



AVALIAÇÃO DA ROBUSTEZ DO DESEMPENHO SETORIAL NO MERCADO BRASILEIRO: UMA APLICAÇÃO DO *DEFLATED SHARPE RATIO*

Aluno Adriano Vinícius Perrud

Mestrando em Contabilidade

Universidade Federal do Mato Grosso do Sul

adrianovsperrud@outlook.com

Professor Jorge Luis Sanchez Arevalo

Doutor em Administração de Empresas

Universidade Federal do Mato Grosso do Sul

jorge.sanchez@ufms.br

RESUMO

O objetivo deste estudo é avaliar a robustez estatística do desempenho ajustado ao risco dos principais setores do mercado acionário brasileiro. Embora gestores frequentemente utilizem o *Sharpe Ratio* para identificar setores com melhor relação risco-retorno, essa métrica pode produzir inferências excessivamente otimistas quando múltiplas alternativas são comparadas ou quando os retornos apresentam distribuições não normais. Para lidar com esse problema, o estudo aplica o *Probabilistic Sharpe Ratio* (PSR) e o *Deflated Sharpe Ratio* (DSR), metodologia proposta por Bailey e López de Prado (2014), que ajusta a inferência para incerteza amostral e viés de seleção decorrente de múltiplos testes. A análise utiliza retornos mensais dos principais índices setoriais da B3 ao longo de dez anos. Inicialmente calcula-se o *Sharpe Ratio* tradicional e, em seguida, aplica-se o PSR e o DSR para avaliar a significância estatística do desempenho observado. Os resultados indicam que, embora alguns setores apresentem *Sharpe Ratio* positivo, nenhum apresenta evidência estatisticamente robusta de desempenho superior após a correção por múltiplas comparações. Esses achados sugerem cautela na interpretação de rankings setoriais baseados exclusivamente no *Sharpe Ratio*. Como contribuição, o estudo demonstra a utilidade do *Deflated Sharpe Ratio* para avaliar a robustez do desempenho setorial em mercados emergentes.

Palavras-chave: Alocação setorial, Desempenho ajustado ao risco, *Deflated Sharpe Ratio*, Mercado acionário brasileiro.



1 INTRODUÇÃO

A avaliação de desempenho ajustado ao risco constitui um dos pilares da teoria e da prática de gestão de portfólios. Entre as métricas mais difundidas na literatura financeira destaca-se o *Sharpe Ratio*, que mensura o retorno excedente de um ativo ou estratégia em relação à taxa livre de risco por unidade de volatilidade (Sharpe, 1966). Devido à sua simplicidade e fácil interpretação, essa métrica tornou-se amplamente utilizada na análise de fundos de investimento, estratégias quantitativas e decisões de alocação de ativos, sendo adotada tanto em estudos acadêmicos quanto na prática profissional de gestão de recursos.

No contexto da gestão de portfólios, a alocação setorial representa um componente relevante das decisões de investimento, uma vez que diferentes setores da economia apresentam dinâmicas distintas de risco e retorno ao longo do tempo. Em mercados emergentes, como o brasileiro, essas diferenças tendem a ser ainda mais pronunciadas devido a fatores como maior volatilidade macroeconômica, mudanças estruturais frequentes e heterogeneidade entre segmentos econômicos (Sahoo & Kumar, 2024; Salotra & Pinsky, 2026; Zhang, et al., 2025). Nesse ambiente, gestores frequentemente utilizam dados históricos para identificar setores que demonstraram melhor desempenho ajustado ao risco, buscando replicar esses padrões em estratégias futuras.

Entretanto, a utilização intensiva de dados históricos para identificar estratégias superiores pode introduzir importantes problemas estatísticos, como *data-snooping* e *overfitting* (Lo & MacKinlay, 1990; Bailey & López de Prado, 2014). Esses fenômenos ocorrem quando múltiplas estratégias, combinações de ativos ou especificações são testadas sobre o mesmo conjunto de dados, aumentando a probabilidade de que alguns resultados aparentemente significativos surjam apenas por acaso. Como consequência, rankings de desempenho baseados exclusivamente em métricas tradicionais podem refletir ruído estatístico, e não evidência de habilidade genuína.

Com o objetivo de mitigar esses problemas, estudos recentes propuseram métricas estatísticas capazes de incorporar explicitamente a incerteza amostral, a não normalidade dos retornos e o viés de seleção decorrente de múltiplos testes. Entre essas abordagens destacam-se o *Probabilistic Sharpe Ratio* (PSR) e o *Deflated Sharpe Ratio* (DSR), propostos por Bailey e López de Prado (2014). Essas métricas permitem avaliar a probabilidade de que um *Sharpe Ratio* observado seja estatisticamente consistente com desempenho superior, levando em consideração características da distribuição dos retornos e o número de estratégias avaliadas.

