LII Simpóso Brasileiro de Pesquisa Operacional
João Pessoa-PB, 3 a 5 de novembro de 2020

Comparação e avaliação de algoritmos genéticos em modelos de index tracking aplicados ao IBOVESPA

Thiago Wanderley de Amorim
GREEFO, Universidade Federal de Pernambuco – Recife-PE
Caixa Postal 7457, CEP 50630-971 - Recife, PE
twa@cin.ufpe.br

Julio Cezar Soares Silva
GREEFO, Universidade Federal de Pernambuco – Recife-PE
Caixa Postal 7457, CEP 50630-971 - Recife, PE
juliocss@greefo.org.br

Diogo Ferreira de Lima Silva
GREEFO, Universidade Federal de Pernambuco – Recife-PE
Caixa Postal 7457, CEP 50630-971 - Recife, PE
diogo@greefo.org.br

Adiel Teixeira de Almeida Filho
Centro de Informática, Universidade Federal de Pernambuco – Recife-PE
Caixa Postal 7457, CEP 50630-971 - Recife, PE
adielfilho@cin.ufpe.br

RESUMO

O problema de *index tracking* consiste em perseguir a performance de um índice com uma quantidade menor de seus ativos componentes. Este problema é útil para a composição de fundos de investimentos passivos, como ETF's. Porém, a limitação da quantidade de ativos aumenta a complexidade computacional do problema, tornando soluções ótimas inviáveis em tempo razoável. Este trabalho apresenta uma comparação entre um modelo de otimização linear e outro não-linear aplicados ao índice IBOVESPA. Foram apresentados dois algoritmos genéticos que objetivam obter soluções boas em tempo aceitável, quando comparados ao CPLEX. O melhor deles foi aplicado ao modelo linear, o qual apresentou maior tempo de solução se comparado ao não-linear. As performances dos portfolios resultantes foram comparadas no período pós-otimização, mostrando que o modelo linear apresentou uma menor medida de *tracking error*, apesar de menor retorno acumulado.

PALAVRAS CHAVE. Otimização de portfolio. Index tracking. Algoritmo genético.

Tópicos Gestão Financeira, Metaheurísticas

ABSTRACT

The index tracking problem consists in trying to mimic the performance of a given index with a smaller number of its assets. The problem is useful in creating passive management funds. However, the imposed limit of composing assets raises the computational complexity of the problem, making optimal solutions in a reasonable time unpractical. This paper presents a comparison between a linear and a non-linear optimization models applied to Brazilian Market Index (IBOVESPA). We present two genetic algorithms aiming good solutions in a feasible time when compared to CPLEX. The algorithm that presented the nearest solution to CPLEX was applied to the linear model, which has taken a longer solving time when confronted with the non-linear problem. The portfolios' performances were analyzed in the out-of-sample period, indicating that the linear model presented a smaller cumulative return, besides the lower tracking error.

KEYWORDS. Portfolio optimization. Index tracking. Genetic Algorithm.
Paper topics Financial Management, Metaheuristics

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

1. Introdução

A fim construir melhores modelos matemáticos para construção de carteiras que refletem o contexto de um determinado investidor, são adicionadas restrições práticas, como taxas de transação proporcionais ao montante [Guastaroba e Speranza 2012], limites de diversificação do portfólio, garantindo que um ativo não possua uma participação na carteira acima ou abaixo de certos níveis [Ruiz-Torrubiano e Suárez 2009,Guastaroba e Speranza 2012,Wang *et al.* 2012,Sant'Anna *et al.* 2017], ou limites máximos e mínimos de erro entre o portfólio e o objetivo [Guastaroba e Speranza 2012]. Por outro lado, essas restrições impactam na complexidade computacional do problema, podendo torná-lo um problema NP-difícil, e fazer com que sua solução seja inalcançável em um tempo de execução aceitável. Um *tracking portfolio* é formado por um subconjunto de ativos que compõem o índice alvo, pois a compra da carteira teórica implica em grandes custos para o gestor de portfólio [Beasley, Meade e Chang 2003,Ruiz-Torrubiano e Suárez 2009,Wang *et al.* 2012,Sant'Anna *et al.* 2017]. O problema de *index tracking* consiste na construção de um *tracking portfólio*, onde o modelo de otimização associado é de natureza combinatória, pois inclui restrições de limite na quantidade de ativos.

A obtenção de soluções exatas para modelos de *index tracking* é computacionalmente inviável com o aumento do tamanho do índice, portanto, são usados métodos heurísticos, como os evolucionários, na tentativa de encontrar soluções boas em um tempo de execução aceitável. O algoritmo genético (GA) é um método heurístico muito popular e com extensa aplicação na resolução de problemas de *index tracking* [Moral-Escudero, Ruiz-Torrubiano e Suárez 2006,Ruiz-Torrubiano e Suárez 2009,Sant'Anna *et al.* 2017]. Criado inicialmente por [John Holland 1975], esta abordagem propõe um algoritmo que tenta replicar o processo de seleção natural observado na natureza, buscando soluções melhores à cada iteração.

Este trabalho contribui para a literatura a partir da aplicação dois modelos de otimização para o problema de *index tracking*, bem como a utilização de heurísticas, no pouco explorado índice de ações brasileiro, analisando as performances em termos de função objetivo e retorno acumulado. Foi feita uma análise sobre a eficiência da aplicação dos modelos de Ruiz-Torrubiano e Suárez [2009] e Wang et al [2012] para a montagem de um *tracking portfolio* do índice da bolsa brasileira (*IBOVESPA*), bem como a performance dos dois algoritmos genéticos, presentes nos artigos de Ruiz-Torrubiano e Suárez [2009] e Sant'Anna et al [2017], os quais mostram-se por vezes mais rápidos e mais eficientes que os modelos completos, encontrando soluções melhores em tempos menores.

