

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
DEPARTAMENTO DE ECONOMIA



MONOGRAFIA DE FINAL DE CURSO

**O ESTUDO DA DIFERENÇA DE PERFORMANCE
ENTRE INVESTIMENTO ATIVO E PASSIVO NO
BRASIL**

João Pedro Jardim Lima

Matrícula: 1910802

Orientador: André Duarte

Coordenador de Monografia: Márcio Garcia

Rio de Janeiro, Novembro de 2023

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO



DEPARTAMENTO DE ECONOMIA
MONOGRAFIA DE FINAL DE CURSO

**O ESTUDO DA DIFERENÇA DE PERFORMANCE
ENTRE INVESTIMENTO ATIVO E PASSIVO NO
BRASIL**

João Pedro Jardim Lima

Matrícula: 1910802

Orientador: André Duarte

Coordenador de Monografia: Márcio Garcia

“Declaro que o presente trabalho é de minha autoria e que não recorri, para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor.”

Rio de Janeiro, Novembro de 2023

“As opiniões expressas neste trabalho são de responsabilidade única e exclusiva do autor”

Agradecimentos

Gostaria de agradecer, primeiramente, a minha família pelo grande apoio e carinho durante toda minha graduação e enquanto estive redigindo esta monografia.

Queria agradecer, também, ao meu orientador, André Duarte, pelas discussões e direcionamento na confecção deste trabalho.

E, aos meus amigos, que fizeram essa jornada comigo e a deixaram um pouco mais leve, obrigado e vamos em frente!

Resumo

Este *paper* busca analisar a diferença de performance entre gestão ativa e passiva no Brasil. Para isso, foram realizados cálculos sobre a performance de 32.603 fundos de investimento, separados em 5.913 fundos de ações e 26.690 fundos multimercados; o desempenho destes fundos foi comparado a performance do índice Bovespa e o fundo passivo BOVA11. Encontramos que, em média, fundos de investimento não são capazes de gerar retornos superiores ao fundo passivo em janelas longas, cerca de 67,9% dos fundos tiveram desempenhos absolutos inferiores ao BOVA11 na janela de 10 anos e 85% em 20 anos. Notou-se, também, que o resultado se manteve após ajustarmos os retornos dos fundos ao risco incorrido por estar investido nestes ativos.

Abstract

This *paper* aims to analyze the performance difference between active and passive management in Brazil. To achieve this, calculations were conducted on the performance of 32,603 investment funds, divided into 5,913 equity funds and 26,609 multimarket funds. The performance of these funds was compared to the Bovespa index and the passive fund BOVA11. It was found that, on average, investment funds are not able to generate returns superior to the passive fund over long periods. Approximately 67.9% of the funds had absolute underperformance compared to BOVA11 in the 10-year window, and 85% in the 20-year window. It was also noted that this result persisted even after adjusting the funds' returns for the risk incurred by investing in these assets.

Sumário

Sumário	5
Lista de Tabelas	6
Lista de Figuras	7
Introdução	8
Motivação	10
Revisão de Literatura	11
1. Bases Teóricas do Investimento Ativo e Passivo	11
2. Análises de Performance no Exterior	13
Metodologia	16
1. Performance dos Fundos de Investimento	20
2. Ativo Livre de Risco	21
3. Retorno de Mercado	22
4. Beta	23
5. Calculando o Alfa do CAPM	23
Resultados Pretendidos	27
Dados	27
1. Análise de Performance Média de Fundos de Investimento	28
2. Tamanho e Performance de Fundos de Investimento	33
3. Survivorship Bias	36
4. Performance de Fundos de Investimento Ativos x Fundos de Investimento Passivos	39
Conclusão	48
Referências	51
Apêndice: Código para Confeccção, Tratamento e Análise da Base de Dados	54

Lista de Tabelas

1	Estatísticas Descritivas de Fundos de Investimento.	28
2	Estatísticas de Retorno Fundos de Investimento de 2003 a 2023.	29
3	Estatísticas de retorno fundos de investimento em ações e fundos mul- timercados de 2003 a 2023.	31
4	Regressão de retorno de fundos de investimento em excesso da taxa livre de risco.	34
5	Estatísticas de retorno fundos de investimento sobreviventes e fundos descontinuados de 2003 a 2023.	37
6	Proporção de Fundos com Desempenho Superior ao Fundo de Índice Passivo.	40
7	Fundos com Desempenho Superior ao Fundo de Índice Passivo em Janelas de 1, 5, 10 e 20 Anos	42
8	Proporção de Fundos com Desempenho Ajustado ao Risco Superior ao Fundo de Índice Passivo.	44
9	Fundos com Desempenho Ajustado ao Risco Superior ao Fundo de Índice Passivo em Janelas de 1, 5, 10 e 20 Anos	45

Lista de Figuras

1	Demonstração de alfa, ou excesso de retorno. Womack and Zhang (2003)	20
2	Funcionamento LFT (Tesouro Direto, n.d.)	22
3	<i>Scatter Plot</i> Performance Mensal Fundos de Investimento e Patrimônio Líquido Defasado.	35
4	Proporção de Fundos com Desempenhos Superiores ao Índice BOVA11 por Ano.	43
5	Proporção de Fundos com Desempenho Ajustado ao Risco Superior ao Índice BOVA11 por Ano.	46

Introdução

Nas últimas décadas, com o surgimento e ganho de popularidade de fundos de índice (ETFs), vem se desenhando um debate na comunidade de investimentos sobre qual estratégia de investimentos é vencedora. De um lado, investidores ativos acreditam que escolhendo ativos específicos (como por exemplo a ação de uma empresa) que entendem ter bom desempenho, serão capazes de superar o mercado ou índice; de outro, investidores passivos acreditam que, ao comprar um índice de ativos (como por exemplo um fundo do índice Bovespa), serão capazes de gerar retornos superiores à gestão ativa no longo prazo.

A indústria de fundos de investimentos brasileira possui aproximadamente de R\$6,8 trilhões¹ em ativos sob gestão. Destes, fundos ativos representam cerca de R\$6,5 trilhões, enquanto fundos passivos representam R\$300 bilhões.

Porém, apesar do tamanho da indústria, pouco estudo foi direcionado a pergunta de se os fundos de investimento ativos no Brasil realmente adicionam, em média, valor para seus cotistas.

Nos Estados Unidos, esse debate é mais maduro, originado na década de 60 e reforçado após a publicação da Hipótese de Mercado Eficiente de Eugene Fama. Estudos do mercado americano, principalmente os mais recentes, chegam a conclusão que, em média, gestores ativos americanos são incapazes de gerar retornos superiores à bolsa americana no longo prazo, após pagamento de taxas. O relatório SPIVA (Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G (2023)), por exemplo, demonstra que cerca de 86% dos fundos de investimento americanos registram desempenhos inferiores ao índice S&P500 em uma janela de 10 anos e 94% em 20 anos.

Este *paper*, busca averiguar, por meio de uma análise sobre a performance de fundos de investimentos brasileiros, se, no Brasil, gestores ativos são capazes de gerar retornos superiores ao mercado e ao fundo de índice no curto e longo prazo. Para isso, foram utilizadas as bases de dados disponibilizadas pela Comissão de Valores Mobiliários (CVM) tangendo fundos de investimentos e calculadas estatísticas sobre a performance média de fundos de investimento, sua performance ajustada ao risco e a proporção de fundos de investimento que obtiveram sucesso em produzir retornos superiores ao índice nas diferentes janelas estudadas de 1, 5, 10 e 20 anos.

¹ Dados coletados em 25/11/2023 em https://www.anbima.com.br/pt_br/informar/estatisticas/fundos-de-investimento/fi-consolidado-diario.htm.

Uma das grandes preocupações deste estudo foi a mitigação do viés de sobrevivência, viés de seleção muito presente em estudos de performance e que consiste, no contexto de análises como esta, na seleção somente de fundos que operaram o período completo analisado, desconsiderando fundos descontinuados e fundos criados ao longo do período estudado.

A implicação da introdução deste viés é uma possível superestimação das performances agregadas dos fundos, já que entende-se que somente fundos bem geridos são capazes de sobreviver no longo prazo.

De forma que neste *paper* buscou-se considerar a amostra completa de fundos que operaram no período, refletindo a verdadeira distribuição de resultados. E, foi realizada uma análise de impactos do viés de sobrevivência. Calculou-se estatísticas sobre a performance média de fundos de investimento, assim como a performance média ajustada ao risco desses fundos, considerando a possibilidade de o estudo ter sido conduzido exclusivamente com fundos que estiveram operacionais durante todo o período de análise.

Adicionalmente, a relação entre tamanho de fundos de investimento e sua performance foi analisada por meio de uma regressão e um *scatter plot*.

Neste estudo serão abordados: o motivo por trás da confecção desta análise e sua importância para o debate entre gestão ativa e passiva no Brasil, na seção de Motivação; as bases teóricas do investimento ativo e passivo, exemplos e resultados de análises de performance nos Estados Unidos, na seção de Revisão de Literatura; o processo de obtenção de dados, a construção e tratamento de base, assim como o *framework* para a subsequente análise dos dados, na seção de Metodologia; hipóteses a serem provadas e os resultados pretendidos, na seção de Resultados Pretendidos; os resultados encontrados, a análise destes e suas implicações, na seção de Dados; e, por fim, as conclusões que podem ser tiradas deste estudo, na seção de Conclusão.

Motivação

Em 2007, Warren Buffett, tido por muitos investidores como o melhor investidor de todos os tempos, realizou uma aposta contra a firma de investimentos americana, Protégé Partners, que a gestão passiva iria *outperform*² a gestão ativa líquido de taxas em um prazo de 10 anos. A aposta totalizava U\$1 milhão, com cada lado aportando U\$500 mil para serem destinados a instituições de caridade a escolha do lado vencedor. Nos termos da aposta, Buffett escolheu como o fundo passivo de índice, o Vanguard S&P 500 Admiral (VFIAX), enquanto a Protégé selecionou 5 fundos de investimento em hedge funds³.

Passados os 10 anos, no final de 2017, o fundo de índice escolhido por Buffett gerou performance acumulada de 125,8%, enquanto a média de performance dos 5 fundos de fundos selecionados pela Protégé foi de 36,3%. Ou seja, caso uma pessoa tivesse investido U\$1.000 no fundo passivo escolhido por Buffett, teria acumulado no fim de 10 anos U\$2.258, enquanto uma pessoa que decidiu investir nos 5 fundos escolhidos pela Protégé, teria acumulado U\$1.363.

Em 2017, eu tive a oportunidade de ir a Omaha, NE, USA, para a assembleia de acionistas da Berkshire Hathaway, empresa de Warren Buffett, onde escutei sobre a aposta. Observando os resultados, me questionei se no Brasil, com nosso maior nível de risco e volatilidade, que permite gestores maiores oportunidades para gerarem retornos superiores ao mercado, esse resultado se manteria.

O tópico da diferença de performance entre fundos ativos e passivos no Brasil ainda é um tema pouco explorado, dado o relativo baixo desenvolvimento do mercado de fundos passivos no país e o menor histórico da indústria de fundos quando comparado com Estados Unidos, por exemplo.

Porém, um estudo focado na diferença de performance entre investimento ativo e passivo no Brasil seria muito útil para pessoas buscando investir seu patrimônio no país, ajudando-os a decidir qual a melhor maneira de alocar seu capital.

² *Outperform*: superar/bater.

³ *Hedge funds*, ou fundos de *hedge*, são *pools* de capital que permitem a utilização de estratégias como a compra, venda e venda-descasada de títulos e valores mobiliários como ações, títulos de renda fixa, títulos conversíveis, moedas, derivativos, entre outros de forma a gerar retorno e reduzir risco.

Revisão de Literatura

O estudo comparativo da performance entre fundos de investimento ativos e fundos de investimento passivos é amplamente pesquisado no exterior, com a literatura em sua maioria focando no mercado dos Estados Unidos. O histórico extenso de fundos de investimento permitiu pesquisadores derivarem diversas análises sobre os dados, como a diferença de performance ajustada ao risco (Elton et al. (1996)), o impacto de custos na diferença de performance entre fundos (Wermers (2000)), o fenômeno de *Hot Hands* (Hendricks et al. (1993)) e o impacto do viés de sobrevivência em análises de performance (Malkiel (1995)), exploraremos algumas dessas ideias, que servirão como base teórica para este *paper*.

Antes disso, no entanto, é importante definirmos os principais conceitos por trás do investimento ativo e passivo, assim como seu debate na academia.

1. Bases Teóricas do Investimento Ativo e Passivo

Em 1970, o economista Eugene Fama publicou sua ilustre pesquisa *Efficient Capital Markets: A Review of Theory and Empirical Work* (“Mercado de Capitais Eficientes: Uma Revisão da Teoria e Trabalhos Empíricos”), onde introduziu a Hipótese do Mercado Eficiente (EMH), que foi a se tornar a fundação teórica do investimento passivo. Fama conceptualizou que mercados são “informacionalmente eficientes”, ou seja, os preços de mercado refletem toda a informação disponível e, além disso, que informações passadas não afetam preços futuros. O economista propôs 3 níveis para sua Hipótese de Mercado Eficiente, eficiência: fraca, semi-forte e forte.

A hipótese de eficiência de mercado fraca baseou-se na ideia que preços de ativos refletem toda informação pública disponível, porém não refletem informações privadas no mercado. A veracidade da hipótese de eficiência fraca, combinada a hipótese de independência entre informações passadas e precificação futura (o chamado *random walk*, ou passeio aleatório), implicaria para Fama a rejeição da efetividade de análise técnica⁴, ou seja, não seria possível gerar retornos excedentes ao mercado utilizando-se da leitura de informações passadas para investir (Fama (1970)).

A eficiência de mercado semi-forte, por sua vez, foi construída em cima da efi-

⁴ Análise técnica: Estudo sobre preços passados para prever movimentações de preço futuras.

ciência fraca. Fama postulou que os preços dos ativos se ajustam instantaneamente de acordo com toda informação pública divulgada no mercado. A validade da hipótese de eficiência de mercado semi-forte implicaria que um investidor não seria capaz de gerar retornos excedentes ao mercado utilizando da leitura de informações passadas para investir, nem explorando assimetrias de informação/análise sobre informações públicas, conhecido como análise fundamentalista.

Por fim, a hipótese de eficiência forte englobaria a hipótese semi-forte e adicionaria a ideia que preços refletem a totalidade de informações não só públicas, mas também das privadas, de forma que entende-se que preços de ações estão sempre em seu valor justo e nenhum tipo de informação é capaz de providenciar um diferencial de performance entre investidores e o mercado, nem a obtenção de *insider information*⁵.

Seguidores da hipótese de mercado eficiente buscam, portanto, comprar índices que representem o mercado e pagar o mínimo de taxas possível, no entendimento que nenhum investidor ativo (ou ao menos o investidor ativo que não tenha informação privilegiada - caso semi-forte) consegue atingir performances superiores ao mercado.

Já no lado do investimento ativo, há um entendimento que informações nem sempre são refletidas nos preços das ações, ou pelo menos, não refletidas imediatamente, de forma que há possibilidades de explorar ineficiências de mercado e gerar retornos superiores à média.

Um argumento utilizado frequentemente pelos críticos da Hipótese de Mercado Eficiente é o componente humano na tomada de decisões de investimento que pode gerar irracionalidades na precificação de ativos (ex: bolhas ou grandes *crashes*). Para a sustentação da EMH, Fama implicitamente assume que seres humanos são racionais e eficientes, o que não parece ser verdade.

Psicólogo e economista, Daniel Kahneman, ganhou um prêmio Nobel no campo das ciências econômicas por suas publicações sobre o comportamento humano, no qual averiguou que nós seres humanos tendemos a utilizar vieses cognitivos, como nossa percepção de justiça e aversão à perda, em nossa tomada de decisão. Uma de suas teorias mais famosas, conhecida como *Prospect Theory*, postula que há uma reação assimétrica dos seres humanos em relação à perda, comparado a uma vitória. Kahneman, em seu *paper* de 1979, também introduz uma outra ideia, o *Endowment Effect*, um viés emoci-

⁵ *Insider Information*/Informação privilegiada: informação interna de uma companhia/evento, que não é sabida pelo público geral.

onal humano de atribuir maior valor aos objetos que possuímos (Kahneman, Daniel and Tversky, Amos (1979)). Por fim, um último viés explorado pelo autor é o *Framing Effect*, que explica como os seres humanos alteram suas decisões baseados em como uma escolha é apresentada. Em suma, os vieses cognitivos que influenciam as tomadas de decisões de seres humanos, poderia levar a irracionalidades de investimentos e especulativas, de forma a ferir a EMH.

2. Análises de Performance no Exterior

Nos últimos 70 anos, houve um grande número de estudos publicados buscando comparar a performance de fundos de investimentos ativos e passivos nos Estados Unidos, e com diferentes conclusões.

O primeiro *paper* nesse sentido foi (Jensen (1968)), que calculou a performance de 115 fundos mútuos de 1945 a 1964 e concluiu que a performance de portfólios ativos, após pagamento de taxas, foi inferior à performance de carteiras de ações com risco equivalente, formadas de maneira aleatória. Uma vez somadas de volta às taxas acruadas pelos gestores no período analisado e corrigidas pela taxa livre de risco, Jensen observou que os esforços de gestores ativos em tentar prever e se aproveitar de possíveis movimentações de preços dos ativos não gerou maiores retornos e que preços de ativos parecem se comportar de forma a capturar toda a informação disponível.

Já estudos como (Ippolito (1989)), após analisar a performances de 143 fundos mútuos de 1965 a 1984, concluíram que a performance desses fundos após pagamento de taxas de administração e performance (mas excluindo taxas cobradas de entrada e saída) foi superior a performance de índices, uma vez ajustado ao risco.

No entanto, estudos mais recentes majoritariamente tendem a concordar que, líquido de taxas, fundos de investimento ativos, em média, não são capazes de produzir retornos superiores a retornos de fundos passivos.

(Malkiel (1995)), que analisa as diferenças de performance entre fundos ativos e passivos entre 1971 e 1991, encontra que, no agregado, fundos ativos obtiveram performances inferiores ao índice do S&P 500, antes e depois do pagamento de taxas. Diferentemente da maioria dos estudos publicados anteriormente, que utilizavam para sua análise comparativa os fundos que sobreviveram o período inteiro da amostra analisada, Malkiel utilizou a base completa de fundos operantes durante o período, incluindo fundos que fo-

ram adicionados ou pereceram ao longo da amostra. O autor argumentou que, fundos que aceitam operar com um maior nível de risco terão uma alta probabilidade de falhar, mas caso sobrevivam, isso implica que o fundo realizou uma grande aposta e foi vitorioso. Fundos que apresentam performances ruins tendem a sair do mercado (quebram, desistem do empreendimento ou são adquiridos), de forma que os fundos restantes, considerados sobreviventes, serão os fundos com performances superiores. E, portanto, estudos de performance que consideram somente os fundos que sobreviveram ao período completo da amostra, tendem a superestimar os retornos da média de fundos. Em seu estudo, Malkiel identificou que a diferença de performance entre fundos operantes durante todo o período analisado e a amostra total de fundos foi de, em média, 150 *basis points* (bips) menor para cada ano analisado.

Outro estudo, (Wermers (2000)), que analisou retornos de fundos mútuos de 1975 a 1994, compartilhou das descobertas de (Malkiel (1995)) de que fundos mútuos, em média, não foram capazes de superar o *benchmark* (o portfólio agregado da CRSP da Nasdaq, NYSE e AMEX) líquidos de taxas, *underperforming*⁶ por 100 bips ao ano. Porém, ao destrinchar o retorno médio de fundos, Wermers encontrou a performance das posições de *equities* superaram o benchmark por 130 bips e, que a diferença de 230 bips entre retornos, foi proveniente de (i) performance relativa inferior de 70 bips dos ativos não *equity* e (ii) das diferenças entre custos da gestão ativa e passiva no valor de 160 bips (fundos ativos mais custosos). Portanto, o autor conclui que, ao considerar somente o portfólio de *equities*, gestores de fundos são capazes de produzir performances superiores ao *benchmark* por um valor quase suficiente para superar os custos associados aos fundos.

No *paper* (Hendricks et al. (1990)) os autores exploram o conceito de persistência no retorno de fundos de investimento. Analisando retornos trimestrais de fundos mútuos com objetivo de investimento em crescimento (*growth funds*) entre 1974 e 1988, os autores buscaram avaliar se a performance relativa dos fundos poderia ser prevista em períodos curtos. Separando os fundos em 8 quartis, de acordo com a performance obtida, Hendricks, Patel e Zeckhauser, testaram a acurácia da habilidade de previsão da performance dos fundos nos períodos subsequentes, ou seja, se os fundos continuariam em seus quartis originais ou se moveram para um próximo quartil. Os autores encontraram robustez nos resultados das previsões de performance de fundos no curto prazo, período de 1 a 8 trimes-

⁶ Underperforming: Desempenho inferior, ou ficar aquém de.

tres. Além disso, os autores fizeram testes, com estratégias de investimento baseadas na identificação *ex-ante* dos fundos que tenderiam a obter performances superiores a média (*Hot Hands*) e fundos que iriam apresentar performances relativas inferiores à média de fundos (*Icy Hands*) e obtiveram retornos em excesso ao *benchmark* de 300 a 400 bips por ano.

Metodologia

Este *paper* busca comparar as performances de fundos de investimento ativos e passivos. Para isso, dados da performance diária de fundos de 2003 até 2023 foram obtidos do portal de dados abertos da CVM, assim como a denominação social e classe de cada fundo.

Os dados de performance diárias, separados em documentos agregados mensalmente, foram consolidados utilizando a linguagem de programação R.

