

Métodos Estocásticos em Otimização de Carteiras: Aspectos Não-Lineares e Genéticos

Orientador André Fonseca - CMCC - UFABC

Professor Colaborador Cristian Coletti - CMCC - UFABC

Aluna: Ana Paula Carvalho

Dissertação Referente ao Programa de
Mestrado em Matemática Aplicada - UFABC

Dezembro de 2009

Conteúdo

1	Introdução	7
2	Distribuições Probabilísticas	9
2.1	Definições Básicas	9
2.2	Distribuições Importantes	18
2.2.1	Modelos de Distribuição Discretas	18
2.2.2	Modelos de Distribuições Contínuas	19
3	Processos Estocásticos	21
3.1	Processo de Markov	22
3.2	O modelo Passeio Aleatório	24
3.3	Movimento Browniano	30
4	Séries Temporais	32
4.1	Modelos de Previsão	32
4.2	Funções de Autocorrelação (FAC) de Autocorrelação Parcial (FACP)	35
5	Modelagem da Volatilidade	42
5.1	Modelos ARCH	43
5.2	Modelos GARCH	45
6	Algoritmos Genéticos	51
6.1	Codificação e Implementação	52
7	Otimização de Carteiras	56
7.1	Teoria do Portifólio	56
7.2	Modelo de Markowitz	58
7.2.1	Modelo de Valor em Risco (VaR)	60
7.3	Fronteira Eficiente	65
8	APLICAÇÕES	67
8.1	FAC e FACP	67
8.2	Agregação Temporal	68
8.3	Análise dos Algoritmos Genéticos	72
8.4	VaR e Algoritmos Genéticos	76
8.5	Aplicação Final	77
9	Considerações Finais	89
10	Referências Bibliográficas	92
11	Apêndices	94
A	Método dos Mínimos Quadrados (MMQ)	94
B	Máxima Verossimilhança	96
C	Teste de Dickey - Fuller	98
D	Média e Variância	100
E	FAC E FACP	101
F	Algoritmo Genético 1	103

Resumo

Este trabalho investiga as series temporais financeiras do mercado brasileiro, utilizando conceitos de Processos Estocásticos, como a modelagem ARIMA e GARCH, e de Algoritmos Genéticos, como as técnicas de cruzamento e mutação.

Como objetivo deve-se encontrar uma resposta ao seguinte problema: "Qual ativo escolho para investimento?", a partir da otimização de carteiras de ações do índice IBrX, isto é, a escolha da proporção ótima de investimento numa carteira, que leva ao menor risco e ao maior retorno possíveis. No decorrer do trabalho, foram apresentados os conceitos e codificações que serão utilizados na aplicação final, a qual consiste em comparar dois algoritmos desenvolvidos: AG1 e AG2.

O primeiro deles, AG1, tem como função de aptidão a função retorno da carteira, e o objetivo do algoritmo é otimizá-la, ou seja, encontrar o seu valor máximo. Já o segundo, AG2, tem como função de aptidão a função VaR, que calcula o risco da carteira, e o objetivo do algoritmo é otimizá-la, ou seja, encontrar seu mínimo.

Como aplicação final, foram aplicadas a esses algoritmos três carteiras, compostas por 12 ações cada. Os resultados obtidos foram que o AG2 é o algoritmo mais eficiente e a carteira 1 é a melhor escolha para investimento.

Palavras-chave: Processos Estocásticos, Algoritmos Genéticos, Otimização de Carteiras.

Abstract

This work investigates financial market time series, using Stochastic Processes concepts, such as ARIMA and GARCH modeling, and also Genetic Algorithms concepts, as crossover and mutation techniques.

The goal is to find an answer to the following problem: "Which assets should be invested?", from the optimization of portfolios of IBrX, that is, choosing the optimum investment in a portfolio, which leads to lower risk and highest return possible. During the work, concepts and encoding used in the final application were presented, comparing two developed algorithms : AG1 and AG2.

The first, AG1, is designed to test the portfolio return function, optimizing it by finding its maximum value. The second, AG2, uses the fitness function VaR, which measures the risk of the portfolio, and finds its minimum as optimization goal.

In the final application, those algorithms were applied in three portfolios, consisting of 12 assets each. These results have led to the conclusion that AG2 is the more efficient algorithm and portfolio 1 is the best choice for investment.

Keywords: Stochastic Process, Genetic Algorithm, Portfolio Optimization.

Agradecimentos

Agradeço a Deus por ter colocado essas pessoas maravilhosas na minha vida, pois acredito que se não fosse por Ele eu não existiria para nenhuma delas e nem elas para mim.

Aos meus pais, Maria José Defavari de Carvalho e Tadeu Sérgio Pinto de Carvalho, por terem me dado todo o suporte e base que uma família deve dar, por terem me dado uma boa educação e sempre exigido o melhor de mim, por tornarem possível o meu ingresso numa boa faculdade e agora de poder concluir o mestrado.

Ao meu irmão, Luís Gustavo Pinto de Carvalho, que me incentivou a pesquisa, me indicou diversos livros, me ajudou no conteúdo da parte financeira, renunciou noites e finais de semana para me ajudar e a minha cunhada, Marina Ferreira Nascimento de Carvalho, que pacientemente nos ouviu falar de mercado financeiro e modelos estatísticos.

Ao meu orientador André Fonseca, por ter dito as palavras certas que me acalmaram na hora do meu desespero e por ter me guiado sempre da melhor forma possível. Ao co-orientador Cristian Coletti, por me ajudar na revisão bibliográfica dos processos estocásticos.

Aos amigos do mestrado, Danilo Peixoto Bellucci, Simone Mata, Michele, Douglas, Wendel, Joyce, Marcio, Anna, Pablo pelas horas de descontração, pelas conversas, pelas risadas, pelos churrascos, pela ajuda com as listas de exercícios para as horas de estudos, mas principalmente pela amizade.

As primas Rebeca Cristina Defavari e Lays Defavari e as amigas Viviane Sanches, Lilia, Ana Cristina e Uiara Correa que foram de grande importância nessa fase da minha vida, sempre me ouvindo quando precisei desabafar, dando-me conselhos quando não sabia o que fazer e que rumo tomar, por me animarem nos momentos de tristeza e rirem comigo nos momentos de alegria.

Ao meu namorado Wagner Cleyton Fonseca, por ter ficado ao meu lado nos momentos de cansaço, nas horas de ansiedade por ter que escrever a dissertação, por me acalmar nas horas difíceis e por ouvir minhas reclamações.

Dedicatória

Dedico esta dissertação a minha família, por sempre ficar ao meu lado e me apoiar nas minhas decisões.

1 Introdução

Neste trabalho será estudado o problema “Qual ativo escolho para investimento?”, dentro da área de otimização de carteira de ativos, o qual tem sido uma pergunta muito difundida no mercado financeiro frente a grande variedade de elementos que se modificam todos os dias. “Quais deles oferecem o menor risco e o maior retorno?”, é outra forma de entender esse problema de otimização.

A primeira parte desta dissertação consiste em revisar algumas distribuições probabilísticas importantes que foram utilizadas ao longo do trabalho e analisar as características dos processos estocásticos em questão, na forma de séries temporais dos ativos, utilizando os modelos clássicos lineares *ARIMA*, que possibilitaram uma introdução aos conceitos estatísticos envolvidos. Em seguida serão estudados os modelos não-lineares *ARCH* e *GARCH*, aprofundando os conceitos inicialmente introduzidos. Para desenvolvimento deste assunto, utiliza-se a obra de Schimdt [1] e Gujarati [2] como base para os estudos, dentre outras obras que estão citadas no decorrer do trabalho.

No decorrer do estudo da modelagem *ARIMA*, pode-se perceber que os modelos lineares não eram os melhores para descrever as séries financeiras, pois tais séries apresentavam heterocedasticidade, ou seja, a variância não era constante no tempo. Assim, para estas séries foram utilizadas as modelagens não-lineares *ARCH* e *GARCH*, que puderam descrevê-las de maneira mais precisa.

Após essa primeira etapa, desenvolveu-se a capacidade de analisar e prever comportamentos emergentes dos dados estudados, através da modelagem, implementação e resolução de problemas introdutórios que unem as áreas de *Processos Estocásticos* e *Mercado Financeiro*. Uma vez feita esta fundamentação teórica, inicia-se a segunda parte do trabalho, onde foram introduzidos e aplicados os conceitos de *Algoritmos Genéticos* ao processo de *Otimização de Carteiras* (aquele que minimiza o risco e maximiza o retorno), considerando as hipóteses do modelo de Markowitz *Markowitz* com medida de risco baseada nos modelos de *desvio absoluto médio* e “*value at risk*”.

Dois algoritmos foram desenvolvidos para que fossem executados como aplicação final deste trabalho. O objetivo do desenvolvimento de tais algoritmos era poder compará-los e decidir qual deles era o mais eficiente, ou seja, o algoritmo que mostrasse a melhor solução para o problema inicial desta dissertação.

O primeiro algoritmo, denominado **AG1**, trouxe como função de aptidão a função retorno da carteira, que nada mais é do que uma combinação linear dos retornos esperados de cada ativo com o seu respectivo peso de investimento, ou seja, a porcentagem do seu capital inicial que será investida naquele ativo. Já o segundo algoritmo, **AG2**, trouxe como função de aptidão a função VaR, que calcula o risco da carteira. O objetivo desses algoritmos é otimizar essas funções de aptidão. No caso do AG1, como a função de aptidão é a função retorno da carteira, quando se fala em otimizar essa função, pensa-se em maximizá-la, o que não acontece quando se fala numa função de risco, função VaR, pois quando se fala em otimizá-la, significa minimizar.

Na parte das aplicações, foi feito um estudo de três carteiras compostas por ações do *IBrX*, índice Brasil, um conjunto das 100 ações mais negociáveis na Bovespa. Esse estudo consiste na aplicação dos algoritmos genéticos utilizando as duas medidas de riscos estudadas nesta dissertação com o objetivo de comparação de resultados. O primeiro passo dessas aplicações consiste em estudar as carteiras, avaliando qual o tipo de modelagem será utilizada, se linear ou não-linear. Em seguida aplica-se o algoritmo para definir qual melhor taxa de mutação e cruzamento deverá ser utilizada para que a carteira seja ótima. Os algoritmos são aplicados um de cada vez, e para cada um é encontrado como resposta uma taxa de mutação e uma taxa de cruzamento consideradas ótimas, ou seja, taxas que otimizam o investimento das carteiras.

A partir disso, o segundo passo da aplicação é iniciado, onde um suposto investimento de R\$ 100.000 é feito em cada uma das três carteiras, a partir dos pesos encontrados pelo algoritmo. Isso é feito para que se possibilite uma comparação entre os algoritmos e assim escolher qual o melhor deles, e consequentemente respondendo a pergunta inicial desta dissertação.

A partir desta suposição, e baseando-se na comparação dos resultados obtidos, foi possível afirmar que o segundo algoritmo (AG2), o qual tinha a função de risco como função de aptidão, foi o melhor dos algoritmos desenvolvidos. Esse resultado foi obtido, devido à consistência das respostas, que foram analisadas de acordo com o maior / menor risco e o mais / menos arriscado.

Como conclusão, em resposta a pergunta inicial proposta neste trabalho, “Qual ativo escolho para

investimento?”, foi possível verificar, a partir da medida de dispersão coeficiente de variação, que a carteira escolhida deveria ser a carteira 1, pois obteve o menor risco, dado um certo retorno.

2 Distribuições Probabilísticas

2.1 Definições Básicas

A construção desta seção foi baseada nas obras de Chung [3], Dimitri [4], James [5] e Levine [6], de onde foi extraída as definições, propriedades e observações que serão de grande importância para o entendimento das vas, que são o objeto de estudo deste trabalho.

Ao falar em modelor probabilístico pensa-se em modelos que mostram valores numéricos de algum experimento aleatório, por exemplo, o preço das ações de uma bolsa de valores. Existem alguns modelos que não mostram esses valores, porém, podem ser associados, de alguma forma, a tais valores numéricos. Por exemplo, se a experiência for a seleção de estudantes de uma determinada população, consideram-se as médias de suas notas. Trabalhar com valores numéricos torna mais fácil a atribuição de probabilidades a eles. É por isso que, ao deparar com modelos que não trazem valores numéricos, estes serão associados através das *variáveis aleatórias*.

Antes de definir variáveis aleatórias precisa-se definir o que é uma probabilidade e consequentemente o espaço onde essas variáveis estarão definidas (*espaço de probabilidade*).

Definição 2.1 *Uma classe \mathcal{F} de subconjuntos de um conjunto não vazio Ω satisfazendo as seguintes condições:*

- $\Omega \in \mathcal{F}$,
- Se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$,
- Se $A_n \in \mathcal{F}$ para $n = 1, 2, \dots$, então $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$,

é chamada de σ -álgebra de subconjuntos de Ω .

Definição 2.2 (Kolmogorov - 1933) *Uma probabilidade é uma função $\mathbb{P}(\cdot)$ a valores reais definida em uma classe \mathcal{F} de eventos de um espaço amostral Ω , que possui as seguintes propriedades:*

(P1) $0 \leq \mathbb{P}(A) \leq 1$, para todo $A \in \mathcal{F}$,

(P2) $\mathbb{P}(\Omega) = 1$,

(P3) *Aditividade enumerável: para qualquer sequência $A_1, A_2, \dots \in \mathcal{F}$ de eventos dois a dois disjuntos,*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Definição 2.3 *Um espaço de probabilidade é um trio $(\Omega, \mathcal{F}, \mathbb{P})$, onde:*

1. Ω é um conjunto não-vazio,
2. \mathcal{F} é uma σ -álgebra de subconjuntos de Ω , e
3. \mathbb{P} é uma probabilidade em \mathcal{F} .

Observação 2.1 *Sobre medida de probabilidade:*

- No caso de Ω ser finito ou infinito enumerável, defini-se a medida de probabilidade na classe \mathcal{F} de todos os subconjuntos de Ω . Neste caso, tem-se $\Omega = \{\omega_1, \omega_2, \dots\}$ onde serão associados a cada $\omega_i, i = 1, 2, \dots$, um número $p(\omega_i)$ tal que $p(\omega_i) > 0$ e $\sum_{i=1}^{\infty} p(\omega_i) = 1$. A Probabilidade de um evento $A \subset \Omega$ é definida por

$$\mathbb{P}(A) = \sum_{i:\omega_i \in A} p(\omega_i).$$

- Por convenção, $\mathbb{P}[\emptyset] = 0$.
- No caso de Ω ser infinito não-enumerável é, em geral, impossível definir uma medida de probabilidade em todos os subconjuntos de Ω . Define-se então uma probabilidade em uma classe mais restrita de subconjuntos de Ω ; apenas esses subconjuntos são denominados eventos.

Pode-se dizer que, *variáveis aleatórias* são funções reais dos resultados de um experimento, que estão definidas num espaço amostral Ω . Por exemplo, se ao tomar a população como sendo o espaço amostral e classificá-la através das idades dos indivíduos, a variável aleatória irá associar a cada indivíduo da população uma idade, em outras palavras, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ é o espaço amostral com n indivíduos e X é a variável aleatória que para cada indivíduo da população associa uma idade ($\omega \rightarrow X(\omega)$).

Definição 2.4 Uma *variável aleatória* (v.a.) X em um espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$ é uma função real definida no espaço Ω tal que $[X \leq x]$ é evento aleatório para todo $x \in \mathbb{R}$; isto é, $X : \Omega \rightarrow \mathbb{R}$ é variável aleatória se $[X \leq x] \in \mathcal{F}, \forall x \in \mathbb{R}$, onde $[X \leq x] = \{\omega \in \Omega / X(\omega) \leq x\} \in \mathcal{F}, \forall x \in \mathbb{R}$.

Denota-se por letras maiúsculas como X, Y, Z , etc., as variáveis aleatórias e com as letras minúsculas como x, y, z , etc. identificaremos seus possíveis valores numéricos, ou seja, números reais.

Proposição 2.1 Se X e Y são v.a., então também serão:

$$X + Y, \quad X - Y, \quad XY, \quad X/Y, (Y \neq 0) \quad e \quad aX + bY, \quad \text{com } a, b \in \mathbb{R}$$

O mesmo vale para uma função f qualquer, ou seja, se X é uma v.a., então $f(X)$ também será.

A maneira mais importante de se caracterizar uma variável aleatória é através das probabilidades dos valores assumidos por ela. Para isso precisa-se de uma função, definida para todos os valores possíveis $x \in X$, que diga quais são suas probabilidades. Sabe-se que, para cada x existe um evento A contido no espaço amostral ($A \subset \Omega$), tal que $\omega \in A \Leftrightarrow X(\omega) = x$. Essa função é denominada função de distribuição acumulada.

Definição 2.5 A *função de distribuição acumulada* de uma variável aleatória X é a função $F = F_X$ definida por:

$$F_X(x) = P(X \leq x) = P(\omega \in \Omega : X(\omega) \leq x), \forall x \in \mathbb{R} \quad (1)$$

Definição 2.6 A variável aleatória X é *discreta* se toma um número finito ou enumerável de valores, isto é, se existe um conjunto finito ou enumerável $\{x_1, x_2, \dots\} \subset \mathbb{R}$ tal que $X(\omega) \in \{x_1, x_2, \dots\}, \forall \omega \in \Omega$.

X	Evento Correspondente
0	$A_0 = \{(r, r, r)\}$
1	$A_1 = \{(c, r, r); (r, c, r); (r, r, c)\}$
2	$A_2 = \{(c, c, r); (c, r, c); (r, c, c)\}$
3	$A_3 = \{(c, c, c)\}$

Tabela 1: Eventos possíveis

Ao considerar uma variável aleatória discreta X , associada a um evento A , no espaço amostral Ω , tem-se que a função $p_X(x_i)$ é chamada de **função de probabilidade** de X . Em outras palavras, é a função que determina a probabilidade do evento $[X = x_i]$, ou seja

$$p_X(x_i) = P(X = x_i), \quad i = 1, 2, \dots \quad (2)$$

Observação 2.2 Uma variável aleatória discreta é definida quando os seus valores possíveis $\{x_i\}_{i \leq 1}$ e as respectivas probabilidades $\{p_X(x_i)\}_{i \leq 1}$ satisfazem:

$$p_X(x_i) > 0, \quad \forall i \quad e \quad \sum_{i=1}^{\infty} p_X(x_i) = 1.$$

Exemplo 2.1 Três moedas.

Lançam-se três moedas. Seja X o número de ocorrências da face cara.

Note que a cardinalidade do espaço amostral do experimento é igual a 8, e seus eventos são: $\{(c, c, c); (c, c, r); (c, r, c); (c, r, r); (r, c, c); (r, c, r); (r, r, c); (r, r, r)\}$, onde **c** significa **cara** e **r** significa **coroa**.

Como X é o número de caras, pode dizer que ele assumirá os valores 0, 1, 2 ou 3, isto é, nenhuma, uma, duas ou três caras respectivamente.

As probabilidades, nesse caso discreto, são determinadas abaixo:

$$p_X(0) = P(X = 0) = \frac{1}{8};$$

$$p_X(1) = P(X = 1) = \frac{3}{8};$$

$$p_X(2) = P(X = 2) = \frac{3}{8};$$

$$p_X(3) = P(X = 3) = \frac{1}{8}.$$

Assim, para cada valor de $x \in X$ tem-se uma função distribuição de probabilidades.

Definição 2.7 A variável aleatória X é (absolutamente) **contínua** se existe uma função $f(x) \geq 0$ tal que:

$$F_X(x) = \int_{-\infty}^x f(t)dt, \quad \forall x \in \mathbb{R}. \quad (3)$$

Neste caso, diz-se que f é a **função de densidade de probabilidade (fdp)** de X , ou simplesmente densidade de X .

Observação 2.3 Uma variável aleatória contínua é definida quando tem-se uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ tal que:

$$f(x) \geq 0 \quad \forall x \in \Omega; \quad \int_{-\infty}^{\infty} f(x)dx = 1.$$

Em particular, pode-se definir a probabilidade de X pertencer ao intervalo de extremos $[a, b]$:

$$\mathbb{P}(a \leq X \leq b) = \int_b^a f(x) dx.$$

Neste trabalho serão estudados apenas dois tipos de variáveis aleatórias: as *discretas* e as *contínuas*. Uma *v.a.* discreta é uma função real do resultado do experimento que pode tomar um número finito ou enumerável de valores. Já a contínua, é uma função que toma um número infinito de valores. Por exemplo, considera-se o experimento onde deve-se escolher ao acaso um ponto a no intervalo $[-1, 1]$. A *v.a.* que associa para cada a o valor numérico a^2 não é discreta. Por outro lado, é discreta a *v.a.* que associa para cada a o seguinte valor numérico:

$$A = \begin{cases} 1, & \text{se } a > 0 \\ 0, & \text{se } a = 0 \\ -1, & \text{se } a < 0 \end{cases}$$

O foco deste trabalho será o estudo das variáveis aleatórias contínuas. Para isso, será preciso desenvolver alguns conceitos que serão de grande importância para os estudos das séries temporais.

Exemplo 2.2 *Tempo de Condução.*

O tempo de condução para o trabalho de Pedro está entre 15 e 20 minutos se o dia está ensolarado, e entre 20 e 25 minutos se o dia está chuvoso, sendo igualmente provável em cada caso. Suponha que um dia é ensolarado, com probabilidade $2 / 3$ e chuvoso, com probabilidade $1 / 3$. Qual é o formato da fdp do tempo de condução de Pedro ao trabalho, visto como uma variável aleatória X ?

Ao dizer que o dia será igualmente provável em cada caso, significa que a fdp de X será uma constante em cada um dos intervalos $[15, 20]$ e $[20, 25]$. Além disso, uma vez que nestes dois intervalos estejam contidos todos os possíveis tempos de condução, a fdp deve ser zero em qualquer outro lugar:

$$f(x) = \begin{cases} c_1, & \text{se } 15 \leq x < 20 \\ c_2, & \text{se } 20 \leq x \leq 25 \\ 0, & \text{caso contrário} \end{cases}$$

Onde c_1 e c_2 são constantes quaisquer.

Essas constantes podem ser determinada usando a equação (3), ou seja:

$$\frac{2}{3} = \mathbb{P}(\text{dia ensolarado}) = \int_{15}^{20} f(x)dx = \int_{15}^{20} c_1 dx = 5c_1 \Rightarrow c_1 = \frac{2}{15}$$

$$\frac{1}{3} = \mathbb{P}(\text{dia chuvoso}) = \int_{20}^{25} f(x)dx = \int_{20}^{25} c_2 dx = 5c_2 \Rightarrow c_2 = \frac{1}{15}$$

Então a fdp de X , onde X é o tempo de condução de Pedro ao trabalho, será:

$$f(x) = \begin{cases} \frac{2}{15}, & \text{se } 15 \leq x < 20 \\ \frac{1}{15}, & \text{se } 20 \leq x \leq 25 \\ 0, & \text{caso contrário} \end{cases}$$

O gráfico da fdp encontra-se na figura 1.

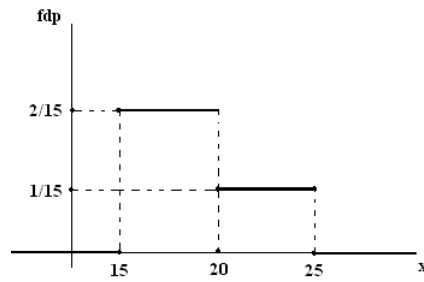


Figura 1: Gráfico da fdp do tempo de condução de Pedro ao trabalho

Observação 2.4 Sobre a variável aleatória X :

1. Se X é discreta, então $[X \leq x] = \bigcup_{i: x_i \leq x} [X = x_i]$, logo

$$F_X(x) = \sum_{i: x_i \leq x} P(X = x_i) = \sum_{i: x_i \leq x} p_X(x_i);$$

2. Se X é contínua, então F_X , sendo uma integral indefinida de f , é contínua. Para $f(x) \geq 0$, tem-se:

$$F_X(x) = \int_{-\infty}^{\infty} f(x) dx = 1.$$

As funções de distribuição fornecem as probabilidades de todos os valores possíveis da *v.a.* X e seria interessante se pudesse resumir estas informações em um único número. Para isso, calcula-se a esperança de X , que é uma média ponderada dos valores possíveis de X .

Definição 2.8 A *esperança* (média, valor esperado) de uma variável aleatória X é definida por:

$$\mu_X = E[X] = \begin{cases} \sum_x x P(X = x) & \text{se } X \text{ é discreta,} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{se } X \text{ é contínua com densidade } f. \end{cases}$$

Observação 2.5 Esperança matemática é a média de todos os valores possíveis de X que são pesados com a função de probabilidade, no caso de *v.a.* discretas, ou com a função densidade de probabilidade (fdp), quando a *v.a.* é contínua.

Exemplo 2.3 Número de Caras.

Calcular o número médio de caras no lançamento de 3 moedas, utilizando os dados do exemplo 2.1.

Através da construção de uma tabela com os dados do exemplo 2.1 acrescentando uma coluna para o cálculo da média, teremos:

Conclui-se então que, em média saem 1,5 caras no lançamento de 3 moedas.

Exemplo 2.4 Roda da Fortuna.

Um jogador gira uma roda da fortuna e observa os resultados. Partindo do princípio de que todos os subintervalos de $[a, b]$ do mesmo comprimento são igualmente prováveis, esta experiência pode ser modelada em termos de uma variável aleatória X cuja fdp é:

X	$p_X(x)$	$x * p_X(x)$
0	1/8	0
1	3/8	3/8
2	3/8	6/8
3	1/8	3/8
	1	12/8 = 1,5

Tabela 2: Cálculo das médias

$$f(x) = \begin{cases} c, & \text{se } a \leq x \leq b \\ 0, & \text{caso contrário,} \end{cases}$$

onde c é uma constante qualquer. Esta variável aleatória é conhecida como uniforme, ou uniformemente distribuída.

Através da equação 3, $c = \frac{1}{b-a}$. Assim, sua esperança será:

$$E[X] = \int_{-\infty}^b 0dx + \int_a^b xcdx + \int_a^{\infty} 0dx$$

$$E[X] = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$$

Quando duas variáveis aleatórias, contínuas ou discretas, X e Y são associadas respectivamente aos eventos E_1 e E_2 , define-se **probabilidade condicional**. Isto é, qual a probabilidade de ocorrer o evento $A \subset E_1$ dado que já ocorreu o evento $B \subset E_2$:

$$\mathbb{P}(X = A|Y = B) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

desde que $\mathbb{P}[B] > 0$.

A média ou esperança de uma variável aleatória condicionada a outra é chamada de **esperança condicional**.

Definição 2.9

1. Se X e Y são variáveis aleatórias discretas, a **função de probabilidade condicional** de X dado $Y = y$ é definida por:

$$E[X|Y = y] = \sum x\mathbb{P}(X|Y = y).$$

2. Se, X e Y são conjuntamente contínuas com função densidade conjunta $f(x, y)$, a **função densidade condicional** de X dado $Y = y$ é definida para todos os valores de y tais que $f_Y(y) > 0$ por:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

A **esperança condicional** de X dado $Y = y$ é, neste caso:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx.$$

X	$p_X(x)$	$x * p_X(x)$	$(x - E[X])$	$(x - E[X])^2$	$(x - E[X])^2 * p_X(x)$
0	1/8	0	0	0	0
1	3/8	3/8	5/8	25/64	25/512
2	3/8	6/8	5/4	25/16	75/128
3	1/8	3/8	12/8	144/64	441/512
	1	1,5			1,5

Tabela 3: Cálculo da Variância e do Desvio Padrão

A **variância** (σ^2) calcula o grau de dispersão (ou de concentração) dos valores que a variável aleatória assume em torno da sua média. Quanto menor a variância, menor o grau de dispersão desses valores em torno da média e quanto maior a variância, maior o grau de dispersão em torno da média. O **desvio padrão** (σ), que é a raiz quadrada da variância, foi definido para acabar com problemas de unidades de medidas, pois ao falar de variância, a unidade de medida deve estar elevada ao quadrado, o que, em alguns casos, não tem sentido algum.

$$X \text{ discreta : } \sigma^2 \equiv Var[X] = \sum (x - E[X])^2 p_X(x). \quad (4)$$

$$X \text{ continua : } \sigma^2 \equiv Var[X] = \int (x - E[X])^2 f dx. \quad (5)$$

Uma outra forma de descrever a expressão da variância, tanto para *v.a.* contínua, quanto para discreta, é a seguinte:

$$\sigma^2 \equiv Var[X] = E[X^2] - (E[X])^2. \quad (6)$$

A seguir alguns exemplos do cálculo da variância para variáveis aleatórias contínuas e discretas:

Exemplo 2.5 Variância e Desvio Padrão.

Cálculo da variância e do desvio padrão da *v.a.* dada no exemplo 2.1.

Para encontrar o desvio padrão, basta extrair a raiz quadrada da variância. Logo, $\sigma = 1,23$.

Exemplo 2.6 Esperança.

Considere agora o exemplo 2.4 da roda da fortuna, lembrando de que nesse exemplo foi calculado a esperança de X .

A variância de X será calculada utilizando a equação 6. Para isso, basta calcular $E[X^2]$, pois já é conhecido o valor de $E[X]$. Então:

$$E[X^2] = \int_{-\infty}^b 0 dx + \int_a^b x^2 c dx + \int_a^{\infty} 0 dx. \quad (7)$$

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{a^2 + ab + b^2}{3}.$$

Agora, substituindo os valores encontrados na equação (6), tem-se:

$$\sigma^2 \equiv Var[X] = E[X^2] - (E[X])^2 = \frac{a^2 + ab + b^2}{3} - \left[\frac{a+b}{2} \right]^2 = \frac{(b-a)^2}{12}.$$

A seguir, serão apresentadas algumas propriedades importantes de esperança e variância, que poderão ser usadas tanto para *v.a.* contínua quanto para a discreta. Essas propriedades poderão ser

verificadas na obra de James [5].

Propriedades 2.1 *Sejam X e Y duas v.a. (contínua ou discreta) e $a, b \in \mathbb{R}$.*

Seja $Y = aX + b$, então:

1. $E[Y] = E[aX + b] = aE[X] + b$
2. $Var[Y] = Var[aX + b] = a^2Var[X]$
3. $Var[X] \geq 0$

Seguem agora os momentos relacionados às variáveis aleatórias contínuas, definidos por:

$$m_n \equiv E[X^n] = \int X^n P(X) dX. \quad (8)$$

De acordo com essa definição, média é o primeiro momento ($m \equiv m_1$) e variância pode ser expressa por meio dos dois primeiros momentos, $\sigma^2 = m_2 - m_1^2$. Dois outros parâmetros importantes, *skewness* (S) e *Curtose* (K), estão relacionados ao terceiro e quarto momentos, respectivamente,

$$S = \frac{E[(X - m)^3]}{\sigma^3}. \quad (9)$$

$$K = \frac{E[(X - m)^4]}{\sigma^4}. \quad (10)$$

Ambos os parâmetros S e K não têm dimensão; $S = 0$ implica que a distribuição é simétrica em torno do valor da média; os valores positivos e negativos de S indicam longa calda positiva e longa calda negativa, respectivamente; K caracteriza o pico distribuição e é geralmente usado para medir o desvio da distribuição normal (equação 20). Em particular, o excesso positivo de K indica maior frequência média e grande desvio do valor médio, que é típico para a distribuição normal; o excesso negativo, de outro lado, indica pequeno desvio do valor médio.

Define-se agora a **função de distribuição conjunta** de duas variáveis aleatórias X e Y , que nada mais é do que a generalização da função densidade de probabilidades (para *v.a.* contínua) ou da função de probabilidade (para *v.a.* discreta), para mais de uma variável aleatória.

Definição 2.10 *Sejam X e Y variáveis aleatórias definidas no mesmo espaço de probabilidade. A função de distribuição acumulada conjunta do par (X, Y) é definida por:*

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y), \quad \forall x, y \in \mathbb{R}.$$

As funções de distribuição conjunta para variáveis aleatórias discretas e contínuas são definidas como segue:

Definição 2.11

1. *Sejam X e Y variáveis aleatórias discretas definidas no mesmo espaço de probabilidade. A função de distribuição conjunta de X e Y é:*

$$p(x, y) = \mathbb{P}(X = x, Y = y), \quad \forall x, y \in \mathbb{R}. \quad (11)$$

2. Sejam X e Y variáveis aleatórias contínuas definidas no mesmo espaço de probabilidade. Dizemos que X e Y são conjuntamente contínuas se existe uma função $f(x, y) \geq 0$, chamada **fdp conjunta**, tal que para quaisquer $x, y \in \mathbb{R}$

$$f(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv. \quad (12)$$

Para estudar a dependência entre variáveis (discretas e contínuas), calcula-se sua **covariância**, que é uma média de como duas variáveis variam conjuntamente. Seus valores se situam no intervalo $(-\infty, +\infty)$; se for positiva significa que existe uma relação diretamente proporcional entre elas e se for negativa implica que elas são inversamente proporcionais.

A covariância entre duas variáveis é medida através da seguinte equação:

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y], \quad (13)$$

quando $\text{cov}(X, Y) = 0$, significa que essas variáveis não são correlacionadas.

Teorema 2.1 Sejam X, Y duas variáveis aleatórias e a, b, c, d constantes reais, teremos [7]:

- $\text{cov}(aX + b, cY + d) = ac * \text{cov}(X, Y)$
- $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{cov}(X, Y)$

Note-se que, se as variáveis forem independentes, ou seja, não forem correlacionadas, a $\text{cov}(X, Y) = 0$ e então $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$. A demonstração desse teorema pode ser encontrada na obra de James [5].

Através de um processo de normalização da covariância, define-se o *coeficiente de correlação*, denotado por ρ :

$$\rho \equiv \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{(\sigma_X^2 \sigma_Y^2)}}, \quad (14)$$

onde $-1 \leq \rho \leq 1$. Quanto mais próximo ρ estiver de 1, maior o grau de dependência proporcional direta entre as variáveis, ou seja, elas estarão diretamente correlacionadas; quanto mais próximo ρ estiver de -1 maior o grau de dependência inversa, ou seja, estarão inversamente correlacionadas; e se $\rho = 0$ então não existe correlação entre elas.

Na figura 2 observa-se exemplos de comportamentos das variáveis nos casos $\rho \cong +1$ (A), $\rho \cong -1$ (B) e quando $\rho = 0$ (C).

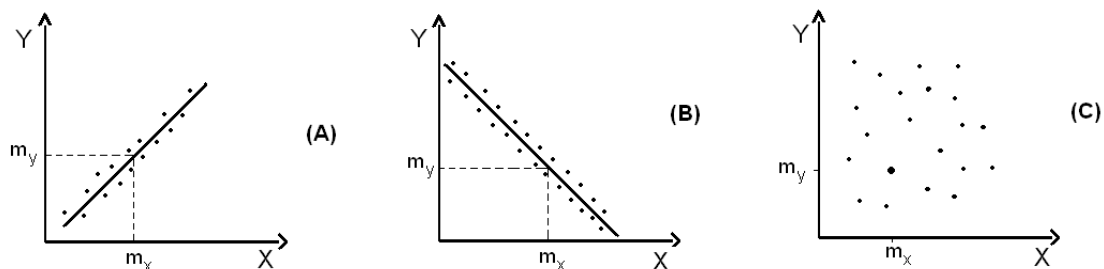


Figura 2: Diagramas de Dispersão

2.2 Distribuições Importantes

Existem várias distribuições probabilísticas importantes em finanças quantitativa, podendo ser classificadas quanto as variáveis aleatórias discretas e contínuas. Entre as variáveis aleatórias discretas destacam-se algumas, tais como: *Distribuição Uniforme*, *Distribuição Binomial*, *Distribuição de Bernoulli* e a *Distribuição de Poisson*; e para as variáveis aleatórias contínuas: *Distribuição Uniforme*, *Distribuição Normal* e a *Distribuição de Cauchy*.

Seguem definições e exemplos com base nas bibliografias de Coletti e Lebensztayn [8] e Ross [9].

2.2.1 Modelos de Distribuição Discretas

Considere X uma variável aleatória discreta.

X tem **distribuição uniforme** sobre o conjunto $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}$ se tiver função de probabilidade dada por:

$$\mathbb{P}(X = x_i) = \frac{1}{n}, \quad i = 1, \dots, n. \quad (15)$$

X representa a escolha ao acaso de um elemento do conjunto $\{x_1, x_2, \dots, x_n\}$.

Quando X tem distribuição uniforme, denota-se da seguinte forma: $X \sim \text{Uniforme}$.

X tem **distribuição Bernoulli**, para $0 \leq p \leq 1$, se tiver função de probabilidade dada por:

$$\mathbb{P}(X = x) = p^x(1-p)^{1-x}, \quad x = 0, 1. \quad (16)$$

X é a função indicadora da ocorrência de sucesso em um ensaio de Bernoulli (experimento que tem apenas dois resultados possíveis, sucesso ou fracasso).

X tem **distribuição binomial**, para $n \geq 1$ inteiro e $0 \leq p \leq 1$, se tiver função de probabilidade dada por:

$$\mathbb{P}(X = x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, \dots, n, \quad (17)$$

onde:

$$\binom{n}{x} = \frac{n!}{(n-x)!x!}.$$

X é o número de sucessos obtidos em n ensaios de Bernoulli independentes com probabilidade de sucesso p em cada ensaio.

Exemplo 2.7 Produtos Defeituosos.

Sabe-se que qualquer produto feito por uma certa máquina sairá defeituoso, com probabilidade 0,1, independentemente de qualquer outro item. Qual é a probabilidade de que em uma amostra de três itens, no máximo um será defeituoso?

Se X é o número de itens defeituosos na amostra, então X é uma variável aleatória binomial com parâmetros $(3, 0, 1)$. Daí, a probabilidade desejada é dada por:

$$\mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \binom{3}{0} (0,1)^0(0,9)^3 + \binom{3}{1} (0,1)^1(0,9)^2 = 0,972.$$

X tem **distribuição de Poisson**, para $\lambda > 0$, se tiver função de probabilidade dada por:

$$\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \dots \quad (18)$$

Exemplo 2.8 *Erros de Digitação.*

Suponha que o número de erros de digitação em uma única página de um livro tem uma distribuição de Poisson com parâmetro $\lambda = 1$. Calcular a probabilidade de que haja pelo menos um erro nesta página.

$$\mathbb{P}(x \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-1} \approx 0,633.$$

2.2.2 Modelos de Distribuições Contínuas

Considere X uma variável aleatória contínua.

X tem **distribuição uniforme** sobre o intervalo $(0, 1)$ se sua função densidade de probabilidade (fdp) é dada por:

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Como $f(x) > 0$ somente quando $x \in (0, 1)$, segue que X deverá assumir algum valor em $(0, 1)$. Além disso, como $f(x)$ é constante para $x \in (0, 1)$, é provável que X esteja próximo de qualquer valor do intervalo $(0, 1)$. Note que, para qualquer $0 < a < b < 1$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx = b - a.$$

Em outras palavras, a probabilidade que X assume em qualquer sub intervalo de $(0, 1)$ se iguala ao comprimento desse sub intervalo.

Diz-se, então, que X é uma variável aleatória uniforme do intervalo (a, b) , um sub intervalo de $(0, 1)$, se sua fdp é dada por:

$$f(x) = \begin{cases} \frac{1}{(b-a)}, & a < x < b, \\ 0, & \text{caso contrário.} \end{cases} \quad (19)$$

Exemplo 2.9 *Distribuição Uniforme.*

Se X tem distribuição uniforme sobre o intervalo $(0, 10)$, calcular a propabilidade de $X < 3$, $X > 7$, $1 < X < 6$.

$$\mathbb{P}(X < 3) = \frac{\int_0^3 dx}{10} = \frac{3}{10} = 0,3,$$

$$\mathbb{P}(X > 7) = \frac{\int_7^{10} dx}{10} = \frac{3}{10} = 0,3,$$

$$\mathbb{P}(1 < X < 6) = \frac{\int_1^6 dx}{10} = \frac{5}{10} = 0,5.$$

X tem **Distribuição Normal** ou **Gaussiana**, com $\sigma > 0$ e $\mu \in \mathbb{R}$, se sua fdp é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)/2\sigma^2}, \quad \forall x \in \mathbb{R}. \quad (20)$$

O gráfico da figura 3 mostra o comportamento da distribuição normal.

Esta distribuição possui algumas características muito importantes, dentre elas estão: seu ponto de máximo coincidindo com a média m ; seus pontos de inflexão são $m + \sigma$ e $m - \sigma$ (note-se que existe uma simetria da curva em relação a média m) e ainda, a esperança e a variância são iguais a m e σ^2 , respectivamente.

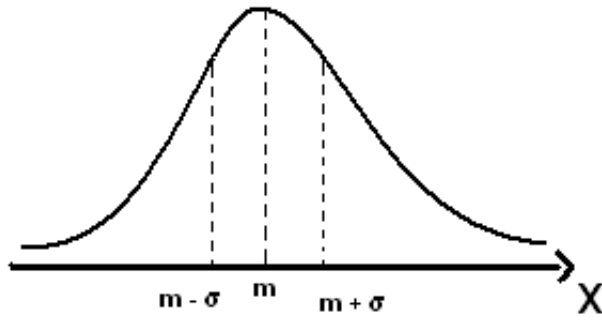


Figura 3: Distribuição Normal

X tem **Distribuição de Cauchy**, com $a \in \mathbb{R}$ e $b > 0$, se sua fdp é dada por:

$$f(x) = \frac{1}{\pi b(1 + [(x - a)/b]^2)}, \quad \forall x \in \mathbb{R}. \quad (21)$$

Todos os momentos da distribuição de Cauchy são infinitos. O caso onde $b = 1$ e $m = 0$ é chamado de distribuição de Cauchy padrão.

3 Processos Estocásticos

Inicialmente, os processos estocásticos foram usados pelos físicos para descrever o movimento de partículas, podendo ser divididos em: *Processos de Tempo Contínuo*, *Processos de Tempo Discreto*, *Processos Estacionários* e *Processos não-Estacionários*. A maioria dos métodos de análise exigem que o processo estocástico gerador dos dados seja um processo estacionário. Para as variáveis financeiras, como preços e retornos, o conceito de Processo Estocástico é usado para descrever seus comportamentos.

Definição 3.1 *Seja T um conjunto arbitrário. Um processo estocástico é uma família $\{X(t), t \in T\}$, tal que, para cada $t \in T$, $X(t)$ é uma variável aleatória.*

Nestas condições, um processo estocástico é uma família de variáveis aleatórias que supomos definidas num mesmo espaço de probabilidades $(\Omega, \mathcal{F}, \mathbb{P})$, onde Ω é um conjunto não-vazio (espaço amostral), \mathcal{F} é uma coleção de subconjuntos contidos em Ω (eventos) e \mathbb{P} é uma função que associa à cada elemento de \mathcal{F} um número real não-negativo (probabilidade associada ao evento) [10].