Apesar do crescente interesse por metodologias que buscam corrigir vieses estatísticos na avaliação de desempenho, sua aplicação em análises empíricas envolvendo alocação setorial em mercados emergentes ainda permanece relativamente limitada. Grande parte da literatura recente concentra-se na avaliação de estratégias quantitativas e fatores de risco por meio de modelos econométricos ou de aprendizado de máquina (Gu et al., 2020; Feng et al., 2020). Ao mesmo tempo, estudos têm demonstrado que resultados aparentemente significativos podem emergir de processos de múltiplos testes e *data-snooping* quando diversas estratégias são avaliadas simultaneamente (Harvey et al., 2016; Bailey et al., 2020). Nesse contexto, análises comparativas de desempenho entre setores frequentemente



continuam baseadas em métricas tradicionais, como o *Sharpe Ratio*, que não consideram explicitamente o risco associado à multiplicidade de comparações.

Diante desse contexto, surge a seguinte questão de pesquisa: o desempenho superior observado entre setores do mercado acionário brasileiro permanece estatisticamente robusto quando se controlam os efeitos de incerteza amostral e múltiplos testes? Assim, o objetivo deste estudo é avaliar a robustez estatística do desempenho ajustado ao risco dos principais setores do mercado acionário brasileiro, utilizando os índices setoriais negociados na B3. Para isso, são aplicadas as métricas *Probabilistic Sharpe Ratio* (PSR) e *Deflated Sharpe Ratio* (DSR), com o propósito de verificar se os desempenhos observados refletem evidência estatística de habilidade genuína ou se podem ser atribuídos a efeitos de *data-snooping* e *overfitting*.

Assim, o estudo visa contribuir para a literatura de avaliação de desempenho financeiro ao aplicar o *Deflated Sharpe Ratio* (DSR) na análise comparativa entre setores de mercado em um contexto de economia emergente. Ao incorporar correções para incerteza amostral e múltiplas comparações, a pesquisa oferece uma abordagem mais rigorosa para a interpretação de rankings setoriais baseados em métricas tradicionais de risco e retorno. Além disso, os resultados fornecem implicações relevantes para gestores de portfólio e analistas de mercado interessados em compreender os limites do *Sharpe Ratio* na avaliação de estratégias de investimento.

2 REVISÃO DA LITERATURA

A literatura de finanças desenvolveu diversas métricas destinadas a avaliar o desempenho de ativos e portfólios ajustado ao risco. Entre as contribuições pioneiras, destaca-se o trabalho de Treynor (1965), que propôs uma medida de desempenho baseada na relação entre retorno excedente e risco sistemático representado pelo beta. Em seguida, Sharpe (1966) introduziu o *Sharpe Ratio*, métrica que relaciona o retorno excedente de um portfólio à sua volatilidade total. Posteriormente, Jensen (1968) ampliou essa abordagem ao introduzir o alfa de Jensen, que mede o desempenho anormal de um portfólio em relação ao retorno esperado de acordo com o *Capital Asset Pricing Model* (CAPM).

Com o desenvolvimento da literatura, surgiram extensões destinadas a capturar aspectos específicos do risco. Sortino e Van der Meer (1991) argumentam que a volatilidade total pode penalizar excessivamente estratégias com retornos assimétricos, propondo a utilização do *downside risk*, que considera apenas a variabilidade negativa dos retornos. De forma complementar, Modigliani e Modigliani (1997) desenvolveram uma medida de desempenho ajustado ao risco que traduz o *Sharpe Ratio* em termos de retorno equivalente ao mercado, facilitando sua interpretação econômica.

Apesar da ampla utilização dessas métricas, estudos recentes têm destacado limitações importantes relacionadas à inferência estatística sobre desempenho financeiro. Um dos principais problemas refere-se ao fenômeno de *data-snooping*, que ocorre quando um grande número de estratégias ou ativos é testado sobre a mesma base de dados, aumentando artificialmente a probabilidade de encontrar resultados aparentemente significativos. Harvey e Liu (2015) argumentam que o processo de *backtesting* em finanças frequentemente sofre desse problema, enquanto Harvey et al. (2016) demonstram que



muitos fatores propostos na literatura de precificação de ativos podem resultar de múltiplos testes sobre a mesma amostra.

Nesse contexto, diversas abordagens metodológicas foram desenvolvidas para aprimorar a inferência estatística associada às métricas tradicionais de desempenho. Bailey e López de Prado (2012) investigam propriedades estatísticas do *Sharpe Ratio* e discutem suas limitações quando múltiplas estratégias são avaliadas simultaneamente. Posteriormente, Bailey e López de Prado (2014) introduzem o *Deflated Sharpe Ratio* (DSR), uma métrica que corrige o *Sharpe Ratio* tradicional ao considerar três elementos fundamentais: vies de seleção decorrente da escolha da melhor estratégia entre várias alternativas, *overfitting* em processos de *backtesting* e a presença de distribuições de retornos não normais.

Além das discussões metodológicas, a literatura também tem investigado a estabilidade de métricas de desempenho em diferentes contextos empíricos. O modelo de cinco fatores de Fama e French (2015) representa um avanço importante na explicação dos retornos de ativos, enquanto estudos recentes mostram que muitos modelos e estratégias apresentam forte deterioração de desempenho quando avaliados fora da amostra. Chincó et al. (2024), por exemplo, demonstram que modelos multifatoriais frequentemente exibem *Sharpe Ratios* elevados na amostra de estimação, mas desempenho significativamente inferior fora da amostra, evidenciando riscos de *overfitting*.

