

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto
Departamento de Economia
Programa de Pós-graduação em Economia - Área: Economia
Aplicada

Flávio Augusto Bassi Vieira

Detecção de falsas estratégias de investimento. Uma análise do
mercado brasileiro através de FWER

Orientador: Prof. Dr. Márcio Polleti Laurini.

Ribeirão Preto

2022

Prof. Dr. Carlos Gilberto Carlotti Junior

Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa

Diretor da Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto

Profa. Dra. Roseli da Silva

Chefe do Departamento de Economia

Prof. Dr. Luciano Nakabashi

Coordenador do Programa de Pós-Graduação em Economia - Área:
Economia Aplicada

FLÁVIO VIEIRA

Detecção de falsas estratégias de investimento. Uma análise do
mercado brasileiro através de FWER

Dissertação de Mestrado submetida ao Pro-
grama de Pós-Graduação em Economia da
Faculdade de Economia, Administração e
Contabilidade de Ribeirão Preto da Univer-
sidade de São Paulo

Orientador: Prof. Dr. Márcio Polleti Laurini.

Ribeirão Preto

2022

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Vieira, Flávio
Detecção de falsas estratégias de investimento. Uma análise do mercado brasileiro através de FWER/ Flávio Augusto Bassi Vieira; Orientador: Prof. Dr. Márcio Polleti Laurini.
Ribeirão Preto, 2022- 61 p. : il.

Dissertação (Mestrado) – Universidade de São Paulo, 2022.

CDU

Agradecimentos

Agradecimentos. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Detecção de falsas estratégias de investimento. Uma análise do mercado brasileiro através de FWER

Resumo

O presente trabalho empregou o método proposto por López de Prado and Lewis (2019) combinado com o estudo de Harvey and Liu (2020) com o intuito de testar a detecção de falsas estratégias de investimento. A metodologia parte da análise dos erros Tipo I e Tipo II sobre múltiplos testes através de FWER com correção de Sidak. Para a validação do método proposto foi construído três cenários em ambiente controlado com o auxílio de movimentos brownianos geométricos além de ser feito testes empíricos com 78 ativos que compuseram o índice Bovespa (IBOV) durante o período de outubro de 2015 até janeiro de 2021 para testar as estratégias de Cruzamento de Médias Móveis, Reversão à Média e Momentum, onde cada estratégia teve um grupo de dez parâmetros diferentes. Os resultados em ambiente controlado comprovam a eficácia do modelo ao detectar que os cenários construídos apenas com resultados aleatórios e maus resultados são falsas estratégias, enquanto o cenário composto apenas por bons resultados foi tido como uma estratégia verdadeira. Em relação aos resultados empíricos, cada estratégia obteve parâmetros que foram dados como estratégias verdadeiras quanto falsas, mostrando que não há uma estratégia que seja falsa em todos os parâmetros, porém houve um consenso entre as três estratégias de serem aceitas como verdadeiras em parâmetros menos sensíveis à volatilidade.

Palavras-chaves: Erro Tipo I, Erro Tipo II, Falsas Estratégias de Investimento, FWER

Detection of false investment strategies. An analysis of the Brazilian market through FWER

Abstract

The present work used the method proposed by López de Prado and Lewis (2019) combined with the study of Harvey and Liu (2020) for the purpose of testing the detection of false investment strategies. The methodology starts from the analysis of Type I and Type II errors over multiple tests through FWER with Sidak correction. For the validation of the proposed method, three scenarios were built in a controlled environment with the aid of geometric Brownian motions, in addition to empirical tests with 78 assets that made up the Bovespa index (IBOV) during the period from October 2015 to January 2021 to test the Moving Averages Crossing, Mean Reversion and Momentum strategies, where each strategy had a group of ten different parameters. The results in a controlled environment prove the effectiveness of the model in detecting that the scenarios built only with random results and bad results are false strategies, while the scenario composed only of good results was considered a true strategy. Regarding the empirical results, each strategy obtained parameters that were given as true and false strategies, showing that there is no strategy that is false in all parameters, but there was a consensus between the three strategies to be accepted as true in parameters less sensitive to volatility.

Key-words: Type I Error, Type II Error, False Investment Strategies, FWER

Lista de ilustrações

Figura 1 – Evolução do índice Bovespa (IBOV) entre outubro de 2015 a janeiro de 2021	20
Figura 2 – Distribuição dos Sharpe Ratio de Cruzamento de Médias Móveis sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*	55
Figura 3 – Distriuição dos Sharpe Ratio de Cruzamento de Médias Móveis com Custo - Destacando a posição do Sharpe Ratio Mediano SR*	56
Figura 4 – Distribuição dos Sharpe Ratio de Momentum sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*	57
Figura 5 – Distribuição dos Sharpe Ratio de Momentum com Custo - Destacando a posição do Sharpe Ratio Mediano SR*	58
Figura 6 – Distribuição dos Sharpe Ratio de Reversão à Média sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*	59
Figura 7 – Distribuição dos Sharpe Ratio de Reversão à Média com Custo - Destacando a posição do Sharpe Ratio Mediano SR*	60

Sumário

	Sumário	9
1	INTRODUÇÃO	11
2	DADOS	18
3	METODOLOGIA	22
3.1	Sharpe Ratio	22
3.1.1	Distribuição Assintótica	23
3.1.2	Familywise Error Rate (FWER)	24
3.2	Correção de Sidak	25
3.3	Clusterização	25
3.4	Erro Tipo I Sobre Múltiplos Testes	27
3.5	Erro Tipo II Sobre Múltiplos Testes	29
4	RESULTADOS	31
4.1	Resultados em Ambiente Controlado	31
4.2	Resultados Empíricos	33
4.2.1	Cruzamento de Médias Móveis	33
4.2.2	Momentum Time Series	35
4.2.3	Reversão à Média	37
5	CONSIDERAÇÕES FINAIS	39
6	CONCLUSÃO	43

Referências	45
APÊNDICES	48
APÊNDICE A – FATORES DE RISCO	49
A.1 HML Factor	49
A.2 IML Factor	50
A.3 Market Factor	51
A.4 SMB Factor	51
A.5 WML Factor	52
A.6 Estimação dos Parâmetros	52
APÊNDICE B – DISTRIBUIÇÃO DOS SHARPE RATIO	54

1 Introdução

O evento de investir traz ao investidor o compromisso de reter seus recursos no período atual com a expectativa de colher benefícios futuros, e por sua vez, uma operação de investimento é aquela que, por meio da análise, promete uma segurança para o principal e um retorno adequado. As operações que não vão ao encontro dessas exigências são especulativas. Isso é, uma das características de um investimento é a segurança. É ponderado todo o risco da operação, considera-se quais as possibilidades desse investimento ser bem sucedido para aí sim ser tomada uma decisão. Em contraparte, ao investir sem ter os pilares acima, estamos lidando com um ambiente de incerteza e que torna o investimento em um movimento especulativo. Nesse caso, o risco é muito maior já que no momento da compra, o especulador nunca tem certeza de que o ativo vai mesmo se valorizar.

Devido a sua alta volatilidade, o mercado de ações é marcado por sempre ter tido a atenção de pesquisadores e investidores que buscam, através de métodos e análises, fazer avaliações das empresas listadas na bolsa em troca de melhorarem seus retornos e performance de suas carteiras de investimento. Outro ponto abordado é a tentativa de reduzir a incerteza de suas operações, isso é, além de tentar maximizar os resultados de suas operações, o investidor tentará minimizar os riscos. O que o tiraria de um ambiente especulativo para um ambiente de investimento

sustentável.

Para isso, um dos métodos mais utilizados são as análises quantitativas, que se baseiam em modelos matemáticos e estatísticos para definição de suas operações, para que esses modelos sejam colocados em práticas eles são testados exaustivamente em busca de evidências de que aquela interpretação da informação resulta em uma estratégia que faça sentido e seja verdadeira. O ponto chave da análise quantitativa é basear-se em evidências, tentando retirar as discricionariedades do processo de investimento.

Uma das vertentes das análises quantitativas é a análise técnica, a qual utiliza representações gráficas do histórico de preços das ações, fórmulas matemáticas simples além de dados de volume e estatísticas. As negociações baseadas em análises técnicas são associadas a anomalias de precificação dos ativos, como momentum e reversão. Moskowitz et al. (2012) define que estratégias baseadas em time series momentum os ativos são comprados (vendidos) caso tiveram uma boa performance no passado (má performance no passado). Para considerar se a performance do ativo no passado é boa ou ruim, é considerado uma janela de um determinado número de dias e retirado sua média, caso essa for positiva (negativa) é visto como uma boa performance (má performance). Enquanto Balvers et al. (2000) descreve que a estratégia de reversão a média refere-se a uma tendência dos preços dos ativos de retornar a um caminho de tendência. Grosso modo, as estratégias de reversão à média

dependem de um raciocínio que é o oposto dos retornos de momento, as estratégias de reversão à média assumem uma correlação negativa. Além de ter as regras de negociação envolvendo o cruzamento de médias móveis pode ser definida usando dois períodos para os parâmetros A e B , sendo $A < B$. Considere que os sinais dessa estratégia são dados em $MA_A - MA_B$, onde x é o número de dias considerado para ser feita a média. Para essa regra é utilizado como pontos de entrada em negociações tanto sinais do tipo $A < B$ quanto do tipo $A > B$, isso é, esta estratégia é baseada como compra (venda) de um ativo quando a média móvel de curto período sobe acima (cai abaixo) da média móvel de longo período. Brock et al. (1992) considera que nesses casos é iniciado um ciclo de tendência de alta (baixa).

Estratégias baseadas nesse tipo de análise tem sido usada e explorada amplamente pelos players do mercado e isso gerou e ainda gera discussões a respeito se ela realmente tem o poder de aumentar a rentabilidade de forma significativa, sustentável e segura. Para que esses modelos sejam colocados em práticas eles são testados exaustivamente em busca de evidências de que aquela interpretação da informação resulta em uma estratégia que faça sentido e seja verdadeira. O ponto chave é basear-se em evidências, tentando retirar as discricionariedades do processo de investimento.