2. Trabalhos relacionados

Desde a publicação do artigo de Markowitz [1952], muitos trabalhos foram desenvolvidos no campo dos problemas de seleção de portfólio, possibilitando avanço em diferentes frentes do problema, como adição de diferentes restrições realistas aos modelos [Konno e Wijayanayake 2001,Guastaroba e Speranza 2012], definições de *tracking error* [Rudolf, Wolter e Zimmermann 1999,Beasley, Meade e Chang 2003]. Uma recente revisão sistemática da literatura indica algoritmos evolucionários multiobjetivo (onde GA se insere) como uma das principais abordagens utilizadas em aplicações de otimização de portfólio, possibilitando a obtenção de boas soluções em tempos viáveis [Almeida-Filho et al. 2020, Dueck e Winker 1992, Guastaroba e Speranza 2012, Lobo et al. 2007, Moral-Escudero et al. 2006, Sant'Anna et al. 2017, Scozzari et al. 2013]. Em meio a essa vasta gama de opções de modelos, experimentos computacionais são importantes seja para testar a adaptabilidade de modelos existentes em diferentes conjuntos de dados, ou para validar novos modelos e algoritmos [Silva, J. C. S. et al. 2018; Silva, J. C. S. S. et al. 2018]

Considerando modelos de *index tracking,* existem na literatura diversas propostas para definições de *tracking errors*, tais como a variância entre o retorno do portfólio e do índice [Roll

Galoá { Este trabalho foi publicado utilizando o Galoá proceedings

1992,Shapcott 1992], diferença média quadrática [Chang *et al.* 2000,Ruiz-Torrubiano e Suárez 2009] e a diferença média absoluta [Wang et al. 2012].

Em um problema de *index tracking*, a aplicação de métodos heurísticos como *kernel search* [Guastaroba e Speranza 2012], *best-extension-by-one* [Mutunge e Haugland 2018], *population heuristic* [Beasley et al. 2003] ou algoritmos genéticos [Ruiz-Torrubiano e Suárez 2009,Sant'Anna *et al.* 2017], permite resultados eficientes em tempos de execução oportunos para modelos complexos computacionalmente, com índices compostos por muitos ativos e um portfólio desejado restringido por uma constante de cardinalidade.

Dentre as diversas abordagens na criação de heurísticas, os algoritmos genéticos possuem extensa aplicação na solução de modelos de *index tracking*. Ruiz-Torrubiano e Suárez [2009] utilizam uma abordagem mista incluindo programação quadrática para valorar as soluções e um GA responsável por buscar novas possíveis soluções de maneira eficiente. O resultado é um algoritmo com bons resultados para problemas reais, mas que apresenta problemas de convergência para índices com muitas ações. Sant'Anna et al [2017] propõem um modelo de GA com uma rotina específica para a criação da população inicial, com o processos de *crossover* e mutação diferentes de Ruiz-Torrubiano e Suárez [2009]. Isso permitiu um algoritmo com resultados consistentes em pouco tempo computacional.

3. Materiais e métodos

Neste trabalho, foram comparados dois modelos de otimização para o problema de *index tracking*: um não-linear [Ruiz-Torrubiano e Suárez 2009], e outro linear [Wang et al. 2012], com diferentes medidas de *tracking error*, analisando as performances de seus portfólios no período *out-of-sample*.

Seja $\{S_i(t)\}_{i=1}^{N}$ a série temporal dos preços para os N ativos que compõem o índice I, o qual a performance deseja ser perseguida. De forma semelhante, $I(t)$ é a série dos preços do índice. O intervalo de tempo das duas séries é o mesmo: $t = 1, 2, 3, \ldots, T$. Definem-se os retornos $\{r_i(t)\}_{i=1}^{N}$ dos ativos e $r_I(t)$ do índice como:

$$r_{i(t)} = \frac{S_{i(t+1)} - S_{i(t)}}{S_{i(t)}} \; e \; r_{I(t)} = \frac{I(t+1) - I(t)}{I(t)}; \; t = 1, 2, \ldots, T \tag{1}$$

O tracking portfolio é composto por ativos pertencentes a ξ, subconjunto dos ativos inclusos no índice. Assumindo que o valor do portfólio em um tempo t é:

$$P(t) = \sum_{i \in \xi} c_i(t) S_i(t) \tag{2}$$

Além disso, assumindo que o montante $\{c_i(t)\}_{i \in \xi}$ investido em cada ativo seja constante no intervalo de tempo de $[t, t+1)$, o retorno do portfólio nesse período pode ser calculado como:

$$r_P(t) = \frac{P(t+1) - P(t)}{P(t)} = \sum_{i \in \xi} w_i(t) r_i(t) \tag{3}$$

onde:

$$w_i(t) = \frac{c_i(t) S_i(t)}{\sum_{j \in \xi} c_j(t) S_j(t)} \tag{4}$$

O *tracking error* durante o período $[1, 2, \ldots, T]$ é definido como o erro médio quadrático da série de retornos do *tracking portfolio*, $r_P(t)$, e os retornos do índice, $r_I(t)$.

$$MSE(r_P, r_I) = \frac{1}{T}\sum_{t=1}^{T}\left(r_P(t) - r_I(t)\right)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i\in\xi} w_i(t)r_i(t) - r_I(t)\right)^2 \quad (5)$$

A expressão (5) é o objetivo que deseja ser minimizado a fim de obter-se a performance ótima para o *tracking portfolio*. Uma das vantagens de usar essa medida como objetivo é tratar-se de um problema quadrático em relação aos pesos $\{w_i(t)\}_{i\in\xi}$ dos ativos, podendo ser minimizado com algoritmos confiáveis e eficientes de programação quadrática. Outro ponto favorável à utilização de (5) como objetivo é evitar os problemas de medidas baseadas apenas na variância do desvio, as quais se tornam zero caso a diferença entre os retornos do índice e do *tracking portfólio* sejam constantes ao longo do tempo.

