



Marcelo Soares Sessim

O Retorno dos Replicantes: Hedge Funds replicados com métodos de seleção e ETFs

Monografia de Final de Curso

Orientador: Álvaro Lima Veiga Filho

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, Junho de 2023

Agradecimentos

É com imensa gratidão que encerro esta jornada acadêmica, e gostaria de expressar meu profundo agradecimento a todas as pessoas que contribuíram para que eu chegasse até aqui. Em primeiro lugar, agradeço à minha mãe, que sempre me incentivou e ofereceu todo o suporte necessário, mesmo nos momentos mais difíceis. À minha irmã, por seu carinho e por ser uma fonte constante de apoio. Ao meu pai, cuja experiência me mostrou por onde trilhar meu caminho. A toda a minha família, que, com amor e compreensão, esteve ao meu lado em cada passo desta caminhada.

Agradeço de coração aos meus amigos, que compartilharam comigo tanto os desafios quanto as conquistas, tornando essa trajetória mais leve e divertida. Aos meus professores, que me guiaram com dedicação e transmitiram conhecimento com paixão, contribuindo imensamente para meu crescimento pessoal e profissional. Um agradecimento especial ao meu orientador, Álvaro, pela orientação precisa e pelo apoio contínuo, sem os quais este trabalho não teria sido possível. Por fim, registro minha profunda gratidão à Quantum, pela generosidade em fornecer os dados de Hedge Funds que foram cruciais para a realização desta pesquisa.

Abstract

Este estudo aplica a metodologia proposta por Duanmu et al. (2019) na análise de retornos de hedge funds brasileiros, com o objetivo de replicar e validar os achados originais em um contexto diferente. Utilizamos uma combinação de técnicas de análise de cluster, regressão lasso e uma versão aprimorada, o adaptive lasso (adalasso), para replicar os retornos de um conjunto diversificado de hedge funds no Brasil. A aplicação dessas metodologias permitiu identificar padrões de desempenho e estrutura de risco que são consistentes com os resultados encontrados por Duanmu et al (2019). Nossos achados indicam que a introdução do adalasso oferece uma melhora na identificação de fatores de risco relevantes, proporcionando uma modelagem mais robusta e menos sensível ao sobreajuste. Este estudo contribui para a literatura ao fornecer uma análise comparativa das técnicas de modelagem de hedge funds em um mercado emergente.

Keywords

LASSO; adaLASSO; Hedge Funds; Risk Factors; Prediction; Finance; Machine Learning; ETFs; Replication.

SUMÁRIO

1	INTRODUÇÃO	7
2	REVISÃO DE LITERATURA	9
3	FORMALIZAÇÃO DO PROBLEMA	11
4	MÉTODOS DE SELEÇÃO	15
4.1	Autometrics	15
4.2	LASSO	17
4.3	adaLASSO	18
5	MEDIDAS DE DESEMPENHO	21
5.1	Mean Excess Monthly Returns	21
5.2	Sharpe Ratios	21
5.3	Sortino Ratio	22
5.4	Information Ratios	22
5.5	Skewness	23
5.6	Attrition Rates	23
6	METODOLOGIA	25
6.1	Cluster Analysis dos ETFs	26
6.2	LASSO e adaLASSO	27
6.3	Replicação	29
6.4	Janelas Preditivas e Rebalanceamento	30

7	DADOS	33
8	RESULTADOS	37
9	CONCLUSÃO	45
	Bibliografia	47

1 Introdução

Os Hedge Funds estabeleceram-se como importantes agentes do mercado financeiro por conta de conseguirem retornos com baixa correlação com outras classes de ativos. O primeiro motivo para tal é que, por definição, são fundos com o objetivo de usar quaisquer estratégias para diminuir a volatilidade e aumentar o retorno, como alavancagem e venda a descoberto. O segundo vem de seus gestores poderem investir em diversas classes de ativos não tradicionais, como derivativos.

Entretanto, os hedge funds apresentam características negativas: altas taxas de administração, baixa liquidez para a venda de suas cotas e falta de transparência dos portfólios. Diante disso, começa um campo de estudo em Finanças voltado à tentativa de replicação dos retornos desses fundos para que o investidor possa proteger-se dessas mazelas. O primeiro impasse para a replicação é criar uma carteira de ativos acessíveis ao investidor e que tenha retornos similares aos do hedge fund. O segundo impasse é entender se o investidor conseguiria identificar e quantificar corretamente os fatores de risco atrelados aos retornos dos hedge funds. Caso os retornos possam ser explicados por "betas alternativos", ou seja, caso o retorno seja recompensa pela exposição aos fatores de risco alternativos do modelo, a replicação é possível. Fung and Hsieh (2004) encontram evidências de que há "betas alternativos" ligados a fatores de risco baseados em ativos ("asset-based style factors").

Esta monografia tem a proposta de utilizar a metodologia de Duanmu et al. (2019). Isto é motivado pelas inovações na área de replicação de hedge funds propostas, conforme descrito pelos autores. Elas são a criação de fatores de risco com ETFs e o uso de técnicas avançadas de seleção de modelos e variáveis. Adicionalmente, propõe-se explorar técnicas alternativas para a seleção e redução de variáveis, focando na aplicação do Adaptive LASSO, uma vez que os estimadores deste modelo mostram-se mais consistentes em comparação ao LASSO tradicional, que é utilizado no estudo mencionado acima. Finalmente, a verificação da aplicabilidade dessa metodologia utilizando dados de Hedge Funds brasileiros é crucial para atestar sua validade externa, beneficiando os stakeholders do mercado financeiro.

2 Revisão de Literatura

A literatura de replicação de Hedge Funds através de fatores de risco segue uma abordagem de análise no estilo de Sharpe (1992). Dito isso, o estudo seminal de Fung e Hsieh(2001) introduz as estratégias baseadas em regras, principalmente criando portfólios de straddles retrospectivos (“lookback straddles”), o que se aproxima mais do que é feito pelos gestores de Hedge Funds.

Já em Fung e Hsieh (2004), são encontradas evidências da capacidade de se replicar um Hedge Fund, visto que boa parte dos retornos pode ser explicada por betas alternativos. Além disso, o alpha de Fung e Hsieh (2004) aparece como um indicador da capacidade de se prever os retornos utilizando os risk factor do modelo. Portanto, ele denota a eficiência do modelo na identificação correta da estrutura de risco a que os retornos estão atrelados.

Hasanhodzic e Lo (2007) aplica a metodologia de fatores de risco e estima seis fatores utilizando os retornos de Hedge Funds individuais para depois analisar o desempenho dos replicantes. Seus resultados são positivos para várias categorias de fundos, mas, para as categorias “Event Driven” e “Emerging Markets”, os replicantes não alcançam os resultados desejados.

A introdução de modelos não lineares para a replicação dos retornos Hedge Funds foi objeto de estudo de Amenc et al.(2010) e de Roncalli e Weisang (2009). Ambos chegaram à conclusão de que há poucas evidências da presença de não-linearidades na distribuição desses retornos independentemente da estratégia seguida pelo fundo. Ademais, os resultados de Amenc et al. (2010) constatam que o uso do racional econômico na seleção dos fatores para cada categoria de Hedge Fund leva a uma melhora na performance dos replicantes fora da amostra.

Giamouridis e Paterlini (2010) e Weber e Peres (2013) criam um maior número de fatores de risco (16 e 30, respectivamente). Então, utilizam técnicas de Statistical Learning, como regressões stepwise e métodos de encolhimento (RIDGE e LASSO) em todo o conjunto de fatores com o intuito de encontrar o melhor conjunto de fatores para cada fundo. Além disso, Weber e Peres (2013) empregam fatores de risco totalmente investíveis (utilizam contratos futuros de portfólios negociados por profissionais para oferecer diversificação de carteira a fundos e investidores institucionais), levam em consideração a variação no tempo dos fatores (utilizando Média Móvel Ponderada Exponencialmente (EWMA)) e também tentam incorporar custos

de transação e métricas de intensidade de transação no modelo.

Duanmu et al. (2019) inovam ao argumentar que ETFs passivos podem ser interpretados como proxies para os fatores de risco dos Hedge Funds por seus padrões de retorno serem gerados de maneira formulada, sem a intervenção humana, de modo que o universo dos ETFs abranja os fatores levados em conta pelos investidores. Outro ponto relevante é a utilização do R^2 ajustado, similarmente a Titman and Tiu (2011) e Bollen (2013) para definir a “replicabilidade” dos fundos de investimento. Os fundos replicáveis são aqueles cuja correspondência de R^2 ajustado dentro da amostra é a melhor. Fora da amostra, os replicantes tiveram, inclusive, desempenho melhor do que o portfólio de fundos de hedge replicáveis, provavelmente por conta da disparidade entre as estruturas de taxas contabilizada. Assim como Bollen (2013), Bali et al. (2011, 2012, 2014) e Titman and Tiu (2011), Duanmu et al. (2019) demonstram que há uma relação entre a exposição a risco sistemático e os retornos dos Hedge Funds.

3 Formalização do Problema

Em finanças, o retorno de um ativo é uma medida de ganho ou perda que um investidor obtém sobre esse ativo em um período de tempo específico. O retorno é calculado com base nos preços de compra e venda do ativo, refletindo a mudança no valor do ativo ao longo do tempo. Matematicamente, o retorno R^t de um ativo entre os tempos $t - 1$ e t pode ser expresso da seguinte forma:

$$R^t = \frac{P^t - P^{t-1}}{P^{t-1}} \times 100$$

onde:

- R^t é o retorno aritmético percentual do ativo no tempo t .
- P^t é o preço do ativo no tempo t .
- P^{t-1} é o preço do ativo no tempo $t - 1$.

Uma carteira de ativos é uma combinação de diversos investimentos, como ações, títulos e fundos, que um investidor possui. O objetivo principal de construir uma carteira é diversificar o risco e potencializar o retorno, que é a medida do ganho ou perda que um investidor obtém em um período de tempo específico. O retorno é um valor fundamental, pois fornece uma indicação do desempenho de um investimento e ajuda na tomada de decisões sobre onde alocar recursos.