Um dos grandes desafios deste *paper* é a confecção e tratamento da base de dados, isto porque as informações de performance diária, classe e denominação social, se encontram em arquivos separados, levando a problemas de integração de bases e *missing data*⁷, e os dados brutos obtidos na CVM contém um número relevante de fundos de másteres⁸ e *feeders*⁹, que têm efeito de duplicação da performance de fundos e alteração da distribuição real das performances dos fundos.

Além disso, como mencionado previamente na seção de Revisão de Literatura, estudos de performance tendem a lidar com problemas de viés de sobrevivência, de forma que é necessário tratar corretamente os dados para extinguir/minimizar este viés.

Estudos como (Jensen (1968)), (Grinblatt and Titman (1989)) e (Hendricks et al. (1990)) incorrem problemas de viés de sobrevivência, uma vez que realizam análises comparativas somente sobre uma amostra de fundos que operaram durante o período completo de suas análises, desconsiderando fundos que “morreram” ao longo do caminho. Muitas vezes, a desconsideração dos fundos não-sobreviventes leva a uma superestimação da performance média de fundos, já que não-sobreviventes tendem a ser fundos que apresentam período relevante de performance inferior (Malkiel (1995)). Além disso, a consideração de fundos que exclusivamente operaram durante todo o período anualizado implica ignorar a entrada de novos fundos, que também podem ter performances diferentes a média (superiores ou inferiores), prejudicando a análise comparativa.

Encontramos problemas de condicionamento de viés de *look-ahead*, uma vez que, inerentemente em estudos de performance, é necessário realizar uma escolha do período

⁷ Missing data: dados faltantes.

⁸ Fundos máster: fundo de investimento em cotas destinado a receber exclusivamente aplicação de outros fundos (ANBIMA, 2022).

⁹ Feeders: Fundos de investimento em cotas de fundo de investimento que aplicam a maior parte de seu capital em cotas de um fundo denominado “máster” (ANBIMA (2022))

mínimo de sobrevivência de um fundo após uma data determinada de análise. Isso implica em um *trade-off* entre complexidade do cálculo dos dados (e necessidade de maior granularidade da base), e viés. Seguindo a metodologia utilizada em (Carhart et al. (2002)), foi utilizado como período mínimo de sobrevivência 1 mês.

A construção da base de dados se iniciou pela consolidação dos dados brutos diários, de valor de cota e do patrimônio gerido dos fundos, em um grande *dataframe*.

Em seguida, foi realizada a integração, com o comando *merge*, da base de dados de cotas consolidada com um *dataframe* de informações cadastrais, composto pelo CNPJ, denominação social e classe dos fundos.

A integração de bases, na primeira tentativa, retornou valores *NA* para denominação social, como confirmado pela utilização da função *is.na* sobre a coluna de 'DENOM_SOCIAL' dos fundos. Para solucionar esse problema a base de dados foi filtrada pelo valor *NA*, utilizando o comando *filter*, e exportada para o Excel. No Excel, foi realizado um segundo processo de integração, no qual a base de informações cadastrais de fundos de investimento foi utilizada para preencher os valores faltantes de denominação social e classe, para esta ação foi utilizado o comando *vlookup (procv)*. Foi realizada a integração da base exportada com a base inicial com o comando *left_join* e, com o comando *coalesce*, foram preenchidos somente os valores *NA* em denominação social e para as informações de classe dos fundos.

Mesmo após a integração de bases pelo método de Excel, ainda havia alguns CNPJs que não possuíam sua contraparte de 'DENOM_SOCIAL' e 'CLASSE' e, portanto, buscou-se uma base além da base de informações cadastrais para prover os valores faltantes. No portal de dados abertos da CVM foram extraídos arquivos referentes ao extrato de informações de cada fundo, que também possuíam informações de denominação social dos fundos e suas classes, de forma que esta base foi importada para o R. Como a base de informações cadastrais, os extratos de informação dos fundos foram integrados à base de fundos por meio do comando *left_join* e os valores faltantes da base de fundos foram preenchidos por meio do comando *coalesce* com os valores da base de extrato de informações. Uma vez concluído que não restava nenhum valor *NA* dentre as denominações sociais dos fundos, prosseguiu-se no tratamento da base.

Nesse ponto, observou-se que havia 48 mil fundos na base, divididos em 15 classes, entre elas: Fundos Multimercado, Fundos de Ações, Fundos de Renda Fixa, Fundos

Referenciados, Fundos de Curto Prazo, etc. Como o objetivo do estudo é comparar a performance de fundos de investimento de ações e multimercados a um índice, buscou-se realizar um filtro na base para remover as classes de fundos indesejadas e selecionar somente os Fundos de Ações e Fundos Multimercado. Porém, antes de realizar a filtragem foram identificados mais de 3,6 mil *missing values* na coluna ‘CLASSE’, que deveriam ser resolvidos.

Diferentemente da denominação social dos fundos, as planilhas de informações cadastrais e de extrato de informações não continham a classe explícita de todos os fundos da base, de forma que para preencher os valores faltantes, foi necessário buscar outras fontes que continham essa informação. Utilizou-se da denominação social dos fundos de investimento.

O Manual para Cadastro de Fundos de Investimento na ANBIMA (ANBIMA (2022)) determina que a denominação social de fundos para fins de registro deverá sempre conter o *Status* (podem ser: FI¹⁰, FC¹¹, FF¹² e FM¹³) e a Classe CVM (Ações, Renda Fixa, Multimercado e Cambial) dos respectivos fundos. De forma que realizou-se uma busca na nomenclatura de cada fundo por sua ‘Classe CVM’, utilizando o comando *grep*, e, de acordo com esta, as células faltantes da coluna ‘CLASSE’ passaram a ser preenchidas. Por fim, o *dataframe* foi filtrado, utilizando o comando *filter*, para selecionar somente fundos cuja ‘CLASSE’ registrava ‘Fundo de Ações’ e ‘Fundo Multimercado’.

Apesar de conter somente fundos multimercados e fundos de ações, objetos do estudo, a base ainda continha fundos de investimento máster e fundos espelhos/*feeders* que, se deixados na base causariam não só dupla contagem nas estatísticas descritivas relacionadas a patrimônio total e patrimônio médio sob gestão, mas também levaria a performances duplicadas e skewness dos resultados.

Para remover os fundos de investimento másteres e espelhos e eliminar o efeito de duplicação de fundos, foi realizada uma filtragem da base novamente pela nomenclatura

¹⁰ FI – Fundo de Investimento: destina-se, em geral, a receber aplicações de investidores de acordo com seu público-alvo, objetivo e política de investimento (ANBIMA (2022)).

¹¹ FC – Fundo de Investimento em Cotas: deve apresentar um valor mínimo de sua carteira em cotas de fundos de investimento, de acordo com a instrução que o rege (ANBIMA (2022)).

¹² FF – Fundo de Investimento: recebe aplicações exclusivamente de outros fundos. Neste caso, o público-alvo necessariamente será fundos de investimento e/ou fundos de investimento em cotas. Se o fundo aceitar aplicação tanto de fundos como de carteiras administradas, deverá ser classificado como FI, e não FF (ANBIMA (2022)).

¹³ FM – Fundo de Investimento em Cotas Master: fundo de investimento em cotas destinado a receber exclusivamente aplicação de outros fundos (ANBIMA (2022)).

dos fundos, agora baseando-se no *Status* de cada fundo. Utilizando do comando *filter*, combinado com o comando *grepl*, foram removidos da base 1.658 fundos, com base no seguinte *Status*: “MASTER”. Diferentemente de fundos másteres, cuja classificação pode ser encontrada no *Status* contido na denominação social de cada fundo, a classificação de fundos espelhos não é identificada na denominação social, necessitando de uma base de dados para prover essas informações adicionais.

A classificação de fundo espelho foi encontrada na base de extrato de informações e foi integrada a base de fundos utilizando o comando *left_join*. A coluna adicionada ‘FUNDO_ESPELHO’ continha 3 diferentes classificações para diferenciar os fundos: ‘S’, no caso do fundo ser identificado como um fundo espelho, ‘N’ no caso do fundo não ser identificado como um fundo espelho e *NA*, caso não fosse aplicável a classificação do fundo como fundo espelho. Entrando em detalhes, a classificação binária entre fundo espelho e não fundo espelho era aplicável somente a fundos de investimento em cotas de fundos de investimento, enquanto a fundos másteres (removidos no passo anterior) e fundos que não investem em cotas de fundos (FIAs e FIMs tradicionais) não lhe aplicavam a classificação como fundos espelhos.

Com base na coluna ‘FUNDO_ESPELHO’ tornamo-nos capazes de filtrar a base. Utilizando o comando *filter*, removemos os fundos espelhos, que se incluídos, iriam trazer *skewness* para os resultados e métricas.

Após a construção da base de dados, buscou-se realizar o cálculo da performance dos fundos. Estudos de performance utilizam 2 modelos principais para analisar o desempenho de fundos de investimento: o *Capital Asset Pricing Model* (CAPM) e o *Carhart 4-factor model*, uma alteração do modelo de 3 fatores publicado por Fama e French (também bastante utilizado). Para este estudo, utilizaremos o CAPM.

O *Capital Asset Pricing Model* (CAPM), como o nome diz, busca precificar um ativo, determinando uma taxa de retorno teórica que seria apropriada para o ativo tomando em conta a suscetibilidade do ativo em relação ao risco de mercado, o retorno esperado do mercado e a taxa livre de risco (Kenton W. (2023)). A fórmula do CAPM é dada por:

$$E(R_i) = R_f + \beta_i * (E(R_m) - R_f) \quad (1)$$

Onde $E(R_i)$ representa o retorno teórico esperado do ativo, R_f representa o retorno do ativo livre de risco, β_i representa a sensibilidade do ativo a variações do mercado, e $E(R_m)$ representa o retorno esperado do mercado.

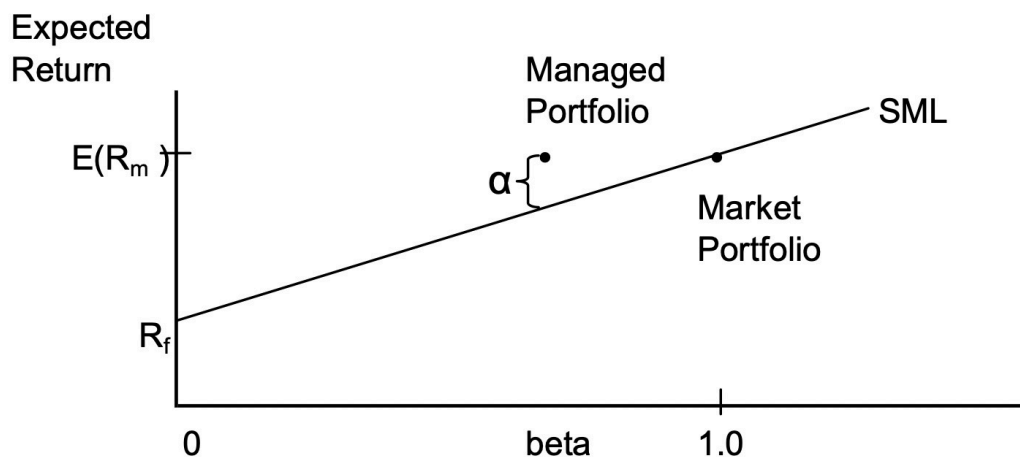


Figura 1 – Demonstração de alfa, ou excesso de retorno. Womack and Zhang (2003)

De outra maneira, o CAPM permite calcular o prêmio de risco teórico que deveria ser exigido por estar investido no ativo.

Uma vez que é possível comparar a performance do fundo de investimento com o retorno teórico esperado do fundo dado o risco incorrido pelo gestor, pode-se, portanto, utilizar o modelo de precificação de ativos para avaliar a performance de fundos de investimento. A diferença entre o retorno teórico e o retorno efetivo atingido pelo fundo é entendido como o alfa (α) atribuído ao investimento ativo, caso o alfa seja positivo, significa que o gestor foi capaz de acionar valor além do esperado para um ativo com o aquele nível de risco, caso seja negativo ou zero, o gestor foi incapaz de adicionar valor.

Uma vez estabelecido o modelo a ser utilizado para auferir e segregar a performance dos fundos de investimento, partimos para a obtenção de dados para realizar os cálculos.

1. Performance dos Fundos de Investimento

Primeiramente, buscou-se calcular a performance efetiva dos fundos de investimento durante o período analisado. Para calculá-la, foram utilizados os valores de cota para cada fundo, encontrados na coluna 'VL_QUOTA' da base de fundos.

Seguindo estudos como (Carhart et al. (2002)), foi decidida a realização de cálculos mensais de performance dos fundos.

Agrupando os fundos por seu CNPJ e a última data de cada mês, com o comando *group_by*, foi realizado um cálculo de variação, seguindo a fórmula abaixo para encontrar a performance mensal de cada fundo e registrá-la em uma nova coluna na base de dados.

$$PerformanceMensal = \frac{VL_QUOTA_t}{VL_QUOTA_{t-1}} \quad (2)$$

O cálculo de performance por meio da variação de cotas permite calcularmos o retorno do fundo de investimento líquido de taxas, uma vez que os valores referentes a taxa de administração e taxa de performance de cada fundo são provisionados e pagos durante o período de operação dos fundos, conforme seu regulamento, invariavelmente traduzindo-se na cota de cada fundo. Uma prática comum em fundos de investimento é o provisionamento diário de taxa de administração e taxa de performance, para pagamento semestral.

2. Ativo Livre de Risco

Em estudos de performance nos Estados Unidos, a proxy utilizada para a taxa livre de risco são os títulos do Tesouro Americano de 10 anos (*10-year treasury bills*).

O título do tesouro, em especial o título de 10 anos, é o *benchmark* títulos corporativos e títulos de agências como a Fannie Mae e Freddie Mac¹⁴, e suas movimentações impactam juros hipotecários, empréstimos automotivos, pessoais, de educação (*student loans*) entre muitos outros métodos de financiamento da economia americana (Franck (2018)).

O título é tido como livre de risco, devido ao fato de ter como devedor o governo dos Estados Unidos, instituição que, entende-se, têm baixíssima probabilidade de *default*¹⁵.

Para este estudo, buscou-se por um ativo financeiro brasileiro que servisse de proxy ao título americano para classificar-se como ativo livre de risco.

O ativo escolhido foi a Letra Financeira do Tesouro (LFT). Como o título de 10 anos emitido pelo Tesouro Americano, a LFT também possui risco soberano, mas neste caso, referente a capacidade e disposição do Tesouro Brasileiro em honrar pontualmente e de forma integral os pagamentos de sua dívida (Tesouro Nacional (2023)).

Porém, diferentemente do *treasury bill* americano, a Letra Financeira do Tesouro é um título pós-fixado, ou seja, este é corrigido a uma taxa de retorno que não é pré-acordada na compra do título; o retorno do título só pode ser auferido no vencimento da aplicação.

¹⁴ Fannie Mae e Freddie Mac: Companhias apoiadas pelo governo federal americano que provêm liquidez para o sistema hipotecário. As companhias compram hipotecas de credores, podendo mantê-las em portfólios ou agrupá-las e vendê-las como ativos (Federal Housing Financing Agency (n.d.)).

¹⁵ Default: Não pagamento ou pagamento em atraso ao prazo contratual da dívida.

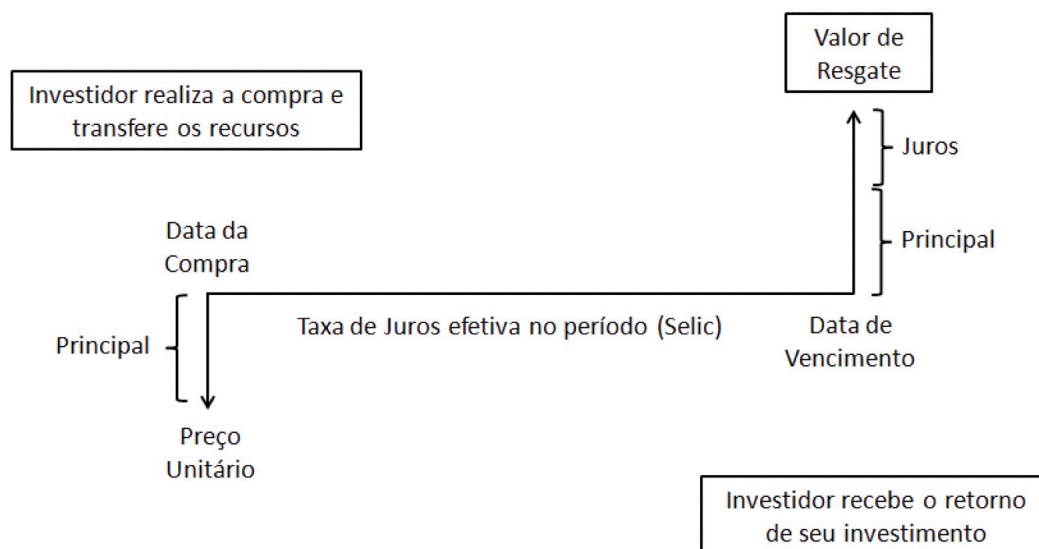


Figura 2 – Funcionamento LFT (Tesouro Direto, n.d.)

Os títulos pós-fixados são atrelados a um indexador, no caso das LFTs, o indexador da taxa do título é a Taxa Selic¹⁶.

A LFT foi escolhida como ativo livre de risco dado que foi o título governamental que possuía o maior prazo de vencimento no histórico do Tesouro Direto desde 2003. A LFT de vencimento em 3 anos foi acompanhada mensalmente de 2003 até 2023 e seus retornos foram integrados à base.

3. Retorno de Mercado

Como mencionado na seção de revisão de literatura, nos Estados Unidos, predominantemente, utiliza-se do índice do S&P 500 para medir e representar o retorno de mercado. Para fins desse estudo, utilizarei o índice Bovespa, a mais conhecida representação da bolsa brasileira.

Detalhando-o, o índice Bovespa inclui os ativos, não-*Penny Stock*¹⁷, que representam 85% do número de negócios e volume financeiro do mercado de capitais. A representatividade das companhias dentro do Ibovespa é calculada ponderando o valor de mercado do *free float*¹⁸, porém institui-se limites de concentração por empresa: até 20% de representação do índice baseando-se no cálculo ponderado por valor de mercado do *free float*

¹⁶ Taxa Selic: Taxa básica de juros da economia.

¹⁷ *Penny Stock*: Ações negociadas na bolsa de valores a valores muito baixos, muitas vezes centavos por ação.

¹⁸ *Free float*: Uma fração das ações de companhias listadas na bolsa de valores que estão livres para negociação.

e limite de até 2x a representatividade do ativo na carteira caso o método de ponderação do índice fosse o índice de negociabilidade¹⁹(B3 (n.d.)).

Apesar das limitações de concentração, o valor de mercado do *free float* é o grande determinante do tamanho proporcional das empresas dentro do índice. As implicações desta metodologia de balanceamento são que poucas empresas podem representar mais de 41% do índice, como é o caso hoje²⁰, e, portanto, movimentações no índice Bovespa não necessariamente representam o movimento do mercado como um todo.

4. Beta

Por fim, o beta (β) calculado tomando em conta a covariância entre o retorno de cada fundo de investimento e o retorno do índice Bovespa no período analisado, dividido pela variação do índice.

$$\beta = \frac{COV(R_i, R_{Ibovespa})}{Var_{R_{Ibovespa}}} \quad (3)$$

5. Calculando o Alfa do CAPM

Agora, buscarei detalhar a obtenção dos dados e tratamento da base para a realização dos cálculos mencionados acima.

O histórico das taxas da LFT de 3 anos foram retirados da página do Tesouro Direto. Arquivos anuais, contendo retorno diário do título, foram baixados e importados.

No Excel, criou-se uma lista agregando os dados diários de preço de compra e venda dos títulos de 2003 até 2023. Uma vez formada a lista, foram filtrados os valores de compra e venda referentes ao último dia de cada mês, de forma a calcular a rentabilidade mensal da taxa livre de risco. Seguindo a metodologia provida pelo tesouro nacional, calculou-se a rentabilidade do título:

$$R_f = \frac{PreodeVenda}{PreodeCompra} - 1 \quad (4)$$

Assumiu-se uma compra do título no último dia do mês anterior e uma venda do título no último dia do mês corrente. Os dados de rentabilidade mensal dos títulos foram integrados à base de fundo na coluna 'RISK-FREE RATE'.

¹⁹ Índice de Negociabilidade: Medida que indica o volume de negociação de um ativo na bolsa de valores (Mais Retorno, 2020).

²⁰ Coletado em 29/11/2023 em https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carreira.htm. Vale, Petrobras, Itaú, Bradesco e Eletrobras representam 41,06% do índice Bovespa.

Já o retorno de mercado foi obtido utilizando a ferramenta de obtenção de dados financeiros, Economática. A cotação diária do índice Bovespa foi armazenada. No R, foram selecionadas, como os valores de cota, a cotação do último dia de cada mês, e foi realizado um cálculo de variação mensal. Ao fim desse processo, os dados de performance do índice foram incorporados à base de performance de fundos na coluna ‘RETORNO DE MERCADO’.

Uma vez obtido o retorno mensal dos mercados e o retorno dos fundos, buscou-se calcular o beta de cada fundo. Tomou-se a covariância do retorno mensal dos fundos e do retorno mensal dos mercados da amostra inteira de performance e dividiu-se pela variação da amostra completa de performance do índice.

O cálculo do alfa (α) foi realizado por meio da modificação da equação do *Capital Asset Pricing Model*. Sendo o alfa o excesso de retorno do fundo acima do retorno esperado para ativos com o mesmo nível de risco, buscou-se calcular portanto a diferença entre o retorno efetivo atingido (R_i) e o retorno esperado ($E(R_i)$) do fundo.

$$\alpha = (R_i) - (R_f + \beta_i * (E(R_m) - R_f)) \quad (5)$$

Seguindo estudos como (Carhart et al. (2002)) e (Malkiel (1995)), buscou-se criar um índice de fundos ativos, agregando a performance de todos os fundos e calculando uma performance média que, posteriormente, seria comparada com a performance dos fundos de índice passivos. O método de agregação da performance dos fundos pode ser feito de duas maneiras: *equal-weighted* ou *value-weighted*.