Ao dizer que esses modelos são testados exaustivamente quer dizer que eles passam por um processo de ser colocado à prova em um ambiente controlado e observar como ele se comporta.

Uma das ferramentas que permitem a condução de tais experimentos é o backtest, o qual oportuniza a simulação de como uma estratégia de investimentos teria performado sobre um cenário histórico, isso é, backtests são uma simulação histórica de como uma estratégia teria se saído no passado e, portanto, a ideia é que a performance histórica guie a performance futura.

Porém, não há a possibilidade de alterarmos a estrutura das variáveis e a partir disso criarmos novas séries históricas de tempo onde realizar um backtest independente seria possível. Em outras palavras, não estamos lidando com um experimento controlável.

Como resultado, esses métodos de análise de estratégias não podem nos ajudar a derivar os mecanismos de causa e efeito precisos que tornam uma estratégia bem sucedida. Segundo Bailey et al. (2014) um pesquisador pode, facilmente, causar overfit em um backtest conduzindo várias simulações históricas e selecionando a estratégia de melhor desempenho.

A ocorrência de um backtest apresentar overfitting ocorre quando um pesquisador faz uma descoberta falsa (encontra um falso positivo) como resultado da seleção do melhor resultado de uma multiplicidade de backtests conduzidos no mesmo conjunto de dados. Uma falsa descoberta ocorre quando um teste estatístico rejeita uma hipótese nula verdadeira, ocasionando um erro Tipo I. A probabilidade de obtenção de um falso positivo é definida pelo nível de significância. (Nesse trabalho é utilizado 5%)

Bailey and López de Prado (2012) introduziram o teorema de

falsas estratégias e demonstram como um Sharpe Ratio estimado pode ser usado para rejeitar falsas descobertas em retornos não normais enquanto há controle de viés de seleção sob testes múltiplos (SBuMT). Lopez de Prado (2020) provém uma metodologia prática que nos permite calcular e relatar a probabilidade de que uma estratégia de investimento seja um falso positivo, enquanto controla o viés de seleção em testes múltiplos (SBuMT) utilizando a metodologia Familywise Error Rate (FWER) para caracterizar um erro Tipo I.

Além de caracterizar o erro Tipo I, é associado a taxa do erro Tipo II à nossa taxa otimizada do erro Tipo I. Harvey and Liu (2020) cita que em um teste de hipótese única, a taxa de erro Tipo II em um determinado valor de parâmetro é calculada como a probabilidade de não rejeitar a hipótese nula nesse valor. Em testes múltiplos, o cálculo da taxa de erro Tipo II é menos direto porque, em vez de um valor de parâmetro único, precisamos especificar um vetor de parâmetros diferentes de zero, onde cada parâmetro corresponde a um único teste sob a hipótese alternativa.

Harvey and Liu (2020) argumenta que fazer a análise do erro Tipo II associada à taxa otimizada do erro Tipo I contribui em três pontos. Primeiro, apresenta uma estrutura que oferece uma definição intuitiva de poder de teste. Segundo estimar de forma flexível o poder do teste. E, por fim, apresenta como o poder do teste pode mudar materialmente nossa interpretação de importantes descobertas de pesquisa.

No contexto apresentado, empregamos o método proposto por López de Prado and Lewis (2019) combinado com o estudo de Harvey and Liu (2020) para realizar várias verificações de robustez e fornecer respostas em relação as estratégias de momentum, reversão a média e de cruzamento de médias móveis se realmente são estratégias de investimentos válidas ou se são falsas estratégias, isso em um ambiente com e sem custos de transação, além de testar estratégias criadas em ambiente controlado com o auxílio de movimentos brownianos geométricos para ser feita a validação da metodologia proposta. De forma detalhada, é gerado um conjunto de experimentos de cada estratégia e feito a clusterização desse conjunto para identificar qual o número de experimentos independentes dentro desse conjunto. É encontrado o sharpe ratio máximo além de ser calculado sua assimetria e curtose e para fins de comparação é utilizado o sharpe ratio mediano. E então é utilizada a metodologia de Familywise Error com a correção de Sidak para encontrar a probabilidade de um Erro tipo I sobre múltiplos testes, isso é, testar se a estratégia de fato é uma estratégia válida e por fim, derivar o Erro tipo II sobre múltiplos testes e consequentemente encontrar o poder do teste.

Para o exercício de validação da metodologia foi utilizado movimento browniano geométrico para realizar três ensaios cada um contendo 80 experimentos com um T igual a 2 mil dias de retorno, isso é, é feito a simulação de três estratégias diferentes onde é feito o backtest para 80 ativos diferentes com 2 mil dias de

retornos contados. Em relação aos resultados empíricos, é testado as três estratégias propostas utilizando as ações que compuseram o Índice Bovespa (IBOV) da B3 e que tiverem histórico de preços no período compreendido entre outubro de 2015 até janeiro de 2021. Dessa forma, foi considerado 1308 dias de retornos além de 78 ativos, isso é, 78 experimentos. Na seção a seguir será feito um melhor detalhamento a respeito dos dados utilizados

Em relação aos resultados em ambiente controlado, foi testado três cenários, sendo o primeiro composto apenas com experimentos bem sucedidos, o segundo contou com experimentos aleatórios e o terceiro apenas com experimentos mal sucedidos e o terceiro. Encontramos evidências empíricas de que a metodologia proposta por López de Prado and Lewis (2019) é válida ao mostrar que a estratégia do primeiro cenário foi tida como uma estratégia verdadeira e a do terceiro cenário apresentou um alto α_K , sendo uma falsa estratégia. O segundo cenário, por ter sido formado de forma aleatória, foi também dado como uma falsa estratégia.

O restante desse trabalho é apresentado a seguir. A segunda seção expõe o conjunto de dados considerado e apresenta as regras técnicas de negociação utilizadas. A terceira seção faz diz respeito a metodologia proposta, onde é explicado cada integrante e passo que é dado para a construção da análise. Os resultados empíricos para cada estratégia são apresentados na quarta seção e, por fim, a quinta seção traz as conclusões e considerações finais.

2 Dados

No sentido apresentado acima, este trabalho propõe um exercício principal, que consiste em testar se estratégias de Cruzamento de Médias Móveis, Reversão a Média e Momentum são de fato estratégias ou são falsas estratégias. Além de um exercício de validação da metodologia proposta através da geração de experimentos utilizando movimento browniano geométrico. Esta seção apresentará os dados necessários para a realização de ambos os exercícios.

Para ser feito o exercício de validação da metodologia proposta utilizando movimento browniano geométrico foi realizado três ensaios diferentes, cada um contendo 80 experimentos com um T igual a 2 mil dias de retorno. O primeiro ensaio foi selecionado apenas com retornos positivos, isso é, foi feita a seleção de 80 experimentos relacionados a mesma estratégia fictícia que foram muito bem sucedidos. O segundo ensaio, não foi selecionado um modelo ideal de retorno e sim, contou com total aleatoriedade para simular uma estratégia que teria retornos positivos em determinado ativo e retornos negativos em outro determinado ativo. No terceiro e último cenário, foi selecionado 80 experimentos que tiveram retornos negativos para simular uma estratégia que falhou em todos os ativos testados. Dados esses fatos, é esperado que a metodologia reconheça que o primeiro seja de fato uma estratégia, enquanto o terceiro cenário seja uma falsa

estratégia e que, o segundo cenário, seja avaliado pelos resultados.

Após a validação do modelo através do primeiro exercício, é testado as três estratégias propostas utilizando as seguintes ações:

ABEV3, ALPA4, ALSO3, AMAR3, B3SA3, BBAS3, BBDC3, BBDC4, BBSE3, BEEF3, BRAP4, BRFS3, BRKM5, BRML3, CCRO3, CESP6, CIEL3, CMIG4, COGN3, CPFE3, CPLE6, CSAN3, CSMG3, CSNA3, CVCB3, CYRE3, ECOR3, EGIE3, ELET3, ELET6, EMBR3, ENBR3, ENEV3, EQTL3, EZTC3, FLRY3, GGBR4, GOAU4, GOLL4, HYPE3, ITSA4, ITUB4, JBSS3, JHSF3, KLBN11, LAME3, LAME4, LCAM3, LIGT3, LREN3, MDIA3, MEAL3, MGLU3, MRFG3, MRVE3, MULT3, PETR3, PETR4, PRIO3, PSSA3, QUAL3, RADL3, RAIL3, RAPT4, RENT3, SANB11, SBSP3, SULA11, TAEE11, TIMS3, TOTS3, TRPL4, UGPA3, USIM5, VALE3, VIVT3, WEGE3, YDUQ3.

As ações foram escolhidas a partir da composição do índice Bovespa entre o período de outubro de 2015 a janeiro de 2021, contabilizando um total de 1308 dias de retornos analisados. Além de que foi excluído do trabalho as ações que tinham mais de 40% de *missing values* no período analisado, isso é, caso uma ação não tenha participado do IBOV por pelo menos 60% do período analisado, ela será excluída do trabalho. Assim como mostra a figura 1 abaixo, o período escolhido para análise foi marcado por uma tendência de alta no índice Bovespa até o rompimento devido a crise da COVID-19 e sua recuperação em

seguida. Mesmo em um ciclo de alta, o período apresenta movimentos contra cíclicos que corroboram para a mudança de sinal em uma operação em estratégias com parâmetros mais sensíveis à volatilidade. Portanto a janela analisada trás um ambiente em que, tanto estratégias muito sensíveis à volatilidade quanto aquelas menos sensíveis irão sofrer mudanças de sinal, excluindo qualquer parâmetro que poderia ser confundido com uma estratégia *Buy and Hold*.



