



UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS - CFM  
DEPARTAMENTO DE FÍSICA

## **Estudo de Flutuações e Correlações em Séries Financeiras**

Renata Jaria Santos

FLORIANÓPOLIS

2008

UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS - CFM  
DEPARTAMENTO DE FÍSICA

## **Estudo de Flutuações e Correlações em Séries Financeiras**

**Renata Faria Santos**

*Dissertação apresentada ao Curso de Pós-graduação em Física da UFSC em preenchimento parcial dos requisitos para obtenção do título de Mestre em Física.*

ORIENTADOR: *Prof. Dr. Wagner Figueiredo*

FLORIANÓPOLIS

2008

Santos, Renata Faria

Estudo de Flutuações e Correlações em Séries Financeiras /  
Renata Faria Santos. — Florianópolis, 2008

81 p.

Dissertação (Mestrado) — Curso de Pós-graduação em Física da UFSC, 2008.

Orientador: Prof. Dr. Wagner Figueiredo

1. Econofísica 2. Processos Estocásticos 3. Flutuações de Preços

I. Título.

---

## Agradecimentos

---

Amizade, preocupação e apoio recebidos foram fundamentais nesse período de desenvolvimento deste trabalho, por isso gostaria de agradecer:

- Primeiramente, aos meus pais por tudo que fizeram e fazem por mim; meu imenso carinho e respeito.
- Ao Daniel, pelo amor, carinho e por me incentivar nos momentos de desânimo...
- e a sua querida família, pelo carinho e preocupação.
- Ao Prof. Wagner, agradeço pela paciência, dedicação e por ter aceitado me orientar em um tema ainda não estudado neste curso.
- Ao Prof. Renato Vicente, do EACH, USP, pelas sugestões de possíveis temas e pelas preciosas dicas que me ajudaram a iniciar esta dissertação.
- Aos amigos ...
- Aos colegas de trabalho e de curso, que tanto me ajudaram com sugestões, dicas e esclarecimentos.
- Para concluir, eu gostaria de agradecer a todos que direta ou indiretamente contribuíram para o desenvolvimetro deste trabalho.

Este trabalho foi financiado pela Coordenação de  
Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

# Resumo

Neste trabalho, utilizamos alguns conceitos da física, em especial, da mecânica estatística, que são aplicados no estudo de dados financeiros. Nos concentramos nos processos estocásticos e propriedades estatísticas que descrevem os retornos de preços de ações e no estudo das correlações entre elas. Utilizamos a Teoria da Matriz Aleatória para testar incertezas estatísticas presentes na matriz de correlação, obtida usando-se séries temporais dos retornos diários de preços de um conjunto de ações muito negociadas na Bolsa de Valores de São Paulo (Bovespa) no período de 2001 a 2007. Obtemos a densidade de autovalores da matriz de correlação e comparamos a estatística desses autovalores com a dos autovalores de uma matriz de correlação aleatória, construída a partir de séries temporais mutuamente descorrelacionadas. Analisamos a distribuição das componentes dos autovetores e observamos que os autovetores correspondentes aos menores autovalores, que não são explicados pela Teoria da Matriz Aleatória, carregam consigo informação econômica real, enquanto o maior autovalor corresponde a uma influência comum do mercado sobre todas as ações. Utilizamos o conceito de Espaço Ultramétrico e Teoria de Grafos para construir árvores de extensão mínima a fim de obtermos informações sobre o agrupamento das ações e extrair informações comuns compartilhadas entre elas. Consideramos diferentes intervalos de tempo ao obtermos a árvore na tentativa de se entender as flutuações dos preços ao longo do tempo.

# Abstract

In this work, we use some physical concepts, particularly, from statistical mechanics, which are applied in the study of financial data. We focus our attention in stochastic process and statistical properties, which describe price returns and in the study of cross correlations among stocks. We use the Random Matrix Theory to test statistical uncertainties in the estimation of the correlation matrix constructed for daily price returns of a set of stocks highly traded at São Paulo Stock Exchange (Bovespa) for the 7-yr period 2001-2007. We find the density of eigenvalues from the correlation matrix and we compare its statistics with that of a random correlation matrix - a correlation matrix constructed from mutually uncorrelated series. We analyze the distribution of eigenvector components of the correlation matrix and we observe that eigenvectors belonging to the smallest eigenvalues, which deviates from the Random Matrix Theory, carry real economic information, while the largest eigenvalue of the correlation matrix accounts the influence of the whole market that is shared by all stocks. We use concepts of Ultrametric Distance and Graph Theory to obtain minimum spanning trees in order to get useful information on the stocks clustering and to extract common information shared by them. We consider different time horizons to obtain trees in an effort to understand how price fluctuations evolve in time.

---

## Sumário

---

Agradecimentos	ii
Lista de Figuras	viii
Lista de Tabelas	x
Capítulo 1. Introdução	1
1.1 Física e Finanças - Introdução Histórica à Econofísica . . . . .	1
1.2 Motivação . . . . .	3
1.3 Foco Atual . . . . .	3
Capítulo 2. Definições Gerais e Introdução aos Processos Estocásticos	7
2.1 Informações Básicas sobre Ações . . . . .	7
2.2 Distribuições de Probabilidade, Momentos e Funções Características . . . . .	8
2.3 Modelos Estocásticos da Dinâmica de Preços . . . . .	18
2.3.1 Movimento Browniano Geométrico . . . . .	20
2.3.2 Outras Propostas . . . . .	21
2.4 Correlação Temporal em Séries Financeiras . . . . .	22
2.5 Teoria da Matriz Aleatória . . . . .	24
2.5.1 Teoria da Matriz Aleatória na Análise de Séries Financeiras . . . . .	26
2.6 Teoria de Grafos . . . . .	28
2.6.1 Distância entre Ações . . . . .	29



2.6.2	Espaços Ultramétricos . . . . .	30
Capítulo 3. Análise dos Dados		35
3.1	Análise dos coeficientes de correlação . . . . .	39
3.1.1	Teoria da Matriz Aleatória . . . . .	42
3.1.2	Árvore de Extensão Mínima . . . . .	52
Capítulo 4. Conclusões e Perspectivas		66
Apêndices		69
Apêndice A. Movimento Browniano		70
A.1	Distribuição de probabilidade da velocidade . . . . .	73
Apêndice B. Processo de Wiener		76
B.1	Fórmula de Itô . . . . .	77
Referências		80

---

## Lista de Figuras

---

2.1	Densidade de probabilidade Gaussiana (a) e Log-normal (b). Os parâmetros utilizados estão ilustrados nas figuras. . . . .	16
2.2	Densidade de probabilidade de Lévy (a) e t-Student (b) com os casos especiais. Os parâmetros utilizados estão indicados em cada figura. . . . .	17
2.3	Exemplo de aplicação do Algoritmo de Kruskal. Passo 1. . . . .	33
2.4	Passo 2. . . . .	33
2.5	Passo 3. . . . .	33
2.6	Passo 4. . . . .	33
2.7	Passo 5. . . . .	33
2.8	Passo 6. . . . .	33
2.9	Passo 7. . . . .	34
2.10	Passo 8. . . . .	34
2.11	Passo 9. . . . .	34
2.12	Passo 10. . . . .	34
2.13	Passo 11. . . . .	34
2.14	Passo 12. Árvore de extensão mínima resultante. . . . .	34
3.1	Evolução temporal do preço $S(t)$ , em reais, da ação ARCZ6 no período de janeiro de 2001 a junho de 2007. Está indicado o intervalo correspondente a cada ano estudado. . . . .	38

3.2	Série temporal do retorno de preço da ação ARCZ6 de janeiro de 2001 a junho de 2007. . . . .	38
3.3	Distribuição de probabilidade do retorno de preço para ações selecionadas de setores econômicos diferentes. . . . .	39
3.4	Distribuição de probabilidade do retorno de preço para ações selecionadas de quatro setores econômicos estudados. . . . .	40
3.5	Séries do logaritmo dos preços para as ações SBSP3 e VALE3 e BRTO4 e BRTP4, respectivamente. . . . .	41
3.6	Densidades das componentes dos autovetores correspondentes a alguns autovalores no intervalo $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$ comparadas à $\rho_{rm}(u)$ . . . . .	47
3.7	Densidade de probabilidade das componentes dos autovetores da matriz de correlação comparada à densidade $\rho_{rm}(u)$ . . . . .	47
3.8	Contribuição $X_s^k$ para alguns autovetores da matriz de correlação. . . . .	49
3.9	Contribuição $X_s^k$ para o autovalor $\lambda_{43}$ . Os setores econômicos são indicados por s, ver Figura 3.8. . . . .	50
3.10	Árvore de extensão mínima para o ano de 2001. . . . .	54
3.11	Árvore de extensão mínima para o ano de 2002. . . . .	55
3.12	Árvore de extensão mínima para o ano de 2003. . . . .	56
3.13	Árvore de extensão mínima para o ano de 2004. . . . .	58
3.14	Árvore de extensão mínima para o ano de 2005. . . . .	59
3.15	Árvore de extensão mínima para o ano de 2006. . . . .	60
3.16	Árvore de extensão mínima para o ano de 2007. . . . .	61
3.17	Árvore obtida para o todo o período estudado, de janeiro de 2001 a junho de 2007. . . . .	62

---

## Lista de Tabelas

---

- 3.1 Lista das 43 empresas estudadas. A primeira coluna identifica as empresas pelo nome de pregão, a segunda coluna indica o símbolo de negociação da empresa que é composto por 4 letras e um número: 3 indica ações do tipo ON, 4 PN, 5 PNA e 6 PNB. A terceira coluna mostra o setor e a quarta o sub-setor econômico ao qual a companhia pertence. . . . . 36
- 3.2 Coeficientes de correlação mínimo e máximo obtidos para cada ano estudado. São mostrados os  $\rho_{ij}^{max}$  obtidos para companhias  $i$  e  $j$  diferentes. . . . 42

---

# CAPÍTULO 1

## Introdução

---

### 1.1 Física e Finanças - Introdução Histórica à Econofísica

O contato de físicos com Finanças é de longa data. Isaac Newton perdeu muito de sua fortuna na Bolsa de Londres e reclamou que enquanto podia calcular precisamente o caminho de corpos celestes em minutos e centímetros, ele era incapaz de prever o quão alta ou baixa a cotação de um ativo poderia chegar.

Carl Friederich Gauss (1777-1855) teve muito êxito em suas operações financeiras. Isto é evidenciado por sua fortuna deixada de 170000 Taler (unidade monetária local na época) ao morrer, enquanto seu salário era de 1000 Taler. Acredita-se que ele derivou a distribuição normal de probabilidade (Gaussiana) ao estimar o risco ao fornecer crédito a seus vizinhos [1].

Mas podemos considerar como uma raiz da área conhecida como Econofísica a primeira formalização da caminhada aleatória (*random walk*). Esse trabalho não se deu em uma publicação de Einstein, mas em uma tese de doutorado de Bachelier. O matemático francês Louis Bachelier defendeu sua tese na Academia de Paris em 29 de março de 1900 intitulada "*Théorie de la Spéculation*" (Teoria da Especulação)[2]. Sua tese

lida com apreçamento de opções<sup>1 2</sup> em mercados especulativos, uma atividade que hoje é extremamente importante nos mercados financeiros em que derivativos<sup>3</sup> são negociados em vários mercados. Bachelier determinou a probabilidade de mudanças de preços ao escrever a equação que atualmente é conhecida como equação de Chapman-Kolmogorov, e ao reconhecer que o processo de Wiener satisfaz a equação de difusão (este ponto foi redescoberto por Einstein em seu artigo sobre o Movimento Browniano em 1905)[4].

Outra contribuição importante nessa área foi dada pelo economista Vilfredo Pareto (1848-1923). Ele investigou o caráter estatístico das riquezas individuais em uma economia estável usando um modelo de distribuição segundo uma lei de potência. Pareto usou uma lei de potência para descrever a distribuição de riquezas dentre os indivíduos, uma vez que ela descrevia muito bem o fato de uma grande parte da riqueza de qualquer sociedade pertencer a uma pequena fração das pessoas dessa sociedade. Este estudo serviu de base para modelos da teoria de probabilidade [4].

Analogias à física desempenharam um papel importante no desenvolvimento da Teoria Econômica durante o século XIX, e alguns fundadores da Teoria Econômica Neoclássica, incluindo Irving Fisher, tiveram uma formação básica em física; Fisher foi aluno de Willard Gibbs [5].

---

<sup>1</sup>Opção é um instrumento financeiro que confere a seu titular o direito de comprar ou vender um ativo-objeto a um preço determinado. Opções são instrumentos derivativos, isso significa que o valor de uma opção e suas características de negociação estão ligados ao ativo adjacente às opções. Uma opção da Petrobrás PN está ligada ao ativo PETR4, Petrobrás PN.

<sup>2</sup>Ativos são títulos, valores mobiliários e outros instrumentos financeiros de emissor público ou privado.

<sup>3</sup>Derivativos são ativos financeiros ou valores mobiliários cujo valor e características de negociação derivam do ativo que lhes servem de referência. [3]

## 1.2 Motivação

*A* partir de 1970, uma série de mudanças significativas ocorreram no mercado financeiro. Especialmente em 1973, moedas começaram a ser negociadas no mercado financeiro e seus valores determinados em mercados de diferentes países, um mercado financeiro ativo 24 horas por dia em todo o mundo. Nesse mesmo período, Robert C. Merton e Myron S. Scholes, com a colaboração de Fischer Black, desenvolveram um trabalho sobre apreçamento de derivativos. Fisher Black e Myron Scholes reduziram o problema de apreçamento de opções a um problema de difusão. Esse trabalho deu origem a novas formas de tratar os instrumentos financeiros e facilitou a administração do risco. Desde então, o volume de negociações em mercados financeiros em todo o mundo cresceu a uma taxa impressionante. O volume de negociações em 1995 foi 80 vezes maior que em 1973, e um crescimento ainda mais expressivo se deu nos mercados de derivativos.

Uma segunda grande transformação se deu nos anos 80 com a expansão da negociação eletrônica para bolsas de valores em diversas regiões do mundo. Isso gerou uma quantidade muito grande de dados disponíveis eletronicamente. A enorme expansão dos mercados financeiros requer grandes investimentos em dinheiro e em recursos humanos para atingir uma qualificação confiável e para que haja uma minimização dos riscos para as instituições financeiras envolvidas [4].

## 1.3 Foco Atual

*N*o final do século passado alguns físicos da área da Mecânica Estatística passaram a fazer pesquisas relacionadas com as flutuações observadas no mercado

financeiro. Essa área de pesquisa tem se desenvolvido muito nos últimos anos devido ao volume enorme de dados disponíveis. Há na natureza poucos sistemas físicos com tantos dados disponíveis, em várias escalas de tempo.

O conceito de sistemas complexos em física não tem uma definição muito precisa, mas uma característica que permeia todos eles é a presença da não-linearidade e de um número muito grande de unidades interagentes. O comportamento coletivo desses sistemas exhibe padrões espaciais e temporais em suas propriedades dinâmicas.

Mercados financeiros exibem várias das propriedades que caracterizam sistemas complexos. As regras que os governam são estáveis e a evolução temporal do sistema é continuamente monitorada. É possível desenvolver modelos e testar sua validade, além de fazer previsões a partir dos dados disponíveis, de modo que o sistema financeiro se apresenta como um verdadeiro laboratório para aplicar as idéias desenvolvidas na área de Mecânica Estatística fora do equilíbrio, particularmente aquelas relativas aos processos estocásticos.

O fato de haver um grande número de participantes nos mercados, com maneiras diversas de agir e interesses diferentes, implica em um comportamento não determinístico para o sistema. O estudo estatístico consiste em, a partir de observações das flutuações de preços no passado, obter informações sobre a frequência de possíveis mudanças nos preços dos ativos. Se assumirmos que essa frequência reflete algum mecanismo próprio dos mercados, então podemos esperar que ela irá permanecer estacionária. Essa suposição é muito razoável, por exemplo, para descrever o comportamento de dados, mas é menos justificável no caso de mercados financeiros - apesar de o grande número de participantes conferir ao sistema uma certa regularidade. Está claro que os mercados financeiros não se comportam da mesma maneira que há 30 anos atrás, mas o estudo estatístico está baseado na idéia segundo a qual as evoluções ocorrem lentamente (na escala de vários anos), exceto ao ocorrer, por exemplo, crises. Dessa maneira, a observação do



passado recente é útil para descrever um futuro não tão distante [6]. Essa nova área foi batizada por Mantegna de Econofísica [4].

Uma das áreas mais ativas em Finanças é o apreçamento de derivativos quando certas suposições do Modelo de Black e Scholes são alteradas. No caso mais simples, um ativo é descrito por um processo estocástico e um derivativo é obtido com base no ativo do qual deriva, no valor deste e em suas propriedades estatísticas. Este problema apresenta pelo menos dois aspectos diferentes: um está relacionado à natureza do processo aleatório do ativo e o outro está relacionado às soluções do problema de apreçamento da opção, que se baseia na suposição da ocorrência de certos processos aleatórios.

Outra área de interesse se refere à completa caracterização dos processos estocásticos das mudanças de preços de ativos financeiros. Vários estudos têm sido realizados concentrando-se nos diferentes aspectos do processo estocástico analisado como, por exemplo, a forma da distribuição das mudanças de preços, memória temporal e propriedades estatísticas de ordens maiores. Esta é uma área muito ativa e esforços estão sendo feitos a fim de desenvolver um modelo estocástico mais satisfatório que descreva todos os aspectos dos dados empíricos (ver referências contidas em [4]).

Outro campo de interesse é o desenvolvimento de modelos teóricos capazes de descrever as características dos mercados financeiros reais. Muitos modelos já foram propostos e algumas das principais propriedades da dinâmica estocástica do preço das ações são reproduzidas por eles como, por exemplo, a forma não gaussiana da cauda da distribuição das mudanças de preços.

Além dos trabalhos que analisam e modelam os sistemas financeiros, há estudos sobre a distribuição de riquezas de empresas e estudos das propriedades estatísticas de seu crescimento. As propriedades estatísticas de resultados econômicos de organizações complexas tais como países e universidades também têm sido estudadas [4].

Neste trabalho, focalizamos nosso estudo nos processos estocásticos e propriedades estatísticas que descrevem os retornos de preços de ações, assuntos discutidos

---

no Capítulo 2, e no estudo das correlações desses retornos. No Capítulo 3, utilizamos conceitos da Teoria da Matriz Aleatória para testar incertezas estatísticas encontradas na matriz de correlação obtida a partir dos retornos de preços diários de um conjunto de ações muito negociadas na Bolsa de Valores de São Paulo. Usamos conceitos de Espaço Ultramétrico e Teoria de Grafos para obter informações quanto ao agrupamento das ações e extrair informações comuns compartilhadas entre elas. Obtemos árvores de extensão mínima de diferentes períodos na tentativa de entender as flutuações dos preços ao longo do tempo. O entendimento das flutuações dos preços é muito importante para a escolha das ações que irão compor uma carteira de investimento.

---

## CAPÍTULO 2

### Definições Gerais e Introdução aos Processos Estocásticos

---

#### 2.1 Informações Básicas sobre Ações

Uma ação representa a menor parcela do capital social de uma empresa (sociedade anônima, sociedade por ações ou companhia).

Todas as ações negociadas atualmente no Brasil são nominativas (N) ou escriturais (E). Para ações nominativas, presume-se a propriedade para aquele que constar no Livro Registro das Ações Nominativas. Já para as ações escriturais, dispensa-se a emissão de títulos de propriedade; circulam e se transferem mediante extratos dos bancos depositários [7].

Quanto às suas características, podem ser classificadas em duas categorias: ordinárias (O) e preferenciais (P). Ações ordinárias são aquelas que garantem a seu possuidor o direito a voto nas assembleias dos acionistas, e ações preferenciais garantem a preferência a seu detentor no recebimento de dividendos<sup>1</sup> em relação às ordinárias.

As ações também podem apresentar características específicas definidas em sua emissão que as fazem ser classificadas por letras A, B ou C, vindas logo após

---

<sup>1</sup>Dividendos representam a mais importante forma de remuneração do acionista. A Lei das Sociedades Anônimas determina que pelo menos 25% do lucro líquido da empresa seja distribuído na forma de dividendos, após a realização das reservas de lucros permitidas em lei [3].

sua designação principal, como, por exemplo, PNA (ações preferenciais nominativas série A). Não existe uma regra geral para delimitar as características de cada classe, que são definidas pela empresa emissora em seu estatuto social [8].

## 2.2 Distribuições de Probabilidade, Momentos e Funções Características

Para descrever um processo descrito por uma variável aleatória  $X$  para a qual o resultado é um número real<sup>2</sup>  $x$ , usamos a densidade de probabilidade  $P(x)$ . A probabilidade de a variável aleatória  $X$  assumir um valor no intervalo  $[a, b]$  é

$$\int_a^b P(x)dx, \quad (2.1)$$

sendo que  $P(x)$  deve ter as propriedades

$$P(x) \geq 0, \quad (2.2)$$

$$\int_{-\infty}^{\infty} P(x)dx = 1. \quad (2.3)$$

---

<sup>2</sup>Nesta dissertação, usamos a convenção de que variáveis aleatórias são designadas por letras maiúsculas, enquanto os valores que elas tomam são designados por letras minúsculas.

Os momentos  $\mu_n$  da distribuição  $P(x)$  são definidos como

$$\mu_n \equiv \langle x^n \rangle = \int x^n P(x) dx. \quad (2.4)$$

A média  $\mu$  é o primeiro momento ( $n = 1$ ) e a dispersão ou variância está relacionada ao segundo momento ( $\sigma^2 = \mu_2 - \mu_1^2$ ). A eq. (2.4) é válida somente se a integral converge, o que requer que  $P(x)$  decresça suficientemente rápido para grandes valores de  $|x|$ .

O conhecimento de todos os momentos de uma distribuição equivale ao conhecimento da própria distribuição<sup>3</sup>  $P(x)$ . Na análise de dados coletados em experimentos reais, na maioria das vezes, apenas o primeiro e segundo momentos são considerados. Por exemplo, no caso de séries temporais, para se avaliar momentos de ordem maior do que dois, necessitamos de séries muito longas; caso contrário há uma certa perda de precisão ao se determinar a distribuição.

A função característica de uma variável aleatória  $X$  é definida como a transformada de Fourier da densidade de probabilidade associada a ela

$$\tilde{P}(z) \equiv \int e^{izx} P(x) dx = \langle e^{izx} \rangle. \quad (2.5)$$

Definimos os cumulantes  $c_n$  de uma distribuição pelas derivadas sucessivas do logaritmo da função característica:

$$c_n = (-i)^n \left. \frac{d^n}{dz^n} \log \tilde{P}(z) \right|_{z=0}. \quad (2.6)$$

---

<sup>3</sup>Isto não é rigorosamente correto, uma vez que há exemplos de distribuições diferentes que possuem os mesmos momentos. Veja, por exemplo, as distribuições na página 9 da Referência [6].

