

Dissertação de Mestrado

SAZONALIDADES INTRA-DIÁRIAS E  
NÃO-ESTACIONARIEDADE DE VOLUME NEGOCIADO  
EM MERCADOS FINANCEIROS: PROPRIEDADES  
ESTATÍSTICAS E MATRICIAIS.

Michelle Baú Graczyk

Centro Brasileiro de Pesquisas Físicas

Rio de Janeiro, 2016

---

Dissertação de Mestrado ao Centro Brasileiro  
de Pesquisas Físicas sob orientação do  
Pesquisador Doutor *Sílvio Manuel Duarte Queirós*  
para a obtenção do título de  
Mestre em Física por **Michelle Baú Graczyk**

2016

*Para a minha mãe, Daniele Baú, com muito amor.*

# Agradecimentos

Obrigada meu Deus, por tomar minha vida em suas mãos, me guiar e me dar forças para seguir o meu caminho.

Ao meu Orientador, Pesq. Doutor. Silvio M. Duarte Queirós, pela dedicação, pelas conversas, pelos ensinamentos, pela paciência, por acreditar em mim e por tornar o meu mestrado mais feliz.

À CAPES e a FAPERJ (Bolsa Nota 10) pelo apoio financeiro.

Ao meu tripé, Mãe, Babi e Leila, muito obrigada por serem meu alicerce desde sempre. Obrigada por alegrarem a minha vida e me apoiarem sempre.

Mãe, obrigada por abraçar os meus sonhos e os sonhos da Babi, e toma-los como seus também. Obrigada por nos fazer acreditar que éramos capazes e nos incentivar a sempre darmos mais um passo à frente. Muito obrigada por acreditar em mim, mesmo quando em muitas vezes eu mesma não acreditei.

Victor, muito obrigada por me fazer sorrir nos momentos mais duros deste mestrado. Obrigada por ser tão companheiro e amigo.

Pri e Jairo, muito obrigada pelas conversas que tanto me tranquilizavam, pelas dúvidas tiradas em R e pela amizade que tanto me faz bem.

Muito obrigada aos meus familiares e amigos que fazem a minha vida ser mais colorida.

## Resumo

Nesta tese se apresenta o estudo das propriedades intra-diárias e de não-estacionariedade do volume negociado em mercados financeiros, sistema que é considerado um caso paradigmático de complexidade. Os dados correspondem aos títulos que compuseram o índice Dow Jones Industrial Average no segundo semestre de 2004 sendo que o período total de análise vai do primeiro semestre de 2003 ao primeiro semestre de 2014. O trabalho se encontra dividido em duas partes principais: análise individual das empresas e análise do comportamento conjunto através da matriz de correlação.

Na primeira parte se comprova a existência de um perfil convexo do volume médio bem como para a variância. Para o caso da assimetria, o perfil encontrado é côncavo. Já no caso da curtose a forma apresentada assemelha-se a uma função com um perfil semelhante a  $\sim$ . Estes resultados permitiram verificar que existiu uma alteração na forma como o volume médio reage à abertura de mercado com a crise de 2008 e que o perfil intra-diário do volume médio tem vindo a perder a sua concavidade. Comparando as diferentes ordens estatísticas verificamos uma diferença substancial entre os períodos de negociação matutina e vespertina.

Na segunda parte, por diagonalização da matriz de correlação verificou-se que existe um modo de mercado associado ao maior autovalor e que domina a correlação entre o volume negociado. Esse valor próprio — é menor do que o modo de mercado das flutuações de preço — é crescente no tempo e essa evolução pode ser dividida em duas regiões antes de almoço e

após almoço apontando para a existência de dois regimes de negociação dentro da sessão. O autovetor que corresponde ao maior autovalor é próximo de um vetor uniforme com a coerência entre os dois vetores a aumentar ao longo do tempo. Os períodos em que o valor do primeiro autovalor apresenta máximos locais corresponde aos instantes em que a projeção do respetivo autovetor sobre o vetor uniforme apresenta valores menores. Também na análise da matriz de correlação foi possível verificar o impacto da crise financeira de 2008 já que a projeção dos primeiros autovetores ao longo de um dia do primeiro semestre dos dados com os semestres seguintes apresenta uma clara mudança de comportamento a partir dessa data.

## Abstract

This thesis reports the study of intraday and non-stationarity properties of trading volume in financial markets, a system that is considered a paradigmatic case of complexity. The data we use correspond to the securities that made up the Dow Jones Industrial Average in the second half of 2004 that we span from the first half of 2003 to the first half of 2014. The work is divided into two main parts: the analysis of the statistical properties of the trading volume the companies treating them independently and the joint-behaviour of the trading volume of all the companies using a correlation matrix approach.

In the first part, we assert existence of a convex profile for both the average volume and the variance. For the skewness, the profile is concave and for the kurtosis the shape resembles an inverted tilde. From this analysis we a verify there was a change in the way the average volume reacts to the opening of the trading session in the wake of the 2008 financial crisis. Moreover, we show that the famous  $\cup$ -shape has been losing its concavity. Last, by comparing different order statistics we observe a substantial difference between the morning and afternoon periods of the trading.

In the second part of the thesis, by diagonalising the correlation matrix of the trading volume it was found that there is a market mode associated with the largest eigenvalue of that correlation matrix. That eigenvalue itself — which is less the market mode than the market price of such fluctuations — is increasing in time and its evolution can be divided into two regions

before lunch, after lunch and pointing to the existence of two trading regimes during the session. The eigenvector corresponding to the largest eigenvalue is close to a uniform vector with the coherence between the two vectors to increasing over time. The periods in which the first eigenvalue value has local maxima correspond to the trading times at which the projection of the respective eigenvector onto the uniform vector has smaller values. Also within the context of the correlation matrix analysis, one emphasises the impact of the financial crisis of 2008 by computing the projection of a vector composed of the first eigenvectors over a day of the first semester onto a similar vectors for the following semesters. This projection clearly changes its behaviour from the 2008 crisis onwards.

# Conteúdo

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>   | <b>1</b>  |
| 1.1      | Objetivo desta tese e estrutura . . . . .                               | 4         |
| <b>2</b> | <b>Dados e Métodos</b>  | <b>10</b> |
| 2.1      | Os dados analisados . . . . .   | 10        |
| 2.2      | Momentos Estatísticos . . . . .   | 13        |
| 2.3      | Teoria da Matriz de Correlação . . . . .                                | 17        |
| <b>3</b> | <b>Sazonalidades individuais</b>  | <b>35</b> |
| 3.1      | Relação da Média do Volume com o Tempo . . . . .                        | 35        |
| 3.1.1    | Evolução da Concavidade e da Paridade do Volume . . . . .               | 43        |
| 3.1.2    | Relação entre a Paridade/Concavidade e a Média do Volume . . . . .      | 46        |
| 3.1.3    | Relação entre a Concavidade/Paridade e as Flutuações de Preço . . . . . | 48        |
| 3.1.4    | Relação entre Concavidade/Paridade e a Volatilidade . . . . .           | 51        |
| 3.2      | Variância do Volume . . . . .   | 54        |
| 3.3      | Assimetria . . . . .  | 59        |

|          |   |            |
|----------|---|------------|
| 3.4      | Curtose . . . . .   | 66         |
| <b>4</b> | <b>Sazonalidades coletivas</b>  | <b>76</b>  |
| 4.1      | Análise da Componente Principal . . . . .                               | 77         |
| 4.1.1    | Matrizes Completas e Matrizes Condicionadas . . . . .                   | 79         |
| 4.1.2    | Matrizes Completas e Matrizes Condicionadas em Valor Absoluto . . . . . | 87         |
| 4.2      | Análise à luz da Teoria da Matriz Randômica . . . . .                   | 95         |
| <b>5</b> | <b>Notas finais e Perspetivas Futuras</b>                               | <b>117</b> |
| 5.1      | Questões futuras . . . . .  | 126        |
| <b>A</b> | <b>Descrição da fórmula da volatilidade</b>                             | <b>129</b> |
| <b>B</b> | <b>Cenário Econômico entre 2004 até 2013</b>                            | <b>132</b> |
| <b>C</b> | <b>Gráficos típicos para as empresas individuais</b>                    | <b>141</b> |
| C.1      | Assimetria . . . . .  | 142        |
| C.2      | Curtose . . . . .   | 143        |

# Lista de Figuras

|     |   |    |
|-----|---|----|
| 3.1 | Volume médio transacionado por horário intra-diário para algumas empresas. . . . .  | 36 |
| 3.2 | Volume transacionado por horário intra-diário de todas as empresas e somente da empresa Intel Corp.. . . . .  | 38 |
| 3.3 | Média do volume transacionado intra-diário sobre todas as empresas para cada semestre. A curva pontilhada vermelha corresponde ao ajuste $\mathcal{M}_i = h_i^{-a} + constante$ . As linhas horizontais laranja correspondem aos valores médios dos dois períodos $a = 0.29$ e $0.37$ , respetivamente. . . . . | 39 |
| 3.4 | Paridade do volume transacionado para cada semestre para as empresas indicadas nas legendas. A linha vermelha indica a média sobre todas as empresas e as linhas tracejadas azuis indicam o valor médio da média sobre todas as empresas e deste somado e subtraído o desvio padrão. . . . .                    | 43 |
| 3.5 | Concavidade do volume transacionado para cada semestre para as empresas indicadas nas legendas. A linha vermelha indica a média sobre todas as empresas e as linhas tracejadas azuis indicam o valor médio da média sobre todas as empresas e deste somado e subtraído o desvio padrão. . . . .                 | 45 |
| 3.6 | Paridade do volume transacionado para cada semestre em função do volume médio semestral para as empresas indicadas nas legendas. . . . .  | 47 |
| 3.7 | Concavidade do volume transacionado para cada semestre em função da média do volume semestral. . . . .  | 48 |

|      |   |    |
|------|---|----|
| 3.8  | Paridade do volume transacionado para cada semestre em função da flutuação do preço do ativo. A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na variação de preço. . . . .   | 49 |
| 3.9  | Concavidade do volume transacionado para cada semestre em função da flutuação do preço do ativo. A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na variação de preço. . . . .  | 51 |
| 3.10 | Paridade do volume transacionado para cada semestre em função do coeficiente de volatilidade para cada semestre calculado através da Eq. (2.2). A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na volatilidade. . . . .  | 53 |
| 3.11 | Concavidade do volume transacionado para cada semestre em função do coeficiente de volatilidade para cada semestre calculado através da Eq. (2.2). A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na volatilidade. . . . .   | 54 |
| 3.12 | Média da Variância em função da Média do volume transacionado por horário intra-diário sobre todas as empresas. A reta tracejada em vermelho corresponde ao ajuste total dos pontos dado por $\mathcal{V}_i = t_i * \tilde{v}_i^2$ , a reta laranja tracejada representa o ajuste dos pontos pertencente ao período matutino dado por $\mathcal{V}_i = m_i * \tilde{v}_i^2$ , enquanto a reta tracejada rosa representa o ajuste dos pontos pertencentes ao período vespertino dado por $\mathcal{V}_i = v_i * \tilde{v}_i^2$ . . . . . | 57 |
| 3.13 | Variância do volume transacionado por horário intra-diário de todas as empresas e somente da empresa Intel Corp.. . . . .   | 59 |

- 3.14 Média da Assimetria em função da Média do volume transacionado por horário intra-diário sobre todas as empresas. A reta tracejada em vermelho corresponde ao ajuste total dos pontos dado por  $\mathcal{A}_i = t_i \bar{v}_i + i_i$ , a reta laranja tracejada representa o ajuste dos pontos pertencente ao período matutino dado por  $\mathcal{A}_i = m_i * \bar{v}_{const} + const'_i$  enquanto a reta tracejada rosa representa o ajuste dos pontos pertencentes ao período vespertino dado por  $\mathcal{A}_i = v_i * \bar{v}_i + const''_i$ . . . . . 60
- 3.15 Assimetria do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário. 63
- 3.16 Assimetria do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Assimetria sobre todas as empresas para cada horário. As linhas azuis corresponde ao ajuste  $\mathcal{A}_{ci} = c_i * h_i + cte_i$  e  $\mathcal{A}_{di} = d_i * h_i + cte'_i$ , sendo  $c_i$  e  $cte_i$  os coeficientes de ajuste da reta crescente e  $d_i$  e  $cte'_i$  os coeficiente de ajuste da reta decrescente para um dado semestre  $i$ . A linha vermelha tracejada define a região de normalidade. A curva azul clara representa a curva do coeficiente de ajuste  $c$  e a curva azul escura representa a curva do coeficiente de ajuste  $d$ . . . . . 65
- 3.17 Média da curtose em função da média do volume transacionado. A reta tracejada em vermelho é o ajuste linear de todos os pontos  $\mathcal{C}_i = t_i * \bar{v}_i + const_i$ , a reta laranja tracejada(rosa) o ajuste dos pontos do período matutino(vespertino):  $\mathcal{C}_{i_{m(ve)}} = a2_{i_{m(ve)}}^* \bar{v}_i^2 + a1_{i_{m(ve)}}^* \bar{v}_i + a0_{i_{m(ve)}}$ . . . . . 68
- 3.18 Média da Curtose do volume transacionado sobre todas as empresas em função da Média do Volume transacionado (calculadas para o mesmo horário). . . . . 70
- 3.19 Curtose do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário. . 71
- 3.20 Gráfico da Curtose do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário. A linha azul corresponde ao ajuste  $\mathcal{C}_i = h_i^{-\alpha_i} + constante$  e a linha azul escura corresponde ao ajuste  $\mathcal{C}_i = C_{max_i} - (h_i - h_{max_i})^{\alpha_{f_i}}$  para cada semestre  $i$ . . . . 74