O conjunto T , quando tomado pelos conjuntos \mathbb{Z} e \mathbb{N} é denominado um *processo de tempo discreto* e quando tomado pelo conjunto dos reais \mathbb{R} ou $[0, \infty]$, é denominado um *processo de tempo contínuo*. Como para cada $t \in T$ $X(t)$ é uma variável aleatória definida sobre Ω , diz-se que $X(t)$ é uma função de dois argumentos, isto é, $X = X(t, \omega)$ com $t \in T$ e $\omega \in \Omega$. Na figura 4 nota-se que essa interpretação, onde para cada $t \in T$ há uma v.a. $X(t, \omega)$ com uma função densidade de probabilidade (fdp), $P(X)$. Por outro lado, para cada $\omega \in \Omega$ fixado, haverá uma função de t , ou seja, uma *Série Temporal*.

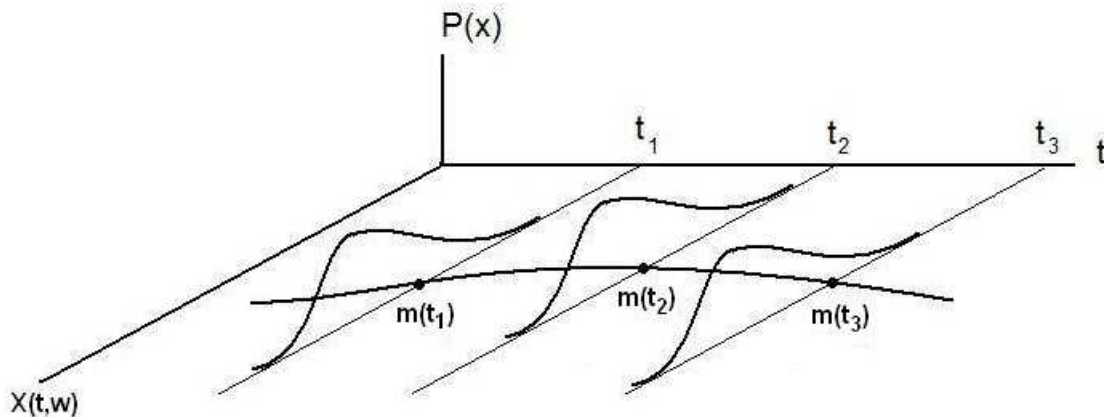


Figura 4: Processo Estocástico

Uma maneira de estudar os processos estocásticos é determinando seus momentos, no entanto de acordo com Morettin e Toloí [10], esse estudo é restringido para momentos de baixa ordem, ou seja, média, variância e covariância. A forma de calcular tais momentos foi visto na seção 2 e a partir deles, será possível definir os *Processos Estacionários*.

Intuitivamente, tratado de processos estacionários refere-se a processos que desenvolvem no tempo. Em outras palavras, as características de $X(t + \tau)$, $\forall \tau$, são as mesmas de $X(t)$.

Definição 3.2 *Um processo estocástico $\{X(t), t \in T\}$ diz-se fracamente estacionário, ou simplesmente estacionário, se e somente se:*

1. $E[X(t)] = \mu$, constante para todo o $t \in T$;
2. $Var[X(t)] = \sigma^2$, para todo o $t \in T$;
3. $Cov(X(t), X(t - 1))$ não depende de t .

Supõe-se que $\mu = 0$, sem perda de generalidade; caso contrário, considera-se o processo $\{X(t) - \mu\}$.

Se as variáveis aleatórias $X(t)$ tem-se a mesma distribuição, teremos uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Neste caso, o processo $X(t)$ é estacionário. E, ainda, se $E[X(t)] = \mu$ e $Var[X(t)] = \sigma^2$, para todo $t \geq 1$, então:

$$\gamma_\tau = Cov(X(t), X(t - \tau)) = \begin{cases} \sigma^2, & \text{se } \tau = 0 \\ 0, & \text{se } \tau \neq 0 \end{cases} \quad (22)$$

Definição 3.3 Diz-se que $\{\epsilon(t), t \in Z\}$ é um ruído branco se as variáveis aleatórias $\epsilon(t)$ são não correlacionadas, isto é, $Cov(\epsilon(t), \epsilon(s)) = 0$, para $t \neq s$.

Este processo será estacionário se $E[\epsilon(t)] = \mu_\epsilon$ e $Var[\epsilon(t)] = \sigma_\epsilon^2$, para todo t . Segue que, a função de autocovariância de $\epsilon(t)$ é dada pela equação (22).

Para simplificação dos cálculos será suposto de agora em diante que $\mu_\epsilon = 0$.

3.1 Processo de Markov

O *Processo de Markov* é um processo estocástico no qual somente o valor atual da variável é relevante para prever a evolução futura do processo. Isso significa que, valores históricos ou mesmo o caminho pelo qual a variável atingiu o seu valor atual são irrelevantes para prever o seu valor futuro. Assume-se que preços de ativos em geral, como ações, seguem um processo de Markov. Dentro dessa premissa, o preço atual de uma ação reflete todas as informações históricas bem como as expectativas a respeito do preço futuro dessa ação.

Definição 3.4 Seja S um conjunto finito ou infinito enumerável. Seja $(X_n)_{n \geq 0}$ um processo estocástico com espaço de estados S . O processo $(X_n)_{n \geq 0}$ será um *Processo de Markov homogêneo* se:

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n],$$

para toda sequência x_0, x_1, \dots, x_{n+1} tal que $\mathbb{P}[X_n = x_n, \dots, X_0 = x_0] > 0$.

Considere um processo estocástico $\{X_n\}_{n \geq 0} = \{X_0, X_1, X_2, \dots\}$ que assume um número finito ou enumerável de valores. Note que, se $X_n = i$, então o processo no tempo n está no estado i . Assim, de acordo com Ross [9], um processo estocástico é uma **Cadeia de Markov** se a probabilidade condicional de um estado futuro só depende do estado presente, ou seja:

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i) = P_{ij}, \quad (23)$$

onde P_{ij} é a probabilidade de transição do estado i para o estado j .

Note-se que a Cadeia de Markov é um Processo Markoviano. Se o processo for uma cadeia de Markov homogênea, ou seja, $\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n] = p(x_n, x_{n+1})$, então $p(x_n, x_{n+1})$ será dado por uma *Matriz de Transição* P , dada por:

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & & & \cdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix}$$

Exemplo 3.1 Transformando um processo em uma cadeia de Markov.

Suponha que chover ou não chover hoje depende das condições do tempo dos últimos dois dias. Especificamente, suponha que se choveu nos últimos dois dias, então choverá amanhã com probabilidade 0,7; se choveu hoje, mas não ontem, então choverá amanhã com probabilidade 0,5; se choveu

ontem, mas não hoje, então choverá amanhã com probabilidade 0,4; se não tiver chovido nos últimos dois dias, então choverá amanhã com probabilidade 0,2.

Se o estado do tempo (chover ou não chover) for deixado no dia n , será preciso saber apenas se chove ou não chove no dia n para que esse processo seja uma cadeia de Markov, o que caracteriza que o modelo anterior não é uma cadeia de Markov. No entanto, pode-se transformar este modelo em uma cadeia de Markov, dizendo que o estado do tempo em qualquer momento é determinado pelas condições meteorológicas do dia atual e do dia anterior. Em outras palavras, o processo estará no:

- Estado 0 - se choveu tanto hoje como ontem,
- Estado 1 - se choveu hoje, mas não ontem,
- Estado 2 - se choveu ontem, mas não hoje,
- Estado 3 - se não choveu nem ontem nem hoje.

O problema é modelado, então, por uma cadeia de Markov com a seguinte matriz de probabilidades de transição:

$$P = \begin{bmatrix} 0,7 & 0 & 0,3 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,4 & 0 & 0,6 \\ 0 & 0,2 & 0 & 0,8 \end{bmatrix}$$

Exemplo 3.2 *Previsão do Tempo.*

Suponha que a chance de chuva amanhã dependa das condições meteorológicas de estar ou não chovendo hoje, e não em condições de tempo passado. Suponha também que, se chover hoje, então choverá amanhã com probabilidade α , e se não chover hoje, então choverá amanhã com probabilidade β .

Diz-se que o processo está no estado 0 quando chove e no estado 1 quando não chove. A matriz de transição nesse caso é:

$$P = \begin{bmatrix} \alpha & 1 - \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Definida a probabilidade de transição de um passo (P_{ij}), passa-se a definir a probabilidade de transição de n passos (P_{ij}^n), ou seja:

$$P_{ij}^n = \mathbb{P}(X_{n+k} = j | X_k = i), \quad n \geq 0, \quad i, j \geq 0$$

A **Equação de Chapman-Kolmogorov** que calcula essa probabilidade de transição de n passos é dada por:

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m, \quad \forall n, m \geq 0, \quad e \quad \forall i, j \tag{24}$$

A demonstração dessa fórmula encontra-se na obra de Ross [9].

Matricialmente, tem-se a seguinte equação:

$$P^{n+m} = P^n * P^m,$$

onde (*) representa multiplicação matricial.

Exemplo 3.3 *Aplicação da Matriz de Transição.*

Considere o exemplo anterior em que o estado do tempo (chover ou não) é considerado uma cadeia de dois estados de Markov. Se $\alpha = 0,7$ e $\beta = 0,4$, será calculada a probabilidade de que vai chover quatro dias a partir de hoje, uma vez que está chovendo hoje.

Primeiramente deve-se reescrever a matriz de transição de Markov com os devidos valores de α e β :

$$P = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}$$

Então:

$$P^{(2)} = P^2 \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} * \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} = \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix}$$

$$P^{(4)} = (P^2)^2 \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix} * \begin{bmatrix} 0,61 & 0,39 \\ 0,52 & 0,48 \end{bmatrix} = \begin{bmatrix} 0,5749 & 0,4251 \\ 0,5668 & 0,4332 \end{bmatrix}$$

A probabilidade desejada é $P_{00}^4 = 0,5749$.

3.2 O modelo Passeio Aleatório

Uma simples explicação do passeio aleatório, de acordo com Chung [3], seria o movimento de uma partícula ao longo de uma linha, onde cada passo dessa partícula é de uma unidade, seja para a direita ou para a esquerda, com probabilidade p e $q = 1 - p$, respectivamente, com $0 < p < 1$. Supondo que cada passo que é dado por essa partícula tenha uma unidade de tempo, o n -ésimo passo estará no tempo n . Ainda, as possíveis posições dessa partícula pertencem ao conjunto dos números inteiros \mathbb{Z} .

Para chegarmos a uma forma matemática de representar esse passeio aleatório, deve-se representar o n -ésimo passo, como segue:

$$\xi_n = \begin{cases} +1, & \text{com probabilidade } p \\ -1, & \text{com probabilidade } q, \end{cases}$$

onde os ξ_n representam o n -ésimo passo e são variáveis aleatórias independentes. Se a posição inicial por X_0 , então a posição no tempo n será:

$$X_n = X_0 + \xi_1 + \dots + \xi_n \tag{25}$$

Pode-se dizer, de acordo com Chung [3], que o passeio aleatório é representado pela sequência de variáveis aleatórias $\{X_n, n \geq 0\}$ que é um processo estocástico em tempo discreto.

Dando continuidade, todo o desenvolvimento do passeio aleatório será estudado. Para isso, serão mostrados alguns problemas relevantes, propostos por Chung [3], que levarão ao desenvolvimento total.

Problema 3.1 *Considere o intervalo $[0, c]$, onde $c = a + b$ e $a \geq 1, b \geq 1$. Se a partícula inicia no ponto a , qual é a probabilidade de que ela atingirá um ponto da extremidade do intervalo antes do outro?*

Este é um famoso problema discutido por Fermat e Pascal e resolvido, por Montmart, onde dois jogadores, Pedro e Paulo, jogam uma série de jogos em que Pedro ganha com probabilidade p e Paulo ganha com probabilidade q , e os resultados dos jogos sucessivos são assumidos como independentes. O perdedor paga um real por vez para o vencedor. Agora, se Pedro tem a reais e Paulo tem b reais, desde o início, e eles continuam a jogar até que um deles venha a falir, qual é a probabilidade de que Pedro falirá?

Nesta formulação a posição da partícula no tempo n torna-se o número de reais que Pedro tem depois de n partidas. Cada passo para a direita ele ganha 1 real e cada passo para a esquerda ele

perde 1 real. Se a partícula atinge 0 antes de atingir c , então Pedro perde todo seu capital inicial e está falido, por outro lado, se a partícula atinge c antes de atingir 0, então Paulo perde todo seu capital e está falido. O jogo termina quando uma dessas eventualidades ocorrer. Esta é a razão do nome histórico “O problema da ruína do jogador”.

A resolução deste problema vai depender dos valores atribuídos para o seguinte intervalo: $1 \leq j \leq c - 1$, onde u_j é a probabilidade da partícula atingir o 0 antes do c , começando por j . O problema é encontrar o valor de u_a , com a arbitrário, por isso precisa-se de todos os valores de u_j . Para tanto, deve-se encontrar uma relação entre u_a e u_j através da equação abaixo:

$$u_j = pu_{j+1} + qu_{j-1}, \quad 1 \leq j \leq c - 1, \quad (26)$$

onde as condições iniciais são: $u_0 = 1$ e $u_c = 0$.

Para melhor entendimento da equação 26, deve-se pensar na partícula como estando na posição j e considerar o que vai acontecer depois de dar um passo. Então ela terá probabilidade p quando estiver na posição $j + 1$, em que a hipótese da probabilidade de se chegar a 0 antes de c será u_{j+1} ; do mesmo modo, ela terá probabilidade q quando estiver na posição $j - 1$, a partir da hipótese da probabilidade de se chegar a c antes de 0 será u_{j-1} . Daí a probabilidade total u_j é igual à soma dos dois termos do lado direito da equação 26. Para os casos extremos, onde $j = 1$ e $j = c - 1$, utiliza-se os valores iniciais dados anteriormente (u_0 e u_c).

Como $p + q = 1$, substitui-se o lado esquerdo da equação 26 por $pu_j + qu_j$, e após alguns cálculos:

$$q(u_j - u_{j-1}) = p(u_{j+1} - u_j).$$

Se $\frac{q}{p} = r$ e $d_j = u_j - u_{j+1}$, então:

$$d_j = rd_{j-1}.$$

Com $d_j = r^j d_0$, tem-se:

$$1 = u_0 - u_c = \sum_{j=0}^{c-1} (u_j - u_{j+1}) = \sum_{j=0}^{c-1} d_j = \sum_{j=0}^{c-1} r^j d_0 = \frac{1 - r^c}{1 - r} d_0, \quad (27)$$

onde $r \neq 1$.

Continuando:

$$u_j = u_j - u_c = \sum_{i=j}^{c-1} (u_i - u_{i+1}) = \sum_{i=j}^{c-1} d_i = \sum_{i=j}^{c-1} r^i d_0 = \frac{r^j - r^c}{1 - r} d_0 \quad (28)$$

Segue que:

$$u_j = \frac{r^j - r^c}{1 - r^c}, \quad 0 \leq j \leq c \quad (29)$$

Considerando $r \neq 1$, a partir das equações 27 e 28:

$$\begin{aligned} 1 &= cd_0, \\ u_j &= (c - j)d_0 = \frac{c-j}{c}, \end{aligned} \quad (30)$$

$$u_a = \frac{b}{c}.$$

A primeira parte do problema foi resolvida, o valor de u_a foi encontrado. Agora, deve-se encontrar v_j , que é a probabilidade da partícula atingir c antes de 0, começando em j .

Partindo da explicação retro, mostra-se que a equação 26 será válida quando u for substituído pelo v e as condições iniciais de contorno forem trocadas por: $v_0 = 0$ e $v_c = 1$. Assim, pode-se encontrar

todos os valores de v_j por um método semelhante. Existem outras formas mais simples de se resolver esse problema, uma delas seria pensar em termos de jogadores. Se mudar p para q (r por $1/r$) e, ao mesmo tempo o j por $c - j$ (pois quando Pedro tem j reais, Paulo tem $(c - j)$ reais), então os seus papéis seriam trocados e assim u_j passaria a ser v_j . Ao fazer essas alterações nas equações 29 e 30, obtem-se:

$$v_j = \frac{1-r^j}{1-r^c}, \text{ se } p \neq q,$$

$$v_j = \frac{j}{c}, \text{ se } p = q.$$

Observa-se que em ambos os casos $u_j + v_j = 1$, para $0 \leq j \leq c$.

Problema 3.2 *Se a partícula começa dentro do intervalo $[0, c]$, qual é a probabilidade de que ela abandone o intervalo?*

Para resolução do problema (3.2) utilizaremos o teorema a seguir, cuja demonstração pode ser encontrada em [3].

Teorema 3.1 *Para qualquer passeio aleatório (com p arbitrário), a partícula irá quase certamente (ou ainda, com probabilidade 1) não permanecer em qualquer intervalo finito para sempre.*

Como conseqüência, define-se uma variável aleatória que denota o tempo de espera até que a partícula atinja a fronteira do intervalo, muitas vezes referida como “tempo de absorção” se os pontos de fronteira são considerados “barreiras de absorção”, ou seja, a partícula ser supostamente presa dentro do intervalo, logo que os atinge. Em termos de jogadores, também é conhecida como a “duração do jogo”. Considere para $1 \leq j \leq c - 1$, S_j como sendo a primeira vez que a partícula atinge 0 ou c , começando por j e denote sua esperança $E[S_j]$ por e_j .

A resposta ao problema 3.2 demonstra que S_j é quase certamente finito, portanto, é uma variável aleatória tomando valores inteiros positivos.

Seja agora um conjunto de relações para e_j , análogas as do u_j vista anteriormente.

$$e_j = pe_{j+1} + qe_{j-1} + 1, \quad 1 \leq j \leq c - 1,$$

onde as condições iniciais serão: $e_0 = 0$ e $e_c = 0$.

Considere somente a solução geral para $p = q = 1/2$. Para isso, seja $f_j = e_j - e_{j+1}$. Então:

$$\begin{aligned} f_j &= f_{j-1} + 2, \\ f_j &= f_0 + 2j, \\ 0 &= \sum_{j=0}^{c-1} f_j = c(f_0 + c - 1). \end{aligned}$$

Sendo $f_0 = 1 - c$, após alguns cálculos tem-se:

$$e_j = \sum_{i=j}^{c-1} f_i = \sum_{i=j}^{c-1} (1 - c + 2i) = j(c - j). \quad (31)$$

Uma vez que o passeio aleatório é simétrico, o tempo de absorção esperado deve ser o mesmo quando a partícula está a uma distância j de 0, ou de c (ou ainda, a uma distância $c - j$ de 0), portanto, é claro que $e_j = e_{c-j}$, o que se verifica na equação 31.

A partir de agora, é possível tirar algumas conclusões sobre as fórmulas anteriormente apresentadas. Primeiro, converte-se o intervalo $[0, c]$ para $[0, \infty)$. A partir das equações 29 e 30 teremos:

$$\lim_{c \rightarrow \infty} u_j = \begin{cases} r^j, & \text{se } r < 1, \\ 1, & \text{se } r \geq 1 \end{cases}$$

Intuitivamente, este limite deve significar a probabilidade de que a partícula atingirá 0 antes de "chegar a $+\infty$ " a partir de j , ou então a probabilidade de que Pedro será arruinado quando ele jogar contra um Paulo infinitamente rico, ou que não pode ser arruinado. Assim, ele simplesmente representa a probabilidade de que a partícula nunca vai chegar a 0 a partir de j , ou a eventual ruína de Pedro quando o seu capital é de j reais. Esta interpretação é correta e fornece a resposta para o problema seguinte.

Problema 3.3 *Se a partícula começa a partir de um a (≥ 1), qual é a probabilidade de que ela nunca atingirá o 0?*

A resposta é 1 se $p \leq q$ e $(q/p)^a$ se $p > q$. Observe que, quando $p \leq q$ a partícula é pelo menos tão provável que vá para a esquerda como para a direita, porém a primeira conclusão é mais plausível.

No caso $p > q$ as implicações para o jogo são curiosas. Se Pedro tem uma certa vantagem, então mesmo que tenha apenas 1 real e está jogando contra o Paulo, ele ainda tem uma chance de $1 - q/p$ de escapar da falência para sempre. Com efeito, se ele puder provar que, neste caso feliz, Pedro irá ganhar muito, X_n denotará sua fortuna depois de n jogos.

$$\mathbb{P}\{X_n \rightarrow +\infty | X_n \neq 0 \quad \forall n\} = 1.$$

Quando $p = q = 1/2$, o argumento acima não se aplica, e neste caso não há simetria entre direita e esquerda, nossa conclusão pode-se afirmar com mais força com o seguinte teorema. A prova deste teorema pode ser encontrada na obra de Chung [3].

Teorema 3.2 *A partir de qualquer ponto de um passeio aleatório simétrico, a partícula irá quase certamente atingir qualquer ponto qualquer número de vezes.*

Diz-se, resumidamente, que a partícula irá atingir qualquer ponto de \mathbb{Z} e que o passeio aleatório é recorrente (ou persistente). Essas noções são estendidas para as cadeias de Markov.

Em termos de jogo, o Teorema 3.2 tem a seguinte implicação: se o jogo é justo, então, quase que certamente, Pedro vence qualquer montante definido com antecedência como sendo seu objetivo, desde que ele possa se dar ao luxo de entrar em uma dívida com um valor arbitrariamente grande. O Teorema 3.2 garante apenas que ele acabará por ganhar, por exemplo, \$1.000.000 sem ter qualquer garantia de quanto ele pode ter perdido antes que ele ganhe essa meta. Uma previsão mais realista é dado na equação 30, que pode ser reescrita como:

$$u_a = \frac{b}{a+b} \quad e \quad v_a = \frac{a}{a+b}, \tag{32}$$

que diz que a chance de Pedro ganhar a sua meta b antes de perde a totalidade do seu capital, está em exata proporção inversa de a para b . Assim, se ele tem \$100, a sua chance de ganhar \$1.000.000 é igual a $100/1.000.100$, ou cerca de 1 em 10.000.

Vamos mencionar um outro método para derivar a equação 32. No caso $p = q$, $E(\xi_n) = 0$ para todo n e, conseqüentemente, tem-se da equação 25 que:

$$E[X_n] = E[X_0] + E[\xi_1] + \dots + E[\xi_n] = a.$$

Em termos de jogadores, isto significa que o capital esperado de Pedro permanece constante durante todo o jogo. Agora considere a duração do jogo em S_a , ou seja, a primeira vez que a partícula atinge 0 ou c começando por a . Isto significa que, S_a é uma variável aleatória que toma valores inteiros positivos. Note-se que X_{S_a} leva apenas dois valores 0 e c por definição.

$$\mathbb{P}(X_{S_a} = 0) = \rho, \quad \mathbb{P}(X_{S_a} = c) = 1 - \rho.$$

Então,

$$E[X_{S_a}] = \rho * 0 + (1 - \rho) * c = (1 - \rho)c.$$

E isso significa que,

$$\rho = 1 - \frac{a}{c} = \frac{b}{a + b}$$

Em poucas palavras, o argumento acima diz que o jogo continua sendo honesto inclusive no momento da sua rescisão.

Agora passa-se a descrever um processo que levará do limitado passeio aleatório simétrico ao **Movimento Browniano**. De acordo com [3], o botânico Inglês Brown observou em 1826, que partículas microscópicas suspensas em um líquido estão sujeitas a contínuos impactos moleculares e a executar movimentos de ziguezague.

Einstein e Smoluchovski constataram que, apesar de sua aparente irregularidade esses movimentos podem ser analisados pelas leis da probabilidade, na verdade, o deslocamento ao longo de um período de tempo segue uma distribuição normal. O resultado de Einstein (1906) elevou-se a uma derivação do teorema do limite central pelo método de equações diferenciais. O estudo do movimento browniano como um processo estocástico foi realizada por Wiener em 1923, precedido de um trabalho heurístico de Bachelier's, e logo foi desenvolvido em seu edifício moderno por Paul Lévy e seus seguidores. Juntamente com o processo de Poisson, ele constitui uma das duas espécies fundamentais de processos estocásticos, tanto na teoria e aplicação. É possível dar uma idéia de como o movimento Browniano pode ser alcançado através do passeio aleatório e descrever algumas de suas propriedades básicas.

A partícula em movimento, observado por Brown, mudou de curso no espaço tridimensional, mas é possível pensar em sua projeção sobre um eixo de coordenadas. Como alguns numerosos impactos são recebidos por segundo, encurta-se a unidade de tempo, mas também a unidade de comprimento, de tal forma a levar o modelo correto. Seja δ a nova unidade de tempo, em outras palavras o tempo entre dois impactos sucessivos. Então t/λ são passos tomados no tempo antigo t . Cada etapa é ainda uma variável aleatória simétrica, mas agora supondo que o passo é de magnitude $\sqrt{\delta}$, ou seja, para todos os k :

$$\mathbb{P}(\xi_k = \sqrt{\delta}) = \mathbb{P}(\xi_k = -\sqrt{\delta}) = \frac{1}{2}.$$

E então,

$$E[\xi_k] = 0, \quad \sigma^2(\xi_k) = \frac{1}{2}(\sqrt{\delta})^2 + \frac{1}{2}(-\sqrt{\delta})^2 = \delta.$$

Para $X_0 = 0$, da equação 25:

$$X_t = \sum_{k=1}^{t/\delta} \xi_k.$$

Se δ é muito menor que t , t/δ é grande e pode ser pensado como um número inteiro.

Uma outra forma de analisar o caminho aleatório é através do estudo da equação de difusão. A equação de difusão descreve um movimento que surge como resultado de um objeto ou organismo fazer muitos movimentos de curto prazo em direções aleatórias. A descrição difusiva do movimento aleatório emerge como um limite contínuo de tais passeios aleatórios quando o comprimento Δx de cada etapa e o tempo Δt necessários para cada etapa vai para zero de tal forma que o raio $(\Delta x)^2/\Delta t$ permanece constante. Para entender como isso funciona, é útil considerar um exemplo simples em uma dimensão espacial. Suponha que um organismo move uma distância Δx ao longo de uma linha para a esquerda com probabilidade $1/2$ para cada tempo Δt . Suponha que $p(x, t)$ é a probabilidade de que o organismo está na posição x em tempo t . Para chegar a esse ponto, nesse momento, ele deve ter dado um passo para a esquerda no tempo $t - \Delta t$ e depois mudou para a direita, ou deu um passo para a direita e mudou para a esquerda. Tem-se então:

$$p(x, t) = \frac{1}{2}p(x + \Delta x, t - \Delta t) + \frac{1}{2}p(x - \Delta x, t - \Delta t). \quad (33)$$

Se subtrair $p(x, t - \Delta t)$ de ambos os lados e dividir por Δt obtem-se:

$$\frac{p(x, t) - p(x, t - \Delta t)}{\Delta t} = \frac{1}{2\Delta t} [p(x + \Delta x, t - \Delta t) - 2p(x, t - \Delta t) + p(x - \Delta x, t - \Delta t)]. \quad (34)$$

Supondo agora, que seja imposto o seguinte escalonamento difuso:

$$\frac{(\Delta x)^2}{\Delta t} = 2D. \quad (35)$$

Tem-se que:

$$\frac{p(x, t) - p(x, t - \Delta t)}{\Delta t} = \frac{D}{(\Delta x)^2} [p(x + \Delta x, t - \Delta t) - 2p(x, t - \Delta t) + p(x - \Delta x, t - \Delta t)]. \quad (36)$$

Tomando o limite de (36), com $\Delta x, \Delta t \rightarrow 0$, teremos que, enquanto (35) permanecer em vigor ela produzirá a equação de difusão:

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2}. \quad (37)$$

Note-se que o escalonamento na equação 35, onde $2D$ é o quadrado da distância Δx movido pelo organismo em uma unidade de tempo Δt , produz um coeficiente que é igual a 1/2 do quadrado da distância percorrida por unidade de tempo. Esta interpretação do coeficiente de difusão D é válida em qualquer número de dimensões.

A equação 37 descreve a localização provável de um único organismo. Se ela for resolvida por $p(x, t)$ que corresponde a um único organismo a partir de tempo $t = 0$ na posição $x = z$ obtem-se:

$$p(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-(x-z)^2/4Dt}. \quad (38)$$

Se for iniciada com uma coleção de organismos em $t = 0$ com densidade $\mu_0(x)$, então a densidade esperada $\mu(x, t)$ em tempo T é obtida através da média da equação 38:

$$\mu(x, t) = \frac{1}{\sqrt{4\pi Dt}} \int_{-\infty}^{\infty} e^{-(x-z)^2/4Dt} \mu_0(y) dy. \quad (39)$$

A expressão da equação 38 é a solução fundamental para a equação de difusão 37, e assim, se o organismo estiver apenas se movendo e não morrer ou se reproduzir, tem-se:

$$\frac{\partial \mu}{\partial t} = D \frac{\partial^2 \mu}{\partial x^2}. \quad (40)$$

Esta é a equação de difusão para a densidade $\mu(x, t)$, que é a parte de difusão da reação típica de modelos de difusão. Em dimensões superiores do espaço, a equação 40 torna-se:

$$\frac{\partial p}{\partial t} = D \nabla^2 \mu = D \left(\frac{\partial^2 \mu}{\partial x^2} + \frac{\partial^2 \mu}{\partial y^2} \right).$$

O operador ∇^2 é conhecido como o Laplaciano ou Operador de Laplace e é por vezes denotado por Δ . O coeficiente D ainda representa 1/2 da distância média ao quadrado percorrida por um organismo na unidade de tempo de todas as dimensões do espaço. Para os modelos de n dimensões, a expressão correspondente a equação 38 é:

$$\frac{1}{\sqrt{(4\pi Dt)^{n/2}}} e^{-|x-z|^2/4Dt}.$$

quando $|x| = \sqrt{(x_1^2 + \dots + x_n^2)}$, se torna a distância euclidiana.

3.3 Movimento Browniano

Imagine uma nuvem de fumaça em um céu completamente sem vento. Com o passar do tempo, essa nuvem vai se espalhar ao longo do céu e a concentração de fumaça irá variar de uma forma harmoniosa. No entanto, se uma única partícula de fumaça for observada, o seu caminho acaba por ser extremamente grosseiro, devido a colisões com outras partículas. Isto exemplifica aspectos do fenômeno chamado de difusão, onde a errática trajetória da partícula num nível microscópico, dá origem a um comportamento muito suave da densidade de todo o conjunto de partículas [11].

De acordo com Ross [9], o processo de Wiener (ou **Movimento Browniano**), é um dos processos estocásticos mais úteis na teoria de probabilidade aplicada. Esse fenômeno, descoberto pelo Inglês botânico Robert Brown, exibe o movimento de uma pequena partícula, que está totalmente imersa em um líquido ou gás. Desde então, o processo tem sido utilizado beneficentemente em áreas como a estatística para testes de ajustes, para análise dos níveis de preços no mercado de ações, e na mecânica quântica.

A primeira explicação do fenômeno do movimento browniano foi dada por Einstein em 1905. Ele mostrou que o movimento browniano poderia ser explicado supondo que a partícula imersa está continuamente sujeita ao bombardeamento das moléculas do meio envolvente. No entanto, a definição sucinta deste processo estocástico foi dada por Wiener em uma série de documentos originários de 1918. Abaixo, segue a definição de acordo com Tomasz [11].

Definição 3.5 *O Processo de Wiener ou Movimento Browniano é um processo estocástico $W(t)$ com valores em \mathbb{R} definidos para $t \in [0, \infty)$ tal que:*

1. $W(0) = 0$;
2. os caminhos $t \rightarrow W(t)$ são contínuos, quase que certamente,
3. para toda sequência $x_1, \dots, x_n \in \mathbb{R}$, com $0 < t_1 < \dots < t_n$ e conjuntos da forma $A_1 = (-\infty, x_1]$; ...; $A_n = (-\infty, x_n]$ tais que:

$$P\{W(t_1) \in A_1, \dots, W(t_n) \in A_n\} = \int_{A_1} \dots \int_{A_n} p(t_1, 0, x_1)p(t_2 - t_1, x_1, x_2) \dots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 dx_2 \dots dx_n,$$

onde,

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}, \quad (41)$$

para cada $x, y \in \mathbb{R}$ e $t > 0$ a equação 41 é denominada *Transição de Densidade*.

Exemplo 3.4 *Esperança e Variância.*

Mostre que,

$$f_{W(t)}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}},$$

é a função densidade de probabilidade do processo $W(t)$. Encontre a esperança e a variância de $W(t)$.

A partir do item 3 da definição 3.5 do movimento Browniano, tem-se que:

$$f_{W(t)}(x) = p(t, 0, x),$$

é a densidade de $W(t)$. Então, integrando por partes, a esperança será:

$$E[W(t)] = \int_{-\infty}^{\infty} xp(t, 0, x)dx = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2t}} dx = -\frac{t}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \frac{d}{dx} e^{-\frac{x^2}{2t}} dx = -\frac{t}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \Big|_{-\infty}^{\infty} = 0$$

E a variância será:

$$\begin{aligned} E[(W(t))^2] &= \int_{-\infty}^{\infty} x^2 p(t, 0, x) dx = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2t}} dx = -\frac{t}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x \frac{d}{dx} e^{-\frac{x^2}{2t}} dx = \\ &= -\frac{t}{\sqrt{2\pi t}} x e^{-\frac{x^2}{2t}} \Big|_{-\infty}^{\infty} + \frac{t}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2t}} dx = 0 + \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = t \end{aligned}$$

onde $u = \frac{x}{\sqrt{t}}$.

Para resolução desta integral foi usada uma dica dada por [11], onde:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

4 Séries Temporais

De acordo com Morettin e Toloi [10], *série temporal* é um conjunto de dados numéricos obtidos durante períodos regulares ao longo do tempo, como por exemplo: o preço diário de fechamento de determinada ação na Bolsa de Valores, publicações mensais do IPC (Índice de Preços ao Consumidor), entre outros. O objetivo principal de qualquer série temporal é fornecer uma descrição concisa de uma série histórica. Esta descrição pode consistir em um resumo estatístico, porém é mais provável que os dados sejam incluídos em representações gráficas. [12].

Nas séries temporais estacionárias, os fatores que influenciaram padrões da atividade no passado e no presente continuarão a fazê-lo no futuro, mais ou menos da mesma maneira. A análise dessas séries estacionárias tem como objetivo identificar e isolar esses fatores para fins de previsão, bem como o planejamento e o controle gerencial. Os métodos utilizados para fazer previsões de séries temporais envolvem a projeção dos valores futuros de uma variável, com base em observações do presente e do passado dessa variável [10].

As séries temporais podem ser definidas da seguinte forma:

$$[x(t)] \equiv x = \{x(0), x(1), \dots, x(T)\}. \quad (42)$$

Não se pode esquecer que, para as séries temporais estacionárias a média, variância e autocovariância permanecerão as mesmas independentemente do período de tempo em que sejam medidas, então:

$$\begin{aligned} E[x] &= m = \mu \\ Var[x] &= E[(x - E[x])^2] = E[x^2] - (E[x])^2 = \sigma^2 \\ Cov[x, y] &= E[(x - E[x])(y - E[y])] = E[xy] - E[x]E[y] \end{aligned}$$

Para facilitar os cálculos e o desenvolvimento de programas futuros, no *Apêndice D* encontra-se os programas desenvolvidos no Matlab para calcular a média e a variância de uma série dada.

Agora, serão apresentados alguns modelos cujo objetivo é a previsão dessas séries.

4.1 Modelos de Previsão

- **Modelo Autorregressivo - AR(p):**

Os modelos autorregressivos, de ordem p, têm a forma:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \dots + a_p y(t-p) + \epsilon(t) \quad \text{com } t > p \quad (43)$$

Onde $\epsilon(t)$ é o ruído branco.

De acordo com Diggle [12], o *ruído branco* é o exemplo mais simples de uma série estacionária, onde pode-se identificar algumas propriedades muito interessantes a seu respeito, pois é um termo de erro estocástico que tem média zero, variância constante e não é autocorrelacionado, isto é, sua covariância é zero.

$$E[\epsilon(t)] = 0,$$

$$Var[\epsilon(t)] = \sigma^2 \quad \forall t,$$

$$Cov[\epsilon(t), \epsilon(s)] = 0 \quad \text{se } t \neq s.$$

Exemplo 4.1 *AR(1)*.

$$AR(1) : \begin{cases} y(0) = \epsilon(0) \\ y(t) = a_1 y(t-1) + \epsilon(t) \end{cases}$$

Considerando $|a_1| < 1$ tem-se, para t variando de 1 a t :

$$\begin{cases} y(1) = a_1 y(0) + \epsilon(1) = a_1 \epsilon(0) + \epsilon(1) \\ y(2) = a_1 y(1) + \epsilon(2) = a_1^2 \epsilon(0) + a_1 \epsilon(1) + \epsilon(2) \\ \dots \\ y(t) = \sum_{i=0}^t (a_1)^i \epsilon(t-i). \end{cases}$$

Onde,

$$E[y(t)] = \sum_{i=0}^t (a_1)^i E[\epsilon(t-i)] = 0.$$

$$Var[y(t)] = \sum_{i=0}^t (a_1)^{2i} \underbrace{Var[\epsilon(t-i)]}_{\sigma^2} = \sigma^2 \underbrace{\sum_{i=0}^t (a_1)^{2i}}_{S_n} = \sigma^2 \frac{1 - (a_1)^{2n}}{1 - (a_1)^2} \rightarrow \frac{\sigma^2}{1 - (a_1)^2} \quad \text{com } n \rightarrow \infty.$$

Onde S_n é a formula da soma de uma PG, isto é:

$$S_n = \frac{a_1(1 - q^n)}{1 - q}.$$

Nessas condições, diz-se que $y(t)$ não se afasta da média, ou seja, $y(t)$ é reversível a média, pois não há memória do ruído antigo.

Note que no modelo AR(1), se $a_1 = 1$ tem-se o chamado *Caminho Aleatório*, ou *Random Walk*:

$$\begin{cases} y(0) = \epsilon(0) \\ y(t) = y(t-1) + \epsilon(t); \quad t > 0. \end{cases}$$

Nesse caso,

$$y(t) = \sum_{i=0}^t \epsilon(t-i).$$

$$Var[y(t)] = \sigma^2(t+1).$$

Ou seja, $y(t)$ não é reversível a média.

Considere agora a série de diferenças para o *Random Walk*:

$$I(1) : \begin{cases} x(0) = \epsilon(0) \\ x(t) = y(t) - y(t-1) = \epsilon(t); \quad t > 0. \end{cases}$$

Onde $Var[x(T)] = \sigma^2$.

Note que agora $x(t)$ fica reversível à média.

- **Modelo Média Móvel Puro - MA(q):**

Os modelos de médias móveis puros, de ordem q, têm a forma:

$$y(t) = b_1 \epsilon(t-1) + \dots + b_q \epsilon(t-q) + \epsilon(t). \quad (44)$$

Exemplo 4.2 $MA(1)$.

$$MA(1) : \begin{cases} \epsilon(0) = 0 \\ y(t) = \epsilon(t) + b_1\epsilon(t-1). \end{cases}$$

Abrindo a expressão acima para t variando de 1 a t , tem-se:

$$\begin{cases} y(1) = \epsilon(1) \\ y(2) = \epsilon(2) + b_1\epsilon(1) = \epsilon(2) + b_1y(1) \\ y(3) = \epsilon(3) + b_1\epsilon(2) = \epsilon(3) + b_1(y(2) - b_1y(1)) \\ y(4) = \epsilon(4) + b_1\epsilon(3) = \epsilon(4) + b_1(y(3) - b_1y(2) - b_1^2y(1)) \\ \dots \end{cases}$$

Enfim, pode-se dizer que $MA(1) \equiv AR(\infty)$. Essa noção de aproximar $MA(q)$ pelo $AR(p)$ sendo p suficientemente grande, chama-se *inversão*.

Teorema 4.1 *Essa inversão para $MA(q)$ é equivalente a equação:*

$$1 + b_1z + b_2z^2 + \dots + b_qz^q = 0, \quad (45)$$

onde esta, tem todas as soluções complexas fora do círculo unitário.

- **Modelo Autorregressivo e de Média Móvel - ARMA(p,q):**

Basicamente o modelo ARMA é a combinação das equações (43) e (44). Então, os modelos autorregressivos com médias móveis de ordem (p,q) tem a forma:

$$y(t) = a_1y(t-1) + \dots + a_py(t-p) + b_1\epsilon(t-1) + \dots + b_q\epsilon(t-q) + \epsilon(t) \quad (46)$$

Observação 4.1 *Este modelo é reversível à média.*

É interessante notar que a modelagem ARMA pode descrever uma série temporal usando menos parâmetros do que as modelagens AR e MA separadamente, ou seja, ao modelar uma série temporal usando somente termos autorregressivos (ou médias móveis) isso pode exigir um número muito grande de parâmetros o que não aconteceria se fosse utilizado uma modelagem ARMA. Por exemplo, suponha que uma série possa ser modelada por um $AR(10)$ ou $ARMA(1,1)$, isto significa que no $AR(10)$ precisaria encontrar a_1, \dots, a_{10} e na modelagem $ARMA(1,1)$ precisaria somente encontrar a_1 e b_1 .

- **Modelo Autorregressivo Integrado de Médias Móveis - ARIMA(p,d,q):**

Os modelos de séries temporais discutidos até agora, se baseiam na hipótese de que as séries temporais envolvidas são estacionárias, ou seja, a média e a variância são constantes e sua covariância é invariável no tempo. Mas sabe-se que muitas séries temporais econômicas são não-estacionárias. Pode-se obter série estacionária de uma não-estacionária, a partir da diferença de seus termos, que é denominado com um parâmetro de integração d . Assim, $I(d)$ representa a quantidade de vezes que a série foi integrada para se chegar a uma série estacionária.

Portanto, se for preciso diferenciar uma série temporal d vezes para torná-la estacionária e então aplicar o modelo $ARMA(p,q)$, diz-se que a série temporal original é **ARIMA(p,d,q)**, ou seja, é uma

série temporal **autorregressiva integrada de média móvel**, em que p indica o número de termos autorregressivos, d o número de vezes em que a série tem que ser integrada para se tornar estacionária e q o número de termos de média móvel.

Uma questão interessante é como saber, olhando para uma série temporal, se ela segue um processo puramente AR (qual o valor de p) ou um processo puramente MA (qual o valor de q) ou um processo ARMA (qual o valor de p e q), ou ainda um processo ARIMA. É para isso que usa-se a *Metodologia de Box-Jenkins (BJ)* que de acordo com Morettin e Toloi [10], consiste em *identificar*, ou seja, descobrir os valores apropriados para p e q , a partir da análise das funções de autocorrelação e autocorrelação parcial (FAC e FACP); *estimar*, ou seja, onde os parâmetros do modelo identificado são estimados e finalmente *verificar* ou *diagnosticar*, ou seja, através de uma análise de resíduos pode-se descobrir se o modelo é adequado para os fins em vista. Caso o modelo não seja adequado, este ciclo é repetido.

É importante destacar que para usar a metodologia de Box-Jenkins, deve-se ter, ou uma série temporal estacionária ou uma série temporal que seja estacionária depois de uma ou mais diferenciações.

4.2 Funções de Autocorrelação (FAC) de Autocorrelação Parcial (FACP)

A **Função de Autocorrelação (FAC)** mede a correlação entre $y(t)$ e $y(t - k)$, ou seja, o quanto um pode influenciar no outro. Quando a FAC tende a zero, significa que não existe correlação entre os termos. A não existência de correlação implica em estacionariedade, caso contrario, não-estacionariedade.