Figura 1 – Evolução do índice Bovespa (IBOV) entre outubro de 2015 a janeiro de 2021

Dessa forma, cada estratégia é testada levando em consideração dez diferentes parâmetros, onde o objetivo é analisar como cada estratégia performa com diferentes níveis de sensibilidade à volatilidade. Portanto, para a estratégia de Cruzamento de Médias Móveis serão utilizados os seguintes parâmetros de médias móveis: 5-10, 5-22, 5-30, 7-15, 7-30, 7-45, 10-22, 10-50, 15-30, 15-90. Para a estratégia de Momentum Time Series são utilizadas as seguintes médias: 1, 3, 5, 7, 9, 11, 13, 15, 17 e 20. Por fim, para a estratégia de Reversão à Média, são utilizadas as seguintes médias: 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50.

Para completar, é inserido ao trabalho um custo de transação de 0,1% por cada vez que uma operação sofre uma alteração, por exemplo, caso uma estratégia que esta comprada em determinado ativo dê um sinal de venda, será cobrado 0,1% para inverter a posição e esse valor será descontado da rentabilidade que a estratégia estava apresentando aquele dia, ou seja, caso houver um lucro de 1,0% no dia da troca, o valor correto do lucro será de 0,9% e caso houver um prejuízo de 1,0%, o prejuízo passará a ser de 1,1%.

3 Metodologia

Nessa seção é exposto o setup utilizado para a análise de falsas estratégias de investimento através do uso de Familywise error.

Bailey and López de Prado (2012) demonstraram como um Sharpe Ratio estimado pode ser usado para rejeitar falsas descobertas sobre retornos não normais controlados por SBuMT (viés de seleção sobre múltiplos testes). A primeira parte da seção é explicado a teoria por trás do Sharpe Ratio, depois é passado pela teoria de familywise error e explicado o motivo pelo seu uso, em seguida é explicado qual método de clusterização é utilizado para encontrar o número de eventos independentes em cada estratégia e por fim é exibido a teoria dos erros tipo I e II sobre múltiplos testes.

3.1 Sharpe Ratio

Sharpe (1966) desenvolveu o índice que mensura a quantidade adicional de retorno que um investidor recebe por incremento de unidades de risco. Sua definição original se da pela diferença entre o retorno do investimento e o retorno livre de risco, dividido pelo desvio padrão do investimento (volatilidade),

$$SR = \frac{r_v - r_f}{\sigma_v}$$

A ideia principal por trás do índice é quanto maior for ser valor,

maior será o retorno em relação à um nível de risco. López de Prado and Lewis (2019) considera uma estratégia de investimento com excesso de retornos $\{r_v\}$, o qual são independentes e identicamente distribuídos IID, $r_v \sim N(\mu, \sigma^2)$, além de tratar o retorno livre de risco constante,

$$SR = \frac{\mu}{\sigma}$$

e como μ e σ são parâmetros desconhecidos, o SR é estimado da seguinte forma

$$\hat{SR} = \frac{E[\{r_v\}]}{\sqrt{V[\{r_v\}]}}$$

3.1.1 Distribuição Assintótica

De acordo com a seção anterior, os parâmetros para termos o Sharpe Ratio devem ser estimados estatisticamente e portanto estão sujeitos a erros de estimativa. Lo (2002) utiliza a teoria assintótica para entender qual precisão o índice é medido. Lo obtém expressões explícitas para a distribuição estatística do índice de Sharpe,

$$(\hat{SR} - SR) \rightarrow N\left[0, \frac{1 + \frac{1}{2}SR^2}{T}\right]$$

o processo de geração de retornos são distribuídos de forma independente e idêntica, isso é, Lo parte da hipótese de que os

retornos possuem distribuição Normal. Porém, assumir que os retornos são IID normais podem levar a uma subestimação grosseira da probabilidade de falso positivo. Pensando nessa questão, Mertens (2002) deriva a distribuição assintótica do \hat{SR} como

$$(\hat{SR} - SR) \rightarrow N \left[0, \frac{1 + \frac{1}{2}SR^2 - \gamma_3SR + \frac{\gamma_4 - 3}{4}SR^2}{T} \right]$$

onde γ_3 é a assimetria e γ_4 é a kurtose de $\{r_t\}$

3.1.2 Familywise Error Rate (FWER)

Neyman and Pearson (1933) propuseram uma estrutura em que rejeitamos a hipótese nula H_0 com confiança $(1 - \alpha)$ quando observamos um evento que, caso a hipótese nula fosse verdadeira, só poderia ocorrer com probabilidade α . Então, a probabilidade de rejeitar falsamente a hipótese nula (erro tipo I) é α . Isso é também conhecido como a probabilidade de um falso positivo.

Lopez de Prado (2020) argumenta que acima não considera a possibilidade de ser conduzido experimento sobre múltiplos testes. Após um grupo de K testes independentes, a hipótese nula H_0 é rejeitada com confiança $(1 - \alpha)^K$. Portanto, a probabilidade de cometer pelo menos um erro de tipo I é o complemento, $\alpha_K = 1 - (1 - \alpha)^K$, também conhecido como a taxa de familywise error (FWER). De forma geral, o conceito de Familywise Error se refere à probabilidade de fazer uma ou mais descobertas falsas ao realizar vários testes de hipótese.

Frane (2015) descreve que o FWER exerce um controle mais rigoroso sobre falsas descobertas do que demais métodos por limitar a probabilidade de pelo menos uma descoberta falsa.

3.2 Correção de Sidak

Šidák (1967) desenvolveu o método de correção para testes múltiplos, isso é, o problema de comparações múltiplas é neutralizado e pode ser aproximado para o primeiro termo de uma expansão de Taylor $\alpha \approx \alpha_K/K$. Em um experimento com testes múltiplos, se todas as hipóteses nulas são verdadeiras, Sidak fornece controle de erro familiar que é exato para testes que são estocasticamente independentes.

Suponha que definimos um FWER sobre K diferentes hipóteses nulas e um nível alpha familiar de α , cada hipótese nula será rejeitada com um valor p inferior

$$\alpha = 1 - (1 - \alpha_K)^{1/K}$$

Dados que os testes são independentes entre si e todas as hipóteses nulas são verdadeiras, o teste produz uma taxa de erro familiar do tipo I de exatamente α e portanto rejeitamos a hipótese nula com confiança $(1 - \alpha_K)$.

3.3 Clusterização

Nas seções anteriores é mostrado o processo da metodologia do Familywise Error além do método de correção de testes múltiplos

desevolvido por Šidák (1967). Para o estudo ser completo é necessário o conhecimento do número de eventos independentes dentro de um conjunto de testes de uma estratégia, e para isso é utilizado a clusterização.

A clusterização é uma tarefa presente na análise exploratória de dados e seu uso se dá para associar um conjunto de dados de forma que os dados dentro de um cluster sejam mais parecidos entre si do que em outros clusters.

Estivill-Castro (2002) argumenta em seu trabalho que há diversos algoritmos de clusterização porque a notação de cluster não pode ser definida precisamente. Em seu trabalho, López de Prado and Lewis (2019) utiliza o algoritmo chamado K-Means o qual é um agrupamento baseado em no centroíde que são representados por um vetor central. Porém esse algoritmo necessita que o número k de clusters seja fornecido como input do modelo o que é considerado uma grande desvantagem.

Para o trabalho exposto é utilizado o algoritmo clusterização (Affinity Propagation) onde o objetivo é segregar grupos com traços semelhantes e atribuí-los a clusters. Além disso, o modelo escolhido é um algoritmo de aprendizado de máquina não supervisionado que é particularmente adequado para problemas em que não sabemos o número ideal de clusters.

Dueck and Frey (2007) descreve que a maioria das técnicas para identificar clusters mantém o controle de um conjunto fixo de candidatos enquanto procura por boas soluções. Os algoritmos de

propagação de afinidade, entretanto, consideram simultaneamente todos os pontos de dados como candidatos.

3.4 Erro Tipo I Sobre Múltiplos Testes

O Erro Tipo I é o erro de rejeitar uma hipótese nula quando ela for de fato verdadeira, isso é, é a chance de termos um resultado "falso positivo". Dentro do ambiente pesquisado é podemos considerar uma estratégia como verdadeira e vencedora quando na realidade se trata de uma enganação e ao usá-la estamos contando com a sorte. De forma geral, é o erro de aceitar uma hipótese alternativa quando os resultados podem ser atribuídos ao acaso.

Ao introduzir o conceito de múltiplos testes estamos dizendo que os testes estatísticos são repetidos diversas e o resultado disso é um aumento potencial no Erro Tipo I. Considere uma estratégia de investimentos com uma série de retornos diários de tamanho T . É estimado o Sharpe Ratio, \hat{SR} e submete-se ao testes de hipótese, onde $H_0 : SR = 0$ e $H_1 : SR > 0$. Através de tais hipóteses é determinado a probabilidade de um falso positivo quando aplica-se esse teste múltiplas vezes.

Em Bailey and López de Prado (2012) é derivado a probabilidade de que o verdadeiro Sharpe Ratio exceda um SR^* dado, sobre as hipóteses gerais de que os retornos são estacionários e ergódicos. Caso ocorra o caso em que $\hat{SR} = SR^*$, então a estatística $\hat{z}[SR^*]$ é assintoticamente distribuída como uma Normal Padrão,

$$\hat{z}[SR^*] = \frac{(\hat{SR} - SR^*)\sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3\hat{SR} + \frac{\hat{\gamma}_4-1}{4}\hat{SR}^2}} \rightarrow Z,$$

onde $\hat{\gamma}_3$ é a assimetria dos retornos e $\hat{\gamma}_4$ é a kurtose, além de que o \hat{SR} é o Sharpe Ratio estimado não anualizado e SR^* é o Sharpe Ratio mediano dos eventos analisados.

