

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**ESCOLA DE ADMINISTRAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA/UFRGS**

**Leonardo Riegel Sant’Anna**

**Index Tracking com Controle do Número de Ativos e  
Aplicação com Uso de Algoritmos Genéticos.**

**Porto Alegre**

**Janeiro/2014**

## CIP - Catalogação na Publicação

Riegel Sant'Anna, Leonardo  
Index Tracking com Controle do Número de Ativos e  
Aplicação com Uso de Algoritmos Genéticos / Leonardo  
Riegel Sant'Anna. -- 2014.  
77 f.

Orientador: Tiago Pascoal Filomena.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Escola de Administração, Programa  
de Pós-Graduação em Administração, Porto Alegre, BR-RS,  
2014.

1. Index Track. 2. Otimização de Carteiras. 3.  
Programação Quadrática e Inteira. 4. Administração de  
Ativos. I. Pascoal Filomena, Tiago, orient. II.  
Título.

**Leonardo Riegel Sant'Anna**

**Index Tracking com Controle do Número de Ativos e  
Aplicação com Uso de Algoritmos Genéticos.**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Administração pelo Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul.

**Orientador: Prof. Tiago Pascoal Filomena,  
PhD.**

**Porto Alegre**

**Janeiro/2014**

**Leonardo Riegel Sant'Anna**

**Index Tracking com Controle do Número de Ativos e  
Aplicação com Uso de Algoritmos Genéticos.**

Dissertação apresentada como requisito para a obtenção do título de Mestre em Administração pelo Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul.

**Orientador: Prof. Tiago Pascoal Filomena,  
PhD.**

BANCA EXAMINADORA:

---

**Prof. Dr. Denis Borenstein – PPGA/UFRGS**

---

**Prof. Dr. João Frois Caldeira – PPGA/UFRGS**

---

**Prof. Dr. Vilmar Trevisan – PPG Matemática Aplicada/UFRGS**

**Porto Alegre**

**Janeiro/2014**

## RESUMO

Nesta dissertação, discute-se o problema de otimização de carteiras de investimento para estratégia passiva de *Index Tracking*. Os objetivos principais são (i) apresentar um modelo de otimização de *Index Tracking* e (ii) a solucionar esse modelo com uso do método heurístico de Algoritmos Genéticos (AG) para formação de carteiras com número reduzido de ativos. O índice de referência utilizado é o Ibovespa, para o período de Janeiro/2009 a Julho/2012, com um total de 890 observações diárias de preços. A partir de uma amostra de 67 ativos, são formadas carteiras sem limite de ativos e limitadas a 40, 30, 20, 10 e 05 ativos; os intervalos de rebalanceamento das carteiras são 20, 40 e 60 períodos (dias úteis), ou seja, rebalanceamento mensal, bimestral e trimestral. É verificado que, para essa amostra, não é possível formar carteiras de 20 ou menos ativos via otimização direta com o *solver* Cplex com menos de 1 hora de processamento e *gap* abaixo de 5%. Com uso da heurística de Algoritmos Genéticos, são formadas carteiras de 10 e 05 ativos com tempo de processamento em torno de 5 minutos; nesse caso, o *gap* médio fica abaixo de 10% para ambos os tipos de carteira. E, com tempo de processamento do AG um pouco maior, em torno de 8 minutos, o algoritmo fornece soluções para carteiras de 10 e 05 ativos com *gap* médio abaixo de 5%.

**Palavras-chave:** *Index Track*, Otimização de Carteiras, Programação Quadrática e Inteira, Administração de Ativos.

## ABSTRACT

In this master's thesis it is discussed the portfolio optimization problem using the passive investment strategy of Index Tracking. The main goals are (i) to present an optimization model for the Index Tracking problem and (ii) to solve this model using the heuristic approach of Genetic Algorithms (GA) to create portfolios with reduced amount of stocks. The benchmark used is the Ibovespa Index (main reference for the Brazilian Stock Market), during the period from January/2009 to July/2012 (using a total of 890 daily stock prices). The sample contains 67 assets, and the model is used to build portfolios without limit in the amount of assets and portfolios limited to 40, 30, 20, 10 and 05 assets; the ranges of time to rebalance the portfolios are 20, 40, and 60 trading days, which means to rebalance monthly, bimonthly and quarterly. The results show that, considering this sample, it is not possible to build portfolios with 20 stocks (or less than 20) through direct optimization using the solver Cplex with computational processing time less than 1 hour and results with gap below 5%. On the other hand, using the Genetic Algorithms heuristic approach, portfolios limited to 10 and 05 stocks are built with computational time close to 5 minutes; for both types of portfolio, the solutions provided by the GA have average gap below 10%. Also, with a computational time slightly bigger, close to 8 minutes, the algorithm provides solutions with average gap below 5% for portfolios limited to 10 and 05 stocks.

**Keywords:** Index Track, Portfolio Optimization, Quadratic Integer Programming, Asset Management.

## ÍNDICE DE FIGURAS

<b>Figura 01</b> – Tempo Médio, Tempo Máximo e <i>Gap</i> Médio de solução das 36 carteiras geradas para Rebalanceamento Mensal (20 períodos) .....	31
<b>Figura 02</b> – Tempo Médio, Tempo Máximo e <i>Gap</i> Médio de solução das 18 carteiras geradas para Rebalanceamento Bimestral (40 períodos) .....	31
<b>Figura 03</b> – Tempo Médio, Tempo Máximo e <i>Gap</i> Médio de solução das 12 carteiras geradas para Rebalanceamento Trimestral (60 períodos) .....	31
<b>Figura 04</b> – Formato de cada indivíduo utilizado na heurística .....	45
<b>Figura 05</b> – Descrição do processo de <i>crossover</i> entre dois indivíduos .....	46
<b>Figura 06</b> – Descrição do processo de mutação entre dois indivíduos .....	46
<b>Figura 07</b> – Fluxo resumido de cada Fase de Evolução do Algoritmo Genético .....	47
<b>Figura 08</b> – Rendimento por carteira para Carteiras de 05 e 10 ativos – Rebalanceamento Trimestral .....	52
<b>Figura 09</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 20 períodos, Rebalanceamento Mensal, Carteiras de 20 ou mais ativos .....	69
<b>Figura 10</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 40 períodos, Rebalanceamento Bimestral, Carteiras de 20 ou mais ativos .....	70
<b>Figura 11</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 60 períodos, Rebalanceamento Trimestral, Carteiras de 20 ou mais ativos .....	71
<b>Figura 12</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 20 períodos, Rebalanceamento Mensal, Carteiras de 05 e 10 ativos .....	72
<b>Figura 13</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 40 períodos, Rebalanceamento Bimestral, Carteiras de 05 e 10 ativos .....	73
<b>Figura 14</b> – Projeção de resultados das carteiras com intervalo fora da amostra de 60 períodos, Rebalanceamento Trimestral, Carteiras de 05 e 10 ativos .....	74

## ÍNDICE DE TABELAS

<b>Tabela 01</b> – Resultados para formação de Carteiras Sem Limite de Ativos com otimização via Cplex .....	32
<b>Tabela 02</b> – Resultados para formação de Carteiras com Limite de 40 Ativos com otimização via Cplex .....	33
<b>Tabela 03</b> – Resultados para formação de Carteiras com Limite de 30 Ativos com otimização via Cplex .....	33
<b>Tabela 04</b> – Resultados para formação de Carteiras com Limite de 20 Ativos com otimização via Cplex .....	34
<b>Tabela 05</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 20 períodos (Rebalanceamento Mensal) .....	35
<b>Tabela 06</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 40 períodos (Rebalanceamento Bimestral) .....	35
<b>Tabela 07</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 60 períodos (Rebalanceamento Trimestral) .....	36
<b>Tabela 08</b> – Comparação de resultados para Carteiras com intervalo <i>out-of-sample</i> de 20 períodos e limite de 20 ativos, para $T = 100$ e $T = 150$ .....	37
<b>Tabela 09</b> – Parâmetros do Cplex alterados nos novos testes computacionais .....	40
<b>Tabela 10</b> – Resultados parciais dos testes com alteração de parâmetros do Cplex .....	40
<b>Tabela 11</b> – Resultados para testes com alteração de parâmetros do Cplex .....	41
<b>Tabela 12</b> – Parâmetros adotados para os Testes Iniciais do algoritmo .....	49
<b>Tabela 13</b> – Resultados para formação de Carteiras com Limite de 05 Ativos .....	50
<b>Tabela 14</b> – Resultados para formação de Carteiras com Limite de 10 Ativos .....	51
<b>Tabela 15</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 20 períodos .....	53
<b>Tabela 16</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 40 períodos .....	53
<b>Tabela 17</b> – Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 60 períodos .....	53
<b>Tabela 18</b> – Resultados para verificação de <i>gap</i> das soluções obtidas com o algoritmo genético .....	55



<b>Tabela 19</b> – Descrição dos novos testes do algoritmo com carteiras de limites de 05 e 10 ativos, com alteração de parâmetros e do tempo de execução .....	57
<b>Tabela 20</b> – Resultados para testes com alterações de parâmetros no algoritmo – Formação de carteiras com limite de 05 ativos .....	58
<b>Tabela 21</b> – Resultados para testes com alterações de parâmetros no algoritmo – Formação de carteiras com limite de 10 ativos .....	59
<b>Tabela 22</b> – Lista de ativos da amostra de dados utilizada .....	75
<b>Tabela 23</b> – Descrição dos ativos usados para formação das carteiras de 05 ativos, considerando rebalanceamento trimestral .....	76
<b>Tabela 24</b> – Descrição dos ativos usados para formação das carteiras de 10 ativos, considerando rebalanceamento trimestral .....	77

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	11
1.1. Objetivos.....	12
1.2. Justificativa.....	14
1.3. Organização da Dissertação.....	16
<b>2. REVISÃO BIBLIOGRÁFICA</b> .....	18
2.1. Otimização de portfólios de investimento de <i>Index Tracking</i> .....	18
2.2. Uso de Algoritmos Genéticos para o problema de <i>Index Tracking</i> .....	22
2.3. Considerações sobre a Revisão Bibliográfica .....	23
<b>3. METODOLOGIA</b> .....	25
3.1. Modelo de Otimização .....	25
3.2. Base de dados e descrição do problema .....	27
<b>4. OTIMIZAÇÃO COM APLICAÇÃO DIRETA DO SOLVER CPLEX</b> .....	29
4.1. Testes computacionais iniciais com uso do <i>solver</i> Cplex.....	30
4.2. Testes com alteração da constante $T$ (tamanho da amostra <i>in-sample</i> ) .....	36
4.3. Alteração de parâmetros do <i>solver</i> Cplex .....	38
4.4. Considerações sobre os testes iniciais de otimização com o <i>solver</i> Cplex .....	42
<b>5. MÉTODO HEURÍSTICO: ALGORITMOS GENÉTICOS</b> .....	44
5.1. Método de Algoritmos Genéticos adaptado ao problema de <i>Index Tracking</i> .....	44
5.2. Resultados iniciais da aplicação do algoritmo.....	49
5.3. Verificação dos valores de <i>gap</i> dos resultados obtidos com o algoritmo .....	54
5.4. Alterações de parâmetros e do tempo de execução do algoritmo .....	56
5.5. Considerações sobre a aplicação da heurística de Algoritmos Genéticos .....	59
<b>6. CONCLUSÕES</b> .....	61
6.1. Pesquisas Futuras .....	64
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	66
<b>8. APÊNDICES</b> .....	69

## 1. INTRODUÇÃO

O início da Teoria Moderna de Portfólio ocorreu na década de 1950 com Markowitz (1952), com a proposta de um modelo de média-variância para seleção de carteira. Nesse modelo, a variância dos retornos dos ativos é tomada como medida de risco, e é feita a otimização do *trade-off* entre risco e retorno esperado. Com a evolução das propostas sobre formação de portfólios, verifica-se que há basicamente dois tipos de fundos de investimentos: fundos de administração ativa e fundos de administração passiva (Roll, 1992; Beasley, Meade e Chang, 2003; e Maringer e Oyewumi, 2007, entre outros).

Fundos de administração ativa possuem, em geral, gestão mais flexível e maior quantidade de operações de compra e venda. Com isto, tendem a apresentar maiores custos de gestão e de transações (associados com maior volume de operações). Estratégias ativas assumem que os mercados não são perfeitamente eficientes, e buscam explorar desequilíbrios resultantes dessa ineficiência.

Já fundos de administração passiva caracterizam-se por uma gestão menos flexível e normalmente seguem determinados critérios previamente definidos (por exemplo, a tentativa de reprodução do rendimento de um determinado índice do mercado). São fundos que tendem a apresentar custos de gestão e transação menores, pois envolvem menor quantidade de operações e alterações na carteira, e constituem essencialmente uma estratégia de investimento de médio e longo prazo. A fundamentação teórica desses fundos está na hipótese de que os mercados são eficientes e seus rendimentos não podem ser superados no longo prazo (e, por consequência, os rendimentos dos índices de referência dos mercados também não podem ser superados).

A eficiência de mercados foi proposta por Fama (1970). Considera-se que um mercado de capitais é eficiente em termos de informação se ele cumpre a prerrogativa de utilizar todas as informações disponíveis no processo de definição dos preços dos ativos. Basicamente, nesse caso, cada indivíduo analisa as informações que estão disponíveis e toma suas decisões como uma resposta a essas informações (considerando também as preferências individuais). Assim, os preços de mercado incorporam essas informações disponíveis de variadas formas e, portanto, refletem todas as informações disponíveis no mercado (pois os preços de mercado representam o conjunto das decisões individuais).

Dessa forma, fundos de investimento de *index tracking* (também conhecidos como “*index funds*”, ou “Fundos de Índice” no Brasil) são uma forma de investimento passivo, em que o objetivo é reproduzir a evolução do rendimento de um determinado índice de mercado (por exemplo: índice Ibovespa). Além disso, a estratégia de formação de carteiras de *index tracking*

pode ser usada para composição de um ETF (*Exchange-Traded Fund*, que se trata de um tipo de ativo que tem se popularizado fortemente nos últimos anos) ou, ainda, para tentativa de reprodução de um indicador de mercado (como inflação).

A primeira opção para se aplicar a estratégia de *index tracking* seria realizar uma réplica total do índice, ou seja, compor um portfólio com os mesmos ativos que formam o índice e seus respectivos pesos. Porém, essa estratégia apresenta claras desvantagens, principalmente em relação a custos de transação, que serão maiores devido à grande quantidade de ativos que será adicionada ao portfólio (o que implicará maiores custos na medida em que a carteira sofrerá rebalanceamentos ao longo do tempo); além disso, outra desvantagem é que serão atribuídos pesos bastante baixos a alguns ativos da carteira, implicando em custos que poderiam ser evitados (Barro e Canestrelli, 2009; e Canakgoz e Beasley, 2009, entre outros autores).

Com essas desvantagens de se realizar uma réplica total do índice, os modelos de *index tracking* apresentam comumente, entre outras restrições, a limitação da quantidade de ativos que podem compor a carteira. O objetivo principal é reduzir os custos com compra, venda e gestão de ativos, através de limitação das transações executadas para formar o portfólio. Porém, nesse caso, dificilmente se obtém a reprodução exata do índice (Barro e Canestrelli, 2009). Devido a essa dificuldade de reprodução exata de um índice, pretende-se realizar nessa dissertação a aplicação de um modelo de otimização que tem como objetivo minimizar o erro (diferença) do rendimento da carteira formada em relação ao rendimento do índice (minimização do *tracking error*); e, para limitar os custos de transação e gestão, é utilizada uma restrição de controle da quantidade de ativos que compõem o portfólio.

### 1.1. Objetivos

Diante das questões expostas acima, a dissertação que está sendo proposta terá como objetivo central contribuir com a literatura de *index tracking* em dois pontos centrais:

- i. Apresentar um modelo de *index tracking* a partir da função objetivo de aproximação de retorno apresentada por Gaivoronski, Krylov e van der Wijst (2005), para formação de carteiras com um número reduzido de ativos;
- ii. Propor um modelo heurístico baseado em Algoritmos Genéticos para solução do modelo com controle no número de ativos.

Ou seja, busca-se inicialmente apresentar uma modelagem para o problema de *index tracking*, para em seguida realizar o uso de um método de heurística para solução do modelo. O objetivo da heurística é possibilitar a obtenção de soluções de forma mais rápida especialmente para formação de carteiras com quantidades pequenas de ativos.

Será utilizado, no decorrer da dissertação, o índice Ibovespa como referência, com uso de uma amostra de 67 ativos (sendo que o índice possui atualmente 69 ativos, porém dois ativos foram retirados da amostra por falta de dados no período selecionado). Foi adotada uma amostra de dados para o período de janeiro/2009 a julho/2012, e serão realizados testes para formação de seis tipos de carteiras: carteiras sem limite de quantidade de ativos e com limites de 40, 30, 20, 10 e 05 ativos (com intervalos de rebalanceamento de 20, 40 e 60 dias úteis, ou seja, basicamente rebalanceamento mensal, bimestral e trimestral).

Os testes discutidos no Capítulo 4 demonstram que, para formação de carteiras com limite de 20 ativos, o tempo computacional necessário para solução próxima à otimalidade extrapola 1 hora de processamento com uso do *solver* Cplex. Por esse motivo, a partir dos dois objetivos centrais expostos acima, o objetivo específico da dissertação é aplicar uma heurística que possibilite formar carteiras com 05 e 10 ativos com tempo de processamento máximo em torno de 10 minutos (em linha com os tempos de solução de Beasley, Meade e Chang, 2003).

Stoyan e Kwon (2010) apresentam pontos relevantes para construção de portfólios de *index tracking*. No presente trabalho, serão abordados alguns desses pontos, tais como (i) objetivo de minimizar o erro entre rendimento do índice e portfólio, (ii) formar portfólios que tenham quantidade reduzida de ativos (em relação à quantidade de ativos que compõe o índice), e (iii) tratar de controle e minimização de custos de transação, o que será feito através da delimitação de quantidade de ativos que podem compor as carteiras e do aumento do período de rebalanceamento, o que também altera os custos de transação. Trata-se de objetivos secundários que também deverão ser cumpridos pelo modelo.

Porém, esses mesmos autores citam outros pontos relevantes para modelos de *index tracking* que não serão abordados. Dentre esses fatores, pode-se mencionar principalmente a aplicação de restrições no rebalanceamento entre as carteiras (através, por exemplo, de verificação e limitação de *turnover*), como forma de controle de custos de transação; outra importante questão que pode ser mencionada refere-se à limitação de pesos mínimo e máximo para cada ativo adicionado ao portfólio. Essas são questões relevantes que, ao serem aplicadas,

contribuem para uma maior limitação do espaço de soluções do problema, dificultando ainda mais a sua solução.

Como esses fatores não serão incluídos no modelo de otimização a ser proposto na dissertação, acabam se constituindo em limitações da pesquisa realizada. Apesar disso, essas questões são amenizadas por outros fatores. O fato de não haver controle de rebalanceamento (o que seria uma ferramenta de controle de custo) é parcialmente compensado com a limitação da quantidade de ativos das carteiras a ser formadas, visto que essa limitação tem objetivo claro de limitar os custos de transação e gestão.

Há ainda outras limitações da pesquisa a ser realizada, como o uso de apenas um software como *solver* (pacote de otimização a ser utilizado nos processos de solução do problema – nessa dissertação, será utilizado o *solver* Cplex). O uso de outros softwares de otimização seria benéfico no sentido de se verificar possíveis discrepâncias principalmente em relação a tempo de solução. Porém, essa é uma questão que não será incluída no presente trabalho, considerando que o Cplex trata-se do *solver* atualmente mais difundido e utilizado comercialmente.

## 1.2. Justificativa

A adoção de fundos de investimento de *index tracking* justifica-se essencialmente por constituir uma forma de investimento que, baseando-se na hipótese de mercados eficientes, tende a apresentar resultados superiores que fundos de investimento ativos. Frino e Gallagher (2001), por exemplo, demonstram que investimentos em estratégias passivas apresentam, em média, melhor desempenho do que investimentos em estratégias ativas no longo prazo. Fama e French (2010) também realizaram um estudo que demonstra que fundos ativos de investimento têm tendência a ficar abaixo do rendimento de índices do mercado.

Ou seja, com base na literatura existente, pode-se perceber a importância de estratégias de investimento passivas, em que a modelagem de otimização de *index tracking* pode ser aplicada para formação de fundos de índice. Além disso, modelagens de otimização de *index tracking* podem também ser utilizadas para formação dos chamados ETFs (*Exchange-Traded Funds*), que se tratam de ativos negociados no mercado de ações e que representam fundos de índice. Uma breve revisão literária que corrobora o forte crescimento de ETFs no mercado financeiro nos últimos anos pode ser encontrada, por exemplo, em Borges, Eid Jr. e Yoshinaga (2012). Por fim, também é possível que seja utilizado um modelo de *index tracking* para formação de carteiras que busquem seguir um determinado indicador, como a inflação, o que ajuda a corroborar a

relevância de se desenvolver modelos de otimização para realização de *tracking* de um índice ou outro indicador de mercado, como a que é apresentada na presente dissertação. Devido a limitações de acesso a bases de dados mais amplas, a presente dissertação lidará apenas com o índice brasileiro Ibovespa, considerando a impossibilidade de se obter dados de ações de índices de maior tamanho como, por exemplo, o índice americano S&P 500.

E, a partir da proposta de um modelo de *index tracking*, é necessária a adoção de um método de heurística para resolver a otimização, pois o tempo computacional necessário para solução desse problema na otimalidade (ou próximo da otimalidade) é bastante elevado para formação de carteiras de número reduzido de ativos. Trata-se de um modelo que não converge para a resposta ótima; conforme afirmam Coleman, Li e Henniger (2006), esse problema possui complexidade computacional *NP-hard*, com crescimento exponencial do tempo necessário para solução na otimalidade na medida em que se diminui a quantidade limite de ativos da carteira. Como diferencial, nessa dissertação, será realizada uma análise dos resultados da heurística verificando-se os *gaps* das soluções, o que não é feito de forma recorrente pela literatura, conforme citado posteriormente no Capítulo 2.

