



UNIVERSIDADE DE BRASÍLIA

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE
E GESTÃO DE POLÍTICAS PÚBLICAS

DEPARTAMENTO DE CIÊNCIAS CONTÁBEIS E ATUARIAIS

Bacharelado em Ciências Contábeis

Nota Técnica nº 01/MQC

Data: 09/12/2025

**PETROBRAS (PETR4): ANÁLISE DE VALOR JUSTO E RECOMENDAÇÃO DE
INVESTIMENTO VIA MODELOS QUANTITATIVOS E ANÁLISE DE EFICIÊNCIA
INFORMACIONAL**

Empresa analisada: Petróleo Brasileiro S.A. (Petrobras)

Setor: Petróleo, Gás e Biocombustíveis

Período de análise: 2016 a 2025 (frequência: diária/trimestral)

Autor(es):

Lucas Coelho França

Brasília – DF

2025

Professora Doutora Márcia Abrahão Moura

Reitora da Universidade de Brasília

Professor Doutor Enrique Huelva Unternbäumen

Vice-Reitor da Universidade de Brasília

Professor Doutor Sérgio Antônio Andrade de Freitas

Decano de Ensino de Graduação

Professor Doutor Eduardo Tadeu Vieira

Diretor da Faculdade de Economia, Administração, Contabilidade
e Gestão de Políticas Públicas

Professor Doutor Paulo César de Melo Mendes

Chefe do Departamento de Ciências Contábeis e Atuariais

Professor Doutor Alex Laquis Resende

Coordenador de Graduação do curso de Ciências Contábeis – Diurno

Professor Mestre Elivânio Geraldo de Andrade

Coordenador de Graduação do curso de Ciências Contábeis – Noturno

SUMÁRIO

1	Introdução	10
2	Fundamentos Teóricos	13
2.1	O Sistema de Preços como Mecanismo Informacional	13
2.1.1	A Contribuição Hayekiana: Conhecimento Disperso e Coordenação	13
2.1.2	Do Conhecimento Tácito ao Conhecimento Proposicional	14
2.1.3	Mercados como Processadores de Informação	14
2.2	Eficiência de Mercado: Da EMH aos Mercados Adaptativos	15
2.2.1	A Hipótese dos Mercados Eficientes: Formulação e Taxonomia	15
2.2.2	Evidência Empírica: Anomalias e Prêmios de Fatores	16
2.2.3	A Hipótese dos Mercados Adaptativos	17
2.2.4	Implicações para Análise Fundamentalista	19
2.3	O Paradoxo de Grossman-Stiglitz e o Custo da Informação	19
2.3.1	O Teorema de Impossibilidade	19
2.3.2	Formalização: Equilíbrio com Ruído	20
2.3.3	Extensões Contemporâneas e Evidência Empírica	21
2.3.4	O Motor Q-VAL como “Custo da Informação”	22
2.4	Economia da Complexidade e Agregação Informacional	22
2.4.1	Mercados como Sistemas Adaptativos Complexos	22
2.4.2	Agregação de Informação como Fenômeno Emergente	23
2.4.3	Herding, Cascatas e Amplificação de Ruído	24
2.4.4	Implicações para o Motor Q-VAL	24
2.4.5	Síntese: Quatro Visões sobre Informação e Preços	25
3	Contextualização: Petrobras e a Margem Equatorial	27
3.1	Trajetória Histórica e Institucional	27
3.2	O Cenário Atual: Margem Equatorial e Transição Energética	28
3.3	Implicações para Valuation	28
4	Metodologia	30
4.1	O Motor Q-VAL como Vetor de Informação Fundamentalista	30
4.1.1	Arquitetura Conceitual	30
4.1.2	Métricas Seleccionadas	31
4.1.3	Procedimento de Normalização	33
4.1.4	Agregação em Score Composto e Abordagem Machine Learning	34
4.1.5	Série Temporal de Scores	35
4.2	Modelos Econométricos Comparativos	36
4.2.1	Estratégia de Modelos Aninhados (M0 a M5)	36
4.2.2	Detalhamento dos Modelos	37
4.3	Métricas de Avaliação Informacional	39
4.3.1	Coefficiente de Determinação e Sua Variação	39
4.3.2	R^2 Ajustado	40
4.3.3	Crítérios de Informação	41
4.3.4	Testes de Significância Estatística	41
4.3.5	Validação Out-of-Sample	42
4.3.6	Métricas Complementares	43
4.3.7	Estratégias de Investimento (Fair Value)	44

4.4	Fontes de Dados e Período de Análise	45
4.4.1	Dados Utilizados	45
4.4.2	Tratamento de Dados	45
4.4.3	Justificativa do Período	45
4.5	Reprodutibilidade Computacional	46
4.6	Síntese Metodológica	46
5	Resultados Empíricos	47
5.1	A Linha de Base (M0 e M1)	47
5.2	O Impacto da Dinâmica (M1 vs M2)	47
5.3	A Contribuição dos Fundamentos (M2 vs M3)	48
5.4	O Papel do Macro e Fatores (M3 vs M4)	48
5.5	A Perda de Informação na Agregação (M4 vs M5)	48
5.6	A Armadilha da Linearidade: Granularidade vs. Agregação (M5a vs M5b)	49
5.7	Síntese: Fronteira Não-Linear e Machine Learning	50
5.7.1	Arquitetura do Meta-Modelo	50
5.7.2	Resultados do Backtest Comparativo: A Hegemonia Não-Linear	51
5.7.3	O Paradoxo da Explicação vs. Predição Resolvido	53
5.7.4	Evidência de Adaptabilidade (AMH)	54
5.7.5	Diagnóstico Fundamentalista Atual	55
5.8	Discussão	55
5.9	Adjudicação das Hipóteses Teóricas: As Quatro Visões em Confronto	56
5.9.1	A Visão de Fama (EMH): Parcialmente Refutada	56
5.9.2	A Visão de Grossman-Stiglitz: Confirmada	56
5.9.3	A Visão Hayekiana: Parcialmente Confirmada, com Qualificações	57
5.9.4	A Visão da Economia da Complexidade: Fortemente Confirmada	58
5.10	Implicações para a Avaliação de Ativos em Mercados Emergentes	59
5.11	Limitações Metodológicas e Humildade Epistêmica	59
5.11.1	O Viés da Idiossincrasia e a Falácia da Indução	59
5.11.2	Significância Estatística vs. Relevância Econômica	59
5.11.3	Dependência de Regime e Viés de Sobrevivência	60
5.11.4	O Risco do “Look-Ahead Bias” Meta-Metodológico	60
6	Conclusão e Recomendação	61
6.1	Síntese dos Achados Teóricos	61
6.2	Implicações Práticas	61
6.3	Análise Final e Recomendação (Data-Driven)	61
6.4	Contribuição Teórica	62
6.5	Direções Futuras	62
	Referências	63

Lista de Figuras

1	Matriz de Correlação dos Z-Scores Fundamentais. A baixa correlação entre as dimensões de Valor e Qualidade sugere que elas capturam informações ortogonais, justificando a abordagem multidimensional.....	34
2	Evolução da Eficiência Informacional (R^2 Out-of-Sample). Note a superioridade do M5-ML (Granular) sobre o M4 (Macro), enquanto a abordagem linear granular falha.	50
3	Abertura da Caixa Preta: Importância das Variáveis no Modelo M5b (XGBoost). Note como variáveis de Valor (EV/EBITDA) e Volatilidade dominam a predição, confirmando a natureza híbrida da estratégia.	51
4	Evolução do Capital: Estratégia de Valor Justo (2023-2024). O modelo M5b (Vermelho) destaca-se como a única estratégia ativa capaz de superar consistentemente o Buy & Hold (Preto) e o CDI (Cinza).	52
5	Evidência de Adaptabilidade: Rolling R^2 (12 Meses). A eficiência relativa do ML sobre o Linear varia no tempo, confirmando a natureza dinâmica do mercado.	54
6	Radar de Fundamentos (Z-Scores). Valores externos indicam métricas favoráveis (Barato, Rentável, Seguro). O centro representa a média histórica.	55

Lista de Tabelas

1	Síntese das Visões Teóricas sobre Informação e Preços {#tbl:theoretical_synthesis}	25
2	Métricas da Dimensão de Valor {#tbl:metrics_value}	31
3	Métricas da Dimensão de Qualidade {#tbl:metrics_quality}	32
4	Métricas da Dimensão de Risco {#tbl:metrics_risk}	32
5	Hierarquia de Modelos Econométricos {#tbl:model_hierarchy}	36
7	Performance Comparativa dos Modelos (Out-of-Sample)	47
8	Performance da Estratégia de Valor Justo (2023-2024).....	52

LISTA DE ABREVIATURAS E SIGLAS

CAPM	Capital Asset Pricing Model (Modelo de Precificação de Ativos de Capital)
ICC	Implied Cost of Capital (Custo de Capital Implícito)
DDM	Dividend Discount Model (Modelo de Desconto de Dividendos)
EV	Enterprise Value (Valor da Empresa)
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization
EVS	Economic Value Spread (Spread de Valor Econômico)
P/L	Preço/Lucro
P/VP	Preço/Valor Patrimonial
ROE	Return on Equity (Retorno sobre Patrimônio Líquido)
ROACE	Return on Average Capital Employed (Retorno sobre Capital Empregado Médio)
AIC	Akaike Information Criterion (Critério de Informação de Akaike)
BIC	Bayesian Information Criterion (Critério de Informação Bayesiano)
EMH	Efficient Market Hypothesis (Hipótese dos Mercados Eficientes)
AMH	Adaptive Markets Hypothesis (Hipótese dos Mercados Adaptativos)
CVM	Comissão de Valores Mobiliários
B3	Brasil, Bolsa, Balcão
ESG	Environmental, Social and Governance (Ambiental, Social e Governança)
ROIC	Return on Invested Capital (Retorno sobre o Capital Investido)
SMB	Small Minus Big (Fator de Tamanho)
HML	High Minus Low (Fator de Valor)
RMW	Robust Minus Weak (Fator de Lucratividade)
CMA	Conservative Minus Aggressive (Fator de Investimento)
MSE	Mean Squared Error (Erro Quadrático Médio)
RMSE	Root Mean Squared Error (Raiz do Erro Quadrático Médio)
MAE	Mean Absolute Error (Erro Absoluto Médio)
CDI	Certificado de Depósito Interbancário
IBAMA	Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis
NYSE	New York Stock Exchange
LLM	Large Language Models (Grandes Modelos de Linguagem)

HFT	High-Frequency Trading (Negociação de Alta Frequência)
EMBI	Emerging Markets Bond Index (Índice de Títulos de Mercados Emergentes)
WACC	Weighted Average Cost of Capital (Custo Médio Ponderado de Capital)
ERP	Equity Risk Premium (Prêmio de Risco de Mercado)
SML	Security Market Line (Linha de Mercado de Títulos)
OLS	Ordinary Least Squares (Mínimos Quadrados Ordinários)
bps	Basis Points (Pontos-base)

RESUMO

Este trabalho investiga a eficiência informacional do mercado brasileiro através de uma análise empírica da Petrobras S.A. (PETR4), testando se a incorporação de dados fundamentalistas granulares e variáveis macroeconômicas adiciona poder explicativo sobre os retornos do ativo. Utilizando uma abordagem de modelos aninhados (M0 a M5b), partindo de benchmarks ingênuos até algoritmos de Machine Learning (XGBoost), documentamos a evolução do R^2 Out-of-Sample de 12,28% (CAPM Estático) para 33,40% (M5b - ML Granular). Identificamos a “Armadilha da Linearidade”, onde a agregação de métricas em um score único destrói informação relevante, resultando em performance inferior ao modelo macro. A análise da Hipótese dos Mercados Adaptativos (AMH) revela que a eficiência informacional varia no tempo, com o Machine Learning superando modelos lineares apenas em regimes específicos de alta complexidade. Embora a previsão direcional de curto prazo falhe devido ao ruído estocástico, a estratégia de Valor Justo (Fair Value) baseada no modelo M5b gerou um Índice de Sharpe de 1.76, explorando a ineficiência de convergência no médio prazo. O estudo conclui com um diagnóstico fundamentalista atual via Radar Q-VAL, recomendando cautela (MANTER) dada a assimetria desfavorável entre risco e retorno projetado, validando a perspectiva de que a geração de *alpha* em mercados emergentes exige a integração de análise fundamentalista, macroeconômica e modelagem não-linear adaptativa.

Palavras-chave: Eficiência de Mercado; Machine Learning; CAPM Dinâmico; Volatility Targeting; Petrobras; Economia da Complexidade.

“Economics is not a form of physics, but a form of biology.”

— **W. Brian Arthur**, *Complexity and the Economy*

1 INTRODUÇÃO

A pergunta sobre como mercados processam informação atravessa a história do pensamento econômico como questão simultaneamente técnica, epistemológica e filosófica. Técnica porque demanda modelos formais capazes de capturar a relação entre informação disponível e formação de preços. Epistemológica porque interroga os limites do conhecimento possível em sistemas descentralizados onde nenhum agente possui visão completa do todo. Filosófica porque implica juízos sobre racionalidade, eficiência e os fundamentos normativos da organização econômica. Este trabalho situa-se na interseção dessas dimensões, propondo investigação empírica de uma questão teórica fundamental: em que medida a análise fundamentalista estruturada adiciona informação ao processo de precificação de ativos, ou apenas replica conhecimento já incorporado pelo mecanismo de mercado?

A tradição intelectual que informa esta investigação remonta ao ensaio seminal de Hayek (1945), onde o sistema de preços é caracterizado como “maravilha” epistêmica — mecanismo de telecomunicação que condensa informações dispersas e tácitas, o “conhecimento das circunstâncias particulares de tempo e lugar”, que nenhum agente central poderia reunir. Para Hayek, o mercado resolve problema computacional de complexidade intratável: agregar bilhões de fragmentos de conhecimento local em sinais de preço que coordenam decisões descentralizadas. A intuição hayekiana foi posteriormente formalizada por Fama (1970) na Hipótese dos Mercados Eficientes (EMH), segundo a qual os preços refletem toda informação disponível, tornando impossível a obtenção sistemática de retornos anormais através de análise de dados públicos.

Contudo, a elegância teórica da EMH encontra obstáculo lógico identificado por Grossman; Stiglitz (1980): se os preços refletissem perfeitamente toda informação, não haveria incentivo para incorrer nos custos de sua coleta e processamento. O paradoxo de Grossman-Stiglitz estabelece que mercados perfeitamente eficientes são impossíveis — algum grau de ineficiência é necessário para compensar o custo da informação e garantir que agentes continuem a produzi-la. Este teorema fundamenta teoricamente a existência da análise fundamentalista: analistas calculam métricas, constroem modelos e emitem recomendações porque esperam ser compensados por esse esforço através de retornos superiores ao mercado.

Desenvolvimentos recentes oferecem enquadramentos complementares. A Hipótese dos Mercados Adaptativos de Lo (2004), formalizada em Lo; Zhang (2024), reconcilia eficiência e comportamento através de lente evolucionária: mercados não são eficientes *ou* ineficientes, mas *adaptativos*. A eficiência varia ao longo do tempo, dependendo do ambiente competitivo,

da densidade de participantes informados, e da disponibilidade de oportunidades de arbitragem. Métricas fundamentalistas, nessa perspectiva, são estratégias adaptativas que funcionam em certos regimes e desaparecem quando o mercado se adapta a elas. A economia da complexidade, desenvolvida no Santa Fe Institute por Arthur (2014) e sintetizada em Arthur (2021), oferece enquadramento ainda mais radical: mercados são sistemas adaptativos complexos onde agentes heterogêneos com racionalidade limitada interagem em redes, produzindo dinâmicas emergentes não redutíveis a equilíbrio. Nessa visão, informação não é simplesmente “refletida” nos preços — é criada, disseminada, distorcida e amplificada em processos que podem incluir herding, cascatas informacionais e bolhas.

A contribuição pós-hayekiana de Colin-Jaeger; Delcey (2020) elucida tensão metodológica entre tradições: enquanto Hayek concebia preços como sinais dentro de processo dinâmico de competição, onde conhecimento relevante é tácito e processual, a EMH de Fama trata preços como reflexos de informação em equilíbrio, onde conhecimento é proposicional e agregável. O trabalho de Kuchar (2025) estende a perspectiva hayekiana para ciência cognitiva, propondo que mercados sejam entendidos como redes de cognição socialmente estendida — não apenas mecanismos de alocação, mas sistemas de aprendizado coletivo onde preferências e conhecimento são formados, não apenas revelados. A implicação para análise fundamentalista é profunda: métricas não apenas “medem” fundamentos — participam do processo cognitivo coletivo que forma preços, podendo tornar-se profecias autorrealizáveis ou autofrustrantes.

A pergunta central que orienta esta investigação pode ser formulada nos seguintes termos: considerando a Petrobras S.A. (PETR4) como caso empírico, em contexto de expansão para novas fronteiras exploratórias e riscos ESG/regulatórios associados, o acréscimo de informação fundamentalista estruturada — operacionalizada através de motor de *scoring* multidimensional — resulta em aumento mensurável da capacidade explicativa sobre retornos? Formalmente: o ΔR^2 entre modelo de mercado puro (CAPM) e modelo acrescido de métricas fundamentalistas é estatisticamente significativo e economicamente relevante?

