

Revisitando Fama-French sob a perspectiva de Aprendizado de Máquina

Palavras-Chave: Modelos de Fatores, Aprendizado de Máquina, Finanças Quantitativas

Autores(as):

NATALIA THEREZINHA NEGRI, IMECC – UNICAMP

Prof. Dr. CARLOS TRUCÍOS (orientador), IMECC – UNICAMP

INTRODUÇÃO:

Os modelos de fatores estão presentes em economia e finanças há mais de 50 anos, mas, com a popularização e interesse por ETFs (Exchange Traded Funds) nos últimos anos, a pesquisa teórica e aplicada neste assunto tem se intensificado, existindo uma diversa gama de fatores, conhecidos como "factor zoo", bem como metodologias para relacionar estes fatores aos ativos financeiros de interesse.

Entre os modelos mais influentes para previsão de ativos estão os modelos de 3 e 5 fatores Fama-French, que têm sido amplamente utilizados e testados ao longo das últimas décadas. Esses modelos expandem a teoria do Capital Asset Pricing Model (CAPM), incorporando novos fatores ao tradicional fator de mercado ($R_m - R_f$), incluindo tamanho (SMB - Small Minus Big), valor (HML - High Minus Low), rentabilidade (RMW - Robust Minus Weak) e investimento (CMA - Conservative Minus Aggressive).

Assim, os modelos de três e cinco fatores são representados, respectivamente, pelas seguintes equações:

1. Modelo de três fatores

$$R_i - R_f = \alpha + \beta_1(R_m - R_f) + \beta_2SMB + \beta_3HML + \epsilon$$

2. Modelo de cinco fatores:

$$R_i - R_f = \alpha + \beta_1(R_m - R_f) + \beta_2SMB + \beta_3HML + \beta_4RMW + \beta_5CMA + \epsilon$$

Nessas equações, R_i representa o retorno do ativo, R_f é a taxa livre de risco, R_m é o retorno do mercado, α é o termo de intercepto e ϵ é o termo de erro aleatório.

O fator SMB captura a diferença nos retornos entre empresas de pequeno e grande porte, enquanto o fator HML mede a diferença nos retornos entre empresas com alto e baixo valor de mercado. No modelo de cinco fatores, o fator RMW captura a diferença entre os retornos de empresas com alta (robusta) e baixa (fraca) lucratividade operacional, enquanto o fator CMA mede a diferença nos retornos entre empresas que investem de forma conservadora e agressiva.

Tradicionalmente, utilizam-se o método de regressão linear múltipla para estimar os coeficientes e prever os retornos nos modelos de Fama-French. No entanto, com o avanço das técnicas de aprendizado de máquina, surgiu a oportunidade de revisitar esses modelos sob uma nova perspectiva. Neste trabalho, exploramos a aplicação de técnicas de aprendizado de máquina supervisionado aos modelos de fatores Fama-French, com o objetivo de investigar se essas novas abordagens podem trazer ganhos significativos na previsão de retornos acionários.

METODOLOGIA:

Para esta análise, foram considerados dados de retornos mensais de 451 ações da bolsa de valores dos EUA, obtidos da Economática, assim como os fatores de Fama-French, retirados do site de Kenneth French (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library). O período de estudo compreende de janeiro de 2000 a junho de 2023, totalizando 281 observações por ação.

As técnicas avaliadas incluem Regressão Linear Múltipla (MLR), Regressão Penalizada (Lasso, Ridge e ElasticNet), Support Vector Regression (SVR), Árvore de Regressão Simples (DT) e Regressão de Floresta Aleatória (RF). Esses métodos foram aplicados considerando três e cinco fatores de Fama-French. Para otimizar a escolha dos hiperparâmetros, utilizamos o GridSearchCV nas Regressões Penalizadas e o RandomSearchCV exclusivamente para o SVR, ambos com validação cruzada. Para cada previsão, o modelo foi treinado utilizando uma janela de 120 observações, através da estratégia de Rolling Window forecast, método amplamente utilizado neste tipo de estudos.

O desempenho dos métodos foi avaliado utilizando a Raiz do Erro Quadrático Médio (RMSE), que mede a raiz da média das diferenças quadráticas entre os valores previstos e reais. Uma das vantagens do RMSE é sua fácil interpretação, bem como o fato de ser amplamente utilizada por outros autores em contextos similares.