Portanto, o objetivo central da heurística é possibilitar a obtenção de respostas para formação de portfólios especialmente com quantidades menores de ativos com tempo computacional mais baixo (carteiras de 10 e 05 ativos, para o caso da amostra de dados adotada). A necessidade de se resolver o problema com um tempo reduzido deve-se ao entendimento de que, em uma situação real de uma empresa que aplica a modelagem de *index tracking*, decisões sobre a gestão da carteira devem ser tomadas com certa velocidade. Certamente, deve-se considerar que não é possível que um modelo seja resolvido aplicando-se apenas otimização direta via um *solver* como o Cplex, pois esse problema demoraria diversas horas para ser solucionado. Em uma situação de alteração da carteira virtual de um índice, por exemplo, normalmente a nova composição desse índice só será definitivamente conhecida pelo mercado apenas no dia anterior ao início da vigência das alterações. Dessa forma, com o rebalanceamento do índice, a tomada de decisão sobre o rebalanceamento da carteira de *index tracking* também necessita ocorrer de forma ágil.

Assim, com o uso de uma heurística que resolva o modelo de otimização com rapidez (em torno de 10 minutos, por exemplo), é possível que a carteira de *index tracking* possa ser alterada com maior facilidade e agilidade. Além disso, pode ser necessária a realização de diversos testes para a tomada de decisão sobre o rebalanceamento, para se verificar, por exemplo, se é preferível que a nova carteira de *tracking* seja formada com pouca quantidade de ativos, ou com quantidades um pouco maiores, com base nas previsões de erro de *tracking* futuro.

O método heurístico selecionado para desenvolvimento da dissertação será de Algoritmos Genéticos. Essa escolha baseia-se essencialmente nas boas soluções obtidas por autores que utilizaram esse método para o problema de *index tracking* (como Beasley, Meade e Chang, 2003; Oh, Kim e Min, 2005; e Jeurissen e van den Berg, 2008). Na revisão bibliográfica, é apresentada uma discussão um pouco maior destes estudos.

### 1.3. Organização da Dissertação

No Capítulo 2, é apresentada uma revisão literária de trabalhos referentes ao tema de *index tracking*, com ênfase na análise de artigos contemporâneos sobre o tema. Na seção 2.1, constam exemplos de estudos utilizando diversos métodos de solução, com uso de programação quadrática e inteira, aplicação de método de cointegração e utilização de alguns métodos heurísticos, principalmente algoritmos “evolucionários”. Pode-se perceber, com base nessa literatura contemporânea, que se trata de um problema bastante trabalhado por diversos autores e que pode ser solucionado de diferentes formas. Já na seção 2.2, são citados alguns exemplos de uso do método heurístico de Algoritmos Genéticos para solução de *index tracking*; trata-se de referências relevantes para o desenvolvimento dessa dissertação. E, na seção 2.3, é realizado um breve resumo das análises bibliográficas realizadas.

No Capítulo 3, é apresentado inicialmente, na seção 3.1, o modelo de otimização a ser utilizado. E na seção 3.2, são feitas as descrições da base de dados adotada e da metodologia de aplicação dos testes computacionais.

O Capítulo 4 apresenta a forma de aplicação do modelo para solução direta com uso do *solver* de otimização Cplex, a fim de verificar as condições da modelagem para solução próxima à otimalidade. Na seção 4.1, são apresentados os testes iniciais realizados a fim de se verificar a qualidade do modelo de otimização e também, de se verificar as limitações de solução do modelo em termos de tempo computacional com uso direto do Cplex. Após isso, as seções 4.2 e 4.3 apresentam os resultados para novos testes realizados com alteração do tamanho da amostra de dados *in-sample* e com alterações em parâmetros de configuração do *solver*, de forma a se verificar a possibilidade de resolver o modelo de otimização em um tempo menor. A seção 4.4 apresenta um breve resumo das conclusões obtidas com os testes realizados.

Já no Capítulo 5, é feita a descrição da aplicação do método heurístico de Algoritmos Genéticos. A seção 5.1 apresenta as principais características desse método, adaptando-o ao problema de *index tracking*. Na seção 5.2, são apresentados os resultados iniciais para aplicação



da heurística. E, na seção 5.3, são apresentados os resultados para novos testes computacionais realizados com alterações de parâmetros da heurística. Na seção 5.4, é feita uma breve descrição das conclusões obtidas.

E, finalmente, o Capítulo 6 apresenta as conclusões obtidas com a realização da pesquisa, além de sugestões de pesquisas futuras sobre o tema. O Capítulo 7 apresenta as Referências bibliográficas utilizadas, e o Capítulo 8 trata-se de um Apêndice com informações sobre os testes computacionais realizados.

## 2. REVISÃO BIBLIOGRÁFICA

Para melhor apresentação das referências bibliográficas, este capítulo está dividido em duas partes: na seção 2.1, são apresentados alguns exemplos de trabalhos a respeito de *index tracking*, que buscam solucionar esse problema de formas variadas. Na seção 2.2, são apresentados alguns exemplos de artigos com uso específico de Algoritmos Genéticos para solução do problema. E, finalmente, a seção 2.3 apresenta algumas considerações principais a respeito da revisão bibliográfica realizada.

### 2.1. Otimização de portfólios de investimento de *Index Tracking*

O problema de *index tracking* é apresentado na literatura contemporânea com modelos que empregam diferentes métodos de solução como, por exemplo, uso de programação quadrática, cointegração e diversas técnicas de heurística, entre outras formas.

Gaivoronski, Krylov e van der Wijst (2005) apresentam diferentes abordagens para tratar do problema de *index tracking*, incluindo questões relativas à análise de risco do portfólio (relação com o erro entre carteira e índice), rebalanceamento das carteiras (análise de portfólios dinâmicos) e restrição de custos de transação. Os autores discutem algumas abordagens do problema, realizam análises sobre diferentes medidas de risco e desempenho e também sobre rebalanceamento do portfólio (com incorporação de novas informações ao modelo após a formação da carteira inicial). Diversos experimentos foram realizados pelos autores com dados da Oslo Stock Exchange, e foi verificado, principalmente, que o desempenho de *tracking* das carteiras geralmente é melhor quanto maior é o período de dados utilizado para a otimização (dados *in-sample*); além disso, a qualidade de *tracking* das carteiras é influenciada pela quantidade de ativos incluídos na carteira e pela forma de realização de rebalanceamento do portfólio ao longo do tempo.

Diversos artigos apresentam uso de métodos heurísticos de “evolução diferencial”, em que se baseia a proposta de algoritmos genéticos. Por exemplo, pode-se citar Maringer e Oyewumi (2007), que apresentam uma modelagem do problema de *index tracking* com controle da quantidade de ativos a ser incluídos na carteira e também com controle de peso máximo de cada ativo, e aplicam um método de heurística de “evolução diferencial”. Testes numéricos foram realizados com o índice Down Jones Industrial Average. Os testes permitiram concluir que há tendência de que sejam obtidos menores erros de *tracking* com aumento da quantidade de ativos da carteira, sendo que essa melhora no erro é decrescente na medida em que se aumenta a

quantidade de ativos. Scozzari et al (2012) também propõem um modelo baseado em heurística de evolução (Evolução Diferencial e Busca Combinatória – DECS); é utilizada formulação de programação quadrática e inteira, com restrições que seguem as normas UCITS (Undertaking for Collective Investments in Transferable Securities, conforme definições da European Union Directive). Nas restrições utilizadas, há o controle do *turnover* entre carteiras (definição de estratégia de rebalanceamento) e o controle do limite de percentual que pode ser investido em cada ativo (na medida em que se diminui a quantidade máxima de ativos que podem formar cada carteira, essa restrição tende a ser mais exigida de forma a evitar “sobrecarga” de investimento em poucos ativos).

Outra proposta de heurística é feita por Krink, Mittnik e Paterlini (2009), que utilizam dois métodos de forma conjunta: heurística de “evolução diferencial” e busca combinatória (formando um algoritmo híbrido), buscando obter soluções mais rápidas sem perder eficiência em relação às soluções ótimas. A modelagem proposta inclui controle de rebalanceamento e de custos de transação (restrições não lineares). Foram realizados testes com dois índices de referência (Dow Jones 65 e Nikkei 225), com uso de duas funções objetivo diferentes. Os autores concluíram que o fornecimento de boas soluções iniciais ao algoritmo pode auxiliar no encontro de melhores resultados – embora isso não seja fator determinante. Guastaroba e Speranza (2012) tratam dos problemas de *index tracking* e também de *enchanced index tracking*, propondo uma formulação integrada com uso de programação linear e inteira (*mixed-integer linear programming* - MILP) e uma heurística denominada “Kernel Search” para solução do problema. Em relação ao problema de *index tracking*, os autores empregaram um modelo de minimização do erro de *tracking*, com restrições de custos de transação, controle de valores mínimo e máximo a ser investidos em cada ativo, consideração de custos fixos e controle de rebalanceamento. Foram utilizados oito índices para os testes computacionais, com amostras envolvendo entre 31 e 2.151 ativos, com controle dos tempos computacionais de solução (que variam de forma acentuada, entre poucos segundos para instâncias menores até 5.400 segundos para instâncias maiores).

Corielli e Marcellino (2006) apresentam um exemplo de aplicação de modelos fatoriais para o problema de *index tracking* (uso de métodos estatísticos). Os autores assumem inicialmente que os preços dos ativos têm movimentos direcionados por um modelo fatorial, considerando a existência de propriedades dinâmicas dos componentes do índice e também a estrutura de pesos dos ativos para formação do índice. A abordagem considera que o índice e a carteira de *index tracking* possuem a mesma estrutura de fatores, e apresentam uma forma de construção da carteira de *tracking* a partir de determinadas características desses fatores e de suas respectivas

correlações. A abordagem foi aplicada utilizando o índice EuroStoxx50 como referência, sem análise de tempos computacionais para os testes realizados.

Outra abordagem dada ao problema de *index tracking* envolve uso de cointegração. Dunis e Ho (2005) utilizam método de cointegração para *index tracking* por esse método incorporar relações de longo prazo entre séries. Conforme ressaltam os autores, métodos quantitativos tradicionais de formação de carteiras baseiam-se em análise de correlação; porém, como correlações possuem mudanças constantes no decorrer do tempo, essa metodologia torna-se mais exigente de rebalanceamentos, o que implica maiores custos. Em contraste, cointegração ameniza a necessidade de rebalanceamentos constantes ao medir “comovimentos” de longo prazo nos preços dos ativos, em vez de utilizar os retornos dos ativos. Foram realizados testes para rebalanceamento com intervalos de um mês, três meses, seis meses e um ano. Os autores consideram que os resultados obtidos demonstraram forte cointegração entre as carteiras formadas e o índice, sendo de melhor resultado a estratégia de rebalanceamento a cada três meses.

Caldeira e Portugal (2010) também se baseiam em cointegração; conforme esses autores, carteiras definidas a partir de análise de média-variância apresentam *tracking error* não estacionário, podendo sofrer variações mais acentuadas, o que exigiria mais rebalanceamentos. Dessa forma, sugerem uso de cointegração para formação de carteiras para *index tracking* e também para estratégia de investimento *long-short*, para formação de carteiras mais estáveis. É utilizada cointegração dos preços dos ativos para desenvolvimento e aplicação da metodologia proposta, pois o uso de cointegração possibilita a percepção de tendências estocásticas nas séries de preços dos ativos, formando carteiras mais estáveis. Os autores apresentaram resultados satisfatórios para formação de carteiras de *index tracking* com rebalanceamento mensal, comprovando a efetividade dos resultados.

Yu, Zhang e Zhou (2006) propõem um modelo de *index tracking* consistente com a modelagem tradicional de média-variância de Markowitz, utilizando *downside risk* como medida de risco. Nesse trabalho, o *tracking* de um índice relaciona-se com limitar a probabilidade de que o retorno do portfólio tenha baixas além do retorno do índice (que é utilizado como *downside risk*), e aplicam as condições de otimalidade Karush-Kuhn-Tucker para obter a forma analítica da carteira ótima. Já Yao, Zhang e Zhou (2006) apresentam uma proposta de modelagem para simular o rendimento de um índice ou também de uma determinada taxa de crescimento constante. Para ambos os casos, o objetivo central é formar uma carteira com poucos ativos. Porém, nessa abordagem, os ativos componentes da carteira são parâmetros exógenos, ou seja, não são determinados durante a otimização. Com o uso de técnica SLQ (*Stochastic Linear Quadratic Control*), a solução para a otimização é obtida através de uso de programação SDP (do

inglês, *semidefinite programming*). O período de dados *in-sample* é de 60 observações diárias, e os testes realizados envolvem rebalanceamento do portfólio em intervalos diários, de cinco dias e de vinte dias. Para o caso de *intex tracking*, os testes adotaram portfólio com cinco ativos para otimização (ativos pré-determinados aleatoriamente, com o índice S&P 500 como referência). Os autores concluem que o desempenho do *tracking* do índice é eficiente independente de o mercado apresentar movimentos de alta ou baixa, apesar do curto período de dados *in-sample* e da forma de seleção dos ativos da carteira.

O uso de um algoritmo de decomposição específico para SMIP (do inglês, *Stochastic Mixed-Integer Programming*) é feito por Barro e Canestrelli (2009) e Stoyan e Kwon (2010). Barro e Canestrelli (2009) apresentam um estudo de problemas de *index tracking* multiestágios. São apresentadas algumas propostas de modelos estáticos e dinâmicos, com foco para desenvolvimento de uma modelagem dinâmica com restrição de custos de transação e introdução de componente de liquidez no portfólio (com controle de *turnover* entre as carteiras – o que corresponde a uma estratégia de rebalanceamento). Para solução, é utilizada uma abordagem de "dupla decomposição" do problema, dividindo-o em dois subproblemas menores e mais simples de ser solucionados (um dos problemas recebe o componente estocástico do modelo; e o outro, o componente dinâmico). Stoyan e Kwon (2010) também apresentam uma proposta de modelagem com SMIP, de uma forma similar a Barro e Canestrelli (2009), com uso de restrição para controle da quantidade de ativos e controle do rebalanceamento dos portfólios; a solução é derivada igualmente com algoritmo de decomposição, dividindo o problema em partes menores para diminuição do tempo de solução.

Chen e Know (2012) buscam demonstrar o desenvolvimento de um modelo de otimização robusta para seleção de carteira. O uso de otimização robusta é justificado pela possibilidade de se gerar soluções ótimas para *worst-case scenarios*, dadas algumas condições de incerteza que podem ser incluídas no modelo; outra vantagem é a possibilidade de controle do grau de "conservadorismo" do modelo. O trabalho possui foco em modelos de um único estágio, sem rebalanceamento. Testes computacionais foram realizados com o índice S&P 100. Segundo os autores, essa abordagem apresenta algumas vantagens em relação a propostas alternativas do problema de *index tracking*, como o fato de utilizar programação linear e uma modelagem que permite certa flexibilidade na definição de pesos dos ativos na carteira.

O uso de otimização quadrática com reduzido número de ativos para o problema de *index tracking* é feito, por exemplo, em Jansen e van Dijk (2002); porém, nesse trabalho, a limitação de quantidade de ativos não é feita através de restrição. Os autores propõem uso de função quadrática, porém sem restrição de quantidade de ativos. A partir de um conjunto inicial de ativos

(exemplo: 100 ativos, caso do índice S&P 100 utilizado pelos autores), é formada uma carteira sem limitação de tamanho; com isso, são realizadas diversas otimizações em sequência diminuindo a amostra de ativos, utilizando-se sempre os ativos de maior peso para realização da otimização seguinte, até a formação de uma carteira de tamanho previamente definido (exemplo: 20 ativos). Como alternativa, os autores propõem um “método de diversidade”, em que a função objetivo quadrática é formada de modo a criar um *trade-off* entre *tracking error* e diversidade do portfólio (sendo a diversidade expressada pela quantidade de ativos do portfólio, determinada por uma constante incluída na função objetivo). Coleman, Li e Henniger (2006) também empregam programação quadrática e inteira; a proposta de modelagem para *index tracking* utiliza função objetivo para minimização do erro e função linear "descontínua" na restrição de quantidade de ativos. Considerando que o problema possui complexidade computacional NP-hard, os autores propõem um método de solução similar ao de Jansen e van Dijk (2002), com execução de otimizações em sequência diminuindo a amostra de ativos a cada otimização, até a obtenção de uma carteira final de tamanho previamente estabelecido.

## 2.2. Uso de Algoritmos Genéticos para o problema de *Index Tracking*

A literatura sobre *index tracking* apresenta o uso de alguns métodos heurísticos para solução do problema, devido à sua grande complexidade computacional. O principal objetivo do uso de heurísticas é obter soluções de forma mais ágil sem perda de qualidade especialmente na formação de carteiras de número reduzido de ativos. Como pôde ser verificado acima, um método heurístico bastante recorrente é o de evolução diferencial (considerando que o método de algoritmos genéticos se trata de uma heurística baseada na mecânica de seleção natural e evolução diferencial de populações), utilizado por autores como Maringer e Oyewumi (2007), Scozzari et al (2012) e Krink, Mittnik e Paterlini (2009). Outros autores que fazem uso de Algoritmos Genéticos são Beasley, Meade e Chang (2003), Oh, Kim e Min (2005) e Jeurissen e van den Berg (2008).

Em relação ao uso de Algoritmos Genéticos (AG) para solução do problema de *index tracking*, Beasley, Meade e Chang (2003) e Oh, Kim e Min (2005) são dois trabalhos bastante referenciados na literatura contemporânea sobre esse assunto. Beasley, Meade e Chang (2003) apresentam uma proposta de formulação de *index tracking* com restrição de custos de transação e de rebalanceamento; os autores desenvolvem um método chamado de “heurística populacional” baseado em AG para solução do problema, considerando sua complexidade e necessidade de se obter soluções de forma mais rápida. Foram realizados diversos testes com diferentes índices e quantidades de ativos para testar a heurística em diferentes condições. Com uso de amostra *in-*

*sample* de 290 observações, foram utilizados os índices Hang Seng (amostra de 31 ativos usados), DAX (amostra de 85 ativos usados), FTSE (89 ativos), S&P 100 (98 ativos) e Nikkei (225 ativos). Nos testes, foram formadas carteiras de 10 ativos para cada índice. O tempo de execução médio do algoritmo ficou em 6,7 minutos (sendo que o menor tempo foi 1,7 minutos e o maior foram 15,7 minutos).

Oh, Kim e Min (2005) desenvolveram um AG para solução de *index tracking* separando o algoritmo em basicamente duas partes. O primeiro passo do algoritmo trata de selecionar os ativos que irão compor o portfólio. A seleção é feita com base em indicadores econômicos das empresas; para essa análise, as empresas são divididas em blocos de acordo com seus respectivos setores industriais. Após ter sido feita a composição da carteira, a segunda parte do algoritmo trata de realizar a minimização de uma função objetivo utilizando como amostra apenas os ativos selecionados para compor a carteira. Os resultados do AG são confrontados com os resultados da aplicação de um modelo convencional de minimização de erro. Testes empregaram análise do índice KOSPI 200 (referência da Korea Stock Exchange) para o período de Janeiro/1999 a Dezembro/2001. A eficiência do AG é demonstrada via análise dos valores de erro de *tracking* dos portfólios gerados pelo algoritmo, que em média são menores que os erros dos portfólios obtidos pelo método convencional.

Outro exemplo de uso de AG é Jeurissen e van den Berg (2008). Nesse caso, são utilizados, em combinação, o software Matlab (para codificação do algoritmo, com uso de função quadrática) e o software Microsoft Excel (para verificar o desempenho de cada carteira formada através do cálculo do *tracking error*, que corresponde ao *fitness* da carteira). Os autores realizaram testes com o índice Dutch AEX, buscando formar carteiras de 10 ativos a partir da amostra de 25 ativos componentes do índice. Uma das conclusões principais é que, dependendo dos parâmetros adotados para solução do AG, há a tendência de que o algoritmo encontre a solução ótima com no máximo 23 gerações no caso da amostra utilizada.

### 2.3. Considerações sobre a Revisão Bibliográfica

Considerando a revisão literária realizada, nota-se a relevância do uso de algoritmos evolucionários e algoritmos genéticos para solução do problema de *index tracking* (Beasley, Meade e Chang, 2003; Oh, Kim e Min, 2005; Maringer e Oyewumi, 2007; Jeurissen e van den Berg, 2008; Canakgoz e Beasley, 2009; Krink, Mittnik e Paterlini, 2009; Scozzari et al, 2012).

Além disso, a análise mais detalhada de alguns desses estudos permitiu perceber-se que o método de algoritmos genéticos pode gerar resultados bastante satisfatórios, como no caso, por exemplo, de Jeurissen e van den Berg (2008); no estudo desenvolvido por esses autores, apesar a base de dados de ativos ser relativamente pequena, há a tendência de que o algoritmo encontre a solução ótima para o problema de minimização de erro de *tracking*.

Já Oh, Kim e Min (2005) realizam a aplicação de algoritmos genéticos de uma forma um pouco mais específica; nessa aplicação, antes da execução do algoritmo, é feita uma pré-seleção de ativos que poderão compor a carteira, partindo-se da base de ativos adotada. Após isso, o algoritmo é executado para se encontrar a melhor que pode ser formada com os ativos selecionados anteriormente. Os autores concluem a respeito da qualidade dos resultados em termos de erro de *tracking*, porém não realizam uma análise das soluções da heurística em termos de *gap* das soluções obtidas; dessa forma, não concluem sobre se as respostas obtidas são respostas ótimas para o problema.

Da mesma forma, foi possível verificar a qualidade das respostas obtidas por Beasley, Meade e Chang (2003). Pela análise da literatura, foi possível verificar que esse estudo trata-se de uma das referências mais citadas sobre o assunto, com um desenvolvimento detalhado de algoritmos genéticos. Esses autores apresentam uma otimização mais complexa, com inclusão de custos de transação na análise, e a qualidade das respostas é verificada pelos valores de erro de *tracking* e também de tempo de solução. Porém, novamente, esses autores também não verificam o *gap* das soluções obtidas em relação às soluções ótimas. Sobre os tempos de solução, os autores solucionaram o problema de *index tracking* com rapidez, usando em média em torno de 10 minutos de processamento. A velocidade de obtenção das soluções, combinada com os baixos valores de erro, demonstra novamente que o método de algoritmos genéticos pode ser bastante eficiente.

