do ajuste. Atenta-se para o fato de que os testes para a especificação de variáveis causais não se referem à parte dinâmica do modelo, e não tratam da inclusão de lags da variável dependente e de erros estruturados¹².

1) Teste das variáveis causais excluídas

Neste teste verifica-se a necessidade de inclusão de cada uma das variáveis (escolhidas previamente para análise) mas que ainda não estão presentes no modelo. Se quaisquer destas variáveis são consideradas significantes, deve-se incluí-las no modelo (talvez sequencialmente) e “rodar” a mesma bateria de testes para verificar se a inclusão foi vantajosa.

2) Teste de tendência temporal

Este teste corresponde à inclusão de uma variável do tipo $X_t=t$ no modelo. Esta variável é útil em casos onde a série dependente não é estacionária.

3) Teste da defasagem das variáveis exógenas (causais)

¹² Estes são realizados nos testes para a “dinâmica”.

Na hipótese alternativa inclui-se um lag adicional das variáveis causais já presentes no modelo atual.

4) Teste para a presença de funções não lineares das variáveis exógenas

Neste teste inclui-se o quadrado de cada variável exógena já presente no modelo. Todos os quadrados das variáveis exógenas são incluídos de uma só vez e, portanto, é necessário buscar quais (ou qual) quadrados são realmente significantes.

5) Teste do fator comum

Este teste é realizado só quando o modelo inclui erros estruturados. Sob a hipótese alternativa, a autorregressão dos erros é eliminada, e todos os lags da variável dependente e das causais são adicionados ao modelo. Se a hipótese nula é rejeitada, existe evidência de que um modelo mais geral deveria ser considerado, ao invés do modelo de Cochrane-Orcutt. O grande problema é descobrir em que direção deve-se generalizar o modelo corrente, e não existe uma resposta única para esta questão.

3.4.3 - TESTES BASEADOS NA AUTOCORRELAÇÃO DOS RESÍDUOS¹³

Como pode ser visto, o processo de construção de um modelo de regressão dinâmica deve levar em conta diversos diagnósticos com o objetivo de verificar se o modelo atual é apropriado. Em particular, deve-se sempre examinar o gráfico das autocorrelações dos resíduos. Se estas são significantes para alguns lags, alguma característica da variável dependente não foi capturada pelo modelo atual. Por exemplo, no caso de dados mensais, se a autocorrelação dos resíduos é significativa no lag 12, a observação situada num período genérico $t-12$ meses é relevante para explicar a observação no período t , e sua inclusão no modelo possivelmente resultará num decréscimo dos erros de previsão do modelo.

¹³ Podem ser feitos ainda testes para verificar a existência de variações na variância dos resíduos, ou seja, procura-se detectar a heterocedasticidade da série de resíduos.

Segundo Barros e Souza (1995), a existência de autocorrelações significantes nos resíduos pode então indicar uma das seguintes situações: 1) deve-se incluir mais lags da variável dependente ou 2) deve-se incluir lags adicionais das variáveis exógenas já presentes no modelo ou incluir novas variáveis causais.

Enfim, é importante saber que, em qualquer das situações mencionadas, o fato dos resíduos apresentarem autorrelações significantes indica que algum tipo de estrutura presente na série Y_t não foi captada pelo modelo em consideração.

