

Implementação do Modelo Fama e French 5 fatores aplicado ao Mercado de ações brasileiro

FERNANDO G. ZANCHITTA¹

¹Aeronautics Institute of Technology (e-mail: fernando.zanchitta@ga.ita.br)

ABSTRACT O trabalho aplica o Modelo de 5 fatores de Fama e French (2015) sobre o cenário do Mercado de ações brasileiro. Para isso, realiza a avaliação do modelo para analisar a relevância de cada fator, bem como o poder explicativo para o retorno de uma carteira de ações selecionadas no período entre os anos de 2001 e 2022, e que compõem a B3. Ademais, foi realizada o teste da performance de uma estratégia de investimento utilizando de Regressão Linear aplicando os fatores de risco como variáveis independentes. Obteve-se um resultado abaixo da performance histórica do portfólio.

INDEX TERMS Precificação de Ativos; Anomalias de Mercado; Modelo de três e cinco fatores de Fama French; Regressão linear.

I. INTRODUÇÃO

A. CONTEXTO

A validade dos modelos de CAPM (Modelo de Precificação de ativos) proposto por Sharpe (1964) e Linter(1965) aplicados ao cenário de Finanças Moderno vem sofrendo bastante críticas, e vários modelos alternativos de precificação de ativos foram propostos. Esses modelos tentam capturar a relação risco retorno e explicar os dados de *Cross-Section* da média dos retornos das ações. Entretanto, a maioria dos modelos não captura as relações corretas das anomalias nos testes empíricos.

Determinou-se empiricamente, a significância estatística das variáveis: Tamanho (*Market Capitalization*), *Book-to-Market (B/M)*, Investimentos(*Capex*) e Lucro Operacional (*EBITDA*). Banz (1981) encontrou relações do tamanho das empresas que indicam que empresas menores tendem a ter, na média, retornos maiores. Fama e French (1992) determinou o efeito do valor, representado pelo indicador *Book-to-Market*, e estabelece uma relação positiva entre essa variável e o retorno. Titman, Wei and Xie(2004) mostram que empresas que aumentam seus investimento em capital tendem a ter um retorno ajustado ao risco negativo no futuro. Por fim, Novy-Marx(2013) desenvolve que há uma relação positiva entre empresas lucrativas e retornos esperados.

Para determinar se as variáveis acima possuem fatores explicativos necessários, é necessário testar a habilidade de prever retornos ajustados ao risco. Com isso, testes foram feitos de forma a consolidar uma metodologia que gera um

modelo de precificação de ativos baseados em fatores, como o modelo de 3 Fatores (Fama e French, 1993), ao encontrar um intercepto estatisticamente igual a zero para a regressão, usando os Fatores: Risco, Tamanho e Valor. Após 30 anos, o Modelo de 3 fatores tornou-se extremamente consolidado e ganhou muita atenção pela sua capacidade de gerar um retornos acumulados acima do que o tradicional CAPM gerava, para o mercado americano, principalmente no período de 1963-1993.

Outros modelos surgiram com o passar dos anos, na medida que o tradicional FF3F perdia capacidade de explicar retornos e anomalias de Mercado. Assim os autores propuseram a abordagem de 5 Fatores (2014) que inclui as outras duas variáveis mencionadas acima (Investimento e Lucratividade) [2]. O modelo *FF5F* encontrou resultados superiores ao FF3F e soube capturar melhor a variação do retorno dos ativos e anomalias de mercado [2].

B. REVISÃO DA LITERATURA

1) Modelo de Fama French

A partir de evidências experimentais, foi possível explicar que o retorno de um portfólio está correlacionado com as anomalias de tamanho de mercado, o índice *Book-to-Market*, além do prêmio de Mercado. Fama e French (1993) propuseram o modelo de 3 fatores expresso pela equação:

$$r = R_f + \beta_i(R_m - R_f) + b_s \cdot SMB + b_v \cdot HML + \epsilon \quad (1)$$

O fator tamanho *SMB* representa a diferença de retorno de um portfólio de ações com grandes empresas, de alto valor de mercado, em relação ao retorno de um portfólio de ações com pequenas empresas. O fator valor (*HML*) indica a diferença de retorno de um portfólio de ações com alto índice *Book-to-Market* em relação a um portfólio de ações com baixo índice.

Ademais, Jegadeesh e Titman (1993) evidenciaram que estratégias que comprem ações que performam bem e vende ações que performam mal, em uma janela de análise anterior de 3 a 12 meses, gera retornos consistentes (*Momentum*) e que não são devidos aos riscos sistemáticos nem ao atraso de reação de investidores. Cahart (1997) passou a testar o *Momentum* em seus modelos para tentar explicar seus retornos, e percebeu que o 3-Fatores não explicava essa anomalia. Com isso criaram o modelo de 4 Fatores, adicionando o fator *WML* (*Winner minus Loser*) representado na equação:

$$R_i - R_f = \beta_i(R_m - R_f) + b_s \cdot SMB + b_v \cdot HML + w_i \cdot WML + \epsilon \quad (2)$$