Diante dessas evidências, torna-se claro que a utilização isolada do *Sharpe Ratio* pode levar a interpretações equivocadas sobre a existência de habilidade genuína em estratégias de investimento ou no desempenho relativo entre ativos e setores. A crescente complexidade dos mercados financeiros e a multiplicidade de estratégias testadas exigem ferramentas estatísticas capazes de corrigir vieses associados a múltiplos testes e à incerteza amostral.

Apesar dos avanços recentes na literatura sobre avaliação de desempenho ajustado ao risco, grande parte dos estudos empíricos ainda continua baseada em métricas tradicionais como o *Sharpe Ratio*, frequentemente sem considerar explicitamente problemas de incerteza amostral, dependência temporal ou multiplicidade de testes. Estudos recentes demonstram que o *Sharpe* estimado pode ser sensível a erros de estimação, distribuição não normal dos retornos e diferenças entre desempenho dentro e fora da amostra (Feng et al., 2024; Senescall & Low, 2024; Kao et al., 2025). Essa limitação torna-se particularmente relevante em contextos nos quais múltiplos ativos, estratégias ou setores são comparados simultaneamente, aumentando a probabilidade de identificar desempenhos aparentemente superiores apenas por acaso.

À luz dessas contribuições teóricas e empíricas, torna-se relevante investigar se a aparente superioridade de determinados setores, frequentemente identificada por métricas tradicionais de desempenho, permanece estatisticamente robusta quando se consideram correções para incerteza amostral e múltiplos testes. Em particular, a literatura sugere que rankings baseados exclusivamente no *Sharpe Ratio* podem refletir ruído estatístico decorrente de *data-snooping*, especialmente quando diversas alternativas são avaliadas simultaneamente. Nesse contexto, o uso de métricas inferenciais, como o *Probabilistic Sharpe Ratio* (PSR) e o *Deflated Sharpe Ratio* (DSR), permite avaliar se o desempenho observado é consistente com habilidade genuína ou se pode ser explicado por efeitos aleatórios associados à multiplicidade de testes.



Assim, com base nesses argumentos, são formuladas as seguintes hipóteses de pesquisa:

H1: Não há evidência estatística de habilidade genuína no desempenho setorial após correções para incerteza amostral e múltiplos testes.

H2: Pelo menos um setor apresenta desempenho ajustado ao risco estatisticamente robusto após a aplicação das métricas PSR e DSR.

H3: O *Sharpe Ratio* tradicional não é suficiente para inferência estatística robusta sobre desempenho ajustado ao risco.

H4: O *Sharpe Ratio* tradicional tende a superestimar a evidência de desempenho superior quando comparado às métricas ajustadas PSR e DSR.

3 PROCEDIMENTOS METODOLÓGICOS

O estudo caracteriza-se como uma pesquisa quantitativa e empírica, com enfoque descritivo e inferencial, voltada à avaliação do desempenho ajustado ao risco de índices setoriais do mercado acionário brasileiro. A metodologia baseia-se no delineamento proposto por Bailey e López de Prado (2014), com ênfase no uso do *Deflated Sharpe Ratio* (DSR) para correção de vieses associados à não-normalidade dos retornos e à seleção *ex-post* do melhor desempenho entre múltiplas alternativas.

3.1 Base de Dados e Critérios de Seleção

A seleção dos setores para análise seguiu critérios de representatividade, liquidez e disponibilidade de dados históricos. Foram selecionados os seis principais índices setoriais da B3 que, em conjunto, representam a estrutura produtiva da economia brasileira e possuem séries históricas consistentes para o período de estudo.

Tabela 1: Setores e Índices Analisados

Setor	Índice	Participação no Ibovespa (%)
Consumo	ICON	13,81
Energia Elétrica	IEE	8,92
Financeiro	IFNC	28,52
Imobiliário	IMOB	1,40
Materiais Básicos	IMAT	18,50
Utilidades Públicas	UTIL	4,03

Fonte: B3 - Brasil Bolsa Balcão (2025).

Esses setores representam coletivamente cerca de 75% da capitalização do Ibovespa. Cada setor foi representado por seu respectivo índice setorial oficial da B3, garantindo a representatividade e replicabilidade da análise.

Além da seleção dos setores, a definição do período amostral constitui uma decisão metodológica crítica em estudos financeiros, pois influencia diretamente a robustez das



inferências sobre o desempenho e a estabilidade de estratégias de investimento (Sharpe, 1994; Bailey & López de Prado, 2012). Para este estudo, foi definido um período de dez anos completos, de janeiro de 2015 a dezembro de 2024, com dados na frequência mensal, totalizando 120 observações temporais para cada série analisada.

Este intervalo de dez anos captura ciclos econômicos completos e distintos, incluindo a recessão doméstica de 2015 a 2016, o período de recuperação, o choque da pandemia de COVID-19 e a fase recente de ajustes macroeconômicos. Tal diversidade de cenários permite avaliar a performance dos ativos sob diferentes condições de mercado, assegurando que os resultados não reflitam apenas um contexto econômico específico.