Mantendo-se os pesos do portfólio constantes durante o período de otimização, os valores que os fazem minimizar o valor de (5) podem ser encontrados ao solucionar o seguinte problema de otimização inteira-mista quadrática [Ruiz-Torrubiano e Suárez 2009]:

$$Min_w \left[\frac{1}{2}w' \cdot H \cdot w - g' \cdot w\right] \quad (6)$$

Com as seguintes definições:

$$H_{ij} = \frac{1}{T}\sum_{t} r_i(t)r_j(t)\,;\; i,j = 1,\dots,N \quad (7)$$

$$g_i = \frac{1}{T}\sum_{t} r_i(t)r_I(t)\,;\; i = 1,\dots,N \quad (8)$$

Sujeito a:.

$$\sum_{i=1}^{N} w_i = 1 \quad (9)$$

$$w_i \geq 0; \forall i = 1,2,\dots,N \quad (10)$$

$$\sum_{i=1}^{N} z_i \leq K \quad (11)$$

$$a_i z_i \leq w_i \leq b_i z_i; a_i, b_i \geq 0;\; \forall i = 1,\dots,N \quad (12)$$

$$z_i \in \{0,1\};\; \forall i = 1,\dots,N \quad (13)$$

onde $\boldsymbol{w}$ é a matriz $N \times 1$ de pesos dos ativos. Foram incluídas as variáveis binárias z_i, as quais tem valor 1 se o ativo está incluído no tracking portfolio e 0 quando não. A equação (9) assegura que todo o capital disponível será investido no portfólio. A restrição (10) garante que os pesos de todos os ativos sejam positivos, não permitindo ao modelo operações de venda à descoberto. A equação (11) estabelece o limite de ativos a serem incluídos na carteira. Por fim, a equação (12) impõe limites de participação de um ativo, isto é, em caso de inclusão no portfólio, o peso de um ativo i deve se manter acima de a_i e abaixo de b_i, garantindo uma maior diversificação ao portfólio. Nota-se ainda que o resultado da otimização não é o valor mínimo do objetivo do investidor, mas com a otimização, é revelada a composição ótima da carteira, sendo assim possível a obtenção do valor da equação (5), por cálculos posteriores.

A otimização de (6) com as restrições (9)-(13) é feita para o período $[1,\dots,T]$, com o objetivo de construir portfolios com boas performances em um período $[T+1,\dots,T+L]$. Para isso, assume-se que portfolios ótimos em $[1,\dots,T]$ terão uma boa performance em um intervalo futuro L. Esta afirmação é razoável para um L pequeno, mas quando $T+L \to \infty$ sua validade passa a ser cada vez mais questionada. Por último, o período de otimização $[1,\dots,T]$ é chamado de *in-sample*. Este período é usado com um treino do modelo, para a formação de um portfólio com boa performance no período $[T+1,\dots,T+L]$ desejado, chamado de *out-of-sample*.

O *tracking error* utilizado no modelo a ser comparado é o desvio médio absoluto (DMA):

$$DMA = TE = \frac{1}{T}\sum_{t=1}^{T}\left|R^t - \sum_{i=1}^{N} r_i^t\, x_i\right| \tag{14}$$

O conjunto $x = \{x_1, x_2, \dots, x_N\}$ é o *tracking portfolio*, aonde x_i é o peso do ativo i no portfólio, o qual permanecerá inalterado durante todo o período de investimento. O problema de otimização é definido como [Wang et al. 2012]:

$$Min_{x,q^+,q^-,Z} T.E. = \frac{1}{T}\sum_{t=1}^{T}(q_t^+ + q_t^-) \tag{15}$$

Sujeito a:

$$q_t^+ - q_t^- = R^t - \sum_{i=1}^{N} r_i^t\, x_i \tag{16}$$

$$\sum_{i=1}^{N} z_i = K \tag{17}$$

$$L_i z_i \le x_i \le U_i z_i;\ i = 1, 2, \dots, N \tag{18}$$

$$\sum_{i=1}^{N} x_i = 1 \tag{19}$$

$$q_t^+, q_t^- \ge 0;\ i = 1, 2, \dots, T \tag{20}$$

$$z_i \in \{0,1\};\ i = 1, 2, \dots, N \tag{21}$$

Novamente são introduzidas ao modelo variáveis binárias (21), com $z_i = 1$ caso o ativo i esteja incluído no portfólio e $z_i = 0$ em caso contrário. A restrição (17) determina que a quantidade de ativos inclusos na carteira seja exatamente K. Para garantir que todo o capital disponível será investido, foi adicionada a restrição (19), enquanto (18) garante que a qualquer ativo participante da solução tenha uma participação mínima de L_i e máxima de U_i.

Quanto maior o número de ativos que compõem um índice, mais tempo será preciso para que um método de solução exato encontre a solução ótima dos modelos, visto que a restrição de cardinalidade é altamente não-linear [Ruiz-Torrubiano e Suárez 2009]. Na tentativa de contornar esse problema, foi usado um algoritmo genético responsável por gerar possíveis soluções ao tratar da natureza combinatória da restrição de cardinalidade e repassar de forma implícita à um modelo relaxado. Isto é, em um índice fictício de 10 ativos e um *tracking portfolio* com 3 ativos, o GA é responsável por gerar possíveis soluções formadas por ativos a serem investidos, como o conjunto {2,5,7}, em que cada número corresponde à um ativo que compõe o índice, enquanto o modelo relaxado será composto apenas por informações sobre os ativos em questão e o próprio índice. O valor de sua função objetivo será comparado com a de outros possíveis *tracking portfolios* gerados pelo algoritmo genético. Cada possível solução gerada pelo algoritmo é chamada de indivíduo.