Para chegar na sua fórmula, usa-se a equação de autocovariância, denotada por γ_k , dada abaixo:

$$\gamma_k = cov[y(t), y(t - k)] = E[(y(t) - \mu)(y(t - k) - \mu)] = E[y(t)y(t - k)] - \mu^2. \quad (47)$$

Lembrendo que a autocovariância no instante zero, γ_0 , é igual a variância de $y(t)$. A partir da equação 47, obtem-se a função de autocorrelação, denotada por ρ_k :

$$FAC_k = \rho_k = \frac{\gamma_k}{\gamma_0} = \sum_{t=k+1}^T \frac{[(y(t) - \mu)(y(t - k) - \mu)]}{(T - k)\gamma_0}. \quad (48)$$

Propriedades 4.1

1. $\rho_0 = 1$
2. $\rho_k \in (-1, 1)$; $k \neq 0$.

A **Função de Autocorrelação Parcial (FACP)** é definida como o último coeficiente da autorregressão de ordem k da série $Y(t) = y(t) - \mu$. Da mesma forma, quando a FACP tende a zero, significa que não existe correlação entre os termos.

Em outras palavras, se :

$$Y(t) = a_1Y(t - 1) + \dots + a_kY(t - k) + \epsilon(t).$$

Então:

$$FACP_k = a_k$$

Para obter o termo a_k , deve-se usar a relação matricial de *Yule-Walker* [10].

Relação entre FAC e FACP

Para $Y(t) = y(t) - \mu$:

$$Y(t - k)Y(t) = a_1Y(t - k)Y(t - 1) + \dots + a_kY(t - k)Y(t - k) + Y(t - k)\epsilon(t).$$

Tomando a média:

$$E[Y(t-k)Y(t)] = \sum_{i=1}^k a_i \underbrace{E[Y(t-k)Y(t-i)]}_{\text{cov}[y(t')y(t'+k+i)]} + \underbrace{E[Y(t-k)\epsilon(t)]}_{=0}. \quad (49)$$

Note que, $E[Y(t-k)Y(t)] = E[(y(t) - \mu)(y(t-k) - \mu)] = \text{cov}[y(t), y(t-k)] = \gamma_k$.

Note também que, para $t' = t - k$ pode-se reescrever $\text{cov}[y(t')y(t'+k+i)]$ como γ_{k-i} .

Como $E[Y(t-k)\epsilon(t)] = E[(y(t-k) - \mu)\epsilon(t)]$ e sabe-se que $E[y(t)] = \mu = 0$, conclui-se que $E[Y(t-k)\epsilon(t)] = 0$.

Feitos esses cálculos, pode-se reescrever a equação 49 como:

$$\gamma_k = \sum_{i=1}^k a_i \gamma_{k-i}.$$

Dividindo tudo por γ_0 :

$$\rho_k = \sum_{i=1}^k a_i \rho_{k-i}.$$

Abrindo o somatório:

$$\rho_k = a_1 \rho_{k-1} + a_2 \rho_{k-2} + \dots + a_k \rho_0.$$

Mas $\rho_0 = 1$, então:

$$\underbrace{a_k}_{FACP_k} = \underbrace{\rho_k}_{FAC_k} - a_1 \rho_{k-1} - \dots - a_{k-1} \rho_1. \quad (50)$$

Como é visto na relação 50, a FACP de ordem k retrata a autocorrelação de ordem k sem a influência das autocorrelações de ordem inferiores a k .

Foram desenvolvidos programas no Matlab que calculam a FAC e a FACP, e um outro programa que exhibe os seus gráficos. Tais programas podem ser encontrados no *Apêndice E*.

Exemplo 4.3 *Calcular a FAC e a FACP do modelo AR(1) e MA(1).*

Para AR(1), considera-se $|a_1| < 1$.

$$AR(1) : y(t) = a_1 y(t-1) + \epsilon(t).$$

$$E[y(t)] = \mu = 0.$$

$$Var[y(t)] = \gamma_0 \approx \sigma^2.$$

Para facilitar nossa visualização, foi simulado o modelo AR(1) de tal forma que $a_1 = 0,5$, $\mu = 0$ e $\sigma^2 = 1$, como mostra a figura (5).

Para $k \neq 0$, considera-se $Y(t) = y(t) - \mu = y(t)$, pois $\mu = 0$.

$$y(t+k) = a_1 y(t+k-1) + \epsilon(t+k).$$

Multiplicando tudo por $y(t)$:

$$y(t)y(t+k) = a_1 y(t)y(t+k-1) + y(t)\epsilon(t+k).$$

Tomando a média e dividindo por γ_0 :

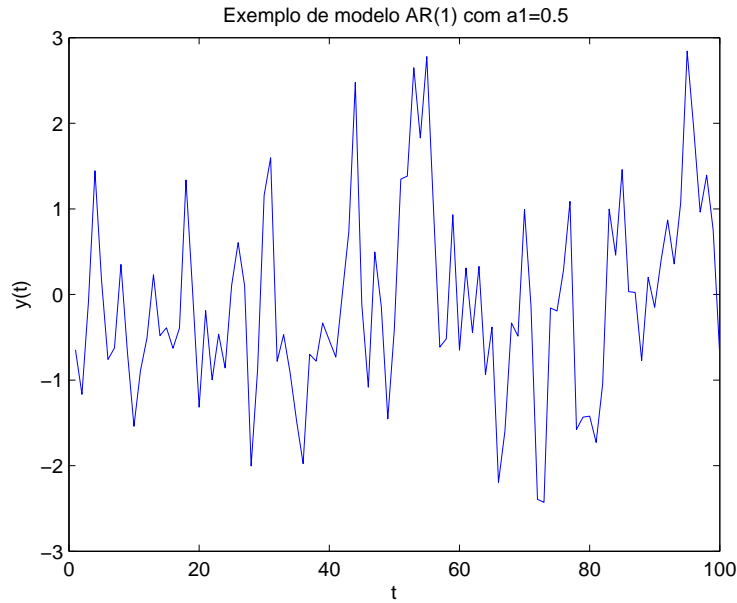


Figura 5: AR(1) com $a_1 = 0,5$, $\mu = 0$ e $\sigma^2 = 1$

$$\rho_k = a_1 \rho_{k-1}.$$

Por indução:

$$\rho_k = a_1 \rho_{k-1} = a_1^2 \rho_{k-2} = \dots = a_1^k \rho_0.$$

Portanto, $FAC_k = a_1^k$.

A partir da série simulada, percebe-se graficamente que a hipótese (comportamento exponencial) é verdadeira, como mostra a figura 6.

Como visto anteriormente, a FACP é o último coeficiente da autorregressão da série $Y(t) = y(t) - \mu$, mas $\mu = 0$, então basta tomar o último coeficiente de $Y(t) = y(t)$, ou seja, $\mathbf{FACP} = a_1$. Como a ordem da série é $k = 1$, só considera a FACP nesse valor, isto significa que para $k > 1$ a $FACP_k = 0$.

Da mesma forma, encontra-se a FACP da série simulada e a figura 7 mostra a confirmação do comportamento esperado.

Demonstra-se o mesmo comportamento para $p > 1$, isto é, na modelagem AR(p) a $FACP_k = 0$, para $k > p$ e $FAC_k \rightarrow 0$ exponencialmente, quando $k \rightarrow \infty$. Isso significa que p é definido pela FACP.

Agora a FAC e FACP do modelo MA(1) será calculada.

Para MA(1), considere $|b_1| < 1$.

$$MA(1) : y(t) = \epsilon(t) + b_1 \epsilon(t-1).$$

Onde $E[y(t)] = \mu = 0$.

Para facilitar a visualização, simula-se o modelo MA(1) de tal forma que $b_1 = 0,5$, $\mu = 0$ e $\sigma^2 = 1$, como mostra a figura 8.

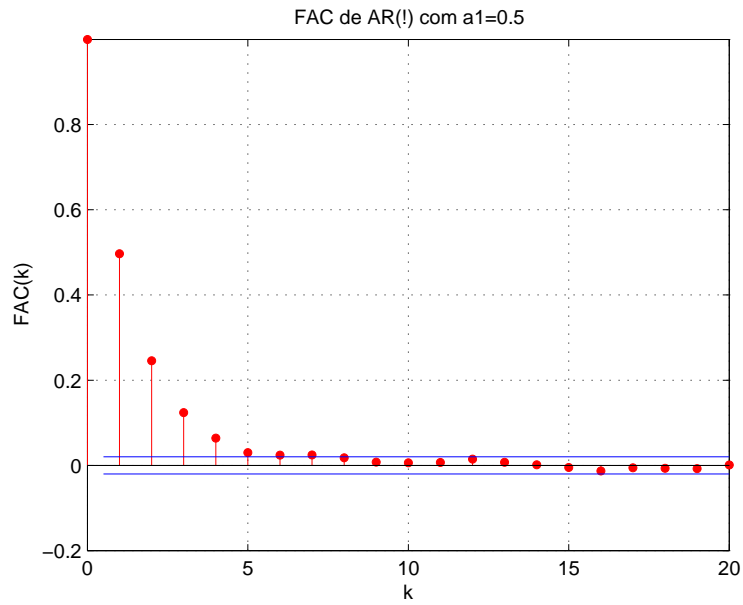


Figura 6: FAC de AR(1) com $a_1 = 0,5$

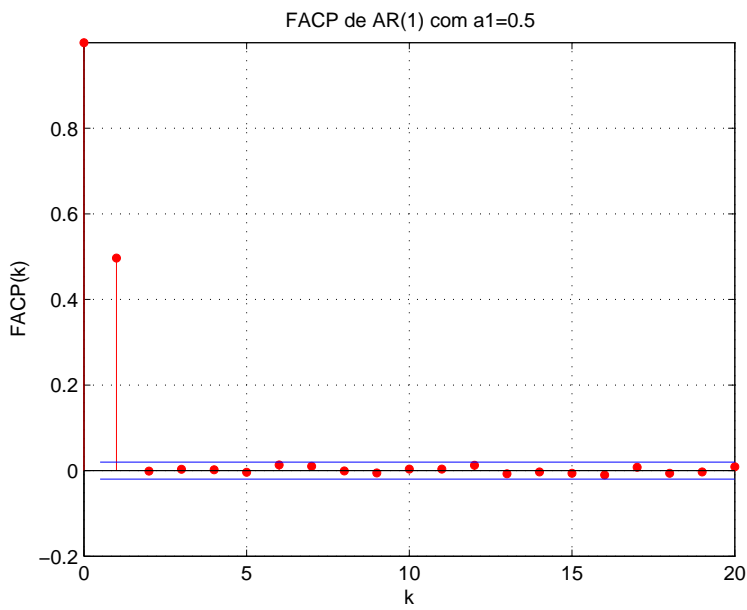


Figura 7: FACP de AR(1) com $a_1 = 0,5$

Primeiramente multiplica-se a série $y(t)$ por ela mesma, em seguida tira-se a média, de tal forma que obtenha-se o valor de γ_0 .

$$y(t)y(t) = b_1 y(t)\epsilon(t-1) + y(t)\epsilon(t).$$

$$y(t)y(t) = b_1 [b_1 \epsilon(t-1) + \epsilon(t)] \epsilon(t-1) + [b_1 \epsilon(t-1) + \epsilon(t)] \epsilon(t).$$

Tomando a média:

$$\gamma_0 = \sigma^2 (b_1^2 + 1).$$

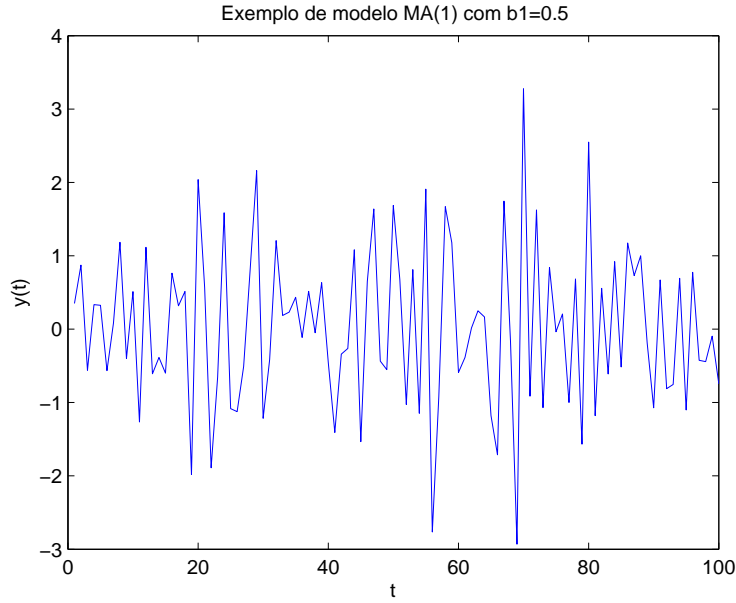


Figura 8: MA(1) com $b_1 = 0,5$, $\mu = 0$ e $\sigma^2 = 1$

Agora multiplica-se $y(t)$ por $y(t-1)$, em seguida calcula-se a média de tal forma que se obtenha o valor de γ_1 .

$$y(t-1)y(t) = b_1y(t-1)\epsilon(t-1) + y(t-1)\epsilon(t)$$

$$y(t)y(t) = b_1[b_1\epsilon(t-2) + \epsilon(t-1)]\epsilon(t-1) + [b_1\epsilon(t-2) + \epsilon(t-1)]\epsilon(t)$$

Tomando a média:

$$\gamma_1 = b_1\sigma^2$$

Então a FAC_1 será:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{b_1}{b_1^2 + 1}$$

Refazendo o mesmo procedimento, agora para $y(t)$ multiplicado por $y(t-k)$, tem-se que:

$$\gamma_k = 0$$

Portando, $FAC_k = \rho_k = 0$, para todo $k > 1$.

A partir da série simulada, percebe-se graficamente que a hipótese é verdadeira, como mostra a figura 9.

Para o cálculo da FACP usa-se as equações de Yule-Waller, que se encontram na obra de Morettin e Toloi [10], e é possível mostrar que $FACP_k$ é uma exponencial. Para visualização do que acontece com o gráfico da FACP, pode-se observar na figura 10 que foi gerada a partir da série simulada.

Demonstra-se o mesmo comportamento para $q > 1$, isto é, na modelagem MA(q) a $FAC_k = 0$, para $k > q$ e $FACP_k \rightarrow 0$ quando $k \rightarrow \infty$ exponencialmente. Isso significa que q é definido pela FAC.

Abaixo o resumo da fase de identificação da metodologia BJ:

Feita a fase de identificação, passa-se para a fase em que estima-se os parâmetros do modelo. Para isso, utiliza-se o *Método dos Mínimos Quadrados (MMQ)*, no caso de funções lineares (nos parâmetros), e *Máxima Verossimilhança*, no caso de funções de distribuição de probabilidade (Vide Apêndices A e

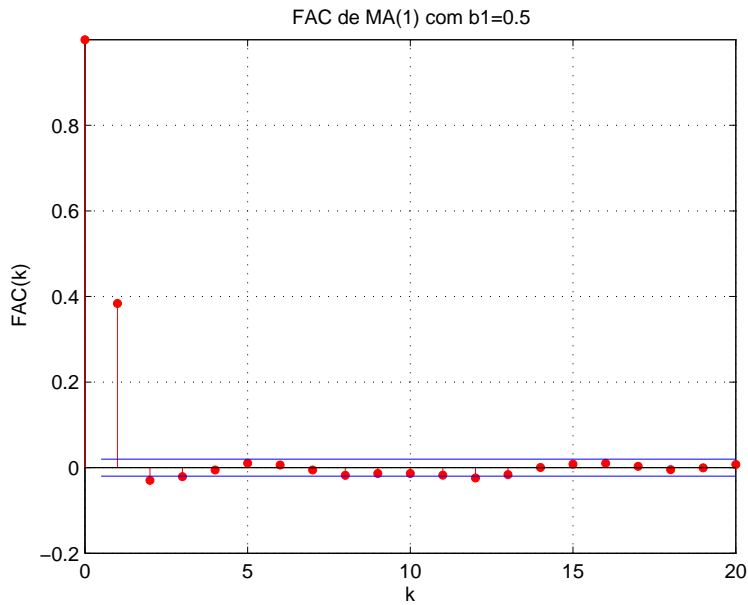


Figura 9: FAC de MA(1) com $b_1 = 0,5$

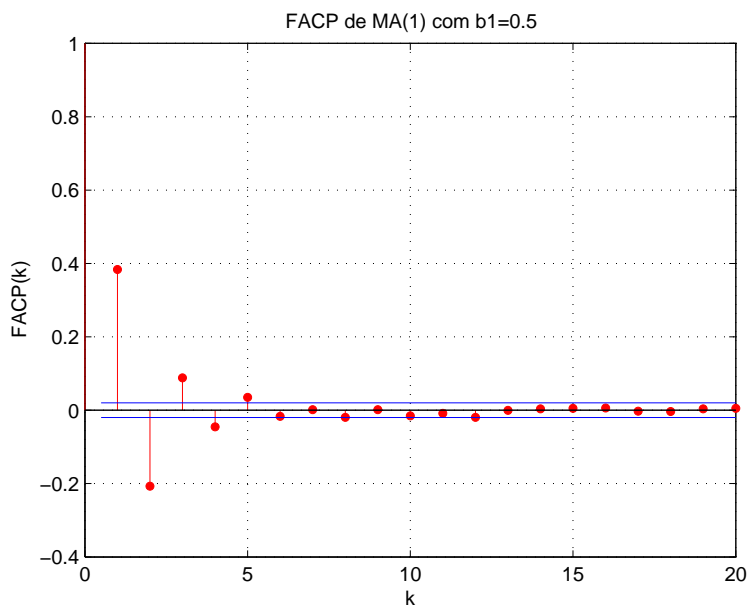


Figura 10: FACP de MA(1) com $b_1 = 0,5$

	AR(p)	MA(q)	ARMA(p,q)
FAC_k	decrece exponencialmente	se anula quando $k > q$	é igual a FAC_k do AR(p), para $k > p - q$
$FACP_k$	se anula quando $k > p$	decrece exponencialmente	é igual a $FACP_k$ do MA(q), para $k > p - q$

Tabela 4: Resumo dos Resultados

B).

Exemplo 4.4 Regressão ARIMA(2,0,1).

O programa gera 100 pontos de uma série temporal para simular o modelo ARIMA(2,0,1). A

figura (11) nos mostra a série simulada.

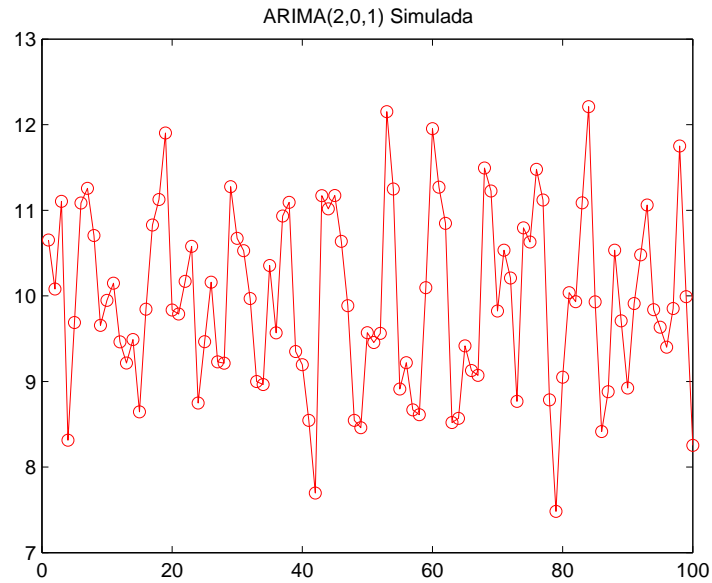


Figura 11: ARIMA(2,0,1) Simulada

É importante lembrar que, num modelo real os valores de p , d e q teriam que ser estimados, mas nesse exemplo sabe-se que são (2,0,1) respectivamente. Neste caso, exibem-se os gráficos da FAC e FACP para verificar se a regressão escolhida foi boa. Os resultados podem ser vistos nas figuras 12 e 13, respectivamente.

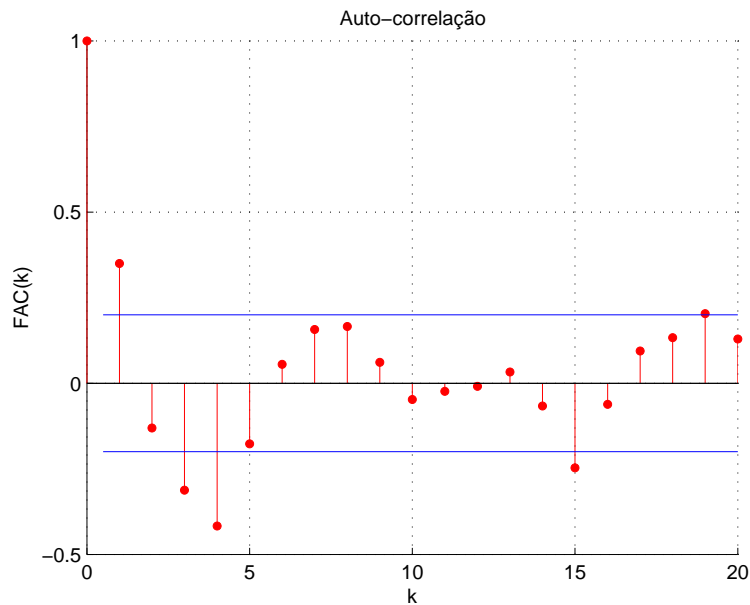


Figura 12: FAC - ARIMA(2,0,1)

Os valores da média e variância são, respectivamente: $\mu = -7,1054e - 016$ e $\sigma^2 = 0,9976$.

Na figura 14 encontra-se o gráfico da série simulada e da série ajustada pelo modelo. Já na figura 15 pode-se perceber que o resíduo da modelagem ARIMA(2,0,1) (diferença entre a série original e a gerada pelo modelo) se aproxima do ruído branco.

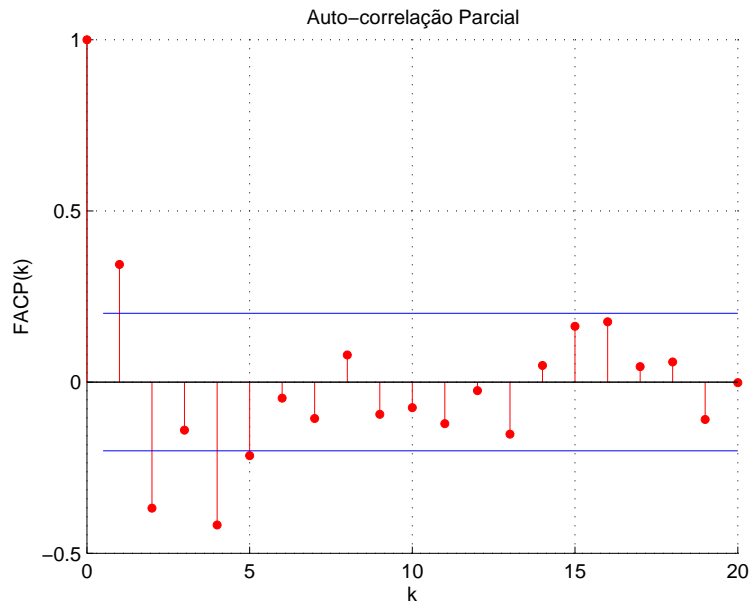


Figura 13: FACP - ARIMA(2,0,1)

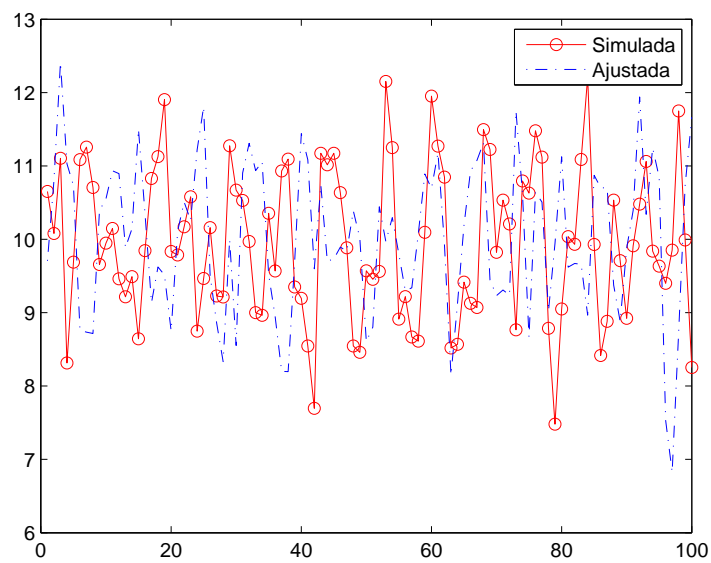


Figura 14: ARIMA(2,0,1)

5 Modelagem da Volatilidade

Serão apresentados agora alguns modelos apropriados para séries financeiras que apresentam a variância condicional evoluindo com o tempo. Foi visto anteriormente que os modelos do tipo ARIMA são lineares, com média zero e variância constante, usados para modelagem de séries estacionárias. Agora, serão apresentados os modelos não-lineares onde a média é zero, porém a variância pode não ser constante.

Seja a seguinte série de retornos:

$$X_t = \ln(P_t) - \ln(P_{t-1}).$$

E sejam:

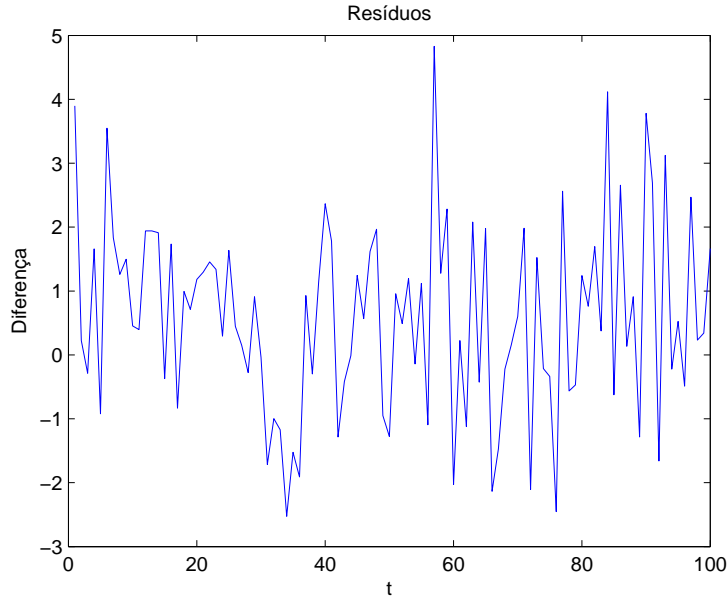


Figura 15: Resíduos da modelagem ARIMA(2,0,1)

$$\mu_t = E(X_t|F_{t-1}) \quad e \quad h_t = Var(X_t|F_{t-1}),$$

a média e a variância condicional de X_t , onde F_t denotará a informação até o instante t . Em alguns casos será suposto que $\mu_t = 0$, tendo assim $h_t = E(X_t^2|F_{t-1})$.

5.1 Modelos ARCH

De acordo com [13], os modelos ARCH (Modelos Autorregressivos com Heteroscedasticidade) foram introduzidos por Engle (1982), com o objetivo de estimar a variância da inflação. Basicamente, tem-se que o retorno X_t é não-correlacionado serialmente, mas a volatilidade (variância condicional) depende de retornos passados por meio de uma função quadrática.

Definição 5.1 Um modelo ARCH(r) é definido por:

$$X_t = \sqrt{h_t}\epsilon_t \quad \text{onde} \quad h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2, \quad (51)$$

onde, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i > 0$.

Na análise de modelos não-lineares os ruídos brancos (ϵ_t) são em geral supostos i.i.d. e o modelo tem a forma:

$$X_t = g(\epsilon_{t-1}, \epsilon_{t-2}, \dots) + \epsilon_t h(\epsilon_{t-1}, \epsilon_{t-2}, \dots),$$

de modo que $g(\cdot)$ representa a média condicional e $h^2(\cdot)$ é a variância condicional. Se $g(\cdot)$ for não-linear, o modelo diz-se *não-linear na média*, enquanto se $h(\cdot)$ for não-linear, o modelo diz-se *não-linear na variância*.

Note que o modelo:

$$X_t = \epsilon_t + \alpha \epsilon_{t-1}^2,$$

é não-linear na média, pois $g(\cdot) = \alpha \epsilon_{t-1}^2$ e $h(\cdot) = 1$. Agora, note que o modelo ARCH(1), que veremos logo abaixo, é não-linear na variância, pois teremos $g(\cdot) = 0$ e $h(\cdot) = \epsilon_t \sqrt{\alpha X_{t-1}^2}$.

Serão investigadas algumas propriedades dos modelos ARCH a partir do caso especial onde $r = 1$, ou seja,

$$X_t = \sqrt{h_t}\epsilon_t \quad \text{onde} \quad h_t = \alpha_0 + \alpha_1 X_{t-1}^2 \quad \text{com} \quad \alpha_0 > 0, \quad \alpha_1 \geq 0. \quad (52)$$

Calcula-se a média e a variância incondicionais da série:

- $E(X_t) = E[E(X_t|F_{t-1})] = 0$;
- $Var(X_t) = E(X_t^2) = E[E(X_t^2|F_{t-1})] = E(\alpha_0 + \alpha_1 X_{t-1}^2) = \alpha_0 + \alpha_1 E(X_{t-1}^2)$.

Se o processo X_t for estacionário de segunda ordem, então, para todo t , $E(X_t^2) = E(X_{t-1}^2) = Var(X_t)$, do que decorre:

$$Var(X_t) = \frac{\alpha_0}{1 - \alpha_1}.$$

Como $Var(X_t) > 0$, deve-se ter $0 \leq \alpha_1 < 1$.

- $Cov(X_t, X_{t+k}) = E(X_t X_{t+k}) = E[E(X_t X_{t+k}|F_{t+k-1})] = E[X_t E(\sqrt{h_{t+k}}\epsilon_{t+k}|F_{t+k-1})] = 0$, para $k > 0$, pois X_t está em F_{t+k-1} e $E(\epsilon_{t+k}|F_{t+k-1}) = 0$.

Para calcular a curtose, supondo que X_t siga um modelo ARCH(1), é necessário calcular o momento de quarta ordem de X_t , supondo que os ruídos brancos, ϵ_t , sejam normais para facilidade dos cálculos. Então, após alguns cálculos, tem-se como curtose de X_t :

$$K = \frac{\mu_4}{[Var(X_t)]^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3.$$

Como a curtose é maior do que 3, pode-se dizer que os retornos apresentam caudas longas.

Note que, uma desvantagem do modelo ARCH é que este trata retornos positivos e negativos de forma similiar, já que quadrados dos retornos entram na fórmula da volatilidade. Na prática, sabe-se que a volatilidade reage de forma diferente a retornos positivos e negativos. Também, devido ao fato de tem-se retornos ao quadrado, alguns grandes e isolados podem conduzir a super-previsões.

Utilizando o modelo ARCH(1) e calculando $X_t^2 - h_t$, te-se que:

$$X_t^2 - (\alpha_0 + \alpha_1 X_{t-1}^2) = h_t(\epsilon_t^2 - 1),$$

ou seja,

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t, \quad \text{onde} \quad v_t = h_t(\epsilon_t^2 - 1) = h_t(X - 1), \quad (53)$$

onde X é uma v.a. com distribuição $\chi^2(1)$, o que mostra um modelo AR(1) para X_t^2 , mas com erros não-gaussianos. Note que, v_t é uma sequência de v.a. de média zero, não-correlacionadas, mas com variância não-constante.

Da equação 53 tem-se que a FAC de X_t^2 é dada por:

$$\rho_{X^2}(k) = \alpha_1^k, \quad k > 0.$$

Para um modelo ARCH(r) tem-se:

$$X_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + v_t,$$

onde os v_t são como no caso $r = 1$. Ou seja, tem-se um modelo AR(p) para X_t^2 , com inovações não-gaussianas. Além disso, pode-se demonstrar que os retornos X_t também formam um ruído branco, com variância dada por:

$$Var(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^r \alpha_i}.$$

De acordo com Morettin [13], da mesma forma feita na modelagem ARIMA, deve-se considerar três estágios: identificação, estimação e verificação.

A identificação começa com o primeiro passo na construção de modelos ARCH, que é tentar ajustar modelos ARMA, para remover a correlação serial na série, se esta existir. Se este for o caso:

$$\phi(B)X_t = \theta(b)a_t,$$

sendo que $a_t \approx ARCH(r)$. Quando for feita referencia a X_t , estará sendo suposto que a série é não-correlacionada, ou então que ela é o resíduo da aplicação de um modelo ARMA à série original.

Para verificar se a série apresenta heteroscedasticidade condicional, pode-se utilizar dois testes, examinando-se a série X_t^2 .

- Teste de Box-Pierce-Ljung para X_t^2 .
- Teste de multiplicadores de Lagrange.

Esses testes podem ser vistos mais detalhadamente na obra de Morettin [13] ou ainda no livro de Engle (1982).

Os estimadores dos parâmetros do modelo são obtidos pelo método de máxima verossimilhança condicional. Para maiores detalhes, ver *Apêndice B*.

Uma maneira de verificar se o modelo é adequado é calcular a estatística Q de Ljung-Box para a sequência \tilde{X}_t . Além disso, calculam-se os coeficientes de assimetria e curtose estimados a fazer um grafico $Q \times Q$ para avaliar a suposição de normalidade.

Para verificar se ainda existe heteroscedasticidade condicional nos resíduos, pode-se aplicar o teste do Multiplicador de lagrange para a sequência \tilde{X}_t^2 .

5.2 Modelos GARCH

De acordo com [13], foi Bollerslev (1986,1987, 1988) quem sugeriu uma generalização do modelo ARCH, o chamado modelo GARCH ("generalized ARCH"). Foi visto que um modelo ARMA pode ser mais economico, no sentido de apresentar menos parâmetros do que um modelo AR ou MA puro. Do mesmo modo, um modelo GARCH pode ser usado para descrever a volatilidade com menos parâmetros do que um modelo ARCH.

Um grande diferencial da modelagem GARCH para a modelagem ARCH é que, os modelos GARCH incluem informações sobre a variância condicional. Pode-se notar na equação 54 que, os β_i são os coeficientes da variância condicional no período $t - j$.

Definição 5.2 Um modelo $GARCH(r,s)$ é definido por:

$$X_t = \sqrt{h_t}\epsilon_t \quad \text{com} \quad h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j}, \quad (54)$$

onde, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$ e $q = \max(r, s)$.

Os coeficientes positivos são uma condição suficiente, mas não necessária, para que $h_t > 0$.

A partir de Morettin [13], pode-se ver que no caso de modelos ARCH, usualmente supe-se que os ϵ_t são normais ou seguem uma distribuição t de Student.

Denota-se então, $v_t = X_t^2 - h_t$, de modo que, substituindo na equação 54 obtem-se:

$$X_t^2 = \alpha_0 + \sum_{i=1}^q (\alpha_i + \beta_i) X_{t-i}^2 + v_t - \sum_{j=1}^s \beta_j v_{t-j},$$

ou seja, um modelo ARMA(q,s) para X_t^2 , onde v_t não é um processo i.i.d.

Um modelo bastante usado na prática é o GARCH(1,1), para o qual a volatilidade é expressa como

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1}, \quad (55)$$

com $0 \leq \alpha_1, \beta_1 < 1, \alpha_1 + \beta_1 < 1$.

De acordo com Morettin [13], tem-se que as vantagens e desvantagens dos modelos GARCH são as mesmas dos modelos ARCH. E ainda, as volatilidades altas são precedidas de retornos ou volatilidades grandes, que é o caso das séries financeiras.

Pode-se calcular facilmente a curtose do modelo 55, obtendo:

$$k = \frac{E(X_t^4)}{[E(X_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

Note que, com o denominador positivo, teremos que X_t segue um modelo GARCH e suas caudas serão mais longas do que as da normal.

A identificação da ordem de um modelo GARCH a ser ajustada a uma série real é difícil. Baseado em Moerttin [13], percebe-se que é recomendado o uso de modelos de ordem baixa, como (1,1), (1,2), (2,1) ou (2,2), para depois escolher o modelo com base em alguns critérios, tais como, valores de assimetria e curtose, da log-verossimilhança e de alguma função perda, como

$$\sum_{t=1}^N (X_t^2 - h_t)^2.$$

Os estimadores dos parâmetros do modelo 54 são obtidos pelo método de máxima verossimilhança condicional (ver *Apêndice B*).

As estimativas dos parâmetros são obtidas por meio de métodos numéricos de maximização.

Previsões da volatilidade, usando um modelo GARCH, podem ser calculadas de forma similar àquelas do modelo ARMA. As previsões, com origem em t , considerando um modelo GARCH(1,1) da forma 55, são dadas por:

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 X_t^2 + \beta_1 h_t.$$

Para $l > 1$:

$$\hat{h}_t(l) = \alpha_0 + \alpha_1 \underbrace{\hat{X}_t^2(l-1)}_{X_t = \sqrt{\hat{h}_t} \epsilon_t} + \beta_1 \hat{h}_t(l-1) = \alpha_0 + \alpha_1 \hat{h}_t(l-1) \hat{\epsilon}_t^2(l-1) + \beta_1 \hat{h}_t(l-1).$$

Substituindo $\hat{\epsilon}_t^2(l-1)$ por $E(\epsilon_{t+l-1}^2) = 1$, tem-se:

$$\hat{h}_t(l) = \alpha_0 + (\alpha_1 + \beta_1) \hat{h}_t(l-1), \quad l > 1.$$

Exemplo 5.1 *Série temporal simulada.*

Será usado um programa desenvolvido no MATLAB para fazer a análise da série temporal simulada com a modelagem GARCH(1,1). Seja o modelo GARCH(1,1):

$$X_t = \sqrt{\hat{h}_t} \epsilon_t \quad \text{com} \quad h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1}.$$

Observa-se a série simulada na figura 16.

Ajustando a série temporal $x(t)$ a um processo GARCH(1,1) através do método da máxima verossimilhança, obtem-se os seguintes valores estimados para os coeficientes: $\alpha_0 = 0,0084$, $\alpha_1 = 0,2009$ e $\beta_1 = 0,7587$.

Agora os dados, as variâncias e o resíduo estimado (diferença entre a série simulada e a ajustada pelo modelo) serão plotados na 17.

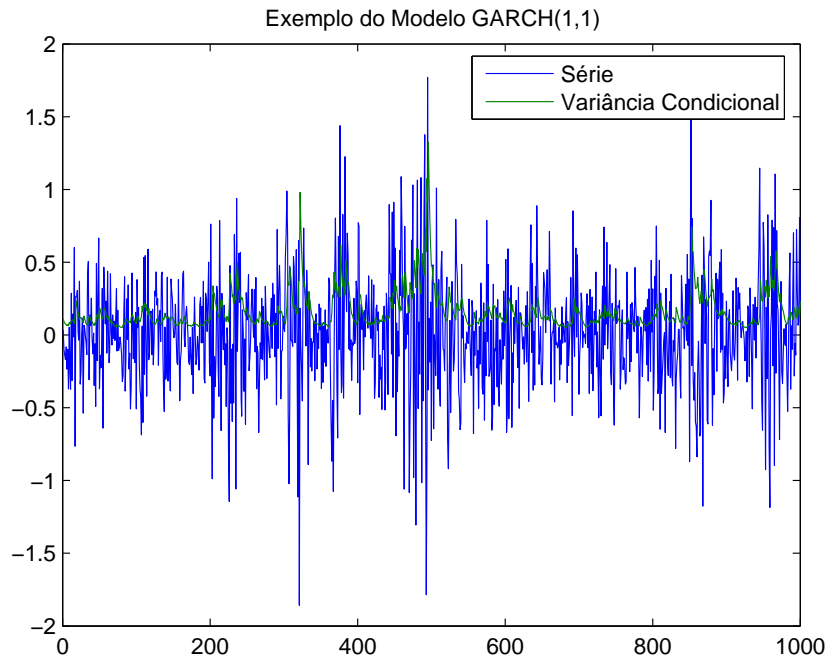


Figura 16: GARCH(1,1)

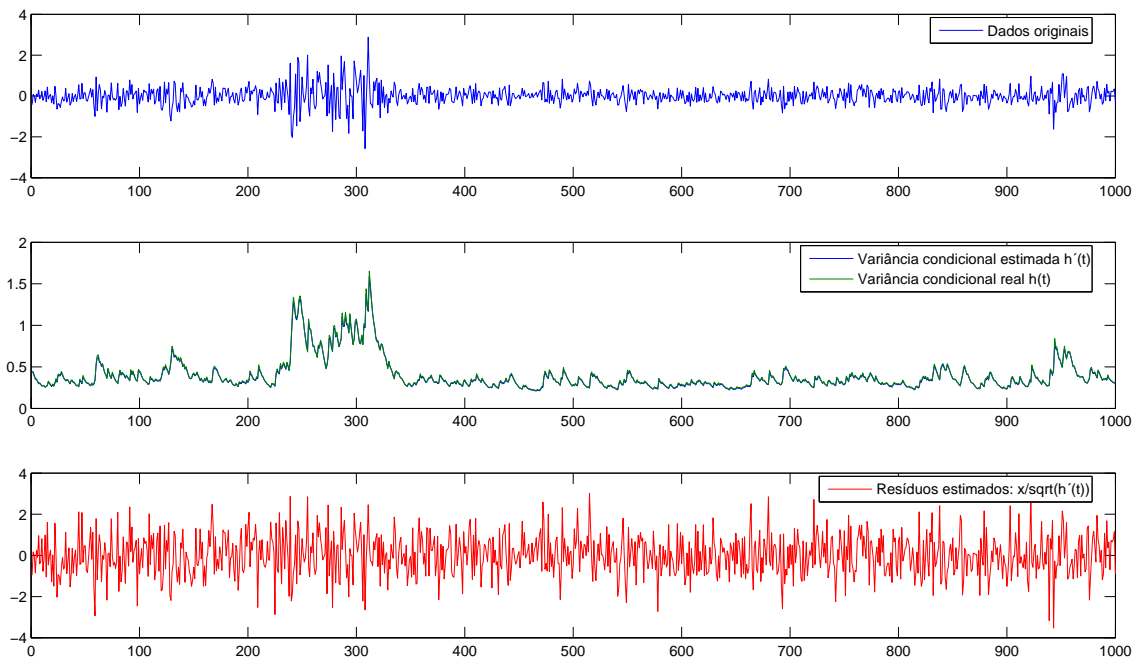


Figura 17: Resultados do modelo GARCH(1,1)

Exemplo 5.2 *Série de retornos da taxa de câmbio da moeda britânica.*

Na figura 18 encontra-se a série de preços da taxa de câmbio da moeda britânica, em relação ao dólar. Como a modelagem GARCH usa os dados de séries de retorno, estes serão deverão ser convertidos e para isso, será usado um programa do MATLAB que contém os dados de observações diárias

sobre o preço da taxa de câmbio da moeda britânica em relação ao dólar. Os resultados obtidos foram plotados na figura 19.



