

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-Graduação em Estatística

Delmário Prates Alves

**Regressão Linear Integrada ao
Modelo de Markowitz:
Quantificando a Incerteza na
Fronteira Eficiente de Portfólios e
sua Relação com Fatores
Macroeconômicos**

Belo Horizonte

2025

Delmário Prates Alves

**Regressão Linear Integrada ao Modelo de Markowitz:
Quantificando a Incerteza na Fronteira Eficiente de
Portfólios e sua Relação com Fatores Macroeconômicos**

Monografia de Especialização apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Especialista em Estatística.

Orientador: Prof. Dr. Guilherme Lopes de Oliveira

Belo Horizonte

2025

	Alves, Delmário Prates.
A474r	<p>Regressão linear integrada ao modelo de Markowitz: [recurso eletrônico] quantificando a incerteza na fronteira eficiente de portfólios e sua relação com fatores macroeconômicos / Delmário Prates Alves – 2025. 1 recurso online (51 f. il., color.) : pdf.</p> <p>Orientador: Guilherme Lopes de Oliveira.</p> <p>Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. Referências: f. 49-51.</p> <p>1. Estatística. 2. Análise de regressão. 3. Gestão de portfólio 4. Análise de investimentos. I. Oliveira, Guilherme Lopes de. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.</p> <p style="text-align: right;">CDU 519.2(043)</p>



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 340ª. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE DELMÁRIO PRATES ALVES.

Aos vinte e sete dias do mês de março de 2025, às 10:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística Computacional Aplicada, para julgar a apresentação do trabalho de fim de curso do aluno **Delmário Prates Alves**, intitulado: “*Regressão Linear Integrada ao Modelo de Markowitz: Quantificando a Incerteza na Fronteira Eficiente de Portfólios e sua Relação com Fatores Macroeconômicos*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Guilherme Lopes de Oliveira – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente à candidata pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 27 de março de 2025.

Documento assinado digitalmente



GUILHERME LOPES DE OLIVEIRA
Data: 27/03/2025 11:55:53-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Guilherme Lopes de Oliveira (orientador)
DECOM/CEFET-MG

Roberto da Costa

Quinino:80871291720

Assinado de forma digital por
Roberto da Costa
Quinino:80871291720
Dados: 2025.03.27 11:16:56 -03'00'

Prof. Dr. Roberto da Costa Quinino
DEST/UFMG

Resumo

Este estudo investiga como a análise de regressão linear pode complementar o modelo de Markowitz na seleção e otimização de carteiras de investimentos. Foram ajustados modelos de regressão para prever os retornos de dois ativos com base em variáveis macroeconômicas, e os resultados foram integrados à teoria de Markowitz para construir uma nova fronteira eficiente. Os ativos escolhidos da base de dados da bolsa de valores do Brasil (IBOVESPA) foram o BBAS3 e IVVB11, em escala mensal de 2017 a 2024, sendo os dados de 2024 usados apenas para avaliação das previsões de retorno do portfólio. Dentre os indicadores macroeconômicos foram considerados o índice IBOVESPA e cotação do dólar. A comparação entre os retornos históricos, previstos e reais demonstrou que o uso de modelagem estatística melhora a precisão das previsões, reduzindo o erro médio (RMSE) e aproximando os retornos estimados dos valores efetivamente observados. Os resultados indicam que a inclusão de modelos de regressão pode oferecer estimativas mais precisas do que o simples uso de médias históricas (RMSE= 0,009 versus RMSE=0,014), reforçando a importância da previsão econômica na tomada de decisão. No entanto, destaca-se que previsões eficazes exigem estimativas confiáveis dos indicadores macroeconômicos futuros.

Palavras-chave: estatística; regressão linear; Markowitz; teoria moderna do portfólio; previsão de retornos; fronteira eficiente.

Abstract

This study investigates how linear regression analysis can complement the Markowitz model in the selection and optimization of investment portfolios. Regression models were adjusted to predict the returns of two assets based on macroeconomic variables, and the results were integrated into Markowitz's theory to construct a new efficient frontier. The assets chosen from the Brazilian Stock Exchange (IBOVESPA) database were BBAS3 and IVVB11, on a monthly scale from 2017 to 2024, with the 2024 data being used only to evaluate the portfolio return predictions. Among the macroeconomic indicators, the IBOVESPA index and the dollar exchange rate were considered. The comparison between historical, predicted and actual returns demonstrated that the use of statistical modeling improves the accuracy of forecasts, reducing the mean error (RMSE) and bringing estimated returns closer to actually observed values. The results indicate that the inclusion of regression models can offer more accurate estimates than the simple use of historical averages (RMSE= 0.009 versus RMSE=0.014), reinforcing the importance of economic forecasting in decision making. However, it should be noted that effective forecasts require reliable estimates of future macroeconomic indicators.

Keywords: statistic; linear regression; Markowitz; modern portfolio theory; return forecast; efficient frontier.

Lista de Figuras

2.1	Gráfico da fronteira eficiente dos ativos de risco A e B hipotetizados . . .	18
3.1	Distribuição e Dispersão dos Retornos	26
3.2	Matriz de Correlação dos Ativos e Indicadores	27
3.3	Relação entre BBAS3 e indicadores	28
3.4	Relação entre IVVB11 e indicadores	28
3.5	Fronteira eficiente da carteira baseada nos retornos históricos	31
3.6	Estimativa dos Coeficientes da Regressão Linear para o Retorno de BBAS3	33
3.7	Análise dos Resíduos do Modelo de Regressão para o Retorno de BBAS3	34
3.8	Análise dos Resíduos do Modelo de Regressão para o Retorno de IVVB11 (modelo1)	35
3.9	Estimativa dos Coeficientes da Regressão Linear para o Retorno de IVVB11	36
3.10	Análise dos Resíduos do Modelo de Regressão para o Retorno de IVVB11 (modelo_autoreg)	38
3.11	Fronteira eficiente da carteira baseada nos retornos previstos	39
3.12	Fronteira eficiente da carteira baseada nos retornos reais	40
3.13	Fronteira eficiente: Histórica x Prevista x Real	41
3.14	Retorno Real, Histórico e Previsto com Intervalo de Confiança	44

Lista de Tabelas

3.1	Comparação dos Modelos de Regressão para o IVVB11	36
3.2	Visão parcial dos dados do dataframe com a simulação de diferentes composições da carteira	42

Sumário

1	Introdução	9
2	Metodologia	12
2.1	Dados	12
2.2	Modelo de Markowitz	14
2.2.1	Principais tópicos de estatística utilizados pelo Modelo	14
2.2.2	O Modelo de Markowitz e a sua aplicação	15
2.3	Modelo de Regressão Linear	19
2.4	Procedimentos Metodológicos	24
3	Resultados	26
3.1	Análise exploratória dos dados	26
3.2	Aplicação do Modelo de Markowitz baseada nos retornos históricos (2017-2023)	29
3.3	Aplicação do Modelo de Markowitz baseada nos retornos previstos para 2024, via regressão linear	32
3.3.1	Modelos de Regressão Ajustados	32
3.3.2	Fronteira eficiente baseada nos retornos previstos para 2024	38
3.4	Aplicação do Modelo de Markowitz baseada nos retornos reais de 2024	39
3.5	Comparação dos gráficos das fronteiras eficientes baseadas nos dados históricos, previstos e reais	40
3.6	Comparação dos retornos dos modelos	41
4	Considerações Finais	46
	Referências	49

1 Introdução

A decisão de investir envolve um equilíbrio entre risco e retorno. Nesse contexto, a Estatística se mostra uma ferramenta indispensável. O uso da Estatística é de suma importância no mercado de capitais, propiciando bases sólidas para o desenvolvimento de técnicas financeiras (VIRGILLITO, 2017). Ao quantificar as incertezas e projetar resultados futuros, ela permite que os investidores tomem decisões mais embasadas e organizem seus portfólios, ou seja, o conjunto de ativos financeiros que possuem, de forma mais eficiente.

Um ativo pode ser definido como qualquer recurso ou bem financeiro, como ações, títulos ou imóveis, que possui valor econômico e potencial de retorno (GITMAN e ZUTTER, 2015). Já um portfólio ou carteira representa a combinação desses ativos, estrategicamente composta para balancear risco e retorno, conforme os objetivos do investidor (MARKOWITZ, 1952). Para auxiliar os investidores a encontrar o equilíbrio ideal entre esses dois fatores, a Teoria Moderna de Portfólio (TMP) se tornou uma referência. Desenvolvida por Harry Markowitz na década de 1950, a TPM utiliza ferramentas estatísticas para otimizar a composição de um portfólio de investimentos. Ao considerar o retorno esperado, a volatilidade e a correlação entre os ativos, essa teoria permite identificar a combinação de investimentos que oferece o melhor retorno para um determinado nível de risco.

Markowitz introduziu o conceito de que todo investimento apresenta um retorno esperado e uma variabilidade em torno desse retorno, que ele denominou variância. Para ele, o retorno é desejável, enquanto a variância, que representa o risco, é indesejável. Seu objetivo era construir um modelo que maximizasse o retorno esperado, minimizando ao mesmo tempo a variância da carteira. Para construir seu modelo, Markowitz estabeleceu algumas premissas sobre o comportamento dos investidores (SÁ, 1999; MACEDO JR, 2003), como:

- Aversão ao risco: Investidores preferem menor risco para o mesmo retorno.
- Busca por maior retorno: Diante de riscos iguais, preferem maior retorno.
- Relevância da variabilidade: A variação dos retornos indica o nível de risco.

- Simplicidade da avaliação: Investidores consideram apenas retorno esperado e variância.
- Irrelevância de custos: Custos de transação e impostos são ignorados no modelo.

A partir dessas informações, a TMP permite construir a chamada "fronteira eficiente", que representa o conjunto de todas as carteiras que oferecem o maior retorno possível para um determinado nível de risco. A escolha da carteira ideal dentro dessa fronteira dependerá da tolerância ao risco de cada investidor.

Um dos pontos a se considerar nessa teoria é que nem sempre a rentabilidade histórica dos ativos pode se repetir no futuro. Para aprimorar essa abordagem, estudos recentes têm explorado a integração de outros modelos estatísticos, como a regressão linear, para melhorar a estimação dos retornos esperados e minimizar incertezas. Brodie et al. (2007), por exemplo, propõem uma reformulação do modelo de Markowitz como um problema de regressão linear com restrições, introduzindo penalização para incentivar portfólios mais esparsos e reduzir a sobrealocação de ativos. Já Ferreira (2015) aplicou regressão linear para testar a validade do modelo CAPM (que é uma extensão do modelo de Markowitz) em diferentes mercados, evidenciando a importância da modelagem estatística na previsão de retornos financeiros.

Nesse sentido, a motivação para este estudo parte do cenário hipotético de um investidor que busca otimizar sua carteira composta por ativos financeiros, levando em consideração indicadores econômicos que podem influenciar seus retornos. O objetivo é identificar a melhor alocação possível que maximize o retorno esperado e minimize o risco associado. Para atingir esse objetivo, este trabalho propõe um estudo de como a incorporação da regressão linear ao modelo de Markowitz pode trazer contribuições valiosas para a análise e otimização de carteiras de investimentos. Enquanto a Teoria Moderna do Portfólio se concentra em maximizar o retorno esperado para um dado nível de risco através da diversificação, a regressão linear oferece uma abordagem estatística para entender as relações entre variáveis econômicas e o desempenho dos ativos.

Por meio da regressão linear, é possível identificar como fatores macroeconômicos, como taxas de juros, inflação e câmbio, influenciam diretamente o retorno de cada ativo. Isso permite refinar as estimativas de risco e retorno, podendo tornar o modelo de Markowitz mais robusto e realista. Dessa forma, o uso da regressão linear pode ajudar a identificar tendências ou variáveis externas que podem impactar o portfólio. Além disso, a abordagem da regressão linear pode permitir a mensuração e consideração da variabilidade histórica dos indicadores externos, viabilizando a simulação e previsão de cenários adversos baseados na incerteza futura sobre tais fatores.

Com base na integração da regressão linear ao modelo de Markowitz, este trabalho tem como objetivo geral explorar como a análise de regressão linear pode complementar o modelo de Markowitz na seleção e otimização de carteiras de investimentos, avaliando

sua eficácia na melhoria da eficiência do portfólio e na interpretação das relações entre variáveis econômicas e retornos financeiros. Como objetivos específicos, temos:

- Analisar os retornos históricos dos ativos e aplicar o modelo de Markowitz, criando a fronteira eficiente;
- Ajustar modelos de regressão linear, com base nos retornos históricos e suas relações com fatores macroeconômicos, para estimar os retornos previstos dos ativos;
- Integrar os retornos previstos pelos modelos de regressão ao modelo de Markowitz para criar uma nova fronteira eficiente;
- Comparar o desempenho das carteiras (portfólios) com base nos retornos previstos e históricos, avaliando a proximidade com os retornos reais.

Para o cumprimento de tais objetivos, no próximo capítulo é apresentada a metodologia utilizada neste estudo. Em seguida são apresentados a aplicação e os resultados obtidos, englobando a análise descritiva dos dados, a aplicação do modelo de Markowitz e a integração da regressão linear à este modelo. Ao final do trabalho são feitas as considerações finais.

2 Metodologia

Neste capítulo, apresentamos a metodologia utilizada neste estudo, detalhando os dados empregados e as abordagens adotadas para a análise.

2.1 Dados

Conforme mencionado anteriormente, a motivação para este estudo surgiu de um cenário hipotético no qual um investidor busca otimizar sua carteira composta por dois ativos financeiros, considerando indicadores econômicos que podem influenciar seus retornos. O objetivo é identificar a melhor alocação possível que maximize o retorno esperado e minimize o risco associado. Para encontrar essa melhor alocação é adotado o modelo de Markowitz.