|     |   |    |
|-----|---|----|
| 4.1 | Mapa de Calor das Matrizes de Correlação para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:29 horas e 16:01 horas. . . . .  | 81 |
| 4.2 | Mapa de Calor das Matrizes de Correlação 40% para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:00 horas e 16:01 horas. . . . .  | 82 |
| 4.3 | Mapa de Calor das Matrizes de Correlação 20% para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:00 horas e 16:01 horas. . . . .  | 83 |
| 4.4 | Curva da Média do Valor de Correlação para as Matriz de Correlação, Matriz Correlação 20% e Matriz Correlação 40%. . . . .  | 86 |
| 4.5 | Curva da Média do Valor Absoluto de Correlação para as Matriz de Correlação, Matriz Correlação 20% e Matriz Correlação 40%. . . . .   | 87 |
| 4.6 | Mapa de Calor das Matrizes de Correlação e Matrizes de Correlação 20% em Valor Absoluto para os horários 09:31 horas, 13:01 horas e 16:01 horas. . . . .  | 90 |
| 4.7 | Mapa de Calor das Matrizes de Correlação e Matrizes de Correlação 20% em ordem decrescente para os horários indicados. . . . .  | 92 |
| 4.8 | Mapa de Calor das Matrizes de Correlação, e Matrizes de Correlação 20% em ordem decrescente do valor absoluto das correlações para os horários indicados. . . . .   | 95 |
| 4.9 | Porcentagem dos autovalores máximos(mínimos) em relação ao autovalor máximo(mínimo) das Matrizes de Correlação (turquesa), Matrizes de Correlação 40%(rosa) e Matrizes de Correlação 20% (laranja). Na direita baixa: Pcs do Autovalor mais significativo para cada horário. A linha tracejada vermelha corresponde ao ajuste linear de todo o período intra-diário enquanto a linha tracejada laranja corresponde ao ajuste linear do período matutino e a linha tracejada rosa corresponde ao ajuste linear do período vespertino. Os erros dos declives são da mesma ordem do seu valor apontando para existência de dois patamares. . . . . | 97 |

|      |  |     |
|------|--|-----|
| 4.10 | Primeiro autovalor mais significativo, representado pela curva vermelha, do segundo autovalor mais significativo, representado pela curva azul, e do terceiro autovalor mais significativo, representado pela curva verde, em função dos horários intra-diários. . . . .   | 99  |
| 4.11 | Produto Interno dos três primeiros autovetores mais significativos e Desvio Padrão em unidades de Média. Os pontos vermelhos/azuis/verde representam o produto interno entre os auto vetores associados aos primeiros/segundos/terceiros autovetores mais significativos. . . . .  | 103 |
| 4.12 | Os pontos vermelhos/azuis/verdes representam o produto interno entre os autovetores associados aos primeiros/segundos/terceiros autovalores mais significativos. . . . .   | 104 |
| 4.13 | Projeção dos primeiros três versores de autovalores $\vec{A}_i$ ao longo dos semestres em sobreposição. Os pontos vermelhos/azuis/verdes representam o produto interno entre os versores associados ao primeiro/segundo/terceiro autovalor mais significativo. Quatro pontos referentes ao primeiro autovalor foram omitidos para que se tivesse uma ampliação do gráfico. Esses pontos correspondem aos semestres 105, 106, 107 e 108 e valem 0.867, 0.858, 0.862 e 0.843 respectivamente. . . . .                  | 106 |
| 4.14 | Projeção do autovetor mais significativo para cada horário no Versor Uniforme. A linha tracejada vermelha corresponde ao ajuste linear de todo o período intra-diário enquanto a linha tracejada laranja corresponde ao ajuste linear do período matutino e a linha tracejada rosa corresponde ao ajuste linear do período vespertino. Note que no período vespertino o declive e o erro são da mesma ordem de grandeza apontando assim para a independência da projeção nos horários desta parte da sessão. . . . . | 110 |
| 4.15 | Produto Interno entre os três primeiros autovetores mais significativos. Os pontos vermelhos representam o produto interno entre os primeiros autovetores mais significativos, os pontos em azuis representam o produto interno entre os segundos autovetores mais significativos e os pontos em verde representam o produto interno entre os terceiros autovetores mais significativos. . . . .   | 112 |

|      |   |     |
|------|---|-----|
| 4.16 | Em cima: Peso dos cinco primeiros Autovalores mais Significativos; em baixo: Peso dos dezessete primeiros Autovalores mais Significativos. A reta vermelha representa o ajuste linear intra-diário, a reta laranja representa o ajuste linear dos dados matutinos, a reta rosa representa o ajuste linear dos dados vespertinos. . . . .                        | 115 |
| 4.17 | Pesos das somas dos dezessete primeiros (em verde e em verde escuro) e dos cinco primeiros (em azul e em azul escuro) Autovalores mais Significativos e do Peso do primeiro Autovalor mais Significativo (em vermelho e em vermelho escuro). As cores claras representam o período matutino enquanto as cores escuras representam o período vespertino. . . . . | 116 |
| C.1  | Variância do volume em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .   | 142 |
| C.2  | Variância do volume em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .   | 143 |
| C.3  | Assimetria do volume transacionado em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .  | 144 |
| C.4  | Assimetria do volume transacionado em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .  | 145 |
| C.5  | Curtose do volume transacionado em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .   | 146 |
| C.6  | Curtose do volume transacionado em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos. . . . .   | 147 |

# Lista de Tabelas

|     |  |    |
|-----|--|----|
| 2.1 | Empresas analisadas e seus respectivos símbolos ( <i>ticker symbol</i> ) usados no mercado de ações. . . . . | 12 |
|-----|--|----|

# Capítulo 1

## Introdução

Muito embora sempre tenha existido por parte dos físicos o interesse no estudo de sistemas que não têm qualquer hipótese de serem definidas através de um Hamiltoniano, esse interesse se tornou comum na comunidade de Física Estatística que se dedica ao estudo dos chamados Sistemas Complexos. A classificação de complexo surge em contraposição à abordagem estabelecida pela Termodinâmica; segundo os postulados fundamentais deste ramo da Física toda a informação necessária para a descrição do estado de um sistema microscópico está totalmente contida em uma função de poucas quantidades extensivas [1]. Assim, um sistema complexo será tal que não verifica esta propriedade. Ou seja, para um sistema complexo, a equação fundamental deve ser também função de uma série de detalhes, nomeadamente pormenores de natureza microscópica que são desprezáveis na descrição de sistemas simples. Em último caso, este tipo de relação implica na existência de mecanismos de influência mútua a diferentes escalas de tal forma que o comportamento do sistema como um todo influencia diretamente o comportamento

individual dos seus elementos e vice-versa.

Entre aqueles unanimemente entendidos como complexos, os sistemas financeiros representam um dos casos de maior sucesso na extensão dos limites de aplicabilidade de conceitos e técnicas nascidas no âmbito da Mecânica Estatística. Os motivos para que um sistema financeiro possa ser considerado complexo não são difíceis de entender. Por exemplo, pode se afirmar que se este tipo de sistemas fosse simples, se teria a validação da teoria do mercado eficiente no qual toda a informação a propósito de um ativo está contida no seu preço. Claramente os mercados apresentam constantes ineficiências e quando se pretende explicitar o seu estado é fácil verificar que muitos fatores pesam sobre o estado de um mercado, sejam eles externos (*e.g.*, situação política, medidas econômicas por bancos centrais, relações internacionais) ou internos como gestão de portfólios, interação entre o preço de ações e o setor de atividade ou a relação do ativo com outros instrumentos financeiros, nomeadamente a relação entre o preço do crude e as ações de companhias petrolíferas e do setor energético em geral.

Embora a maioria do trabalho com vista à caracterização de mercados financeiros em um contexto de complexidade e de aplicações de Física a Finança Quantitativa se tem focado em propriedades de flutuações de preços e da volatilidade, existem outras quantidades que são importantes na descrição do comportamento dinâmico e estatístico. De entre essas quantidades existe uma que claramente se destaca: o **volume negociado**,  $v$ , definido como a quantidade de ações que são transacionadas durante um dado período de tempo. A razão principal de sua importância é relativamente simples: para que exista alteração de preço de uma ação negociada

em um mercado financeiro é necessário que um comprador e um vendedor estejam de acordo em negociar um certo volume  $v$  por um preço  $\Pi$  diferente do preço anterior.

Por razões diversas, uma grande maioria dos estudos de Finança Quantitativa — que engloba a chamada Econofísica — se tem focado em grandes mercados como os mercados americanos que são definidos como líquidos. Nesse caso, existe tradicionalmente um grande número de potenciais compradores e vendedores altamente concentrados nos primeiros níveis dos livros de ordem e por conseguinte grandes variações de preço são associadas com fortes pressões compradoras(vendedoras). Por este motivo emergiu a percepção, entretanto difundida como dito financeiro de que “é necessário volume para que o preço se altere” [2]. Em outras palavras, a emergência de um desequilíbrio entre o número de licitantes e ofertantes e que é altamente passível de ser causado por algum fluxo de informação corresponde ao fator que leva um agente de mercado a tomar uma determinada decisão. Desta maneira pode se entender o volume negociado como uma representação de informação no sistema.

Esta relação acaba por estar na origem do que separa duas teorias financeiras distintas e antagônicas que ligam flutuações de preço, volume negociado e informação. Estas são a Hipótese de Mistura de Distribuições (MDH) e a Hipótese de Chegada Sequencial de Informação (SAIH). A primeira, foi introduzida qualitativamente por Clark [3] conjecturando que a informação chega simultaneamente a todos os agentes do mercado e que a dinâmica das quantidades financeiras é dominada por eventos latentes — que levam à difusão de informação — resultando em uma distribuição conjunta em que a volatilidade e o volume negociado são descritas

por distribuições Log-Normal. Vale dizer que em seu cenário, Clark acaba por introduzir um conceito que é particularmente caro aos físicos que estudam sistemas complexos: a diferença entre o tempo físico (marcado pelo relógio) e o tempo próprio que é definido pelos eventos, que neste caso é chegada de informação. Este tipo de distinção é crucial na descoberta de várias leis em fenômenos complexos como terremotos e avalanches [4].

Contrariamente, o cenário SAIH introduzido por Copeland [5], conjectura que a informação chega aos agentes no mercado de forma assíncrona de tal maneira que o estado estacionário de um mercado é precedido por uma sequência de estados estacionários locais.

Apesar de estudos mais recentes sobre a dinâmica de livros de ordem transmitir a ideia da inverosimilhança da máxima “para que o preço se altere é necessário volume”, a verdade é que uma análise simples da relação entre o volume negociado e as flutuações de preço permite verificar que 76% das variações de preço diárias com magnitude superior a 4 desvios padrão (que é da ordem de 1%) estão relacionadas com valores de volume negociado superiores a 150% do valor médio do volume diário negociado. Temos assim mais uma indicação da relevância do volume negociado na descrição efetiva de um mercado financeiro.

## **1.1 Objetivo desta tese e estrutura**

Com o advento da informatização foi possível a aquisição e tratamento de enormes quantidades de dados referentes aos mais diversos tipos de sistemas incluindo mercados financeiros. A

ampliação do acesso a dados financeiros significou em grande medida uma mudança radical na escala de análise dos mercados. Se na década de 80 fruto do aparecimento do microcomputador foi já possível realizar estudos à escala da hora o estado atual das tecnologias de informação permite a aquisição e tratamento de um conjunto de dados gigantesco (hoje em dia pomposamente tratados como *big data*) o que incluiu a análise da dinâmica de todo um mercado financeiro ordem a ordem, negócio a negócio. Esta situação possibilitou a descoberta ou a confirmação de novas propriedades estatísticas de quantidades financeiras habitualmente designadas de  *fatos estilizados*. Neste caso se destaca a forma em lei de potência assintótica da maioria das distribuições de probabilidade ou o decaimento lento das funções de auto-correlação ou a natureza da matriz das correlações entre as flutuações de preço dos diferentes títulos [6][7]. A estas acrescentam propriedades relacionados com a não-estacionariedade dos dados financeiros [9, 10]. Neste campo existe desde os anos 80 uma verificação quantitativa de que a amplitude das flutuações de preço é sustentadamente mais elevada no início e no fim das sessões[11] conferindo a esta quantidade um perfil em forma de U que foi posteriormente verificado também para o volume negociado [12]. Este efeito é atribuído a diferentes causas como as ações levadas a cabo pelos chamados *chartistas* e provedores de liquidez (*market makers*), ou agentes de mercado que tentam fazer lucro sobre oscilações do preço ao longo do dia e que não estão interessados numa exposição ao risco das flutuações *overnight*. Dessa forma, esses agentes começam o dia com dinheiro, estabelecem as suas posições pela manhã e tendencialmente as fecham terminando novamente com um portfolio basicamente composto por dinheiro. Uma outra hipótese aven-

tada se prende com regras de liquidação implementadas em certos mercados que favorecem o surgimento desse perfil.

Existem teorias para explicar a concentração de negociações na abertura e encerramento do mercado, como por exemplo, as hipóteses baseadas no modelo de Admati & Pfleider [2], que consideram que a abertura e o fechamento se distinguem pelo fato de os preços caírem depois e um pouco antes do encerramento diário do mercado tornando impossível continuar as negociações, podendo causar assim um aumento na liquidez ao abrir e ao fechar. Outra hipótese é que a concentração de negociações no fim do dia se deve as regras de liquidação seguidas por vários mercados sob as quais todas as negociações realizadas em um determinado dia serão somente liquidadas posteriormente nos próximos dias, gerando, por conseguinte, uma tendência de que as negociações sejam concentradas no final onde há o maior número de intervalos de tempo hábeis para diferentes operadores de liquidez. Outra hipótese para a concentração na abertura é que se provedores de liquidez percebem a existência de demanda de liquidez após o fechamento (after-market), então tenderão a satisfazer essa demanda o mais breve possível.

Os estudos até agora existente sobre o perfil intra-diário de quantidades financeiras apresentam duas características: se centram na análise da média e consideram o processo como completamente estacionário independentemente da duração do período considerado. No que diz respeito ao primeiro caso e para as flutuações de preço, a análise de um eventual perfil intra-diário dos momentos estatísticos de ordem superior foi levada a cabo por Allez & Bouchaud [13]. Adicionalmente, os mesmo autores procuraram entender o comportamento das

relações entre as flutuações de preços de diferentes empresas por meio do estudo da matriz de correlação ao longo do dia de negócios.

Habitualmente, as análises de dados (financeiros) pressupõem um grau elevado de estacionariedade das séries temporais pelo que são sujeitas a tratamentos diversos. No caso da presente tese, me foquei precisamente no estudo de não-estacionariedades. Desta forma não foi feito qualquer tratamento prévio dos dados, todavia, tendo em linha de conta o fator inflacionário que introduz uma não-estacionariedade de baixa frequência nos dados. Por esse motivo, não foram analisados períodos superiores a um semestre, de tal forma que esses efeitos possam ser considerados desprezíveis, pelo menos em primeira ordem, ao mesmo tempo que garante uma estatística com significado: cada empresa apresenta sempre cerca de 2386 registros, algo que já estaria sobre risco caso se optasse por analisar intervalos trimestrais. Uma outra opção plausível seria a utilização de intervalos de duração variável seguindo a premissa c os métodos apresentados primeiramente em [14] e aplicados a dados financeiros em [9]. No entanto, sob o ponto de vista macroscópico, uma divisão semestral representa uma escala plausível, capaz de fornecer resultados relevantes.

