

Regimes Focados em Alocação e Aplicações em Investimento Dinâmico por Fatores no Mercado Acionário Brasileiro: Replicação e Adaptações Metodológicas

Arthur Formiga Kevin Lima Luisa Obara Luis Poliseli
Rodrigo Kandir Théo Ogier

11 de janeiro de 2026

Resumo

Este trabalho apresenta uma replicação e adaptação, para o mercado acionário brasileiro, de uma estrutura híbrida de identificação e previsão de regimes *focados em alocação* (*allocation-focused regimes*) aplicada a investimento dinâmico por fatores. A noção central é definir regimes não como estados macroeconômicos amplos, mas como períodos persistentes nos quais uma estratégia (ou fator) apresenta desempenho relativo sistematicamente superior a outro, tornando o regime diretamente acionável para decisão de alocação. A metodologia adotada desacopla (i) a etapa de rotulagem em-sample, conduzida por um modelo estatístico com penalização explícita de trocas de regime (*jump penalty*), e (ii) a etapa de previsão out-of-sample, conduzida por um classificador de árvores por gradiente (XGBoost) treinado sobre um conjunto enriquecido de *features* financeiras e macro. A ligação entre as etapas é realizada por seleção de hiperparâmetros orientada por desempenho: escolhe-se a penalização de salto que maximiza retorno ajustado ao risco em janelas de validação, e aplica-se o par rotulador-classificador a janelas subsequentes em um procedimento rolante. No Brasil, a implementação exigiu adaptações substantivas: (a) restringimos a amostra a 2000–2024 em virtude de indisponibilidade, baixa granularidade e confiabilidade heterogênea de dados históricos; (b) construímos portfólios fatoriais long-only com rebalanceamentos mensais ou anuais e pesos por valor de mercado, a partir de filtros de liquidez e consistência de preços/volumes; e (c) limitamos o bloco macroeconômico do XGBoost às variáveis câmbio, preço do petróleo e juros de 3 meses, reconhecendo que a ausência de proxies desejáveis (por exemplo, índice de volatilidade e juros longos) reduz eficiência preditiva. Discutimos ainda limitações prováveis e relevantes,

com ênfase em vieses de antecipação (*forward bias*) e seus mecanismos potenciais na construção do universo investível, no alinhamento temporal de fundamentos e no reuso de estatísticas calculadas com informação futura, o que pode inflar artificialmente a performance fora da amostra e, ao mesmo tempo, degradar a robustez econômica do sinal quando implantado em tempo real. Apresentamos resultados fora da amostra para estratégias ativas (fator versus EW) e complementares (rotação entre pares), enfatizando a estabilidade do hiperparâmetro λ^* ao longo das janelas e a relação entre acurácia de classificação e desempenho econômico.

1 Introdução

Mercados financeiros exibem dinâmica não estacionária, com dependências temporais, quebras estruturais e mudanças na relação entre sinais e retornos. Embora a literatura de *regime switching* frequentemente defina regimes como estados macroeconômicos agregados (por exemplo, alta/baixa volatilidade, expansão/recessão, inflação elevada/baixa), a utilidade prática para alocação entre estratégias concorrentes depende menos de caracterizações gerais e mais da previsibilidade do *desempenho relativo* entre alternativas de investimento. Em carteiras multi-estratégia, o problema decisório relevante é: dado um conjunto de estratégias admissíveis, quando realocar capital de forma sistemática de uma estratégia para outra, minimizando exposição a períodos em que a estratégia escolhida tende a subperformar sua alternativa?

A abordagem replicada neste artigo formaliza regimes diretamente a partir dessa pergunta. Em vez de inferir um “estado do mundo” e, apenas depois, mapear tal estado para uma decisão de alocação, definimos regimes como segmentos temporais em que um portfólio hipotético long-short entre duas estratégias (ou entre um fator e um benchmark) apresenta sinal persistente de desempenho relativo. Essa definição tem implicações importantes: (i) o regime já nasce alinhado com a ação de alocar, o que reduz a lacuna entre modelagem e execução; (ii) a variável alvo da previsão torna-se o rótulo do regime de *outperformance* versus *underperformance*, potencialmente mais estável e previsível do que retornos absolutos; e (iii) a avaliação final pode ser realizada em termos de métricas diretamente alinhadas ao objetivo do investidor (Sharpe, CAGR, estabilidade de alocação sob custos), em vez de medidas puramente estatísticas de ajuste.

A replicação para o Brasil enfrenta desafios estruturais: dados menos longos e mais ruidosos, mudanças frequentes em regras de mercado e microestrutura, e disponibilidade incompleta de variáveis macro de alta qualidade e frequência compatível. Por isso, adotamos uma janela amostral menor (2000–2024), aplicamos filtros de liquidez e padronização de tickers, construímos portfólios fatoriais long-only com rebalanceamento definido e pesos por valor de mercado, e implementamos um pipeline híbrido com regularização

temporal explícita e previsão supervisionada com XGBoost. O resultado é um arcabouço metodológico coerente com o desenho original, porém adaptado às restrições brasileiras e acompanhado de uma discussão crítica de limitações, particularmente as associadas a *forward bias*, cujos efeitos podem ser substanciais em ambientes com defasagens contábeis, revisões de bases e sobrevivência de ativos.