O modelo de 5 fatores (2015) direcionou na captura de Tamanho, Valor, Lucratividade e padrões de Investimentos das Firms para o retorno médio das ações. Essa abordagem performou melhor que o *FF3F* (1993). O principal problema encontrado nessa abordagem foi a falha em capturar o retorno médio de ações pequenas cujos retornos se comportam como firmas que investem muito mesmo com baixa Lucratividade. A equação que descreve o modelo:

$$R_i - R_f = \beta_i(R_m - R_f) + s_i \cdot SMB + h_i \cdot HML + r_i \cdot RMW + c_i \cdot CMA + \epsilon \quad (3)$$

O modelo de cinco fatores foi posteriormente aplicado sob o cenário de diferentes mercados internacionais (Fama French 2017) e com isso foi observado que o retorno das ações para as regiões da América do Norte, Europa e Ásia-Pacífico. Observou-se que os retornos aumentaram com o índice *Book-to-Market* e Lucratividade, e foram correlacionados negativamente com relação ao Investimento. Para o Japão, a relação entre retornos e *B/M* foi forte, mas o retorno médio mostra pouca relação com a Lucratividade e Investimento. Os problemas dessas abordagens foram semelhantes aos encontrados para o mercado americano (2015), em relação as ações pequenas (*microcaps*), cujos retornos se comportam como os de empresas que investem agressivamente, ainda que possuam baixa lucratividade.

2) Aplicações no Mercado Brasileiro

Os modelos de Fama French e Cahart vem sido testados na perspectiva do cenário do mercado brasileiro, por grande parte dos estudos no ramo de precificação de ativos. Durante essas análises, o modelo de fatores geralmente é comparando com o CAPM. Nessa perspectiva, Rodrigues (2000) e Málaga e Securato (2004) avaliaram o impacto dos fatores mercado, tamanho e índice book-to-market sobre os retornos de ações listadas na Bolsa de Valores de São Paulo. O desempenho do modelo de três fatores se mostrou superior ao do CAPM

em ambos os estudos. Lucena e Pinto (2007) e Chague(2007) também introduziram trabalhos sobre a validação desse modelo no cenário brasileiro, e seus resultados corroboraram com os anteriores.

No entanto, da mesma forma que é possível encontrar em estudos internacionais, algumas pesquisas ainda possuem certos resultados conflitantes. Roger e Securato (2009), mesmo seu trabalho confirmando o poder preditivo do modelo de 3 fatores sobre o CAPM, eles capturaram que os fatores que maior impulsionam o poder explicativo são o Tamanho e o Beta.

II. OBJETIVO

O problema envolve a avaliação desse modelo no cenário do Mercado de Ações brasileiro sob a perspectiva da detecção do prêmio de risco e da significância estatística de cada uma das anomalias, bem como a capacidade do modelo em descrever os retornos médios dos ativos.

- Aplicar a metodologia de Fatores no Mercado brasileiro, identificando os prêmios de risco e interceptos.
- Validar a Hipótese do funcionamento do modelo.
- Montar carteira simplificada e realizar Backtesting de estratégia de sinais de compra.

III. METODOLOGIA

A. CONSTRUÇÃO DE PORTFÓLIOS E AMOSTRAS

Para calcular os valores dos fatores é preciso seguir uma metodologia robusta que reflita com precisão o cenário do mercado brasileiro em cada um dos anos de análise, a fim de validar corretamente o *Backtesting* gerado. A população objeto de análise abrangeu todas as ações listadas na B3 de janeiro de 2001 até dezembro de 2022. Os critérios utilizados na seleção das amostras foram a disponibilidade de valores:

- Valor de mercado no último dia útil do ano anterior.
- Valor de mercado e patrimônio líquido positivos no último dia útil do ano anterior.
- Ativo total no último dia útil do ano anterior.
- Lucro Operacional no último dia útil do ano anterior.
- cotações mensais consecutivas das ações nos 12 meses anteriores.
- Ações de empresas financeiras foram excluídas das amostras devido às suas estruturas financeiras diferente das demais empresas (como alto grau de endividamento, por exemplo).
- Ações preferenciais e Units foram retiradas, a fim de reduzir a redundância da escolha de ativos nas amostras.

Neste projeto, não foi preciso fazer a construção dos fatores de risco, uma vez que os valores foram adquiridos a partir de um provedor de dados público (NEFIN).

B. CONSTRUÇÃO DE PORTFÓLIOS DE FATORES

O Retorno dos ativos foram ajustados aos proventos (Dividendos e Juros sobre Capital próprio), e foram transformados