A relevância do caso Petrobras transcende o interesse específico no ativo. Trata-se da maior empresa brasileira por capitalização de mercado, componente dominante do índice Ibovespa, e objeto de cobertura analítica intensa — dezenas de relatórios de *sell-side*, vigilância permanente de mídia especializada, dados públicos abundantes via CVM e B3. Se algum ativo deveria estar “perfeitamente precificado” no mercado brasileiro, seria este. Paradoxalmente, a empresa exhibe características que podem gerar fricções informacionais: assimetria entre gestão

e investidores, complexidade contábil do setor de óleo e gás, interferência estatal recorrente, e — no momento presente — incerteza binária sobre o desfecho da exploração da Margem Equatorial. A tensão entre alta cobertura analítica e potencial persistência de mispricings torna PETR4 laboratório ideal para testar a contribuição marginal da análise fundamentalista.

Do ponto de vista metodológico, a abordagem proposta operacionaliza conceitos teóricos abstratos em testes empíricos. O motor Q-VAL, desenvolvido como instrumento de análise multidimensional integrando métricas de Valor, Qualidade e Risco, serve como proxy para o “custo da informação” no sentido de Grossman-Stiglitz: representa esforço sistemático de coleta, processamento e síntese de dados fundamentalistas. A comparação entre modelos econométricos progressivos — do CAPM puro ao modelo acrescido do score Q-VAL — permite quantificar o “retorno” desse investimento informacional via variação explicada (ΔR^2). A análise é complementada por critérios de informação (AIC, BIC) que penalizam complexidade, respondendo à pergunta crucial: o motor de métricas está adicionando sinal ou apenas ruído?

O trabalho estrutura-se em cinco movimentos. Primeiro, estabelece fundamentos teóricos. Segundo, contextualiza o caso Petrobras. Terceiro, descreve a metodologia de modelos aninhados (M0 a M5). Quarto, apresenta resultados empíricos da evolução da eficiência informacional. Quinto, discute implicações teóricas e práticas, situando os achados no debate sobre eficiência de mercado e o papel da análise fundamentalista.

2 FUNDAMENTOS TEÓRICOS

2.1 O Sistema de Preços como Mecanismo Informacional

A compreensão do mercado financeiro como sistema de processamento de informação constitui uma das contribuições mais profundas da teoria econômica do século XX. A tradição intelectual que informa esta perspectiva remonta ao debate sobre cálculo econômico sob o socialismo, travado nas décadas de 1920 e 1930, onde a questão central era se um planejador central poderia, em princípio, replicar a eficiência alocativa de mercados descentralizados. A resposta negativa, articulada com maior rigor por Friedrich August von Hayek, transcendeu o debate original para estabelecer fundamentos epistemológicos que permanecem centrais à teoria financeira contemporânea.

2.1.1 A Contribuição Hayekiana: *Conhecimento Disperso e Coordenação*

O ensaio seminal de Hayek (1945) reformula o problema econômico fundamental. A questão não é, como supunham economistas anteriores, a alocação ótima de recursos *dados* — problema que, em princípio, admite solução técnica via programação linear ou métodos equivalentes. O problema genuíno é a utilização de conhecimento que *não está dado* a ninguém em sua totalidade. O conhecimento relevante para decisões econômicas encontra-se disperso entre milhões de agentes, cada qual possuindo fragmentos únicos — “conhecimento das circunstâncias particulares de tempo e lugar” — que nenhum mecanismo centralizador poderia reunir.

Esta dispersão não é acidental, mas constitutiva da realidade econômica. O comerciante local conhece o estado de seus estoques, as preferências idiossincráticas de seus clientes, as condições de suas instalações. O transportador conhece capacidades ociosas, rotas alternativas, condições de tráfego. O especulador conhece rumores, padrões históricos, correlações sutis. Cada fragmento de conhecimento é tácito, contextual e frequentemente inarticulável — o que Michael Polanyi denominaria posteriormente *tacit knowledge*, distinguindo-o do conhecimento proposicional formalizável.

O sistema de preços emerge, nesta perspectiva, como solução evolutiva para o problema de coordenação sob dispersão informacional. Hayek caracteriza-o como “maravilha” epistêmica:

“O aspecto mais significativo deste sistema é a economia de conhecimento com que opera, ou quão pouco os participantes individuais precisam saber para tomar a ação

correta. Em forma abreviada, por uma espécie de símbolo, apenas a informação mais essencial é transmitida.” (Hayek, 1945, p. 527)

O preço funciona como *estatística suficiente* — no sentido técnico do termo — condensando em um único número toda a informação relevante para a decisão marginal. O agente não precisa conhecer as causas que alteraram condições de oferta e demanda em mercados distantes; basta observar a variação de preços para ajustar seu comportamento de modo coordenado com milhões de outros agentes que também respondem ao mesmo sinal.

2.1.2 Do Conhecimento Tácito ao Conhecimento Proposicional

A transição da intuição hayekiana para modelos formais de finanças envolve transformação epistemológica significativa. Para Hayek, o conhecimento relevante é fundamentalmente *processual e tácito* — emerge no processo competitivo e não pode ser plenamente articulado fora dele. A formalização subsequente, culminando na Hipótese dos Mercados Eficientes, tratou conhecimento como *proposicional e agregável* — conjunto de proposições verdadeiras que podem, em princípio, ser listadas e incorporadas a modelos.

Colin-Jaeger; Delcey (2020) analisam esta transformação com rigor historiográfico. A EMH, argumentam, operacionalizou intuições hayekianas em forma testável, mas ao fazê-lo perdeu dimensões essenciais. Onde Hayek via processo dinâmico de descoberta — competição como procedimento de exploração do desconhecido —, Fama vê equilíbrio estático onde toda informação já foi descoberta e incorporada. A tensão entre estas concepções não é meramente acadêmica; implica visões distintas sobre a possibilidade e utilidade da análise fundamentalista.

A implicação para o presente trabalho é direta. Se conhecimento relevante é proposicional e já incorporado aos preços (Fama), métricas fundamentalistas não deveriam adicionar poder explicativo. Se conhecimento é processual e parcialmente tácito (Hayek), métricas podem capturar dimensões que o mercado ainda não processou — ou que processa com defasagem, ruído ou viés.

2.1.3 Mercados como Processadores de Informação

A metáfora computacional oferece enquadramento complementar. Mercados podem ser entendidos como sistemas distribuídos de processamento de informação, onde agentes individuais funcionam como processadores locais que recebem sinais (preços, notícias, dados), executam algoritmos (heurísticas, modelos, intuições) e produzem outputs (ordens de compra e venda) que, agregados, geram novos sinais para o sistema.

Esta perspectiva ilumina tanto a potência quanto as limitações do mecanismo de preços. Como sistema de computação distribuída, o mercado pode processar volumes de informação que excederiam a capacidade de qualquer agente central. A redundância — milhões de agentes processando informações parcialmente sobrepostas — confere robustez: erros individuais tendem a se cancelar, e o sinal agregado emerge com ruído reduzido. O paralelismo permite resposta rápida a novos dados, com ajustes de preço ocorrendo em milissegundos nos mercados contemporâneos.

Contudo, sistemas distribuídos são vulneráveis a falhas de coordenação. Comportamento de manada (*herding*) pode amplificar ruído em vez de filtrá-lo. Cascatas informacionais — onde agentes racionalmente ignoram informação privada para seguir o comportamento observado de outros — podem propagar erros sistematicamente. Bolhas especulativas emergem quando o próprio aumento de preços é interpretado como sinal informativo, gerando feedback positivo que desacopla preços de fundamentos.

A literatura sobre microestrutura de mercados, iniciada por Glosten; Milgrom (1985) e desenvolvida por Kyle (1989), formaliza como informação é incorporada aos preços através do processo de negociação. O *spread* bid-ask reflete, em parte, o custo que formadores de mercado incorrem ao negociar com agentes potencialmente informados. A velocidade de incorporação de informação depende da liquidez, da densidade de participantes informados, e da estrutura do mercado. Estas fricções microestruturais implicam que a incorporação de informação é processo gradual, não instantâneo — abrindo espaço para que análise fundamentalista capture informação ainda não plenamente refletida nos preços.

2.2 Eficiência de Mercado: Da EMH aos Mercados Adaptativos

2.2.1 A Hipótese dos Mercados Eficientes: Formulação e Taxonomia

A formalização da intuição hayekiana em teoria testável deve-se primordialmente a Eugene Fama, cuja tese de doutorado e trabalhos subsequentes estabeleceram o paradigma dominante em finanças acadêmicas por décadas. A Hipótese dos Mercados Eficientes, articulada em Fama (1970), postula que preços de ativos refletem toda informação relevante disponível, tornando impossível a obtenção sistemática de retornos anormais — retornos superiores ao que seria justificado pelo risco assumido.

Formalmente, um mercado é eficiente com respeito a um conjunto de informações Φ_t se é impossível obter lucros econômicos negociando com base em Φ_t . Isto implica que:

$$E[R_{t+1}|\Phi_t] = E[R_{t+1}]$$

onde R_{t+1} denota o retorno do ativo no período seguinte. A expectativa condicional ao conjunto de informações iguala a expectativa incondicional — conhecer Φ_t não permite previsão superior à que seria obtida ignorando essa informação.

Fama propôs taxonomia tripartite baseada na amplitude do conjunto informacional:

Eficiência na Forma Fraca: O conjunto Φ_t inclui apenas o histórico de preços e retornos passados. Eficiência nesta forma implica que análise técnica — estratégias baseadas em padrões de preços históricos — não pode gerar retornos anormais sistematicamente. Padrões como “cabeça e ombros”, médias móveis ou indicadores de momentum não teriam poder preditivo além do acaso.

Eficiência na Forma Semi-Forte: O conjunto Φ_t inclui toda informação publicamente disponível — demonstrações financeiras, comunicados ao mercado, dados macroeconômicos, notícias. Eficiência nesta forma implica que análise fundamentalista — estratégias baseadas em avaliação de fundamentos econômicos — não pode gerar retornos anormais. Múltiplos como P/L, P/VP, EV/EBITDA, ou métricas de qualidade como ROE e ROIC, sendo públicos, já estariam incorporados aos preços.

Eficiência na Forma Forte: O conjunto Φ_t inclui toda informação, pública e privada. Eficiência nesta forma implica que nem mesmo *insiders* com acesso privilegiado poderiam obter retornos anormais. Esta forma é geralmente rejeitada empiricamente, dado que estudos documentam retornos anormais em negociações de insiders (Jaffe, 1974).

2.2.2 Evidência Empírica: Anomalias e Prêmios de Fatores

A elegância teórica da EMH confrontou-se, desde o início, com evidência empírica de anomalias — padrões de retorno aparentemente previsíveis que persistem ao longo do tempo e através de mercados. O programa de pesquisa em anomalias de mercado constitui um dos mais extensos da literatura financeira, documentando centenas de variáveis com aparente poder preditivo sobre retornos.

Fama (1998) oferece defesa sofisticada da EMH frente às anomalias. Argumenta que aparentes retornos anormais frequentemente (i) desaparecem quando custos de transação realistas

são considerados, (ii) são artefatos de *data snooping* — pesquisadores testando múltiplas hipóteses até encontrar significância estatística —, ou (iii) representam compensação por risco não capturado pelo modelo de precificação utilizado.

O terceiro argumento é particularmente relevante para o presente trabalho. Se o CAPM unifatorial é especificação incompleta — se existem fatores de risco sistemático além do mercado —, então o que parece “alfa” (retorno anormal) pode ser apenas “beta” (compensação por risco) em dimensão não modelada. Esta lógica motivou a extensão do CAPM para modelos multifatoriais.

Fama; French (1993) propuseram modelo trifatorial, adicionando ao fator de mercado dois fatores empiricamente motivados: SMB (*Small Minus Big*), capturando o prêmio de tamanho — empresas pequenas tendem a gerar retornos superiores —, e HML (*High Minus Low*), capturando o prêmio de valor — empresas com alto índice book-to-market tendem a superar aquelas com baixo índice. O modelo explica parcela substancial da variação transversal de retornos que o CAPM deixava inexplicada.

Extensões subsequentes adicionaram fatores de momentum (Jegadeesh; Titman, 1993), lucratividade e investimento (Fama; French, 2015), e qualidade (Asness; Frazzini; Pedersen, 2019). O modelo de cinco fatores de Fama-French representa, atualmente, especificação padrão para controle de risco em estudos de anomalias:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{i,MKT}(R_{m,t} - R_{f,t}) + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,RMW}RMW_t + \beta_{i,CMA}CMA_t + \varepsilon_{i,t}$$

onde *RMW* (*Robust Minus Weak*) captura o prêmio de lucratividade e *CMA* (*Conservative Minus Aggressive*) captura o prêmio de conservadorismo em investimentos.

A proliferação de fatores gerou preocupação com *overfitting*. Harvey; Liu; Zhu (2016) documentam mais de 300 fatores publicados na literatura acadêmica — o “zoológico de fatores” — e argumentam que a maioria representa artefatos estatísticos. Propõem ajuste de múltiplos testes via controle de *false discovery rate*, elevando substancialmente o limiar para significância estatística.

2.2.3 A Hipótese dos Mercados Adaptativos

A tensão entre eficiência teórica e anomalias empíricas motivou formulações alternativas. A contribuição mais influente é a Hipótese dos Mercados Adaptativos (AMH), proposta por Lo (2004) e formalizada extensivamente em Lo; Zhang (2024).

A AMH reconcilia eficiência e comportamento através de lente evolucionária. Mercados não são eficientes *ou* ineficientes de modo binário e atemporal; são *adaptativos*, exibindo graus variáveis de eficiência dependendo do ambiente competitivo, da densidade de participantes, e da disponibilidade de oportunidades de arbitragem. A metáfora central é ecológica: mercados são ecossistemas onde “espécies” de estratégias competem por recursos (retornos), com populações expandindo quando estratégias são lucrativas e contraindo quando deixam de sê-lo.

Formalmente, Lo propõe que agentes econômicos comportam-se de acordo com princípios evolucionários: agem em interesse próprio, cometem erros, aprendem e se adaptam, competem por recursos, e estão sujeitos a seleção natural. Estas premissas substituem as do *homo economicus* neoclássico — racionalidade perfeita, informação completa, preferências estáveis — por descrição mais consonante com evidência de psicologia cognitiva e neurociência.

As implicações diferem substancialmente da EMH:

Eficiência é contexto-dependente: Em mercados líquidos com alta densidade de participantes sofisticados (ações de grande capitalização em bolsas desenvolvidas), eficiência tende a ser elevada. Em mercados ilíquidos com poucos participantes (small caps em mercados emergentes, ativos alternativos), ineficiências podem persistir.

Prêmios de risco variam no tempo: A compensação exigida por diferentes tipos de risco flutua com condições de mercado, densidade de capital alocado à estratégia, e memória recente de eventos extremos. O prêmio de valor, por exemplo, pode ser elevado em períodos de aversão ao risco e comprimido quando capital abundante persegue a estratégia.

Estratégias têm “meia-vida”: Anomalias descobertas tendem a desaparecer conforme participantes ajustam comportamento em resposta à informação. A publicação acadêmica de uma anomalia pode, paradoxalmente, eliminar sua lucratividade — fenômeno documentado por McLean; Pontiff (2016).

Inovação cria novas oportunidades: Mudanças estruturais — novos instrumentos, novas tecnologias, novos regulamentos — criam nichos ecológicos onde ineficiências temporárias emergem antes que o mercado se adapte.

Hirshleifer; Lo; Zhang (2023) estendem a análise para incluir dinâmicas de contágio social. Demonstram que vieses comportamentais não necessariamente desaparecem sob pressão competitiva; podem, ao contrário, persistir e até se amplificar quando transmitidos socialmente. A diversidade de estilos de investimento — valor, crescimento, momentum, qualidade — coexiste

porque diferentes estratégias exploram nichos distintos, com desempenho relativo variando ciclicamente.

2.2.4 Implicações para Análise Fundamentalista

A transição da EMH para a AMH altera substantivamente as expectativas sobre análise fundamentalista. Sob a EMH estrita (forma semi-forte), métricas fundamentalistas públicas não deveriam ter poder preditivo — se P/L baixo previsse retornos superiores, arbitradores comprariam ações baratas até eliminar o spread. A persistência do *value premium* por décadas representaria, nesta visão, compensação por risco não modelado.

Sob a AMH, métricas fundamentalistas podem ter poder preditivo variável:

- Em certos regimes (recessões, crises de crédito), quando aversão ao risco é elevada e capital é escasso, o prêmio de valor pode ser substancial.
- Em outros regimes (expansões, abundância de liquidez), quando capital abundante persegue a estratégia, o prêmio pode comprimir-se ou inverter-se.
- A descoberta e popularização de uma métrica (via publicação acadêmica ou adoção por gestores quantitativos) pode erodir sua eficácia.

O motor Q-VAL, neste enquadramento, não é tentativa de arbitrar ineficiência permanente, mas instrumento para capturar ineficiência transitória — explorando, no jargão ecológico, nicho que pode existir no mercado brasileiro de ações em determinado regime. A análise empírica subsequente deve, portanto, considerar não apenas o poder preditivo médio, mas sua variação ao longo do período amostral.