O modelo de Markowitz utiliza dados passados para o cálculo do retorno esperado. No entanto, sabemos que nem sempre a rentabilidade passada pode se manter no futuro. Consideremos então, em um cenário real, que o investidor teria acesso a projeções futuras dos indicadores econômicos, permitindo estimar os retornos esperados dos ativos antes da tomada de decisão. O investidor poderia criar modelos de regressão ajustados com base no período histórico para calcular os retornos previstos dos ativos.

Para este estudo foram escolhidos dados mensais de 2 ativos e 6 índices ou indicadores econômicos, no período entre janeiro/2017 a dezembro/2024, no total de 96 registros, obtidos no site <<https://br.investing.com>> em 05/01/2025.

Para estimar os retornos previstos dos ativos no cenário futuro, foram ajustados modelos de regressão linear múltipla, nos quais os retornos dos ativos foram modelados em função de indicadores econômicos. Esses modelos foram criados com os dados históricos (2017-2023) e, posteriormente, aplicados aos valores dos indicadores econômicos observados em 2024, gerando estimativas dos retornos futuros.

Dessa forma, o estudo ilustra como um investidor poderia utilizar previsões econômicas para otimizar sua carteira de investimentos. Em um cenário real, as projeções de 2024 seriam baseadas em expectativas de mercado, mas, para fins deste trabalho, utilizamos os dados efetivamente observados em 2024 para simular o impacto da informação prospectiva na tomada de decisão.

Em relação aos ativos, foram selecionadas as ações do Banco do Brasil (BBAS3) como representante do mercado brasileiro e o ETF (Fundo de Índice) IVVB11, que acompanha o índice S&P500, como representante do mercado americano.

O BBAS3, sendo uma ação de um grande banco brasileiro, é sensível às condições econômicas nacionais, como taxas de juros e crescimento econômico. Já o IVVB11, ao acompanhar o S&P500, oferece exposição a uma ampla gama de empresas americanas e proporciona uma diversificação geográfica. Ao analisar a relação entre o retorno e o risco desses dois ativos, buscamos construir um portfólio que maximize o retorno esperado para um determinado nível de risco. Conforme veremos adiante, o Modelo de Markowitz preconiza que os ativos sejam pouco correlacionados porque a diversificação reduz o risco total do portfólio sem necessariamente diminuir o retorno esperado. A escolha desses dois ativos levou em consideração essa condição.

Já a seleção dos indicadores econômicos e financeiros utilizados neste estudo foi feita com o objetivo de capturar variáveis relevantes que influenciam o desempenho dos ativos e a dinâmica do mercado. Foram considerados índices de referência, taxas macroeconômicas e variáveis cambiais que afetam tanto o mercado brasileiro quanto o internacional. Como indicadores ou índices econômicos foram escolhidos: S&P500, NASDAQ, SELIC, IBOVESPA, IPCA e DOLAR.

A escolha dos indicadores econômicos foi feita com o objetivo de capturar as principais forças que movem os mercados financeiros, tanto no Brasil quanto nos Estados Unidos.

O S&P500 é um dos principais índices acionários dos Estados Unidos, composto pelas 500 maiores empresas listadas nas bolsas NYSE (Bolsa de Valores de Nova York) e NASDAQ, sendo utilizado como referência para medir o desempenho do mercado americano (S&P GLOBAL, 2025). O NASDAQ também representa o mercado dos EUA, mas é voltado principalmente para empresas de tecnologia, como Apple, Microsoft e Google, sendo um termômetro desse setor (NASDAQ, 2025). O IBOVESPA é o principal índice da Bolsa de Valores do Brasil (B3), refletindo o desempenho das ações mais negociadas e representativas do mercado brasileiro (B3, 2025).

Além dos índices acionários, foram considerados indicadores macroeconômicos que afetam diretamente os investimentos. A taxa Selic é a taxa básica de juros da economia brasileira, definida pelo Banco Central do Brasil, influenciando o custo do crédito, o retorno de investimentos de renda fixa e a atratividade da renda variável (BANCO CENTRAL DO BRASIL, 2025). O IPCA é o índice oficial de inflação no Brasil, calculado pelo IBGE, e serve como referência para políticas econômicas e ajustes de preços (IBGE, 2025). Por fim, a cotação do dólar representa a taxa de câmbio entre o real e a moeda americana, impactando empresas exportadoras, importadores e ativos dolarizados, como o ativo IVVB11 (BANCO CENTRAL DO BRASIL, 2025).

Essa combinação de indicadores permite uma análise mais abrangente da relação

entre os ativos e o cenário econômico, considerando tanto fatores internos quanto externos.

2.2 Modelo de Markowitz

2.2.1 Principais tópicos de estatística utilizados pelo Modelo

O cálculo do valor esperado e da variância estão entre os principais tópicos estatísticos utilizados pelo Modelo de Markowitz. Abaixo citamos a definição de valor esperado e a propriedade do cálculo da variância de uma combinação linear.

Começemos com a definição dada por Morettin e Bussab (2017) :

Definição 1 (Valor esperado). Dada a variável aleatória X discreta, assumindo os valores x_1, \dots, x_n, \dots , chamamos valor esperado ou esperança matemática de X ao valor

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i,$$

sendo que a soma “infinita” dessa fórmula deve ser “convergente”.

Em outras palavras, conforme Carvalho e Morgado (2015) , o valor esperado de um resultado aleatório numérico é definido como sendo a média ponderada de seus possíveis valores em que os pesos são as respectivas probabilidades. Isto é, se os possíveis valores para o resultado são x_1, x_2, \dots, x_n , com probabilidades p_1, p_2, \dots, p_n , seu valor esperado é $x_1 p_1 + x_2 p_2 + \dots + x_n p_n$ (note que a soma de todos os pesos é igual a 1). Ainda, se todos os valores possíveis x_1, x_2, \dots, x_n forem equiprováveis, ou seja, tenham a mesma probabilidade de ocorrer, $E(X) = \frac{1}{n} \sum_{i=1}^n x_i$ representa a média aritmética simples dos n possíveis valores.

Veremos agora uma propriedade importantíssima da variância que será usada na Teoria Moderna do Portfólio. As propriedades relativas ao valor esperado e à variância utilizadas na demonstração abaixo podem ser encontradas em ALVES (2020).

Propriedade 1. Seja $Y = \sum_{i=1}^n \alpha_i X_i$ uma combinação de variável aleatória. A variância σ_Y^2 dessa combinação linear será dada por

$$\sigma_Y^2 = \sum_{i=1}^n \alpha_i^2 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \alpha_i \alpha_j \text{Cov}(X_i, X_j)$$

Demonstração. Vamos mostrar o cálculo da variância para o caso particular $Y = \alpha_1 X_1 + \alpha_2 X_2$, que será o utilizado no desenvolvimento deste trabalho, sendo que a demonstração para o caso geral pode ser feita aplicando indução em n . Assim, temos:

$$\begin{aligned}
\sigma_Y^2 &= E \{ [Y - E(Y)]^2 \} \\
&= E \{ [\alpha_1 X_1 + \alpha_2 X_2 - E(\alpha_1 X_1 + \alpha_2 X_2)]^2 \} \\
&= E \{ [\alpha_1 X_1 + \alpha_2 X_2 - E(\alpha_1 X_1) - E(\alpha_2 X_2)]^2 \} \\
&= E \{ [\alpha_1 X_1 + \alpha_2 X_2 - \alpha_1 E(X_1) - \alpha_2 E(X_2)]^2 \} \\
&= E \{ [\alpha_1 (X_1 - E(X_1)) + \alpha_2 (X_2 - E(X_2))]^2 \} \\
&= E \{ \alpha_1^2 (X_1 - E(X_1))^2 + 2\alpha_1 \alpha_2 (X_1 - E(X_1))(X_2 - E(X_2)) + \alpha_2^2 (X_2 - E(X_2))^2 \} \\
&= \alpha_1^2 E(X_1 - E(X_1))^2 + \alpha_2^2 E(X_2 - E(X_2))^2 + 2\alpha_1 \alpha_2 E(X_1 - E(X_1))(X_2 - E(X_2)) \\
&= \alpha_1^2 Var(X_1) + \alpha_2^2 Var(X_2) + 2\alpha_1 \alpha_2 Cov(X_1, X_2)
\end{aligned}$$

□

Como podemos ver, se $Cov(X_i, X_j) = 0$, ou seja, se as variáveis não são correlacionadas, a variância da soma é a própria soma das variâncias. Quando existe dependência ou quando a covariância não for nula, uma nova variável aleatória contribui com o dobro da covariância, podendo aumentar ou diminuir a variabilidade do sistema.

2.2.2 O Modelo de Markowitz e a sua aplicação

A teoria de carteiras de Markowitz objetiva determinar o conjunto de carteiras que irão compor a chamada fronteira eficiente, conjunto este onde, para determinado nível de risco, se obtém o retorno máximo. Para isso é preciso estudar como se comporta o retorno esperado e o risco de uma combinação de ativos.

O retorno esperado de uma carteira é dado pela média ponderada dos retornos esperados dos ativos individuais. Assim, a expressão do retorno esperado de uma carteira p que combine os ativos 1 a n é dada por:

$$E(R_p) = \sum_{i=1}^n \alpha_i E(R_i),$$

onde α_i é a proporção aplicada no ativo i , com $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, e $E(R_i)$ é o retorno esperado do ativo i .

Uma das maneiras de obter os dados do retorno esperado parte da observação de uma amostra de retornos históricos. Presumimos que cada um dos retornos representa um desenho independente da distribuição de probabilidades histórica. Por isso, a cada retorno é atribuída uma probabilidade idêntica. Quando se utiliza uma probabilidade fixa na equação acima, obtém-se a média simples das observações, frequentemente empregada para prever o retorno médio (BODIE, KANE e MARCUS, 2014). Dessa forma, o retorno esperado do ativo i pode ser representado pela média histórica \bar{R}_i , e

o retorno esperado da carteira torna-se:

$$E(R_p) = \sum_{i=1}^n \alpha_i \bar{R}_i.$$

Como o retorno da carteira é uma combinação linear dos retornos dos ativos que a compõem, sua variância também pode ser obtida pela variância dessa combinação linear. Isso ocorre porque, matematicamente, a variância de uma combinação linear de variáveis aleatórias depende não só das variâncias individuais, mas também das covariâncias entre elas. Assim, a fórmula da variância da carteira, com base na Propriedade 1, é dada por:

$$\sigma_p^2 = \sum_{i=1}^n \alpha_i^2 \text{Var}(R_i) + 2 \sum_{1 \leq i < j \leq n} \alpha_i \alpha_j \text{Cov}(R_i, R_j),$$

onde $\text{Var}(R_i)$ e α_i são, respectivamente, a variância dos retornos e a proporção aplicadas no ativo i , com $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, e $\text{Cov}(R_i, R_j)$ é a covariância dos retornos históricos dos ativos i e j , com $i \neq j$.

Com base nesta expressão, vemos que a variância de uma carteira é a soma das contribuições das variâncias dos ativos componentes mais um termo que envolve a covariância (e, conseqüentemente, a correlação) entre os retornos dos ativos componentes. Quando a correlação entre os ativos componentes for pequena ou mesmo negativa, a tendência dos retornos dos ativos se compensarem será maior. Isso diminuirá o risco da carteira. Podemos observar que a variância da carteira é menor quando o coeficiente de correlação é menor.

Dessa forma, o problema geral da Teoria de Markowitz se apresenta da seguinte forma:

“Determinar os percentuais investidos nos vários ativos que irão compor as carteiras de tal forma que para cada nível de retorno esperado $E(R_p)$ da carteira o risco seja mínimo, sujeito à condição de $\sum_{i=1}^n \alpha_i = 1$, ou seja, que todos os recursos sejam investidos nos ativos da carteira” (SÁ, 1999).

Para melhor ilustrarmos como funciona a determinação da fronteira eficiente, vamos exemplificar com uma carteira constituída somente por dois ativos, A e B, e determinar a expressão do retorno esperado e do risco dessa carteira. Com relação ao retorno esperado da carteira, a fórmula que permite o seu cálculo é simplesmente a média ponderada do retorno esperado de cada ativo:

$$E(R_p) = \alpha_A \bar{R}_A + \alpha_B \bar{R}_B,$$

onde α_A e α_B representam a proporção dos recursos investidos em cada um dos ativos A e B que compõem a carteira, com $\alpha_A + \alpha_B = 1$, e \bar{R}_A e \bar{R}_B são os retornos médios

dos ativos A e B , respectivamente. Já a variância da carteira é dada por:

$$\sigma_p^2 = \alpha_A^2 Var(R_A) + \alpha_B^2 Var(R_B) + 2\alpha_A\alpha_B Cov(R_A, R_B),$$

onde $Var(R_A)$ e $Var(R_B)$ são as variâncias dos retornos dos ativos A e B , respectivamente, e $Cov(R_A, R_B)$ é a covariância dos retornos dos ativos A e B .

Suponha que se queira aplicar a totalidade dos recursos financeiros existentes nos dois ativos, A e B , cujas rentabilidades históricas (média histórica) sejam 1,66% e 11,2%, respectivamente. Suponhamos ainda que a variância de A e B seja, respectivamente, 22,99 e 11,20, e a covariância entre eles seja 0,30.