Tal como referido anteriormente, o volume negociado apresenta sazonalidades intra-diárias em sua série temporal, quanto mais não seja em consequência do ritmo pelo qual o mercado é operado, o ritmo humano, apesar de cada vez mais os robôs operadores representam uma fatia importante do volume total. O mercado financeiro abre de manhã e fecha ao fim da tarde, permanecendo fechado durante os fins de semana e feriados. Além disso, os salários são pagos

mensalmente, e portanto, os portfólios, ou carteiras, são reequilibrados também mensalmente. A presença de tais periodicidades foi estudada sobre a volatilidade por Andersen & Bollerslev [15] concluindo sobre a existência de um comportamento conhecido por “U-effect” [2, 15] sobre o volume médio das ações negociadas por Jain e Joh [?], que observou o mesmo tipo de padrão que ficou conhecido como U-shaped que descreve o padrão intra-diário em forma de U das médias das observáveis das ações individuais: a média da observável aumenta assim que o mercado é aberto e então passa a diminuir progressivamente, tendo o seu mínimo perto da hora do almoço voltando a aumentar progressivamente até o fechamento do mercado [2] sugerindo também que os retornos na abertura e fechamento do mercado são mais variáveis que durante o período em que o mercado está aberto.

Tendo-se já feito notar da importância do volume negociado, como uma putativa medida do fluxo de informação, na descrição do mercado financeiro e a exibição de um perfil intra-diário por esta quantidade, é relevante analisar como se comportarão os momentos de ordem superior do volume negociado ao longo do dia. Mormente, pode se tentar entender como é que o volume negociado por um determinado título se encontra correlacionado com o volume negociado por outra companhia, ou seguindo a assunção da relação volume-informação, de que forma a informação relativa a empresas diferentes está relacionada.

Neste sentido, o estudo que apresento em minha tese se enquadra no mesmo contexto do estudo apresentado na referência [13]. No entanto, existem diferenças substanciais que vão além de análise de quantidades completamente diferentes. Em seu estudo sobre perfis intra-diários

de flutuações de preço os autores de [13] consideram o comportamento do mercado financeiro como estacionário. Relembrando que um mercado financeiro é um dos componentes de um sistema econômico e que este vive fases de expansão e retração a hipótese de estacionariedade do comportamento intra-diário é demasiado forte e merece pelo menos a avaliação de sua significância estatística. Assim, os resultados que compõe, esta tese compreendem este elemento de análise completamente inédito.

O restante desta tese se encontra organizada da seguinte forma: No Capítulo 2 é feita uma apresentação dos dados que serão objeto de análise assim como descrição dos métodos que foram utilizados, *i.e.*, as medidas estatísticas empregues e as formas de tratamento da matriz de correlação sobretudo quando se está perante conjuntos de dados evidentemente finitos. No Capítulo 3 apresento os resultados sobre o perfil intra-diário e não-estacionariedade dos momentos estatísticos do volume negociado de ações individuais com o objetivo de explorar as propriedades do padrão  $\cup$  existente para o volume médio e de estudar se existem outros padrões intra-diários para a dinâmica dos momentos estatísticos de ordem superior; o Capítulo 4 é dedicado aos resultados sobre a evolução intra-diária da matriz de correlação e são apontados alguns resultados sobre a não-estacionariedade dessas matrizes. Um resumo dos resultados obtidos, conclusões que destes se podem extrair das análises efetuadas bem como considerações sobre futuros trabalhos são remetidas para o Capítulo 5. Ao longo destes capítulos apresentarei principalmente resultados de média aritmética para todas as empresas e nos Apêndices se apresentam resultados específicos para diferentes empresas sobretudo como ilustração.

# Capítulo 2

## Dados e Métodos

### 2.1 Os dados analisados

Os dados utilizados no tratamento correspondem às 30 companhias que compunham o índice Dow Jones Industrial Average durante os segundo semestre de 2004. Essa companhias encontram-se indicadas na Tabela 2.1.

O sufixo .N representa que essas ações são cotadas pela Bolsa de Nova Iorque (NYSE) e o sufixo .OQ representa que essas ações são cotadas pela NASDAQ Stock Market . As duas operam no período entre 09:30 horas até as 16:00 horas, porém as duas empresas negociadas pela NASDAQ apresentam um período de negociação pré-abertura e pós-fechamento. Essas negociações fora do período regular não foram consideradas já que após o horário de fechamento e antes do horário de abertura a curva passa a apresentar um perfil muito diferente daquele apresentado durante o período comum.

| <b>Empresa</b>                        | <b>Ticker symbol</b> |
|---------------------------------------|----------------------|
| Alcoa Inc.                            | AA.N                 |
| American International Group Inc.     | AIG.N                |
| American Express Co.                  | AXP.N                |
| Boeing Co.                            | BA                   |
| Citigroup Inc.                        | C.N                  |
| Catterpillar Inc.                     | CAT.N                |
| E.I. DuPont de Nemours & Co.          | DD.N                 |
| Walt Disney Co.                       | DIS.N                |
| General Eletric Co.                   | GE.N                 |
| General Motors Corp.                  | GM.N                 |
| Home Depot Inc.                       | HD.N                 |
| Honeywell International Inc.          | HO.N                 |
| HewlettPackard Co.                    | HPQ.N                |
| International Business Machines Corp. | IBM.N                |
| Intel Corp.                           | INTC.OQ              |
| Johnson & Johnson                     | JNJ.N                |
| JPMorgan Case & Co.                   | JPM.N                |
| Coca-Cola Co.                         | KO.N                 |
| McDonald's Corp.                      | MCD.N                |
| 3M Co.                                | MMM.N                |
| Altria Group Inc.                     | MO.N                 |
| Merc & Co.                            | MRK.N                |
| Microsoft Corp.                       | MSFT.OQ              |
| Pfizer Inc.                           | PFE.N                |

|                                      |              |
|--------------------------------------|--------------|
| Procter & Gamble Co.                 | PG.N         |
| SBC Communications Inc. -> AT&T Inc. | SBC.N -> T.N |
| United Technologies Corp.            | UTX.N        |
| Verizon Communications Inc.          | VZ.N         |
| Wal-Mart Stores Inc.                 | WMT.N        |
| Exxon Mobil Corp.                    | XOM.N        |

Tabela 2.1: Empresas analisadas e seus respectivos símbolos (*ticker symbol*) usados no mercado de ações.

Estes dados foram escolhidos, pois correspondiam à nossa base de dados inicial fornecida pela Olsen Data contendo os valores de preço e volume dessas empresas registrados minuto a minuto. Posteriormente, esses dados foram complementados com dados equivalentes, mas para o período compreendido entre 4 de Janeiro de 2004 e 30 de Junho de 2013 que foram gentilmente cedidos pela Cátedra de Finança Quantitativa da École Centrale de Paris — Supélec. Durante todo o período as 30 companhias podem ser entendidas como empresas de referência ou *blue-chip*.

Duas empresas tiveram um tratamento diferenciado quando foi montada a tabela de dados durante o período do primeiro semestre de 2004 até o primeiro Semestre de 2013.

A primeira foi a empresa General Motors Corp. que começou a enfrentar uma crise financeira em 2006 e esta foi agravada após o advento da crise mundial de 2008. No primeiro semestre de 2009 a General Motors Corp. apresentou um pedido de concordata fazendo com que o governo americano tomasse controle de 61% para que ela não entrasse em falência. Por esse motivo,

durante o segundo semestre de 2009 e o primeiro semestre de 2010 não houve oferta de ações da GM. Assim, no banco de dados utilizados nesta tese, e para os cálculos das propriedades estudadas aqui, durante esse período os dados da GM são considerados como "NA", *i.e.*, como ausência de valor.

A segunda empresa foi a SBC Communications, do Texas, que em 18 de novembro de 2005 teve sua fusão concluída após comprar a empresa AT&T Corp.. A partir de então, a empresa resultante da fusão passa a ser chamada de AT&T e, no dia primeiro de novembro de 2005, o símbolo na Bolsa de Nova York da empresa resultante da fusão foi alterada de SBC para T tradicionalmente usada pela AT&T. Por este motivo, a partir do dia 03 de dezembro de 2005, passamos a usar os dados da AT&T.

## 2.2 Momentos Estatísticos

Como já referido, os principais resultados da tese dividem-se em duas partes, sendo a primeira centrada na variação dos momentos estatísticos ao longo do dia de negócios.

Genericamente as principais características estatísticas de um conjunto de  $N$  dados,  $\{x\}$ , são definidas pelos seus 4 primeiros cumulantes obtidos a partir nos momentos de ordem  $n$

$$\langle x^n \rangle \equiv \frac{1}{N} \sum_{i=1}^N x_i^n, \quad (2.1)$$

que no limite em que  $N \rightarrow \infty$  e sendo  $x$  uma variável estocástica real corresponderá a

$$\langle x^n \rangle \equiv \int x^n p(x) dx \quad (2.2)$$

onde  $p(x)$  corresponde à função de densidade de probabilidade onde se assume  $x$  continuamente distribuída.

Assim temos:

- o primeiro cumulante, ou *média*,

$$\mu = \langle x \rangle, \quad (2.3)$$

que corresponde ao valor típico da variável  $x$ ;

- o segundo cumulante, ou a *variância*,

$$\sigma^2 \equiv \langle x^2 \rangle - \langle x \rangle^2, \quad (2.4)$$

que estabelece a dispersão dos dados relativamente ao seu valor médio;

- o terceiro cumulante,

$$\langle \langle x^3 \rangle \rangle \equiv \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3, \quad (2.5)$$

donde se calcula a *assimetria*,

$$\gamma_1 \equiv \frac{\langle \langle x^3 \rangle \rangle}{\sigma^3}. \quad (2.6)$$

- e o quarto cumulante,

$$\langle\langle x^4 \rangle\rangle \equiv \langle x^4 \rangle - 4 \langle x^3 \rangle \langle x \rangle - 3 \langle x^2 \rangle^2 + 12 \langle x^2 \rangle \langle x \rangle^2 - 6 \langle x \rangle^4, \quad (2.7)$$

que permite o cálculo da curtose,

$$\gamma_2 \equiv \frac{\langle\langle x^4 \rangle\rangle}{\sigma^2}, \quad (2.8)$$

que oferece uma medida da agucidade dos dados.

Sob o ponto de vista prático, quando se lida com um número finito de dados o cálculo de  $\gamma_1$  e  $\gamma_2$  pode estar sujeito a um erro significativo. Técnicas estatísticas como o *bootstrapping* ou o *jackknifing* [16] e que se baseiam na realização de médias de momentos sobre sub-conjuntos de dados formados a partir do conjunto em mãos são relativamente lentos. Nesse sentido e com vista a ter uma medida simples de quantidades de assimetria e curtose utilizamos as mesmas expressões de Allez & Bouchaud [13],

- Assimetria:

$$\tilde{x}_i^3(t; l) = \frac{6}{\sqrt{\tilde{x}_i^2(t; l)}} [\tilde{x}_i(t; l) - m_i(t; l)], \quad (2.9)$$

- Curtose:

$$\tilde{x}_i^4(t; l) = 24 \left( 1 - \sqrt{\frac{\pi}{2}} \frac{\frac{1}{D} \sum_{d=ds}^{de} |x_i(d; t; l) - \tilde{x}_i(t; l)|}{\sqrt{\tilde{x}_i^2(t; l)}}} \right) + \langle \tilde{x}_i^3(t; l) \rangle^2. \quad (2.10)$$

onde  $l$  é a frequência dos dados equivalente neste trabalho a 1 minuto,  $t$  representa o horário,  $d$  representa o dia,  $\tilde{x}_i(t; l)$  e  $\tilde{x}_i^2(t; l)$  representam a média e a variância tal como dada pelas Eqs. (2.3) e (2.4),  $D$  representa a quantidade de dias do período estudado,  $d_s$  o dia que começa o período,  $d_e$  o dia que termina o período e  $m$  representa a *mediana* definida como sendo o valor para o qual a densidade de probabilidade acumulada vale 1/2. A vantagem desta abordagem é que apenas está envolvido o cálculo dos dois primeiros momentos e da mediana.

Estas expressões podem ser obtidas assumindo que a distribuição da variável estocástica é definida utilizando uma série de Gram-Charlier tomando como distribuição de referência a distribuição normal

$$\mathcal{N}(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \quad (2.11)$$

ou seja,

$$p(x) = \exp\left[\sum_{s=1}^{\infty} (-1)^s \frac{(\langle\langle x^s \rangle\rangle - \langle(x^s)\rangle_{\mathcal{N}})}{s!} \frac{d^s}{dx^s}\right] \mathcal{N}(x). \quad (2.12)$$

Vale realçar que a assimetria  $\xi$  corresponde a um múltiplo do chamado segundo coeficiente de assimetria de Pearson,

$$\xi_P = 3 \frac{\mu - m}{\sigma}. \quad (2.13)$$

Para além destas medidas existem outras definições de assimetria e curtose, nomeadamente as diferenças e razões entre quantis [17].

## 2.3 Teoria da Matriz de Correlação

A dificuldade em entender as propriedades das matrizes com entradas aleatórias vem sendo abordada desde 1950 pela Física Nuclear a partir do trabalho de Wigner e posteriormente por Dyson e Mehta, sendo os resultados bem conhecidos. Neste contexto, o problema estava em entender os níveis energéticos de um núcleo complexo quando o modelo falhava ao tentar explicar os dados experimentais.

Foi então que Wigner supôs que as interações entre os constituintes do núcleo são tão complexas que podem ser modeladas como aleatórias. Por conseguinte, ele assumiu que a Hamiltoniana que descreve um núcleo pesado tem como a matriz de representação  $H$ , sendo os elementos  $H_{ij}$  números aleatórios independentes entre si. Com base neste pressuposto, Wigner derivou [18, 19, 20] propriedades com as estatísticas dos valores próprios de  $H$  que estavam em acordo com os dados experimentais.

As previsões da Teoria das Matrizes Aleatórias representam uma média ao longo de todas as interações possíveis [21, 22] e os desvios em relação a essas previsões universais identificam um sistema específico, com propriedades não aleatórias, fornecendo pistas sobre as interações subjacentes [23, 24].

Dentro das classes das matrizes, Wigner as considerou reais e simétricas e seus elementos eram distribuídos de acordo com uma Distribuição de Probabilidades Gaussiana chamada de

Ensemble Ortogonal Gaussiano. Para essas matrizes, Wigner mostrou

$$P_{GOS}(s) = \frac{\pi s}{2} \exp\left(-\frac{\pi}{4}s^2\right) \quad (2.14)$$

como sendo a distribuição de níveis de energia de espaçamento  $s$ .