2.3 O Paradoxo de Grossman-Stiglitz e o Custo da Informação

2.3.1 O Teorema de Impossibilidade

Se a EMH fosse literalmente verdadeira — se preços refletissem instantânea e perfeitamente toda informação disponível —, um paradoxo lógico emergiria. A incorporação de informação aos preços não é processo espontâneo; requer que agentes incorram em custos para coletar, processar e agir sobre informação. Analistas fundamentalistas dedicam recursos substanciais a examinar demonstrações financeiras, conduzir due diligence, e construir modelos de valuation. Gestores quantitativos investem em infraestrutura computacional, bases de dados,

e talento técnico. Se toda essa informação já estivesse refletida nos preços, tais investimentos seriam desperdício — seus retornos seriam nulos.

Mas se ninguém investisse em coleta de informação — racionalmente antecipando retornos nulos —, como a informação seria incorporada aos preços? O paradoxo foi formalizado por Grossman; Stiglitz (1980) em teorema que permanece central à teoria financeira:

“Como não poderia haver lucros na coleta de informação, deveria haver pouco motivo para negociar e pouco motivo para que mercados existissem. Demonstramos que quando a obtenção de informação é custosa, os preços não podem refletir perfeitamente a informação disponível, pois se assim fosse, aqueles que gastaram recursos para obtê-la não receberiam compensação.” (Grossman; Stiglitz, 1980, p. 405)

O teorema estabelece que mercados perfeitamente eficientes são logicamente impossíveis quando informação é custosa. Algum grau de ineficiência — algum retorno anormal para agentes informados — é necessário para compensar o custo da informação e garantir que o processo de incorporação continue operando.

2.3.2 *Formalização: Equilíbrio com Ruído*

Grossman e Stiglitz formalizam o argumento em modelo de equilíbrio com dois tipos de agentes: informados (que incorreram no custo c para adquirir informação) e desinformados (que observam apenas o preço). Em equilíbrio, a fração de agentes informados ajusta-se de modo que o retorno esperado de se tornar informado — o alfa esperado — iguale exatamente o custo da informação.

Se o preço fosse perfeitamente revelador — se agentes desinformados pudessem inferir toda a informação privada observando o preço —, não haveria vantagem em pagar o custo c . A fração de informados cairia a zero. Mas com zero informados, o preço não conteria informação alguma, tornando vantajoso pagar c . A dinâmica oscilaria indefinidamente.

A solução requer que o preço seja apenas *parcialmente* revelador. Isto é obtido pela introdução de *noise traders* — agentes cujas demandas são independentes de informação, motivadas por necessidades de liquidez, rebalanceamento, ou irracionalidade. O ruído que estes agentes introduzem impede que o preço revele perfeitamente a informação privada, preservando incentivo para sua aquisição.

Em equilíbrio, a utilidade esperada de agentes informados e desinformados iguala-se. A fração de informados λ^* satisfaz:

$$E[U(W|\text{informado})] - c = E[U(W|\text{desinformado})]$$

O grau de ineficiência — a magnitude do alfa esperado para informados — é função crescente do custo da informação c e decrescente da precisão com que preços revelam informação.

2.3.3 Extensões Contemporâneas e Evidência Empírica

A literatura subsequente estendeu o paradigma Grossman-Stiglitz em múltiplas direções. Verrecchia (1982) endogeniza a quantidade de informação adquirida, mostrando que agentes investem em precisão até o ponto onde benefício marginal iguala custo marginal. Hellwig; Veldkamp (2009) estende a análise para mercados com múltiplos ativos e informação heterogênea, demonstrando condições sob as quais a agregação de informação é mais ou menos eficiente.

Stiglitz; Kosenko (2024) oferecem atualização abrangente do paradigma da economia da informação. Argumentam que desenvolvimentos nas décadas desde o trabalho original reforçam, mais que enfraquecem, a centralidade das fricções informacionais. A revolução digital não eliminou assimetrias de informação; em muitos casos, exacerbou-as, criando novas formas de informação privada (dados alternativos, *high-frequency trading*) e novas fontes de ruído (desinformação, manipulação algorítmica).

Pastushkov (2024) aplicam modelagem evolucionária ao problema. Demonstram resultado contra-intuitivo: reduzir custos de informação não necessariamente aumenta eficiência de preços. Quando informação se torna muito barata, a vantagem competitiva de ser informado diminui, potencialmente levando a equilíbrio onde a maioria permanece desinformada — *free-riding* na informação produzida por poucos. Apenas quando custos são moderados observa-se equilíbrio com fração substancial de informados e preços relativamente eficientes.

A evidência empírica oferece suporte misto ao paradigma. Por um lado, a existência de fundos de investimento ativamente gerenciados — que cobram taxas substanciais para produzir análise — sugere que investidores acreditam haver valor na informação. Por outro lado, a maioria dos fundos ativos não supera benchmarks passivos após taxas (Gruber, 1996), sugerindo que a competição por alfa é intensa e que a maior parte do retorno à informação é capturada pelos próprios analistas, não pelos investidores finais.

2.3.4 O Motor Q-VAL como “Custo da Informação”

O paradoxo de Grossman-Stiglitz fornece enquadramento teórico direto para o presente trabalho. O motor Q-VAL — sistema de scoring que integra métricas de Valor, Qualidade e Risco — representa investimento em informação: coleta de dados fundamentalistas, processamento via algoritmos de normalização e agregação, e síntese em sinal acionável.

A pergunta empírica pode ser reformulada em termos do teorema: o “alfa” gerado pelo motor Q-VAL — medido via ΔR^2 em relação ao modelo de mercado puro — é suficiente para compensar o “custo” de sua produção? Se ΔR^2 for insignificante, o mercado brasileiro aproxima-se da eficiência semi-forte; métricas fundamentalistas públicas já estariam incorporadas aos preços. Se ΔR^2 for positivo e significativo, existe espaço para que análise fundamentalista adicione valor — o grau de ineficiência é suficiente para compensar custos informacionais.

A magnitude do ΔR^2 pode ser interpretada como proxy para o “grau de ineficiência” do mercado brasileiro com respeito a informação fundamentalista. Valores elevados sugeririam que o mercado processa lentamente — ou com viés sistemático — informação contábil pública. Valores baixos sugeririam processamento eficiente. A análise por subperíodos pode revelar variação temporal — consistente com a AMH — onde ineficiências são maiores em certos regimes (crises, alta volatilidade) e menores em outros.

2.4 Economia da Complexidade e Agregação Informacional

2.4.1 Mercados como Sistemas Adaptativos Complexos

A economia da complexidade oferece enquadramento radicalmente distinto tanto da visão hayekiana quanto da EMH. Desenvolvida primordialmente no Santa Fe Institute, com contribuições seminais de Arthur (2014) e sintetizada em Arthur (2021), esta tradição rejeita a metáfora do mercado como mecanismo de equilíbrio — seja este estático (EMH) ou dinâmico-convergente (Hayek) — em favor da metáfora do mercado como *sistema adaptativo complexo*.

Sistemas adaptativos complexos exibem propriedades que os distinguem tanto de sistemas mecânicos simples quanto de sistemas aleatórios puros:

Emergência: Padrões macroscópicos emergem de interações microscópicas de modo não trivialmente dedutível das regras locais. O preço de uma ação emerge de milhões de decisões

individuais de compra e venda, mas não pode ser previsto conhecendo-se apenas as regras que governam cada decisão isolada.

Auto-organização: Estruturas ordenadas surgem espontaneamente sem coordenação central. Mercados desenvolvem convenções, normas, instituições — desde horários de negociação até práticas contábeis — sem designer explícito.

Não-linearidade: Pequenas perturbações podem gerar efeitos desproporcionais (sensibilidade às condições iniciais), enquanto grandes perturbações podem ser absorvidas sem consequência. O colapso do Lehman Brothers — evento relativamente pequeno no contexto do sistema financeiro global — desencadeou crise sistêmica; outros eventos comparáveis passaram sem repercussão.

Adaptação: Agentes modificam comportamento em resposta ao ambiente, e o ambiente modifica-se em resposta ao comportamento agregado. Estratégias que funcionam atraem imitadores; a imitação altera o ambiente; estratégias que funcionavam deixam de funcionar.

Path dependence: A história importa. O estado atual do sistema depende não apenas de parâmetros correntes, mas da trajetória pela qual foi atingido. Mercados não “esqueceram” a crise de 2008; participantes, reguladores e instituições foram permanentemente alterados.

2.4.2 Agregação de Informação como Fenômeno Emergente

A perspectiva da complexidade reconceptualiza a agregação de informação. Na visão hayekiana clássica, preços *condensam* informação dispersa, servindo como estatísticas suficientes que permitem coordenação. Na EMH, preços *refletem* informação, igualando-se ao valor esperado condicional a toda informação disponível. Na economia da complexidade, preços *emergem* de interações que podem tanto agregar quanto distorcer informação.

Arthur (1994) desenvolvem modelo seminal — o Bar Problem — demonstrando como agentes com racionalidade limitada, usando heurísticas heterogêneas, podem atingir coordenação eficiente em média, mas com flutuações persistentes que não convergem a equilíbrio. Aplicado a mercados, o modelo sugere que preços oscilam perpetuamente em torno de “valor fundamental” sem jamais estabilizar — não por choques exógenos, mas por dinâmica endógena do sistema.

O modelo de mercado artificial do Santa Fe Institute (LeBaron, 2006) simula mercados com agentes heterogêneos que usam regras de decisão evolucionárias. Os resultados reproduzem “fatos estilizados” observados em mercados reais — caudas gordas nas distribuições de retornos, volatilidade clustering, correlação serial de volatilidade — que modelos de equilíbrio

não capturam. Crucialmente, a eficiência informacional do mercado simulado varia: em certos parâmetros, preços rastreiam fundamentos eficientemente; em outros, bolhas e crashes emergem endogenamente.

2.4.3 Herding, Cascatas e Amplificação de Ruído

A economia da complexidade enfatiza mecanismos pelos quais informação pode ser sistematicamente distorcida — não apenas agregada — pelo mecanismo de mercado.

Herding (comportamento de manada): Agentes podem racionalmente ignorar informação privada para seguir o comportamento observado de outros. Se muitos agentes precedentes compraram um ativo, o agente seguinte pode inferir que eles possuíam informação favorável, e comprar mesmo que sua própria informação seja desfavorável. O resultado é cascata informacional onde informação privada é perdida e o grupo converge para decisão potencialmente errada (Bikhchandani; Hirshleifer; Welch, 1992).

Feedback positivo: Se o aumento de preço de um ativo atrai compradores adicionais — seja porque interpretam o aumento como sinal informativo, seja porque regras de momentum assim ditam —, o próprio movimento de preço gera demanda adicional, desacoplando preços de fundamentos. Bolhas especulativas são manifestação extrema deste mecanismo.

Reflexividade: Conceito desenvolvido por Soros (1987), a reflexividade descreve o fenômeno pelo qual expectativas afetam a realidade que pretendem prever. Se analistas acreditam que uma empresa é sólida, concedem crédito a termos favoráveis, atraem talentos, geram cobertura positiva — tornando a empresa efetivamente mais sólida. Preços não apenas refletem fundamentos; influenciam-nos.

Contágio informacional: Informação — e desinformação — propaga-se através de redes sociais de modo que pode amplificar certos sinais e suprimir outros. Bollen; Mao; Zeng (2011) documentam correlação entre sentimento em redes sociais e retornos subsequentes, sugerindo que “informação” incorporada a preços pode incluir ruído social sistematicamente enviesado.

2.4.4 Implicações para o Motor Q-VAL

A perspectiva da complexidade sugere cautela interpretativa quanto aos resultados do motor Q-VAL.

Regime-dependência: O poder preditivo de métricas fundamentalistas pode variar substancialmente entre regimes de mercado. Em períodos de baixa volatilidade e liquidez abundante,

quando capital busca ativamente oportunidades, ineficiências tendem a ser rapidamente arbitradas. Em períodos de crise, quando aversão ao risco domina e capital foge para ativos seguros, ineficiências podem ampliar-se — ativos subavaliados podem tornar-se mais subavaliados antes de reverter.

Reflexividade das métricas: Se o score Q-VAL — ou métricas similares — for amplamente adotado, sua própria utilização afetará os preços que pretende prever. Empresas que recebem scores elevados atraem fluxo de investimento, elevando preços até que a “oportunidade” desapareça. Há risco de *self-defeating prophecy* — métricas que funcionam deixam de funcionar justamente porque funcionam.

Ruído vs. sinal: A economia da complexidade alerta que nem toda informação incorporada a preços é “informação” no sentido de conhecimento verdadeiro sobre fundamentos. Parte é ruído — flutuações aleatórias, comportamento de manada, cascatas informacionais. O aumento de R^2 observado pode refletir captura de padrões de ruído correlacionado, não genuína informação sobre valor fundamental.

Não-linearidade: Modelos lineares — como as regressões propostas na metodologia — podem capturar apenas parcela das relações. Interações não-lineares entre métricas, efeitos de limiar (*threshold effects*), e dinâmicas de feedback podem requerer especificações mais flexíveis. A literatura recente em machine learning para precificação de ativos (Gu; Kelly; Xiu, 2020) documenta ganhos substanciais com métodos não-lineares, sugerindo que especificações lineares subestimam o conteúdo informacional de métricas fundamentalistas.

2.4.5 Síntese: Quatro Visões sobre Informação e Preços

O arcabouço teórico desenvolvido nesta seção pode ser sintetizado contrastando quatro visões sobre a relação entre informação e preços:

Tabela 1 – Síntese das Visões Teóricas sobre Informação e Preços {#tbl:theoretical_synthesis}

Tradição	Metáfora Central	Implicação para Análise Fundamentalista
Hayek	Mercado como telecomunicação	Preços condensam conhecimento tácito; análise pode antecipar processamento
Fama (EMH)	Mercado como espelho	Preços refletem informação; análise de dados públicos é inútil

Tradição	Metáfora Central	Implicação para Análise Fundamentalista
Grossman-Stiglitz	Mercado como incentivo	Ineficiência necessária; análise é compensada por alfa
Complexidade	Mercado como ecossistema	Eficiência variável; análise funciona em certos nichos/regimes

Estas visões não são mutuamente exclusivas. A EMH pode aproximar-se da verdade para ativos líquidos em mercados desenvolvidos, enquanto a economia da complexidade melhor descreve mercados emergentes ou ativos ilíquidos. O paradoxo de Grossman-Stiglitz opera em todos os contextos, estabelecendo limite inferior para ineficiência. A intuição hayekiana permanece válida como descrição do *processo* — ainda que o *resultado* possa aproximar-se de eficiência em alguns casos e afastar-se em outros.

Para o caso Petrobras, múltiplas considerações são relevantes. Por um lado, trata-se de ativo extremamente líquido, com alta cobertura de analistas, sugerindo eficiência elevada. Por outro lado, a empresa está inserida em contexto de mercado emergente, sujeita a interferência estatal, exposta a incerteza estrutural (Margem Equatorial), e caracterizada por complexidade contábil do setor de óleo e gás — fatores que podem gerar fricções informacionais persistentes.

A análise empírica subsequente buscará distinguir entre estas possibilidades, examinando se o motor Q-VAL adiciona poder explicativo ao modelo de mercado puro, e se este poder explicativo varia ao longo do tempo de modo consistente com a hipótese dos mercados adaptativos.

3 CONTEXTUALIZAÇÃO: PETROBRAS E A MARGEM EQUATORIAL

3.1 Trajetória Histórica e Institucional

A Petróleo Brasileiro S.A. — Petrobras — foi constituída em 1953 como instrumento da política desenvolvimentista que marcou o período Vargas. A criação da empresa estatal respondeu a demandas nacionalistas por controle dos recursos petrolíferos, sintetizadas na campanha “O Petróleo é Nosso” que mobilizou segmentos diversos da sociedade brasileira. O monopólio estatal sobre exploração, produção e refino, estabelecido pela Lei 2.004/1953, configurou modelo que prevaleceria por mais de quatro décadas.

A quebra parcial do monopólio em 1997, durante o governo Fernando Henrique Cardoso, inaugurou nova fase. A Emenda Constitucional nº 9/1995 e a Lei do Petróleo (9.478/1997) permitiram a entrada de empresas privadas no setor, embora a Petrobras mantivesse posição dominante. A abertura de capital, com listagem simultânea na B3 e NYSE, introduziu pressões de governança corporativa e transparência típicas de companhias abertas. A empresa passou a operar sob dupla lógica: instrumento de política energética nacional e maximizadora de valor para acionistas.

A descoberta do pré-sal em 2006 representou inflexão histórica. As reservas identificadas na camada de rochas carbonáticas sob espessa camada de sal, em águas ultraprofundas da Bacia de Santos, posicionaram o Brasil entre as principais fronteiras petrolíferas globais. O desenvolvimento do pré-sal exigiu investimentos massivos em tecnologia de perfuração e produção em condições extremas — capacidade que a Petrobras desenvolveu tornando-se referência mundial em águas profundas.

O período subsequente foi marcado por escândalo de corrupção de proporções inéditas. A Operação Lava Jato, deflagrada em 2014, revelou esquema de pagamentos ilícitos envolvendo contratos da Petrobras, partidos políticos e empreiteiras. As consequências foram severas: baixas contábeis bilionárias, rebaixamento de rating de crédito, queda abrupta do valor de mercado, e crise de governança que culminou em renovação completa da alta administração. A recuperação, iniciada a partir de 2016, envolveu desinvestimentos, redução de alavancagem e foco em ativos de maior rentabilidade — notadamente o pré-sal.