O retorno de uma carteira, em um determinado período de tempo, é dado pela média ponderada dos retornos dos ativos que a compõem, com os pesos correspondendo à proporção do valor total da carteira alocada em cada ativo. Assim, o retorno de uma carteira $R_p^{(t)}$ pode ser representado pela seguinte fórmula:

$$R_p^{(t)} = \sum_{i=1}^n \beta_i R_i^{(t)}$$

onde:

- $R_p^{(t)}$ é o retorno aritmético da carteira no tempo t .
- β_i é a fração do valor total da carteira alocada no ativo i .

-
- $R_i^{(t)}$ é o retorno aritmético do ativo i no tempo t .
 - n é o número de ativos na carteira.

Neste trabalho, o objetivo é construir uma carteira de ETFs (Exchange-Traded Funds) que consiga replicar os retornos de Hedge Funds. Os ETFs são fundos que acompanham índices, setores ou outras classes de ativos, e podem ser negociados em bolsa de valores como ações. Os retornos serão calculados em base diária, ou seja, em intervalos de um dia.

Para replicar os retornos dos Hedge Funds, consideramos a seguinte equação que modela o retorno de uma carteira de ETFs:

$$r_i^{(t)} = \beta_1 \text{ETF_ret}_1^{(t)} + \beta_2 \text{ETF_ret}_2^{(t)} + \dots + \beta_j \text{ETF_ret}_j^{(t)}$$

onde:

- $r_i^{(t)}$ é o retorno aritmético da carteira no tempo t .
- β_j são os coeficientes que representam a alocação de cada ETF j .
- $\text{ETF_ret}_j^{(t)}$ é o retorno aritmético da ETF j no tempo t .

O cerne da questão está em selecionar os ETFs e encontrar os valores de β_j que maximizem a semelhança do retorno da carteira formada por ETFs com o retorno reportado por cada Hedge Fund. Como a porcentagem dos recursos alocados em cada ativo da carteira deve somar 1, deveria haver uma restrição no momento de implementar as metodologias de estimação dos parâmetros β_j . Para evitar essa complicação, optamos por inserir um ativo de renda fixa ($r_f^{(t)}$) na nossa carteira com seu respectivo β_0 :

$$r_i^{(t)} = \beta_0 r_f^{(t)} + \beta_1 \text{ETF_ret}_1^{(t)} + \beta_2 \text{ETF_ret}_2^{(t)} + \dots + \beta_j \text{ETF_ret}_j^{(t)}$$

Com uma simples manipulação algébrica, encontramos a equação que não necessita de restrição para algum dos parâmetros das ETFs:

$$r_i^{(t)} - r_f^{(t)} = \beta_1 (\text{ETF_ret}_1^{(t)} - r_f^{(t)}) + \beta_2 (\text{ETF_ret}_2^{(t)} - r_f^{(t)}) + \dots + \beta_j (\text{ETF_ret}_j^{(t)} - r_f^{(t)})$$

A equação acima pode ser reescrita de outro modo que facilita a compreensão de que, agora, o β_0 omitido faz com que a soma dos β s seja igual a 1:

$$r_i^{(t)} = \left(1 - \sum_{j=1}^n \beta_j\right) r_f^{(t)} + \sum_{j=1}^n \beta_j \text{ETF_ret}_j^{(t)}$$

onde $r_i^{(t)}$ é o retorno aritmético ajustado da carteira no tempo t .

4 Métodos de Seleção

Dada a formalização do problema apresentada no capítulo anterior, o próximo passo é lidar com o número de ETFs e a colinearidade inerente a seus retornos. Métodos de estimação convencionais, como o OLS (Ordinary Least Squares), enfrentam uma limitação significativa para este estudo: o número de variáveis explicativas excede o número de observações disponíveis. Portanto, é necessário utilizar métodos de seleção de variáveis para garantir a convergência do modelo. Neste capítulo, discutiremos o LASSO e o adaLASSO, ambos utilizados nas estimativas, além do Autometrics, uma metodologia avançada de stepwise regression.

4.1 Autometrics

Esta seção segue Epprecht et al. (2017). O Autometrics é uma metodologia de regressão stepwise para seleção de variáveis, fundamentada no algoritmo GETS (General-to-Specific). Este algoritmo seleciona automaticamente modelos empíricos a partir dos dados observados, explorando todos os caminhos possíveis de redução, começando com um modelo geral e eliminando variáveis insignificantes até restarem apenas as variáveis relevantes. A metodologia Autometrics passa por cinco estágios principais:

1. Formulação do GUM

- O primeiro estágio envolve a especificação do Modelo Geral Não Restrito (GUM) pelo usuário, com base em teoria do assunto, conhecimento institucional, contingências históricas, disponibilidade de dados e informações de medição. O GUM deve incluir variáveis potencialmente relevantes e excluir efeitos irrelevantes, garantindo ortogonalidade entre os regressores.

2. Testes de Especificação

- Após a estimativa do GUM por mínimos quadrados ordinários, a segunda etapa consiste em testar o modelo para identificar erros de especificação. Os testes garantem que o modelo seja congruente, ajustando o nível de significância ou excluindo testes conforme necessário durante a simplificação do GUM.

3. Reduções Pré-busca

- Na terceira etapa, ocorre uma simplificação preliminar do GUM, eliminando variáveis altamente irrelevantes com base em testes F e t. A cada redução, são confirmados os testes de diagnóstico para assegurar a congruência, resultando em um novo GUM que servirá de base para as etapas seguintes.

4. Caminhos Múltiplos de Redução

- O algoritmo então implementa uma busca por múltiplos caminhos, iniciando de todos os pontos de exclusão possíveis. As buscas filtram repetidamente variáveis relevantes usando testes t e F em bloco, verificando a congruência do modelo a cada etapa. O caminho é encerrado quando todas as variáveis restantes são significativas ou um teste de diagnóstico falha, resultando no modelo terminal desse caminho.

5. Seleção do Modelo Final

- Após explorar todos os caminhos e encontrar os modelos terminais distintos, utiliza-se o procedimento de *encompassing* para testar entre eles, retendo apenas as especificações sobreviventes. Os modelos terminais são testados contra sua união para encontrar um modelo abrangente não dominado. O processo continua até que se obtenha um modelo final único, selecionado por um critério de informação predefinido.

Além de compartilhar essas etapas com o algoritmo PcGets, o Autometrics introduz melhorias, como a busca em árvore e a capacidade de detectar *outliers* utilizando saturação de indicadores de impulso (*IIS*). Isso permite ao Autometrics lidar com mais variáveis candidatas do que observações, aplicando o algoritmo de blocos cruzados quando necessário.

Os dados encontrados por Epprecht et al. (2017) mostram que, quando há colinearidade nos dados e grande número de variáveis em relação ao número de observações, é preferível usar LASSO e adaLASSO em detrimento do Autometrics. Como observado no estudo, os modelos estimados pelo Autometrics apresentaram resultados insatisfatórios fora da amostra quando as variáveis candidatas eram correlacionadas, enquanto LASSO e adaLASSO conseguiram selecionar modelos com estatísticas satisfatórias em termos de R^2 fora da amostra e RMSE.

4.2 LASSO

Como mencionado anteriormente, utilizaremos o método LASSO para a seleção de variáveis. Tibshirani (1996) propôs este modelo como uma alternativa para a regressão Ridge, especialmente para casos com parâmetros potencialmente iguais a 0 em alta dimensão. No LASSO, utiliza-se a penalidade ℓ_1 , que consiste em penalizar o módulo dos β 's. Desse modo, os coeficientes LASSO minimizarão a seguinte equação:

$$\hat{\beta}_{\text{LASSO}}^{(t)} = \arg \min_{\beta^{(t)}} \left\{ \sum_{i=1}^n \left(y_i^{(t)} - \beta_0^{(t)} - \sum_{j=1}^p x_{ij}^{(t)} \beta_j^{(t)} \right)^2 + \lambda^{(t)} \sum_{j=1}^p |\beta_j^{(t)}| \right\}$$

Onde:

- $y_i^{(t)}$ é o valor observado da variável dependente na i -ésima observação no tempo t ,
- $\beta_0^{(t)}$ é o intercepto do modelo no tempo t ,
- $x_{ij}^{(t)}$ representa o valor da j -ésima variável independente na i -ésima observação no tempo t ,
- $\beta_j^{(t)}$ é o coeficiente associado à j -ésima variável independente no tempo t ,
- $\lambda^{(t)}$ é o parâmetro de penalização no tempo t que controla o nível de shrinkage (redução) dos coeficientes,
- $|\beta_j^{(t)}|$ é o valor absoluto do coeficiente $\beta_j^{(t)}$, sendo a penalização ℓ_1 aplicada sobre esses valores para promover a seleção de variáveis no tempo t .

A seleção das variáveis ocorre por conta da penalização ℓ_1 . Através desta metodologia de otimização, torna-se possível reduzir o modelo para que não contenha as variáveis explicativas que revelam pouca ou nenhuma informação sobre a variável dependente. Isso ajuda a lidar melhor com colinearidades, proporcionando um ganho significativo na interpretabilidade do modelo resultante e na capacidade de identificar efeitos significativos em um conjunto menor de variáveis.

4.3 adaLASSO

Um impasse ao trabalhar com o LASSO está na consistência de sinal do estimador LASSO. De acordo com Zhao e Yu (2006), ela requer a validade da "Irrepresentable Condition" e, para tal, deve haver um limite para a correlação entre as variáveis selecionadas e não selecionadas.

