Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 8, n. 1, 2021.

Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Portfólios baseados em Clusterização Espectral e *Factor Investing*

Lucas Siviero Sibemberg[1]
UFRGS, Porto Alegre, RS
Luiz Emilio Allem[2]
UFRGS, Porto Alegre, RS
Carlos Hoppen[3]
UFRGS, Porto Alegre, RS

O número de investidores em bolsas de valores nunca foi tão grande. Em contrapartida, novos entrantes acabam tendo menos conhecimento sobre o mercado para montar um portfólio que apresente um bom desempenho. Dessa forma estratégias que tomem pouco tempo do investidor, como por exemplo, delegar para algum fundo de investimentos a seleção de ativos, começam a ficar mais populares. Nesse trabalho propomos uma abordagem combinando clusterização espectral e *factor investing* para criação de um portfólio que pode ser feita uma vez no início de cada ano. A seguir apresentamos o procedimento de escolha de ativos que compõem o portfólio. Observamos que todos os dados utilizados ao longo desse trabalho foram retirados do *Yahoo Finance* e podem ser acessados em https://github.com/Lucassib/Gestao_de_Portfolio.git.

Assim, seja $\mathcal{X} = \{x_1, \dots, x_n\}$ o conjunto de ativos que são as opções para a seleção do portfólio, $k \in \mathbb{N}$ fixo e P denota o portfólio escolhido.

1. Divide os ativos em k classes, isto é, $\{A_i \neq \emptyset\}_{i=1}^{k}$, $A_i \cap A_j = \emptyset$, se $i \neq j$ e $\cup_{i=1}^{k} A_k = \mathcal{X}$.

2. Define $P = \{x : y(x) = \min_j(y(x_j)), x_j \in A_i, \text{ para } i \in 1, \dots, k\}$, onde y é uma métrica.

Para a divisão dos ativos em k classes utilizamos uma heurística espectral muito famosa, introduzida por [2]. Nessa abordagem associamos um conjunto de dados $\mathcal{X} = \{x_1, \dots, x_n\}$ com uma matriz $A = (a_{ij})$ de similaridade, onde a_{ij} indica a similaridade entre os dados x_i e x_j. Na heurística dividimos o conjunto de dados em k classes a partir de autovetores associados a uma matriz obtida com base em A. Vale ressaltar que o objetivo é que dados que pertençam a uma mesma classe sejam mais similares.

Para a montagem do portfólio, definimos $A = A^m = (a_{ij}^m)$, em que $a_{ij}^m = \frac{1+w_{ij}^m}{2}$, onde w_{ij}^m é a correlação entre os ativos i e j relacionada ao desempenho diário no ano m. Note que quanto maior é a_{ij}^m maior foi a correlação dos ativos i e j e, portanto, maior a chance de eles estarem em uma mesma classe e, por fim, dificilmente estarão no mesmo portfólio construído pela nossa estratégia.

A partir da saída do algoritmo de clusterização espectral há k classes distintas com ativos de $\mathcal{X}$, onde $\mathcal{X}$ é o conjunto de ações negociadas na Bovespa. A próxima etapa foi escolher um ativo de cada classe para compor o portfólio. Existem critérios bem estabelecidos entre economistas para fazer essa seleção, conhecidos como critérios de *factor investing* [1].

O investimento em fatores (*factor investing*) é muito discutido no mundo dos investimentos. Um fator é uma característica relacionada a alguma métrica que é importante para explicar riscos e retornos dos ativo [1] e já há estratégias famosas baseadas em fatores. Nesse trabalho utilizamos

[1] lucas.siviero@ufrgs.br
[2] emilio.allem@ufrgs.br
[3] choppen@ufrgs.br

a métrica $y_1 = \frac{P}{L}$ se $L > 0$ e $y_1 = 0$ caso contrário, onde P é o preço da ação no último dia considerado e L é o lucro acumulado por ação da empresa nos últimos doze meses. Também utilizamos a métrica $y_2 = \frac{\text{PL}}{L_t}$ se $L > 0$ e 0 caso contrário, onde PL é o patrimônio líquido da empresa e L_t seu lucro acumulado nos últimos doze meses. Note que podemos entender $y_1 > 0$ como a quantidade de anos necessários para a empresa retornar ao acionista o valor investido em lucro líquido, assim, quanto menor é y_1, mais rápido é o retorno ao acionista. Já y_2 é o patrimônio líquido por cada real lucrado, portanto é razoável que quanto menor, mais eficiente é a empresa.