Dessa forma, o valor esperado dessa carteira é dado por:

$$E(R_p) = 1,66\alpha_A + 11,2\alpha_B.$$

Já a variância dessa carteira é dada por:

$$\sigma_p^2 = 22,99\alpha_A^2 + 11,2\alpha_B^2 + 0,6\alpha_A\alpha_B$$

Aqui no nosso exemplo, o problema geral da Teoria de Markowitz consistiria em encontrar os valores de α_A e α_B de tal forma a minimizar

$$\sigma_p^2 = 22,99\alpha_A^2 + 11,2\alpha_B^2 + 0,6\alpha_A\alpha_B$$

sujeito às restrições: $\alpha_A, \alpha_B \geq 0$ e $\alpha_A + \alpha_B = 1$.

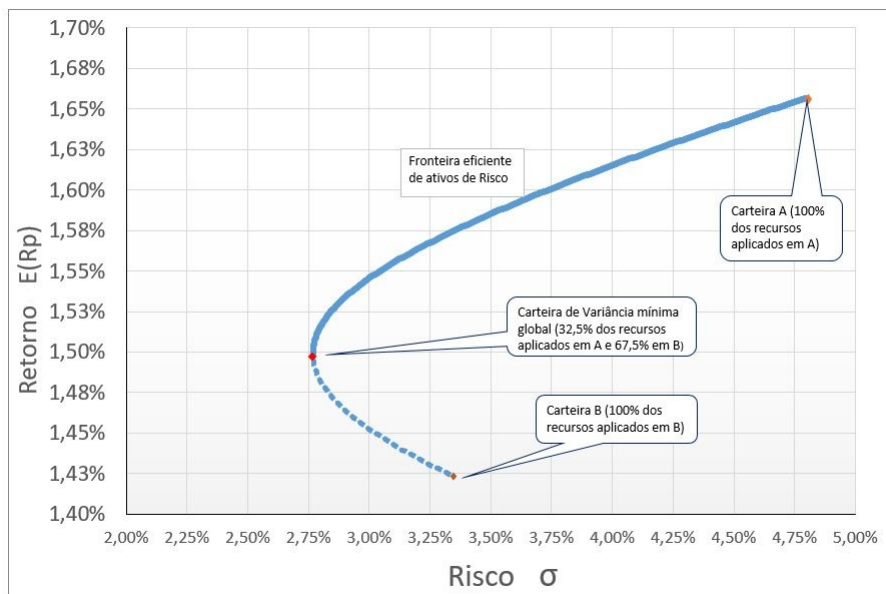
Este cálculo pode ser facilmente obtido com o apoio do software *MS Excel*, por meio do seu suplemento *Solver*, ou mesmo com o uso de programação nas linguagens *R* ou *Python*. Não é escopo deste trabalho detalhar a utilização da programação quadrática e linear na seleção das carteiras de investimento, especificamente na seleção das carteiras eficientes. Para maior detalhamento e aprofundamento neste assunto, sugerimos a leitura do trabalho desenvolvido por Siervo (SIERVO, 2017).

Com base nestas duas últimas expressões citadas, podemos esboçar um gráfico do Retorno \times Risco (desvio-padrão), calculados com base nas combinações possíveis dos valores de α_A e α_B , mostrado na Figura 2.1. Podemos notar no gráfico que, no nosso exemplo com dois ativos, a carteira de variância mínima global é representada por aquela em que 32,5% dos recursos são aplicados no ativo A e 67,5% no ativo B , percentuais esses obtidos a partir do *Solver* ou do *R*, conforme citado anteriormente. Fazendo esta aplicação, obteríamos um valor esperado para a carteira de 1,5%, com uma variância mínima de 7,66% (desvio-padrão de 2,76%).

Outro ponto que podemos observar na Figura 2.1 é a fronteira eficiente dos ativos. Essa fronteira é formada pelas carteiras eficientes, que são aquelas onde é impossível encontrar outra com o mesmo desvio padrão e com maior retorno esperado. No

gráfico apresentado, a fronteira eficiente é identificada pela curva contínua destacada, iniciando na "Carteira de Variância mínima global", que representa a carteira com o menor risco possível. A partir desse ponto, qualquer combinação acima dessa carteira e seguindo a curva contínua representa uma opção eficiente, pois aumenta o retorno esperado à medida que o investidor assume mais risco. Combinações abaixo ou à direita desse ponto (indicadas por uma curva tracejada) são consideradas ineficientes, pois há carteiras que oferecem maior retorno para o mesmo risco ou menor risco para o mesmo retorno esperado.

Figura 2.1: Gráfico da fronteira eficiente dos ativos de risco A e B hipotetizados



Fonte: autor

Tradicionalmente, a construção da fronteira eficiente pode ser realizada em planilhas eletrônicas como o Excel, utilizando ferramentas de otimização como o **Solver**. No entanto, neste trabalho, optamos por utilizar a linguagem R (R CORE TEAM, 2024), utilizando o ambiente integrado de desenvolvimento RStudio (RSTUDIO TEAM, 2024), que oferece diversas vantagens em relação a essa abordagem tradicional. A flexibilidade do R permite a implementação de modelos mais complexos, a utilização de diferentes métodos de otimização e a geração de gráficos personalizados.

Além disso, a linguagem de programação R oferece uma ampla gama de pacotes e funções estatísticas, facilitando a análise de dados. Neste trabalho, a otimização foi realizada utilizando o método de programação não-linear com restrições (GRG) via função `nloptr()` no R. Essa abordagem permitiu encontrar a combinação ótima de ativos que minimiza o risco para diferentes níveis de retorno esperado, resultando na fronteira eficiente. Ao utilizar o pacote `nloptr` (JOHNSON, 2023), conseguimos obter resultados

precisos e personalizados, superando algumas limitações das planilhas eletrônicas. No entanto, a planilha Excel ainda foi utilizada em alguns momentos para a comparação e validação dos cálculos de otimização feitos no R.

Para visualizar a fronteira eficiente, foram gerados gráficos utilizando a função `plot()`, permitindo ilustrar a relação entre risco e retorno dos portfólios eficientes.

2.3 Modelo de Regressão Linear

A regressão linear é um dos modelos estatísticos mais utilizados para analisar a relação entre variáveis, permitindo prever a variável resposta Y com base em uma ou mais variáveis explicativas X . A relação entre Y e X é assumida como linear nos parâmetros, podendo ser expressa como uma regressão linear simples (com uma variável explicativa) ou múltipla (com duas ou mais variáveis explicativas).

A regressão linear simples modela a relação entre uma variável resposta Y e uma variável explicativa X por meio da equação (MORETTIN e SINGER, 2022):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

onde:

- Y_i é a variável resposta para a i -ésima observação;
- X_i é a variável explicativa;
- β_0 é o intercepto e β_1 é o coeficiente angular a serem estimados;
- ε_i representa o erro aleatório, assumido com média zero e variância σ^2 .

Segundo Morettin e Singer (2022), em geral, supõe-se que a média (ou valor esperado) dos erros é nula, o que significa, de modo genérico, que existe uma compensação entre erros positivos e negativos e que, conseqüentemente, o objetivo da análise é modelar o valor esperado da variável resposta

$$E(Y_i) = \beta_0 + \beta_1 X_i.$$

A estimativa dos coeficientes pode ser feita pelo método dos mínimos quadrados ordinários (MQO), minimizando a soma dos quadrados dos resíduos e_i :

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2.$$

Quando há mais de uma variável explicativa, utilizamos a regressão linear múltipla, representada por (Morettin e Singer, 2022):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i,$$

onde:

- Y_i é a variável resposta;
- $X_{i1}, X_{i2}, \dots, X_{ip}$ são as variáveis explicativas;
- $\beta_0, \beta_1, \dots, \beta_p$ são os coeficientes a serem estimados.

Como podemos observar, a regressão linear múltipla permite modelar a influência simultânea de múltiplos fatores sobre uma variável resposta.

Para facilitar o desenvolvimento metodológico, expressamos a regressão linear múltipla na forma matricial (MORETTIN e SINGER, 2022):

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,$$

onde:

- \mathbf{Y} é o vetor de variáveis respostas $(y_1, \dots, y_n)^T$;
- \mathbf{X} é a matriz de variáveis explicativas, incluindo a constante 1 para o intercepto;
- β é o vetor de coeficientes $(\beta_0, \beta_1, \dots, \beta_p)^T$;
- ε é o vetor de erros aleatórios $(\varepsilon_1, \dots, \varepsilon_n)^T$.

Os coeficientes podem ser estimados pelo método dos mínimos quadrados, minimizando a soma dos quadrados dos erros:

$$Q(\beta_0, \dots, \beta_p) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(Y_i - \sum_{j=0}^p \beta_j X_{ij} \right)^2.$$

A solução para encontrar $\hat{\beta}$ é dada por:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

Segundo Morettin e Singer, 2022, uma vez ajustado o modelo, convém avaliar a qualidade do ajuste e um dos indicadores mais utilizados para essa finalidade é o coeficiente de determinação R^2 , que é uma medida que indica a proporção da variabilidade total da variável resposta explicada pelo modelo de regressão, sendo definido por:

$$R^2 = \frac{SQT_{ot} - SQ_{Res}}{SQT_{ot}} = \frac{SQ_{Reg}}{SQT_{ot}} = 1 - \frac{SQ_{Res}}{SQT_{ot}},$$

em que:

- $SQT_{ot} = \sum_{i=1}^n (Y_i - \bar{Y})^2$: soma total dos quadrados,
- $SQ_{Res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$: soma dos quadrados dos resíduos,
- $SQ_{Reg} = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$: soma dos quadrados da regressão.

Quanto mais próximo de 1 for o valor de R^2 , maior será a proporção da variabilidade explicada pelo modelo, indicando um bom ajuste.

Já o AIC (Akaike Information Criterion) é um critério utilizado para comparar modelos estatísticos, considerando simultaneamente a qualidade do ajuste e a complexidade do modelo. Segundo Burnham e Anderson, 2002, o AIC é definido por:

$$AIC = 2k - 2 \ln(\hat{L}),$$

em que k é o número de parâmetros do modelo e \hat{L} é o valor máximo da função de verossimilhança. No caso da regressão linear com erros normalmente distribuídos, o AIC pode ser aproximado por:

$$AIC = n \cdot \ln \left(\frac{SQ_{Res}}{n} \right) + 2k,$$

sendo n o número de observações. Nesse contexto, valores menores de AIC indicam modelos preferíveis, pois sinalizam um bom equilíbrio entre ajuste e parcimônia.

Nos modelos de regressão linear, algumas suposições são fundamentais para garantir a validade das estimativas e dos testes estatísticos de significância do modelo e dos efeitos individuais da(s) variável(is) explicativa(s) (MORETTIN e SINGER, 2022):

- Média zero do erro: $E(\varepsilon) = 0$, garantindo que os coeficientes estimados sejam não viesados.
- Homocedasticidade: a variância dos erros é constante ($Var(\varepsilon) = \sigma^2$), ou seja, a dispersão dos erros não varia ao longo das observações.
- Não correlação dos erros: os erros ε_i são não correlacionados entre si.
- Distribuição normal dos erros: assume-se que os erros seguem uma distribuição normal $\varepsilon_i \sim N(0, \sigma^2)$, permitindo realizar testes estatísticos confiáveis.

Essas propriedades garantem estimativas eficientes e não viesadas dos coeficientes da regressão, além de permitir a construção de intervalos de confiança e testes de hipóteses. Quando essas suposições não são atendidas, métodos alternativos podem ser aplicados para ajustar o modelo e garantir a validade das análises estatísticas.

Se os resíduos apresentarem autocorrelação, ou seja, quando os erros de um período influenciam os seguintes, o que pode ser identificado, por exemplo, por meio do teste

de Durbin-Watson, segundo Morettin e Singer (2022), o modelo nessa situação poderá ser ajustados pelo método de mínimos quadrados generalizados (GLS) ou por métodos de séries temporais. Ainda, segundo eles, uma maneira de contornar esse problema seria modificar componentes aleatórios do modelos para incorporar essa possível autocorrelação nos erros, ou seja, introduzir um termo autoregressivo no modelo, onde os resíduos seguem a equação:

$$e_i = \rho e_{i-1} + u_i, \quad i = 1, \dots, n$$

onde:

- e_i representa os resíduos do modelo no instante i ,
- ρ é o coeficiente de autocorrelação,
- $u_i \sim N(0, \sigma^2)$ são variáveis aleatórias independentes.

Esse modelo assume que o erro e_i pode ser expresso em termos de um erro anterior e_{i-1} , acrescido de um termo aleatório u_i , garantindo que a estrutura de dependência seja incorporada na modelagem, sendo similar à ideia de modelos de séries temporais de médias móveis.

Outra solução para autocorrelação é o uso de defasagens das variáveis explicativas (modelos com variáveis lag). Nesses modelos, incluem-se versões defasadas das variáveis preditoras, permitindo capturar relações temporais entre as variáveis. A inclusão de defasagens da variável resposta como sendo variáveis explicativas é também comum, sendo similar à ideia de modelos de séries temporais autoregressivos. Todas as estratégias citadas podem ser avaliadas durante o processo de modelagem, dependendo da característica de dependência exibida pelos dados e/ou pelos resíduos de um modelo inicial.

Outro problema comum é a heterocedasticidade, que ocorre quando a variância dos erros não é constante ao longo do tempo. Aqui também o modelo de mínimos quadrados generalizados é uma ferramenta útil para lidar com essa situação, pois ajusta os pesos das observações para corrigir a heterocedasticidade. Outra forma de correção desse problema que pode ser investigada é a transformação da variável resposta usando, por exemplo, a abordagem de Box-Cox (CORDEIRO, DEMÉTRIO e MORAL, 2024).