Assim, com a análise dessa revisão literária, decidiu-se pelo uso de Algoritmos Genéticos para solução do modelo proposto na presente dissertação em desenvolvimento. Um estudo mais amplo poderia ser feito no sentido de se resolver o mesmo modelo de otimização utilizando diferentes métodos heurísticos e estatísticos, além de algoritmos genéticos. Nesse caso, isso se constitui em uma limitação dessa dissertação, ao não realizar a comparação de diferentes métodos para concluir sobre quais métodos poderiam ser mais adequados ao problema. Apesar disso, a literatura aponta para o fato de que a aplicação de algoritmos genéticos é adequada ao problema e pode gerar resultados bastante satisfatórios.



### 3. METODOLOGIA

O problema de *index tracking* possui um espaço de solução bastante largo e de crescimento exponencial. Por exemplo, no caso de se realizar *tracking* do índice Ibovespa (composto por 69 ativos) com uma carteira de 20 ativos, a quantidade de portfólios que pode ser formada fica em torno de  $[69!/(20! \times 49!)] = 116 \times 10^{15}$ . Por esse motivo, diversos autores utilizam heurísticas como forma de solução do problema para carteiras com baixa quantidade de ativos (por exemplo, 05 ou 10 ativos no caso de se utilizar o Ibovespa). A alta complexidade aponta para a dificuldade de se obter soluções ótimas devido ao tempo necessário para solução do modelo.

Para demonstrar a formulação do problema e o modo como será aplicado, são apresentados inicialmente a seguir o modelo de otimização (seção 3.1) e a descrição da base de dados adotada (seção 3.2). Para desenvolvimento do modelo, será utilizada a seguinte notação:

- $x_i$  = variável de peso do ativo  $i$  no portfólio, sendo  $0 \leq x \leq 1$ ;
- $z_i$  = variável binária para cada ativo  $i$ ;
- $r_{i,t}$  = parâmetro: retorno do ativo  $i$  no período  $t$ ;
- $R_t$  = parâmetro: retorno do índice no período  $t$ ;
- $t$  = parâmetro: momento de tempo;
- $T$  = parâmetro: número total de momentos de tempo;
- $K$  = parâmetro: limite de quantidade de ativos na carteira;
- $\vartheta$  = parâmetro: limite mínimo de erro entre carteira e índice em cada  $t$ ;
- $\theta$  = parâmetro: limite máximo de erro entre carteira e índice em cada  $t$ .

#### 3.1. Modelo de Otimização

O modelo de otimização utilizado nesse trabalho é elaborado a partir a função objetivo de aproximação de retorno proposta por Gaivoronski, Krylov e van der Wijst (2005). A função objetivo busca minimizar a variância da diferença entre o retorno do portfólio e do índice, conforme abaixo:

$$\text{Minimizar} \quad \frac{1}{T} \sum_{t=1}^T \left[ \left( \sum_{i=1}^I x_i \cdot r_{i,t} \right) - R_t \right]^2 \quad (1)$$

$$\text{S. t.} \quad \sum_{i=1}^I (x_i \cdot r_{i,t}) - R_t \geq \vartheta \quad , \forall t \quad (2)$$

$$\sum_{i=1}^I (x_i \cdot r_{i,t}) - R_t \leq \theta \quad , \forall t \quad (3)$$

$$\sum_{i=1}^I x_i = 1 \quad (4)$$

$$\sum_{i=1}^I z_i \leq K \quad (5)$$

$$x_i \leq z_i \quad , \forall i \quad (6)$$

$$x_i \geq 0 \quad , \forall i \quad (7)$$

$$z_i = 0; 1 \quad , \forall i \quad (8)$$

O objetivo do modelo é minimizar o erro de *tracking* da carteira formada em relação ao índice. O retorno do índice em cada momento ( $R_t$ ) é calculado a partir do valor médio do Ibovespa nos momentos  $t$  e  $t - 1$ ; da mesma forma, o retorno de cada ativo em cada momento de tempo ( $r_{i,t}$ ) é calculado a partir do preço médio de cada ativo, conforme abaixo:

$$R_t = \frac{\text{índice\_médio}_t}{\text{índice\_médio}_{t-1}} - 1 \quad \quad r_{i,t} = \frac{\text{preço\_médio}_{i,t}}{\text{preço\_médio}_{i,t-1}} - 1$$

As restrições (2) e (3) limitam a diferença de rendimento entre carteira e índice a um valor mínimo e a um valor máximo de erro para cada momento  $t$  – parâmetros  $\vartheta$  e  $\theta$  previamente definidos. A restrição (4) estabelece que 100% do valor disponível são aplicados no portfólio. As restrições (5) e (6), em conjunto, definem o limite de ativos que podem ser alocados na carteira (quantidade definida por  $K$ ). As restrições (7) e (8) definem, respectivamente, que o valor de cada  $x_i$  deve ser positivo (e menor ou igual a 1) e que  $z_i$  é variável binária (com valor 1 se o ativo  $i$  é incluído na carteira, e valor 0 caso contrário).

### 3.2. Base de dados e descrição do problema

Para realização dos testes computacionais, será adotado o índice Ibovespa. Os ativos selecionados são os 67 ativos componentes da carteira artificial do Ibovespa no período de maio/2012 a agosto/2012 (conforme Tabela 22 no Apêndice C) – considerando que a carteira artificial do índice possui 69 ativos; porém, dois ativos foram retirados por falta de dados nesse período (ativos excluídos: ALLL3 e OIBR3). Nos testes realizados, são formadas carteiras sem limite de quantidade de ativos e com limites de 40, 30, 20, 10 e 05 ativos. Após testes prévios já feitos, foi constatado que, para formação de carteiras com 20 ativos ou menos, não é possível obter-se soluções ao menos próximas das soluções ótimas (com *gap* inferior a 5%) com tempo de processamento inferior a uma hora. Assim, o objetivo da aplicação de um método heurístico é essencialmente buscar formar carteiras de 10 e 05 ativos em tempos abaixo de 1 hora, sem perda de qualidade da solução (soluções que sejam ao menos próximas das soluções ótimas, com *gap* abaixo de 10%).

Inicialmente, é adotado  $T = 150$  (período de análise *in-sample*). Trata-se de um período longo, em torno de sete a oito meses, já que cada mês possui, aproximadamente, 20 dias úteis. Assim, segue-se a proposta de Gaivoronski, Krylov e van der Wijst (2005) sobre o uso de períodos *in-sample* mais longos por apresentarem tendência a gerar melhores resultados; a adoção de  $T = 150$  implica que, para cada execução do modelo de otimização, são adotados os dados históricos das ações (retornos calculados a partir dos preços médios) nos 150 períodos de tempo anteriores ao período de tempo de projeção do resultado. A carteira formada a partir desse conjunto de dados será projetada sobre os retornos de  $n$  períodos de tempo subsequentes (neste estudo,  $n$  será igual a 20, 40 e 60). Ou seja, para formar a primeira carteira de investimento para  $n = 60$ , por exemplo, é usado intervalo de dados *in-sample* de  $1 \leq t \leq 150$ , e o portfólio formado é projetado sobre os retornos no intervalo de tempo fora da amostra de  $150 \leq t \leq 210$  (intervalo *out-of-sample* para projeção dos resultados em *rolling horizon*). A segunda carteira será formada com dados de  $61 \leq t \leq 210$ , e a projeção será feita em  $210 \leq t \leq 270$ , e assim sucessivamente. Os intervalos de rebalanceamento serão de 20, 40 e 60 observações (ou seja, tem-se atualização das carteiras por mês, bimestre e trimestre). Posteriormente, é adotado também  $T = 100$ , de forma a verificar como os resultados podem ou não ser influenciados pela alteração de intervalo de tempo *in-sample*.

A base de dados adotada é composta pelos preços médios diários dos ativos para o intervalo de 13/jan/2009 e 14/jul/2012, com um total de 890 observações diárias de preços. A

projeção dos resultados para os períodos fora da amostra (projeções *out-of-sample*) ocorre entre agosto/2009 e julho/2012. Serão formadas 36 carteiras de investimentos para  $n = 20$ , 18 para  $n = 40$  e 12 para  $n = 60$ .

Em princípio, os valores dos parâmetros  $\vartheta$  e  $\theta$  são definidos individualmente para cada carteira formada no caso das execuções do modelo de solução próxima à otimalidade (o que será feito para formação das carteiras de maior quantidade de ativos); são adotados sempre valores iguais em módulo, alterando apenas o sinal (ex.  $\vartheta = -0,002$  e  $\theta = 0,002$ ), buscando-se formar o menor intervalo possível entre esses dois parâmetros para que o modelo tenha solução. Com testes prévios já realizados, definiu-se inicialmente que a constante  $\vartheta$  varia entre  $-0,0065$  e  $-0,0015$ , e a constante  $\theta$  varia entre  $0,0015$  e  $0,0065$ . Nos testes para aplicação do método de Algoritmos Genéticos, os valores adotados são  $\vartheta = -0,01$  e  $\theta = 0,01$  (conforme descrito no Capítulo 5).

#### 4. OTIMIZAÇÃO COM APLICAÇÃO DIRETA DO SOLVER CPLEX

Com definição do modelo de otimização e base de dados a ser utilizada, o próximo passo consiste em realizar a otimização diretamente com o uso de um *solver* apropriado. Dessa forma, pode-se verificar inicialmente a funcionalidade do modelo; além disso, pode-se também constatar as limitações do modelo em termos de tempo computacional necessário para solução, de acordo com o limite de ativos para formação das carteiras.

Em relação à expectativa de aumento do tempo de processamento conforme é diminuída a quantidade de ativos para formação das carteiras, esse aumento de tempo é consequência da diminuição do espaço de possíveis soluções do modelo. Por exemplo, para o caso de formação de uma carteira de 30 ativos (considerando uma amostra de 69 ativos), pode-se formar um total de  $[69!/(30! \times 39!)] = 3,16 \times 10^{19}$  portfólios; para formação de carteiras de 20 ativos, pode-se formar  $1,16 \times 10^{17}$  portfólios; para carteiras de 10 ativos, pode-se formar  $3,40 \times 10^{11}$  portfólios; e, para carteiras de 5 ativos, pode-se formar  $1,12 \times 10^7$  portfólios. Ou seja, com diminuição da quantidade de ativos, o espaço de possíveis soluções torna-se mais restrito. Conforme instruções do Cplex, a solução de problemas que usam restrição linear e inteira (“*Mixed-Integer Programming*”) é feita através da aplicação de algoritmo de *branch-and-cut*, em que o *solver* forma diversos subproblemas para solucionar a otimização. Na medida em que o espaço de soluções torna-se mais restrito, aumenta a quantidade de subproblemas, e a árvore de soluções formada para solucionar o modelo é ampliada, exigindo maior uso de memória física e capacidade de processamento do computador. Por esse motivo, explica-se a expectativa de que a formação de carteiras de menor tamanho tende a exigir maior tempo de processamento.

Assim, a seção 4.1 apresenta os resultados iniciais da aplicação do modelo; na seção 4.2, tem-se os resultados para alteração do parâmetro  $T$ ; já na seção 4.3, são apresentados os resultados para testes com alterações nas configurações do *solver* adotado, a fim de se concluir se alterações no *solver* podem influenciar os tempos computacionais de solução. E, na seção 4.4, são apresentadas as considerações principais sobre testes iniciais com aplicação direta do Cplex.

Para realização de todos os testes de otimização, foi usado um computador com processador Intel® Core™ i7-3770 @ 3.40GHz, 8GB RAM e Windows® 7. Adotou-se também a linguagem de programação AMPL e *solver* (pacote de otimização) Cplex (versão 12.05). A escolha do *solver* Cplex justifica-se por ser um dos pacotes de otimização mais difundidos e utilizados comercialmente, além de ser adequado para programação quadrática e inteira.

#### 4.1. Testes computacionais iniciais com uso do *solver* Cplex

Os testes iniciais realizados buscam verificar a qualidade do modelo proposto e analisar os tempos computacionais necessários para formação de carteiras de maior quantidade de ativos. Considerando o objetivo de formar desde portfólios sem limite de ativos até com limite de apenas 05 ativos, esses testes foram realizados para confirmar a necessidade de uso de heurística para formação de carteiras menores (05 ou 10 ativos para o caso da amostra adotada), dada a complexidade do modelo (conforme já citado).

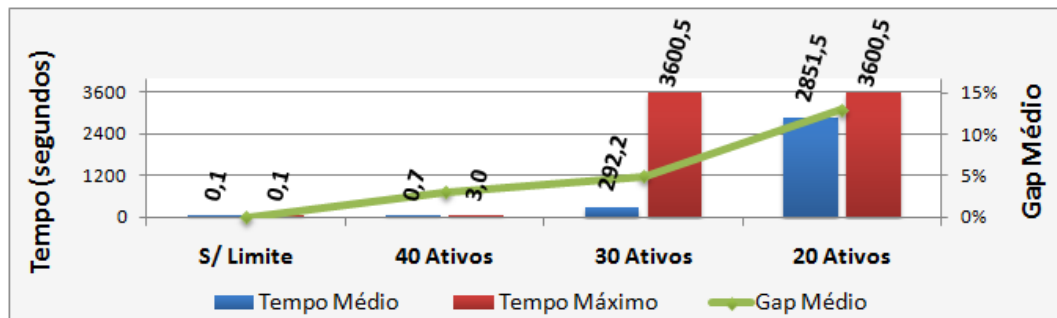
Foi estabelecido, para cada uma das otimizações (otimização de cada carteira via Cplex), tempo limite de execução do *solver* de 3.600 segundos (com meta de se obter *gap* máximo de 5% para cada carteira). O *gap* de resposta é estimado pelo *solver* Cplex e corresponde à diferença percentual entre a solução obtida e a solução ótima estimada (isto é, refere-se à tolerância da resposta obtida em relação à resposta ótima; em outras palavras, com *gap* de 5% por exemplo, temos 95% de confiança de que a resposta obtida é a resposta ótima). Ainda para esses testes iniciais, adotou-se o parâmetro  $T = 150$  (período de análise *in-sample*), de acordo com Gaivoronski, Krylov e van der Wijst (2005).

Os resultados são apresentados a seguir nas Figuras 01 a 03, com resumo dos tempos computacionais; nas Figuras 09 a 11 (Apêndice A), com as projeções das carteiras nos períodos fora da amostra; e nas Tabelas 01 a 07, com resumo dos resultados e estatísticas sobre as carteiras formadas. Os diversos resultados obtidos corroboram a qualidade do modelo de otimização, realizando o *tracking* do Ibovespa de forma satisfatória e com controle de custos (o que pode ser analisado através dos valores de *turnover*); além disso, as informações estatísticas (como volatilidade e valores de correlação) demonstram a forte relação entre as carteiras formadas e o índice Ibovespa.

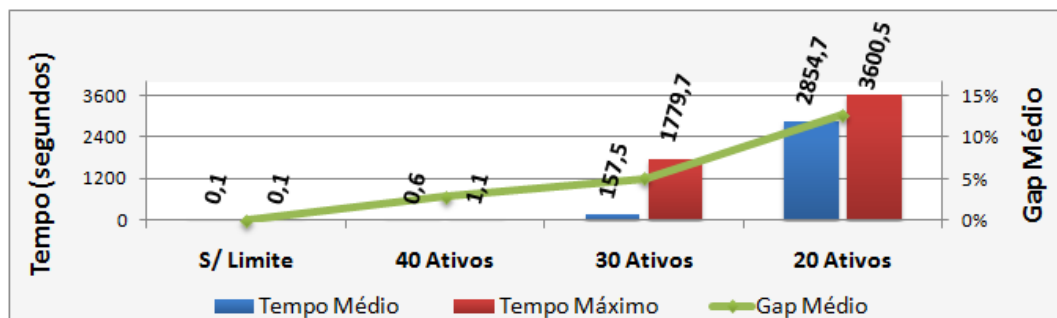
Em relação aos tempos de solução e valores de *gap* (conforme Figuras 01 a 03 e Tabelas 01 a 04 abaixo), foi confirmado que o problema deixa de ter condições de ser resolvido próximo da otimalidade com a diminuição da quantidade de ativos das carteiras. Nota-se, nas Figuras 01 a 03 que, já nas carteiras limitadas a 30 ativos (para rebalanceamento de 20 períodos), o tempo máximo de solução chega ao limite estabelecido de 1 hora, com o *gap* em torno de 5%. Com carteiras de 20 ativos, torna-se mais evidente a dificuldade de tempo de solução; para os três casos de rebalanceamento, o tempo médio fica acima de 2.500 segundos (mais de 40 minutos) e o *gap* médio permanece acima de 11%.

Assim, confirma-se que não é possível formar carteiras com menos de 20 ativos através de otimização direta via *solver* Cplex, pois os tempos de solução têm crescimento exponencial na

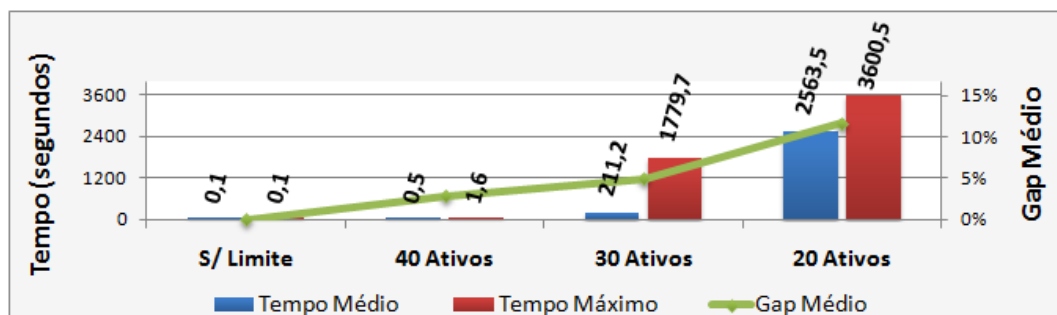
medida em que se diminui a quantidade limite de ativos das carteiras. Dessa forma, demonstra-se a necessidade de aplicação de uma heurística para tornar possível a obtenção de soluções para carteiras com menos de 20 ativos em tempos bastante inferiores a 1 hora e com soluções que apresentem *gap* ao menos abaixo de 10%.



**Figura 01:** Tempo Médio, Tempo Máximo e *Gap* Médio de solução das 36 carteiras geradas para Rebalanceamento Mensal (20 períodos). Fonte: elaborado pelo autor.



**Figura 02:** Tempo Médio, Tempo Máximo e *Gap* Médio de solução das 18 carteiras geradas para Rebalanceamento Bimestral (40 períodos). Fonte: elaborado pelo autor.



**Figura 03:** Tempo Médio, Tempo Máximo e *Gap* Médio de solução das 12 carteiras geradas para Rebalanceamento Trimestral (60 períodos). Fonte: elaborado pelo autor.

Deve-se considerar que, para formação de carteiras de 30 ativos, embora o tempo de solução médio esteja abaixo de 5 minutos, o tempo máximo atinge 1 hora para intervalo *out-of-sample* de 20 períodos e aproximadamente 30 minutos para intervalos de 40 e 60 períodos, o que são tempos bastante elevados. Porém, dado que o intervalo de rebalanceamento é elevado (implicando custos mais baixos, especialmente para intervalo *out-of-sample* de 60 períodos, como pode ser verificado pelos valores de *turnover* mensal) e que os valores de *gap* não ultrapassam significativamente 5% nas projeções de 40 e 60 períodos, pode-se tratar esses fatores como atenuantes do problema de tempo de solução. Porém, apesar disso, é necessário que seja aplicado um método de heurística para que as soluções possam ser obtidas de forma mais ágil para o caso de carteiras com menos de 20 ativos.

**Tabela 01:** Resultados para formação de Carteiras Sem Limite de Ativos com otimização via Cplex.

Carteiras Sem Limite de Ativos						
Intervalo out-of-sample	20 períodos		40 períodos		60 períodos	
Resultados	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap
Média	0.111	0.00%	0.110	0.00%	0.112	0.00%
Mínimo	0.094	0.00%	0.094	0.00%	0.094	0.00%
Máximo	0.125	0.00%	0.125	0.00%	0.125	0.00%
Desvio Padrão	0.008	0.00%	0.008	0.00%	0.009	0.00%
Tracking Error (período out-of-sample)						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	0.085%		0.085%		0.282%	
Mínimo	-1.044%		-2.148%		-1.643%	
Máximo	1.014%		1.510%		3.007%	
Desvio Padrão	0.584%		0.893%		1.253%	
Turnover Mensal						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	17.404%		11.438%		9.423%	
Mínimo	7.128%		6.869%		5.224%	
Máximo	36.902%		17.169%		14.542%	
Desvio Padrão	6.505%		2.962%		2.565%	

Fonte: elaborado pelo autor.



**Tabela 02:** Resultados para formação de Carteiras com Limite de 40 Ativos com otimização via Cplex.

Carteiras com Limite de 40 Ativos						
Intervalo out-of-sample	20 períodos		40 períodos		60 períodos	
Valores de saída	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap
Média	0.659	3.02%	0.568	2.83%	0.549	2.83%
Mínimo	0.109	0.00%	0.109	0.00%	0.109	0.00%
Máximo	3.026	5.22%	1.123	4.82%	1.560	5.11%
Desvio Padrão	0.488	1.66%	0.258	1.65%	0.395	2.03%
Tracking Error (período out-of-sample)						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	0.104%		0.092%		0.319%	
Mínimo	-1.051%		-2.148%		-1.643%	
Máximo	1.093%		1.510%		3.138%	
Desvio Padrão	0.589%		0.883%		1.247%	
Turnover Mensal						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	20.628%		12.928%		10.060%	
Mínimo	7.457%		7.025%		5.224%	
Máximo	38.515%		17.732%		14.542%	
Desvio Padrão	7.355%		2.965%		2.415%	

Fonte: elaborado pelo autor.

**Tabela 03:** Resultados para formação de Carteiras com Limite de 30 Ativos com otimização via Cplex.

Carteiras com Limite de 30 Ativos						
Intervalo out-of-sample	20 períodos		40 períodos		60 períodos	
Valores de saída	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap
Média	292.197	4.97%	157.519	5.01%	211.165	4.96%
Mínimo	0.515	3.77%	0.608	4.47%	0.515	3.94%
Máximo	3600.490	5.36%	1779.710	5.26%	1779.710	5.35%
Desvio Padrão	797.356	0.33%	411.909	0.17%	503.257	0.35%
Tracking Error (período out-of-sample)						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	0.079%		0.101%		0.418%	
Mínimo	-1.670%		-2.670%		-2.338%	
Máximo	1.268%		1.424%		3.274%	
Desvio Padrão	0.657%		0.927%		1.458%	
Turnover Mensal						
Out-of-sample	20 períodos		40 períodos		60 períodos	
Média	23.672%		14.425%		11.409%	
Mínimo	11.889%		8.021%		6.358%	
Máximo	40.514%		20.343%		16.014%	
Desvio Padrão	7.262%		3.399%		2.652%	

Fonte: elaborado pelo autor.