Os cumulantes  $c_n$  são combinações polinomiais dos momentos  $\mu_p$  com  $p \leq n$  [9]. Por exemplo:

$$\begin{aligned} c_1 &= \mu_1, \\ c_2 &= \mu_2 - \mu_1^2 = \sigma^2, \\ c_3 &= \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3, \\ c_4 &= \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4. \end{aligned} \tag{2.7}$$

É útil normalizar os cumulantes por uma potência da variância apropriada, de maneira que obtemos quantidades adimensionais. Os cumulantes normalizados  $\lambda_n$  são definidos [6] por

$$\lambda_n \equiv c_n / \sigma^n. \tag{2.8}$$

Destacamos aqui o terceiro e quarto cumulantes normalizados, chamados de obliquidade<sup>4</sup> (*skewness*) ( $\zeta$ ) e curtose<sup>5</sup> (*kurtosis*) ( $\kappa$ ), respectivamente

$$\zeta \equiv \frac{\langle (x - \mu)^3 \rangle}{\sigma^3}, \tag{2.9}$$

$$\kappa \equiv \frac{\langle (x - \mu)^4 \rangle}{\sigma^4} - 3. \tag{2.10}$$

A definição acima de cumulantes pode parecer arbitrária, mas essas quantidades possuem propriedades interessantes. Por exemplo, cumulantes se adicionam quando variáveis aleatórias independentes são somadas. Além disso, uma distribuição Gaussiana é caracterizada pelo fato de todos os seus cumulantes de ordem maior que dois serem identicamente nulos. Conseqüentemente, os cumulantes, em especial  $\kappa$ , podem ser interpretados como uma medida da distância entre uma dada distribuição  $P(x)$  e uma Gaussiana.

<sup>4</sup>Obliquidade é uma medida da assimetria de uma determinada distribuição de probabilidade.

<sup>5</sup>Curtose é uma medida de dispersão que caracteriza o "achatamento" da curva da função de distribuição. Note que é possível encontrar  $\kappa + 3$ , e não  $\kappa$ , sendo chamado de curtose.

Se uma distribuição possui  $\kappa > 0$ , isto significa que relativamente a uma distribuição Gaussiana, ela apresenta um pico mais acentuado no valor médio e uma cauda que decresce mais suavemente. Por outro lado, se  $\kappa < 0$ , a distribuição é mais achatada em torno do valor médio e sua cauda decresce mais abruptamente.

Uma densidade de probabilidade Gaussiana de média  $\mu$  e variância  $\sigma$  é definida como

$$P(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \quad (2.11)$$

Uma variável com uma distribuição Gaussiana é peculiar porque 'grandes desvios' são extremamente raros. A quantidade  $\exp[-(x-\mu)^2/2\sigma^2]$  decai tão rapidamente para grandes valores de  $x$  que desvios algumas vezes maiores que  $\sigma$  são quase impossíveis. Por exemplo, uma variável Gaussiana se desvia de seu valor mais provável por  $2\sigma$  apenas em 5% das vezes, enquanto flutuações de  $10\sigma$  possuem uma probabilidade de ocorrência de  $2 \times 10^{-23}$ .

Definimos uma densidade de probabilidade para a distribuição log-normal [6] por

$$P(x) \equiv \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad (2.12)$$

válida somente para  $x > 0$ .

As distribuições normal e log-normal estão relacionadas entre si. Se  $X$  possui distribuição log-normal com parâmetros  $\mu$  e  $\sigma^2$ , então  $\ln X$  possui distribuição normal com parâmetros  $\mu$  e  $\sigma^2$ . A log-normal é aplicável quando a quantidade de interesse deve ser positiva. Essa distribuição é utilizada em matemática financeira para modelar a distribuição de riquezas, por exemplo.

Quando a probabilidade de se medir um valor particular de uma dada quantidade variar inversamente como uma potência desse valor, essa quantidade possui comportamento tipo lei de potência, também conhecida na literatura como distribuição de

Pareto [10]. A densidade de probabilidade é definida por

$$P(x) \equiv Cx^{-\alpha}, \quad (2.13)$$

sendo  $\alpha$  o expoente da lei de potência. Deve haver um valor mínimo  $x_{min}$  para o qual a distribuição seja ainda válida, uma vez que ela diverge à medida que  $x \rightarrow 0$ . A constante  $C$  é determinada através da condição de normalização,

$$C = (\alpha - 1)x_{min}^{\alpha-1}, \quad \text{para } \alpha > 1. \quad (2.14)$$

Note que para  $\alpha = 1$ , a distribuição não é normalizável.

A média, ou primeiro momento, é dada por

$$\mu_1 = \int_{x_{min}}^{\infty} xP(x)dx = \frac{\alpha - 1}{\alpha - 2}x_{min}, \quad (2.15)$$

sendo que leis de potência com  $1 < \alpha \leq 2$  não possuem média finita.

O segundo momento é dado por

$$\mu_2 = \frac{C}{3 - \alpha} \left[ x^{-\alpha+3} \right]_{x_{min}}^{\infty} = \frac{\alpha - 1}{\alpha - 3}x_{min}^2, \quad (2.16)$$

que diverge para  $1 < \alpha \leq 3$ . De maneira geral, todos os momentos  $\mu_n$  são finitos para  $n < \alpha - 1$  enquanto que todos os momentos de ordem maior divergem.

A distribuição de riqueza dos indivíduos em um país e a intensidade das manchas solares são exemplos de distribuições que seguem leis de potência. Outra situ-



ação na qual pode aparecer distribuição do tipo lei de potência é em transições de fase. Alguns sistemas podem apresentar apenas uma escala de comprimento ou de tempo que os governam. Um exemplo clássico é o de um magneto, que possui um comprimento característico que mede o tamanho típico das correlações magnéticas. Sob certas circunstâncias, esta escala de comprimento pode divergir, deixando o sistema sem uma escala característica. Neste caso, o comportamento de certas quantidades macroscópicas seguem leis de potência. No caso de sistemas ferromagnéticos, esse comportamento surge nas vizinhanças de seu ponto crítico.

Outra distribuição importante é a de Lévy, que aparece no contexto do Teorema Central do Limite <sup>6</sup>, devido a sua propriedade de estabilidade <sup>7</sup> sob adição. A cauda de certas distribuições de Lévy decai menos abruptamente que a da Gaussiana e é, conseqüentemente, mais útil para descrever fenômenos para os quais valores muito grandes ou muito pequenos de uma quantidade podem ser observados, como, por exemplo, riquezas pessoais, tamanho de fundos de pensão, amplitudes de terremotos ou qualquer outra catástrofe natural.

Lévy [11] resolveu o problema de determinar uma classe inteira de distribuições estáveis. A forma mais geral para a função característica de um processo estável [4] é

$$\tilde{P}(z) = \begin{cases} \exp \{ i\mu z - c|z|^\alpha [1 - i\beta \frac{z}{|z|} \tan(\frac{\pi}{2}\alpha)] \} & [\alpha \neq 1], \\ \exp \{ i\mu z - c|z|[1 + i\beta \frac{z}{|z|} \frac{2}{\pi} \ln |z|] \} & [\alpha = 1], \end{cases} \quad (2.17)$$

sendo  $\mu$  qualquer número real,  $0 < \alpha \leq 2$ ,  $c$  um fator de escala positivo, que é uma medida da largura da distribuição, e  $\beta$  um parâmetro de assimetria variando no intervalo  $[-1, 1]$ .

O caso simétrico ocorre para  $\beta = 0$ .

---

<sup>6</sup>O Teorema Central do Limite afirma que se a soma (combinação linear) de variáveis aleatórias independentes possui variância finita, então a distribuição da soma tenderá a uma distribuição Gaussiana, à medida que o número de variáveis aleatórias torna-se muito grande. Para o caso em que não haja independência ou variância finita das variáveis aleatórias, então há generalizações desse teorema [4].

<sup>7</sup>Se um conjunto de variáveis aleatórias independentes possui uma distribuição estável, então uma combinação linear dessas variáveis terá a mesma distribuição, mas com parâmetros de escala diferentes. Gaussiana, Lorentziana e Lévy simétrica são exemplos de distribuições estáveis.

A forma analítica da distribuição estável de Lévy é conhecida apenas para alguns valores de  $\alpha$  e  $\beta$ :

- $\alpha = 1, \beta = 0$  (Cauchy ou Lorentziana)

$$P_{\alpha=1}(x) = \frac{\gamma}{\pi(x^2 + \gamma^2)} \quad , \quad (2.18)$$

- $\alpha = 1/2, \beta = 1$  (Lévy-Smirnov)

$$P_{\alpha=1/2}(x) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/2(x-x_0)}}{(x-x_0)^{3/2}} \quad , \quad (2.19)$$

sendo  $x_0$  um parâmetro de desvio da curva.

- $\alpha = 2$  (Gaussiana)

Uma importante propriedade das distribuições de Lévy é o seu comportamento do tipo lei de potência para grandes valores de seu argumento [6]:

$$P_\alpha(x) \sim \frac{\alpha A_\pm^\alpha}{|x|^{1+\alpha}} \quad \text{para } x \rightarrow \pm\infty, \quad (2.20)$$

sendo  $A_\pm^\alpha$  duas constantes chamadas de parâmetros de escala que indicam a ordem de magnitude de grandes flutuações (positivas ou negativas),  $\beta \equiv (A_+^\alpha - A_-^\alpha)/(A_+^\alpha + A_-^\alpha)$  e  $\alpha$  um expoente  $0 \leq \alpha \leq 2$ . Pode-se observar, entretanto, comportamentos de lei de potência para  $\alpha \geq 2$ , mas essas distribuições não correspondem ao comportamento assintótico de uma distribuição de Lévy.

Em particular, todos os processos de Lévy estáveis com  $\alpha < 2$  possuem variância infinita. Consequentemente, processos estocásticos estáveis não-Gaussianos não

possuem escala característica - a variância é infinita! No caso de  $\alpha \leq 1$ , a densidade de probabilidade decai tão lentamente que mesmo a média não é mais finita<sup>8</sup>.

Outra densidade de probabilidade é a distribuição t-Student [4], que possui também caudas do tipo lei de potência:

$$P(x) \equiv \frac{1}{\sqrt{\pi\gamma}} \frac{\Gamma[(1+\gamma)/2]}{\Gamma(\gamma/2)} \frac{1}{(1+x^2/\gamma)^{(1+\gamma)/2}}, \quad (2.21)$$

que coincide com a distribuição de Lorentz para  $\gamma = 1$ , e para  $\gamma \rightarrow \infty$  esta distribuição é Gaussiana. Um processo estocástico caracterizado por esta distribuição pode ter momentos finitos ou infinitos, dependendo do parâmetro  $\gamma$  selecionado.

A Fig. (2.1) ilustra as densidades de probabilidade Gaussiana (a) e Log-normal (b). A Fig. (2.2) mostra as densidades de probabilidade de Lévy (a) e t-Student (b), ambas com casos especiais que se reduzem à Lorentziana e à Gaussiana. Notem como as densidades diferem entre si na cauda e em torno do centro da distribuição. Nas legendas de cada densidade ilustrada estão os parâmetros e expoentes utilizados.

---

<sup>8</sup>A mediana e o valor mais provável, entretanto, existem.

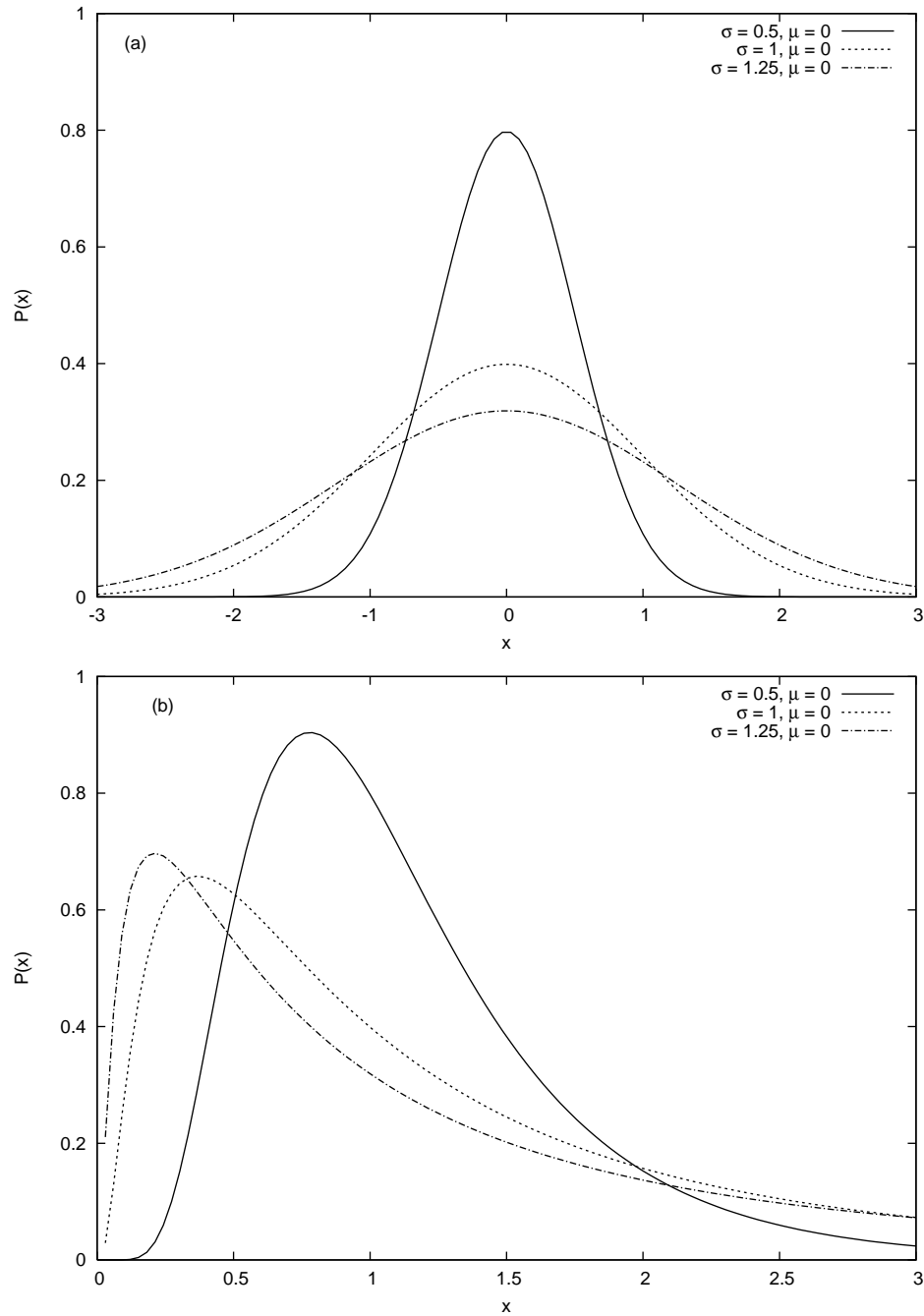


Figura 2.1: Densidade de probabilidade Gaussiana (a) e Log-normal (b). Os parâmetros utilizados estão ilustrados nas figuras.

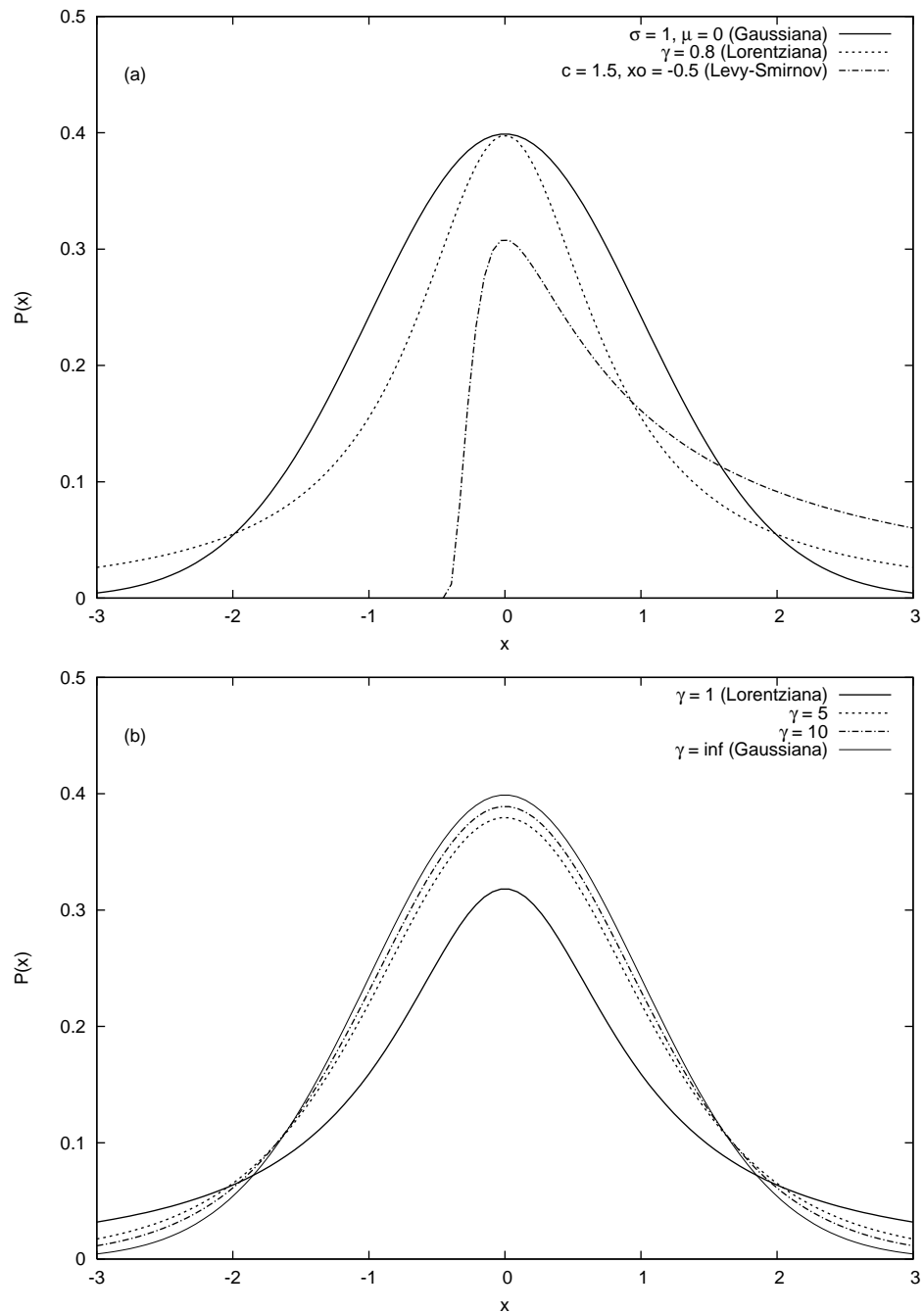


Figura 2.2: Densidade de probabilidade de Lévy (a) e t-Student (b) com os casos especiais. Os parâmetros utilizados estão indicados em cada figura.

## 2.3 Modelos Estocásticos da Dinâmica de Preços

Definindo  $S_i(t)$  como o preço do ativo financeiro  $i$  no instante  $t$ , vamos considerar duas variáveis estocásticas: a mudança de preço  $Z_i(t)$  e a diferença do logaritmo natural dos preços  $R_i(t)$  [4],

$$Z_i(t) \equiv S_i(t) - S_i(t - \Delta t), \quad (2.22)$$

$$R_i(t) \equiv \ln S_i(t) - \ln S_i(t - \Delta t). \quad (2.23)$$

sendo  $\Delta t$  um intervalo de tempo escolhido.

Definimos a eq. (2.23) como o retorno de preços. Usaremos essa variável em nosso estudo, considerando  $\Delta t = 1$  dia. A vantagem do uso dessa variável estocástica é que a correção média das mudanças de escala são incorporadas sem a necessidade de fatores de desconto, pois o intervalo de tempo escolhido em nosso estudo é pequeno. Assim, os preços utilizados não precisam ser recalculados a fim de se levar em consideração a inflação do período. A principal desvantagem de se trabalhar com  $Z_i(t)$  é que essa variável é muito afetada por mudanças de escala.

Uma vez que a ação representa a menor parcela do capital social de uma empresa, seu preço de alguma maneira deve refletir o valor dessa empresa. Note que o valor presente de uma companhia depende não apenas de sua situação atual, mas também de sua performance futura. Aqui surge o problema básico de apreçamento de um ativo: estamos tentando fazer previsões futuras baseadas em informações atuais. Assim, se uma nova informação é revelada, e que poderá afetar o desempenho futuro da ação,

então a dinâmica de preços também poderá ser afetada por essa expectativa, ou seja, o preço futuro de uma ação estará sujeito a um certo grau de incerteza. Isto é refletido no comportamento errático que os preços exibem quando são vistos em função do tempo. Podemos escrever uma equação que ilustra a taxa de variação de preço de um ativo

$$\frac{dS}{dt} = [\mu + \sigma \xi(t)]S, \quad (2.24)$$

sendo que o termo entre colchetes se divide em duas componentes: a taxa de retorno média previsível  $\mu$  e um termo de flutuação ou ruído  $\sigma \xi(t)$ , responsável pela aleatoriedade do preço da ação.

O modelo estocástico mais comum da dinâmica de preços assume que o  $\ln S_i(t)$  é um processo difusivo, e que os incrementos do  $\ln S_i(t)$  são distribuídos de acordo com uma distribuição Gaussiana [4]. Esse modelo, conhecido como Movimento Browniano Geométrico e descrito na Subseção 2.3.1, fornece uma primeira aproximação do comportamento observado em dados empíricos. Entretanto, desvios sistemáticos das previsões do modelo são observados, e as distribuições das séries temporais reais apresentam picos mais acentuados em relação ao valor médio e caudas que decrescem menos abruptamente que uma distribuição Gaussiana. Por esse motivo, outros modelos foram propostos numa tentativa de descrever melhor o comportamento de dados reais. Um exemplo de modelo alternativo é discutido na Subseção 2.3.2.

### 2.3.1 Movimento Browniano Geométrico

O movimento Browniano geométrico faz duas hipóteses fundamentais sobre um processo estocástico [1]:

1. Acontecimentos sucessivos da variável estocástica são estatisticamente independentes;
2. O retorno de preços, ou mudanças relativas da variável estocástica são dados por uma função densidade de probabilidade Gaussiana, e assim, a função densidade da variável estocástica, isto é, dos preços, é Log-normal.

O movimento Browniano geométrico é definido<sup>9</sup> como a solução da equação diferencial estocástica [12]

$$dS = \mu S dt + \sigma S dW, \quad (2.25)$$

sendo  $\mu$  e  $\sigma$  constantes, sujeitas a uma condição genérica inicial  $S(t_0) = S_0$ , e  $dW$  representa o processo estocástico de Wiener (ver Apêndice B). Fazendo a mudança de variável,  $G = \ln S$  e aplicando a Fórmula de Itô, eq.(B.11) dada no Apêndice B com  $a = \mu S$  e  $b = \sigma S$  e  $F(S) = \ln S$ , temos

$$dG = \left( \mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dW. \quad (2.26)$$

O  $\ln S$  segue um processo de Wiener generalizado com uma tendência efetiva  $\mu - \frac{1}{2} \sigma^2$  e desvio padrão  $\sigma$ . Note que tanto  $dS$  como  $dG$  são afetados pela mesma fonte de incerteza: o processo estocástico  $dW(t)$ . Após integração de  $dG$ , obtemos a equação

$$G(t) = G_0 + \left( \mu - \frac{1}{2} \sigma^2 \right) (t - t_0) + \sigma [W(t) - W(t_0)], \quad (2.27)$$

---

<sup>9</sup>Note a semelhança dessa equação com a equação de Langevin dada no Apêndice A e discutida de forma mais geral no Apêndice B, eq. (B.7).



sendo  $G_0 = \ln S_0$ . Podemos obter, assim, a solução da eq. (2.25)

$$S(t) = S_0 \exp \left\{ \left( \mu - \frac{1}{2} \sigma^2 \right) (t - t_0) + \sigma [W(t) - W(t_0)] \right\}. \quad (2.28)$$

Da eq. (2.27), vemos que  $G(t) - G_0$  se distribui segundo a densidade de probabilidade Gaussiana  $\mathcal{N}((\mu - \sigma^2/2)\tau, \sigma\sqrt{\tau})$ , sendo  $\tau = t - t_0$ . Assim, o movimento Browniano geométrico com valor inicial  $S(t_0) = S_0$  possui a seguinte distribuição Log-normal:

$$P(S(t)) = \frac{1}{S\sigma\sqrt{2\pi\tau}} \exp \left\{ - \frac{\left[ \ln\left(\frac{S}{S_0}\right) - (\mu - \frac{1}{2}\sigma^2)\tau \right]^2}{2\sigma^2\tau} \right\}. \quad (2.29)$$

### 2.3.2 Outras Propostas

Baseado em suposições teóricas e análise de dados, vários modelos alternativos ao movimento Browniano geométrico foram propostos [4]. Os modelos diferem entre si não apenas com respeito à forma da função densidade de probabilidade, mas também com respeito a certas propriedades, tais como:

- segundo momento e momentos de ordem maior da distribuição finitos ou infinitos,
- caráter contínuo ou descontínuo de  $S(t)$  - ou  $\ln S(t)$ ,
- comportamento de escala do processo estocástico.