2 Definição de regimes focados em alocação

Considere duas estratégias admissíveis, denotadas por S_1 e S_2 , com retornos diários $r_{1,t}$ e $r_{2,t}$. Define-se o retorno do portfólio relativo (hipotético) long-short como

$$\ell_t = r_{1,t} - r_{2,t}. \quad (1)$$

Um regime focado em alocação é, por construção, um estado latente $s_t \in \{0, 1\}$ que particiona o tempo em períodos nos quais S_1 tende a dominar S_2 ($s_t = 1$) versus períodos nos quais S_2 tende a dominar S_1 ($s_t = 0$). A operacionalização desse conceito depende de duas propriedades: (i) os regimes devem ser *persistentes* o suficiente para serem aprendidos e explorados economicamente, evitando rótulos que alternem de forma errática a cada dia; e (ii) o processo de rotulagem não deve embutir informação futura por construção, preservando uma separação temporal rigorosa entre identificação e teste.

A persistência é crítica porque, em finanças, retornos diários exibem ruído elevado e reversões de curtíssimo prazo; se os rótulos forem equivalentes ao sinal de ℓ_t sem regularização, obtém-se um alvo com alta variância condicional e baixa capacidade preditiva fora da amostra. Por isso, na etapa de identificação em-sample utilizamos uma regularização temporal explícita via penalidade de troca de regime, o que cria regimes mais contínuos e coerentes com a noção econômica de “períodos” de dominância relativa.

3 Universo de ativos e construção dos portfólios fatoriais no Brasil

A implementação brasileira parte de preços ajustados diários de ações, organizados em matriz $P_{i,t}$, com i representando o ticker e t dias úteis. A amostra efetiva é restrita a 2000–2024 por limitações de disponibilidade e qualidade de dados: em mercados emergentes, a cobertura histórica tende a ser curta e sujeita a rupturas de série, mudanças de classificação e descontinuidades de negociação, o que torna a extensão do horizonte (como em estudos com EUA desde meados do século XX) metodologicamente frágil e, muitas vezes, enganosa do ponto de vista inferencial. A restrição do período não é meramente um detalhe: ela altera o conjunto de regimes plausíveis, o aprendizado do

classificador e o próprio trade-off entre viés e variância na validação rolante.

3.1 Filtros de liquidez e padronização de tickers

Antes da construção dos fatores, definimos um universo investível. O pipeline impõe: (i) presença mínima de observações de preço ao longo do painel (alta proporção de dias com preço não ausente), (ii) preço médio mínimo para reduzir microcaps e séries com problemas de cotação, (iii) filtro de volume médio diário acima de um limiar para aproximar negociabilidade, e (iv) padronização de tickers por raiz, com preferência por classes mais líquidas quando aplicável, para reduzir duplicidade e minimizar distorções por estrutura de capital. Após esses filtros, valores ausentes remanescentes são preenchidos por carregamento do último preço (*forward fill*) com a finalidade de permitir cálculo de retornos e rebalanceamentos sem que a rotina quebre por lacunas ocasionais.

Embora tais filtros sejam necessários para aproximar viabilidade operacional, eles são também um ponto sensível para vieses: medidas como “presença mínima” e “volume médio” são frequentemente calculadas usando todo o período disponível e, se não forem ancoradas em janelas históricas estritamente anteriores ao ponto de decisão, podem incorporar informação futura sobre sobrevivência e negociabilidade, alterando simultaneamente o conjunto investível e a dinâmica dos regimes inferidos.

3.2 Retornos e convenções de rebalanceamento

A partir dos preços ajustados, definimos retornos diários por ativo

$$r_{i,t} = \frac{P_{i,t}}{P_{i,t-1}} - 1, \quad (2)$$

com tratamento de ausências pontuais de forma conservadora para evitar instabilidades numéricas na agregação de carteiras.

Os portfólios fatoriais são long-only e seguem duas ideias dominantes: (i) seleção por quantis em sinais fundamentalistas ou de retorno passado e (ii) ponderação por valor de mercado dentro do subconjunto selecionado, aproximando uma implementação replicável e reduzindo concentração em ativos pequenos. A frequência de rebalanceamento é mensal para fatores cujos sinais são mensais (por exemplo, *size* e *value* via BM com atualização trimestral e *forward-fill* mensal), e anual para o fator *growth* baseado em P/E formado em junho e mantido até o próximo junho, uma convenção compatível com a disponibilidade típica de fundamentos anuais e com o objetivo de reduzir turnover excessivo.