No método *equal-weighted* (EW), ou de pesos equivalentes, a performance média dos fundos é calculada por uma média aritmética da performance dos fundos no período determinado. De forma que todos os fundos têm o mesmo peso na amostra.

$$PerformanceMensalMedia = \frac{1}{n} * \sum_{n=1}^n R_i \quad (6)$$

Já no método *value-weighted* (VW), a média da performance dos fundos é realizada atribuindo graus de importância às performances da base de dados, pesando pelo patrimônio dos fundos no período analisado.

$$PerformanceMensalMedia = \frac{1}{n} * \sum_{n=1}^n w_i * R_i \quad (7)$$

Na academia há uma grande discussão sobre qual método de agregação é superior. Estudos como (Carhart (1997)), demonstraram que o grupo de fundos não-sobreviventes

é composto, em média, por fundos de menor porte, de forma que a agregação da performance individual de fundos por *equal weight* (mesmo peso) faz com que fundos menores tenham o mesmo impacto no índice de performance que fundos grandes e diminuam o retorno agregado dos fundos de investimento. Por outro lado, entende-se que quão maior o tamanho de um fundo, menor é seu aquário de oportunidades, de forma que torna-se cada vez mais difícil para fundos maiores gerar retornos superiores à média de mercado e, portanto, em uma metodologia de agregação *value-weighted* (pesado por tamanho), fundos grandes poderiam influenciar negativamente o retorno agregado dos fundos.

Neste estudo foi feito o cálculo da performance média utilizando o método de ponderação *equal-weight*, porém buscou-se entender o impacto de tamanho na performance de fundos de investimento. Para isso, foi construída uma regressão de forma a analisar impactos de valor do patrimônio dos fundos, e outras métricas descritivas relacionadas ao patrimônio dos fundos, sobre sua performance.

Seguindo (Carhart et al. (2002)), o cálculo do índice de performances foi realizado mensalmente, de forma que a média de performance dos fundos contém todos os fundos que estiveram operantes durante o mês completo. O rebalanceamento do índice é feito mensalmente, excluindo os fundos que deixaram de operar e que, portanto, não possuem mais dados e incluindo os fundos novos, que operaram o mês completo e tiveram sua performance registrada. Dessa forma, faz-se possível considerar a performance de fundos que eventualmente foram descontinuados durante o período completo analisado e minimizar o viés de sobrevivência.

Além do índice da performance geral de fundos ativos, foram feitos índices de performance separados para Fundos Multimercados e Fundos de Ação. As performances dos respectivos índices ativos foram, então, comparadas com as performances do índice passivo no período completo analisado e, em diferentes períodos.

Além disso, de forma a analisar o impacto da utilização correta dos dados na análise comparativa, buscou-se calcular o tamanho do viés de sobrevivência, assumindo a realização do estudo selecionando somente os fundos que sobreviveram ao período total da análise. Em outras palavras, foi comparado o índice de performance de fundos ativos obtido considerando tanto os fundos sobreviventes, como os que não sobreviveram o período completo da amostra, e os fundos criados durante os períodos analisados, a um índice de performance composto somente de fundos que sobreviveram ao período total da análise.

Para isso, foi realizado um filtro na base de dados, de forma a segregar somente os fundos operantes no último mês registrado da pesquisa, julho de 2023. Colocou-se os CNPJs desses fundos em uma lista. Concomitantemente a base de dados foi filtrada para segregar somente os fundos operantes durante o primeiro mês do primeiro ano do estudo, janeiro de 2003.

Utilizou-se da lista de CNPJs de fundos ativos em 2023 para filtrar a base de fundos de 2003, de forma a selecionar somente os fundos existentes durante todo o período da amostra.

Com os CNPJs desses fundos, a base completa foi filtrada com intuito de acompanharmos a performance somente dos fundos sobreviventes (chamados de *survivors*). Foi criado, portanto, um novo índice de performance de fundos ativos, seguindo as mesmas instruções de (Carhart et al. (2002)), porém consistindo da média de performance somente dos fundos que sobreviveram ao período completo da pesquisa. Foram realizados, também, índices individuais para os fundos multimercados e fundos de ações. E, por fim, a performance mensal dos índices de *survivors* foi comparada com a performance mensal do índice da amostra completa de fundos para mensurarmos o tamanho do viés de sobrevivência.

Por fim, buscou-se realizar uma análise sobre a proporção de gestores capazes de obter performances superiores ao fundo passivo, BOVA11. Para isso, calculou-se a performance absoluta acumulada considerando janelas de 1, 5, 10 e 20 anos para todos os fundos. Estes retornos foram, então, comparados com o retorno acumulado do fundo de índice no período e os resultados registrados. A análise foi feita não só considerando a amostra completa de fundos, mas também buscou-se entender a proporção de fundos de ações e de fundos multimercados capazes de obter desempenhos superiores ao fundo passivo. Novamente, buscando proteger a análise de viés e seguindo exemplos como (Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G (2023)), foram incluídos fundos que permaneceram ao longo do período. Além da análise de performances absolutas nas janelas de 1, 5, 10 e 20 anos, foram calculadas as proporções de fundos que obtiveram desempenhos superiores ao BOVA11, ajustados ao risco.

Resultados Pretendidos

Conforme mencionado anteriormente, o Brasil carrega um maior nível de risco, quando comparado com países desenvolvidos, como os Estados Unidos. Esse maior nível de risco pode proporcionar maiores oportunidades para gestores ativos gerarem retornos superiores ao mercado e realizá-los com menor volatilidade. De forma que, quando comparamos com os Estados Unidos, espera-se que, no Brasil, haja uma menor proporção de fundos com performances inferiores ao mercado ou performances acumuladas similares, porém com um menor nível de risco.

Dados

A base de dados, com informações coletadas do Portal de Dados Abertos da CVM de 2003 até 2023, conteve 32.603 fundos de investimento. Destes, 26.690 classificaram-se como Fundo Multimercado e 5.913 classificaram-se como Fundo de Ações. Ao longo dos 21 anos analisados no estudo, foram criados 32.022 fundos e 15.634 foram descontinuados. Em janeiro de 2003 haviam 581 fundos operantes, já em julho de 2023, somavam-se 17.754 fundos.

A Tabela 1 explicita algumas estatísticas dos fundos de investimento analisados. O ano de 2021 teve o maior número de fundos criados, cerca de 3,5 mil fundos; seguidos por 2004, com 2,76 mil fundos criados, 2020 com 2.6 mil e 2022, com o surgimento de 2,5 mil fundos. Analisando os fundos descontinuados, destaca-se que o ano de 2022 apresentou o maior registro de descontinuação, com 1,3 mil fundos sendo encerrados, seguidos pelos anos de 2015, 2008, 2016 e 2014. Durante o período analisado, o AUM médio dos fundos variou entre R\$173,3 milhões e R\$112,5 milhões.

No período analisado, o tempo médio de sobrevivência dos fundos de investimento foi de 59,7 meses, ou cerca de 5 anos. A mediana de sobrevivência dos fundos foi de 40 meses.

Fundos de ações tiveram a maior longevidade média, 60,8 meses, porém não muito distinto da média de 59,4 meses dos fundos multimercados. Olhando para a mediana, fundos multimercados tiveram longevidade superior, 40 meses, comparado aos 38 meses de longevidade dos fundos de ações. O período mínimo de sobrevivência foi de 2 meses

Ano	Número de Fundos	Fundos Criados	Fundos Descontinuados	AUM Médio
2003	618	37	93	173.343.868
2004	3.288	2.763	79	135.605.513
2005	3.756	548	468	134.456.737
2006	4.016	729	393	146.277.560
2007	5.617	1.993	395	127.048.903
2008	6.543	1.321	1.144	112.489.592
2009	6.325	926	700	124.487.401
2010	7.056	1.431	521	122.468.945
2011	7.962	1.431	792	123.940.790
2012	8.646	1.472	814	125.679.393
2013	9.538	1.705	843	126.101.335
2014	9.781	1.086	1.098	128.629.852
2015	9.684	1.001	1.168	136.771.443
2016	9.695	1.181	1.105	145.252.229
2017	9.984	1.392	792	150.741.820
2018	10.579	1.397	769	156.235.501
2019	11.716	1.909	736	157.057.621
2020	13.623	2.638	681	154.405.869
2021	16.487	3.543	936	146.651.450
2022	18.098	2.542	1.322	138.020.252
2023	17.754	977	785	138.193.962

Tabela 1 – Estatísticas Descritivas de Fundos de Investimento.

e o período máximo foi de 247 meses, período completo analisado.

1. Análise de Performance Média de Fundos de Investimento

Os retornos mensais e os alfas dos 32.603 fundos de investimento na amostra ao longo do período de 2003 a 2023. O alfa dos fundos foi registrado líquido de taxas, conforme o pagamento das devidas taxas de administração e performance de cada fundo e seu impacto no valor de cota. Os alfas mensais de cada fundo foram, então, agregados e

anualizados; seus resultados foram disponibilizados na Tabela 2.

Ano	Retorno	Retorno Ibovespa	Alfa	Média de Fundos em Op.
2003	69,57%	97,34%	-8,36%	553
2004	5,70%	17,81%	1,22%	2.216
2005	19,31%	27,71%	-0,88%	3.273
2006	20,02%	32,93%	1,68%	3.490
2007	16,83%	43,65%	-0,08%	4.496
2008	-5,21%	-41,22%	-2,16%	5.577
2009	26,83%	82,66%	2,30%	5.477
2010	8,83%	1,04%	2,05%	6.181
2011	3,96%	-18,11%	0,65%	6.986
2012	12,02%	7,40%	3,96%	7.544
2013	2,52%	-15,50%	-0,12%	8.402
2014	4,06%	-2,91%	-3,06%	8.817
2015	7,44%	-13,31%	-0,12%	8.703
2016	14,39%	38,94%	-4,53%	8.594
2017	12,79%	26,86%	-0,78%	9.013
2018	7,26%	15,03%	-0,62%	9.577
2019	14,22%	31,58%	2,36%	10.386
2020	5,79%	2,92%	1,59%	12.091
2021	-1,52%	-11,93%	-1,84%	14.515
2022	3,04%	4,69%	-6,67%	16.384
2023	6,71%	11,13%	-1,21%	17.014
Retorno Acum.	847,98%	982,21%	-14,44%	-

Tabela 2 – Estatísticas de Retorno Fundos de Investimento de 2003 a 2023.

Observa-se na Tabela 2 que, durante os anos de 2003 e 2023, o Ibovespa teve performance positiva em 15 dos 21 anos analisados, enquanto o índice agregado de fundos gerou um retorno positivo em 19 desses anos.

Apesar disso, no entanto, o Ibovespa apresentou um retorno acumulado de 982,2%, um retorno anual médio composto de aproximadamente 12,01% a.a., enquanto o índice agregado de fundos apresentou um retorno realizado acumulado de 847,98%, ou cerca de 11,30% a.a.

A diferença de retorno entre o índice agregado de fundos e o Ibovespa é explicada pela magnitude de avanço do índice nos anos que este foi positivo. Dos 15 anos de performance positiva do Ibovespa, em 12 o Ibovespa apresentou retornos superiores ao índice agregado de fundos. O delta de retorno médio durante esses anos foi de 1.780 bips a favor do Ibovespa. Esse delta gerado foi suficientemente grande para encobrir os 9 anos de performance inferior ao índice agregado de fundos, nos quais houve uma diferença média de retorno 1.439 bips a favor do índice de fundos.

Uma vez comparado o retorno absoluto do índice agregado de fundos com o Ibovespa, faz-se necessário o cálculo do alfa para medir o retorno dos fundos ajustado ao risco.

Observa-se que o índice agregado de fundos demonstrou alfas positivos em somente 8 anos dos 21 anos analisados. O maior alfa apresentado pelo índice agregado de fundos foi em 2012, o retorno registrado foi cerca de 396 bips superior ao retorno esperado dado o nível de risco dos fundos e, o menor, foi em 2003, cerca de 836 bips inferior ao retorno esperado.

De 2003 até 2023, o índice agregado de fundos de investimento atingiu um alfa acumulado de -14,44%, um desempenho inferior ao retorno esperado para fundos com seu determinado nível de risco, líquido de taxas, em cerca de 74 bips ao ano.

Com isso, é possível concluirmos que os gestores ativos, em média, não foram capazes, após o pagamento de taxas, de gerar retornos acima do retorno esperado para ativos com aquele nível de risco. Ou seja, em média, gestores foram incapazes de gerar alfa para seus cotistas.

Após a realização de uma análise sobre a performance agregada dos fundos de investimento, buscou-se segregar a amostra de fundos em Fundos Multimercado e Fundos de Ações para compararmos os resultados.

Como apontado previamente, durante os 21 anos analisados neste paper, observaram-se 26.690 fundos multimercados e 5.913 fundos de ações. O AUM médio gerido por fundos de ações foi de R\$102.667.633, enquanto o de fundos multimercados foi de R\$98.924.553.

Ano	Retorno Ações	Retorno Multimercados	Retorno Ibovespa	Alfa Ações	Alfa Multimercados
2003	70,71%	54,43%	97,34%	-8,72%	-3,12%
2004	23,72%	7,22%	17,81%	5,85%	-2,25%
2005	25,56%	17,71%	27,71%	-0,63%	-0,95%
2006	32,25%	17,32%	32,93%	1,97%	1,61%
2007	38,43%	12,76%	43,65%	0,58%	-0,24%
2008	-39,57%	4,31%	-41,22%	-7,00%	-1,06%
2009	75,57%	16,31%	82,66%	4,72%	1,68%
2010	5,88%	9,40%	1,04%	3,06%	1,78%
2011	-12,96%	8,64%	-18,11%	-0,57%	0,96%
2012	13,89%	11,44%	7,40%	5,29%	3,61%
2013	-5,15%	4,41%	-15,50%	5,15%	-1,40%
2014	-6,17%	6,50%	-2,91%	-7,03%	-2,08%
2015	-9,57%	11,55%	-13,31%	-3,16%	0,61%
2016	27,25%	11,56%	38,94%	-4,96%	-4,44%
2017	27,00%	10,22%	26,86%	3,18%	-1,51%
2018	13,74%	5,95%	15,03%	0,21%	-0,78%
2019	37,58%	10,23%	31,58%	9,05%	1,15%
2020	5,47%	4,78%	2,92%	3,21%	1,29%
2021	-9,49%	0,28%	-11,93%	-1,38%	-1,92%
2022	-7,37%	5,38%	4,69%	-13,16%	-5,13%
2023	14,36%	5,01%	11,13%	3,80%	-2,30%
Retorno Acum.	1012,72%	758,18%	982,21%	-3,69%	-13,95%

Tabela 3 – Estatísticas de retorno fundos de investimento em ações e fundos multimercados de 2003 a 2023.

De 2003 até 2023, o Ibovespa apresentou um retorno acumulado de 982,2%. Neste período, o índice agregado de fundos multimercados registrou retornos positivos em todos os anos do estudo, porém acumularam um retorno de 758,2%, cerca de 123 bips ao ano menor. Já o índice agregado de fundos de ações apresentou retornos positivos em 14 dos

21 anos estudados e um retorno acumulado de 1.012,7%, cerca de 15 bips por ano superior ao índice Bovespa e 138 bips ao ano superior ao índice agregado de fundos multimercados.

Da mesma forma que foi feito na com o índice agregado de fundos geral, é necessário observarmos não só o retorno absoluto dos fundos e do mercado, mas também tomar em conta o risco incorrido para o atingimento desses retornos. Portanto, buscou-se calcular o alfa dos índices agregados de fundos de ações e multimercados.

Durante os 21 anos analisados, o índice de fundos multimercados gerou retornos acima do retorno esperado para ativos com o mesmo nível de risco em 8 desses anos, enquanto o índice de fundos de ações gerou alfa em 12. O maior alfa apresentado pelo índice de fundos de ações foi em 2019, cerca de 905 bips acima do retorno esperado para ativos com o mesmo nível de risco, seguidos por 2004 (585 bips), 2012 (529 bips) e 2013 (515 bips). Já o índice de fundos multimercados registrou alfas mais tímidos, com o seu maior descolamento em 2012 (361 bips), seguidos por 2010 (178 bips) e 2009 (168 bips).

Quando observamos retornos abaixo ao retorno esperado dado o nível de risco incorrido de investir nesses ativos, o índice de fundos de ações apesar de apresentar retornos negativos em somente 9 dos 21 anos, possui as diferenças mais agudas. O menor alfa registrado pelo índice de fundos de ações foi em 2022, 1.316 bips abaixo do retorno esperado para ativos com seu nível de risco, seguidos por 2003 (-872 bips), 2014 (-703 bips), 2008 (700 bips) e 2016 (-496 bips). Enquanto isso, o menor alfa apresentado pelo índice de fundos multimercados foi de 513 bips abaixo do retorno esperado para ativos com o mesmo nível de risco, registrado em 2022, seguidos também por 2016 (-444 bips) e 2003 (-312 bips). Observa-se, portanto que, apesar de descolamentos acima do retorno esperado ajustado ao risco mais tímidos, fundos multimercados em um determinado ano também apresentam *underperformance* menor a esse retorno esperado.

Por fim, ao olharmos para o alfa acumulado, pode-se notar que, de 2003 até 2003, tanto o índice agregado de fundos de ações, como o índice agregado de fundos multimercados tiveram alfas acumulados negativos.

O alfa gerado pelo índice agregado de fundos multimercados foi de -13,95% nos 21 anos estudados, cerca de 71 bips ao ano abaixo do retorno esperado para ativos com seu nível de risco. Já o alfa gerado pelo índice de fundos de ações foi de -3,69%, cerca de 18 bips ao ano inferior ao retorno ajustado ao risco incorrido de estar investido nesses ativos. Comparando os resultados, observa-se que investidores olhando tanto pela ótica

de retorno absoluto, quanto pela ótica de retorno ajustado ao risco, estariam melhor amparados investindo no índice agregado de fundos de ações, comparado ao índice de fundos multimercados.

Porém, observando os retornos produzidos por ambas as classes de ativos, chegamos à conclusão que tanto os gestores de fundos multimercados, quanto os gestores de fundos de ações, em média, não são capazes de gerar alfa para seus cotistas, líquidos de taxas.

2. Tamanho e Performance de Fundos de Investimento

Após a realização de uma análise geral sobre a performance de fundos de investimentos, buscamos entender a relação entre tamanho e performance de fundos.

Como mencionado previamente na seção de Metodologia, estudos como (Carhart (1997)) provam que o grupo de fundos descontinuados ao longo de análises de performance é normalmente composto por fundos de menor porte e, no contexto prático, é entendido que um número maior de ativos sob gestão contribuem para maior uma maior alavancagem operacional do negócio de gestão de ativos, uma vez que custos de operação de fundos grandes e pequenos não são tão distintos, o que permite maiores lucros e remuneração da equipe de investimento, levando a atração e retenção de bons talentos, e portanto, performance.

Porém, acredita-se que a partir de um certo tamanho, a liquidez de ativos passa a tornar-se um problema, uma vez que, para montar e desmontar seus investimentos, fundos grandes necessitam de muito mais tempo de forma a não influenciar a trajetória de preço dos ativos significativamente, o que pode inviabilizar investimentos e diminuir o aquário de oportunidades.

Para tentar medir efeitos de tamanho sobre a performance de fundos, foi realizada uma regressão de variáveis explanatórias, como beta do fundo, excesso de retorno de mercado acima da taxa livre de risco, o *lag* do valor do patrimônio líquido do fundo, a classe do fundo, a relação entre o *lag* do valor do patrimônio líquido e a classe, o *lag* da performance mensal do fundo e a relação entre o *lag* do valor do patrimônio líquido e a performance mensal defasada sobre o excesso de performance de fundos acima da taxa livre de risco.

$$\begin{aligned}
 R_i - R_{LFT} = & \beta * (E(R_{Ibovespa}) - R_{LFT}) + lag(Patrim.Liq.) \\
 & + Classe + lag(R_i) + lag(Patrim.Liq.)xlag(R_i) \\
 & + Classexlag(R_i) + lag(Patrim.Liq.)xClasse + Idade
 \end{aligned}
 \tag{8}$$

	Estimativa		Erro Padrão
(Intercept)	1,83E-01	.	1,09E-01
BETA_FUNDO	-9,22E+00	***	1,05E-01
MKT_PERF	-1,09E+00	*	5,23E-01
lag(VL_PATRIM_LIQ)	1,31E-10	*	6,43E-11
factor(CLASSE) Fundo Multimercado	-1,20E+00	***	1,00E-01
lag(PERF_MONTH)	9,60E+01	***	1,26E+00
IDADE	4,07E-03	***	5,42E-04
BETA_FUNDO:MKT_PERF	9,47E+02	***	1,22E+00
lag(VL_PATRIM_LIQ) : factor(CLASSE) Fundo Multimercado	-1,89E-10	*	8,10E-11
lag(VL_PATRIM_LIQ) : lag(PERF_MONTH)	-5,06E-10		7,25E-10
factor(CLASSE) Fundo Multimercado : lag(PERF_MONTH)	-5,58E+01	***	1,41E+00

Tabela 4 – Regressão de retorno de fundos de investimento em excesso da taxa livre de risco.

Observa-se pela tabela 4 que a maior parte dos coeficientes introduzidos para as variáveis independentes são estatisticamente significativos ao nível de significância de, pelo menos, 5%. Prosseguiremos para uma análise sobre as variáveis explicativas descritivas.