Outro problema relevante é a violação da normalidade dos erros, o que pode afetar a validade dos testes de significância dos coeficientes. Esse problema pode ser identificado com o teste de Shapiro-Wilk e a inspeção visual do QQ-Plot. O QQ-Plot (Quantile-Quantile Plot) é um gráfico que compara os quantis dos resíduos observados com os quantis de uma distribuição normal teórica. Caso os erros não sejam normalmente distribuídos, algumas estratégias podem ser utilizadas, tal como a aplicação de transformações nos dados, como logaritmo.

Uma característica interessante dos modelos de regressão é a disponibilidade de estimativas de incerteza sobre os parâmetros e previsões do modelo, o que pode ser feito via intervalos de confiança, por exemplo. O intervalo de confiança para a média condicional da variável resposta Y dado um conjunto de variáveis explicativas $X_{i1}, X_{i2}, \dots, X_{ip}$ é um intervalo que contém, com um nível de confiança γ , o verdadeiro valor médio da resposta esperada para determinado conjunto de valores explicativos. Esse intervalo é dado por (adaptado de MORETTIN e BUSSAB, 2017):

$$IC(\hat{Y}_i; \gamma) = \hat{Y}_i \pm t_\gamma(n - p - 1)S_e \sqrt{\frac{1}{n} + \sum_{j=1}^p \frac{(X_{ij} - \bar{X}_j)^2}{\sum (X_{ij} - \bar{X}_j)^2}}$$

onde:

- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_p X_{ip}$ é a estimativa pontual da média condicional;
- S_e é o erro padrão da regressão;
- $t_\gamma(n - p - 1)$ é o quantil da distribuição t de Student com $n - p - 1$ graus de liberdade para o nível de confiança γ ;
- \bar{X}_j representa a média dos valores da variável explicativa X_j ;
- n é o número total de observações;
- p é o número de variáveis explicativas no modelo.

Esse intervalo de confiança leva em consideração a incerteza associada à estimação dos coeficientes $\beta_0, \beta_1, \dots, \beta_p$, refletindo a variabilidade das variáveis explicativas no modelo.

Uma distinção importante é a diferença entre intervalos de confiança e intervalos de previsão (CORDEIRO, DEMÉTRIO e MORAL, 2024). Enquanto o primeiro refere-se à média esperada dos valores previstos, o intervalo de previsão se relaciona com a incerteza associada a um único valor futuro. Dessa forma, intervalos de previsão são geralmente mais amplos que os intervalos de confiança, pois incorporam tanto a incerteza na estimativa do modelo quanto a variabilidade individual dos dados (BRUCE e BRUCE, 2019). Neste trabalho, optamos pelos intervalos de confiança em vez de intervalos de previsão, pois estamos mais interessados na média dos retornos do que em prever o retorno de um único investimento no futuro. Essa abordagem é fundamental para construir a fronteira eficiente de Markowitz, que busca encontrar o melhor equilíbrio entre risco e retorno para um portfólio

Os cálculos relativos à análise de regressão linear deste trabalho também foram feitos com base no R . Algumas funções foram utilizadas, como a função `lm()` para

criar modelos de regressão linear e a função `predict()` para fazer previsões, bem como foram aplicados alguns testes citados anteriormente para verificação das suposições do modelo.

2.4 Procedimentos Metodológicos

Os procedimentos metodológicos adotados para alcançar os objetivos propostos nesse trabalho foram estruturados em quatro etapas, conforme descrito a seguir:

Aplicação do Modelo de Markowitz baseada nos retornos históricos (2017-2023)

Nesta etapa, foi aplicado o modelo clássico de Markowitz para construir uma primeira fronteira eficiente utilizando exclusivamente dados históricos dos ativos BBAS3 e IVVB11, abrangendo o período de 2017 a 2023. Para isso, foram calculados os retornos médios históricos dos ativos e estimada a matriz de covariância correspondente, resultando na fronteira eficiente inicial que serve como base comparativa para as etapas subsequentes.

Aplicação do Modelo de Markowitz baseada nos retornos previstos para 2024, via regressão linear

Nesta fase, partiu-se da situação hipotética de um investidor que teria à sua disposição apenas os dados históricos dos ativos até o ano de 2023, além de previsões dos indicadores econômicos para o ano de 2024. Sob esse contexto, foram ajustados modelos de regressão linear para cada ativo (BBAS3 e IVVB11) utilizando como variável resposta os retornos históricos de 2017 a 2023. Indicadores econômicos e termos com possíveis defasagens temporais foram utilizados como variáveis explicativas. Com base nesses modelos ajustados, foram realizadas previsões dos retornos dos ativos para o ano de 2024, empregando os valores reais dos indicadores econômicos para este período. Aqui, devemos ressaltar que num cenário real o investidor não teria à disposição os valores reais dos indicadores econômicos de 2024, mas sim previsões.

A partir das previsões geradas pelo modelo, uma nova fronteira eficiente foi construída, mantendo-se a matriz de covariância histórica (calculada anteriormente), mas substituindo as médias históricas pelos retornos médios previstos obtidos a partir das regressões ajustadas, somente para o ano de 2024. A matriz de covariância foi mantida inalterada, pois reflete a relação histórica entre os ativos e sua movimentação ao longo do tempo, não sendo influenciada diretamente pelas previsões de retorno.

Aplicação do Modelo de Markowitz baseada nos retornos reais de 2024

Nesta etapa, com base nos retornos reais observados apenas em 2024 para os ativos BBAS3 e IVVB11, uma terceira fronteira eficiente foi gerada. Novamente, utilizou-se a matriz de covariância histórica dos retornos (2017-2023), porém agora empregando as médias reais dos retornos observados em 2024, com o objetivo de permitir comparações entre desempenho real e previsto para este ano.

Comparação dos gráficos das fronteiras eficientes e dos retornos dos modelos

Por fim, foi realizada a comparação gráfica das fronteiras eficientes construídas nas etapas anteriores (histórica, prevista e real), permitindo avaliar visualmente as diferenças entre os cenários analisados.

Adicionalmente, foi efetuada uma comparação quantitativa dos retornos previstos com os retornos reais, por meio do cálculo da raiz quadrada do erro quadrático médio (RMSE), possibilitando avaliar numericamente a precisão e proximidade das previsões obtidas via regressão linear em relação ao desempenho efetivamente observado no ano de 2024. O objetivo dessa comparação foi analisar, sob a perspectiva do investidor, se o uso dos retornos previstos por meio da regressão linear resultaria em estimativas mais próximas dos retornos reais observados no ano de 2024 do que o uso dos retornos médios históricos do período 2017-2023.

3 Resultados

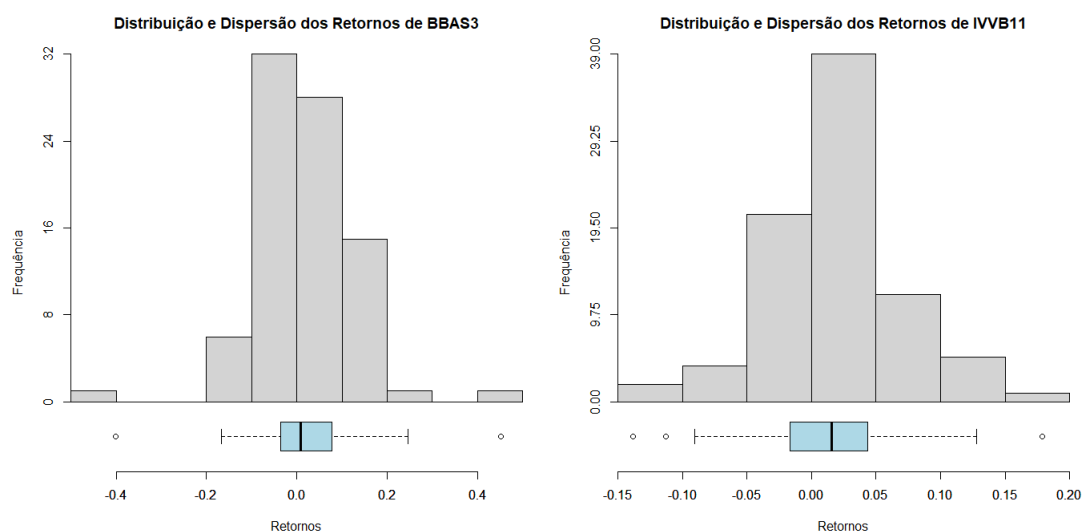
Neste capítulo, inicialmente apresentamos a análise exploratória dos dados utilizados. Em seguida, apresentamos a aplicação dos procedimentos metodológicos citados no capítulo anterior.

3.1 Análise exploratória dos dados

Para essa análise exploratória foram considerados os dados relativos ao período 2017-2023, que serão os dados utilizados para se criar a fronteira eficiente histórica da carteira, baseada nas médias históricas, e para criar os modelos de regressão.

A análise exploratória iniciou-se com uma investigação detalhada dos retornos dos ativos BBAS3 e IVVB11, conforme apresentados na Figura 3.1. Histogramas e boxplots foram construídos para visualizar a distribuição desses retornos. Observamos que, em geral, os retornos se distribuíam de forma aproximadamente simétrica, com alguns outliers.

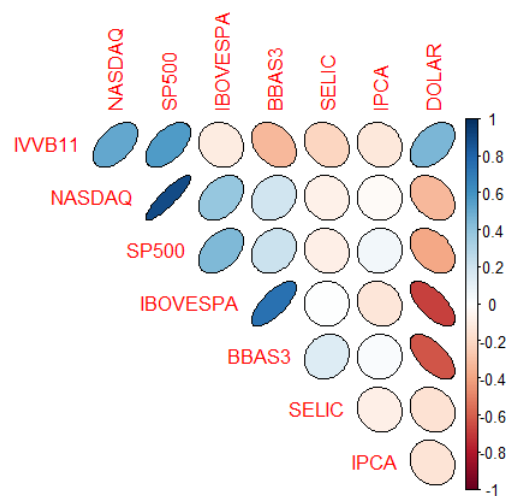
Figura 3.1: Distribuição e Dispersão dos Retornos



Fonte: autor

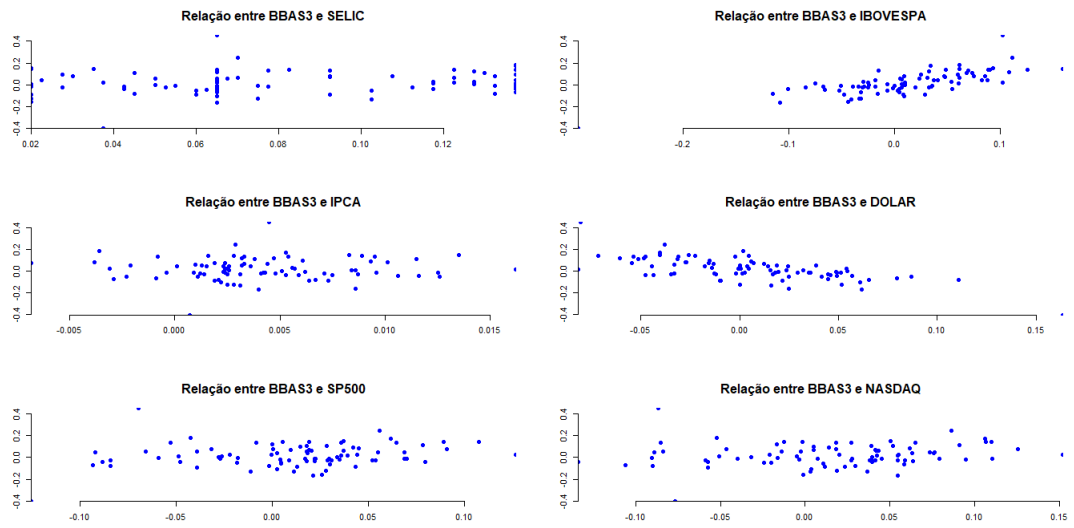
A relação entre os ativos e os indicadores econômicos foi explorada através da matriz de correlação de Spearman (Figura 3.2) e de gráficos de dispersão (Figuras 3.3 e 3.4). A análise visual dos gráficos de dispersão sugeriu uma forte correlação positiva entre o BBAS3 e o IBOVESPA, com um coeficiente de correlação de 0,74, indicando que os dois ativos tendem a se mover na mesma direção. Por outro lado, a correlação entre o BBAS3 e o dólar foi negativa (-0,63), sugerindo uma relação inversa. O ativo IVVB11, por sua vez, apresentou uma correlação moderada e positiva com os índices S&P 500 (0,56) e Nasdaq (0,51), como esperado, dada a sua natureza como um ETF que replica o mercado americano.