3.3 Fator *Value* via Book-to-Market (BM)

O fator *value* é construído a partir de BM, obtido como inverso do múltiplo preço-valor patrimonial (P/B) quando positivo. Observações trimestrais são mapeadas para o fim do trimestre e então reamostradas para frequência mensal por *forward-fill*. Em cada data de rebalanceamento mensal, define-se o universo válido como a interseção entre tickers com BM disponível, valor de mercado positivo e preços presentes. Ordena-se BM de forma decrescente e seleciona-se o quartil superior (top 25%), com imposição de um número mínimo de ativos para evitar carteiras degeneradas. Os pesos são proporcionais ao valor de mercado no instante de formação, e os retornos diários do fator são a soma ponderada dos retornos dos constituintes no subperíodo até o próximo rebalanceamento.

3.4 Fator *Growth* via P/E com rebalanceamento anual em junho

O fator *growth* utiliza P/E como sinal de alta precificação relativa. Computa-se P/E por

$$PE_{i,t} = \frac{\text{MktCap}_{i,t}}{\text{Earnings}_{i,y(t)}}, \quad (3)$$

onde $\text{Earnings}_{i,y(t)}$ são lucros anuais associados ao ano calendário $y(t)$. A implementação realiza *forward-fill* do lucro anual por ticker para preencher anos com dados faltantes dentro do painel de market cap mensal. Em cada mês de junho, forma-se o universo do fator selecionando o quartil superior de P/E entre ativos com P/E definido (lucro positivo e market cap válido). A carteira é então mantida por doze meses, com pesos por valor de mercado na data de formação, e retorna diariamente até o próximo junho.

Essa escolha anual é tecnicamente relevante: reduz turnover e evita que flutuações de curtíssimo prazo no preço dominem a composição, mas introduz um ponto de atenção central no Brasil: a defasagem de divulgação de resultados anuais e a possibilidade de que o lucro do próprio ano esteja sendo usado sem atraso de reporte, constituindo uma fonte clássica de *forward bias* se não for tratada com um calendário de divulgação realista.

3.5 Fatores *Size: Small e Big*

O fator *small* é a carteira do quartil inferior de valor de mercado em cada rebalanceamento mensal, com pesos proporcionais ao valor de mercado dentro do subconjunto, e retornos diários até o próximo rebalance. O fator *big* é construído analogamente como o quartil superior de valor de mercado.

3.6 Fatores de retorno passado: *Momentum* e *Mean Reversion*

Para *momentum*, utiliza-se um retorno acumulado em janela de $t - 12$ a $t - 2$ meses (excluindo o último mês para reduzir reversões de curtíssimo prazo). Em cada data mensal, calcula-se, para cada ativo com observações suficientes, o retorno composto na janela; seleciona-se o quartil superior e limita-se o número máximo de tickers, ponderando-se por valor de mercado na data de formação. O fator de *mean reversion* (ou reversão) é construído simetricamente, selecionando o quartil inferior do mesmo retorno acumulado.

3.7 Benchmark *Equally-Weighted* (EW)

O benchmark EW é construído mensalmente como a média simples dos retornos dos ativos com preço mensal não ausente na data de rebalanceamento e com retornos diários disponíveis no subperíodo. A carteira é rebalanceada a cada mês e mantém pesos iguais até o próximo rebalanceamento.

4 Estrutura metodológica: rotulagem com regularização temporal e previsão com XGBoost

A estratégia híbrida possui duas camadas: uma camada de identificação em-sample que produz rótulos de regime com persistência controlada, e uma camada de previsão supervisionada que mapeia *features* contemporâneas para o regime do próximo período.

4.1 Rotulagem em-sample por modelo com *jump penalty*

Denote por $\mathbf{x}_t \in \mathbb{R}^d$ o vetor de *features* de rotulagem no tempo t , construído a partir de estatísticas suavizadas do long-short ℓ_t . A etapa de identificação busca atribuir $s_t \in \{0, 1\}$ minimizando uma função objetivo que combina erro de agrupamento e penalidade por trocas de regime:

$$\min_{\{\mu_0, \mu_1\}, \{s_t\}} \sum_{t=1}^T \|\mathbf{x}_t - \mu_{s_t}\|_2^2 + \lambda \sum_{t=2}^T \mathbb{I}\{s_t \neq s_{t-1}\}, \quad (4)$$

onde μ_k são centróides de cada regime e $\lambda > 0$ controla persistência temporal. A implementação segue um procedimento iterativo: inicializa centróides via *k-means* e atualiza rótulos com um algoritmo de caminho ótimo sob penalidade de troca; em seguida, atualiza centróides pela média dos pontos atribuídos a cada classe e itera até convergência. O conjunto de *features* de rotulagem é propositalmente curto, baseado

em médias móveis exponenciais do long-short em janelas curtas, forçando os rótulos a refletirem dinâmica local do desempenho relativo.

Após rotular, os estados são reordenados para consistência: o regime “favorável” é definido como aquele com maior média de ℓ_t no treino, garantindo que $s_t = 1$ sempre signifique “fator domina benchmark” (ou “estratégia 1 domina estratégia 2”) dentro de cada janela, sem recorrer a informação futura, pois a comparação é feita estritamente em-sample.