Figura 18: Taxa de câmbio da moeda Britânica

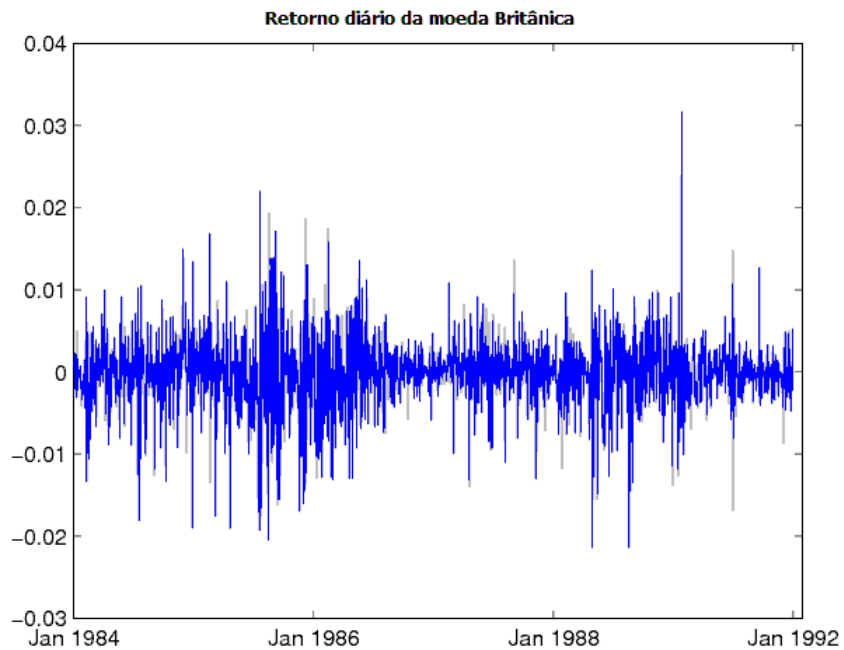


Figura 19: Retorno diário da moeda Britânica

O próximo passo é checar se existe correlação na série de retorno utilizando as funções FAC e FACP.

É possível observar na figura 20, os resultados da análise feita pela FAC, ou seja, que a amostra dos retornos está dentro do intervalo de confiança delimitado pelo desvio padrão, o que significa que não existe correlação entre eles.

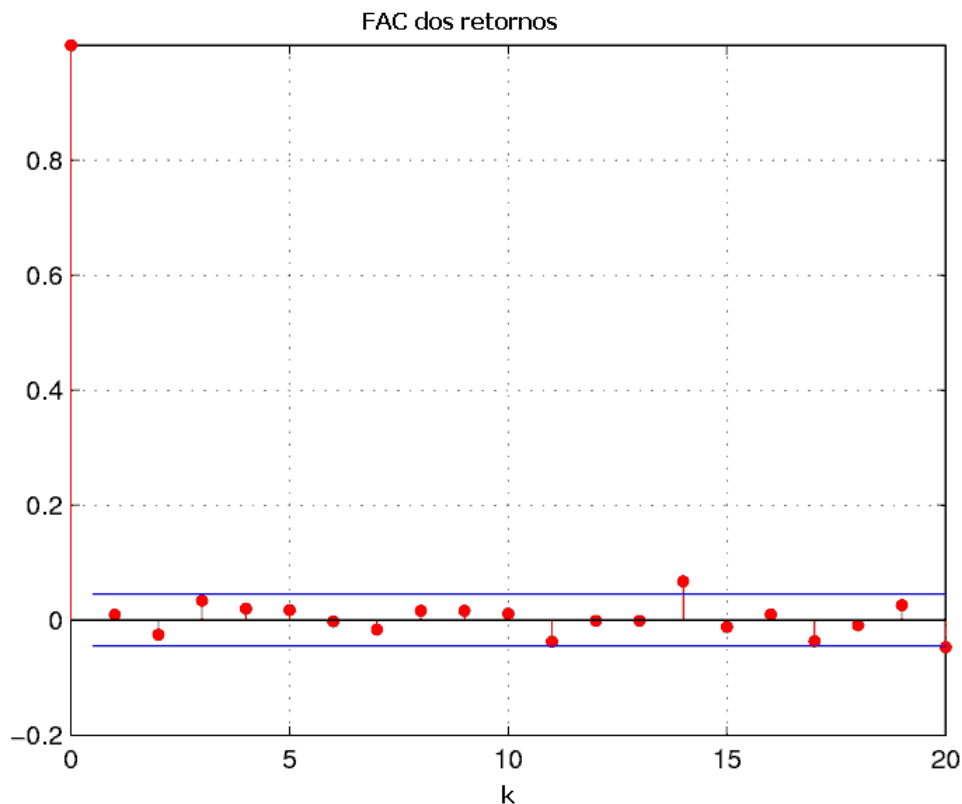


Figura 20: FAC dos Retornos

Analogamente, os mesmos cálculos serão feitos a partir da análise da FACP. A figura 21 mostra que os dados não estão correlacionados.

Agora a análise da FAC será feita para o quadrado dos retornos, e pode-se observar na figura 22 que existe correlação entre os dados de retornos da série, ou seja, a série dos quadrados dos retornos apresenta heterocedasticidade.

Uma outra forma de se verificar a heterocedasticidade é através da aplicação do teste baseado no trabalho de Engle, autor do modelo ARCH em 1982. Esse teste também pode ser aplicado na generalização GARCH, criada por Bollerslev em 1986. O teste LM (Langrange Multiplier) de Engle consiste em negar a afirmação (chamada de hipótese nula):

“Se os quadrados dos retornos com atraso k não são autocorrelacionados, então eles são assintoticamente distribuídos por uma função distribuição de probabilidades (fdp) Chi-quadrado com k graus de liberdade.”

A tabela 5 representa os resultados para o teste de Engle com atrasos $k = 10, 15, 20$. O valor P representa o nível de significância entre o teste de Engle e a distribuição χ^2 associada. Rejeitamos a hipótese nula se $P < 0.1$ (significância de 10%).

Pode-se ver na tabela 5 que os resultados obtidos pelo programa do MATLAB conferem, ou seja, a hipótese foi rejeitada o que significa que existe heterocedasticidade. Então, as figuras 20, 21 e 22 mostraram de maneira exata o que acontece com os dados.

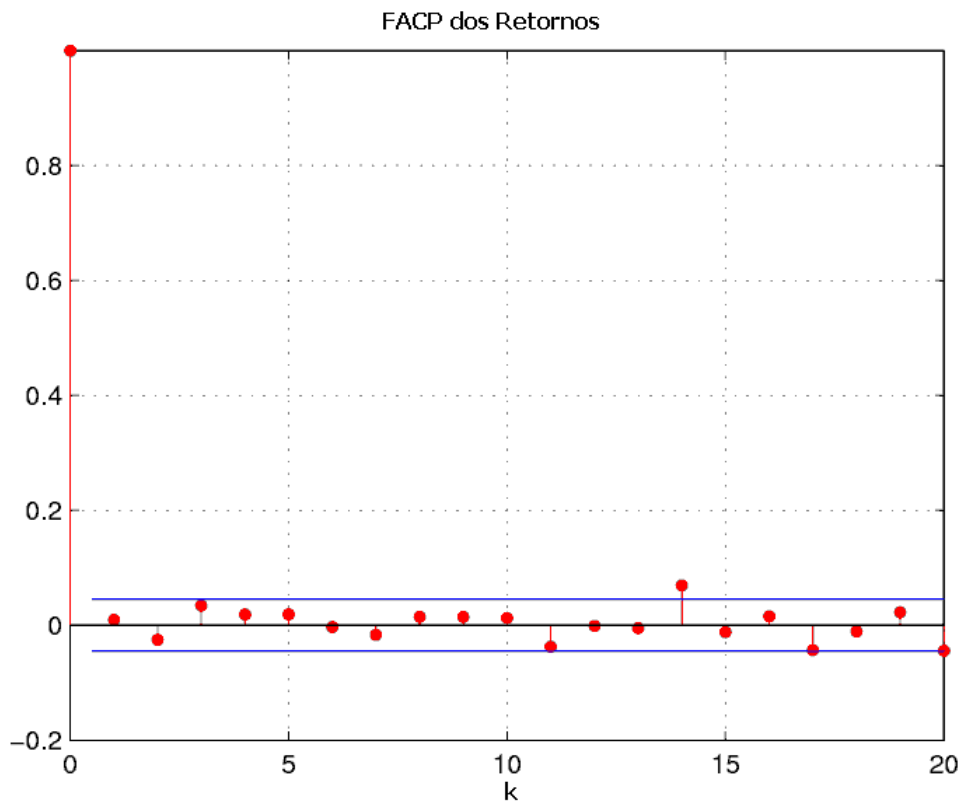


Figura 21: FACP dos Retornos

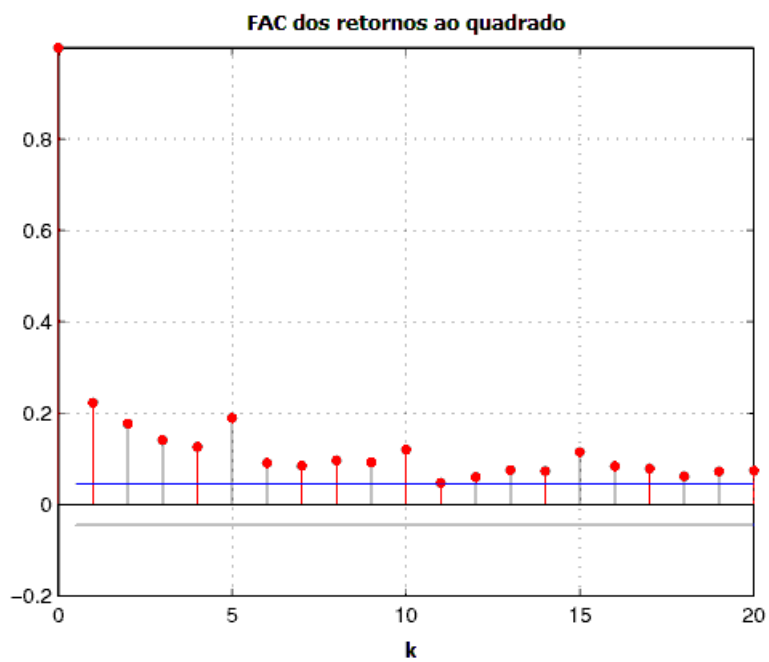


Figura 22: FAC - Retorno ao quadrado

O próximo passo é fazer graficamente as comparações dos resultados obtidos, ou seja, comparar o gráfico da série de retorno original, da série de retorno calculada através do GARCH(1,1) e do desvio padrão da série calculada. Segue na figura 23 esses três gráficos.

k	P	Teste ARCH	χ^2
10	0	192.3783	18.3070
15	0	201.4652	24.9958
20	0	203.3018	31.4104

Tabela 5: Teste de Engle

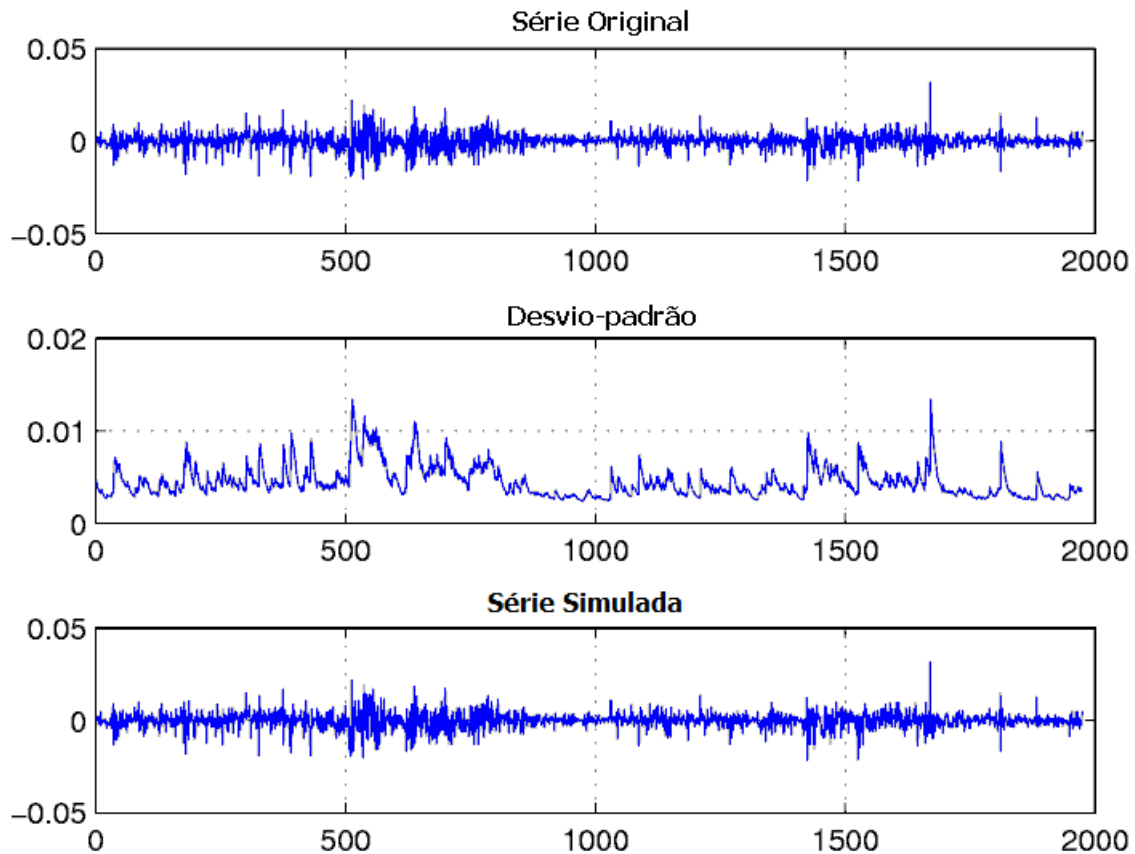


Figura 23: Resultados da Moeda Britânica

6 Algoritmos Genéticos

Uma das primeiras referências aos algoritmos genéticos (AG) foi feita por John Holland em 1975, que utiliza conceitos inspirado em evolução biológica, tais como herança, a seleção, cruzamento e mutação. É fato que ao longo de muitas gerações, as populações evoluíram de acordo com o princípio da seleção natural e da sobrevivência do mais forte, e é através da simulação desse processo que os algoritmos genéticos poderão desenvolver soluções para problemas reais, se forem devidamente codificados [14, 15].

Na natureza, indivíduos de uma população competem entre si em busca de recursos para sobreviver. Membros da mesma espécie também competem entre si para atrair uma fêmea ou um macho. Esses indivíduos, que serão os mais bem sucedidos, terão um número relativamente grande de descendentes, o que significa que os genes dos mais bem adaptados se espalharão para gerações futuras. Essas combinações de boas características dos diferentes ancestrais acabarão na produção de uma descendência muito boa, e ainda, serão melhores do que qualquer um de seus pais.

De acordo com Whitley [15], Algoritmos Genéticos são uma família de modelos computacionais inspirados pela evolução, vistos como otimizadores de funções, que têm como finalidade codificar uma solução potencial para um problema específico e aplicá-los aos operadores de recombinação de modo

a preservar as informações críticas. Os AGs usam uma analogia ligada ao comportamento natural das espécies, onde o membro da população que menos se adequou ao ambiente terá menos chances de ser selecionado para reprodução. Uma nova população de possíveis soluções será produzida, através da seleção dos melhores indivíduos da geração atual para se acasalarem e então produzi-la. Esta nova geração, por sua vez, terá a maior proporção das características de seus pais, que eram os melhores membros da geração anterior. Desta forma, percebe-se que as características boas se espalharão por toda a população e ainda serão misturadas e mudadas por outras novas e melhores características conforme as gerações evoluem [16].

O propósito do AG não é simplesmente avaliar ou melhorar uma única solução (ou indivíduo), mas sim analisar e modificar uma série de soluções simultaneamente. Apesar de ter sido criado para casos estocásticos e possuir algumas limitações, tais como convergência prematura, dependência de parâmetros e a dificuldade de encontrar um critério de parada eficiente, o AG tem sido estudado intensivamente e aplicado em vários campos [17].

A execução do algoritmo genético é um processo de dois estágios. Ele começa com a população atual da qual é feita uma seleção, que cria uma população intermediária. Em seguida, os operadores de recombinação (ou cruzamento) são aplicados à população intermediária para criar uma nova população. O processo de evolução da população atual para a nova população constitui uma *geração* na execução de um algoritmo genético [15].

Antes de usar os AGs precisa-se adequar o problema com uma *codificação* (ou representação). Necessita-se também de uma *função de aptidão*, a qual atribui uma pontuação para cada solução (ou indivíduo) codificada. Durante o processo, os “pais” devem ser selecionados para a *reprodução* e em seguida, serão recombinados para gerarem seus descendentes.

Na próxima seção será definida a codificação mencionada acima e também será explicado o seu significado de acordo com o uso nesse trabalho. Para tal, esta seção foi baseada na obra de Beasley, Bull e Martin [16].

6.1 Codificação e Implementação

A partir de uma explicação geral da implementação do AG, as codificações serão apresentadas. É importante lembrar de que os objetos de estudo deste trabalho são as ações do IBrX (índice da BOVESPA). O primeiro passo para a implementação de qualquer algoritmo genético é gerar uma *população inicial*, que no neste caso seria o conjunto de algumas possibilidades de aplicação numa certa *carteira* de ações, onde cada sequência de possibilidade é referida como *indivíduo* ou ainda *cromossomo*, representando proporções de investimento nas ações extraídas do IBrX que compõem essa carteira. De acordo com Dallagnol [14], esses cromossomos geralmente são representados por cadeias binárias, mas há outras codificações possíveis, tal como a codificação real que é a usada em nosso trabalho, onde cada *gene* (proporção de investimento em cada ação) do cromossomo é um número real. Os *genes* desses cromossomos respondem à pergunta “Qual ativo escolho para investimento ?”

Em seguida, avaliam-se esses cromossomos a partir da **função de aptidão** e atribui-se a eles possibilidades de reprodução de tal forma que os cromossomos que representam uma melhor solução para o problema-alvo tenham mais oportunidades de reprodução do que os outros [15]. Para o problema desenvolvido neste trabalho, essas possibilidades de reprodução seriam as melhores proporções de investimento em cada ação, baseadas em algum critério sobre risco e retorno dessa carteira. Note que ao falar de probabilidades, ou pesos, significa que a soma desses pesos tem que totalizar um. O conjunto de cromossomos selecionados pelas melhores possibilidades de investimento (ou pesos, que representam seus genes), é o conjunto que será denotado por “*pais*”, a partir do qual obtém-se uma geração ou nova população.

Para gerar uma nova população, aplicam-se os mecanismos de **cruzamento** e/ou **mutação** nos cromossomos selecionados (“pais”) até satisfazer algum critério de parada, que pode ser um número fixo de iterações ou uma solução (cromossomo) encontrada que satisfaça algum critério previamente selecionado [14]. As noções de cruzamento e mutação serão exploradas mais adiante.

Cada cruzamento envolve dois pais (cromossomos selecionados através da função de aptidão). Em seguida, seus genes (unidades do cromossomo) são combinados para produzir dois novos cromossomos ou “descendentes”. Os *genes*, estão se referindo aos pesos de investimento em cada uma das ações e

os *descendentes* se referem a duas novas possibilidades de investimento na mesma carteira, compondo a nova geração, que poderão ser “pais” de acordo com sua pontuação. É assim que a idéia biológica da evolução se processa.

- ***Função de Aptidão***

A ***função de aptidão*** deve ser definida para cada problema particularmente. Dado um cromossomo específico, a função de aptidão retornará um único número chamado *pontuação*, o qual será proporcional à utilidade ou habilidade do indivíduo que o cromossomo está representando. Essa pontuação é o elo entre o AG e o mundo externo. A avaliação é feita através de uma função que melhor representa o problema e tem por objetivo fornecer uma medida de aptidão de cada indivíduo na população corrente, que irá dirigir o processo de seleção.

A função de aptidão é a função que deve-se otimizar (por exemplo, minimizar o risco ou maximizar o retorno de uma carteira de ações). Geralmente, para problemas da área financeira, é preciso lidar com diversos mínimos/máximos locais, por isso, a função de aptidão deve ser bem escolhida, junto com os outros critérios e restrições do AG, para a obtenção do mínimo/máximo global. O processo de otimização estabelecido pelos AGs, é muito utilizado quando os métodos tradicionais (com o cálculo do gradiente) não se aplicam.

Procura-se uma função de aptidão que seja suave e regular, de modo que os cromossomos com aptidão razoável estejam próximos dos cromossomos com aptidão ligeiramente melhor. Para funções de aptidão que não possuem essas características, é necessário um controle maior dos critérios de cruzamento, mutação e de parada dos AGs.

Um caso onde não seria possível construir uma função de aptidão é quando o “valor real” de um cromossomo não é sempre uma quantidade útil para orientar a pesquisa genética, ou seja, esse seria um cromossomo inválido e, portanto, seu “valor real” seria zero. Um exemplo desse problema é a codificação do AG para horários escolares. Considerando um número finito de salas e professores disponíveis (restrições), tem-se que para um determinado número de salas deve ser dado um número de aulas (codificação). Uma hipótese de atribuição de aula e professor para as salas que viola as restrições seria: ou uma sala sendo ocupada por duas aulas diferentes ao mesmo tempo, ou uma aula ou professor estarem em duas salas ao mesmo tempo, ou ainda, uma sala não sendo programada para todas as aulas que se supõe receber.

Para um AG ser eficaz, é preciso construir uma função de aptidão, cuja a aptidão de um cromossomo inválido é visto em termos de uma baixa pontuação, levando aos cromossomos válidos. É preciso saber onde os cromossomos válidos estão para nos assegurarmos de que aos cromossomos próximos podem ser atribuídos valores de aptidão boa, e aos cromossomos distantes serão dados valores de aptidão menores. Mas, se não se sabe onde os cromossomos válidos estão, isto não pode ser feito.

Cramer sugeriu que se o objetivo natural do problema é tudo ou nada, melhores resultados podem ser obtidos se considerarmos significados sub-objetivos, e premiá-los de acordo com algum critério. No problema dos horários escolares, por exemplo, pode-se dar uma recompensa para cada uma das salas que tem suas aulas atribuídas em um caminho válido.

No caso específico do problema de otimização de carteiras, que será detalhado adiante, a função de aptidão que será usada pode ser a função retorno da carteira (maximizar), ou a função risco da carteira (minimizar). Então, o objetivo da função é encontrar os cromossomos (proporção de investimento em cada ação da carteira) que geram, na compra dos ativos dessa carteira, um melhor resultado (maior retorno ou menor risco).

Depois de calcular a função de aptidão para os cromossomos da população atual, a seleção é realizada. A probabilidade de que os genes da população atual sejam transferidos à próxima população é proporcional à sua aptidão. Após a seleção ter sido realizada, a construção da população intermediária está completa e as recombinações para a próxima geração ocorrem. Isto pode ser visto como a criação da próxima população vinda da população intermediária. Para isso serão utilizados os mecanismos de *cruzamento e mutação*.

- *Cruzamento*

O mecanismo de *Cruzamento* pode selecionar dois indivíduos (“pais”), de acordo com uma taxa ou proporção definidas, e extrair os genes de seus cromossomos numa posição aleatória, produzindo assim dois segmentos da parte inicial e dois segmentos da parte final dos “pais”. Os segmentos são transferidos e permutados, de tal forma que a parte final de um dos “pais” se combine com a parte inicial do outro “pai” para produzir, assim, dois novos cromossomos “descendentes”. Cada um dos dois “descendentes” herdou genes de cada “pai”. Este método é conhecido como *cruzamento de um ponto*. Pode-se entender melhor esse processo olhando a figura 24.

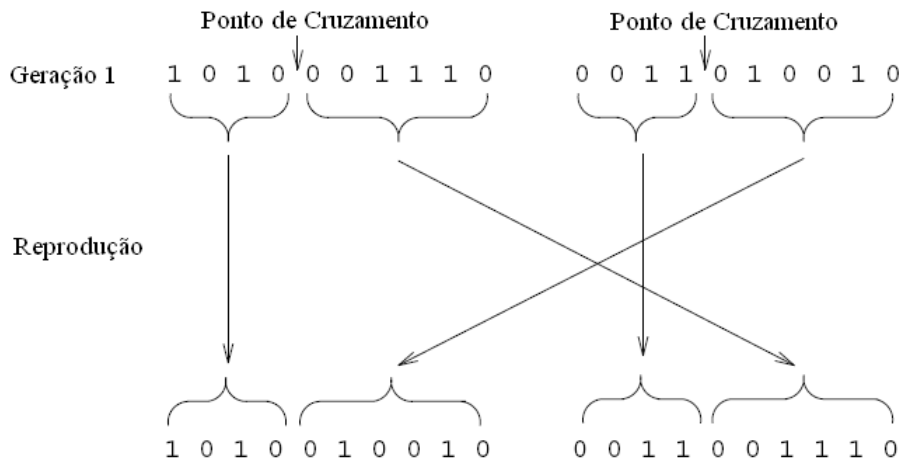


Figura 24: Cruzamento de um ponto

- *Mutação*

O mecanismo de *Mutação* é aplicado a cada “descendente” individualmente depois do cruzamento, com uma taxa ou proporção definidas (geralmente pequena, como 1% ou 5%), alterando aleatoriamente os cromossomos. Assim, de acordo com a taxa escolhida, dois genes dos “descendentes” são aleatoriamente alterados, adicionando-se um valor pequeno para um e subtraindo-se o mesmo valor do outro [15]. A figura 25 mostra um exemplo, onde um gene do cromossomo, após o cruzamento, sofre mutação. Este processo recebe o nome de *mutação simples*.

Geralmente o método do cruzamento é o mais importante dos dois, pois explora rapidamente o seu espaço de busca. A mutação fornece uma pequena amostra de uma pesquisa aleatória e ajuda a assegurar que nenhum ponto do espaço de busca tenha a probabilidade zero de ser examinado e que tenha chance de atingir o mínimo/máximo global.

O motivo pelo qual utiliza-se a mutação é para evitar a perda permanente de qualquer informação importante do cromossomo. Depois de várias gerações, é possível que o processo conduza todos os cromossomos aos mesmos valores para cada gene (a população é composta pelo mesmo indivíduo, isto é, não há mais diversidade das soluções). Se isso acontecer o algoritmo genético pode convergir para um mínimo/máximo local que não é global. Diz-se que o algoritmo convergiu “prematuramente”. Sem um operador de mutação, de acordo com a função de aptidão escolhida, pode-se não ter como introduzir ou reintroduzir informações necessárias para a obtenção do mínimo/máximo global.

É preciso também definir critérios de parada, para que o AG não continue seu processo indefinidamente. No caso do problema proposto neste trabalho, além de um número máximo de gerações permitidas, utiliza-se um critério chamado “*Stall Generation*”. Esse critério consiste em parar o AG quando a melhor avaliação (pontuação) de uma geração se repetir num determinado número de vezes nas próximas gerações. Se esse critério for satisfeito, e essa melhor avaliação não for satisfatória, pode ocorrer o fato de não haver mais informação necessária nos cromossomos para a procura do

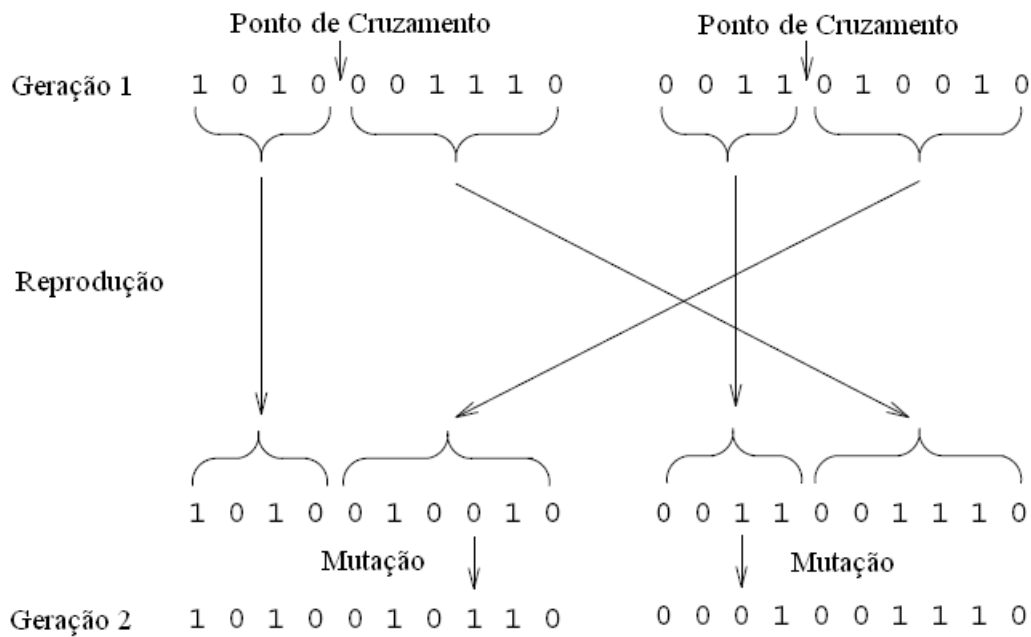


Figura 25: Mutaçãõ simples

mínimo/máximo global. Dessa forma pode-se rever as taxas de cruzamento e mutaçãõ e recommear o processo.

Para Beasley, Bull e Martin [16], se o AG for implementado corretamente, entãõ a populaçãõ evoluirá ao longo de sucessivas gerações para que as aptidões dos melhores indivíduos em cada geraçãõ aumentem sempre, tendendo a uma otimizaçãõ global. Para isso define-se convergência como sendo a progressãõ para o aumento da uniformidade numa populaçãõ em relaçãõ à sua funçãõ de aptidãõ. Diz-se que uma populaçãõ converge quando todos os seus cromossomos têm a mesma pontuaçãõ.

O AG estarã sempre sujeito a erros estocásticos. Mesmo na ausênciã de seleçãõ (por exemplo, se a funçãõ de aptidãõ é constante), os indivíduos da populaçãõ continuarãõ a convergir para algum ponto no espaço de soluçãõ apõs os processos de cruzamento e mutaçãõ. Se, por acaso, um gene torna-se predominante na populaçãõ, entãõ ele tem a mesma probabilidade de se tornar mais predominante na prõxima geraçãõ, da mesma forma para os genes menos predominantes. Se um aumento na predominância é sustentado durante várias gerações sucessivas e a populaçãõ é finita, entãõ um gene pode se espalhar para todos os membros da populaçãõ. Percebe-se a imensa importãncia da definiçãõ da funçãõ de aptidãõ, suas restrições, das taxas de cruzamento e mutaçãõ, e finalmente dos critériõs de parada do AG, para o processo de otimizaçãõ ser realizado com sucesso.

7 Otimização de Carteiras

7.1 Teoria do Portifólio

Falar em otimização de carteiras, é o mesmo que falar na teoria do portfólio [18], que trata essencialmente da composição de uma carteira ótima de ativos, ou seja, a melhor combinação de ativos, tendo por objetivo principal maximizar a utilidade do investidor pela relação risco/retorno. Em outras palavras, escolhe-se uma carteira de ativos que traga o menor risco com o maior retorno.

O retorno esperado de uma carteira de ativos é definido pela média ponderada do retorno de cada ativo em relação a sua participação no total da carteira [19]. Logo, o **retorno esperado** da carteira pode ser obtido pela seguinte expressão:

$$\bar{R}_p = \sum_{j=1}^n W_j * R_j, \quad (56)$$

onde \bar{R}_p é o retorno esperado ponderado da carteira; W_j representa a proporção do capital total investido no ativo j ; n o número total de ativos que compõem a carteira e R_j é o retorno esperado do ativo j .

Por exemplo, admita que uma carteira seja composta por duas ações (X e Y). O retorno esperado da ação X é de 20% e o da ação Y é de 40%. Suponha que 40% da carteira estejam aplicados na ação X e 60% na ação Y . Então, a partir da equação 56 tem-se que o retorno esperado dessa carteira é:

$$\bar{R}_p = \sum_{j=1}^2 W_j * R_j = [W_1 * R_1] + [W_2 * R_2] = [0,40 * 0,20] + [0,60 * 0,40] = 0,32,$$

isto é, ao aplicar 40% na ação X e 60% na ação Y obtém-se um retorno de 32%. É interessante notar que, se toda a carteira estivesse representada pela ação X , o retorno esperado seria de 20%, ou 40% se toda a carteira estivesse representada pela ação Y . Logo, o retorno esperado de toda a carteira depende da proporção investida em cada ativo que a compõe.

Para Assaf [18], a mensuração do risco de um investimento ocorre por meio de um critério probabilístico, ou seja, é delineada uma distribuição de probabilidades dos resultados esperados para que assim possa-se verificar a variabilidade desses resultados em relação a média. Os gráficos da figura 26 dão uma intuição do que esperar.

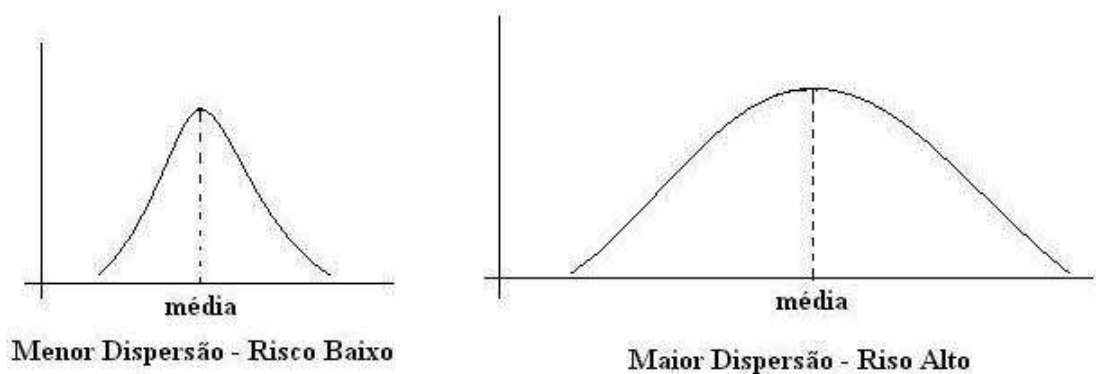


Figura 26: Risco Baixo e Risco Alto

Admita, por exemplo, que dois ativos tenham os seguintes percentuais de retorno e probabilidades de ocorrência:

Os retornos esperados de cada ativo estão calculados a seguir:

$$\bar{R}_A = (7\% * 10\%) + (12\% * 30\%) + (15\% * 40\%) + (20\% * 20\%) = 14,3\%$$

Ativo A	Ativo A	Ativo B	Ativo B
Retorno	Probabilidade	Retorno	Probabilidade
7%	10%	-5%	20%
12%	30%	0%	30%
15%	40%	10%	40%
20%	20%	30%	10%

Tabela 6: Retornos e Probabilidades de ocorrência

$$\bar{R}_B = (-5\% * 20\%) + (0\% * 30\%) + (10\% * 40\%) + (30\% * 10\%) = 6,0\%$$

A figura 27 mostra a representação gráfica da distribuição dos retornos desses dois ativos.

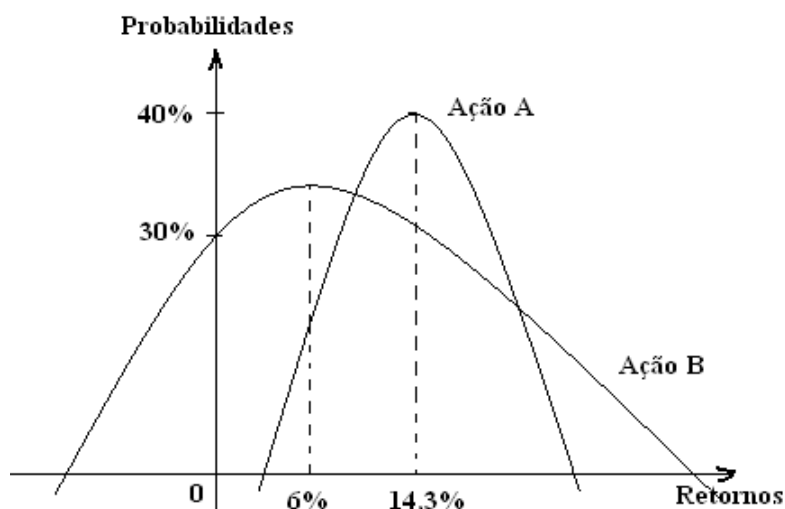


Figura 27: Retornos das ações A e B

Os gráficos da figura 27 revelam que a dispersão a média do ativo B é superior à do ativo A, indicando maior risco. Isso mostra que, ao obter um maior retorno, tem-se um menor risco, ou seja, nesse caso o ativo A é preferível ao B. Pode-se verificar esse fato através do cálculo dos desvios padrões de cada ativo, pois assim tem-se um resultado numérico que mostra a variabilidade dos retornos em relação à média.

O estudo apresentado acima foi feito para carteiras compostas de um ativo. Agora será apresentado como avaliar os riscos de carteiras compostas com mais de um ativo. De acordo com Assaf [18], a orientação é de selecionar ativos diversificados ao montar uma carteira, ou seja, diversificar o setor/indústria dos ativos, que levará na redução do risco dos investimentos, produzindo um retorno mais aceitável.

Em 1952, **Harry Markowitz** escreveu sobre a **Seleção do Portfólio**, que de acordo com Brealey e Myers [20], dizia que deve-se escolher ativos diversificados para compor uma carteira, mostrando para os investidores que assim o desvio-padrão dos retornos poderia diminuir. E mais, o risco pode ser eliminado se essas ações se comportarem de maneira contrária, ou seja, seus coeficientes de correlação forem de sinais opostos e muito próximos de um. Markowitz trabalhou para desenvolver os princípios fundamentais da construção dos portfólios, que foram a base para todo o estudo sobre a relação entre retorno e risco que viriam a seguir.

A existência de aplicações negativamente correlacionadas indica a existência de carteiras com investimentos que produzem retornos inversamente proporcionais, isto é, quando um retorno decrescer o outro se elevará na mesma intensidade, anulando os reflexos negativos produzidos dentro da carteira. Esse comportamento mostra a eliminação total do risco da carteira, pois os resultados desfavoráveis

Ano	Ativo X	Ativo Y	Cálculo do retorno da carteira	Retorno esperado da carteira (k_p)
2004	8%	16%	$(0,50 * 8\%) + (0,50 * 16\%)$	12%
2005	10%	14%	$(0,50 * 10\%) + (0,50 * 14\%)$	12%
2006	12%	12%	$(0,50 * 12\%) + (0,50 * 12\%)$	12%
2007	14%	10%	$(0,50 * 14\%) + (0,50 * 10\%)$	12%
2008	16%	8%	$(0,50 * 16\%) + (0,50 * 8\%)$	12%

Tabela 7: Retorno Esperado da Carteira

serão compensados pelo desempenho positivo dos outros [21].

A expressão para calcular o risco (desvio-padrão) de uma carteira contendo n ativos é:

$$\sigma_p = \sqrt{\sum_{i=1}^n \sum_{j=1}^n W_i W_j \rho_{i,j} \sigma_i \sigma_j},$$

onde $W_{i,j}$ representa a proporção do capital total investido no ativo i, j ; $\rho_{i,j}$ representa o coeficiente de correlação dos ativos i, j e $\sigma_{i,j}$ representa o desvio-padrão de cada ativo i, j .

Será apresentado agora um exemplo tirado do livro do Gitman [21] onde existem dois ativos inversamente correlacionados. É importante lembrar que o livro do Gitman [21] foi publicado em 2003, por isso os retornos esperados são dos anos de 2004 à 2008.

Considerando os dois ativos X e Y , e analisase o retorno esperado da carteira formada por eles, supondo que esta contenha 50% de cada ativo:

Agora calcular-se o valor esperado dos retornos da carteira de 2004 à 2008:

$$\bar{k}_p = \frac{12\% + 12\% + 12\% + 12\% + 12\%}{5} = \frac{60\%}{5} = 12\%$$

Em seguida, calcula-se o desvio padrão (risco) dos retornos esperados da carteira:

$$\sigma_{k_p} = \sqrt{\frac{(12\% - 12\%)^2 + (12\% - 12\%)^2 + (12\% - 12\%)^2 + (12\% - 12\%)^2 + (12\% - 12\%)^2}{5 - 1}} = \frac{0\%}{4} = 0$$

Percebe-se através dos cálculos feitos acima que o retorno esperado é constante, o que resulta num desvio padrão igual a zero. Isso significa que o risco é nulo, porém é importante destacar que essa correlação negativa perfeita é praticamente impossível de acontecer na vida real.

Um aspecto relevante da teoria de portfólio é que o risco de um ativo mantido fora de uma carteira é diferente de seu risco quando incluído na carteira. No estudo da diversificação, o risco de um ativo é avaliado pela sua contribuição ao risco total da carteira. Elevando-se, de maneira diversificada, o número de títulos de uma carteira, é possível promover-se a redução de seu risco, porém a uma taxa decrescente. É importante ressaltar, que a partir de um determinado número de títulos, a redução do risco praticamente deixa de existir.

7.2 Modelo de Markowitz

O trabalho pioneiro de Markowitz, de acordo com Gilli e Kellezi [22], na área de otimização de carteiras foi a proposição do modelo **média-variância (MV)** em 1952, uma ferramenta quantitativa que permite fazer a alocação de seus investimentos entre ativos diferentes, considerando o “trade-off” entre o risco, medido pela variância dos retornos futuros de ativos, e o retorno. Esse “trade off” consiste em maximizar o retorno dado um nível de risco previamente especificado, ou minimizar o risco dado um nível de retorno. Em outras palavras, define-se o nível de risco do portfólio e encontra-se a composição que forneça o retorno máximo, ou define-se o retorno e determina-se a composição que corresponda ao risco mínimo. Para isso, é traçado uma curva com a relação risco \times retorno, denominada *fronteira eficiente*.

Muitas pesquisas foram feitas na área de otimização de portfólios mas não foi encontrado modelo de otimização de carteira que invalidasse os conceitos postulados por Markowitz (1952). Devido a importância deste trabalho foi concedido ao autor o prêmio Nobel em Economia no ano de 1990.

A proposição de Markowitz permitiu que os agentes do mercado, pela primeira vez, utilizassem os conceitos de risco e retorno de forma conjunta na avaliação de investimentos. Apesar da grande aceitação e disseminação o modelo de Markowitz tem sofrido algumas críticas. Dentre elas está a utilização da variância como medida de risco. Apesar de ser a medida de risco mais popular, ela não é a única, podendo ser substituída por outra medida, tal como *Value at Risk* (VaR).

A teoria da escolha do Portfólio explica como investidores racionais irão usar o princípio da diversificação para otimizar as suas carteiras de investimentos. O conceito de diversificação de Markowitz permite que, ao se selecionar ativos com correlação perfeitamente negativa, seja eliminado todo o risco da carteira. É importante acrescentar que a diversificação quando utilizada com o propósito de redução do risco, não é uma decisão aleatória, ela deve ser elaborada observando-se as correlações dos retornos dos ativos, de maneira a se estabelecer a melhor composição possível de uma carteira.