Operacionalização das variáveis (características) subjacentes aos portfólios

Variáveis	Fórmula	Descrição
Tamanho (Valor de Mercado)	$VM_{i,t} = P_{i,t} \times N_{i,t}$	$VM_{i,t}$ é o valor de mercado da ação i no tempo t ; $P_{i,t}$ é o preço da ação i no tempo t ; $N_{i,t}$ é o número de ações i em circulação no tempo t .
Índice Book-to-Market	$B/M_{i,t} = \frac{PL_{i,dez(t-1)}}{VM_{i,dez(t-1)}}$	$B/M_{i,t}$ é o índice <i>book-to-market</i> da empresa i , no tempo t ; $PL_{i,dez(t-1)}$ é o valor do patrimônio líquido da empresa i em 31 de dezembro do ano $t-1$, em relação ao ano de formação do portfólio; $VM_{i,dez(t-1)}$ é o valor de mercado da empresa i em 31 de dezembro do ano $t-1$, em relação ao ano de formação do portfólio.
Lucratividade	$Lucr_{i,t} = \frac{Lop_{i,dez(t-1)}}{PL_{i,dez(t-1)}}$	$Lucr_{i,t}$ é a lucratividade da empresa i , no tempo t ; $Lop_{i,dez(t-1)}$ é o valor do lucro operacional da empresa i em 31 de dezembro do ano $t-1$; $PL_{i,dez(t-1)}$ é o valor do patrimônio líquido da empresa i em 31 de dezembro do ano $t-1$.
Investimento	$Inv_{i,t} = \frac{(AT_{i,dez(t-1)} - AT_{i,dez(t-2)})}{AT_{i,dez(t-1)}}$	$Inv_{i,t}$ é o valor do investimento da empresa i , no tempo t ; $AT_{i,dez(t-1)}$ é o valor do ativo total da empresa i em 31 de dezembro do ano $t-1$; $AT_{i,dez(t-2)}$ é o valor do ativo total da empresa i em 31 de dezembro do ano $t-2$, sendo t o ano de formação do portfólio.

Fonte: Adaptado de Fama e French (1993, 2015) e Carhart (1997).

FIGURE 1. Variáveis financeiras calculadas.

em logaritmo neperiano:

$$R_{p,t} = \sum_{i=1}^n \frac{VM_{i,t}}{VM_{p,t}} (R_{i,t}) \quad (4)$$

Em que $R_{p,t}$ é o retorno do portfólio p em t , $VM_{i,t}$ é o valor de mercado da ação i em t , $VM_{p,t}$ é o valor de mercado do portfólio p em t , e $R_{i,t}$ é a taxa de retorno do ativo em t .

A operacionalização das variáveis que permeiam o cálculo dos fatores foi feito seguindo a Figura 1. Para criação dos fatores de risco utilizados na pesquisa (variáveis independentes na regressão), fez-se uso dos procedimentos descritos em Fama e French (1993,2015) e Cahart (1997). Para formação de portfólios, os ativos foram classificados de forma crescente pelo seu valor de mercado. Uma vez estimado o valor de mercado de cada ativo, eles foram divididos pela sua mediana nos grupos (i) *Big (B)* e (ii) *Small (S)*, sendo o primeiro grupo composto pelos ativos acima da mediana, e vice versa.

Após a classificação anterior, também no mesmo mês t , os ativos foram reordenados com base em seus valores de índice *Book-To-Market*, momento, investimento, e lucratividade, formando os grupos *Baixo (low)*, *médio (medium)* e *alto (high)*. Ou seja, os fatores foram obtidos mantendo-se o tamanho como a variável principal, alterando somente a segunda variável de ordenação de modo a formar portfólios com a combinação de duas características das firmas, seguindo-se um critério 2x3.

Para a reordenação das segundas variáveis, foi dividido o conjunto *low*, *medium*, *high* nos percentis 30/40/30. Após esses procedimentos foram obtidos 30 portfólios no total, os quais foram utilizados na construção dos fatores de risco (variáveis independentes) dos modelos de precificação estimados no primeiro passo da metodologia de Fama e MacBeth

(1973). A construção de portfólios a partir dos Fatores de Risco está descrita na Figura 2.

C. ESTIMAÇÃO DOS MODELOS

O modelo de cinco fatores foi estimado conforme a metodologia Fama e MacBeth (1973), que compreende a realização de regressões em dois passos. No primeiro passo os excessos dos retornos dos portfólios (variáveis dependentes) formados com bases nas anomalias mantendo constante o tamanho, variando a segunda característica, foram regredidos sobre os fatores de risco (*MKT*, *SMB*, *HML*, *WML*, *RMW*, *CMA*), seguindo a equação 3.

Esse passo envolveu a estimação dos modelos de séries temporais para obtenção das sensibilidades dos retornos dos portfólios às variações nos retornos dos fatores dos modelos de três, quatro e cinco fatores.

A motivação é verificar a capacidade do modelo na explicação dos retornos médios dos portfólios. Foi feito o teste GRS, verifica se os interceptos dos modelos de regressão de séries temporais são, em conjunto, estatisticamente iguais a zero (hipótese nula), ou diferentes de zero (hipótese alternativa).

No segundo passo, o objetivo foi verificar se existem no mercado acionário brasileiro, prêmios associados aos coeficientes de risco sistêmico estimados no primeiro passo. Ao analisar o impacto desses coeficientes de risco, verifica-se a premiação de risco que representam. Foram estimadas as regressões *cross-section* dos excessos de retorno, descritas na equação 5.

$$r_{t,n} = \gamma_{t,0} + \sum_{k=1}^K \gamma_{t,k} \hat{\beta}_{n,k} + \epsilon_{t,n} \quad (5)$$

Em seguida, as estimativas obtidas na regressão foram analisadas por meio de estatísticas descritivas.