4.2 Previsão supervisionada de regimes com XGBoost

Dada a série de rótulos em-sample, treinamos um classificador XGBoost com objetivo logístico binário para prever o regime do próximo dia. O alvo é defasado:

$$y_t = s_{t+1}. \quad (5)$$

O modelo aprende $p_t = \mathbb{P}(y_t = 1 \mid \tilde{\mathbf{x}}_t)$ e convertamos p_t em decisão de alocação por um limiar τ :

$$\hat{w}_t^{(f)} = \mathbb{I}\{p_t \geq \tau\}, \quad \hat{w}_t^{(b)} = 1 - \hat{w}_t^{(f)}, \quad (6)$$

com retorno bruto do portfólio ativo

$$r_t^{\text{gross}} = \hat{w}_t^{(f)} r_{f,t} + \hat{w}_t^{(b)} r_{b,t}. \quad (7)$$

A implementação incorpora histerese com duas bandas e permanência mínima para reduzir alternâncias espúrias, aproximando plausibilidade operacional sob custos e fricções.

4.3 Conjunto de *features* de previsão e bloco macro brasileiro

O conjunto de *features* de previsão combina estatísticas do long-short em múltiplas escalas (médias exponenciais em spans curtos e mensais e medidas de assimetria/instabilidade via componentes negativas suavizadas) e variáveis macroeconômicas. Por restrições de dados, utilizamos apenas câmbio, preço do petróleo e juros de 3 meses. A incompletude desse bloco macro é material: a ausência de índices de volatilidade e de juros longos reduz eficiência preditiva e pode induzir o classificador a depender excessivamente de proxies, elevando risco de instabilidade fora da amostra.

4.4 Seleção de hiperparâmetros orientada por desempenho e validação rolante

O hiperparâmetro crucial na rotulagem é a penalidade de salto λ . Valores baixos produzem rótulos reativos; valores altos produzem regimes persistentes demais. Escolhemos λ pelo desempenho econômico fora do treino, isto é, pelo Sharpe do portfólio ativo em validação, com penalização de turnover. O procedimento é rolante, com atualização trimestral: em cada iteração, recalibramos rotulador e classificador em janelas de treino/validação e testamos em uma janela subsequente, registrando λ^* e métricas de desempenho por janela.

4.5 Custos de transação e retorno líquido

O retorno líquido incorpora custos proporcionais à variação de posição:

$$r_t^{\text{net}} = r_t^{\text{gross}} - c \cdot |\Delta \hat{w}_t^{(f)}|, \quad (8)$$

onde c é o custo por mudança de posição em pontos-base.

5 Resultados empíricos fora da amostra

Esta seção consolida os resultados fora da amostra (OOS) para estratégias ativas (fator versus EW) e complementares (rotação entre pares), com ênfase em três dimensões: (i) desempenho ajustado ao risco (Sharpe, volatilidade anualizada, CAGR), (ii) qualidade estatística do classificador (acurácia média OOS) e (iii) estabilidade temporal da etapa de rotulagem, avaliada pela prevalência de λ^* por janela e sua relação com o desempenho (correlação entre acurácia e Sharpe por janela).

5.1 Estratégias ativas: fator versus EW

A Tabela 1 apresenta o resumo agregado das estratégias ativas. Observa-se que λ^* é extremamente estável e concentrado em 3 (mediana 3,0 e modo 3 em todas as estratégias), o que sugere que, dado o desenho de *features* de rotulagem e a penalidade por turnover embutida na seleção, a solução ótima recorrentemente favorece regimes com persistência moderada. Essa regularidade é relevante porque indica que o procedimento não está “caçando” hiperparâmetros idiossincráticos por janela, fenômeno típico de pipelines frágeis. Ao mesmo tempo, a estabilidade de λ^* não implica estabilidade econômica dos ganhos: a dispersão de Sharpes e CAGRs entre fatores indica heterogeneidade na previsibilidade do desempenho relativo, compatível com a hipótese de que certos prêmios

fatoriais no Brasil são mais dependentes de condições não observadas (por exemplo, estados de risco global) do que outros.

Tabela 1: Resumo agregado OOS — Estratégias ativas (fator vs. EW).