4 ANÁLISE DE DADOS E RESULTADOS

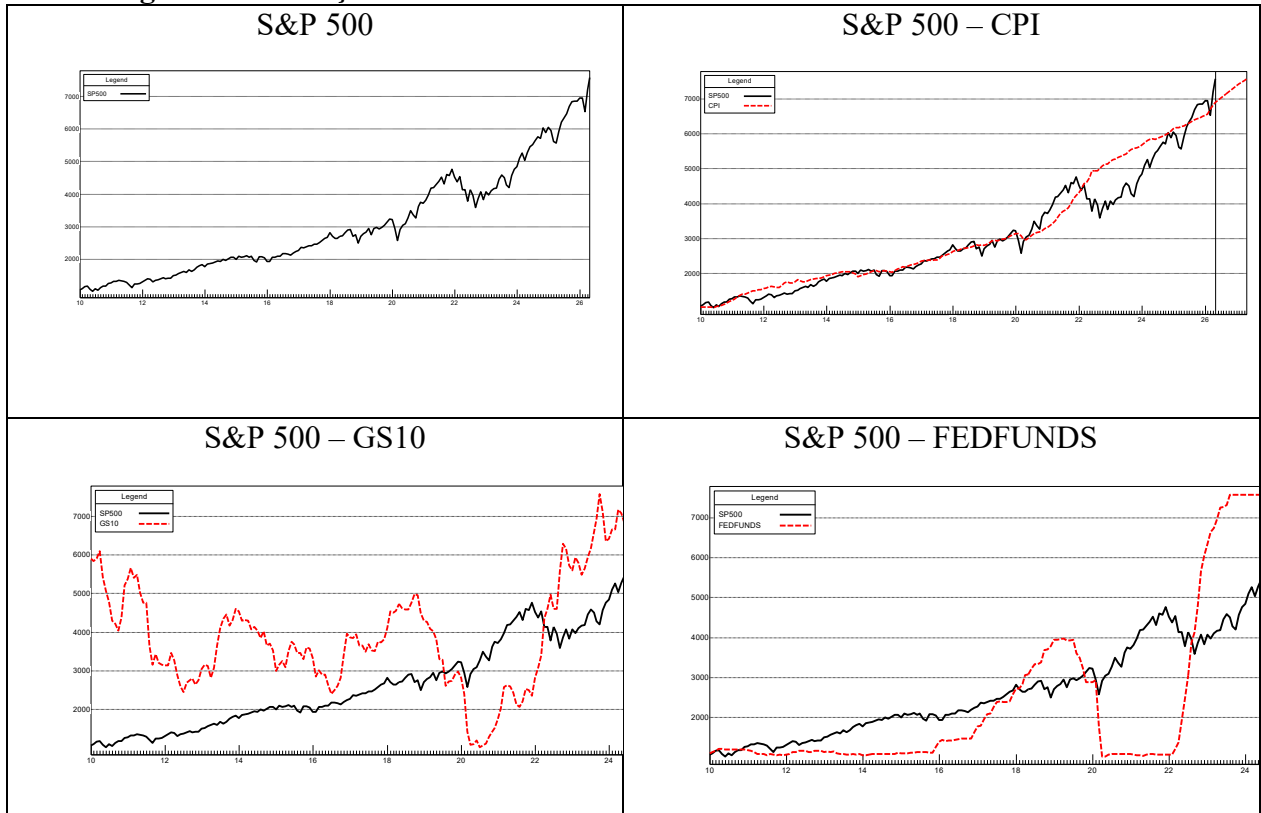
A análise foi realizada utilizando o software Forecast Pro for Windows (FPW), aplicando-se o Método de Regressão Dinâmica (MRD) descrito no capítulo anterior. É apresentada a seguir a evolução das variáveis e análise do modelo final. Além disso, pode-se ver as projeções geradas por esse modelo.

4.1 AS VARIÁVEIS

Na figura 2 a seguir é possível ver a variável de interesse deste trabalho monográfico a evolução histórica da mesma com as demais variáveis a serem testadas. Na primeira imagem tem-se o próprio S&P 500, mostrando uma tendência de alta durante todo o período da amostra. Nos gráficos observa-se dois momentos de maior volatilidade 2020 e 2022, pandemia e o ciclo de aperto monetário.

No gráfico do CPI é visto um crescimento em linha com o do S&P 500, como uma subida mais estável. Já nos gráficos da taxa de juros FEDFUNDS e GS10 é observado uma maior volatilidade e menor aderência ao índice analisado.

Figura 2: Evolução histórica das variáveis – Janeiro de 2010 a Maio de 2026



Fonte: Elaboração própria.

4.2 EQUAÇÃO DE PREVISÃO

Na aplicação do Método de Regressão Dinâmica (MRD), foram testadas as seguintes variáveis: Fed Funds Rate (FEDFUNDS), *Treasury 10y* (GS10) e Consumer Price Index (CPI). Após a aplicação da metodologia expressa no capítulo 3 (figura 1), o modelo obtido através do MRD foi:

$$\ln(\text{SP500})_t = -0,53^{14} + 0,15\ln(\text{CPI})_t^{15} + 0,96\ln(\text{SP500})_{t-1} \quad (4.1)$$

Onde:

SP500_t = Índice S&P 500 no tempo t ;

CPI_t = Consumer Price Index (CPI) no tempo t ;

SP500_{t-1} = Índice S&P 500 no tempo $t-1$;

Com essa equação, pode-se analisar o comportamento da variável dependente (S&P 500) em t através das variáveis explicativas CPI e o S&P 500 defasado em um período. Nota-se que o modelo está em formato LOG-LOG, ou seja, os coeficientes de regressão podem ser interpretados como elasticidades¹⁶.