Algoritmo Genético 1

O GA usado segue a mesma rotina desenvolvida por Ruiz-Torrubiano e Suárez[2009], com o tempo de execução limitado :

Passo 1: Criação de uma população inicial com P indivíduos.
Passo 2: Calcula-se a função objetivo de todos os indivíduos.
Passo 3: Até que o critério de convergência ou o tempo limite seja atingido:
  Passo 3.1: Removem-se aleatoriamente dois indivíduos da população, chamados de pais.

Passo 3.2: Aplica-se o operador RAR como *crossover*, gerando uma criança.
Passo 3.3: Aplica-se uma mutação na criança e calcula-se seu valor da função objetivo.
Passo 3.4: Devolvem-se a população o pai com menor função objetivo e a criança, removendo o pai de maior valor.

Dado um indivíduo, a avaliação de sua função objetivo é feita a partir dos pesos de cada ativo, retornados pela solução do modelo relaxado, sem as variáveis binárias e os demais ativos não incluídos no índice. Após o cálculo da função objetiva de todos os indivíduos da população, será realizada uma iteração, responsável por gerar uma nova possível solução a partir de outras duas inclusas na população, e substituir a pior delas pelo novo indivíduo gerado.

O *crossover* é o operador responsável pela formação de uma nova possível solução, chamada de criança, a partir de outras duas já existentes, nomeadas pais, de forma a carregar alguma informação dos seus pais. O método adotado, chamado de RAR (*Random Assorting Recombination*), segue os seguintes passos, dados dois pais:

Passo 1: Criar os conjuntos auxiliares (A, B, C, D, E, G), onde:
A: Ativos presentes em ambos os pais.
B: Ativos que não estão presentes em nenhum dos dois pais.
$C \equiv D$: Ativos presentes em apenas um dos pais.
E: Conjunto vazio.
Passo 2: Construir o conjunto G com:
w cópias dos elementos em A e B.
1 cópia dos elementos em C e D.
Passo 3: Criar a criança:
Passo 3.1: Retirar um elemento de G (sem repetição) até que a criança esteja completa ou o conjunto G vazio.
- Se esse elemento se origina de A ou C e não é elemento de E, então incluí-lo na criança.
- Se o elemento se origina de B ou D, inclui-lo no conjunto E.

Se a criança não estiver completa, incluir outros ativos aleatoriamente.

Após a aplicação do crossover, é efetuada sobre a criança gerada uma mutação simples, que consiste apenas na troca de um ativo incluso na criança por outro ainda não pertencente ao conjunto. A escolha do ativo que participará do portfólio e do que deixará de fazer parte é feita aleatoriamente. Como exemplo, dada a criança {2,6,8,9,10}, em um índice de 42 ativos, uma possível mutação seria {2,6,8,10,40}, ou ainda {2,6,9,10,22}.

Algoritmo Genético 2

Foi feita uma adaptação no GA proposto por Sant'Anna et al[2017]. A rotina desenvolvida neste trabalho consiste em:

Passo 1: Criação de uma população inicial criada aleatoriamente e com P indivíduos.
Passo 2: Calcula-se o valor da função objetivo de toda a população.
Passo 3: Até que o critério de convergência ou o limite de tempo seja atingido:
Passo 3.1: Escolhe-se aleatoriamente dois pais, sem removê-los da população.
Passo 3.2: Realiza-se o *crossover*, com uma taxa de ocorrência de T_{cross}, gerando duas crianças.
Passo 3.3: Valida-se a cardinalidade das crianças geradas, realizando ajustes caso necessário.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

Passo 3.4: Realiza-se a mutação, com uma taxa de ocorrência de T_{mut}, e se adicionam as crianças à população, a qual terá tamanho de $P + 2$ indivíduos.
Passo 3.5: Calcula-se o valor da função objetivo das duas crianças.
Passo 3.6: É criada uma nova população de tamanho *P*, eliminando os dois indivíduos com maiores funções objetivo da população anterior. Este novo conjunto será a população inicial da próxima iteração

Foi adotada para este GA a representação binária de cada indivíduo. Com base nisso, cada indivíduo é um vetor binário de tamanho *N*, onde *N* é a quantidade de ativos presentes no índice, e o *r-ésimo* ativo que compõe o índice é representado pelo elemento de mesma ordem no vetor, com valor 1 caso esteja incluído em uma possível solução, e 0 em caso negativo. Desta forma, em um índice de 10 ativos e um *tracking portfólio* com 5 ações, o indivíduo anteriormente representado {1,2,4,6,8} é agora escrito como [1, 1, 0, 1, 0, 1, 0, 1, 0, 0].