Operacionalização dos fatores RHS dos modelos de três, quatro e cinco fatores

Fatores RHS	Fórmulas	Modelos
Fator SMB. Diferença entre os retornos médios dos portfólios <i>small</i> e <i>big</i> .	$SMB = \bar{R}_S - \bar{R}_B$	3 fatores
	$\bar{R}_S = \frac{SL_t + SM_t + SH_t}{3}, \bar{R}_B = \frac{BL_t + BM_t + BH_t}{3}$	4 fatores
	$SMB = \bar{R}_S - \bar{R}_B$ $\bar{R}_S = \frac{SL_t + SN_{BMt} + SH_t + SW_t + SN_{LUCt} + SR_t + SC_t + SN_{INVt} + SA_t}{9}$ $\bar{R}_B = \frac{BL_t + BN_{BMt} + BH_t + BW_t + BN_{LUCt} + BR_t + BC_t + BN_{INVt} + BA_t}{9}$	5 fatores
Fator HML. Diferença entre os retornos médios dos portfólios <i>high</i> e <i>low book-to-market</i> .	$HML_t = \bar{R}_H - \bar{R}_L$ $\bar{R}_H = \frac{SH_t + BH_t}{2}, \bar{R}_L = \frac{SL_t + BL_t}{2}$	3, 4 e 5 fatores
Fator MKT – Prêmio de Mercado. Diferença entre os retornos mensais médios, ponderados pelo valor de mercado mensal, de todas as ações da amostra analisada em cada ano do período amostral e o retorno mensal do Certificado de Depósito Interbancário – CDI.	$MKT = R_{carteira\ de\ mercado,t} - R_{CDI,t}$ $R_{carteira\ de\ mercado,t} = proxy\ para\ a\ carteira\ de\ mercado.$ $R_{CDI,t} = proxy\ para\ o\ retorno\ do\ ativo\ livre\ de\ risco\ (CDI).$	3, 4 e 5 fatores
Fator WML – momento. Diferença entre os retornos médios dos portfólios de ações vencedoras (<i>winners</i>) e os retornos médios dos portfólios de ações perdedoras (<i>losers</i>). As ações vencedoras são aquelas com os maiores retornos acumulados nos últimos onze meses; e as perdedoras, os menores retornos.	$WML_t = \bar{R}_W - \bar{R}_L$ $\bar{R}_W = \frac{SW_t + BW_t}{2}, \bar{R}_L = \frac{SL_t + BL_t}{2}$	4 fatores
Fator RMW – lucratividade. Diferença entre os retornos médios dos portfólios de ações com desempenho operacional robusto e os com desempenho operacional fraco.	$RMW_t = \bar{R}_R - \bar{R}_W$ $\bar{R}_R = \frac{SR_t + BR_t}{2}, \bar{R}_W = \frac{SW_t + BW_t}{2}$	5 fatores
Fator CMA – investimento. Diferença entre os retornos médios dos portfólios de empresas que realizaram investimentos conservadores e os das empresas que realizaram investimentos agressivos	$CMA_t = \bar{R}_C - \bar{R}_A$ $\bar{R}_C = \frac{SC_t + BC_t}{2}, \bar{R}_A = \frac{SA_t + BA_t}{2}$	5 fatores

Fonte: Elaborada pelos autores com base em Fama e French (1993, 2015) e Carhart (1997).

FIGURE 2. Cálculo dos portfólios dos Fatores.

D. TESTE DE COLINEARIDADE

O teste de colinearidade é realizado durante a análise de regressão para identificar a presença de alta correlação linear entre as variáveis independentes (também conhecidas como variáveis de preditoras). A colinearidade ocorre quando duas ou mais variáveis independentes estão altamente correlacionadas entre si, o que pode afetar a qualidade e a interpretação dos resultados da regressão.

A presença de colinearidade pode levar a problemas, como coeficientes de regressão imprecisos, estimativas pouco confiáveis dos efeitos das variáveis independentes e dificuldade em interpretar corretamente o papel de cada variável na explicação da variabilidade da variável dependente.

Um método para identificar fatores redundantes entre si é utilizar a regressão de cada fator em relação aos demais como

$$f_{t,k} = a_k + \sum_{j \neq k} \delta_{k,j} f_{t,j} + \epsilon_{t,k}. \quad (6)$$

E. MODELO PREDITIVO DE SINAIS DE COMPRA

Foi desenvolvido um modelo de Regressão Linear com os atributos sendo somente o valor dos fatores a cada dia e o preço de fechamento do portfólio. A variável alvo predita é categórica, representando dois estados: 0 (Comprado), 1 (Não Comprado).

Para simplificar o problema foram adotadas as seguintes medidas:

- Tamanho da compra fixo na variável *ShareSize*.
- Modelo *Long Only*.
- Portfólios compostos somente por um ativo.
- Capital inicial é fixo.

Os resultados desse modelo foram submetidos a um processo de *Backtesting* para avaliar sua performance em comparação a cotação histórica da ação (sempre comprada).

F. AQUISIÇÃO DE DADOS

Os dados dos fatores foram adquiridos pelo NEFIN (*Center for Research in Financial Economics of the University of São*

Paulo) e são disponibilizados publicamente desde o primeiro dia útil de 2001 até o último dia útil de 2022.

Já os valores do índice e das ações do *IBOV* foram adquirida a partir do provedor privado *Comdinheiro*. A aquisição desses dados foi feita a partir da *API* do provedor.

G. PRÉ-PROCESSAMENTO DOS DADOS

O pré processamento dos dados foi feito utilizando múltiplas requisições da *API* do *Comdinheiro* em conjunto com a manipulação dos dados usando *Pandas*. A Figura 3 representa o esquema empregado para manipulação e aquisição dos dados.