**Tabela 04:** Resultados para formação de Carteiras com Limite de 20 Ativos com otimização via Cplex.

Carteiras com Limite de 20 Ativos						
Intervalo out-of-sample	20 períodos		40 períodos		60 períodos	
Valores de saída	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap
<b>Média</b>	2851.541	12.96%	2854.720	12.66%	2563.509	11.61%
<b>Mínimo</b>	403.668	5.00%	466.271	5.00%	403.668	5.00%
<b>Máximo</b>	3600.520	31.88%	3600.520	25.43%	3600.500	23.07%
<b>Desvio Padrão</b>	1138.275	7.63%	1157.697	6.91%	1352.043	7.30%
Tracking Error (período out-of-sample)						
Out-of-sample	20 períodos		40 períodos		60 períodos	
<b>Média</b>	0.065%		0.186%		0.340%	
<b>Mínimo</b>	-1.787%		-2.215%		-2.207%	
<b>Máximo</b>	1.236%		1.405%		2.352%	
<b>Desvio Padrão</b>	0.799%		0.927%		1.395%	
Turnover Mensal						
Out-of-sample	20 períodos		40 períodos		60 períodos	
<b>Média</b>	30.420%		17.137%		11.981%	
<b>Mínimo</b>	12.624%		10.330%		6.805%	
<b>Máximo</b>	53.419%		22.320%		18.768%	
<b>Desvio Padrão</b>	11.396%		3.694%		3.741%	

Fonte: elaborado pelo autor.

A partir desses testes iniciais, foram também analisados outros resultados. Em relação à projeção em *rolling horizon* das carteiras, as Figuras 09, 10 e 11 (Apêndice A) demonstram que as carteiras formadas cumprem o papel de realizar o *tracking* do índice especialmente nos casos de rebalanceamento bimestral e trimestral dos portfólios.

Verifica-se também a tendência de que as carteiras apresentem, em média, erro positivo em relação ao índice, ou seja, o rendimento médio das carteiras tende a ser superior ao rendimento do índice (consequência da modelagem de otimização empregada); dessa forma, a curva do índice em cada gráfico permanece abaixo das curvas das carteiras. Esse fato pode ser percebido também nas Tabelas 01 a 04, em que se percebe que, para os quatro tipos de carteiras formadas, o erro médio de *tracking* permanece baixo, porém positivo. Para todos os casos, o *tracking error* é crescente conforme aumenta o intervalo de rebalanceamento (aumentando também a volatilidade).

Percebe-se ainda esse movimento de erro positivo em relação ao índice nas Tabelas 05 a 07 a seguir, com resultados estatísticos calculados a partir dos retornos compostos das carteiras. Nota-se que o retorno médio das carteiras permanece sempre positivo, apesar de os retornos acumulados serem negativos especificamente nesse período da amostra de dados adotada (exceto

para o caso de projeção de 60 períodos com carteira de 30 ativos, em que o retorno acumulado é positivo, tendo, por consequência, maior erro de *tracking*).

Em relação aos custos, conforme informações apresentadas nas Tabelas 01 a 04, para cada tipo de carteira formada, a vantagem da realização de rebalanceamento de 60 períodos, em relação a 20 e 40 períodos, pode ser percebida claramente pelos valores médios mensais de *turnover*. Considerando que um dos objetivos principais da estratégia de *index tracking* é ter custos baixos, o rebalanceamento em 60 períodos apresenta claro destaque nos quatro tipos de carteiras, uma vez que o maior período de rebalanceamento não demonstra piora de Volatilidade em relação ao índice (conforme nota-se nas Tabelas 05 a 07 abaixo).

**Tabela 05:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 20 períodos (Rebalanceamento Mensal).

Estatísticas	Ibovespa	Rebalanceamento de 20 Períodos			
		Sem Limite de Ativos	Limite de 40 Ativos	Limite de 30 Ativos	Limite de 20 Ativos
Retorno Médio Por Carteira	0.003%	0.007%	0.008%	0.007%	0.006%
Retorno Acumulado	-3.757%	-0.865%	-0.177%	-0.974%	-1.483%
Volatilidade	1.251%	1.253%	1.253%	1.255%	1.275%
Volatilidade Anual	19.855%	19.894%	19.890%	19.916%	20.246%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987
Assimetria	-0.392	-0.395	-0.388	-0.390	-0.397
Curtose	1.755	1.706	1.701	1.678	1.576

Fonte: elaborado pelo autor.

**Tabela 06:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 40 períodos (Rebalanceamento Bimestral).

Estatísticas	Ibovespa	Rebalanceamento de 40 Períodos			
		Sem Limite de Ativos	Limite de 40 Ativos	Limite de 30 Ativos	Limite de 20 Ativos
Retorno Médio Por Carteira	0.003%	0.005%	0.005%	0.005%	0.008%
Retorno Acumulado	-3.757%	-2.365%	-2.264%	-2.137%	-0.335%
Volatilidade	1.251%	1.252%	1.253%	1.251%	1.274%
Volatilidade Anual	19.855%	19.883%	19.891%	19.854%	20.220%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987
Assimetria	1.755	1.793	1.798	1.718	1.812
Curtose	-0.392	-0.391	-0.387	-0.378	-0.400

Fonte: elaborado pelo autor.

**Tabela 07:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 60 períodos (Rebalanceamento Trimestral).

Estatísticas	Ibovespa	Rebalanceamento de 60 Períodos			
		Sem Limite de Ativos	Limite de 40 Ativos	Limite de 30 Ativos	Limite de 20 Ativos
Retorno Médio Por Carteira	0.003%	0.007%	0.007%	0.009%	0.008%
Retorno Acumulado	-3.757%	-0.729%	-0.308%	0.961%	-0.197%
Volatilidade	1.251%	1.255%	1.254%	1.256%	1.276%
Volatilidade Anual	19.855%	19.916%	19.901%	19.938%	20.258%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987
Assimetria	-0.392	-0.402	-0.400	-0.371	-0.383
Curtose	1.755	1.795	1.810	1.739	1.803

Fonte: elaborado pelo autor.

As estatísticas apresentadas nas Tabelas 05 a 07 auxiliam a corroborar a boa qualidade das soluções obtidas inicialmente. Para os três casos de período de rebalanceamento dos portfólios, verificam-se valores de Retorno Médio das carteiras próximos dos valores do Ibovespa. Da mesma forma, os Retornos Acumulados também permanecem próximos do índice para os três intervalos *out-of-sample* projetados. Os valores de Volatilidade e Volatilidade Anual aproximam-se do índice em todos os casos, assim como os valores de Correlação, Assimetria e Curtose. Em todos os casos, a Volatilidade tende a aumentar um pouco e a Correlação tende a diminuir um pouco com menor quantidade de ativos compondo os portfólios, o que aponta para aumento do risco na medida em que a quantidade de ativos na carteira é reduzida.

A partir da apresentação desses resultados prévios, fica demonstrada inicialmente a qualidade das soluções que podem ser obtidas com a aplicação do modelo proposto. Além disso, verifica-se a clara necessidade de uso de heurística para solução do problema com carteiras de menos de 20 ativos.

#### 4.2. Testes com alteração da constante $T$ (tamanho da amostra *in-sample*)

Após execução de testes iniciais com o parâmetro  $T = 150$ , foram realizados novos testes com alteração da amostra *in-sample* para  $T = 100$ . De acordo com Gaivoronski, Krylov e van der Wijst (2005), o uso de maiores amostras tende a gerar carteiras mais estáveis como resultado. Dessa forma, o objetivo de se testar o uso de  $T = 100$  consiste em verificar se essa alteração pode alterar a qualidade das soluções iniciais obtidas, bem como resultar também em tempos

computacionais de solução menores (na medida em que o tamanho do problema de otimização diminui com essa alteração em  $T$ ).

A Tabela 08 a seguir demonstra sucintamente os resultados dos testes com adoção de  $T = 100$ . Para obtenção dos resultados abaixo, foram realizadas otimizações com carteira de limite de 20 ativos e período de rebalanceamento mensal (20 dias úteis); com esse período de rebalanceamento, conforme já citado, é possível formar um total de 36 carteiras para a amostra de dados adotada nessa dissertação. A Tabela 08 demonstra, assim, uma comparação entre os resultados para  $T = 100$  e  $T = 150$  em termos de tempo de solução, *gap*, erro de *tracking* e *turnover* mensal.

**Tabela 08:** Comparação de resultados para Carteiras com intervalo *out-of-sample* de 20 períodos e limite de 20 ativos, para  $T = 100$  e  $T = 150$ .

Carteiras com Limite de 20 Ativos - Rebalanceamento 20 Períodos				
Tamanho da Amostra In-sample	$T = 100$		$T = 150$	
Valores de saída	Tempo de solução (segundos)	Gap	Tempo de solução (segundos)	Gap
<b>Média</b>	1873.082	7.02%	2851.541	12.96%
<b>Mínimo</b>	14.524	5.00%	403.668	5.00%
<b>Máximo</b>	3600.570	20.79%	3600.520	31.88%
<b>Desvio Padrão</b>	1503.609	4.38%	1138.275	7.63%
Tracking Error (período out-of-sample)				
<b>Média</b>	0.149%		0.065%	
<b>Mínimo</b>	-1.184%		-1.787%	
<b>Máximo</b>	1.677%		1.236%	
<b>Desvio Padrão</b>	0.695%		0.799%	
Turnover Mensal				
<b>Média</b>	36.166%		30.420%	
<b>Mínimo</b>	16.205%		12.624%	
<b>Máximo</b>	59.034%		53.419%	
<b>Desvio Padrão</b>	9.984%		11.396%	

**Fonte:** elaborado pelo autor.

Empiricamente, poder-se-ia esperar que a diminuição de  $T$  tenderia a formar carteiras mais instáveis, de acordo com Gaivoronski, Krylov e van der Wijst (2005) – o que poderia ser verificado por maior variação nos valores de erro de *tracking*. Pela Tabela 08, vê-se que, de fato, o erro médio de *tracking* aumenta com a diminuição de  $T$ . Nota-se também que os valores de *turnover* são maiores para  $T = 100$ , demonstrando que as carteiras sofrem mais alterações entre si

nos momentos de rebalanceamento. Esses fatos ajudam a se concluir que as carteiras formadas com diminuição de  $T$  são mais instáveis e geram aumento do erro de *tracking*.

Já em relação aos tempos de solução e valores de *gap*, pode-se concluir que a alteração em  $T$  não implica grandes alterações nos resultados. Embora o tempo e *gap* médios sejam menores com redução de  $T$ , os valores de tempo chegam facilmente ao limite adotado de 3.600 segundos, com *gap* bastante acima de 5%.

Dessa forma, a alteração em  $T$  permite a conclusão de que, de uma forma geral, a diminuição da amostra de dados *in-sample* tende a gerar carteiras mais instáveis; porém, esse fator não é, de certa forma, “compensado” por uma queda considerável nos tempos computacionais de otimização e nos valores de *gap*. Nota-se que os tempos médios de solução sofrem uma queda com diminuição de  $T$ ; porém, permanece inviável a possibilidade de se formar carteiras com menos de 20 ativos, pois o tempo pode ultrapassar 1 hora de processamento com valores elevados de *gap*. Assim, como conclusão, pode-se afirmar que é preferível a manutenção de  $T = 150$ , de forma a se obter carteiras mais estáveis que permitam realização do *tracking* mesmo para períodos mais longos de rebalanceamento, já que a perda de estabilidade não é compensada por tempos menores de solução.

### 4.3. Alteração de parâmetros do *solver* Cplex

Conforme demonstrado nas seções 4.1 e 4.2, os testes iniciais realizados apresentaram tempos de solução elevados e inviáveis para formação de carteiras limitadas menos de 20 ativos. Dessa forma, foram realizados novos testes computacionais para tentativa de se obter tempos mais baixos através de alteração de parâmetros do *solver* Cplex, considerando que o Cplex pode ter configurações alteradas.

Após uma análise de parâmetros do *solver*, foram verificadas algumas instruções do modo de processamento do Cplex que, se alteradas, poderiam ter influência sobre o tempo de solução, resolvendo os problemas mais rapidamente. Dessa forma, alguns desses parâmetros foram selecionados para realização de novos testes de otimização. Abaixo, são apresentados esses parâmetros selecionados com uma breve descrição.

- “*coeffreduce*”: essa instrução é aplicada na fase de pré-processamento do problema e busca aumentar o desempenho do Cplex no caso de programação inteira, influenciando o tempo de

solução dos subproblemas na execução do algoritmo *branch-and-bound*. Ao desabilitar essa instrução (*coeffreduce=1*), podem ser obtidos tempos de solução menores para a solução de um modelo;

- “*probe*”: essa instrução define se o Cplex realiza, no momento de pré-processamento, uma espécie de sondagem em problemas de programação linear e inteira, com objetivo de reduzir o tamanho do problema antes de iniciar a fazer de processamento e solução. Por padrão do *solver*, essa instrução é executada algumas vezes, e desabilitar essa instrução (*probe=-1*) pode influenciar os tempos de solução;
- “*mipemphasis*”: essa instrução afeta a aplicação, pelo Cplex, do método de *branch-and-cut* para solução do problema. Por padrão, o *solver* realiza um “balanceamento” entre a busca por possíveis soluções e a busca pela solução ótima. Pode-se, então, alterar a configuração desse parâmetro para que o *solver* trabalhe com ênfase em encontrar novas possíveis soluções (*mipemphasis=1*), o que pode ser adequado para problemas em que a solução ótima não é exigida;
- “*fpheur*”: Essa instrução determina a aplicação, pelo Cplex, da chamada “*feasibility pump heuristic*”. Por definição padrão, o *solver* determina os melhores momentos de uso dessa heurística. Ao desabilitar-se essa instrução (*fpheur=-1*), pode ser possível se obter respostas de forma mais rápida especialmente em casos em que não é exigida uma solução ótima;
- “*mipsearch*”: essa instrução determina se o Cplex utiliza busca por *branch-and-bound* ou busca dinâmica para o processo de solução dos problemas. Por definição padrão, o *solver* realiza análise do problema e decide automaticamente pela melhor opção. Testes foram realizados para determinar que o *solver* utilize apenas o tipo de busca por *branch-and-bound* (*mipsearch=1*) ou busca dinâmica (*mipsearch=2*);
- “*rinsheur*”: essa diretiva determina o uso da chamada heurística de Rins (do inglês “*relaxation induced neighborhood search heuristic*”). Por definição padrão, o *solver* aplica essa heurística algumas vezes durante o desenvolvimento do problema. Dessa forma, desabilitar essa instrução (*rinsheur=-1*) pode influenciar os tempos de solução finais.

Após seleção dessas instruções, foram então realizados testes computacionais iniciais para verificar se elas podem afetar o tempo de solução dos problemas. Para isso, foram realizados testes em que as carteiras de investimento são limitadas a 20 ativos. Adotando  $n = 20$  (ou seja, amostra *out-of-sample* de 20 períodos), podemos formar um total de 36 carteiras de investimentos para o período de análise. Dessas 36 carteiras, foram selecionadas duas aleatoriamente, e então foi

realizada otimização dessas duas carteiras utilizando-se oito diferentes conjuntos de instruções (oito testes), conforme descrição na tabela abaixo:

**Tabela 09:** Parâmetros do Cplex alterados nos novos testes computacionais.

Descrição dos parâmetros alterados em cada um dos oito novos testes			
<b>1º Teste:</b>	fpheur = -1	rinsheur = -1	
<b>2º Teste:</b>	fpheur = -1	rinsheur = -1	mipemphasis = 1
<b>3º Teste:</b>	fpheur = -1	rinsheur = -1	mipsearch = 1
<b>4º Teste:</b>	fpheur = -1	rinsheur = -1	mipsearch = 2
<b>5º Teste:</b>	rinheur = -1		
<b>6º Teste:</b>	fpheur = -1		
<b>7º Teste:</b>	probe = -1		
<b>8º Teste:</b>	coeffreduce = 1		

**Fonte:** elaborado pelo autor.

Os resultados obtidos com uso desses conjuntos de instruções podem ser analisados pela Tabela 10 abaixo, com o tempo de tempo de solução (em segundos) de cada uma das duas carteiras testadas, para cada combinação de instruções definida na Tabela 09 acima. Com base na Tabela 10, nota-se que esses testes realizados com alterações de parâmetros do Cplex não geraram modificações consideráveis nos tempos finais de solução, exceto para dois casos. No sétimo e oitavo testes, tem-se redução mais significativa do tempo final de solução da otimização para as duas carteiras aleatórias selecionadas.

**Tabela 10:** Resultados parciais dos testes com alteração de parâmetros do Cplex.

Instruções utilizadas	Tempo de solução (em segundos)	
	Carteira A	Carteira B
<b>Solução original (sem alterações no Cplex)</b>	1982,55	1712,92
<b>1º Teste:</b> fpheur = -1      rinsheur = -1	2131,85	1845,69
<b>2º Teste:</b> fpheur = -1      rinsheur = -1      mipemphasis = 1	3597,10	1613,52
<b>3º Teste:</b> fpheur = -1      rinsheur = -1      mipsearch = 1	2037,84	1854,74
<b>4º Teste:</b> fpheur = -1      rinsheur = -1      mipsearch = 2	2137,82	1811,62
<b>5º Teste:</b> rinheur = -1	2180,47	1845,54
<b>6º Teste:</b> fpheur = -1	2003,26	1742,08
<b>7º Teste:</b> probe = -1	1852,20	1589,96
<b>8º Teste:</b> coeffreduce = 1	1294,12	1010,36

**Fonte:** elaborado pelo autor.



Assim, essas duas instruções (7º Teste: “*probe=-1*”; e 8º Teste: “*coeffreduce=1*”) foram adotadas e aplicadas para otimização de todas as 36 carteiras de investimentos que podem ser formadas, considerando intervalo de rebalanceamento de 20 períodos. Dessa forma, torna-se possível verificar de forma mais clara se essas instruções têm, de fato, impacto sobre os tempos de solução do modelo de otimização.

Após execução dos testes com otimização do total de 36 carteiras de investimento com uso da instrução “*coeffreduce=1*” (mantendo a instrução “*probe*” com seu valor padrão, sem sofrer alteração) e, após, com uso da instrução “*probe=-1*” (mantendo a instrução “*coeffreduce*” com seu valor padrão, sem sofrer alteração), a Tabela 11 abaixo resume os resultados obtidos em termos de tempo de solução, *gap* e erro de *tracking*.

**Tabela 11:** Resultados para testes com alteração de parâmetros do Cplex.

		Soluções originais	8º Teste: 'coeffreduce = 1'	7º Teste: 'probe = -1'
Tempo de solução (em segundos)	Médio	2851.54	2809.96	2978.53
	Mínimo	403.67	435.70	519.80
	Máximo	3600.52	3600.52	3600.32
	Desvio Padrão	1138.28	1231.49	1016.16
Gap	Médio	12.96%	12.45%	13.47%
	Mínimo	5.00%	5.00%	5.00%
	Máximo	31.88%	24.96%	36.24%
	Desvio Padrão	7.63%	6.75%	8.26%
Tracking Error	Médio	0.065%	0.079%	0.061%
	Mínimo	-1.787%	-2.282%	-2.242%
	Máximo	1.236%	1.236%	1.236%
	Desvio Padrão	0.799%	0.841%	0.839%

**Fonte:** elaborado pelo autor.

Percebe-se que, apesar de testes iniciais terem indicado que cada uma das duas instruções destacadas (“*probe=-1*”; “*coeffreduce=1*”) poderia afetar os tempos de solução das otimizações, os resultados na Tabela 11 apontam para conclusão em sentido contrário. Nota-se que os tempos de solução médios para as otimizações feitas inicialmente sem alterações no Cplex são bastante similares aos tempos médios das soluções para cada um dos casos de alteração de parâmetros. Da mesma forma, o tempo de solução máximo é o mesmo (3600 segundos). Apenas o tempo mínimo tem alteração mais significativa, sendo que as alterações no Cplex resultaram em aumento do tempo mínimo.

Da mesma forma, também os valores de *gap* e de erro de *tracking* são bastante similares. Não há grandes variações entre os resultados das otimizações sem alterações no Cplex e os resultados das otimizações com alterações de parâmetros. Isso atesta que os resultados são bastante similares e mudanças nas configurações do Cplex não geram grandes diferenças nos resultados obtidos.

De forma geral, esses resultados demonstram a ineficácia do uso dessas instruções para tentativa de diminuição dos tempos de solução para esse problema de otimização que está sendo trabalhado. Assim, tem-se mais um indicativo da necessidade de uso de um método heurístico para solução dos problemas de otimização com número reduzido de ativos, pois alterações no *solver* não são suficientes para permitir formação de carteiras com menos de 20 ativos.

#### 4.4. Considerações sobre os testes iniciais de otimização com o *solver* Cplex

Com o desenvolvimento das seções acima, buscou-se demonstrar a aplicação inicial do modelo de otimização com os objetivos de “validar” o modelo e, também, de verificar as limitações do uso direto do *solver* Cplex. Para isso, foram desenvolvidas as seções 4.1, com testes iniciais de otimização, e as seções 4.2 e 4.3, com alterações que poderiam ter algum impacto na qualidade e no tempo de solução das otimizações.

Inicialmente, foi possível verificar a qualidade do modelo de otimização na seção 4.1 em termos de realização do *tracking* do índice. Considerando as carteiras formadas com diferentes quantidades de ativos, foi possível constatar que a diminuição da quantidade limite de ativos nas carteiras não gerou alterações significativas nos valores de erro de *tracking*. Mesmo no caso de carteiras de 20 ativos, têm-se valores médios de erro de *tracking* pequenos, com as carteiras acompanhando o desenvolvimento do índice (conforme Figuras 09 a 11 no Apêndice A).