Estratégia	CAGR	Sharpe	Acurácia	Corr(Acc,Sharpe)
Value vs EW	0,432	0,939	0,569	0,092
Growth vs EW	0,181	0,262	0,429	0,829
Momentum vs EW	0,348	0,785	0,549	0,474
Reversal vs EW	0,362	0,745	0,554	0,246
Small vs EW	0,203	0,332	0,554	0,240
Big vs EW	0,219	0,392	0,563	0,392

Em termos econômicos, a estratégia *Value vs EW* destaca-se pela combinação de CAGR elevado (0,4329) e Sharpe OOS próximo a 0,94, com volatilidade anualizada de 28,16%. É importante notar que a acurácia média OOS (0,570) é apenas moderadamente superior a 0,5, e a correlação entre acurácia e Sharpe por janela (0,092) é baixa. Esse padrão é consistente com uma característica estrutural de problemas de alocação binária em finanças: acurácia média pode ser uma métrica pobre quando a assimetria do payoff é alta, pois poucas decisões corretas em regimes de forte divergência entre fator e benchmark podem dominar o desempenho, mesmo com acurácia global apenas marginalmente acima do aleatório. A estratégia *Momentum vs EW* e *Reversal vs EW* também exibem Sharpes OOS robustos (0,785 e 0,746), sugerindo que, no período testado, existe previsibilidade operacional no desempenho relativo desses fatores em relação a EW, ainda que em magnitudes inferiores ao *value*. Já *Small vs EW* e *Big vs EW* apresentam Sharpes mais baixos, coerentes com a hipótese de que, em mercados com microestrutura mais ruidosa, a decomposição por tamanho long-only é menos “regimeável” sem incorporar variáveis macro mais informativas e sem um tratamento fino de custos/impacto.

A estratégia *Growth vs EW* merece uma leitura distinta. Embora seu CAGR seja positivo (0,1815), o Sharpe OOS é baixo (0,263) e a acurácia média é inferior a 0,5 (0,429). Entretanto, a correlação entre acurácia e Sharpe por janela é extremamente alta (0,830), sugerindo que, quando o classificador de fato aprende algo útil em determinadas janelas, isso se traduz fortemente em performance; o problema é que o classificador falha em manter qualidade de forma consistente. Além disso, há janelas com Sharpes por janela numericamente extremos e negativamente muito grandes (por exemplo, valores próximos de $-39,598$), o que é um sintoma clássico de instabilidade computacional associada a volatilidade quase nula em retornos líquidos dentro da janela (denominador do Sharpe aproximando zero), ou a períodos com uma alocação quase determinística gerando retornos comprimidos. Do ponto de vista metodológico, isso é um alerta:

estratégias baseadas em P/E anual são particularmente sensíveis a *forward bias* de divulgação e a mudanças discretas na composição do universo em junho, e esses efeitos podem produzir janelas com distribuição de retornos anômala.

5.2 Retornos anuais e distribuição temporal do desempenho

Para dar concretude temporal ao resultado agregado, as Tabelas 2–7 reportam retornos por ano (teste) para as estratégias ativas. Esses números evidenciam duas propriedades relevantes. Primeiro, o retorno anual é altamente disperso e episódico, o que reforça a importância de regimes: anos com forte diferencial entre fator e EW geram “saltos” de performance quando o classificador acerta o estado dominante. Segundo, a presença de anos com retornos extraordinariamente altos (por exemplo, 2016 em múltiplas estratégias) é compatível com períodos de recuperação e reprecificação ampla, nos quais sinais relativos entre estilos podem tornar-se particularmente pronunciados; isso também significa que resultados agregados podem ser dominados por poucos anos, exigindo cautela na extrapolação.

Tabela 2: Retorno anual (teste) — Active VALUE vs EW.

Ano	Retorno
2011	−15,22%
2012	20,60%
2013	59,11%
2014	61,42%
2015	84,85%
2016	591,70%
2017	149,76%
2018	300,36%
2019	99,41%
2020	155,99%
2021	80,12%
2022	187,48%
2023	215,41%
2024	−8,07%
2025	19,12%

5.3 Estratégias complementares: rotação entre pares

A Tabela 8 resume as estratégias complementares. O resultado mais forte aparece em *Momentum vs Reversal*, com CAGR de 0,5770, Sharpe OOS de 1,1567 e volatilidade anualizada de 32,21%. Esse desempenho é consistente com a tese de complementaridade: momentum e reversão capturam extremos opostos de dinâmica de preços e tendem a

Tabela 3: Retorno anual (teste) — Active GROWTH vs EW.

Ano	Retorno
2013	9,78%
2014	4,56%
2015	−1,00%
2016	65,89%
2017	91,07%
2018	18,69%
2019	250,78%
2020	58,19%
2021	75,09%
2022	−7,40%
2023	59,76%
2024	−24,81%
2025	24,92%

Tabela 4: Retorno anual (teste) — Active MOMENTUM vs EW.

Ano	Retorno
2011	17,14%
2012	130,29%
2013	3,44%
2014	−3,33%
2015	1,60%
2016	177,49%
2017	315,97%
2018	158,11%
2019	98,98%
2020	28,31%
2021	170,28%
2022	99,31%
2023	221,82%
2024	−5,21%
2025	22,79%

dominar em períodos distintos; ao definir regimes diretamente sobre o long-short entre ambos, o alvo do classificador torna-se particularmente “alocável”, e o ganho vem menos de prever retornos absolutos e mais de prever a direção do diferencial. O par *Big vs Small* apresenta desempenho intermediário (Sharpe 0,638), sugerindo que a rotação entre estilos de tamanho captura parte de mudanças de ambiente, mas possivelmente carece de um bloco macro mais rico para discriminar estados de funding e risco global que afetam desproporcionalmente small caps.