Para superar este empecilho, Zou (2006) propôs o modelo adaLASSO (Adaptive LASSO), que inclui os pesos ω_j na penalização. A partir disto, lida-se com uma condição para a consistência mais fraca, a "Weighted Irrepresentable Condition". O problema de minimização para encontrar os $\hat{\beta}_{\text{adaLASSO}}^{(t)}$ é montado da seguinte forma:

$$\hat{\beta}_{\text{adaLASSO}}^{(t)} = \arg \min_{\beta^{(t)}} \left\{ \sum_{i=1}^n \left(y_i^{(t)} - \beta_0^{(t)} - \sum_{j=1}^p x_{ij}^{(t)} \beta_j^{(t)} \right)^2 + \lambda^{(t)} \sum_{j=1}^p \omega_j^{(t)} |\beta_j^{(t)}| \right\}$$

Onde:

- $y_i^{(t)}$ é o valor observado da variável dependente na i -ésima observação no tempo t ,
- $\beta_0^{(t)}$ é o intercepto do modelo no tempo t ,
- $x_{ij}^{(t)}$ representa o valor da j -ésima variável independente na i -ésima observação no tempo t ,
- $\beta_j^{(t)}$ é o coeficiente associado à j -ésima variável independente no tempo t ,
- $\lambda^{(t)}$ é o parâmetro de penalização no tempo t que controla o nível de shrinkage dos coeficientes,
- $\omega_j^{(t)}$ são os pesos aplicados na penalização de cada coeficiente no tempo t , adaptativos de acordo com a relevância prévia estimada de cada variável.

Portanto, em casos como o estudado nesta pesquisa, cuja gama de variáveis é grande e a correlação entre elas é relevante, o modelo LASSO pode não apresentar consistência nos seus resultados. De acordo com a literatura citada acima, nota-se que o uso do adaLASSO adequa-se melhor ao escopo deste estudo. Assim, serão gerados resultados com ambos os modelos para que haja comparação de seus resultados.

Para encontrar os ω_j , há uma etapa prévia em que utiliza-se o MQO, em pequena dimensão, ou o LASSO tradicional, em alta dimensão. Através desta etapa, serão encontrados os estimadores ω_j . Com estes ω_j , os pesos são usualmente encontrados com a seguinte fórmula:

$$\omega_j^{(t)} = \frac{1}{\left(\hat{\beta}_{j,\text{RIDGE}}\right)^{(t)} + \frac{1}{\sqrt{n^{(t)}}}}$$

Onde:

- ω_j é o peso aplicado ao coeficiente β_j no modelo adaLASSO,
- $\hat{\beta}_{j,\text{RIDGE}}$ são os estimadores iniciais dos coeficientes, obtidos pelo método RIDGE,
- n é o tamanho da amostra, usado para ajustar os pesos, onde \sqrt{n} ajuda a diminuir a influência de $\hat{\beta}_{j,\text{RIDGE}}$ à medida que o tamanho da amostra aumenta, promovendo uma maior penalização para coeficientes baseados em estimadores menos certos.

5 Medidas de Desempenho

Neste capítulo, serão elaboradas explicações de algumas medidas de desempenho relevantes para a área de replicação de Hedge Funds.

5.1 Mean Excess Monthly Returns

Mean Excess Monthly Returns ou Média do Excesso de Retornos Mensais é uma medida simples para comparar os retornos de um investimento a outro. Para tal, ao comparar o retorno do replicador do fundo com o próprio fundo, esta medida, mesmo que simples, pode dar informações relevantes. O cálculo consiste em subtrair da média mensal do retorno do replicador a média mensal do retorno do fundo.

5.2 Sharpe Ratios

O Sharpe Ratio (1994) é uma medida que calcula a recompensa em termos de cada ponto percentual de volatilidade que se tem em dado portfólio. Para tal, o cálculo é feito a partir da seguinte fórmula:

$$\text{Índice de Sharpe} = \frac{R_p - R_f}{\sigma_p}$$

Onde:

- R_p é o retorno médio do portfólio,
- R_f é o retorno de um ativo livre de risco, como por exemplo, títulos do Tesouro ou CDI,
- σ_p é o desvio padrão do retorno do portfólio, representando o risco total do portfólio.

Assim como em Weber et al. (2013), neste estudo será utilizado o Sharpe Ratio para comparar o desempenho do replicador com o desempenho do fundo.

5.3 Sortino Ratio

O Sortino Ratio (1994) tem o mesmo numerador que o Sharpe Ratio, mas, no denominador, considera-se somente a volatilidade para baixo ("downside volatility"). Dessa forma, procura-se calcular a recompensa em relação às quedas do retorno do portfólio. Assim, ativos que oferecem maior proteção a retornos negativos terão melhor performance de acordo com este indicador.

$$\text{Sortino Ratio} = \frac{R_p - R_f}{\sigma_d}$$

Onde:

- R_p é o retorno médio do portfólio,
- R_f é o retorno de um ativo livre de risco, como por exemplo, títulos do Tesouro ou CDI,
- σ_d é o desvio padrão dos retornos negativos, medindo apenas a volatilidade associada aos retornos abaixo de um determinado limite.

5.4 Information Ratios

A Information Ratio (Razão de Informação) é uma medida do retorno ajustado ao risco de um portfólio em relação a um benchmark ou retorno alvo. É calculada da seguinte forma:

$$\text{Information Ratio} = \frac{R_p - R_b}{\sigma_p}$$

Onde:

- R_p é o retorno médio do portfólio,
- R_b é o retorno do benchmark ou índice comparável,
- σ_p é o desvio padrão dos retornos excedentes do portfólio sobre o benchmark, representando o risco do portfólio relativo a esse benchmark.

5.5 Skewness

Skewness é uma medida estatística que descreve a assimetria na distribuição dos retornos de um ativo financeiro em torno de sua média. A fórmula é:

$$\text{Skewness} = \frac{E[(X - \mu)^3]}{\sigma^3}$$

Onde:

- $E[(X - \mu)^3]$ é o valor esperado (ou a média) dos desvios ao cubo da média (μ) dos retornos,
- σ é o desvio padrão dos retornos,
- Uma assimetria positiva indica que a cauda no lado direito da distribuição é mais longa ou mais gorda que a esquerda, sugerindo uma maior probabilidade de altos retornos. Por outro lado, uma assimetria negativa indica um potencial para retornos mais baixos.

5.6 Attrition Rates

A Taxa de Atrito, frequentemente usada no contexto da análise da base de clientes, mede a taxa na qual os clientes deixam de fazer negócios com uma entidade. É calculada da seguinte forma:

$$\text{Attrition Ratio} = \frac{\text{Número de Clientes Perdidos}}{\text{Número Total de Clientes no Início}}$$

Onde:

- Número de Clientes Perdidos é o número de clientes que pararam de comprar ou assinar o serviço durante um certo período,
- Número Total de Clientes no Início é o total de clientes no início do período.

6 Metodologia

Neste capítulo, a metodologia adotada será explicada. O racional utilizado aqui segue de perto o que foi adotado em Duanmu et al (2019). A primeira subseção aborda a redução do número de ETFs a serem utilizados de forma a atenuar o efeito da colinearidade entre eles. A segunda subseção descreve como serão utilizados o LASSO e o adaLASSO. A terceira subseção aborda a replicação. Na quarta, abordamos como foram feitas as janelas preditivas e o rebalanceamento dos coeficientes dos modelos.

Como descrito no capítulo 3, o retorno de uma carteira é uma combinação linear do retorno dos ativos que a compõe, onde cada coeficiente representa a proporção do valor total investido no ativo. Desse modo, a regressão a ser montada segue a seguinte fórmula:

$$r_{i,\text{gross}}^{(t)} - r_f^{(t)} = \beta_1^{(t)}(\text{ETF}_1^{(t)} - r_f^{(t)}) + \beta_2^{(t)}(\text{ETF}_2^{(t)} - r_f^{(t)}) + \dots + \beta_j^{(t)}(\text{ETF}_j^{(t)} - r_f^{(t)}) + \epsilon_i^{(t)},$$

- $r_{i,\text{gross}}^{(t)}$ é o retorno aritmético bruto mensal do fundo i ,
- $r_f^{(t)}$ é a taxa livre de risco representada pelo retorno diário do CDI,
- $\beta_1^{(t)}, \beta_2^{(t)}, \dots, \beta_j^{(t)}$ são os coeficientes de regressão que medem a sensibilidade do retorno do fundo i em relação ao retorno dos ETFs correspondentes, refletindo a exposição do fundo a cada ETF,
- $\text{ETF}_1^{(t)}, \text{ETF}_2^{(t)}, \dots, \text{ETF}_j^{(t)}$ representam os retornos aritmético dos ETFs,
- $\epsilon_i^{(t)}$ é o termo de erro, representando a parte do retorno do fundo i que não é explicada pelos retornos dos ETFs selecionados.

Getmansky et al (2004) evidenciam que os retornos de fundos de hedge apresentam correlação serial, atribuível tanto às posições em ativos ilíquidos quanto à suavização deliberada realizada pelos gestores ao reportarem esses retornos. Com base nesses achados, implementamos uma correção de suavização do tipo MA(2) nos retornos dos fundos de hedge utilizados em nossos modelos de regressão.

Ao subtrair a taxa livre de risco tanto do retorno do Hedge Fund quanto do retorno dos ETFs, implicitamente fazemos com que a soma dos betas seja 1. Neste ponto, separa-nos da literatura de referência (Duanmu et al (2019)), que supõe a

falta de necessidade da soma dos coeficientes ser igual a 1. Ademais, existe o $\beta_0^{(t)}$ referente à porcentagem do capital investido na taxa livre de risco, cujo cálculo é relevante para os Hedge Funds que serão replicados.

$$r_{i,\text{gross}}^{(t)} = \beta_0^{(t)} r f^{(t)} + \beta_1^{(t)} \text{ETF}_1^{(t)} + \beta_2^{(t)} \text{ETF}_2^{(t)} + \dots + \beta_{100}^{(t)} \text{ETF}_{100}^{(t)} + \epsilon_i^{(t)}, \quad (3)$$

- $r_{i,\text{gross}}^{(t)}$ é o retorno aritmético bruto mensal do fundo i ,
- $r f^{(t)}$ é a taxa livre de risco representada pelo retorno diário do CDI,
- $\beta_0^{(t)}$ é o coeficiente que mede a parcela do retorno do fundo i atribuível diretamente à taxa livre de risco, representando a porcentagem do capital do fundo investido em ativos de risco zero,
- $\beta_i^{(t)}$ para $i = 1, 2, \dots, 100$ são os coeficientes de regressão que medem a sensibilidade do retorno do fundo i em relação ao retorno dos ETFs correspondentes, refletindo a exposição do fundo a esses ETFs,
- $\text{ETF}_i^{(t)}$ representa o retorno aritmético dos ETFs.