Para estudar o modelo de **média-variância (MV)** algumas considerações iniciais [23] serão feitas.

Seja Y a variável cujo valor é decidido ao acaso. Supõe-se que Y tenha um número finito de valores y_1, y_2, \dots, y_N e seja p_1 a probabilidade de $Y = y_1$; p_2 a probabilidade de $Y = y_2$, e assim sucessivamente. O valor esperado de Y é definido por:

$$E[Y] = p_1 y_1 + p_2 y_2 + \dots + p_N y_N.$$

A variância de Y é definida por:

$$V[Y] = p_1 (y_1 - E[Y])^2 + p_2 (y_2 - E[Y])^2 + \dots + p_N (y_N - E[Y])^2.$$

Supondo as seguintes variáveis aleatórias: R_1, R_2, \dots, R_n , onde R é uma combinação linear dos R_i , então:

$$R = a_1 R_1 + a_2 R_2 + \dots + a_n R_n,$$

também será uma variável aleatória.

É muito importante para este trabalho ver como a esperança e a variância da combinação linear R estão se relacionando com a distribuição de probabilidades de R_1, R_2, \dots, R_n . Logo abaixo mostram-se essas relações, porém as demonstrações poderão ser encontradas em bibliografias complementares.

O valor esperado da combinação linear R nada mais é do que a combinação linear dos valores esperados, isto é:

$$E[R] = a_1 E[R_1] + a_2 E[R_2] + \dots + a_n E[R_n].$$

A variância não é tão simples de se calcular. Para isso será preciso definir covariância. Então:

$$\sigma_{ij} = E[(R_i - E[R_i])(R_j - E[R_j])].$$

E mais, a covariância entre R_i e R_j pode ser escrita em função da correlação, ou seja:

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j.$$

Enfim, a variância da combinação linear R será dada por:

$$V[R] = \sum_{i=1}^N a_i^2 V[W_i] + 2 \sum_{i=1}^N \sum_{j>1}^N a_i a_j \sigma_{ij},$$

onde W_i seria a porcentagem investida numa ação i .

Substituindo $V[R_i] = \sigma_{ii}$ tem-se:

$$V[R] = \sum_{i=1}^N \sum_{j=1}^N a_i a_j \sigma_{ij}.$$

Em termos financeiros, toma-se R_i como sendo o retorno de uma ação. Seja μ_i o valor esperado do retorno R_i , σ_{ij} a covariância entre os retornos R_i e R_j e seja W_i a porcentagem investida na ação i . Então o rendimento da carteira de ações é dado por:

$$R = \sum R_i W_i.$$

E conseqüentemente, o valor esperado do rendimento dessa carteira e sua variância são:

$$E[R] = \sum_{i=1}^N W_i \mu_i,$$

$$V[R] = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} W_i W_j.$$

Para probabilidades fixas, acredita-se que o investidor tem várias combinações de E (média) e V (variância), dependendo da sua escolha da carteira X_1, \dots, X_N . Supondo que todas as possíveis combinações de (E, V) estejam presentes na figura 28, o modelo MV diz que o investidor deveria escolher a carteira que tenha um V mínimo para um dado E e um máximo E para um dado V .

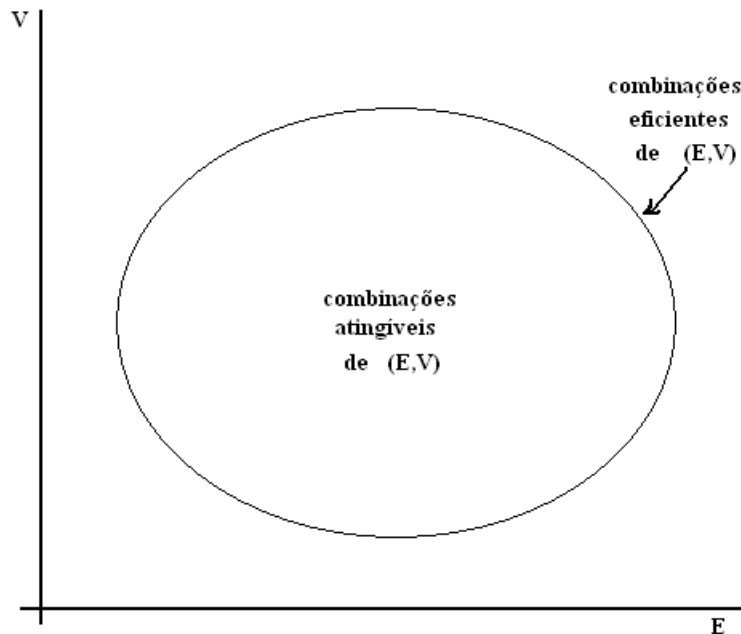


Figura 28: Possíveis combinações de (E, V)

As técnicas para se calcular as carteiras eficientes e eficientes combinações de (E, V) associadas a um dado μ_i e σ_{ij} serão vistos mais adiante ao estudar a fronteira eficiente.

7.2.1 Modelo de Valor em Risco (VaR)

A principal inovação proposta por Markowitz foi a interpretação do retorno de uma carteira como uma variável aleatória cujo risco é avaliado por meio de sua variância, como foi visto anteriormente. A partir dessa abordagem, foi possível compreender que a diversificação reduz o risco na medida em que carteiras que contêm ativos negativamente correlacionados possuem variância (risco) inferior à soma das variâncias individuais. Porém essa abordagem tinha um problema, pois a variância do retorno de uma carteira não é inteiramente adequada para avaliar situações de perdas extremas. E para tratar a variância de forma simplificada, o modelo pressupõe que a distribuição de probabilidades dos retornos seja elíptica, o que na prática não acontece [24].

De acordo com Ribeiro e Ferreira [24], a prática da gestão de risco mostrou que considerar média e variância do retorno de uma carteira como medida de risco era insatisfatório. Foi então que surgiu o **Valor em Risco (VaR)**.

A que perda máxima em um único dia o investidor está sujeito quando aplica em um fundo de investimento? Essa é uma questão cuja resposta todo o investidor gostaria de conhecer antes de aplicar seus recursos em uma carteira que tenha certo risco. Uma possível resposta seria basear-se nos resultados diários obtidos pela carteira durante certo período e determinar se existirá uma perda máxima diária que poderá ocorrer em seu investimento, o VAR. Por exemplo, se for informado que o VaR desse investimento é de 8%, isso significaria que a perda máxima diária à qual o investidor está exposto é de 8%.

Porém, o VaR apenas não diz muito, é preciso saber outros detalhes do cálculo, para ter uma informação mais completa do risco. Por exemplo, saber qual a probabilidade de que essa perda máxima venha a ocorrer, pois é diferente saber que essa probabilidade é de 5% ou 50%. A essa probabilidade corresponde um nível de significância do VaR, o qual é de 99% quando a probabilidade de perder um valor superior ao VaR é de 1% ou 95% quando esta probabilidade de risco é de 5%, etc. Portanto, a informação do VAR deve ser acompanhada dessa probabilidade, para que ela seja mais efetiva e ajude o investidor a tomar sua decisão conscientemente.

Voltando ao exemplo inicial onde o VaR é de 8%, considere que o nível de significância seja 95%, dessa forma o investidor saberá que poderá perder no máximo 8% e a chance de que ocorra uma perda superior a essa é de 5%. Em outras palavras, quanto menor a significância maior o risco de o investimento apresentar variação negativa superior à prevista.

Conceitualmente, a obtenção do VaR é simples, porém, na prática, sua determinação depende de técnicas estatísticas de estimação o que introduz incertezas em seu cálculo. Há diversos métodos para estimação do VaR. Um deles é chamado *paramétrico*, desenvolvido por Philippe Jorion [25], o qual parte da premissa que a distribuição de probabilidade do retorno da carteira é normal, bem determinado a partir de sua média e variância, o que possibilita obter uma expressão analítica simples. Um outro método é a *Simulação de Monte Carlo*, o qual pressupõe que o histórico passado dos retornos reflete, de maneira adequada, o que deverá ocorrer no futuro. E por fim, o método da *série histórica*, onde para calcular o VaR é preciso selecionar uma amostra de retornos para a carteira, ordenar estes valores e escolher um valor de acordo com a função M (um estimador de percentil) em que α é o nível de confiança desejado [24].

A definição para Valor de Risco (VaR) de acordo com Jorion [25], é que VaR é uma medida da pior perda esperada ao longo de um determinado intervalo de tempo, em condições normais de mercado dado um determinado nível de confiança. Por exemplo, um banco pode dizer que o valor VaR diário de uma carteira de negociação é de R\$1 milhão em um nível de confiança de 99%. Em outras palavras, sob condições normais de mercado, somente um por cento das vezes a perda diária excederá R\$1 milhão.

Se o retorno de uma carteira segue uma distribuição normal com média μ e desvio padrão σ , então o valor do VaR em um nível de confiança de 99% pode ser obtido da seguinte forma:

1. Encontrar o valor de α correspondente a um nível de confiança de 99% na tabela normal;
2. Calcular o valor VaR da seguinte forma: $VaR = \alpha\sigma - \mu$. Se for preciso normalizar (ou seja, média $\mu = 0$ e desvio padrão $\sigma = 1$) basta encontrar $z = (x - \mu)/\sigma$, onde z é o novo valor da tabela normal para o dado nível de confiança.

Note que como VaR é o retorno da carteira no pior dos casos, ele deveria corresponder a menor cauda da distribuição. No entanto, são sempre usados números positivos para representar perdas, logo, ambos os valores α e VaR serão positivos.

Suponha que o valor inicial da carteira é P_0 e a taxa de retorno R é normalmente distribuída com média μ e desvio padrão σ , então o valor VaR em um nível de confiança c , onde α é o número correspondente a c na tabela normal, será:

$$VaR = P_0(\alpha\sigma - \mu).$$

Se o tempo alvo é curto, será assumido que o valor médio de R é zero, então a fórmula pode ser simplificada por:

$$VaR = P_0 \alpha \sigma$$

Se uma carteira possui N ações, o retorno dessa carteira é uma combinação linear dos retornos de cada ação.

Define-se o peso $w_i = P_i/P_0$, com P_0 sendo o valor do retorno da carteira e P_i o valor do retorno da ação i . Seja w o vetor coluna dos pesos, e R o vetor coluna da taxa de retorno, então a taxa de retorno dessa carteira será:

$$R_p = w^T R,$$

onde w^T é a matriz transposta da matriz w dos pesos.

A variância para R_p será:

$$\sigma_p^2 = w^T \Sigma w,$$

onde Σ é a matriz covariância.

O valor VaR será então:

$$VaR_p = \alpha \sigma_p P_0 = \alpha \sqrt{x^T \Sigma x}$$

De acordo com Pearson [26], os revendedores de derivativos (bancos comerciais e de intestimento) que desenvolveram o VaR na década de 1980, enfrentaram o problema de que sua carteira de derivativos e sua "carteira" de negociação tinham crescido ao ponto que os riscos inerentes ao mercado eram de preocupação significativa. E daí veio a pergunta: como esses riscos podem ser medidos, descritos e comunicados ao conselho de administração e a diretoria? As posições eram tão numerosas que não poderiam ser facilmente enumeradas e descritas. Mesmo que isso pudesse ser feito, seria útil se o conselho administrativo e a diretoria pudessem entender todas as posições e os instrumentos, e os riscos de cada um. Naturalmente, os riscos poderiam ser medidos por sensibilidades da carteira, ou seja, quanto o valor da carteira muda quando várias alterações subjacentes às taxas de mercado ou dos preços mudam. Mesmo que esses conceitos possam ser bem explicados, a exposição a diferentes tipos de risco de mercado não pode ser agregada de forma significativa sem um quadro estatístico. Então o VaR ofereceu uma maneira de fazer isso e, portanto, ajudou a superar os problemas na medição e transmissão de informações de risco.

Uma outra definição para o valor de risco (VaR) seria: uma simples medida estatística das possíveis perdas da carteira devido ao risco de mercado, que o torna útil para medir e comparar os riscos de mercado de carteiras diferentes, e para comparar o risco da carteira mesmo em momentos diferentes. A ideia é que, prejuízos maiores do que o VaR, sofrem apenas com uma pequena probabilidade. Associados a cada medida de VaR estão: uma probabilidade α , ou um nível de confiança $1 - \alpha$, e um período, ou horizonte de tempo h . O nível de confiança $1 - \alpha$ é simplesmente a perda que será excedida com uma probabilidade de apenas α por cento ao longo do período de comprimento h ; equivalentemente, a perda será menor do que o VaR com probabilidade $1 - \alpha$ [26].

Para melhor entendimento do conceito de VaR, considera-se o exemplo de uma carteira exposta a alterações nos mercados de ações dos E.U.A. e do Reino Unido. O portfólio é composto por \$110 milhões de dólares investidos em uma carteira bem diversificada de ações de grande capitalização, juntamente com os índices de contratos futuros dos E.U.A. (S&P 500) e do Reino Unido (FT-SE 100). A carteira de títulos E.U.A. é bem diversificada, e os seus retornos são altamente correlacionados com os retornos dos índices do S&P 500. Para simplificar, supõe-se que os retornos da carteira estão perfeitamente correlacionados com as mudanças no índice do S&P 500, ou seja, à medida que o índice S&P 500 aumenta a carteira aumenta na mesma proporção e vice-versa.

Foi estimado que o desvio padrão, das taxas mensais do retorno da carteira, está abaixo do índice S&P 500 que é $\sigma_1 = 0,061(6,1\%)$, o desvio padrão das taxas mensais de retorno da carteira está

abaixo do índice FT-SE 100 que é $\sigma_2 = 0,065(6,5\%)$, e a correlação entre as taxas mensais de retorno estima-se que seja $\rho = 0,55$. O retorno esperado das mudanças nos índices S&P 500 e FT-SE 100 são estimadas em $\mu_1 = 0,01(1\%)$ e $\mu_2 = 0,0125(1,25\%)$ por mês, respectivamente. E ainda, a carteira das ações americana paga os dividendos a uma taxa de $1,4\%$ ao ano, ou $1,4/12 = 0,1167\%$ ao mês.

Para calcular o VaR precisa-se escolher um período h e o nível de confiança $(1 - \alpha)$. Supondo que o período seja 1 mês e o nível de confiança $(1 - \alpha) = 95\%$ ou $\alpha = 5\%$. A partir disso e das informações dadas acima, fica fácil de calcular o VaR se for assumido que os retornos dos índices S&P 500 e FT-SE 100 são normalmente distribuídos. Assim o retorno da carteira também será normalmente distribuído e o retorno esperado e a variância da carteira podem ser calculados usando resultados matemáticos. E assim, como a distribuição normal é determinada através do valor esperado e da variância, poderá saber a distribuição do lucro e da perda durante um mês.

Por exemplo, supondo que a distribuição dos possíveis lucros e perdas da carteira possam ser aproximados pela função densidade de probabilidade como mostra a figura 29. A distribuição descrita por essa fdp tem uma média de \$1,2759 milhões de dólares e um desvio padrão de \$5,6845 milhões de dólares. Uma propriedade da distribuição normal é que o valor crítico, ou corte, é igual a 1,645 desvios padrão abaixo da média, o que deixa 5% da probabilidade no lado esquerdo da calda. Se esse corte for denotado por 5% *quantil* da distribuição de lucro ou perda, como segue:

$$5\% \text{quantil} = \text{média da carteira} - [1,645 * (\text{desvio padrão da carteira})] = 1,2759 - (1,645 * 5,6845) = -8,0752.$$

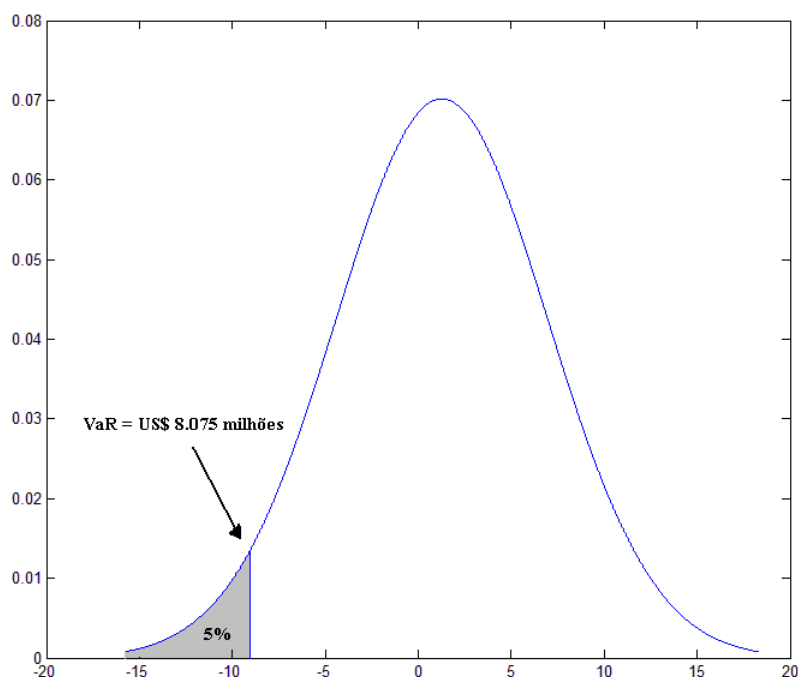


Figura 29: Valor de Risco (VaR)

Ou seja, o lucro diário será menor que $-\$8,0752$ milhões de dólares com uma probabilidade de 5%. Então, como o VaR 5% é definido como perda que excederá com probabilidade de 5%, o VaR é o negativo desse quantil, ou $\$8,0752$ milhões de dólares. Esse VaR também pode ser visto na figura 29.

Quando existem duas posições, o valor esperado da carteira, incluindo os dividendos, será:

$$E[\Delta V] = W_1\mu_1 + W_2\mu_2 + D.$$

Onde, ΔV é a mudança do valor da carteira, W_1 e W_2 são as quantidades em dólares investidas nas duas posições, e $D = \$110(0,014/12)$ milhões é o dividendo a ser recebido durante o próximo mês. Usando o fato de que a carteira é equivalente à posição de $\$54.457$ milhões investidos na carteira do

índice S&P 500 e \$48.319 milhões investidos na carteira do índice FT-SE 100, tem-se $W_1 = 54.357$ milhões e $W_2 = 48.319$ milhões. A variância das mudanças mensais da carteira depende dos desvios-padrão das mudanças no valor padronizado das posições, da correlação, e do tamanho da posição, e isso é dado pela fórmula:

$$Var[\Delta V] = W_1^2 \sigma_1^2 + W_2^2 \sigma_2^2 + 2W_1 W_2 \rho_{12} \sigma_1 \sigma_2.$$

A partir dos dados do exemplo e usando as fórmulas, o valor esperado e a variância das mudanças na carteira são:

$$E[\Delta V] = 54.357(0,01) + 48.319(0,0125) + 110(0,014/12) = 1,2759.$$

$$Var[\Delta V] = (54.357)^2(0,061)^2 + (48.319)^2(0,065)^2 + 2(54.357)(48.319)(0,061)(0,065)(0,55) = 32.3138.$$

Uma alternativa é $V = 110$ milhões de dólares denotar o valor da carteira e $r = \Delta V/V$ o retorno da carteira, assim o valor esperado e a variância do retorno da carteira serão:

$$E[r] = \frac{54.357}{110}(0,01) + \frac{48.319}{110}(0,0125) + \frac{0,014}{12} = 0,01160.$$

$$Var[r] = \left(\frac{54.357}{110}\right)^2 (0,061)^2 + \left(\frac{48.319}{110}\right)^2 (0,065)^2 + 2\left(\frac{54.357}{110}\right)\left(\frac{48.319}{110}\right)(0,061)(0,065)(0,55) = 0,002670.$$

O desvio-padrão é simplesmente a raiz quadrada da variância, ou seja, $dv[\Delta V] = \$5.6845$ milhões e $dv[r] = 0.0517(5.17\%)$.

Usando o fato de que os resultados menores ou iguais a 1.645 desvios-padrão abaixo da média ocorrem apenas 5% das vezes, pode-se calcular o VaR:

$$VaR = -(E[\Delta V] - 1,645 * dv[\Delta V]) = -(1.2759 - 1,645 * 5.6845) = 8.0752.$$

Como fração do valor inicial da carteira:

$$VaR = -(E[r] - 1,645 * dv[r]) = -(0,01160 - 1,645 * 0,05168) = 0,0734.$$

Ou ainda, 7,34% do valor inicial da carteira.

No cálculo da estimativa do VaR, às vezes, é assumido que a variação esperada no valor da carteira é zero. Se esta hipótese for feita, o VaR é de $1,645(\$5.6845) = \9.351 milhões, ou $1,645(0,05168) = 0,0850$, ou 8,5%. A suposição de que a mudança esperada no valor da carteira seja zero é comum quando o horizonte temporal da estimativa VaR é de um dia.

Na interpretação da estimativa do VaR, é fundamental ter em mente o período e o nível de confiança $1 - \alpha$, as estimativas diferentes serão obtidas se forem feitas diferentes escolhas desses parâmetros. Por exemplo, para calcular o VaR utilizando um nível de confiança de 99%, pode-se usar o fato de que, para a distribuição normal, os resultados iguais ou inferiores a 2,326 desvios padrão abaixo da média ocorrem apenas 1% das vezes. Assim, com um período mensal, a estimativa do VaR com nível de confiança 99% é:

$$VaR = -\left(E\left[\frac{\Delta V}{V}\right]\right) - 2,326 * dv\left(\left[\frac{\Delta V}{V}\right]\right) = -(0,01160 - 2,326 * 0,05168) = 0,1086,$$

ou 10,86% do valor inicial. A escolha do período pode ter um impacto ainda maior, para o VaR calculado, utilizando esta abordagem, é aproximadamente proporcional à raiz quadrada da duração do período, porque as variações de retorno são aproximadamente proporcionais à duração do período.

Carteira	W_X	W_Y	Retorno Esperado(\overline{R}_P)	Risco(σ_P)
A	0%	100%	12,0%	22,0%
B	20%	80%	13,6%	20,3%
C	40%	60%	15,2%	21,4%
D	60%	40%	16,8%	24,9%
E	80%	20%	18,4%	29,9%
F	100%	0%	20,0%	36,0%

Tabela 8: Retorno e Risco Esperado da Carteira

7.3 Fronteira Eficiente

Supondo dois ativos, X e Y , onde os retornos esperados de X e Y são 20% e 12% e os riscos são 36% e 22%, respectivamente. Serão feitas algumas combinações desses dois ativos para montar algumas possíveis soluções de carteiras, onde para cada uma serão calculados seu retorno esperado e seu risco, como pode ser visto na tabela 8.

Plotando os resultados obtidos de Retorno Esperado e Risco, obtem-se o gráfico da figura 30.

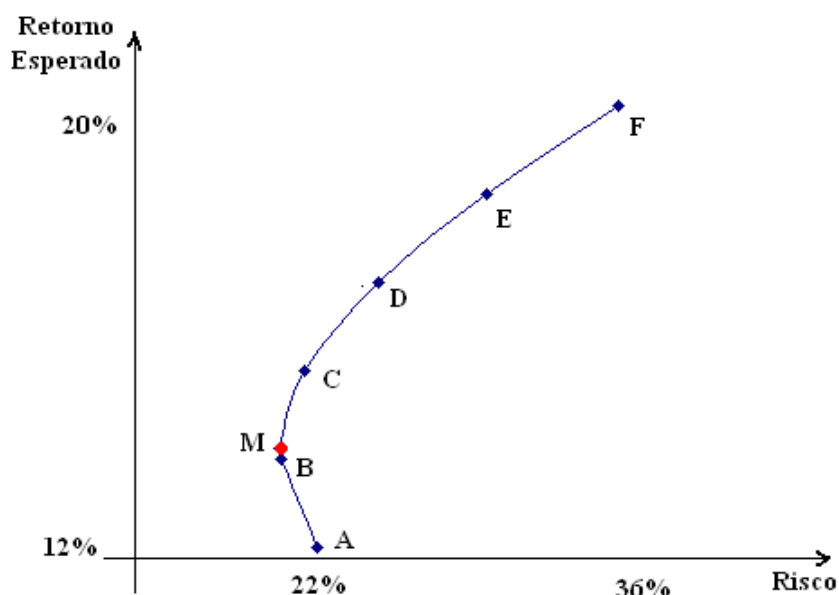


Figura 30: Retorno Esperado e Risco de carteiras com diferentes composições

Nota-se na figura 30 a presença do ponto M , que representa a carteira de ativos com o menor risco possível, denominado *variância mínima*. Quanto mais uma carteira se distancia desse M (desvio-padrão do retorno), maior é o risco que esta carteira apresentará, e conseqüentemente terá um retorno maior.

Para dois ativos (X e Y), a carteira de variância mínima (ponto M) pode ser determinada a partir da seguinte expressão:

$$W_X^* = \frac{\sigma_Y^2 - (\rho_{X,Y} * \sigma_X * \sigma_Y)}{(\sigma_X^2 + \sigma_Y^2) - (2\rho_{X,Y} * \sigma_X * \sigma_Y)},$$

onde $W_Y^* = 100\% - W_X^*$.

Depois de encontrar a variância mínima de X e Y , deve-se encontrar seus retornos esperados e riscos. Usando os dados do exemplo acima, tem-se:

$$W_X^* = \frac{0,22^2 - (0,20 * 0,36 * 0,22)}{(0,36^2 + 0,22^2) - (2 * 0,20 * 0,36 * 0,22)} = 0,2225(22,25\%).$$

$$W_Y^* = 100\% - 22,25\% = 77,75\%.$$

Isto significa que, a participação do ativo X (W_X^*) na carteira M é de 22,25% e a participação do ativo Y (W_Y^*) na carteira M é de 77,75%. Basta calcular o retorno esperado e o risco dessa carteira M , ou seja:

$$\bar{R}_p = (0,20 * 0,2225) + (0,12 * 0,7775) = 0,1378 = 13,78\%.$$

$$\sigma_p = \sqrt{(0,36^2 * 0,2225^2) + (0,22^2 * 0,7775^2) + (2 * 0,2225 * 0,7775 * 0,20 * 0,36 * 0,22)} = 0,2020 = 20,2\%.$$

Esses resultados podem ser encontrados no gráfico da figura 30.

A *Fronteira Eficiente* é o segmento determinado pelas carteiras A a F do gráfico da figura 30, que representa a seleção de carteiras de investimento mais atraentes para um investidor racional (aquele que é averso ao risco, ou seja, prefere investir numa carteira cujo risco seja menor e conseqüentemente seu retorno também), incluindo todas as carteiras possíveis de serem construídas.

Em outras palavras, na fronteira eficiente é possível selecionar uma carteira que apresenta o menor risco para um determinado retorno [18], basta saber que tipo de investimento se quer fazer, ou seja, se é desejado um retorno maior, não importando que trará um risco alto ou se é desejado um risco menor, o que acarretará num retorno menor. Pode-se dizer que esta curva representa o conjunto das melhores opções de carteira de investimento [27].

8 APLICAÇÕES

8.1 FAC e FACP

Para efeito de aprimoramento da definição da FAC e da FACP foram feitos alguns testes com algumas séries de dados extraídos do sítio da BOVESPA. A partir dos dados do IBrX e com auxílio das funções FAC (autocorr) e FACP (parcorr) elaboradas no matlab (*Apêndice E*), os gráficos das séries foram avaliados.

Primeiramente foram plotados os gráficos de cada série, em seguida os gráficos dos retornos e para finalizar os gráficos da FAC e FACP das séries para analisar seu comportamento.

A figura 31 mostra os gráficos da série e do retorno da Petrobrás e a figura 32 mostra os gráficos da FAC e FACP.

A partir dos gráficos da FAC e FACP da série do retorno da Petrobrás percebe-se que os dados estão convergindo para zero, o que diz que não existe correlação entre os termos dessa série, ou seja, a série dos retornos é estacionária. Em contrapartida, a FAC do retorno ao quadrado explodiu, de acordo com o gráfico da figura 33, isso significa que a série do retorno ao quadrados da Petrobrás apresenta heterocedasticidade, ou seja, não é estacionária.

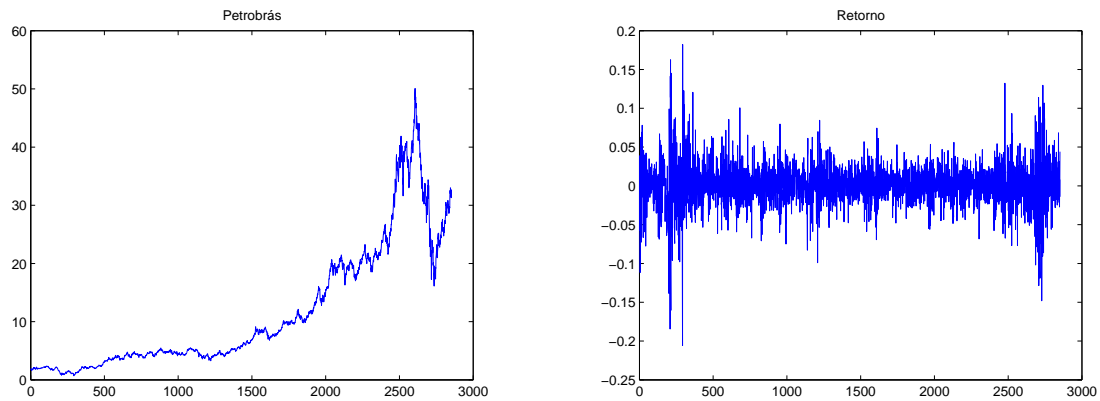


Figura 31: Série e Retorno da Pétrobras

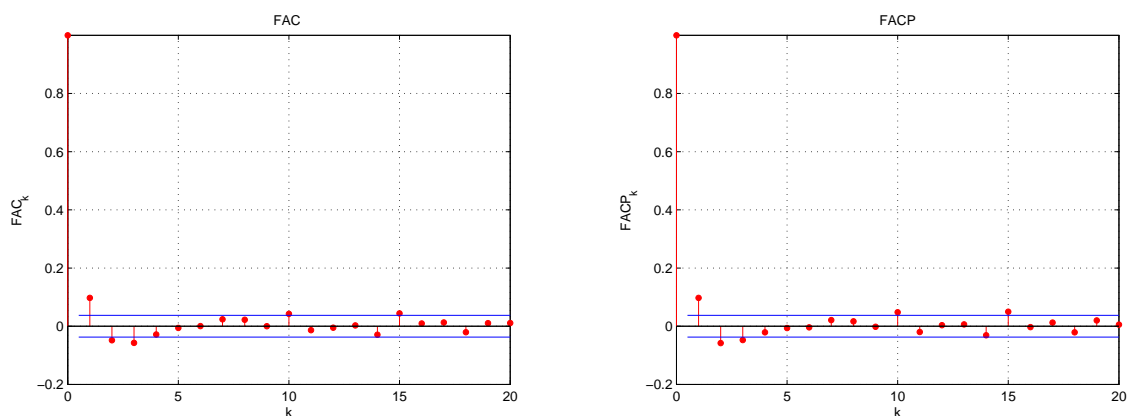


Figura 32: FAC e FACP do Retorno da Pétrobras

Esse mesmo comportamento pode ser encontrado nas séries da Vivo e da Comgás. Basta verificarmos os gráficos das figuras 35, 36, 38, 39.

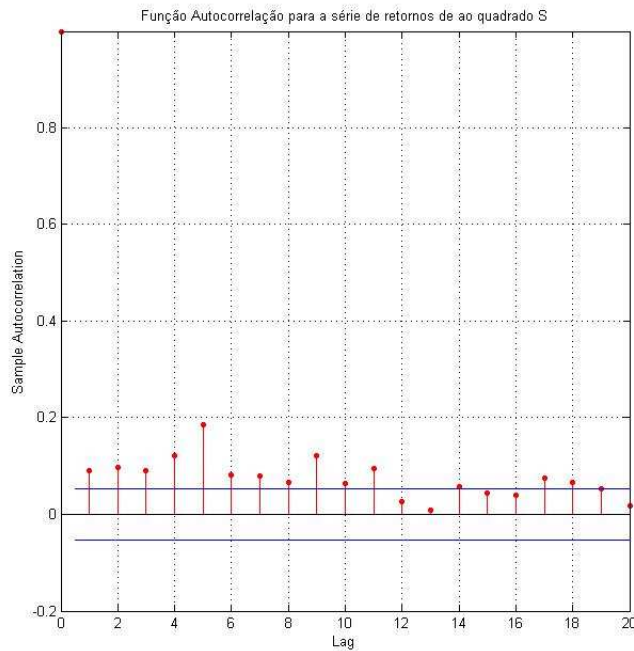


Figura 33: FAC do Retorno ao quadrado da Pétrobras

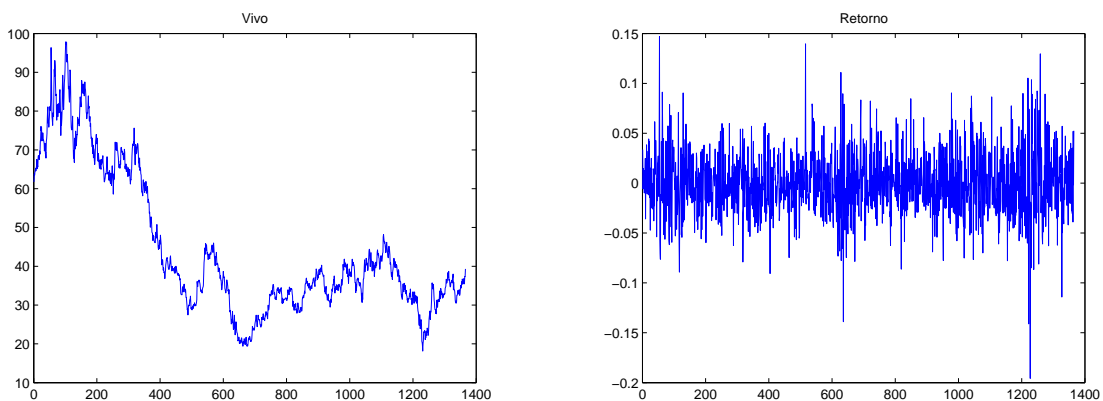


Figura 34: Série e Retorno da Vivo

8.2 Agregação Temporal

Nesta fase do trabalho será feita uma análise de séries financeiras, usando modelos de baixa frequência e de alta frequência. Para tal estudo, se tomará como base o artigo de Drust e Nijman [28] que se encontra na bibliografia.

Quando se fala em modelos de baixa frequência, refere-se, por exemplo, a um estudo semanal de um fenômeno diário. Se o modelo de alta frequência (diário) tiver heterocedasticidade condicional, observa-se que o modelo de baixa frequência (semanal) também possui essa característica e os parâmetros em sua equação de variância condicional dependem da média, variância e da curtose do modelo de alta frequência.

A proposta dos autores deste artigo é demonstrar os resultados da agregação temporal (ou diminuição de frequência) observados na prática, sobre um número finito de períodos, e observar como se comporta a heterocedasticidade condicional se o intervalo de tempo da amostra vai para infinito. Demonstra-se também que todo modelo GARCH se agrega para um modelo GARCH, isto é, os modelos GARCH

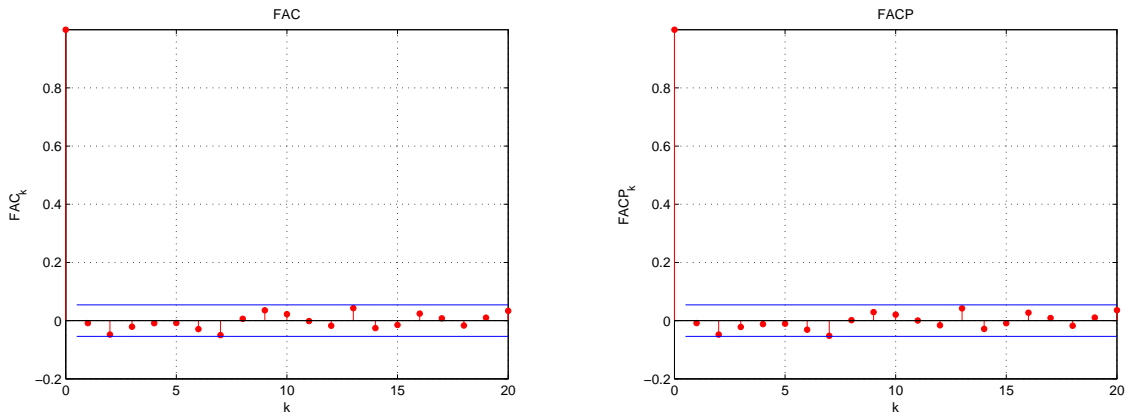


Figura 35: FAC e FACP do Retorno da Vivo

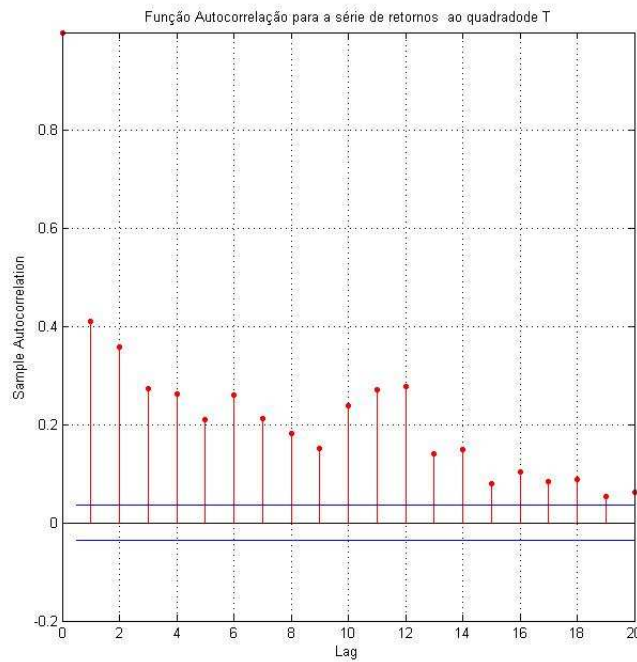


Figura 36: FAC do Retorno ao quadrado da Vivo

são fechados pela agregação temporal.

Definições e Notações:

Serão estabelecidos algumas notações e definições novas, para que se possa utilizar o exemplo do artigo de Drost e Nijman [28].

Primeiramente, seja $\{\epsilon_t, t \in Z\}$, a sequência de erros aleatórios com o quarto momento finito. Defini-se os seguintes operadores:

$$A(L) = 1 + \sum_{i=1}^q \alpha_i L^i \quad e \quad B(L) = 1 - \sum_{i=1}^p \beta_i L^i. \quad (57)$$

Tem-se que a sequência $\{h_t, t \in Z\}$ será a solução estacionária de:

$$B(L)h_t = \Psi + (A(L) - 1)\epsilon_t^2,$$

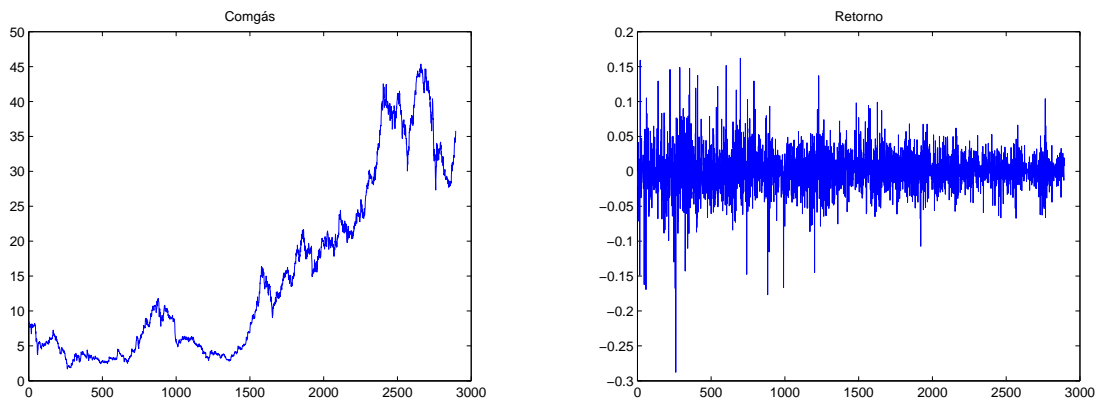


Figura 37: Série e Retorno da Comgás

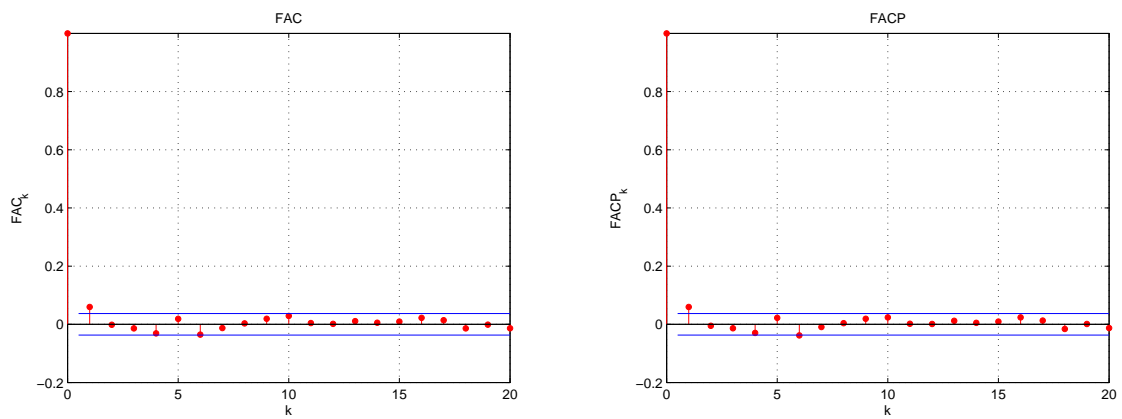


Figura 38: FAC e FACP do Retorno da Comgás

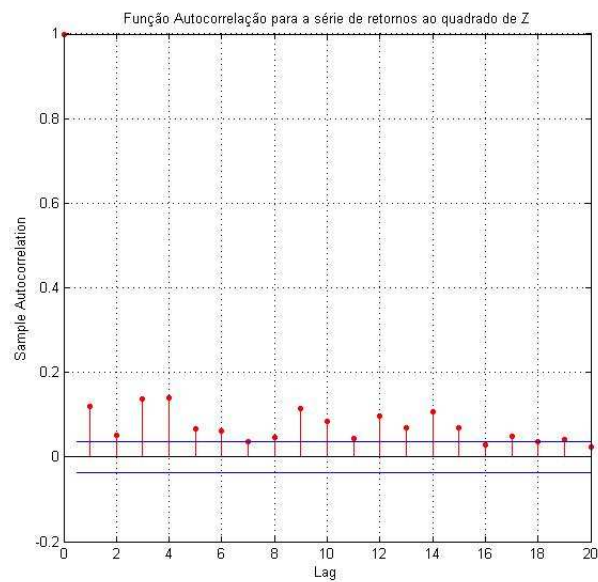


Figura 39: FAC do Retorno ao quadrado da Comgás

com $\Psi > 0$, $\beta_i > 0$, $\alpha_i > 0$, $\forall i$.