O par *Value vs Growth* apresenta um contraste metodológico importante. Apesar de

Tabela 5: Retorno anual (teste) — Active REVERSAL vs EW.

Ano	Retorno
2011	11,86%
2012	35,48%
2013	100,02%
2014	−8,72%
2015	184,59%
2016	946,97%
2017	150,05%
2018	54,02%
2019	148,10%
2020	172,24%
2021	39,69%
2022	−13,51%
2023	115,27%
2024	−24,46%
2025	34,01%

Tabela 6: Retorno anual (teste) — Active SMALL vs EW.

Ano	Retorno
2011	23,28%
2012	37,66%
2013	6,99%
2014	−34,71%
2015	−30,32%
2016	503,38%
2017	132,87%
2018	54,62%
2019	113,51%
2020	102,24%
2021	72,66%
2022	17,25%
2023	80,40%
2024	−30,48%
2025	10,34%

um resumo agregado plausível (CAGR 0,2605, Sharpe 0,536, acurácia média 0,657), os resultados por janela exibem ocorrências de Sharpes numericamente absurdos (ordem de 10^{16} em magnitude) e acurácias iguais a 1,000 em múltiplas janelas, o que é incompatível com um processo estocástico realista e aponta para degenerescência numérica e/ou vazamento de informação em pontos específicos do pipeline (por exemplo, alinhamento temporal de P/E e BM, universos que se tornam quase determinísticos, ou retornos com variância quase nula em janelas curtas). Do ponto de vista de rigor, esse tipo de

Tabela 7: Retorno anual (teste) — Active BIG vs EW.

Ano	Retorno
2011	−12,69%
2012	27,02%
2013	24,26%
2014	25,08%
2015	−18,41%
2016	206,50%
2017	97,89%
2018	77,94%
2019	103,08%
2020	108,71%
2021	53,15%
2022	41,32%
2023	136,01%
2024	−20,32%
2025	22,03%

artefato não pode ser interpretado como evidência econômica; ao contrário, deve ser tratado como diagnóstico de falha experimental e reforça a centralidade da discussão de *forward bias* e de auditoria temporal.

Tabela 8: Resumo agregado OOS — Estratégias complementares (rotação entre pares).

Estratégia	CAGR	Sharpe	Acurácia	Corr(Acc,Sharpe)
Value vs Growth	0,260	0,536	0,657	−0,794
Momentum vs Reversal	0,577	1,156	0,545	0,524
Big vs Small	0,299	0,638	0,562	0,418

Os retornos anuais das estratégias complementares, nas Tabelas 9–11, reforçam que a superioridade do par *Momentum vs Reversal* não decorre de um único ano isolado, mas de uma sequência prolongada de anos com retornos elevados, apesar de dispersão e episódios de queda. Esse padrão é o que se espera quando o sinal de regimes é genuinamente acionável: não elimina drawdowns, mas melhora a relação retorno/risco ao reduzir exposição sistemática aos períodos em que o lado “errado” do par domina.

5.4 Estabilidade de λ^* e a relação entre acurácia e desempenho

A concentração de λ^* em 3 é uma evidência empírica de que, no conjunto de dados e no desenho de validação, a penalidade de troca ótima é recorrentemente de baixa magnitude, mas não trivial, promovendo regimes relativamente persistentes sem “congelar” o estado. Esse resultado é consistente com o fato de que as *features* de rotulagem já são suavizadas

Tabela 9: Retorno anual (teste) — Complementary VALUE vs GROWTH.

Ano	Retorno
2013	41,38%
2014	19,26%
2015	0,00%
2016	140,88%
2017	102,22%
2018	252,69%
2019	148,44%
2020	22,13%
2021	13,68%
2022	38,05%
2023	113,15%
2024	-17,96%
2025	9,20%

Tabela 10: Retorno anual (teste) — Complementary MOMENTUM vs REVERSAL.

Ano	Retorno
2011	31,21%
2012	213,36%
2013	58,79%
2014	54,53%
2015	305,73%
2016	760,61%
2017	246,01%
2018	177,71%
2019	182,21%
2020	16,07%
2021	144,54%
2022	93,16%
2023	242,01%
2024	39,96%
2025	33,92%

por EWMA's; portanto, a penalidade adicional necessária para impedir alternância é moderada. Em termos de interpretação, $\lambda^* \approx 3$ implica que o modelo aceita trocar de regime quando há evidência suficientemente forte de mudança no agrupamento das *features*, mas penaliza oscilações de curtíssimo prazo, o que é exatamente a propriedade desejável para uma estratégia sujeita a custos de transação.