Inicialmente, foi feita a busca pela *API* para obter o conjunto histórico de ações que compõem o índice *Ibovespa* dentro da janela de tempo de 2001 até 2022. Com essa lista de ações, buscou-se individualmente para cada ativo os dados diários referente as cotações (*OPHLC*).

Em paralelo, foram estabelecidas as informações e critérios para buscar nessas ações a fim de que sejam calculados os Fatores para delimitação das carteiras de cada fator. O conjunto das ações com dados mensais dos filtros de seleção é chamado de *Universo*.

A partir disso, é possível calcular cada uma das variáveis financeiras, e posteriormente cada uma das carteiras de fatores seguindo a metodologia das seções anteriores. Foi gerado a carteira *Momentum*, *Valor*, *Tamanho*. A organização dos arquivos do projeto pode ser descrita da seguinte forma:

- As funções de manipulação de portfólio foram escritas no arquivo *portfolio.py*.
- Os métodos de avaliação do modelo foram escritas no arquivo *evaluation.py*.
- A análise estatística dos Fatores no mercado brasileiro foi feita no *statsfactor.ipynb*.
- O modelo preditivo dos portfólios foi feito no notebook *projeto.ipynb*.
- A análise exploratória foi realizada no arquivo *exploratory.ipynb*.

IV. RESULTADOS E DISCUSSÕES

A. AVALIAÇÃO DO MODELO DE FATORES NO MERCADO BRASILEIRO

Nessa etapa, criou-se o objeto *Portfolio*, nele contendo algumas seleções de portfólios baseados nos critérios do *Factor Investing*. Dentre eles, foram criadas as carteiras *Momento*, *Tamanho* e *Valor*. Os resultados dos retornos médios anuais da carteira *Valor* podem ser observados na figura 4.

A avaliação do modelo de 5 fatores no mercado brasileiro foi feita seguindo a metodologia de Estimção de modelos proposta por Fama e MacBeth (1973) [14].

Aplicando a primeira regressão, obteve-se os resultados da tabela 1.

O valor médio do intercepto foi de $6.523e - 11$, ou seja, muito próximo de zero. Isso sugere que, na ausência das variáveis independentes, a variável dependente não tem um valor médio significativo. Em outras palavras, a contribuição das variáveis independentes é muito importante para explicar

TABLE 1. Resultado dos Betas da Primeira Regressão aplicada a cada ativo.

Ação	Intercepto	MKT	SMB	HML	WML	IML
BBAS3	2.985e-13	-3.689e-11	2.723e-10	6.970e-11	6.100e-11	-2.705e-10
BBDC4	9.106e-13	7.366e-11	-7.278e-11	-3.734e-12	1.575e-10	-1.049e-11
BRAP4	3.348e-10	8.569e-9	-3.556e-8	4.776e-8	4.395e-8	-5.269e-9
CMIG4	6.176e-11	-3.400e-10	-6.426e-9	3.907e-9	1.725e-10	-5.409e-10
CSNA3	4.777e-11	-2.327e-9	4.008e-9	5.657e-9	5.235e-9	-3.546e-9

TABLE 2. Resultado do teste de Colinearidade entre as variáveis independentes/.

	Intercepto	MKT	HML	IML	SMB	WML
MKT	0.0	nan	0.306(**)	-0.99	0.447(**)	-0.127
HML	0.0 (*)	0.114 (**)	nan	0.116 (**)	0.121 (**)	-0.028
IML	0.0	-0.165	0.051 (**)	nan	0.739 (**)	0.033 (**)
SMB	-0.0	0.079 (**)	0.057 (**)	0.782 (**)	nan	-0.142
WML	0.001 (**)	-0.069	-0.04	0.109 (**)	-0.437	nan

e prever o valor da variável dependente, enquanto o efeito isolado do intercepto é mínimo.

1) Segunda Regressão

A verificação da segunda regressão *cross-section* foi feita de modo a gerar os coeficientes de risco: $\gamma_{t,k}$. Os resultados estão apresentados na figura 5

Em resumo, a análise dos gammas permite avaliar a exposição dos portfólios aos fatores de risco específicos e determinar se esses fatores têm um impacto significativo nos retornos dos ativos, além de verificar a existência de premiação de risco associada aos coeficientes de risco estimados, e os investidores podem ser recompensados por assumir esses riscos adicionais.

O resultado dos coeficientes de risco não permite observar um impacto constante de um fator no retorno dos ativos, evidenciando oscilações ao longo dos anos.

2) Teste de Colinearidade

Os resultados do teste de colinearidade entre os fatores gerou os resultados da tabela 2 Nós formatamos essas figuras juntamente com os limiares de p-valor e as exportamos em uma tabela de resumo. Os níveis de significância dos coeficientes são codificados da seguinte forma:

$$0 < (***) < 0.001 < (**) < 0.01 < (*) < 0.05$$

Nesse sentido é possível observar certa colinearidade do fator *WML* em relação aos demais, de forma que pode ser redundante para algumas classes de ações.