Ficou assim constatado que, independente do período de rebalanceamento adotado, não é possível a formação de carteiras com menos de 20 ativos sem que os tempos computacionais fiquem abaixo de 1 hora de processamento com *gap* abaixo de 5%. Para carteiras com limite de 20 ativos (para a amostra adotada de 67 ativos), embora os valores de erro de *tracking* sejam bons, têm-se *gap* médio acima de 11% e tempo de solução que atinge facilmente 1 hora. Assim, evidencia-se inicialmente a necessidade de uso de um método heurístico para formação de carteiras com menos de 20 ativos.

Com essa constatação inicial de limitação do uso direto do *solver*, foi realizada a seguir a alteração do parâmetro  $T$  nas otimizações. A diminuição desse parâmetro, de  $T = 150$  para

$T = 100$ , levou a pequenas variações nos valores de *tracking error*, com aumento do erro médio, demonstrando assim que as novas carteiras são mais instáveis – em linha com Gaivoronski, Krylov e van der Wijst (2005).

Conforme a Tabela 08, em relação a tempos de solução e *gap*, nota-se que a diminuição de  $T$  gera decréscimo do tempo de solução médio e do *gap* médio das respostas. Apesar disso, o *gap* médio permanece acima de 5%, com tempos de solução ainda atingindo 1 hora de processamento. Assim, percebe-se que não há melhora de qualidade das soluções quando se diminui o tamanho da amostra *in-sample*; essa diminuição tende a gerar maiores valores de erro de *tracking* ao mesmo tempo em que o tempo computacional permanece atingindo 1 hora de processamento com *gap* acima de 5%.

Já na seção 4.3, foi realizada uma análise de parâmetros do *solver* Cplex, a fim de se verificar se alterações de parâmetros do *solver* poderiam impactar os tempos de otimização. Os testes realizados com alterações de diferentes parâmetros do *solver* não geraram alterações consideráveis em termos de tempo de processamento. Pode-se concluir, pela Tabela 11, que as alterações realizadas não resultaram em diminuição do tempo médio de solução; dessa forma, percebe-se que alterações no Cplex não são suficientes para resolver o problema de otimização com carteiras de 20 ativos ou menos.

Assim, foi possível verificar inicialmente que o modelo apresenta soluções de qualidade em termos de erro de *tracking*. Porém, evidencia-se que não é possível formar carteiras com menos de 20 ativos para a amostra de dados adotada. As alterações realizadas na quantidade de dados dentro da amostra (período *in-sample*) demonstraram que há perda gradual de qualidade nas soluções, sem que o tempo de solução diminua de forma considerável; além disso, alterações no *solver* Cplex permitiram a constatação de que o *solver* já possui funcionamento eficiente, no sentido de que alterações de parâmetros do *solver* não resultam em melhora no tempo de processamento. Dessa forma, constata-se a necessidade de uso de um método heurístico para resolver esse problema de otimização com formação de carteiras de menos de 20 ativos com tempos de solução abaixo de 10 minutos e valores de *gap* mais baixos (ao menos abaixo de 10%).

## 5. MÉTODO HEURÍSTICO: ALGORITMOS GENÉTICOS

A partir dos resultados apresentados anteriormente, ficou demonstrando a necessidade de um método heurístico para que se consiga solucionar o problema com tempos computacionais menores e carteiras de menor quantidade de ativos. Para preencher esta lacuna, optou-se por adotar o método heurístico de Algoritmos Genéticos (AG) nesta dissertação. Assim, este capítulo apresenta algumas características principais do método de AG e sua adaptação para aplicação ao problema proposto de *index tracking*.

Na seção 5.1, são apresentadas as características principais desse método heurístico, adaptando-o ao problema de *index tracking*; a seção 5.2 apresenta os resultados obtidos para aplicação inicial do algoritmo ao problema; na seção 5.3, é demonstrado como foi verificada a qualidade das respostas geradas pelo algoritmo em termos de *gap* das soluções. Na seção 5.4, são feitos novos testes com alteração de parâmetros do AG e, também, alteração do tempo de execução do algoritmo, de forma a se verificar a possibilidade de se obter soluções melhores do que as apresentadas na seção 5.2. E, na seção 5.5, têm-se as considerações principais sobre a aplicação do método heurístico. Para aplicação do AG, foi utilizado novamente o *solver* Cplex. O algoritmo foi feito com linguagem de programação C++, por se tratar de uma das linguagens de programação mais utilizadas atualmente.

### 5.1. Método de Algoritmos Genéticos adaptado ao problema de *Index Tracking*

Conforme Goldberg (1989), o método heurístico de Algoritmos Genéticos (AG) é uma técnica de otimização proposta por John H. Holland em 1975 (em “Adaptation in Natural and Artificial Systems: an introductory analysis with applications to biology, control and artificial intelligence”). Em essência, essa técnica funciona como um algoritmo de busca baseado na evolução dos melhores indivíduos entre um conjunto populacional (utilizando preceitos de biologia e da teoria evolucionista de seleção natural).

No AG, essencialmente, tem-se uma população formada por uma quantidade determinada de indivíduos (ou cromossomos) que correspondem a possíveis soluções do modelo – ou seja, um conjunto de soluções iniciais. Com essa população, podem ser praticados diversos tipos de processos evolutivos, como, por exemplo, cruzamentos entre indivíduos (*crossovers*), ou mutações para alteração de cada indivíduo, para gerar novos indivíduos que possam ter melhores valores de aptidão (ou *fitness*). Assim, ocorre a evolução da população, com geração constante de novos indivíduos, esperando-se que os novos indivíduos tenham melhores valores de aptidão em

relação à população anterior. A obtenção de indivíduos de melhor aptidão significa que estão sendo obtidas melhores soluções para o problema.

Na adaptação de AG para o problema de *index tracking*, serão adotados os seguintes conceitos:

- População: conjunto de  $N$  indivíduos (ou cromossomos). Cada indivíduo corresponde a uma possível solução para o modelo;
- Indivíduo (ou cromossomo): trata-se de cada portfólio formado, ou seja, uma possível solução da otimização. Cada indivíduo corresponde a um vetor binário de 67 posições (ou genes) – ou seja, cada posição corresponde à variável  $z_i$  do modelo, para todo  $i = 1, 2, 3, \dots, 67$ , indicando se o ativo  $i$  está ou não na carteira. Por exemplo, uma carteira formada pelos ativos 01, 02, 03, 04 e 05, dentro do conjunto de 67 ativos da amostra, será representada conforme demonstrado na Figura 04 abaixo:

	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$	...	$z_{65}$	$z_{66}$	$z_{67}$
Indivíduo 1	1	1	1	0	0	0	0	0	0	0	...	0	1	1

**Figura 04:** Formato de cada indivíduo utilizado na heurística. **Fonte:** elaborado pelo autor.

- Fase de Evolução: cada fase em que os indivíduos de uma população  $N$  passam por processos evolutivos (como cruzamentos e mutações), com o objetivo de gerar novos indivíduos e formar uma nova população de indivíduos de melhor aptidão. O algoritmo deverá executar diversas fases de evolução em sequência, de forma a aprimorar os indivíduos de cada população;
- Cruzamento (ou *crossover*): processo em que dois indivíduos são parcialmente alterados para que cada um desses indivíduos receba parte dos genes do outro indivíduo. A determinação de quais genes serão “trocados” entre os indivíduos será feita de forma aleatória. No desenvolvimento do AG nessa dissertação, todos os  $N$  indivíduos da população serão combinados dois a dois para execução de cruzamento. Serão realizados cruzamentos simples, em que uma posição de corte é selecionada aleatoriamente. Na Figura 05 abaixo, tem-se um exemplo em que dois Indivíduos 1 e 2 sofrem um processo de *crossover* de forma a gerar dois novos indivíduos 3 e 4 (em que, como já citado, a “posição de corte” é selecionada aleatoriamente):

	<i>Posição de "Corte"</i>													
	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$	...	$z_{65}$	$z_{66}$	$z_{67}$
Indivíduo 1	1	1	1	0	0	0	0	0	0	0	...	1	1	0
Indivíduo 2	0	0	0	1	1	0	0	0	0	1	...	0	0	1
Indivíduo 3	1	1	1	0	0	0	0	0	0	1	...	0	0	1
Indivíduo 4	0	0	0	1	1	0	0	0	0	0	...	1	1	0

**Figura 05:** Descrição do processo de *crossover* entre dois indivíduos. **Fonte:** elaborado pelo autor.

- **Mutação:** processo em que cada indivíduo tem parte de seus genes alterada, gerando um novo indivíduo. No AG aplicado nessa dissertação, cada um dos dois novos indivíduos gerados no processo de cruzamento (conforme citado acima) terá um determinado percentual de probabilidade de sofrer mutação. No processo de mutação “simples”, é selecionada aleatoriamente uma posição do vetor binário com valor 0, e essa posição é alterada para o valor 1; e, ao mesmo tempo, é selecionada uma posição de valor 1, e esse valor é alterado para 0, mantendo assim o indivíduo com a mesma quantidade de ativos. Para o caso de mutações “duplas”, serão selecionadas duas posições de valor 1, que serão alteradas para 0, e duas posições de valor 0, que serão alteradas para 1. Segue abaixo na Figura 06 um exemplo em que um Indivíduo 1 sofre uma mutação simples, com alterações dos Ativos 03 e 65:

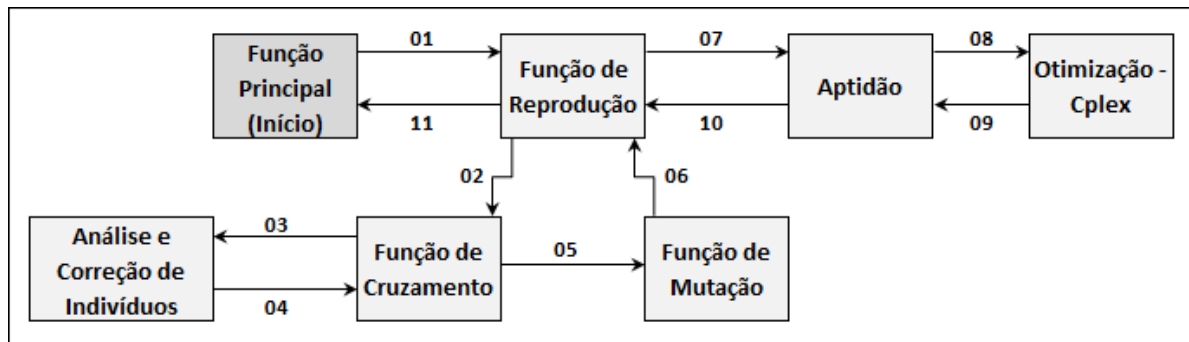
	<i>1ª posição de alteração</i> →														<i>2ª posição de alteração</i> →													
	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$	...	$z_{65}$	$z_{66}$	$z_{67}$														
Indivíduo 1	1	1	1	0	0	0	0	0	0	0	...	0	1	1														
Indivíduo 2	1	1	0	0	0	0	0	0	0	0	...	1	1	1														

**Figura 06:** Descrição do processo de mutação entre dois indivíduos. **Fonte:** elaborado pelo autor.

- **Taxas de ocorrência de cruzamentos e de mutações:** taxas correspondentes a parâmetros pré-determinados. Correspondem à probabilidade de ocorrência de cruzamentos e de mutações. Nessa proposta, pretende-se adotar probabilidade de ocorrência de 100% para cruzamentos, e um percentual menor que 100% para mutações. O objetivo é garantir que ao menos os cruzamentos ocorram em todos os casos, provocando a geração de uma quantidade maior de indivíduos diferentes. Dessa forma, com maior variedade de indivíduos, há a tendência maior de que o algoritmo não fique preso em uma solução de ótimo local;

- Aptidão (ou *fitness*) de cada indivíduo: resultado da função objetivo da otimização desse indivíduo. Ou seja, cada indivíduo terá um correspondente valor de aptidão (que corresponde ao resultado da função objetivo) e menor valor de aptidão indica que o indivíduo é melhor apto e corresponde a uma solução melhor para o problema.

Dados esses conceitos, o método de AG será aplicado através da execução de cada processo de Fase de Evolução por uma quantidade pré-determinada de vezes. Cada execução de uma Fase de Evolução corresponde à realização dos seguintes passos (resumidos no fluxograma básico apresentado a seguir na Figura 07):



**Figura 07:** Fluxo resumido de cada Fase de Evolução do Algoritmo Genético. **Fonte:** elaborado pelo autor.

- 1º. Função Principal (Início): recebe uma população inicial de  $N$  indivíduos e a definição da quantidade de Fases de Evolução que devem ser executadas; após, envia a população inicial para a Função de Reprodução;
- 2º. Função de Reprodução: recebe uma população de  $N$  indivíduos (representados pelos vetores binários) e forma todas as combinações possíveis desses indivíduos dois a dois. Cada par de indivíduos é enviado para a Função de *Crossover*, em que sofrerão cruzamento com 100% de probabilidade. Após isso, os dois novos indivíduos formados são enviados para uma Função de Análise e Correção de Indivíduos, para que os novos indivíduos contêm a quantidade correta de ativos da carteira (05 ou 10 ativos). A correção é feita da seguinte forma: caso um novo indivíduo tenha sete ativos, por exemplo, para o caso de formação de carteiras de cinco ativos, então dois ativos serão selecionados aleatoriamente e retirados da carteira, para que fique com cinco ativos. Da mesma forma, caso seja necessário acrescentar dois ativos (em vez de excluir), então dois ativos serão selecionados aleatoriamente e incluídos na carteira;

- 3°. Função de Mutação: essa função recebe os dois novos indivíduos formados na Função de *Crossover*; cada indivíduo terá um determinado percentual de probabilidade de sofrer mutação (sendo esse percentual pré-determinado). Dessa forma, há a possibilidade de que indivíduos não passem por mutação. Após o processo de mutação, esses indivíduos são enviados de volta à Função de Reprodução. Dessa forma, a partir de uma população de  $N$  indivíduos inicialmente recebida na Função de Reprodução, tem-se um novo conjunto de  $N \times 4$  indivíduos (indivíduos iniciais e novos indivíduos gerados pelos cruzamentos e mutações);
- 4°. Função de Cálculo de Aptidão: recebe os  $N \times 4$  indivíduos da Função de Reprodução e realiza otimização via *solver* Cplex de cada um desses indivíduos, para definição dos pesos dos ativos de cada carteira (pois, agora, o *solver* recebe os valores de  $z_i$  como constantes, e não como variáveis – dessa forma, a complexidade do modelo é diminuída na medida em que a definição de  $z_i$  como constante elimina a restrição com variável inteira). Dessa forma, obtêm-se, para cada indivíduo, um resultado para a função objetivo, correspondente ao valor de aptidão do indivíduo;
- 5°. Todos os indivíduos, com respectivos valores de aptidão, são retornados para a Função de Reprodução, que os ordena e forma uma nova população de  $N$  indivíduos com os melhores valores de aptidão (aplicando assim o conceito de elitismo, pois os melhores indivíduos são mantidos para a população seguinte). Dessa forma, inicia-se um novo ciclo de Fase de Evolução. Após a execução da quantidade pré-determinada de Fases de Evolução, obtêm-se uma carteira final, correspondente à carteira de melhor valor de aptidão.

Considerando essa descrição do algoritmo, é necessário ainda definir alguns parâmetros que devem ser estabelecidos para execução do AG. Esses parâmetros deverão ser alterados durante os testes computacionais para tentativa de geração das melhores respostas possíveis, obtendo-se, se possível, as respostas ótimas para as carteiras de otimização. Para implantação do algoritmo, foram definidos os seguintes parâmetros:

- *Taxa\_Crossover*: taxa de ocorrência de cruzamentos (*crossover*): probabilidade de dois indivíduos passarem pelo processo de cruzamento. Essa taxa será sempre de 100%;
- *Taxa\_Mut*: taxa de ocorrência de mutações; poderá ser adotada taxa de 100% ou inferior a 100%;
- *Num\_Mut*: definição se serão aplicadas mutações “simples” ou “duplas”;



- *Num\_Evol*: definição da quantidade de ciclos de reprodução (Fases de Evolução) que deverão ser executados para obtenção da resposta final;
- *Num\_Popul*: correspondente ao valor de  $N$ , ou seja, ao tamanho da população;
- *Tam\_Cart*: tamanho da carteira, ou seja, definição se a carteira terá 05 ativos, 10 ativos, etc.

## 5.2. Resultados iniciais da aplicação do algoritmo

Conforme já citado anteriormente, o objetivo principal da implantação do método heurístico de AG é possibilitar a obtenção de carteiras de reduzida quantidade de ativos em um tempo computacional baixo. Para o caso da amostra adotada de 67 ativos, espera-se formar carteiras com limite de 05 ou 10 ativos.

Com base na literatura sobre o assunto (principalmente Beasley, Meade e Chang, 2003; e Oh, Kim e Min, 2005), optou-se inicialmente nesses testes por definir em torno de 5 minutos como um tempo de execução bastante aceitável para o algoritmo. E, para esse tempo de execução, espera-se obter respostas com, ao menos, *gap* médio abaixo de 10%.

Para os testes iniciais para formação de carteiras de 05 e 10 ativos, os valores adotados para os parâmetros do algoritmo estão expressos abaixo na Tabela 12. A definição dos valores de *Num\_Popul* e *Num\_Evol* foi feita de forma combinada para que o tempo de execução do AG ficasse em torno de 5 minutos. Além disso, os parâmetros  $\vartheta$  e  $\theta$  do modelo de otimização foram sempre mantidos constantes com os valores de  $\vartheta = -0,01$  e  $\theta = 0,01$ .

**Tabela 12:** Parâmetros adotados para os **Testes Iniciais** do algoritmo.

Tamanho da Carteira (Parâmetro " <i>Tam_Cart</i> ")	05 Ativos	10 Ativos
<i>Num_Popul</i>	20	20
<i>Taxa_Crossover</i>	100%	100%
<i>Taxa_Mut</i>	85%	85%
<i>Num_Mut</i>	1	1
<i>Num_Evol</i>	30	25

**Fonte:** elaborado pelo autor.

Novamente, para cada um dos tipos de carteiras obtidas (limite de 05 e 10 ativos), têm-se períodos de rebalanceamento mensal, bimestral e trimestral; para cada um desses períodos, pode-

se formar, respectivamente, 36 carteiras, 18 carteiras e 12 carteiras. Os resultados obtidos com o AG podem ser verificados nas Tabelas 13 e 14 abaixo.

Como resultado inicial, verificou-se que tanto as carteiras de 05 quanto de 10 ativos apresentam valores médios de erro de *tracking* parecidos, exceto para o caso de rebalanceamento de 40 períodos. Para esse caso de 40 períodos, tem-se erro de *tracking* bastante elevado, com grande oscilação (alto desvio-padrão), apontando que essas respostas iniciais podem não ser numericamente boas, ou seja, podem ser respostas com *gap* bastante elevado. Além disso, nota-se que os valores de erro de *tracking* são maiores para carteiras de 05 ativos do que para 10 ativos, sendo esse resultado já esperado na medida em que há clara tendência de que o *tracking* piore com diminuição do tamanho das carteiras.

Já em relação aos valores de *turnover* mensal, comparando-se esses resultados com os obtidos para carteiras de 20 ativos ou mais, nota-se que não há grandes variações. No caso de rebalanceamento de 60 períodos, para o caso de 20 ativos, por exemplo, obteve-se *turnover* médio de 11,9% (conforme Tabela 04); e, para 10 e 05 ativos, foram obtidos, respectivamente, valores de *turnover* médios mensais de 13,1% e 15,6% (que se trata de valores baixos se for considerado que o tamanho das carteiras é bastante reduzido). Assim, nota-se mais uma vez a vantagem da estratégia de rebalanceamento de 60 períodos em termos de custos de transação, representados pelos baixos valores de *turnover* mensais.

**Tabela 13:** Resultados para formação de Carteiras com Limite de 05 Ativos.

Carteiras com Limite de 05 Ativos			
Intervalo out-of-sample	20 períodos	40 períodos	60 períodos
<b>Tracking Error (período out-of-sample)</b>			
<b>Média</b>	0.370%	4.512%	1.210%
<b>Mínimo</b>	-3.725%	-10.741%	-4.280%
<b>Máximo</b>	3.246%	25.944%	7.389%
<b>Desvio Padrão</b>	1.772%	10.081%	3.459%
<b>Turnover Mensal</b>			
<b>Média</b>	34.185%	22.749%	15.598%
<b>Mínimo</b>	1.241%	6.840%	3.878%
<b>Máximo</b>	78.719%	43.699%	27.661%
<b>Desvio Padrão</b>	20.699%	12.019%	7.859%

**Fonte:** elaborado pelo autor.

**Tabela 14:** Resultados para formação de Carteiras com Limite de 10 Ativos.

Carteiras com Limite de 10 Ativos			
Intervalo out-of-sample	20 períodos	40 períodos	60 períodos
<b>Tracking Error (período out-of-sample)</b>			
<b>Média</b>	0.345%	4.022%	0.464%
<b>Mínimo</b>	-2.420%	-10.964%	-2.693%
<b>Máximo</b>	3.148%	24.948%	5.106%
<b>Desvio Padrão</b>	1.262%	9.929%	2.250%
<b>Turnover Mensal</b>			
<b>Média</b>	36.142%	18.438%	13.126%
<b>Mínimo</b>	7.982%	6.108%	6.181%
<b>Máximo</b>	81.263%	29.741%	20.067%
<b>Desvio Padrão</b>	16.006%	6.737%	5.587%

**Fonte:** elaborado pelo autor.

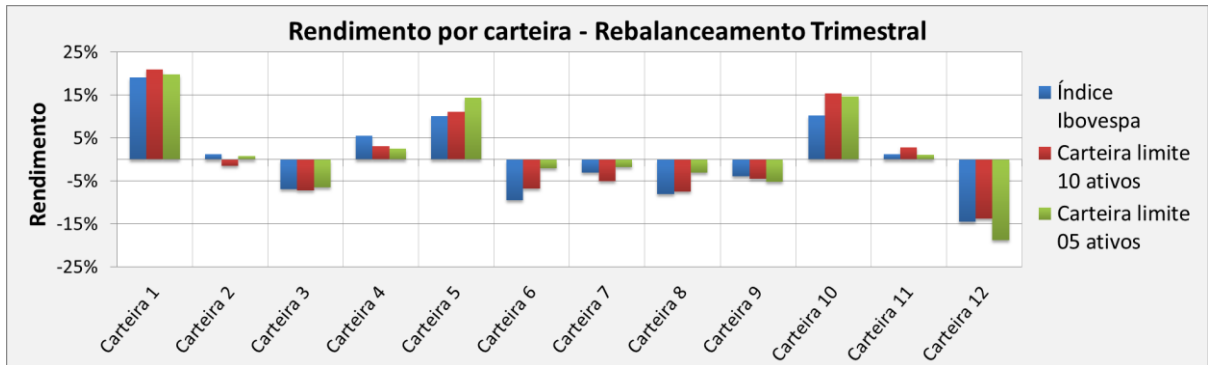
Os resultados desses testes também podem ser analisados a partir da projeção das carteiras conforme Figuras 12, 13 e 14 (Apêndice B). Nota-se que as curvas relativas às carteiras de 10 ativos e especialmente de 05 ativos estão mais “descoladas” da curva do índice.