A correlação entre acurácia e Sharpe por janela varia substancialmente entre estratégias. Em *Value vs EW*, a correlação próxima de zero sugere que a acurácia média é pouco informativa sobre o payoff, um cenário comum quando o retorno do long-short

Tabela 11: Retorno anual (teste) — Complementary BIG vs SMALL.

Ano	Retorno
2011	64,69%
2012	58,62%
2013	8,55%
2014	30,49%
2015	−24,08%
2016	373,77%
2017	165,26%
2018	57,06%
2019	92,23%
2020	91,69%
2021	148,50%
2022	90,89%
2023	115,62%
2024	−22,13%
2025	24,83%

é altamente heterocedástico e concentrado em episódios. Em *Growth vs EW*, a correlação alta indica que, quando o classificador acerta, o ganho é proporcional, mas a inconsistência global aponta para instabilidade do sinal e possivelmente para problemas de alinhamento temporal do fator. Em estratégias complementares, a interpretação é ainda mais delicada: no par *Value vs Growth*, a correlação negativa forte coexistindo com episódios de acurácia perfeita é incompatível com uma leitura econômica direta e reforça que métricas de classificação podem ser dominadas por artefatos quando o pipeline produz janelas degeneradas.

6 Discussão crítica e limitações

A replicação em um mercado emergente exige explicitar riscos metodológicos com potencial de magnitude material. Além das limitações de dados e do conjunto macro truncado, há uma categoria de riscos particularmente relevante: vieses de antecipação (*forward bias*) e seus efeitos indiretos sobre regimes, treinamento e validação.

Primeiro, há *forward bias* na definição do universo investível quando filtros de liquidez são calculados com estatísticas agregadas sobre a amostra inteira. Exigir presença mínima ao longo de todo o painel ou volume médio no período seleciona ativos cuja negociabilidade futura já é conhecida ex post, removendo do passado ações que, em tempo real, teriam entrado no universo e depois saído por perda de liquidez, eventos corporativos ou descontinuidade de série. Esse mecanismo altera não apenas o nível de retorno, mas o próprio alvo de regimes, porque muda a composição do benchmark e dos

fatores exatamente nos períodos em que regimes tendem a ser mais pronunciados.

Segundo, há *forward bias* potencial no alinhamento temporal de fundamentos. Para BM trimestral e lucros anuais, a data de competência contábil não coincide com a data de disponibilidade para o investidor, e revisões podem ocorrer. Se o pipeline usa o valor “do trimestre” como se fosse conhecido no fechamento do trimestre, ou usa o lucro “do ano” como se fosse conhecido em junho sem defasagem mínima de reporte, o sinal torna-se parcialmente *look-ahead*. Em modelos flexíveis, esse vazamento pode se manifestar como previsibilidade artificial e, por consequência, como janelas com métricas extremas e não realistas.

Terceiro, o uso de *forward-fill* em variáveis macro e em séries de fundamentos, embora necessário para casar frequências, pode inadvertidamente colar valores futuros em datas anteriores se a origem da série já vem com carimbos temporais que representam competência e não divulgação. Em ambientes com dados heterogêneos, esse é um vetor comum de vazamento sutil.

Quarto, a validação rolante com janelas sobrepostas exige disciplina estrita: qualquer escolha de hiperparâmetro, limiar e regra de histerese deve depender apenas de validação estritamente anterior ao teste. Caso contrário, pode surgir um viés indireto de parametrização que não é visível como “uso do futuro” em uma linha única, mas como consequência do desenho experimental.

Quinto, os próprios resultados fornecem evidência de que parte do pipeline pode gerar degenerescência numérica em janelas específicas, como Sharpes por janela absurdamente grandes em magnitude e acurácia perfeita em múltiplas janelas no par *Value vs Growth*. Esse padrão é típico quando a variância do retorno na janela colapsa (Sharpe explode), quando há alocação quase determinística combinada com retornos comprimidos, ou quando o alvo de classificação se torna trivial por vazamento de informação. Em qualquer dessas hipóteses, a interpretação econômica deve ser suspensa até que se audite o alinhamento temporal das variáveis e o mecanismo de construção do universo.

Finalmente, o horizonte amostral 2000–2024, embora pragmaticamente necessário, implica que o conjunto de regimes observados é menor e que transições podem ser dominadas por alguns episódios extremos. Isso eleva a sensibilidade do resultado agregado a poucos anos e reforça a necessidade de testes adicionais de robustez, como *stress tests* por subperíodos e imposição explícita de defasagens de reporte em fundamentos.

7 Conclusão

Replicamos para o mercado acionário brasileiro uma estrutura de regimes focados em alocação com duas etapas desacopladas: identificação em-sample com regularização temporal via *jump penalty* e previsão out-of-sample com XGBoost sobre um conjunto en-

riquecido de *features*. Construímos uma família de fatores long-only (value, growth, size, momentum e reversão) e um benchmark EW com regras explícitas de rebalanceamento e ponderação, e implementamos estratégias ativas (fator vs. EW) e complementares (rotação entre pares). A calibração foi conduzida em procedimento rolante com atualização trimestral e seleção de hiperparâmetros orientada por desempenho, incluindo penalização de turnover e regras de histerese para estabilização operacional.