O modelo final indica que há uma relação estimada do S&P 500 do mês anterior no mês seguinte, relação positiva e estatisticamente significativa, o que é esperado para mercados mais maduros. Já a interpretação do CPI, índice de preços ao consumidor americano, indica que, neste modelo, 1% de variação nesse índice está associado a um aumento de 0,15% no S&P 500¹⁷. Ressalta-se que essa relação positiva pode refletir uma tendência de crescimento ao longo do tempo para ambas as variáveis. As variáveis FEDFUNDS e GS10 foram testadas através dos testes diagnósticos e se mostraram não significativas para o modelo.

A hipótese de ausência de autocorrelação dos erros é fundamental para avaliar a fidedignidade do modelo. Através do teste de Ljung-Box, foi aceita a hipótese de erros não correlacionados (estatística de Ljung-Box igual a 21,03 com p-valor igual a 0,72), como pode-se ver na figura 3. Vê-se nesta figura (Função de Autocorrelação dos Erros) que os coeficientes de correlação dos erros do modelo estimado, para diferentes defasagens do erro, estão dentro da faixa de -0,2 a 0,2, configurando uma correlação fraca ou próxima de zero.

¹⁴ Significante ao nível de significância de 10%.

¹⁵ Significante ao nível de significância de 8%.

¹⁶ Para maiores detalhes, ver Zanini (2000).

¹⁷ Importante salientar que não se está falando numa relação de causalidade, mas apenas numa associação ou relação estimada ou correlação condicional ao modelo em análise. Como S&P 500 e CPI apresentam tendência crescente ao longo do tempo, a relação positiva pode apenas refletir uma tendência comum entre séries em nível.

Figura 3: Função de Autocorrelação dos erros do modelo

Fonte: Elaboração própria

Na tabela 3 pode-se verificar algumas medidas que refletem o desempenho preditivo do modelo. Observa-se que o modelo estimado explica 99,37% da evolução do S&P5 00, podendo refletir principalmente a influência do próprio índice defasado, e apresenta um erro médio absoluto percentual igual a 3,17%, ou seja, ao se prever o próximo mês, o modelo erra, em média, cerca de 3% para baixo ou para cima. É importante ressaltar que um R^2 ajustado próximo de 99% deve ser tratado com cuidado, porque séries financeiras em nível, especialmente índices acionários, tendem a ser fortemente persistentes. Assim, um R^2 elevado pode refletir principalmente a presença do S&P 500 defasado na equação, e não necessariamente forte capacidade preditiva econômica.

Tabela 3 – Estatísticas de Desempenho do Modelo (%)

R^2 ajustado	MAPE
99,37	3,17

Fonte: Elaboração própria

4.3 PREVISÕES

As previsões do S&P 500 foram calculadas através do Método de Regressão Dinâmica (MRD), equação 4.1. Atenta-se que uma das vantagens de se trabalhar com o MRD, é que diferentes cenários podem ser gerados para a variável explicativa presente na equação, verificando se seu impacto na variável de interesse. Neste trabalho, a variável CPI foi autoprojeta através de um modelo univariado selecionado numa competição entre o Método de Amortecimento Exponencial (MAE)¹⁸ e o Método de Box & Jenkins (MBJ)¹⁹. Foi realizada uma análise recursiva fora da amostra (out-of-sample rolling evaluation)²⁰, selecionando como melhor modelo aquele que minimizou o erro absoluto acumulado fora da amostra. Atenta-se que o método vencedor foi o de Box & Jenkins, estimando-se um ARIMA (1,1,2)(0,1,1) com transformação logarítmica da variável para se projetar o CPI para o horizonte de doze meses. Para realizar esta análise recursiva fora da amostra, foram retirados os últimos 12 meses (junho de 2025 a maio de 2026) da variável CPI, fazendo-se um deslocamento da origem de previsão no período fora da amostra. Ao final do processo de análise, foram geradas 12 previsões para 1 mês à frente, 11 previsões para 2 meses à frente, 10 previsões para 3 meses à frente e assim sucessivamente até 1 previsão para 12 meses à frente. Em sendo assim, ao final do processo foi gerado um total de 78 previsões. Foi escolhido como melhor modelo para projetar o CPI aquele que minimizou o erro acumulado fora da amostra. Importante ressaltar que este procedimento de análise foi utilizado apenas para selecionar o modelo que seria utilizado para gerar os valores futuros o CPI de forma a se alimentar o Modelo de Regressão Dinâmica estimado (equação 4.1). Não foi feita nenhuma comparação do MRD com outros métodos.

Na tabela 4 a seguir, pode-se observar as previsões para o S&P 500 com o respectivo intervalo de confiança de 95%.

¹⁸ Montgomery & Johson (1990).