6.1 Cluster Analysis dos ETFs

Para reduzir ainda mais os efeitos adversos da colinearidade entre os retornos dos ETFs, optou-se por fazer um cluster analysis deles além de usar o LASSO e o adaLASSO posteriormente. A colinearidade impacta negativamente as regressões pois leva a estimativas imprecisas dos coeficientes, infla a variância dos estimadores e, geralmente, leva a um baixo desempenho preditivo. Desse modo, ainda que o LASSO e o adaLASSO consigam lidar com parte deste problema por conta do encolhimento dos coeficientes e da seleção de variáveis, é necessário utilizar este outro artifício.

O cluster analysis é feito antes de rodar o LASSO e o adaLASSO para que haja uma pré-seleção dos ETFs com retornos mais relevantes. Dentro de cada cluster, será escolhido o ETF cuja distância entre ele e o centro do cluster seja a menor. Para calcular tal distância, utilizamos a medida SDI de Sun et al(2012) e selcionamos o ETF com o menor SDI, que vai servir como uma proxy para os outros ETFs do cluster. A fórmula do SDI é a seguinte:

$$\text{SDI}_i^{(t)} = 1 - \text{corr}(r_i^{(t)}, \mu_I^{(t)})$$

Onde:

- $\text{corr}(r_i^{(t)}, \mu_I^{(t)})$ é a correlação do retorno do ETF i com o retorno médio dos ETFs do mesmo cluster I ,
- $\mu_I^{(t)} = \frac{\sum_{i \in I} r_i^{(t)}}{\text{count}(i \in I)}$ é a média dos retornos do cluster I .

Assim como em Duanmu et al (2019), o número ótimo de clusters é escolhido iterativamente. Criamos modelos fazendo variar de 2 a 100 o limite de clusters. Os retornos dos ETFs escolhidos em cada um dos 99 modelos para cada Hedge Fund serão usados como variáveis explicativas nos modelos LASSO e adaLASSO. Finalmente, o modelo replicante é o modelo com menor BIC dentre os 99.

6.2 LASSO e adaLASSO

Após o cluster analysis, rodamos o LASSO e o adaLASSO, como foi descrito no capítulo 4. Diante do que foi dito acima e do formalização do LASSO no Capítulo 4, a equação de minimização do LASSO será a seguinte:

$$\hat{\beta}_{\text{LASSO}}^{(t)} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(r_{i,\text{gross}}^{(t)} - \beta_0^{(t)} r f^{(t)} - \sum_{j=1}^p \beta_j^{(t)} \text{ETF}_j^{(t)} \right)^2 + \lambda \sum_{j=0}^p |\beta_j^{(t)}| \right\}$$

Onde:

- $r_{i,\text{gross}}^{(t)}$ é o retorno aritmético bruto na i -ésimo Hedge Fund,
- $r f^{(t)}$ é a taxa livre de risco,
- $\text{ETF}_j^{(t)}$ representa o retorno aritmético do j -ésimo ETF selecionado no cluster analysis,
- $\beta_0^{(t)}$ é o coeficiente associado à taxa livre de risco, representando o impacto direto de $r f^{(t)}$ sobre o retorno do fundo,
- $\beta_j^{(t)}$ para $j = 1, 2, \dots, 100$ é o coeficiente associado ao retorno do j -ésimo ETF,

-
- λ é o parâmetro de penalização que controla o nível de shrinkage (redução) dos coeficientes,
 - $|\beta_j^{(t)}|$ é o valor absoluto do coeficiente $\beta_j^{(t)}$, sendo a penalização ℓ_1 aplicada sobre esses valores para promover a seleção de variáveis e a exclusão de coeficientes irrelevantes.

Além disso, a equação de minimização do adaLASSO fica da seguinte forma:

$$\hat{\beta}_{\text{adaLASSO}}^{(t)} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (r_{i,\text{gross}}^{(t)} - \beta_0^{(t)} r f^{(t)} - \sum_{j=1}^p \beta_j^{(t)} \text{ETF}_j^{(t)})^2 + \lambda \sum_{j=0}^p \omega_j^{(t)} |\beta_j^{(t)}| \right\}$$

- $r_{i,\text{gross}}^{(t)}$ é o retorno aritmético bruto mensal do fundo i ,
- $r f^{(t)}$ é a taxa livre de risco representada pelo retorno diário do CDI,
- $\beta_0^{(t)}$ é o coeficiente associado diretamente à taxa livre de risco, representando a porcentagem do capital do fundo investido em ativos de risco zero,
- $\beta_j^{(t)}$ para $j = 1, 2, \dots, 100$ são os coeficientes de regressão que medem a sensibilidade do retorno do fundo i em relação ao retorno dos ETFs correspondentes, refletindo a exposição do fundo a esses ETFs,
- $\text{ETF}_j^{(t)}$ representa o retorno aritmético dos ETFs selecionados no cluster analysis para $j = 1, 2, \dots, 100$,
- λ é o parâmetro de penalização que controla o nível de shrinkage (redução) dos coeficientes,
- $\omega_j^{(t)}$ são os pesos aplicados na penalização de cada coeficiente, adaptativos de acordo com a relevância prévia estimada de cada variável, inclusive para $\beta_0^{(t)}$.

Vale notar que, como foi escrito no capítulo referente ao adaLASSO, os pesos $\omega_j^{(t)}$ são calculados deste modo:

$$\omega_j^{(t)} = \frac{1}{\hat{\beta}_{j,\text{RIDGE}}^{(t)} + \frac{1}{\sqrt{n^{(t)}}}}$$

Onde:

- $\omega_j^{(t)}$ é o peso aplicado ao coeficiente $\beta_j^{(t)}$ no modelo adaLASSO,
- $\left(\hat{\beta}_{j,\text{RIDGE}}\right)^{(t)}$ são os estimadores iniciais dos coeficientes, obtidos pelo método RIDGE,
- $n^{(t)}$ é o tamanho da amostra, usado para ajustar os pesos, onde $\sqrt{n^{(t)}}$ ajuda a diminuir a influência de $\left(\hat{\beta}_{j,\text{RIDGE}}\right)^{(t)}$ à medida que o tamanho da amostra aumenta, promovendo uma maior penalização para coeficientes baseados em estimadores menos certos.

6.3 Replicação

Consideramos os valores do *BIC* (Bayesian Information Criteria) das regressões correspondentes como proxies internos da "qualidade geral" da nossa metodologia. Interpretamos um R^2 ajustado mais elevado e valores de *BIC* mais baixos como indicativos do sucesso da nossa metodologia em captar os fatores de risco dos Hedge Funds e, conseqüentemente, do potencial para replicar os retornos desses fundos usando ETFs.

Entretanto, o objetivo principal é testar o poder preditivo da metodologia, de modo a validar o poder explicativo interno demonstrado pelos altos valores de R^2 e baixos valores de *BIC*. O LASSO e adaLASSO são rodados in-sample para um período de 1 ano levando em conta retornos diários das ETFs e Hedge Funds. Após verificar quais coeficientes não são nulos, rodamos um OLS com as variáveis selecionadas pelo LASSO e adaLASSO.

Como rodamos o cluster analysis limitando o número máximo de clusters a 100, na terceira etapa, escolhemos, dentre os 99 modelos possíveis, o que teve o melhor desempenho no ano in-sample para ser o replicante de um fundo na primeira janela preditiva. Finalmente, falta medir o desempenho deste replicante para o próximo mês. O desempenho do replicante do Hedge Fund é determinado da seguinte maneira:

$$\text{ReplicantRet}_i^{(t)} = \sum_{j=1}^{n^{(t)}} \beta_j^{(t-1)} (ETF_j^{(t)} - r_f^{(t)}), \quad (6.1)$$

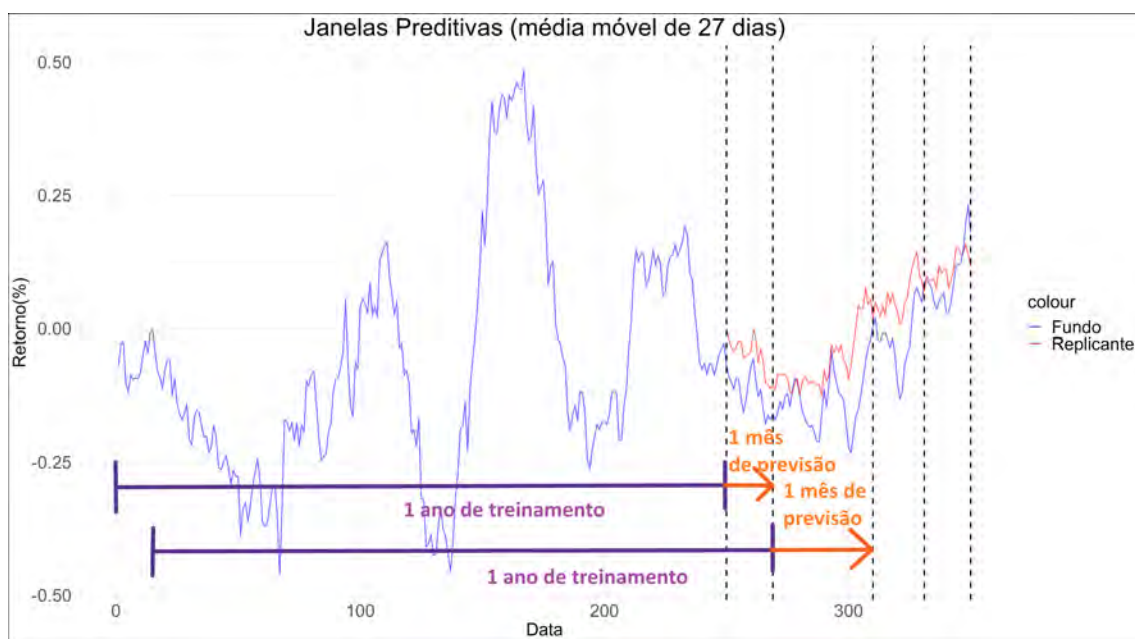
- $\text{ReplicantRet}_i^{(t)}$ é o retorno aritmético do clone do fundo de cobertura no tempo

-
- t ,
- $r_f^{(t)}$ é a taxa livre de risco no tempo t ,
 - $n^{(t)}$ é o número total de ETFs selecionados na regressão LASSO ou adaLASSO para o fundo de cobertura,
 - $\beta_j^{(t-1)}$ é o coeficiente do ETF j selecionado na janela de dois anos pela regressão LASSO ou adaLASSO,
 - $ETF_j^{(t)}$ é o retorno aritmético do ETF j no tempo t .