A probabilidade de ocorrer um Familywise Error Tipo I é

$$P[\max_k \{\hat{z}[0]_k\}_{k=1,\dots,K} > z_\alpha | H_0] = 1 - (1 - \alpha)^K = \alpha_K$$

De forma sucinta o Erro Tipo I sobre múltiplos testes (α_K) pode ser derivada de acordo com os seguintes passos:

1. É aplicado o algoritmo de clusterização sobre os eventos analisados para assim, encontrar o número de eventos independentes $E[K]$
2. Estima-se $\hat{z}[0] = \max_k \{\hat{z}[0]_k\}_{k=1,\dots,K}$ sobre os retornos clusterizados
3. Calcula o Erro Tipo I para um único teste, $\alpha = 1 - Z[\hat{z}[0]]$
4. Corrige para múltiplos testes, $\alpha_K = 1 - (1 - \alpha)^K$

$$\text{Resultando em } \alpha_K = 1 - Z[\hat{z}[0]]^{E[K]}$$

3.5 Erro Tipo II Sobre Múltiplos Testes

O erro Tipo II é o erro de não rejeitar uma hipótese nula quando de fato for falsa, isso é, é a chance de termos um resultado "falso negativo". Do modo contrário do que ocorre ao Erro Tipo I, nesse caso podemos considerar uma estratégia de investimento falsa quando na realidade se trata de uma estrutura capaz de ser vencedora.

Seguindo o modelo descrito no Erro Tipo I Sobre Múltiplos Testes, suponha que a hipótese alternativa seja $H_1: SR > 0$ para a melhor estratégia seja verdadeira, e $SR = SR^*$. Portanto, o poder do teste associado com um FWER α_K é

$$\begin{aligned} & P[\max_k \{\hat{z}[0]_k\}_{k=1,\dots,K} > z_\alpha | SR = SR^*] \\ &= 1 - Z \left[z_\alpha - \frac{SR^* \sqrt{T-1}}{\sqrt{1 - \hat{\gamma}_3 SR + \frac{\hat{\gamma}_4 - 1}{4} \hat{S}R^2}} \right] \\ &= 1 - \beta \end{aligned}$$

onde $z_\alpha = Z^{-1}[(1 - \alpha_K)^{1/K}]$. É considerável a diferença entre o erro Tipo I e o erro Tipo II. No caso do erro Tipo I, onde temos o caso falso positivo, estamos envolvidos na probabilidade de cometer ao menos um erro, isso é, um único erro será uma falha. Na outra via, para o erro Tipo II, onde temos o caso falso

negativo, estamos interessados na probabilidade de que todos os positivos sejam perdidos.

Lopez de Prado (2020) argumenta que o poder do teste terá um aumento com o Sharpe Ratio mediano (SR^*), com a assimetria e com o comprimento da amostra, porém há uma correlação negativa com a curtose. Caso essa aumente, o poder do teste será deprimido. Sabemos, portanto, que a probabilidade de cometermos um erro Tipo II é β . Então, após K experimentos independentes, nossa probabilidade será $\beta_K = \beta^K$.

Por fim, a interpretação do erro Tipo II em um FWER é que alcançar um poder familiar de $(1 - \beta_K)$ requer que a verdadeira razão de Sharpe Ratio exceda SR^* . Em outras palavras, o teste não é poderoso o suficiente para detectar estratégias verdadeiras com um Sharpe ratio abaixo do SR^* implícito. Os autores não deixam claro qual seria um valor de β_K ideal para considerar se o teste é poderoso ou não para detectar estratégias verdadeiras e portanto, nesse trabalho será considerado um valor de β_K menor que 0,90, isso é, estratégias em que menos de 10% dos Sharpe ratio não exceda o SR^* serão considerados com poder insuficiente.

4 Resultados

Nesta seção, a priori, é apresentado os resultados obtidos em controlado e em seguida é revisado cada estratégia que foi proposta no trabalho e seu resultado encontrado levando em considerações os cenários sem custos de transação e aquele que há 0,1% de custo de transação.

4.1 Resultados em Ambiente Controlado

Através do uso de movimentos brownianos geométricos, foi realizado três ensaios diferentes, cada um contendo 80 experimentos com um T igual a 2 mil. O intuito dessa análise é testar como o modelo performaria em casos extremos, como por exemplo, termos uma estratégia com apenas retornos bons ou apenas retornos ruins.

Dessa forma, o primeiro ensaio foi composto apenas de experimentos bem sucedidos, onde tivemos um Sharpe Ratio mínimo de 0.05897, mediano de 0.10182 e máximo de 0.15972. O segundo teve experimentos aleatórios, contando com resultados muito bons quanto muito ruins e obteve o Sharpe Ratio mínimo de -0.03695, mediano de 0.01103 e máximo de 0.06512. Por fim, o último ensaio foi composto apenas de experimentos com resultado ruim e obteve um Sharpe Ratio mínimo de -0.05658, mediano de -0.01461 e máximo de 0.01058.

Utilizando dos dados acima e do modelo proposto, temos os

seguintes resultados para os três cenários

Tabela I Ambiente Controlado

	α_K	β_K
Cenário 1	0.0499	0,8699
Cenário 2	0.0594	0.8052
Cenário 3	0.7143	0.7086

De fato, nos resultados obtidos dentro de um ambiente controlado, a única estratégia que teve um α_K menor que 5% foi o cenário 1, que foi gerado a partir de apenas experimentos bem sucedidos. E para complementar, a estratégia também teve um poder do teste capaz de detectar estratégias verdadeiras.

Portanto, o modelo proposto é capaz de detectar se uma estratégia é verdadeira ou falsa.

4.2 Resultados Empíricos

4.2.1 Cruzamento de Médias Móveis

Rememorando a funcionalidade da estratégia de cruzamento de médias móveis, temos o estabelecimento das médias com prazos distintos onde a menor aponta para a tendência dos preços no curto prazo e a maior aponta para a tendência de longo prazo. Espera-se que a média de prazo menor seja mais impactada com a volatilidade do mercado. Portanto, se a média curta romper a média longa para cima, temos a possibilidade de uma tendência de alta nos preços e então é realizado a compra do ativo objeto. Já quando ocorre o contrário, isso é, quando a média curta rompe a média longa para baixo, existe a possibilidade de os preços sofrerem uma queda e assim, é realizado a venda do ativo objeto.

Dessa forma, geramos um conjunto de dez conjuntos com diferentes combinações de valores de média curta e longa com 78 ativos do índice Bovespa durante 1308 dias de retorno. Nota-se que para cada valor de média curta, é testado diferentes valores de média longa. A ideia é observar como uma mesma média curta se comporta com médias longas sentindo cada vez menos a volatilidade do mercado.

Em relação à estratégia de Cruzamento de Médias Móveis apenas as seguintes combinações de médias obtiveram um α_K abaixo dos 5%; nos cenários sem custos de transação tivemos as médias de 5-10, e 10-50. Porém, no caso da primeira combinação, o teste não é poderoso o suficiente para detectar estratégias verdadeiras

Tabela II Cruzamento de Médias Móveis

Médias Móveis	Sem custos de Transação		Com custos de Transação	
	α_K	β_K	α_K	β_K
5 - 10	0,0026	0,9939	0,0015	0,9989
5 - 22	0,1360	0,3469	0,1061	0,5526
5 - 30	0,1891	0,2096	0,1692	0,3298
7 - 15	0,2451	0,2480	0,2095	0,4044
7 - 30	0,1830	0,1716	0,1700	0,2393
7 - 45	0,0532	0,5220	0,0452	0,6117
10 - 22	0,0920	0,2677	0,0825	0,3826
10 - 50	0,0281	0,6926	0,0303	0,6989
15 - 30	0,0859	0,4622	0,0777	0,5469
15 - 90	0,0573	0,9027	0,0510	0,9243

pois o número de Sharpe ratio que supera o SR* implícito é menor que 1%. Ao se tratar de um ambiente com custos de transação, temos as seguintes combinações com α_K menor que 5%: 5-10, 7-45 e 10-50 e novamente as médias de 5 e 10 períodos não geraram um teste poderoso o suficiente para detectar estratégias verdadeiras. Portanto, no universo testado dentro das estratégias de cruzamento de médias móveis e com custos de transação de 0,1%, apenas as combinações de 7-45 e 10-50 são estratégias verdadeiras.

4.2.2 Momentum Time Series

A estratégia de momentum time series é visto como um indicador baseado na média dos preços considerando um intervalo de tempo pré definido. Nos resultados abaixo é testado um conjunto de dez diferentes intervalos. Como explicado anteriormente, caso a média móvel for positiva, é esperado uma valorização dos preços do ativo objeto e portanto é feito uma operação de compra. Por outro lado, caso a média do intervalo analisado seja negativa, é esperado que o ativo objeto sofra uma deterioração nos preços e ,portanto, é aberto uma operação de venda.

Tabela III Momentum Time Series

Média Móvel	Sem custos de Transação		Com custos de Transação	
	α_K	β_K	α_K	β_K
1	0,0157	0,9963	0,0075	0,9999
3	0,0146	0,9937	0,0053	0,9998
5	0,0038	0,9964	0,0012	0,9999
7	0,1947	0,7912	0,1558	0,9432
9	0,1895	0,6173	0,1154	0,9075
11	0,4649	0,0544	0,3824	0,3080
13	0,2544	0,3150	0,1961	0,6690
15	0,1330	0,3866	0,0723	0,7881
17	0,1365	0,3916	0,1169	0,6189
20	0,0448	0,7737	0,0342	0,8986

As estratégias baseadas em Momentum Time Series em um ambiente sem custos de transação tiveram as seguintes periodicidades com um α_K abaixo dos 5%: 1, 3, 5 e 20 períodos. Porém, as estratégias baseadas em 1, 3 e 5 períodos obtiveram um valor de β_K mais que 99%, isso é, menos de 1% dos Sharpe Ratio dessas estratégias tiveram valor acima do SR* e, portanto, o teste não é poderoso o suficiente para detectar estratégias verdadeiras. Ao adicionar custos de transação de 0,1% a cada troca de sinal na

estratégias, novamente as periodicidades que tiveram α_K abaixo dos 5% foram de 1,3,5 e 20 períodos. E, novamente, as estratégias de 1,3 e 5 períodos não tiveram um teste poderoso o suficiente para ara detectar estratégias verdadeiras. Desse modo, em um ambiente de estratégias baseadas em Momentum Time Series, apenas uma estratégia com 20 períodos e com custos de transação é uma estratégia verdadeira

4.2.3 Reversão à Média

As operações que utilizam as estratégias de reversão à média se balizam no princípio que os preços tendem a reverter a uma média recente. A grosso modo, a reversão média depende de um raciocínio que é o oposto da estratégia de Momentum Time Series, pois esse assume uma correlação positiva entre os retornos, enquanto aquele assume uma correlação negativa entre os retornos.