Assume-se que $B(L)$ e $B(L) + 1 - A(L)$ possuem raízes fora do círculo unitário, logo são inversíveis. Os autores trabalham com três definições do GARCH:

Definição 1: (GARCH Forte)

A sequência $\{\epsilon_t, t \in Z\}$ é definida para gerar um processo GARCH(p,q) forte se $\Psi, A(L)$ e $B(L)$ possam ser escolhidos tais que:

$$\xi_t = \frac{\epsilon_t}{\sqrt{h_t}} \quad i.i.d.D(0, 1),$$

onde $D(0, 1)$ especifica a distribuição de média zero e variância um.

Definição 2: (GARCH Semi-Forte)

A sequência $\{\epsilon_t, t \in Z\}$ é definida para gerar um processo GARCH(p,q) semi-forte se $\Psi, A(L)$ e $B(L)$ possam ser escolhidos tais que:

$$E[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = 0 \quad e \quad E[\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = h_t.$$

Definição 3: (GARCH Fraco)

A sequência $\{\epsilon_t, t \in Z\}$ é definida para gerar um processo GARCH(p,q) fraco se $\Psi, A(L)$ e $B(L)$ possam ser escolhidos tais que:

$$P[\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = 0 \quad e \quad P[\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \dots] = h_t,$$

onde $P[x_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots]$ denota o melhor preditor linear de x_t em termos de $1, \epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-1}^2, \epsilon_{t-2}^2, \dots$, isto é:

$$E(x_t - P[x_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots])\epsilon_{t-i}^r, \quad i \geq 1 \quad e \quad r = 0, 1, 2.$$

Observe que essas três definições do GARCH requerem:

$$\sum_{i=1}^p \beta_i + \sum_{i=1}^q \alpha_i < 1.$$

Agregando o GARCH(1,1):

Considera-se agora a agregação temporal do modelo GARCH(1,1). Pode-se ver que o modelo GARCH(1,1) é fechado sobre a agregação temporal, ou seja, após a mudança da frequência através da agregação temporal o modelo não deixará de ser GARCH(1,1). Serão usados os modelos de ordem (1,1), porém os de maior ordem são estudados da mesma forma.

Exemplo: (GARCH(1,1), ações)

A classe do processo GARCH(1,1) fraco é fechado a partir da agregação temporal. Mais precisamente, se $\{y_t, t \in Z\}$ é fraco e:

$$h_t = \psi + \beta h_{t-1} + \alpha y_{t-1}^2. \tag{58}$$

Então $\{y_{tm}, t \in Z\}$ será fraco e:

$$h_{(m)tm} = \psi_{(m)} + \beta(m)h_{(m)tm-m} + \alpha_{(m)}y_{tm-m}^2, \tag{59}$$

$$\psi_{(m)} = \psi \frac{1 - (\beta + \alpha)^m}{1 - (\beta + \alpha)}, \quad \alpha_{(m)} = (\beta + \alpha)^m - \beta_{(m)}.$$

Onde $\beta_{(m)} \in (0, 1)$ é solução da equação quadrática:

$$\frac{\beta_{(m)}}{1 + \beta_{(m)}^2} = \frac{\beta(\beta + \alpha)^{m-1}}{1 + \alpha^2 \frac{1 - (\beta + \alpha)^{2m-2}}{1 - (\beta + \alpha)^2} + \beta^2(\beta + \alpha)^{2m-2}}.$$

Supondo $m = 2$.

Trabalha-se então com $\{y_{2t}, t \in Z\}$. Da equação 59:

$$h_{(2)2t} = \psi_{(2)} + \beta_{(2)}h_{(2)2t-2} + \alpha_{(2)}y_{2t-2}^2.$$

Denota-se $\eta_t = y_t^2 - h_t$ onde η_t é i.i.d. com $N(0,1)$. Substituindo η_t em (58):

$$y_t^2 - \eta_t = \psi + \beta(y_{t-1}^2 - \eta_{t-1}) + \alpha y_{t-1}^2 \Rightarrow y_t^2 = \psi + (\beta + \alpha)y_{t-1}^2 - \beta\eta_{t-1} + \eta_t. \quad (60)$$

Atrasando o período da equação 60 uma vez e substituindo nela própria, tem-se:

$$y_t^2 = \psi(1 + \beta + \alpha) + (\beta + \alpha)^2 y_{t-2}^2 - \beta(\beta + \alpha)\eta_{t-2} + \alpha\eta_{t-1} + \eta_t.$$

Substituindo $v_t = -\beta(\beta + \alpha)\eta_{t-2} + \alpha\eta_{t-1} + \eta_t$ para facilitar a escrita, tem-se:

$$y_t^2 = \psi(1 + \beta + \alpha) + (\beta + \alpha)^2 y_{t-2}^2 + v_t. \quad (61)$$

A equação 61 determina a parte autorregressiva do modelo de baixa frequência.

Levando-se em consideração a definição dos autores de melhor preditor linear, e tomando λ tal que v_t seja não correlacionado para índices pares:

$$P[y_t^2 | y_{t-2}, y_{t-4}, \dots] - \lambda P[y_{t-2}^2 | y_{t-4}, y_{t-6}, \dots] = h_t - \lambda h_{t-2} = \psi(1 + \beta + \alpha) + [(\beta + \alpha)^2 - \lambda] y_{t-2}^2. \quad (62)$$

Pode-se perceber, pela relação 62, que os novos coeficientes dependem dos coeficientes do modelo de alta frequência como era previsto, ou seja:

$$\psi_{(2)} = \psi(1 + \beta + \alpha), \quad \alpha_{(2)} = (\beta + \alpha)^2 - \lambda \quad e \quad \beta_{(2)} = \lambda.$$

A figura 40 mostra o que acontece quando é feita uma agregação temporal. As seis curvas indicam uma redução a metade dos intervalos, quando caminha da direita para a esquerda, o que indica uma diminuição da frequência. Quando α e β chegam à origem, o modelo perde a heterocedasticidade, ou seja, deixa de ser GARCH, pois conforme o período vai para infinito o processo GARCH(1,1) tende a ARCH(1) ($\beta_{(m)} \rightarrow 0$) e depois perde a heterocedasticidade ($\alpha_{(m)}, \beta_{(m)} \rightarrow 0$).

8.3 Análise dos Algoritmos Genéticos

Para iniciar o estudo da metodologia dos algoritmos genéticos, em busca da solução do nosso problema: “Qual conjunto de ativos escolho para investimento?”, montamos uma carteira que contém 6 ações do IBrX, entre elas: Banco do Brasil, Petrobrás, Vivo, Comgás, Bom Bril e Dólar. Todas as informações foram tiradas do *site* da BOVESPA, através da utilização de um programa chamado “Economática”. Os dados dessa séries estão datados de 09/1997 a 05/2009, data presente da coleta desses dados.

A partir da utilização do programa *Retorno*, que desenvolvemos com o auxílio do Matlab e que pode ser encontrado no *Apêndice F*, obtemos os gráficos da FAC dos retornos de cada um dos ativos que compõem a carteira e em seguida os gráficos da FAC dos retornos ao quadrado, onde a partir desses, podemos ver que as séries apresentam heterocedasticidade. Para confirmação da nossa hipótese

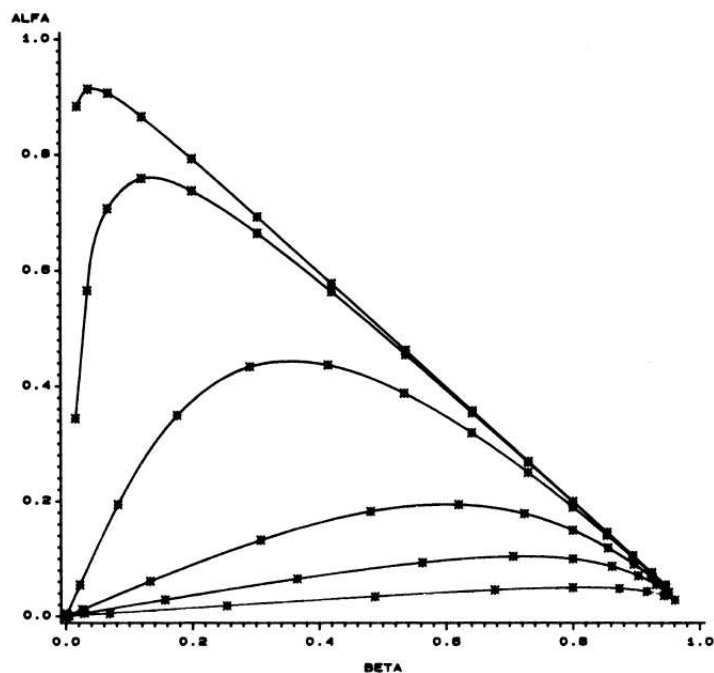


Figura 40: Agregação Temporal

Ativo	Banco do Brasil	Bom Bril	Comgás	Dólar	Petrobrás	Vivo
Retorno Esperado (\bar{R}_P)	-0.53%	2.06%	1.62%	0.24%	-0.66%	1.87%

Tabela 9: Retornos Esperados

a partir do gráficos, aplicamos o teste de Engle, que também se encontra no programa *Retorno*, e a partir dos resultados obtidos confirmamos a presença de heterocedasticidade, pois a hipótese nula do teste foi rejeitada.

O próximo passo do programa é gerar uma matriz com as seis séries em colunas e calcular a covariância dessa matriz, com o intuito de avaliar cada uma dessas ações a partir da sua variância. Ao terminar de rodar esse programa, obtém-se como resposta os retornos esperados de cada uma das ações que compunham a carteira e também o gráfico da fronteira eficiente, que combina os retornos esperados com o risco de cada uma das séries. Esse gráfico pode ser visto na figura 41 e os retornos obtidos podem ser encontrados na tabela 9:

O primeiro algoritmo genético que foi desenvolvido, o qual será denominado de **AG1**, tem como função de aptidão a função dos retornos e otimizá-la é obter seu valor máximo. Ele foi implementado utilizando-se duas fases: a primeira fase (**FASE1**), consiste na sua execução com taxas de cruzamento e mutação definidas como padrão pela literatura, respectivamente, com os valores de 100% e 1%.

Após obter os retornos esperados, formula-se a função retorno da carteira com os dados obtidos na tabela 9, que é a função aptidão do **AG1**. Então, utilizando o programa *MarkowAG* no *Apêndice F*), que irá gerar os pesos propostos pelo **AG1** para ser aplicado em cada ação da carteira. Como os pesos representam proporções de investimento, deve-se lembrar da restrição estabelecida no programa *Markowrestric* (também no *Apêndice F*), de que a soma desses pesos não pode ultrapassar um. Eles representam a primeira exploração do espaço de soluções, com o objetivo de “calibrar” o **AG1**. Então, após a **FASE1** determinam-se as melhores taxas de cruzamento e de mutação, obtidas pelo processo de varredura e observadas no gráfico 42. Passando para a segunda fase (**FASE2**), executa-se o **AG1** com essas taxas ótimas.

A partir dos gráficos da figura 42, pode-se ver que as taxas ótimas sugeridas pelo programa foram: 90% para mutação e 70% para cruzamento. Inicia-se a **FASE2**, executando o **AG1** com essas taxas.

A tabela 10 mostra todos os dados gerados pelo **AG1**: os retornos de cada ação, os pesos (w_{1i})

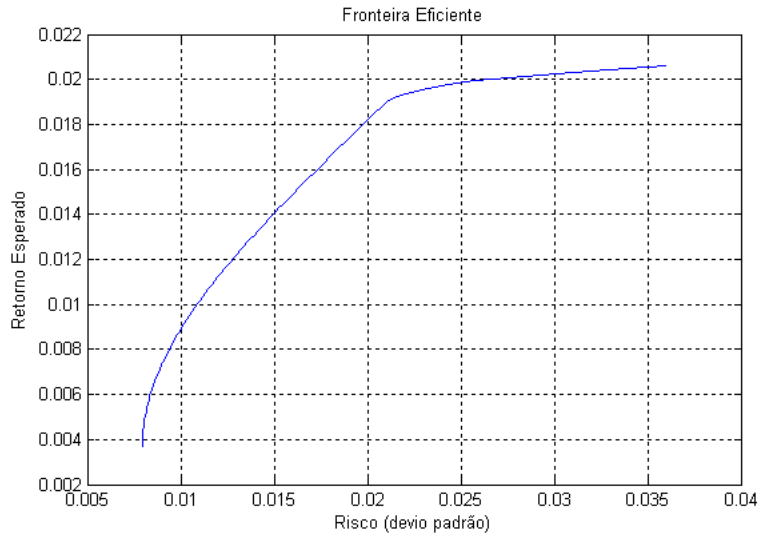


Figura 41: Fronteira Eficiente

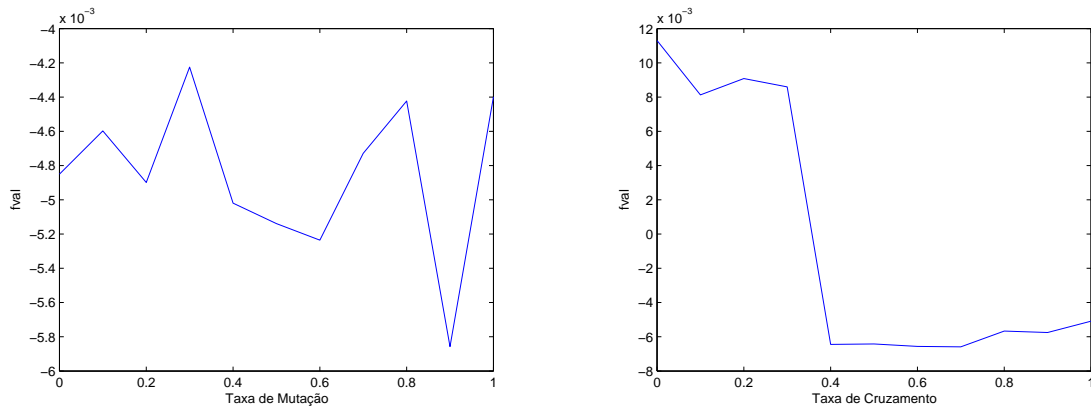


Figura 42: Taxas de Mutação e Cruzamento

dados antes do algoritmo fornecer as taxas de cruzamento e mutação ideais (*FASE1*) e os pesos (w_{1j}) calculados pelo programa com as taxas sugeridas (*FASE2*). Os índices i e j representarão daqui pra frente os pesos das fases *FASE1* e *FASE2*, respectivamente e os números representarão o algoritmo em questão, AG1 ou AG2. Essa notação será mantida para as próximas aplicações.

Ações	Retorno (R_i)	Peso <i>FASE1</i> (w_{1i})	Peso <i>FASE2</i> (w_{1j})
Banco do Brasil	-0,53%	50,71%	2,24%
Bom Bril	2,06%	0,56%	0,06%
Comgás	1,62%	4,62%	0,03%
Dólar	0,24%	1,09%	0,05%
Petrobrás	-0,66%	42,05%	97,59%
Vivo	1,87%	1,17%	0,04%

Tabela 10: Pesos da FASE1 e FASE2

Um outro gráfico que o programa exibe, figura 43, mostra a avaliação média, a avaliação do melhor e do pior indivíduo (conjunto de pesos) e a distância média entre eles. Pode-se ver que, na *FASE1* do AG1, como a taxa de mutação é pequena, a avaliação dos indivíduos converge rapidamente para a avaliação média, da mesma forma que a distância média entre eles converge para 0, demonstrando

que toda a população se resume a um único indivíduo.

Já na figura 44, que representa a *FASE2* do *AG1*, não há convergência destas medidas, indicando uma maior diversidade da população em estudo. Consequência da mudança significativa da taxa de mutação de 1% para 90%.

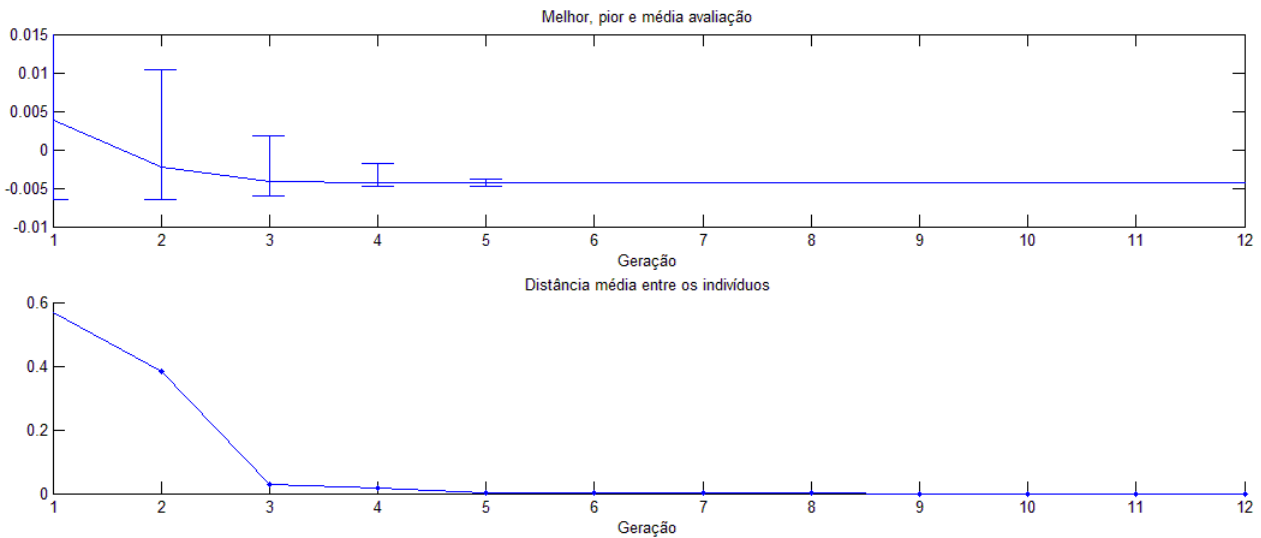


Figura 43: Avaliação e Distância média entre os indivíduos: *AG1*, *FASE1*

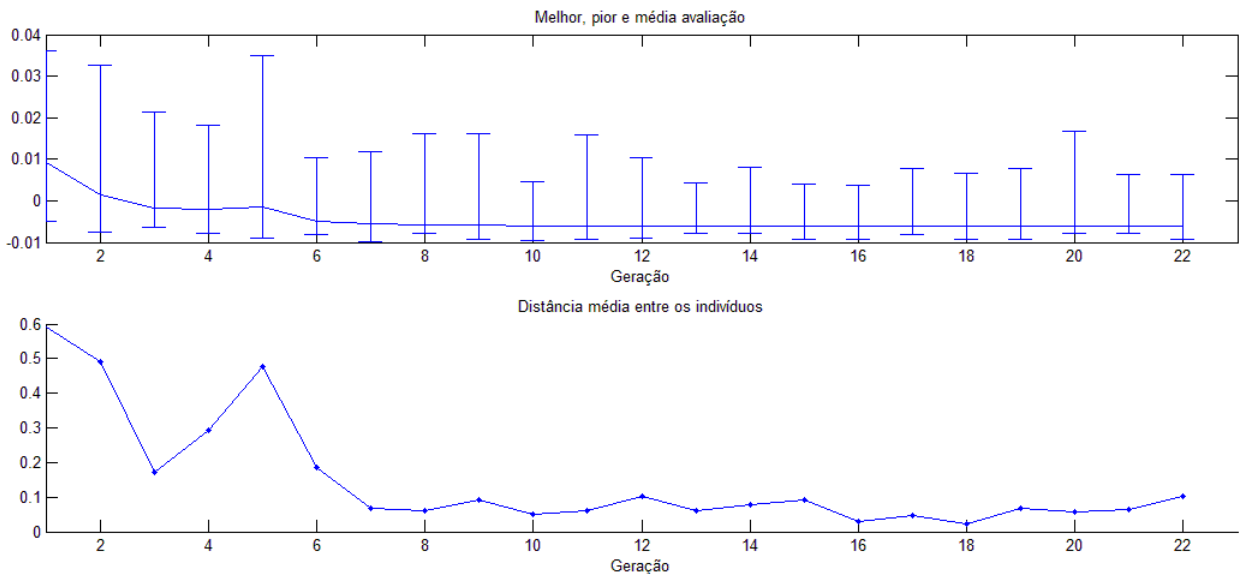


Figura 44: Avaliação e Distância média entre os indivíduos: *AG1*, *FASE2*

Nas tabelas 11 e 12 estão demonstrados alguns cálculos feitos no *Excel* a partir das informações geradas pelo *AG1* nas etapas *FASE1* e *FASE2*. Foram calculados os Retornos Esperados da carteira e seu Risco com os pesos de cada etapa.

Como pode ser visto nas tabelas 11 e 12, foram obtidos retornos negativos de 0,44% com um risco de 0,003% a partir dos pesos fornecidos pelo *AG1* na *FASE1* e 0,65% com um risco de 0,000093% a partir dos pesos fornecidos pelo *AG1* na *FASE2*. Nesse caso houve alteração no resultado do problema (diferença dos retornos), mas ambos os casos têm um risco praticamente zero. Lembrando que no *AG1*, a função de aptidão envolve somente o retorno da carteira e não o seu risco.

Ações	Retorno (R_i)	Peso <i>FASE1</i> (w_{1i})	$R_i * w_{1i}$	$\sqrt{(R_i - \bar{R})^2 * w_{1i}}$
Banco do Brasil	-0,53%	50,71%	-0,2688%	0,000045%
Bom Bril	2,06%	0,56%	0,0115%	0,000349%
Comgás	1,62%	4,62%	0,0748%	0,00195%
Dólar	0,24%	1,09%	0,0026%	0,000050%
Petrobrás	-0,66%	42,05%	-0,2775%	0,000212%
Vivo	1,87%	1,17%	0,0219%	0,000622%
Total	-	-	-0,4354%	0,003%

Tabela 11: Retorno e Risco na FASE1

Ações	Retorno (R_i)	Peso <i>FASE2</i> (w_j)	$R_i * w_j$	$\sqrt{(R_i - \bar{R})^2 * w_j}$
Banco do Brasil	-0,53%	2,24%	-0,0119%	0,000003%
Bom Bril	2,06%	0,06%	0,0012%	0,000044%
Comgás	1,62%	0,03%	0,0005%	0,000016%
Dólar	0,24%	0,05%	0,0001%	0,000004%
Petrobrás	-0,66%	97,59%	-0,6441%	0,000000%
Vivo	1,87%	0,04%	0,0007%	0,000025%
Total	-	-	-0,6534%	0,000093%

Tabela 12: Retorno e Risco na FASE2

8.4 VaR e Algoritmos Genéticos

O segundo algoritmo genético desenvolvido, o qual será denominado de **AG2**, tem como função de aptidão a função VaR, e otimizá-la é obter seu valor mínimo. Como no caso do **AG1** será trabalhado duas etapas: *FASE1* e *FASE2*.

Para efeito de teste do *AG2*, iniciamos o nosso problema montando uma carteira que continha três ações do IBrX, entre elas Banco do Brasil, Comgás e Bom Bril. Todas as informações foram tiradas do *site* da BOVESPA, através da utilização de um programa chamado “Economática”. Os dados dessas séries estão datados de 09/1997 a 05/2009, data presente da coleta desses dados.

A partir da utilização do programa *RetornoVaR*, que foi desenvolvido com o auxílio do Matlab e que pode ser encontrado no *Apêndice G*, foi gerado um vetor com os valores esperados dos retornos de cada um desses ativos, juntamente com a matriz de autocovariância. Em seguida foi rodado o *AG2*, *FASE1* (programa *VaRAG* também encontrado no *Apêndice G*), que fornece resultados para as taxas de cruzamento e mutações padrão já citadas. Depois será iniciado o processo de varredura gerando os gráficos que fornecem as taxas ótimas para a carteira. Esses valores podem ser observados na figura 45.

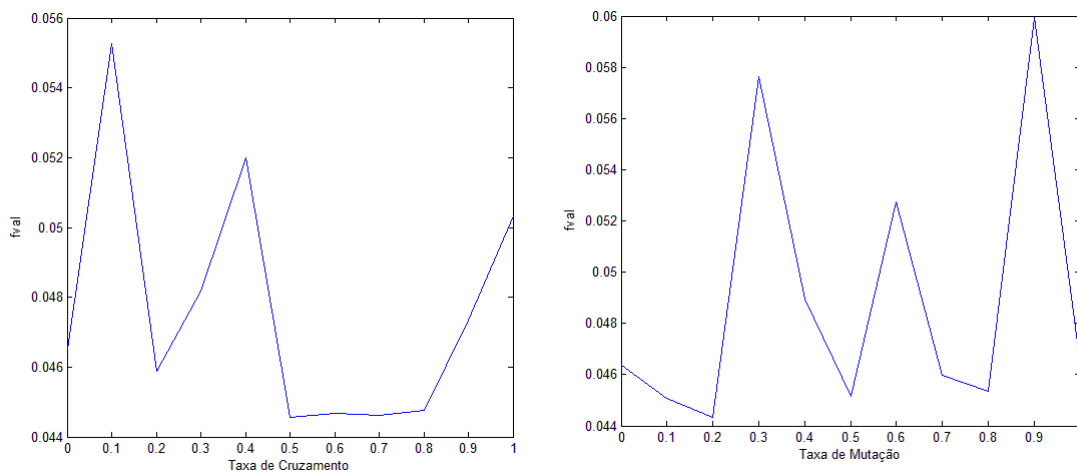


Figura 45: Taxa de Cruzamento e Mutação

Ações	Peso <i>FASE1</i> (w_{2i})	Peso <i>FASE2</i> (w_{2j})
Banco do Brasil	67,87%	26,61%
Bom Bril	25,11%	34,44%
Comgás	7,82%	38,95%

Tabela 13: Pesos - AG2, FASE1 e FASE2

A partir desses gráficos, pode ser visto que as taxas que melhor ajustam o modelo são 50% de cruzamento e 20% de mutação. O próximo passo, *FASE2*, do *AG2*, será substituir esses valores no programa *VaRAG* para que este retorne os pesos ideais de investimento em cada ativo que compõe a carteira.

A tabela 13 mostra os pesos gerados pelo *AG2*, *FASE1* (w_{2i}) e os pesos gerados pelo *AG2*, *FASE2* (w_{2j}), ou seja, com 50% de cruzamento e 20% de mutação.

Em seguida, são exibidos os gráficos da avaliação média, a avaliação do melhor e do pior indivíduo (conjunto de pesos) e o da distância média entre eles para o *AG2*. Nas figuras 46 e 47 percebe-se, como no caso do *AG1*, o papel fundamental da taxa de mutação para garantir a diversidade da população em questão. Na *FASE2*, percebe-se que não ocorre a convergência da avaliação do pior e melhor indivíduo para a avaliação média, nem a convergência da distância média para 0.

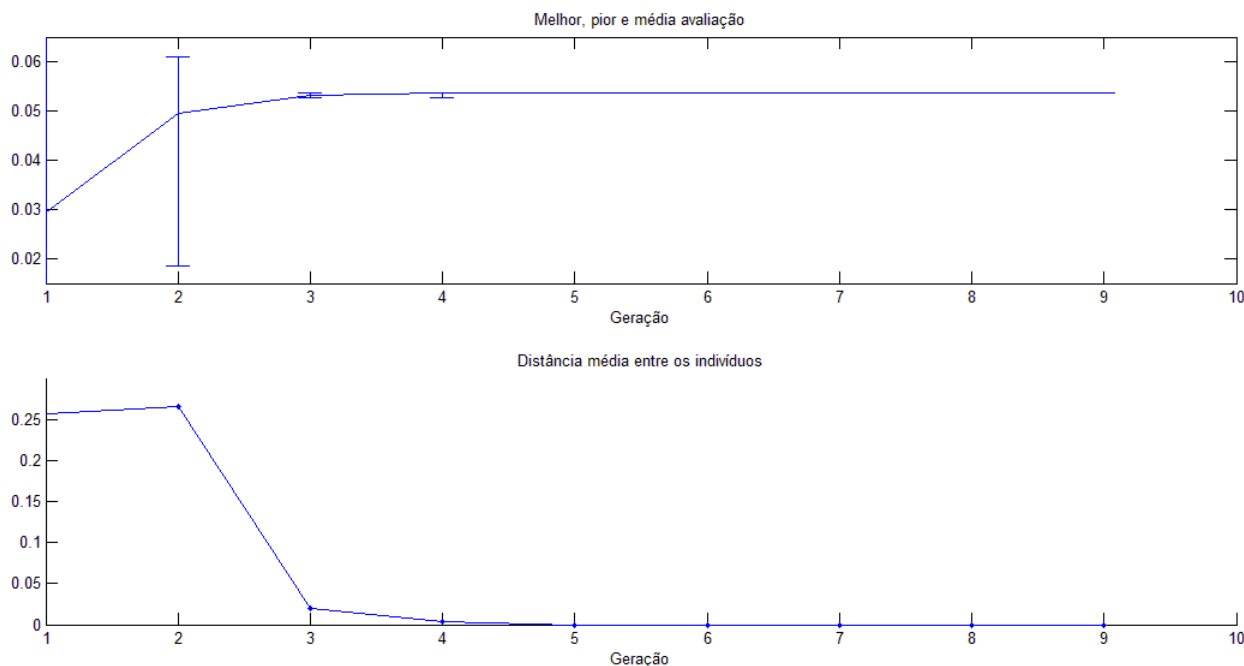


Figura 46: Avaliação e Distância média entre os indivíduos: *AG2*, *FASE1*

8.5 Aplicação Final

Uma vez feitos todos os testes dos programas, foram escolhidas três carteiras, com 12 ações cada uma, para que elas fossem analisadas a partir dos programas *AG1* e *AG2*, e que tais algoritmos fossem comparados.

As carteiras foram selecionadas a partir do *site* da BOVESPA e o critério de seleção que foi utilizado para montá-las foi de acordo com seus pesos no IBRX, ou seja, na carteira 1 estão as ações que têm a maior participação no IBRX (maiores pesos), na carteira 2 estão as de pesos intermediários e na carteira 3 as de menores pesos. Os dados foram coletados de 01/01/2004 a 17/11/2009. Assim, a carteira 1 foi composta pelas seguintes ações: Ambev, Banco do Brasil, Bradesco, Cemig, Embraer,

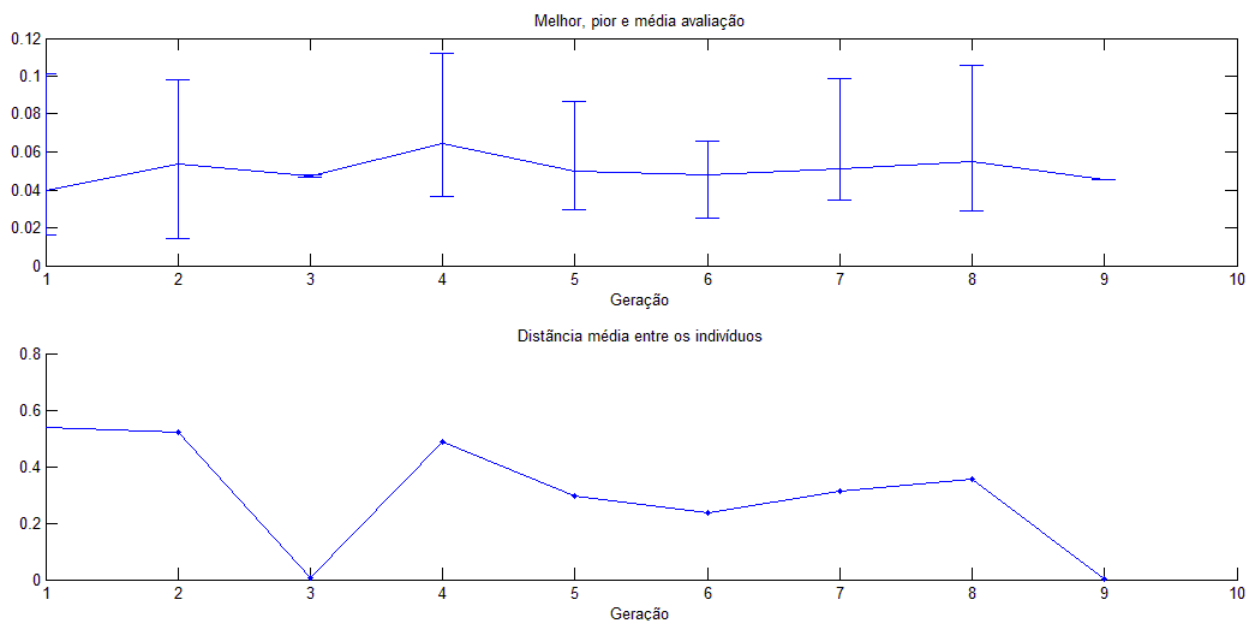


Figura 47: Avaliação e Distância média entre os indivíduos: *AG2*, *FASE2*

Gerdau, Itausa, Itaú Unibanco, Petrobrás, Telemar, Usiminas e Vale. A carteira 2 foi composta por: ALL, Bradespar, Brasil T Par, CC Rodovias, Eletrobrás, Gerdau Met, Pão de Açúcar, Sabesp, Tractebel, Trans Paulista, Ultrapar e Vivo. E, finalmente, a carteira 3 foi composta por: Brasken, Celesc, Comgás, Confab, Brasil Telecom, Fosfertil, Light S.A., Lojas Americanas, NET, Suzano Papel, Telesp e TIM.

O primeiro passo foi aplicar o *AG1* (*Apêndice F*), também denominado por algoritmo de Markowitz. Como já foi dito anteriormente, inicialmente o programa *Retorno* exibe os gráficos da FAC dos retornos de cada um dos ativos que compõe as carteiras e em seguida os gráficos da FAC dos retornos ao quadrado, e a partir desses pode-se ver que as séries apresentam heterocedasticidade. Para verificação desta hipótese, foi aplicado o teste de Engle que confirma a presença de heterocedasticidade.

Em seguida o programa gera os retornos de cada uma das ações compostas em cada uma das carteiras e exibe as fronteiras eficientes de cada uma das carteiras.

A tabela 14, mostra os retornos de cada uma das ações das carteiras 1, 2 e 3 respectivamente. A partir dos retornos dados pelo programa calcula-se os retornos esperados de cada carteira e, conseqüentemente seus riscos. Nas figuras 48, 49 e 50 pode-se ver os gráficos das fronteiras eficientes de cada uma das carteiras.

Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
Ações	Retorno (\bar{R})	Ações	Retorno (\bar{R})	Ações	Retorno (\bar{R})
Ambev	6,84%	ALL	1,20%	Brasken	-1,32%
Banco do Brasil	2,79%	Bradespar	-0,92%	Celesc	-0,57%
Bradesco	-0,32%	Brasil T Par	-3,60%	Comgás	-0,90%
Cemig	0,10%	CC Rodovias	5,99%	Confab	1,42%
Embraer	-2,36%	Eletrobrás	-1,89%	Brasil Telecom	3,81%
Gerdau	-1,90%	Gerdau Met	-0,08%	Fosfertil	-1,31%
Itaú Unibanco	-2,01%	Pão de Açúcar	-1,90%	Light S.A.	5,57%
Itausa	3,08%	Sabesp	-1,47%	Lojas Americanas	-8,63%
Petrobrás	2,61%	Tractebel	-0,41%	NET	1,40%
Telemar	-3,71%	Trans Paulista	-0,25%	Suzano Papel	-1,46%
Usiminas	6,54%	Ultrapar	0,45%	Telesp	1,04%
Vale	2,79%	Vivo	-5,40%	TIM	-2,30%

Tabela 14: Retornos Esperados da Carteiras 1, 2 e 3

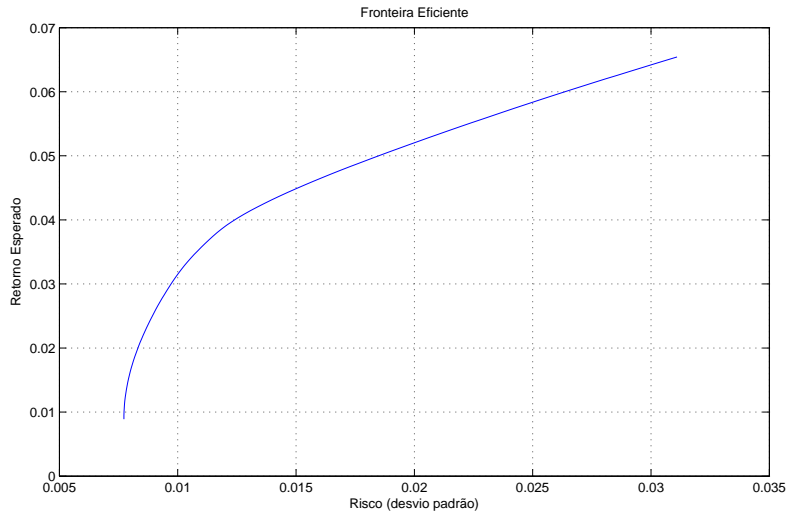


Figura 48: Fronteira Eficiente - Carteira 1

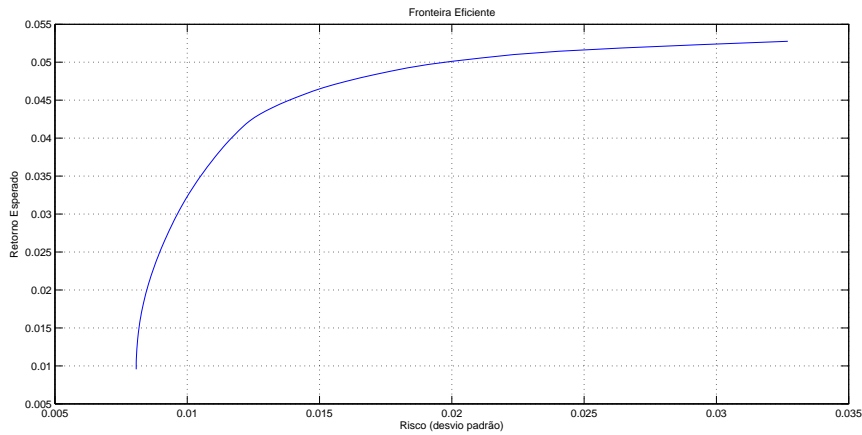


Figura 49: Fronteira Eficiente - Carteira 2

A partir do *AG1* foram aplicados os mecanismos de mutação e de cruzamento para se obter os pesos de aplicação em cada ação de cada uma das três carteiras. Os primeiros pesos a serem gerados, como definidos na *FASE1*, apresentam taxa de mutação de 1% e taxa de cruzamento de 100%. Esses valores são considerados padrões e serão usados sempre que o programa é rodado na *FASE1*, para depois iniciar o processo de varredura e gerar os gráficos das taxas de mutação e cruzamento, dos quais obtém-se as taxas ótimas para a *FASE2*, que serão aplicadas em cada uma das três carteiras.

A tabela 15 mostra os pesos da *FASE1* gerados pelo *AG1* de cada uma das ações das três carteiras. A numeração dos índices, de agora em diante, representará a carteira em questão e as letras *i* e *j*, continuarão a representar as *FASE1* e *FASE2*, respectivamente, por exemplo: w_{1i} representa o peso da *FASE1* da carteira 1, w_{2i} representa o peso da *FASE1* da carteira 2 e w_{3i} representa o peso da *FASE1* da carteira 3. Essa notação seguirá até o final.

As figuras 51, 52 e 53 mostram as taxas de mutação e cruzamento sugeridas pelo *AG1*. A partir disso, essas taxas substituirão as taxas padrões e a *FASE2* será rodada, para que o programa retorne os novos pesos.

A partir dos gráficos das figuras 51, 52 e 53, pode ser visto que as taxas ótimas foram: para a carteira 1: 10% para mutação e 50% para cruzamento, para a carteira 2: 80% para mutação e 20% para cruzamento e para a carteira 3: 80% para mutação e 20% para cruzamento.

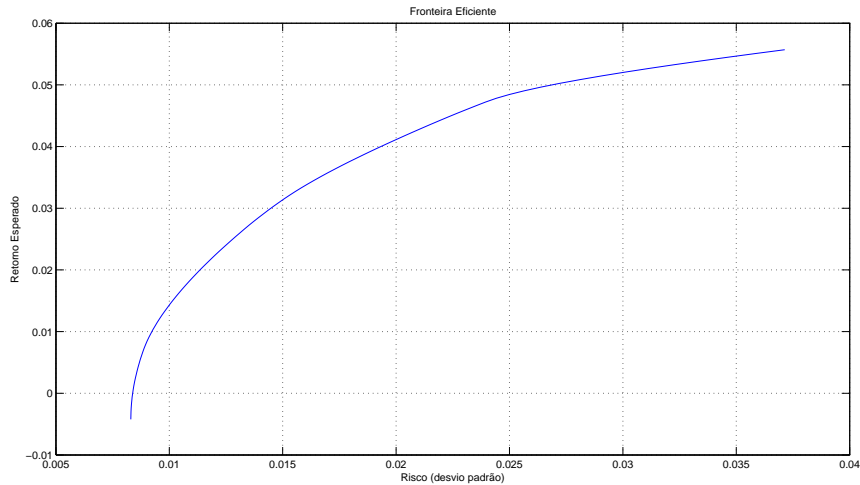


Figura 50: Fronteira Eficiente - Carteira 3

Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
Ações	Peso <i>FASE1</i> (w_{1i})	Ações	Peso <i>FASE1</i> (w_{2i})	Ações	Peso <i>FASE1</i> (w_{3i})
Ambev	4,42%	ALL	6,84%	Brasken	14,73%
Banco do Brasil	9,98%	Bradespar	6,14%	Celesc	9,75%
Bradesco	0,36%	Brasil T Par	23,43%	Comgás	12,67%
Cemig	3,47%	CC Rodovias	8,31%	Confab	8,83%
Embraer	0,32%	Eletrobrás	3,19%	Brasil Telecom	5,75%
Gerdau	29,68%	Gerdau Met	25,16%	Fosfertil	12,97%
Itaú Unibanco	1,29%	Pão de Açúcar	5,77%	Light S.A.	0,67%
Itausa	8,19%	Sabesp	5,01%	Lojas Americanas	0,38%
Petrobrás	20,55%	Tractebel	0,68%	NET	13,66%
Telemar	2,73%	Trans Paulista	12,07%	Suzano Papel	10,72%
Usiminas	9,26%	Ultrapar	0,62%	Telesp	9,88%
Vale	9,85%	Vivo	0,67%	TIM	0,02%

Tabela 15: Tabela dos pesos: *AG1*, *FASE1*

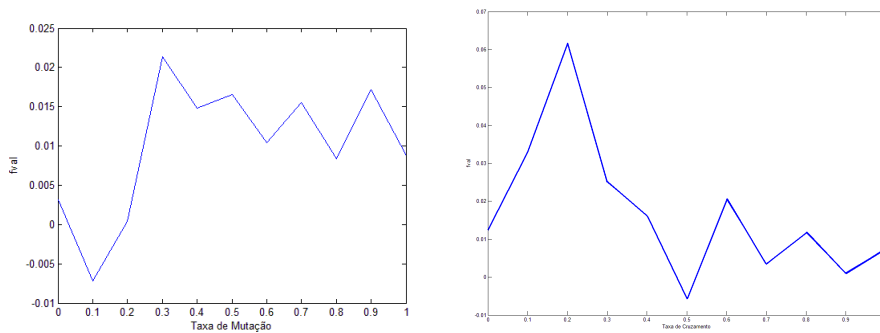


Figura 51: Taxas de Mutação e Cruzamento - Carteira 1 - *AG1*

A tabela 16 mostra os pesos obtidos pelo *AG1* na *FASE2*.