Fundamentando esta escolha, Bailey e López de Prado (2012, 2014) argumentam que a confiabilidade de métricas como o Índice de Sharpe depende criticamente do comprimento e da qualidade da série histórica. Uma amostra longa é essencial para: (1) mitigar a incerteza inerente à estimativa do prêmio de risco; (2) proporcionar maior confiança estatística na avaliação de desempenho, conforme preconizado pelo *Probabilistic Sharpe Ratio* (PSR); e (3) reduzir os riscos de conclusões enviesadas por *overfitting* ou por eventos transitórios, problemas destacados pelo *Deflated Sharpe Ratio* (DSR).

A utilização de dados mensais equilibra a captura de tendências de médio e longo prazos com a minimização do ruído de alta frequência, enquanto a anualização desses resultados permite a comparação padronizada entre os setores. Dessa forma, o desenho amostral adotado não apenas garante representatividade econômica, mas também atende aos rigorosos requisitos estatísticos necessários para uma avaliação de desempenho financeiro moderna e confiável.

No que se refere à taxa livre de risco, adotou-se como *proxy* a Taxa SELIC, coletada junto ao Banco Central do Brasil. Esta escolha alinha-se com o *framework* conceitual estabelecido por Sharpe (1966, 1975, 1994), que fundamenta o cálculo do retorno excedente como diferença entre o retorno do investimento e o retorno de um ativo livre de risco. Conforme Sharpe (1975) definiu em seu trabalho seminal, "a recompensa proporcionada por um portfólio é medida por seu retorno excedente médio - o prêmio obtido por assumir risco" (p. 30).

A SELIC representa a taxa básica de juros da economia brasileira, caracterizando-se pela elevada liquidez e mínimo risco de crédito, atributos essenciais para uma *proxy* adequada. Bailey e López de Prado (2012, 2014) enfatizam que a seleção apropriada da taxa livre de risco é crucial para evitar distorções no *Sharpe Ratio*, particularmente em contextos de avaliação de desempenho ajustado ao risco. A conversão para base mensal equivalente seguiu a convenção de juros compostos, assegurando compatibilidade temporal com os retornos setoriais e permitindo comparações consistentes com *benchmarks* internacionais.

3.2 Procedimentos para Cálculo do Deflated Sharpe Ratio

Para o cálculo do *Deflated Sharpe Ratio* (DSR) é necessário seguir uma sequência de etapas que envolvem a estimação dos retornos setoriais, dos retornos excedentes, do *Sharpe Ratio* tradicional e de seus ajustes estatísticos. Os retornos dos índices



setoriais são calculados utilizando a abordagem de retorno logarítmico, amplamente adotada na literatura financeira por suas propriedades estatísticas. O retorno do setor i no período t é definido como:

$$R_{i,t} = \ln \left(\frac{P_{i,t}}{P_{i,t-1}} \right) \quad (1)$$

Onde $R_{i,t}$ representa o retorno do setor i no período t , $P_{i,t}$ corresponde ao preço do índice setorial i no período t , e $P_{i,t-1}$ denota o valor do índice setorial i no final do período anterior $t-1$. A partir desses retornos, obtêm-se os retornos excedentes mediante a dedução da taxa livre de risco:

$$R_{i,t}^{ex} = R_{i,t} - R_{f,t} \quad (2)$$

Sendo $R_{f,t}$ a taxa livre de risco no período t , representada pela SELIC convertida para base mensal equivalente.

Neste estudo, cada índice setorial é interpretado como uma estratégia candidata de investimento, de modo que a comparação entre setores envolve implicitamente um processo de seleção do melhor desempenho entre múltiplas alternativas. Com os retornos excedentes definidos, procede-se à estimação do *Sharpe Ratio*, que sintetiza o retorno excedente por unidade de risco:

$$SR_i = \frac{\bar{R}_i^{ex}}{\sigma_i} \quad (3)$$

Onde $\bar{R}_i^{ex}$ representa o retorno excedente médio do setor i e σ_i denota o desvio padrão dos retornos excedentes do setor i , conforme a formulação:

$$\sigma_i = \sqrt{\frac{1}{T-1} \sum_{t=1}^T (\bar{R}_{i,t}^{ex} - \bar{R}_i^{ex})^2} \quad (4)$$

Apesar de sua ampla utilização como métrica de desempenho ajustado ao risco, o *Sharpe Ratio* apresenta limitações relevantes em contextos de não-normalidade e múltiplas comparações. Diante dessas limitações, são estimados momentos estatísticos de



ordem superior, notadamente os coeficientes de assimetria e curtose. Com base nesses momentos, ajusta-se o erro padrão do *Sharpe Ratio*, conforme a formulação proposta por Bailey e López de Prado (2014). Assim, na sequência, com o objetivo de adequar a inferência estatística às características da distribuição empírica dos retornos, procedeu-se à estimação dos coeficientes de assimetria da distribuição dos retornos.

$$\gamma_{3,i} = \frac{\frac{1}{T} \sum_{t=1}^T (R_{i,t}^{ex} - \bar{R}_i^{ex})^3}{\sigma_i^3} \quad (5)$$

e o coeficiente de curtose, que captura a probabilidade de eventos extremos:

$$\gamma_{4,i} = \frac{\frac{1}{T} \sum_{t=1}^T (R_{i,t}^{ex} - \bar{R}_i^{ex})^4}{\sigma_i^4} \quad (6)$$

Esses momentos são fundamentais para ajustar as inferências estatísticas às características distribucionais reais dos retornos financeiros, que frequentemente divergem da normalidade assumida pelas medidas tradicionais de desempenho.