Para nossa análise, utilizamos a média móvel de um intervalo pré-definido com base. Faz-se a média aritmética dos preços de fechamento dos últimos dias negociados, considerando uma quantidade de fechamentos específica, que se torna o parâmetro do indicador.

Tabela IV Reversão à Média

Média Móvel	Sem custos de Transação		Com custos de Transação	
	α_K	β_K	α_K	β_K
5	0,0254	0,7078	0,0121	0,9103
10	0,0685	0,2725	0,0548	0,3741
15	0,0119	0,7650	0,0087	0,9272
20	0,0058	0,8371	0,0046	0,9563
25	0,0128	0,5097	0,0097	0,7493
30	0,0560	0,1780	0,0510	0,9528
35	0,0596	0,2618	0,0498	0,5560
40	0,0950	0,0750	0,0816	0,1528
45	0,0199	0,4524	0,0170	0,7149
50	0,0063	0,6974	0,0048	0,8099

Por fim, as estratégias baseadas em Reversão à Médias foram as que tiveram o maior número de estratégias consideradas verdadeiras. Em um ambiente sem custos de transação, as médias móveis de 5, 15, 20, 25, 45 e 50 tiveram um α_K abaixo de 5% e também tiveram um valor de β_K aceitável e assim, todas essas 6

periodicidades são estratégias verdadeiras. Ao adicionar custos de transação, as estratégias que tiveram o α_K abaixo de 5% foram: 5, 15, 20, 25, 35, 45 e 50. Desse grupo de estratégias, as de 5, 15, 20 não tiveram um teste poderoso o suficiente para detectar estratégias verdadeiras. Portanto, dentro do universo de estratégias baseadas em Reversão à Média testadas e com custos de transação, apenas as estratégias com periodicidade de 25, 35, 45 e 50 são verdadeiras.

5 Considerações Finais

Segundo López de Prado and Lewis (2019), a literatura estatística sobre múltiplos testes trabalham com duas diferentes definições para o erro Tipo I. A primeira, *Familywise Error Rate* (FWER) são testes projetados para controlar um único falso positivo, isso é, defina-se como a probabilidade de que pelo menos um falso positivo ocorra. A segunda forma, se trata da *False Discovery Rate* (FDR), a qual é definida como o valor esperado da proporção de falsos positivos para positivos previstos. Os testes baseados em FDR são projetados para gerar erros do Tipo I a uma taxa constante, proporcional ao número de positivos previstos.

Ainda de acordo com López de Prado and Lewis (2019), o uso de FDR não é incentivado. Tal razão se fundamenta no fato de que um investidor não alocará recursos para todas as estratégias com resultados positivos previstos dentro de uma família de testes, onde uma proporção deles provavelmente será falsa. Em vez disso, os investidores são apresentados apenas à melhor estratégia de uma família de milhões de alternativas. Um exemplo claro disso é o caso de uma estratégia baseado no cruzamento de médias móveis utilizando 7 períodos para a média curta e 30 períodos para a média longa. Essa estratégia é um exemplar único e caso for falsa, todos aqueles que a estão usando irão fracassar. Dessa forma, a definição do erro tipo I que mais se adere ao contexto de

aplicações financeiras é a FWER.

Portanto, este trabalho foi construído baseando-se em Lopez de Prado (2020) e fundamentado na definição FWER erro Tipo I como sugere López de Prado and Lewis (2019), além de incluir a análise do erro Tipo II proposta por Harvey and Liu (2020) para fornecer respostas em relação as estratégias de investimentos baseadas em Momentum Time Series, Reversão a Média e de Cruzamento de Médias Móveis. A ideia é testar se tais estratégias são válidas ou se são falsas estratégias. Para aumentar a robustez dos resultados, foi testado para cada estratégia 10 conjuntos de retornos aplicando diferentes combinações de parâmetros que causariam diferentes quantidades de entrada e saída no ativo objeto, além de ser usado um custo de transação de 0,1% que é cobrado a cada mudança de posição na estratégia. Para verificar a eficácia do modelo utilizado, é feito, com o auxílio de movimentos brownianos geométricos, testes em ambiente controlado.

Os resultados obtidos dentro do ambiente controlado comprovam que o modelo é eficaz para testar estratégias de investimento e definir se são falsas ou não. O experimentou contou com três cenários distintos. O cenário composto por apenas estratégias mal sucedidas e o cenário que contou com a aleatoriedade tiveram um α_K maior que 0,05 e, portanto, já foram tidos como falsas estratégias. O cenário que contou com apenas experimentos bem sucedidos teve um α_K menor que 0,05 e mais de 10% dos Sharpe Ratio maiores que o SR*. Dessa forma, por ter um poder de teste suficiente para detectar estratégias verdadeiras, esse cenário é

considerado uma estratégia verdadeira.

Ao direcionar o teste para estratégias de investimento e dados coletados no mercado brasileiro, temos os seguintes resultados. No universo de Cruzamento de Médias Móveis, ao desconsiderar custos de transação, tivemos duas ramificações da estratégia usando médias móveis de 5-10 e 10-50 com α_K menor que 0,05, porém a primeira não teve um poder de teste suficiente. Ao adicionar custos de transação, tivemos uma terceira ramificação com α_K menor que 0,05 que foi a estratégia utilizando as médias 7-45. Nesse cenário, apenas as médias de 7-45 e 10-50 foram capazes de serem consideradas estratégias verdadeiras. No universo de estratégias de Momentum Time Series tivemos novamente um número de ramificações com α_K menor que 0,05, porém que não tiveram um poder do teste suficiente, o que foi o caso das estratégias utilizando 1, 3 e 5 períodos tanto com custo quando sem. A única ramificação considerando custos de transação que foi dada como uma estratégia verdadeira foi a de 20 períodos. As estratégias baseadas em Reversão à Média tiveram o maior índice de aceite de estratégias verdadeiras. Em um ambiente sem custos de transação, seis ramificações foram aceitas, que são as de 5, 15, 20, 25, 45 e 50 períodos. Ao adicionar custos de transação, sete estratégias tiveram um α_K menor que 0,05, e quatro dessas tiveram um poder do teste suficiente para detectar estratégias verdadeiras e, portanto, consideradas estratégias verdadeiras e são elas as de 25, 35, 45 e 50 períodos.

Por fim, é interessante notar que, ao adicionar custos de

transação, houve uma queda nos α_K o que aumentou em alguns casos o número de estratégias aceitas dentro do erro Tipo I, enquanto houve um aumento do valor de β_K e ocasionando o fato de algumas estratégias não serem aceitas dentro do erro Tipo II. Para futuros trabalhos, seria interessante fazer a investigação do por que ocorre esse movimento.

6 Conclusão

A indústria de investimentos seguirá criando algoritmos baseados em análises quantitativas para traçar suas estratégias de compra e venda de ativos e montagem de portfólios. Pensando nesse padrão e evolução, este trabalho foi dedicado a trabalhar em uma abordagem, apresentada por Lopez de Prado (2020), de como testar se uma estratégia é de fato verdadeira ou se trata de uma estratégia falsa ou charlatanismo.

Foi testado três universos diferentes de estratégia, Cruzamento de Médias Móveis, Momentum Time Series e Reversão à Média. Para a primeira estratégia, dos dez grupos testados, apenas dois foram considerados estratégias verdadeiras, que são as estratégias com médias móveis de 7-45 e 10-50 períodos. Para a segunda, dos dez grupos testados apenas a estratégia de 20 períodos foi considerada uma estratégia verdadeira. E por fim, a terceira estratégia obteve o maior número de ramificações dados como verdadeiras estratégias, que são as de 25, 35, 45 e 50 períodos. Importante notar que nos três universos analisados, houve um padrão de considerar verdadeiras as estratégias que sofreriam menos com a volatilidade do mercado na decisão de inverter a posição. Note que, em Cruzamento de Médias Móveis, a posição demoraria a ser invertida dado que as médias curtas está bem distante das médias longas. Não bastaria apenas uma anomalia para que invertesse essa posição. Em relação a Momentum,

apenas a estratégia considerando o maior número de períodos foi dada como verdadeira, isso é, a posição seria invertida apenas quando a média dos últimos 20 dias negociados mudasse de lado.

Por fim, é possível notar o mesmo padrão ocorrendo com Reversão à Média, porém com um número maior de estratégias verdadeiras.

Portanto, concluímos que não podemos determinar se um universo inteiro de estratégias é falso ou verdadeiro, sempre deverá ser levado em consideração os parâmetros utilizados para a sua montagem. Porém, podemos observar que, estruturas criadas para sofrer maiores picos de volatilidade sem mudarem de posição possuem uma maior chance de serem verdadeiras frente àquelas que alteram suas posições de comprado para vendido com mais exaustão.

Referências

- Acharya, V. V. and Pedersen, L. H. (2005). Asset pricing with liquidity risk. *Journal of Financial Economics*, 77(2):375–410.
- Bailey, D. H., Borwein, J. M., Lopez de Prado, M., and Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism. *Notices of the American Mathematical Society*, 61(5):458 – 471.
- Bailey, D. H. and López de Prado, M. L. (2012). The sharpe ratio efficient frontier. *Journal of Risk*, 15(2):3–44.
- Balvers, R., Wu, Y., and Gilliland, E. (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *Journal of Finance*, 55(2):745–772.
- Brock, W., Lakonishok, J., and LeBaron, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, 47(5):1731–1764.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1):57–82.
- Dueck, D. and Frey, B. J. (2007). Non-metric affinity propagation for unsupervised image categorization. *Proceedings of the IEEE International Conference on Computer Vision*.
- Estivill-Castro, V. (2002). Why so many clustering algorithms. *ACM SIGKDD Explorations Newsletter*, 4(1):65–75.