Em sua análise pioneira de séries temporais do preço do algodão, Mandelbrot [13] observou que a distribuição dos retornos não era muito bem descrita por uma distribuição Gaussiana, principalmente na cauda. Mandelbrot propôs que a distribuição do retorno de preços é consistente com uma distribuição de Lévy estável.

Estudos recentes de dados reais do índice Standard and Poor's 500 (S & P 500) analisam tipicamente  $10^6$ - $10^7$  pontos, em contraste com aproximadamente 2000 pontos analisados no trabalho clássico de Mandelbrot [13]. É mostrado na referência [14] que a parte central da distribuição dos retornos do índice S & P 500 é bem ajustada por uma distribuição de Lévy, mas o comportamento assintótico dessa distribuição decresce mais abruptamente do que o previsto.

Apesar de existirem muitos modelos para descrever a distribuição do retorno de preços para diferentes ativos financeiros, resultados conclusivos ainda não foram obtidos e requerem uma quantidade de dados muito grande para se estudar eventos raros que dão origem às caudas da distribuição [15].

## 2.4 Correlação Temporal em Séries Financeiras

Nos mercados de capitais, muitas ações são negociadas simultaneamente. Uma maneira de detectar similaridades e diferenças na evolução temporal síncrona de um par de ações  $i$  e  $j$  é estudar o coeficiente de correlação  $\rho_{ij}$  entre retornos de preços (eq. (2.23)).

O coeficiente de correlação é definido [4] por:

$$\rho_{ij}(\Delta t) = \frac{\langle R_i R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{(\langle R_i^2 \rangle - \langle R_i \rangle^2)(\langle R_j^2 \rangle - \langle R_j \rangle^2)}} \quad i, j = 1, \dots, n, \quad (2.30)$$

sendo  $n$  o número de ações,  $i, j$  os rótulos das ações e  $\Delta t$  o intervalo de tempo escolhido para o cálculo dos retornos. Utilizamos os preços de fechamento para se calcular o retorno dos preços.

Os coeficientes de correlação podem assumir valores no intervalo de  $-1$  a  $1$ , com três valores especiais:

$$\rho_{ij} = \begin{cases} 1, & \text{mudanças no preço completamente correlacionadas,} \\ 0, & \text{não correlacionadas,} \\ -1, & \text{mudanças completamente anti-correlacionadas.} \end{cases} \quad (2.31)$$

A matriz de correlação é uma matriz  $n \times n$  cujos elementos são os coeficientes de correlação  $\rho_{ij}$ . Essa matriz é uma excelente ferramenta utilizada no estudo do comportamento coletivo de sistemas caracterizados pela presença de um grande número de elementos interagindo entre si em um dado intervalo de tempo. Séries temporais com muitas variáveis são detectadas e obtidas em experimentos e pelo monitoramento de um grande número de sistemas físicos, biológicos e econômicos [16]. Entretanto, ao se analisar séries temporais concentrando-se no estudo de correlações cruzadas entre os elementos do sistema, depara-se com incertezas estatísticas, que resultam do fato das séries terem comprimento finito, do ruído devido à interação do sistema com o ambiente e também da complexidade intrínseca das interações entre os elementos do sistema. Uma tentativa de superar esses problemas é filtrar informações confiáveis estatisticamente da matriz de correlação [17].

No caso de séries temporais do retorno de preço de ações, a matriz de correlação pode conter informações sobre os setores econômicos das companhias consideradas. Análises teóricas e empíricas recentes têm mostrado que essas informações podem ser detectadas usando-se uma variedade de métodos [16]. Neste estudo, vamos considerar métodos baseados na Teoria da Matriz Aleatória, na Teoria de Grafos e utilizar os conceitos de espaço ultramétrico.

## 2.5 Teoria da Matriz Aleatória

*M*atrizes aleatórias apareceram pela primeira vez nos anos 30 do século passado, mas não atraíram muita atenção nessa época [18]. A Teoria da Matriz Aleatória preocupa-se com a seguinte questão: considere uma grande matriz cujos elementos são variáveis aleatórias distribuídas segundo uma dada densidade de probabilidade. O que podemos dizer sobre a probabilidade de alguns de seus autovalores ou de seus autovetores? Esta questão é pertinente para a compreensão do comportamento estatístico de ressonâncias de nêutrons a baixas velocidades em física nuclear, questão proposta em 1950 e intensivamente estudada por físicos.

Naquela época, estava disponível uma grande quantidade de dados espectroscópicos sobre os níveis de energia de grandes núcleos, mas eram muito complexos para serem explicados através dos modelos existentes, uma vez que a natureza exata das interações entre os nucleons era desconhecida. A Teoria da Matriz Aleatória foi desenvolvida inicialmente a fim de lidar com a estatística dos níveis de energia dos sistemas quânticos complexos. Com a suposição de que o Hamiltoniano que descreve um núcleo pesado é dado por uma matriz simétrica real com elementos aleatórios independentes distribuídos segundo uma dada densidade de probabilidade, uma série de previsões importantes foram feitas e testadas com sucesso no estudo de núcleos complexos (ver referências citadas em [19]). As previsões feitas pela Teoria da Matriz Aleatória representam uma média sobre todas as possíveis interações. Desvios das previsões da teoria indicam propriedades não-aleatórias específicas do sistema considerado, fornecendo indicações sobre a natureza das interações.

Mais tarde, a Teoria da Matriz Aleatória ganhou importância em outras áreas da física e da matemática, como por exemplo, na caracterização de sistemas caóticos, condutividade em metais desordenados, distribuição dos valores da função Zeta de Riemann,

enumeração de permutações que possuem certas particularidades, cromodinâmica quântica, dentre outros [18].

No estudo da estatística dos autovalores de matrizes aleatórias, três classes de matrizes com variáveis distribuídas segundo uma densidade de probabilidade Gaussiana desempenham papéis importantes [20]:

- *Ensemble* Ortogonal Gaussiano (GOE): matrizes são simétricas e seus elementos são números reais,
- *Ensemble* Unitário Gaussiano (GUE): matrizes com elementos Hermitianos complexos,
- *Ensemble* Simplético Gaussiano (GSE): matrizes com elementos Hermitianos auto-duais.

Neste trabalho, consideramos apenas matrizes reais simétricas.

Uma forma de se construir uma matriz simétrica real  $M$  ( $n \times n$ ), cujos elementos  $x_{ij}$  são distribuídos independentemente por densidades de probabilidade Gaussianas é dada por

$$P(x_{ii}) = \frac{1}{\sqrt{2\pi}} e^{-x_{ii}^2/2} \quad \text{com } \mu = 0, \quad \sigma = 1, \quad (2.32)$$

$$P(x_{ij}) = \frac{1}{\sqrt{\pi}} e^{-x_{ij}^2} \quad \text{com } \mu = 0, \quad \sigma = 1/\sqrt{2}. \quad (2.33)$$

Assim, a distribuição conjunta dos  $n(n-1)/2$  elementos independentes de  $M$  é o produto das distribuições individuais e pode ser escrita em uma forma composta

$$P(M) = A_n e^{-tr(M^2)/2}, \quad (2.34)$$

sendo  $A_n$  uma constante de normalização.

## 2.5.1 Teoria da Matriz Aleatória na Análise de Séries Financeiras

As dificuldades em analisar o peso e o significado do coeficiente de correlação  $\rho_{ij}$  calculado são devidas a muitas razões [21], dentre as quais:

- Condições de mercado mudam com o tempo e correlações existentes entre pares de ações podem não ser estacionárias,
- O comprimento finito das séries temporais usadas para obter as correlações introduz ruído.

Se usarmos séries temporais longas a fim de driblar o problema devido ao comprimento finito, nossas estimativas serão afetadas pela não estacionariedade dos coeficientes de correlação. Por essas razões, medidas empíricas dos coeficientes vão conter ruído, e é difícil, em geral, estimar os coeficientes que não resultam da aleatoriedade, a partir da matriz de correlação.

Uma maneira de identificar ações que permanecem correlacionadas (em média) no intervalo de tempo estudado é testar a estatística da matriz de correlação (obtida usando as séries reais) em relação a uma matriz de correlação aleatória, construída a partir de séries temporais mutuamente descorrelacionadas. Se as propriedades da matriz de correlação dos dados reais obedecerem às das matrizes aleatórias, então isso significa que as medidas empíricas da matriz de correlação serão aleatórias. Por outro lado, se as propriedades da matriz de correlação diferirem daquelas da matriz aleatória, então informações sobre correlações reais podem ser estimadas.

Consideremos a matriz de correlação aleatória dada por [21]

$$R = \frac{1}{T}AA^T, \quad (2.35)$$

onde  $A$  é uma matriz  $n \times T$  contendo  $n$  séries temporais de  $T$  elementos aleatórios com média zero e variância 1, extraídas de uma distribuição Gaussiana, que são mutuamente não-correlacionados.  $A^T$  é a transposta de  $A$ .

As propriedades estatísticas das matrizes aleatórias tais como  $R$  são bem conhecidas. Particularmente, no limite  $n \rightarrow \infty$ ,  $T \rightarrow \infty$ , de maneira que a razão  $Q \equiv \frac{T}{n} > 1$  é fixa, a função densidade de probabilidade dos autovalores  $\lambda$  da matriz de correlação aleatória  $R$  é dada por

$$P_R(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \quad (2.36)$$

para  $\lambda$  no intervalo  $\lambda_{min} \leq \lambda \leq \lambda_{max}$ , sendo  $\lambda_{min}$  e  $\lambda_{max}$ , respectivamente, os autovalores mínimo e máximo da matriz  $R$ , dados por

$$\lambda_{min}^{max} = \sigma^2 \left( 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \right). \quad (2.37)$$

Para  $n$  e  $T$  finitos, o corte abrupto observado em  $P_R(\lambda)$  é substituído por um decaimento um pouco mais suave.

A eq. (2.36) é exata para o caso de elementos da matriz  $R$  distribuídos segundo uma densidade de probabilidade Gaussiana. Numericamente, observa-se que séries temporais com elementos distribuídos segundo leis de potência com expoentes idênticos aos encontrados para séries de retornos de preços, mostram boa concordância com a eq. (2.36) [22].

## 2.6 Teoria de Grafos

A Teoria de Grafos <sup>10</sup> tem sido amplamente utilizada na análise de sistemas complexos [17]. Ela é usada para extrair informações de sistemas que apresentam elementos correlacionados entre si. Considere um sistema de  $n$  elementos. O coeficiente de correlação pode ser interpretado como a magnitude da ligação conectando pares de elementos. Em outras palavras, o sistema pode ser descrito por uma rede de elementos conectados entre si, por um peso associado às ligações de acordo com o coeficiente de correlação.

A idéia é extrair um sub-grafo da rede completa considerando os pesos das ligações entre os elementos. O objetivo deste procedimento é transferir as informações contidas na matriz de correlação do sistema em uma estrutura topológica do sub-grafo e então explorar técnicas da Teoria de Grafos para analisar tais informações.

Um sub-grafo da rede completa amplamente utilizado é a árvore de extensão mínima (*minimal spanning tree*). O número de ligações da árvore é  $n - 1$  para um sistema de  $n$  elementos e a árvore é um grafo conectado.

Estudos feitos utilizando árvores de extensão mínima revelam ser este um método excelente para estudar sistemas financeiros [24]. Um resultado interessante obtido ao se investigar as propriedades da árvore de extensão mínima está relacionado ao comportamento do mercado nas proximidades de uma quebra nas Bolsas como, por exemplo, a segunda-feira negra de 1987, onde uma mudança topológica na árvore foi observada [25], [26].

---

<sup>10</sup>Grafo é um conjunto de pontos ou vértices e um conjunto de linhas unindo todos ou alguns desses pontos [23].



### 2.6.1 Distância entre Ações

O coeficiente de correlação  $\rho_{ij}$  quantifica o grau de sincronização de um par de ações  $i$  e  $j$ . Este conceito é útil de duas maneiras diferentes: ele permite definir uma distância métrica que fornece uma medida relativa entre ações de uma dada carteira e fornece um método para extrair informações econômicas contidas nas séries temporais dos retornos de preço das ações [4].

O coeficiente de correlação não pode ser usado como uma distância métrica entre duas ações, pois pode assumir valores negativos. Essa distância é determinada a partir da distância Euclidiana entre vetores, que no nosso caso, são as séries temporais dos retornos  $\vec{R}_i$ . Os três axiomas que definem uma distância métrica Euclidiana são:

$$\begin{aligned}
 (i) \quad d_{ij} &= 0 \iff i = j, \\
 (ii) \quad d_{ij} &= d_{ji}, \\
 (iii) \quad d_{ij} &\leq d_{ik} + d_{kj}.
 \end{aligned}
 \tag{2.38}$$

Entretanto, uma métrica generalizada pode ser obtida usando-se como distância uma função apropriada do coeficiente de correlação. Normalizando cada série temporal, temos:

$$\tilde{R}_i = \frac{R_i - \langle R_i \rangle}{\sigma_{R_i}},
 \tag{2.39}$$

sendo  $\sigma_{R_i}$  o desvio padrão do retorno. A distância entre vetores é representada por  $d_{ij} = |\tilde{R}_i - \tilde{R}_j|$ . Desta forma,

$$d_{ij}^2 = |\tilde{R}_i - \tilde{R}_j|^2 = |\tilde{R}_i|^2 + |\tilde{R}_j|^2 - 2\tilde{R}_i\tilde{R}_j = 2 - 2\rho_{ij},
 \tag{2.40}$$

onde  $\rho_{ij} = \widetilde{R}_i \widetilde{R}_j$  e  $\rho_{ii} = |\widetilde{R}_i|^2 = 1$ . Assim, a distância entre duas ações em função de seus coeficientes de correlação pode ser escrita como

$$d_{ij} = \sqrt{2(1 - \rho_{ij})}. \quad (2.41)$$

Com essa escolha,  $d_{ij}$  satisfaz os axiomas de uma métrica Euclidiana (eq. (2.38)), *(i)* é satisfeita se a correlação é completa ( $\rho_{ij} = 1$  apenas se duas ações forem descritas por processos estocásticos idênticos, quer dizer, se  $i = j$ ), a condição *(ii)* é válida porque a matriz de correlação ou a matriz de distância é simétrica por definição, e a condição *(iii)* é naturalmente satisfeita pela própria definição de distância métrica Euclidiana.

O conhecimento da matriz de distância dos  $n$  objetos é geralmente usada para decompor o conjunto em subconjuntos de objetos mais estreitamente correlacionados. Para obter tal configuração, assume-se a hipótese sobre o espaço dos  $n$  objetos, segundo a qual uma distância ultramétrica sub-dominante descreve o arranjo das ações em uma carteira.

## 2.6.2 Espaços Ultramétricos

Um espaço ultramétrico é um espaço no qual a distância entre dois objetos é dada pela distância ultramétrica. Uma distância ultramétrica  $d_{ij}$  deve satisfazer as propriedades *(i)* e *(ii)* da eq. (2.38), enquanto a desigualdade triangular *(iii)* é substituída por uma desigualdade mais forte, chamada de desigualdade ultramétrica [4]

$$d_{ij} \leq \max\{d_{ik}, d_{kj}\}. \quad (2.42)$$

Espaços ultramétricos fornecem uma maneira de descrever sistemas complexos hierarquicamente estruturados, uma vez que o conceito de ultrametricidade está diretamente ligada ao conceito de hierarquia.

Uma vez que a distância métrica entre  $n$  objetos existe, muitos espaços ultramétricos podem ser obtidos através de diferentes separações de um dado conjunto de objetos. Dentre todas as estruturas ultramétricas possíveis associadas com a distância métrica  $d_{ij}$ , uma se destaca pela sua simplicidade e propriedades importantes: a ultramétrica sub-dominante.

Considerando a existência de uma espaço métrico, no qual  $n$  objetos estão ligados entre si, a estrutura ultramétrica sub-dominante pode ser determinada calculando-se a árvore de extensão mínima conectando os objetos. Em um grafo conectado cujas ligações possuem pesos, a árvore de extensão mínima é uma árvore que minimiza a soma total das distâncias entre os  $n$  objetos. A investigação de medidas ultramétricas sub-dominantes nos permite determinar de uma maneira única a hierarquia do sistema. O método utilizado para a construção da árvore de extensão mínima conectando os elementos do sistema, conhecido como Algoritmo de Kruskal, é simples e direto.

O Algoritmo de Kruskal pode ser resumido da seguinte maneira [23]:

**Passo 1** *Comece com um grafo completamente desconectado  $T$  de  $n$  vértices.*

**Passo 2** *Construa uma lista ordenada de maneira crescente de pares de ações  $L_{ord}$  com todos os pares possíveis de acordo com as distâncias  $d_{ij}$ . O primeiro par de  $L_{ord}$  possui a menor distância.*

**Passo 3** *Começando do topo da lista, adicione ligações a  $T$  de maneira que essas ligações não formem laços.*

**Passo 4** Repita o **Passo 3**, analisando  $L_{ord}$  e escolhendo pares, até que  $n - 1$  ligações sejam feitas. Assim  $T$  é a árvore de extensão mínima do grafo.

Nas figuras a seguir (Figuras 2.3 a 2.14), apresentamos um exemplo de construção da árvore de extensão mínima usando o algoritmo de Kruskal.

No Capítulo 3, aplicaremos esta técnica na análise de séries temporais dos retornos de preço de uma carteira selecionada de ações negociadas na Bovespa.

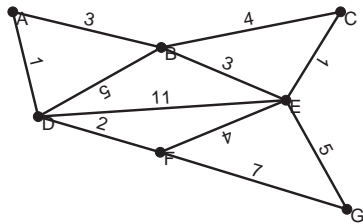


Figura 2.3: Grafo inicial. Números indicam os valores das ligações entre os vértices. Outras possíveis ligações não foram consideradas para simplificar o exemplo.

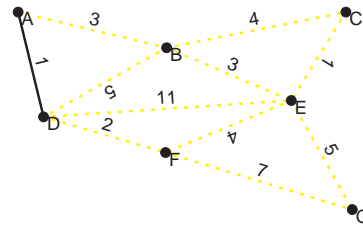


Figura 2.4: As ligações AD e CE são as menores. O algoritmo seleciona arbitrariamente AD.

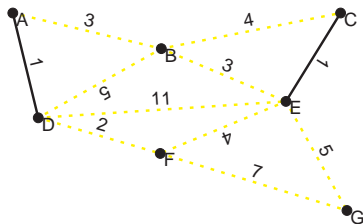


Figura 2.5: CE é escolhido.

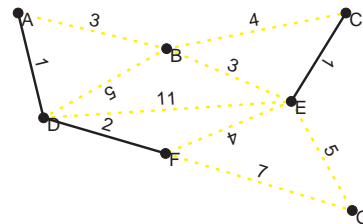


Figura 2.6: DF é selecionado.

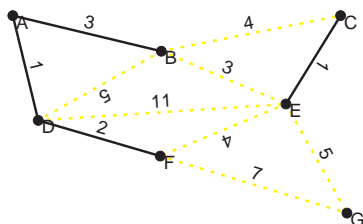


Figura 2.7: Os menores arcos são AB e BE. AB é escolhido arbitrariamente.

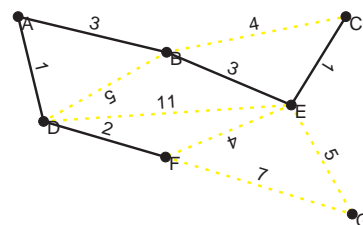


Figura 2.8: Em seguida, BE é selecionado.

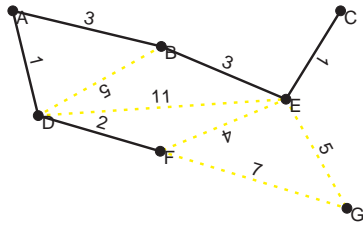


Figura 2.9: Os menores arcos são BC e EF. BC não é adicionado ao grafo, pois formaria o laço BCE.

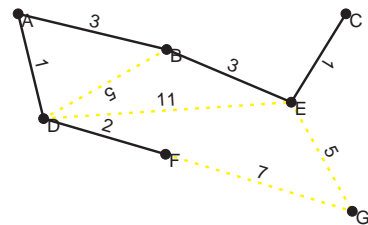


Figura 2.10: EF também é excluído, uma vez que formaria o laço ABEFD.

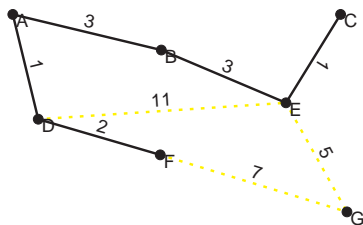


Figura 2.11: Considerando-se os arcos BD e EG, BD não é adicionado, pois formaria o laço ABD.

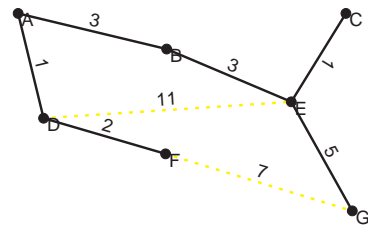


Figura 2.12: EG é selecionado.

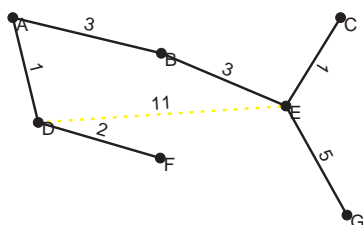


Figura 2.13: FG é excluído para não formar laço.

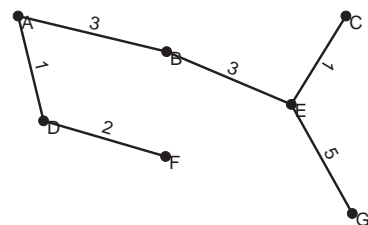


Figura 2.14: DE é excluído do processo e temos a árvore resultante de extensão mínima.

---

## CAPÍTULO 3

### Análise dos Dados

---

*I*nvestigamos as propriedades estatísticas do retorno de preços para um conjunto de 43 ações negociadas na Bolsa de Valores de São Paulo. Os dados foram obtidos gratuitamente na página [www.analistedemercado.com.br](http://www.analistedemercado.com.br) e compreendem o período de 02-01-2001 a 18-06-2007, totalizando 1603 dias de negociações. Essas ações estão dentre aquelas mais negociadas na Bovespa e pertencem a 8 setores econômicos diferentes de acordo com a classificação que se encontra na página [www.bovespa.com.br](http://www.bovespa.com.br).