O fator CLASSE, nesse caso uma *dummy* para a classe de fundos ‘Multimercado’ apresenta um coeficiente negativo, apontando para o fato que, mantendo as outras variáveis do modelo constantes, fundos multimercados produzem uma performance mensal menor que ‘Fundos de Ações’ (fator Classe ‘Multimercado’ = 0).

Regredindo a performance mensal dos fundos sobre a defasagem de suas performances mensais, podemos notar uma relação positiva e altamente significativa. Cada 100 bips de aumento na performance mensal defasada do fundo, está associado a um aumento de, em média, 9,6 bips na performance mensal do fundo. A idade do fundo também se mostrou altamente significativa e carregando uma relação positiva com a performance mensal em excesso da taxa livre de risco do fundo, cada aumento de 1 mês na idade do fundo está associado a um aumento de, em média, 0,0004 bips no excesso de performance do fundo.

Abordando a relação entre tamanho e performance, encontramos relação significativa considerando um intervalo de confiança de 95% entre o lag do patrimônio líquido do

fundo e o excesso de retorno mensal do fundo sobre a taxa livre de risco. Um aumento de R\$1 milhão no ativo sob gestão, apontaria para um aumento $1,3e-06$ na performance mensal em excesso a taxa livre de risco. E, ao relacionar o lag do patrimônio líquido do fundo com sua classe, podemos notar que há uma relação inversa e estatisticamente significativa entre a interação e o excesso de performance de fundos, ou seja, conforme maior seja o patrimônio líquido defasado de fundos multimercados, em média, a performance mensal do fundo é marginalmente menor. Um aumento de R\$1 milhão do patrimônio líquido defasado do multimercado, aponta para uma redução de $-1,9e-04$ bips.

Porém, ao relacionar o *lag* do patrimônio líquido do fundo com a performance mensal defasada do fundo, observa-se uma relação negativa, mas não estatisticamente significativa.

Buscando investigar mais a relação entre tamanho e performance, foi construído um *scatter plot* utilizando como o eixo vertical a performance mensal dos fundos de investimento e, como o eixo horizontal, o patrimônio líquido dos fundos.

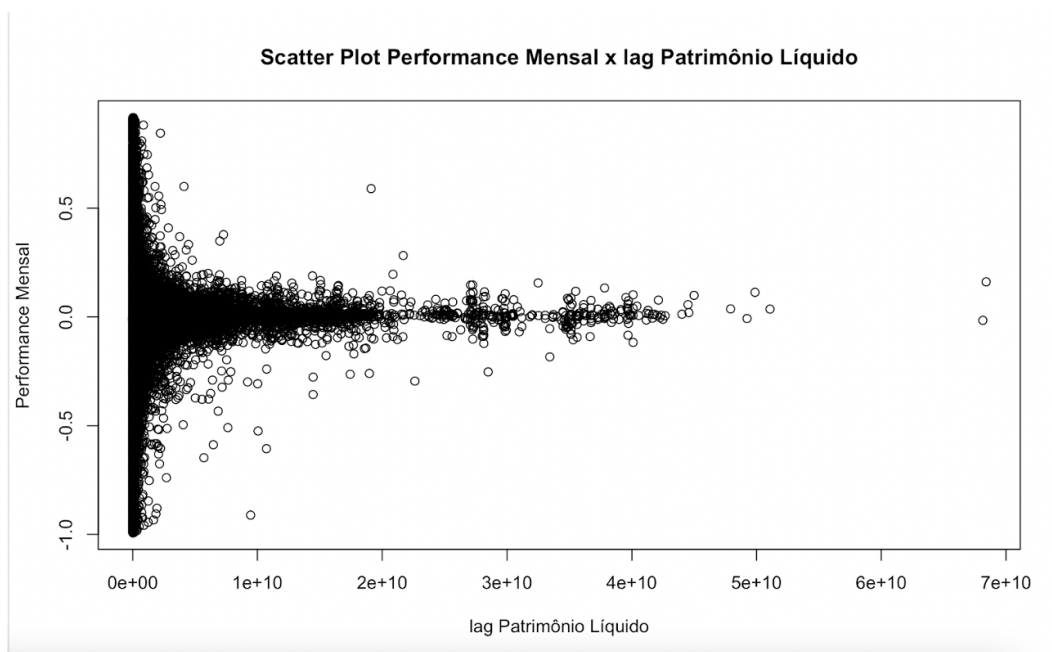


Figura 3 – *Scatter Plot* Performance Mensal Fundos de Investimento e Patrimônio Líquido Defasado.

O *scatter plot* aponta para uma maior volatilidade de resultados considerando fundos menores, entre R\$0 a R\$5 bilhões de patrimônio. Nesta categoria, podemos ver uma grande quantidade de fundos com performances significativamente superiores e inferiores a 0, porém há uma convergência mais rápida de fundos cuja performance mensal foi

positiva para o eixo. Conforme, fundos aumentam de tamanho, acompanha-se que a performance destes se aproxima cada vez mais ao eixo.

3. Survivorship Bias

Conforme abordado na Seção de Revisão de Literatura, estudos sobre retornos de fundos de investimento incorrem em problemas de viés de sobrevivência.

Bases de dados, normalmente, disponibilizam somente informações sobre fundos em atividade, excluindo fundos que tiveram suas operações descontinuadas (seja por desligamento do fundo ou aquisição por outros fundos). A consequência dessa ação pode ser significativa para estudos de performance, uma vez que considerar a performance somente de fundos hoje operantes pode enviesar a pesquisa, inflando resultados, dado que estas bases estão indiretamente mantendo somente os fundos que tiveram “*fitness*” suficiente para operar durante o período todo da amostra e continuar operante nos dias de hoje.

Outro erro incorrido em estudos de performance é a decisão de acompanhar somente fundos que operaram o período todo da amostra, ignorando fundos criados e descontinuados ao longo do período anualizado. Essa decisão, como a de manter somente fundos chamados “sobreviventes”, pode enviesar a amostra.

Para mitigar esses vieses, fez-se necessário considerar não somente os fundos hoje sobreviventes, mas também fundos criados e que pereceram ao longo do caminho. Conforme explicado na seção de Metodologia, foi criado um índice agregado das performances e alfas gerados pelos fundos de ações e multimercados rebalanceado mensalmente de forma a acompanhar todos os fundos operantes no período, incluindo fundos que em datas posteriores deixaram de existir. O índice acompanhou a performance desses fundos de maneira mensal até o último mês completo de performance registrada; no mês da descontinuação do fundo, caso não houvesse mais retorno gerado durante um mês completo, havia um rebalanceamento do índice agregado de fundos.

Buscou-se auferir o tamanho do viés de sobrevivência incorrido caso o estudo fosse realizado tomando em conta somente os fundos operantes durante o período completo da pesquisa. Para isso, nossa base de fundos completa foi filtrada de forma a segregar esses fundos sobreviventes, suas performances foram calculadas e foi criado um índice agregando a performance média desses fundos. Os resultados foram registrados na tabela 5.

Ano	Alfa Sobreviventes	Alfa Descontinuados	Fundos Sobreviventes	Média Fundos Descontinuados
2003	-7,73%	-8,65%	125	423
2004	12,27%	-0,23%	125	1.694
2005	0,84%	-1,19%	125	2.473
2006	2,48%	1,43%	125	2.543
2007	3,37%	-0,50%	125	3.299
2008	-5,97%	-2,75%	125	4.045
2009	-0,43%	2,17%	125	3.696
2010	-0,36%	1,63%	125	4.020
2011	0,49%	-0,27%	125	4.367
2012	1,50%	3,06%	125	4.466
2013	5,62%	-1,27%	125	4.748
2014	-8,33%	-4,26%	125	4.602
2015	-2,89%	-2,47%	125	4.014
2016	2,79%	-5,62%	125	3.307
2017	0,67%	-2,93%	125	2.888
2018	5,95%	-3,05%	125	2.578
2019	2,93%	-1,48%	125	2.303
2020	4,23%	-2,72%	125	2.199
2021	1,28%	-7,63%	125	2.097
2022	-2,33%	-15,28%	125	1.418
2023	0,18%	-11,13%	125	456
Retorno Acum.	15,38%	-48,53%	-	-

Tabela 5 – Estatísticas de retorno fundos de investimento sobreviventes e fundos descontinuados de 2003 a 2023.

Conforme pode ser observado na Tabela 5, acima, dos 581 fundos operantes em janeiro de 2003, somente 125 ainda estão ativos hoje.

Nos 21 anos analisados, a performance acumulada do índice agregado dos fundos de investimento presentes no ano de 2003 e que sobreviveram o período completo da amostra

foi de 1169,44%, um retorno anual composto de 12,86%. Já a performance acumulada do índice de fundos que inclui tanto os fundos criados e descontinuados ao longo da amostra (a amostra completa de fundos) foi de 847,98% (11,30% a.a.).

Para colocar em perspectiva, R\$1 milhão investidos em 2003 no índice agregado de fundos de investimento de sobreviventes tornar-se-iam cerca de R\$11,7 milhões em 2023, já esse mesmo montante investido no índice agregado de fundos tornar-se-iam cerca de R\$8,5 milhões, uma diferença de 38%, o investidor receberia R\$3,2 milhões a menos.

Ao analisarmos os retornos desses fundos em excesso ao retorno esperado por seu nível de risco, pode-se observar uma imagem ainda mais díspar. O alfa para o grupo de fundos sobreviventes acumulado durante o período analisado foi de 15,38%, taxa anual média composta em excesso ao retorno esperado dado o nível de risco desses fundos de 0,68% a.a.. Já o alfa agregado produzido pela amostra inteira de fundos, acumulado no período foi de -14,44%, uma diferença de performance acumulada de 2.982 bips. Fundos sobreviventes tiveram uma performance em excesso ao nível de risco 142 bips maior por ano.

Portanto, pode-se notar que estudos de performance que consideram somente os fundos que operaram o período completo da amostra introduzem vieses significativos, que amontam para uma performance superestimada por parte dos fundos analisados de 142 bips ao ano, comparado a uma amostra de fundos que contém tanto os fundos sobreviventes, como também fundos criados e fundos descontinuados ao longo do período analisado.

Um outro dado interessante de acompanharmos é a performance do grupo de fundos não-sobreviventes. Durante o período de 2003 a 2023, este grupo apresentou um alfa acumulado de -48,53%, demonstrando um grande desempenho inferior em relação a seu retorno esperado dado o nível de risco incorrido por estes fundos. Comparando com o grupo de sobreviventes, os não sobreviventes demonstraram uma alfa acumulado de cerca de 6.391 bips inferior. Em performances absolutas, o grupo de sobreviventes foi capaz de entregar um retorno agregado acumulado de 1169,44%, enquanto o grupo de sobreviventes retornou somente 445,42%. Com R\$1 milhão investidos em 2003, um investidor nos fundos sobreviventes chegaria em 2023 com R\$11,7 milhões, enquanto o investidor no grupo de fundos não-sobreviventes terminaria 2023 somente com R\$4,5 milhões, valor cerca de 62% menor.

4. Performance de Fundos de Investimento Ativos x Fundos de Investimento Passivos

Agora, buscaremos comparar a performance de fundos de investimento ativos com os fundos passivos.

Como mencionado na seção de Revisão de Literatura, investidores passivos entendem que a tarefa de gerar alfa por parte de fundos ativos é difícil (se não, impossível) principalmente após incumbência de custos de transações e taxas. Estes, portanto, buscam investir em fundos que buscam replicar ao índice de mercado, no entendimento que este tem trajetória ascendente no longo prazo, e minimizando o pagamento de taxas.

Nos Estados Unidos, o fundo de índice mais popular é o o SPDR S&P 500 ETF (SPY), com cerca de U\$407 bilhões em ativos sob gestão²¹.

Estudos de performance nos Estados Unidos demonstram que, em períodos longos de 10 a 15 anos, mais de 90% dos fundos de investimento têm desempenho inferior ao S&P500. Diferentemente de fundos ativos, que utilizam estratégia de *market timing* e *stock picking*, o SPY busca construir sua carteira de forma a replicar o índice S&P 500; sua taxa de administração é de 0,09% a.a. (State Street Global Advisors SPDR).

No Brasil, o iShares Ibovespa Fundo de Índice, cujo *ticker*²² é BOVA11, é um dos fundos de índice mais populares, com cerca de R\$12,6 bilhões sob gestão²³. O fundo busca replicar a carteira do índice Bovespa, índice de referência do mercado, e possui uma taxa de administração de 0,10% (Blackrock). O BOVA11 foi utilizado como o índice passivo de referência neste *paper*.

Seguindo a metodologia utilizada no SPIVA Scorecard (Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G (2023)), buscou-se averiguar a porcentagem de fundos que foram capazes de gerar desempenhos superiores ao índice passivo em períodos de 1, 5, 10 e 20 anos. Vale mencionar, no entanto, que o BOVA11 foi criado somente no fim de novembro de 2008, de forma que informações de performances anteriores a essa data foram extrapoladas descontando a performance mensal do índice Bovespa pela diferença média entre a performance do índice e a performance do fundo de novembro de 2008 até o julho de 2023.

²¹ Informação coletada em 07/11/2023 em <https://www.ssga.com/us/en/intermediary/etfs/funds/spdr-sp-500-etf-trust-spy>.

²² Ticker: Termo utilizado para referenciar o nome abreviado, ou símbolo, de um ativo na bolsa de valores.

²³ Informação coletada em 03/10/2023.

Na tabela 6, podemos observar a proporção de fundos capazes de produzir performances superiores ao índice passivo nas diferentes janelas.

Categoria	1 Ano	5 Anos	10 Anos	20 Anos
Todos os Fundos	18,94%	32,57%	19,84%	17,05%
Fundos de Ações	58,20%	38,41%	20,98%	17,68%
Fundos Multimercados	9,99%	31,49%	19,56%	8,33%

Tabela 6 – Proporção de Fundos com Desempenho Superior ao Fundo de Índice Passivo.

Em uma primeira análise da tabela, nota-se que somente cerca de 18,94% dos fundos de investimento, isto é, fundos multimercados e fundos de ações, apresentaram desempenhos superiores ao fundo de índice, BOVA11, do começo de 2023 até julho de 2023. O grupo composto somente por fundos de ações foram os que, proporcionalmente, melhor performaram, com 58,20% dos fundos apresentando retornos superiores ao fundo passivo, enquanto o grupo de fundos multimercados (maioria e, por isso, reduzem a proporção global) teve somente 9,99% dos participantes superando o BOVA11.

Ao olharmos para uma janela de 5 anos, podemos observar que 32,57% dos fundos de investimentos atingiram performances superiores ao fundo de índice. Novamente, ao fazermos a separação por classe de fundo, vemos que o grupo de fundos de ações performaram proporcionalmente melhor, com 38,41% dos fundos superando o BOVA11 no período, enquanto somente cerca de 31,49% dos fundos multimercados o fizeram. Nota-se que movendo da janela mais curta, de 1 ano, para a janela mais longa de 5 anos, a proporção de fundos de ações capazes de bater o fundo do índice passou de mais da metade para 38,41%, uma queda de cerca de 1.979 bips. Já a proporção de fundos multimercados capazes de bater o fundo de índice aumentou de 9,99% para 31,49%, ganho de 2.150 bips.

Em um período de 10 anos, o número de fundos capazes de obter performances superiores ao fundo passivo cai 1.273 bips, para 19,84%. Nesta janela mais longa, no entanto, observa-se que o grupo de fundos multimercado performa proporcionalmente melhor que o grupo dos fundos de ações durante o período, com cerca de 20,98% dos multimercados apresentando desempenhos superiores ao fundo de índice, versus 19,56% dos fundos de ações.

Por fim, em uma janela de 20 anos, somente 17,05% dos fundos de investimento são capazes de produzirem retornos superiores ao fundo de índice passivo. Ou seja, 82,95%

dos fundos não são capazes de gerar retornos superiores ao fundo passivo em um período de 20 anos. Separando por classe, observa-se que na janela de 20 anos, somente 17,68% dos fundos de ações foram capazes de gerar retornos acima do fundo passivo e somente 8,33% dos fundos multimercados tiveram performances superiores.

Traçando uma comparação com os Estados Unidos, nota-se que o Brasil tem uma maior proporção de fundos com desempenhos absolutos superiores ao índice passivo.

Segundo o relatório (Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G (2023)), em períodos de 5 anos, somente 10,9% dos fundos mútuos foram capazes de gerar retornos superiores ao S&P 500. Em períodos de 10 anos, essa proporção caiu para 9,8% dos fundos e, em 20 anos, somente 6,9% dos fundos produziram retornos absolutos superiores ao S&P. Uma das razões que pode levar a essa diferença na proporção de fundos que produzem retornos superiores ao índice passivo é a maior volatilidade do mercado brasileiro, que pode ser aproveitada por gestores para gerar performances significativas acima do mercado.

Na tabela 7, foram registrados os 5 fundos de ações e 5 fundos multimercados que melhor performaram nas janelas analisadas.

Na figura 4, podemos acompanhar a evolução da proporção de fundos com desempenhos superiores ao fundo passivo nos períodos de 1, 5, 10 e 20 anos.

Observa-se que há grande variação na proporção de fundos capazes de superar o fundo de índice em janelas mais curtas, 1 e 5 anos. A proporção média de fundos que atingiram performances superiores ao BOVA11 em uma janela de 1 ano é de 44,42% e, em 5 anos, 33,81%; o desvio padrão nos períodos foi de 33,65% e 19,46%, respectivamente. O menor número de *data points* em datas mais curtas parece ser a razão da maior variação entre as proporções de fundos registradas na figura 4, dado que quanto menor a janela, mais importante é a performance em um dado ano para o retorno acumulado.

A janela de 10 anos se mostra mais estável, em média, 32,08% dos fundos de investimento mostraram-se capazes de superar o fundo passivo, com desvio padrão de 4,83%. Quando separamos os fundos em classes, observamos resultados similares, em média, 28,77% dos fundos de ações e 31,90% dos fundos multimercados foram capazes de atingir retornos melhores do que o BOVA11.

Novamente realizando uma comparação com o mercado americano, esses resultados comprovam uma maior capacidade de gestores brasileiros de gerar desempenhos superi-

Categoria	1 Ano	5 Anos	10 Anos	20 Anos
Fundos de Ações	Tolda FIA IE	Grou Provence FIA IE	Santander FIA C. Magazine Luiza	Alaska Poland FIA BDR Nível I
	Indie PIPE II FIA	Grou SPO FIA IE	Organon FIC FIA	Dynamo Cougar FIC FIA
	Polo Special Situations V	Guarany FIA IE	FIA IE PGB	Argos FIA IE
	TSO I FIA	Bartelle FIA IE	Bartelle FIA IE	FI Elo Ações IE
	FUNNDG FIA BDR Nível 1	FIA Saline IE	Nova Milano FIA IE	FIA IP Seleção
Fundos Multimercados	Megmo FIM CP	Morro de São Paulo FIM	Equis High FIM CP	Provence Equity FIM IE
	Special Situation I FIM	HM FIM CP IE	Concórdia TI FIM LP	CSHG Rose FIM IE
	Excelsior FIM CP	FI Lameirão Multim. CP	FIM Gold CP IE	FPRV1 Sabiá CP
	ACP Partners FIM	XMZ FIM CP IE	CSHG Roka FIM CP IE	-
	Alamo FIM CP	Prime FIM CP	Klimt FIM CP	-

Tabela 7 – Fundos com Desempenho Superior ao Fundo de Índice Passivo em Janelas de 1, 5, 10 e 20 Anos

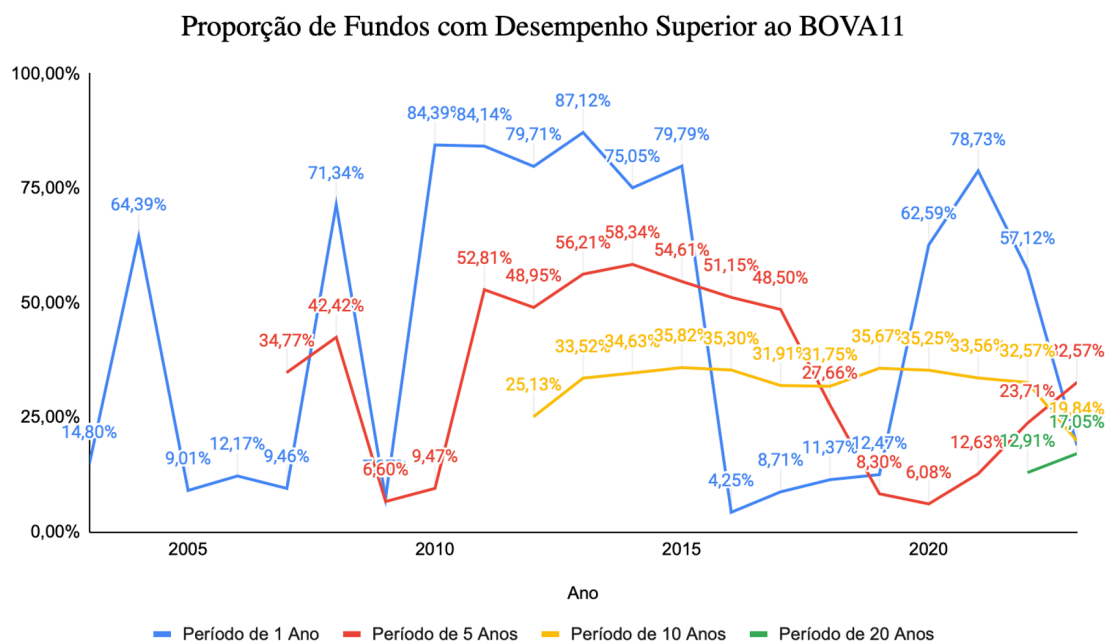


Figura 4 – Proporção de Fundos com Desempenhos Superiores ao Índice BOVA11 por Ano.

ores ao índice. Enquanto somente 9,8% dos gestores ativos americanos são capazes de superar o S&P500, 32% dos gestores brasileiros superam o BOVA11 em um período de 10 anos.