Os desvios em relação a Teoria das Matrizes Aleatórias indicam propriedades que são específicas do sistema e que decorrem da presença de modos coletivos. O comportamento coletivo pode ser detectado através da análise da distribuição dos valores próprios da matriz de correlação.

A partir da Teoria das Matrizes Aleatórias, pôde-se analisar as propriedades estatísticas dos valores próprios  $\lambda_\alpha$  das matrizes de correlação  $H$  entre os volumes médios das  $N$  empresas através da densidade de valores próprios

$$\rho_N(\lambda) = \frac{1}{N} \sum_{\alpha=1}^N \delta(\lambda - \lambda_\alpha), \quad (2.15)$$

onde  $\lambda_\alpha$  trata-se dos autovalores das matrizes de correlação simétrica  $H$  pertencente ao conjunto estatístico criticamente analisado. Agora, pode se introduzir a transformada de Stieltjes sendo  $z$  um número complexo:

$$G_H(z) = \frac{1}{N} \text{Tr}[(zI - H)^{-1}], \quad (2.16)$$

que tem como espectro

$$\rho_N(\lambda) = \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \Im(G_H(\lambda - i\epsilon)). \quad (2.17)$$

No limite onde  $N \rightarrow \infty$ , a densidade dos autovalores tende, quase certamente, a uma densidade única bem definida, significando que a densidade de autovalores se torna independente de uma específica matriz  $H$ , torna-se  $H$  uma amostra típica do ensemble. Esta propriedade matemática corresponde à propriedade de ergodicidade, permitindo que a densidade de valores próprios assintoticamente obtida tem correspondência com a densidade de valores próprios obtida em uma única realização. Esta propriedade é de suma importância dado que na maioria dos sistemas complexos, mercados financeiros incluídos, não é possível repetir a história de eventos.

Duas outras transformadas também muito importantes são: a chamada função Blue, que é a funcional inversa de  $G(z)$  ou seja,

$$B[G(z)] = G[B(z)] = z,$$

e a  $S$ -transformada. Pode-se correlacionar a  $R$ -transformada com a função Blue através da equação [25]:

$$R(z) = B(z) - z^{-1} \quad (2.18)$$

que tem como propriedade, para qualquer número real arbitrário  $a$ ,

$$R_{aH}(z) = a R_H(az). \quad (2.19)$$

Além disso,  $R(z)$  pode ser expandido para grandes valores de  $z$  como  $R(z) = \sum_{k=1}^{\infty} c_k z^{k-1}$ , sendo que os coeficientes  $c_k$  podem ser tomados como cumulantes, como por exemplo  $c_1 = \int d\lambda \lambda \rho(\lambda)$ . Assim, se  $c_1 = 0$ , tem-se que  $c_2 = \int d\lambda \lambda^2 \rho(\lambda)$ .

Quanto à  $S$ -transformada, ela se correlaciona com a função Blue através da definição [25]

$$S(z) = -\frac{1+z}{z} \eta^{-1}(1+z) \quad (2.20)$$

com  $\eta(y) \equiv -\frac{1}{y} G(\frac{1}{y})$ .

Para o casos das matrizes de correlação aqui analisadas, tratam-se de matrizes de correlação construídas a partir de séries temporais finitas e independentes, ou seja, que não são correlacionadas. Esse tipo de matriz é chamada de “matriz livre”, pois apresenta variáveis aleatórias independentes. Pode-se tomar um exemplo: tome duas matrizes  $H_1$  e  $H_2$  independentes escolhidas dentro do conjunto Ensemble Ortogonal Gaussiano, desta maneira  $H_1$  e  $H_2$  são matrizes simétricas e todas as suas entradas são variáveis mutualmente independentes e identicamente distribuídas. Como a medida do ensemble de matrizes aleatórias é invariante sob uma transformação ortogonal, tem-se que a matriz de rotação  $O_1$  que diagonaliza  $H_1$  é uma matriz de rotação aleatória do grupo ortogonal  $O(N)$ . Assim, a rotação  $O_1^T O_2$  das bases próprias de  $H_1$  para as bases próprias de  $H_2$  é também aleatória e  $H_1$  e  $H_2$  são mutualmente livres.

Considerando esta propriedade e as propriedades as quais generaliza a regra de convolução para a soma de duas variáveis aleatórias independentes: matrizes livres permite que se obtenha o espectro da soma entre elas desde que as mesmas sejam mutualmente livres, assim, se  $R_A(z)$  e  $R_B(z)$  são  $R$ -transformadas de duas matrizes livres  $A$  e  $B$ , então  $R_{A+B}(z) = R_A(z) + R_B(z)$  quando o logaritmo da função característica é aditivo e, quando é multiplicativo, tem-se a  $S$ -transformada  $S_{A+B}(z) = S_A(z) S_B(z)$ ; pode voltar-se à abordagem de Wigner, mais clara-

mente, o semi-círculo para matrizes Gaussianas Ortogonais.

Considerando, tal como dito acima,  $H_1$  e  $H_2$  matrizes mutualmente livres, e considerando a estabilidade das variáveis gaussianas sob adição, tem-se que  $(H_1 + H_2) / \sqrt{2}$  se trata do mesmo ensemble e, por conseguinte, se obtém a igualdade:

$$R_{\sqrt{2}H}(z) = R_{H_1+H_2}(z) = R_{H_1}(z) + R_{H_2}(z) = 2 R_H(z). \quad (2.21)$$

Tomando  $a = \sqrt{2}$ , assumindo a normalização padrão  $\text{Tr } H^2 = 1$  e como  $R_{aH}(z) = a R_H(a z)$ , já visto anteriormente em (2.16), tem-se que:

$$2R_H(z) = \sqrt{2} R_H(\sqrt{2} z) \rightarrow R_H(z) = z. \quad (2.22)$$

Uma outra maneira usando o Teorema do Limite Central para verificar que  $R(z) = z$  é de fato a  $R$ -transformada do semi-círculo de Wigner é supor que  $H_i$  com  $i = 1, \dots, N$  trata-se de pequenas matrizes aleatórias sem traço nas quais cada elemento tem uma variância igual a  $\epsilon^2$  com  $\epsilon \rightarrow 0$  de tal maneira que é possível fazer a seguinte expansão:

$$G(z) = \frac{1}{z} + 0 + \epsilon^2 \frac{1}{z^3} + O\left(\frac{\epsilon^3}{z^4}\right) \rightarrow \frac{1}{Z} \approx G - \epsilon^2 G^3, \quad (2.23)$$

então, tomando o limite:

$$B(z) \approx \frac{1}{z - \epsilon^2 z^3} \rightarrow R(z) = B(z) - \frac{1}{z} \approx \epsilon^2 z + O(\epsilon^3 z^2). \quad (2.24)$$

Se as  $N$  matrizes são tomadas como mutualmente livres, com  $\epsilon = \frac{1}{\sqrt{N}}$  e  $N \rightarrow \infty$ , então a  $R$ -transformada da soma de cada matriz será:

$$R(z) = N\epsilon^2 z + O(N\epsilon z^2) \quad (2.25)$$

mas quando  $N \rightarrow \infty$ , por consequência,

$$R(z) \rightarrow z. \quad (2.26)$$

O resultado é um espectro de Wigner, com correções computáveis, proveniente da soma das  $N$  “pequenas” matrizes centradas para um grande valor limítrofe.

Considerando a correlação empírica de uma verdadeira matriz de correlação a qual a mesma trata-se da matriz identidade, ou seja  $C = I$ , e sendo  $E$  a soma da linha  $u$  das matrizes  $\delta E_{ij}^t = (r_i^t r_j^t)/T$  sendo estas variáveis aleatórias de variância unitária. Desta maneira,  $\delta E^t$  tem apenas um autovalor igual a  $q$ , para  $N$  grande, associado com a direção  $r^t$  e  $N - 1$  autovalores nulos correspondentes ao hiperplano perpendicular a  $r^t$ , sendo, por tanto, os diferentes  $\delta E^t$  mutualmente livres tornando possível usar a  $R$ -transformada. Fazendo:

$$\delta G^t(z) = \frac{1}{N} \left( \frac{1}{z - q} + \frac{N - 1}{z} \right) \quad (2.27)$$

e invertendo  $\delta G(z)$  para a primeira ordem em  $1/N$ , a transformada Blue leva a

$$\delta B(z) = \frac{1}{z} + \frac{q}{N(1 - qz)} \rightarrow \delta R(z) = \frac{q}{N(1 - qz)}. \quad (2.28)$$

Usando, então, a propriedade de adição da  $R$ -transformada descrita acima, se obtém:

$$B_E(z) = \frac{1}{z} + \frac{1}{(1-qz)} \rightarrow G_E(z) = \frac{(z+q-1) - \sqrt{(z+q-1)^2 - 4zq}}{2zq}, \quad (2.29)$$

reproduzindo, para  $q < 1$ , o resultado de Marcenko & Pastur para a densidade de autovalores dada por:

$$\rho_E(\lambda) = \frac{\sqrt{4\lambda q - (\lambda + q - 1)^2}}{2\pi\lambda q} \quad (2.30)$$

com  $\lambda \in [(1 - \sqrt{q})^2, (1 + \sqrt{q})^2]$ .

No limite onde  $N \rightarrow \infty$ , não há valores próprios fora do intervalo mencionado acima e, quando  $q = \frac{1}{Q}$  ou  $T \gg N$ , sendo  $Q = \frac{T}{N}$ , a densidade de autovalores converge para  $\delta(\lambda - 1)$  pois, quando  $q > 1$ , necessariamente alguns autovalores nulos aparecem no espectro, como já visto anteriormente, e este fato leva a :

$$\rho_E(\lambda) = (1 - Q)\delta(\lambda) + \frac{\sqrt{4\lambda q - (\lambda + q - 1)^2}}{2\pi\lambda q}. \quad (2.31)$$

Se considerado ainda uma matriz de correlação empírica igual à matriz identidade porém computada usando uma média móvel ponderada exponencial e tomando  $\epsilon < 1$ , como é feito frequentemente no mundo financeiro, tem-se

$$E_{ij} = \epsilon \sum_{t'=-\infty}^{t-1} (1 - \epsilon)^{t-t'} r_i^{t'} r_j^{t'} \quad (2.32)$$

que é um ensemble que satisfaz  $E_{ij} = (1 - \epsilon)E_{ij} + \epsilon r_i^0 r_j^0$  e, revertendo o resolvente de  $E_0$ , é possível encontrar a transformada elementar Blue:

$$B_0(z) = \frac{1}{z} + R_0(z), \quad (2.33)$$

com

$$R_0(z) = \frac{q}{N(1 - qx)}, \quad (2.34)$$

onde  $q$  passa a ser  $q = N\epsilon$ . Usando a Eq. (2.16) é possível achar, para a primeira ordem de  $1/N$ , a igualdade para  $R(z)$ :

$$R(z) + zR'(z) + \frac{q}{1 - qz} = 0 \rightarrow R(z) = -\frac{\ln(1 - qz)}{qz} \quad (2.35)$$

e, para achar a densidade de autovalores, usa-se a equação do espectro (2.17) resultando em:

$$\rho(\lambda) = \frac{1}{\pi} \ln G(\lambda) \quad (2.36)$$

onde  $G(\lambda)$  resolve a equação

$$\lambda q G = q - \ln(1 - q G). \quad (2.37)$$

Da Eq. (2.15):

$$\lambda_{\pm} = B(z_{\pm}) \quad (2.38)$$

onde  $B'(z_{\pm}) = 0$  e, avaliando  $B(z)$  quando  $B'(z) = 0$ , pode-se escrever a equação que tem como solução o espectro das bordas:

$$\lambda_{\pm} = \ln(\lambda_{\pm}) + q + 1. \quad (2.39)$$

Nota-se que quando  $q$  é nulo, o espectro volta a ser aquele esperado,  $\delta(\lambda - 1)$ , porém a medida que o ruído aumenta, ou o tempo característico diminui, a borda inferior se aproxima

a zero tão rapidamente quanto a  $\lambda_{..} \sim \exp(-\frac{1}{Q})$ . Embora, para estas matrizes, não existam valores próprios exatamente iguais a zero, o menor valor próprio é exponencialmente próximo a zero quando  $N \gg T$ . ou seja, quando  $Q \rightarrow 0$ .

As matrizes de correlação consideradas aqui, e também como ocorre de maneira geral no mundo real, são matrizes de correlação verdadeiras  $C$  porém com uma estrutura não trivial diferente da matriz identidade  $I$ .

Desta maneira, estende-se o resultado de Marcenko-Paseur obtido para uma matriz empírica  $E$  para uma matriz geral  $C$  de modo a tornar possível a descrição do espectro  $\rho_C$  mesmo com informações parciais de quando  $Q = \frac{T}{N} < \infty$ . No entanto, para uma matriz geral  $C$ , os diferentes projetores  $r_i^t r_j^t$  não podem ser usados com o objetivo de definir direções não correlacionadas para valores de tempo  $t$  diferentes mesmo se as variáveis aleatórias  $r_i^t$  não são correlacionadas no tempo e, por conseguinte, a abordagem acima baseada nas R-transformada não pode ser usada.

No entanto, assumindo que  $r_i^t$  tratam-se de variáveis gaussianas, a matriz empírica  $E$  pode ser escrita como

$$C^{1/2} \hat{X} [C^{1/2} \hat{X}]^T \quad (2.40)$$

sendo  $\hat{X}$  uma matriz retangular  $N \times T$  de variáveis aleatórias gaussianas não correlacionadas com variância unitária. Como os autovalores de (28) são os mesmos de  $C \hat{X} \hat{X}^T$ , é permitido o uso da transformada  $S$ , mencionada anteriormente, considerando  $A = C$  e  $B =$

$\hat{X}\hat{X}^T$  mutuamente livres e construindo o espectro de  $B$  através da lei de Marcenko-Pastur permite-se obter a equação auto-consistente para o resolvente de  $E$  [43, 44]:

$$G_E(z) = \int d\lambda \rho_C(\lambda) \frac{1}{z - \lambda(1 - q + qzG_E(z))}. \quad (2.41)$$

Através do resultado encontrado em (2.26) pode-se verificar se  $\rho_C(\lambda) = \delta(\lambda - 1)$ , de maneira equivalente.

$$zG_E(z) = ZG_C(z), \quad (2.42)$$

onde  $Z = \frac{z}{1 - q(zG_E(z) - 1)}$ , sendo possível, através destes resultados, avaliar  $-G_E(0) = \text{Tr } E^{-1}$  como uma igualdade a  $\text{Tr } C^{-1}/(1 - q)$ .