Novamente a avaliação média, a avaliação do melhor e do pior indivíduo (conjunto de pesos) e a distância média entre eles é exibida. Nas figuras 54, 55 e 56 pode-se ver, respectivamente, a avaliação dos indivíduos na *FASE1* (figuras superiores) e *FASE2* (figuras inferiores) para as carteiras 1, 2 e 3.

Mais uma vez pode ser observada a importância da mutação após o cruzamento dos indivíduos. Note que os dois gráficos na parte superior da figura 54 foram gerados com as taxas de mutação (1%) e

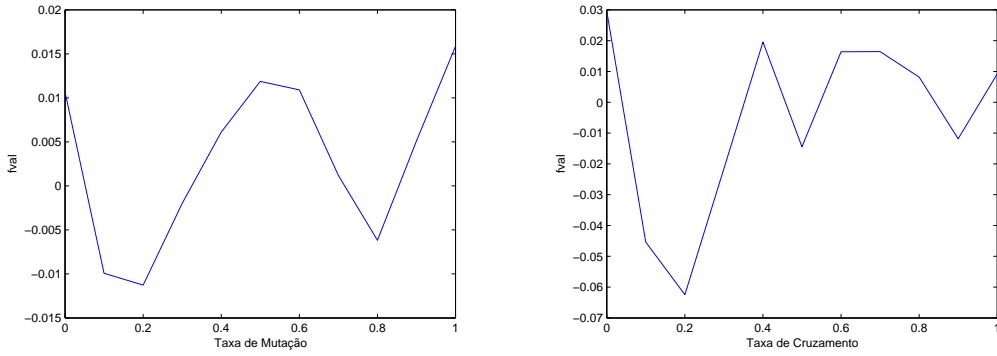


Figura 52: Taxas de Mutação e Cruzamento - Carteira 2 - *AG1*

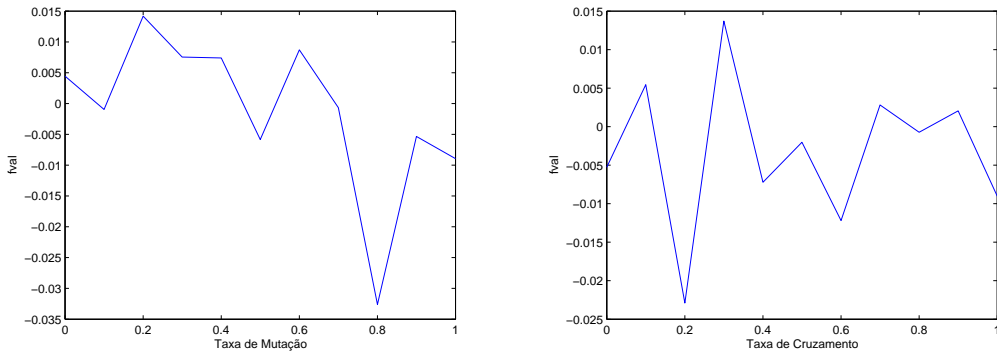


Figura 53: Taxas de Mutação e Cruzamento - Carteira 3 - *AG1*

Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
	Mut 10%-Cruz 50%		Mut 80%-Cruz 20%		Mut 80%-Cruz 20%
Ações	Peso <i>FASE2</i> (w_i)	Ações	Peso <i>FASE2</i> (w_j)	Ações	Peso <i>FASE2</i> (w_k)
Ambev	5,49%	ALL	4,78%	Brasken	16,54%
Banco do Brasil	5,69%	Bradespar	18,83%	Celesc	9,64%
Bradesco	3,54%	Brasil T Par	17,10%	Comgás	13,08%
Cemig	8,81%	CC Rodovias	11,23%	Confab	0,85%
Embraer	1,45%	Eletróbás	1,83%	Brasil Telecom	2,05%
Gerdau	12,26%	Gerdau Met	5,99%	Fosfertil	4,88%
Itaú Unibanco	6,82%	Pão de Açúcar	4,57%	Light S.A.	8,25%
Itausa	16,67%	Sabesp	3,96%	Lojas Americanas	0,93%
Petrobás	11,21%	Tractebel	4,24%	NET	4,42%
Telemar	17,03%	Trans Paulista	18,34%	Suzano Papel	12,25%
Usiminas	7,35%	Ultrapar	6,56%	Telesp	8,95%
Vale	3,68%	Vivo	2,57%	TIM	24,43%

Tabela 16: Tabela dos pesos: *AG1*, *FASE2*

cruzamento (100%) padrões. Pode-se ver que as avaliações dos indivíduos convergem para a avaliação média rapidamente, o que mostra a existencia de pouca diversidade entre esses indivíduos e que a distância média entre eles converge rapidamente para zero, o que diz que existe pouca chance de que esses indivíduos explorem outros mínimos locais.

O contrário acontece nos dois gráficos na parte inferior da figura 54, pois nesse caso a taxa de mutação dos indivíduos é maior (10%), causando um aumento na sua diversidade e fazendo com que a distância média entre eles oscile sem convergir para o zero. Isso significa que existe mais chance destes indivíduos explorem outros mínimos locais, levando a um mínimo global.

Esse mesmo comportamento é observado nas figuras (55) e (56), que representam as avaliações e distância média das carteiras 2 e 3, respectivamente.

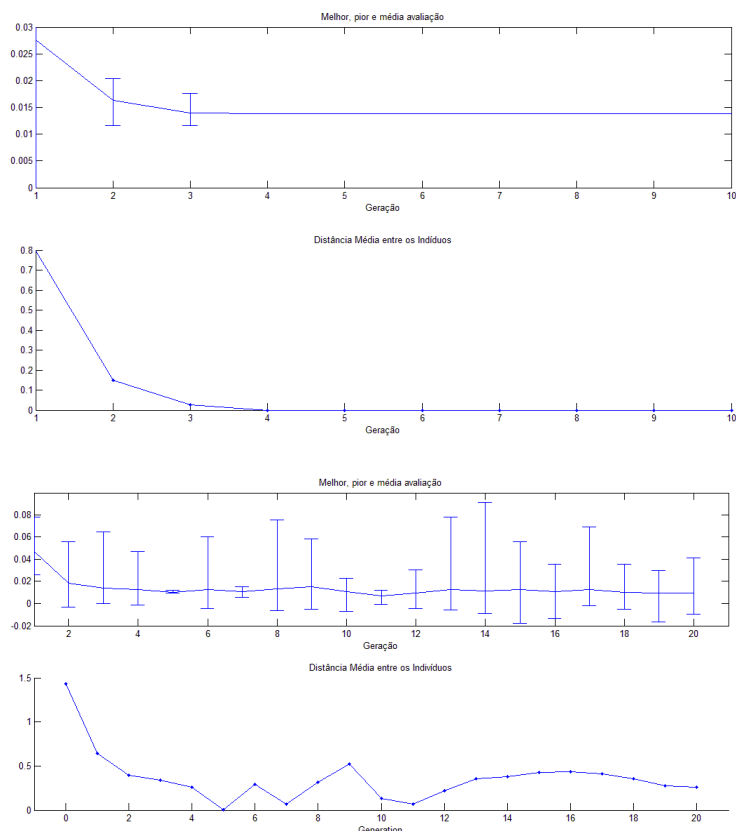


Figura 54: Avaliação e Distância média entre os indivíduos - Carteira 1 - *FASE1* (superior) e *FASE2* (inferior)

Para poder comparar as metodologias utilizadas nos dois algoritmos genéticos, foi suposto um investimento inicial de R\$100.000 em cada uma das carteiras, o qual ficaria aplicado por um ano, de 02/01/2008 a 02/01/2009. Nas tabelas 17, 18 e 19, encontram-se os resumos dos cálculos feitos com o auxílio do *Excel*. Na segunda coluna, encontram-se as variações (x) do valor de fechamento de cada ação, para que seja possível saber se elas obtiveram lucro ou perda no decorrer desse ano e na terceira e quarta colunas, encontram-se os pesos fornecidos pelo *AG1* na *FASE1* e *FASE2*.

A partir desses dados, pode-se calcular o total em reais resultante desse investimento em cada uma das carteiras e, conseqüentemente se houve ganho ou perda. Para isso deve-se descobrir o valor, em reais, investido em cada ação da carteira (multiplicando o peso (w) ao valor do investimento inicial) e em seguida, multiplicar esse valor à variação (x) de cada uma das carteiras durante esse ano. A soma dessas quantias em reais é o quanto o investimento inicial rendeu ao longo desse ano, ou seja, o capital final do investimento. Se for dividido o capital final pelo capital inicial e do resultado subtrair uma unidade, será obtido como resposta a variação (ganho ou perda em %) desse investimento. Todo esse processo é feito para os pesos da *FASE1* (w_i) e *FASE2* (w_j) de cada uma das três carteiras, ou seja, para os pesos com as taxas de mutação e cruzamento padrões e com as taxas ótimas sugeridas pelo programa.

A tabela 20 mostra os resultados obtidos com os cálculos mencionados anteriormente.

Percebe-se que as três carteiras tiveram uma perda do investimento inicial. É interessante nesse caso, dizer que o mercado de ações BOVESPA teve uma perda de 35% (variação do valor de fechamento da BOVESPA), isto diz que se fosse aplicado esse valor por um ano, utilizando as sugestões geradas pelo *AG1* (algoritmo genético de Markowitz), a perda teria sido menor do que se fosse aplicado na BOVESPA, exceto no caso da carteira 1 a partir dos pesos da *FASE1*.

É importante destacar que não se pode tomar conclusões precipitadas a respeito das carteiras, é preciso antes saber quais são seus riscos e retornos. Para isso a tabela 21 mostra os retornos

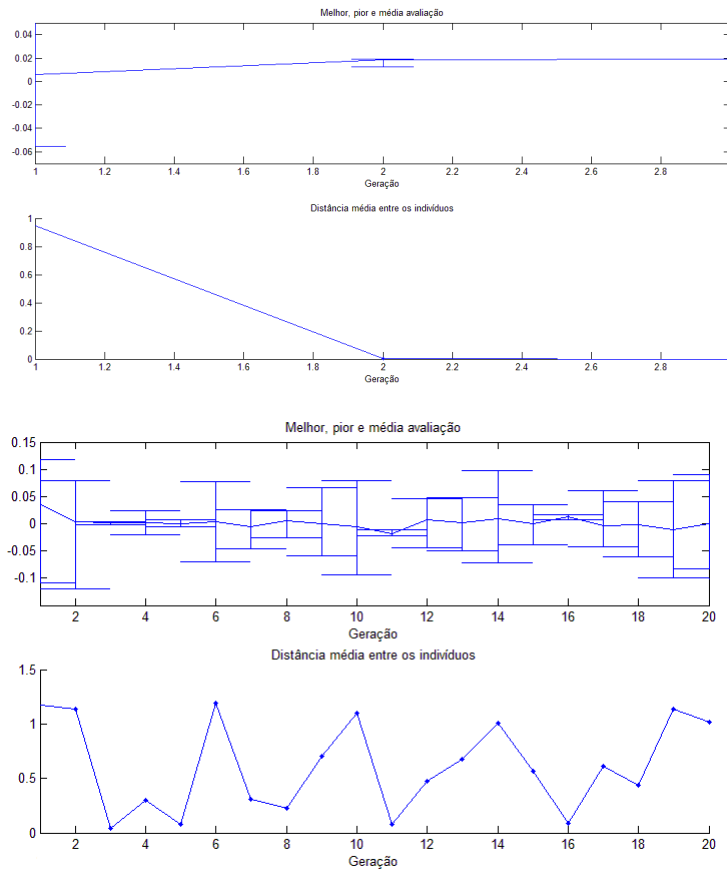


Figura 55: Avaliação e Distância média entre os indivíduos - Carteira 2 - *FASE1* (superior) e *FASE2* (inferior)

Ações	Varição(x)	$w_{1i} * 100.000 * x$	$w_{1j} * 100.000 * x$
Ambev	87%	R\$3.850	R\$4.782, 26
Banco do Brasil	53%	R\$5.299, 81	R\$5.690
Bradesco	73%	R\$263, 44	R\$2.590, 53
Cemig	110%	R\$3.802, 35	R\$9.653, 79
Embraer	52%	R\$167, 55	R\$759, 19
Gedau	67%	R\$19.737, 55	R\$8.153, 05
Itaú Unibanco	87%	R\$1.124, 94	R\$5.947, 35
Itausa	87%	R\$7.166, 13	R\$14.586, 01
Petrobrás	59%	R\$12.090, 51	R\$6.595, 36
Telemar	111%	R\$3.036, 30	R\$18.940, 71
Usiminas	57%	R\$5.303, 98	R\$4.209, 96
Vale	55%	R\$5.434, 28	R\$2.030, 27
TOTAL		R\$57.836, 58	R\$70.875, 69

Tabela 17: Variação do investimento para a carteira 1 na *FASE1* e *FASE2*

esperados das carteiras 1, 2 e 3 juntamente com seus respectivos desvios padrão (riscos). Esses dados foram calculados através das fórmulas da seção (7) e com o auxílio do *Excel*. Os retornos esperados calculados a partir dos pesos iniciais serão denotados por \bar{R}_i e os retornos esperados calculados a partir dos pesos finais serão denotados por \bar{R}_f . Da mesma forma, σ_i representará o desvio padrão do retorno esperado da *FASE1* e σ_f representará o desvio padrão do retorno esperado da *FASE2*.

Agora será utilizado o *AG2* (Apêndice G), onde a função de aptidão (ou avaliação) será a função VaR e pretende-se minimizar seu valor. Para isso foi suposto que o nível de confiança usado será o de 99% e pela tabela normal foi encontrado que $\alpha = -2.33$ porém, como já foi explicado anteriormente,

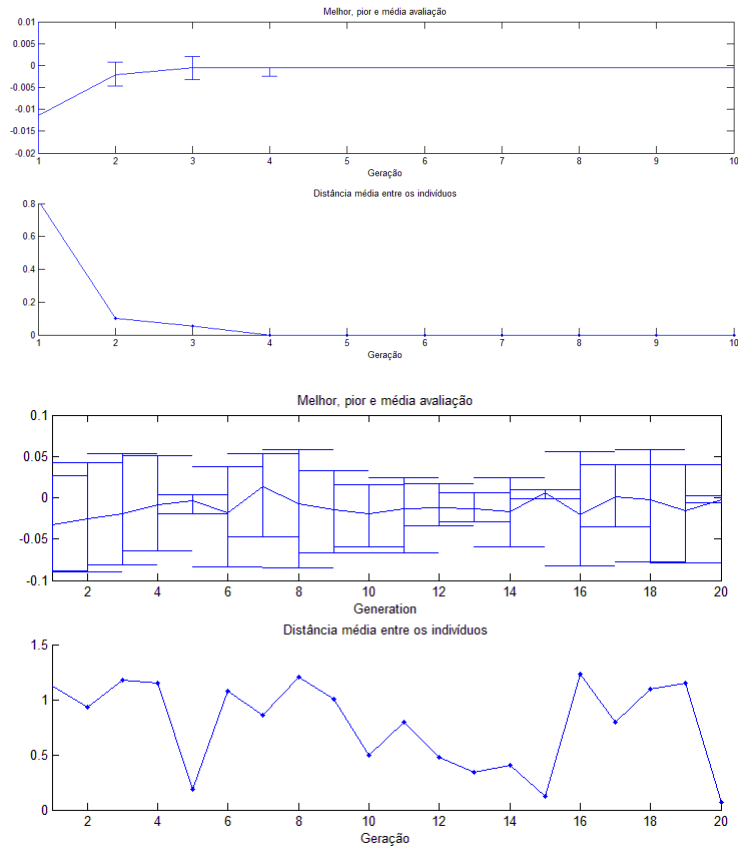


Figura 56: Avaliação e Distância média entre os indivíduos - Carteira 3 - *FASE1* (superior) e *FASE2* (inferior)

Ações	Variação(x)	$w_{2i} * 100.000 * x$	$w_{2j} * 100.000 * x$
ALL	47%	R\$3.213, 91	R\$2.247, 44
Bradespar	47%	R\$2.866, 31	R\$8.788, 98
Brasil T Par	80%	R\$18.828, 75	R\$13.744, 60
CC Rodovias	92%	R\$7.673, 47	R\$10.366, 39
Eletrobrás	113%	R\$3.598, 26	R\$2.066, 27
Gerdau Met	65%	R\$16.274, 15	R\$3.875, 53
Pão de Açúcar	100%	R\$5.753, 50	R\$4.560, 13
Sabesp	71%	R\$3.573, 20	R\$2.826, 18
Tractebel	101%	R\$683, 97	R\$4.260, 84
Trans Paulista	126%	R\$15.248, 26	R\$23.165, 93
Ultrapar	88%	R\$544, 77	R\$5.761, 99
Vivo	81%	R\$539, 50	R\$2.069, 15
TOTAL		R\$78.798, 05	R\$83.733, 42

Tabela 18: Variação do investimento para a carteira 2 na *FASE1* e *FASE2*

será usado o valor positivo de α para representar essa perda.

Nesse caso os retornos de cada ação não são exibidos, são armazenados no Matlab para que os outros programas o utilizem. Ao rodar o programa *VaRAG* obtém-se os pesos da *FASE1* e depois inicia-se o processo de varredura para as taxas de cruzamento e mutação.

Na tabela 22 exibi os pesos da *FASE1*, das ações de cada uma das três carteiras, obtidos pelo *AG2*. E em seguida, nas figuras 57, 58 e 59 pode-se ver os gráfico com as taxas de mutação e cruzamento ótimas para cada carteira.

A partir dos gráficos das figuras 57, 58 e 59, pode-se ver que as taxas ótimas obtidas pelo *AG2* foram: para a carteira 1 - 10% para mutação e 60% para cruzamento, para a carteira 2 - 50% para

Ações	Variação(x)	$w_{3i} * 100.000 * x$	$w_{3j} * 100.000 * x$
Brasken	42%	R\$6.157, 71	R\$6.914, 37
Celesc	88%	R\$8.537, 06	R\$8.440, 74
Comgás	87%	R\$11.006, 44	R\$11.362, 61
Confab	84%	R\$7.427, 53	R\$714, 99
Brasil Telecom	113%	R\$6.485, 89	R\$2.312, 36
Fosfertil	62%	R\$8.099, 15	R\$3.047, 33
Light S.A.	96%	R\$640, 99	R\$7.892, 80
Lojas Americanas	43%	R\$162, 15	R\$396, 84
NET	68%	R\$9.244, 14	R\$2.991, 15
Suzano Papel	45%	R\$4.803, 17	R\$5.488, 69
Telesp	114%	R\$11.290, 55	R\$10.227, 78
TIM	51%	R\$5.434, 28	R\$12.362, 46
TOTAL		R\$73.864, 90	R\$72.152, 12

Tabela 19: Variação do investimento para a carteira 3 na *FASE1* e *FASE2*

	Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
		Mut 10%-Cruz 50%		Mut 80%-Cruz 20%		Mut 80%-Cruz 20%
Capital Inicial	R\$100.000	R\$100.000	R\$100.000	R\$100.000	R\$100.000	R\$100.000
Capital Final	R\$57.863, 58	R\$70.875, 69	R\$78.798, 05	R\$83.733, 42	R\$73.864, 90	R\$72.152, 12
Varição (%)	-42%	-29%	-21%	-16%	-26%	-28%

Tabela 20: Resumo dos resultados do *AG1*, *FASE1* e *FASE2*

	Carteira 1	Carteira 2	Carteira 3
\bar{R}_i	1, 39%	1, 89%	-0, 05%
σ_i	2, 77%	2, 52%	1, 65%
\bar{R}_f	0, 68%	0, 59%	-0, 57%
σ_f	0, 02%	3, 42%	16, 03%

Tabela 21: Retorno e Risco pelo *AG1*, na *FASE1* e *FASE2*

Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
Ações	Peso inicial(w_{1i})	Ações	Peso inicial(w_{2i})	Ações	Peso inicial(w_{3i})
Ambev	0%	ALL	0%	Brasken	2, 85%
Banco do Brasil	0%	Bradespar	0%	Celesc	4, 84%
Bradesco	0%	Brasil T Par	7, 47%	Comgás	0, 41%
Cemig	64, 91%	CC Rodovias	15, 03%	Confab	5, 36%
Embraer	8, 55%	Eletróbrás	0%	Brasil Telecom	7, 66%
Gerdau	0%	Gerdau Met	26, 13%	Fosfertil	2, 73%
Itaú Unibanco	0%	Pão de Açúcar	32, 79%	Light S.A.	9, 51%
Itausa	0%	Sabesp	0%	Lojas Americanas	2, 92%
Petrobrás	26, 65%	Tractebel	0%	NET	6, 09%
Telemar	0%	Trans Paulista	0%	Suzano Papel	16, 90%
Usiminas	0%	Ultrapar	18, 87%	Telesp	5, 70%
Vale	0%	Vivo	0%	TIM	35, 04%

Tabela 22: Tabela dos pesos: *AG2*, *FASE1*

mutação e 70% para cruzamento e para a carteira 3 - 70% para mutação e 50% para cruzamento. O próximo passo, *FASE2*, consiste em rodar o *AG2* (programa *VaRAG*) novamente com essas taxas ótimas. Os pesos obtidos estão na tabela 23.

Nas figuras 60, 61 e 62 pode-se ver a avaliação média, a avaliação do melhor e do pior indivíduo (conjunto de pesos) e a distância média entre eles para a *FASE1* e *FASE2* do *AG2*.

A informação introduzida pela maior taxa de mutação é confirmada, como nos casos anteriores. Como explicado anteriormente, essa informação gera uma oscilação na distância média entre os in-

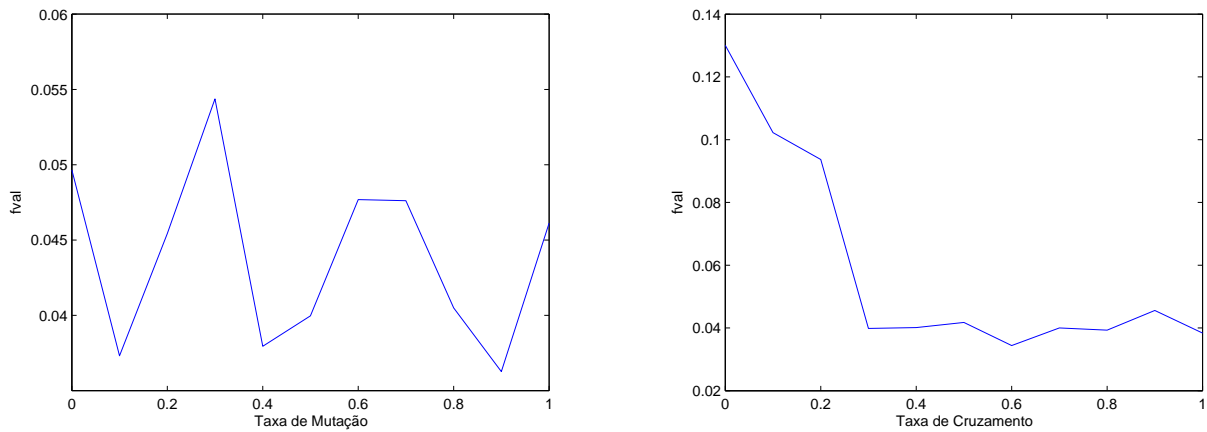


Figura 57: Taxas de Mutação e Cruzamento - Carteira 1 - AG2

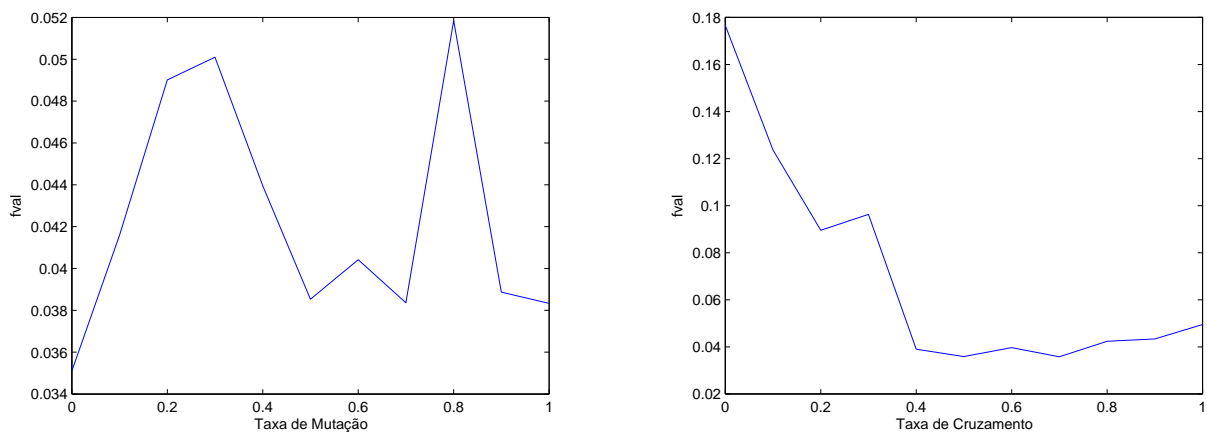


Figura 58: Taxas de Mutação e Cruzamento - Carteira 2 - AG2

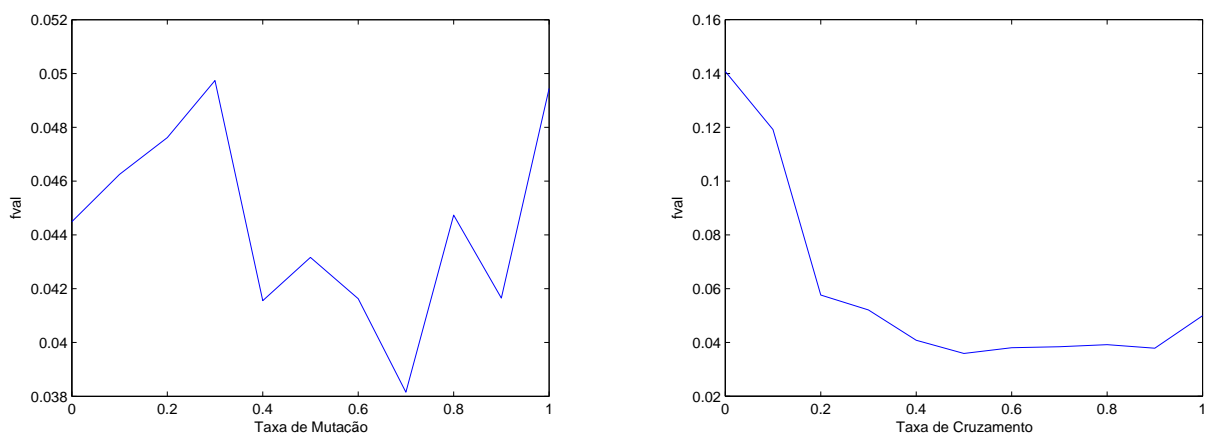


Figura 59: Taxas de Mutação e Cruzamento - Carteira 3 - AG2

divíduos que pode ser interpretada também como oscilação na diversidade da população, onde uma maior diversidade representa mais possibilidades de investimentos na carteira em questão.

Da mesma forma como foi feito para o AG1, foi suposto um investimento de R\$100.000 em cada uma das carteiras e esse investimento foi aplicado por um ano, de 02/01/2008 a 02/01/2009. Os

Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
	Mut 10%-Cruz 60%		Mut 50%-Cruz 70%		Mut 70%-Cruz 50%
Ações	Peso $FASE2 (w_{1j})$	Ações	Peso $FASE2 (w_{2j})$	Ações	Peso $FASE2 (w_{3j})$
Ambev	0%	ALL	0,55%	Brasken	12,23%
Banco do Brasil	0%	Bradespar	0,39%	Celesc	3,04%
Bradesco	16,03%	Brasil T Par	10,68%	Comgás	2,62%
Cemig	54,06%	CC Rodovias	18,94%	Confab	5,03%
Embraer	25%	Eletrobrás	20,07%	Brasil Telecom	11,93%
Gerdau	0%	Gerdau Met	3,84%	Fosfertil	8,41%
Itaú Unibanco	0%	Pão de Açúcar	9%	Light S.A.	12,82%
Itausa	0%	Sabesp	12,03%	Lojas Americanas	8,84%
Petrobrás	0,10%	Tractebel	2,21%	NET	1,01%
Telemar	0,39%	Trans Paulista	8,90%	Suzano Papel	12,47%
Usiminas	4,52%	Ultrapar	2,92%	Telesp	6,76%
Vale	0%	Vivo	10,43%	TIM	14,84%

Tabela 23: Tabela dos pesos: $AG2$, $FASE2$

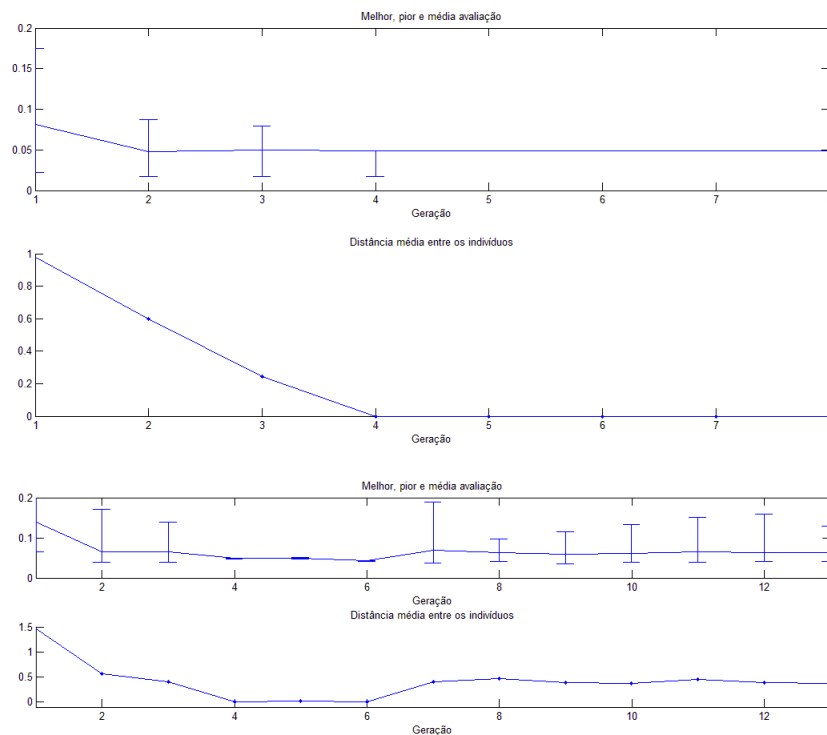


Figura 60: Avaliação e Distância média entre os indivíduos - Carteira 1 - $FASE1$ (superior) e $FASE2$ (inferior)

mesmos cálculos foram feitos para as três carteiras e na tabela 24 mostra-se os resultados obtidos.

Novamente foi obtido uma perda de capital ao deixar o dinheiro investido por um ano nas carteiras propostas. Porém, é importante ressaltar que apostando no $AG2$ e investindo nessas carteiras, a perda teria sido inferior a perda do mercado (BOVESPA).

Ressaltando, mais uma vez, a necessidade de levar em conta os retornos e riscos. A tabela 25 mostra os retornos esperados das carteiras 1, 2 e 3 juntamente com seus respectivos desvios padrão (riscos), fornecidos pela função VaR do programa. Esses dados foram calculados através das fórmulas da seção 7 e com o auxílio do *Excel*. As notações utilizadas serão as mesmas do $AG1$.

Para fazer comparações entre as carteiras utilizando os resultados obtidos dos riscos e dos retornos esperados, precisa-se introduzir uma medida de dispersão relativa, muito útil na comparação, cujos retornos esperados das carteiras são diferentes, chamada de *coeficiente de variação*. De acordo com Gitman [21] essa medida é definida como:

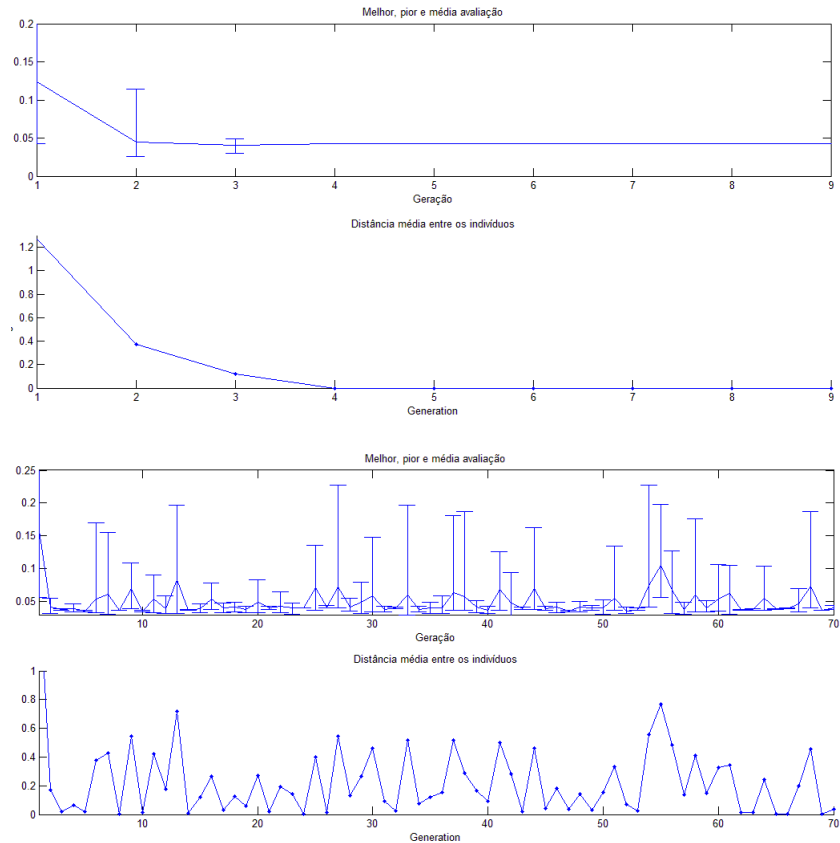


Figura 61: Avaliação e Distância média entre os indivíduos - Carteira 2 - *FASE1* (superior) e *FASE2* (inferior)

	Carteira 1	Carteira 1	Carteira 2	Carteira 2	Carteira 3	Carteira 3
		Mut 10%-Cruz 60%		Mut 50%-Cruz 70%		Mut 70%-Cruz 50%
Capital Inicial	R\$100.000	R\$100.000	R\$100.000	R\$100.000	R\$100.000	R\$100.000
Capital Final	R\$91.282, 91	R\$87.139, 36	R\$86.059, 91	R\$93.619, 61	R\$66.912, 84	R\$70.532, 46
Variação (%)	-9%	-13%	-14%	-6%	-33%	-29%

Tabela 24: Resumo dos resultados do *AG2*, *FASE1* e *FASE2*

$$CV = \frac{\sigma}{R}.$$

Quanto maior o coeficiente de variação, maior o risco. É importante destacar que o coeficiente de variação não pode ser negativo, por isso deve-se usar os valores absolutos dos retornos.

Com isso, tem-se os coeficientes de variação, da *FASE1* e *FASE2*, das três carteiras:

- Para o *AG1*:

Pode-se concluir que, para o *AG1*, utilizando os pesos da *FASE1*, a carteira 3 é a mais arriscada de todas e a carteira 2 é a menos arriscada. Por outro lado, utilizando os pesos da *FASE2*, a carteira 3 é a mais arriscada de todas e a carteira 1 é a menos arriscada.

- Para o *AG2*:

No caso do *AG2*, utilizando os pesos da *FASE1*, a carteira 3 é a mais arriscada de todas e a carteira 1 é a menos arriscada. Por outro lado, utilizando os pesos da *FASE2*, a carteira 1 é a mais arriscada de todas e a carteira 3 é a menos arriscada.

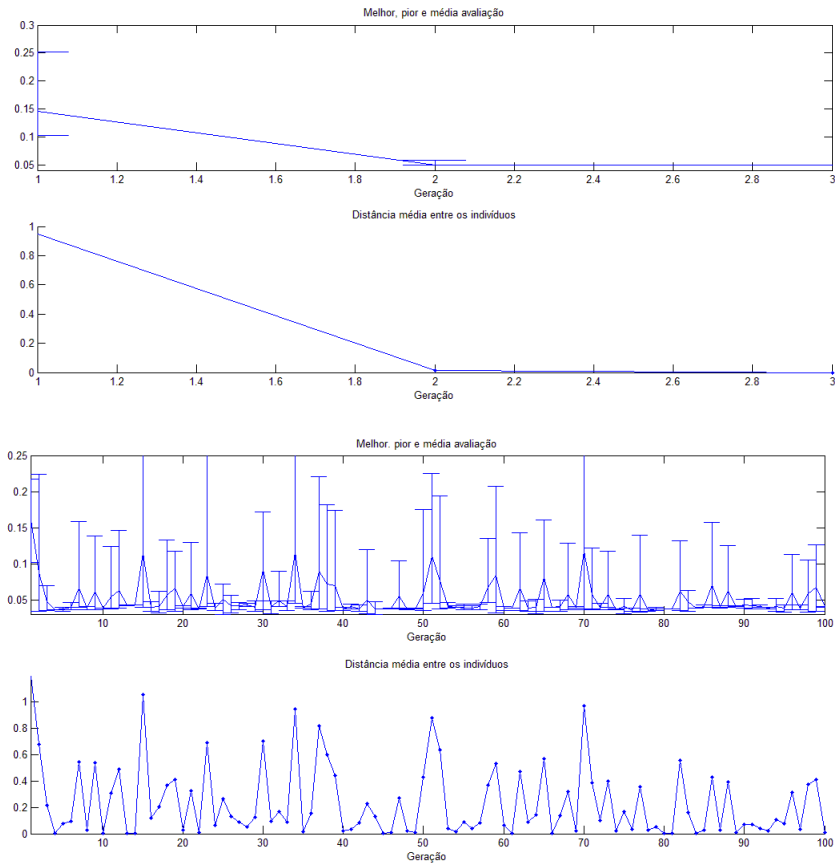


Figura 62: Avaliação e Distância média entre os indivíduos - Carteira 3 - *FASE1* (superior) e *FASE2* (inferior)

	Carteira 1	Carteira 2	Carteira 3
\overline{R}_i	0,09%	0,07%	0,06%
σ_i	4,93%	4,25%	4,95%
\overline{R}_f	0,07%	0,07%	0,08%
σ_f	4,21%	4,10%	3,98%

Tabela 25: Retorno e Risco pelo *AG2*, na *FASE1* e *FASE2*

9 Considerações Finais

O Algoritmo Genético (AG) tem o propósito de analisar e modificar uma série de soluções quantitativamente diferentes, objetivando a otimização de sua função de aptidão, mesmo quando as técnicas tradicionais de otimização não se aplicam. Por isso foi escolhido como ferramenta para analisar o problema: “Qual ativo escolho para investimento?”, referindo a carteiras de ações compostas através do índice IBrX da BOVESPA. Nesse contexto, tem-se um grande número de ações, 36 ações no total e, para cada ação, tem-se uma série histórica de 6 anos de seus valores diários. Antes da codificação do problema e da construção do AG, foi fundamental uma introdução à teoria dos Processos Estocásticos, especificamente aos modelos ARIMA e GARCH, que permitiram analisar e compreender certas medidas e comportamentos dessa imensa massa de dados que foi utilizada. Depois, para a construção de uma função de aptidão satisfatória do AG, foi preciso estudar a chamada análise de risco, na área de otimização de carteiras, analisando as metodologias sobre retorno versus risco. Foram construídos dois modelos de AG para avaliar o investimento em três carteiras, explorando as taxas de cruzamento e de mutação em cada caso, verificando as diferenças entre populações com baixa e alta diversidade e

	Carteira 1	Carteira 2	Carteira 3
<i>CV FASE1</i>	1,99	1,33	33,4
<i>CV FASE2</i>	0,03	5,80	28,12

Tabela 26: Coeficientes de Variação AG1

	Carteira 1	Carteira 2	Carteira 3
<i>CV FASE1</i>	54,78	60,71	82,5
<i>CV FASE2</i>	60,14	58,57	49,75

Tabela 27: Coeficientes de Variação AG2

suas conseqüências na convergência dos AGs.

A partir das aplicações realizadas e dos resultados obtidos, foi possível observar que as metodologias empregadas na construção das funções de aptidão e suas restrições, bem como o estudo das taxas de cruzamento e de mutação, mais os critérios de parada, foram todos satisfatórios, o que enriqueceu muito o conhecimento dentro da área de modelagem matemática e computacional aplicada ao mercado financeiro brasileiro. No primeiro AG desenvolvido (**AG1**) aplicado às três carteiras, a função de aptidão é a função dos retornos e otimizá-la é obter seu valor máximo. Já no segundo AG aplicado às três séries (**AG2**), foi usado como função de aptidão a função VaR, e otimizá-la é obter seu valor mínimo. Foi estudada a consistência dessas duas abordagens frente aos parâmetros aplicados e utilizada duas fases para cada AG. A primeira fase (**FASE1**), consiste na execução do AG com taxas de cruzamento e mutação padrões, definidas pela literatura. Em seguida, obteve-se as melhores taxas de cruzamento e de mutação, através do processo de varredura e observadas nos gráficos das aplicações, em seguida passa-se para a segunda fase (**FASE2**), em que foi executado cada AG com essas taxas ótimas.

Ao observar-se as tabelas dos resultados das aplicações, iniciando pela aplicação do *AG1*, pode ser visto que na *FASE1* a carteira que teve a menor perda foi a carteira 2, com -21% , contra a carteira 3 que obteve a maior perda, com -26% . É interessante lembrar que a carteira 2 era composta por ações com pesos intermediários no IBrX e a carteira 3 era composta por ações com os menores pesos no IBrX. A partir do cálculo do coeficiente de variação, pode ser visto que a carteira 2, além de ter a menor perda, é a menos arriscada das três, o que não é o caso da mais arriscada, carteira 3, que perde para a carteira 1 com -42% . É interessante ressaltar, que ao comparar as perdas das carteiras com o mercado (BOVESPA), pode ser concluído que foi a carteira 1 que obteve o pior resultado, pois o mercado nesse período teve uma perda de 35% .

Por outro lado, ainda aplicando o *AG1*, mas na *FASE2*, ou seja, depois de utilizar taxas de cruzamento e mutação ótimas nos AGs, pode ser visto que a carteira que obteve a menor perda foi a carteira 2, com -16% , contra a carteira 1 que obteve a maior perda, com -29% . Observa-se agora que a mais arriscada das três é a carteira 3 e a menos arriscada é a carteira 1. Pode-se concluir que a carteira 3 foi a mais arriscada na *FASE1* e *FASE2*, apontando uma certa consistência dos resultados dos AGs.