Porém, a conclusão que encontramos olhando para retornos absolutos dos fundos é que, como no estudo do mercado americano, gestores, em média, são incapazes de gerar retornos absolutos superiores ao fundo passivo para seus cotistas. Em média, 67,92% dos gestores não superam o fundo de índice na janela de 10 anos e 85,02% não são capazes de superar o fundo passivo na janela de 20 anos.

Apesar de nossa conclusão para retornos absolutos dos gestores versus o fundo de índice, buscou-se analisar, também, a performance dos gestores brasileiros ajustada ao risco. Para isso, utilizou-se novamente da métrica do alfa. Na tabela 8, foram registrados os resultados.

Podemos observar que, do início de 2023 até julho de 2023, cerca de 39,80% dos fundos de investimento apresentaram retornos ajustados ao risco superiores ao fundo de índice, BOVA11. O grupo formado por fundos de ações performou proporcionalmente melhor que o grupo de fundos multimercados, com 62,64% dos fundos de ações superando o fundo passivo, enquanto somente 34,60% dos fundos multimercados foram capazes de

Categoria	1 Ano	5 Anos	10 Anos	20 Anos
Todos os Fundos	39,80%	37,97%	23,70%	17,23%
Fundos de Ações	62,64%	40,33%	20,23%	17,88%
Fundos Multimercados	34,60%	37,53%	24,54%	8,33%

Tabela 8 – Proporção de Fundos com Desempenho Ajustado ao Risco Superior ao Fundo de Índice Passivo.

fazê-lo.

Nota-se que na janela de 5 anos, a proporção de fundos com retornos ajustados ao risco superiores ao fundo de índice permanece praticamente a mesma, 37,97%. Ao separarmos por classe, vemos que a proporção de fundos de ações com desempenhos superiores ao fundo passivo cai significativamente (2.231 bips), para 40,33% dos fundos, enquanto a proporção de multimercados aumenta 293 bips, para 37,53%.

Na janela de 10 anos, a proporção de fundos capazes de superar o BOVA11 ajustando pelo risco incorrido por estar investido nos fundos cai cerca de 1.427 bips, para 23,70%. Diferentemente de outras janelas, o grupo de fundos multimercados desempenhou proporcionalmente melhor que o grupo de fundos de ações, 24,54% versus 20,23%. Vemos uma queda de 2.010 bips na proporção de fundos de ações que superaram o índice entre as janelas de 5 e 10 anos, e uma queda de 1.299 bips na proporção de fundos multimercados.

Por fim, olhando para o período de 20 anos, a proporção de fundos cujo desempenho foi superior ao BOVA11 foi de 17,23%. Observou-se que 17,88% dos fundos de ações e 8,33% dos fundos multimercados tiveram performances ajustadas ao risco superiores ao fundo passivo.

Observa-se que nas janelas de 1, 5, 10 e 20 anos, a proporção de fundos com retornos superiores ao BOVA11 foi maior considerando retornos ajustados ao risco que a proporção medida por retornos absolutos. Isso parece apontar para uma capacidade dos fundos de apresentarem retornos (maiores, similares, ou menores), porém com menor nível de risco comparado ao fundo passivo.

Na tabela 9, foram registrados os 5 fundos de ações e 5 fundos multimercados que melhor performaram nas janelas analisadas.

Novamente buscamos olhar para a evolução da proporção de fundos com desempenhos superiores ao fundo passivo, compartilhada na figura 5.

Como observado em nossa análise sobre retorno absoluto, há grande variação na

Categoria	1 Ano	5 Anos	10 Anos	20 Anos
Fundos de Ações	Tolda FIA IE	Grou Provence FIA IE	Santander FIA C. Magazine Luiza	Argos FIA IE
	Indie PIPE II FIA	Grou SPO FIA IE	Organon FIC FIA	Dynamo Cougar FIC FIA
	Polo Special Situations V	Guarany FIA IE	Bartelle FIA IE	Alaska Poland FIA BDR Nível I
	TSO I FIA	Bartelle FIA IE	FIA IE PGB	FI Elo Ações IE
	FUNDG FIA BDR Nível 1	Tempranillo FIA IE	Nova Milano FIA IE	FIA IP Seleção
Fundos Multimercados	Megmo FIM CP	Morro de São Paulo FIM	Equis High FIM CP	Provence Equity FIM IE
	Special Situation I FIM	HM FIM CP IE	Concórdia TI FIM LP	CSHG Rose FIM IE
	Excelsior FIM CP	FI Lameirão Multim. CP	FIM Gold CP IE	FPRV1 Sabiá CP
	ACP Partners FIM	XMZ FIM CP IE	CSHG Roka FIM CP IE	-
	Alamo FIM CP	Prime FIM CP	R&C FIM	-

Tabela 9 – Fundos com Desempenho Ajustado ao Risco Superior ao Fundo de Índice Passivo em Janelas de 1, 5, 10 e 20 Anos

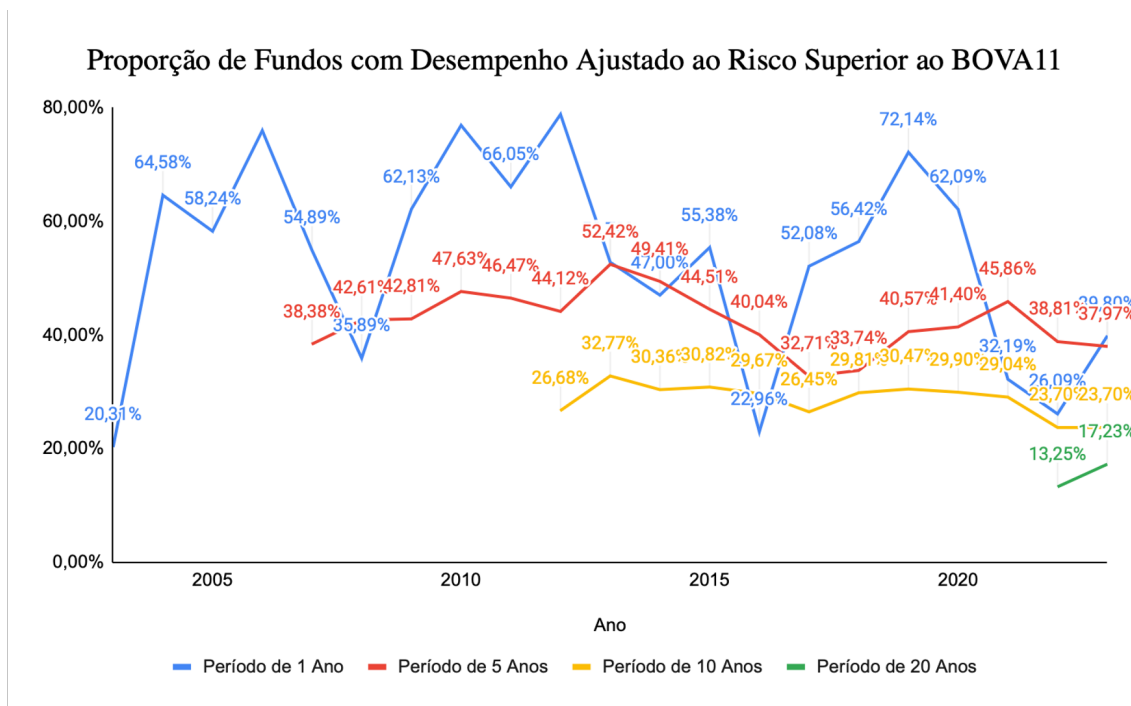


Figura 5 – Proporção de Fundos com Desempenho Ajustado ao Risco Superior ao Índice BOVA11 por Ano.

proporção de fundos capazes de gerar desempenhos superiores ao índice passivo em janelas curtas, de 1 ano. A proporção média de fundos que atingiram performances ajustadas ao risco superiores ao BOVA11 em uma janela de 1 ano foi de 52,98% e o desvio padrão no período foi de 17,72%. Diferentemente da estimativa de 5 anos por retorno absoluto, observa-se que a proporção média de fundos cujo retorno ajustado ao risco conseguiram superar o BOVA11 ficou relativamente estável 42,32%; e 5,25%, respectivamente.

Comparando com as estatísticas de retornos absolutos, a proporção média de fundos com retornos superiores ao BOVA11, ajustados ao risco, foi 856 bips maior para janela de 1 ano e 851 bips maior para a janela de 5 anos. O desvio padrão das estatísticas de retorno ajustado ao risco também foram menores comparadas às de retorno absoluto, com um desvio padrão de 17,72% para a janela de 1 ano (1.593 bips menor) e 5,25% (1.422 bips menor).

A janela de 10 anos, novamente, se mostrou estável, em média, 28,61% dos fundos de investimento mostraram-se capazes de superar o fundo passivo ajustado ao risco, com desvio padrão de 2,86%. Os fundos multimercados performaram, proporcionalmente, um pouco melhor que os fundos de ações, com, em média, 28,32% dos multimercados atin-

gindo retornos ajustados ao risco melhores que o fundo passivo, enquanto 27,08% dos fundos de ações foram capazes de fazê-lo. A proporção de fundos com retornos maiores que o retorno do BOVA11 ajustados ao risco foi, em média, 347 bips menor que a proporção de fundos cujo retorno absoluto superou o fundo passivo, no período.

Na janela de 20 anos, observou-se que, em média, 15,24% dos fundos foram capazes de superar os retornos do BOVA11 ajustados ao risco. A proporção foi cerca de 27 bips maior que a proporção de fundos cujos retornos absolutos superaram o fundo passivo no período.

Podemos notar, portanto, que a proporção média de fundos capazes de superar ao BOVA11 em períodos de 1, 5 e 20 anos foi superior considerando retornos ajustados ao risco, que retornos absolutos. Conclui-se portanto que nessas janelas, fundos ações e multimercados apresentaram, em média, retornos similares ou até um pouco menores, porém os entregaram com menor nível de risco.

Os resultados de (Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G (2023)) apontaram que, nos Estados Unidos, somente 3,57% dos fundos de investimento foram capazes de gerar desempenhos superiores ao S&P500, em um período de 10 anos, e 3,15% em períodos de 20 anos, uma vez feitos ajustes ao risco.

Os resultados obtidos em nossa análise comprovam que, mesmo após ajustarmos os retornos dos fundos de investimento ao risco incorrido por estar investidos nestes ativos, gestores brasileiros demonstraram uma maior capacidade de gerar desempenhos superiores ao índice. Na estatística de 10 anos, a proporção de fundos brasileiros capazes de superar o fundo passivo, ajustado por risco, é cerca de 8x maior que a proporção dos gestores americanos e, em 20 anos, 4,8x maior.

No entanto, chegamos à mesma conclusão atingida pelos estudos do mercado americano, gestores, em média, são incapazes de gerar retornos absolutos ou ajustados ao risco superiores ao fundo passivo para seus cotistas. Em média, 67,9% (medido por retorno absoluto) a 71,4% (medido por retorno ajustado ao risco) dos gestores não produzem desempenhos superiores ao BOVA11 na janela de 10 anos e, 85% (medido por retorno absoluto) a 84,8% (medido por retorno ajustado ao risco) não são capazes de superar o fundo passivo na janela de 20 anos.

Conclusão

As evidências coletadas e apresentadas neste *paper* demonstram que, no Brasil, em média, fundos de investimento não são capazes de gerar desempenhos superiores ao índice Bovespa. Observou-se que, de janeiro de 2003 até julho de 2023, o índice agregado de fundos de investimentos atingiu um retorno absoluto acumulado de 847,98%, uma taxa de retorno anualizada cerca de 81 bips ao ano inferior ao índice, que registrou retorno acumulado de 982,21%. O alfa gerado pelo índice agregado de fundos no período foi negativo em -14,44%, apontando para um desempenho ajustado ao risco da média de fundos 74 bips ao ano inferior ao esperado para ativos com o mesmo nível de risco.

Ao separarmos o índice agregado de fundos de investimento em fundos de ações e fundos multimercados, foi possível vermos que os índices para ambas as classes de fundos apresentaram alfas negativos. Ou seja, gestores ativos independentemente da classe de fundos, em média, não conseguiram gerar resultados superiores ao retorno já esperado pelo risco incorrido nestes ativos. O índice de fundos de ações performou, em média, melhor que o índice multimercado, gerando uma performance absoluta acumulada de 1012,72%, 25.454 bips superior que a performance de 758,18% do índice de multimercados, e um alfa acumulado de -3,69%, cerca de 1.026 bips superior ao alfa de -13,95% dos multimercados.

Notou-se, também, que ao ignorar fundos de investimentos que pereceram antes do término do estudo na análise de performance, assim como desconsiderar fundos criados ao longo do período, superestima o desempenho médio agregado de fundos e introduz viés de sobrevivência no estudo em até 2.982 bips, ou 142 bips ao ano.

Adicionalmente, identificamos uma maior proporção de gestores brasileiros capazes de superar o fundo passivo em janelas de 5 a 20 anos, quando comparado com fundos ativos americanos. Encontrou-se que 32,57% dos gestores ativos foram capazes de obter desempenhos absolutos superiores ao índice passivo, BOVA11, em uma janela de 5 anos, 19,84% em 10 anos e 17,05% em 20 anos, cerca de 2.167 bips, 1.004 bips e 1.015 bips acima da proporção de fundos ativos americanos capazes de superar o S&P.

Por fim, após a realização de uma análise de retorno ajustado ao risco de estar investido nos fundos estudados, encontramos que a proporção de fundos que apresentaram retornos ajustados ao risco superiores que o BOVA11 foi maior que a proporção de fundos com retornos absolutos superiores ao fundo passivo em todas as janelas, apontando para

uma capacidade de fundos de ações e multimercados de entregarem retornos parecidos com o BOVA11, porém com um menor nível de risco. A proporção de fundos de investimento capazes de gerar desempenhos ajustado ao risco superiores ao fundo passivo foi de 37,97% em 5 anos, 23,70% em 10 anos e 17,23% em 20 anos, cerca de 540 bips, 386 bips e 18 bips acima das proporções de fundos cujo retorno absoluto foi superior ao BOVA11. E, comparando com o mercado americano, foi possível identificar que a proporção de gestores brasileiros que performaram melhor que o índice passivo, uma vez ajustado ao risco, foi cerca de 8x maior em 10 anos e 4,8x maior em 20 anos que a proporção de gestores americanos capazes de superar seu índice passivo.

No entanto, ao final, chegamos à mesma conclusão do estudo americano, em média, fundos de investimento não são capazes de gerar retornos superiores ao fundo passivo em janelas mais longas. Cerca de 67,9% (medido por retorno absoluto) a 71,4% (medido por retorno ajustado ao risco) dos gestores não produzem desempenhos superiores ao BOVA11 na janela de 10 anos e, 85% (medido por retorno absoluto) a 84,8% (medido por retorno ajustado ao risco) não são capazes de superar o fundo passivo na janela de 20 anos.

Concluo, traçando possíveis extensões desta análise e endereçando as limitações observadas nesta monografia.

Primeiramente, apesar da extração de dados ser realizada da Comissão de Valores Mobiliários, uma das limitações encontradas no estudo foi o carecimento de uma base completa de fundos de investimento que fosse atualizada retroativamente. A base de dados CVM registra informações cadastrais, valores diários de cota e patrimônio líquido de fundos, no entanto, na ocorrência de modificações das informações cadastrais dos fundos, só é realizada uma atualização retroativa em até 1 ano prévio, de forma que fundos com históricos mais longos não obtêm atualizações cadastrais para anos anteriores a data limite. Isso implica, no caso de alterações na estrutura de fundos, como a “masterização” do fundo (alterar a estrutura do fundo para que este se torne um fundo máster), mudanças de CNPJ e transferências de histórico, que cálculos de performance retroativas não consigam capturar estas mudanças. Em estudos futuros, sugeriria a obtenção de bases adicionais buscando complementar a base de dados da CVM e a realização de atualizações retroativas sobre os dados.

Adicionalmente, estudos podem ser realizados procurando obter uma granularidade ainda maior sobre os dados, explorando as diferenças de performance de fundos de inves-

timento *intra*-classes, como por exemplo comparando performance de fundos de ações que investem somente em ativos locais, com fundos cujo mandato lhes permite explorar outros mercados.

Referências

- ANBIMA. Manual para cadastro de fundos de investimento na ANBIMA. 2022.
- ANBIMA. Consolidado Diário Fundos de Investimento. https://www.anbima.com.br/pt_br/informar/estatisticas/fundos-de-investimento/fi-consolidado-diario.htm, 2023.
- B3. Carteira Teórica do IBovespa - 29/11/23. https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm.
- B3. Índice Bovespa (Ibovespa B3). https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm, n.d.
- Blackrock. iShares Ibovespa Fundo de Índice - BOVA11. <https://www.blackrock.com/br/products/251816/ishares-ibovespa-fundo-de-ndice-fund>.
- Mark M Carhart. On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82, 1997.
- Mark M Carhart, Jennifer N Carpenter, Anthony W Lynch, and David K Musto. Mutual Fund Survivorship. *The Review of Financial Studies*, 15(5):1439–1463, 2002.
- Edwin J Elton, Martin J Gruber, and Christopher R Blake. The Persistence of Risk-adjusted Mutual Fund Performance. *Journal of Business*, pages 133–157, 1996.
- Eugene F Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417, 1970.
- Federal Housing Financing Agency. FANNIE MAE and FREDDIE MAC. <https://www.fhfa.gov/SupervisionRegulation/FannieMaeandFreddieMac>, n.d.
- Thomas Franck. Why Wall Street cares so much about 10-year Treasury rates. <https://www.cnbc.com/2018/10/05/why-wall-street-cares-so-much-about-10-year-treasury-rates.html>, 2018.

-
- Mark Grinblatt and Sheridan Titman. Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings. *Journal of business*, pages 393–416, 1989.
- Darryll Hendricks, Jayendu Patel, and Richard Zeckhauser. Hot Hands in Mutual Funds: the Persistence of Performance, 1974-87. 1990.
- Darryll Hendricks, Jayendu Patel, and Richard Zeckhauser. Hot Hands in Mutual Funds: Short-run Persistence of Relative Performance, 1974–1988. *The Journal of Finance*, 48(1):93–130, 1993.
- Richard A Ippolito. Efficiency with Costly Information: A Study of Mutual Fund Performance, 1965–1984. *The Quarterly Journal of Economics*, 104(1):1–23, 1989.
- Richard A Ippolito. On Studies of Mutual Fund Performance, 1962–1991. *Financial Analysts Journal*, 49(1):42–50, 1993.
- Michael C Jensen. The Performance of Mutual Funds in the Period 1945-1964. *The Journal of finance*, 23(2):389–416, 1968.
- Kahneman, Daniel and Tversky, Amos. Prospect Theory: An Analysis of Decision Under Risk. pages 99–127, 1979.
- Kenton W. What Is the Capital Asset Pricing Model (CAPM)? www.investopedia.com/terms/c/capm.asp, 2023.
- Lazzara C, and Ganti A, and Di Gioia D, and Stoddart G. SPIVA® U.S. Scorecard. 2023.
- Grinblatt M and Titman S. The Persistence of Mutual Fund Performance. *The Journal of Finance*, 47(5):1977–1984, 1992.
- Burton G Malkiel. Returns from Investing in Equity Mutual Funds 1971 to 1991. *The Journal of Finance*, 50(2):549–572, 1995.
- William F Sharpe. Mutual Fund Performance. *The Journal of business*, 39(1):119–138, 1966.
- State Street Global Advisors SPDR. SPDR® SP 500® ETF Trust. <https://www.ssga.com/us/en/intermediary/etfs/funds/spdr-sp-500-etf-trust-spy>.

Tesouro Direto. Cálculo da Rentabilidade dos Títulos Públicos ofertados no Tesouro Direto □ Tesouro Selic (LFT). https://www.tesourodireto.com.br/data/files/91/60/D8/C9/73B2D610393A62D6894D49A8/Calculo_Rentabilidade_TesouroSelic.pdf.

Tesouro Nacional. Classificação de Risco. <https://www.gov.br/tesouronacional/pt-br/divida-publica-federal/sobre-a-divida-publica/classificacao-de-risco-da-republica-soberana-do-brasil>, 2023.

Russ Wermers. Mutual Fund performance: An Empirical Decomposition into Stock-picking Talent, Style, Transactions Costs, and Expenses. *The Journal of Finance*, 55 (4):1655–1695, 2000.

Kent L Womack and Ying Zhang. Understanding Risk and Return, the CAPM, and the Fama-French Three-factor Model. *Available at SSRN 481881*, 2003.