Com base nesses momentos, ajusta-se o erro padrão do *Sharpe Ratio*, conforme a formulação proposta por Bailey e López de Prado (2014). Uma vez ajustado o erro padrão do *Sharpe Ratio*, torna-se possível a construção do *Probabilistic Sharpe Ratio* (PSR), que fornece uma interpretação probabilística do desempenho observado. O PSR é definido como:

$$DSR = PSR(SR^*) = \Phi\left(\frac{SR - SR^*}{\sigma(SR)}\right) \quad (7)$$

Onde o erro padrão do *Sharpe* estimado é dado por:

$$\sigma(SR) = \sqrt{\frac{1 - S \cdot SR + \frac{K+2}{4} \cdot SR^2}{T-1}} \quad (8)$$

O PSR fornece, portanto, uma interpretação probabilística do *Sharpe* estimado, indicando a probabilidade de que o desempenho ajustado ao risco seja superior a um limiar predefinido. Todavia, quando múltiplas alternativas são avaliadas simultaneamente, a simples



aplicação do PSR pode levar a inferências excessivamente otimistas. Para corrigir esse viés, o *Deflated Sharpe Ratio* (DSR) redefine o limiar de comparação do PSR, incorporando o efeito da seleção *ex-post* do melhor desempenho entre múltiplas alternativas.

O DSR ajusta o PSR para o viés de seleção decorrente da escolha do melhor desempenho entre múltiplas alternativas. O parâmetro SR^* representa o *Sharpe* mínimo esperado sob a hipótese nula de ausência de habilidade, ajustado pelo número de alternativas testadas. Em outras palavras, trata-se do nível de desempenho que pode surgir exclusivamente por sorte quando se seleciona *ex-post* o melhor resultado entre múltiplos setores. O limiar SR^* pode ser calculado conforme descrito a seguir:

$$SR^* = \sqrt{V(SR_n)} \left((1 - \gamma) \Phi^{-1} \left(1 - \frac{1}{N} \right) + \gamma \Phi^{-1} \left(1 - \frac{1}{N_e} \right) \right) \quad (9)$$

Onde N é o número total de setores analisados, N_e representa o número efetivo de testes independentes, obtido a partir da matriz de correlação entre os setores, γ é a constante de Euler–Mascheroni.

O *Deflated Sharpe Ratio* (DSR) pode ser interpretado como a probabilidade de que o desempenho ajustado ao risco observado reflita habilidade econômica genuína, e não seja consequência de flutuações amostrais ou de vieses estatísticos associados ao *data-snooping*. Viés que ocorre quando múltiplas estratégias ou alternativas são testadas simultaneamente e apenas o melhor resultado *ex-post* é selecionado, levando à superestimação do desempenho e à confusão entre sorte e habilidade. Ao incorporar correções para múltiplos testes e dependência entre séries, o DSR fornece uma inferência estatística mais robusta sobre a significância do desempenho observado.

4 ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

A Tabela 2 apresenta as estatísticas descritivas e o desempenho ajustado ao risco dos índices setoriais analisados. Observa-se que os setores de Materiais Básicos (IMAT) e Utilidades Públicas (UTIL) apresentam os maiores retornos excedentes médios no período, bem como os maiores valores de *Sharpe Ratio*. Por outro lado, os setores de Consumo (ICON) e Imobiliário (IMOB) exibem retornos excedentes médios negativos e *Sharpe Ratio* negativo, indicando desempenho inferior ajustado ao risco.

Sob uma análise tradicional baseada exclusivamente no *Sharpe Ratio*, esses resultados sugeririam a superioridade relativa dos setores IMAT e UTIL em termos de desempenho ajustado ao risco no período analisado.

Tabela 2: Estatísticas descritivas e Sharpe tradicional

Setor / Índice	Retorno (%)	Risco (%)	Sharpe	Assimetria	Curtose
Consumo (ICON)	-0,79	7,08	-0,11	-1,27	5,65



Energia Elétrica (IEE)	0,17	5,92	0,03	-0,53	3,40
Financeiro (IFNC)	0,03	8,07	0,00	-0,77	4,12
Imobiliário (IMOB)	-0,43	9,23	-0,05	-1,47	8,00
Materiais Básicos (IMAT)	0,51	7,55	0,07	-0,49	0,43
Utilidades Públicas (UTIL)	0,41	6,45	0,06	-0,69	4,81

Fonte: Dados da Pesquisa (2026).