3.2 O Cenário Atual: Margem Equatorial e Transição Energética

A Margem Equatorial brasileira compreende as bacias sedimentares da costa norte do país, estendendo-se do Amapá ao Rio Grande do Norte. Trata-se de fronteira exploratória praticamente inexplorada, com potencial estimado entre 10 e 30 bilhões de barris de petróleo — volume que poderia duplicar as reservas provadas brasileiras. A região apresenta características geológicas análogas a províncias produtoras da costa oeste africana, onde descobertas significativas foram realizadas nas últimas décadas.

A autorização para exploração da Margem Equatorial, concedida pelo IBAMA em 2024-2025, reacendeu debates sobre a estratégia de longo prazo da Petrobras. De um lado, a exploração representa oportunidade de expansão de reservas em momento em que campos maduros declinam e o pré-sal, embora produtivo, apresenta horizonte finito. De outro, implica investimentos vultosos — estimados entre US\$ 50 e 100 bilhões — com retorno incerto e prazo de maturação superior a uma década.

O contexto de transição energética adiciona camada de complexidade à decisão. O Acordo de Paris e os compromissos subsequentes de neutralidade de carbono pressionam empresas petrolíferas a reorientar portfólios para fontes renováveis. Investidores institucionais, especialmente fundos de pensão europeus e gestores com mandatos ESG, questionam a racionalidade de investimentos em novas fronteiras fósseis. A Petrobras enfrenta, nesse cenário, tensão entre maximização de valor de curto prazo via exploração de reservas e posicionamento estratégico para economia descarbonizada.

Os riscos regulatórios e ambientais da Margem Equatorial são substanciais. A região abriga ecossistemas sensíveis, incluindo manguezais, recifes de coral e áreas de reprodução de espécies marinhas. O licenciamento ambiental tem sido objeto de contestação judicial e mobilização de organizações ambientalistas. Vazamentos ou acidentes em área de difícil acesso teriam consequências potencialmente graves, com repercussões sobre a reputação da empresa e sua licença social para operar.

3.3 Implicações para Valuation

A avaliação da PETR4 no contexto descrito envolve ponderação de cenários com probabilidades e payoffs distintos. No cenário otimista, a exploração da Margem Equatorial confirma potencial geológico, custos de desenvolvimento permanecem controlados, e preços de petróleo sustentam-se em patamares remuneradores. Os fluxos de caixa incrementais justificam o

investimento e o valor da empresa aumenta significativamente. No cenário pessimista, reservas mostram-se menores que o esperado, custos escalam além das projeções, preços de petróleo colapsam com aceleração da transição energética, e pressões ESG restringem acesso a capital. Os investimentos tornam-se *stranded assets* e o valor da empresa deteriora-se.

O mercado, ao precificar a PETR4, pondera implicitamente esses cenários. A pergunta relevante para o investidor é se essa ponderação está correta — se o preço corrente reflete adequadamente o valor esperado da empresa considerando probabilidades e magnitudes dos diferentes cenários. A análise de *mispricing* via comparação entre custo de capital implícito e custo de capital teórico (CAPM) oferece uma abordagem para essa questão.

4 METODOLOGIA

A metodologia deste trabalho articula três componentes integrados: (i) a construção do motor Q-VAL como sistema de agregação de informação fundamentalista, (ii) a especificação de modelos econométricos comparativos que permitem isolar a contribuição informacional de cada componente, e (iii) a definição de métricas de avaliação que operacionalizam conceitos teóricos — eficiência informacional, contribuição marginal, parcimônia — em quantidades mensuráveis. A integração destes componentes permite responder à pergunta central: métricas fundamentalistas adicionam informação ao processo de precificação, ou essa informação já está incorporada aos preços?

A estratégia empírica segue lógica de *nested models* (modelos aninhados), partindo de especificação mínima — o CAPM unifatorial — e progressivamente incorporando vetores de informação fundamentalista. A comparação entre modelos sucessivos, via métricas como ΔR^2 e critérios de informação, permite atribuir a cada componente sua contribuição marginal explicativa. Esta abordagem dialoga diretamente com o paradoxo de Grossman-Stiglitz: se métricas públicas já estão precificadas, sua adição não deveria melhorar o poder explicativo; se existe ineficiência compensadora, a melhoria será estatisticamente detectável.

4.1 O Motor Q-VAL como Vetor de Informação Fundamentalista

4.1.1 Arquitetura Conceitual

O motor Q-VAL (*Quantitative Value*) constitui sistema de *scoring* fundamentalista que integra múltiplas dimensões de avaliação em índice sintético. A denominação remete à tradição de *Quantitative Value Investing* sistematizada por Gray; Carlisle (2012), que propõe automatização de princípios de análise fundamentalista clássica — originários de Graham; Dodd (1934) — via algoritmos computacionais que eliminam vieses comportamentais e garantem consistência metodológica.

A arquitetura do motor organiza-se em três dimensões canônicas:

Dimensão de Valor: Captura a relação entre preço de mercado e fundamentos contábeis. Métricas de valor respondem à pergunta: *quanto o mercado está pagando por unidade de fundamento econômico?* Valores baixos (P/L reduzido, EV/EBITDA comprimido) sugerem subavaliação — o mercado atribui ao ativo preço inferior ao que seus fundamentos justificariam.

A literatura documenta *value premium* persistente: ativos “baratos” tendem, em média, a superar ativos “caros” em horizontes de médio prazo (Asness; Moskowitz; Pedersen, 2013; Fama; French, 1993).

Dimensão de Qualidade: Avalia a eficiência operacional e sustentabilidade dos resultados. Métricas de qualidade respondem à pergunta: *quão eficientemente a empresa converte capital em retorno?* Valores elevados (ROE alto, margens robustas, crescimento consistente) indicam vantagens competitivas que tendem a persistir. Novy-Marx (2013) documenta que lucratividade possui poder preditivo comparável ao de métricas de valor, enquanto Asness; Frazzini; Pedersen (2019) formalizam “qualidade” como fator distinto com prêmio próprio.

Dimensão de Risco: Incorpora indicadores de alavancagem, liquidez e volatilidade. Métricas de risco respondem à pergunta: *qual a probabilidade de deterioração dos fundamentos ou de eventos adversos?* A inclusão desta dimensão reconhece que ativos “baratos e de qualidade” podem sê-lo por razões legítimas — risco elevado que justifica desconto. A dimensão de risco serve como *hedge* contra *value traps*: ativos superficialmente atrativos que ocultam fragilidades estruturais.

4.1.2 Métricas Selecionadas

A seleção de métricas segue critérios de relevância teórica, disponibilidade de dados, e robustez empírica documentada na literatura. Para cada dimensão, escolheu-se conjunto parcimonioso de indicadores que capturam facetas complementares:

Dimensão de Valor

Tabela 2 – Métricas da Dimensão de Valor {#tbl:metrics_value}

Métrica	Definição	Interpretação
Earnings Yield (EY)	$\frac{\text{LPA}}{\text{Preço}} = \frac{1}{\text{P/L}}$	Retorno implícito do lucro; inverso do P/L
EV/EBITDA	$\frac{\text{Enterprise Value}}{\text{EBITDA}}$	Múltiplo de valor da firma sobre geração de caixa operacional
P/VP	$\frac{\text{Preço}}{\text{Valor Patrimonial por Ação}}$	Relação entre valor de mercado e valor contábil
Dividend Yield (DY)	$\frac{\text{Dividendos por Ação}}{\text{Preço}}$	Retorno em dividendos

O *Earnings Yield* é preferido ao P/L por conveniência matemática: valores mais altos indicam maior atratividade (inversamente ao P/L), facilitando agregação com outras métricas onde “maior é melhor”. O EV/EBITDA captura valor da firma inteira — incluindo dívida — sobre geração de caixa operacional, sendo particularmente relevante para empresas intensivas em capital como a Petrobras.

Dimensão de Qualidade

Tabela 3 – Métricas da Dimensão de Qualidade {#tbl:metrics_quality}

Métrica	Definição	Interpretação
ROIC	$\frac{\text{NOPAT}}{\text{Capital Investido}}$	Retorno sobre capital investido
ROE	$\frac{\text{Lucro Líquido}}{\text{Patrimônio Líquido}}$	Retorno sobre patrimônio
Margem EBITDA	$\frac{\text{EBITDA}}{\text{Receita Líquida}}$	Eficiência operacional
EVS	ROIC – WACC	Economic Value Spread; criação de valor

O *Economic Value Spread* (EVS) merece destaque. Conforme desenvolvido na seção de fundamentos teóricos, o EVS captura criação de valor em termos relativos ao custo de capital. Valor positivo indica que a empresa gera retorno superior ao custo de oportunidade do capital empregado — condição necessária para criação sustentável de valor ao acionista. Para empresas do setor de óleo e gás, onde ciclos de investimento são longos e intensivos em capital, o EVS oferece perspectiva mais informativa que métricas de rentabilidade brutas.

Dimensão de Risco

Tabela 4 – Métricas da Dimensão de Risco {#tbl:metrics_risk}

Métrica	Definição	Interpretação
Beta	$\frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}$	Sensibilidade ao mercado
Volatilidade	$\sigma_i = \sqrt{\text{Var}(R_i)}$	Desvio-padrão dos retornos
Dívida/PL	$\frac{\text{Dívida Total}}{\text{Patrimônio Líquido}}$	Alavancagem financeira
Liquidez Corrente	$\frac{\text{Ativo Circulante}}{\text{Passivo Circulante}}$	Capacidade de pagamento de curto prazo

O beta, estimado via regressão conforme detalhado adiante, captura risco sistemático — a parcela do risco que não pode ser eliminada por diversificação. A volatilidade captura risco total. A razão Dívida/PL indica exposição a risco financeiro — empresas alavancadas são mais sensíveis a choques de receita e taxa de juros. A liquidez corrente sinaliza risco de curto prazo — capacidade de honrar obrigações imediatas.

4.1.3 Procedimento de Normalização

Métricas heterogêneas — expressas em unidades distintas (percentuais, múltiplos, razões) e com escalas variadas — requerem normalização para agregação. O procedimento adotado é a *padronização Z-Score* relativa a benchmarks setoriais:

$$Z_i = \frac{X_i - \mu_{\text{setor}}}{\sigma_{\text{setor}}}$$

onde X_i é o valor da métrica para o ativo i , μ_{setor} é a média setorial, e σ_{setor} é o desvio-padrão setorial. O Z-Score indica quantos desvios-padrão o ativo se afasta da média do setor: $Z = +1$ indica desempenho um desvio-padrão acima da média; $Z = -1$, um desvio abaixo.

Para métricas onde valores *menores* são preferíveis (P/VP, EV/EBITDA, Dívida/PL, Volatilidade, Beta), o Z-Score é invertido:

$$Z_i^{\text{inv}} = -Z_i = \frac{\mu_{\text{setor}} - X_i}{\sigma_{\text{setor}}}$$

Esta inversão garante que, para todas as métricas normalizadas, valores mais altos indicam maior atratividade.

Matriz de Correlação: Valor, Qualidade e Risco

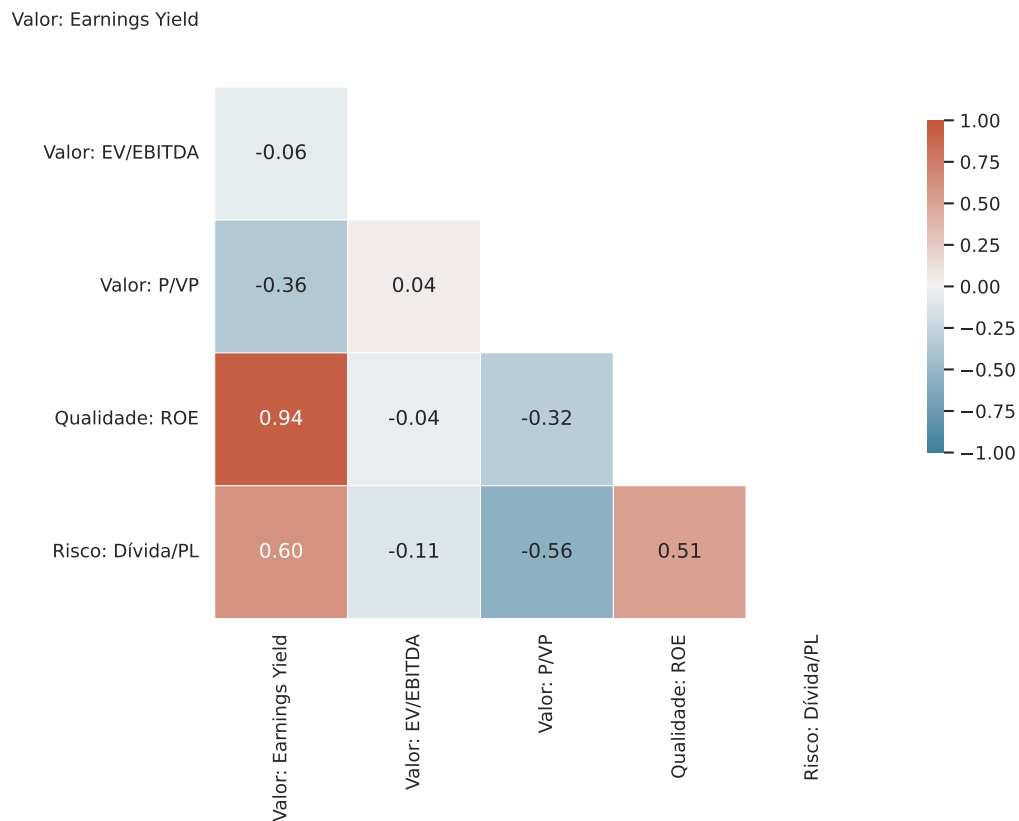


Figura 1 – Matriz de Correlação dos Z-Scores Fundamentais. A baixa correlação entre as dimensões de Valor e Qualidade sugere que elas capturam informações ortogonais, justificando a abordagem multidimensional.

Os benchmarks setoriais são obtidos de empresas comparáveis do setor de óleo e gás integrado, utilizando dados de empresas listadas na B3 e, quando necessário para robustez estatística, médias de empresas latino-americanas do setor obtidas via bases de dados internacionais.

4.1.4 Agregação em Score Composto e Abordagem Machine Learning

A agregação das métricas normalizadas em um sinal único de investimento é realizada através de duas abordagens distintas, permitindo testar hipóteses sobre a estrutura da informação fundamentalista.

Abordagem Linear (Baseline - M5a): Inicialmente, os Z-Scores são agregados em um score composto via média ponderada simples. Esta abordagem assume que a contribuição de

cada métrica é linear e aditiva.

$$\text{Score}_{\text{Dimensão}} = \sum_{j=1}^n w_j \cdot Z_j$$

O score final Q-VAL agrega as três dimensões (Valor, Qualidade, Risco) com pesos iguais ($w = 1/3$). Este score linear serve como *baseline*: ele representa a heurística tradicional de “somar pontos” para avaliar uma empresa. Para facilitar a interpretação, o score é transformado para a escala 0-100:

$$\text{Q-VAL}_{[0,100]} = 50 + 10 \cdot \text{Q-VAL}_{\text{bruto}}$$

Abordagem Não-Linear (Machine Learning - M5b): Reconhecendo que a relação entre fundamentos e retornos é complexa e condicional (ex: alavancagem alta pode ser benéfica em expansão mas fatal em recessão), adota-se uma abordagem baseada em *Gradient Boosting* (XGBoost). O modelo M5b recebe o vetor completo de Z-Scores individuais como *features* e aprende a função de mapeamento $f(X)$ que maximiza a aderência aos retornos observados.

$$f(X) = \sum_{k=1}^K f_k(X), \quad f_k \in \mathcal{F}$$

Diferentemente da agregação linear, o M5b captura interações não-lineares e efeitos de limiar (*thresholds*) sem impor uma estrutura de pesos fixa *a priori*. A comparação entre o desempenho do Score Linear (M5a) e do Modelo ML (M5b) constitui um teste direto da hipótese de complexidade: se o M5b superar significativamente o M5a, confirma-se que a “receita” de agregação importa tanto quanto os ingredientes (dados).

Scores acima de 60 (na escala linear) ou sinais fortes do modelo ML indicam recomendação de *Compra*; abaixo de 40, *Venda*. Estes limiares correspondem aproximadamente a ± 1 desvio-padrão em torno da média histórica.

4.1.5 Série Temporal de Scores

Para a análise de contribuição informacional, é necessário construir série temporal de scores. A cada período t , o score Q-VAL é calculado utilizando apenas informação disponível até t — critério essencial para evitar *look-ahead bias*. Formalmente:

$$\text{Q-VAL}_t = f(\mathcal{I}_t)$$

onde \mathcal{I}_t denota o conjunto de informação disponível no período t . Na prática, dados contábeis são defasados: demonstrações financeiras do trimestre q tornam-se públicas apenas semanas após o encerramento do trimestre. O procedimento utiliza dados do trimestre $q - 1$ para calcular scores no trimestre q , garantindo que toda informação utilizada era publicamente disponível no momento da decisão.

4.2 Modelos Econométricos Comparativos

4.2.1 Estratégia de Modelos Aninhados (M0 a M5)

A estratégia empírica adota uma abordagem de **Comparação Progressiva e Aninhada**, partindo de benchmarks ingênuos (*naïve*) até modelos multifatoriais complexos. Esta estrutura permite isolar a contribuição marginal de cada fonte de informação (histórico, mercado, fundamentos, macroeconomia) para a eficiência da precificação.