Apêndice: Código para Confecção, Tratamento e Análise da Base de Dados

```
##### DETERMINACAO DO DIRETÓRIO #####
```

```
setwd("~/Documents/Monografia/inf_diario_pasta")
```

```
##### PACOTE DE FUNCOES #####
```

```
library(data.table)
```

```
library(plyr)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(tidyr)
```

```
library(readr)
```

```
library(usethis)
```

```
library(tidyverse)
```

```
library(readxl)
```

```
library(openxlsx)
```

```
library(fuzzyjoin)
```

```
library(stringi)
```

```
library(scales)
```

```
library(lubridate)
```

```
library(zoo)
```

```
library/slider)
```

```
##### IMPORTACAO DE ARQUIVOS #####
```

```
arquivos =
```

```
  list.files("~/Documents/Monografia/inf_diario_pasta/",
```

```
pattern = "*.csv")

nomes=read_delim("~/Documents/Monografia/cad_fi-7.csv",
  delim = ";", escape_double = FALSE, trim_ws = TRUE)

lista_df = list()

colunas_a_serem_mantidas = c("CNPJ_FUNDO", "DT_COMPTC",
  "VL_QUOTA", "VL_PATRIM_LIQ")

for (file in arquivos) {
  tryCatch({
    dt_temp = fread(file, header = TRUE, sep = ";")
    colunas_a_serem_mantidas =
      intersect(colunas_a_serem_mantidas, names(dt_temp))
    if (length(colunas_a_serem_mantidas) > 0) {
      dt_temp = dt_temp[, names(dt_temp) %in%
        colunas_a_serem_mantidas, with = FALSE]
    }
    lista_df[[file]] = as.data.frame(dt_temp)
  }, error = function(e) {
    cat("Error reading file:", file, "\n")
  })
}

dataframe_inf_diaria = bind_rows(lista_df)

dataframe_nomes = nomes[,c(2:3,13)]

##### CONSOLIDACAO DE BASE E TRATAMENTO DE DADOS #####

dataframe_inf_diaria_e_nomes =
```



```
left_join(dataframe_inf_diaria , dataframe_nomes ,
by="CNPJ_FUNDO")

dataframe_inf_diaria_e_nomes =
  dataframe_inf_diaria_e_nomes%>%distinct(CNPJ_FUNDO,
DT_COMPTC, .keep_all = TRUE)

fundos_com_NA = dataframe_inf_diaria_e_nomes%>%
filter(is.na(DENOM_SOCIAL))

fundos_unicos = fundos_com_NA%>%distinct(CNPJ_FUNDO)

resolvido_NAs=read_delim("~/Documents/Monografia /
NA_DENOM_SOCIAL_Solved_Fundos.csv", delim = ";",
escape_double = FALSE, trim_ws = TRUE)

base_de_fundos_NA_resolvido =
  left_join(dataframe_inf_diaria_e_nomes, resolvido_NAs,
by= "CNPJ_FUNDO")

base_de_fundos_NA_resolvido =
  base_de_fundos_NA_resolvido%>%distinct(CNPJ_FUNDO,
DT_COMPTC, .keep_all = TRUE)

base_de_fundos_NA_resolvido$DENOM_SOCIAL =
  coalesce(base_de_fundos_NA_resolvido$DENOM_SOCIAL.x,
base_de_fundos_NA_resolvido$DENOM_SOCIAL.y)

base_de_fundos_NA_resolvido$CLASSE =
  coalesce(base_de_fundos_NA_resolvido$CLASSE.x,
base_de_fundos_NA_resolvido$CLASSE.y)
```

```
base_de_fundos_inf_diaria_e_cadastral =
  subset(base_de_fundos_NA_resolvido, select =
    -c(CLASSE.x, CLASSE.y, DENOM_SOCIAL.x, DENOM_SOCIAL.y))

##### IMPORTACAO INFORMACOES CADASTRAIS E JUNCAO A BASE
#####

setwd("~/Documents/Monografia/extrato_informacoes")

arquivos_dos_extratos =
  list.files("~/Documents/Monografia/extrato_informacoes/",
    pattern = "*.csv")

colunas_a_serem_mantidas_2 = c("CNPJ_FUNDO",
  "DENOM_SOCIAL", "FUNDO_ESPELHO")

lista_df_2 = list()

for (file in arquivos_dos_extratos) {
  tryCatch({
    dt_temp_2 = fread(file, header = TRUE, sep = ";",
      encoding = "UTF-8")
    colunas_a_serem_mantidas_2 =
      intersect(colunas_a_serem_mantidas_2,
        names(dt_temp_2))
    if (length(colunas_a_serem_mantidas_2) > 0) {
      dt_temp_2 = dt_temp_2[, names(dt_temp_2) %in%
        colunas_a_serem_mantidas_2, with = FALSE]
    }
    lista_df_2[[file]] = as.data.frame(dt_temp_2)
  }, error = function(e) {
    cat("Error reading file:", file, "\n")
  })
}
```

```
    })
  }

dataframe_extrato = bind_rows(lista_df_2)

dataframe_fundos =
  left_join(base_de_fundos_inf_diaria_e_cadastral ,
    dataframe_extrato , by = "CNPJ_FUNDO")

dataframe_agregado = dataframe_fundos

##### TRATAMENTO DE DADOS #####

dataframe_agregado =
  dataframe_agregado %>% distinct(CNPJ_FUNDO, DT_COMPTC,
    FUNDO_ESPELHO, .keep_all = TRUE)

dataframe_agregado$DENOM_SOCIAL =
  coalesce(dataframe_agregado$DENOM_SOCIAL.x,
    dataframe_agregado$DENOM_SOCIAL.y)

dataframe_agregado = subset(dataframe_agregado , select =
  -c(DENOM_SOCIAL.x, DENOM_SOCIAL.y))

base_de_fundos = dataframe_agregado %>%
  distinct(CNPJ_FUNDO, DT_COMPTC, .keep_all = TRUE)

base_de_fundos_adj = base_de_fundos %>%
  mutate(DENOM_SOCIAL_1 = iconv(DENOM_SOCIAL, to =
    "UTF-8", from = "ISO-8859-1", sub = "byte"))

base_de_fundos_adj = base_de_fundos_adj %>%
```

```
mutate(CLASSE_1 = iconv(CLASSE, to = "UTF-8", from =
"ISO-8859-1", sub = "byte"))
```

```
base_de_fundos_adj = subset(base_de_fundos_adj, select =
-c(DENOM_SOCIAL, CLASSE))
```

```
colnames(base_de_fundos_adj) = c("CNPJ_FUNDO",
"DT_COMPTC", "VL_QUOTA", "VL_PATRIM_LIQ",
"FUNDO_ESPELHO", "DENOM_SOCIAL", "CLASSE")
```

```
fundos_com_NA =
base_de_fundos_adj%>%filter(is.na(DENOM_SOCIAL))
```

```
fundos_unicos = fundos_com_NA%>%distinct(CNPJ_FUNDO)
```

```
fundos_unicos$DENOM_SOCIAL = c("Gas_Mix_Fundo_de_
Investimento_em_Cotas_de_Fundo_de_Investimento_em_
Ações", "ANAS_II_FIC_FIM_CP_IE", "ITAÚ_HEDGE_PLUS_
OSTRUM_FIC_FIM")
```

```
fundos_unicos$CLASSE = c("Fundo_de_Ações", "Fundo_
Multimercado", "Fundo_Multimercado")
```

```
base_de_fundos_combinada = left_join(base_de_fundos_adj,
fundos_unicos, by= "CNPJ_FUNDO")
```

```
base_de_fundos_combinada$DENOM_SOCIAL =
coalesce(base_de_fundos_combinada$DENOM_SOCIAL.x,
base_de_fundos_combinada$DENOM_SOCIAL.y)
```

```
base_de_fundos_combinada$CLASSE =
coalesce(base_de_fundos_combinada$CLASSE.x,
```

```

base_de_fundos_combinada$CLASSE.y)

base_de_fundos_combinada =
  subset(base_de_fundos_combinada, select =
    -c(DENOM_SOCIAL.x, DENOM_SOCIAL.y, CLASSE.x, CLASSE.y))

classeNAs=base_de_fundos_combinada%>%filter(is.na(CLASSE))

classeNAs=classeNAs%>%filter(!duplicated(CNPJ_FUNDO))

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FIA",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo de
  Ações"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("AÇÕES",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo de
  Ações"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("ACOES",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo de
  Ações"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FMIA",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo de
  Ações"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) &
  grepl("A\xc7\xd5ES", classeNAs$DENOM_SOCIAL,
  ignore.case = TRUE)]="Fundo de Ações"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) &
  grepl("MULTIMERCADO", classeNAs$DENOM_SOCIAL,
  ignore.case = TRUE)]="Fundo Multimercado"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("RENDA
  FIXA", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundo de Renda Fixa"

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FIM",

```

```
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
grepl("CAMBIAL", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="Fundo_Cambial"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
grepl("REFERENCIADO", classeNAs$DENOM_SOCIAL,
ignore.case = TRUE)]="Fundo_Referenciado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
grepl("INCENTIVADO", classeNAs$DENOM_SOCIAL,
ignore.case = TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("IAÁes",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
Ações"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("MUTUO_DE_
PRIV", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="FMP-FGTS"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_RENDA_
", classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
de_Renda_Fixa"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_CREDITO_
", classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CRÉD",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_FGTS_",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="FMP-FGTS"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_FIM",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_RF_",
```

```

    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
    Renda_Fixa"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CRED_
    PRIV", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CREDITO_
    PRIVADO", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
    grepl("CR\`c9DITO_PRIVADO", classeNAs$DENOM_SOCIAL,
    ignore.case = TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CURTO_
    PRAZO", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_de_Renda_Fixa"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
    grepl("EXTERIOR", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_FIXA",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
    Renda_Fixa"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FIXED_
    INCOME", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_de_Renda_Fixa"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CR\`c9D_
    PRIV", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("...CRD_
    PRIV", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CARTEIRA_
    ATIVA", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_de_Ações"

```

```

classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("DI",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  Referenciado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("CREDIT",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("ETF",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  de
  Ações"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("S&P",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  de
  Ações"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FITVM",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  de
  Ações"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("HEDGE",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  Multimercado"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FUNDO
  DE
  INVESTIMENTO
  FIN", classeNAs$DENOM_SOCIAL, ignore.case
  = TRUE)]="Fundo
  de
  Investimento
  Financeiro"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FIF",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  de
  Investimento
  Financeiro"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
  grepl("APOSENTADORIA", classeNAs$DENOM_SOCIAL,
  ignore.case = TRUE)]="Fundo
  de
  Previdenciário"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FAPI",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
  de
  Previdenciário"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("
  FINANCEIRO", classeNAs$DENOM_SOCIAL, ignore.case =

```



```

TRUE)]="Fundo_de_Investimento_Financeiro"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("DI",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_DI"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("_
  FMP-FGTS_", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="FMP-FGTS"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDO_DE_
  INVESTIMENTO", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDO_DE_
  INVEST", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) &
  grepl("FINANCEIRO", classeNAs$DENOM_SOCIAL, ignore.case
  = TRUE)]="Fundos_de_Investimento_Financeiro"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDOS_
  DE_INVESTIMENTO", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("MULT",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_
  Multimercado"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDOS_
  DE_INVEST", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDOS_
  INV", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FDOS_DE_
  INVEST", classeNAs$DENOM_SOCIAL, ignore.case =
  TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FDO_DE_
  INVEST", classeNAs$DENOM_SOCIAL, ignore.case =

```

```

TRUE)]="Fundos de Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FAQ DE
FI", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="Fundos de Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FAQFI",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos de
Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FACFI",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos de
Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FDOS DE
INVEST.", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="Fundos de Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDOS
DE INV", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="Fundos de Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("QUOTAS
DE FI", classeNAs$DENOM_SOCIAL, ignore.case =
TRUE)]="Fundos de Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FICFIM",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
Multimercado"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("MACRO",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
Multimercado"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("MM",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo
Multimercado"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("MIA",
classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo de
Ações"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("ÍNDICE",

```

```

    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
    Índice"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("INDICE",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
    Índice"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
    grepl("IBOVESPA", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundo_de_Índice"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FINAN",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos_de_
    Investimento_Financeiro"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FAPI",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundo_de_
    Previdenciário"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FIF",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos_de_
    Investimento_Financeiro"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FDO_DE_
    INV", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FDOS_DE_
    INV", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) &
    grepl("FDO.DE.INV", classeNAs$DENOM_SOCIAL, ignore.case
    = TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("FUNDO_
    INVESTIMENTO", classeNAs$DENOM_SOCIAL, ignore.case =
    TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[ is.na(classeNAs$CLASSE) & grepl("_FI_",
    classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos_de_
    Investimento"

```

```

classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("FUNDO_DE_
  APLICACAO_DE_QUOTAS", classeNAs$DENOM_SOCIAL,
  ignore.case = TRUE)]="Fundos_de_Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE) & grepl("QUOTAS",
  classeNAs$DENOM_SOCIAL, ignore.case = TRUE)]="Fundos_de_
  Investimento"
classeNAs$CLASSE[is.na(classeNAs$CLASSE)]="Fundos_de_
  Investimento"

checkNA=classeNAs%>%filter(is.na(CLASSE))

checkNA=checkNA%>%filter(!duplicated(CNPJ_FUNDO))

filtro_classe_NAs=classeNAs[,c(1,7)]

base_de_fundos_classe =
  left_join(base_de_fundos_combinada, filtro_classe_NAs,
  by = "CNPJ_FUNDO")

base_de_fundos_classe$CLASSE =
  coalesce(base_de_fundos_classe$CLASSE.x,
  base_de_fundos_classe$CLASSE.y)

base_de_fundos_classe = subset(base_de_fundos_classe,
  select = -c(CLASSE.x, CLASSE.y))

fwrite(base_de_fundos_classe,
  "~/Documents/Monografia/Base_fundos_classe.csv", sep =
  ";")

base_de_fundos_classe=read_delim("~/Documents/Monografia/
  Base_fundos_classe.csv", delim = ";", escape_double =

```

```

FALSE, trim_ws = TRUE)

fundos_de_invest = base_de_fundos_classe%>% filter(CLASSE
  == "Fundos_de_Investimento")
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "Fundos_de_Investimento" & grepl("AÇOES",
  base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
  TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "Fundos_de_Investimento" & grepl("ACOES",
  base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
  TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "Fundos_de_Investimento" & grepl("MULTIM",
  base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
  TRUE)] = "Fundo_Multimercado"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "Fundos_de_Investimento" & grepl("LONG",
  base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
  TRUE)] = "Fundo_de_Ações"
fundos_de_invest = base_de_fundos_classe%>% filter(CLASSE
  == "Fundos_de_Investimento") %>%
  filter(!duplicated(CNPJ_FUNDO))
fundos_de_invest = base_de_fundos_classe%>% filter(CLASSE
  == "0") %>% filter(!duplicated(CNPJ_FUNDO))

base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "0" & grepl("FINANC",
  base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
  TRUE)] = "Fundo_de_Investimento_Financeiro"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
  == "0" & grepl("ACOES",

```

```

base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("AÇÕES",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("FICFIA",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("FICFITVM",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("FIQFITVM",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Ações"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("FINAN",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Investimento_Financeiro"
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0" & grepl("FIF",
base_de_fundos_classe$DENOM_SOCIAL, ignore.case =
TRUE)] = "Fundo_de_Investimento_Financeiro"
fundos_de_invest = base_de_fundos_classe%>% filter(CLASSE
== "0") %>% filter(!duplicated(CNPJ_FUNDO))
base_de_fundos_classe$CLASSE[base_de_fundos_classe$CLASSE
== "0"] = "Fundo_de_Investimento"
unique(base_de_fundos_classe$CLASSE)

```

```
##### SEPARACAO DE FUNDOS PARA REMOCAO DE FUNDOS QUE NAO
CLASSIFICAM COMO FUNDOS DE ACOES E FUNDOS MULTIMERCADOS
#####
```

```
fundos_acoes_e_mult =
  base_de_fundos_classe%>%filter(CLASSE == "Fundo_
Multimercado" | CLASSE == "Fundo_de_Ações")

base_de_fundos_unicos =
  fundos_acoes_e_mult%>%filter(!duplicated(CNPJ_FUNDO))

### SEPARACAO DE FUNDOS PARA REMOCAO DE FUNDOS ESPELHOS E
FUNDOS MASTER ##

fundos_acoes_e_mult_FICs = fundos_acoes_e_mult %>%
  filter(!across(DENOM_SOCIAL, ~grepl('MASTER|MÁSTER', .,
ignore.case = TRUE)))

fundos_acoes_e_mult_FICs_sem_espelho =
  fundos_acoes_e_mult_FICs%>%filter(FUNDO_ESPELHO != "S"
| is.na(FUNDO_ESPELHO))

base_de_fundos_FICs_unicos =
  fundos_acoes_e_mult_FICs_sem_espelho%>%
  filter(!duplicated(CNPJ_FUNDO))

fwrite(fundos_acoes_e_mult_FICs_sem_espelho ,
  "~/Documents/Monografia/Base_Filtrada_Classe.csv", sep
= ";")

fundos_acoes_e_mult_FICs_sem_espelho=read_delim("~/Docu-
ments/Monografia/Base_Filtrada_Classe.csv",
```



```
limite_inferior = quantile(performance_fundos$PERF_MONTH,  
  percentil_limite_inferior)
```

```
performance_fundos_ajustada = performance_fundos %>%  
  filter(PERF_MONTH >= limite_inferior , PERF_MONTH <=  
  limite_superior)
```

```
##### ADICAO RETORNO DO MERCADO #####
```

```
ibovespa=read_excel("~/Documents/Monografia/Ibovespa.xlsx")
```

```
ibovespa$YEAR_MONTH = ibovespa$Data
```

```
ibovespa = ibovespa %>%  
  mutate(Data = sub("Jan", "01", Data),  
    Data = sub("Fev", "02", Data),  
    Data = sub("Mar", "03", Data),  
    Data = sub("Abr", "04", Data),  
    Data = sub("Mai", "05", Data),  
    Data = sub("Jun", "06", Data),  
    Data = sub("Jul", "07", Data),  
    Data = sub("Ago", "08", Data),  
    Data = sub("Set", "09", Data),  
    Data = sub("Out", "10", Data),  
    Data = sub("Nov", "11", Data),  
    Data = sub("Dez", "12", Data))
```

```
ibovespa = ibovespa %>% mutate(MES = substr(Data, 1, 2))
```

```
ibovespa = ibovespa %>% mutate(ANO = substr(Data, 4, 7))
```

```

ibovespa = ibovespa %>% mutate(YEAR_MONTH = paste(ANO,
  MES, sep = "-")) %>% select(YEAR_MONTH , Fechamento)

ibovespa = ibovespa %>% mutate(MKT_RETURN = Fechamento /
  lag(Fechamento) -1)

performance_fundos_e_mercado =
  left_join(performance_fundos_ajustada , ibovespa , by =
  "YEAR_MONTH")

##### CALCULO DO BETA #####

performance_fundos_e_mercado = performance_fun-
dos_e_mercado %>% group_by(CNPJ_FUNDO) %>%
  mutate(BETA_FUNDO = cov(PERF_MONTH,
  MKT_RETURN) / var(MKT_RETURN))

checkNA = performance_fundos_e_mercado %>%
  filter(is.na(BETA_FUNDO)) %>% select(CNPJ_FUNDO)

beta_calc = performance_fundos_e_mercado[performance_fun-
dos_e_mercado$CNPJ_FUNDO %in% checkNA$CNPJ_FUNDO,]

performance_fundos_e_mercado = performance_fun-
dos_e_mercado[!(performance_fundos_e_mercado$CNPJ_FUNDO
  %in% checkNA$CNPJ_FUNDO) ,]

##### ADICAO RISK FREE RATE #####

LFTs_Performance=read_excel("~/Documents/Monografia/
  COMBINED_LFTs_Performance_Mensal.xlsx")