¹⁹ Box & Jenkins (1994).

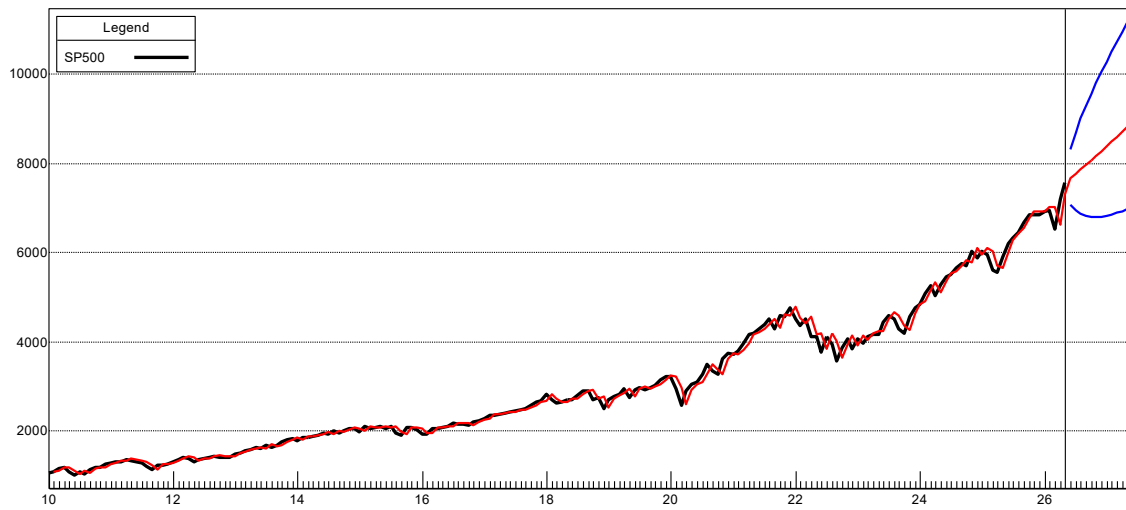
²⁰ Goodrich & Stellwagen (1999).

Tabela 4 – Previsão do S&P 500 – Intervalo de Confiança de 95%

Período	Limite Inferior	Previsão	Limite Superior
2026-06	7076,20	7673,54	8321,30
2026-07	6941,07	7768,89	8695,45
2026-08	6867,89	7865,97	9009,09
2026-09	6827,42	7964,87	9291,81
2026-10	6808,39	8065,75	9555,31
2026-11	6804,90	8168,59	9805,57
2026-12	6813,29	8273,21	10045,96
2027-01	6831,38	8379,73	10279,02
2027-02	6857,49	8488,03	10506,28
2027-03	6889,84	8597,35	10728,02
2027-04	6927,79	8707,94	10945,53
2027-05	6970,83	8820,05	11159,82

Fonte: Elaboração própria,

Atenta-se que uma das vantagens de se trabalhar com modelos estatísticos é que as previsões não são pontuais, mas probabilísticas. Isto significa que não se calcula apenas um número, a previsão, mas um intervalo de confiança que possui determinada chance de conter o valor esperado. Analisando-se a tabela 4, verifica-se, com um nível de confiança igual a 95%, um intervalo amplo, o que pode ser um reflexo da volatilidade desta variável no decorrer do tempo. Outro ponto a se destacar é que, como dito anteriormente, estas previsões são resultado de um determinado cenário para a variável CPI e pode se alterar dependendo de cenários alternativos utilizados. Na figura 4 a seguir pode-se ver as previsões na forma probabilística.

Figura 4: Previsão do S&P 500

Fonte: Elaboração própria. Nota: ____ Valor Real ____ Valor Ajustado/Previsões ____ Intervalo de Confiança

5 CONCLUSÃO

Este trabalho teve como objetivo analisar a evolução histórica do S&P 500 com o objetivo de calcular o comportamento futuro do índice, utilizando dados de janeiro de 2010 até maio de 2026. O S&P 500 pode ser considerado um dos principais índices do mundo, pois reúne empresas de grande capitalização do mercado acionário mais relevante em termos globais

O trabalho usou como método a Análise de Regressão Dinâmica que, a partir de uma estrutura corrente, aplica um conjunto de testes diagnósticos baseados em Lagrange Multiplier para se chegar ao modelo final. As análises captaram um componente de dinâmica na evolução do S&P 500 através da correlação estimada com o valor deste índice no mês anterior. Os testes diagnósticos indicaram ainda uma relação positiva com o Consumer Price Index (CPI). As variáveis FEDFUNDS e GS10 não tiveram significância estatística e, por isso, não entraram na equação final.