É importante notar que, enquanto for numericamente estável o mapeamento entre o real espectro  $\rho_C$  e o espectro empírico  $\rho_E$ , o mapeamento inverso será instável, um pouco como a inversão de uma transformada Laplaciana. Assim, para construir o espectro de  $C$  dado o espectro de  $E$ , não deve-se inverter diretamente o mapeamento acima mas sim usar uma aproximação paramétrica de  $\rho_C$  para ajustar o  $\rho_E$  observado.

Outro fato que deve-se notar é que, quando  $C$  apresenta autovalores isolados, os resultados acima não podem ser aplicados pois estes só descrevem partes contínuas do espectro.

Como pode ser visto até aqui, os ensembles de Marcenko-Pastur são maximamente aleatórios já que nenhuma informação prévia é computada sobre a estrutura das matrizes. Isto não é um bom aspecto para o estudo aqui apresentado já que o mercado financeiro é altamente influenciável por ações externas pertinentes a economia e a política, por exemplo.

Por conseguinte, há a existência de pelo menos um fator comum para todas as ações das empresas contabilizadas fazendo com que o coeficiente médio de correlação sobre todos os pares seja positivo e, por este motivo, a hipótese nula mais razoável a ser aceita aqui é que a matriz real  $C$  deve ter seus coeficientes assim descritos:

$$C_{ii} = 1, \quad C_{ij} = \bar{\rho}, \quad \forall i \neq j.$$

Para isto, adiciona-se à matriz de correlação empírica uma matriz de perturbação que tenha um alto autovalor  $N\bar{\rho}$  e  $N - 1$  autovalores nulos de tal modo que, quando  $N\bar{\rho} \gg 1$ , a matriz de correlação empírica terá também um alto autovalor próximo a  $N\bar{\rho}$  e, por tanto, muito acima do limite superior  $\lambda_+$  de Marcenko-Pastrur.

Para um caso mais geral,  $N\bar{\rho}$  não é tão maior que a unidade e, por tanto, apresenta  $k$  autovalores especiais chamados de “spikes” [26]. Este cenário mais geral também é visto para as matrizes de Wigner quando adicionada uma matriz de perturbação de rank  $k$ . Se os elementos aleatórios  $H_{ij}$  não tem uma média  $\bar{h}$  igual a zero, a perturbação tem um autovalor  $N\bar{h}$  diferente de zero e  $N - 1$  autovalores nulos.

Este cenário tem uma transição de fase aguda entre o regime onde esse rank de perturbação é fraco e “dissolvido” no mar de Wigner e o regime onde a perturbação é forte o suficiente para escapar deste mar. Esta transição corresponde a uma condensação do autovetor correspondente ao maior autovalor do ranking de perturbação. Mais claramente, usando a transformada  $R$  e assumindo que o valor não nulo do ranking de perturbação vale  $\Lambda$  e tom  $\vec{e}_1 = (1, 0, \dots, 0)$  como

seu autovetor correspondente, então a função resolvente  $G_\Lambda$  da perturbação e a função Blue  $B_\Lambda$  da perturbação podem ser escritas como:

$$G_\Lambda(z) = \frac{N-1}{Nz} + \frac{1}{N} \frac{1}{z-\Lambda} \rightarrow B_\Lambda(z) \approx \frac{1}{z} + \frac{1}{N} \frac{\Lambda}{1-\Lambda z}, \quad (2.43)$$

sendo esta perturbação livre em relação as matrizes de Wigner. O próximo passo então é usar a  $R$  transformada da soma para computar o resolvente correto  $G$ :

$$R_{H+\Lambda} = z + \frac{1}{N} \frac{\Lambda}{1-\Lambda z} \rightarrow z \approx G + \frac{1}{G} + \frac{1}{N} \frac{\Lambda}{1-\Lambda G}, \quad (2.44)$$

sendo o termo de correção da ordem de  $\frac{1}{N}$  e pode-se substituir  $G$  pelo resolvente de Wigner  $G_W$  para a primeira ordem. Esta correção só ira sobreviver no limite de  $N$  grande se  $\Lambda \times G_W(z) = 1$  tiver uma solução não trivial de modo a divergência compensar o fator  $1/N$ . O valor correspondente de  $z$  define o autovalor isolado, levando a [27, 42]:

$$z = \lambda_{max} = \Lambda + \frac{1}{\Lambda} \quad \text{para } \Lambda > 1,$$

$$\lambda_{max} = 2 \quad \text{para } \Lambda \leq 1.$$

Ou seja, quando  $\Lambda = 1$ , o maior autovalor salta para fora do mar de Wigner e, usando a teoria de perturbação, pode-se calcular a sobreposição entre o maior autovetor  $\vec{V}_{max}$  e  $\vec{e}_1$  [27]:

$$(\vec{V}_{max} \cdot \vec{e}_1)^2 = 1 - \Lambda^{-2} \quad \text{para } (\Lambda > 1)$$

expondo que a coerência entre o maior autovalor da matriz e o maior autovalor da perturbação vai diminuindo de maneira progressiva quando  $\Lambda \rightarrow 1^+$ .

Este fenômeno, de uma maneira semelhante, pode ser visto no caso de matrizes de correlação: para um ranking de perturbação do tipo descrito acima cujo valor próprio é  $\Lambda = N\rho$ , o critério para que um autovalor isolado seja expulso do mar de Marcenko-Pastur é:

$$\lambda_{max} = \Lambda + \frac{\Lambda q}{\Lambda - 1} \quad (\Lambda > 1 + \sqrt{q}),$$

$$\lambda_{max} = (1 + \sqrt{q})^2 \quad (\Lambda \leq 1 + \sqrt{q}).$$

No limite onde  $\Lambda \rightarrow \infty$ , tem-se  $\lambda_{max} \approx \Lambda + q + O(\Lambda^{-1})$  e, para o rank  $k$  de perturbação, todos os autovalores que satisfazem  $\Lambda_x > 1 + \sqrt{q}$  onde  $1 \leq x \leq k$  vão acabar sendo isolados acima do mar de Marcenko-Pastur e todos os outros desaparecem abaixo de  $\lambda_+$ . Além disso,

todos os autovalores isolados tem uma flutuação gaussiana da ordem de  $T^{-1/2}$ , como pode ser encontrado na Ref. [28].

Quanto à matriz de correlação entre os volumes transacionados aqui estudados, tem-se que, os elementos  $C^{ij}$  estão restrito por construção dentro do domínio  $-1 \leq C^{ij} \leq 1$  de tal modo que, quando  $C^{ij} = 1$  significa que as empresas estão perfeitamente correlacionadas, quando  $C^{ij} = -1$  significa que as empresas estão perfeitamente anticorrelacionadas e quando  $C^{ij} = 0$ , significa que as empresas não possuem nenhum tipo de correlação entre si.

Como já citado acima, as dificuldades em analisar a importância e o significado dos coeficientes de uma matriz empírica de correlação são o comprimento finito das séries temporais introduzindo um ruído de medida e o fato de que as condições de mercado mudam a todo tempo fazendo com que as correlações não sejam estacionárias.

Se forem usadas séries temporais longas para contornar o problema do tamanho finito, os resultados serão afetados pela não estacionariedade das correlações fazendo com que a matriz de correlação empírica contenham contribuições aleatórias e, então, estimar as correlações que não são resultados desta aleatoriedade torna-se um problema muito difícil.

Assim, para identificar quais destas empresa continuam correlacionadas em média num determinado período é preciso testar a "hipótese nula" da matriz de correlação aleatória construída a partir de séries temporais que não são correlacionadas entre si, também já comentada acima: se as propriedades da matriz  $C^{(v)}$  estão em conformidade com as propriedades de uma matriz

aleatória então o conteúdo de  $C^{(v)}$  medido empiricamente também será aleatório.

Mas, por outro lado e também já citado acima, os desvios da matriz  $C^{(v)}$  em relação a uma matriz de correlação aleatória transmitem informações singulares sobre as correlações ditas "genuínas". Por este motivo, o estudo aqui dirigido se atem a análise tanto do ruído quando se compara a matriz de correlação  $C^{(v)}$  com uma matriz de correlação randômica quanto das informações que se obtém dos desvios consequentes desta mesma comparação.

Ao se montar a matriz de correlação para cada semestre usando a equação (2.1), obteve-se matrizes que apresentavam alguns de seus auto valores negativos consequentes do ruído encontrado pela finitude das séries temporais de cada empresa utilizadas.

Para filtrar o ruído dessas matrizes, aplicou-se o método descrito por Rebonato & Jackel [29] conhecido como Decomposição Espectral dentro da Análise da Componente Principal que previamente já foi discutido na seção 2.1:

Seja  $S$  o auto-sistema de uma matriz real e simétrica  $U$  e  $\{\lambda_i\}$  o conjunto de autovalores de  $U$  de modo que

$$U \cdot S = \Lambda \cdot S \quad (2.45)$$

onde  $\Lambda = \text{diag}(\lambda_i)$  e define os elementos não nulos da matriz diagonal  $\Lambda'$  como

$$\Lambda' : \lambda'_i = \begin{cases} \lambda_i & \lambda_i \geq 0 \\ 0 & \lambda_i < 0 \end{cases}$$

Se  $M$  não for semidefinida positiva, ela tem ao menos um autovalor negativo e, por tanto, este  $\lambda'_i$  deverá ser nulo.

Pode-se também definir os elementos não nulos de uma matriz de escalonamento  $\mathbb{T}$  a partir do auto-sistema  $S$  por :

$$\mathbb{T} : t_i = \left[ \sum_m s_{im}^2 \lambda'_m \right]^{-1} \quad (2.46)$$

Tomando

$$B' := S\sqrt{A'} \quad (2.47)$$

e

$$B := \sqrt{\mathbb{T}}B' = \sqrt{\mathbb{T}}S\sqrt{A'} \quad (2.48)$$

tem-se, por construção que:

$$\hat{U} := BB^T \quad (2.49)$$

sendo  $\hat{U}$  tanto semi-definida positiva e tendo os elementos da diagonal principal unitários.

Assim, o método se resume em:

1. Primeiro se calcula os autovalores  $\lambda_i$  e os auto-vetores  $s_i$  de  $M$ ,
2. Em seguida tomar todos os autovalores  $\lambda_i$  negativos como nulos,

3. Multiplica-se os auto-vetores  $s_i$  com os seus autovalores associados corrigidos  $\lambda'_i$  de modo a arranjar o resultados em colunas de  $B'$  e,
4. Por último, se obtém  $B$  a partir da normalização dos vetores de linhas de  $B'$ .

Através deste procedimento é possível obter uma matriz de correlação aceitável e que ao mesmo tempo é similar à calculada anteriormente.

Depois de feita as correções, foram montados histogramas para cada matriz de correlação correspondente a um determinado horário. O resultado foram histogramas que obedeciam uma curva gaussiana porém com uma cauda maior à direita, condizente ao comportamento predito pela teoria de Wigner [30] e condizente à distribuição fornecida pela Teoria das Matrizes Aleatórias [31].

Para avaliar a distribuição dos autovalores, assim como o limite inferior e o superior teórico para os mesmos, usou-se a função de densidade de probabilidade de autovalores para uma matriz de correlação randômica no limite onde  $N \rightarrow \infty$ , sendo  $N$  o número de empresas, e  $L \rightarrow \infty$ , sendo  $L$  a quantidade de dados, tais que  $Q \equiv \frac{L}{N}$  ( $Q > 1$ ), fixo:

$$P(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (2.50)$$

para um autovalor dentro do limite  $\lambda_- \leq \lambda \leq \lambda_+$  onde  $\lambda_+$  e  $\lambda_-$  são os autovalores teóricos máximo e mínimo, respectivamente, da matriz de correlação  $C^{(n)}$  dados por:

$$\lambda_{\pm} = 1 - \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \quad (2.51)$$

Apesar de se ter  $Q > 1$ , as matrizes de correlação  $C^{(v)}$  calculadas aqui através da equação

$$c^{ij}(t; l) \equiv \frac{\frac{1}{D} \sum_{d=d_s}^{d_e} v_i(d, t; l)v_j(d, t; l) - \bar{v}_i(t; l)\bar{v}_j(t; l)}{\bar{v}_i^2(t; l)\bar{v}_j^2(t; l)}. \quad (2.52)$$

na qual  $upsilon_i(d, t; l)$  corresponde ao volume negociado num dado dia  $d$ , num dado horário  $t$  e de frequência de  $l$  de 1 minuto,  $\bar{v}_i(t; l)$  a média do volume negociado e  $\bar{v}_i^2(t; l)$  a sua variância;  $D$  é o número de dias do período considerado,  $d_s$  é o dia que começa o período e  $d_e$  o dia que termina o período, possuem um valor finito e pequeno de empresas e finito, mas não tão pequeno, de dados. Para este caso, espera-se que o ponto abrupto de corte de  $P$  que existe quando  $N$  e  $L$  tendem ao infinito seja substituído por uma borda que decai rapidamente [32].

Porém, apenas os valores teóricos máximos e mínimos foram calculados já que com 30 empresas a distribuição obtida não aferiria um resultado conclusivo.

## Capítulo 3

### Sazonalidades individuais

Após uma introdução teórica apresento então os resultados da minha análise. Como já referi, os dados foram separados em semestres contíguos, tendo ao todo 19 semestres: primeiro semestre de 2004 (PS2004), segundo semestre de 2004 (SS2004), primeiro semestre de 2005 (PS2005), segundo semestre de 2005 (SS2005),..., até o primeiro semestre de 2013 (PS2013).

#### 3.1 Relação da Média do Volume com o Tempo

Para cada semestre foram feitos gráficos da média do volume por horário em função do horário intra-diário Figura 3.1. Em todas as figuras o intervalo [1,391] nas abcissas representa os minutos da sessão, sendo que o valor 1 corresponde 09:31 horas e o instante 391 às 16:01. Além disso, para distinguir os período matutino e vespertino se utilizaram pontos azuis claros para o período matutino de 09:31 até 12:45, enquanto e pontos azuis para o vespertino de 12:46 até 16:01.