No entanto, essa interpretação desconsidera duas questões estatísticas relevantes: (i) a incerteza associada à estimação do *Sharpe Ratio* em amostras finitas e com retornos potencialmente não normais; e (ii) o viés de seleção decorrente da comparação simultânea de múltiplos setores, situação na qual o melhor desempenho observado pode emergir apenas por acaso.

Com o objetivo de incorporar essas correções estatísticas, na Tabela 3 se verifica os resultados do *Probabilistic Sharpe Ratio* (PSR) e do *Deflated Sharpe Ratio* (DSR). O PSR mensura a probabilidade de que o *Sharpe Ratio* verdadeiro seja superior a zero, considerando explicitamente o tamanho da amostra, bem como a assimetria e a curtose dos retornos. Os resultados indicam que os setores Materiais Básicos (76,45%) e Utilidades Públicas (75,23%) apresentam as maiores probabilidades de Sharpe positivo, seguidos pelo setor Energia Elétrica (62,52%). O setor Financeiro apresenta probabilidade próxima ao limiar de indiferença estatística (51,87%), enquanto os setores Imobiliário (29,82%) e Consumo (9,67%) exibem evidência estatística limitada de desempenho ajustado ao risco positivo.

Tabela 3: Avaliação estatística do desempenho (PSR e DSR)

Setor / Índice	Sharpe	PSR (%)	DSR (%)
Consumo (ICON)	-0,11	9,67	0,93
Energia Elétrica (IEE)	0,03	62,52	25,45
Financeiro (IFNC)	0,00	51,87	17,38
Imobiliário (IMOB)	-0,05	29,82	6,05
Materiais Básicos (IMAT)	0,07	76,45	40,13
Utilidades Públicas (UTIL)	0,06	75,23	38,91

Fonte: Dados da Pesquisa (2026).

Nota: O DSR corresponde ao PSR avaliado em um limiar de *Sharpe* mínimo SR^* , ajustado por múltiplas tentativas. Valores elevados indicam maior evidência estatística de desempenho ajustado ao risco puro.

Embora o PSR já incorpore correções relacionadas à incerteza amostral e à não normalidade dos retornos, ele não considera explicitamente o viés de seleção associado à escolha do melhor desempenho entre múltiplas alternativas. Para lidar com esse problema, utiliza-se o *Deflated Sharpe Ratio* (DSR), que ajusta a inferência estatística ao número de estratégias avaliadas e à variabilidade observada entre os *Sharpe* estimados. Após a aplicação dessa correção adicional, os resultados tornam-se substancialmente mais conservadores. Embora os setores Materiais Básicos (IMAT) e Utilidades Públicas (UTIL) permaneçam como aqueles com maior desempenho relativo, seus valores de DSR (40,13% e 38,91%, respectivamente) indicam que essa superioridade não apresenta robustez estatística após o ajuste para múltiplas comparações.



Em particular, nenhum dos setores analisados apresenta valores de DSR superiores ao limiar convencional de 95%, frequentemente utilizado para rejeitar a hipótese nula de ausência de habilidade. Mesmo os maiores valores observados situam-se abaixo de 50%, sugerindo evidência estatística limitada de desempenho ajustado ao risco genuíno. Assim, apesar das diferenças observadas nos valores do *Sharpe Ratio* tradicional, o desempenho superior aparente não pode ser distinguido estatisticamente de resultados que poderiam emergir por acaso, considerando o número de alternativas analisadas e a dependência entre elas.

Em síntese, os resultados evidenciam uma diferença clara entre o desempenho setorial avaliado por métricas tradicionais e a robustez estatística desse desempenho quando se consideram ajustes para incerteza amostral e viés de seleção. Sob a ótica do *Sharpe Ratio* tradicional, os setores Materiais Básicos (IMAT) e Utilidades Públicas (UTIL) apresentam a melhor relação risco-retorno no período analisado. Entretanto, a aplicação do PSR e, sobretudo, do DSR demonstra que essa superioridade é significativamente atenuada quando se consideram correções estatísticas mais rigorosas.

Além disso, o fato de nenhum setor apresentar valores de DSR superiores ao limiar convencional de 95% sugere que o desempenho observado é compatível com flutuações amostrais e com o processo de seleção *ex-post* do melhor resultado entre múltiplas alternativas. Em termos econômicos, isso indica que diferenças aparentes de desempenho entre setores podem refletir predominantemente a exposição a fatores comuns de mercado e a efeitos transitórios do ciclo econômico, e não necessariamente vantagens estruturais persistentes.

A seguir, os parâmetros reportados na Tabela 4 permitem compreender por que a correção promovida pelo *Deflated Sharpe Ratio* (DSR) resulta em inferências mais conservadoras. Em particular, o número de setores analisados ($N = 6$), aliado ao tamanho da amostra ($T = 120$) e à variância observada entre os *Sharpe* estimados, implica um limiar mínimo de *Sharpe* (SR^*) relativamente elevado. Esse limiar representa o nível de desempenho que poderia emergir apenas por efeito de sorte quando se seleciona *ex-post* o melhor resultado entre múltiplas alternativas correlacionadas. Dessa forma, o fato de os *Sharpe* observados situarem-se abaixo desse patamar reforça a interpretação de que as diferenças de desempenho entre os setores não constituem evidência estatística suficiente de habilidade setorial genuína, mas refletem o comportamento esperado sob a hipótese nula de ausência de habilidade.