```

```
LFTs_Performance = LFTs_Performance %>% mutate(MES =
  substr(Dia, 6, 7),
  YEAR_MONTH = paste(ANO, MES, sep = "-")) %>%
  select(YEAR_MONTH, 'Performance Mensal')

performance_fundos_e_mercado_e_risk_free =
  left_join(performance_fundos_e_mercado,
  LFTs_Performance, by = "YEAR_MONTH")

checkNA = performance_fundos_e_mercado_e_risk_free %>%
  filter(is.na('Performance Mensal'))

##### TRATAMENTO BASE COMPLETA #####

base_completa = performance_fundos_e_mercado_e_risk_free
%>%
  group_by(CNPJ_FUNDO) %>%
  mutate(AVG_SIZE = mean(VL_PATRIM_LIQ))%>%
  select(YEAR_MONTH, CNPJ_FUNDO, DENOM_SOCIAL, CLASSE,
  FUNDO_ESPELHO, VL_QUOTA, VL_PATRIM_LIQ, AVG_SIZE,
  PERF_MONTH, MKT_RETURN, BETA_FUNDO, 'Performance
  Mensal')

colnames(base_completa) = c("YEAR_MONTH", "CNPJ_FUNDO",
  "DENOM_SOCIAL", "CLASSE", "FUNDO_ESPELHO", "VL_QUOTA",
  "VL_PATRIM_LIQ", "AVG_SIZE", "PERF_MONTH", "MKT_PERF",
  "BETA_FUNDO", "RF_PERF")

##### CALCULO RETORNO ESPERADO E ALFA #####

base_completa = base_completa %>% mutate(RETORNO_ESP =
  BETA_FUNDO*(MKT_PERF - RF_PERF) + RF_PERF,
```

```
ALFA = PERF_MONTH - RETORNO_ESP)

checkNA = base_completa %>% filter(is.na(ALFA))

base_unicos = base_completa %>%
  filter(!duplicated(CNPJ_FUNDO))

base_unicos_acoes = base_completa %>% filter(CLASSE ==
  "Fundo_de_Ações") %>% filter(!duplicated(CNPJ_FUNDO))

base_unicos_multi = base_completa %>% filter(CLASSE ==
  "Fundo_Multimercado") %>%
  filter(!duplicated(CNPJ_FUNDO))

base_completa = base_completa %>% group_by(CNPJ_FUNDO,
  YEAR_MONTH) %>% distinct()

fwrite(base_completa ,
  "~/Documents/Monografia/Base_Completa_Final.csv", sep =
  ";")

##### CONFECCAO TABELA DE PERFORMANCE FUNDOS DE
  INVESTIMENTO #####

base_completa = base_completa %>%
  mutate_all(str_squish)

base_completa = subset(base_completa ,
  !duplicated(base_completa[c("CNPJ_FUNDO",
  "YEAR_MONTH")]))

base_completa = base_completa %>%
```

```
mutate(across(c("VL_QUOTA", "VL_PATRIM_LIQ", "AVG_SIZE",
               "PERF_MONTH", "MKT_PERF", "BETA_FUNDO", "RF_PERF",
               "RETORNO_ESP", "ALFA"), as.numeric))
```

```
base_completa_performance = base_completa %>% mutate(YEAR
  = substr(YEAR_MONTH, 1, 4)) %>%
  group_by(YEAR_MONTH) %>%
  mutate(NUM_FUNDOS = n_distinct(CNPJ_FUNDO),
         PERFORMANCE_MEDIA = mean(PERF_MONTH),
         ALFA_MEDIO = mean(ALFA))%>%
  ungroup()
```

```
indice_anual_base_completa = base_completa_performance%>%
  filter(!duplicated(YEAR_MONTH))%>%
  group_by(YEAR) %>%
  mutate(PERFORMANCE_ANUAL_MEDIA = prod(1 +
    PERFORMANCE_MEDIA) -1,
         ALFA_ANUAL_MEDIO = prod(1 + ALFA_MEDIO) -1,
         MKT_PERF_ANUAL = prod(1+MKT_PERF) -1,
         NUM_FUNDOS_MEDIO = mean(NUM_FUNDOS))%>%
  filter(!duplicated(YEAR)) %>%
  select(YEAR, PERFORMANCE_ANUAL_MEDIA, ALFA_ANUAL_MEDIO,
         MKT_PERF_ANUAL, NUM_FUNDOS_MEDIO)
```

```
performance_fundos_acumulada =
  prod(indice_anual_base_completa$PERFORMANCE_ANUAL_MEDIA
  +1) -1
print(performance_fundos_acumulada)
```

```
alfa_fundos_acumulado =
  prod(indice_anual_base_completa$ALFA_ANUAL_MEDIO +1) -1
print(alfa_fundos_acumulado)
```

```

mkt_performance_acumulada =
  prod(indice_anual_base_completa$MKT_PERF_ANUAL +1) -1
print(mkt_performance_acumulada)

unicos = base_completa %>% distinct(CNPJ_FUNDO)

unicos_acoes = base_completa %>% filter(CLASSE == "Fundo_
de_Ações") %>% distinct(CNPJ_FUNDO)

unicos_multim = base_completa %>% filter(CLASSE != "Fundo_
de_Ações") %>% distinct(CNPJ_FUNDO)

##### INDICE SEPARADO - ACOES #####

base_acoes = base_completa %>% filter(CLASSE == "Fundo_
de_Ações") %>%
mutate(YEAR = substr(YEAR_MONTH, 1, 4)) %>%
group_by(YEAR_MONTH) %>%
mutate(NUM_FUNDOS = n_distinct(CNPJ_FUNDO),
PERFORMANCE_MEDIA = mean(PERF_MONTH),
ALFA_MEDIO = mean(ALFA))%>%
ungroup()

indice_anual_acoes = base_acoes%>%
filter(!duplicated(YEAR_MONTH))%>%
group_by(YEAR) %>%
mutate(PERFORMANCE_ANUAL_MEDIA = prod(1 +
PERFORMANCE_MEDIA) -1,
ALFA_ANUAL_MEDIO = prod(1 + ALFA_MEDIO) -1,
MKT_PERF_ANUAL = prod(1+MKT_PERF) -1,
NUM_FUNDOS_MEDIO = mean(NUM_FUNDOS))%>%

```

```

filter(!duplicated(YEAR)) %>%
select(YEAR, PERFORMANCE_ANUAL_MEDIA, ALFA_ANUAL_MEDIO,
       MKT_PERF_ANUAL, NUM_FUNDOS_MEDIO)

acoes_performance_acumulada =
  prod(indice_anual_acoes$PERFORMANCE_ANUAL_MEDIA +1) -1
print(acoes_performance_acumulada)

acoes_alfa_acumulado =
  prod(indice_anual_acoes$ALFA_ANUAL_MEDIO +1) -1
print(acoes_alfa_acumulado)

mkt_performance_acumulada =
  prod(indice_anual_acoes$MKT_PERF_ANUAL +1) -1
print(mkt_performance_acumulada)

aum_acoes = base_acoes %>% group_by(CNPJ_FUNDO) %>%
  mutate(AUM_MEDIO = mean(VL_PATRIM_LIQ)) %>%
  ungroup() %>%
  filter(!duplicated(CNPJ_FUNDO)) %>%
  mutate(AUM_TOTAL_MEDIO = mean(AUM_MEDIO))

##### INDICE SEPARADO – MULTIMERCADOS #####

base_multimercados = base_completa %>% filter(CLASSE ==
  "Fundo_Multimercado") %>%
  mutate(YEAR = substr(YEAR_MONTH, 1, 4)) %>%
  group_by(YEAR_MONTH) %>%
  mutate(NUM_FUNDOS = n_distinct(CNPJ_FUNDO),
         PERFORMANCE_MEDIA = mean(PERF_MONIH),
         ALFA_MEDIO = mean(ALFA))%>%
  ungroup()

```

```

indice_anual_multimercado = base_multimercados %>%
  filter(!duplicated(YEAR_MONTH)) %>%
  group_by(YEAR) %>%
  mutate(PERFORMANCE_ANUAL_MEDIO = prod(1 +
    PERFORMANCE_MEDIO) -1,
    ALFA_ANUAL_MEDIO = prod(1 + ALFA_MEDIO) -1,
    MKT_PERF_ANUAL = prod(1+MKT_PERF) -1,
    NUM_FUNDOS_MEDIO = mean(NUM_FUNDOS)) %>%
  filter(!duplicated(YEAR)) %>%
  select(YEAR, PERFORMANCE_ANUAL_MEDIO, ALFA_ANUAL_MEDIO,
    MKT_PERF_ANUAL, NUM_FUNDOS_MEDIO)

```

```

multimercado_performance_acumulada =
  prod(indice_anual_multimercado$PERFORMANCE_ANUAL_MEDIO
    +1) -1
print(multimercado_performance_acumulada)

```

```

multimercado_alfa_acumulado =
  prod(indice_anual_multimercado$ALFA_ANUAL_MEDIO +1) -1
print(multimercado_alfa_acumulado)

```

```

mkt_performance_acumulada =
  prod(indice_anual_multimercado$MKT_PERF_ANUAL +1) -1
print(mkt_performance_acumulada)

```

```

aum_multim = base_multimercados %>% group_by(CNPJ_FUNDO)
  %>%
  mutate(AUM_MEDIO = mean(VL_PATRIM_LIQ)) %>%
  ungroup() %>%
  filter(!duplicated(CNPJ_FUNDO)) %>%
  mutate(AUM_TOTAL_MEDIO = mean(AUM_MEDIO))

```



```
##### TABELA ACOES E MULTIMERCADOS #####
```

```
tabela_indices = left_join(indice_anual_acoes ,
  indice_anual_multimercado , by = "YEAR")

colnames(tabela_indices) = c("ANO", "PERF_ACOES",
  "ALFA_ACOES", "PERF_IBOV", "NUM_FUND_MEDIO_ACOES",
  "PERF_MULT", "ALFA_MULT", "PERF_IBOV2",
  "NUM_FUNDS_MEDIO_MULT")

tabela_indices = tabela_indices %>% select(ANO,
  PERF_ACOES, PERF_MULT, PERF_IBOV, ALFA_ACOES, ALFA_MULT)

fwrite(tabela_indices ,
  "~/Documents/Monografia/Tabela_Indices.csv", sep = ";")

base_completa = base_completa %>%
  mutate(across(c("VL_QUOTA", "VL_PATRIM_LIQ", "AVG_SIZE",
    "PERF_MONIH", "MKT_PERF", "BETA_FUNDO", "RF_PERF",
    "RETORNO_ESP", "ALFA"), as.numeric))
```

```
##### TABELA DE ENTRADAS E SAIDAS #####
```

```
primeira_vista_fundos = base_completa %>%
  group_by(CNPJ_FUNDO)%>%
  filter(YEAR_MONTH == min(YEAR_MONTH))%>%
  group_by(YEAR_MONTH) %>%
  summarise(Count = n())

ultima_vista_fundos = base_completa %>%
  group_by(CNPJ_FUNDO)%>%
```

```

    filter(YEAR_MONTH == max(YEAR_MONTH))%>%
    group_by(YEAR_MONTH)%>%
    summarise(Count = n())

entradas_e_saidas = merge(primeira_vista_fundos ,
    ultima_vista_fundos , by = "YEAR_MONTH" , all = TRUE)

indice_estatisticas = base_completa %>%
    mutate(YEAR=substr(YEAR_MONTH, 1, 4)) %>%
    group_by(YEAR)%>%
    mutate(NUM_FUNDS = n_distinct(CNPJ_FUNDO))%>%
    mutate(AUM_MEDIO = mean(AVG_SIZE)) %>%
    filter(!duplicated(YEAR))%>%
    select(YEAR, NUM_FUNDS, AUM_MEDIO)

##### TEMPO MEDIO DE SOBREVIVENCIA #####

idade_media_todos_os_fundos = base_completa %>%
    group_by(CNPJ_FUNDO) %>%
    summarise(DATA_MIN = min(YEAR_MONTH) , DATA_MAX =
        max(YEAR_MONTH) , NUM_APARICOES = n()) %>%
    mutate(SOBREVIVENCIA = mean(NUM_APARICOES) ,
        MED_SOBREVIVENCIA = median(NUM_APARICOES))

idade_media_acoes = base_completa %>%
    filter(CLASSE == "Fundo de Ações") %>%
    group_by(CNPJ_FUNDO) %>%
    summarise(DATA_MIN = min(YEAR_MONTH) ,
        DATA_MAX = max(YEAR_MONTH) ,
        NUM_APARICOES = n()) %>%
    mutate(SOBREVIVENCIA = mean(NUM_APARICOES) ,
        MED_SOBREVIVENCIA = median(NUM_APARICOES))

```

```

idade_media_multimercado= base_completa %>%
  filter(CLASSE != "Fundo de Ações") %>%
  group_by(CNPJ_FUNDO) %>%
  summarise(DATA_MIN = min(YEAR_MONTH) ,
            DATA_MAX = max(YEAR_MONTH) ,
            NUM_APARICOES = n() ) %>%
  mutate(SOBREVIVENCIA = mean(NUM_APARICOES) ,
         MED_SOBREVIVENCIA = median(NUM_APARICOES))

##### TAMANHO #####

base_completa = base_completa %>% group_by(CNPJ_FUNDO) %>%
  mutate(IDADE = row_number())%>%
  ungroup()

modelo_mqo = lm(PERF_MONTH - RF_PERF ~
  BETA_FUNDO*(MKT_PERF - RF_PERF) + (lag(VL_PATRIM_LIQ))
  + (lag(VL_PATRIM_LIQ))*factor(CLASSE) + factor(CLASSE)
  + lag(PERF_MONTH) +
  (lag(VL_PATRIM_LIQ))*lag(PERF_MONTH) + IDADE +
  lag(PERF_MONTH)*factor(CLASSE) , base_completa)

summary(modelo_mqo)

plot(lag(base_completa$VL_PATRIM_LIQ) ,
     base_completa$PERF_MONTH, main = "Scatter Plot
     Performance Mensal x lag Patrimônio Líquido", xlab =
     "lag Patrimônio Líquido", ylab = "Performance Mensal")

##### VIES DE SOBREVIVENCIA #####

```

```
sobreviventes_2023 = base_completa %>% filter(YEAR_MONTH
  == "2023-07") %>% select(CNPJ_FUNDO)
```

```
fundos_2003 = base_completa %>% filter(YEAR_MONTH ==
  "2003-01")
```

```
fundos_2003_acoes = fundos_2003 %>%
  filter(!duplicated(CNPJ_FUNDO)) %>% filter(CLASSE ==
  "Fundo de Ações")
```

```
fundos_2003_multimercado = fundos_2003 %>%
  filter(!duplicated(CNPJ_FUNDO)) %>% filter(CLASSE ==
  "Fundo Multimercado")
```

```
fundos_2003_a_2023 = fundos_2003[fundos_2003$CNPJ_FUNDO
  %in% sobreviventes_2023$CNPJ_FUNDO,]
```

```
survivors = base_completa[base_completa$CNPJ_FUNDO %in%
  fundos_2003_a_2023$CNPJ_FUNDO,]
```

```
indice_survivors = survivors %>%
  mutate(YEAR = substr(YEAR_MONTH, 1, 4)) %>%
  group_by(YEAR_MONTH) %>%
  mutate(NUM_FUNDOS = n_distinct(CNPJ_FUNDO),
    PERFORMANCE_MEDIA = mean(PERF_MONTH),
    ALFA_MEDIO = mean(ALFA)) %>%
  ungroup() %>%
  filter(!duplicated(YEAR_MONTH)) %>%
  group_by(YEAR) %>%
  mutate(PERFORMANCE_ANUAL_MEDIA = prod(1 +
    PERFORMANCE_MEDIA) - 1,
    ALFA_ANUAL_MEDIO = prod(1 + ALFA_MEDIO) - 1,
```

```

    MKT_PERF_ANUAL = prod(1+MKT_PERF) -1,
    NUM_FUNDOS_MEDIO = mean(NUM_FUNDOS) %>%
filter(! duplicated (YEAR)) %>%
select(YEAR, PERFORMANCE_ANUAL_MEDIA, ALFA_ANUAL_MEDIO,
       MKT_PERF_ANUAL, NUM_FUNDOS_MEDIO)

survivors_performance_acumulada =
  prod(indice_survivors$PERFORMANCE_ANUAL_MEDIA +1) -1
print(survivors_performance_acumulada)

survivors_alfa_acumulado =
  prod(indice_survivors$ALFA_ANUAL_MEDIO +1) -1
print(survivors_alfa_acumulado)

mkt_performance_acumulada =
  prod(indice_survivors$MKT_PERF_ANUAL +1) -1
print(mkt_performance_acumulada)

##### NAO-SOBREVIVENTES #####

nonsurvivors = base_completa[!(base_completa$CNPJ_FUNDO
  %in% sobreviventes_2023$CNPJ_FUNDO),]

indice_non_survivors = nonsurvivors %>%
  mutate(YEAR =  substr (YEAR_MONTH, 1, 4)) %>%
  group_by(YEAR_MONTH) %>%
  mutate(NUM_FUNDOS = n_distinct(CNPJ_FUNDO),
         PERFORMANCE_MEDIA =  mean (PERF_MONTH),
         ALFA_MEDIO =  mean (ALFA)) %>%
  ungroup() %>%
  filter(! duplicated (YEAR_MONTH)) %>%
  group_by(YEAR) %>%

```

```

mutate (PERFORMANCE_ANUAL_MEDIA = prod(1 +
  PERFORMANCE_MEDIA) -1,
  ALFA_ANUAL_MEDIO = prod(1 + ALFA_MEDIO) -1,
  MKT_PERF_ANUAL = prod(1+MKT_PERF) -1,
  NUM_FUNDOS_MEDIO = mean(NUM_FUNDOS) %>%
filter(!duplicated(YEAR)) %>%
select(YEAR, PERFORMANCE_ANUAL_MEDIA, ALFA_ANUAL_MEDIO,
  MKT_PERF_ANUAL, NUM_FUNDOS_MEDIO)

```

```

Non_survivors_performance_acumulada =
  prod(indice_non_survivors$PERFORMANCE_ANUAL_MEDIA +1) -1
print(Non_survivors_performance_acumulada)

```

```

Non_survivors_alfa_acumulado =
  prod(indice_non_survivors$ALFA_ANUAL_MEDIO +1) -1
print(Non_survivors_alfa_acumulado)

```

```

mkt_performance_acumulada =
  prod(indice_non_survivors$MKT_PERF_ANUAL +1) -1
print(mkt_performance_acumulada)

```

TABELA SOBREVIVENTES E NAO SOBREVIVENTES

```

tabela_survivors_non_survivors =
  left_join(indice_survivors , indice_non_survivors , by =
  "YEAR")

```

```

colnames(tabela_survivors_non_survivors) = c("ANO",
  "PERF_SURVIVORS", "ALFA_SURVIVORS", "PERF_IBOV",
  "NUM_FUND_MEDIO_SURVIVORS", "PERF_NONSURVIVORS",
  "ALFA_NONSURVIVORS", "PERF_IBOV2",
  "NUM_FUNDS_MEDIO_NONSURVIVORS")

```

```
tabela_survivors_non_survivors =
  tabela_survivors_non_survivors %>% select(ANO,
  PERF_SURVIVORS, PERF_NONSURVIVORS, ALFA_SURVIVORS,
  ALFA_NONSURVIVORS, NUM_FUND_MEDIO_SURVIVORS,
  NUM_FUNDS_MEDIO_NONSURVIVORS)

fwrite(tabela_survivors_non_survivors, "~/Documents/Monografia/Tabela_Survivors_Non_Survivors.csv", sep = ";")

##### IMPORTAÇÃO DE BASE #####

ibovespa=read_excel("~/Documents/Monografia/Ibovespa.xlsx")

BOVA11SA =
  read_csv("~/Documents/Monografia/BOVA11.SA.csv",
  col_types = cols(Close = col_number()))

COTABOVA11=read_excel("~/Documents/Monografia/COTA_BOVA11.xlsx")

##### IBOVESPA #####

ibovespa$YEAR_MONTH = ibovespa$Data

ibovespa = ibovespa %>%
  mutate(Data = sub("Jan", "01", Data),
  Data = sub("Fev", "02", Data),
  Data = sub("Mar", "03", Data),
  Data = sub("Abr", "04", Data),
  Data = sub("Mai", "05", Data),
  Data = sub("Jun", "06", Data),
```

```

Data = sub("Jul", "07", Data),
Data = sub("Ago", "08", Data),
Data = sub("Set", "09", Data),
Data = sub("Out", "10", Data),
Data = sub("Nov", "11", Data),
Data = sub("Dez", "12", Data))

```

```

ibovespa = ibovespa %>%
mutate(MES = substr(Data, 1, 2),
      ANO = substr(Data, 4, 7),
      YEAR_MONTH = paste(ANO, MES, sep = "-"))%>%
select(YEAR_MONTH , Fechamento)

```

```

ibovespa = ibovespa %>% arrange(YEAR_MONTH) %>%
mutate(MKT_RETURN = Fechamento / lag(Fechamento) -1)

```

```
##### BOVA11 #####
```

```
COTABOVA11$Data = COTABOVA11$`a partir de`
```

```

COTABOVA11 = COTABOVA11 %>%
mutate(Data = sub(" jan", "-01-", Data),
      Data = sub(" fev", "-02-", Data),
      Data = sub(" mar", "-03-", Data),
      Data = sub(" abr", "-04-", Data),
      Data = sub(" mai", "-05-", Data),
      Data = sub(" jun", "-06-", Data),
      Data = sub(" jul", "-07-", Data),
      Data = sub(" ago", "-08-", Data),
      Data = sub(" set", "-09-", Data),
      Data = sub(" out", "-10-", Data),
      Data = sub(" nov", "-11-", Data),

```



```

Data = sub("dez", "-12-", Data),
DAY = substr(Data, 1, 2),
MONTH = substr(Data, 4, 5),
YEAR = substr(Data, 7, 10)

```

```

COTABOVA11 = COTABOVA11 %>%
  mutate(Data_Combinada = as.Date(paste(YEAR, MONTH, DAY,
    sep = "-"), format = "%Y-%m-%d"),
    YEAR_MONTH = substr(Data_Combinada, 1, 7)) %>%
  arrange(Data_Combinada)%>%
  group_by(YEAR_MONTH)%>%
  filter(Data_Combinada == max(Data_Combinada)) %>%
  ungroup() %>%
  arrange(YEAR_MONTH)%>%
  mutate(BOVA11_PERF = 'Cota por ação' / lag('Cota por
    ação') - 1)

bova11 = COTABOVA11 %>% group_by(YEAR)%>%
  mutate(BOVA11_ANUAL_PERF = prod(1+BOVA11_PERF) - 1)%>%
  filter(YEAR_MONTH != "2008-11", YEAR_MONTH != "2008-12",
    YEAR_MONTH != "2023-11", YEAR_MONTH != "2023-10",
    YEAR_MONTH != "2023-09", YEAR_MONTH != "2023-08")

ibovespa_mesmo_periodo = ibovespa[ibovespa$YEAR_MONTH %in%
  bova11$YEAR_MONTH,]

perf_absoluta_bova11 = prod(bova11$BOVA11_PERF + 1) - 1
perf_absoluta_ibov =
  prod(ibovespa_mesmo_periodo$MKT_RETURN + 1) - 1
diferenca_bova_ibov =
  perf_absoluta_bova11 / perf_absoluta_ibov - 1

```

```

ibovespa_2003_a_2008 = ibovespa %>% filter(YEAR_MONTH >=
  "2003-01", YEAR_MONTH <= "2008-12") %>%
  mutate(PERF_BOVA11 = MKT_RETURN*(1-0.04388161)) %>%
  select(YEAR_MONTH, PERF_BOVA11)

```

```

bova11 = bova11 %>% ungroup() %>%
  mutate(PERF_BOVA11 = BOVA11_PERF) %>%
  select(YEAR_MONTH, PERF_BOVA11)

```

```

bova11_2003_2023 = rbind.data.frame(ibovespa_2003_a_2008,
  bova11)

```

```

check_bova_ibov = ibovespa[ibovespa$YEAR_MONTH %in%
  bova11_2003_2023$YEAR_MONTH,]

```

```

prod(bova11_2003_2023$PERF_BOVA11 +1) -1

```

```

prod(check_bova_ibov$MKT_RETURN +1) -1

```

```

bova11_2003_2023 = bova11_2003_2023 %>%
  mutate(YEAR = substr(YEAR_MONTH, 1, 4)) %>%
  group_by(YEAR) %>%
  mutate(ANNUAL_PERF_BOVA11 = prod(1+PERF_BOVA11)-1)