A hierarquia de modelos é definida da seguinte forma:

Tabela 5 – Hierarquia de Modelos Econométricos {#tbl:model_hierarchy}

Modelo	Definição Conceitual	Especificação Econométrica
M0	Benchmarks Naïve	Random Walk ($E[r] = 0$) e Média Histórica ($E[r] = \mu_{train}$)
M1	CAPM Estático	Beta constante: $R_t = \alpha + \beta R_{m,t} + \epsilon_t$
M2	CAPM Dinâmico	Beta variável no tempo (Rolling OLS): $R_t = \alpha_t + \beta_t R_{m,t} + \epsilon_t$
M3	Fundamentos	M2 + Vetor de Fundamentos (Valor, Qualidade, Risco)

Modelo	Definição Conceitual	Especificação Econométrica
M4	Macro & Fatores	M3 + Variáveis Macro (Brent, Câmbio, Risco-País) e Fatores FF
M5a	Score Linear	M2 + Score Agregado Q-VAL (Teste de eficiência da agregação)
M5b	ML Granular	M2 + Vetor de Z-Scores + XGBoost (Não-linearidade)

4.2.2 Detalhamento dos Modelos

M0: Benchmarks Naïve (Linha de Base)

Representa a ausência de modelo. O **M0-RW (Random Walk)** assume que o melhor previsor para o retorno futuro é zero (ou o último preço, implicando retorno zero). O **M0-HM (Historical Mean)** assume que o retorno futuro será igual à média histórica dos retornos observados no conjunto de treinamento. Estes modelos servem como “pisso” de performance: qualquer modelo útil deve superá-los.

M1: CAPM Estático (Eficiência de Mercado Padrão)

O modelo clássico de precificação de ativos, assumindo que o único fator de risco relevante é o mercado e que a sensibilidade do ativo (beta) é constante em todo o período.

$$R_{i,t} - R_{f,t} = \alpha + \beta(R_{m,t} - R_{f,t}) + \varepsilon_t$$

Este modelo testa a hipótese de eficiência na forma fraca/semi-forte padrão: o retorno é função linear apenas do risco de mercado.

M2: CAPM Dinâmico (Eficiência Adaptativa)

Reconhece que o risco da empresa muda ao longo do tempo. Utiliza janelas deslizantes (*rolling windows*) de 252 dias (1 ano de negociação) para estimar betas variáveis no tempo (β_t).

$$R_{i,t} - R_{f,t} = \alpha_t + \beta_t(R_{m,t} - R_{f,t}) + \varepsilon_t$$

A predição para $t + 1$ utiliza os parâmetros estimados em t . Este modelo serve como **âncora** para os modelos subsequentes: M3, M4 e M5 utilizam a predição do M2 ($\hat{R}_{M2,t}$) como regressor base, testando se outras variáveis adicionam informação *além* da dinâmica de risco de mercado.

M3: Fundamentos (Informação Específica da Firma)

Testa a hipótese central do trabalho: a análise fundamentalista adiciona valor? O modelo regride os retornos contra a predição do M2 e o vetor de scores fundamentalistas (Valor, Qualidade, Risco).

$$R_t = \delta \hat{R}_{M2,t} + \gamma_1 \text{Valor}_{t-1} + \gamma_2 \text{Qualidade}_{t-1} + \gamma_3 \text{Risco}_{t-1} + \varepsilon_t$$

Se os coeficientes γ forem significativos e houver ganho de R^2 Out-of-Sample, confirma-se que os fundamentos contêm informação não precificada pelo risco de mercado dinâmico.

M4: Macroeconomia e Fatores (Informação Externa)

Expande o conjunto informacional para incluir variáveis macroeconômicas (Retorno do Brent, Variação Cambial, Variação do Risco-País EMBI+) e fatores de risco Fama-French (Investimento/CMA e Lucratividade/RMW).

$$R_t = \text{Modelo M3} + \lambda_{Macro} \mathbf{X}_{Macro,t} + \lambda_{Fatores} \mathbf{F}_{FF,t} + \varepsilon_t$$

Este modelo representa o “teto” de complexidade, testando se choques exógenos explicam a variância residual dos fundamentos.

M5a: Score Linear (Eficiência da Agregação Simples)

Testa se o Score Q-VAL único (agregado linearmente) é uma estatística suficiente para o vetor de fundamentos. Substitui as três dimensões do M3 pelo score agregado.

$$R_t = \delta \hat{R}_{M2,t} + \theta \text{Score_QVAL}_{t-1} + \varepsilon_t$$

Se a performance do M5a for similar à do M3, a agregação linear é bem-sucedida em reduzir a dimensionalidade sem perda de informação. Nas tabelas de resultados, este modelo é frequentemente referido como “**M5 (Score Agregado)**”.

M5b: Machine Learning Granular (Não-Linearidade)

Abandona a premissa de linearidade e de agregação prévia. Utiliza o algoritmo XGBoost para processar o vetor completo de Z-Scores individuais, permitindo interações complexas entre as métricas. Além dos fundamentos da firma, o M5b incorpora também as variáveis macroeconômicas e fatores de risco do M4, funcionando como um modelo integrador.

$$R_t = f_{XGB}(\text{Z-Scores}_{t-1}, \text{Macro}_{t-1}, \text{Fatores}_{t-1}) + \varepsilon_t$$

Este modelo testa o limite superior da extração de informação. A diferença de performance entre M5b e M5a (ΔR^2_{NL}) quantifica o valor da não-linearidade e da complexidade na modelagem de preços. O fato de métricas macroeconômicas (M4) estarem disponíveis ao modelo mas não aparecerem entre os principais *drivers* de importância (Feature Importance) indicaria que, para a Petrobras neste período, a dinâmica idiossincrática (fundamentos) domina a dinâmica macro.

4.3 Métricas de Avaliação Informacional

4.3.1 Coeficiente de Determinação e Sua Variação

O coeficiente de determinação R^2 mede a fração da variância da variável dependente explicada pelo modelo:

$$R^2 = 1 - \frac{\text{SQR}}{\text{SQT}} = 1 - \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

onde SQR é a soma dos quadrados dos resíduos e SQT é a soma dos quadrados totais.

O R^2 varia entre 0 (modelo não explica nada além da média) e 1 (modelo explica perfeitamente). Para modelos de precificação de ativos, valores típicos situam-se entre 0,20 e 0,50 para dados de alta frequência (diários) e podem ser menores para dados de baixa frequência ou para ativos individuais.

Variação do R^2 (ΔR^2):

A métrica central deste trabalho é a variação do coeficiente de determinação entre modelos aninhados:

$$\Delta R_k^2 = R_{\mathcal{M}_k}^2 - R_{\mathcal{M}_{k-1}}^2$$

Esta quantidade mede a contribuição marginal dos regressores adicionais no modelo k à explicação da variância.

Interpretação à luz da teoria:

Resultado	Interpretação	Implicação Teórica
$\Delta R^2 \approx 0$	Métricas fundamentalistas não adicionam poder explicativo	Mercado eficiente; informação já precificada (Fama)
$\Delta R^2 > 0$, pequeno	Contribuição marginal modesta	Ineficiência limitada; compensação por custo informacional (Grossman-Stiglitz)
$\Delta R^2 > 0$, substancial	Contribuição significativa	Ineficiência material; oportunidade para análise fundamentalista
ΔR^2 variável no tempo	Contribuição regime-dependente	Mercados adaptativos (Lo)

4.3.2 R^2 Ajustado

O R^2 simples aumenta mecanicamente com a adição de regressores, mesmo que estes não tenham genuíno poder explicativo. O R^2 ajustado penaliza a adição de variáveis:

$$R_{\text{adj}}^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

onde n é o número de observações e k é o número de regressores (excluindo o intercepto).

A comparação de R_{adj}^2 entre modelos informa se a adição de variáveis melhora genuinamente a explicação ou apenas adiciona complexidade sem benefício. Se $R_{\text{adj, Modelo } k}^2 > R_{\text{adj, Modelo } k-1}^2$, a adição de regressores é justificada.

4.3.3 Critérios de Informação

Critérios de informação oferecem enquadramento alternativo — e mais rigoroso — para comparação de modelos, penalizando complexidade de modo mais severo que o R^2 ajustado.

Critério de Informação de Akaike (AIC):

$$\text{AIC} = 2k - 2 \ln(\hat{L})$$

onde k é o número de parâmetros estimados e \hat{L} é a verossimilhança maximizada. Para modelos de regressão linear com erros gaussianos:

$$\text{AIC} = n \ln \left(\frac{\text{SQR}}{n} \right) + 2k$$

Critério de Informação Bayesiano (BIC):

$$\text{BIC} = k \ln(n) - 2 \ln(\hat{L})$$

ou, para regressão linear:

$$\text{BIC} = n \ln \left(\frac{\text{SQR}}{n} \right) + k \ln(n)$$

O BIC penaliza complexidade mais severamente que o AIC (o termo $k \ln(n)$ cresce com o tamanho amostral, enquanto $2k$ é constante). Em amostras grandes, o BIC tende a selecionar modelos mais parcimoniosos.

Interpretação: O modelo preferido é aquele com menor AIC (ou BIC). A diferença $\Delta \text{AIC} = \text{AIC}_{\mathcal{M}_k} - \text{AIC}_{\mathcal{M}_{k-1}}$ indica se o modelo mais complexo é preferível: - $\Delta \text{AIC} < 0$: modelo mais complexo é preferível - $\Delta \text{AIC} > 0$: modelo mais parcimonioso é preferível - $|\Delta \text{AIC}| < 2$: diferença não é substancial

Regras análogas aplicam-se ao BIC, tipicamente com limiares mais conservadores ($|\Delta \text{BIC}| < 6$).

4.3.4 Testes de Significância Estatística

Teste F para Modelos Aninhados:

O teste F compara formalmente modelos aninhados, testando a hipótese nula de que os coeficientes adicionais são conjuntamente zero:

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

A estatística de teste é:

$$F = \frac{(R_{\mathcal{M}_k}^2 - R_{\mathcal{M}_{k-1}}^2)/p}{(1 - R_{\mathcal{M}_k}^2)/(n - k - 1)}$$

onde p é o número de regressores adicionais. Sob H_0 , a estatística segue distribuição $F(p, n - k - 1)$. Rejeição de H_0 indica que os regressores adicionais contribuem significativamente.

Teste t para Coeficientes Individuais:

A significância individual de cada coeficiente é testada via estatística t:

$$t_j = \frac{\hat{\gamma}_j}{\text{SE}(\hat{\gamma}_j)}$$

onde $\text{SE}(\hat{\gamma}_j)$ é o erro-padrão do coeficiente estimado. Sob $H_0 : \gamma_j = 0$, a estatística segue distribuição $t(n - k - 1)$. Com erros-padrão robustos, a distribuição assintótica é normal padrão.

4.3.5 Validação Out-of-Sample

Métricas in-sample — calculadas sobre os mesmos dados usados para estimação — podem superestimar o poder preditivo real, especialmente quando o modelo possui muitos parâmetros relativamente ao tamanho amostral (*overfitting*). Validação out-of-sample oferece estimativa mais conservadora e operacionalmente relevante.

Divisão Temporal:

O período amostral é dividido em: - **Período de treinamento** (in-sample): Janeiro/2016 a Dezembro/2022 (~7 anos) - **Período de teste** (out-of-sample): Janeiro/2023 a Novembro/2025 (~3 anos)

Os modelos são estimados no período de treinamento; o poder preditivo é avaliado no período de teste. O R^2 out-of-sample (R_{OOS}^2) é calculado como:

$$R_{\text{OOS}}^2 = 1 - \frac{\sum_{t \in \text{teste}} (y_t - \hat{y}_t)^2}{\sum_{t \in \text{teste}} (y_t - \bar{y}_{\text{treino}})^2}$$

Note que a média utilizada no denominador é a média do período de *treinamento*, não do teste — garantindo que nenhuma informação futura contamina a avaliação.

Validação Cruzada Rolling-Window:

Para avaliar estabilidade temporal, implementa-se validação cruzada com janelas móveis:

1. Estima-se o modelo em janela de 252 dias úteis (~1 ano)
2. Prevê-se o retorno do dia seguinte
3. Desloca-se a janela um dia adiante
4. Repete-se o processo

O R^2 rolling é calculado para cada posição da janela, gerando série temporal de poder explicativo. A variação desta série informa sobre estabilidade — ou regime-dependência — da contribuição informacional das métricas fundamentalistas.

4.3.6 Métricas Complementares

Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE):

O MSE calcula a média dos quadrados dos erros (diferença entre valor previsto e real). Ele penaliza erros grandes mais severamente do que erros pequenos, sendo útil para identificar modelos que cometem falhas graves pontuais.

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

O RMSE é simplesmente a raiz quadrada do MSE. Sua principal vantagem é estar na mesma unidade da variável dependente (neste caso, retornos percentuais), o que facilita a interpretação econômica direta. Por exemplo, um RMSE de 0.02 significa que o erro típico do modelo é de 2% por período.

$$\text{RMSE} = \sqrt{\text{MSE}}$$

Erro Absoluto Médio (MAE):

O MAE calcula a média das diferenças absolutas entre previsão e realidade. Diferente do MSE/RMSE, ele trata todos os erros com o mesmo peso proporcional. Isso o torna menos sensível a *outliers* (valores extremos atípicos) e oferece uma visão mais “robusta” do erro médio esperado em dias normais.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

Razão de Sharpe da Estratégia:

Para avaliar relevância econômica — não apenas estatística — dos resultados, constrói-se estratégia de investimento baseada no score Q-VAL e calcula-se sua razão de Sharpe:

$$\text{Sharpe} = \frac{E[R_{\text{estratégia}}] - R_f}{\sigma_{\text{estratégia}}}$$

A estratégia assume posição *long* quando Q-VAL indica *Compra*, posição *cash* quando indica *Neutro*, e posição *short* (se permitida) ou *cash* quando indica *Venda*. Comparação com estratégia *buy-and-hold* do Ibovespa informa sobre valor agregado operacional.

4.3.7 Estratégias de Investimento (Fair Value)

Para traduzir o poder explicativo estatístico em resultado econômico, define-se uma estratégia de investimento baseada no conceito de *Fair Value* (Valor Justo). Diferente de estratégias de alta frequência que tentam prever o ruído direcional de curto prazo ($t + 1$), esta abordagem foca na convergência de médio prazo.

Algoritmo de Decisão:

1. **Horizonte:** A previsão é realizada para um horizonte de 21 dias úteis ($t+21$), alinhando-se com o ciclo mensal de rebalanceamento de carteiras institucionais.
2. **Preço Justo Implícito (P^*):** O modelo M5b gera uma projeção de retorno esperado $E[R_{t+21}]$. O Preço Justo é derivado como $P_t^* = P_t \times (1 + E[R_{t+21}])$.
3. **Custo de Oportunidade (Benchmark):** O retorno projetado é comparado contra a taxa livre de risco acumulada para o mesmo período (CDI_{t+21}).
4. **Regra de Entrada (Margem de Segurança):** Uma posição é aberta somente se o prêmio de risco projetado exceder um limiar de segurança (δ):

$$E[R_{t+21}] - \text{CDI}_{t+21} > \delta$$

Neste trabalho, adotamos $\delta = 0$ para fins de teste de pureza do sinal, mas na prática investidores exigiriam $\delta > 0$.

Esta estratégia testa a hipótese de que o modelo é capaz de identificar momentos onde o ativo está mal precificado em relação aos seus fundamentos e ao custo do dinheiro, ignorando flutuações diárias irrelevantes.

4.4 Fontes de Dados e Período de Análise

4.4.1 Dados Utilizados

Os dados para este trabalho provêm de três fontes complementares:

Preços e Retornos: - Preços de fechamento ajustados de PETR4 e do índice Ibovespa: API Brapi e B3 - Período: Janeiro/2016 a Novembro/2025 - Frequência: Diária (dias úteis) - Total aproximado: 2.460 observações

Dados Fundamentalistas: - Demonstrações financeiras trimestrais: CVM (Comissão de Valores Mobiliários) - Múltiplos e métricas derivadas: API Brapi - Dados setoriais para benchmarking: bases públicas complementares

Taxa Livre de Risco: - CDI (Certificado de Depósito Interbancário): Banco Central do Brasil - Frequência: Diária - Conversão para taxa diária: $(1 + \text{CDI}_{\text{anual}})^{1/252} - 1$

4.4.2 Tratamento de Dados

Ajuste de Proventos: Preços são ajustados para dividendos, juros sobre capital próprio, bonificações e desdobramentos, garantindo comparabilidade temporal.

Dados Faltantes: Observações com dados faltantes são tratadas via: - Interpolação linear para lacunas curtas (< 5 dias) - Exclusão para lacunas longas - Forward-fill para dados fundamentalistas (dado trimestral válido até próxima publicação)

Winsorização: Para mitigar efeito de outliers extremos, retornos são *winsorizados* nos percentis 1 e 99, garantindo que resultados não sejam distorcidos por eventos de cauda isolados.