A Tabela 3.1 lista as 43 empresas estudadas, símbolos de negociação, setores e sub-setores aos quais pertencem. Sub-setores econômicos são mostrados a fim de especificar a atividade das empresas pertencentes a um mesmo setor. Cada símbolo é composto por quatro letras que indicam a empresa e um número que indica o tipo da ação: 3 representa ações ordinárias nominativas (ON), 4 indica ações preferenciais nominativas (PN), 5 ações preferenciais nominativas série A (PNA) e 6 representa ações preferenciais nominativas série B (PNB). Em nosso estudo, estamos considerando 3 ações do setor Petróleo, Gás e Biocombustíveis, 11 do setor Minerais Básicos, 1 papel do setor Bens Industriais, 4 do setor Consumo não-Cíclico, 2 do Consumo Cíclico, 7 ações de Telecomunicações, 11 de Utilidade Pública e 4 papéis do setor Financeiro e Outros.

<b>Nome de Pregão</b>	<b>Símbolo</b>	<b>Setor econômico</b>	<b>Sub-setor</b>	
Petrobras	PETR3, PETR4	Petróleo, Gás e Biocombustíveis	Petróleo, Gás e Biocombustíveis	
Ipiranga Pet	PTIP4			
Vale R Doce	VALE3, VALE5	Materiais Básicos	Mineração	
Acesita	ACES4		Siderurgia e Metalurgia	
Gerdau	GGBR4			
Gerdau Met	GOAU4			
Sid Nacional	CSNA3			
Usiminas	USIM5			
Braskem	BRKM5		Químicos	
Aracruz	ARCZ6		Madeira e Papel	
Klabin S/A	KLBN4			
Votorantim	VCPA4			
Embraer	EMBR3		Bens Industriais	Material de Transporte
Sadia	SDIA4		Consumo não-Cíclico	Alimentos
Ambev	AMBV4	Bebidas		
Souza Cruz	CRUZ3	Fumo		
P.Açucar - CBD	PCAR4	Comércio		
Lojas Americ	LAME4	Consumo Cíclico		Comércio
NET	NETC4		Mídia	
Brasil T Par	BRTP3, BRTP4	Telecomunicações	Telefonia Fixa	
Brasil Telec	BRTO4			
Telemar	TNLP3, TNLP4			
Telemig Part	TMCP4		Telefonia Móvel	
Vivo	VIVO4			
Celesc	CLSC6	Utilidade Pública	Energia Elétrica	
Cemig	CMIG3, CMIG4			
Copel	CPL6			
Cesp	Cesp6			
Eletrobras	ELET3, ELET6			
Light S/A	LIGT3			
Tran Paulist	TRPL4		Água e Saneamento	
Sabesp	SBSP3			
Comgas	CGAS5		Gás	
Bradesco	BBDC4		Financeiro e Outros	Intermediários Financeiros
Itaubanco	ITAU4			
Itausa	ITSA4	<i>Holdings</i> Diversificadas		
Bradespar	BRAP4			

Tabela 3.1: Lista das 43 empresas estudadas. A primeira coluna identifica as empresas pelo nome de pregão, a segunda coluna indica o símbolo de negociação da empresa que é composto por 4 letras e um número: 3 indica ações do tipo ON, 4 PN, 5 PNA e 6 PNB. A terceira coluna mostra o setor e a quarta o sub-setor econômico ao qual a companhia pertence.



---

A Figura 3.1 ilustra a série temporal do preço da ação Aracruz (ARCZ6).  $S(t)$  é o preço, em reais, da ação numa determinada data, de janeiro de 2001 a junho de 2007. Vale ressaltar que os preços não foram corrigidos, ou seja, não foram trazidos para o valor presente considerando a inflação do período. A partir dos preços de fechamento, calculamos os retornos através da eq. (2.23), que são mostrados na Figura 3.2 usando-se o intervalo  $\Delta t = 1$  dia.

Selecionamos séries do retorno de preços para ações de cada um dos setores econômicos: Petróleo, Gás e Biocombustíveis (PETR4), Materiais Básicos (ARCZ6), Bens Industriais (EMBR3), Consumo não-Cíclico (PCAR4), Consumo Cíclico (LAME4), Telecomunicações (BRTO4), Utilidade Pública (CGAS5), Financeiros e Outros (ITAU4). Dividimos o intervalo total dos valores possíveis do retorno em caixas igualmente espaçadas de largura  $\Delta R = 1/1603$  e calculamos a densidade de probabilidade dos retornos. O resultado está ilustrado nas Figuras 3.3 e 3.4, que mostram a distribuição de probabilidade do retorno de preços. Ajustamos uma distribuição Gaussiana sobre os dados. Observamos que a distribuição para os dados reais decai mais suavemente que a distribuição Gaussiana, enquanto que o pico da distribuição é mais pronunciado, confirmando resultados obtidos na literatura (ver citações e discussão em [4]). Os coeficientes para a distribuição Gaussiana são ajustados com um nível de confiança de 95%.

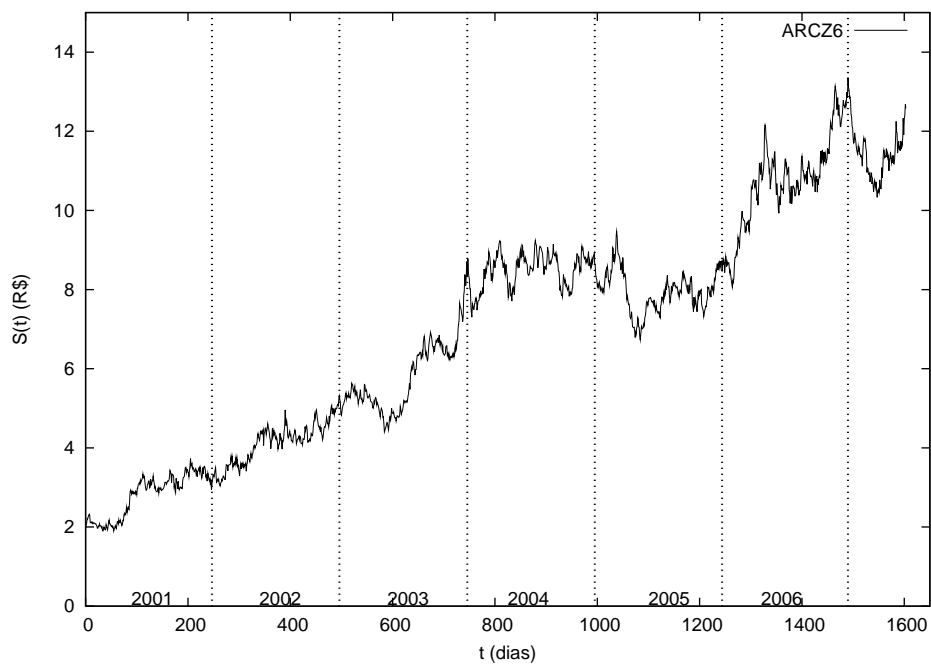


Figura 3.1: Evolução temporal do preço  $S(t)$ , em reais, da ação ARCZ6 no período de janeiro de 2001 a junho de 2007. Está indicado o intervalo correspondente a cada ano estudado.

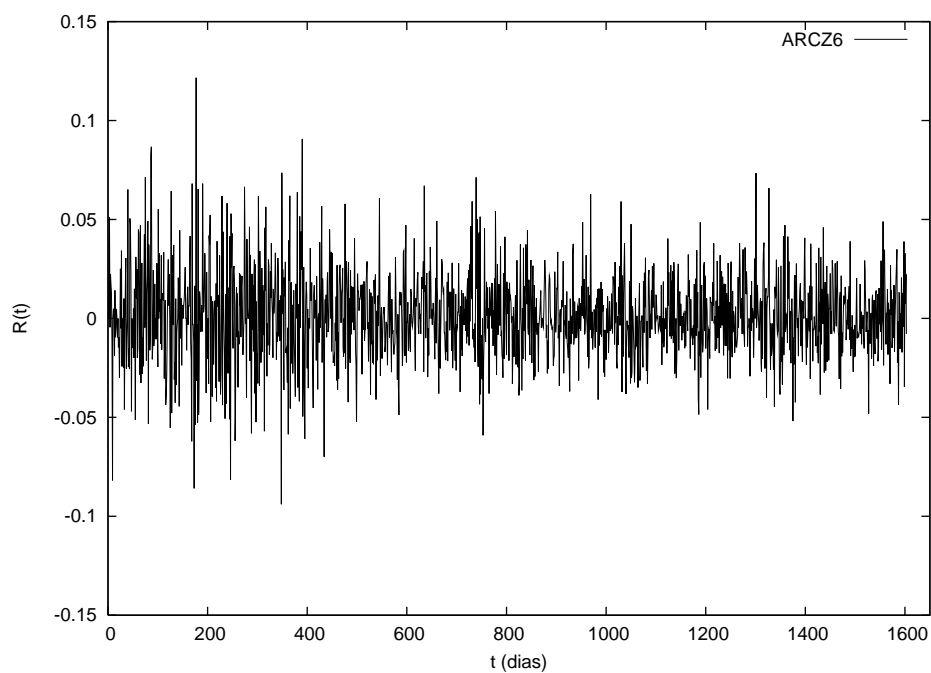


Figura 3.2: Série temporal do retorno de preço da ação ARCZ6 de janeiro de 2001 a junho de 2007.

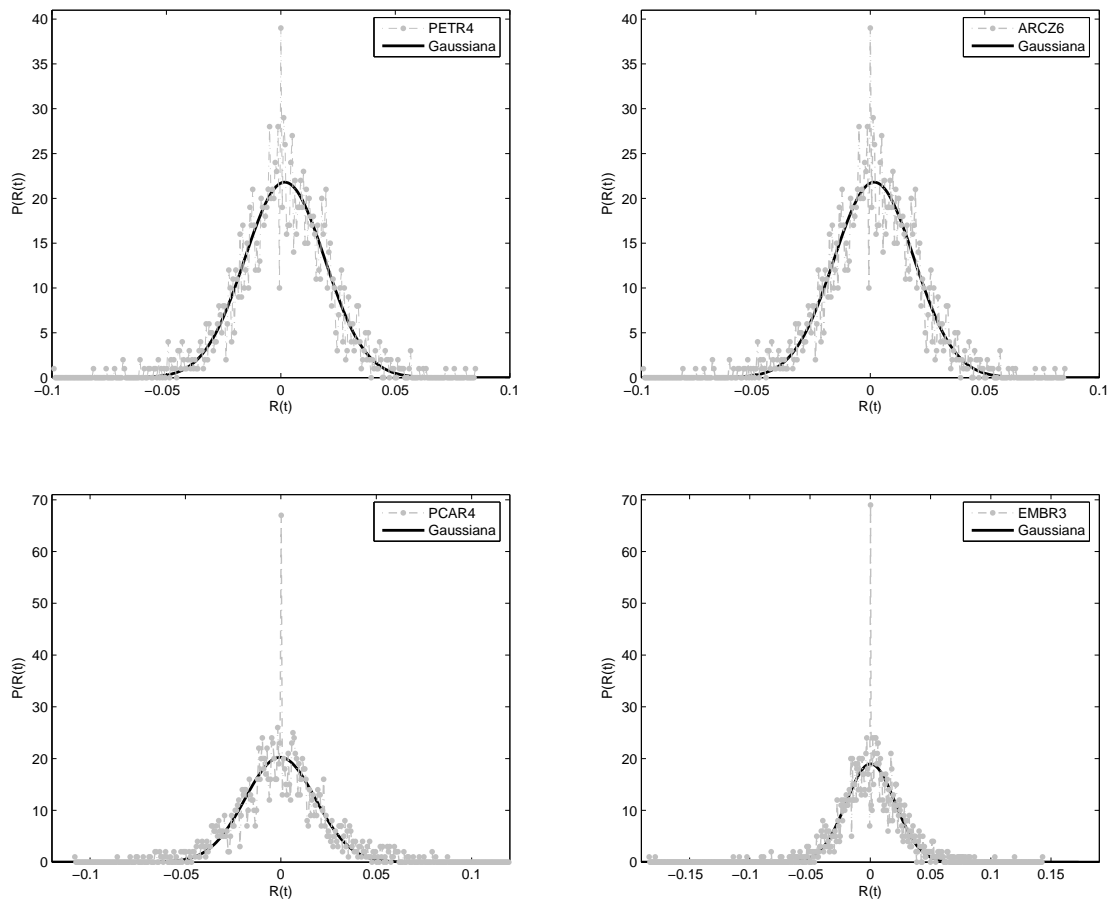


Figura 3.3: Distribuição de probabilidade do retorno de preço para ações selecionadas dos setores Petróleo, Gás e Biocombustíveis (PETR4), Materiais Básicos (ARCZ6), Bens Industriais (EMBR3) e Consumo não-Cíclico (PCAR4).

### 3.1 Análise dos coeficientes de correlação

Usando a eq. (2.30), calculamos os coeficientes de correlação entre retornos síncronos para todas as séries temporais. A matriz dos coeficientes de correlação é uma matriz simétrica  $n \times n$ , sendo  $n = 43$  papéis estudados, cujos elementos são os coeficientes  $\rho_{ij}$  e  $\rho_{ii} = 1$  na diagonal principal. Temos um conjunto de  $n(n-1)/2$  coeficientes distintos. O cálculo foi feito para todo o período compreendido entre janeiro de 2001 e

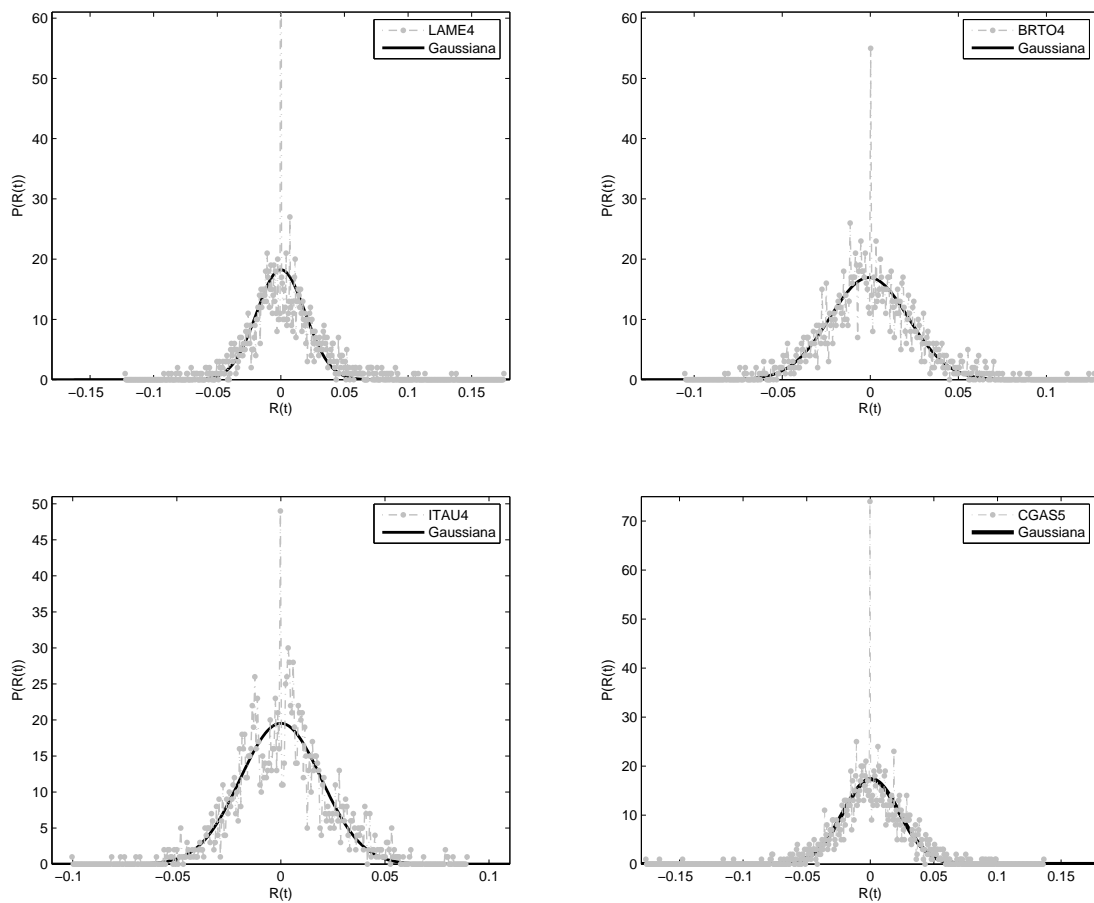


Figura 3.4: Distribuição de probabilidade do retorno de preço para ações escolhidas dos setores Consumo Cíclico (LAME4), Telecomunicações (BRTO4), Utilidade Pública (CGAS5), Financeiros e Outros (ITAU4).

junho de 2007 e para cada ano separadamente. A Tabela 3.2 mostra os coeficientes de correlação mínimo  $\rho_{ij}^{min}$  e máximo  $\rho_{ij}^{max}$  obtidos para cada ano estudado. Vale ressaltar que os  $\rho_{ij}^{max}$  obtidos para todos os anos são entre papéis da mesma companhia. Por exemplo, para o ano de 2001, o maior valor do coeficiente de correlação é  $\rho_{ij}^{max} = 0,92$  entre as séries PETR4 e PETR3, mas escolhemos mostrar os  $\rho_{ij}^{max}$  obtidos entre empresas diferentes.

A Figura 3.5 mostra duas séries do logaritmo natural dos preços, respectivamente, para as ações SBSP3 e VALE3, e BRTO4 e BRTP4. O primeiro par apresenta coeficiente de correlação  $\rho_{ij} = 0,38$  em 2001, anticorrelação nos anos de 2002 ( $\rho_{ij} = -0,14$ ) e 2003 ( $\rho_{ij} = -0,10$ ),  $\rho_{ij} = 0,34$  em 2004,  $\rho_{ij} = 0,27$  em 2005,  $\rho_{ij} = 0,40$  em 2006 e

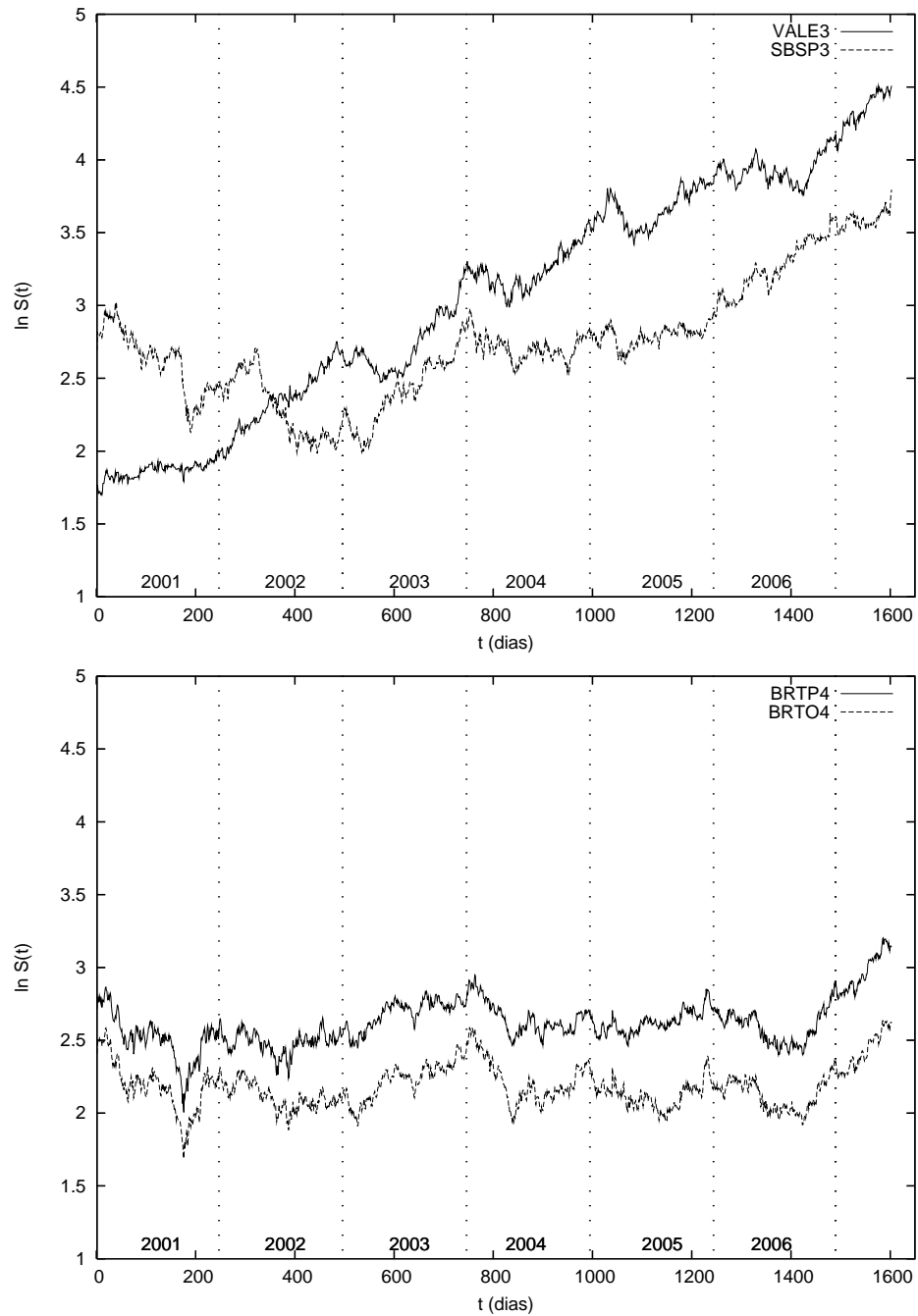


Figura 3.5: Séries do logaritmo dos preços para as ações SBSP3 e VALE3 e BRTO4 e BRTP4, respectivamente.

Ano	$\rho_{ij}^{min}$	i	j	$\rho_{ij}^{max}$	i	j
2001	-0,03	ARCZ6	CGAS5	0,86	BRTO4	BRTP4
2002	-0,14	SBSP3	VALE3	0,89	BRTP4	BRTO4
2003	-0,13	SBSP3	VALE5	0,85	BRTP4	BRTO4
2004	0,04	NETC4	PCAR4	0,86	GOAU4	GGBR4
2005	0,04	NETC4	PCAR4	0,87	GOAU4	GGBR4
2006	0,05	CRUZ3	EMBR3	0,91	GOAU4	GGBR4
2007	0,03	CRUZ3	EMBR3	0,93	GOAU4	GGBR4
01-07	0,03	ARCZ6	NETC4	0,85	BRTP4	BRTO4

Tabela 3.2: Coeficientes de correlação mínimo e máximo obtidos para cada ano estudado. São mostrados os  $\rho_{ij}^{max}$  obtidos para companhias  $i$  e  $j$  diferentes.

$\rho_{ij} = 0,41$  em 2007. Já as ações BRTO4 e BRTP4 apresentam coeficientes de correlação superiores a  $\rho_{ij} = 0,82$  em todos os anos estudados.