RESULTADOS E DISCUSSÃO:

Para comparação, foi feita o RMSE de cada ação, utilizando 159 previsões, as tabelas a seguir fazem a contagem de ações que cada método teve o menor RMSE em número absoluto em comparação com os outros.

Tabela 1: Contagem de vezes que cada método foi o melhor

Método	Contagem
Ridge com 5 fatores	156
ElasticNet com 3 fatores	90
ElasticNet com 5 fatores	74
Ridge com 3 fatores	60
SVR com 3 fatores	35
SVR com 5 fatores	28
Lasso com 3 fatores	4
MLR com 3 fatores	3
MLR com 5 fatores	1
Outros Métodos	0

Pela tabela 1 nota-se que os métodos de Ridge, ElasticNet e SVR tiveram desempenhos melhores que a regressão linear múltipla para qualquer número de fatores. Observa-se que o método de regressão penalizada Ridge com 5 fatores apresentou menor RMSE em 156 de 451 ações, em comparação com os outros métodos. O ElasticNet com 3 fatores também se destacou, sendo o melhor em 90 ações. Entretanto, os métodos de árvores de regressão não tiveram o menor erro em nenhuma ação.

Os resultados da tabela 2 indicam que os modelos com 3 fatores são, em número absoluto, melhores que os modelos de 5 fatores em 4 de 7 métodos, especialmente em técnicas como MLR e Lasso, onde a diferença é maior. No entanto, métodos como Ridge e RF apresentaram melhor desempenho com 5 fatores, o que sugere que a inclusão de mais fatores pode ser benéfica para esses modelos.

Tabela 2: Contagem de vezes que cada modelo de Fama French foi o melhor

Métodos	Com 5f	Com 3f
MLR 5f vs MLR 3f	44	407
Lasso 5f vs Lasso 3f	24	427
Elastic 5f vs Elastic 3f	204	247
Rid 5f vs Rid 3f	310	141
SVR 5f vs SVR 3f	241	210
DT 5f vs DT 3f	212	239
RF 5f vs RF 3f	346	105

Tabela 3: Comparação entre os métodos com o modelo de 5 fatores (Contagem 5f) e entre os métodos com 3 fatores (Contagem 3f), contando quantas vezes cada um foi o melhor

Método	Contagem 5f	Contagem 3f
Ridge	254	206
ElasticNet	159	190
SVR	37	47
MLR	1	3
Lasso	0	5
DT	0	0
RF	0	0

Na tabela 3 separa-se a comparação dos métodos por modelo, sendo a primeira coluna a contagem considerando apenas os modelos de 5 fatores e a segunda coluna apenas considerando os modelos de 3 fatores, vemos que os métodos Ridge e ElasticNet são melhores que o método MLR em ambos modelos. Especificamente no modelo 5f, apenas em uma ação o método de regressão teve menor RMSE, em comparação, em 254 ações o método Ridge foi melhor, já com 3 fatores, a regressão foi melhor em três ações e o Ridge em 206.

Em relação ao método de regressão com penalização Ridge com 5 fatores, como exemplificação, o maior RMSE de um ativo encontrado foi de 33,052 e o menor foi de 4,307, considerados altos mas, em geral, apresenta os menores erros.

CONCLUSÕES:

Avaliando os valores de RMSE de cada método, conclui-se que o método de regressão com penalização Ridge utilizando o modelo de 5 fatores, apresentou melhor desempenho em 156 de 451 ações, número superior aos outros métodos. Além disso, para 4 dos 7 métodos, o modelo com 3 fatores aparenta prever melhor que o modelo com 5 fatores em número absoluto. Enquanto isso, para ambos os modelos de Fama French, a penalização Ridge teve melhor desempenho. Concluído assim, que as técnicas de aprendizado supervisionada como Ridge, ElasticNet e SVR tiveram, em número absoluto, desempenho melhores que a Regressão Linear Múltipla (MLR).

BIBLIOGRAFIA

- DIALLO, B.; BAGUDU, A.; ZHANG, Q. **Fama–French Three versus Five: Which Model is Better? A Machine Learning Approach**. Journal of Forecasting, p. 1–15, 2023.
- COQUERET, G.; GUIDA, T. **Machine Learning for Factor Investing: Python Version**. 1. ed. Boca Raton: CRC Press, 2023.