Os resultados fora da amostra sugerem que a abordagem é economicamente promissora quando aplicada a fatores cujo diferencial de retorno em relação ao benchmark é suficientemente regime-dependente e quando o pipeline não é dominado por fragilidades de dados. Entre as estratégias ativas, *Value vs EW* apresenta o melhor balanço risco-retorno (CAGR 0,4329, Sharpe 0,940 e volatilidade anualizada 28,16%), enquanto *Momentum vs EW* e *Reversal vs EW* exibem Sharpes robustos (0,785 e 0,746), coerentes com previsibilidade operacional do desempenho relativo. Em contraste, *Growth vs EW* apresenta baixo Sharpe agregado (0,263) e sinais de instabilidade por janela, sugerindo que, com dados brasileiros e bloco macro truncado, o fator baseado em P/E anual é particularmente sensível a alinhamento temporal e a regimes não observados. Entre as estratégias complementares, o par *Momentum vs Reversal* se destaca com Sharpe OOS de 1,157 e CAGR 0,577, oferecendo a evidência empírica mais forte de que regimes definidos diretamente sobre o long-short entre estratégias complementares podem ser mais “alocáveis” do que regimes definidos contra um benchmark amplo.

Um achado metodológico transversal é a estabilidade de λ^* , com modo e mediana iguais a 3 em todas as estratégias. Essa concentração indica que o procedimento de seleção recorrentemente favorece uma persistência moderada de regimes, compatível com a suavização das *features* de rotulagem por EWMA e com o objetivo econômico de reduzir alternância sob custos. Todavia, a relação entre acurácia e desempenho é heterogênea e, em alguns casos, pouco informativa: a correlação próxima de zero em *Value vs EW* reforça que acurácia média não é, por si, um proxy confiável de ganho econômico quando o payoff é episódico e assimétrico; já correlações elevadas em *Growth vs EW* sugerem que, quando o modelo acerta, o ganho aparece, mas a inconsistência global indica falta de robustez do sinal.

Ao mesmo tempo, os próprios resultados revelam sintomas de degenerescência e provável vazamento em janelas específicas, em particular no par *Value vs Growth*, onde aparecem acurácias perfeitas e Sharpes por janela numericamente absurdos. Esses artefatos não podem ser tratados como evidência econômica e reforçam a conclusão central deste estudo: a aplicabilidade do framework no Brasil é condicionada à qualidade temporal das bases e à disciplina de backtest. As próximas extensões naturais são (i) impor defasagens realistas de divulgação contábil e reconstruir o universo investível em tempo real para mitigar *forward bias*; (ii) enriquecer substancialmente o bloco macro

com proxies de volatilidade e juros longos para melhorar discriminação de estados de risco; e (iii) incorporar métricas de robustez por subamostras e limites explícitos de estabilidade numérica (por exemplo, tratamento de janelas com volatilidade quase nula). Com essas melhorias, a estrutura de regimes focados em alocação tende a se tornar não apenas um exercício de replicação, mas um arcabouço competitivo e auditável para rotação dinâmica de estilos e fatores no mercado acionário brasileiro.

Referências

Referências

- [1] Yu, C.; Mulvey, J. M.; Nie, Y. *Allocation-Focused Regimes and Applications to Dynamic Factor Investing*. SSRN (working paper), 29 abr. 2025. :contentReference[oaicite:1]index=1
- [2] Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, v. 57, n. 2, p. 357–384, 1989.
- [3] Ang, A.; Bekaert, G. International asset allocation with regime shifts. *The Review of Financial Studies*, v. 15, n. 4, p. 1137–1187, 2002.
- [4] Ang, A.; Timmermann, A. Regime changes and financial markets. *Annual Review of Financial Economics*, v. 4, n. 1, p. 313–337, 2012.
- [5] Sharpe, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, v. 19, n. 3, p. 425–442, 1964.
- [6] Fama, E. F.; French, K. R. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, v. 33, n. 1, p. 3–56, 1993.
- [7] Fama, E. F.; French, K. R. A five-factor asset pricing model. *Journal of Financial Economics*, v. 116, n. 1, p. 1–22, 2015.
- [8] Carhart, M. M. On persistence in mutual fund performance. *The Journal of Finance*, v. 52, n. 1, p. 57–82, 1997.
- [9] Daniel, K.; Moskowitz, T. J. Momentum crashes. *Journal of Financial Economics*, v. 122, n. 2, p. 221–247, 2016.
- [10] Da, Z.; Liu, Q.; Schaumburg, E. A closer look at the short-term return reversal. *Management Science*, v. 60, n. 3, p. 658–674, 2014.
- [11] Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016.
- [12] Estrella, A.; Mishkin, F. S. The yield curve as a predictor of U.S. recessions. *Current Issues in Economics and Finance*, v. 2, n. 7, p. 1–6, 1996.