A seguir, aplicamos métodos baseados na Teoria da Matriz Aleatória, na Teoria de Grafos e vamos utilizar os conceitos de Espaço Ultramétrico a fim de analisar a correlação obtida a partir do retorno dos preços diários.

### 3.1.1 Teoria da Matriz Aleatória

No contexto da administração de carteiras de investimento, a Teoria da Matriz Aleatória é útil porque permite estimar o efeito de incertezas estatísticas da matriz de correlação. A teoria tem sido aplicada na investigação de matrizes de correlação dos retornos dos preços de ativos financeiros e tem sido mostrado que o espectro da matriz de uma carteira típica pode ser dividido em 3 classes de autovalores. O maior autovalor é totalmente incompatível com a eq. (2.36) discutida na Subseção 2.5.1 e descreveria o comportamento comum das ações que compõem a carteira. Este fato implica em outra hipótese, segundo a qual parte da matriz de correlação que é ortogonal ao autovetor correspondente ao maior autovalor é aleatória. Isto corresponde a quantificar a variância da parte não explicada pelo maior autovalor como

$$\sigma^2 = 1 - \frac{\lambda_{max}}{n}, \quad (3.1)$$

sendo  $n$  o número de autovalores encontrados, e usar este valor para recalculer novos  $\lambda_{min}$  e  $\lambda_{max}$  (utilizando novamente a eq.(2.37)), que chamaremos de  $\tilde{\lambda}_{min}$  e  $\tilde{\lambda}_{max}$ .

Com essa suposição, estudos anteriores mostraram que uma parte dos autovalores também é incompatível com a Teoria da Matriz Aleatória, uma vez que estão fora do intervalo  $]\tilde{\lambda}_{min}, \tilde{\lambda}_{max}[$ . Esses valores provavelmente descrevem informações econômicas contidas na matriz de correlação.

Para nossos dados,  $n = 43$ ,  $T = 1603$ ,  $Q = T/n = 37,3$ , e substituindo na eq. (2.36), obtemos  $\lambda_{max} = 1,35$ . Entretanto, ao calcularmos os autovalores da matriz de correlação, encontramos um autovalor  $\lambda_{43} = 16,6$ , aproximadamente 12 vezes maior que  $\lambda_{max}$  previsto, claramente incompatível com a Teoria da Matriz Aleatória. A fim de se levar em consideração o valor do maior autovalor encontrado, usamos a eq. (3.1) para encontrar  $\sigma^2$  e obtemos  $\tilde{\lambda}_{min} = 0,43$  e  $\tilde{\lambda}_{max} = 0,83$ .

É interessante observar que o valor de  $Q$  neste trabalho difere muito dos valores encontrados na literatura, devido ao número reduzido de ações que utilizamos. Apesar disso, os valores encontrados para os autovalores da matriz de correlação, bem como o intervalo  $[\lambda_{min}, \lambda_{max}]$  em nossa análise estão de acordo com aqueles vistos na literatura.

Plerou e colaboradores [21] examinaram retornos de preços calculados com  $\Delta t = 30$  min das  $n = 1000$  maiores empresas norte-americanas, cujas séries contém  $T = 6648$  dados coletados no período entre 1994 e 1995. Obtiveram, através da eq. (2.36),  $0,36 \leq \lambda \leq 1,94$ . Observaram que parte dos autovalores encontrados estão fora do intervalo previsto segundo a Teoria de Matriz Aleatória, tanto para  $\lambda < \lambda_{min}$  como para  $\lambda > \lambda_{max}$ . O maior autovalor encontrado por eles é  $\lambda_{1000} \approx 50$ , que é cerca de 25 vezes maior que  $\lambda_{max} = 1,94$ . Eles também analisaram  $n = 442$  ações norte-americanas, cujas

séries dos retornos diários possuem um comprimento  $T = 1737$ , no período entre 1990 e 1996. Obtiveram resultados semelhantes, sendo  $\lambda_{442} = 46,3$ .

Laloux e colaboradores [27] estudaram a densidade de autovalores da matriz de correlação para um conjunto de  $n = 405$  ativos que compõem o índice Standard and Poor's 500 (S & P 500). As correlações foram calculadas utilizando-se retornos diários de preços no período entre 1990 e 1996, totalizando  $T = 1309$  dias ( $Q = 3,22$ ). Eles compararam a densidade dos autovalores encontrada por eles com a prevista pela Teoria da Matriz Aleatória (eq. (2.36)) e obtiveram 94% do espectro situado abaixo do limite teórico  $\lambda_{max}$ . Enfatizaram que um ajuste melhor da distribuição dos autovalores reais poderia ter sido obtido usando-se um valor de  $Q$  menor, o que poderia ser explicado pela existência de correlações da volatilidade.

## Estatística dos autovetores

Desvios da distribuição dos autovalores da matriz de correlação em relação aos resultados previstos para matrizes aleatórias (eq. (2.36)) sugerem que esses desvios também devem estar presentes na estatística das componentes dos autovetores que correspondem a esses autovalores [27]. Nesta Subseção, vamos testar se os autovalores obtidos para o intervalo  $[\lambda_{min}, \lambda_{max}]$  estão de acordo com as previsões da Teoria da Matriz Aleatória, através da análise da distribuição das componentes dos autovetores da matriz de correlação obtida para o período de 2001 a 2007,  $T = 1603$ .



A distribuição das componentes  $\{u_i^k; i = 1 \dots n\}$  do autovetor  $u_k, \{k = 1 \dots n\}$  da matriz de correlação aleatória  $R$  (eq. (2.35)), descrita na Subseção 2.5.1, deve corresponder a uma distribuição Gaussiana com média 0 e variância 1 [28] dada por

$$\rho_{rm}(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right). \quad (3.2)$$

Comparamos a distribuição das componentes dos autovetores da matriz de correlação real com a eq. (3.2). A Figura 3.6 mostra as distribuições das componentes dos autovetores correspondentes a alguns autovalores no intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$  comparadas ao previsto pela Teoria da Matriz Aleatória. Observamos que essas densidades, bem como as densidades de grande parte dos autovetores cujos autovalores estão nesse intervalo, é bem próxima da prevista pela teoria.

Testamos a concordância da distribuição  $\rho(u_i)$  em relação a  $\rho_{rm}(u)$  para os autovalores no intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$  calculando-se a curtose (eq. (2.10)), que para a Gaussiana é  $\kappa = 3$ . Para as componentes do autovetor correspondente a  $\lambda_{16}$ , obtemos  $\kappa = 3,41$ , e o valor da curtose para alguns dos demais autovetores que estamos assumindo serem aleatórios são próximos a  $\kappa$  obtida para a Gaussiana. Entretanto, isso não é válido para todos os autovetores situados nesse intervalo. Uma possível explicação seria devido ao fato do nosso estudo considerar poucas ações.

A Figura 3.7 ilustra as densidades das componentes dos autovetores correspondentes aos autovalores  $\lambda_7 = 0,208$ ,  $\lambda_{16} = 0,480$ ,  $\lambda_{40} = 0,894$  e  $\lambda_{43} = 16,65$  comparadas a  $\rho_{rm}(u)$ . O autovalor  $\lambda_{16}$  está no intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$ , dado pela teoria que considera o limite  $n \rightarrow \infty, T \rightarrow \infty$ .

Observamos que a distribuição das componentes dos autovetores cujos autovalores  $\lambda_k \leq \tilde{\lambda}_{min}$  difere muito da distribuição teórica dada pela eq. (3.2), como pode ser visto na Figura 3.7, sugerindo que esses autovetores da matriz de correlação podem

conter informações reais. Isto também é válido para a distribuição das componentes dos autovetores cujos  $\lambda_k \geq \tilde{\lambda}_{max}$ .

É interessante notar que a distribuição das componentes do autovetor  $u_{43}$  difere sistematicamente da densidade de probabilidade das componentes aleatórias, sendo que todas aquelas componentes possuem o mesmo sinal, fazendo com que a distribuição esteja deslocada à direita em relação a  $\rho_{rm}(u)$ . Isto sugere que todos os participantes significativos desse autovetor são afetados por um mesmo fator.

Laloux e colaboradores [27] mostraram que a distribuição das componentes dos autovetores se ajusta muito bem à densidade de probabilidade teórica  $\rho_{rm}(u)$ , exceto para os autovetores correspondentes a  $\lambda_k \geq \tilde{\lambda}_{max}$ . Resultados semelhantes também foram obtidos por Plerou e colaboradores [21], que mostraram que os desvios nas densidades de probabilidade das componentes de  $u_k$  quando comparadas a  $\rho_{rm}(u)$  se dá para poucos autovalores (menores e maiores) que se encontram fora do intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$ .

Aplicando-se a Teoria da Matriz Aleatória é possível identificar parte da matriz de correlação que contém informação econômica. Testamos um método [29] usado para mostrar que alguns autovetores cujos autovalores não são explicados por essa teoria descrevem setores econômicos. O conjunto de ações investigadas é dividido em  $S$  setores  $s = 1, 2, \dots, S$  de acordo com os setores econômicos. Definimos uma matriz de projeção  $P$  com elementos  $P_{si} = 1/n_s$ , se a ação pertencer ao setor considerado ou  $P_{si} = 0$ , caso contrário, sendo  $n_s$  o número de ações pertencentes ao setor  $s$ .

Para cada autovetor  $u^k$ , calculamos

$$X_s^k \equiv \sum_{i=1}^{n=43} P_{si} [u_i^k]^2. \quad (3.3)$$

Esta grandeza fornece uma medida da importância de um dado setor  $s$  na composição do autovalor  $u^k$ .

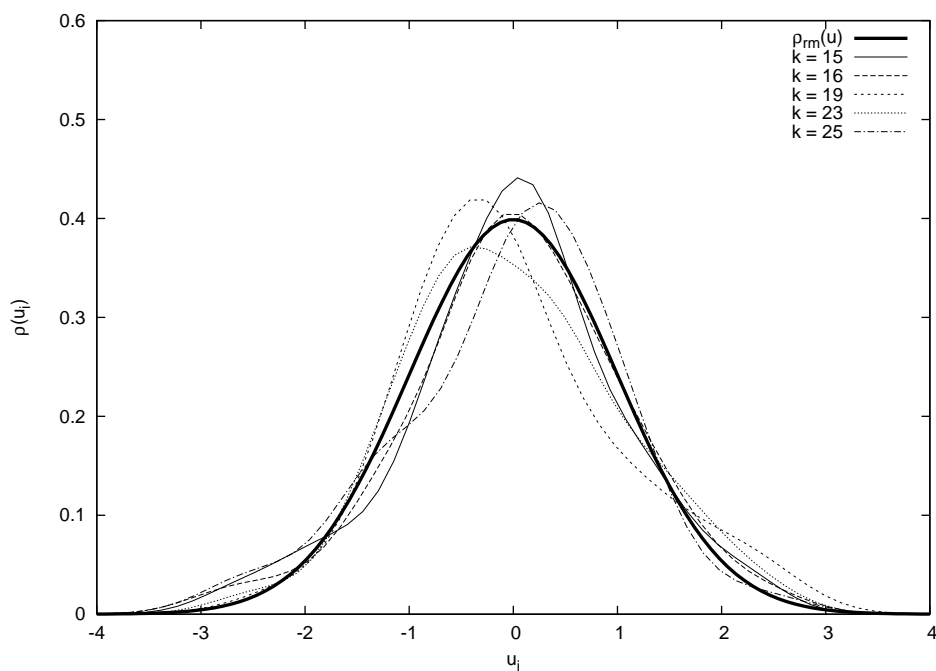


Figura 3.6: Densidades das componentes dos autovetores correspondentes a alguns autovalores no intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$  comparadas a  $\rho_{rm}(u)$ . Estão indicados os índices  $k$  dos autovetores  $u_k$  analisados.

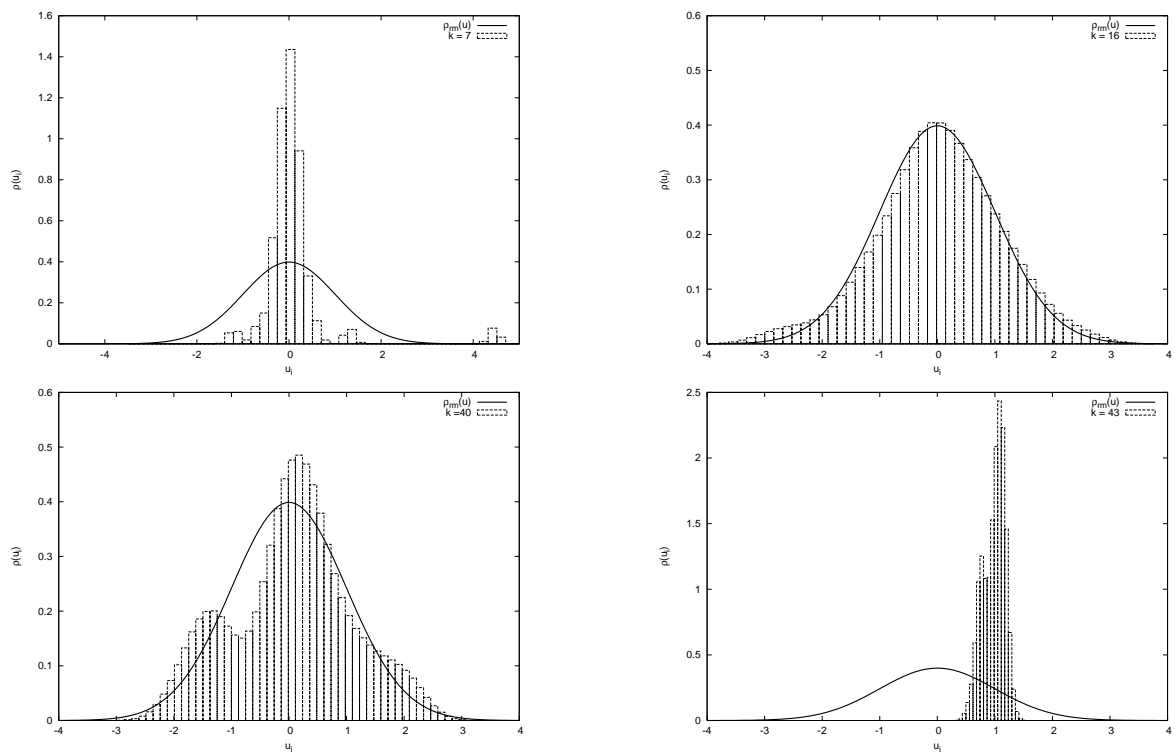


Figura 3.7: Densidade de probabilidade das componentes de alguns autovetores da matriz de correlação comparada à distribuição teórica  $\rho_{rm}(u)$ . Estão ilustradas a distribuição das componentes dos autovetores correspondentes aos autovalores  $\lambda_7$ ,  $\lambda_{16}$ ,  $\lambda_{40}$  e  $\lambda_{43}$ , respectivamente.

A Figura 3.8 mostra a contribuição  $X_s^k$  para os autovetores correspondentes aos autovalores  $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_6$  e  $\lambda_{35}$ , como está indicado na figura. No eixo  $s$  estão indicados os setores econômicos,  $s = 1$  indica o setor Petróleo, Gás e Biocombustíveis,  $s = 2$  Materiais Básicos,  $s = 3$  Bens Industriais,  $s = 4$  Consumo não-Cíclico,  $s = 5$  Consumo Cíclico,  $s = 6$  Telecomunicação,  $s = 7$  Utilidade Pública e  $s = 8$  Financeiro e Outros.

Observamos que o autovetor  $u_{k=1}$  está associado, predominantemente, ao setor Gás, Petróleo e Biocombustíveis, enquanto o setor Utilidade Pública está associado ao autovetor  $u_{k=2}$ . Já o autovetor  $u_{k=3}$  mostra uma grande contribuição do setor Materiais Básicos  $s = 2$ . O setor Telecomunicações está associado ao autovetor  $u_{k=4}$ . O autovetor correspondente ao autovalor  $\lambda_6 = 0,19$  está associado ao setor Financeiro e Outros.

É interessante notar que para os demais autovalores  $\lambda_k \leq \tilde{\lambda}_{min}$ , observamos algum setor econômico predominantemente associado ao autovetor correspondente. Para os autovalores no intervalo  $\tilde{\lambda}_{min} \leq \lambda_k \leq \tilde{\lambda}_{max}$ , nenhum autovetor está associado a um dado setor econômico, como era esperado, uma vez que esses autovetores não carregam informação real, segundo a Teoria da Matriz Aleatória. Já para autovalores  $\lambda_k \geq \tilde{\lambda}_{max}$ , não observamos nenhum setor econômico específico em destaque, mas uma mistura deles, como pode ser visto na Figura 3.8 para o autovetor  $u_{35}$ , que parece associado a dois setores, Bens Industriais  $s = 3$  e Consumo não-Cíclico  $s = 4$ , indicando, provavelmente, que esses são autovetores correspondentes a autovalores fortemente afetados por incertezas estatísticas.

A Figura 3.9 mostra a contribuição  $X_s^k$  para o autovetor correspondente ao maior autovalor  $\lambda_{43} = 16,6$ . Notamos que todos os setores econômicos estão associados de uma maneira similar a esse autovetor.

Uma vez que todas as componentes do autovetor correspondente ao maior autovalor contribuem de uma maneira similar, isto representa uma influência comum a todas as ações. Então, o maior autovalor quantifica a noção qualitativa de que certas notícias (fatos relevantes, variação da taxa de juros), afetam todas as ações da mesma maneira

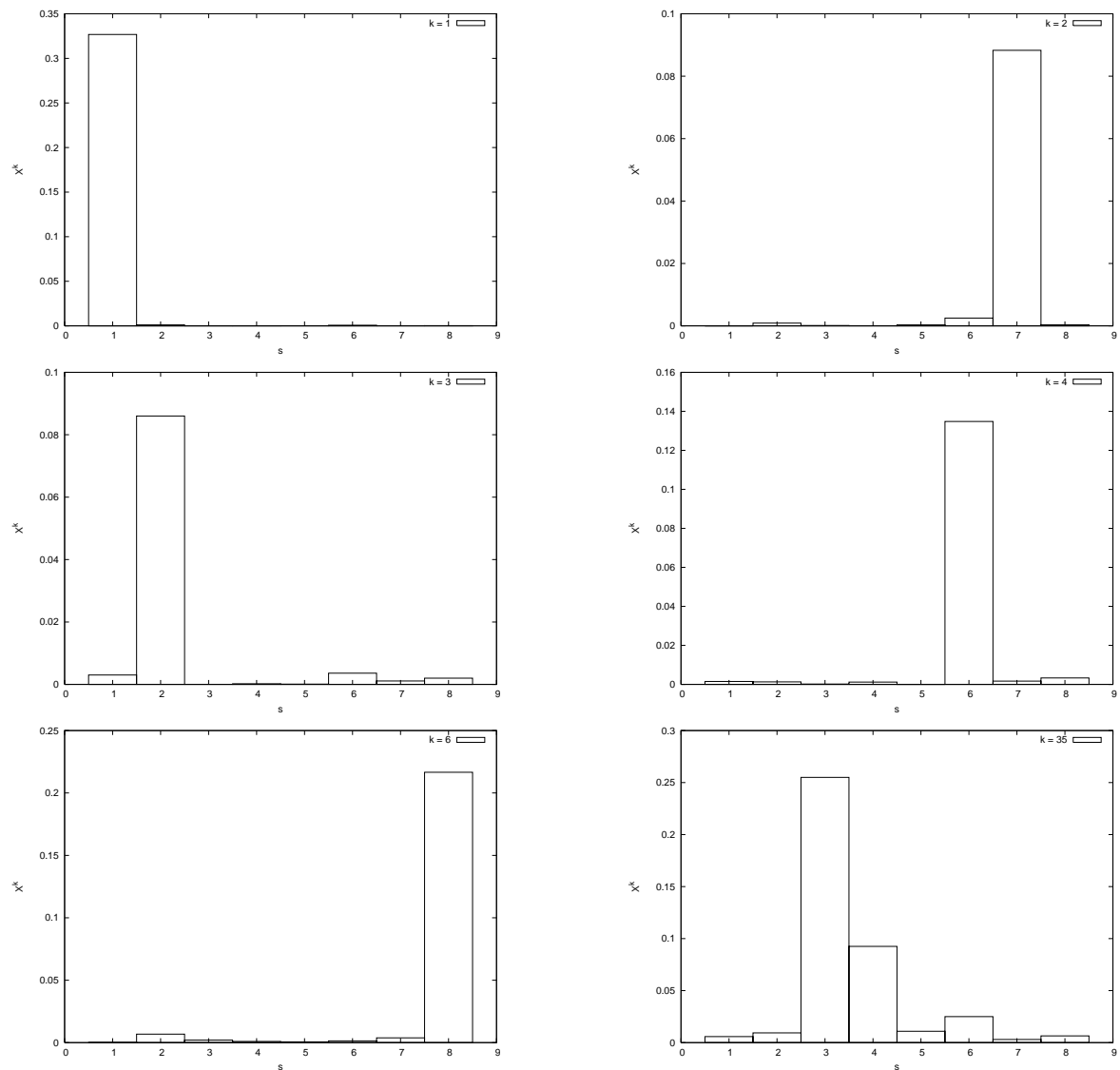


Figura 3.8: Contribuição  $X_s^k$  para alguns autovetores da matriz de correlação. No eixo  $s$  estão indicados os setores econômicos,  $s = 1$  indica o setor Petróleo, Gás e Biocombustíveis,  $s = 2$  Materiais Básicos,  $s = 3$  Bens Industriais,  $s = 4$  Consumo não-Cíclico,  $s = 5$  Consumo Cíclico,  $s = 6$  Telecomunicação,  $s = 7$  Utilidade Pública e  $s = 8$  Financeiro e Outros.

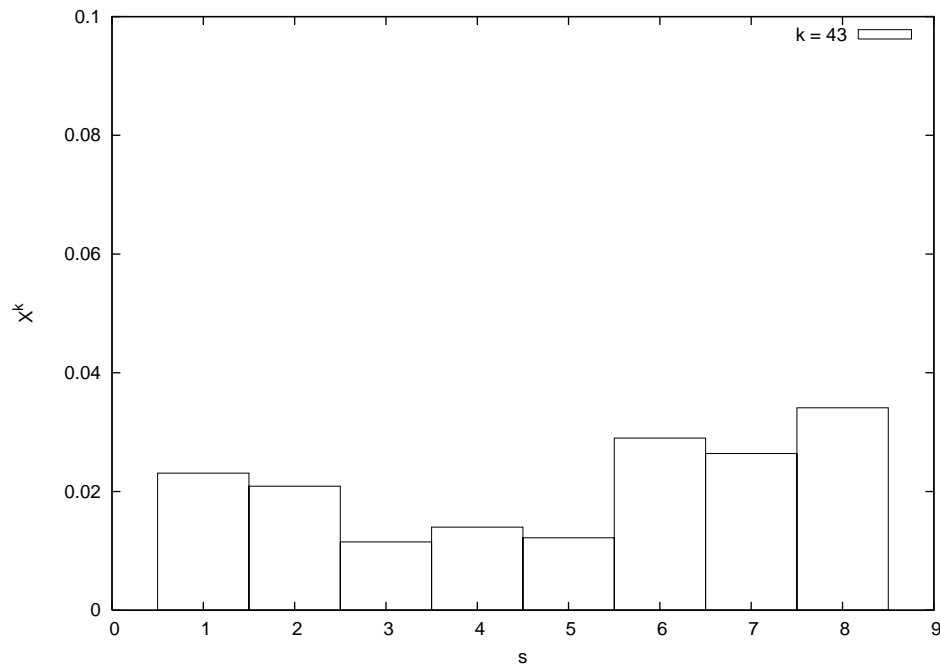


Figura 3.9: Contribuição  $X_s^k$  para o autovalor  $\lambda_{43}$ . Os setores econômicos são indicados por  $s$ , ver Figura 3.8.