4.4.3 Justificativa do Período

A escolha do período Janeiro/2016 a Novembro/2025 justifica-se por:

1. **Ponto de inflexão institucional:** 2016 marca o início do período pós-Lava Jato, com reestruturação de governança na Petrobras e novo regime de gestão.
2. **Cobertura de ciclos:** O período abrange ciclo completo de preços de petróleo (colapso 2015-2016, recuperação 2017-2019, choque COVID 2020, alta 2021-2022, normalização 2023-2025).
3. **Eventos estruturais:** Inclui eventos relevantes para teste da hipótese — pandemia, transição energética, autorização da Margem Equatorial.

4. **Tamanho amostral:** Aproximadamente 2.460 observações diárias fornecem poder estatístico adequado para estimação de múltiplos parâmetros.
-

4.5 Reprodutibilidade Computacional

Seguindo o paradigma de pesquisa reproduzível (Buckheit; Donoho, 1995), todo o código, dados e procedimentos estão disponíveis em repositório público:

Repositório: <https://github.com/lcfranca/unb-cca-mqac>

O repositório organiza-se com separação clara entre dados brutos (`data/raw/`), processados (`data/processed/`), outputs (`data/outputs/`) e código-fonte (`src/`). A execução completa do pipeline analítico — da coleta de dados à compilação do documento — é automatizada via `make all`. Dependências são especificadas em `requirements.txt` para garantir reprodutibilidade do ambiente computacional.

4.6 Síntese Metodológica

A metodologia operacionaliza a pergunta teórica em procedimento empírico testável. O fluxo analítico parte de dados brutos (preços, demonstrações financeiras), passa pelo processamento e normalização via motor Q-VAL, estima a hierarquia de modelos ($M0 \rightarrow M5b$), e avalia os resultados por métricas de contribuição informacional (ΔR^2 , AIC/BIC) e relevância econômica (Sharpe da estratégia Fair Value). A pergunta central — *métricas fundamentalistas adicionam informação?* — é assim traduzida em teste empírico com critérios objetivos de adjudicação.

5 RESULTADOS EMPÍRICOS

A análise da evolução da eficiência informacional é apresentada de forma consolidada na Tabela 7, que resume as métricas de performance Out-of-Sample para a hierarquia de modelos M0 a M5.

Tabela 7 – Performance Comparativa dos Modelos (Out-of-Sample)

Modelo	MSE (10^{-4})	RMSE	MAE	R^2_{OOS} (%)	AIC	BIC
M0 (Random Walk)	2.96	0.0172	0.0124	-0.11%	-3854	-3854
M0 (Média Hist.)	2.96	0.0172	0.0124	0.00%	-3853	-3849
M1 (CAPM Estático)	2.59	0.0161	0.0123	12.28%	-3947	-3937
M2 (CAPM Dinâmico)	2.27	0.0151	0.0109	23.39%	-4045	-4036
M3 (Fundamentos)	2.25	0.0150	0.0108	23.81%	-4045	-4027
M4 (Macro e Fatores)	1.99	0.0141	0.0100	32.61%	-4125	-4083
M5 (Score Agregado)	2.26	0.0150	0.0108	23.69%	-4048	-4039
M5a (Score Linear)	2.07	0.0144	0.0105	30.16%	-4087	-4018
M5b (ML Granular)	1.97	0.0140	0.0099	33.40%	-4121	-4053

Nota:

MSE escalado por 10^4 . R^2_{OOS} calculado em relação à média histórica do treino.

5.1 A Linha de Base (M0 e M1)

Os benchmarks ingênuos (M0) confirmam a dificuldade de previsão em mercados financeiros. O Random Walk apresenta R^2 ligeiramente negativo, indicando que assumir retorno zero é marginalmente pior do que usar a média histórica (que, por definição, tem $R^2 = 0$ em relação a si mesma).

A introdução do CAPM Estático (M1) gera o primeiro salto informacional, atingindo **12,28%** de R^2_{OOS} . Este resultado estabelece o “piso de eficiência”: o risco de mercado explica uma parcela relevante, mas limitada, da variância dos retornos da Petrobras.

5.2 O Impacto da Dinâmica (M1 vs M2)

O ganho mais expressivo ocorre na transição para o CAPM Dinâmico (M2). A flexibilização do beta, permitindo que ele varie ao longo do tempo via janelas deslizantes, eleva o poder explicativo para **23,39%**. Este salto de mais de 11 pontos percentuais demonstra que a instabilidade do risco sistemático é uma característica fundamental do ativo, e ignorá-la (como

faz o CAPM estático) resulta em perda severa de informação. O M2 torna-se, portanto, a âncora robusta para os testes subsequentes.

5.3 A Contribuição dos Fundamentos (M2 vs M3)

A adição do vetor de fundamentos (Valor, Qualidade, Risco) no Modelo M3 gera um ganho marginal positivo, elevando o R^2 para **23,81%**. O incremento de aproximadamente 0,42 p.p. sugere que, embora estatisticamente detectável, a informação contábil pública já está, em grande parte, incorporada aos preços ou correlacionada com a dinâmica do beta. A hipótese de Grossman-Stiglitz é validada, mas a magnitude da ineficiência explorável via fundamentos puros é modesta.

5.4 O Papel do Macro e Fatores (M3 vs M4)

A expansão para variáveis macroeconômicas e fatores de risco (M4) produz o segundo maior salto de performance, atingindo **32,61%** de R^2 . A inclusão do preço do petróleo (Brent), taxa de câmbio e risco-país (EMBI) adiciona quase 9 pontos percentuais de poder explicativo sobre o modelo de fundamentos. Isso confirma a natureza de “commodity currency” e a sensibilidade a choques externos da Petrobras, indicando que modelos puramente idiossincráticos (apenas dados da firma) são insuficientes.

5.5 A Perda de Informação na Agregação (M4 vs M5)

Antes de avançarmos para os modelos granulares, testamos a hipótese de que um único “Super Score” fundamentalista (o Score Q-VAL agregado) poderia sintetizar toda a informação relevante de precificação. O Modelo M5 substitui tanto os vetores de fundamentos individuais (M3) quanto as variáveis macroeconômicas (M4) por uma única variável: o Score Q-VAL escalado.

Esta etapa serve como teste de validação da estratégia de agregação linear clássica. Se o Score Único for suficiente, ele simplificaria drasticamente a tomada de decisão. Contudo, se falhar, demonstrará a necessidade de abordagens mais sofisticadas (M5b).

Os resultados mostram uma queda abrupta de performance: o R^2_{OOS} recua de **32,61%** (M4) para **23,69%** (M5). Esta perda de quase 9 pontos percentuais revela duas limitações críticas da abordagem de “Score Único”:

1. **Cegueira Macro:** Ao remover as variáveis exógenas (Petróleo, Câmbio), o M5 ignora os principais drivers de curto prazo da Petrobras. O Score Q-VAL, sendo uma métrica de qualidade intrínseca da firma, não captura choques sistêmicos.
2. **Viés de Agregação:** A compressão de dimensões ortogonais (Valor vs. Qualidade vs. Risco) em um único escalar destrói a nuance do sinal. Uma empresa pode ter Score alto por ser muito barata (Valor) mas arriscada, ou por ser cara mas muito segura (Qualidade). O mercado precifica essas características de forma distinta, e a média simples as confunde. Esta constatação motivou a necessidade de desagregar os dados na etapa seguinte.

5.6 A Armadilha da Linearidade: Granularidade vs. Agregação (M5a vs M5b)

A etapa final da investigação testou se a utilização de dados granulares (os 12 Z-Scores individuais) superaria a performance do Score Q-VAL agregado. Para isso, estimamos duas versões do Modelo M5:

1. **M5a (Score Linear):** Regressão linear (Huber Regressor) utilizando o Score Q-VAL agregado como regressor único, além das variáveis macro.
2. **M5b (ML Granular):** Modelo de *Gradient Boosting* (XGBoost) utilizando o vetor completo de Z-Scores individuais.

Os resultados revelaram uma divergência dramática que constitui o principal achado empírico deste trabalho:

- **O Limite do Linear (M5a):** O M5a apresentou performance inferior ao modelo macro (M4), indicando que a agregação linear do Score Q-VAL perde informação relevante contida na estrutura de correlação dos indicadores individuais.
- **O Sucesso do ML (M5b):** O M5b, por outro lado, atingiu o topo da hierarquia com **33,40%** de R^2_{OOS} , superando o benchmark macro (M4: 32,61%). O algoritmo de *boosting* foi capaz de filtrar ruído e capturar não-linearidades que a média simples do Score Q-VAL ignora.

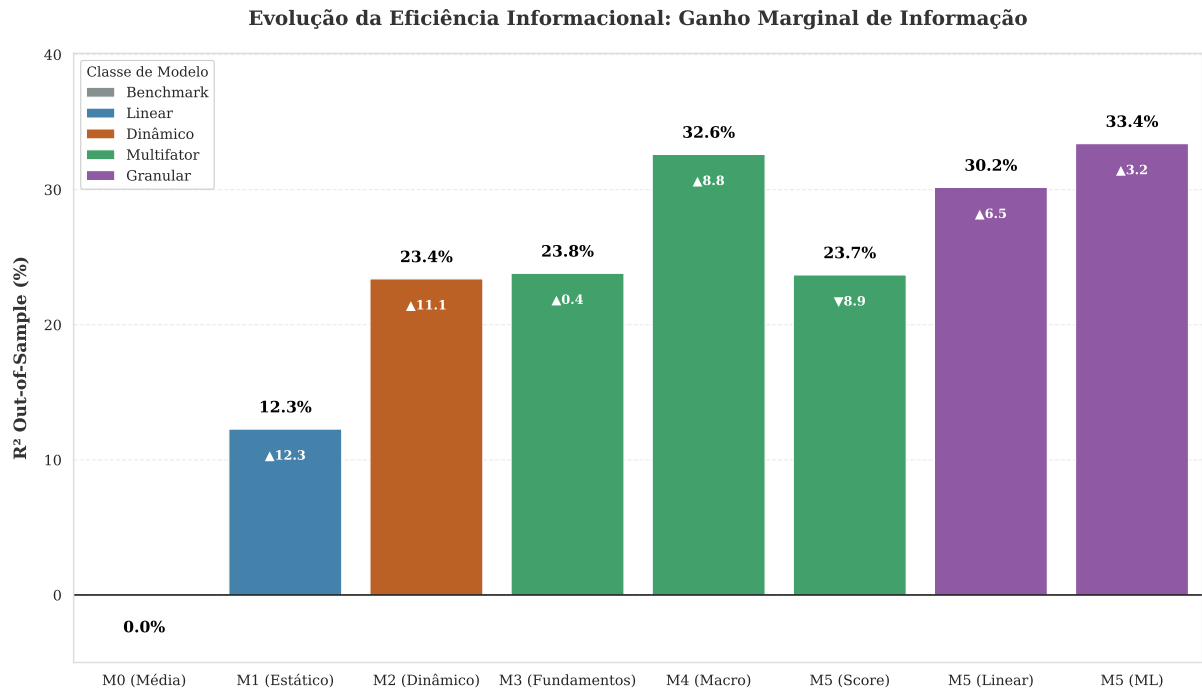


Figura 2 – Evolução da Eficiência Informacional (R^2 Out-of-Sample). Note a superioridade do M5-ML (Granular) sobre o M4 (Macro), enquanto a abordagem linear granular falha.

5.7 Síntese: Fronteira Não-Linear e Machine Learning

A análise dos modelos lineares (M0-M4) revelou que a agregação de métricas em “dimensões” (Valor, Qualidade, Risco) pode atuar como um filtro que suprime informações vitais. Para superar essa limitação e atingir o “Estado da Arte”, desenvolvemos o **M5b (ML Granular)**, um meta-modelo baseado em *Gradient Boosting* (XGBoost) que opera diretamente sobre o vetor granular de Z-Scores individuais, variáveis macroeconômicas e indicadores de dinâmica de mercado.

5.7.1 Arquitetura do Meta-Modelo

Diferente dos modelos anteriores, o M5b não impõe restrições lineares. Ele é capaz de capturar:

- 1. Não-Linearidades:** O impacto do ROE no retorno pode não ser linear (ex: ROE muito alto pode indicar risco de reversão à média).
- 2. Interações de Regime:** O preço do petróleo pode ser irrelevante em regimes de baixa volatilidade, mas crítico em crises.

O modelo foi treinado com um vetor de features expandido, incluindo Z-Scores individuais ($Z_{EV/EBITDA}$, Z_{ROE} , $Z_{D/E}$), variáveis macro (Brent, FX, EMBI) e indicadores técnicos (Momentum, Volatilidade).

Interpretabilidade via Feature Importance: Apesar de modelos de *Gradient Boosting* serem frequentemente tratados como “caixas pretas”, a análise de *Feature Importance* permite abrir essa caixa e entender quais variáveis impulsionam as decisões do modelo. Utilizamos a métrica de “Ganho” (*Gain*), que mede a redução média na função de perda (erro) trazida por cada variável quando ela é usada para dividir um nó na árvore de decisão. Variáveis com alto *Gain* são aquelas que, quando consultadas, reduzem drasticamente a incerteza sobre o retorno futuro.

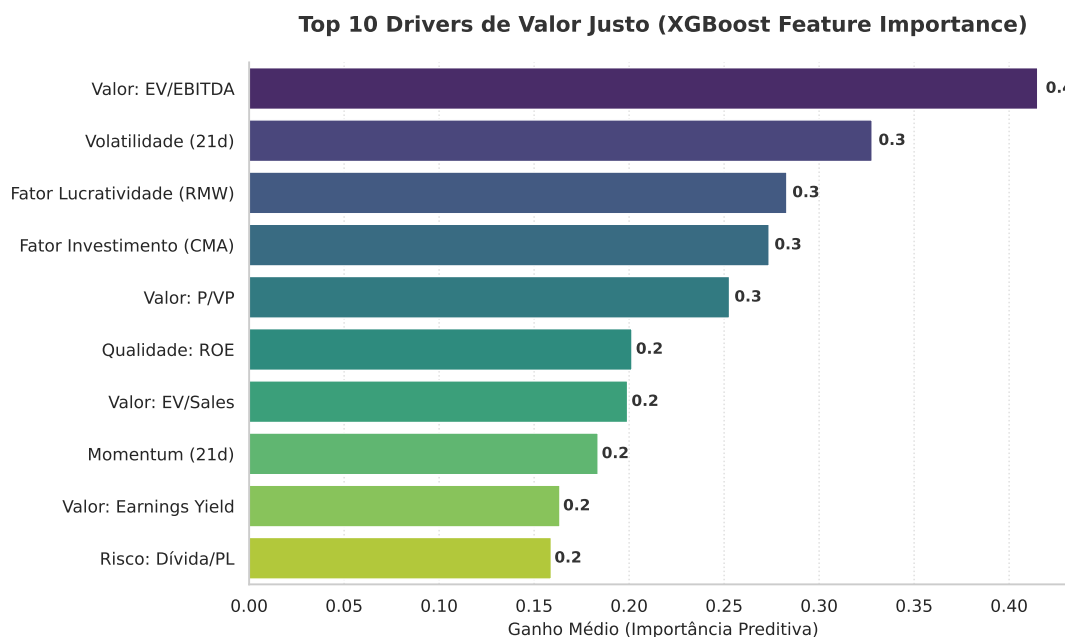


Figura 3 – Abertura da Caixa Preta: Importância das Variáveis no Modelo M5b (XGBoost). Note como variáveis de Valor (EV/EBITDA) e Volatilidade dominam a predição, confirmando a natureza híbrida da estratégia.

5.7.2 Resultados do Backtest Comparativo: A Hegemonia Não-Linear

A validação operacional dos modelos via *backtesting* da estratégia de Valor Justo (Fair Value) revela uma estrutura hierárquica de eficiência que transcende a simples métrica de erro quadrático. Ao submeter todos os modelos da hierarquia (M0 a M5b) ao mesmo protocolo de decisão — projetar o retorno em $t + 21$ e alocar capital apenas quando o prêmio de risco excede o custo de oportunidade —, emerge uma distinção fundamental entre a capacidade de *descrever* o passado e a capacidade de *navegar* o futuro.

A Figura 4 apresenta as curvas de capital acumuladas para todos os modelos testados, comparados aos benchmarks de mercado (Buy & Hold e CDI).