Figura 3.2: Matriz de Correlação dos Ativos e Indicadores



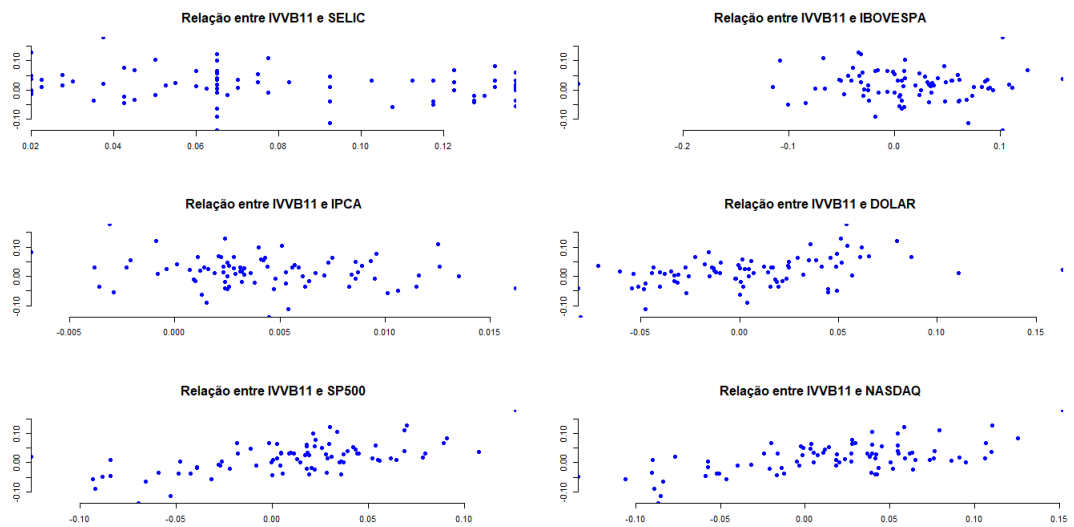
Fonte: autor

Figura 3.3: Relação entre BBAS3 e indicadores



Fonte: autor

Figura 3.4: Relação entre IVVB11 e indicadores



Fonte: autor

As estatísticas descritivas dos retornos revelaram um retorno médio de aproximadamente 1,60% para o IVVB11 e 1,92% para o BBAS3 no período analisado. A amplitude dos retornos, considerando os valores máximos e mínimos, indicou uma maior volatilidade para o ativo BBAS3, o que é consistente com a sua natureza como uma ação individual de uma empresa brasileira, sujeita a riscos específicos. Corroborando com essa análise, calculamos o coeficiente de variação (CV) desses ativos. Segundo Gtiman

(2004), o CV é uma medida relativamente útil na comparação dos riscos de ativos com retornos esperados diferentes. Quanto maior o CV, maior o risco. O valor do CV do ativo BBAS3 foi de 5,5 e do IVVB11 foi de 3,2.

A análise exploratória teve como objetivo inicial entender o comportamento dos retornos dos ativos BBAS3 e IVVB11, bem como a relação desses ativos com os principais indicadores econômicos. Através de gráficos e estatísticas descritivas, foi possível identificar a distribuição dos retornos, a presença de valores atípicos e a força das relações entre as variáveis.

3.2 Aplicação do Modelo de Markowitz baseada nos retornos históricos (2017-2023)

Para a aplicação do Modelo de Markowitz, utilizamos as médias históricas dos retornos dos ativos BBAS3 e IVVB11 no período de 2017 a 2023, juntamente com a matriz de covariância calculada a partir dos retornos passados. Os cálculos a seguir foram feitos com o auxílio por meio do *RStudio*.

Com base nos dados históricos, os valores médios dos retornos dos ativos foram calculados como segue:

- Retorno médio de BBAS3: $\bar{R}_{BBAS3} = 1,92\%$
- Retorno médio de IVVB11: $\bar{R}_{IVVB11} = 1,60\%$

De acordo com a teoria de Markowitz, conforme vimos anteriormente, o retorno esperado para essa carteira é dado por:

$$E(R_p) = \alpha_{BBAS3}\bar{R}_{BBAS3} + \alpha_{IVVB11}\bar{R}_{IVVB11}$$

$$E(R_p) = \alpha_{BBAS3}(1,92\%) + \alpha_{IVVB11}(1,60\%)$$

onde α_{BBAS3} e α_{IVVB11} representam as proporções investidas em cada ativo, enquanto \bar{R}_{BBAS3} e \bar{R}_{IVVB11} são os retornos médios históricos dos ativos BBAS3 e IVVB11, respectivamente.

A matriz de covariância dos ativos foi calculada conforme segue:

$$\begin{bmatrix} 0,011382 & -0,002065 \\ -0,002065 & 0,002671 \end{bmatrix}$$

Para o cálculo do risco, utilizamos a variância e a covariância dos ativos, calculados nessa matriz de covariância:

- Variância de BBAS3: $Var(R_{BBAS3}) = 0,011382$
- Variância de IVVB11: $Var(R_{IVVB11}) = 0,002671$
- Covariância entre BBAS3 e IVVB11: $Cov(R_{BBAS3}, R_{IVVB11}) = -0,002065$

Calculamos ainda a matriz de correlação entre os ativos que apresentou o seguinte resultado:

$$\begin{bmatrix} 1,00 & -0,37 \\ -0,37 & 1,00 \end{bmatrix}$$

Isso indica que os ativos possuem uma correlação negativa, o que pode contribuir para a diversificação da carteira.

A variância do portfólio, que mede o risco da carteira, é dada por:

$$\begin{aligned} \sigma_p^2 &= \alpha_{BBAS3}^2 Var(R_{BBAS3}) + \alpha_{IVVB11}^2 Var(R_{IVVB11}) \\ &\quad + 2\alpha_{BBAS3}\alpha_{IVVB11}Cov(R_{BBAS3}, R_{IVVB11}) \end{aligned}$$

$$\sigma_p^2 = \alpha_{BBAS3}^2(0,011382) + \alpha_{IVVB11}^2(0,002671) + 2\alpha_{BBAS3}\alpha_{IVVB11}(-0,002065)$$

O desvio-padrão (risco) do portfólio é dado pela raiz quadrada da sua variância.

Os valores de α_{BBAS3} e α_{IVVB11} serão determinados por meio da otimização, respeitando as seguintes restrições:

$$\alpha_{BBAS3} + \alpha_{IVVB11} = 1,$$

$$\alpha_{BBAS3} \geq 0, \quad \alpha_{IVVB11} \geq 0.$$

Os pesos para a carteira eficiente, obtidos por meio de otimização, foram:

$$\alpha_{BBAS3} = 26,05\% \quad \text{e} \quad \alpha_{IVVB11} = 73,95\%.$$

Dessa forma, com esses valores dos pesos dos ativos, os resultados da carteira de variância mínima foram:

$$E(R_p) = 1,68\% \quad \text{e} \quad \sigma_p = 3,79\%.$$

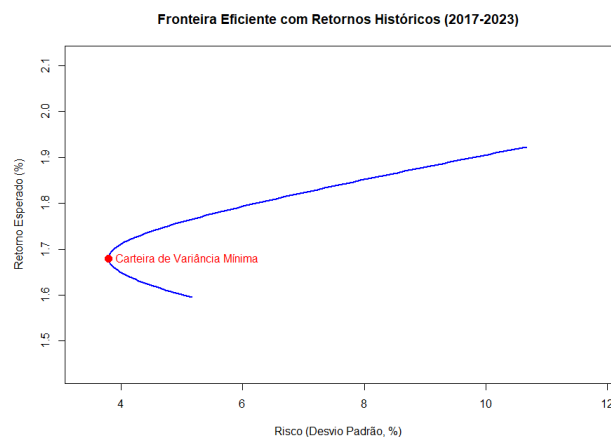
Comparando com os resultados individuais dos ativos:

- BBAS3: Retorno Médio = 1,92%, Desvio-padrão = 10,67%, CV = 5,5
- IVVB11: Retorno Médio = 1,59%, Desvio-padrão = 5,17%, CV = 3,2
- Carteira Otimizada: Retorno Médio = 1,68%, Desvio-padrão = 3,79%, CV = 2,3

O coeficiente de variação (CV) da carteira foi calculado dividindo-se o valor do desvio-padrão da carteira pelo seu respectivo valor esperado. Já os valores do CV dos ativos individuais foram calculados dividindo-se o valor do desvio-padrão de cada ativo pela sua respectiva média histórica. A carteira otimizada apresentou um CV de 2,3, indicando um melhor equilíbrio entre risco e retorno em comparação com os ativos individuais. Isso demonstra a eficiência da alocação de ativos baseada no modelo de Markowitz, reduzindo a volatilidade e melhorando a relação entre risco e retorno.

Na Figura 3.5 observamos o gráfico da fronteira eficiente com base nos dados históricos de 2017-2023. Neste gráfico, a curva azul representa a fronteira eficiente, que contém as combinações de ativos que maximizam o retorno para um dado nível de risco. O ponto vermelho representa a Carteira de Variância Mínima, ou seja, a carteira com o menor risco possível dentro da fronteira eficiente. A fronteira eficiente representa o conjunto de portfólios ótimos, onde cada ponto na curva superior corresponde à melhor combinação de ativos para um determinado nível de retorno esperado.

Figura 3.5: Fronteira eficiente da carteira baseada nos retornos históricos



Fonte: autor

Além da carteira otimizada, analisamos uma carteira com alocação igualitária (50% em BBAS3 e 50% em IVVB11). Os resultados para essa alocação foram:

$$E(R_p) = 1,76\% \quad \text{e} \quad \sigma_p = 4,98\%$$

Ao comparar a carteira otimizada com a de alocação igualitária, observamos que a otimização de Markowitz permitiu reduzir o risco total da carteira em 1,19 pontos percentuais, embora o retorno tenha diminuído ligeiramente. Os resultados obtidos demonstram que a alocação eficiente de Markowitz oferece uma melhor relação risco-retorno do que a distribuição igualitária dos ativos.

3.3 Aplicação do Modelo de Markowitz baseada nos retornos previstos para 2024, via regressão linear

3.3.1 Modelos de Regressão Ajustados

Depois de aplicarmos o Modelo de Markowitz baseado nos retornos históricos, passamos para a modelagem dos modelos de regressão para os dois ativos. Os modelos a seguir foram ajustados considerando os dados históricos dos ativos e dos indicadores econômicos de 2017 à 2023. Esses modelos serão utilizados para estimar os retornos esperados dos ativos para o ano de 2024, a partir dos valores dos indicadores econômicos nesse período.

Modelo para o ativo BBAS3

A análise de regressão para o ativo BBAS3 foi realizada utilizando um modelo de regressão linear múltipla, onde o retorno do ativo foi explicado por três variáveis macroeconômicas: variação do dólar, retorno do índice IBOVESPA e retorno do índice S&P 500. O modelo pode ser representado por:

$$BBAS3_i = \beta_0 + \beta_1 DOLAR_i + \beta_2 IBOVESPA_i + \beta_3 SP500_i + \varepsilon_i$$

A saída do R para a regressão está representada na Figura 3.6. A análise dos coeficientes revela que todas as variáveis explicativas foram estatisticamente significativas ao nível de 5%, conforme indicado pelos valores-p inferiores a 0,05. O coeficiente do Dólar foi negativo, sugerindo uma relação inversa entre essa variável e os retornos de BBAS3. Já o IBOVESPA apresentou um coeficiente positivo, indicando que a valorização do índice está associada a retornos maiores para a ação. O SP500, por sua vez, também teve um coeficiente negativo, o que sugere que aumentos no índice norte-americano podem estar relacionados a retornos menores para BBAS3.

Figura 3.6: Estimativa dos Coeficientes da Regressão Linear para o Retorno de BBAS3

```

Call:
lm(formula = BBAS3 ~ DOLAR + IBOVESPA + SP500, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-0.122025 -0.036857 -0.002732  0.031965  0.216064

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01607    0.00749   2.146 0.034926 *
DOLAR       -0.68506    0.23316  -2.938 0.004313 **
IBOVESPA     1.16504    0.17004   6.852 1.37e-09 ***
SP500       -0.65705    0.16997  -3.866 0.000224 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06247 on 80 degrees of freedom
Multiple R-squared:  0.6695,    Adjusted R-squared:  0.6572
F-statistic: 54.03 on 3 and 80 DF,  p-value: < 2.2e-16

```

Fonte: autor

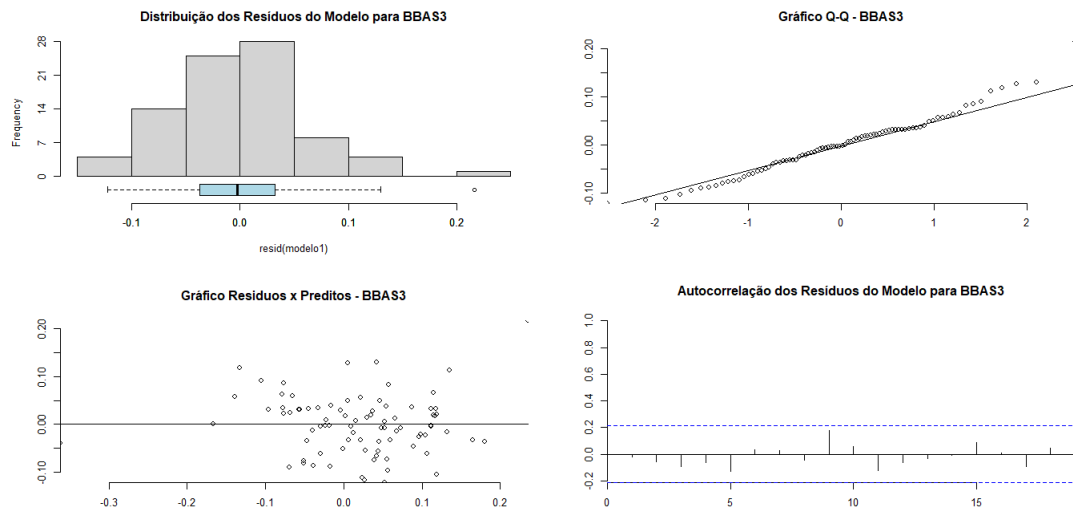
O coeficiente de determinação ajustado (R^2) foi de 0,6572, indicando que aproximadamente 65,7% da variação nos retornos de BBAS3 são explicados pelo modelo. O teste F apresentou um p-valor menor que 0,05, evidenciando a significância global do modelo.

Para avaliar a adequação do modelo, foram realizados testes diagnósticos sobre os resíduos. O teste de Shapiro-Wilk resultou em um valor-p de 0,1191, indicando que não há evidências contra a normalidade dos erros. O teste de Breusch-Pagan apresentou um valor-p de 0,1124, sugerindo que a suposição de homocedasticidade é razoável. O teste de Durbin-Watson forneceu um valor de 2,0401 com um valor-p de 0,5775, não indicando autocorrelação nos resíduos.