Tabela 4: Parâmetros do teste

Parâmetro	Valor	Descrição
Número de estratégias (N)	6	Número de índices setoriais analisados
Tamanho da amostra (T)	120	Observações mensais
Taxa livre de risco	0,751%	SELIC mensal equivalente
Constante de Euler–Mascheroni (γ)	0,5772	Constante matemática
Variância entre estratégias	0,0049	Variância cross-sectional dos Sharpe
Sharpe mínimo exigido (SR^*)	9,36%	Limiar ajustado para múltiplos testes

Fonte: Dados da Pesquisa (2026).



De maneira geral, os resultados indicam que rankings setoriais baseados exclusivamente no *Sharpe Ratio* devem ser interpretados com cautela. Embora alguns setores apresentem desempenho superior sob métricas tradicionais, essa vantagem tende a se reduzir substancialmente quando se consideram correções para incerteza estatística e múltiplas comparações. Nesse sentido, os achados são consistentes com a distribuição nula do máximo *Sharpe Ratio* sob múltiplas tentativas, conforme discutido por Bailey e López de Prado (2014).

5 CONSIDERAÇÕES FINAIS

O presente estudo teve como objetivo avaliar a robustez estatística do desempenho ajustado ao risco de setores da economia brasileira, considerando explicitamente os efeitos de incerteza amostral e viés de seleção decorrentes da comparação simultânea de múltiplas alternativas. Para esse propósito, foram utilizadas, além do *Sharpe Ratio* tradicional, as métricas *Probabilistic Sharpe Ratio* (PSR) e *Deflated Sharpe Ratio* (DSR), conforme a abordagem proposta por Bailey e López de Prado (2014).

Os resultados indicam que a aparente superioridade de determinados setores, quando avaliada exclusivamente por métricas tradicionais de desempenho, não se mantém quando submetida a testes estatísticos mais rigorosos. Embora alguns setores tenham apresentado valores positivos de *Sharpe Ratio* e probabilidades relativamente elevadas no PSR, a aplicação do DSR revela que tais evidências não são estatisticamente robustas quando se consideram simultaneamente o número de estratégias analisadas e a variabilidade entre elas. Dessa forma, a incorporação dessas correções contribui para uma avaliação mais realista do desempenho setorial, reduzindo o risco de conclusões enviesadas por efeitos de sorte estatística ou seleção *ex-post*.

Em relação às hipóteses propostas, os resultados permitem algumas inferências relevantes. A hipótese H1, que estabelece a ausência de evidência estatística de habilidade genuína no desempenho setorial após correções para efeitos de sorte ou seleção *ex-post*, não pôde ser rejeitada, uma vez que nenhum dos setores analisados apresentou valores de DSR superiores ao limiar convencional de significância de 95%. Em contrapartida, a hipótese H2, segundo a qual ao menos um setor apresentaria desempenho ajustado ao risco estatisticamente robusto após a aplicação das correções metodológicas, foi rejeitada.

Os resultados também indicam a rejeição da hipótese H3, uma vez que o *Sharpe Ratio* tradicional, quando utilizado isoladamente, mostrou-se insuficiente para inferências estatísticas robustas sobre desempenho ajustado ao risco. Por fim, os achados oferecem suporte à hipótese H4, ao indicar que o *Sharpe Ratio* tradicional tende a superestimar a evidência de desempenho superior quando comparado a métricas ajustadas como o PSR e o DSR.

Do ponto de vista prático, os resultados apresentam implicações relevantes para investidores, gestores de portfólio e analistas de mercado. Em particular, a evidência de que setores com maior *Sharpe Ratio* não apresentam robustez estatística após correções para múltiplas comparações, sugere cautela no uso de rankings setoriais como base exclusiva para decisões de alocação de ativos. Para investidores institucionais, os achados reforçam a



importância de evitar estratégias de *stock picking* ou *sector rotation* baseadas apenas em desempenho histórico recente, uma vez que o setor aparentemente vencedor pode refletir apenas sorte estatística. Nesse sentido, os resultados indicam a relevância de abordagens mais diversificadas, como alocação setorial balanceada ou estratégias baseadas em fatores de risco amplos. Para gestores ativos, o estudo sugere que a busca por alfa setorial exige evidências adicionais além do *Sharpe* tradicional, incluindo testes de robustez fora da amostra, análise de estabilidade temporal e controle explícito para problemas de *data-snooping*.