6.4 Janelas Preditivas e Rebalanceamento

Neste estudo, a metodologia adotada segue um processo de estimação e seleção de Exchange-Traded Funds (ETFs) relevantes baseado em um ano de retornos diários. Os coeficientes são então estimados para o próximo mês, seguindo uma estratégia de rebalanceamento mensal dos betas, que determinam a alocação em cada ETF. Por exemplo, para cada Hedge Fund, o modelo é estimado para fevereiro e treinado utilizando dados de fevereiro (do ano anterior) a janeiro. Posteriormente, o processo é repetido com treinamento de março a fevereiro do ano seguinte, estimando os coeficientes para março do ano seguinte. Esta abordagem permite capturar as dinâmicas de mercado ao longo do tempo e ajustar as alocações conforme as condições econômicas e financeiras evoluem. A metodologia é facilmente aplicável numa perspectiva de mercado financeiro, possibilitando que os betas sejam atualizados automaticamente conforme novos dados se tornam disponíveis, garantindo uma adaptação contínua às condições do mercado e baixo custo de manutenção. Esta abordagem é um dos pontos em que nos afastamos de Duanmu et al. (2019).

A imagem a seguir ilustra como foi aplicada esta metodologia de janelas preditivas e rebalanceamento dos coeficientes.



Nota: Nota-se que as janelas preditivas variam de tamanho pois utilizamos retornos diários de um mês e cada mês pode ter número diferente de dias úteis. Foi escolhida a média móvel de 27 dias (observações) porque 27 é a média de observações por mês.

7 Dados

Neste estudo, utilizamos dados de Hedge Funds e dados de ETFs. Os dados de Hedge Funds foram cedidos pela Quantum através de sua base de dados. No total, são analisados 147 fundos a partir do dia 2 janeiro de 2018 até 26 de abril de 2023.

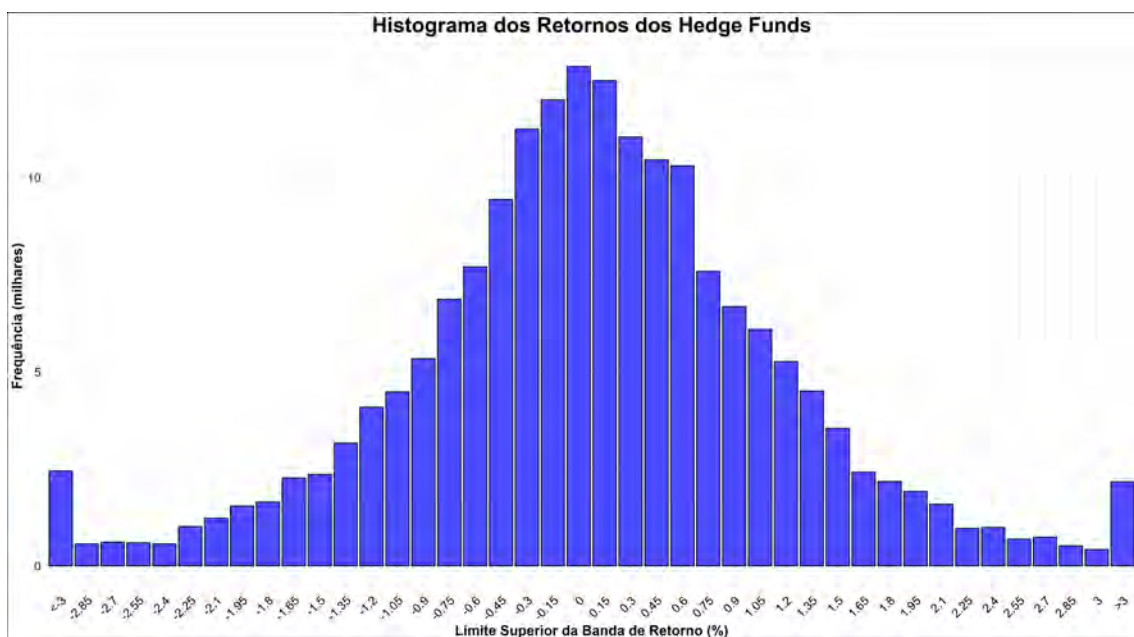
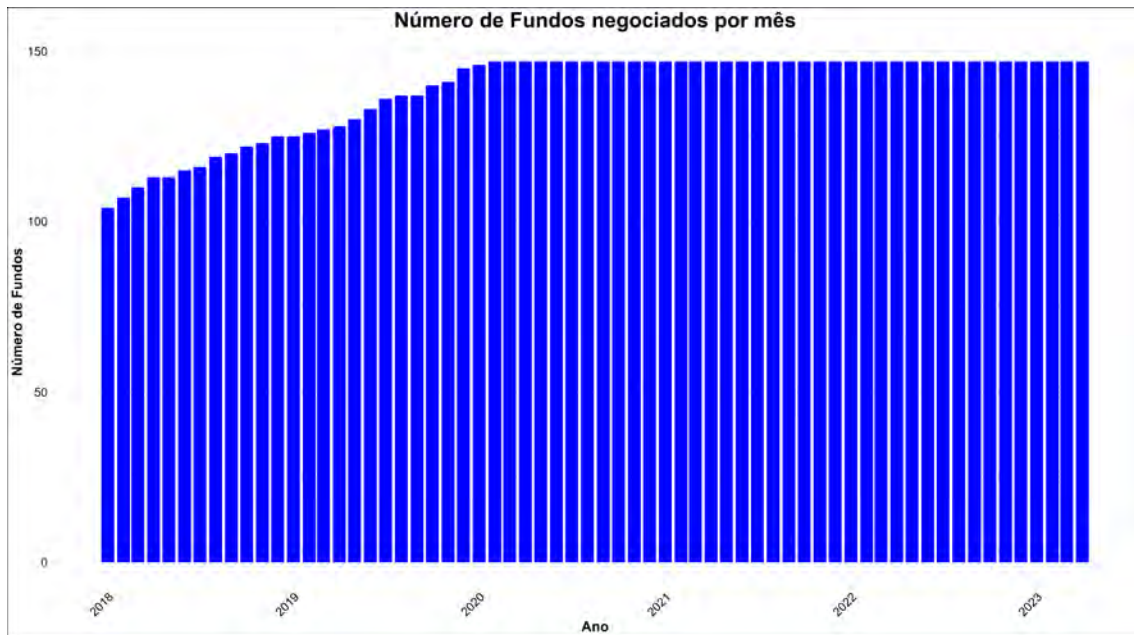
A tabela a seguir apresenta um resumo das estatísticas descritivas para os retornos aritméticos diários de Hedge Funds:

Resumo Estatístico Geral dos Hedge Funds	
Total de Observações	184,674
Total de Fundos	147
Data Inicial	2018-01-02
Data Final	2023-04-26
Retorno Médio (%)	0.514
Desvio Padrão do Retorno (%)	1.28
Retorno Mínimo (%)	-31.5
Retorno Máximo (%)	31.2
Mediana do Retorno (%)	0.288
Total de Retornos Positivos	92,597
Total de Retornos Negativos	92,077
Percentual de Retornos Positivos	50.1%
Percentual de Retornos Negativos	49.9%

Os dados apresentados na tabela evidenciam uma ampla variação nos retornos dos Hedge Funds durante o período analisado. O retorno médio foi ligeiramente positivo, indicando uma ligeira tendência de crescimento ou desempenho favorável em comparação ao esperado. A mediana do retorno também é positiva, refletindo que mais da metade dos fundos tiveram um desempenho levemente positivo. O desvio padrão alto sugere uma volatilidade significativa nos retornos, refletindo a natureza variável dos investimentos em Hedge Funds. A presença de retornos extremos, tanto negativos quanto positivos, destaca a diversidade de resultados obtidos pelos fundos no período de análise. Além disso, a proporção equilibrada entre retornos positivos e negativos sugere uma distribuição quase uniforme de resultados, com 50.1 por cento de retornos positivos e 49.9 por cento de retornos negativos.

A seguir temos dois gráficos. O primeiro deles mostra quantos Hedge Funds apresentam retorno para cada mês do horizonte de tempo analisado. O segundo apresenta um histograma dos retornos aritméticos diários de todos os Hedge Funds.