Para a realização do *crossover*, foi realizado o processo de *1-point order crossover*: escolhe-se um ponto de corte c_{corte} entre 0 e o tamanho *N* do índice, e são geradas duas crianças. A primeira criança será formada por toda a informação do pai 1 presente antes do ponto c_{corte}, e a informação do pai 2 após o ponto de corte. A criança 2 é formada de maneira contraria à primeira, carregando a informação do pai 2 presente antes de c_{corte} e do pai 1 após o ponto. Em um índice de 10 ativos e um *tracking portfolio* com 5 ativos, dados os seguintes pais e o ponto $c_{corte} = 4$, o *crossover* resulta em

Pai1 [0, 1, 1, 0, 1, 0, 0, 1, 0, 1]
Pai2 [1, 1, 0, 0, 0, 1, 1, 0, 0, 1]
$c_{corte} = 4$

Inf. Pai1 [0, 1, 1, 0] + **Inf. Pai2** [0, 1, 1, 0, 0, 1] → [0, 1, 1, 0, 0, 1, 1, 0, 0, 1] Criança1
Inf. Pai2 [1, 1, 0, 0] + **Inf. Pai1** [1, 0, 0, 1, 0, 1] → [1, 1, 0, 0, 1, 0, 0, 1, 0, 1] Criança2

Em contrapartida, esse *crossover* pode gerar crianças com uma quantidade de ativos maior ou menor que a quantidade desejada. Como exemplo, basta que, no pai1, o ativo de ordem 1 fosse incluído no lugar do décimo ativo. O resultado seria que a Criança1 teria uma cardinalidade de 6 ativos, enquanto a Criança2 possuiria apenas 4 ativos. Nessa situação, as duas crianças geradas não têm a quantidade desejada de ativos no portfólio.

Para resolver esse problema é realizada uma validação da criança gerada, aonde é retirado aleatoriamente um ativo caso sua cardinalidade seja maior do que o desejado ou, caso a cardinalidade seja menor, um ativo é incluído de forma aleatória. Esse processo é repetido até que a criança tenha a mesma cardinalidade dos outros indivíduos da população.

Após o *crossover* e a validação das duas crianças geradas, é realizada uma mutação, retirando um ativo já incluso, e adicionando outro que não pertencia ao indivíduo. Tanto a fase de *crossover* quanto a fase de mutação possuem probabilidades de ocorrência T_{cross} e T_{mut}, respectivamente.

Escolhido um modelo e uma cardinalidade de portfólio, o algoritmo realiza 30 instâncias, a fim de obter-se uma média consistente de sua performance. Assim, foram comparados os resultados médios dos dois algoritmos.

Para o algoritmo genético 1, foi definida uma população inicial de $P = 100$, com hiper parâmetro do *crossover* $w = 1$, assim como em Ruiz-Torrubiano e Suárez [2009]. Enquanto os parâmetros adotados para o GA 2 foram $P = 15, T_{cross} = 1, T_{mut} = 0.85$, conforme em Sant'Anna et al[2017]

No modelo1, os parâmetros adotados foram $a_i = a = 1$ *e* $b_i = b = 0.001$. Os parâmetros do modelo 2 escolhidos foram $L_i = L = 0.01$ *e* $U_i = U = 0.5$. Para ambos os modelos. O valor *c* da cardinalidade do portfólio foi variada no intervalo [5,10].

4. Resultados

Os dados usados para validação dos modelos e dos algoritmos genéticos foram retirados da OR-LIBRARY[Beasley et al. 2003], base de dados pública com grande variedade de dados para problemas de pesquisa operacional. Foram usados dados correspondentes a 145 retornos semanais de 5 índices e ativos componentes: Hang Seng (31 ativos), DAX (85 ativos), FTSE (89 ativos), S&P (98 ativos) e Nikkei (225 ativos). Configurou-se limites de tempo para a operação do CPLEX e dos GAs de, respectivamente, 5400s e 300s. Os testes foram realizados em instâncias virtuais do *Google Cloud Platform* com configuração de 8vCPU's com arquitetura *Skylake* e frequência base de 2,0 Ghz e memória de 30 Gb.

Para a comparação dos algoritmos genéticos, foram analisados seus erros relativos à solução do CPLEX, o número de iterações e o tempo de execução.

$$Erro_{relativo} = \frac{Erro_{GA} - Erro_{CPLEX}}{Err_{\ CPLEX}}. \quad (22)$$

Segundo a equação (22), um erro relativo negativo implica em uma solução melhor do que a encontrada pelo CPLEX.

Os resultados para as aplicações do algoritmos genéticos no modelo 1, bem como os dados sobre as funções objetivo e o tempo de otimização do CPLEX estão representados na Tabela 1. A primeira coluna das tabelas informa a cardinalidade do portfólio analisado, enquanto a última coluna mostra a quantas das soluções encontradas pelos GAs que foram iguais ao valor obtido pelo CPLEX.