[30]. Podemos também interpretar o maior autovalor e seu autovetor correspondente como uma resposta coletiva de todo o mercado frente a um estímulo.

Plerou e colaboradores [21] investigaram quantitativamente essa idéia através da comparação da projeção (produto escalar) das séries temporais sobre as componentes do autovetor correspondente ao maior autovalor com uma medida padrão do desempenho do mercado de ações norte-americano - os retornos do índice S&P500. Após a comparação, encontraram um comportamento muito similar entre as duas grandezas. Assim, interpretaram o maior autovalor como uma grandeza que quantifica grandes influências do mercado sobre todas as ações.

## Autovetores correspondentes aos menores autovalores

Vimos que os autovetores correspondentes aos menores autovalores carregam informação econômica. Analisando as componentes desses autovetores, observamos que as componentes significativas são relativas a pares de ações que possuem os maiores coeficientes de correlação  $\rho_{ij}$  da nossa amostra, em outras palavras, destacam-se pares de ações cujos  $\rho_{ij} \gg \langle \rho_{ij} \rangle$ .

Para o autovetor da matriz de correlação (calculada no período compreendido entre os anos de 2001 e 2007) correspondente ao menor autovalor, as maiores componentes (em valor relativo) são 0,68 e  $-0,72$  relativas aos papéis PETR3 e PETR4, par com maior coeficiente de correlação  $\rho_{ij} = 0,94$ . Já para o autovetor  $u_2$ , as maiores componentes  $-0,66$  e  $0,73$  correspondem, respectivamente, a ELET3 e ELET6, ações com o segundo maior coeficiente de correlação  $\rho_{ij} = 0,90$ . As duas maiores componentes do  $u_3$  são relativas aos papéis VALE3 e VALE5 que possuem  $\rho_{ij} = 0,87$  (terceiro maior coeficiente de correlação). Isso é válido para os autovetores correspondentes aos oito menores autovalores da matriz de correlação.

Plerou e colaboradores [21] chegaram a resultados semelhantes. As maiores componentes do autovetor correspondente ao menor autovalor  $u_1$  correspondem às ações estudadas Texas Instruments (TXN) e Micron Technology (MU) com  $\rho_{ij} = 0,64$ , maior coeficiente de correlação da amostra. As maiores componentes do autovetor  $u_2$  se referem às empresas Telefonos de Mexico (TMX) e Grupo Televisa (TV) com  $\rho_{ij} = 0,59$ , segundo maior coeficiente de correlação, enquanto as maiores componentes do autovetor  $u_3$  estão relacionadas às empresas com o terceiro maior  $\rho_{ij}$  da amostra.

### 3.1.2 Árvore de Extensão Mínima

Algumas análises do espaço ultramétrico em dados financeiros nos possibilitam obter informações interessantes quanto ao agrupamento de alguns setores econômicos das ações negociadas, e a comparação das árvores de extensão mínima para diferentes períodos nos possibilita entender as flutuações dos preços ao longo do tempo. Esse entendimento é importante, por exemplo, para a escolha de ações que irão compor carteiras de investimento, uma vez que ao diversificá-las, minimizam-se os riscos.

As árvores de extensão mínima foram obtidas aplicando-se o Algoritmo de Kruskal (Subseção 2.6.2) para a matriz de distâncias ultramétricas construída a partir dos coeficientes de correlação dos retornos de preços diários para as ações da carteira selecionada, com  $n(n - 1)/2$  elementos  $d_{ij}$  distintos. Os resultados estão ilustrados nas Figuras 3.10 a 3.17. As figuras foram construídas utilizando-se o programa Pajek, um programa para análise e visualização de grandes redes [31].

Na Figura 3.10 está ilustrada a árvore de extensão mínima obtida para o ano de 2001. Cada círculo (vértice da árvore) representa uma ação, que é designada pelo símbolo de negociação (ver Tabela 3.1). Os vértices são coloridos, indicando diferentes setores econômicos: azul claro representa o setor Petróleo, Gás e Biocombustíveis, preto ilustra Materiais Básicos, azul escuro setor Bens Industriais, rosa claro representa Consumo não-Cíclico, rosa escuro Consumo Cíclico, amarelo Telecomunicações, vermelho Utilidade Pública, verde indica o setor Financeiro e Outros. Vértices mais próximos entre si indicam menor valor para  $d_{ij}$  entre os papéis considerados. Observamos que 6 das 11 ações do setor Utilidade Pública formam um *cluster* em torno da ação BRAP4, que é do setor Financeiro e Outros. Já 3 das 4 ações do setor Financeiro e Outros estão conectadas entre si e com TNLP4, companhia do setor Telecomunicações. O setor econômico com maior número de ações na carteira escolhida, Materiais Básicos, não apresenta formação



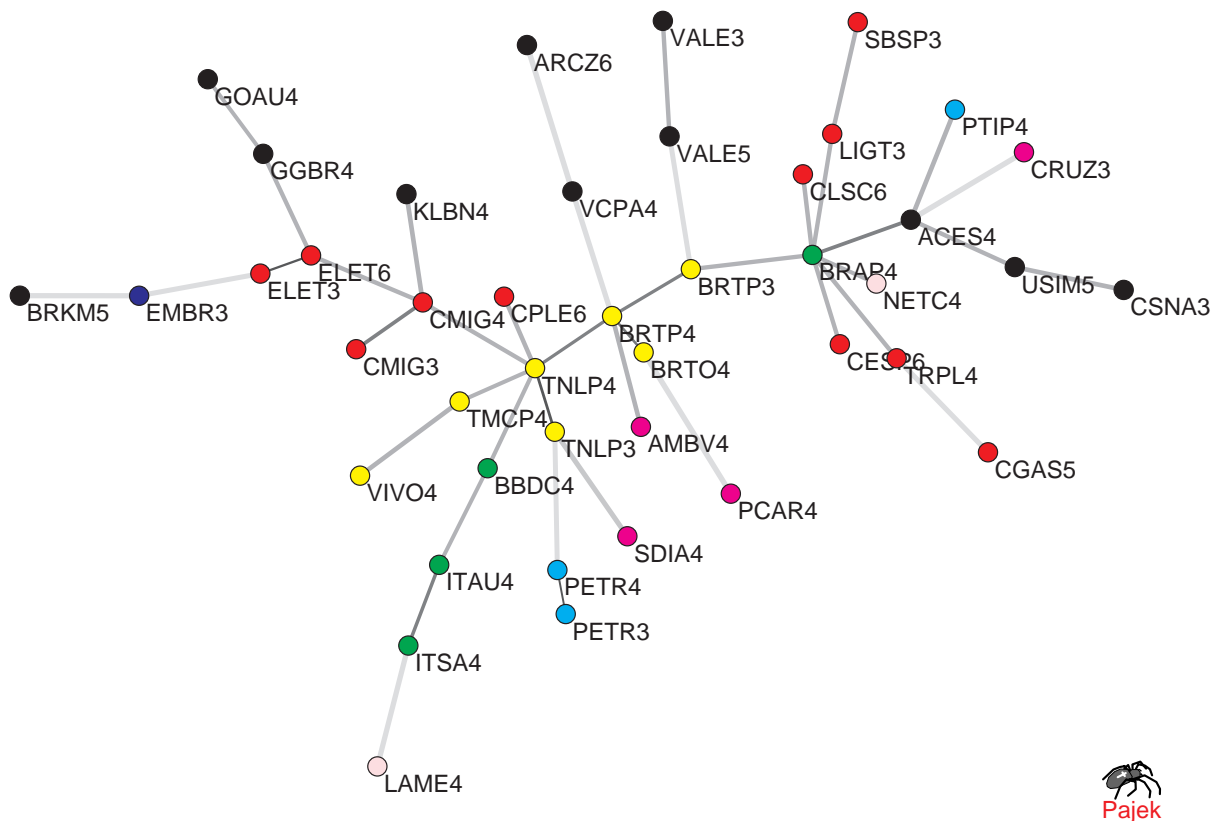


Figura 3.10: Árvore de extensão mínima para o ano de 2001. Círculos indicam ações designadas por seus símbolos de negociação. Cores representam diferentes setores econômicos: azul claro indica o setor Petróleo, Gás e Biocombustíveis, preto ilustra Materiais Básicos, azul escuro setor Bens Industriais, rosa claro Consumo não-Cíclico, rosa escuro Consumo Cíclico, amarelo Telecomunicações, vermelho Utilidade Pública, verde indica o setor Financeiro e Outros. Quanto mais próximos os vértices na árvore, menor é o valor da distância entre as respectivas ações.

de *cluster*. Já ações do setor de Telecomunicações estão agrupadas entre si. Os papéis que apresentam maior número de ligações são BRAP4 e TNLP4. BRAP4 se conecta a 7 papéis, sendo que com ACES4 ocorre a menor distância ( $d = 0,79$ ) e com CLSC6 a maior ligação ( $d = 0,93$ ). TNLP4 se liga a 6 papéis, com valores de distância variando entre  $d = 0,59$  (TNLP3) e  $d = 0,94$  (CPLE6). O menor valor de uma ligação na árvore ocorre entre os papéis PETR3 e PETR4 ( $d = 0,40$ ) e a maior entre VCPA4 e ARCZ6 ( $d = 1,14$ ).

A Figura 3.11 mostra a árvore de extensão mínima para o ano de 2002. A legenda desta figura, assim como das demais árvores ilustradas, é idêntica à legenda da

Figura 3.10. Nessa árvore, a ação TNLP4, do setor Telecomunicações, apresenta o maior número de ligações, 14. As distâncias variam entre  $d = 0,68$  (BRTO4) e  $d = 1,06$  (EMBR3). Notamos a ligação entre 5 das 11 ações pertencentes ao setor Materiais Básicos, VALE3, VALE5 (sub-setor Mineração), ARCZ6, VCPA4 (ambas do sub-setor Madeira e Papel) e CSNA3 (sub-setor Siderurgia e Metalurgia). Um resultado visto aqui, como também na figura anterior, é o agrupamento de 3 das 4 ações do setor Financeiro e Outros. Há a formação de *cluster* entre ações do setor Utilidade Pública, 7 dentre 11 estão agrupadas em torno de ELET6. Outro *cluster* ocorre entre as ações do setor Telecomunicações. A menor distância encontrada entre os vértices é  $d = 0,44$  entre ELET3 e ELET6 e entre TNLP3 e TNLP4, enquanto o maior valor obtido é  $d = 1,22$  entre os papéis LIGT3 e BRAP4.

A árvore de extensão mínima para o ano de 2003 revela resultados semelhantes aos encontrados anteriormente. As ações TNLP4 e CMIG4 apresentam maior número de ligações, 8 e 6, respectivamente. Dentre as ligações com TNLP4, o menor valor de uma ligação é  $d = 0,47$  (TNLP3) e o maior  $d = 1,07$  (USIM5). Há formação de *cluster* entre 3 das 4 ações do setor Financeiro e Outros, resultado visto para as árvores de extensão mínima dos anos de 2001 e 2002 e observamos ligações entre 4 ações do setor Materiais Básicos, VALE3, VALE5, ARCZ6 e VCPA4, que também estavam conectadas entre si na árvore do ano de 2002. Observamos um *cluster* entre ações do setor Utilidade Pública (9 dos 11 papéis estão agrupados) e outro entre as ações pertencentes ao setor Telecomunicações. Os valores para as distâncias entre os papéis estão no intervalo  $0,44(ELET3, ELET6) \leq d \leq 1,21(PCAR4, VALE3)$ .

Para o ano de 2004, novamente a ação TNLP4 apresenta um maior número de ligações, 12. Observamos novamente o *cluster* formado por companhias do setor Telecomunicações e outro formado por empresas do setor Utilidade Pública (8 dentre as 11 ações). Notamos a formação de um *cluster* grande de ações do setor Materiais Básicos (9 dentre os 11 papéis). Novamente as ações ITAU4, ITSA4, BBDC4 (setor Financeiro e Ou-

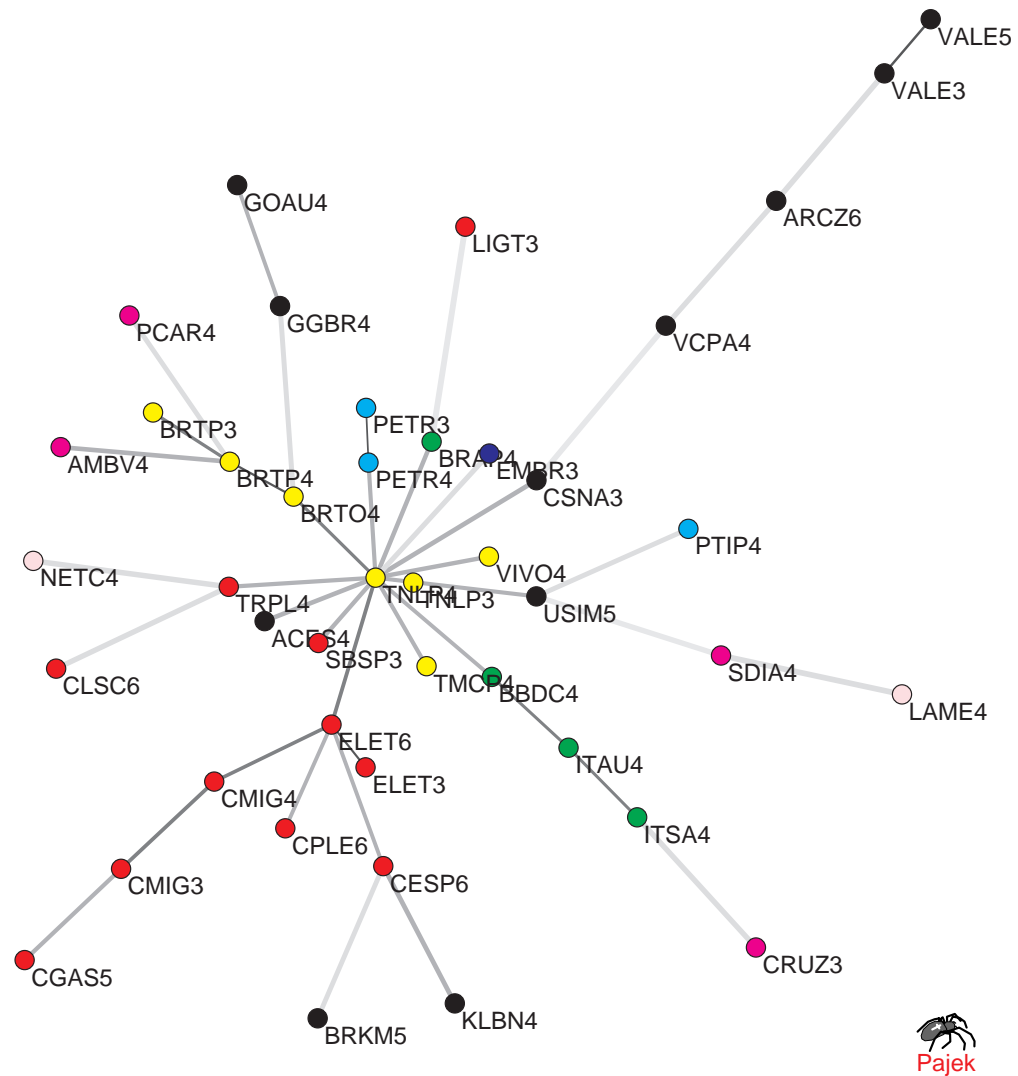


Figura 3.11: Árvore de extensão mínima para o ano de 2002. Mesma legenda da Figura 3.10.

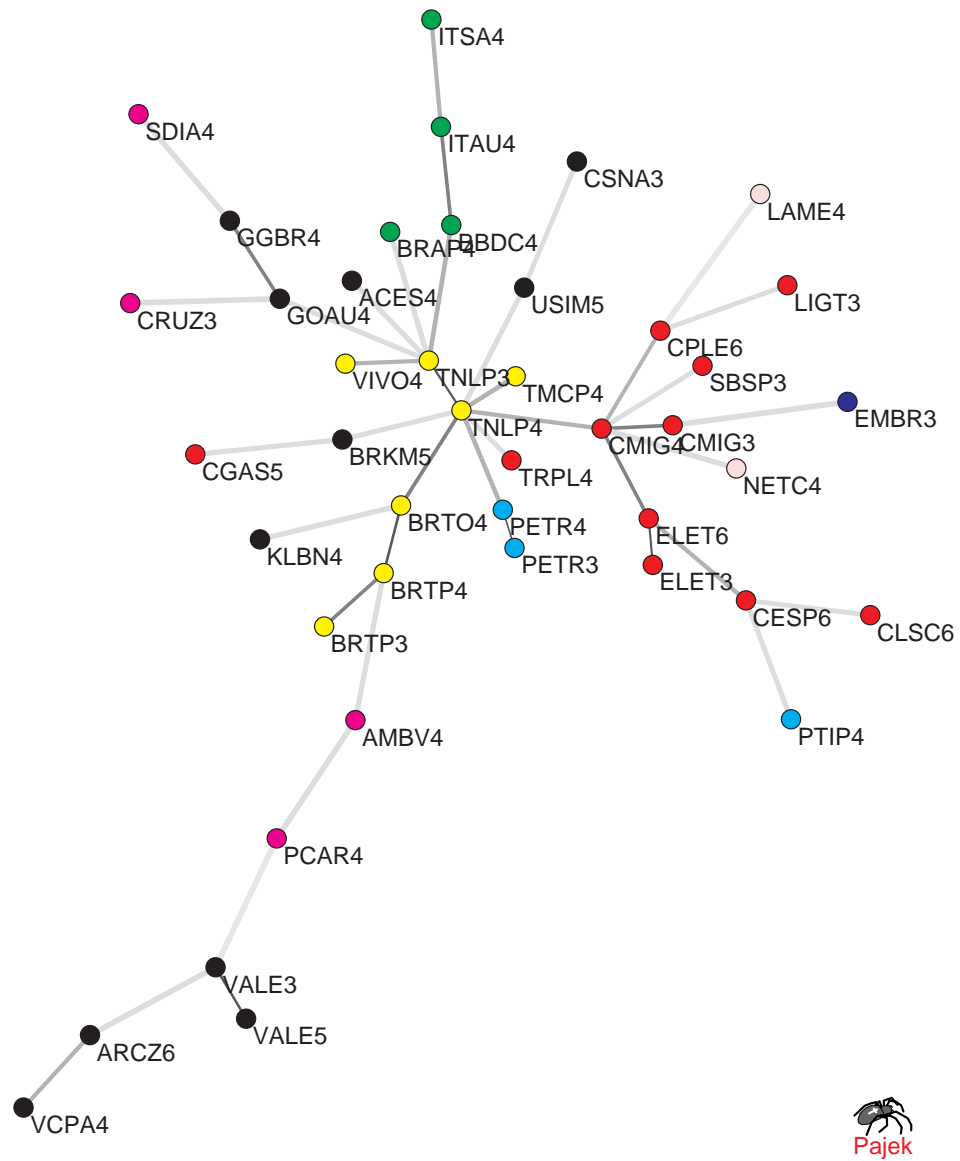


Figura 3.12: Árvore de extensão mínima para o ano de 2003. Legenda idêntica a da Figura 3.10.

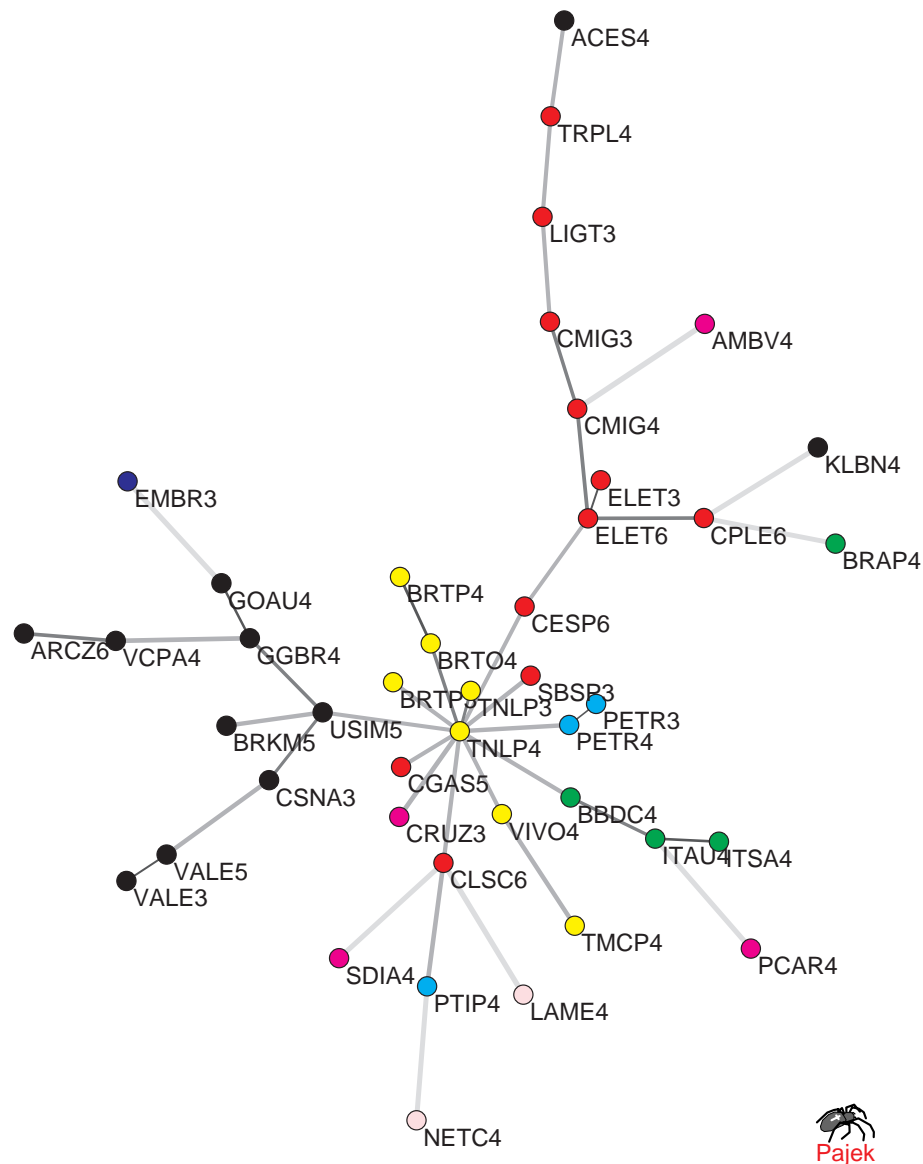


Figura 3.13: Árvore de extensão mínima calculada usando-se como dados de entrada a matriz de distância do ano de 2004. Ver legenda da Figura 3.10.

tros) estão conectadas. Para esta árvore, as distâncias variam entre  $0,33(PETR3, PETR4) \leq d \leq 1,14(AMBV4, CMIG4)$ .