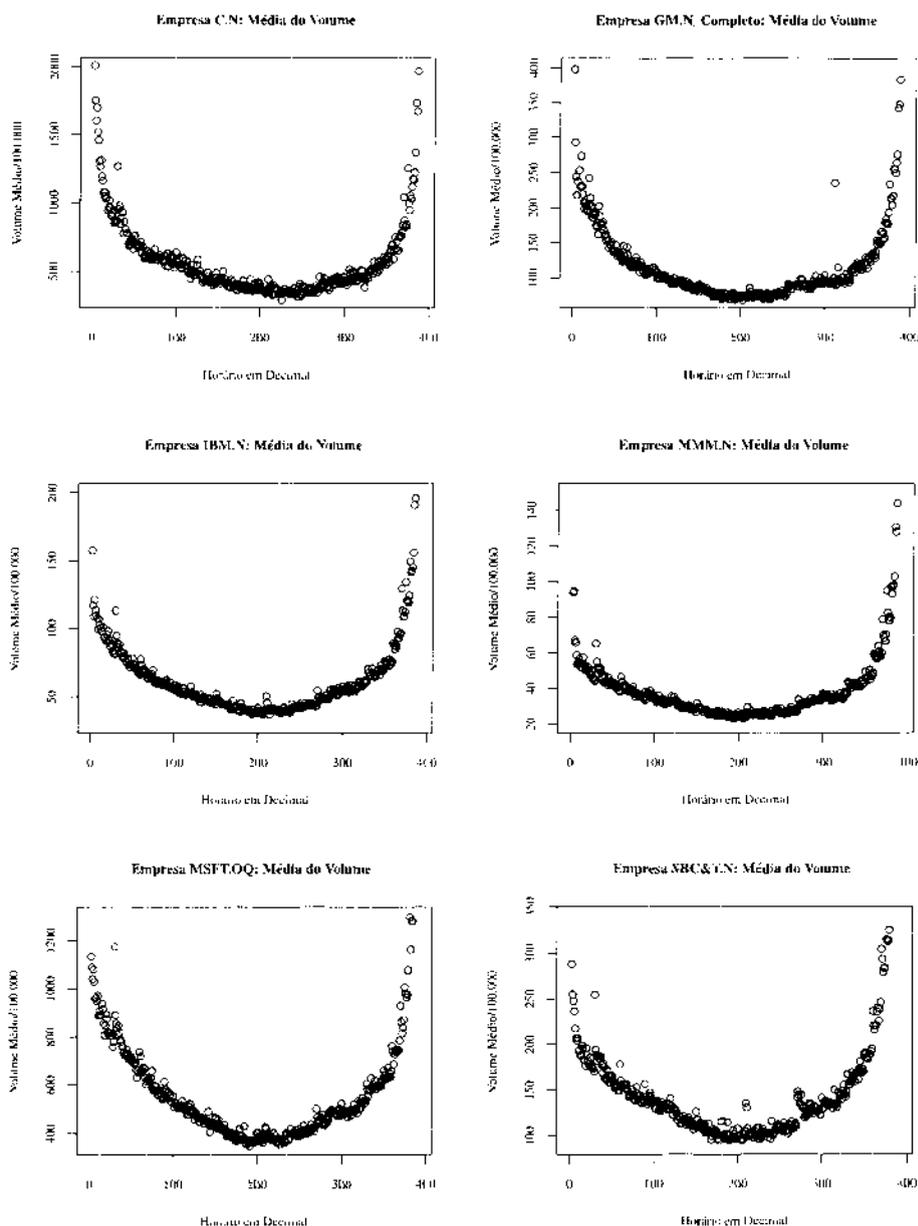


Figura 3.1: Volume médio transacionado por horário intra-diário para algumas empresas.

Para fins de visualização do comportamento individual se apresentam no Apêndice C os resultados específicos para as empresas Citigroup Inc. e Microsoft Corp. por terem seus dados em um intervalo numérico diferente e principalmente por pertencerem a bolsas diferentes. No corpo deste capítulo são apresentados os gráficos com o comportamento médio.

Foi possível observar ao se analisar a Figura 3.1 que o padrão  $\cup$  já visto anteriormente para a volatilidade média [15, 2] também foi encontrado nos gráficos do volume médio por horário em função do tempo. Além disso, os pontos de mínimo para cada empresa correspondem aos horários com média 14 horas, 20 minutos, um horário que se pode considerar como pós-almoço.

Outro ponto que foi possível observar é que, todos os gráficos apresentam um pequeno pico entre os horários 14:00 horas e 14:01 horas que corresponde ao horário de fechamento dos mercados europeus. Este pico também já foi notado quando analisado a volatilidade.

A fim de evidenciar o comportamento matutino e vespertino utilizarei a seguinte representação nas Figuras: os pontos azuis claros pertencem ao período matutino que vai das 09:31 horas (abscissa igual a 1) até às 12:45 horas (abscissa igual a 195) e os pontos azuis escuros pertencem ao período vespertino que vai das 12:46 horas até às 16:01 horas (abscissa igual a 391) Figura 3.2 .

O momento de abertura do mercado é um evento particular caracterizado pela absorção do mercado dos sentimentos e opiniões que os agentes têm sobre cada empresa ativa, assim como a conseqüente movimentação causada pelas informações que circulam no pre/after-market. Por este motivo, este momento de abertura pode ser tratado como um choque no mercado. Em muitos sistemas complexos, a relaxação pós-choque segue uma lei de potência [4], ao contrário da maioria dos sistemas simples onde essa relaxação é exponencial.

Visando descrever o movimento do mercado após a abertura, ou momento de choque, estudou-

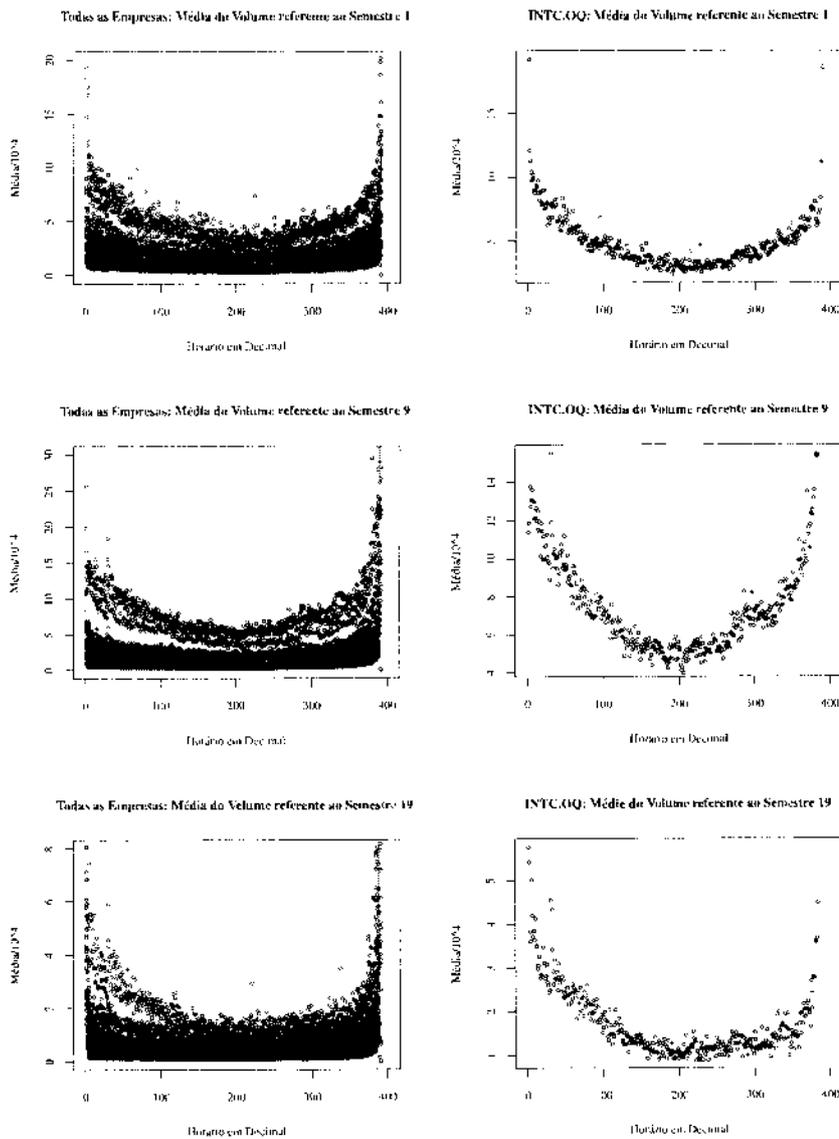


Figura 3.2: Volume transacionado por horário intra-diário de todas as empresas e somente da empresa Intel Corp..

se o decaimento da curva da Média do Volume sobre todas as empresas ao longo do tempo para cada semestre, foram feitos ajustes de lei de potência da forma

$$M_i = h_i^{-a} + \text{constante},$$

sendo  $h_i$  o horário, e com os resultados dos ajustes foi montado um gráfico dos coeficientes  $a$

em função dos semestres, como pode ser visto na Figura 3.3.

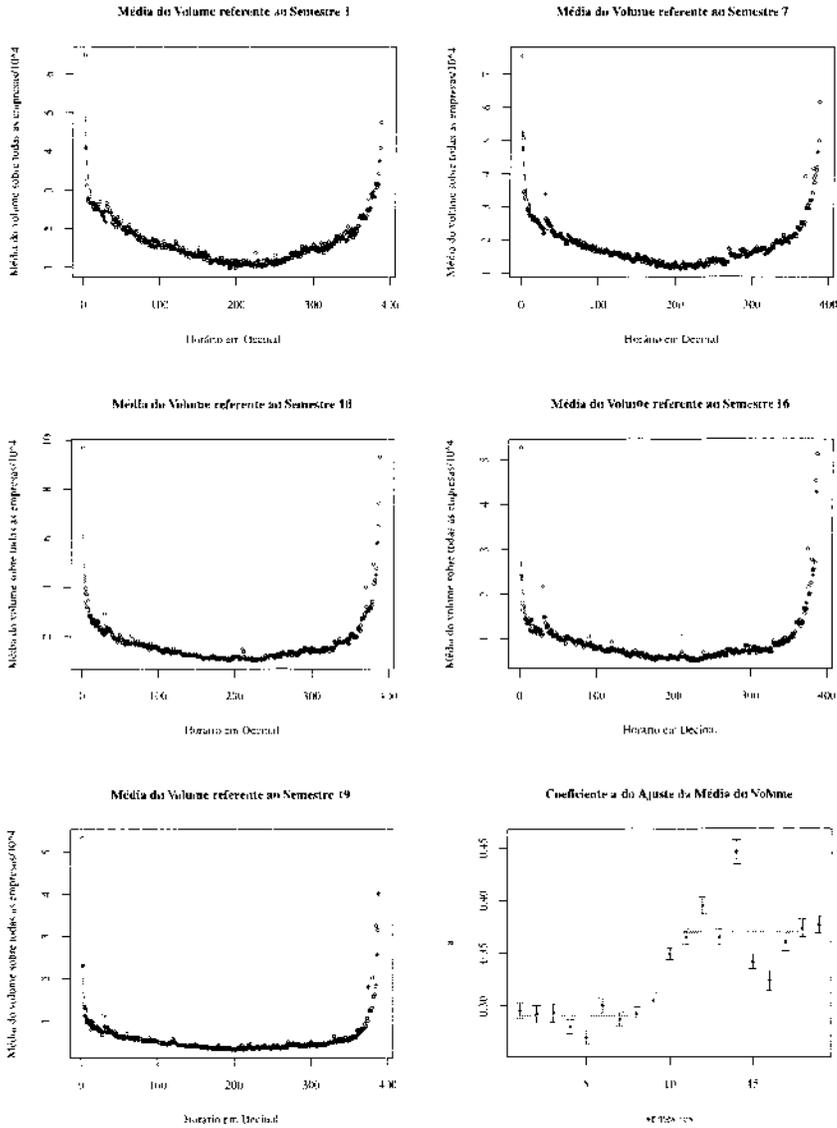


Figura 3.3: Média do volume transacionado intra-diário sobre todas as empresas para cada semestre. A curva pontilhada vermelha corresponde ao ajuste  $M_1 = h_1^{-a} + constante$ . As linhas horizontais laranja correspondem aos valores médios dos dois períodos  $a = 0.29$  e  $0.37$ , respectivamente.

No gráfico do coeficiente  $a$  do ajuste em lei de potência, Figura 3.3, é possível observar que o ponto de menor valor corresponde ao PS2006 enquanto o ponto de maior valor corresponde ao SS2010. São épocas curiosas porque no PS2006, começou a queda dos preços dos imóveis

de forma a arrastar vários bancos para uma situação de insolvência e repercutindo fortemente sobre todas as bolsas de valores do mundo inteiro e o SS2010 o Estados Unidos efetivamente começa a mostrar um crescimento de sua economia deixando no passado a recessão.

Os pontos pertencentes aos semestres que fazem parte do período das bolhas formam uma curva continuamente crescente, ou seja, do PS2006 até o SS2009 tem-se uma curva crescente no gráfico do coeficiente  $a$ , destacando-se o ponto pertencente ao SS2006 período no qual a crise imobiliária já começa ser alarmada e ao SS2008 período no qual a bolha especulativa efetivamente estoura [para mais detalhes vide Apêndice 2].

Além disso, é possível notar que a partir do SS2008 tem-se uma mudança de comportamento. Para analisar esta mudança foi feito o teste t-Student entre o subconjunto de dados formados pelos valores do coeficiente  $a$  dos ajustes do PS2004 até o PS2008 — que tem um valor médio  $a = 0.29$  — e o subconjunto formado pelos valores do coeficiente  $a$  dos ajustes do SS2008 até o PS2013, que tem um valor médio  $a = 0.37$ . Para um intervalo de confiança de 95% (ou nível de significância de 5%), o resultado do teste é de  $t_{calculado} = 5.56$  enquanto que o valor de  $t$  tabelado para 8 graus de liberdade e um intervalo de confiança de 95% é de  $t_{critico} = 1.86$ .

Como  $t_{calculado}$  é maior que o  $t_{critico}$  pode se descartar a hipótese nula de se ter médias dos dois subgrupos iguais concluindo-se que os subgrupos tem comportamentos diferentes e, portanto, pode se alegar que a partir do SS2008 há uma mudança efetiva de comportamento da curva da média em função do tempo. Sendo que o valor obtido para para a primeira parte dos

dados é igual ao valor apresentado em [13] para as relaxações das flutuações de preço após abertura do mercado, é bastante provável que também esse expoente tenha sofrido alterações.