Tabela 1- Validação dos Algoritmos Genéticos para o Modelo 1

Índice (ativos)	GA	Card.	MSE	Tempo CPLEX (s)	Tempo médio GA (s)	No Iterações	Erro Rel. Médio (%)	Maior Erro Rel. (%)	Menor Erro Rel. (%)	Instâncias Iguais CPLEX
Hang Seng (31)	1	K=5	4.13e-5	6.61	13.40	1123	0.00	0.00	0.00	30
		K=7	2.37e-5	7.02	24.08	1657	0.00	0.00	0.00	30
		K=9	1.62e-5	9.06	77.03	4228	0.10	0.99	0.00	27
	2	K=5	4.13e-5	6.61	5.14	199	0.00	0.00	0.00	30
		K=7	2.37e-5	7.02	6.92	224	0.00	0.00	0.00	30
		K=9	1.62e-5	9.06	26.11	656	0.00	0.00	0.00	30
DAX (85)	1	K=5	2.21e-5	3230.02	75.74	4333	0.32	3.35	0.00	27
		K=7	1.37e-5	5400	136.37	6867	0.87	10.87	0.00	26
		K=9	9.22e-6	5400	170.22	7125	0.80	13.51	0.00	28
	2	K=5	2.21e-5	3230.02	71.45	1669	0.09	2.81	0.00	29
		K=7	1.37e-5	5400	69.60	1423	0.00	0.00	0.00	30
		K=9	9.22e-6	5400	84.57	1458	0.00	0.00	0.00	30
FTSE (89)	1	K=5	6.42e-5	5400	141.79	8378	1.44	6.11	0.00	22
		K=7	3.83e-5	5400	192.71	9109	3.53	18.71	0.00	22
		K=9	2.48e-5	5400	300	11995	6.61	28.19	0.00	0
	2	K=5	6.42e-5	5400	184.85	4968	2.51	5.36	0.00	15
		K=7	3.83e-5	5400	164.14	3856	4.23	18.71	0.00	22
		K=9	2.48e-5	5400	300	5720	8.64	33.35	0.00	0
S&P (98)	1	K=5	4.50e-5	5400	108.79	5657	0.69	5.19	0.00	26
		K=7	2.76e-5	5400	175.28	8002	1.95	8.46	0.00	21
		K=9	1.94e-5	5400	237.98	9248	3.78	12.35	0.00	16
	2	K=5	4.50e-5	5400	76.19	1910	0.00	0.00	0.00	30
		K=7	2.76e-5	5400	115.69	2638	0.35	5.26	0.00	28
		K=9	1.94e-5	5400	213.81	4070	3.34	9.44	0.00	16
Nikkei (225)	1	K=5	5.63e-5	5400	296.69	9690	1.63	15.56	-3.08	2
		K=7	3.38e-5	5400	300	8104	13.94	22.62	-6.50	0
		K=9	2.54e-5	5400	300	8115	6.31	24.74	-12.46	0
	2	K=5	5.63e-5	5400	300	4230	3.86	17.74	-3.08	0
		K=7	3.38e-5	5400	300	3910	10.73	32.39	0.32	0
		K=9	2.54e-5	5400	300	3512	1.38	18.11	-10.71	0

Galoá { Este trabalho foi publicado utilizando o Galoá proceedings

A Tabela 2 apresenta os resultados obtidos com os algoritmos genéticos aplicados ao modelo 2.

Tabela 2- Validação dos Algoritmos Genéticos para o Modelo 2

Índice (ativos)	GA	Card.	DMA	Tempo CPLEX (s)	Tempo médio GA (s)	No Iterações	Erro Rel. Médio (%)	Maior Erro Rel. (%)	Menor Erro Rel. (%)	Instâncias Iguais CPLEX
Hang Seng (31)	1	K=5	5.01e-3	1.45	34.44	1156	0.00	0.00	0.00	30
		K=7	3.74e-3	1.72	89.16	2843	0.01	0.12	0.00	28
		K=9	3.10e-3	4.35	90.94	2776	0.00	0.00	0.00	30
	2	K=5	5.01e-3	1.45	13.51	190	0.00	0.00	0.00	30
		K=7	3.74e-3	1.72	47.70	579	0.00	0.00	0.00	30
		K=9	3.10e-3	4.35	37.40	459	0.00	0.00	0.00	30
DAX (85)	1	K=5	3.56e-3	179.72	137.15	4495	0.20	4.10	0.0	27
		K=7	2.89e-3	5400	300	9512	0.03	1.01	-0.15	0
		K=9	2.41e-3	5400	300	8873	-1.23	2.19	-2.22	0
	2	K=5	3.56e-3	179.72	42.40	587	0.00	0.00	0.00	30
		K=7	2.89e-3	5400	300	3936	0.03	0.62	-0.15	0
		K=9	2.41e-3	5400	300	3887	-1.21	6.93	-2.22	0
FTSE (89)	1	K=5	6.24e-3	5400	164.19	5696	-0.49	0.54	-0.97	21
		K=7	5.02e-3	5400	295.72	10000	-2.20	5.75	-4.31	1
		K=9	4.17e-3	5400	300	9288	-3.55	5.28	-7.43	0
	2	K=5	6.24e-3	5400	166.28	2425	-0.22	1.46	-0.97	18
		K=7	5.02e-3	5400	281.36	3968	-2.76	3.59	-4.31	3
		K=9	4.17e-3	5400	297.42	3842	-4.09	5.17	-7.43	1
S&P (98)	1	K=5	5.26e-3	5400	237.77	7797	0.78	2.87	-0.24	10
		K=7	4.23e-3	5400	300.02	9625	-0.38	2.59	-1.55	0
		K=9	3.38e-3	5400	300	9110	3.29	12.64	-0.31	0
	2	K=5	5.26e-3	5400	134.49	1812	0.01	1.17	-0.24	27
		K=7	4.23e-3	5400	300	4033	-1.17	0.83	-1.55	0
		K=9	3.38e-3	5400	300	4109	2.48	8.37	-0.31	0
Nikkei (225)	1	K=5	6.02e-3	5400	300	9978	-1.74	4.56	-6.52	0
		K=7	4.46e-3	5400	300	9509	5.34	12.65	-3.63	0
		K=9	3.83e-3	5400	300	9642	3.43	8.59	-4.69	0
	2	K=5	6.02e-3	5400	300	4195	-1.79	2.73	-6.52	0
		K=7	4.46e-3	5400	300	3966	4.10	10.71	-3.33	0
		K=9	3.83e-3	5400	300	4126	0.80	9.75	-13.73	0

Após a validação, os modelos foram aplicados no índice da bolsa de valores brasileira, IBOVESPA. Os dados utilizados correspondem aos preços diários do de 18/08/2011 até 17/07/2012, retirados de Sant'Anna et al [2017], compreendendo um período de 227 dias, com 226 informações sobre retorno. Os primeiros 113 dias foram usados como período *in-sample*, enquanto os demais 113 dias correspondem ao período *out-of-sample*.

Após a otimização com o CPLEX, foi aplicado o segundo GA, o qual obteve um melhor desempenho se comparado ao primeiro, na solução do modelo 2 para portfólios com mais de 5 ativos, visto que o tempo de operação do solver exato superou o limite de execução do algoritmo genético, o que possibilitou a obtenção de resultados próximos ou melhores às soluções do CPLEX em um tempo de execução menor. Tabela 3 ilustra os resultados obtidos nesse experimento.