Para o caso de 10 ativos, a projeção das carteiras fica bastante próxima ao índice no caso de rebalanceamento de 60 períodos; porém, de forma contrária, fica bastante afastada no caso de rebalanceamento de 40 períodos, refletindo os altos valores de erro de *tracking* verificados na Tabela 14 acima. Para o caso de 05 ativos, a curva de projeção das carteiras apresenta-se afastada da curva do índice nos três casos, também refletindo a maior oscilação de erro de *tracking*, conforme verificado nas tabelas acima de resultados. Porém, apesar do erro maior das carteiras, permanece a clara tendência de que o rendimento acumulado das carteiras é superior ao rendimento acumulado do índice, o que se trata de um ponto positivo do modelo de otimização empregado, pois intuitivamente é preferível que o erro existente seja positivo, e não negativo.

Ainda em relação às projeções nas Figuras 12 a 14, verifica-se a vantagem do rebalanceamento trimestral em relação aos casos de atualização mensal e bimestral. Para carteiras limitadas a 05 ativos, os casos mensal e trimestral têm resultados semelhantes em termos de rendimento acumulado durante o período de análise. E, para carteiras limitadas a 10 ativos, nota-se claramente a vantagem do rebalanceamento trimestral, em que o rendimento acumulado da carteira foi 1,12%, sendo que o rendimento acumulado do índice foi de -3,76% (Tabela 18) – ou seja, diferença acumulada em torno de apenas 4,9 pontos percentuais.

A Figura 08 a seguir ajuda a perceber o bom resultado de rebalanceamento trimestral especialmente para carteiras de 10 ativos. Considerando rebalanceamento trimestral, conforme já citado, é possível formar um total de 12 carteiras para o intervalo de tempo usado nos testes. A

Figura 08 apresenta o rendimento de cada uma dessas 12 carteiras para o índice Ibovespa e para as carteiras de 05 e 10 ativos. Vê-se que o rendimento das otimizações com 10 ativos permanece bastante próximo do rendimento do índice, distanciando-se um pouco no caso das carteiras 6 e 10 – enquanto que a carteira limitada a 05 ativos não apresenta uma aderência tão forte ao índice na maioria das 12 carteiras.



**Figura 08:** Rendimento por carteira para Carteiras de 05 e 10 ativos – Rebalanceamento Trimestral.

**Fonte:** elaborado pelo autor.

Finalmente, os resultados desses testes iniciais também podem ser analisados através dos dados estatísticos apresentados nas Tabelas 15, 16 e 17 a seguir. Para todas as informações, nota-se que os valores para carteiras de 10 e 05 ativos distanciam-se dos valores do índice Ibovespa, conforme esperado inicialmente. Têm-se valores de retorno médio por carteira e retorno acumulado distantes dos valores das carteiras de maiores quantidades de ativos. Da mesma forma, os demais valores também variam bastante, como os valores de correlação com o índice, curtose e assimetria. Essas variações são reflexos do fato de que carteiras de menores quantidades de ativos naturalmente tendem a gerar maiores valores de erro de *tracking*, afastando-se mais da evolução do índice (conforme citam, por exemplo, Gaivoronski, Krylov e van der Wijst, 2005).

Ainda pelas Tabelas 15 a 17, podemos novamente verificar a melhor qualidade da adoção de rebalanceamento trimestral para carteiras de 10 ativos. Além de o retorno acumulado para carteiras de 10 ativos estar próximo do retorno acumulado do índice (diferença em torno de 4,9 pontos percentuais), o retorno médio por carteira, para carteiras limitadas a 10 ativos, é de 0,01%, bastante próximo do retorno médio para os portfólios de 20 ou mais ativos. Esses números ajudam a corroborar a conclusão de que a estratégia de rebalanceamento trimestral, para o caso da amostra de dados adotada, possui clara vantagem em relação às outras estratégias (principalmente no caso de formação de carteiras limitadas a 10 ativos).

**Tabela 15:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 20 períodos.

Estatísticas	Ibovespa	Intervalo Out-of-Sample de 20 Períodos – Rebalanceamento Mensal					
		S/ Limite de Ativos	Limite 40 Ativos	Limite 30 Ativos	Limite 20 Ativos	Limite 10 Ativos	Limite 05 Ativos
Retorno Médio Por Carteira	0.003%	0.007%	0.008%	0.007%	0.006%	0.020%	0.021%
Retorno Acumulado	-3.757%	-0.865%	-0.177%	-0.974%	-1.483%	9.046%	8.867%
Volatilidade	1.251%	1.253%	1.253%	1.255%	1.275%	1.286%	1.326%
Volatilidade Anual	19.855%	19.894%	19.890%	19.916%	20.246%	20.420%	21.055%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987	0.980	0.959
Assimetria	-0.392	-0.395	-0.388	-0.390	-0.397	-0.338	-0.433
Curtose	1.755	1.706	1.701	1.678	1.576	1.555	1.812

Fonte: elaborado pelo autor.

**Tabela 16:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 40 períodos.

Estatísticas	Ibovespa	Intervalo Out-of-Sample de 40 Períodos – Rebalanceamento Bimestral					
		S/ Limite de Ativos	Limite 40 Ativos	Limite 30 Ativos	Limite 20 Ativos	Limite 10 Ativos	Limite 05 Ativos
Retorno Médio Por Carteira	0.003%	0.005%	0.005%	0.005%	0.008%	0.015%	0.028%
Retorno Acumulado	-3.757%	-2.365%	-2.264%	-2.137%	-0.335%	5.031%	14.679%
Volatilidade	1.251%	1.252%	1.253%	1.251%	1.274%	1.270%	1.304%
Volatilidade Anual	19.855%	19.883%	19.891%	19.854%	20.220%	20.156%	20.701%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987	0.977	0.957
Assimetria	1.755	1.793	1.798	1.718	1.812	1.548	2.030
Curtose	-0.392	-0.391	-0.387	-0.378	-0.400	-0.377	-0.502

Fonte: elaborado pelo autor.

**Tabela 17:** Resultados estatísticos para formação de Carteiras com intervalo fora da amostra de 60 períodos.

Estatísticas	Ibovespa	Intervalo Out-of-Sample de 60 Períodos – Rebalanceamento Trimestral					
		S/ Limite de Ativos	Limite 40 Ativos	Limite 30 Ativos	Limite 20 Ativos	Limite 10 Ativos	Limite 05 Ativos
Retorno Médio Por Carteira	0.003%	0.007%	0.007%	0.009%	0.008%	0.010%	0.022%
Retorno Acumulado	-3.757%	-0.729%	-0.308%	0.961%	-0.197%	1.118%	10.115%
Volatilidade	1.251%	1.255%	1.254%	1.256%	1.276%	1.277%	1.343%
Volatilidade Anual	19.855%	19.916%	19.901%	19.938%	20.258%	20.269%	21.323%
Correlação com Ibovespa	1.000	0.992	0.991	0.990	0.987	0.980	0.957
Assimetria	-0.392	-0.402	-0.400	-0.371	-0.383	-0.375	-0.412
Curtose	1.755	1.795	1.810	1.739	1.803	1.544	2.017

Fonte: elaborado pelo autor.

De forma complementar, pode-se ainda fazer uma análise desses resultados observando-se os pesos dos ativos que compõem as carteiras formadas de 05 e 10 ativos. Considerando rebalanceamento trimestral, conforme já informado, é possível formar um total de 12 carteiras para o período de realização dos testes. Na Tabela 23 (Apêndice D), são apresentados os ativos utilizados para formação de todas as 12 carteiras ao longo dos testes, para carteiras limitadas a 05 ativos; e, na Tabela 24, são informados os ativos usados para formação das carteiras limitadas a 10 ativos. E, inicialmente, nota-se a relevância dos ativos PETR4 e VALE5, que são os ativos mais usados tanto para carteiras de 05 quanto de 10 ativos. O ativo PETR4 é incluído em todas as carteiras, e o ativo VALE5 é incluído em 10 carteiras, representando assim as duas empresas de maior peso na composição do índice Ibovespa.

Vê-se também que, para o caso de 05 ativos, as carteiras formadas chegaram a ter ativos com até 33,3% de peso, ou seja, um terço da carteira. E o menor peso utilizado foi de 7,0% atribuído ao ativo CESP6 na formação de uma das 12 carteiras. Em média, o peso dos ativos nas carteiras ficou em torno de 16%, demonstrando que, apesar de haver apenas 05 ativos em cada carteira, não há uma forte concentração de peso em apenas um dos ativos.

Para carteiras de 10 ativos (Tabela 24), os dados demonstram que o maior peso atribuído a um ativo foi o caso de PETR4, que chegou a ter peso de 26,1%. Já o menor peso na composição de uma carteira foi de 3,1%, atribuído ao ativo BTOW3. Em média, os resultados demonstram que os ativos tiveram peso em torno de 7,6%. Novamente, pode-se concluir que, apesar de as carteiras serem relativamente pequenas (apenas 10 ativos), não há forte concentração de peso em alguns ativos, exceto em casos específicos com os ativos mais relevantes ao índice (PETR4 e VALE5).

### 5.3. Verificação dos valores de *gap* dos resultados obtidos com o algoritmo

A fim de se verificar a qualidade das soluções iniciais obtidas com a aplicação do algoritmo, essa dissertação adotou foco em analisar as soluções através de verificação dos *gaps* das respostas.

Considerando rebalanceamento de 20 períodos, pode-se formar um total de 36 carteiras para a amostra de dados adotada. Como há grande dificuldade para se verificar as soluções ótimas para todas essas 36 carteiras para limites de 05 e 10 ativos (devido ao alto tempo de otimização necessário via *solver* Cplex), decidiu-se por verificar os valores de *gap* por “amostragem”. Foram

selecionadas aleatoriamente 12 dessas 36 carteiras; e, para essas 12 carteiras, foi executada otimização direta via Cplex para verificação das respostas ótimas. Após isso, torna-se possível fazer a comparação entre as respostas obtidas via AG e as respostas ótimas para essas 12 carteiras; ou seja, com conhecimento das respostas ótimas, é viável a verificação do *gap* das soluções obtidas com o AG<sup>1</sup>. Dessa forma, pode-se ter uma clara noção da qualidade das respostas que estão sendo obtidas via algoritmo.

Após a obtenção das respostas ótimas para as 12 carteiras previamente selecionadas, foi feita a comparação entre essas soluções ótimas e as soluções obtidas com a heurística. Os resultados de *gap* estão expressos na Tabela 18 a seguir. Evidencia-se assim que, em termos de *gap*, as respostas iniciais obtidas com o AG são razoavelmente boas para as carteiras de 10 ativos e, principalmente, muito boas para carteiras de 05 ativos.

No caso de 10 ativos, temos valores de *gap* mais elevados; porém, o valor médio de *gap* permanece em 7,68%, abaixo de 10%, com *gap* mínimo de apenas 3%. Já para o caso de carteiras de 05 ativos, tem-se *gap* médio de 2,4%, bastante abaixo de 5%. Além disso, em 7 das 12 carteiras testadas, o algoritmo forneceu a resposta ótima para as otimizações. Embora o valor máximo de *gap* seja de 15,1%, trata-se de apenas 1 carteira que teve *gap* acima de 10%. Assim, pode-se considerar que os resultados da heurística são satisfatórios (considerando-se que respostas ao menos próximas das respostas ótimas – com *gap* médio abaixo de 10% – estão sendo obtidas com tempo computacional em torno de 5 minutos para o caso de problemas que seriam resolvidos na otimalidade pelo Cplex apenas com diversas horas de processamento).

**Tabela 18:** Resultados para verificação de *gap* das soluções obtidas com o algoritmo genético.

Carteiras com Limite de 05 e 10 Ativos - Gap das soluções		
Tamanho da carteira	05 Ativos*	10 Ativos**
Gap Médio	2,442%	7,684%
Gap Mínimo	0,000%	3,046%
Gap Máximo	15,128%	17,576%
Desvio Padrão	4,435%	4,046%

\***Carteiras de 05 Ativos:** algoritmo forneceu solução ótima para 7 das 12 carteiras testadas; e 1 carteira ficou com *gap* acima de 10%.

\*\***Carteiras de 10 Ativos:** algoritmo forneceu solução com *gap* acima de 10% para 2 das 12 carteiras testadas.

**Fonte:** elaborado pelo autor.

<sup>1</sup> Forma de cálculo do *Gap* para cada uma das 36 carteiras calculadas:  $Gap = (F_{AG} - F_o) / F_o$

Sendo:  $F_{AG}$  = Resultado da função objetivo do algoritmo genético; e  $F_o$  = Resultado da função objetivo para solução ótima.

Dessa forma, conclui-se intuitivamente que, principalmente para o caso de carteiras de 05 ativos, o algoritmo está fornecendo, em média, soluções bastante próximas das soluções ótimas. Assim, em relação às curvas de projeção das carteiras de 05 ativos nas Figuras 12, 13 e 14 no Apêndice B (que estão “descoladas” das curvas do índice e demais carteiras), conclui-se que, em princípio, não é possível que sejam obtidas soluções muito melhores. Apesar de as curvas estarem afastadas das curvas do índice Ibovespa para os três casos de rebalanceamento (mensal, bimestral e trimestral), pode-se tomar a conclusão de que não é possível que as curvas para 05 ativos se aproximem muito mais do índice; pelos resultados referentes a valores de *gap*, há a tendência de que as respostas obtidas estejam bastante próximas das respostas ótimas e, por consequência, os erros de *tracking* não podem ser diminuídos de forma considerável.

#### 5.4. Alterações de parâmetros e do tempo de execução do algoritmo

Mesmo levando-se em consideração a boa qualidade das respostas iniciais obtidas pelo algoritmo, foram realizados novos testes com alteração de parâmetros do AG e, também, com alteração do tempo de execução do algoritmo. O objetivo desses testes foi verificar de que forma essas alterações poderiam influenciar os resultados. Novamente, os testes foram realizados para as 12 carteiras selecionadas aleatoriamente, para as quais foram encontradas as respostas ótimas (conforme descrito na seção acima); assim, torna-se possível analisar os resultados do AG em termos de *gap* das soluções.

A Tabela 19 abaixo descreve as alterações realizadas no AG. Nessa tabela, constam os parâmetros e o tempo de execução que foram adotados para os testes iniciais (que foram realizados conforme descrito nas seções 5.2 e 5.3) e os novos parâmetros e tempo de execução para três novos testes. No Teste 1 e Teste 2, têm-se alterações dos parâmetros *Num\_Popul*, *Num\_Evol*, *Taxa\_Mut* e *Num\_Mut*; e, nesses Testes 1 e 2, foi mantido tempo de execução do algoritmo em torno de 5 minutos. Já no Teste 3, o tempo de execução do algoritmo foi aumentado para 8 minutos.



**Tabela 19:** Descrição dos novos testes do algoritmo com **carteiras de limites de 05 e 10 ativos**, com alteração de parâmetros e do tempo de execução.

Tamanho da Carteira (Parâmetro " <i>Tam_Cart</i> ")		05 Ativos	10 Ativos	
<b>Teste Inicial</b> (Descrito na seção 5.2)	Tempo de execução: <b>5 minutos</b>	<i>Num_Popul</i>	20	20
		<i>Taxa_Crossover</i>	100%	100%
		<i>Taxa_Mut</i>	85%	85%
		<i>Num_Mut</i>	1	1
		<i>Num_Evol</i>	30	25
<b>Teste 1</b>	Tempo de execução: <b>5 minutos</b>	<i>Num_Popul</i>	15	15
		<i>Taxa_Crossover</i>	100%	100%
		<i>Taxa_Mut</i>	85%	85%
		<i>Num_Mut</i>	1	1
		<i>Num_Evol</i>	35	30
<b>Teste 2</b>	Tempo de execução: <b>5 minutos</b>	<i>Num_Popul</i>	10	10
		<i>Taxa_Crossover</i>	100%	100%
		<i>Taxa_Mut</i>	90%	90%
		<i>Num_Mut</i>	2	2
		<i>Num_Evol</i>	50	45
<b>Teste 3</b>	Tempo de execução: <b>8 minutos</b>	<i>Num_Popul</i>	20	20
		<i>Taxa_Crossover</i>	100%	100%
		<i>Taxa_Mut</i>	80%	80%
		<i>Num_Mut</i>	1	1
		<i>Num_Evol</i>	50	45

**Fonte:** elaborado pelo autor.

Os resultados para os testes propostos na Tabela 19 estão resumidos a seguir na Tabela 20 (Carteiras com limite de 05 ativos) e na Tabela 21 (Carteiras com limite de 10 ativos). De modo geral, a principal conclusão em relação aos Testes 1 e 2 é que as alterações efetuadas nos parâmetros não influenciaram significativamente as soluções do algoritmo. Para o caso de carteiras de 5 ativos, os valores de *gap* máximo e desvio-padrão aumentam nos Testes 1 e 2, apesar de o algoritmo seguir fornecendo soluções ótimas para alguns casos. Da mesma forma, os valores de erro médio de *tracking* aumentam, evidenciando maior oscilação de rendimento das carteiras em relação ao índice. Porém, cabe ressaltar que essa piora nos resultados é bastante pequena, observando-se que, em média, as soluções obtidas permanecem boas especialmente se for considerado que o tempo de solução permanece em torno de 5 minutos, o que é um tempo bastante baixo. Apesar disso, há ligeira vantagem nas soluções do Teste Inicial, demonstrando que o uso de uma população maior (parâmetro *Num\_Popul*) tende a gerar melhores resultados.

Já no caso do Teste 3, em que o tempo de execução do algoritmo é aumentado em 3 minutos, percebe-se que esse aumento de tempo melhora a qualidade das soluções em termos de *gap*. No caso das carteiras de 5 ativos, o *gap* médio das soluções cai para apenas 1%, com *gap*

máximo abaixo de 5%; ou seja, todas as 12 carteiras geradas apresentam *gap* abaixo de 5%, sendo que a solução ótima foi obtida em 8 das 12 carteiras testadas.

**Tabela 20:** Resultados para testes com alterações de parâmetros no algoritmo – Formação de carteiras com limite de **05 ativos**.

<b>Carteiras com Limite de 05 Ativos e Rebalanceamento (Intervalo Out-of-Sample) de 20 Períodos</b>				
	<b>Teste Inicial</b>	<b>Teste 1*</b>	<b>Teste 2**</b>	<b>Teste 3***</b>
<b>Gap das soluções</b>				
<b>Gap Médio</b>	2.44%	4.02%	4.71%	1.05%
<b>Gap Mínimo</b>	0.00%	0.00%	0.00%	0.00%
<b>Gap Máximo</b>	15.13%	23.02%	15.66%	4.73%
<b>Desvio-padrão</b>	4.44%	6.72%	6.34%	1.69%
<b>Tracking Error (período out-of-sample)</b>				
<b>Média</b>	0.370%	0.584%	0.456%	0.366%
<b>Mínimo</b>	-3.725%	-2.802%	-3.914%	-5.967%
<b>Máximo</b>	3.246%	4.939%	4.048%	4.197%
<b>Desvio-padrão</b>	1.772%	2.066%	1.613%	1.856%

\***Teste 1:** algoritmo forneceu solução ótima para 6 das 12 carteiras testadas. Para 2 carteiras, o *gap* ficou acima de 10%.

\*\***Teste 2:** algoritmo forneceu solução ótima para 7 das 12 carteiras testadas. Para 3 carteiras, o *gap* ficou acima de 10%.

\*\*\***Teste 3:** algoritmo forneceu solução ótima para 8 das 12 carteiras testadas.

**Fonte:** elaborado pelo autor.

Para o caso de carteiras de 10 ativos, conforme Tabela 21 abaixo, os valores de *gap* médio oscilam em diferentes sentidos nos Testes 1 e 2; no Teste 1, o *gap* médio aumenta em relação ao Teste Inicial; já em relação ao Teste 2, o *gap* médio diminui. Nesses testes, o algoritmo passou a encontrar soluções ótimas para algumas carteiras, porém houve aumento do desvio-padrão. Assim, a principal conclusão que se pode tomar é que essas alterações de parâmetros não influenciam significativamente os resultados do algoritmo.

Já no caso do Teste 3, em que o tempo de execução do algoritmo é aumentado em mais três minutos, percebe-se que esse aumento de tempo também melhora a qualidade das soluções em termos de *gap* (da mesma forma que ocorre com carteiras de 5 ativos). No caso das carteiras de 10 ativos, o *gap* médio das soluções cai para 3,77%, com *gap* máximo abaixo de 10%; ou seja, todas as 12 carteiras geradas apresentam *gap* abaixo de 10%, além de o algoritmo ter fornecido a solução ótima para 1 carteira. Assim, nota-se novamente que um pequeno aumento no tempo de execução do algoritmo pode ter impacto considerável nas soluções obtidas.

**Tabela 21:** Resultados para testes com alterações de parâmetros no algoritmo – Formação de carteiras com limite de **10 ativos**.

Carteiras com Limite de <b>10 Ativos</b> e Rebalanceamento (Intervalo Out-of-Sample) de 20 Períodos				
	Teste Inicial	Teste 1*	Teste 2**	Teste 3***
<b>Gap das soluções</b>				
<b>Gap Médio</b>	7.68%	7.95%	6.61%	3.78%
<b>Gap Mínimo</b>	3.05%	0.00%	0.00%	0.00%
<b>Gap Máximo</b>	17.58%	19.71%	17.02%	7.86%
<b>Desvio-padrão</b>	4.05%	5.63%	6.45%	2.21%
<b>Tracking Error (período out-of-sample)</b>				
<b>Média</b>	0.345%	0.295%	0.375%	0.318%
<b>Mínimo</b>	-2.420%	-1.794%	-2.132%	-1.960%
<b>Máximo</b>	3.148%	3.197%	3.197%	3.197%
<b>Desvio-padrão</b>	1.262%	1.273%	1.324%	1.284%

\***Teste 1:** algoritmo forneceu solução ótima para 1 das 12 carteiras testadas. Para 4 carteiras, o *gap* ficou acima de 10%.