- Frane, A. V. (2015). Invited debate are per-family type I error rates relevant in social and behavioral science? *Journal of Modern Applied Statistical Methods*, 14(1):12–23.
- Harvey, C. R. and Liu, Y. (2020). False (and Missed) Discoveries in Financial Economics. *Journal of Finance*, 75(5):2503–2553.
- Lo, A. W. (2002). “The Statistics of Sharpe Ratios”: Author’s Response. *Financial Analysts Journal*, 58(6):18–18.
- Lopez de Prado, M. (2020). *Machine Learning for Asset Managers*, volume 44. Cambridge University Press.
- López de Prado, M. and Lewis, M. J. (2019). Detection of false investment strategies using unsupervised learning methods. *Quantitative Finance*, 19(9):1555–1565.
- Mertens, E. (2002). Comments on variance of the iid estimator in lo (2002). Technical report.
- Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2):228–250.
- Neyman, J. and Pearson, E. S. (1933). on the Problem of the Most Efficient Tests of Statistical Hypotheses. *Joint Statistical Papers*, 231(1933):140–185.
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1):119–138.

Šidák, Z. (1967). Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. *Journal of the American Statistical Association*, 62(318):626–633.

Apêndices

APÊNDICE A – Fatores de Risco

Para a construção dos portfólios de ativos baseados em fatores de risco foi utilizado os dados disponíveis no Centro Brasileiro de Pesquisa em Economia Financeira da Universidade de São Paulo (NEFIN). Para uma ação listada no índice bovespa ser elegível a entrar em um dos portfólios, ela deve atender aos seguintes critérios definidos pelo centro:

- A ação é a ação mais negociada da empresa (aquela com o maior volume negociado durante o ano passado);
- A ação foi negociada em mais de 80% dos dias do ano $t-1$ com volume superior a R\$ 500.000,00 por dia. Caso a ação tenha sido listada no ano $t-1$, o período considerado vai do dia da listagem ao último dia do ano;
- O ativo foi inicialmente listado antes de dezembro do ano $t-1$

A.1 HML Factor

High Minus Low (HML) é o retorno de uma carteira comprada em ações com alto índice book-to-market e vendido em ações com baixo índice book-to-market. O modelo argumenta que empresas que apresentam um alto índice book-to-market (alto valor) historicamente superam as ações com baixo índice book-to-market (ações de crescimento). Além de ser também

conhecido como prêmio de valor é um dos fatores usado no modelo de fatores Fama-French.

A construção dos retornos diários do portfólio é dado através dos seguintes passos. Primeiro a cada janeiro do ano t , é classificado (em ordem crescente) as ações elegíveis em 3 quantis (carteiras) de acordo com a relação book-to-market das ações em junho do ano $t-1$. Sendo as carteiras: 1º tercil é a carteira de crescimento, 2º tercil; neutro e, por fim, 3º tercil; valor.

Após isso, é calculado os retornos ponderados da primeira carteira e a terceira carteira e finalmente temos o fator HML onde temos o retorno da carteira de ações de alto valor menos o retorno da carteira de ações de crescimento.

A.2 IML Factor

Illiquid Minus Liquid Factor (IML) é o retorno de uma carteira comprada em ações líquidas e vendida em ações ilíquidas.

A construção dos retornos diários do portfólio é dado através dos seguintes passos. A cada mês t , as ações elegíveis são ordenadas em forma crescente em 3 quantis (carteiras) de acordo com a média móvel de iliquidez dos últimos 12 meses de acordo com o trabalho de Acharya and Pedersen (2005). As 3 carteiras são baseadas em baixa, neutra e alta iliquidez. Em seguida, é calculado os retornos ponderados de forma igual da primeira carteira (baixa iliquidez) e da terceira carteira (alta iliquidez). O Fator IML é o retorno da carteira de ações líquidas menos o

retorno da carteira de ações ilíquidas.

A.3 Market Factor

Market Factor é a diferença entre o retorno diário ponderado pelo valor da carteira de mercado (utilizando as ações elegíveis conforme apresentando anteriormente) e a taxa diária livre de risco. A taxa diária livre de risco é calculada a partir do DI Swap de 30 dias.

Em outras palavras, esse fator explica os retornos de uma carteira em termos da quantidade de risco que ela contém em relação ao mercado.

A.4 SMB Factor

Small Minus Big (SMB) é um dos fatores no modelo de precificação de ações de Fama-French. É o retorno de uma carteira comprada em ações com baixa capitalização de mercado (pequena) e vendida em ações com alta capitalização de mercado (grande), isso é, é o retorno excedente que as empresas de menor capitalização de mercado geram em comparação com as empresas maiores.

A construção dos retornos diários do portfólio é dado através dos seguintes passos. Todo mês de janeiro do ano t é ordenado de forma crescente as ações elegíveis com sua capitalização de mercado em dezembro de $t-1$ e são separadas em 3 tercís (carteiras), onde o 1º tercís é pequena capitalização, 2º é

capitalização neutra e a 3^a é alta capitalização de mercado. O próximo passo é calcular os retornos com pesos iguais da primeira e da terceira carteira, dessa forma temos o fator SMB que é o retorno da carteira de ações pequena menos o retorno da carteira de grandes ações.

A.5 WML Factor

Winners Minus Losers (WML) é o retorno de uma carteira comprada em ações com retornos anteriores elevados e vendida em ações com retornos anteriores baixos. Carhart (1997) estendeu o modelo de três fatores de Fama-French adicionando o WML como um quarto fator.

A construção dos retornos diários do portfólio é dado através dos seguintes passos. A cada mês t é ordenador de forma crescente todas as ações que atendem aos critérios de elegibilidade em 3 tercís (carteiras) de acordo com seus retornos acumulados entre os meses $t-12$ e $t-2$, e cada tercil corresponde à tercil perdedor, neutro e vencedor. Por fim, é calculado os retornos com pesos iguais da primeira carteira (retornos passados baixos) e da terceira carteira (retornos passados altos). O fator WML é o retorno do portfólio dos vencedores menos o retorno do portfólio dos perdedores.

A.6 Estimação dos Parâmetros

Portfólios Baseados em Fatores de Risco

Portfólio	α_K	$1 - \beta_K$
HML Factor	0.1948	0.7134
IML Factor	0.0059	0.9928
Market Factor	0.0125	0.9807
SMB Factor	0.0296	0.9550
WML Factor	0.0032	0.9471

APÊNDICE B – Distribuição dos Sharpe Ratio

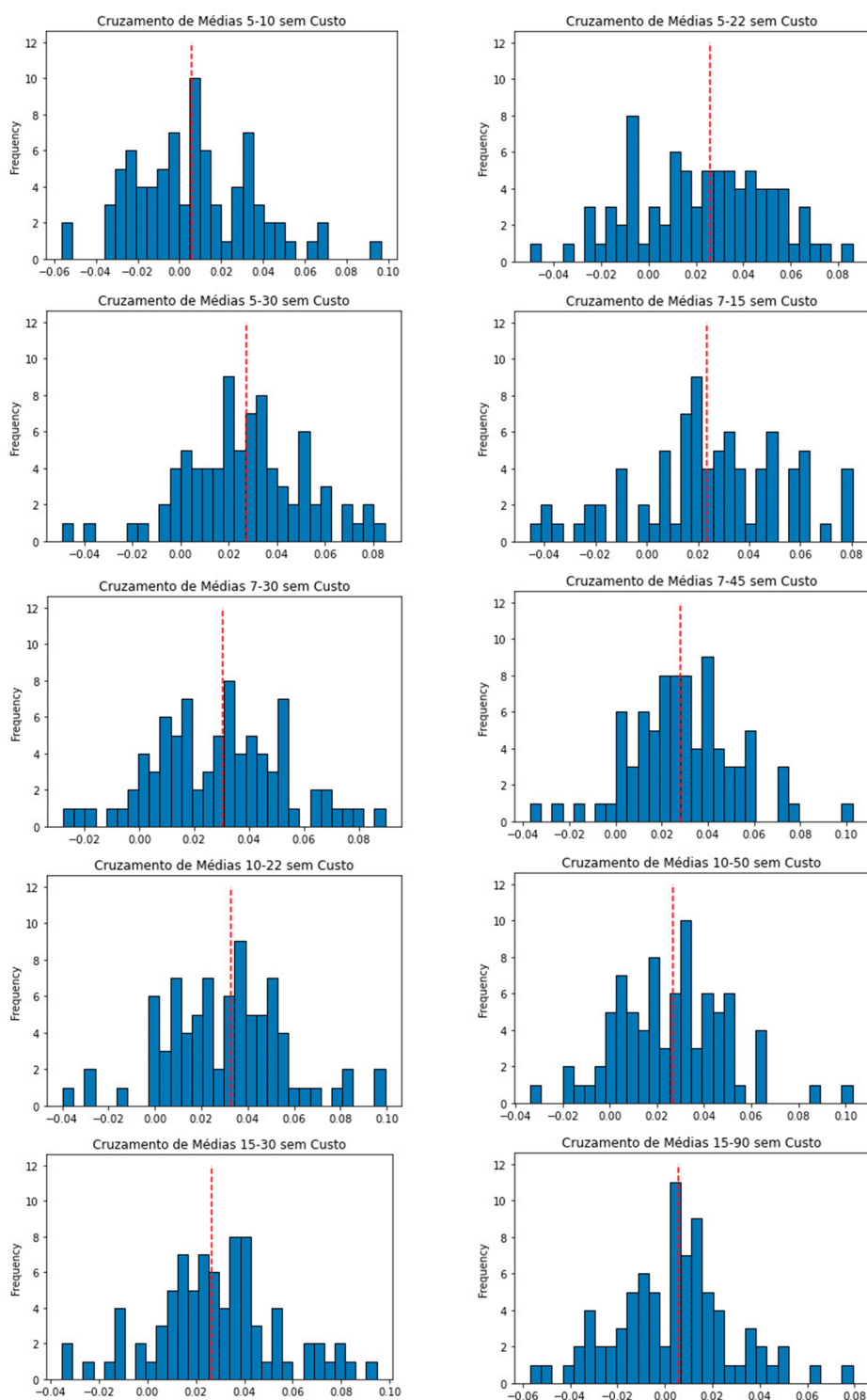


Figura 2 – Distribuição dos Sharpe Ratio de Cruzamento de Médias Móveis sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*