B. MODELO PREDITIVO APLICADO A PERFORMANCE DE CARTEIRA

A Regressão linear foi aplicada numa carteira contendo somente uma ação foi empregada em diversos ativos. Para exemplificar, foram gerados resultados envolvendo o *ITUB3*.

1) Teste ANOVA

O teste ANOVA (análise de variância) do modelo aplicado ao *ITUB3* gerou o resultado da figura 6

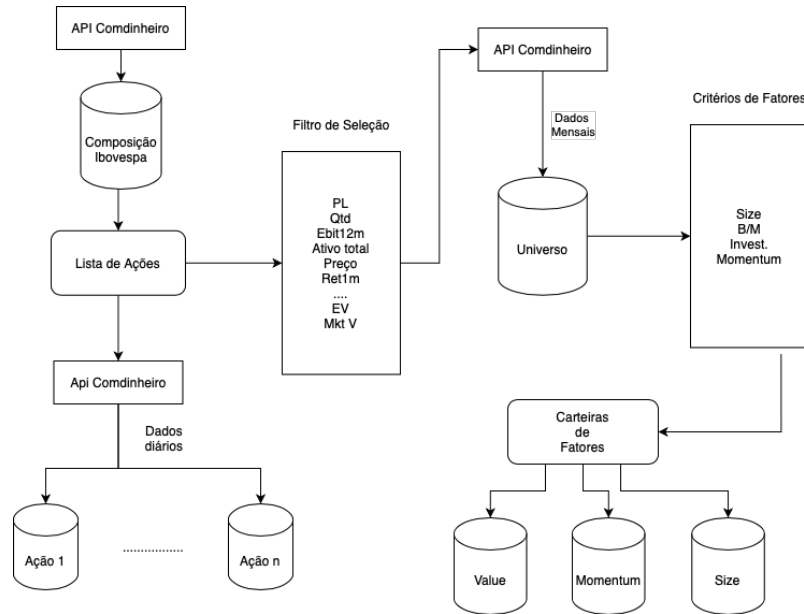


FIGURE 3. Aquisição dos dados da Comdinheiro.

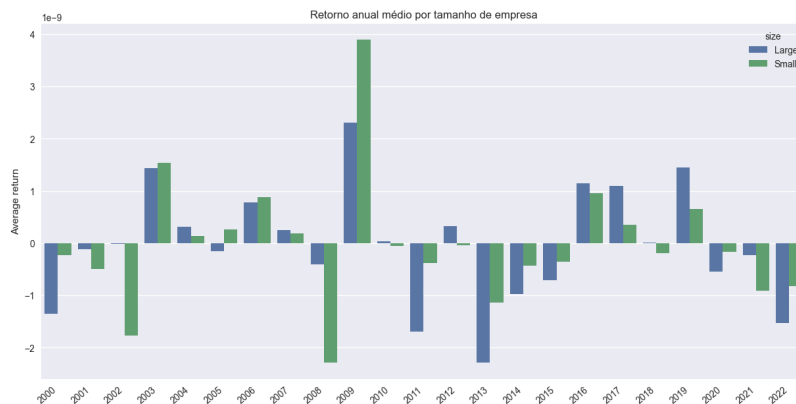


FIGURE 4. Exemplo de retorno médio anual do *Small e Big*.

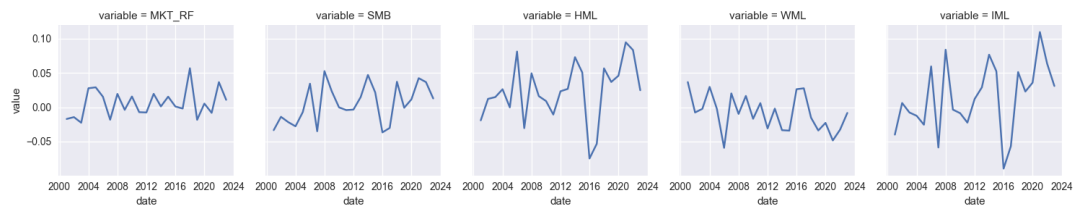


FIGURE 5. Os coeficientes de risco associados a cada um dos fatores ao longo do tempo.

O coeficiente de determinação R^2 de 0.634 em um modelo de regressão indica que aproximadamente 63,4% da variabilidade dos dados de saída pode ser explicada pelas variáveis independentes incluídas no modelo. Em outras palavras, o R^2 é uma medida estatística que avalia o quão bem o modelo se ajusta aos dados.

2) Regressão Parcial

O teste de regressão parcial é uma técnica estatística utilizada para analisar a relação entre uma variável independente específica e uma variável dependente, controlando o efeito das demais variáveis independentes incluídas no modelo. Ele é frequentemente usado em modelos de regressão múltipla para

OLS Regression Results						
=====						
Dep. Variable:	Returns	R-squared:	0.634			
Model:	OLS	Adj. R-squared:	0.633			
Method:	Least Squares	F-statistic:	362.0			
Date:	Mon, 26 Jun 2023	Prob (F-statistic):	5.73e-225			
Time:	21:21:39	Log-Likelihood:	-1618.8			
No. Observations:	1049	AIC:	3250.			
Df Residuals:	1043	BIC:	3279.			
Df Model:	5					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	0.0237	0.035	0.671	0.502	-0.046	0.093
hkt-rf	91.4844	2.365	38.685	0.000	86.844	96.125
MLL	35.2877	4.632	7.617	0.000	26.198	44.378
IML	12.7603	9.123	1.399	0.162	-5.142	30.662
SMB	-30.9075	8.023	-3.853	0.000	-46.650	-15.165
WML	-16.7562	3.215	-5.212	0.000	-23.064	-10.448
=====						
Omnibus:	71.180	Durbin-Watson:	2.121			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	185.645			
Skew:	0.355	Prob(JB):	4.87e-41			
Kurtosis:	4.935	Cond. No.	340.			