\*\***Teste 2:** algoritmo forneceu solução ótima para 2 das 12 carteiras testadas. Para 3 carteiras, o *gap* ficou acima de 10%.

\*\*\***Teste 3:** algoritmo forneceu solução ótima para 1 das 12 carteiras testadas.

**Fonte:** elaborado pelo autor.

## 5.5. Considerações sobre a aplicação da heurística de Algoritmos Genéticos

Conforme se verificou no Capítulo 4 dessa dissertação, existe a necessidade de se aplicar um método heurístico para a solução do problema de *index tracking* com carteiras de número reduzido de ativos. No caso da amostra adotada de 67 ativos, a heurística tornou-se necessária para formação de carteiras de 5 e 10 ativos – pois a obtenção da solução ótima nesse problema apenas com uso do *solver* Cplex necessita de diversas horas de processamento computacional. Dessa forma, o Capítulo 5 foi desenvolvido para aplicação do método heurístico de AG.

Pela aplicação desse método, foi possível solucionar o problema de *index tracking*, para formação de carteiras de 5 e 10 ativos, com tempo de execução do algoritmo em torno de cinco minutos. E, para esse tempo de solução, foi possível constatar que o algoritmo forneceu inicialmente boas soluções em termos de *gap* e também de erro de *tracking*.

Nos testes iniciais realizados, o algoritmo aplicado foi capaz de solucionar o problema fornecendo respostas ótimas para algumas das carteiras testadas com apenas cinco minutos de processamento para o caso de carteiras limitadas a 5 ativos. Para carteiras de 5 ativos, os erros de *tracking* ficaram maiores que para o caso de carteiras de 10 ativos. Essa diferença pode ser

percebida nas projeções das carteiras nas Figuras 12, 13 e 14 (Apêndice B), em que as curvas para portfólio de 5 ativos estão mais afastadas do índice.

Porém, apesar dos erros maiores para carteiras de 5 ativos, a verificação dos valores de *gap* demonstrou que esse erro maior não pode ser melhorado significativamente, considerando que as respostas obtidas já estão bastante próximas das respostas ótimas. Para o caso de carteiras limitadas a 10 ativos, o algoritmo não forneceu soluções ótimas para as carteiras testadas; porém, os resultados demonstram que o *gap* médio ficou abaixo de 10%, com *gap* acima de 10% para apenas 2 de 12 carteiras testadas.

Após a realização de um Teste Inicial (para carteiras de 5 e 10 ativos), foram executados três novos testes com alterações de parâmetros do algoritmo e também alteração do tempo de processamento. Com esses novos testes, foi possível concluir que o algoritmo tende a gerar melhores soluções quando utiliza tamanho maior de população (parâmetro *Num\_Popul*). Porém, essa diferença de qualidade entre as soluções é bastante pequena principalmente no caso de carteiras de 10 ativos. Para o caso de carteiras de 5 ativos, a diminuição do valor do parâmetro *Num\_Popul* levou o algoritmo a gerar soluções um pouco piores, porém ainda com baixos valores de *gap* e de erro de *tracking*.

Já para o caso do Teste 3, em que houve aumento do tempo de execução do AG de cinco para oito minutos, tem-se alterações maiores nos resultados, com diferenças bastante claras principalmente nos valores de *gap*. As Tabelas 20 e 21 mostram que esse aumento de três minutos no tempo de execução permitiu que o algoritmo melhorasse significativamente suas respostas. Vê-se que os valores de *gap* médio diminuem tanto para carteiras de 5 quanto para carteiras de 10 ativos; em nenhum dos casos, o *gap* médio fica acima de 5%. Além disso, o *gap* máximo diminuiu para os dois tipos de carteiras, e nenhuma carteira formada possui *gap* maior que 10%.

Dessa forma, verificou-se que o algoritmo atende as necessidades e cumpre com os objetivos inicialmente propostos. Com a aplicação dessa heurística usando-se a metodologia descrita na seção 5.1, foi possível formar carteiras com número reduzido de ativos e com valores médios de *gap* abaixo de 10%, demonstrando o bom funcionamento do AG com um tempo de processamento bastante baixo (comparando-se com o tempo que seria necessário para resolver esse problema na otimalidade apenas com aplicação direta do *solver* Cplex).

## 6. CONCLUSÕES

Essa dissertação de mestrado foi proposta com o intuito de estudar a estratégia passiva de investimento de *index tracking*. Para tal, os fundamentos que justificam o uso de estratégias passivas estão no conceito de eficiência de mercados apresentada por Fama (1970) e em estudos como Frino e Gallagher (2001) e Fama e French (2010), que demonstram que estratégias passivas tendem a ter melhores resultados que estratégias ativas de investimento no mercado acionário.

Assim, foram apresentados dois objetivos principais: (1) apresentar um modelo de otimização de *index tracking* (realizando minimização do erro de *tracking* de uma carteira em relação ao índice de referência) e (2) aplicar um método heurístico a fim de solucionar esse modelo com tempos relativamente baixos de processamento (em torno de 10 minutos no máximo) e, ao mesmo tempo, gerando soluções próximas da otimalidade (*gap* de menos de 5% em média). Assim, esperava-se poder resolver o problema de otimização de *index tracking* de forma relativamente rápida e eficiente, especialmente para carteiras de reduzida quantidade de ativos. Dessa forma, torna-se possível atender alguns pontos que a literatura apresenta como relevantes em relação ao problema de *index tracking*, como minimizar o erro entre o rendimento de um portfólio e do índice, formar carteiras de reduzido número de ativos (05 ou 10 ativos no caso da amostra de dados utilizada) e minimizar os custos de transação (o que é feito indiretamente através da limitação da quantidade de ativos para formação das carteiras).

A fim de cumprir esses objetivos, foi realizada, inicialmente, uma revisão bibliográfica para análise de métodos de solução que poderiam ser utilizados. Essa revisão literária permitiu a tomada de decisão de se aplicar o método heurístico de Algoritmos Genéticos, aplicado de forma ampla em artigos contemporâneos sobre o assunto, e possibilitando a obtenção de boas soluções (como nos trabalhos de Beasley, Meade e Chang, 2003; e Oh, Kim e Min, 2005). Além disso, a revisão bibliográfica demonstrou algumas das conclusões esperadas após realização dos testes computacionais, como os fatos de que os valores de erro de *tracking* tendem a aumentar com diminuição da quantidade de ativos das carteiras; além disso, a literatura com uso de algoritmos genéticos demonstrou a possibilidade de se obter soluções ótimas através do uso desse método heurístico.

Para realização de todos os testes computacionais, foi adotado o índice Ibovespa como referência, utilizando-se 67 ativos conforme Tabela 22 (Apêndice C). O período adotado foi de Jan/2009 a Jul/2012, compreendendo 890 observações diárias de preços dos ativos; a projeção das carteiras otimizadas no período de Ago/2009 a Jul/2012, com uso essencialmente de 150 observações diárias de preços como amostra *in-sample* para otimização do modelo e formação das

carteiras. Para projeção das carteiras, foram definidos intervalos de rebalanceamento de 20, 40 e 60 períodos (ou seja, basicamente rebalanceamento das carteiras em intervalo mensal, bimestral e trimestral).

Os testes iniciais realizados com o modelo de otimização foram apresentados no Capítulo 4, e demonstraram essencialmente a limitação do uso de otimização direta via *solver* Cplex para solução do modelo em relação ao tempo de processamento necessário. Conforme a literatura consultada, como se trata de um modelo de otimização de programação quadrática que inclui restrição inteira (binária – representando a limitação do tamanho das carteiras), o tempo de processamento cresce exponencialmente na medida em que se diminui o tamanho das carteiras.

Inicialmente, foram formadas carteiras sem limite de ativos e limitadas a 40, 30 e 20 ativos, através de realização de otimização direta com uso do *solver* Cplex. Algumas constatações principais foram que os valores de *tracking error* não diminuem conforme as carteiras são reduzidas a 20 ativos (conforme Tabelas 01 a 04), possibilitando concluir que, para a amostra de dados selecionada, carteiras de 20 ativos são capazes de realizar o *tracking* do índice com bastante qualidade tanto para rebalanceamento mensal, bimestral ou trimestral. Conforme verificado nas Tabelas 05 a 07, o retorno acumulado ao longo dos quase três anos de projeção das carteiras não passa de 3,5 pontos percentuais, no caso de carteiras de 20 ativos e rebalanceamento trimestral. As projeções apresentadas nas Figuras 09 a 11 ajudam a corroborar a qualidade do modelo de otimização e a boa resposta apesar da diminuição da carteira para até 20 ativos. Além disso, verifica-se que o aumento do período de rebalanceamento implica maior erro de *tracking*, porém isso não prejudica consideravelmente os resultados finais de *tracking*.

Apesar disso, foi possível constatar que não é possível formar carteiras de 20 ou menos ativos com tempo de processamento inferior a 1 hora e *gap* inferior a 5%. Já para carteiras de 20 ativos, o tempo de processamento ultrapassa facilmente 1 hora com *gap* médio de solução acima de 11%. Novos testes foram realizados alterando-se o tamanho da amostra de dados *in-sample* e, também, alterando-se parâmetros de configuração do *solver* Cplex (seções 4.2 e 4.3); porém, essas alterações não resultaram em variações significativas nos resultados, em termos de tempo de processamento, comprovando a necessidade de aplicação de um método heurístico para solução do problema de forma mais ágil.

Dadas essas conclusões obtidas a partir do Capítulo 4, o Capítulo 5 apresenta a aplicação do método heurístico de Algoritmos Genéticos. Os testes realizados com o algoritmo buscaram formar carteiras limitadas a 05 e 10 ativos, com tempo de processamento em torno de 5 minutos, e os resultados iniciais demonstraram que o erro de *tracking* aumenta bastante no caso de carteiras

de 05 ativos. As Tabelas 13 e 14 permitem verificar que, para carteiras de 10 ativos, têm-se *tracking error* médio de 0,46% (considerando rebalanceamento trimestral, por exemplo); e, para carteiras de 05 ativos, têm-se *tracking error* médio de 1,21%. As projeções nas Figuras 12 a 14 (Apêndice B) ajudam a corroborar o fato de que o *tracking* com 10 ativos foi bastante eficiente especialmente para o caso de rebalanceamento trimestral. Porém, para o caso de carteiras de 05 ativos, o erro de *tracking* torna-se elevado, acumulando uma diferença total de rendimento acumulado da carteira em relação ao índice (durante todo o período de quase três anos de projeção das carteiras) em torno de 13,87 pontos percentuais (Tabela 17). Assim, a conclusão central nesse ponto é de que, para a amostra de dados adotada, a execução do *tracking* do Ibovespa ainda é possível com 10 ativos, porém torna-se bastante difícil com apenas 05 ativos. Apesar disso, pode considerar como forte atenuante o fato de que o erro tende a ser positivo, de forma que o rendimento acumulado pelas carteiras de *tracking* tende a ficar acima do rendimento do índice.

Apesar dos altos valores de erro especialmente para o caso de carteiras de 05 ativos, a seção 5.3 foi elaborada para verificar o gap das soluções obtidas nos testes iniciais realizados. Para isso, do total de 36 carteiras formadas no período selecionado, foram adotadas 12 carteiras aleatoriamente, para que fossem analisados os *gaps* de suas soluções. Assim, foi possível constatar que o *gap* médio das soluções obtidas para carteiras de 05 ativos foi de apenas 2,44%; para o caso de carteiras de 10 ativos, o *gap* médio foi de 7,68%, mantendo-se abaixo de 10%. Porém, verificou-se que, em alguns casos, o *gap* das soluções permaneceu acima de 10%.

Dessa forma, a seção 5.4 apresenta três novos testes que foram realizados para tentativa de obtenção de soluções de menor *gap*. Foram realizados três novos testes, com alteração de parâmetros do algoritmo e, também, com alteração do tempo de processamento (passando de 5 para 8 minutos). Verificou-se que apenas a alteração de parâmetros do algoritmo, como a diminuição do número de indivíduos da população e o aumento de ciclos de evolução, não é suficiente para alterar significativamente a qualidade das soluções. Porém, para o teste em que o tempo de processamento é elevado para em torno de 8 minutos, nota uma diferença considerável em relação ao *gap* das soluções. Com maior tempo de processamento, o *gap* médio para carteiras de 05 ativos passou de 2,44% para 1,05%, com nenhuma das 12 carteiras verificadas apresentando *gap* acima de 5%; no caso de carteiras de 10 ativos, o *gap* médio passou de 7,68% para 3,78%, com *gap* máximo de 7,86%.

Porém, como contraponto, é preciso ressaltar que as soluções geradas pelo algoritmo são melhores para carteiras de 05 ativos do que para carteiras de 10 ativos, em termos de *gap* das soluções. Dessa forma, nota-se a tendência de que os resultados do algoritmo percam qualidade para formação de carteiras maiores. Assim, recomenda-se o uso da heurística apenas para

formação de carteiras de número bastante reduzido de ações, pois o *solver* Cplex pode formar carteiras maiores com qualidade e rapidez (carteiras de 30 ativos ou mais, para o caso da amostra adotada). Ao mesmo tempo, deve-se também ressaltar que a definição dos parâmetros adotados ao longo desse estudo, principalmente na aplicação do algoritmo, é algo bastante singular; deve-se considerar que a parametrização utilizada no algoritmo adequou-se à base de dados adotada nesse estudo. Por esse motivo, a reprodução do modelo com outras bases de dados tende a exigir novos testes para “calibrar” os parâmetros do algoritmo.

Assim, as conclusões finais são de que o método heurístico aplicado está apresentando respostas com *gap* abaixo de 10% na maioria dos casos, entregando inclusive as respostas ótimas para algumas carteiras. Dessa forma, pode-se perceber a qualidade do método de algoritmos genéticos para solução do problema de *index tracking* (como ocorreu, por exemplo, em Jeurissen e van den Berg, 2008). Nota-se também que, apesar de os valores de erro de *tracking* especialmente para o caso de carteiras de 05 ativos serem um pouco elevados (erro médio acima de 1%), há a tendência de que esse erro não possa ser diminuído de forma considerável, pois a heurística já está fornecendo respostas bastante próximas da otimalidade na maioria dos casos.

Portanto, pode-se concluir que os objetivos principais dessa dissertação foram cumpridos, através da apresentação de um modelo de otimização para minimização do erro de *tracking* de uma carteira em relação ao índice, e também da apresentação de uma heurística capaz de fornecer soluções bastante próximas da otimalidade mesmo com tempo de processamento bastante baixo, de apenas 5 minutos.

### 6.1. Pesquisas Futuras

Pode-se citar que a pesquisa realizada com essa dissertação possui uma limitação bastante forte, que é o uso de apenas um índice de referência para testes do modelo e da heurística. Esse fato justifica-se por uma questão de limitação de base de dados acessível. Portanto, uma sugestão de pesquisa futura seria replicar a aplicação do modelo e do método heurístico para outras bases de dados, especialmente com uso de índices que empregam quantidades maiores de ativos (como, por exemplo, o índice S&P 500). Dessa forma, seria possível verificar se a qualidade das soluções obtidas permaneceria para amostras de dados maiores.

Outras sugestões de pesquisas futuras seriam a aplicação do modelo para tentativa de replicação de um indicador de mercado como, por exemplo, a inflação – o que também não é feito devido à limitação de bases de dados disponíveis; além disso, para replicar a inflação, seria



necessário também alterações no modelo de otimização, para consideração de operações de curto prazo e outras peculiaridades.

Por fim, outra sugestão de pesquisa futura seria a ampliação do modelo de otimização para inclusão de restrições relativas a custos de transação e controle de *turnover*, por exemplo. Dessa forma, seria possível refletir as condições reais de mercado de forma mais detalhada.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- Barro, Diana; Canestrelli, Elio. **Tracking error: a multistage portfolio model**. *Annals of Operations Research*. Vol. 165(1), pp. 47-66, 2009.
- Barro, Diana; Canestrelli, Elio. Tracking error with minimum guarantee constraints. In: Corazza, Marco; Pizzi, Claudio (Eds.). **Mathematical and Statistical Methods for Actuarial Sciences and Finance**. Springer-Verlag: Italia, pp. 13-21, 2010.
- Beasley, J. E.; Meade, N.; Chang, T. J. **An evolutionary heuristic for the index tracking problem**. *European Journal of Operational Research*. Vol. 148, pp. 621-643, 2003.
- Borges, Elaine C.; Eid Jr., Willian; Yoshinaga, Cláudia E. **Exchange Traded Funds versus Fundos Indexados no Brasil**. *Revista de Finanças Aplicadas*. Vol. 1, pp. 1-15, 2012.
- Caldeira, João F.; Portugal, Marcelo S. **Estratégia Long-Short, Neutra ao Mercado, e Index Tracking Baseadas em Portfólios Cointegrados**. *Revista Brasileira de Finanças*. Vol. 8(4), pp. 469-504, 2010.
- Canakgoz, N. A.; Beasley, J. E. **Mixed-integer programming approaches for index tracking and enhanced indexation**. *European Journal of Operational Research*. Vol. 196(1), pp. 384-399, 2009.
- Chen, Chen; Kwon, Roy H. **Robust portfolio selection for index tracking**. *Computers & Operations Research*. Vol. 39, pp. 829-837, 2012.
- Chiang, Wai C. Optimizing Performance. In: Neubert, A. (Ed.). **Indexing for Maximum Investment Results**. GPCo: Chicago, pp. 307-335, 1998.
- Coleman, Thomas F.; Li, Yuying; Henniger, Jay. **Minimizing tracking error while restricting the number of assets**. *Journal of Risk*. Vol. 8(4), pp. 33-56, 2006.
- Corielli, Francesco; Marcellino, Massimiliano. **Factor based index tracking**. *Journal of Banking & Finance*. Vol. 30, pp. 2215-2233, 2006.
- Dunis, Christian L.; Ho, Richard. **Cointegration Portfolios of European Equities for Index Tracking and Market Neutral Strategies**. *Journal of Asset Management*. Vol. 6(1), pp. 33-52, 2005.

- Fama, Eugene. **Efficient Capital Markets: A Review of Theory and Empirical Work**. Journal of Finance. Vol. 25(2), pp. 383-417, 1970.
- Fama, Eugene; French, Kenneth. **Luck versus Skill in the Cross-Section of Mutual Fund Returns**. The Journal of Finance. Vol. 65(5), pp. 1915-1947, 2010.
- Frino, Alex; Gallagher, David R. **Tracking S&P 500 Index Funds**. Journal of Portfolio Management. Vol. 28(1), pp. 44-55, 2001.
- Gaivoronski, A. Alexei; Krylov, Sergiy; van der Wijst, Nico. **Optimal portfolio selection and dynamic benchmark tracking**. European Journal of Operational Research. Vol. 163(1), pp. 115-131, 2005.
- Goldberg, David. E. **Genetic Algorithms in Search, Optimization, and Machine Learning**. New York: Addison-Wesley, 1989.
- Guastaroba, G.; Speranza, M. G. **Kernel Search: An application to the index tracking problem**. European Journal of Operational Research. Vol. 217, pp. 54-68, 2012.
- Jansen, Roel; van Dijk, Ronald. **Optimal Benchmark Tracking with Small Portfolios**. The Journal of Portfolio Management. Vol. 28(2), pp. 33-39, 2002.
- Jeurissen, Roland; van den Berg, Jan. **Optimized Index Tracking using a Hybrid Genetic Algorithm**. IEEE Congress on Evolutionary Computation, pp. 2327-2334, 2008.
- Krink, Thiemo; Mittnik, Stefan; Paterlini, Sandra. **Differential evolution and combinatorial search for constrained index-tracking**. Annals of Operations Research. Vol. 172, pp. 153-176, 2009.
- Liang, Jianfeng. **Tracking Models for Optioned Portfolio Selection**. Cutting-Edge Research Topics on Multiple Criteria Decision Making. Vol. 35, pp. 729-736, 2009.
- Maringer, Dietmar; Oyewumi, Olufemi. **Index Tracking with Constrained Portfolios**. Intelligent Systems in Accounting, Finance and Management. Vol. 15, pp. 57-71, 2007.
- Markowitz, Harry. **Portfolio Selection**. Journal of Finance. Vol. 7(1), pp. 77-91, 1952.
- Oh, Kyong Joo; Kim, Tae Yoon; Min, Sungky. **Using genetic algorithm to support portfolio optimization for index fund management**. Expert Systems with Applications. Vol. 28, pp. 371-379, 2005.

Roll, Richard. **A Mean/Variance Analysis of Tracking Error**. Journal of Portfolio Management. Vol. 18(4), pp. 12-22, 1992.

Ross, Stephen A. **Comment on the Modigliani-Miller Propositions**. Journal of Economic Perspectives. Vol. 2(4), pp. 127-133, 1988.

Scozzari, Andrea; Tardella, Fabio; Paterlini, Sandra; Krink, Thiemo. **Exact and Heuristic Approaches for the Index Tracking Problem with UCITS Constraints**. Annals of Operations Research. Vol. 16(1), 2012.