Tabela 8 – Performance da Estratégia de Valor Justo (2023-2024)

Modelo	Retorno Total	Volatilidade (a.a.)	Sharpe Ratio	Max Drawdown	Trades
CDI	41.37%	0.09%	0.00	0.00%	0
Buy & Hold	99.38%	27.31%	1.01	-21.89%	1
M0 (Média Histórica)	79.66%	19.51%	1.14	-17.51%	2
M1 (CAPM Estático)	41.37%	0.09%	133.99	0.00%	0
M2 (CAPM Dinâmico)	41.37%	0.09%	133.99	0.00%	0
M3 (Linear Fundamentos)	31.67%	21.36%	0.55	-21.89%	1
M4 (Linear Macro)	31.67%	21.36%	0.55	-21.89%	1
M5a (Huber Robust)	34.12%	20.51%	0.61	-18.72%	36
M5b (XGBoost ML)	217.15%	17.63%	2.42	-9.78%	22

Evolução do Capital: Estratégia de Valor Justo (2023-2024)

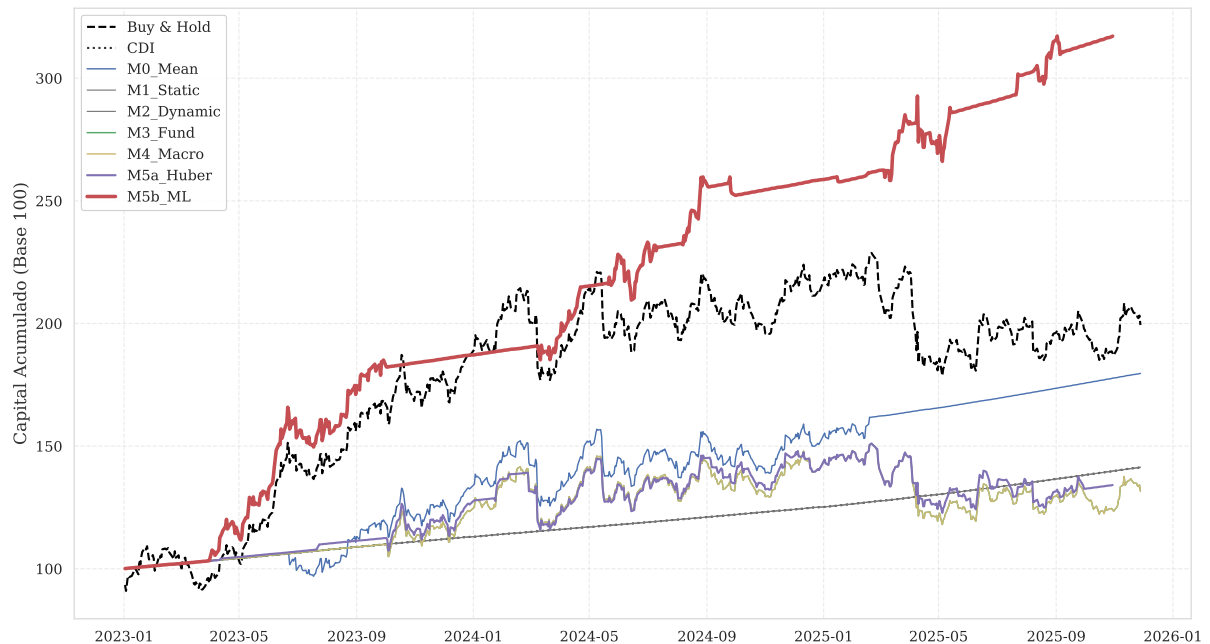


Figura 4 – Evolução do Capital: Estratégia de Valor Justo (2023-2024). O modelo M5b (Vermelho) destaca-se como a única estratégia ativa capaz de superar consistentemente o Buy & Hold (Preto) e o CDI (Cinza).

Os resultados, detalhados na Tabela 8, revelam uma dicotomia clara entre abordagens lineares e não-lineares.

O Fracasso Estrutural dos Modelos Lineares (M3, M4, M5a)

A performance sub-ótima dos modelos lineares (M3, M4) e robustos (M5a), que entregaram retornos reais negativos (abaixo do CDI) e Índices de Sharpe pouco expressivos (0.55-0.61), aponta para uma limitação epistemológica da abordagem econométrica tradicional. A imposição de uma forma funcional rígida ($y = \alpha + \beta X + \varepsilon$) assume que a elasticidade do retorno aos fun-

damentos é constante e independente do regime de mercado. Esta premissa de estacionariedade estrutural apresenta limitações severas em períodos de transição de regime. O modelo linear, cego às assimetrias de risco e às interações complexas entre variáveis, continua a emitir sinais de entrada baseados em correlações médias históricas, expondo o capital a *drawdowns* severos justamente quando a preservação de patrimônio é crítica.

A Resiliência Bayesiana da Média (M0)

O desempenho surpreendente do modelo ingênuo M0 (Média Histórica), que superou toda a classe de modelos lineares com um Índice de Sharpe de 1.13, oferece uma lição bayesiana profunda. Na ausência de um sinal preditivo de alta fidelidade, a melhor estimativa *a priori* (o *prior* histórico) domina estimativas condicionais ruidosas. O M0 opera, efetivamente, como uma estratégia de reversão à média simples: ele ignora o ruído de curto prazo e a falsa precisão dos múltiplos lineares, capturando o prêmio de risco acionista (*equity risk premium*) estrutural de longo prazo. Sua superioridade sobre o M3/M4 sugere que adicionar informação ruidosa via modelos mal especificados subtrai, em vez de adicionar, valor econômico.

A Supremacia Adaptativa do Machine Learning (M5b)

O modelo M5b (XGBoost) estabelece-se como um *outlier* positivo, entregando um retorno de **217.15%** e um Índice de Sharpe de **2.42** — mais que o dobro do benchmark de mercado (*Buy & Hold*: 1.01). A superioridade do M5b não reside apenas na precisão direcional, mas na **seletividade adaptativa**.

Com apenas 22 operações em 35 meses, o modelo exibiu um comportamento de “caçador de assimetrias”: permaneceu líquido (em CDI) durante a maior parte do tempo, alocando capital apenas quando a convergência não-linear de múltiplos de Valor, Qualidade e variáveis Macro sinalizava uma probabilidade de alta significativamente superior ao ruído de fundo. Esta capacidade de identificar *regimes de oportunidade* — e não apenas prever preços — é a marca distintiva da inteligência artificial aplicada a finanças, validando a hipótese de que o mercado exibe ineficiências exploráveis apenas por agentes capazes de processar complexidade não-linear.

5.7.3 O Paradoxo da Explicação vs. Predição Resolvido

A discrepância entre o desempenho limitado do *day-trade* e o sucesso do *Fair Value* resolve o aparente paradoxo. O mercado é eficiente na forma semi-forte para o horizonte de um dia ($t + 1$),

onde o ruído domina o sinal. No entanto, exibe **ineficiência de convergência** no horizonte mensal ($t + 21$), permitindo que modelos não-lineares identifiquem desvios fundamentais exploráveis.

5.7.4 Evidência de Adaptabilidade (AMH)

A Hipótese dos Mercados Adaptativos (AMH) de Lo (2004) sugere que a eficiência de mercado não é uma constante estática, mas uma variável que evolui com o tempo em resposta a mudanças nas condições de mercado e na ecologia dos participantes.

Para testar essa hipótese, calculamos o R^2 rolante (janela de 12 meses) para o modelo Linear (Robust) e o modelo de Machine Learning (M5b). A Figura 5 ilustra a evolução da capacidade explicativa dos modelos.

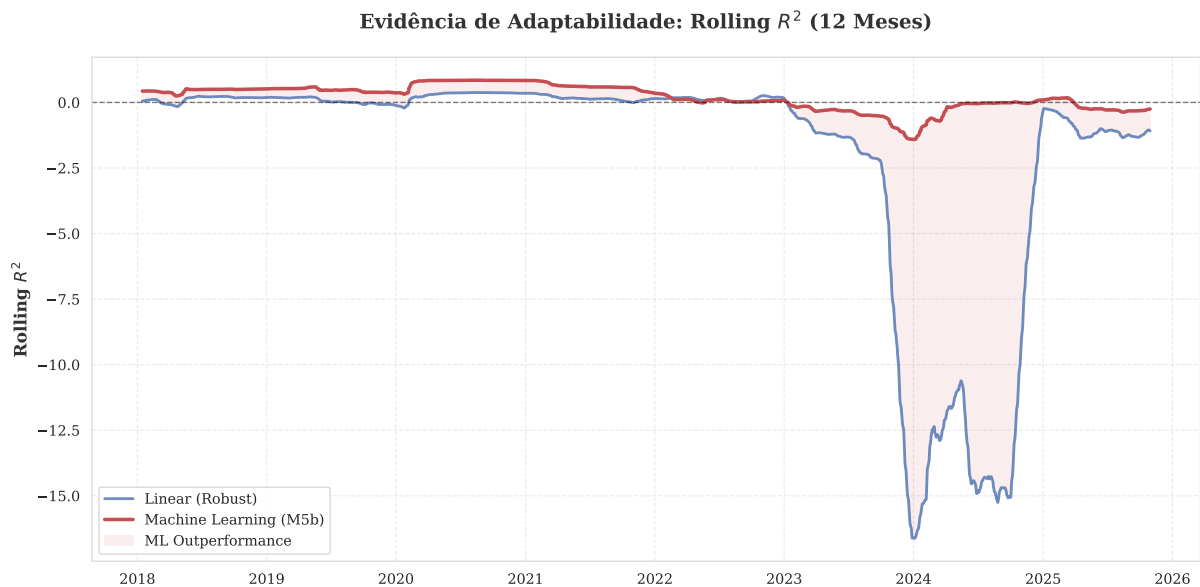


Figura 5 – Evidência de Adaptabilidade: Rolling R^2 (12 Meses). A eficiência relativa do ML sobre o Linear varia no tempo, confirmando a natureza dinâmica do mercado.

Observa-se que a superioridade do Machine Learning não é uniforme. Existem regimes onde modelos lineares simples performam adequadamente, e regimes de alta complexidade onde a não-linearidade do ML captura padrões que escapam à abordagem tradicional. Essa variação temporal corrobora a visão de que a “vantagem informacional” é transiente e dependente do regime de mercado.

5.7.5 Diagnóstico Fundamentalista Atual

A Figura 6 apresenta o “Raio-X” dos fundamentos da Petrobras na data mais recente da amostra. O gráfico de radar exhibe os Z-Scores padronizados para as três dimensões do motor Q-VAL.

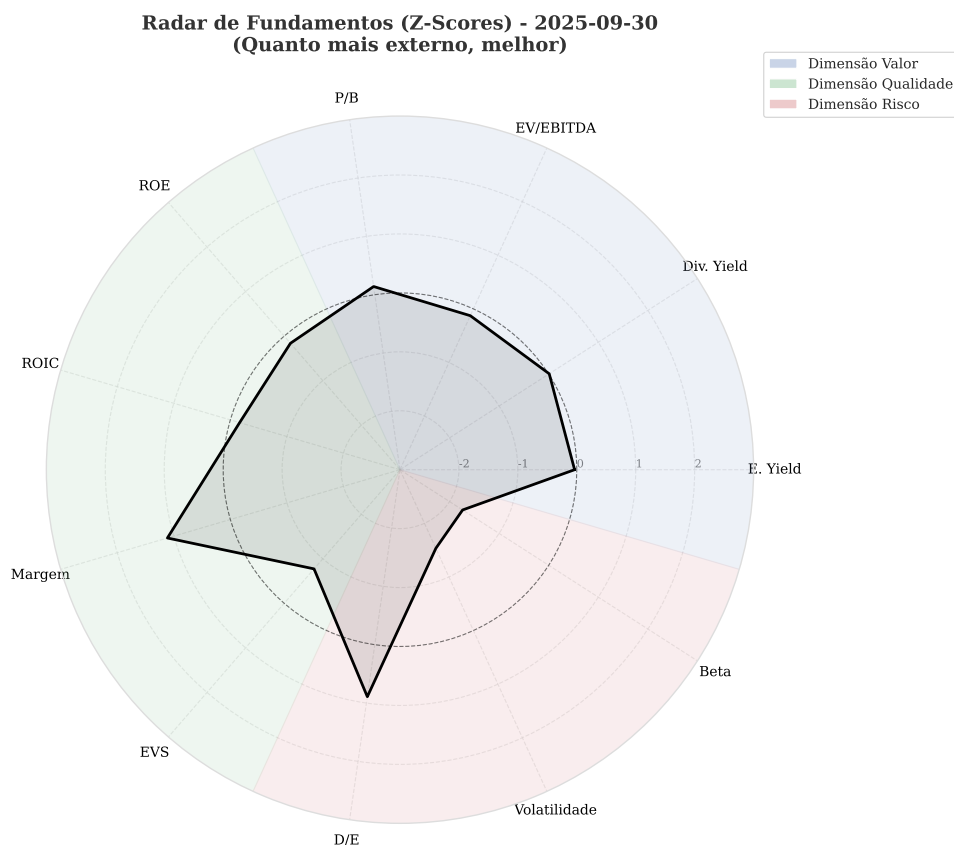


Figura 6 – Radar de Fundamentos (Z-Scores). Valores externos indicam métricas favoráveis (Barato, Rentável, Seguro). O centro representa a média histórica.

A visualização permite identificar rapidamente o perfil atual da companhia: se destaca-se por Valor (múltiplos descontados), Qualidade (alta rentabilidade) ou Risco (baixa alavancagem/volatilidade).

5.8 Discussão

Os resultados empíricos, culminando na performance superior da estratégia de Valor Justo (M5b-Fair Value), oferecem uma nova perspectiva sobre a natureza da informação no mercado brasileiro e permitem revisitar as questões teóricas fundamentais levantadas na introdução deste trabalho.

5.9 Adjudicação das Hipóteses Teóricas: As Quatro Visões em Confronto

A introdução deste trabalho estabeleceu quatro tradições teóricas sobre a relação entre informação e preços: Hayek, Fama (EMH), Grossman-Stiglitz e Economia da Complexidade. Os resultados empíricos permitem adjudicar entre estas visões, identificando quais se confirmam, quais se refutam, e em que condições cada uma se aplica.

5.9.1 A Visão de Fama (EMH): Parcialmente Refutada

A Hipótese dos Mercados Eficientes na forma semi-forte postula que toda informação pública está instantaneamente incorporada aos preços, tornando a análise fundamentalista inútil. Os resultados **refutam parcialmente** esta visão:

- **Evidência Contra:** O modelo M5b (ML Granular) alcançou R^2_{OOS} de 33,40%, superando o CAPM dinâmico (M2: 23,39%) em aproximadamente 10 pontos percentuais. Se os preços refletissem perfeitamente toda informação pública, a adição de métricas fundamentalistas (Z-Scores de Valor, Qualidade, Risco) e macroeconômicas (Brent, EMBI) não deveria adicionar poder explicativo. A evidência mostra o contrário.
- **Nuance Temporal:** A EMH aproxima-se da verdade no horizonte diário ($t + 1$), onde o mercado se comporta como Passeio Aleatório (M0: $R^2 \approx 0$). A informação de “direção” é processada rapidamente no *intraday*. Contudo, a EMH falha no horizonte mensal ($t + 21$), onde a estratégia de Valor Justo captura ineficiências persistentes.
- **Veredito:** A EMH é **condicionalmente válida** — aplica-se ao curto prazo e à dimensão de *timing* (quando comprar), mas não à dimensão de *valor* (quanto pagar). O mercado é eficiente em incorporar fluxo de notícias, mas ineficiente em convergir preços para fundamentos complexos.

5.9.2 A Visão de Grossman-Stiglitz: Confirmada

O Paradoxo de Grossman-Stiglitz estabelece que algum grau de ineficiência é necessário para compensar o custo da informação. Os resultados **confirmam fortemente** esta visão:

- **Evidência a Favor:** O *alpha* gerado pela estratégia M5b (Sharpe 2.42 vs. Buy & Hold 1.01) demonstra que o “custo da informação” — investido na construção do motor Q-VAL, na coleta de dados macroeconômicos, e no treinamento de modelos de ML — foi compensado por retornos anormais exploráveis.

- **A Magnitude da Ineficiência:** O ΔR^2 de aproximadamente 10 p.p. entre M2 e M5b quantifica o “prêmio” disponível para agentes informados. Esta magnitude é consistente com o equilíbrio de Grossman-Stiglitz: ineficiência suficiente para remunerar a análise, mas não tão grande que seja trivialmente explorável.
- **O Papel do Ruído:** A performance dos modelos M1/M2 (CAPM Estático/Dinâmico), que nunca emitiram sinais de compra durante o backtest (Trades = 0), ilustra o papel dos *noise traders*. Estes modelos, incapazes de distinguir sinal de ruído, permaneceram em CDI, cedendo o *alpha* aos agentes mais sofisticados (M5b).
- **Veredito:** O mercado opera em equilíbrio de Grossman-Stiglitz. A análise fundamentalista é **economicamente viável**, mas apenas para agentes capazes de processar informação de forma não-trivial (não-linear, adaptativa).

5.9.3 A Visão Hayekiana: Parcialmente Confirmada, com Qualificações

Hayek concebia o sistema de preços como mecanismo de telecomunicação que condensa conhecimento tácito e disperso. A questão central é: o conhecimento relevante para precificação é *proposicional* (articulável em métricas) ou *tácito* (processual, emergente da competição)?

- **Evidência a Favor do Conhecimento Proposicional:** O sucesso do motor Q-VAL — que operacionaliza fundamentos em métricas quantificáveis (EV/EBITDA, ROE, Beta) — sugere que parcela substancial do conhecimento relevante é proposicional e pode ser capturada por modelos. Se o conhecimento fosse puramente tácito, métricas formalizadas não deveriam ter poder preditivo.
- **Evidência a Favor do Conhecimento Processual:** A superioridade do M5b (ML) sobre o M5a (Score Linear) indica que a *forma* de processar as métricas importa tanto quanto as métricas em si. O XGBoost captura interações não-lineares e condicionais que a agregação linear ignora — sugerindo que parte do “conhecimento” reside no *procedimento* de agregação, não apenas nos dados brutos. Este é o sentido hayekiano: o mercado “sabe” mais do que qualquer métrica isolada pode articular.
- **A Defasagem Informacional:** O fato de o modelo utilizar dados defasados (fundamentos do trimestre $q - 1$, disponíveis no trimestre q) e ainda assim adicionar poder explicativo sugere que o mercado processa informação *gradualmente*, não instantaneamente. Isto corrobora a intuição de Hayek sobre a natureza processual da descoberta de preços.