Sem prejuízo da análise anterior relativa à relaxação, para que se pudesse verificar uma possível evolução temporal do perfil intra-diário do volume médio ao longo dos semestres, foram escolhidas diferentes funções para ajustar a curva da média em função do tempo sendo que todas essas funções tinham como padrão gráfico um perfil  $\cup$  encontrado nas curvas a serem ajustadas. As funções testadas para fazer os ajustes foram:

(a) Razão entre polinômios:

$$f_a(x) = \frac{-a(b((x^2)^n) + cx + d)}{(x + 1)(x - 1)}$$

(b) Razão entre polinômios:

$$f_b(x) = \frac{-a(((x^2)^n) + b)}{(x + 1)(x - 1)}$$

(c) Arco tangente hiperbólica:

$$f_c(x) = \operatorname{arctanh} \left[ \left( \frac{x + bx^4}{c} \right)^2 \right] + a$$

(d) Tangente hiperbólica:

$$f_d(x) = c \tanh(ax + bx^2) + d$$

(e) Polinômio de quarta ordem:

$$f_e(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0.$$

No fim a escolhida foi o ajuste polinomial de quarta ordem por se tratar de uma expressão mais simples de ajustar e se adequar aos objetivos pretendidos. Por conseguinte, ajustes polinomiais de quarta ordem foram feitos para os dados do PS2004 até o PS2013. Para isso, foi feita uma mudança no eixo dos horários de modo que o horário inicial -1 corresponde ao 09:31 e o horário final -1 corresponde ao horário 16:01.

Com os parâmetros  $a_4$ ,  $a_3$ ,  $a_2$ ,  $a_1$  e  $a_0$  obtidos do ajuste, foram calculadas a paridade das curvas através da integral da função de ajuste polinomial de quarta ordem e a concavidade das curvas através da integral da segunda derivada da função de ajuste polinomial de quarta ordem:

- Concavidade do volume transacionado:

$$C = \int_{-1}^1 (2a_2 + 6a_3x + 12a_4x^2) dx$$

- Paridade do volume transacionado:

$$P = \int_0^1 (a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0) dx - \int_{-1}^0 (a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0) dx.$$

Com os valores da paridade e da concavidade de todos os semestres para todas as empresas, foram feitos quatro tipos de análise buscando relações que se apresentam as subseções que se seguem.

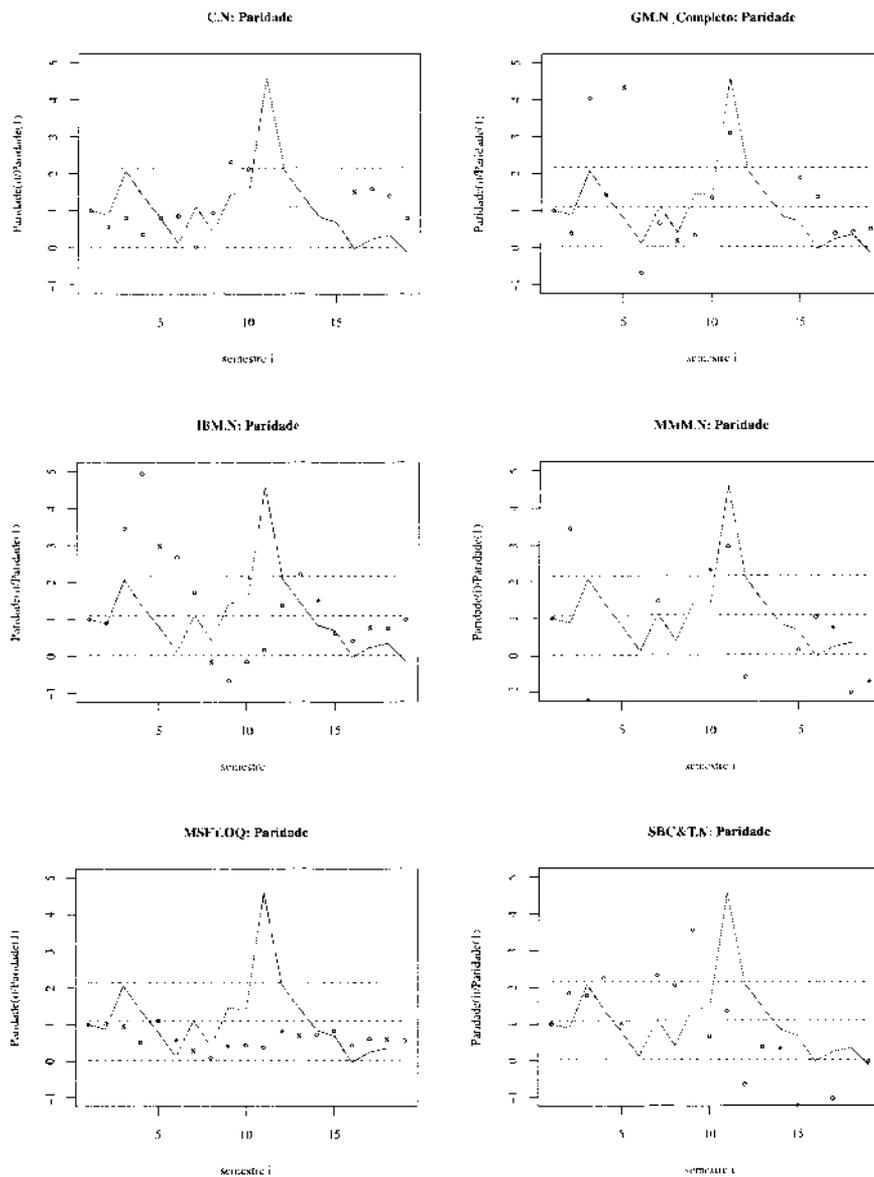


Figura 3.4: Paridade do volume transacionado para cada semestre para as empresas indicadas nas legendas. A linha vermelha indica a média sobre todas as empresas e as linhas tracejadas azuis indicam o valor médio da média sobre todas as empresas e deste somado e subtraído o desvio padrão.

### 3.1.1 Evolução da Concavidade e da Paridade do Volume

Como o mercado financeiro faz parte de um sistema social sendo sensível, por tanto, aos diversos fatores humanos como, por exemplo, ciclos econômicos e o desempenho de um determi-

nado setor, é válido sempre analisar uma eventual não estacionariedade dos fatores estatísticos descritivos.

Por esta razão, gráficos da paridade em função dos semestres, Figura 3.4, e gráficos da concavidade em função dos semestres, Figura 3.5, foram construídos para cada ação.

Para a paridade foi possível observar que existiam grupos de empresas nas quais seu comportamento se aproximava da média calculada da paridade sobre todas as empresas em função do horário: AA, AXP, DIS, GE, GM, HD, IBM, INTC, JNJ, JPM, MCD,MRK, MSFT, PFE, XOM; grupos de empresas nas quais a Paridade parecia decrescer ao longo dos semestres- SBC&T, WMT; e outro que não apresentavam um padrão- AIG, BA, C, CAT, DD, HON, HPQ, JPM, MMM, MO, PG, VZ e UTX .

Além disso, seis empresas- AIG, AXP, C, GM, KO e MMM- apresentaram uma paridade maior que 2 no Primeiro Semestre de 2009 estando, por isso, acima do valor médio somado ao desvio padrão, indicando que estas empresas se diferenciam das demais. Adicionalmente, dos valores totais de paridade, tem-se que 27.4% são valores negativos. Se utilizarmos a abordagem purista de que na realidade a cada instante se calculou o volume negociado por minuto (i.e., uma velocidade média de volume negociado), o valor da paridade indica uma preponderância genérica do período da tarde em relação ao período da manhã relativamente ao volume total negociado em uma sessão.

Para a concavidade, se verificou que todas as empresas, com exceção da Citigroup Inc.,

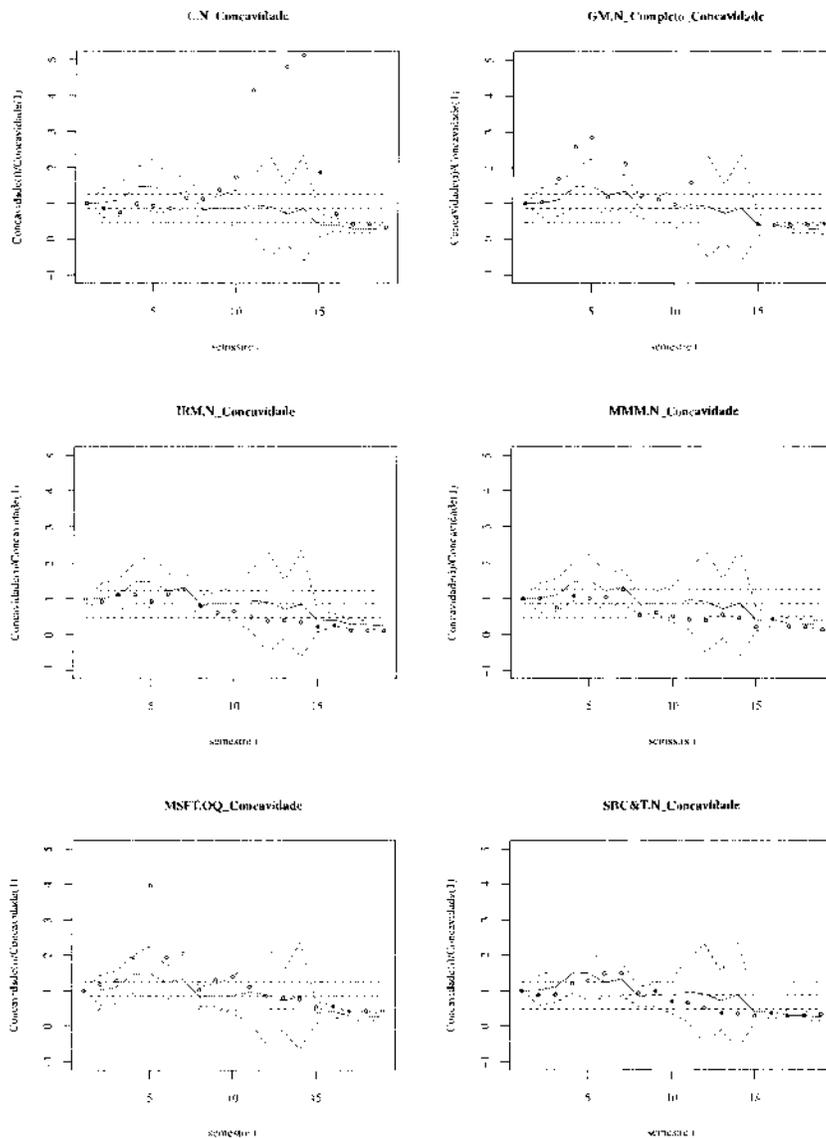


Figura 3.5: Concavidade do volume transacionado para cada semestre para as empresas indicadas nas legendas. A linha vermelha indica a média sobre todas as empresas e as linhas tracejadas azuis indicam o valor médio da média sobre todas as empresas e deste somado e subtraído o desvio padrão.

apresentam valores que flutuam bem próximos a curva média, sendo que todas elas, sem exceção, apresentam uma curva que mostra um decréscimo global da concavidade ao longo dos semestres. Este comportamento da concavidade infere um perfil U cada vez mais plano.

Além disso, pode se observar um pico bem evidente na curva média nos gráficos da paridade

[Figura 3.4] correspondente ao PS2009 com valor paridade de 4.59.

### 3.1.2 Relação entre a Paridade/Concavidade e a Média do Volume

Para procurar uma possível relação entre a paridade e a concavidade com a média do volume, que como visto nas Figura 3.4 e Figura 3.5 não é estacionária, foram feitos gráficos dessas quantidades para cada empresa em função da média do volume por semestre sobre todas as empresas.

Retas de um ajuste linear foram colocadas nos gráficos somente com o propósito de se ter uma idéia do comportamento da curva.

Para a paridade, Figura 3.6, as empresas AA, AIG, BA, C, CAT, GE,GM, HD, IBM, INTC, JNJ, MMM, MO, MRK, MSFT, PFE, WMT e XOM apresentam um comportamento decrescente em sua totalidade e as empresas AXP, DD, DIS, HON, HPQ, JPM, KO, MCD, PG, SBC&T, UTX e VZ apresentam um comportamento crescente em sua totalidade.

Já para a concavidade, Figura 3.7, o resultado para todas as empresas foi o mesmo: os valores fluuavam pouco em torno de uma reta crescente, ou seja, a concavidade em função da média do volume médio apresentou um padrão de reta crescente.

Através dos gráficos da Fig. 3.7 é possível observar que um aumento no volume médio transacionado não pode ser descrito como um simples aumento no coeficiente  $a_0$  mas sim como um aumento da concavidade fazendo com que a curva tenha sua forma em  $\cup$  mais achatada.

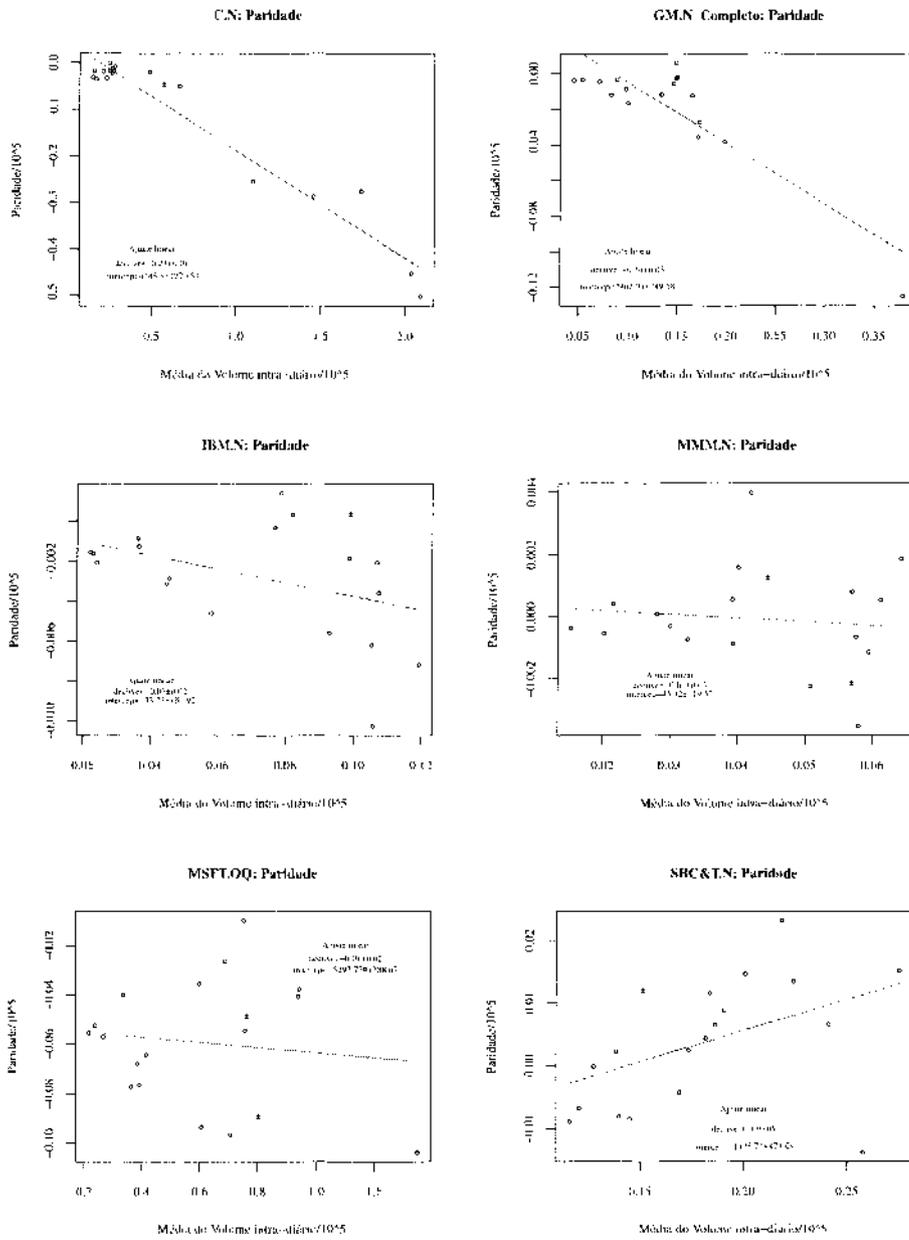


Figura 3.6: Paridade do volume transacionado para cada semestre em função do volume médio semestral para as empresas indicadas nas legendas.