Além dos Hedge Funds, este estudo também analisa dados de Exchange-Traded



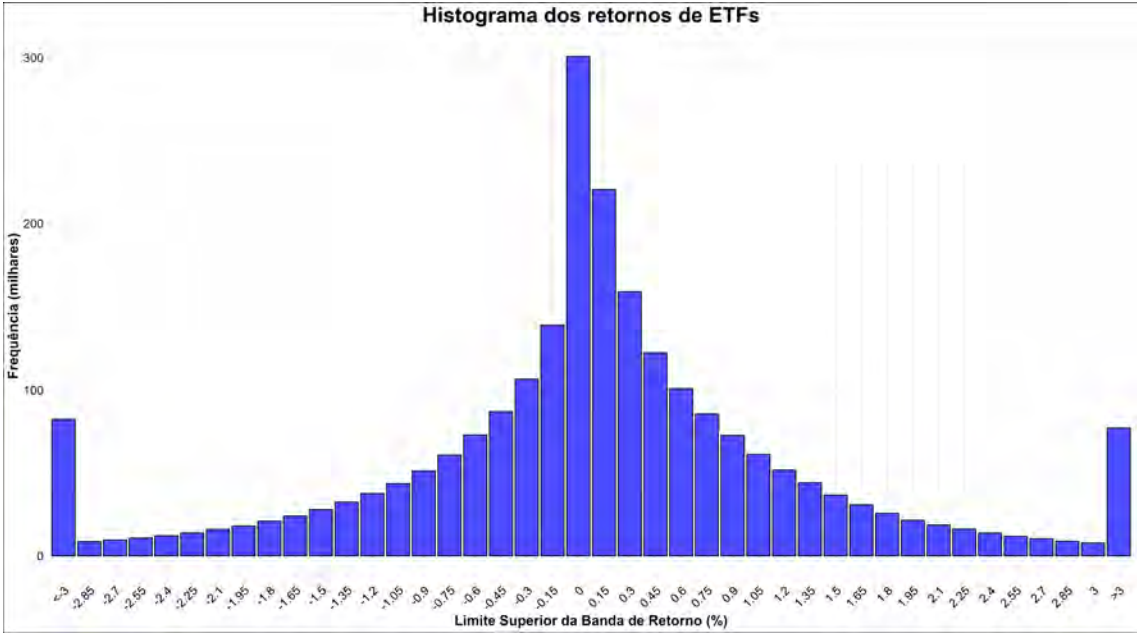
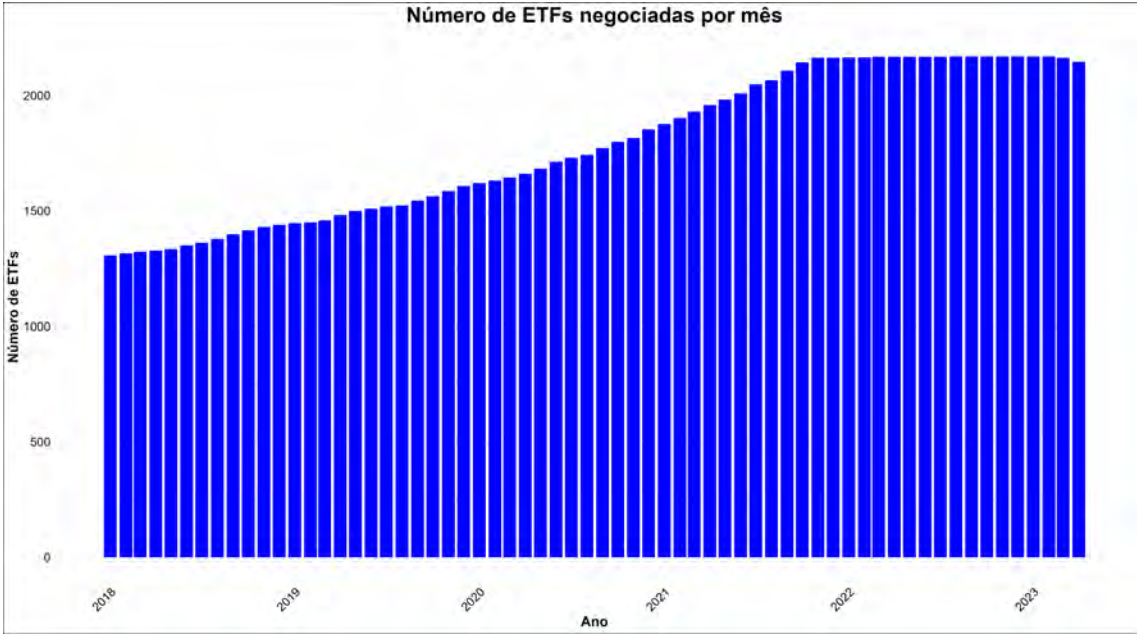
Funds (ETFs), que foram adquiridos do [FirstRateData.com](https://www.firstratedata.com/). Os ETFs abrangem um período similar ao dos Hedge Funds, desde o dia 2 de janeiro de 2018 até 26 de abril de 2023. Esses fundos são conhecidos por sua capacidade de oferecer diversificação instantânea e são amplamente utilizados tanto por investidores individuais quanto institucionais.

A tabela a seguir apresenta um resumo das estatísticas descritivas para os retornos aritméticos diários de ETFs analisados:

Resumo Estatístico Geral dos ETFs	
Total de Observações	2,373,482
Total de ETFs	2,169
Data Inicial	2018-01-02
Data Final	2023-04-26
Retorno Médio (%)	2.38
Desvio Padrão do Retorno (%)	6.49
Retorno Mínimo (%)	-89.8
Retorno Máximo (%)	9,533
Mediana do Retorno (%)	1.03
Total de Retornos Positivos	1,197,307
Total de Retornos Negativos	1,071,716
Percentual de Retornos Positivos	52.8%
Percentual de Retornos Negativos	47.2%

Os ETFs mostraram uma variabilidade significativa nos retornos durante o período analisado. Com um total de 2,373,482 observações, o conjunto de dados é substancialmente grande e inclui 2,169 ETFs diferentes. O retorno médio dos ETFs foi positivo em 0.0238, sugerindo uma ligeira tendência de crescimento no período analisado. A volatilidade, medida pelo desvio padrão, é relativamente alta em 6.49, indicando uma variação significativa nos retornos. Valores extremos nos retornos, com um mínimo de -89.8 e um máximo de 9,533, destacam a diversidade de resultados obtidos por diferentes ETFs. A mediana do retorno também é positiva em 0.0103, o que sugere que, em geral, os ETFs tiveram um desempenho favorável. Além disso, 52.8 por cento dos retornos foram positivos, enquanto 47.2 por cento foram negativos, refletindo uma leve predominância de resultados positivos no período.

A seguir temos dois gráficos. O primeiro deles mostra quantas ETFs apresentam retorno para cada mês do horizonte de tempo analisado. O segundo apresenta um histograma dos retornos aritméticos diários de todas as ETFs.



8 Resultados

As tabelas a seguir fornecem um resumo das métricas de desempenho dos modelos ao longo dos anos de 2019 a 2023. Esses resultados são fundamentais para avaliar a capacidade dos modelos em capturar e replicar características essenciais dos retornos dos fundos por meio da seleção estratégica de ETFs. As métricas apresentadas são médias de todas as estimativas mensais realizadas para todos os fundos, agregadas por ano e modelo, oferecendo uma visão abrangente e consolidada do desempenho dos modelos ao longo do tempo. Além delas, temos um scatter plot comparando o R^2 quando utilizamos o LASSO e o adaLASSO.

Médias das Estatísticas dos Fundos por Ano

Ano	Número de Fundos	Sharpe Ratio (%)	Sortino Ratio (%)
2019	125	-8.15	-6.16
2020	145	-12.5	3.06
2021	147	3.18	13.0
2022	147	-0.760	4.73
2023	147	-2.08	-6.68

Médias das Estatísticas por Modelo e Ano (Parte 1)

Modelo	Ano	Total de ETFs	ETFs Seleccionadas	R^2	R^2 Ajustado
adalasso	2019	1439	18.2	0.485	0.447
lasso	2019	1439	23.7	0.501	0.451
adalasso	2020	1607	19.0	0.532	0.496
lasso	2020	1607	25.5	0.562	0.514
adalasso	2021	1853	12.3	0.428	0.401
lasso	2021	1853	17.7	0.453	0.413
adalasso	2022	2164	13.1	0.494	0.467
lasso	2022	2164	16.0	0.498	0.465
adalasso	2023	2165	14.6	0.572	0.546
lasso	2023	2165	17.1	0.577	0.546

Nota: As estatísticas são calculadas a partir das médias de cada janela preditiva das previsões feitas em um ano para todos os Hedge Funds.

A análise comparativa entre os modelos *adalasso* e *lasso* revelou diferenças significativas em suas abordagens de seleção de variáveis e desempenho preditivo. O *adalasso* demonstrou uma propensão para selecionar um conjunto mais restrito de variáveis, indicando uma abordagem potencialmente mais conservadora na construção do modelo. Além disso, o *lasso* apresentou consistentemente valores mais altos de Erro Quadrático Médio (MSE) na previsão dos retornos dos fundos ao longo de todos os anos analisados, apesar de alcançar um R^2 superior. Este resultado sugere uma maior sensibilidade do *lasso* a sobreajustes ou ruídos nos dados. No gráfico,

Médias das Estatísticas por Modelo e Ano (Parte 2)