Os gráficos da Figura 3.7 corroboraram com a análise dos resíduos:

- Distribuição dos resíduos do modelo para BBAS3: O histograma e o boxplot dos resíduos sugerem uma distribuição aproximadamente normal, sem a presença de valores extremos significativos.
- Gráfico Q-Q dos resíduos: Indica que os resíduos seguem uma distribuição aproximadamente normal, uma vez que os pontos se alinham bem à reta de referência.
- Gráfico de resíduos vs. valores preditos: A análise visual não sugere padrões estruturados nos resíduos, o que corrobora a suposição de homocedasticidade.
- Autocorrelação dos resíduos do modelo para BBAS3: A função de autocorrelação (ACF) mostra que os resíduos não possuem correlação significativa ao longo do tempo, reforçando a validade do modelo ajustado.

Figura 3.7: Análise dos Resíduos do Modelo de Regressão para o Retorno de BBAS3



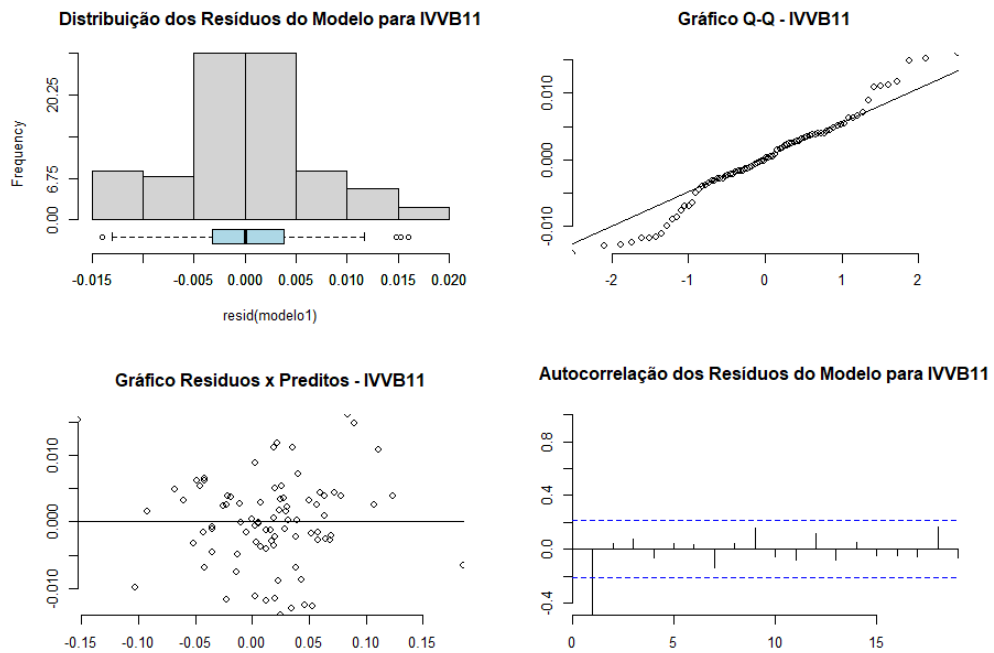
Fonte: autor

Modelo para o ativo IVVB11

Para este ativo, o ajuste do modelo foi mais dificultoso. A equação de regressão para o ativo IVVB11 foi inicialmente ajustada considerando as variáveis explicativas DOLAR e SP500, a qual chamamos no R de *modelo1*. Esse modelo inicial apresentou um coeficiente de determinação ajustado (R^2) de 0,9833, indicando um bom ajuste aos dados. Os coeficientes das variáveis foram estatisticamente significativos ($p < 0,001$), e o teste de normalidade de Shapiro-Wilk ($p = 0,1229$) não rejeitou a hipótese de normalidade dos resíduos.

No entanto, a análise de autocorrelação evidenciou problemas. O teste de Breusch-Godfrey ($p < 0,001$) e o teste de Ljung-Box ($p = 0,00224$) indicaram a presença de autocorrelação nos resíduos, violando a suposição de independência dos erros, evidenciado ainda no gráfico da Figura 3.8. Esse problema podia comprometer a validade das inferências do modelo.

Figura 3.8: Análise dos Resíduos do Modelo de Regressão para o Retorno de IVVB11 (modelo1)



Fonte: autor

Para corrigir essa limitação, foram testados diferentes modelos, conforme ilustrado na Tabela 3.1. Foram testados modelos com variáveis transformadas, termos de defasagem nos resíduos, termos autoregressivos (AR) com relação à variável resposta e também ajuste via método de mínimos quadrados generalizados. Dentre as alternativas avaliadas, o modelo com termo autoregressivo no resíduo se mostrou a melhor opção, tendo o menor AIC e sendo a única em que o gráfico de autocorrelação (ACF) não apresentou maiores problemas. Nesse modelo foi considerado um modelo autoregressivo, no qual os resíduos defasados do *modelo1* foram incorporados como uma nova variável explicativa. O termo autoregressivo foi obtido ao calcular os resíduos do *modelo1* e adicioná-los ao modelo como uma variável adicional ($resid_1$), representando a influência de erros passados sobre os valores atuais. A equação do modelo final ficou:

$$IVVB11_i = \beta_0 + \beta_1 DOLAR_i + \beta_2 SP500_i + \beta_3 resid_{i-1} + \varepsilon_i$$

O ajuste do modelo foi realizado no software *R*, e a saída da regressão está representada na Figura 3.9.

Tabela 3.1: Comparação dos Modelos de Regressão para o IVVB11

Modelo	Descrição dos modelos de regressão linear múltipla	AIC	R ²	Shapiro-Wilk (p-valor)	Breusch-Pagan (p-valor)	Breusch-Godfrey (p-valor)	Ljung-Box (p-valor)
modelo1	Modelo sem termos autocorrelacionados.	-598.13	0.9837	0.1229	0.4698	< 0.001	0.0022
modelo_autoreg	Modelo autoregressivo com defasagem de lag=1 nos resíduos para corrigir autocorrelação residual.	-620.13	0.9877	0.238	0.02474	0.0173	0.06794
modelo_lag	Modelo com inclusão do retorno defasado (lag = 1 mês) da variável dependente IVVB11 como variável preditora adicional.	-589.41	0.9839	0.06633	0.597	< 0.001	0.0016
modelo_gls	Modelo de mínimos quadrados generalizados (GLS) com uma estrutura de correlação autorregressiva de primeira ordem (AR(1)) para os resíduos.	-592.46	0.9837	0.1109	0.04621	N/A	0.0023
modelo_log	Modelo com transformação logarítmica nas variáveis dependente e independentes para estabilizar variância e capturar efeitos proporcionais.	-618.27	0.9869	0.1377	0.5832	< 0.001	< 0.001
modelo_gls_log	Modelo GLS que utiliza uma transformação logarítmica para DOLAR e SP500, considerando variações proporcionais nos dados, além da estrutura AR(1).	-608.33	0.9848	0.1582	0.7266	N/A	< 0.001

Figura 3.9: Estimativa dos Coeficientes da Regressão Linear para o Retorno de IVVB11

```

Call:
lm(formula = IVVB11 ~ DOLAR + SP500 + resid_1, data = dt)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0153285 -0.0028465 -0.0003382  0.0032604  0.0164938

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0002861  0.0006679  -0.428    0.67
DOLAR        1.0085028  0.0160857  62.696 < 2e-16 ***
SP500        1.0272207  0.0146362  70.184 < 2e-16 ***
resid_1     -0.5041048  0.0980015  -5.144 1.87e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.005826 on 80 degrees of freedom
Multiple R-squared:  0.9877,    Adjusted R-squared:  0.9873
F-statistic: 2150 on 3 and 80 DF,  p-value: < 2.2e-16

```

Fonte: autor

A análise dos coeficientes revela que todas as variáveis explicativas foram estatisticamente significativas ao nível de 5%. O coeficiente do Dólar foi positivo, indicando que a valorização dessa variável está associada a um aumento no retorno do ativo IVVB11. Da mesma forma, o SP 500 também apresentou um coeficiente positivo, sugerindo que

variações nesse índice impactam diretamente o retorno do ativo.

O termo autoregressivo (resid_{i-1}) apresentou um coeficiente negativo e estatisticamente significativo, o que sugere que o modelo capturou uma dependência temporal nos resíduos. Esse ajuste foi necessário para corrigir a autocorrelação presente no modelo inicial, como indicado pelos testes estatísticos.

O coeficiente de determinação ajustado (R^2) foi de 0,9873, indicando que aproximadamente 98,73% da variação nos retornos de IVVB11 são explicados pelo modelo. O teste F apresentou um p-valor menor que 0,05, evidenciando a significância global do modelo.

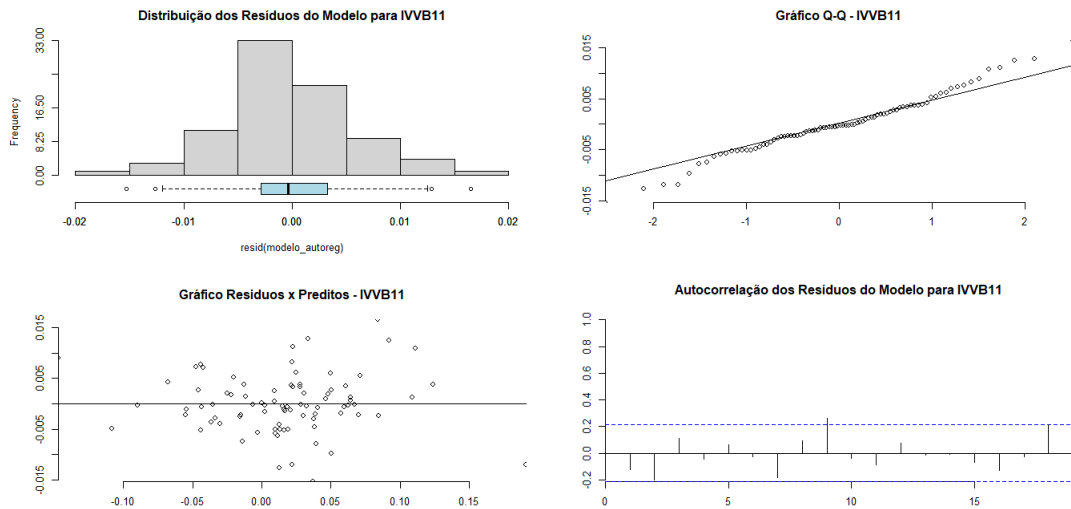
Para avaliar a adequação do modelo, foram realizados testes diagnósticos sobre os resíduos. O teste de Shapiro-Wilk resultou em um valor-p de 0,238, indicando que não há evidências contra a normalidade dos erros. O teste de Breusch-Pagan apresentou um valor-p de 0,02474, sugerindo uma leve indicação de heterocedasticidade. O teste de Durbin-Watson forneceu um valor de 2,2269 com um p-valor de 0,264, indicando que a autocorrelação nos resíduos foi significativamente reduzida.

Os gráficos da Figura 3.10 ilustram a análise dos resíduos:

- Distribuição dos resíduos do modelo para IVVB11: O histograma e o boxplot dos resíduos sugerem uma distribuição aproximadamente normal, sem a presença de valores extremos significativos.
- Gráfico Q-Q dos resíduos: Indica que os resíduos seguem uma distribuição aproximadamente normal, uma vez que os pontos se alinham bem à reta de referência.
- Gráfico de resíduos vs. valores preditos: A análise visual não sugere padrões estruturados nos resíduos, o que corrobora a suposição de homocedasticidade.
- Autocorrelação dos resíduos do modelo para IVVB11: A função de autocorrelação (ACF) mostra que os resíduos não possuem correlação significativa ao longo do tempo, reforçando a validade do modelo ajustado.

Apesar da leve indicação de heterocedasticidade apontada pelo teste de Breusch-Pagan (p-valor de 0,02474), consideramos que essa não foi considerada suficientemente forte para invalidar o modelo, dado que a normalidade dos resíduos foi mantida (p-valor do teste de Shapiro-Wilk = 0,238) e que o critério de informação de Akaike (AIC) do modelo autoregressivo (-620,13) foi significativamente menor do que os demais modelos testados, indicando um melhor ajuste global. Além disso, a análise visual dos resíduos não mostra sinais claros de variação irregular na sua dispersão, reforçando a adequação do modelo para a previsão do retorno do IVVB11.

Figura 3.10: Análise dos Resíduos do Modelo de Regressão para o Retorno de IVVB11 (modelo_autoreg)



Fonte: autor

Embora o modelo com transformação logarítmica tenha apresentado valores de R^2 e AIC semelhantes ao modelo autoregressivo, optou-se por este último porque ele conseguiu capturar melhor a estrutura temporal dos dados, tornando-se mais adequado para fins de previsão. Além disso, no modelo autoregressivo a interpretação dos coeficientes no modelo original é mais direta, sem as complicações associadas às transformações logarítmicas, o que favorece sua aplicação prática na análise e predição de retornos.

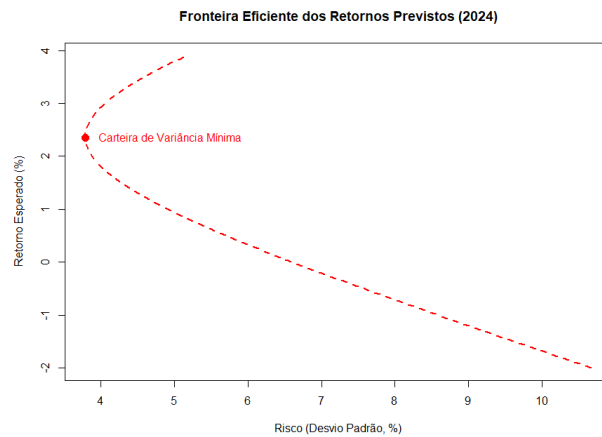
3.3.2 Fronteira eficiente baseada nos retornos previstos para 2024

Após ajustados os modelos de regressão para os ativos BBAS3 e IVVB11, considerando os dados históricos dos ativos e dos indicadores econômicos entre 2017 a 2023, conforme mostrado na Seção 3.3.1, os utilizamos então para estimar os retornos esperados dos ativos para o ano de 2024, a partir dos valores dos indicadores econômicos nesse período.