```

```

##### PERFORMANCE FUNDOS VS INDICE - FORMACAO DE BASE
  PASSIVA #####

```

```

base_fundos = base_completa %>% mutate(YEAR =
  substr(YEAR_MONTH, 1, 4))

```

```

bova11_alfa_calc = base_fundos %>% ungroup() %>%
  filter(!duplicated(YEAR_MONTH)) %>% select(YEAR_MONTH,

```

```

RF_PERF, MKT_PERF)

bova11_alfa = left_join(bova11_2003_2023,
  bova11_alfa_calc, by = "YEAR_MONTH")

bova11_alfa = bova11_alfa %>%
  mutate(across(c("PERF_BOVA11", "MKT_PERF", "RF_PERF"),
    as.numeric))

bova11_alfa = bova11_alfa %>% arrange(YEAR_MONTH) %>%
  mutate(BETA_BOVA11 = cov(PERF_BOVA11,
    MKT_PERF) / var(MKT_PERF),
    RETORNO_ESP_BOVA11 = BETA_BOVA11*(MKT_PERF -
    RF_PERF) + RF_PERF,
    ALFA_BOVA11 = PERF_BOVA11 - RETORNO_ESP_BOVA11)
  %>%
  group_by(YEAR)%>%
  mutate(ANNUAL_ALFA_BOVA11 = prod(ALFA_BOVA11 + 1) - 1)%>%
  select(YEAR_MONTH, YEAR, PERF_BOVA11,
    ANNUAL_PERF_BOVA11, ALFA_BOVA11, ANNUAL_ALFA_BOVA11)

##### PERFORMANCE FUNDOS VS INDICE – FORMACAO DE BASE
  ATIVA #####

performance_base = base_fundos

performance_base = performance_base %>% mutate(YEAR =
  substr(YEAR_MONTH, 1, 4), MONTH = substr(YEAR_MONTH, 6,
  7)) %>%
  group_by(CNPJ_FUNDO, YEAR)%>%
  mutate(ANNUAL_PERF = prod(PERF_MONTH+1)-1)%>%
  mutate(ANNUAL_ALFA = prod(ALFA+1)-1)%>%

```

```
ungroup()

base_fundos_ativo_passivo = left_join(performance_base ,
  bova11_alfa , by = "YEAR_MONTH")

fundos_inicio_ano = base_fundos_ativo_passivo %>%
  group_by(YEAR_MONTH) %>%
  mutate(NUM_FUNDS_START = n()) %>% ungroup() %>%
  filter(MONTH == "01") %>%
  select(YEAR.x, NUM_FUNDS_START) %>%
  filter(!duplicated(YEAR.x))

base_fundos_prob = left_join(base_fundos_ativo_passivo ,
  fundos_inicio_ano , by = "YEAR.x")

##### PERFORMANCE FUNDOS VS INDICE – CALCULO PROPORCAO
##### FUNDOS OVERPERFORMERS #####

fundos_atuacao_completa = base_fundos_prob %>%
  filter(MONTH %in% c("01", "12")) %>%
  group_by(CNPJ_FUNDO, YEAR.x) %>%
  filter(all(c("01", "12") %in% MONTH)) %>%
  ungroup() %>%
  filter(MONTH == "01")

fundos_atuacao_completa_2023 = base_fundos_prob %>%
  filter(YEAR.x == "2023") %>%
  filter(MONTH %in% c("01", "07")) %>%
  group_by(CNPJ_FUNDO, YEAR.x) %>%
  filter(all(c("01", "07") %in% MONTH)) %>%
  ungroup() %>%
  filter(MONTH == "01")
```

```
fundos_atuacao_completa_periodo =  
  rbind.data.frame(fundos_atuacao_completa ,  
  fundos_atuacao_completa_2023)  
  
fund_overperformance = fundos_atuacao_completa_periodo %>%  
  select(YEAR_MONTH, YEAR.x, CNPJ_FUNDO, DENOM_SOCIAL,  
  CLASSE, ANNUAL_PERF, ANNUAL_ALFA, ANNUAL_PERF_BOVA11,  
  ANNUAL_ALFA_BOVA11, NUM_FUNDS_START)  
  
trailing_perf = fund_overperformance %>% arrange(YEAR.x)  
  %>% group_by(CNPJ_FUNDO) %>%  
  mutate(TRAILING_5_YEAR_CUMULATIVE_PERF =  
    rollapply(ANNUAL_PERF + 1, width = 5, FUN =  
      function(x) prod(x) - 1, fill = NA, align = "right"),  
    TRAILING_5_YEAR_CUMULATIVE_PERF_MKT =  
      rollapply(ANNUAL_PERF_BOVA11 + 1, width = 5,  
        FUN = function(x) prod(x) - 1, fill = NA,  
        align = "right")) %>%  
  mutate(TRAILING_10_YEAR_CUMULATIVE_PERF =  
    rollapply(ANNUAL_PERF + 1, width = 10, FUN =  
      function(x) prod(x) - 1, fill = NA, align = "right"),  
    TRAILING_10_YEAR_CUMULATIVE_PERF_MKT =  
      rollapply(ANNUAL_PERF_BOVA11 + 1, width = 10,  
        FUN = function(x) prod(x) - 1, fill = NA,  
        align = "right")) %>%  
  mutate(TRAILING_20_YEAR_CUMULATIVE_PERF =  
    rollapply(ANNUAL_PERF + 1, width = 20, FUN =  
      function(x) prod(x) - 1, fill = NA, align = "right"),  
    TRAILING_20_YEAR_CUMULATIVE_PERF_MKT =  
      rollapply(ANNUAL_PERF_BOVA11 + 1, width = 20,  
        FUN = function(x) prod(x) - 1, fill = NA,
```

```

align = "right"))

prob_base_perf = trailing_perf %>% ungroup() %>%
  arrange(YEAR.x) %>% group_by(YEAR.x) %>%
  mutate(HIGHER_1_YEAR_CUM_PERF = ANNUAL_PERF >
    ANNUAL_PERF_BOVA11,
    HIGHER_5_YEAR_CUM_PERF =
      TRAILING_5_YEAR_CUMULATIVE_PERF >
      TRAILING_5_YEAR_CUMULATIVE_PERF_MKT,
    HIGHER_10_YEAR_CUM_PERF =
      TRAILING_10_YEAR_CUMULATIVE_PERF >
      TRAILING_10_YEAR_CUMULATIVE_PERF_MKT,
    HIGHER_20_YEAR_CUM_PERF =
      TRAILING_20_YEAR_CUMULATIVE_PERF >
      TRAILING_20_YEAR_CUMULATIVE_PERF_MKT)

soma_ano_higher_perf = prob_base_perf %>% group_by(YEAR.x)
  %>%
  summarise(SUM_HIGHER_1_YEAR_PERF =
    sum(HIGHER_1_YEAR_CUM_PERF, na.rm = TRUE),
    SUM_HIGHER_5_YEAR_PERF =
      sum(HIGHER_5_YEAR_CUM_PERF, na.rm = TRUE),
    SUM_HIGHER_10_YEAR_PERF =
      sum(HIGHER_10_YEAR_CUM_PERF, na.rm = TRUE),
    SUM_HIGHER_20_YEAR_PERF =
      sum(HIGHER_20_YEAR_CUM_PERF, na.rm = TRUE),
    CNPJ_ATIVO = n_distinct(CNPJ_FUNDO),
    FUNDOS_INICIO = first(NUM_FUNDS_START))

prob_higher_perf = soma_ano_higher_perf %>% ungroup() %>%
  arrange(YEAR.x) %>%
  mutate(PROB_1_YEAR =

```

```

SUM_HIGHER_1_YEAR_PERF/FUNDOS_INICIO ,
  PROB_5_YEAR =
    SUM_HIGHER_5_YEAR_PERF/lag (FUNDOS_INICIO , 4 ,
      default = NA) ,
  PROB_10_YEAR =
    SUM_HIGHER_10_YEAR_PERF/lag (FUNDOS_INICIO , 9 ,
      default = NA) ,
  PROB_20_YEAR =
    SUM_HIGHER_20_YEAR_PERF/lag (FUNDOS_INICIO , 19 ,
      default = NA))

```

```

fundos_acoes_2023 = trailing_perf %>% filter (YEAR.x ==
  "2023" , CLASSE == "Fundo de Ações")
fundos_multimercados_2023 = trailing_perf %>%
  filter (YEAR.x == "2023" , CLASSE != "Fundo de Ações")

```

```

##### PERFORMANCE ACOES VS INDICE - CALCULO PROPORCAO
FUNDOS OVERPERFORMERS #####

```

```

fundos_de_acoes_base = prob_base_perf %>% filter (CLASSE ==
  "Fundo de Ações")

```

```

fundos_inicio_ano_acoes = base_fundos_ativo_passivo %>%
  filter (CLASSE == "Fundo de Ações")%>%
  group_by (YEAR_MONTH) %>%
  mutate (NUM_FUNDS_START = n ()) %>% ungroup () %>%
  filter (MONTH == "01") %>%
  select (YEAR.x , NUM_FUNDS_START) %>%
  filter (! duplicated (YEAR.x))

```

```

fundo_de_acoes_conj = left_join (fundos_de_acoes_base ,

```

```

fundos_inicio_ano_acoas , by = "YEAR.x")

soma_ano_higher_perf_acoas = fundo_de_acoas_conj %>%
  group_by(YEAR.x) %>%
  summarise(SUM_HIGHER_1_YEAR_PERF =
    sum(HIGHER_1_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_5_YEAR_PERF =
      sum(HIGHER_5_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_10_YEAR_PERF =
      sum(HIGHER_10_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_20_YEAR_PERF =
      sum(HIGHER_20_YEAR_CUM_PERF, na.rm = TRUE) ,
    CNPJ_ATIVO = n_distinct(CNPJ_FUNDO) ,
    FUNDOS_INICIO = first(NUM_FUNDS_START.y))

prob_higher_perf_acoas = soma_ano_higher_perf_acoas %>%
  ungroup() %>% arrange(YEAR.x) %>%
  mutate(PROB_1_YEAR =
    SUM_HIGHER_1_YEAR_PERF/FUNDOS_INICIO ,
    PROB_5_YEAR =
      SUM_HIGHER_5_YEAR_PERF/lag(FUNDOS_INICIO , 4 ,
        default = NA) ,
    PROB_10_YEAR =
      SUM_HIGHER_10_YEAR_PERF/lag(FUNDOS_INICIO , 9 ,
        default = NA) ,
    PROB_20_YEAR =
      SUM_HIGHER_20_YEAR_PERF/lag(FUNDOS_INICIO , 19 ,
        default = NA))

##### PERFORMANCE MULTIMERCADO VS INDICE – CALCULO
PROPORCAO FUNDOS OVERPERFORMERS #####

```



```

multimercados_base = prob_base_perf %>% filter (CLASSE ==
  "Fundo_Multimercado")

fundos_inicio_ano_multimercado = base_fundos_ativo_passivo
  %>% filter (CLASSE == "Fundo_Multimercado")%>%
  group_by (YEAR_MONTH) %>%
  mutate (NUM_FUNDS_START = n()) %>% ungroup () %>%
  filter (MONTH == "01") %>%
  select (YEAR.x, NUM_FUNDS_START) %>%
  filter (! duplicated (YEAR.x))

fundo_multimercado_conj = left_join (multimercados_base ,
  fundos_inicio_ano_multimercado , by = "YEAR.x")

soma_ano_higher_perf_multimercado =
  fundo_multimercado_conj %>% group_by (YEAR.x) %>%
  summarise (SUM_HIGHER_1_YEAR_PERF =
    sum (HIGHER_1_YEAR_CUM_PERF , na.rm = TRUE) ,
    SUM_HIGHER_5_YEAR_PERF =
      sum (HIGHER_5_YEAR_CUM_PERF , na.rm = TRUE) ,
    SUM_HIGHER_10_YEAR_PERF =
      sum (HIGHER_10_YEAR_CUM_PERF , na.rm = TRUE) ,
    SUM_HIGHER_20_YEAR_PERF =
      sum (HIGHER_20_YEAR_CUM_PERF , na.rm = TRUE) ,
    CNPJ_ATIVO = n_distinct (CNPJ_FUNDO) ,
    FUNDOS_INICIO = first (NUM_FUNDS_START.y))

prob_higher_perf_multimercado =
  soma_ano_higher_perf_multimercado %>% ungroup () %>%
  arrange (YEAR.x) %>%
  mutate (PROB_1_YEAR =
    SUM_HIGHER_1_YEAR_PERF / FUNDOS_INICIO ,

```

```

PROB_5_YEAR =
  SUM_HIGHER_5_YEAR_PERF/lag(FUNDOS_INICIO, 4,
    default = NA),
PROB_10_YEAR =
  SUM_HIGHER_10_YEAR_PERF/lag(FUNDOS_INICIO, 9,
    default = NA),
PROB_20_YEAR =
  SUM_HIGHER_20_YEAR_PERF/lag(FUNDOS_INICIO, 19,
    default = NA))

```

PERFORMANCE FUNDOS VS INDICE – CALCULO PROPORCAO
 FUNDOS OVERPERFORMERS AJUSTADO AO RISCO #####

```

trailing_alfa = fund_overperformance %>% arrange(YEAR.x)
%>% group_by(CNPJ_FUNDO) %>%
mutate(TRAILING_5_YEAR_CUMULATIVE_PERF =
  rollapply(ANNUAL_ALFA + 1, width = 5, FUN =
    function(x) prod(x) - 1, fill = NA, align = "right"),
  TRAILING_5_YEAR_CUMULATIVE_PERF_MKT =
    rollapply(ANNUAL_ALFA_BOVA11 + 1, width = 5,
  FUN = function(x) prod(x) - 1, fill = NA,
    align = "right")) %>%
mutate(TRAILING_10_YEAR_CUMULATIVE_PERF =
  rollapply(ANNUAL_ALFA + 1, width = 10, FUN =
    function(x) prod(x) - 1, fill = NA, align = "right"),
  TRAILING_10_YEAR_CUMULATIVE_PERF_MKT =
    rollapply(ANNUAL_ALFA_BOVA11 + 1, width = 10,
  FUN = function(x) prod(x) - 1, fill = NA,
    align = "right")) %>%
mutate(TRAILING_20_YEAR_CUMULATIVE_PERF =
  rollapply(ANNUAL_ALFA + 1, width = 20, FUN =
    function(x) prod(x) - 1, fill = NA, align = "right"),

```

```

TRAILING_20_YEAR_CUMULATIVE_PERF_MKT =
  rollapply(ANNUAL_ALFA_BOVA11 + 1, width = 20,
    FUN = function(x) prod(x) - 1, fill = NA,
    align = "right"))

prob_base_alfa = trailing_alfa %>% ungroup() %>%
  arrange(YEAR.x) %>% group_by(YEAR.x) %>%
  mutate(HIGHER_1_YEAR_CUM_PERF = ANNUAL_ALFA >
    ANNUAL_ALFA_BOVA11,
    HIGHER_5_YEAR_CUM_PERF =
      TRAILING_5_YEAR_CUMULATIVE_PERF >
      TRAILING_5_YEAR_CUMULATIVE_PERF_MKT,
    HIGHER_10_YEAR_CUM_PERF =
      TRAILING_10_YEAR_CUMULATIVE_PERF >
      TRAILING_10_YEAR_CUMULATIVE_PERF_MKT,
    HIGHER_20_YEAR_CUM_PERF =
      TRAILING_20_YEAR_CUMULATIVE_PERF >
      TRAILING_20_YEAR_CUMULATIVE_PERF_MKT)

soma_ano_higher_alfa = prob_base_alfa%>% group_by(YEAR.x)
  %>%
  summarise(SUM_HIGHER_1_YEAR_PERF =
    sum(HIGHER_1_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_5_YEAR_PERF =
      sum(HIGHER_5_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_10_YEAR_PERF =
      sum(HIGHER_10_YEAR_CUM_PERF, na.rm = TRUE) ,
    SUM_HIGHER_20_YEAR_PERF =
      sum(HIGHER_20_YEAR_CUM_PERF, na.rm = TRUE) ,
    CNPJ_ATIVO = n_distinct(CNPJ_FUNDO) ,
    FUNDOS_INICIO = first(NUM_FUNDS_START))

```

```

prob_higher_alfa = soma_ano_higher_alfa %>% ungroup() %>%
  arrange(YEAR.x) %>%
  mutate(PROB_1_YEAR =
    SUM_HIGHER_1_YEAR_PERF/FUNDOS_INICIO ,
    PROB_5_YEAR =
      SUM_HIGHER_5_YEAR_PERF/lag(FUNDOS_INICIO , 4,
        default = NA) ,
    PROB_10_YEAR =
      SUM_HIGHER_10_YEAR_PERF/lag(FUNDOS_INICIO , 9,
        default = NA) ,
    PROB_20_YEAR =
      SUM_HIGHER_20_YEAR_PERF/lag(FUNDOS_INICIO , 19,
        default = NA))

```

```
##### ACOES #####
```

```

fundos_de_acoes_base_alfa = prob_base_alfa %>%
  filter(CLASSE == "Fundo de Ações")

fundos_inicio_ano_acoes = base_fundos_ativo_passivo %>%
  filter(CLASSE == "Fundo de Ações")%>%
  group_by(YEAR_MONIH) %>%
  mutate(NUM_FUNDS_START = n()) %>% ungroup() %>%
  filter(MONTH == "01") %>%
  select(YEAR.x, NUM_FUNDS_START) %>%
  filter(!duplicated(YEAR.x))

fundo_de_acoes_conj_alfa =
  left_join(fundos_de_acoes_base_alfa ,
    fundos_inicio_ano_acoes , by = "YEAR.x")

soma_ano_higher_alfa_acoes = fundo_de_acoes_conj_alfa %>%

```

```

group_by(YEAR.x) %>%
summarise(SUM_HIGHER_1_YEAR_PERF =
  sum(HIGHER_1_YEAR_CUM_PERF, na.rm = TRUE),
  SUM_HIGHER_5_YEAR_PERF =
    sum(HIGHER_5_YEAR_CUM_PERF, na.rm = TRUE),
  SUM_HIGHER_10_YEAR_PERF =
    sum(HIGHER_10_YEAR_CUM_PERF, na.rm = TRUE),
  SUM_HIGHER_20_YEAR_PERF =
    sum(HIGHER_20_YEAR_CUM_PERF, na.rm = TRUE),
  CNPJ_ATIVO = n_distinct(CNPJ_FUNDO),
  FUNDOS_INICIO = first(NUM_FUNDS_START.y))

prob_higher_alfa_acoes = soma_ano_higher_alfa_acoes %>%
  ungroup() %>% arrange(YEAR.x) %>%
  mutate(PROB_1_YEAR =
    SUM_HIGHER_1_YEAR_PERF/FUNDOS_INICIO,
    PROB_5_YEAR =
      SUM_HIGHER_5_YEAR_PERF/lag(FUNDOS_INICIO, 4,
        default = NA),
    PROB_10_YEAR =
      SUM_HIGHER_10_YEAR_PERF/lag(FUNDOS_INICIO, 9,
        default = NA),
    PROB_20_YEAR =
      SUM_HIGHER_20_YEAR_PERF/lag(FUNDOS_INICIO, 19,
        default = NA))

##### MULTIMERCADO #####

multimercados_base_alfa = prob_base_alfa %>% filter(CLASSE
  == "Fundo_Multimercado")

fundos_inicio_ano_multimercado = base_fundos_ativo_passivo

```

```

%>% filter (CLASSE == "Fundo_Multimercado")%>%
group_by (YEAR_MONTH) %>%
mutate (NUM_FUNDS_START = n()) %>% ungroup () %>%
filter (MONTH == "01") %>%
select (YEAR.x, NUM_FUNDS_START) %>%
filter (! duplicated (YEAR.x))

```

```

fundo_multimercado_conj_alfa =
  left_join (multimercados_base_alfa ,
fundos_inicio_ano_multimercado , by = "YEAR.x")

```

```

soma_ano_higher_alfa_multimercado =
  fundo_multimercado_conj_alfa %>% group_by (YEAR.x) %>%
summarise (SUM_HIGHER_1_YEAR_PERF =
  sum (HIGHER_1_YEAR_CUM_PERF , na.rm = TRUE) ,
  SUM_HIGHER_5_YEAR_PERF =
  sum (HIGHER_5_YEAR_CUM_PERF , na.rm = TRUE) ,
  SUM_HIGHER_10_YEAR_PERF =
  sum (HIGHER_10_YEAR_CUM_PERF , na.rm = TRUE) ,
  SUM_HIGHER_20_YEAR_PERF =
  sum (HIGHER_20_YEAR_CUM_PERF , na.rm = TRUE) ,
  CNPJ_ATIVO = n_distinct (CNPJ_FUNDO) ,
  FUNDOS_INICIO = first (NUM_FUNDS_START.y))

```

```

prob_higher_alfa_multimercado =
  soma_ano_higher_alfa_multimercado %>% ungroup () %>%
  arrange (YEAR.x) %>%
  mutate (PROB_1_YEAR =
  SUM_HIGHER_1_YEAR_PERF / FUNDOS_INICIO ,
  PROB_5_YEAR =
  SUM_HIGHER_5_YEAR_PERF / lag (FUNDOS_INICIO , 4 ,
  default = NA) ,

```

```
PROB_10_YEAR =  
  SUM_HIGHER_10_YEAR_PERF / lag(FUNDOS_INICIO, 9,  
    default = NA),  
PROB_20_YEAR =  
  SUM_HIGHER_20_YEAR_PERF / lag(FUNDOS_INICIO, 19,  
    default = NA))
```

```
fundos_acoes_2023_alfa = trailing_alfa %>% filter(YEAR.x  
  == "2023", CLASSE == "Fundo de Ações")  
fundos_multimercados_2023_alfa = trailing_alfa %>%  
  filter(YEAR.x == "2023", CLASSE != "Fundo de Ações")
```