Nós utilizamos $m \in \{2017, 2018, 2019\}$ e escolhemos $k = 10$. Esse é um valor comumente utilizado em estratégias de *factor investing*. Seja $r_i(m)$ a rentabilidade de x_i no ano m. Para cada m, denotamos por $P_1^m = \{x_{j_k}\}_{k=1}^{10}$ é o portfólio gerado pelo algoritmo utilizando $y = y_1$ e $P_2^m = \{x_{j_i}\}_{i=1}^{10}$ é o portfólio gerado pelo algoritmo utilizando $y = y_2$. Definimos $C_1^m = \frac{1}{10}\sum_{k=1}^{10} r_{j_k}(m+1)$ e $C_2^m = \frac{1}{10}\sum_{i=1}^{10} r_{j_i}(m+1)$, como as rentabilidades de P_1^m e P_2^m, respectivamente. Para comparar o desempenho dos portfólios montamos outras quatro carteiras de referência. Baseado no conjunto de ativos dentro do BOVESPA definimos C_3^m (o famoso Índice Bovespa), onde $C_3^m = \sum_{i=1}^{n} p_i(m) r_i(m+1)$, onde $\sum_{i=1}^{n} p_i(m) = 1$, e cada $p_i(m)$ já está pré-definido, e $C_4^m = \frac{1}{n}\sum_{i=1}^{n} r_i(m+1)$. Os últimos dois portfólios P_5^m e P_6^m de *factor investing*, são montados escolhendo os 10 ativos que possuem menor y_1 e y_2 do BOVESPA, respectivamente, onde novamente sua rentabilidade é calculada pela médias das rentabilidades dos ativos de P_5^m, P_6^m, respectivamente. Por fim, observamos que quando olhamos para um período maior, de a anos, a cada ano o portfólio é revisto de acordo com o algoritmo e sua rentabilidade deve ser medida a partir de $C_j = C_j^m \cdot C_j^{m+1} \cdots C_j^{am}$, para $j \in \{1, 2, 3, 4, 5, 6\}$.

Para $m = 2017$, o resultado obtido foi, $P_1^{2017} = \{$SBSP3, BBAS3, BRAP4, EGIE3, MRVE3, SMLS3, CMIG4, IRBR3, BRKM5, VIVT3$\}$ e $C_1^{2018} = 35,82\%$. $P_2^{2017} = \{$CVCB3, ITUB4, BRAP4, ECOR3, NTCO3, SMLS3, CMIG4, IRBR3, BRKM5, FLRY3$\}$ e $C_2^{2018} = 26,98\%$. Enquanto isso, $C_3^{2018} = 17,05\%$, $C_4^{2018} = 18,61\%$, $C_5^{2018} = 26,89\%$, $C_6^{2018} = 13,32\%$. Nos anos que analisamos, o nosso procedimento gerou portfólios com desempenhos melhores que os demais. Na Tabela 1 realizamos uma comparação das rentabilidades dos portfólios. CF é o capital final utilizando a respectiva estratégia após os três anos, considerando um capital inicial de R$ 10.000.

Tabela 1: Comparação das Estratégias.

	$m = 2018$	$m = 2019$	$m = 2020$	CF
C_1^m	35,82%	53,53%	1%	R$21.064,37
C_2^m	26,98%	32,83%	4,06%	R$17.554,15
C_3^m	17,04%	29,96%	-0,14%	R$15.191,74
C_4^m	18,61%	50,12%	0%	R$17.822,99
C_5^m	26,89%	45,18%	-0,33%	R$17.827,25
C_6^m	13,32%	36,93%	-17,96%	R$12.732,06

Nos anos analisados, com as métricas utilizadas, nossa estratégia se mostrou com bom desempenho frente a outras estratégias. Para próximos passos, planejamos estender nossa estratégia para outros anos e outras métricas possíveis. Pelos resultados parciais que tivemos, acreditamos que a estratégia é muito promissora e que é uma boa alternativa para o investidor.

Referências

[1] Bender, J., Briand, R., Melas D., Subramanian R. A. *Foundations of Factor Investing*. 2013.

[2] Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888-905, 2000. DOI: 10.1109/34.868688.