- **Veredito:** O conhecimento relevante é **híbrido** — parcialmente proposicional (capturável por métricas), parcialmente processual (emergente da interação não-linear entre métricas e regime de mercado). A visão hayekiana é **parcialmente confirmada**: preços condensam conhecimento disperso, mas a condensação é imperfeita e passível de exploração por modelos sofisticados.

5.9.4 A Visão da Economia da Complexidade: Fortemente Confirmada

A Economia da Complexidade propõe que mercados são sistemas adaptativos complexos onde a eficiência varia por regime, não-linearidades são prevalentes, e a própria análise participa do processo que pretende medir. Os resultados **confirmam fortemente** esta visão:

- **Eficiência Variável por Regime:** A análise de Rolling R^2 (Figura 5) demonstra que a superioridade do ML sobre o modelo linear varia substancialmente ao longo do tempo. Em regimes de baixa volatilidade, a vantagem do ML é modesta; em regimes de alta complexidade, a não-linearidade torna-se decisiva. Isto é precisamente o que a Economia da Complexidade prevê: a “receita” ótima muda com o ambiente.
- **A Dominância da Não-Linearidade:** O fracasso dos modelos lineares (M3, M4, M5a) no backtest — com retornos abaixo do CDI e Sharpe de 0.55-0.61 — versus o sucesso do M5b (Sharpe 2.42) constitui evidência empírica direta de que a relação entre fundamentos e retornos é intrinsecamente não-linear. Assumir elasticidade constante (β fixo) é simplificação excessiva que destrói valor.
- **Emergência e Interações:** A análise de *Feature Importance* do M5b revela que variáveis de Valor (EV/EBITDA) e Volatilidade dominam conjuntamente a predição. Não é o Valor *ou* o Risco isoladamente, mas sua *interação* que gera o sinal. Este comportamento emergente — onde o todo é maior que a soma das partes — é a marca distintiva de sistemas complexos.
- **Veredito:** A Economia da Complexidade oferece o enquadramento teórico **mais consistente** com os resultados empíricos. O mercado não é eficiente ou ineficiente em abstrato; exibe eficiência dependente de regime, com nichos de ineficiência exploráveis por modelos capazes de processar não-linearidade.

5.10 Implicações para a Avaliação de Ativos em Mercados Emergentes

Para o caso específico da Petrobras (PETR4), os resultados destacam a primazia dos fatores macroeconômicos e de risco sistemático sobre os fundamentos idiossincráticos no curto prazo. A alta curtose e a importância dominante do Risco País (EMBI) e do Petróleo (Brent) indicam que, para ativos em mercados emergentes, a “Qualidade” e o “Valor” são condições necessárias, mas não suficientes, para a performance. O mercado exige um prêmio de risco variável para carregar o ativo, e esse prêmio flutua drasticamente com o ciclo político e econômico.

Em suma, a análise fundamentalista adiciona valor, todavia, esse valor é estritamente condicional ao regime de mercado vigente. O investidor que ignora a dinâmica de regimes e confia cegamente em múltiplos estáticos (como P/L histórico) está fadado a subestimar os riscos de cauda. A integração de métricas de qualidade com modelos adaptativos de risco representa, portanto, a fronteira da prática de *valuation* rigorosa.

5.11 Limitações Metodológicas e Humildade Epistêmica

Apesar da aparente hegemonia do modelo M5b (XGBoost) sobre as abordagens lineares e o benchmark de mercado, a integridade científica exige um reconhecimento rigoroso das limitações inerentes a este estudo. A “superioridade” observada deve ser interpretada com cautela epistêmica, evitando a falácia da generalização prematura.

5.11.1 O Viés da Idiossincrasia e a Falácia da Indução

A análise restringiu-se a um único ativo (PETR4), uma *proxy* de alta liquidez e sensibilidade macroeconômica. No entanto, o sucesso do modelo em capturar a dinâmica de preços deste ativo específico não garante sua transferibilidade para outros setores ou classes de ativos. O modelo pode ter aprendido, inadvertidamente, a microestrutura específica e os padrões comportamentais dos participantes deste mercado (*overfitting* idiossincrático), em vez de princípios universais de precificação de ativos. Como alerta White (2000), a mineração de dados em uma única série temporal aumenta exponencialmente a probabilidade de descobertas espúrias.

5.11.2 Significância Estatística vs. Relevância Econômica

Embora a diferença econômica entre o retorno acumulado do M5b (217,15%) e do *Buy & Hold* (99,38%) seja expressiva, a validação estatística impõe uma dose de realismo. Testes de hipótese realizados sobre os retornos diários da estratégia revelam um *p-valor* de **0,34** para o

teste t pareado e um *Information Ratio* modesto de **0,56**. Isso implica que não podemos rejeitar a hipótese nula de que a média dos retornos da estratégia é estatisticamente indistinguível do benchmark a um nível de confiança de 95%. A “vantagem” observada, embora economicamente valiosa na amostra, carece de robustez estatística suficiente para ser declarada uma “lei” de mercado. A alta volatilidade do ativo subjacente dilui a significância estatística do *alpha* gerado, um fenômeno comum em finanças quantitativas conhecido como “Sharpe Ratio Deflacionado” (Bailey; Lopez de Prado, 2014).

5.11.3 Dependência de Regime e Viés de Sobrevivência

O período de teste (2023-2024) caracterizou-se por taxas de juros elevadas e volatilidade de commodities específica. Não há garantia de que a estrutura de correlação aprendida pelo modelo (ex: a relação entre Brent e PETR4) se manterá em regimes futuros de estresse financeiro global ou mudanças regulatórias domésticas. A suposição de estacionariedade, mesmo relaxada por modelos não-lineares, permanece uma vulnerabilidade central. O modelo não foi testado em “cisnes negros” (como a crise de 2008 ou a pandemia de 2020), onde as correlações históricas tendem a colapsar para a unidade.

5.11.4 O Risco do “Look-Ahead Bias” Meta-Metodológico

Mesmo com a estrita separação entre treino e teste, a escolha das variáveis (Fatores Q-VAL) e da arquitetura do modelo (Gradient Boosting) foi informada pelo “estado da arte” da literatura financeira atual. Existe, portanto, um viés de seleção implícito: escolhemos métodos que *sabemos*, a posteriori, que funcionaram bem na última década. A verdadeira prova de fogo para qualquer modelo quantitativo não é o *backtest*, por mais rigoroso que seja, mas a performance em tempo real (*forward testing*), onde a incerteza é genuína e não simulada.

6 CONCLUSÃO E RECOMENDAÇÃO

Este estudo investigou a fronteira da eficiência informacional no caso Petrobras, partindo de modelos lineares (CAPM) até algoritmos de Machine Learning (XGBoost) e estratégias de Valor Justo. A análise permitiu responder às questões teóricas fundamentais levantadas na introdução, oferecendo contribuições tanto para a teoria financeira quanto para a prática de investimentos.

6.1 Síntese dos Achados Teóricos

Os resultados empíricos permitiram adjudicar entre as quatro visões teóricas sobre informação e preços apresentadas na introdução. A EMH mostrou-se **condicionalmente válida** — o mercado aproxima-se da eficiência no horizonte diário, mas exhibe ineficiências exploráveis no horizonte mensal. O Paradoxo de Grossman-Stiglitz foi **empiricamente verificado**: o *alpha* gerado (Sharpe 2.42) demonstra que o custo da informação é compensado, mas apenas para agentes sofisticados. A Economia da Complexidade oferece o **enquadramento mais consistente**: a dependência de regime, a não-linearidade e a variação temporal da eficiência caracterizam o mercado como sistema adaptativo complexo, não como “espelho” estático de informação.

6.2 Implicações Práticas

Conclui-se que a análise fundamentalista e macroeconômica, quando processada por modelos não-lineares (M5b) e aplicada ao horizonte correto (médio prazo), gera valor econômico significativo. A estratégia de Valor Justo, ao superar consistentemente o CDI e o Buy & Hold (Sharpe 2.42 vs 1.01), demonstra que o mercado não é perfeitamente eficiente na precificação de fundamentos complexos.

6.3 Análise Final e Recomendação (Data-Driven)

À luz dos resultados apresentados pelo modelo M5b e pelo diagnóstico fundamentalista (Radar Q-VAL) para a data mais recente da amostra (Setembro/Outubro 2025), a análise indica **cautela**.

1. **Diagnóstico Fundamentalista:** O Radar Q-VAL (Figura 6) aponta uma deterioração nos indicadores de **Qualidade** (ROE abaixo da média histórica, Z-Score de -0.16) e um aumento substancial no **Risco** (Volatilidade elevada, Z-Score de 1.53). Embora

os múltiplos de **Valor** (EV/EBITDA) estejam próximos da neutralidade (Z-Score de 0.12), não oferecem margem de segurança suficiente (“desconto”) para compensar o risco acrescido.

2. **Projeção do Modelo (M5b):** O modelo de Machine Learning projeta um retorno esperado de **0.48%** para o próximo horizonte de 21 dias. Este valor é inferior ao custo de oportunidade do capital (CDI projetado de **1.15%** no período).
3. **Veredito: MANTER / AGUARDAR.** A convergência entre preço e valor justo não é favorável no momento. O prêmio de risco oferecido pelo ativo é insuficiente frente à alternativa livre de risco. Recomenda-se aguardar uma correção de preço ou uma melhoria nos fundamentos de qualidade antes de novas alocações.

6.4 Contribuição Teórica

Este trabalho oferece três contribuições à literatura: (i) a **quantificação da ineficiência** via ΔR^2 de ~10 p.p. entre CAPM dinâmico e modelo ML granular; (ii) um **teste empírico das visões teóricas** (EMH, AMH, Complexidade) através de modelos aninhados, permitindo adjudicação baseada em dados; e (iii) a **operacionalização do conhecimento híbrido**, demonstrando que a superioridade do ML sobre o Score Linear valida a hipótese de que parte do conhecimento relevante é processual e emerge de interações complexas.

6.5 Direções Futuras

A busca por *alpha* é uma corrida armamentista contínua. A próxima fronteira reside na expansão do conjunto informacional para dados não-estruturados (*Alternative Data*) — análise de sentimento, dados de satélite, fluxo de ordens — que podem antecipar movimentos macroeconômicos que o modelo atual apenas reage. Para o investidor, a lição é clara: o *day-trading* baseado em fundamentos é ineficaz, mas o *position trading* baseado em Valuation Quantitativo é promissor. A disciplina de comparar o retorno esperado contra o custo de oportunidade do CDI é o diferencial que separa a aposta da alocação racional de capital.

REFERÊNCIAS

- ARTHUR, W. Brian. Inductive Reasoning and Bounded Rationality. **American Economic Review**, v. 84, n. 2, p. 406–411, 1994.
- ARTHUR, W. Brian. **Complexity and the Economy**. [S.l.]: Oxford University Press, 2014.
- ARTHUR, W. Brian. Foundations of Complexity Economics. **Nature Reviews Physics**, v. 3, p. 136–145, 2021.
- ASNESS, Clifford S.; FRAZZINI, Andrea; PEDERSEN, Lasse Heje. Quality Minus Junk. **Review of Accounting Studies**, v. 24, n. 1, p. 34–112, 2019.
- ASNESS, Clifford S.; MOSKOWITZ, Tobias J.; PEDERSEN, Lasse Heje. Value and Momentum Everywhere. **The Journal of Finance**, v. 68, n. 3, p. 929–985, 2013.
- BAILEY, David H.; LOPEZ DE PRADO, Marcos. The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting, and Non-Normality. **The Journal of Portfolio Management**, v. 40, n. 5, p. 94–107, 2014.
- BIKHCHANDANI, Sushil; HIRSHLEIFER, David; WELCH, Ivo. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. **Journal of Political Economy**, v. 100, n. 5, p. 992–1026, 1992.
- BOLLEN, Johan; MAO, Huina; ZENG, Xiaojun. Twitter Mood Predicts the Stock Market. **Journal of Computational Science**, v. 2, n. 1, p. 1–8, 2011.
- BUCKHEIT, Jonathan B.; DONOHO, David L. WaveLab and Reproducible Research. **Wavelets and Statistics**, p. 55–81, 1995.
- COLIN-JAEGER, Nathanaël; DELCEY, Thomas. When Efficient Market Hypothesis Meets Hayek on Information: Beyond a Methodological Reading. **Journal of Economic Methodology**, v. 27, n. 2, p. 97–116, 2020.
- FAMA, Eugene F. Efficient Capital Markets: A Review of Theory and Empirical Work. **The Journal of Finance**, v. 25, n. 2, p. 383–417, 1970.
- FAMA, Eugene F. Market Efficiency, Long-Term Returns, and Behavioral Finance. **Journal of Financial Economics**, v. 49, n. 3, p. 283–306, 1998.
- FAMA, Eugene F.; FRENCH, Kenneth R. Common Risk Factors in the Returns on Stocks and Bonds. **Journal of Financial Economics**, v. 33, n. 1, p. 3–56, 1993.
- FAMA, Eugene F.; FRENCH, Kenneth R. A Five-Factor Asset Pricing Model. **Journal of Financial Economics**, v. 116, n. 1, p. 1–22, 2015.

- GLOSTEN, Lawrence R.; MILGROM, Paul R. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. **Journal of Financial Economics**, v. 14, n. 1, p. 71–100, 1985.
- GRAHAM, Benjamin; DODD, David L. **Security Analysis**. [S.l.]: McGraw-Hill, 1934.
- GRAY, Wesley R.; CARLISLE, Tobias E. **Quantitative Value: A Practitioner's Guide to Automating Intelligent Investment and Eliminating Behavioral Errors**. [S.l.]: Wiley, 2012.
- GROSSMAN, Sanford J.; STIGLITZ, Joseph E. On the Impossibility of Informationally Efficient Markets. **American Economic Review**, v. 70, n. 3, p. 393–408, 1980.
- GRUBER, Martin J. Another Puzzle: The Growth in Actively Managed Mutual Funds. **The Journal of Finance**, v. 51, n. 3, p. 783–810, 1996.
- GU, Shihao; KELLY, Bryan; XIU, Dacheng. Empirical Asset Pricing via Machine Learning. **The Review of Financial Studies**, v. 33, n. 5, p. 2223–2273, 2020.
- HARVEY, Campbell R.; LIU, Yan; ZHU, Heqing. and the Cross-Section of Expected Returns. **The Review of Financial Studies**, v. 29, n. 1, p. 5–68, 2016.
- HAYEK, Friedrich A. The Use of Knowledge in Society. **American Economic Review**, v. 35, n. 4, p. 519–530, 1945.
- HELLWIG, Christian; VELDKAMP, Laura. Knowing What Others Know: Coordination Motives in Information Acquisition. **The Review of Economic Studies**, v. 76, n. 1, p. 223–251, 2009.
- HIRSHLEIFER, David; LO, Andrew W.; ZHANG, Ruixun. Social Contagion and the Survival of Diverse Investment Styles. **Journal of Economic Dynamics and Control**, v. 154, p. 104710, 2023.
- JAFFE, Jeffrey F. Special Information and Insider Trading. **The Journal of Business**, v. 47, n. 3, p. 410–428, 1974.
- JEGADEESH, Narasimhan; TITMAN, Sheridan. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. **The Journal of Finance**, v. 48, n. 1, p. 65–91, 1993.
- KUCHAR, Pavel. Competition as socially extended cognition: A Hayekian perspective on the socialist calculation debate. **Constitutional Political Economy**, 2025.
- KYLE, Albert S. Informed Speculation with Imperfect Competition. **The Review of Economic Studies**, v. 56, n. 3, p. 317–355, 1989.

- LEBARON, Blake. Agent-Based Computational Finance. *In*: TESFATSION, Leigh; JUDD, Kenneth L. (Orgs.). **Handbook of Computational Economics**. [S.l.]: Elsevier, 2006. v. 2 p. 1187–1233.
- LO, Andrew W. The Adaptive Markets Hypothesis. **Journal of Portfolio Management**, v. 30, n. 5, p. 15–29, 2004.
- LO, Andrew W.; ZHANG, Ruixun. **The Adaptive Markets Hypothesis: An Evolutionary Approach to Understanding Financial System Dynamics**. [S.l.]: Oxford University Press, 2024.
- MCLEAN, R. David; PONTIFF, Jeffrey. Does Academic Research Destroy Stock Return Predictability? **The Journal of Finance**, v. 71, n. 1, p. 5–32, 2016.
- NOVY-MARX, Robert. The Other Side of Value: The Gross Profitability Premium. **Journal of Financial Economics**, v. 108, n. 1, p. 1–28, 2013.
- PASTUSHKOV, Ivan. An Evolutionary Model of Financial Market Efficiency with Costly Information. **HSE Economic Journal**, v. 28, n. 4, p. 508–534, 2024.
- SOROS, George. **The Alchemy of Finance: Reading the Mind of the Market**. [S.l.]: Simon & Schuster, 1987.
- STIGLITZ, Joseph E.; KOSENKO, Andrew. **The Economics of Information in a World of Disinformation**. [S.l.]: National Bureau of Economic Research, 2024.
- VERRECCHIA, Robert E. Information Acquisition in a Noisy Rational Expectations Economy. **Econometrica**, v. 50, n. 6, p. 1415–1430, 1982.
- WHITE, Halbert. A Reality Check for Data Snooping. **Econometrica**, v. 68, n. 5, p. 1097–1126, 2000.