Em 2005, bem como em 2006, as árvores de extensão mínima mostram a persistência na formação de *clusters* entre papéis de mesmo setor, sendo eles dos setores Telecomunicações, Financeiros e Outros (3 dentre 4 ações), Materiais Básicos (9 em 2005 e

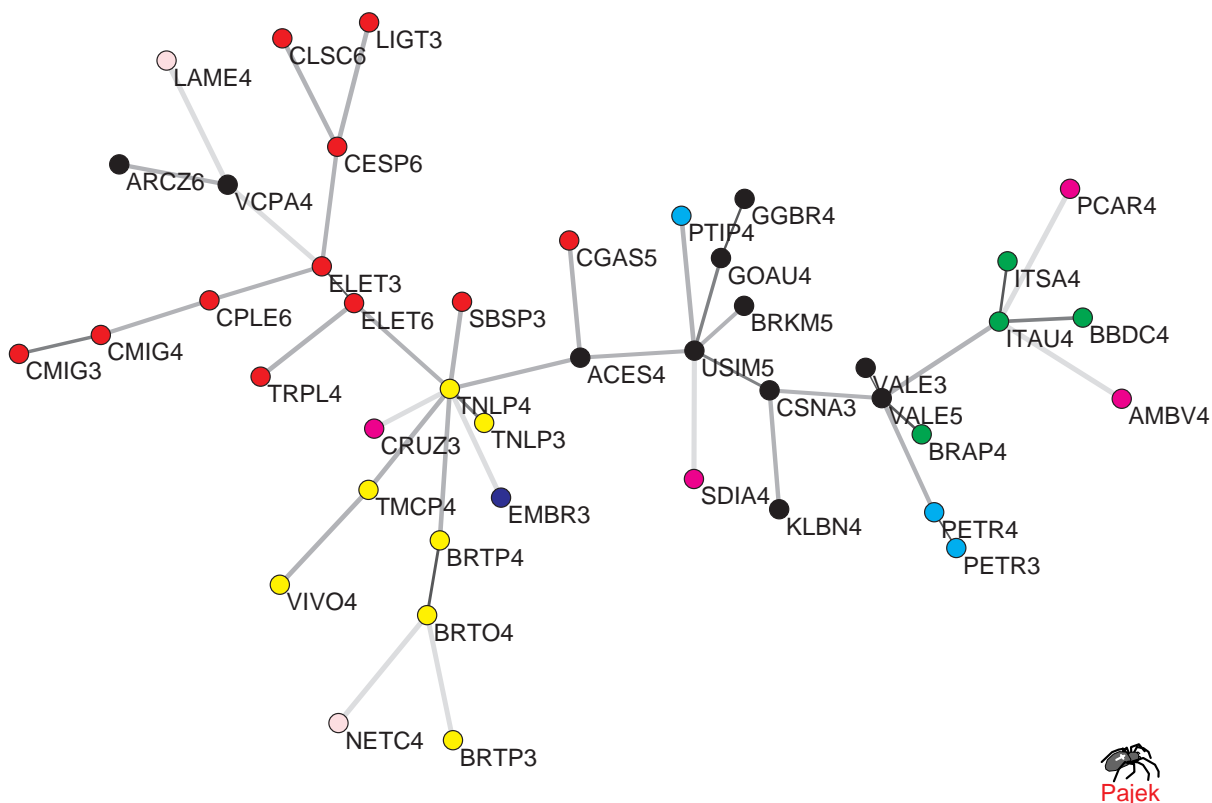


Figura 3.14: Árvore de extensão mínima para o ano de 2005. Mesma legenda da Figura 3.10.

8 em 2006 dentre as 11 ações estudadas desse setor) e Utilidade Pública (9 entre 11). Entretanto, as ações que não formam *cluster* com outras de mesmo setor citadas anteriormente, diferem de um ano para outro. Por exemplo, em 2005 ARCZ6 e VCPA4, ambas do setor Materiais Básicos e sub-setor Madeira e Papel estão ligadas entre si, mas não fazem parte do *cluster* formado por papéis do mesmo setor. Já em 2006, as companhias que não formam *cluster* com as ações do setor Materiais Básicos são ACES4 que está conectada a CMIG3 (Utilidade Pública), KLBNA4 que está ligada a TNLP4 (Telecomunicações) e BRKM5 conectada a BRTO4 (Telecomunicações). Em 2005, a distância mínima encontrada é  $d = 0,31$  (VALE3, VALE5) e a máxima é  $d = 1,08$  (PCAR4, ITAU4). Já em 2006, as distâncias encontradas na árvore variam entre  $0,32(PETR3, PETR4) \leq d \leq 1,14(ITAUA4, EMBR3)$ .

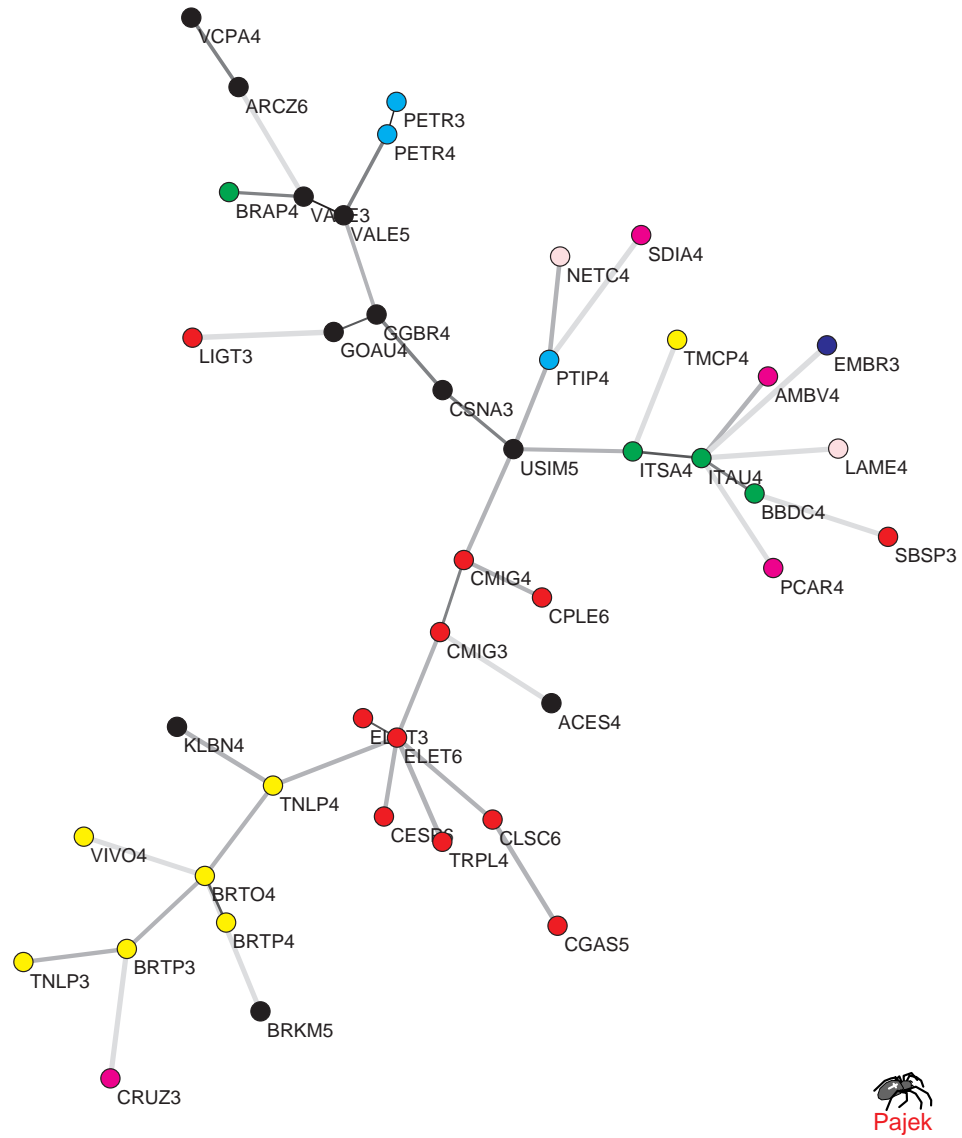


Figura 3.15: Árvore de extensão mínima para o ano de 2006. Legenda idêntica à da Figura 3.10.

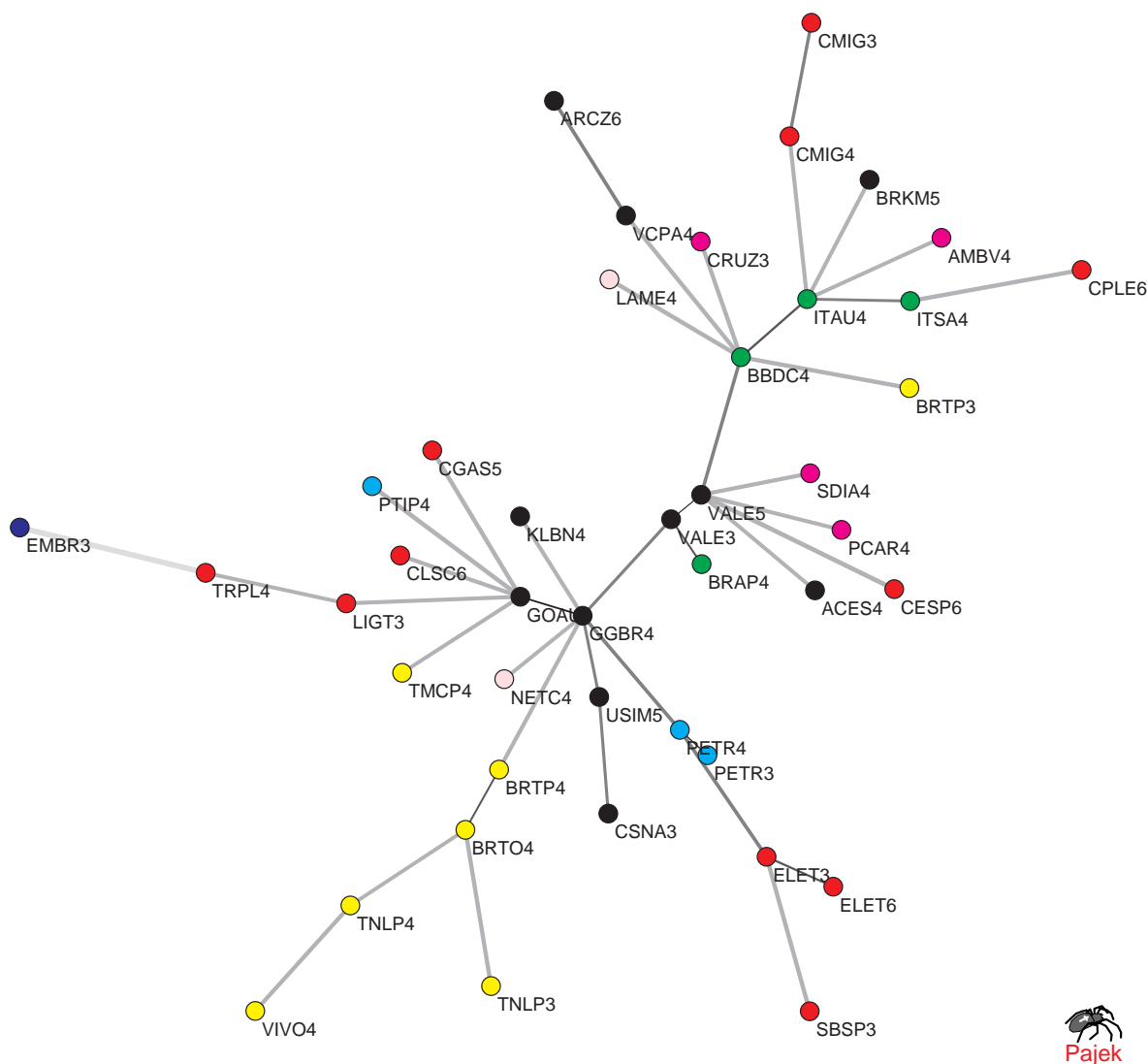


Figura 3.16: Árvore obtida para o período entre janeiro e junho de 2007. Ver legenda da Figura 3.10.

Em 2007, não há *cluster* de ações do setor Utilidade Pública, mas persiste o agrupamento de ações dos setores Materiais Básicos (8 dentre 11 ações), Financeiro e Outros (sub-setor Intermediários Financeiros) e Telecomunicações (5 dentre 7 papéis).

A Figura 3.17 ilustra a árvore de distância mínima para todo o período estudado, de janeiro de 2001 a junho de 2007. Os coeficientes de correlação foram calculados usando-se a média e a variância dos retornos de preços sobre todo o período, foi obtida a matriz de distâncias e aplicado o Algoritmo de Kruskal. Essa árvore mostra a ação



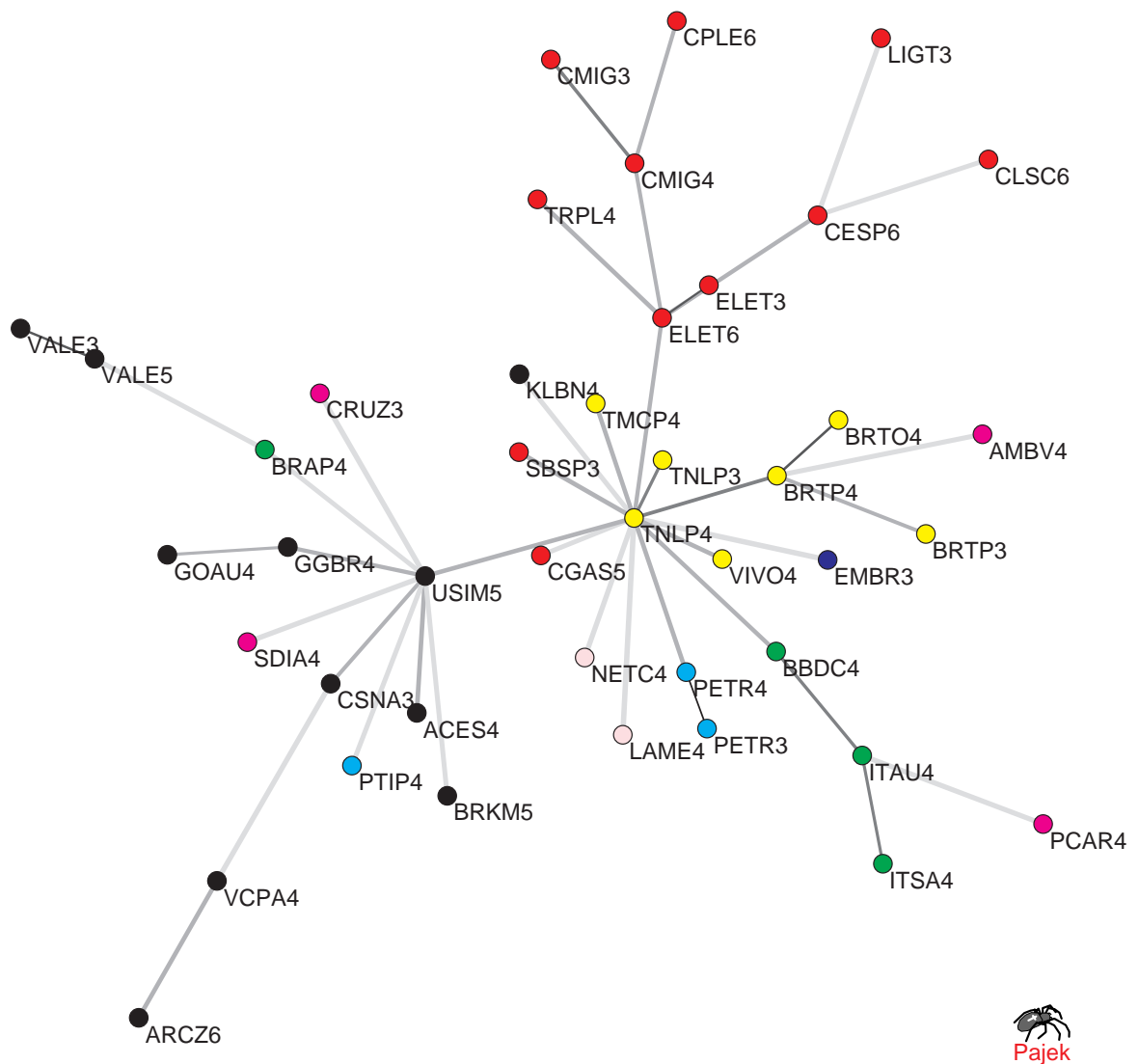


Figura 3.17: Árvore obtida para o todo o período estudado, de janeiro de 2001 a junho de 2007.

TNLP4 como aquela que possui o maior número de ligações, 14, resultado encontrado em árvores obtidas para certos anos estudados. Há formação dos seguintes *clusters*: Telecomunicações, Utilidade Pública (9 dentre 11 ações estudadas desse setor), Financeiro e Outros (papéis do sub-setor Intermediários Financeiros). O menor valor entre dois vértices é  $d = 0,36$  entre os papéis PETR3 e PETR4 enquanto o maior valor obtido é  $d = 1,17$  entre TNLP4 e LAME4.

É interessante observar que há certas ligações que estão presentes nas árvores de distância mínima em todos os anos considerados separadamente e em todo o período estudado. Destacamos as ligações entre ações do setor Financeiros e Outros, sub-setor Intermediários Financeiros, ITAU4, ITSA4 e BBDC4. Em nenhuma das árvores a ação de mesmo setor BRAP4, mas de sub-setor diferente, *Holdings* Diversificadas, está conectada às anteriores. Outras ligações sempre presentes ocorrem entre PETR3 e PETR4, VALE3 e VALE5, papéis da mesma companhia, mas de tipos diferentes e GGBR4 e GOAU4. A ação PTIP4 do setor Petróleo, Gás e Biocombustíveis não está ligada a PETR3, PETR4 em nenhuma das árvores obtidas. Destacamos ainda a formação de *cluster* de ações pertencentes ao setor Telecomunicações em todas as árvores de extensão mínima obtidas. Observamos que não há conexão nem formação de *cluster* entre ações do setor Consumo não-Cíclico nem do setor Consumo Cíclico nos resultados obtidos. Ações que estão intensamente conectadas a outras, como, por exemplo, TNLP4, que se destaca nas árvores de alguns anos estudados, indicam que suas flutuações servem frequentemente de referência para um dado grupo de ações.

Esses resultados indicam que é possível extrair informações comuns que afetam grupos de ações de uma carteira selecionada a partir de séries temporais dos preços das ações. Através do estudo das árvores de extensão mínima observamos que a escala de tempo das variações das informações contidas nessas árvores é maior que um ano, isto é, árvores de extensão mínima entre anos diferentes mantém algumas características, como por exemplo, presença de *clusters* formados por ações de um mesmo setor econômico.

Apesar de o número de empresas investigadas  $n = 43$  e do tamanho das séries temporais  $T = 1603$  serem pequenos, nossos resultados estão, qualitativamente, em acordo com os encontrados na literatura. Mantegna [24] investigou uma carteira composta de  $n = 30$  ações que compõem o índice Dow Jones (Dow Jones Industrial Average - DJIA) e uma carteira com  $n = 443$  ações que fazem parte do índice S&P 500 no período

de julho de 1989 a outubro de 1995. Ambos os índices descrevem principalmente o comportamento da bolsa de valores de Nova York (NYSE). Foram obtidas árvores de extensão mínima para os dois grupos estudados. Uma inspeção da árvore para ações do índice DJIA mostra a existência de 3 grupos de ações. O primeiro grupo é formado por 3 ações pertencentes ao mesmo setor econômico (Energia) e ao mesmo sub-setor (*International Oil*). O segundo grupo é formado por 2 empresas do mesmo setor (Insumos), mas de sub-setores diferentes. O terceiro grupo envolve companhias que estão em setores econômicos relacionados à consumo não durável, bebidas e tabaco.

A mesma investigação foi feita para o conjunto formado por empresas que compõem o índice S&P500. A estrutura da árvore de extensão mínima mostra muitos grupos de ações que são homogêneas do ponto de vista econômico. Com algumas poucas exceções, os grupos são homogêneos com respeito aos setores econômicos e, muitas vezes, em relação aos sub-setores, sugerindo que conjuntos de ações que executam atividades do mesmo setor e sub-setor econômico respondem, de uma maneira estatística, ao mesmo fator econômico. Esses resultados mostram que séries temporais do preço das ações carregam informação econômica importante que pode ser detectada.

Coronnello e colaboradores [16] investigaram séries temporais dos retornos de preços diários para um conjunto de  $n = 100$  ações com grande volume de negociação na NYSE ao longo do ano de 2002. Compararam resultados obtidos através da árvore de extensão mínima com aqueles obtidos analisando-se árvores hierárquicas e, sugerem que a árvore de extensão mínima contém informações adicionais àquelas presentes nas árvores hierárquicas. A topologia das árvores de extensão mínima mostra a existência de várias ações com um grande número de ligações com outros vértices.

Outro estudo interessante foi feito por Coronnello e colaboradores [32] considerando  $n = 92$  ações com grande volume de negociação na bolsa de valores de Londres (LSE) no ano de 2002. Investigaram as propriedades estatísticas do retorno de preços calculados em 2 intervalos de tempo,  $\Delta t = 1$  dia e  $\Delta t = 5$  min. Para dados considerando

retornos diários, eles observaram que duas ações se comportam como um ponto central de conexão na árvore de extensão mínima: Royal Bank of Scotland (RBS) e Shell (SHEL). Notaram que empresas pertencentes a certos setores econômicos formam *clusters* como, por exemplo, empresas do setor Financeiro se agrupam em torno de RBS, mas observaram também que empresas do setor de Serviços se encontram espalhadas pela árvore, não formando *cluster*.

Já para um horizonte de tempo de  $\Delta t = 5$  min, a estrutura da árvore de extensão mínima é um pouco diferente quando comparada com a obtida para um intervalo de  $\Delta t = 1$  dia. A organização de todas as ações em torno dos pontos centrais é mais pronunciada para  $\Delta t = 5$  min, ou seja, o número de ligações que partem de RBS e SHEL é maior quando se leva em consideração um menor intervalo  $\Delta t$ . Entretanto, menos ações do mesmo setor econômico estão agrupadas em torno dela. Coronello e colaboradores [32] concluíram que a eficiência na obtenção de informações sobre setores econômicos depende do horizonte de tempo considerado ao se calcular o retorno dos preços. Especificamente, o sistema é mais hierarquicamente estruturado quando se analisa dados utilizando-se retornos diários, confirmando que o mercado necessita de um certo intervalo de tempo para avaliar o correto grau de correlação cruzada entre pares de ações cujos preços são simultaneamente coletados.

---

## CAPÍTULO 4

### Conclusões e Perspectivas

---

No final do século passado, alguns físicos da área da Mecânica Estatística passaram a fazer pesquisas relacionadas às flutuações observadas no Mercado Financeiro. Essa área de pesquisa tem se desenvolvido muito nos últimos anos, devido ao volume enorme de dados disponíveis, tanto em longos períodos de tempo, na escala de dezenas de anos, como também em escalas curtas de tempo. Não há na natureza sistemas físicos com tantos dados disponíveis, em várias escalas de tempo, de modo que o Mercado Financeiro se apresenta como um verdadeiro laboratório para se aplicar as idéias desenvolvidas na área da mecânica estatística fora do equilíbrio, particularmente aquelas relativas aos processos estocásticos.

No Capítulo 2, estudamos alguns processos estocásticos que descrevem as flutuações de preços de ativos financeiros, bem como propriedades estatísticas aplicáveis a séries temporais. Fizemos uma revisão de métodos como Teoria da Matriz Aleatória e Espaço Ultramétrico utilizados para extrair informações de séries temporais.