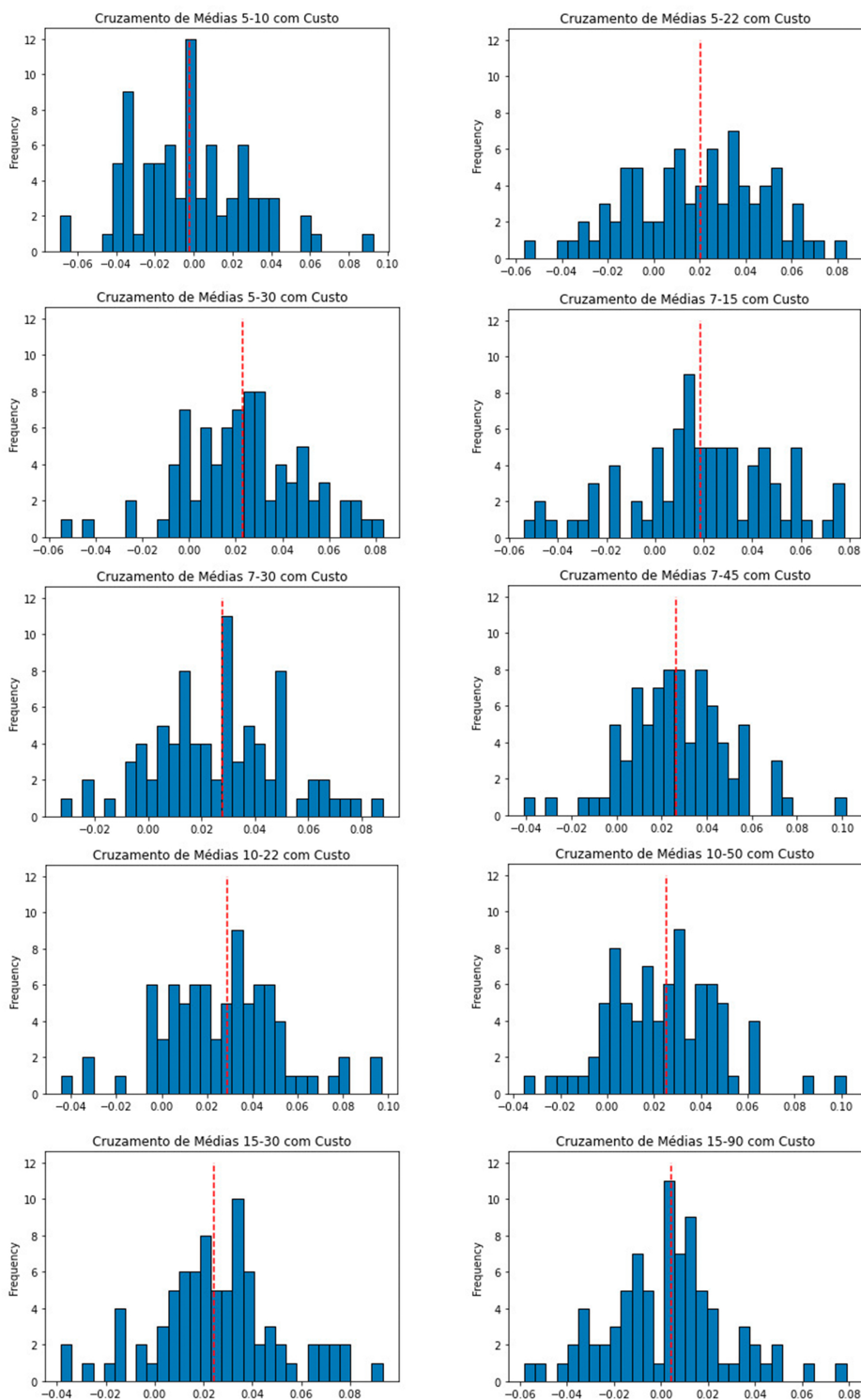


Figura 3 – Distriuição dos Sharpe Ratio de Cruzamento de Médias Móveis com Custo - Destacando a posição do Sharpe Ratio Mediano SR*

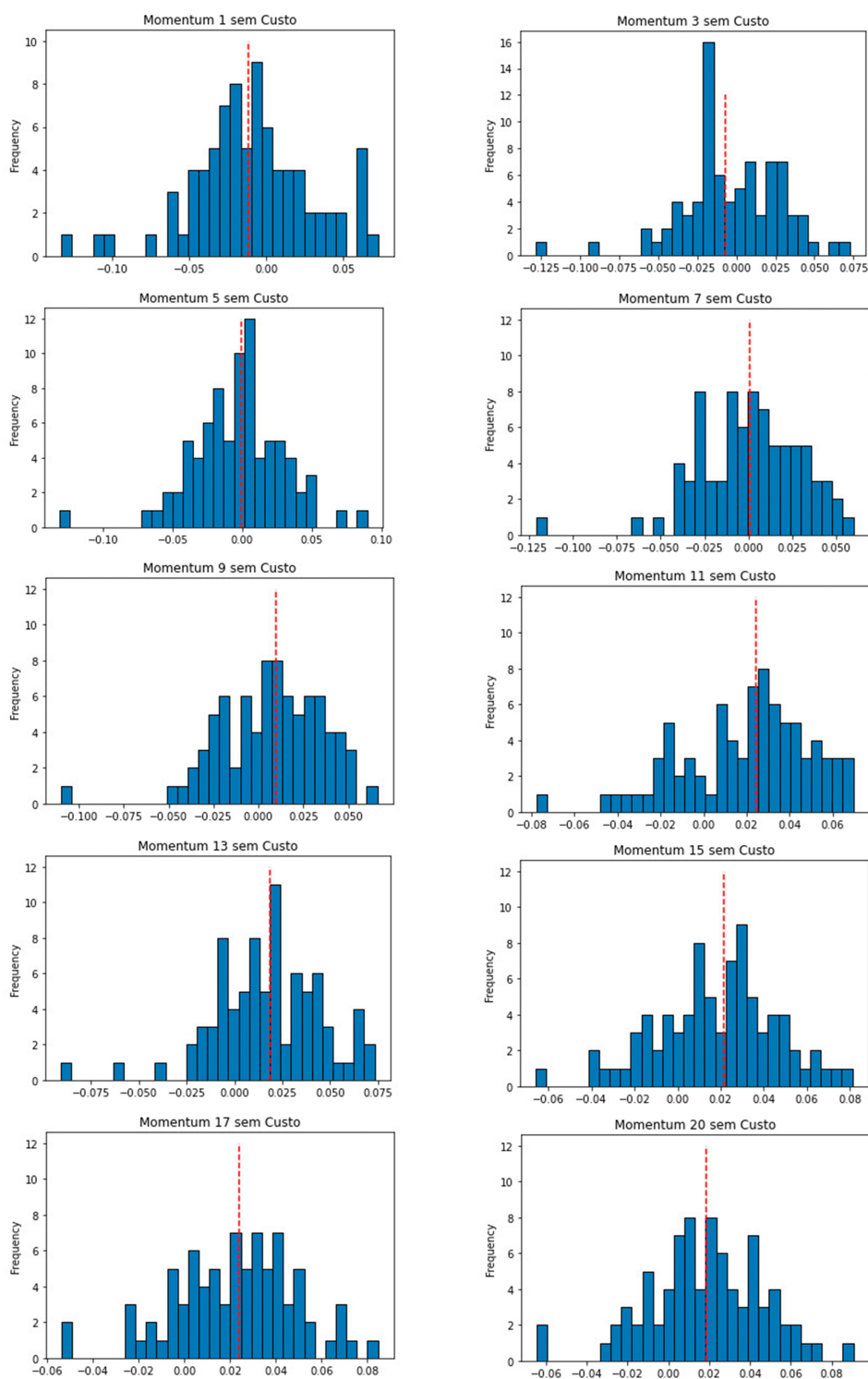


Figura 4 – Distribuição dos Sharpe Ratio de Momentum sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*