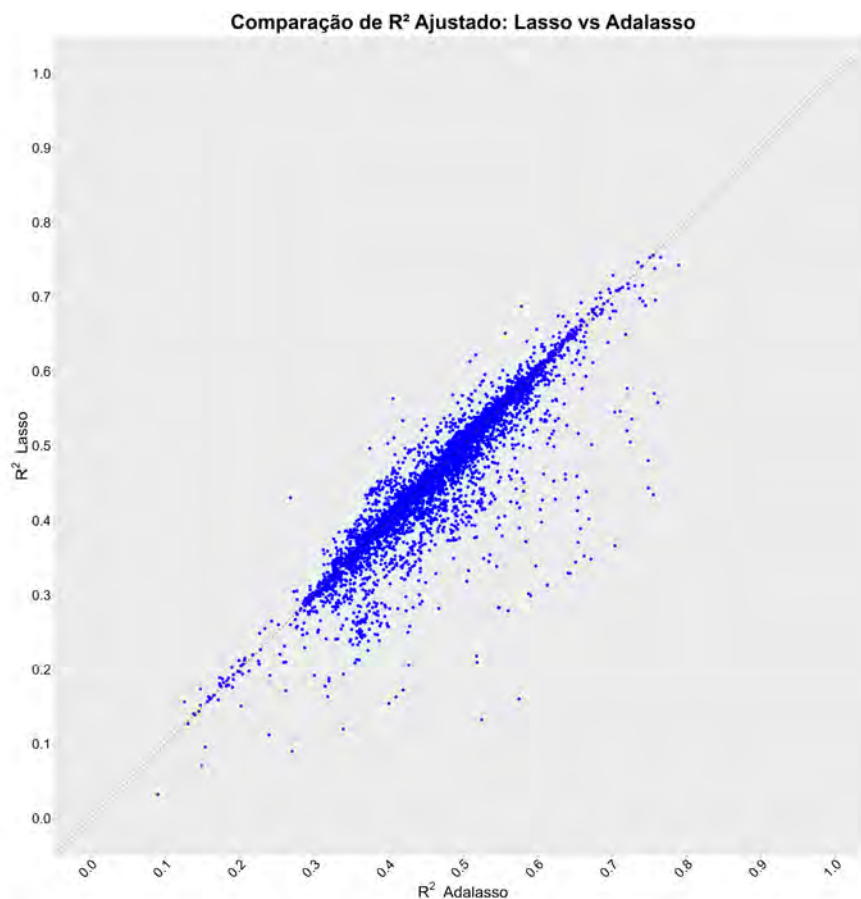
Modelo	Ano	Sharpe Ratio (%)	Sortino Ratio (%)	MSE	MAE	Mean Error
adalasso	2019	-1.92	3.47	0.302	0.408	0.0446
lasso	2019	-0.122	6.47	0.322	0.419	0.0520
adalasso	2020	6.32	18.4	1.20	0.731	0.113
lasso	2020	6.84	20.1	1.47	0.779	0.115
adalasso	2021	0.0113	4.31	0.595	0.568	0.000726
lasso	2021	-0.183	4.37	0.610	0.578	-0.00176
adalasso	2022	1.10	5.00	0.667	0.590	0.0295
lasso	2022	1.65	6.37	0.671	0.589	0.0296
adalasso	2023	-22.4	-23.3	0.528	0.522	-0.0721
lasso	2023	-23.4	-23.9	0.535	0.522	-0.0733

Nota: As estatísticas são calculadas a partir das médias de cada janela preditiva das previsões feitas em um ano para todos os Hedge Funds.

fica evidente que o R^2 ajustado do *adalasso* tende a ser maior, comportamento que não fica explícito na tabela de médias. Essas nuances ressaltam a importância da análise detalhada das métricas de desempenho e das estratégias de seleção de variáveis adotadas por cada modelo

Quando se analisa o desempenho dos modelos replicantes (*adalasso* e *lasso*) em relação aos fundos reais em termos de *Sharpe Ratio* e *Sortino Ratio*, observa-se uma disparidade interessante. Os modelos replicantes geralmente apresentaram índices positivos para essas métricas ao longo dos anos, embora modestos, enquanto os fundos reais consistentemente mostraram resultados negativos ou próximos a zero para ambos os índices. Isso sugere que os modelos replicantes, apesar de suas limitações, conseguiram gerar retornos (ajustados ao risco) superiores aos dos fundos reais. No entanto, é importante notar que esse padrão não se manteve no ano de 2023, onde os fundos reais apresentaram um desempenho significativamente superior aos replicantes. Essa reversão nos resultados merece uma análise posterior mais aprofundada, considerando possíveis mudanças nas condições de mercado ou outras variáveis que possam ter influenciado esse fenômeno. Além disso, vale ressaltar que os modelos replicantes não internalizam os custos de transação, o que é um fator contribuinte para seu melhor desempenho em condições normais.

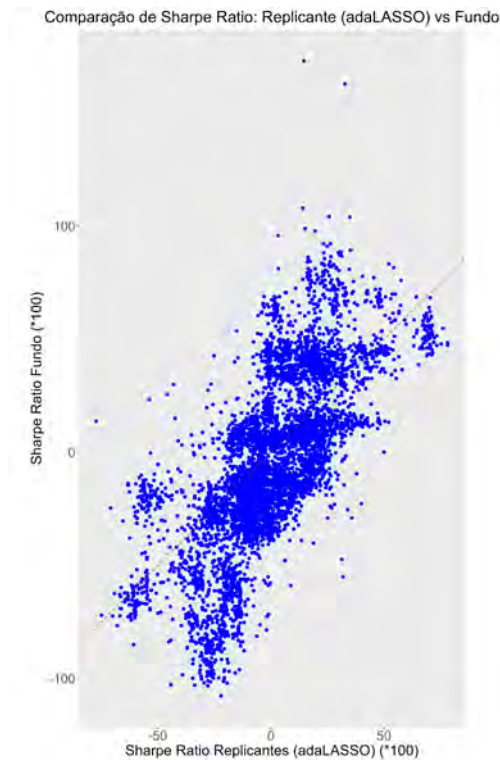
Para complementar, ambos os modelos *adalasso* e *lasso* exibiram um desempenho consistente na replicação dos retornos dos hedge funds, como evidenciado pelo Mean Error e MAE, que mostraram uma diferença média de menos de 1 por cento em relação aos retornos reais para todos os anos analisados. Esse resultado destaca a robustez das metodologias na previsão precisa dos retornos, minimizando discrepâncias significativas com os dados observados, o que é crucial para aplicações



financeiras que demandam precisão e confiabilidade nas projeções de desempenho.

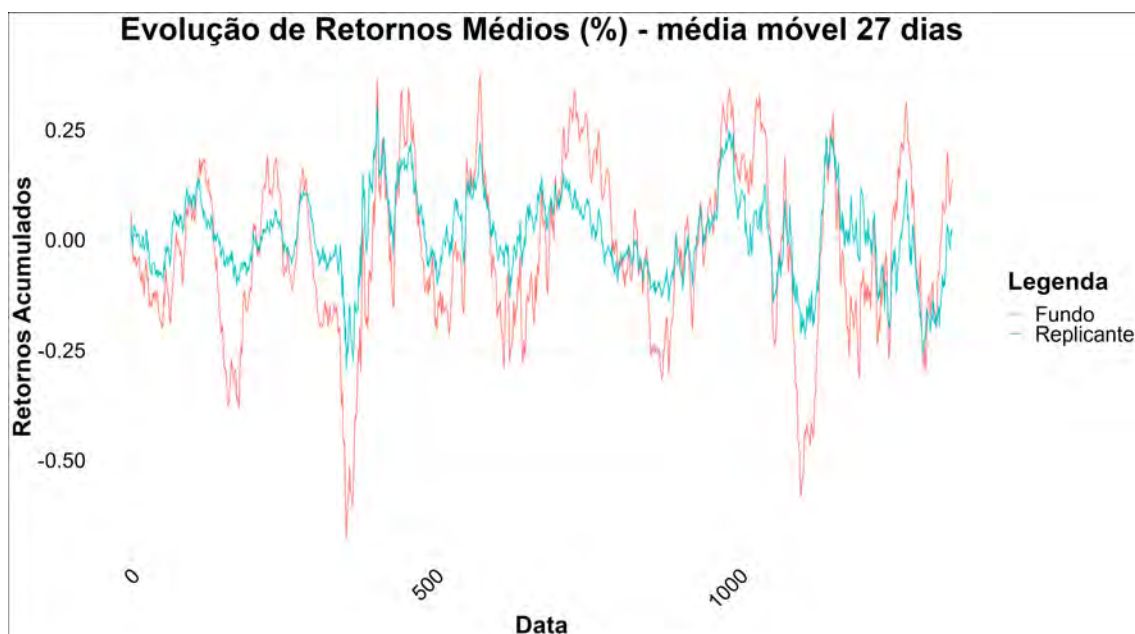
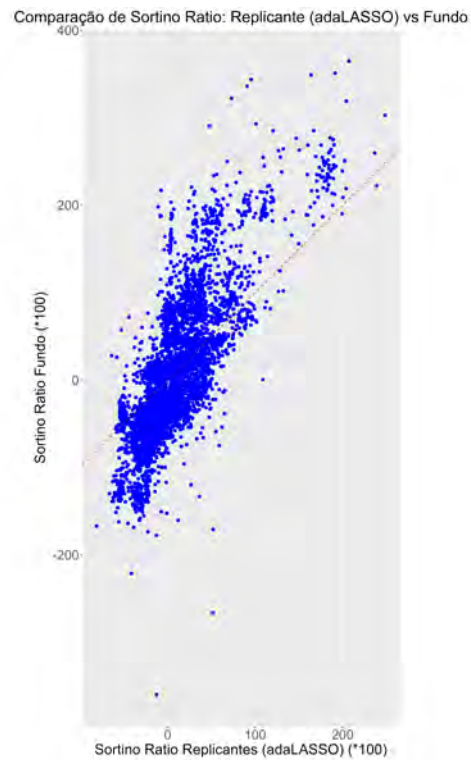
Adicionalmente, ao comparar os índices de Sharpe e Sortino entre os modelos replicantes e os fundos reais, observamos uma volatilidade consideravelmente menor nos replicantes. Em particular, os histogramas que comparam o Sharpe Ratio do *adalasso* com o dos fundos, bem como o Sortino Ratio do *adalasso* com o dos fundos, revelam que os replicantes apresentam uma menor variabilidade nas suas estatísticas de desempenho. Ao optar por escalas livres nos gráficos, conseguimos evidenciar a frequência maior de valores negativos nas estatísticas dos fundos reais em comparação com os replicantes, ressaltando a menor exposição ao risco dos replicantes. Essa constatação é especialmente relevante para investidores que buscam estratégias com menor volatilidade e risco ajustado.

Além disso, um gráfico que mostra a média móvel das médias dos retornos de todos os fundos em comparação com todas as previsões reforça a eficácia das metodologias de replicação utilizadas. Observamos que, de maneira geral, os replicantes conseguem acompanhar razoavelmente a performance dos fundos reais ao longo do tempo, indicando uma habilidade de previsão consistente dentro das janelas predi-



tivas consideradas. No entanto, quando analisamos os gráficos de retornos acumulados, que ilustram a evolução de um investimento hipotético de 1000 reais tanto nos fundos quanto nos replicantes, constatamos uma divergência significativa na performance acumulada. Especificamente, nos últimos quatro gráficos, observamos que, embora os modelos replicantes tenham um bom desempenho preditivo nas janelas temporais analisadas, a replicação dos retornos acumulados ao longo do tempo não apresenta a mesma eficácia.