Foi possível verificar algumas vantagens em se trabalhar com métodos estatísticos matemáticos de previsão. A partir da equação estimada pode-se, por exemplo, gerar cenários distintos para as variáveis explicativas presentes na equação, verificando-se seu impacto no comportamento do S&P 500. Neste trabalho gerou-se um cenário autoprojetoado, ou seja, o CPI foi autoprojetoado através de uma competição entre dois métodos quais sejam o Método de Amortecimento Exponencial e o Método de Box & Jenkins. A partir das projeções da variável explicativa presente na equação, foi possível projetar o S&P 500 para o período de 12 meses à frente (junho de 2026 a maio de 2027) através de uma equação com cerca de 99% de poder de explicação e 3% de erro médio absoluto percentual ao se projetar o mês subsequente. Estimou-se um cenário de alta no índice, movimento esperado pelo modelo ser composto pela tendência recente do próprio índice, ou seja, pela dinâmica captada pelo Método de Regressão Dinâmica (MRD).

Este trabalho aplicou o MRD para a previsão do S&P 500, principal índice de ações americano, sendo o primeiro, a princípio, a utilizar esse método para tal finalidade. Como sugestão de estudos futuros, recomenda-se incluir mais variáveis no conjunto de testes diagnósticos e também a ampliação do período amostral. Pode-se também promover uma competição entre métodos de previsão.

6 REFERÊNCIAS BIBLIOGRÁFICAS

B3, Disponível em: https://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/perfil-pessoas-fisicas/perfil-pessoa-fisica/

B3, Ibovespa: https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm) Acesso em: 2026.

BARROS, M; SOUZA, R, C, Regressão Dinâmica, Núcleo de Estatística Computacional, PUC-Rio, 1995.

BOX, G, E, P; JENKINS, G, M, *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day, 1994.

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM, Federal Funds Effective Rate, St, Louis: Federal Reserve Economic Data (FRED), 2026, Disponível em: <https://fred.stlouisfed.org/series/FEDFUNDS>, Acesso em: 2026.

CHEN, N,F; ROLL, R; ROSS, S, A, Economic Forces and the Stock Market, The Journal of Business, 1986.

COCHRANE, D; ORCUTT, G, H, Application of least squares regression to relationships containing auto-correlated error terms, Journal of the American Statistical Association, 1949.

DUDEWICZ, E, J; MISHRA, S, N, Modern Mathematical Statistics, Wiley, 1988.

FAMA, E, F, Stock Returns, Real Activity, Inflation and Money, The American Economic Review, 1981.

GOODRICH, R,L, Applied Statistical Forecasting, Belmont, Business Forecast Systems, 1989.

GOODRICH, R,L; STELLWAGEN, E,A, *Forecast Pro for Windows*, Business Forecast Systems, 1999.

HAMILTON, J, D, Time Series Analysis, Princeton: Princeton University Press, 1994.

JIAO, Yang, JAKUBOWICZ, Jeremie, "Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks," 2017 IEEE International Conference on Big Data (Big Data), IEEE, 2017.

MAO, Yuexin, et al, "Correlating S&P 500 stocks with Twitter data," Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research, 2012.

MARKOWITZ, Harry, Portfolio Selection, The Journal of Finance, vol, VII, n, 1, March 1952.

MONTGOMERY, D,C,, JOHNSON, L,A,, *Forecasting and Time Series Analysis*, New York, McGraw-Hill Book Co,, 1990.

S&P Global, Disponível em: <https://www.spglobal.com/spdji/pt/indices/equity/sp-500/#overview>

U,S, BUREAU OF LABOR STATISTICS, Consumer Price Index for All Urban Consumers: All Items in U,S, City Average, St, Louis: Federal Reserve Economic Data (FRED), 2026, Disponível em: <https://fred.stlouisfed.org/series/CPIAUCSL>, Acesso em: 2026.

VIEIRA, Mairton Oliveira, Os efeitos da inovação sobre o risco sistemático das empresas de tecnologia listadas no S&P500, TCC (Graduação em Administração) – Universidade Federal de Sergipe, Itabaiana, 2024.

WORLD BANK, Market capitalization of listed domestic companies, Disponível em: <https://data.worldbank.org/indicator/CM,MKT,LCAP,CD>, Acesso em: 2026.

ZANINI, Alexandre, Redes Neurais e Regressão Dinâmica: Um Modelo Híbrido para Previsão de Curto Prazo da Demanda de Gasolina Automotiva no Brasil, Dissertação de Mestrado, PUC-Rio, 2000.