Tabela 3- Resultados da aplicação do GA2 no Modelo 2

Índice (ativos)	Card.	DMA	Tempo CPLEX (s)	Tempo Médio GA (s)	No Iterações	Erro Rel. Médio (%)	Maior Erro Rel. (%)	Menor Erro Rel. (%)	Instâncias Iguais CPLEX
IBOV (67)	K=5	2.24e-3	127.61	-	-	-	-	-	-
	K=6	1.94e-3	401.80	300	6450	0.10	0.13	0.10	0
	K=7	1.75e-3	1160.65	300	6142	-1.20	0.62	-1.47	0
	K=8	1.61e-3	3955.67	300	6452	-0.91	0.63	-1.97	0

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

	K=9	1.44e-3	5400	300	6208	0.30	3.30	-1.62	0
	K=10	1.37e-3	5400	300	5629	-0.56	3.38	-2.55	0

Análises foram realizadas com o objetivo de comparar as performances dos dois modelos durante o período *out-of-sample*. A Tabela 4 contém os resultados obtidos para diferentes metidasde *tracking error*. Para o modelo 2, foi utilizada a melhor das 30 soluções retornadas pelo GA 2 quando $K \geq 6$ e a solução do CPLEX para $K = 5$. O retorno acumulado do índice foi de 83.71%.

Tabela 4- Resultados para os períodos out-of-sample.

		Desvio Médio Absoluto (%)		**Maior Erro Absoluto (%)**		**MSE (%)**		**Maior Erro Quadrático (%)**	
	Card.	**Modelo 1**	**Modelo 2**	**Modelo 1**	**Modelo 2**	**Modelo 1**	**Modelo 2**	**Modelo 1**	**Modelo 2**
IBOV (67)	K=5	4.05e-3	3.77e-3	1.44e-2	1.80e-2	2.57e-5	2.46e-5	2.07e-4	3.22e-4
	K=6	3.20e-3	3.16e-3	1.49e-2	1.35e-2	1.93e-5	1.76e-5	2.21e-4	1.82e-4
	K=7	3.12e-3	3.02e-3	1.28e-2	1.04e-2	1.66e-5	1.48e-5	1.63e-4	1.09e-4
	K=8	3.18e-3	3.04e-3	1.17e-2	9.75e-3	1.63e-5	1.49e-5	1.37e-4	9.50e-5
	K=9	2.88e-3	2.82e-3	9.90e-3	9.45e-3	1.33e-5	1.31e-5	9.81e-5	8.94e-5
	K=10	2.86e-3	2.92e-3	1.29e-2	1.08e-2	1.34e-5	1.32e-5	1.67e-4	1.18e-4

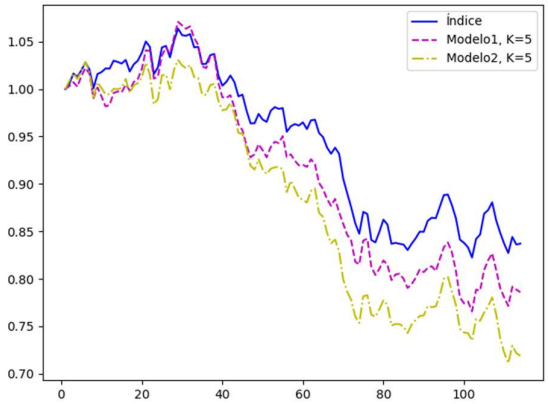
Os resultados do retorno acumulado obtido para cada modelo, assim como as medidas de desvio padrão para erro absoluto e quadrático, são apresentados na Tabela 5.

Tabela 5 Retornos acumulados e desvio padrão dos erros para os modelos 1 e 2

		Retorno Acumulado (%)		**Desvio Padrão Erro Absoluto**		**Desvio Padrão Erro Quadrático**	
	Card.	**Modelo 1**	**Modelo 2**	**Modelo 1**	**Modelo 2**	**Modelo 1**	**Modelo 2**
IBOV (67)	K=5	78.60	71.88	3.04e-3	3.23e-3	3.53e-5	4.50e-5
	K=6	85.63	79.73	3.01e-3	2.76e-3	3.71e-5	2.83e-5
	K=7	84.09	82.47	2.62e-3	2.39e-3	2.86e-5	2.12e-5
	K=8	80.12	80.87	2.48e-3	2.39e-3	2.46e-5	2.14e-5
	K=9	81.61	81.81	2.24e-3	2.27e-3	1.92e-5	1.87e-5
	K=10	83.83	84.06	2.28e-3	2.16e-3	2.22e-5	1.92e-5

A Figura 1 ilustra o retorno acumulado do índice com a solução de cada modelo para a cardinalidade $K = 5$. Observa-se que o modelo consegue seguir o padrão do índice em suas altas e baixas.

Figura 1 Retornos acumulados dos modelos 1, 2 e índice para K=5



4. Conclusões

Na escolha do algoritmo genético a ser implementado nos modelos, o GA 2 mostrou-se mais eficiente, alcançando valores mais próximos ou melhores do que o CPLEX. Portanto, esse algoritmo foi aplicado nas instâncias do modelo 2 com $K \geq 6$, as quais levaram mais tempo para serem resolvidas pelo CPLEX do que o tempo de execução do GA. Foram então comparadas as performances dos portfolios gerados pelos modelos 1 e 2 a fim de identificar qual modelo apresenta melhores soluções no período *out-of-sample*.

A solução do modelo 2, no período *out-of-sample*, resultou em portfólios com função objetivo menores, se comparadas as soluções do modelo 1. Portanto, o modelo linear, apesar de mais complexo e, mesmo com a aplicação do algoritmo genético, mais demorado, retorna soluções melhores.