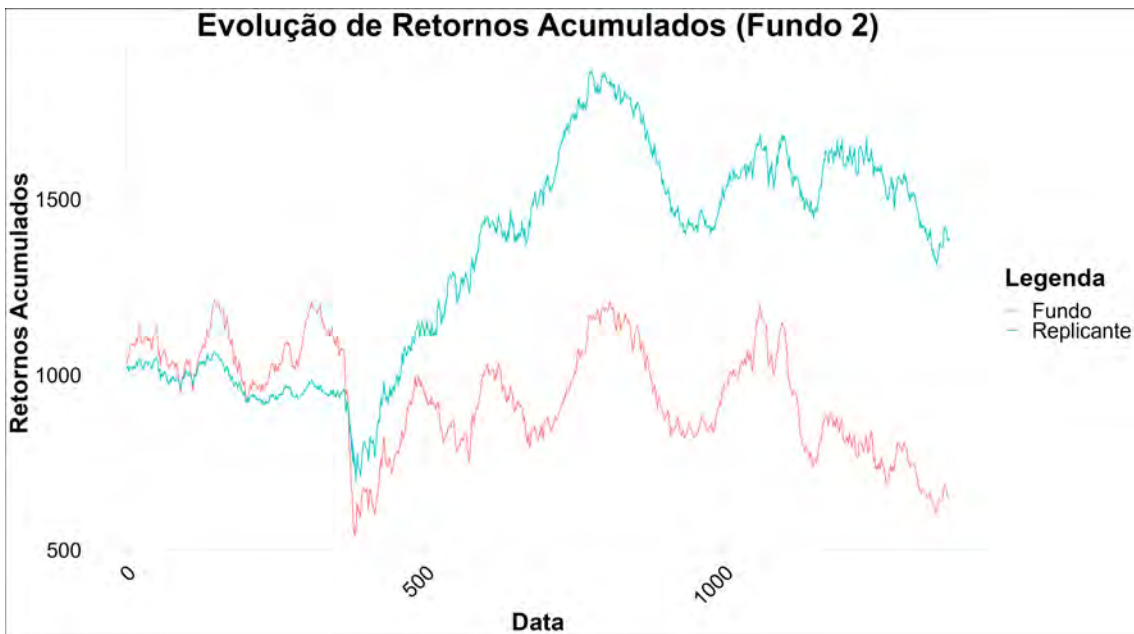
Esse descompasso pode ser atribuído à metodologia de utilização de janelas móveis de um ano para o treinamento dos modelos, que não incorpora todas as variáveis e eventos que impactam os dados históricos completos. Como resultado, os modelos podem falhar em capturar as dinâmicas de longo prazo dos fundos, levando a uma discrepância na replicação dos retornos acumulados. Esse ponto ressalta a importância de considerar diferentes abordagens de treinamento que possam integrar um horizonte temporal mais amplo, possibilitando uma replicação mais fiel e eficaz do desempenho acumulado dos fundos.



Nota: Foi escolhida a média móvel de 27 dias (observações) porque 27 é a média de observações por mês, dado que a base de Hedge Funds tem observações para os dias úteis brasileiros.



Nota: Exemplo de retorno acumulado para Fundo 1.



Nota: Exemplo de retorno acumulado para Fundo 2.



Nota: Exemplo de retorno acumulado para Fundo 3.

9 Conclusão

A replicação dos retornos de Hedge Funds brasileiros, utilizando a metodologia apresentada por Duanmu et al. (2019), apresentou resultados consistentes com os encontrados na literatura. Nesta pesquisa, construímos modelos de fatores de risco baseados em ETFs e empregamos análise de clusters para minimizar o impacto da colinearidade entre as variáveis. Além disso, utilizamos métodos de seleção de variáveis por meio de técnicas de Machine Learning, especificamente o LASSO e o adaLASSO. Essa abordagem foi aplicada em um framework adaptado para a utilização no mercado financeiro.

Apesar de nossa metodologia ter utilizado retornos diários em vez de mensais e um período de um ano de treinamento para prever um mês, em contraste com o período de dois anos de treinamento para prever um ano adotado por Duanmu et al. (2019), os modelos apresentaram um R^2 ajustado semelhante. Isso pode ser atribuído ao fato de que, ao considerar retornos diários, o modelo captura mais fielmente os altos riscos a que os Hedge Funds estão expostos. Além disso, a performance dos modelos replicantes em relação ao Sharpe Ratio e ao Sortino Ratio foi superior à dos Hedge Funds, corroborando os resultados encontrados por Duanmu et al. (2019).

A principal contribuição desta pesquisa reside na comparação entre os desempenhos do LASSO e do adaLASSO. Observou-se que o adaLASSO teve um desempenho superior em termos de Erro Quadrático Médio (MSE) e R^2 ajustado. Essa superioridade pode ser explicada pela natureza mais conservadora do adaLASSO na seleção de variáveis, o que reduz o risco de overfitting. Além disso, a condição de consistência do adaLASSO ("Weighted Irrepresentable Condition"), que é menos restritiva, lida melhor com a colinearidade entre as variáveis explicativas.

Uma das limitações da metodologia escolhida foi a utilização de janelas móveis de um ano para o treinamento dos modelos, o que pode ter levado a uma performance fraca dos replicantes em relação ao retorno acumulado quando comparados aos fundos reais. Essa abordagem não considera todas as variáveis e eventos de longo prazo que impactam os dados históricos completos, resultando em uma replicação menos eficaz dos retornos acumulados dos fundos. Isso destaca a necessidade de explorar diferentes metodologias que integrem um horizonte temporal mais amplo para uma replicação mais fiel do desempenho dos fundos ao longo do tempo.

Como possíveis extensões deste estudo, sugere-se a incorporação dos custos de transação para tornar o modelo mais aplicável na prática, a automação do rebalanceamento dos coeficientes com base em métricas de ajuste do modelo e/ou mudanças nas características dos Hedge Funds, e a inclusão de outros fatores de risco no modelo.

Bibliografia

- Amenc, N., Martellini, L., Meyfredi, J.-C., & Ziemann, V. (2010). Passive hedge fund replication - beyond the linear case. *EUROPEAN FINANCIAL MANAGEMENT*, 16(2), 191–210. <<https://doi.org/10.1111/j.1468-036X.2008.00448.x>>
- Bali, T. G., Brown, S. J., & Caglayan, M. O. (2011). Do hedge funds' exposures to risk factors predict their future returns? *Journal of Financial Economics*, 101, 36–68.
- Bali, T. G., Brown, S. J., & Caglayan, M. O. (2012). Systematic risk and the cross section of hedge fund returns. *Journal of Financial Economics*, 106, 114–131.
- Bali, T. G., Brown, S. J., & Caglayan, M. O. (2014). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, 114, 1–19.
- Bollen, N. P. B. (2013). Zero-r2 hedge funds and market neutrality. *Journal of Financial and Quantitative Analysis*, 48, 519–547.
- Duanmu, J., Li, Y., & Malakhov, A. (2019). Capturing hedge fund risk factor exposures: Hedge fund return replication with etfs. *FINANCIAL REVIEW*, 55(3), 405–431. <<https://doi.org/10.1111/fire.12221>>
- Epprecht, C., Guegan, D., Veiga, Á., & Correa da Rosa, J. (2017). Variable selection and forecasting via automated methods for linear models: Lasso/adalasso and autometrics [Version originale Novembre 2013, révisée en Octobre 2017]. *Documents de travail du Centre d'Economie de la Sorbonne*, 2013.80R (ISSN: 1955-611X). <<https://shs.hal.science/halshs-00917797>>
- Fung, W., & Hsieh, D. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *REVIEW OF FINANCIAL STUDIES*, 14(2), 313–341. <<https://doi.org/10.1093/rfs/14.2.313>>
- Fung, W., & Hsieh, D. (2004). Hedge fund benchmarks: A risk-based approach. *FINANCIAL ANALYSTS JOURNAL*, 60(5), 65–80. <<https://doi.org/10.2469/faj.v60.n5.2657>>
- Getmansky, M., Lo, A. W., & Makarov, I. (2004). An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, 74, 529–609.
- Giamouridis, D., & Paterlini, S. (2010). Regular(ized) hedge fund clones. *Journal of Financial Research*, 33(3), 223–247. <<https://ssrn.com/abstract=1500325>>

-
- Hasanhodzic, J., & Lo, A. W. (2007). Can hedge-fund returns be replicated?: The linear case. *Journal of Investment Management*, 5(2). <<https://ssrn.com/abstract=989605>>
- Roncalli, T., & Weisang, G. (2009). Tracking problems, hedge fund replication and alternative beta. <<https://ssrn.com/abstract=1325190>>
- Sharpe, W. (1994). The sharpe ratio. *Journal of Portfolio Management*, 21(1).
- Sharpe, W. F. (1992). Asset allocation: Management style and performance measurement. *Journal of Portfolio Management*, 18(2). <<https://www.pm-research.com/content/ijpormgmt/18/2/7>>
- Sortino, F. (1994). Performance measurement in a downside risk framework. *Journal of Investing*, 3, 59–65. <<https://doi.org/10.3905/joi.3.3.59>>
- Sun, Z., Wang, A., & Zheng, L. (2012). The road less traveled: Strategy distinctiveness and hedge fund performance. *Review of Financial Studies*, 25, 96–143.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267–288. Retrieved January 23, 2024, from <<http://www.jstor.org/stable/2346178>>
- Titman, S., & Tiu, C. (2011). Do the best hedge funds hedge? *Review of Financial Studies*, 24, 123–168.
- Weber, V., & Peres, F. (2013). Hedge fund replication: Putting the pieces together. <<https://ssrn.com/abstract=2202270>>
- Zhao, P., & Yu, B. (2006). On model selection consistency of lasso. *Journal of Machine Learning Research*, 7(90), 2541–2563. <<http://jmlr.org/papers/v7/zhao06a.html>>
- Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429. <<https://doi.org/10.1198/016214506000000735>>