No Capítulo 3, estudamos as correlações cruzadas entre flutuações de preços de diferentes ações. Calculamos os coeficientes de correlação a partir dos retornos diários de preços de 43 ações muito negociadas na Bovespa no período compreendido entre janeiro de 2001 a junho de 2007, totalizando 1603 dias de negociações. Aplicamos o método da Teoria da Matriz Aleatória para a matriz de correlação calculada usando-se retornos

diários de todo o período estudado. Encontramos o espectro de autovalores e comparamos com os valores teóricos previstos para séries temporais em que é válido o limite  $T \rightarrow \infty$ ,  $n \rightarrow \infty$ , sendo  $n$  o número de ações estudadas e  $T$  o comprimento das séries temporais. O valor encontrado para a razão  $Q = T/n$  difere muito dos valores encontrados na literatura, isto é devido, possivelmente, ao número reduzido de ações estudadas. Apesar disso, os valores encontrados para os autovalores da matriz de correlação, bem como o intervalo  $[\lambda_{min}, \lambda_{max}]$  estão de acordo com os resultados publicados na literatura.

Testamos a distribuição das componentes dos autovetores da matriz de correlação com a distribuição das componentes de uma matriz aleatória. Para uma parte dos autovalores dentro do intervalo teórico e considerados autovalores aleatórios, observamos que a distribuição está de acordo com a densidade de probabilidade obtida para componentes aleatórias. Isto é um indício de que os autovalores situados no intervalo  $[\tilde{\lambda}_{min}, \tilde{\lambda}_{max}]$  são aleatórios e não carregam informação real.

Usando-se a Teoria da Matriz Aleatória é possível identificar parte da matriz de correlação que contém informação econômica real. Usamos um método utilizado para mostrar que alguns autovetores cujos autovalores não são explicados pela Teoria estão associados a setores econômicos, sendo possível medir a importância de um dado setor na composição do autovetor. Observamos que os autovetores cujos autovalores  $\lambda_k \leq \tilde{\lambda}_{min}$  estão associados a um setor econômico. Já para autovalores  $\lambda_k \geq \tilde{\lambda}_{max}$ , não observamos destaque de nenhum setor específico, mas uma mistura deles, indicando que esses são autovetores correspondentes a autovalores fortemente afetados por incertezas estatísticas. Ao se analisar a importância dos setores econômicos na composição do autovetor  $u_{43}$ , correspondente ao maior autovalor, obtemos contribuições similares de todos os setores na caracterização desse autovetor, o que confirma resultados da literatura. Esses resultados interpretam o maior autovalor como uma grandeza que corresponde a uma resposta ou influência do mercado sobre todas as ações.

---

O estudo do Espaço Ultramétrico em dados financeiros nos possibilita obter informações interessantes quanto ao agrupamento das ações analisadas, e a comparação das árvores de extensão mínima para diferentes períodos nos possibilita entender as flutuações dos preços ao longo do tempo. Obtemos árvores para todo o período estudado e para cada ano separadamente. É interessante observar que há certas ligações que estão presentes nas árvores de distância mínima em todos os anos considerados separadamente e em todo o período estudado, e estão relacionadas a empresas de mesmo setor e sub-setor econômico. Ações que estão intensamente conectadas a outras, como, por exemplo, TNLP4, indicam que suas flutuações servem frequentemente de referência para um dado grupo de ações.

Esses resultados mostram que é possível extrair informações comuns que afetam grupos de ações de uma carteira selecionada a partir de séries temporais dos preços das ações. Observamos que a escala de tempo das variações das informações contidas nessas árvores é maior que um ano, isto é, árvores de extensão mínima entre anos diferentes mantém algumas características, como por exemplo, presença de *clusters* formados por ações de um mesmo setor econômico.

Para um próximo trabalho, é necessário utilizar mais séries de retornos, ou seja, um número maior de ações. É interessante usar séries temporais mais longas, a fim de se estudar melhor a estatística da matriz de correlação de uma carteira de ações selecionada. Isto permitiria, por exemplo, comparar propriedades universais das matrizes aleatórias com propriedades da matriz de correlação dos dados, obtendo, assim, resultados mais quantitativos. Apesar de termos estudado poucas ações e utilizado séries pequenas, nossos resultados estão, qualitativamente, de acordo com aqueles encontrados na literatura.

O estudo das matrizes de correlação é muito relevante, uma vez que um aspecto importante do gerenciamento de risco é a estimativa da correlação entre os movimentos de preços de diferentes ativos financeiros. A probabilidade de grandes perdas para

uma dada carteira ou *books* de opções é dominada pelos movimentos correlacionados de seus diferentes constituintes. Assim, o estudo das matrizes de correlação (ou covariância) possui uma longa história em Finanças e é um dos pilares da Teoria de Markowitz para otimização de portfólios: dado um conjunto de ativos financeiros caracterizados por seu retorno médio e risco, qual é o peso ótimo de cada ativo na composição da carteira, de maneira que ela forneça o melhor retorno para um nível fixo de risco, ou o menor risco para um certo retorno?

Mais precisamente, o retorno médio  $R_p$  de uma carteira  $P$  de  $n$  ativos é definida como  $R_p = \sum_{i=1}^n p_i R_i$ , sendo  $p_i$  o capital investido e  $R_i$  o retorno esperado do ativo  $i$ . Similarmente, o risco de uma carteira de investimentos pode ser associado à variância total  $\sigma^2 = \sum_{i,j=1}^n p_i C_{ij} p_j$ , sendo  $C$  a matriz de covariância [27].

Em particular, a composição da carteira menos arriscada possui uma grande contribuição dos autovetores de  $C$  correspondendo aos menores autovalores da matriz. Seria interessante aplicar esses resultados e utilizar a Teoria de Otimização de Markowitz a fim de montar carteiras, bem como administrar os riscos de investimento.



---

## APÊNDICE A

### Movimento Browniano

---

Discutimos aqui o movimento browniano descrevendo uma partícula de massa  $m$  imersa num líquido. Para mais detalhes ver referência [9].

Considere uma partícula sujeita a uma força viscosa proporcional a sua velocidade e a forças de caráter aleatório devidas ao impacto das moléculas do líquido com a partícula. Considerando o caso de um movimento unidimensional ao longo do eixo  $x$ , a equação de movimento é

$$m \frac{dv}{dt} = -\alpha v + F(t), \quad (\text{A.1})$$

sendo  $v = \frac{dx}{dt}$  a velocidade da partícula,  $\alpha$  uma constante e  $F(t)$  é a força aleatória que possui as seguintes propriedades:

$$\langle F(t) \rangle = 0, \quad (\text{A.2})$$

pois, em média, a força devida à colisão aleatória com as moléculas é nula e

$$\langle F(t)F(t') \rangle = B\delta(t - t'), \quad (\text{A.3})$$

uma vez que estamos assumindo impactos independentes. A eq. (A.1), suplementada pelas propriedades (A.2) e (A.3) é denominada equação de Langevin.

Podemos reescrever a equação de Langevin dividindo a eq. (A.1) por  $m$ .

Temos assim

$$\frac{dv}{dt} = -\gamma v + \zeta(t), \quad (\text{A.4})$$

onde  $\gamma = \alpha/m$  e  $\zeta(t) = F(t)/m$ . O ruído  $\zeta(t)$  é uma variável estocástica, isto é, uma variável aleatória dependente do tempo, que possui as propriedades:

$$\langle \zeta(t) \rangle = 0, \quad (\text{A.5})$$

$$\langle \zeta(t)\zeta(t') \rangle = \Gamma \delta(t - t'), \quad (\text{A.6})$$

sendo que  $\Gamma = B/m^2$ .

Para encontrar a solução geral da equação diferencial (A.4), vamos escrever que  $v(t) = u(t)e^{-\gamma t}$ , sendo  $u(t)$  uma função de  $t$  a ser determinada. Substituindo na eq.(A.4) temos,

$$\frac{du}{dt} = e^{\gamma t} \zeta(t), \quad (\text{A.7})$$

cuja solução é

$$u = u_0 + \int_0^t e^{\gamma t'} \zeta(t') dt'. \quad (\text{A.8})$$

Portanto,

$$v = v_0 e^{-\gamma t} + e^{-\gamma t} \int_0^t e^{\gamma t'} \zeta(t') dt', \quad (\text{A.9})$$

sendo  $v_0$  a velocidade da partícula no instante  $t = 0$ . Essa solução é válida para qualquer tipo de ruído  $\zeta(t)$ .

Vamos calcular o deslocamento médio da partícula. O deslocamento  $x(t)$  é dado por

$$x = x_0 + \int_0^t v(t') dt', \quad (\text{A.10})$$

sendo  $x_0$  a posição da partícula no instante  $t = 0$ . Tendo em vista a eq.(A.9), temos

$$x = x_0 + v_0 \int_0^t e^{-\gamma t'} dt' + \int_0^t e^{-\gamma t'} \int_0^{t'} \zeta(t'') e^{\gamma t''} dt'' dt'. \quad (\text{A.11})$$

Resolvendo a primeira integral e invertendo a ordem das integrais na última parcela, e integrando-se em  $t'$ , obtemos o resultado

$$x = x_0 + v_0 \frac{1}{\gamma} (1 - e^{-\gamma t}) + \frac{1}{\gamma} \int_0^t \zeta(t'') (1 - e^{\gamma(t''-t)}) dt'', \quad (\text{A.12})$$

que é válido para qualquer ruído  $\zeta(t)$ .

Levando-se em conta a propriedade (A.5), obtemos o deslocamento médio da partícula

$$\langle x \rangle = x_0 + v_0 \frac{1}{\gamma} (1 - e^{-\gamma t}). \quad (\text{A.13})$$

Usando as equações (A.12) e (A.13), temos que

$$x - \langle x \rangle = \frac{1}{\gamma} \int_0^t \zeta(t'') (1 - e^{\gamma(t''-t)}) dt''. \quad (\text{A.14})$$

Elevando-se essa equação ao quadrado e usando a propriedade (A.6), obtemos

$$(x - \langle x \rangle)^2 = \frac{\Gamma}{\gamma^2} \int_0^t (1 - e^{\gamma(t'-t)})^2 dt', \quad (\text{A.15})$$

e, resolvendo-se essa integral, temos, finalmente, o desvio quadrático médio

$$\langle x^2 \rangle - \langle x \rangle^2 = \frac{\Gamma}{\gamma^2} \left[ t - \frac{2}{\gamma}(1 - e^{-\gamma t}) + \frac{1}{2\gamma}(1 - e^{-2\gamma t}) \right]. \quad (\text{A.16})$$

Para tempos longos ( $t \gg 1/\gamma$ ), o termo dominante é o primeiro, de modo que nesse regime o desvio quadrático médio é proporcional a  $t$ , isto é,

$$\langle x^2 \rangle - \langle x \rangle^2 = Dt, \quad (\text{A.17})$$

sendo  $D = \Gamma/\gamma^2 = B/\alpha^2$  o coeficiente de difusão.

## A.1 Distribuição de probabilidade da velocidade

Para achar a distribuição de probabilidade relativa a  $v(t)$ , começamos por discretizar o tempo em intervalos iguais a  $\tau$ , escrevendo  $t = n\tau$ . Dessa maneira, a equação de Langevin (A.4) pode ser escrita na seguinte forma

$$v_{n+1} = av_n + \sqrt{\tau\Gamma}\xi_n, \quad (\text{A.18})$$

sendo que  $a = (1 - \tau\gamma)$  e as variáveis aleatórias  $\xi_j$  possuindo as propriedades

$$\begin{aligned} \langle \xi_j \rangle &= 0, \\ \langle \xi_j \xi_k \rangle &= \delta_{jk}. \end{aligned} \quad (\text{A.19})$$

A partir da eq. (A.18), podemos escrever que

$$v_n = \sum_{l=0}^{n-1} w_l, \quad (\text{A.20})$$

sendo  $w_l$  variáveis aleatórias independentes definidas por

$$w_l \equiv a^l \sqrt{\tau\Gamma} \xi_{n-1-l}, \quad (\text{A.21})$$

onde consideramos, por simplicidade, a condição inicial  $v_0 = 0$ .

A função característica relativa à variável  $v_n$  é

$$\tilde{P}_n(z) = \langle e^{izv_n} \rangle = \prod_{l=0}^{n-1} \langle e^{izw_l} \rangle. \quad (\text{A.22})$$

Supondo que a distribuição de probabilidades da variável  $\xi_l$  seja uma Gaussiana de média zero e variância 1,  $\mathcal{N}(0, 1)$ , segue-se que a distribuição da variável  $w_l$  também será uma Gaussiana de média zero, mas com variância  $a^{2l} \tau\Gamma$ . Portanto,

$$\langle e^{izw_l} \rangle = e^{-a^{2l} \tau\Gamma z^2 / 2}, \quad (\text{A.23})$$

de onde obtemos

$$\tilde{P}_n(z) = e^{-b_n z^2 / 2}, \quad (\text{A.24})$$

sendo que

$$b_n = \tau\Gamma \sum_{l=0}^{n-1} a^{2l} = \frac{1 - a^{2n}}{1 - a^2} \tau\Gamma. \quad (\text{A.25})$$

Portanto, a densidade de probabilidade da variável estocástica  $v(t)$  é obtida tomando-se a anti-transformada de Fourier da eq.(A.24). Desta forma, obtemos

$$P_n(v_n) = \frac{1}{\sqrt{2\pi b_n}} \exp\left(-\frac{v_n^2}{2b_n}\right). \quad (\text{A.26})$$

Tomando-se o limite em que  $\tau \rightarrow 0$  e  $n \rightarrow \infty$ , com  $n\tau = t$  fixo, a densidade de probabilidade da variável  $v$  no instante  $t$  é

$$P(v, t) = \frac{1}{\sqrt{2\pi b(t)}} \exp\left(-\frac{v^2}{2b(t)}\right), \quad (\text{A.27})$$

onde

$$b(t) = \frac{\Gamma}{2\gamma}(1 - e^{-\gamma t}). \quad (\text{A.28})$$

Logo, a velocidade  $v(t)$  de uma partícula livre num meio viscoso e sujeita a forças aleatórias varia de acordo com a equação de Langevin (A.4) e se distribui segundo uma densidade de probabilidade Gaussiana.

---

## APÊNDICE B

### Processo de Wiener

---

Uma teoria matemática rigorosa para o movimento Browniano foi proposta por Wiener em 1923. Aqui vamos fazer uma breve revisão dos conceitos e definições envolvidos.

**Definição 1** *O movimento Browniano padrão ou processo de Wiener  $(W(t), t \geq 0)$  é um processo estocástico com as seguintes propriedades [12]:*

1.  $W(0) = 0$ ;
2. *Os incrementos  $W(t) - W(s)$  são estacionários (a função densidade de probabilidade é invariante) e independentes;*
3. *Para todo  $t > 0$ ,  $W(t)$  possui uma distribuição normal com média zero e desvio padrão  $\sqrt{t}$ ,  $\mathcal{N}(0, \sqrt{t})$ ;*
4. *As trajetórias são contínuas.*

A condição de estacionariedade implica que a função densidade de probabilidade de  $W(t) - W(s)$ , para  $t \geq s$ , depende apenas da diferença  $t - s$ . Esse fato, combinado com a propriedade Gaussiana de  $W(t)$ , implica que os incrementos  $W(t) - W(s)$  são distribuídos de acordo com  $\mathcal{N}(0, \sqrt{t-s})$  para  $t > s$ . Isto segue da definição que  $W(t)$  possui  $\langle W(t) \rangle = 0$  para todo  $t \geq 0$ .

O movimento Browniano possui a importante propriedade de ser estatisticamente auto-afim:

$$W(at) \stackrel{d}{=} a^{1/2} W(t), \tag{B.1}$$

para  $a > 0$ . Aqui  $\stackrel{d}{=}$  significa igual no sentido de distribuição de probabilidade, ou seja, os dois processos estocásticos  $W(at)$  e  $a^{1/2}W(t)$  possuem a mesma distribuição finita  $P(x_1, t_1; \dots, x_n, t_n)$  para qualquer escolha de  $t_i, i = 1, \dots, n$  e  $n \geq 1$ . Auto-afinidade significa que qualquer porção da trajetória do movimento Browniano, quando propriamente reescalada é (estatisticamente) indistinguível de toda a trajetória. Por exemplo, se ampliarmos uma pequena região da trajetória de um movimento Browniano (não importa o quão pequena), reescalando o eixo dos tempos por um fator  $a$  e o eixo vertical por  $\sqrt{a}$ , obtemos uma curva similar (estatisticamente) à trajetória original.

Apesar da derivada de  $W(t)$  não existir como um processo estocástico regular, é possível dar um sentido para  $dW/dt$  como um processo generalizado (no sentido de funções generalizadas ou distribuições). Neste caso, a derivada de  $W(t)$  é chamada de ruído branco  $\xi(t)$ :

$$\xi(t) \equiv \frac{dW}{dt}. \quad (\text{B.2})$$

Na física, o ruído branco  $\xi(t)$  é simplesmente 'definido' como uma 'função que flutua rapidamente' (na verdade, um processo estocástico generalizado) que satisfaz as seguintes condições :

$$\langle \xi(t) \rangle = 0, \quad (\text{B.3})$$

e

$$\langle \xi(t)\xi(t') \rangle = \delta(t - t'). \quad (\text{B.4})$$

## B.1 Fórmula de Itô

Vamos retornar à equação de Langevin (A.4), reescrevendo-se  $\zeta(t) = \sigma \xi(t)$ .

Temos então:

$$\frac{dv}{dt} = -\gamma v + \sigma \xi(t), \quad (\text{B.5})$$



onde  $\xi(t)$  é um ruído branco. Introduzindo-se o processo de Wiener  $\xi(t) = dW/dt$ , temos

$$dv = -\gamma v dt + \sigma dW. \quad (\text{B.6})$$

Essa equação é um caso particular de uma equação diferencial estocástica mais geral da forma,

$$dX = a(X, t)dt + b(X, t)dW, \quad (\text{B.7})$$

sendo  $a(x, t), b(x, t)$  funções conhecidas e  $X$  uma variável aleatória com uma dada densidade de probabilidade  $P(x)$ .

Suponhamos que temos um novo processo estocástico, processo  $G$ , definido por

$$G(t) = F(X(t), t), \quad (\text{B.8})$$

para uma dada função  $F(x, t)$ , sendo que  $X(t)$  satisfaz a eq. (B.7). Queremos encontrar a dinâmica local seguida por  $G(t)$ , ou seja, a equação diferencial estocástica cujas soluções correspondem ao processo  $G(t)$ . A resposta é dada pela fórmula de Itô e, para obtê-la, vamos considerar a expansão de Taylor da função  $F(X, t)$ :

$$\begin{aligned} dG &= \frac{\partial F}{\partial t} dt + \frac{\partial F}{\partial x} dX + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} (dX)^2 + \\ &+ \frac{1}{2} \frac{\partial^2 F}{\partial t^2} (dt)^2 + \frac{1}{2} \frac{\partial^2 F}{\partial x \partial t} dt dX + \dots \end{aligned} \quad (\text{B.9})$$

Mas

$$(dX)^2 = b^2 dt + o(dt^{3/2}), \quad (\text{B.10})$$

pois  $dW^2 = dt$  e  $dt dW = O(dt^{3/2})$ . Reescrevendo  $dG$  até termos de ordem  $dt$ , obtemos

$$\begin{aligned} dG = & \left[ \frac{\partial F}{\partial t} + a(X, t) \frac{\partial F}{\partial x} + \frac{1}{2} b^2(X, t) \frac{\partial^2 F}{\partial x^2} \right] dt \\ & + b(X, t) \frac{\partial F}{\partial x} dW, \end{aligned} \tag{B.11}$$

que é a Fórmula de Itô.

---

## Referências

---

- [1] J. Voit, *The Statistical Mechanics of Financial Markets* (Springer, 2001).
- [2] L. Bachelier, "Théorie de la Spéculation", Ph. D. Thesis, Annales Scientifiques de l'Ecole Normale Supérieure III-17, 21-86 (1900).
- [3] Dicionário de Finanças na página [www.bovespa.com.br](http://www.bovespa.com.br).
- [4] R. Mantegna, H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge Univ. Press, Cambridge, 1999).
- [5] J. D. Farmer, M. Shubik e E. Smith, *Physics Today*, September 2005.
- [6] J. P. Bouchaud e M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge Univ. Press, Cambridge, 2003).
- [7] Comissão Nacional de Bolsas, *Mercado de Capitais* (Elsevier, 2005).
- [8] G. Oliveira e M. Pacheco, *Mercado Financeiro* (Editora Fundamento Educacional, 2005).
- [9] T. Tomé e M. J. de Oliveira, *Dinâmica Estocástica e Irreversibilidade* (Editora da Universidade de São Paulo, 2001).
- [10] M. E. J. Newman, arXiv:cond-mat/0412004, vol 2, no. **9** (2005).
- [11] P. Lévy, *Calcul des Probabilités* (Gauthier-Villars, Paris, 1925).
- [12] G. L. Vasconcelos, *Braz. J. Phys.* vol. **34** no.3B, p.1039-1065 (2004).
- [13] B. Mandelbrot, *J. Business*, **36** (1963) 294.
- [14] R. N. Mantegna and H. E. Stanley, *Nature* **376**, 46 (1995).
- [15] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, H. E. Stanley, *Phys. Rev. E* **60**, 5305 (1999).
- [16] C. Coronello, M. Tumminello, F. Lillo, S. Miccichè, R. N. Mantegna, arXiv:physics/0609036v1 (2006).

- [17] M. Tumminello, C. Coronello, F. Lillo, S. Miccichè, R. N. Mantegna, *IJBC*, **17**, 2319-2329 (2007).
- [18] M.L. Mehta, *Random Matrices* (Academic Press, New York, 3rd ed., 2004).
- [19] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral e H. E. Stanley, *Phys. Rev. Lett.* **83**, 1471 (1999).
- [20] S. N. Majumdar, arXiv:cond-mat/0701193v1 (2007).
- [21] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, e H. E. Stanley, *Phys. Rev. E* **65**, 066126 (2002).
- [22] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, M. Meyer e H. E. Stanley, *Phys. Rev. Lett.* **60**, 6519 (1999).
- [23] N. Christofides, *Graph Theory - An Algorithmic Approach* (Academic Press, London, 1975).
- [24] R. N. Mantegna, *Eur. Phys. J. B* **11**, 193 (1999).
- [25] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, e A. Kanto, *Phys. Rev. E* **68**, 056110 (2003).
- [26] J.-P. Onnela, A. Chakraborti, K. Kaski, e J. Kertész, *Physica A* **324**, 247-252 (2003).
- [27] L. Laloux, P. Cizeau, J.-P. Bouchaud, e M. Potters, *Phys. Rev. Lett.* **83**, 1467-1470 (1999).
- [28] T. Guhr, A. Müller-Groeling e H. A. Weidenmüller, *Phys. Rep.* **299**, 190 (1998).
- [29] P. Gopikrishnan, B. Rosenow, V. Plerou, e H.E. Stanley, *Phys. Rev. E* **64**, 035106 (2001).
- [30] J. Campbell, A. W. Lo, e C. MacKinlay, *The Econometrics of Financial Markets* (Princeton University Press, Princeton, 1997).
- [31] V. Batagelj e A. Mrvar, *Pajek: Program for Large Network Analysis* (2005).  
*Homepage:* <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
- [32] C. Coronello, M. Tumminello, F. Lillo, S. Miccichè, R.N. Mantegna, *Acta Phys. Pol. B* **36**, 2653, (2005).