FIGURE 6. Resultados do Teste ANOVA no portfólio contendo ITUB3.

TABLE 3. Resultado do Backtesting da estratégia de investimentos aplicada ao ITUB3.

Análise	Estratégia	Ativo
Annual Return	-3.4%	6.5%
Cumulative Returns	-15%	9.1%
Annual Volatility	9.6%	29%
Sharpe Ratio	-0.36	0.21
Sortino Ratio	-0.50	0.32

entender o impacto individual de cada variável independente no resultado. A regressão parcial do modelo aplicado a ITUB3 gerou a figura 7

O coeficiente de regressão parcial indica a magnitude e a direção do impacto da variável independente em questão na variável dependente, controlando o efeito das demais variáveis independentes. Pode-se observar um efeito maior do fator *MKT* na variável alvo. Os demais fatores influenciaram pouco os retornos da ação.

C. PERFORMANCE HISTÓRICA

Aplicando o *Backtesting* no conjunto de teste do modelo, que atua a partir do ano de 2018 até o final de 2022, gera-se os retornos acumulados apresentados na figura 8. Os resultados dessa estratégia podem ser descritos na tabela 3, com retornos anualizados menores e volatilidade menores que a performance histórica do papel (considerando estratégia sempre comprada desde o início).

É possível observar que os retornos acumulados da estratégia foram piores que os retornos do ativo no mesmo período. Isso se deve à dificuldade de se montar um modelo preditivo para esse tipo de atividade (predição de movimento das ações).

Em definição, o problema das previsões que se baseiam em dados passados para inferir padrões sobre flutuações subseqüentes. A esperança mais ou menos explícita de qualquer previsor é que o passado se revele uma boa aproximação do futuro. Por conta disso, em geral, as previsões têm re-

sultados ruins. Surpreendentemente, isso não depende muito da sofisticação da ferramenta econométrica. Na verdade, suposições heurísticas muitas vezes são difíceis de superar.

V. CONCLUSÃO

O estudo apresentado visou a análise e a aplicação de estratégias de investimento baseadas em fatores no mercado de ações brasileiro, abrangendo o período de 2001 a 2022. Utilizamos uma ampla gama de métodos, incluindo manipulação de dados, análise estatística, e modelagem preditiva.

Os dados utilizados foram adquiridos através do NEFIN e do provedor privado Comdinheiro. Posteriormente, foram pré-processados e analisados, levando à construção de carteiras baseadas nos fatores de Momentum, Valor e Tamanho.

A avaliação do modelo de fatores no mercado brasileiro seguiu a metodologia de Estimação de modelos proposta por Fama e MacBeth (1973). Descobrimos que a contribuição das variáveis independentes é muito importante para explicar e prever o valor da variável dependente, enquanto o efeito isolado do intercepto é mínimo.

A segunda regressão permitiu avaliar a exposição dos portfólios aos fatores de risco específicos e determinar se esses fatores têm um impacto significativo nos retornos dos ativos. Os resultados indicaram oscilações ao longo dos anos, não permitindo observar um impacto constante de um fator no retorno dos ativos.

Um teste de colinearidade revelou a possibilidade de redundância do fator WML para algumas classes de ações.

O estudo também explorou a aplicação de um modelo preditivo a performances de carteira individual, no exemplo específico da ITUB3. Apesar da análise de variância indicar que aproximadamente 63,4% da variabilidade dos dados de saída pode ser explicada pelas variáveis independentes, os resultados não superaram os retornos reais do ativo no período analisado. Isso ressalta a complexidade e a incerteza inerentes à previsão de movimentos de ações.

Embora a previsão baseada em dados passados possa ser vista como uma boa aproximação do futuro, este estudo demonstra que a realidade é muitas vezes mais complexa. Ainda assim, a análise de fatores e a modelagem preditiva podem oferecer insights valiosos para a compreensão das tendências e comportamentos do mercado, o que pode ser útil para a formulação de estratégias de investimento.

VI. PRÓXIMOS PASSOS

Para verificar se os retornos acumulados da estratégia de investimento apresentam melhora, mudanças na separação dos conjuntos de dados de treino e teste utilizando outras metodologias de séries temporais e validação cruzada podem ser efetivos. Além disso, pode ser relevante aumentar o número de atributos no modelo de aprendizado, com a adição de indicadores fundamentalistas. Além disso, Propõem-se o teste da estratégia de investimento aplicado a uma carteira composta por mais de um ativo.