Observando os resultados obtidos pelo *AG2*, pode ser visto que a carteira 3 continua sendo a mais arriscada na *FASE1*. Ela obteve a maior perda, com 25% . Observa-se que, desta vez, a carteira que foi menos arriscada também obteve a menor perda das três carteiras na *FASE1*, ou seja, a carteira 1. Analisando agora as perdas e os riscos na *FASE2*, pode-se ver que a mais arriscada foi a carteira 1, porém a maior perda foi da carteira 3, a menos arriscada.

O que pode-se concluir a partir das análises realizadas e dos resultados obtidos é que o *AG2* obteve um resultado melhor. Comparando as perdas, pode-se ver que em todos os casos, exceto na carteira 3, na *FASE2*, o *AG2* sempre obteve os melhores resultados. Incluindo agora as carteiras mais ou menos arriscadas, confirmando a observação ao analisar a carteira 3, que no *AG1* além de ser a mais arriscada na *FASE1* e *FASE2*, obteve uma das maiores perdas. Após a aplicação do *AG2* na *FASE2*, ela deixou de ser a mais arriscada e ainda sua perda foi menor do que a calculada pelo *AG1*.

Pode-se afirmar que o *AG2* (cuja função de aptidão é função VaR) é o mais eficiente, minimizando o risco das carterias. Se fossem utilizados os resultados do *AG2* para investir nas três carteiras estudadas durante o período em questão, os resultados teriam sido melhores do que o mercado. E ainda, pode-se concluir que a melhor carteira de todas é a carteira 1, pois a partir do coeficiente de variação calculado, a carteira 1 obteve o menor resultado, implicando em um menor risco, ou seja, em resposta a pergunta inicial deste trabalho, a carteira 1 é a escolhida.

Em trabalho futuro, pretende-se avaliar o comportamento dos AGs, frente à mudança da frequência das séries históricas. Já se sabe, pela teoria estudada, que o modelo GARCH (utilizado para análise das séries financeiras) é fechado para a agregação temporal, isto é, a mudança de frequência de um modelo GARCH, respeitando-se certas condições, é um novo modelo GARCH, com parâmetros determinados pelos parâmetros do modelo original. E os AGs, como eles se comportam com a mudança da série histórica diária para quinzenal ou mensal? Pode-se esperar que a solução ótima (mínimo/máximo global) de um modelo com frequência mensal seja uma solução de mínimo/máximo local da série histórica original, com frequência diária? É uma questão que pretende-se investigar.

Também pretende-se investigar a combinação das funções de aptidão dos *AG1* e *AG2*, o chamado Algoritmo Genético Multi-Objetivo, onde num mesmo AG, deseja-se ao mesmo tempo, maximizar o retorno e minimizar o risco.

10 Referências Bibliográficas

- [1] A.B. Schimdt. *Quantitative Finance fo Physicists: An introduction*. Elsevier Academic Press, 2005.
- [2] D.N. Gujarati. *Econometria Básica*. Elsevier, 2000.
- [3] K.L. Chung. *Elementary probability theory with stochastic processes and an introduction to mathematical finance*. Massachusetts Institute of Technology, 2000.
- [4] J.N. Tsitsiklis P.B. Dimitri. *Introduction to Probability*. Massachusetts Institute of Technology, 2000.
- [5] B.R. James. *Probabilidade: um curso em nível intermediário*. IMPA, 2006.
- [6] M.L. ; Stephan D. Levine, D.M. ; Berenson. *Estatística: Teoria e Aplicações*. LTC, 2000.
- [7] D. Stirzaker. *Elementary Probability*. Cambridge, 2003.
- [8] E. Lebensztayn C.F. Coletti. *Probabilidade: Teoria e Exercícios*. IME-USP, 2008.
- [9] S.M. Ross. *Introduction to Probability Models*. ELSEVIER, 2002.
- [10] C.M.C. Toloí P.A. Morettin. *Análise de Séries Temporais*. Blucher, 2006.
- [11] Z. Brezezniak ; T. Zastawniak. *Basic Stochastic Processes*. Springer, 2007.
- [12] P. J. Diggle. *Time Series. A Biostatistical Introduction*. Oxford Science Publications, 2004.
- [13] P.A. Morettin. *Econometria Financeira: Um curso em Séries Temporais*. Blucher, 2008.
- [14] J. van den Berg V.A.F. Dallagnol and L. Mous. Portfolio management using value at risk: A comparison between genetic algorithms and particle swarm optimization. *International Journal of Intelligent Systems*, 24:766–792, 2009.
- [15] D. Whitley. A genetic algorithm tutorial. *Computer Science Department*, 44:93–103, 1993.
- [16] R.R. Martin D. Beasley, D.R. Bull. An overview of genetic algorithm: Part 1, fundamentals. *University Computing*, 15:58–69, 1993.
- [17] D.Z. Zheng L. Wang, L. Zhang. A class of hypothesis-test-based genetic algorithms for flow shop scheduling with stochastic processing time. *Springer-Verlag London Limited*, 25:1157–1163, 2005.
- [18] A.Assaf. *Mercado Financeiro*. Atlas, 2005.
- [19] M.G. Speranza R. Mansini. Heuristic algorithms for the portifolio selection problem with minimum transaction lots. *European Journal of Operational Research*, 114:219–233, 1999.
- [20] S.C. Myers R.A. Brealey. *Principles of Corporate Finance*. McGraw-Hill, 2003.
- [21] L.J. Gitman. *Princípios de Administração Financeira*. Pearson Addison Wesley, 2003.
- [22] Manfred Gilli and Evis Kéllezi. A heuristic approach to portfolio optimization. Fame research paper series, International Center for Financial Asset Management and Engineering, 2000.
- [23] H.M. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952.
- [24] C.O. Ribeiro; L.A.S. Ferreira. Uma contribuição ao problema de composição de carteiras de mínimo valor em risco. *Gestão & Produção*, 12:295–304, 2005.

- [25] P. Jorion. *Value at Risk*. McGraw-Hill, 2001.
- [26] N.D. Pearson. *Risk Budgeting. Portfolio Problem Solving with Value-at-Risk*. John Wiley & Sons, Inc., 2002.
- [27] J.E. Beasley Y.M Sharaiha T.J. Chang, N. Meade. Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.*, 27:1271–1302, 2000.
- [28] T.E. Nijman F.C. Drost. Temporal aggregation of garch processes. *Econometrica*, 61:909–927, 1993.
- [29] W.A. Fuller D.A. Dickey. Distribution of the estimators for autoregressive time series with a unit root. *Journal of Statistical Physics*, 44:67–93, 1986.
- [30] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

11 Apêndices

A Método dos Mínimos Quadrados (MMQ)

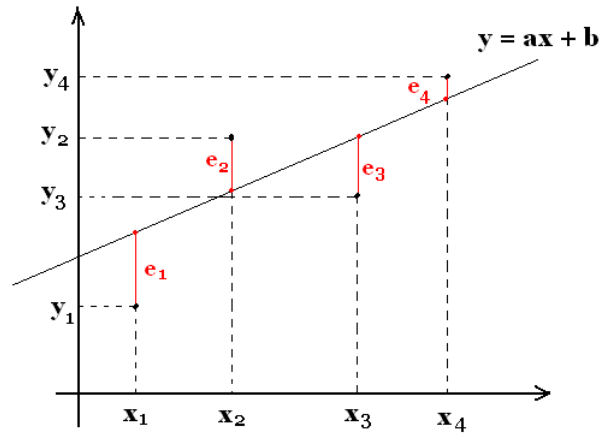


Figura 63: MQQ

Seja S a soma dos quadrados dos erros encontrados da figura 63.

$$S = e_1^2 + e_2^2 + e_3^2 + e_4^2$$

O Método dos Mínimos quadrados consiste em encontrar os coeficientes da reta $y = ax + b$, de tal forma que, (a, b) minimizem S .

$$e_1^2 = [(ax_1 + b) - y_1]^2, \quad \dots, \quad e_4^2 = [(ax_4 + b) - y_4]^2.$$

Como o objetivo é encontrar os coeficientes que minimizam S , deve-se calcular as derivadas parciais do erro e em relação a a e b e igualá-las a zero.

$$\begin{cases} \frac{\partial S}{\partial a} = 2[(ax_1 + b) - y_1]x_1 + \dots + 2[(ax_4 + b) - y_4]x_4 \\ \frac{\partial S}{\partial b} = 2[(ax_1 + b) - y_1] + \dots + 2[(ax_4 + b) - y_4] \end{cases}$$
$$\begin{cases} \frac{\partial S}{\partial a} = 0 \Rightarrow \sum_{i=1}^4 ax_i^2 + bx_i - x_i y_i = 0 \\ \frac{\partial S}{\partial b} = 0 \Rightarrow \sum_{i=1}^4 ax_i + b - y_i = 0 \end{cases}$$

Por fim, tem-se as **Equações Normais**, que vão nos auxiliar a determinar os coeficientes a e b .

$$\begin{cases} a \sum_{i=1}^4 x_i^2 + b \sum_{i=1}^4 x_i = \sum_{i=1}^4 x_i y_i \\ a \sum_{i=1}^4 x_i + b \sum_{i=1}^4 1 = \sum_{i=1}^4 y_i \end{cases}$$

Exemplo:

Sejam $x = [1,2,3,4]$ e $y = [1,3,2,4]$. Calcula-se os coeficientes da reta $y = ax + b$ que é aproximada por esses pontos. Utilizando as Equações Normais.

$$\begin{cases} 30a + 10b = 29 \\ 10a + 4b = 10 \end{cases}$$

Resolvendo o sistema tem-se que $a = \frac{4}{5}$ e $b = \frac{1}{2}$. Logo, a reta que melhor aproxima esses pontos é $y = \frac{4}{5}x + \frac{1}{2}$.

É importante observar que a equação que melhor aproxima os pontos nem sempre é uma reta, ou seja, pode ser não-linear.

B Máxima Verossimilhança

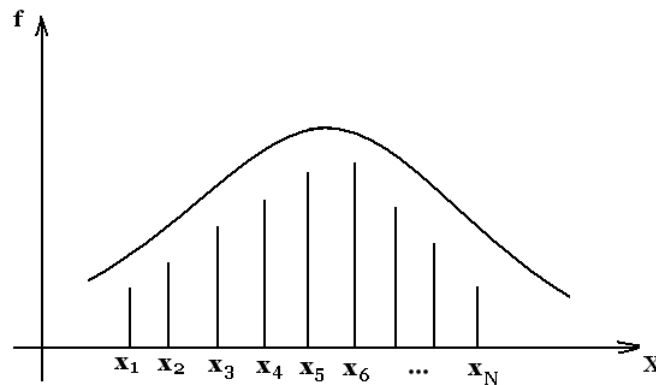


Figura 64: Máxima Verossimilhança

Tome os pontos $\{x_1, \dots, x_n\}$ e a distribuição $P(x, \theta)$ vetorial. A partir disso, deve-se encontrar a **Função de Verossimilhança**:

$$f_{\theta}(x_i) = P(x_i|\theta).$$

Em termos de logaritmo, tem-se:

$$L(\theta) = \sum_{i=1}^N \log(f_{\theta}(x_i)).$$

Agora, deve-se encontrar θ que maximiza $L(\theta)$.

Exemplo: Ruído Branco Gaussiano

Sejam $\vec{x} = \{x_1, \dots, x_n\}$ com $\vec{x} \sim N(\mu, \sigma^2)$. Deve-se encontrar $\theta = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$.

Como é uma distribuição normal, tem-se que a função de verossimilhança será:

$$f_{\theta}(x) = P(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Em termos de logaritmo:

$$L(\theta) = \sum_{i=1}^N \log(f_{\theta}(x_i)).$$

$$\begin{cases} \frac{\partial L(\theta)}{\partial \mu} = \sum_{i=1}^N \frac{\partial}{\partial \mu} \log(f_{\theta}(x_i)) \\ \frac{\partial L(\theta)}{\partial \sigma^2} = \sum_{i=1}^N \frac{\partial}{\partial \sigma^2} \log(f_{\theta}(x_i)) \end{cases}$$

Resolvendo para μ :

$$\frac{\partial}{\partial \mu} \left[\log \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right] = \frac{\partial}{\partial \mu} \left[\log \frac{1}{\sigma\sqrt{2\pi}} - \frac{(x_i-\mu)^2}{2\sigma^2} \right] = \frac{\partial}{\partial \mu} \left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right) = -\frac{2(x_i-\mu)}{2\sigma^2} (-1)$$

Portanto:

$$\frac{\partial L(\theta)}{\partial \mu} = \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2}.$$

Resolvendo agora para σ^2 :

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \left[\log \frac{1}{\sigma \sqrt{2\pi}} - \frac{(x_i - \mu)^2}{2\sigma^2} \right] &= \frac{\partial}{\partial \sigma^2} \left[\log 1 - \log(\sigma \sqrt{2\pi}) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] \stackrel{\mu=\sigma^2}{=} \frac{\partial}{\partial \mu} \left[-\log(\sqrt{\mu} \sqrt{2\pi}) - \frac{(x_i - \mu)^2}{2\mu} \right] = \\ &= -\frac{1}{\sqrt{\mu} \sqrt{2\pi}} \sqrt{2\pi} \frac{1}{2} \mu^{-\frac{1}{2}} + \frac{(x_i - \mu)^2}{2\mu} \mu^{-2} = -\frac{1}{2\mu} + \frac{(x_i - \mu)^2}{2} \frac{\partial}{\partial \sigma^2} \log(f_{\theta}(x_i)) = -\frac{1}{2\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4}. \end{aligned}$$

Portanto:

$$\frac{\partial L(\theta)}{\partial \sigma^2} = \sum_{i=1}^N -\frac{1}{2\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4}.$$

Como o objetivo é encontrar θ que maximiza $L(\theta)$, tem-se que igualar essas derivadas a zero.

Primeiramente para μ :

$$\frac{\partial L(\theta)}{\partial \mu} = 0 \quad \Rightarrow \quad \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2} = 0.$$

Após alguns cálculos:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i.$$

E agora, para σ^2 :

$$\frac{\partial L(\theta)}{\partial \sigma^2} = 0 \quad \Rightarrow \quad \sum_{i=1}^N -\frac{1}{2\sigma^2} + \frac{(x_i - \mu)^2}{2\sigma^4} = 0.$$

Após alguns cálculos:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2.$$

C Teste de Dickey - Fuller

Dickey e Fuller [29] desenvolveram um teste para verificar a existência de raiz unitária em uma série temporal. O procedimento para a realização desse teste consiste em regredir a série temporal contra seus valores defasados de uma unidade, em seguida testa-se a significância estatística do coeficiente α_1 estimado por mínimos quadrados, utilizando um teste de hipóteses.

As hipóteses estabelecidas podem ser de dois tipos: *hipótese nula* (H_0), quando se admite não haver diferença entre a informação fornecida pela evidência amostral e a afirmação da hipótese ou *hipótese alternativa* (H_1), quando se admite haver diferença entre a informação fornecida pela evidência amostral e a afirmação da hipótese estabelecida para a realização do teste. Assim, o processo do teste de hipóteses consiste em rejeitar ou não a hipótese nula H_0 .

O teste da raiz unitária consiste em dizer se a série temporal é estacionária ou não, pois para ser não-estacionária a série temporal deve ter raiz unitária, ou seja, $|\alpha_1| = 1$. As hipóteses do teste da raiz unitária podem ser formuladas da seguinte forma:

$$\begin{cases} H_0 : |\alpha_1| = 1 \\ H_1 : |\alpha_1| < 1 \end{cases}$$

Isso significa que, para que a série temporal seja não-estacionária deve-se aceitar a hipótese nula.

Para apresentar o processo da raiz unitária, esta seção será baseada na obra de James Hamilton [30]. Assim, considera-se a seguinte série temporal:

$$y_t = \mu + \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots = \mu + \psi(L)\epsilon_t, \quad (63)$$

onde $\sum_{j=0}^{\infty} |\psi_j| < \infty$, as raízes de $\psi(L) = 0$ estão fora do círculo unitário, e $\{\epsilon_t\}$ é um ruído branco com média zero e variância σ^2 .

Considere agora o processo da *raiz unitária*:

$$(1 - L)y_t = \gamma + \psi(L)\epsilon_t, \quad (64)$$

onde $\psi(1) \neq 0$. Para o processo da raiz unitária, a representação estacionária da equação 63 descreve algumas mudanças na série. Considere também que a média de $(1 - L)y_t$ seja denotada por γ ao invés de μ .

Denota-se o operador da primeira diferença $(1 - L)$ por Δ , então:

$$\Delta y_t \equiv y_t - y_{t-1}.$$

Suponha que $\psi(1) = 1$ na equação 64. Então:

$$y_t = y_{t-1} + \gamma + \epsilon_t.$$

Este processo é conhecido como caminho aleatório com "drift" γ .

Na definição do processo de raiz unitária 64, foi assumido que $\psi(1) \neq 0$, onde $\psi(1)$ denota o polinômio 65 para $z = 1$:

$$\psi(z) = 1 + \psi_1 z^1 + \psi_2 z^2 + \dots \quad (65)$$

Para verificar o porque essa restrição deve fazer parte da definição, supõe-se que a série original y_t é de fato estacionária, representada da seguinte forma:

$$y_t = \mu + \chi(L)\epsilon_t.$$

Diferenciando esta série:

$$(1 - L)y_t = (1 - L)\chi(L)\epsilon_t \equiv \psi(L)\epsilon_t,$$

onde $\psi(L) \equiv (1 - L)\chi(L)$. Isso diz que, se a série original é estacionária, ao diferenciar uma vez ela continuará sendo. No entanto, o operador de médias móveis $\psi(L)$ que caracteriza Δy_t tem uma

propriedade que $\psi(1) = (1 - 1)\chi(1) = 0$. Quando estipula-se que $\psi(1) \neq 0$ na equação 64, descarta-se a possibilidade de que a série original seja estacionária.

Algumas vezes é conveniente trabalhar com uma representação um pouco diferente do processo da raiz unitária 64. Considere então:

$$y_t = \alpha + \gamma_t + \mu_t, \quad (66)$$

onde μ_t segue um processo ARMA com média zero, isto é:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)\mu_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)\epsilon_t. \quad (67)$$

e o operador de média móvel $(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$ é invertível. Suponha agora que o operador auto-regressivo em 67 seja fatorado da seguinte forma:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)\dots(1 - \lambda_p L).$$

Se todos os $\lambda_1, \dots, \lambda_p$ estiverem dentro do círculo unitário, então a equação 67 pode ser escrita da seguinte forma:

$$\mu_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_1 L)(1 - \lambda_2 L)\dots(1 - \lambda_p L)}\epsilon_t \equiv \psi(L)\epsilon_t,$$

com $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e as raízes de $\psi(z) = 0$ estão fora do círculo unitário.

Suponha agora que $\lambda_1 = 1$, ou seja, $\lambda_1, \dots, \lambda_p$ estão no círculo unitário. Então a equação 67 poderá ser escrita como:

$$(1 - L)\mu_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_2 L)(1 - \lambda_3 L)\dots(1 - \lambda_p L)}\epsilon_t \equiv \psi^*(L)\epsilon_t,$$

com $\sum_{j=0}^{\infty} |\psi_j^*| < \infty$ e as raízes de $\psi^*(z) = 0$ estão fora do círculo unitário. Então, se a equação 66 é a primeira diferença, o resultado será:

$$(1 - L)y_t = (1 - L)\alpha + [\gamma_t - \gamma(t - 1)] + (1 - L)\mu_t = 0 + \gamma + \psi^*(L)\epsilon_t,$$

o qual está fora da forma do processo de raiz unitária 64.

A representação na equação 66 explica o uso do termo "processo de raiz unitária". Uma das raízes ou dos valores de (λ_1) do polinômio auto-regressivo 65 é unitária, e todos os outros estão dentro do círculo unitário.

D Média e Variância

Média

```
function m = media (x)
% Cálculo da média de uma série
%acumula os valores para o cálculo da média da série original
m=0;
for (i=1:length(x))
    m=m+x(i);
end
m=m/length(x);
```

Variância

```
function v = var(x)
% Cálculo da variância de uma série
% acumula os valores para o cálculo da variância da série original
mi=media(x);
v=0;
for (i=1:length(x))
    v=v+(x(i)-mi)^2;
end
v=v/length(x);
```

E FAC E FACP

FAC - Função de Autocorrelação

```
function f = fac(x,k)
% Calcula a auto-correlação de ordem k de uma série
% obs: fac(x,0)=1 para qualquer série
if ((k>=0)&&(k<length(x)))
    mi=media(x); % calcula a média da série
    sigma2=var(x); % calcula a variância da série
    f=0; % zera a auto-correlação de ordem k

    if (sigma2~=0)
        for i=k+1:length(x)
            f=f+(x(i)-mi)*(x(i-k)-mi);
        end
        %calcula a auto-covariância de ordem k
        f=f/(length(x)-k);
        %calcula a auto-correlação de ordem k
        f=f/sigma2;
    else
        display('Variância zero');
    end
else
    display('Não é possível calcular auto-correlação dessa ordem');
end
```

FACP - Função de Autocorrelação Parcial

```
function fp = facp(x,k)
% Calcula da autocorrelação parcial
% Utiliza a relação matricial Yule-Walker:  $T_{ij} * \phi_i = \rho_i$ 
% T: matriz de auto-correlação;
%phi: vetor de auto-correlação parcial;
%rho: vetor de auto-correlação
% Utiliza também o algoritmo de Levinson-Durbin para resolução de sistema
% matricial Toeplitz
%a autocorrelação parcial de ordem 0 não existe.
if ((k<1) || (k>=length(x)))
    display('Não é possível calcular auto-correlação dessa ordem');
else
    fp(1)=fac(x,1);
    %a autocorrelação parcial de ordem 1 é a autocorrelação de ordem 1
    T(1,1)=fac(x,1); % matriz de auto-correlação
    for ordem=2:k
        aux1=0;
        aux2=0;
        for i=1:ordem-1
            aux1=aux1+T(ordem-1,i)*fac(x,ordem-i);
            aux2=aux2+T(ordem-1,i)*fac(x,i);
        end
        T(ordem,ordem)=(fac(x,ordem)-aux1)/(1-aux2);
        fp(ordem)=T(ordem,ordem);
    end
end
```

```

        for i=1:ordem-1
            T(ordem,i)=T(ordem-1,i)-T(ordem,ordem)*T(ordem-1,ordem-i);
        end
    end
end
end

```

Exibe os gráficos da FAC e FACP

```

function [] = exhibi(x,k)
% Chama e exibe a auto-correlação e auto-correlação parcial de uma série
% Calcula a autocorrelação da série
for ordem=1:k
    f(ordem)=fac(x,ordem);
end
% Calcula a autocorrelação parcial da série
fp=facp(x,k);

% Variável independente: ordem das auto-correlações e auto-correlações
% parciais
c=1:k;

% Gera gráficos
subplot(1,2,1)
stem(c,f(c));
xlabel('k');
ylabel('FAC(k)');
title('Auto-correlação');

subplot(1,2,2)
stem(c,fp(c));
xlabel('k');
ylabel('FACP(k)');
title('Auto-correlação Parcial');

```

F Algoritmo Genético 1

Para fazer os testes do algoritmo genético utiliza-se uma carteira formada por seis séries, entre elas estão: Vivo, Petrobrás, Dolar, Comgás, Bom Bril e Banco do Brasil. Todos os dados dessas séries foram retirados do *site* da BOVESPA.

Retorno

Esse programa denominado *Retorno*, prevê os retornos dos ativos através da modelagem GARCH(1,1) e calcula a fronteira eficiente. Para isso carrega-se as seis séries no programa e logo em seguida exibem-se os gráficos da FAC dos retornos e a FAC dos retornos ao quadrado para verificar a presença de heterocedasticidade.

Feito isso, aplica-se o teste de Engle para confirmação da presença da heterocedasticidade. Feito o teste, o programa fornece os coeficientes do ARCH(1).

Uma vez encontrados os coeficientes, o modelo calcula os retornos das séries e exibe o último deles, ou seja, o retorno esperado. Para o cálculo da fronteira eficiente é preciso truncar essas séries para que elas tenham o mesmo tamanho. A partir daí, os dados são colocados em uma matriz e é calculada a sua covariância, ou seja, os riscos relacionados aos retornos, para que a fronteira eficiente seja traçada.

```
% Prevê retornos de ativos através da modelagem GARCH(1,1)
% Calcula a Fronteira Eficiente
x=load('bb.mat','-ascii');
y=load('bombril.mat','-ascii');
z=load('congas','-ascii');
t=load('dolar.mat','-ascii');
r=load('petrobras.mat','-ascii');
s=load('vivo.mat','-ascii');

xret=price2ret(x);
yret=price2ret(y);
zret=price2ret(z);
tret=price2ret(t);
rret=price2ret(r);
sret=price2ret(s);

% Exibi os gráficos da autocorrelação dos retornos

figure();
autocorr(xret)
title('Função Autocorrelação para a série de retornos de X');

figure();
autocorr(yret)
title('Função Autocorrelação para a série de retornos de Y');

figure();
autocorr(zret)
title('Função Autocorrelação para a série de retornos de Z');

figure();
autocorr(tret)
title('Função Autocorrelação para a série de retornos de T');
```

```

figure();
autocorr(rret)
title('Função Autocorrelação para a série de retornos de R');

figure();
autocorr(sret)
title('Função Autocorrelação para a série de retornos de S');

% Exibi os gráficos da autocorrelação dos quadrados dos retornos

figure();
autocorr(xret.^2)
title('Função Autocorrelação para a série de retornos ao quadrado de X');

figure();
autocorr(yret.^2)
title('Função Autocorrelação para a série de retornos ao quadrado de Y');

figure();
autocorr(zret.^2)
title('Função Autocorrelação para a série de retornos ao quadrado de Z');

figure();
autocorr(tret.^2)
title('Função Autocorrelação para a série de retornos ao quadrado de T');

figure();
autocorr(rret.^2)
title('Função Autocorrelação para a série de retornos de ao quadrado R');

figure();
autocorr(sret.^2)
title('Função Autocorrelação para a série de retornos de ao quadrado S');

% Modelagem ARCH(1)

% Teste de Engle para a série X
[H,pValue,Stat,CriticalValue] = archtest(xret-mean(xret),[10 15 20]',0.05);
[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(xret);
garchdisp(coeff,errors);

% Teste de Engle para a série Y
[H,pValue,Stat,CriticalValue] = archtest(yret-mean(yret),[10 15 20]',0.05);
[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(yret);
garchdisp(coeff,errors);

% Teste de Engle para a série Z
[H,pValue,Stat,CriticalValue] = archtest(zret-mean(zret),[10 15 20]',0.05);

```



```

[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(zret);
garchdisp(coeff,errors);

% Teste de Engle para a série T
[H,pValue,Stat,CriticalValue] = archtest(tret-mean(tret),[10 15 20]',0.05);
[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(tret);
garchdisp(coeff,errors);

% Teste de Engle para a série R
[H,pValue,Stat,CriticalValue] = archtest(rret-mean(rret),[10 15 20]',0.05);
[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(rret);
garchdisp(coeff,errors);

% Teste de Engle para a série S
[H,pValue,Stat,CriticalValue] = archtest(sret-mean(sret),[10 15 20]',0.05);
[H,pValue,Stat,CriticalValue]
% Estimar parâmetros
[coeff,errors,LLF,innovations,sigmas,summary] = garchfit(sret);
garchdisp(coeff,errors);

% Simulação
% X
specx = garchset('C',0.0015003,'K',2.577e-005,'GARCH',0.90222,'ARCH',
0.070082,'Display','off');
[xepsilon,xsigma,xsim] = garchsim(specx,length(x)+1,1);
% exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série X: %.4f\n\n',xsim(end))

%Y
specy = garchset('C',-0.00075309,'K',0.00022158,'GARCH',0.64407,'ARCH',
0.19512,'Display','off');
[yepsilon,ysigma,ysim] = garchsim(specy,length(y)+1,1);
%exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série Y: %.4f\n\n',ysim(end))

%Z
specz = garchset('C',0.00010659,'K',4.2494e-006,'GARCH',0.94773,'ARCH',
0.049673,'Display','off');
[zepsilon,zsigma,zsim] = garchsim(specz,length(z)+1,1);
%exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série Z: %.4f\n\n',zsim(end))

%T
spect = garchset('C',-1.0872e-005,'K',4.0648e-006,'GARCH',0.7996,'ARCH',
0.15752,'Display','off');
[tepsilon,tsigma,tsim] = garchsim(spect,length(t)+1,1);

```

```

% exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série T: %.4f\n\n',tsim(end))

%R
specr = garchset('C',0.0018138,'K',1.5521e-005,'GARCH',0.88144,'ARCH',
                0.095857,'Display','off');
[repsilon,rsigma,rsim] = garchsim(specr,length(r)+1,1);
% exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série R: %.4f\n\n',rsim(end))

%S
specs = garchset('C',-0.00013648,'K',4.9281e-005,'GARCH',0.90105,'ARCH',
                0.052836,'Display','off');
[sepsilon,ssigma,ssim] = garchsim(specs,length(s)+1,1);
% exibi o último retorno, que é a previsão
fprintf(1,'\n\nRetorno Estimado para a série S: %.4f\n\n',ssim(end))

% Recorta (trunca) as séries para ficarem com o mesmo tamanho

m=min([length(xsim),length(ysim),length(zsim),length(tsim),length(rsim),
        length(ssim)]);
for i=1:m
    xsimtrunc(i)=xsim(length(x)-m+i+1);
    ysimtrunc(i)=ysim(length(y)-m+i+1);
    zsimtrunc(i)=zsim(length(z)-m+i+1);
    tsimtrunc(i)=tsim(length(t)-m+i+1);
    rsimtrunc(i)=rsim(length(r)-m+i+1);
    ssimtrunc(i)=ssim(length(s)-m+i+1);
end

%M é a matriz onde cada coluna é uma série temporal
M=[xsimtrunc' ysimtrunc' zsimtrunc' tsimtrunc' rsimtrunc' ssimtrunc'];

%C é a matriz de covariância
C=cov(M);

% Pega os retornos finais de cada série.
ret=[xsim(end) ysim(end) zsim(end) tsim(end) rsim(end) ssim(end)];

fprintf(1,'Fronteira Eficiente.\n\n')
frontcon (ret, C, 100); % exibi o gráfico.

% exibi riscos, retornos e pesos associados a cada portfólio.
% [PortRisk, PortReturn, PortWts]=frontcon (ret, C, 10)

```

Markow

Esse programa, *Markow*, mostra a função dos retornos esperados da carteira, que foram calculados no programa *Retorno*. Essa função dos retornos será a nossa função de aptidão para a implementação do algoritmo genético.

```
function z = Markow(x)
% função retorno da carteira

z = x(1) * 0.0368 + x(2) * 0.0095 + x(3)* 0.0318 + x(4)* 0.0020 +
    x(5)* -0.0172 + x(6)* -0.0278;
```

Markowrestric

O programa *Markowrestric* apresenta as restrições do algoritmo genético, ou seja, a soma dos pesos aplicados em cada série da carteira tem que ser um e cada um desses pesos tem que ser maior ou igual a zero.

```
function [c,ceq] = Markowrestric(x)
% c * x <=0 para desigualdades
% ceq * x =0 para igualdades

% Restrições:
% x1+x2+x3+x4+x5+x6=1 => x1+x2+x3+x4+x5+x6-1=0
% x1>=0 => -x1<=0
% x2>=0 => -x2<=0
% x3>=0 => -x3<=0
% x4>=0 => -x4<=0
% x5>=0 => -x5<=0
% x6>=0 => -x6<=0

c = [-x(1);-x(2);-x(3);-x(4);-x(5);-x(6)];
ceq = [x(1)+x(2)+x(3)+x(4)+x(5)+x(6)-1];
```

MarkowAG

O programa *MarkowAG* consiste em avaliar uma população inicial de ações (carteira) de tal forma que obtenha-se como resultado final uma sequência de pesos que indicam o quanto deve-se investir em cada uma das ações que compõem a nossa carteira para que se obtenha um resultado ótimo, ou seja, melhor retorno com menor risco.

Primeiramente o programa avalia cada uma dessas ações a partir da função de aptidão que foi definida no programa *Markow*. Após essas ações serem avaliadas, o programa vai gerar duas sequências de pesos que estarão diretamente relacionados com a sequência das ações. A partir disso, ele aplica o mecanismo de cruzamento e o mecanismo de mutação, para que a próxima geração seja criada. Isso ocorre sucessivamente até que o critério de parada, definido no início, diga quando encerrá-lo.

Esse programa deve ser rodado ao menos duas vezes, onde na primeira, é estabelecido como valores iniciais para as taxas de cruzamento e mutação 100% e 1%, e eles fornecerão os gráficos da taxa de mutação e da taxa de cruzamento que melhor se ajustarão a essa carteira. O segundo passo é rodar o programa novamente, porém com as taxas encontradas nos gráficos. Uma outra forma de testar as taxas de cruzamento e mutação seria atribuindo alguns valores a elas, e avaliar os resultados obtidos, ou seja, calcular quanto a carteira traria de retorno com os pesos sugeridos pelo programa.

Para efeito de avaliação do programa, são gerados dois outros gráficos. Um deles mostra as avaliações melhor, pior e média e o outro mostra a distância média entre cada indivíduo da população.

```
% Exemplo de função de avaliação com restrições
% Defini parâmetros
```

```

ObjectiveFunction = @Markow;           % define função de avaliação
nvars = 6;                             % número de variáveis
LB = [0 0 0 0 0 0];                   % limite inferior do domínio
UB = [1 1 1 1 1 1];                   % limite superior do domínio
ConstraintFunction = @Markowrestric;    % restrições

% Opções do Algoritmo Genético

% Critério de parada para gerações com mesma avaliação (stall)
options = gaoptimset('StallGenLimit',5);

% Tipo de Cruzamento
% Tipos possíveis: @crossoverheuristic @crossoversscattered
% @crossoverintermediate @crossoverssinglepoint @crossoverstwopoint
% @crossoverarithmetic

options = gaoptimset(options,'CrossoverFcn',@crossoverssinglepoint);

% Taxa de Cruzamento 100%
options = gaoptimset(options,'CrossoverFraction',1);

% Taxa de Mutação 1%
options = gaoptimset(options,'MutationFcn', {@mutationuniform,0.01})

% Mutation Feasible: função de mutação usada em problemas com restrições.
options = gaoptimset(options,'MutationFcn',@mutationadaptfeasible);

% Taxa de Migração 0%
options = gaoptimset(options,'MigrationFraction',0);

% Clonagem (elite) 0 (número inteiro)
options = gaoptimset(options,'EliteCount',0);

% Fronteira Pareto 0%
options = gaoptimset(options,'ParetoFraction',0);

% Tamanho de cada geração
options = gaoptimset(options,'PopulationSize',100);

% Exibi gráficos
options = gaoptimset(options,'PlotFcns',{@gaplotscores,@gaplotdistance},
    'Display','iter');

% Defini ponto inicial
X0 = [1 1 1 1 1 1];
options = gaoptimset(options,'InitialPopulation',X0);

[x,fval] = ga(ObjectiveFunction,nvars,[],[],[],[],LB,UB,ConstraintFunction,
    options)

%Comparando taxas de mutação
record=[];

```

```

for n=0:0.1:1
disp(sprintf('\nTaxa de mutação: %2.2f\n',n));
options = gaoptimset(options,'MutationFcn', {@mutationuniform, n},'PlotFcns', []);
[x fval]=ga(ObjectiveFunction,nvars, [], [], [], [], LB,UB,ConstraintFunction,options);
record = [record; fval];
end
figure();
plot(0:0.1:1,record)
xlabel('Taxa de Mutação')
ylabel('fval');

%Comparando taxas de cruzamento
record=[];
for n=0:0.1:1
    disp(sprintf('\nTaxa de cruzamento: %2.2f\n',n));
options = gaoptimset(options,'CrossoverFraction',n,'PlotFcns', []);
[x fval]=ga(ObjectiveFunction,nvars, [], [], [], [], LB,UB,ConstraintFunction,options);
record = [record; fval];
end
figure();
plot(0:0.1:1,record)
xlabel('Taxa de Cruzamento')
ylabel('fval');

```

G Algoritmo Genético 2

Para testar esse programa, é rodado utilizando uma carteira composta por três ações, entre elas estão Banco do Brasil, Bom Bril e Comgás. Esse teste consiste em gerar os retornos dos ativos e como medida de risco, usar o VaR, estudado na seção 6.

RetornoVaR

Esse programa consiste em calcular o valor esperado dos retornos de cada ativo que compõe a carteira e a matriz covariância. Diferentemente do programa *Retorno* do algoritmo genético 1, este salva os dados no matlab, ou seja, eles não são exibidos, porém, podem ser exibidos separadamente caso seja desejado. Os dados que serão salvos são: o vetor com os valores esperados dos retornos de cada ativo denominado *Med* e a matriz de auto-covariância denominada *C*.

```
% Calcula valor esperado e
% matriz de Auto-Covariância dos Retornos

x=load('bb.mat','-ascii');
y=load('bombril.mat','-ascii');
z=load('congas','-ascii');
xret=price2ret(x);
yret=price2ret(y);
zret=price2ret(z);

% Recorta (trunca) as séries para ficarem com o mesmo tamanho
m=min([length(xret),length(yret),length(zret)]);
for i=1:m-1
    xtrunc(i)=xret(length(x)-m+i);
    ytrunc(i)=yret(length(y)-m+i);
    ztrunc(i)=zret(length(z)-m+i);
end

% Vetor com os valores esperados dos retornos
Med=[mean(xtrunc) mean(ytrunc) mean(ztrunc)];

save('Med.mat','Med','-ASCII');

%M é a matriz onde cada coluna é uma série temporal
M=[xtrunc' ytrunc' ztrunc'];

%C é a matriz de auto-covariância
C=cov(M);

save('AutoCov.mat','C','-ASCII');
```

VaRRestrict

Esse programa tem a finalidade de restringir a soma dos pesos que serão fornecidos pelo algoritmo. Da mesma forma como foi explicado no algoritmo genético, os pesos são porcentagens que serão investidas em cada ativo da carteira, logo sua soma tem que ser igual a um e cada um deles deve ser maior que zero.

```
function [c,ceq] = VarRestrict(x)
```

```

mi=load('Med.mat','-ASCII'); % vetor com valores esperados dos retornos

% c * x <=0 para desigualdades
% ceq * x =0 para igualdades

% Restrições:
% x1+x2+x3=1 => x1+x2+x3-1=0
% x1 mi1 + x2 mi2 + x3mi3 >=0 => -(x1 mi1 + x2 mi2 + x3 mi3) <=0
% x1>=0 => -x1<=0
% x2>=0 => -x2<=0
% x3>=0 => -x3<=0

c = [-x(1);-x(2);-x(3);-(x(1)*mi(1)+x(2)*mi(2)+x(3)*mi(3))];
ceq = [x(1)+x(2)+x(3)-1];

```

VaR

Esse programa calcula a função do valor em risco, que será a função de aptidão para o algoritmo genético. Como foi definido na seção 6, para utilizar o VaR como medida de risco é preciso que seja dado um nível de confiança para encontrar o valor do índice α que compõe a fórmula do seu cálculo. Nesse caso considera-se esse nível de confiança 99% assim, de acordo com a tabela normal, o valor do α será 2,33. Além disso, considera-se que a taxa de retorno seja normalmente distribuída com média zero e desvio padrão σ . Para isso será feito o uso da matriz auto-covariância calculada no programa *Retorno VaR*.

```

function z = VaR(x)
alpha=2.33; % valor z da tabela normal que resulta num nível de confiança de 99%
M=load('AutoCov.mat','-ASCII'); % lê matriz de auto-cavariâncias

z = alpha*sqrt(x*M*x'); % função Var

```

VaRAG

Da mesma forma que o programa *MarkowAG*, este programa consiste na aplicação do algoritmo genético. A diferença entre eles é que nesse caso a função de aptidão utilizada é a função VaR, que é uma outra medida de avaliação do risco.

```

% Defini parâmetros
ObjectiveFunction = @VaR; % define função de avaliação
nvars = 3; % número de variáveis
LB = [0 0 0]; % limite inferior do domínio
UB = [1 1 1]; % limite superior do domínio
ConstraintFunction = @VarRestrict; % restrições

% Opções do Algoritmo Genético

% Critério de parada para gerações com mesma avaliação (stall)
options = gaoptimset('StallGenLimit',5);

% Forma da criação da população inicial
% @gacreationuniform @gacreationlinearfeasible

```

```

options = gaoptimset(options,'CreationFcn',@gacreationuniform);

% Tipo de Cruzamento
% Tipos possíveis: @crossoverheuristic @crossoversscattered
% @crossoverintermediate @crossoverssinglepoint @crossoverstwopoint
% @crossoverarithmic
% options = gaoptimset(options,'CrossoverFcn',@crossoverssinglepoint);
options = gaoptimset(options,'CrossoverFcn',@crossoversscattered);

% Taxa de Cruzamento
options = gaoptimset(options,'CrossoverFraction',1);

% Taxa de Mutação 1%
options = gaoptimset(options,'MutationFcn', {@mutationuniform,0.01})

% Mutation Feasible: função de mutação usada em problemas com restrições.
options = gaoptimset(options,'MutationFcn',@mutationadaptfeasible);

% Taxa de Migração 0%
options = gaoptimset(options,'MigrationFraction',0);

% Clonagem (elite) 0 (número inteiro)
options = gaoptimset(options,'EliteCount',0);

% Fronteira Pareto 0%
options = gaoptimset(options,'ParetoFraction',0);

% Tamanho de cada geração
options = gaoptimset(options,'PopulationSize',50);

% Exibi gráficos
options = gaoptimset(options,'PlotFcns',{@gaplotrange,@gaplotdistance},
    'Display','iter');

% Defini ponto inicial
% X0 = [0 0 0 0 0 0];
% options = gaoptimset(options,'InitialPopulation',X0);

[x,fval] = ga(ObjectiveFunction,nvars,[],[],[],[],LB,UB,ConstraintFunction,
    options)

% Comparando taxas de cruzamento
record=[];
for n=0:0.1:1
    disp(sprintf('\nTaxa de cruzamento: %2.2f\n',n));
options = gaoptimset(options,'CrossoverFraction',n,'PlotFcns',[]);
    [x fval]=ga(ObjectiveFunction,nvars,[],[],[],[],LB,UB,ConstraintFunction,
        options);
record = [record; fval];
end
figure();
plot(0:0.1:1,record)
xlabel('Taxa de Cruzamento')

```



```

ylabel('fval');

% Comparando taxas de mutação
record=[];
for n=0:0.1:1
    disp(sprintf('\nTaxa de mutação: %2.2f\n',n));
    options = gaoptimset(options,'MutationFcn', {@mutationuniform, n},
        'PlotFcns', []);
    [x fval]=ga(ObjectiveFunction,nvars, [], [], [], [], LB,UB,ConstraintFunction,
        options);
    record = [record; fval];
end
figure();
plot(0:0.1:1,record)
xlabel('Taxa de Mutação')
ylabel('fval');

```