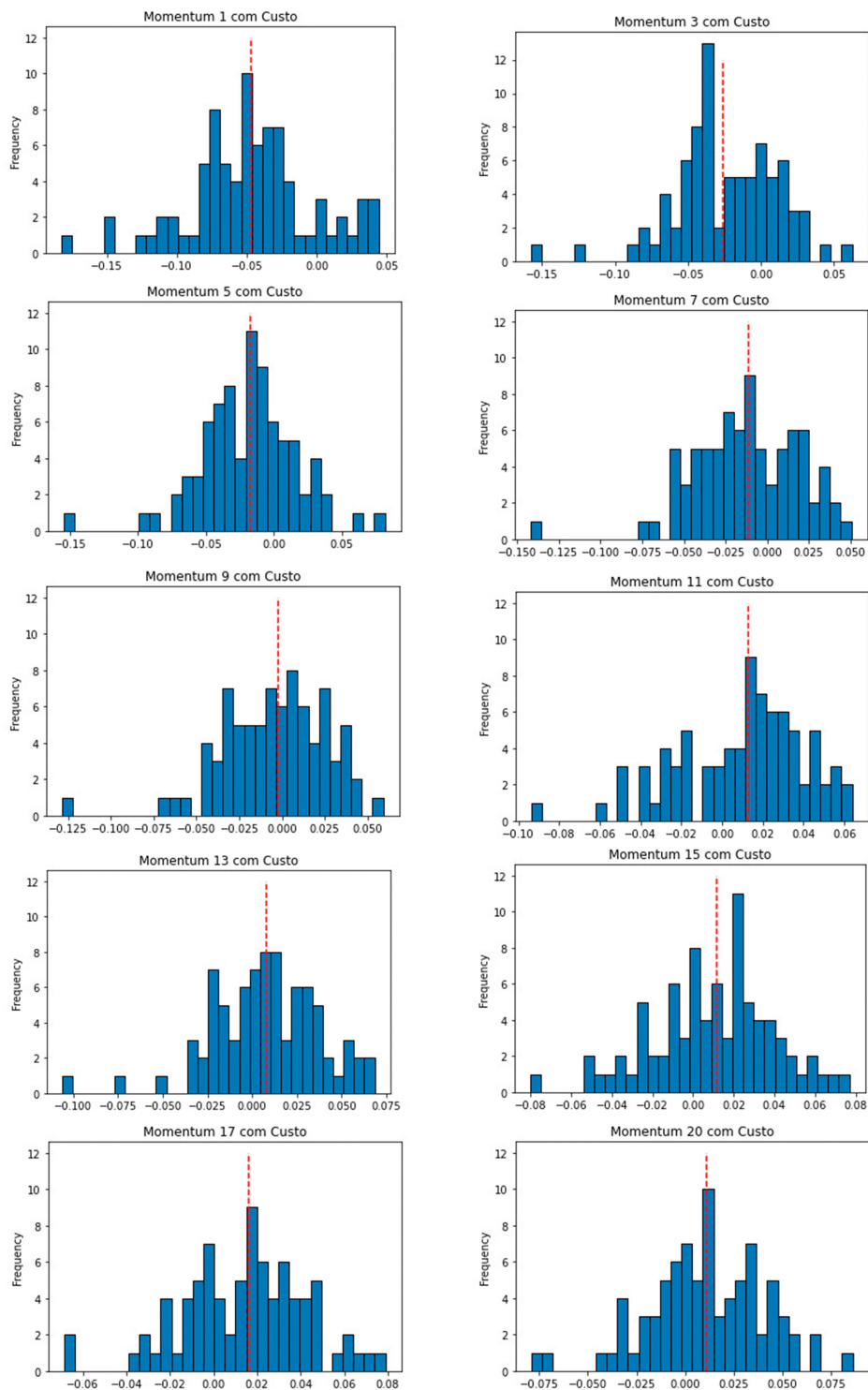


Figura 5 – Distribuição dos Sharpe Ratio de Momentum com Custo - Destacando a posição do Sharpe Ratio Mediano SR*

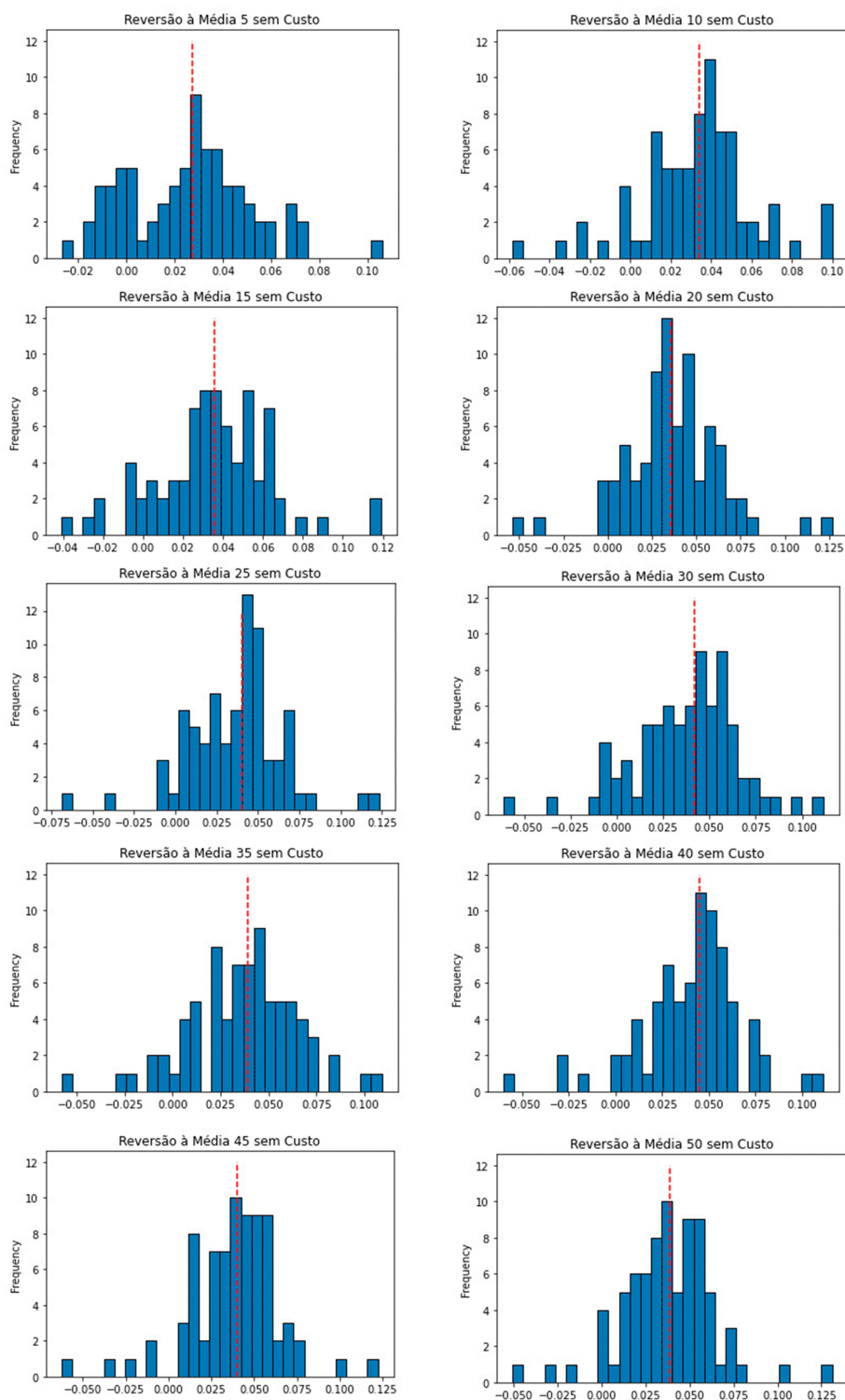


Figura 6 – Distribuição dos Sharpe Ratio de Reversão à Média sem Custo - Destacando a posição do Sharpe Ratio Mediano SR*

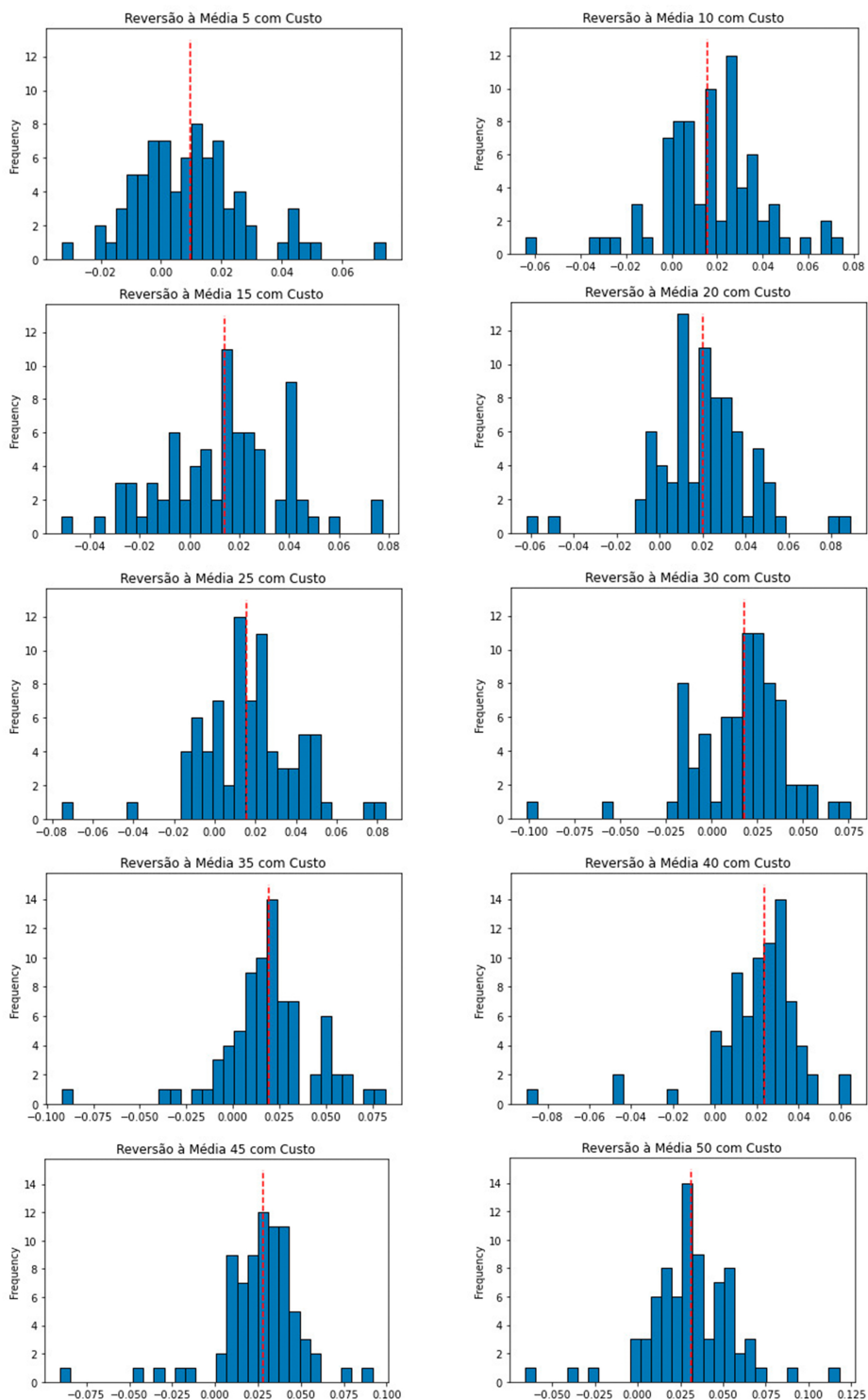


Figura 7 – Distribuição dos Sharpe Ratio de Reversão à Média com Custo - Destacando a posição do Sharpe Ratio Mediano SR*