No entanto, o primeiro modelo apresentou performances mais próximas ao índice em um gráfico de retorno acumulado, como na Figura 1. Esse comportamento se repetiu para portfólios contendo entre 6 e 10 ativos, e também não foi observada relação entre o retorno acumulado e o desvio padrão dos *tracking errors*. Desta forma, mesmo que apresente uma melhor função objetivo, o modelo 2 não se apresenta como a melhor resposta à uma aplicação real do modelo por parte de um investidor, visto que este modelo resultaria em uma performance ao final do período *out-of-sample* pior do que a entregue pelo modelo 1.

O cálculo do *tracking error* considera apenas o valor absoluto dos retornos, enquanto ao calcular o retorno acumulado esse sinal é considerado. Desta forma, é possível que o modelo 1 apresente um retorno acumulado mais distante do índice se comparado ao modelo 2, em uma situação em que o retorno absoluto do modelo 2 é mais próximo do índice, mas ainda assim negativo, enquanto o do modelo 1 é positivo. Nesta ocasião, o *tracking error* do modelo 1 será menor do que a medida do modelo 2.

Em suma, o modelo 1 apresentou retornos acumulados mais próximos do índice, apesar de *tracking errors* mais altos, exceto para portfólios com 10 ativos. Essa inversão pode indicar uma melhor performance do modelo 2 para portfólios com mais de 11 ações incluídas. A análise de portfolios com maior cardinalidade pode ser tema de trabalhos futuros, bem como a comparação do algoritmo genético desenvolvido com outros métodos heurísticos como *Kernel Search* [Guastaroba e Speranza 2012], ou ainda a aplicação de medidas de *tracking error* que impliquem em um maior retorno acumulado.

Agradecimentos

Este trabalho foi parcialmente apoiado pela Facepe, CNPq e Capes.

Referências

Almeida-Filho, A. T. De, De Lima Silva, D. F. e Ferreira, L. (22 jun 2020). Financial modelling with multiple criteria decision making: A systematic literature review. *Journal of the Operational Research Society*, p. 1–19.

Beasley, J. E., Meade, N. e Chang, T. J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research*, v. 148, n. 3, p. 621–643.

Chang, T. J., Meade, N., Beasley, J. E. e Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers and Operations Research*, v. 27, n. 13, p. 1271–1302.

Dueck, G. e Winker, P. (1992). New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis*, v. 8, n. 3, p. 159–178.

Guastaroba, G. e Speranza, M. G. (2012). Kernel Search: An application to the index tracking problem. *European Journal of Operational Research*, v. 217, n. 1, p. 54–68.

Holland, J. (1975). Adaptation in natural and artificial systems. Ann Arbor: The University of Michigan Press.

KONNO, H. e WIJAYANAYAKE, A. (2001). Minimal Cost Index Tracking Under Nonlinear

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings

Transaction Costs and Minimal Transaction Unit Constraints. *International Journal of Theoretical and Applied Finance*, v. 04, n. 06, p. 939–957.
Lobo, M. S., Fazel, M. e Boyd, S. (2007). Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, v. 152, n. 1, p. 341–365.
Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, v. 7, n. 1, p. 77–91.
Moral-Escudero, R., Ruiz-Torrubiano, R. e Suárez, A. (2006). Selection of optimal investment portfolios with cardinality constraints. *2006 IEEE Congress on Evolutionary Computation, CEC 2006*, p. 2382–2388.
Mutunge, P. e Haugland, D. (2018). Minimizing the tracking error of cardinality constrained portfolios. *Computers and Operations Research*, v. 90, p. 33–41.
Roll, R. (31 jul 1992). A Mean/Variance Analysis of Tracking Error. *The Journal of Portfolio Management*, v. 18, n. 4, p. 13–22.
Rudolf, M., Wolter, H. J. e Zimmermann, H. (1999). A linear model for tracking error minimization. *Journal of Banking and Finance*, v. 23, n. 1, p. 85–103.
Ruiz-Torrubiano, R. e Suárez, A. (2009). A hybrid optimization approach to index tracking. *Annals of Operations Research*, v. 166, n. 1, p. 57–71.
Sant'Anna, L. R., Filomena, T. P., Guedes, P. C. e Borenstein, D. (2017). Index tracking with controlled number of assets using a hybrid heuristic combining genetic algorithm and non-linear programming. *Annals of Operations Research*, v. 258, n. 2, p. 849–867.
Scozzari, A., Tardella, F., Paterlini, S. e Krink, T. (2013). Exact and heuristic approaches for the index tracking problem with UCITS constraints. *Annals of Operations Research*, v. 205, n. 1, p. 235–250.
Shapcott, J. (1992). Index Tracking : Genetic Algorithms for Investment Portfolio Selection. *Computing*, p. 1–24.
Silva, J. C. S., Feitosa, I. S. C. de S., Silva, D. F. de L. e De Almeida-Filho, A. (2018). Experimentos com GRASP para um modelo de seleção de carteiras de investimentos | Galoá Proceedings. *SBPO*,
Silva, J. C. S. S., Feitosa, I. S. C. de S., Silva, D. F. de L. e De Almeida-Filho, A. (2018). Avaliação de diferentes formulações do problema de seleção de carteiras para investimentos em ativos da Bovespa | Galoá Proceedings. *SBPO*,
Wang, M., Xu, C., Xu, F. e Xue, H. (2012). A mixed 0-1 LP for index tracking problem with CVaR risk constraints. *Annals of Operations Research*, v. 196, n. 1, p. 591–609.

Galoá { Este trabalho foi publicado utilizando o **Galoá** proceedings