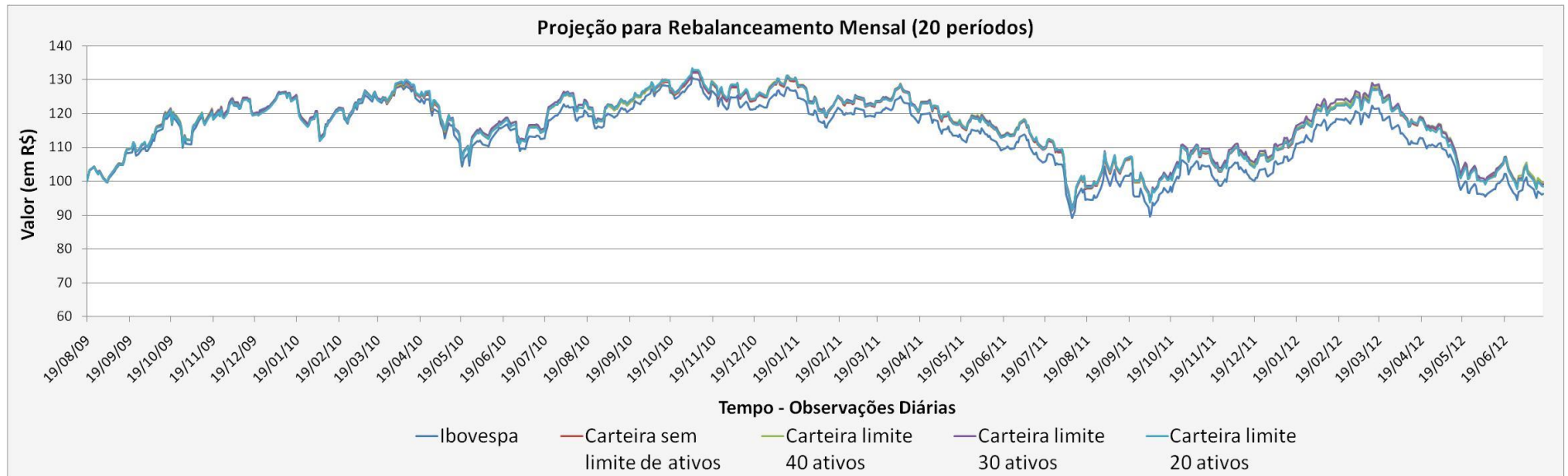
Stoyan, Stephen J.; Kwon, Roy H. **A two-stage stochastic mixed-integer programming approach to the index tracking problem**. Optimization and Engineering. Vol. 11(2), pp. 247-275, 2010.

Yao, David D.; Zhang, Shuzhong; Zhou, Xun Yu. **Tracking a Financial Benchmark Using a Few Assets**. Operations Research. Vol. 54(2), pp. 232-246, 2006.

Yu, Lian; Zhang, Shuzhong; Zhou, Xun Yu. A downside risk analysis based on financial index tracking models. In: Shiryayev, A. N.; Grossinho, M. R.; Oliveira, P. E.; Esquivel, M. L. (Eds.). **Stochastic Finance**. Springer: Berlin, pp. 213-236, 2006.

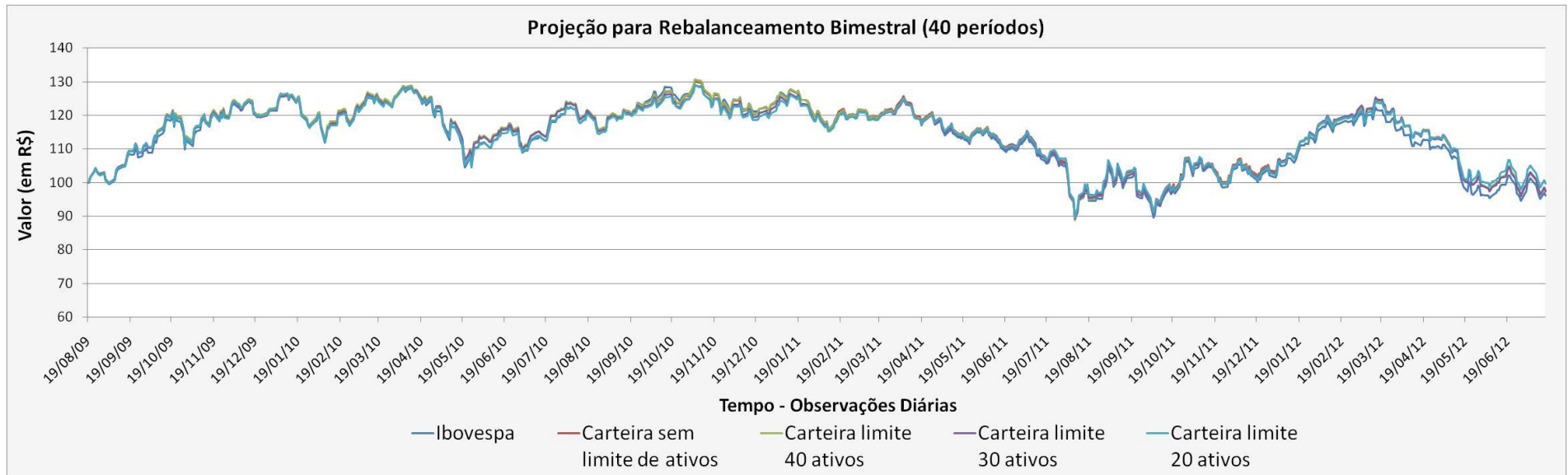
## 8. APÊNDICES

**APÊNDICE A** – Evolução das Carteiras (em R\$) para projeções com intervalos *out-of-sample* de 20, 40 e 60 observações diárias (Rebalanceamento Mensal, Bimestral e Trimestral) – Período de projeção: 19/08/2009 a 17/07/2012 – Gráficos com Índice, Carteira Sem Limite de Ativos e Carteiras com Limites de 40, 30 e 20 Ativos.



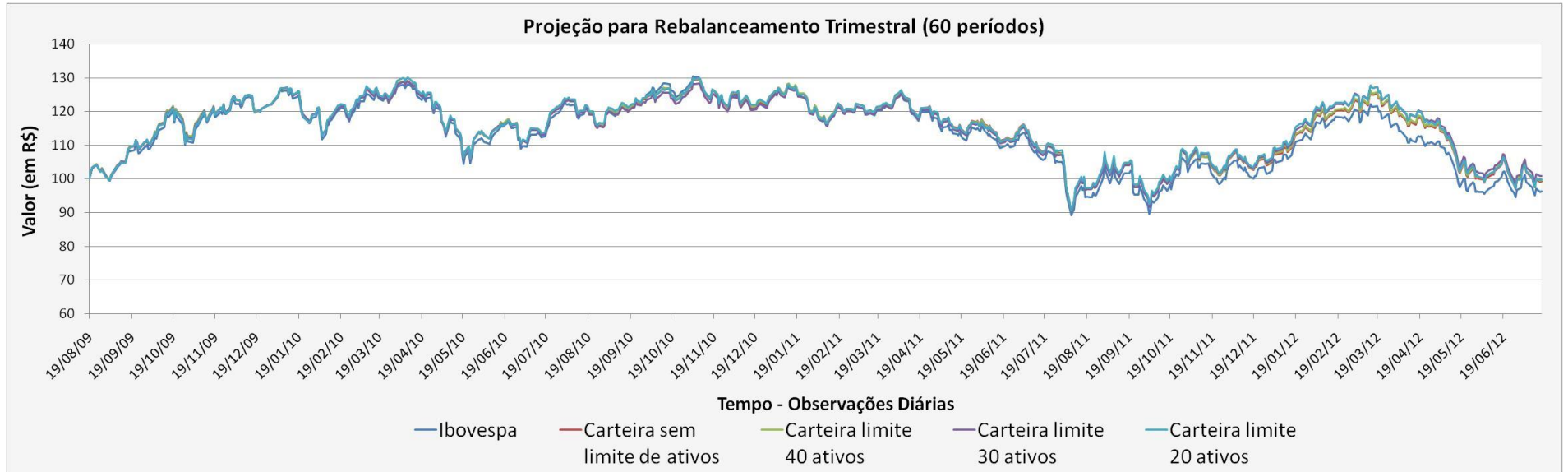
**Figura 09:** Projeção de resultados das carteiras com intervalo fora da amostra de 20 períodos, Rebalanceamento Mensal, Carteiras de 20 ou mais ativos.

**Fonte:** elaborado pelo autor.



**Figura 10:** Projeção de resultados das carteiras com intervalo fora da amostra de 40 períodos, Rebalanceamento Bimestral, Carteiras de 20 ou mais ativos.

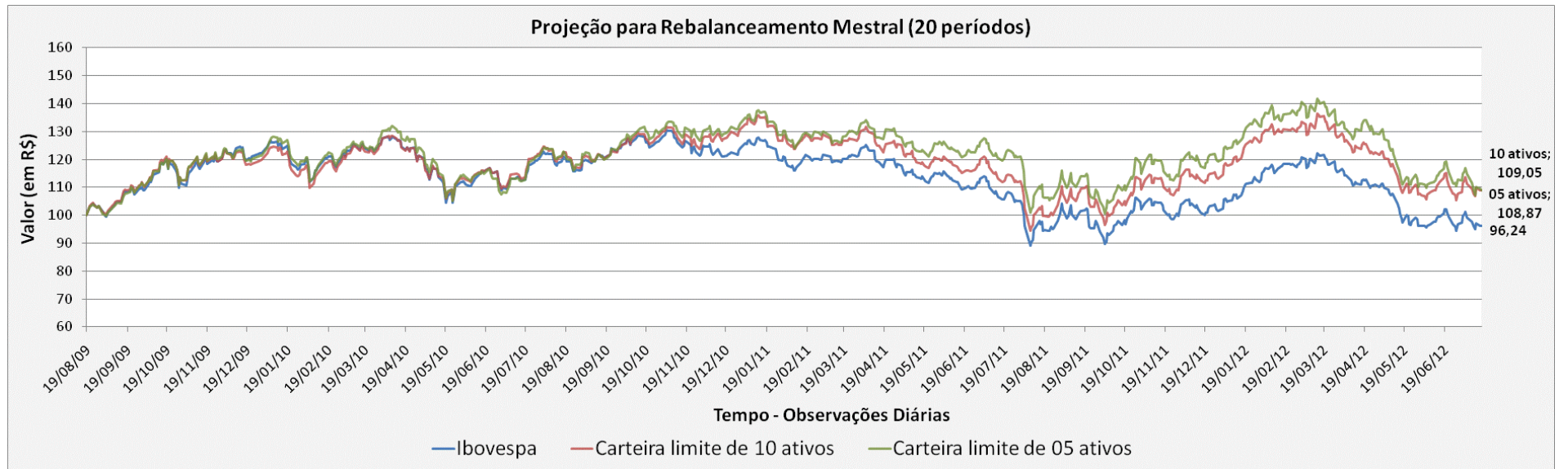
**Fonte:** elaborado pelo autor.



**Figura 11:** Projeção de resultados das carteiras com intervalo fora da amostra de 60 períodos, Rebalanceamento Trimestral, Carteiras de 20 ou mais ativos.

**Fonte:** elaborado pelo autor.

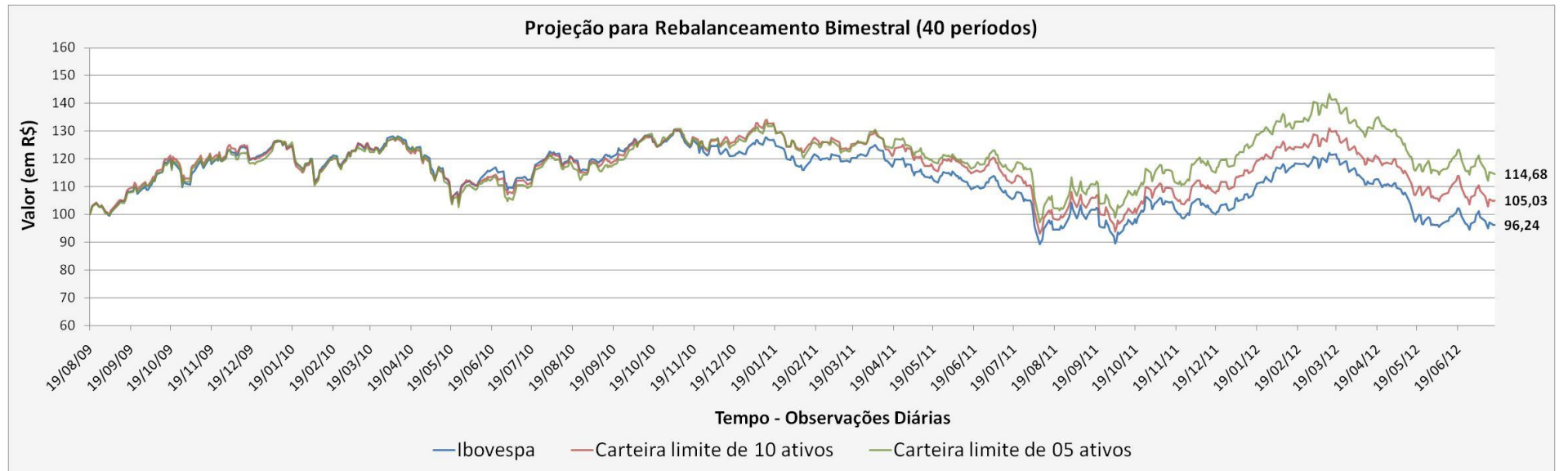
**APÊNDICE B** – Evolução das Carteiras (em R\$) para projeções com intervalos *out-of-sample* de 20, 40 e 60 observações diárias (Rebalanceamento Mensal, Bimestral e Trimestral) – Período de projeção: 19/08/2009 a 17/07/2012 – Gráficos com Índice e Carteiras com Limites de 10 e 05 Ativos.



**Figura 12:** Projeção de resultados das carteiras com intervalo fora da amostra de 20 períodos, Rebalanceamento Mensal, Carteiras de 05 e 10 ativos.

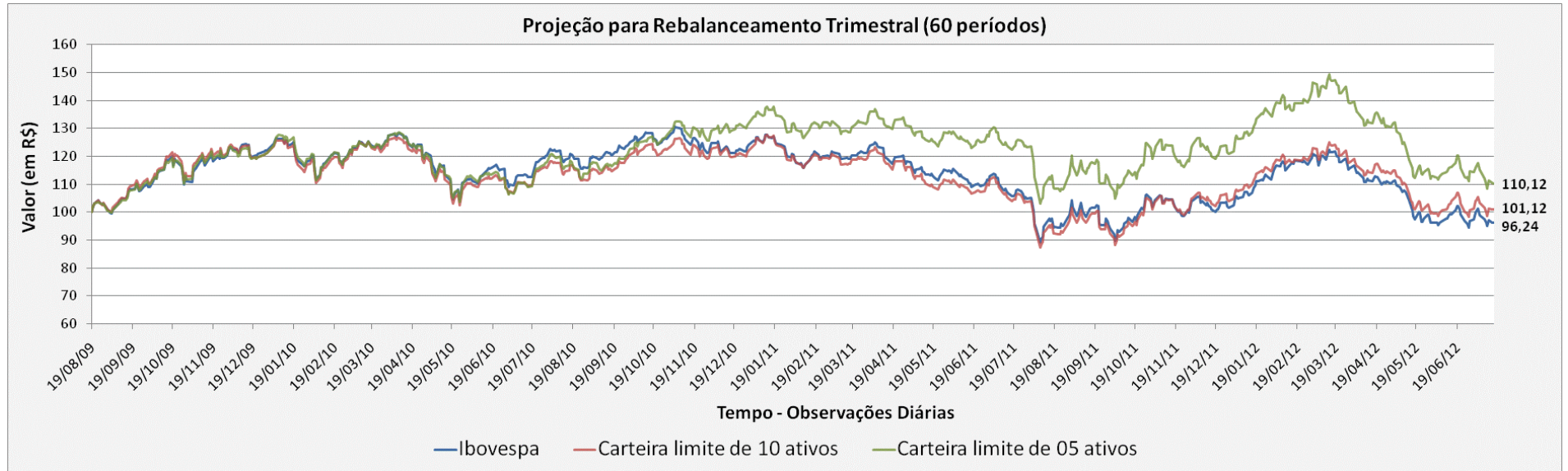
**Fonte:** elaborado pelo autor.





**Figura 13:** Projeção de resultados das carteiras com intervalo fora da amostra de 40 períodos, Rebalanceamento Bimestral, Carteiras de 05 e 10 ativos.

**Fonte:** elaborado pelo autor.



**Figura 14:** Projeção de resultados das carteiras com intervalo fora da amostra de 60 períodos, Rebalanceamento Trimestral, Carteiras de 05 e 10 ativos.

**Fonte:** elaborado pelo autor.

**APÊNDICE C** – Lista de ativos que compõem a amostra de dados utilizada.

**Tabela 22:** Lista de ativos da amostra de dados utilizada.

Relação de ativos da amostra utilizada					
01.	AMBV4	24.	ELET6	47.	OIBR4
02.	BTOW3	25.	ELPL4	48.	PCAR4
03.	BVMF3	26.	EMBR3	49.	PDGR3
04.	BRML3	27.	FIBR3	50.	PETR3
05.	BBDC4	28.	GFS3A	51.	PETR4
06.	BRAP4	29.	GGBR4	52.	RDCD3
07.	BBAS3	30.	GOAU4	53.	RSID3
08.	BRKM5	31.	GOLL4	54.	SBSP3
09.	BRFS3	32.	HYPE3	55.	SANB11
10.	BISA3	33.	ITSA4	56.	CSNA3
11.	CCRO3	34.	ITUB4	57.	CRUZ3
12.	CMIG4	35.	JBSS3	58.	SUZB5
13.	CESP6	36.	KLBN4	59.	VIVT4
14.	CTIP3	37.	LIGT3	60.	TIMP3
15.	HGTX3	38.	LLXL3	61.	TRPL4
16.	CIEL3	39.	RENT3	62.	UGPA3
17.	CPLE6	40.	LAME4	63.	USIM3
18.	CSAN3	41.	LREN3	64.	USIM5
19.	CPFE3	42.	MRFG3	65.	VAGR3
20.	CYRE3	43.	MMXM3	66.	VALE3
21.	DASA3	44.	MRVE3	67.	VALE5
22.	DTEX3	45.	NATU3		
23.	ELET3	46.	OGXP3		

**Fonte:** elaborado pelo autor.

**APÊNDICE D** – Descrição dos ativos usados para formação das Carteiras de 05 e 10 ativos (caso de Rebalanceamento Trimestral – considerando que, para rebalanceamento trimestral, é possível formar um total de 12 carteiras para o intervalo de tempo da amostra de dados adotada).

**Tabela 23:** Descrição dos ativos usados para formação das **carteiras de 05 ativos**, considerando **rebalanceamento trimestral**.

Descrição dos ativos usados para formação de carteiras de 05 ativos					
	Ativo	Quantidade de vezes que o ativo foi usado	Peso Mínimo	Peso Médio	Peso Máximo
1.	PETR4	12	17,0%	24,8%	32,8%
2.	VALE5	10	24,4%	29,0%	33,3%
3.	ITSA4	06	18,9%	22,4%	25,5%
4.	ITUB4	05	19,2%	21,4%	22,9%
5.	BVMF3	03	10,4%	11,6%	12,6%
6.	CYRE3	03	8,7%	12,2%	15,3%
7.	CPFE3	02	14,0%	15,4%	16,7%
8.	ELPL4	02	8,3%	11,4%	14,6%
9.	LAME4	02	10,4%	12,4%	14,3%
10.	CRUZ3	02	15,2%	15,9%	16,6%
11.	BISA3	01	9,6%	9,6%	9,6%
12.	CMIG4	01	10,2%	10,2%	10,2%
13.	CESP6	01	7,0%	7,0%	7,0%
14.	GFS3	01	12,4%	12,4%	12,4%
15.	GGBR4	01	16,5%	16,5%	16,5%
16.	GOAU4	01	14,5%	14,5%	14,5%
17.	MRVE3	01	12,1%	12,1%	12,1%
18.	PDGR3	01	15,6%	15,6%	15,6%
19.	RSID3	01	14,4%	14,4%	14,4%
20.	SBSP3	01	12,0%	12,0%	12,0%
21.	CSNA3	01	12,8%	12,8%	12,8%
22.	TRPL4	01	24,1%	24,1%	24,1%
23.	VALE3	01	28,9%	28,9%	28,9%
	Média		14,6%	15,9%	17,2%

**Fonte:** elaborado pelo autor.

**Tabela 24:** Descrição dos ativos usados para formação das carteiras de 10 ativos, considerando rebalanceamento trimestral.

Descrição dos ativos usados para formação de carteiras de 10 ativos					
	Ativo	Quantidade de vezes que o ativo foi usado	Peso Mínimo	Peso Médio	Peso Máximo
1.	PETR4	12	13,1%	19,4%	26,1%
2.	VALE5	10	18,7%	21,5%	25,1%
3.	GGBR4	07	6,4%	7,7%	9,1%
4.	ITUB4	07	11,0%	15,2%	18,9%
5.	BVMF3	06	5,2%	6,6%	7,4%
6.	CPLE6	05	5,8%	7,4%	8,6%
7.	ITSA4	05	11,3%	13,9%	15,5%
8.	OGXP3	05	4,6%	5,5%	6,5%
9.	RSID3	05	3,8%	6,1%	7,9%
10.	USIM5	05	4,6%	5,9%	8,1%
11.	CESP6	04	5,2%	5,9%	6,9%
12.	MRVE3	04	4,5%	7,6%	10,1%
13.	CMIG4	03	7,8%	9,6%	12,6%
14.	CPFE3	03	7,0%	8,6%	11,7%
15.	CYRE3	03	3,7%	7,1%	10,4%
16.	LREN3	03	4,3%	5,3%	6,5%
17.	CSNA3	03	8,0%	10,6%	12,3%
18.	CRUZ3	03	6,7%	8,4%	10,3%
19.	BTOW3	02	3,1%	3,3%	3,5%
20.	BRFS3	02	5,9%	6,3%	6,8%
21.	GFSA3	02	3,6%	5,4%	7,2%
22.	GOAU4	02	4,4%	6,5%	8,6%
23.	LAME4	02	4,6%	4,9%	5,1%
24.	NATU3	02	3,9%	4,3%	4,6%
25.	OIBR4	02	4,3%	4,5%	4,8%
26.	VAGR3	02	3,2%	3,5%	3,7%
27.	VALE3	02	15,2%	15,8%	16,5%
28.	BRML3	01	5,6%	5,6%	5,6%
29.	BRKM5	01	4,2%	4,2%	4,2%
30.	BISA3	01	4,4%	4,4%	4,4%
31.	DTEX3	01	4,8%	4,8%	4,8%
32.	GOLL4	01	6,4%	6,4%	6,4%
33.	KLBN4	01	4,4%	4,4%	4,4%
34.	PCAR4	01	6,2%	6,2%	6,2%
35.	RDCD3	01	4,3%	4,3%	4,3%
36.	SBSP3	01	7,1%	7,1%	7,1%
	Média		6,3%	7,6%	9,0%

**Fonte:** elaborado pelo autor.