A partir dessas previsões, uma nova fronteira eficiente foi construída para os retornos no ano de 2024, representada na Figura 3.11. O cálculo seguiu a metodologia do modelo de Markowitz, conforme explicada nos capítulos anteriores. Para garantir a coerência da comparação entre as fronteiras histórica, prevista e real, foi utilizada a mesma matriz de covariância em todas as otimizações. Essa escolha metodológica se justifica porque a intenção do estudo era analisar as diferenças nos retornos esperados

nessas duas situações, tanto com base na média histórica e na média prevista, sem interferências de mudanças na dispersão e correlação entre os ativos.

Figura 3.11: Fronteira eficiente da carteira baseada nos retornos previstos

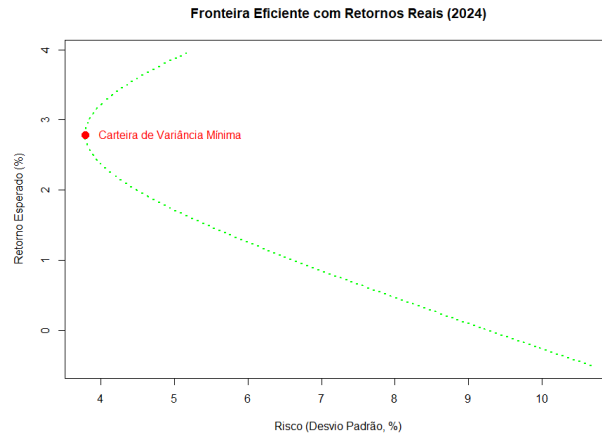


Fonte: autor

3.4 Aplicação do Modelo de Markowitz baseada nos retornos reais de 2024

Como base nos dados reais de 2024, criamos ainda uma terceira fronteira eficiente, representada na Figura 3.12, mas agora baseada na média dos retornos reais dos dois ativos nesse ano. Para este caso, mais uma vez, seguimos a mesma metodologia do modelo de Markowitz explicada nos capítulos anteriores, mas utilizando a média real dos retornos dos ativos em 2024 e a matriz de covariância original baseada nos dados históricos de 2017 a 2023.

Figura 3.12: Fronteira eficiente da carteira baseada nos retornos reais



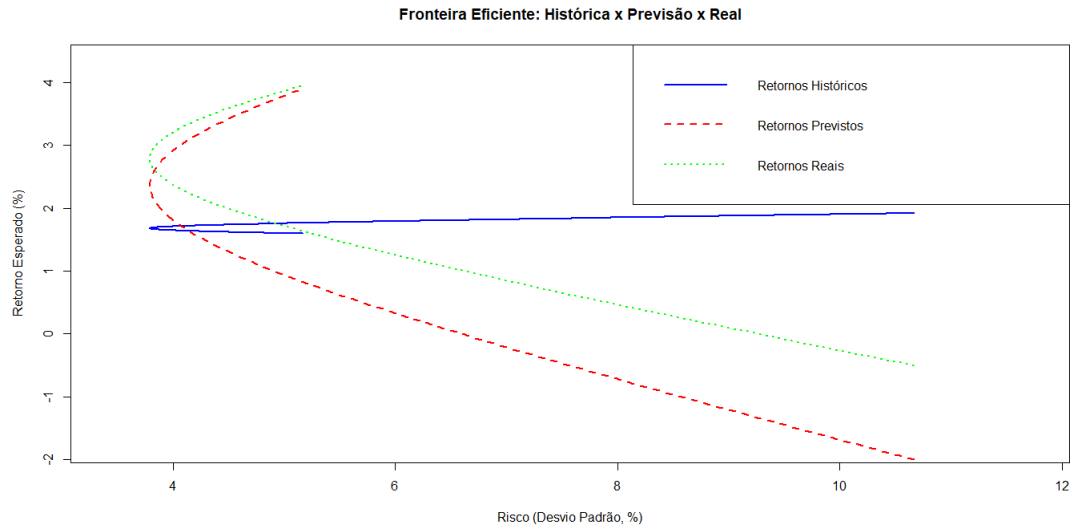
Fonte: autor

3.5 Comparação dos gráficos das fronteiras eficientes baseadas nos dados históricos, previstos e reais

O gráfico com as três fronteiras pode ser observado na Figura 3.13. Nesse gráfico podemos observar três cenários distintos:

- Fronteira Histórica (linha azul sólida): Representa a otimização da carteira com base na média dos retornos históricos dos ativos (2017-2023).
- Fronteira Prevista (linha vermelha tracejada): Baseada na média dos retornos estimados pelo modelo de regressão para 2024.
- Fronteira Real (linha verde pontilhada): Obtida considerando a média dos retornos efetivamente observados no ano de 2024.

Figura 3.13: Fronteira eficiente: Histórica x Prevista x Real



Fonte: autor

Conforme podemos observar no gráfico, o risco mínimo observado nas três fronteiras é o mesmo. Como a matriz de covariância define a relação estatística entre os ativos e foi mantida constante, a carteira de variância mínima (o ponto mais à esquerda das curvas) permanece a mesma entre as três versões da fronteira eficiente. Isso assegura que a análise foque unicamente no impacto das diferenças nos retornos.

Adicionalmente, podemos avaliar os valores dos retornos nos três cenários:

- Retorno da Carteira de Variância Mínima (baseado na média histórica): 1,68%
- Retorno da Carteira de Variância Mínima (baseado na média prevista): 2,37%
- Retorno da Carteira de Variância Mínima (baseado na média real): 2,79%

Esses resultados indicam que a carteira baseada na média prevista teria apresentado um retorno esperado superior ao da carteira baseada nos dados históricos. Além disso o retorno baseado na média prevista foi mais próximo do retorno real que o baseado na média histórica.

3.6 Comparação dos retornos dos modelos

Olhando o gráfico da Figura 3.13 anterior, observamos que visualmente os retornos previstos das carteiras, calculados via regressão, aparentam estar mais próximos dos retornos reais, em comparação com os retornos calculados com base nas médias históricas. Para tentarmos ratificar esse análise, realizamos uma comparação do desempenho

das carteiras calculadas com base na média dos retornos previstos e na média dos retornos históricos. Essa validação foi feita por meio do cálculo da raiz quadrada do erro quadrático médio (RMSE), permitindo avaliar a proximidade entre os retornos estimados e os retornos reais observados. A análise realizada teve como objetivo comparar os retornos de diferentes alocações entre os ativos BBAS3 e IVVB11 sob três perspectivas distintas: retornos históricos, previstos e reais.

Para avaliar a relação entre o retorno da carteira e a alocação percentual em cada ativo, no R foi criado um dataframe de simulação contendo diferentes composições da carteira. Uma visão parcial dos dados deste dataframe pode ser visto na Tabela 3.2. Foram gerados percentuais variando de 0% a 100% para o ativo BBAS3, com o restante sendo alocado automaticamente no ativo IVVB11, garantindo que a soma das alocações fosse sempre 100%.

Tabela 3.2: Visão parcial dos dados do dataframe com a simulação de diferentes composições da carteira

Percentual_BBAS3	Percentual_IVVB11	R_real	R_historico	R_previsto	R_ic_max	R_ic_min
0	100	0.0396	0.0159	0.0390	0.0408	0.0372
1	99	0.0391	0.0160	0.0384	0.0405	0.0364
2	98	0.0387	0.0160	0.0378	0.0401	0.0356
3	97	0.0382	0.0160	0.0372	0.0397	0.0348
4	96	0.0378	0.0161	0.0367	0.0393	0.0340
5	95	0.0373	0.0161	0.0361	0.0389	0.0332
6	94	0.0369	0.0161	0.0355	0.0386	0.0324
7	93	0.0364	0.0162	0.0349	0.0382	0.0316
8	92	0.0360	0.0162	0.0343	0.0378	0.0308
9	91	0.0355	0.0162	0.0337	0.0374	0.0300
10	90	0.0351	0.0163	0.0331	0.0371	0.0292
11	89	0.0347	0.0163	0.0325	0.0367	0.0284
12	88	0.0342	0.0163	0.0319	0.0363	0.0276
13	87	0.0338	0.0164	0.0313	0.0359	0.0268
14	86	0.0333	0.0164	0.0308	0.0355	0.0260
			∴ (continuação) ∴			
92	8	-0.0015	0.0190	-0.0153	0.0060	-0.0366
93	7	-0.0019	0.0190	-0.0159	0.0057	-0.0374
94	6	-0.0024	0.0190	-0.0164	0.0053	-0.0382
95	5	-0.0028	0.0191	-0.0170	0.0049	-0.0390
96	4	-0.0032	0.0191	-0.0176	0.0045	-0.0398
97	3	-0.0037	0.0191	-0.0182	0.0042	-0.0406
98	2	-0.0041	0.0192	-0.0188	0.0038	-0.0414
99	1	-0.0046	0.0192	-0.0194	0.0034	-0.0422
100	0	-0.0050	0.0192	-0.0200	0.0030	-0.0430

Com base nesses percentuais, os retornos das carteiras foram calculados considerando os diferentes cenários de retorno:

$$R_p = \alpha_{\text{BBAS3}} \bar{R}_{\text{BBAS3}} + \alpha_{\text{IVVB11}} \bar{R}_{\text{IVVB11}}$$

onde:

- α_{BBAS3} e α_{IVVB11} representam as proporções investidas em cada ativo;

- \bar{R}_{BBAS3} e \bar{R}_{IVVB11} correspondem aos retornos médios históricos (2017-2023), previstos (2024) ou reais (2024) dos ativos BBAS3 e IVVB11, respectivamente.

No caso dos dados baseados nas previsões do modelo de regressão, além dos retornos médios, foram calculadas as médias dos limites superior e inferior do intervalo de 95% de confiança para os retornos previstos de 2024, proporcionando uma faixa de variação esperada para as previsões.

Dessa forma, a Tabela 3.2 apresenta os retornos das carteiras resultantes das diferentes composições entre os ativos BBAS3 e IVVB11, conforme os percentuais alocados nas duas primeiras colunas. Os valores das colunas de retorno foram calculados a partir da equação acima, considerando diferentes formas de estimar o retorno esperado dos ativos. Especificamente:

- **R_real**: Retorno da carteira com base na média dos retornos reais observados dos ativos BBAS3 e IVVB11 no ano de 2024, refletindo o desempenho efetivo do mercado.
- **R_historico**: Retorno da carteira com base na média dos retornos históricos dos ativos calculados no período de 2017 a 2023, representando um cenário baseado apenas em dados passados.
- **R_previsto**: Retorno da carteira com base na média dos retornos previstos para 2024, estimados via modelos de regressão ajustados.
- **R_ic_max**: Retorno da carteira com base na média dos limites superiores do intervalo de confiança para a média dos retornos previstos de 2024, fornecendo uma estimativa otimista do desempenho esperado.
- **R_ic_min**: Retorno da carteira com base na média dos limites inferiores do intervalo de confiança para a média dos retornos previstos de 2024, representando um cenário mais conservador.

Essa abordagem permite comparar como diferentes métodos de estimativa de retorno afetam a construção da carteira, avaliando se o uso de previsões baseadas em regressão melhora a precisão da tomada de decisão em relação ao uso de médias históricas.

Após a simulação dos retornos para cada percentual de alocação, foi calculado o RMSE para comparar a precisão dos retornos históricos e previstos em relação aos retornos reais. Na prática, em relação à Tabela 3.2, foi calculado o RMSE dos valores das colunas R_historico e R_previsto em relação aos valores da coluna R_real. O RMSE mede a diferença média entre os valores estimados e os valores observados, sendo calculado como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{\text{Real},i} - R_{\text{Estimado},i})^2}$$

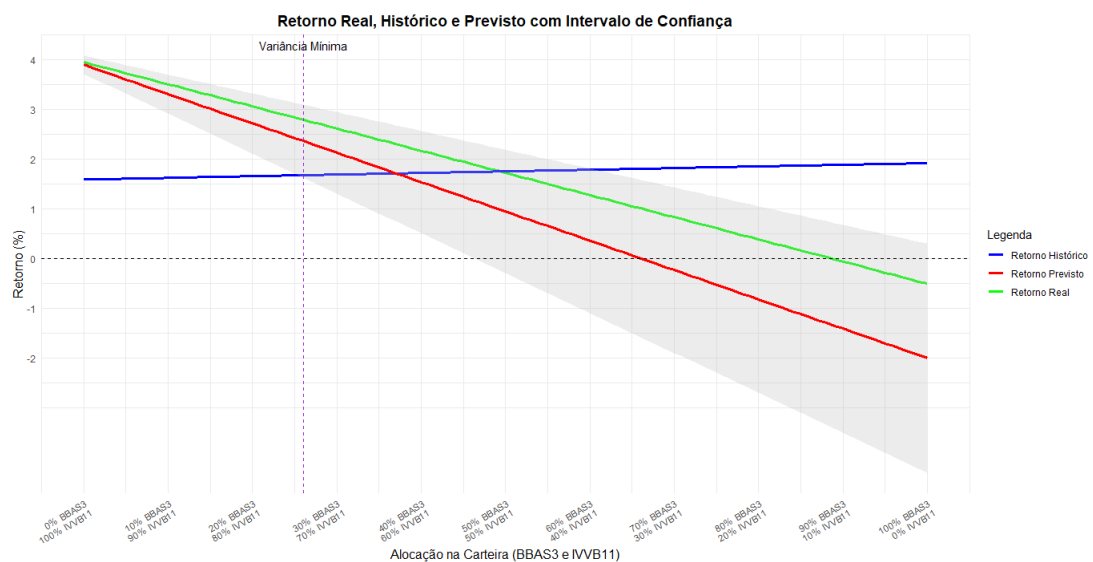
Os valores obtidos foram:

- RMSE para Retornos Históricos: 0,013959
- RMSE para Retornos Previstos: 0,008815

Esses resultados indicam que os retornos previstos apresentaram um RMSE menor em relação aos retornos reais quando comparados aos retornos históricos, sugerindo que a modelagem estatística utilizada conseguiu capturar melhor o comportamento do mercado.