Do ponto de vista acadêmico, os resultados dialogam com uma literatura recente que tem questionado a robustez de estratégias baseadas exclusivamente em desempenho histórico e métricas tradicionais de risco-retorno. Estudos recentes indicam que uma parcela significativa das estratégias documentadas na literatura pode refletir efeitos de amostragem ou correlação entre alternativas testadas, perdendo significância estatística quando se consideram correções para múltiplos testes (Feng et al., 2020; Gu et al., 2020; Harvey et al., 2016). Nesse contexto, a principal contribuição empírica deste estudo consiste em demonstrar que a aplicação do *Deflated Sharpe Ratio* (DSR) ao contexto da alocação setorial no mercado acionário brasileiro produz resultados consistentes com essa evidência internacional, sugerindo que grande parte das aparentes oportunidades de desempenho superior pode ser explicada por sorte estatística, e não por habilidade persistente.

Por fim, os resultados obtidos também abrem espaço para diversas extensões de pesquisa. Estudos futuros podem investigar se a robustez do desempenho setorial se altera quando se consideram janelas temporais móveis, permitindo avaliar a estabilidade do DSR ao longo de diferentes regimes econômicos. Outra possibilidade consiste na realização de testes fora da amostra, examinando se setores com maior DSR em determinado período mantêm desempenho superior em períodos subsequentes. Adicionalmente, pesquisas futuras podem integrar o DSR a modelos de alocação de carteiras setoriais, avaliando se estratégias de investimento que incorporam restrições de robustez estatística apresentam desempenho superior em termos de risco e retorno.

REFERÊNCIAS

B3 – Brasil, Bolsa, Balcão. (2025). *Composição da carteira do Ibovespa*. https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm

B3 – Brasil, Bolsa, Balcão. (2025). *Índices setoriais*. https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-segmentos-e-setoriais/

Banco Central do Brasil. (2025). *Histórico de taxas de juros*. <https://www.bcb.gov.br/controleinflacao/historicotaxasjuros>

Bailey, D. H., & López de Prado, M. (2012). The Sharpe ratio efficient frontier. *Journal of Risk*, 15(2), 3–44.



Bailey, D. H., & López de Prado, M. (2014). The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality. *The Journal of Portfolio Management*, 40(5), 94–107.

Chinco, A., Clark-Joseph, A., & Ye, M. (2024). In-sample and out-of-sample Sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics*, 155, 103837. <https://doi.org/10.1016/j.jfineco.2024.103837>

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3), 1327–1370. <https://doi.org/10.1111/jofi.12883>

Feng, G., Polson, N., & Xu, J. (2020). Deep learning in asset pricing. *Journal of Financial Economics*, 134(3), 567–595.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273. <https://doi.org/10.1093/rfs/hhaa009>

Harvey, C. R., & Liu, Y. (2015). Backtesting. *The Journal of Portfolio Management*, 42(1), 13–28. <https://doi.org/10.3905/jpm.2015.42.1.013>

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ...and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68. <https://doi.org/10.1093/rfs/hhv059>

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5), 2019–2133.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23(2), 389–416. <https://doi.org/10.1111/j.1540-6261.1968.tb00815.x>

Jiang, Y., Wang, Y., & Raza, M. W. (2022). The Sharpe ratio's upper bound of portfolios in the presence of a benchmark: Application to the US financial market. *Journal of Mathematical Finance*, 12, 566–581. <https://doi.org/10.4236/jmf.2022.123030>

Kao, L. J., Lee, C. F., & Lee, H. H. (2025). Estimated Sharpe ratio of asset returns with fat tails: Theory and empirical evidence. *Review of Quantitative Finance and Accounting*. <https://doi.org/10.1007/s11156-025-01474-6>

Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *The Journal of Portfolio Management*, 23(2), 45–54. <https://doi.org/10.3905/jpm.23.2.45>



Pham, N. V., & Duong, K. D. (2022). The performance measurement of generalized Sharpe ratio and economic performance measure: A hedge funds example. *Universal Journal of Accounting and Finance*, 10(1), 124–131. <https://doi.org/10.13189/ujaf.2022.100113>

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, 39(1), 119–138.

Sharpe, W. F. (1975). Adjusting for risk in portfolio performance measurement. *The Journal of Portfolio Management*, 1(2), 29–34.

Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, 21(1), 49–58.

Sahoo, S., & Kumar, S. (2024). Volatility spillover among the sectors of emerging and developed markets: A hedging perspective. *Cogent Economics & Finance*, 12(1). <https://doi.org/10.1080/23322039.2024.2316048>

Salotra, G., & Pinsky, E. (2026). Sector rotation strategies in the TSX 60: A comprehensive analysis of risk-adjusted returns, machine learning applications, and out-of-sample validation. *Journal of Risk and Financial Management*, 19(1), 70. <https://doi.org/10.3390/jrfm19010070>

Senescall, M., & Low, R. K. Y. (2024). Quantitative portfolio management: Review and outlook. *Mathematics*, 12(18), 2897. <https://doi.org/10.3390/math12182897>

Sortino, F. A., & Van der Meer, R. (1991). Downside risk. *The Journal of Portfolio Management*, 17(4), 27–31. <https://doi.org/10.3905/jpm.1991.409343>

Treynor, J. L. (1965). How to rate management of investment funds. *Harvard Business Review*, 43(1), 63–75.

Zhang, Y., et al. (2025). Quantile spillover effects and sector dynamics in U.S. stock markets: Normal vs. extreme market conditions. *Finance Research Letters*, 83, 107608.