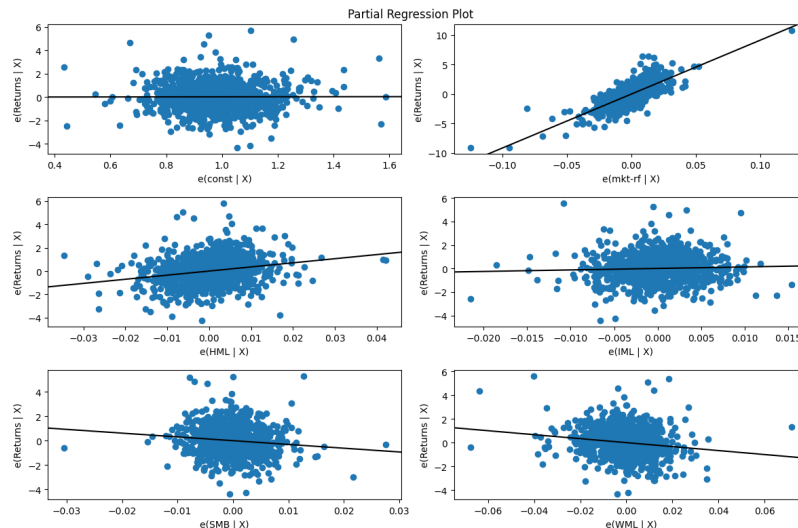


FIGURE 7. Resultados da Regressão parcial aplicada no ITUB3 .

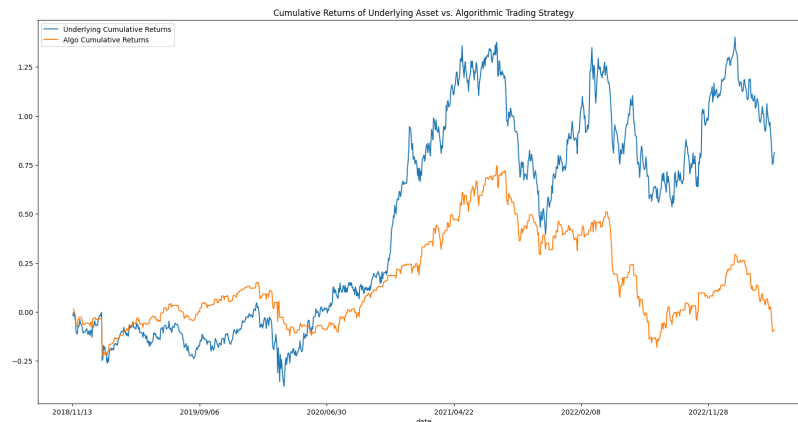


FIGURE 8. Performance Acumulada do modelo em ITUB3 versus sua cotação histórica no período.

REFERENCES

- [1] Jegadeesh, N., Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48(1), 65-91. <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>
- [2] Fama, Eugene F. and French, Kenneth R., A Five-Factor Asset Pricing Model (September 2014). Fama-Miller Working Paper, Available at SSRN: <https://ssrn.com/abstract=2287202> or <http://dx.doi.org/10.2139/ssrn.2287202>
- [3] Fama, E. F., and French, K. R. (1996). Multifactor explanation of asset pricing anomalies. *Journal of Finance*, 56(1), 55-84. <https://doi.org/10.1111/j.1540-6261.1996.tb05202.x>
- [4] Fama, E. F., and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- [5] Fama, E. F., and French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, 123(3), 441-463. <https://doi.org/10.1016/j.jfineco.2016.11.004>
- [6] Rodrigues, M. R. A. (2000). O efeito valor, o efeito tamanho e o modelo multifatorial: evidências do caso brasileiro. *Anais do ENANPAD*, Florianópolis, Santa Catarina, Brasil, 24.
- [7] Málaga, F. K., and Securato, J. R. (2004). Aplicação do modelo de três fatores de Fama e French no mercado acionário brasileiro – um estudo empírico do período 1995-2003. *Anais do ENANPAD*, Curitiba, PR, Brasil, 28.
- [8] Leonardo, T. B. S., MODELO DE CINCO FATORES FAMA-FRENCH: TESTE NO MERCADO BRASILEIRO. 2019
- [9] Clarice, C. M. Pricing Assets with Fama and French 5-Factor Model: a Brazilian market novelty. 2015.
- [10] Diallo, Boubacar and Bagudu, Aliyu and Zhang, Qi, A Machine Learning Approach to the Fama-French Three- and Five-Factor Models (August 21, 2019). Available at SSRN: <https://ssrn.com/abstract=3440840> or <http://dx.doi.org/10.2139/ssrn.3440840>
- [11] Bui Thanh Khoa, Tran Trong Huynh, "Long Short-Term Memory Recurrent Neural Network for Predicting the Return of Rate Underframe the Fama-French 5 Factor", *Discrete Dynamics in Nature and Society*, vol. 2022, Article ID 3936122, 8 pages, 2022. <https://doi.org/10.1155/2022/3936122>
- [12] Banz, R. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3-18. [https://doi.org/10.1016/0304-405X\(81\)90018-0](https://doi.org/10.1016/0304-405X(81)90018-0)
- [13] Claudia, F. M., Laíse, F. C (2021). Performance of the Fama-French five-factor model in the pricing of anomalies in the Brazilian market. <https://doi.org/10.5007/2175-8069.2021.e78962>
- [14] Fama, E. F., and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), 607-636.
- [15] Titman, S., Wei, K. J., and Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, 39(4), 677-700.

- [16] Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of financial economics*, 108(1), 1-28.

...