A Figura 3.14 ilustra a relação entre os retornos e a alocação na carteira para os diferentes cenários analisados. Ao analisar o gráfico, podemos observar que a curva baseada nas previsões (linha vermelha) apresenta um formato diferente da curva histórica (linha azul). Isso indica que os modelos estatísticos utilizados para fazer as previsões capturaram mudanças importantes no mercado que não estavam presentes nos dados históricos. A curva real (linha verde), por sua vez, se encontra próxima à curva prevista e dentro do intervalo de previsão, confirmando a precisão das nossas previsões.

Figura 3.14: Retorno Real, Histórico e Previsto com Intervalo de Confiança



Fonte: autor

Uma outra observação que podemos fazer é sobre o intervalo de confiança mostrado na região sombreada do gráfico que se expande conforme a alocação no ativo IVVB11 diminui. Esse comportamento pode ser explicado pelo erro padrão dos modelos de

regressão utilizados para estimar os retornos. O modelo de regressão do BBAS3 apresentou um erro padrão significativamente maior (0,062) em comparação ao do IVVB11 (0,005), tornando suas previsões menos precisas. Como a incerteza da previsão está diretamente associada ao erro padrão, conforme a carteira se torna mais dependente do BBAS3, a variabilidade do retorno previsto aumenta, resultando em um intervalo de confiança mais amplo.

Além disso, observamos ainda que o retorno histórico se mostrou bem diferente das curvas de retorno previsto e retorno real. Isso ocorre porque os retornos médios históricos foram calculados sobre um período mais longo, capturando padrões de retorno que podem não refletir as condições recentes do mercado. Os valores médios indicam que o retorno histórico do BBAS3 foi de aproximadamente 1,92%, enquanto tanto os retornos reais (-0,50%) quanto os previstos (-1,99%) para esse ativo foram negativos, revelando uma queda recente não capturada pelo histórico mais amplo. Já o IVVB11 apresentou um comportamento mais estável, com retornos históricos (1,59%), reais (3,95%) e previstos (3,90%) relativamente próximos. Essa diferença nos retornos históricos pode indicar mudanças nos ativos ou eventos recentes que afetaram os retornos observados e previstos.

4 Considerações Finais

Este estudo teve como objetivo analisar como a regressão linear pode complementar o modelo de Markowitz na seleção e otimização de carteiras de investimentos. Para isso, foram investigadas as relações entre os retornos dos ativos BBAS3 e IVVB11 e variáveis macroeconômicas, ajustando modelos de regressão para prever seus retornos esperados. Com essas previsões, foi possível integrar os resultados ao modelo de Markowitz e construir uma nova fronteira eficiente, que foi comparada com a fronteira baseada nos retornos históricos e com os retornos reais observados.

Conforme mencionado anteriormente, a decisão de investir envolve um equilíbrio entre risco e retorno. Nesse cenário, a Estatística se apresenta como uma ferramenta indispensável, conforme já mostrado em trabalhos anteriores, como no desenvolvido por ALVES (2020). Naquele trabalho, o objetivo foi destacar a importância da Estatística na decisão de investimentos, utilizando o modelo de Markowitz como ferramenta de apoio ao investidor iniciante na seleção de carteiras, por meio de planilhas eletrônicas. No presente trabalho, a mesma base teórica foi retomada sob uma perspectiva mais analítica, incorporando técnicas estatísticas adicionais, como modelos de regressão linear, e utilizando a linguagem R como ferramenta para ajustes, previsões e construção da fronteira eficiente. Essa abordagem permitiu integrar variáveis macroeconômicas à estimativa dos retornos esperados, bem como avaliar a acurácia das previsões por meio de métricas como o RMSE. Assim, os dois estudos se complementam ao evidenciar a aplicabilidade da Estatística no apoio à tomada de decisão no contexto financeiro.

Dessa forma, os objetivos do presente estudo foram alcançados, demonstrando que a regressão linear pode ser uma ferramenta valiosa para complementar o modelo de Markowitz. No entanto, vale ressaltar que as previsões foram realizadas com base em dados já conhecidos para os indicadores macroeconômicos. Na prática, um investidor precisaria contar com projeções econômicas precisas e confiáveis para que os retornos estimados se confirmem no futuro. Além disso, em modelos que incluem termos autoregressivos, eles deveriam ser calculados ao longo do processo de predição.

Os resultados mostraram que a incorporação da regressão linear permitiu interpretar melhor as relações entre variáveis econômicas e retornos financeiros, além de aprimorar a otimização do portfólio. A comparação entre as fronteiras revelou que

os retornos previstos diferiram dos históricos, reforçando a importância de modelos estatísticos, como os de regressão, para auxiliar a tomada de decisão. Vale destacar também que a abordagem de previsão dos retornos com base em modelos apresenta a vantagem de permitir a identificação e estimação da incerteza em torno do retorno médio, levando à tomada de decisão mais robusta. O modelo clássico de Markowitz, ao utilizar dados históricos (2017-2023) ou um ano específico (2024), estima a variabilidade dos retornos por meio da matriz de covariância, permitindo avaliar o risco das carteiras. No entanto, essa abordagem não incorpora diretamente simulações condicionais ao ambiente macroeconômico, o que pode ser complementado por modelos de regressão que permitem estimar retornos esperados com base em variáveis econômicas e projetar diferentes cenários para auxiliar na tomada de decisão. A eficácia dessa abordagem foi evidenciada pelos resultados do RMSE, que mostraram que os retornos previstos ficaram mais próximos dos valores reais. Isso reforça que o uso de modelos de regressão pode fornecer estimativas mais precisas do que simplesmente considerar médias históricas.

Cabe destacar ainda que, ao comparar graficamente as fronteiras eficientes construídas, a fronteira eficiente obtida a partir dos dados históricos revelou-se mais achatada e distinta das demais, particularmente quando comparada àquelas obtidas com retornos previstos e reais. Essa diferença talvez se justifique pelo período adotado de 7 anos, o que pode ter influenciado a variabilidade e tendências captadas pelos modelos. Contudo, a escolha de utilizar dados mensais de 2017 a 2023 teve como objetivo capturar diferentes ciclos econômicos, refletindo variações em fatores macroeconômicos, como taxas de juros, inflação e crescimento econômico, proporcionando uma visão mais realista do comportamento dos ativos e sua relação com o ambiente econômico, sendo essencial tanto para a construção da fronteira eficiente quanto para o ajuste dos modelos de regressão.

Elton, Gruber e Brown (2012) destacam que períodos mais longos reduzem o impacto de flutuações de curto prazo e proporcionam estimativas mais confiáveis. No contexto da regressão, um período de sete anos permite maior estabilidade nos coeficientes estimados, minimizando distorções que poderiam surgir em períodos muito curtos. Além disso, com a inclusão de eventos extremos, como a pandemia de COVID-19, buscou-se evitar a subestimação de riscos reais, conforme argumentado por Taleb (2007). Embora a pandemia tenha gerado um choque significativo no mercado econômico brasileiro e mundial (BANCO MUNDIAL, 2022), a consideração de um período mais amplo ajuda a diluir seu impacto excessivo nos resultados do modelo, procurando garantir previsões mais robustas e alinhadas ao comportamento histórico do mercado.

Para trabalhos futuros, seria interessante expandir a análise para incluir mais ativos, utilizar outros períodos de dados históricos e testar abordagens mais avançadas como, por exemplo, modelos de séries temporais ou outros métodos para análise de dados

em painel. Além disso, avaliar a carteira sob diferentes cenários econômicos gerados por simulação, por exemplo, poderia proporcionar ideias adicionais sobre a robustez do modelo. Dessa forma, este estudo reforça a importância da modelagem estatística na gestão de investimentos e na busca por estratégias mais eficientes por parte do investidor.

Referências

ALVES, D. P. **A Estatística como auxílio na seleção de carteiras de investimento: uma aplicação por meio do modelo de Markowitz**. Dissertação (Mestrado) — Universidade de Brasília – ProfMat, 2020.

B3 – BOLSA, BRASIL, BALCÃO. **IBOVESPA – Índice Bovespa**. Disponível em: <https://www.b3.com.br>. Acesso em: 1 fev. 2025.

BANCO CENTRAL DO BRASIL. **Cotações de moedas estrangeiras – Dólar comercial e turismo**. Disponível em: <https://www.bcb.gov.br>. Acesso em: 1 fev. 2025.

BANCO CENTRAL DO BRASIL. **Taxa Selic – Sistema Especial de Liquidação e Custódia**. Disponível em: <https://www.bcb.gov.br>. Acesso em: 1 fev. 2025.

BANCO MUNDIAL. **Relatório de Desenvolvimento Mundial 2022: Os Impactos Econômicos da Crise da COVID-19**. Washington, D.C.: Banco Mundial, 2022. Disponível em: <https://www.worldbank.org/pt/publication/wdr2022/brief/chapter-1-introduction-the-economic-impacts-of-the-covid-19-crisis>. Acesso em: 18 mar. 2025.

BODIE, Zvi; KANE, Alex; MARCUS, Alan. **Fundamentos de Investimentos**. São Paulo: AMGH Editora, 2014.

BRODIE, J.; DAHL, J.; BOYD, S.; PELGER, M. **Sparse and stable Markowitz portfolios**. *arXiv preprint*, 2007. Disponível em: <https://arxiv.org/abs/0708.0046>. Acesso em: 31 jan. 2025.

BRUCE, Andrew; BRUCE, Peter. **Estatística prática para cientistas de dados: 50 conceitos essenciais**. Rio de Janeiro: Alta Books, 2019.

BURNHAM, Kenneth P.; ANDERSON, David R. **Model selection and multimodel inference: a practical information-theoretic approach**. 2. ed. New York: Springer, 2002.

CARVALHO, P.C.P; MORGADO, A.C. **Matemática Discreta**. SBM (Coleção PROF-MAT, 2ª edição, 2015).

CORDEIRO, Gauss Moutinho; DEMÉTRIO, Clarice Garcia Borges; MORAL, Rafael de Andrade. **Modelos lineares generalizados e aplicações**. São Paulo: Blucher, 2024. Acesso em: 03 mar. 2025.

ELTON, Edwin; GRUBER, Martin; BROWN, Stephen. **Moderna teoria de carteiras e análise de investimentos**. Elsevier Brasil, 2012.

FERREIRA, J. **Teste empírico do CAPM no mercado financeiro português**. 2015. Disponível em: https://bibliotecadigital.ipb.pt/bitstream/10198/11953/1/Ferreira_Jos%C3%A9.pdf. Acesso em: 31 jan. 2025.

GITMAN, Lawrence J. **Princípios de administração Financeira**. 10ª Edição. São Paulo: Person Addison Wesley, 2004.

GITMAN, L. J.; ZUTTER, C. J. **Principles of managerial finance**. 14. ed. Boston: Pearson, 2015.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **IPCA – Índice Nacional de Preços ao Consumidor Amplo**. Disponível em: <https://www.ibge.gov.br>. Acesso em: 1 fev. 2025.

JOHNSON, S. G. **The NLOpt nonlinear-optimization package**. 2023. Disponível em: <https://nlopt.readthedocs.io/en/latest/>. Acesso em: 13 mar. 2025.

MARKOWITZ, H. M. **Portfolio selection**. *The Journal of Finance*, Hoboken, v. 7, n. 1, p. 77–91, 1952.

MORETTIN, Pedro Alberto; BUSSAB, Wilton Oliveira. **Estatística Básica**. 9ª edição. São Paulo: Editora Saraiva, 2017.

MORETTIN, Pedro Alberto; SINGER, Júlio da Motta. **Estatística e ciência de dados**. Rio de Janeiro: LTC, 2022.

NASDAQ. **NASDAQ Composite Index**. Disponível em: <https://www.nasdaq.com>. Acesso em: 1 fev. 2025.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. Disponível em: <https://www.R-project.org/>. Acesso em: 13 mar. 2025.

RSTUDIO TEAM. **RStudio: Integrated Development Environment for R**. Boston, MA: Posit Software, PBC, 2024. Disponível em: <http://www.posit.co/>. Acesso

em: 13 mar. 2025.

SÁ, Geraldo Tosta de. **Administração de investimentos: teoria de carteiras e gerenciamento do risco**. Qualitymark, 1999.

SIERVO, Juliano Squarone Di. **Aplicação de programação linear na seleção de carteiras de investimento**. Dissertação de mestrado, Universidade Federal de São Carlos, Sorocaba, 2017.

S&P GLOBAL. **S&P 500 Index**. Disponível em: <https://www.spglobal.com>. Acesso em: 1 fev. 2025.

TALEB, Nassim Nicholas. **The Black Swan: The Impact of the Highly Improbable**. New York: Random House, 2007.

VIRGILLITO, Salvatore Benito. **Estatística aplicada**. São Paulo: Saraiva, 1. ed., 2017.