Nota-se que os primeiros semestres apresentam concavidade muito baixa, praticamente nula.

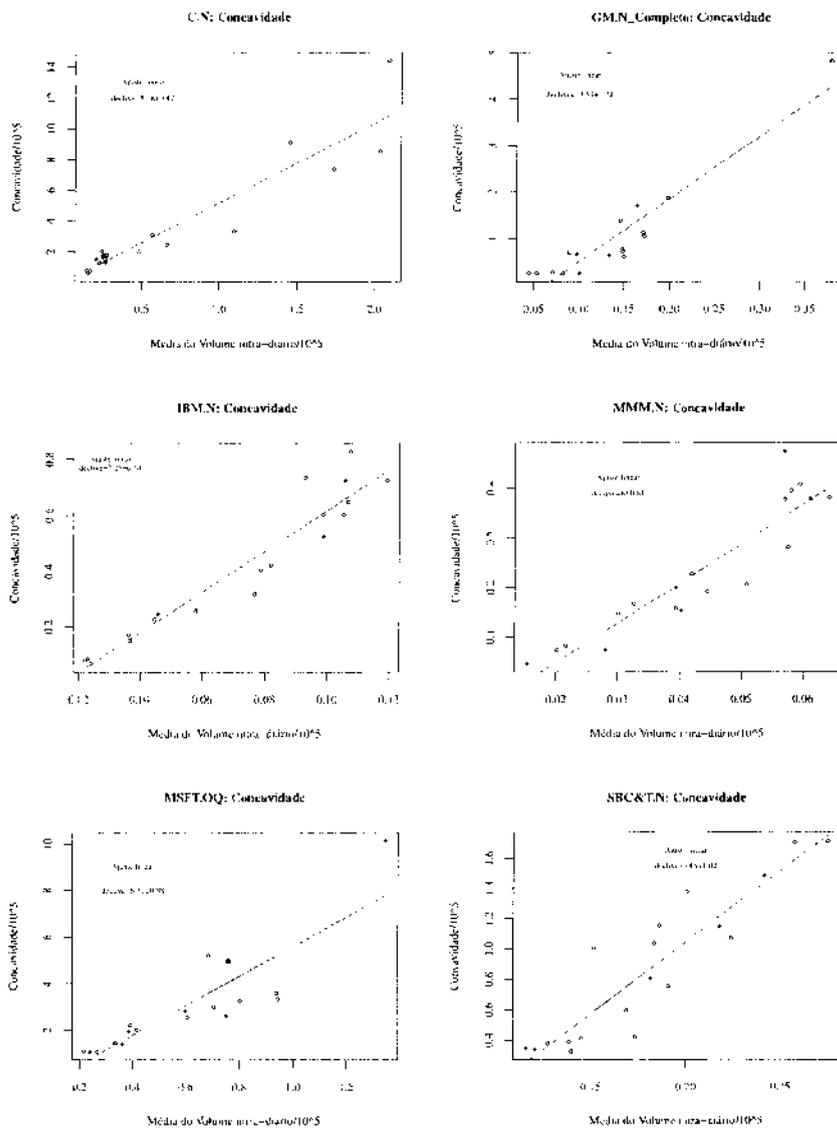


Figura 3.7: Concavidade do volume transacionado para cada semestre em função da média do volume semestral.

### 3.1.3 Relação entre a Concavidade/Paridade e as Flutuações de Preço

Como as flutuações de preços estão relacionados com o volume transacionado, foram feitos gráficos da Paridade em função da flutuação de preço do ativo, Figura 3.8, e da Concavidade em função da flutuação de preço do ativo, Figura 3.9.

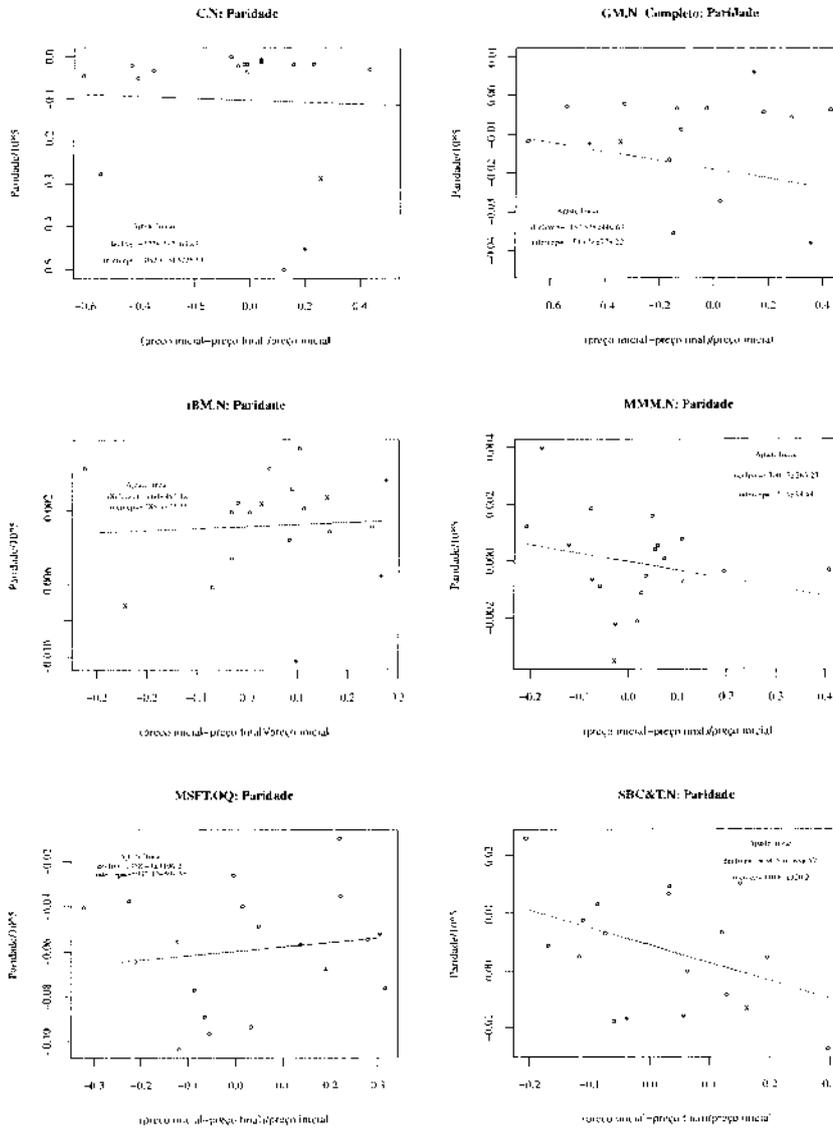


Figura 3.8: Paridade do volume transacionado para cada semestre em função da flutuação do preço do ativo. A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na variação de preço.

Os gráficos dos valores da concavidade quando da Paridade de cada empresa em função da razão

$$r(s) \equiv \frac{\Pi_f(s) - \Pi_i(s)}{\Pi_i(s)}, \tag{3.1}$$

onde  $\Pi_f$  representa o preço final e  $\Pi_i$  o preço inicial em cada semestre ,apresentaram em sua

totalidade pontos bem dispersos, porém com uma característica curiosa para a maioria delas é que o último ponto se destaca dos demais.

Para a paridade em função da flutuação do preço, é possível verificar três tipos de comportamento:

- crescente: como por exemplo a Microsoft Corp.;
- decrescente: como por exemplo a AT&T Inc. e
- aparentemente insensível: como por exemplo a Citigroup Inc..

Esta classificação de comportamento também pode ser tomada para a concavidade em função da flutuação do preço:

- crescente: como por exemplo a AT&T Inc.;
- decrescente: como por exemplo a International Business Machines Corp. e
- aparentemente insensível: como por exemplo a Citigroup Inc..

É possível notar também que, com exceção da empresa 3M Co., que a relação entre a concavidade e a flutuação de preço é inversa a entre a paridade e a flutuação de preço: empresas que apresentam um comportamento crescente em uma, na outra apresenta um comportamento decrescente. Empresas com uma concavidade aparentemente insensível em relação à flutuação de preço, continuam a sê-lo para a paridade em função da flutuação de preço.

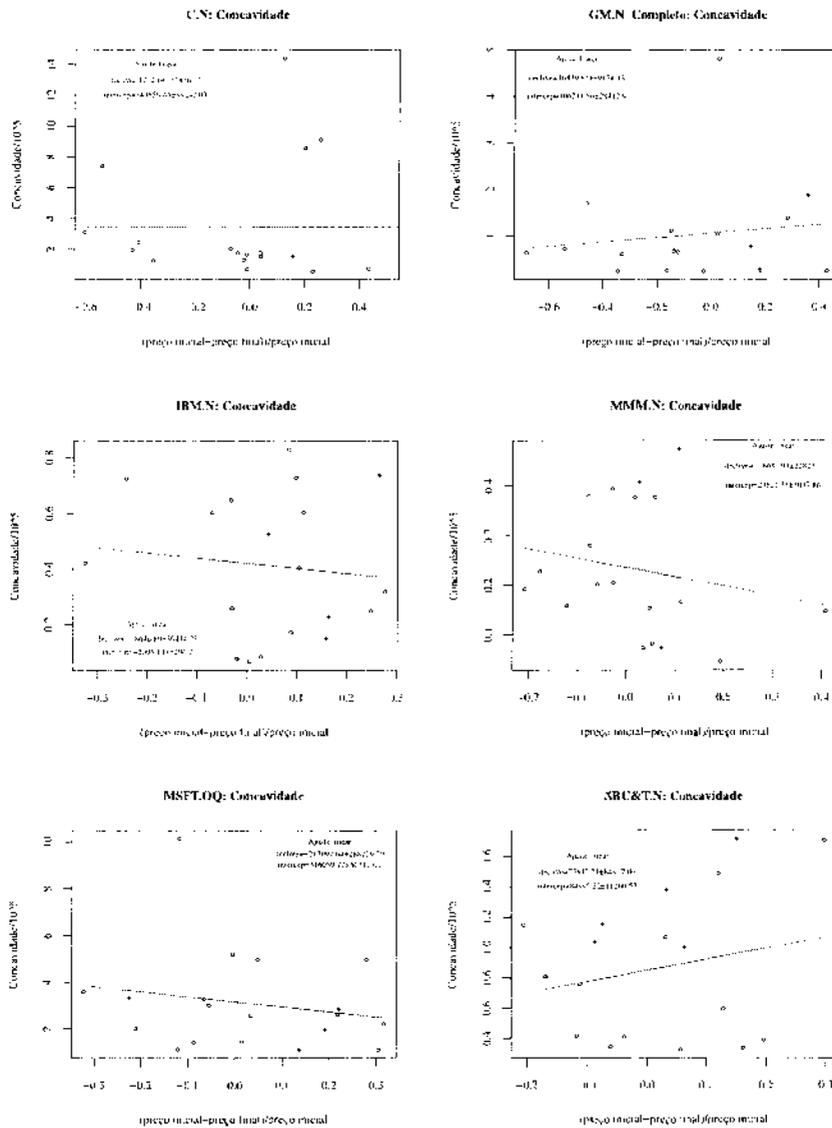


Figura 3.9: Concavidade do volume transacionado para cada semestre em função da flutuação do preço do ativo. A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na variação de preço.

### 3.1.4 Relação entre Concavidade/Paridade e a Volatilidade

A Volatilidade pode ser quantitativamente descrita por medidas relacionadas com a dispersão dos retornos e associada à variedade de opiniões existentes no mercado que tem por base a

existência de agentes com diferentes níveis de informação. Para calcular esta medida, foi usado o coeficiente de Volatilidade calculado através do Estimador de Volatilidade High-Low-Open-Close

$$\sigma_{HLOC} = \sqrt{\frac{252}{n} \sum_{i=1}^n \left[ \frac{1}{2} \left( \ln \frac{H_i}{L_i} \right)^2 - (2 \ln 2 - 1) \left( \ln \frac{C_i}{O_i} \right)^2 \right]}. \quad (3.2)$$

onde  $H_i$  equivale ao maior preço dentro do intervalo de 1 minuto,  $L_i$  o menor preço,  $O_i$  o preço de abertura do intervalo e  $C_i$  o preço de fechamento do intervalo, sendo 252 dias por semestre, obtido considerando que, para pequenas flutuação de preço,

$$r_i = \ln \frac{\Pi_{i+1}}{\Pi_i}$$

pode ser aproximado à  $r_i = \frac{\Pi_{i+1} - \Pi_i}{\Pi_i}$ , como pode ser visto de maneira detalhada no Apêndice 1.

Para analisar como essa medida se relaciona com a paridade e a concavidade, foram construídos gráficos tanto para a paridade, Figura 3.10, quanto para concavidade, Figura 3.11, em função dos coeficientes de volatilidade.

Aqui mais uma vez, o propósito do ajuste linear foi dar uma idéia do comportamento da distribuição.

Tanto para a paridade quanto para a concavidade, não houve nenhum comportamento padrão das curvas e nem certa tendência.

Este resultado discorda com a Hipótese de Mistura a qual afirma que o volume e a volatilidade estão diretamente relacionados e que têm a mesma origem, a introdução de informação

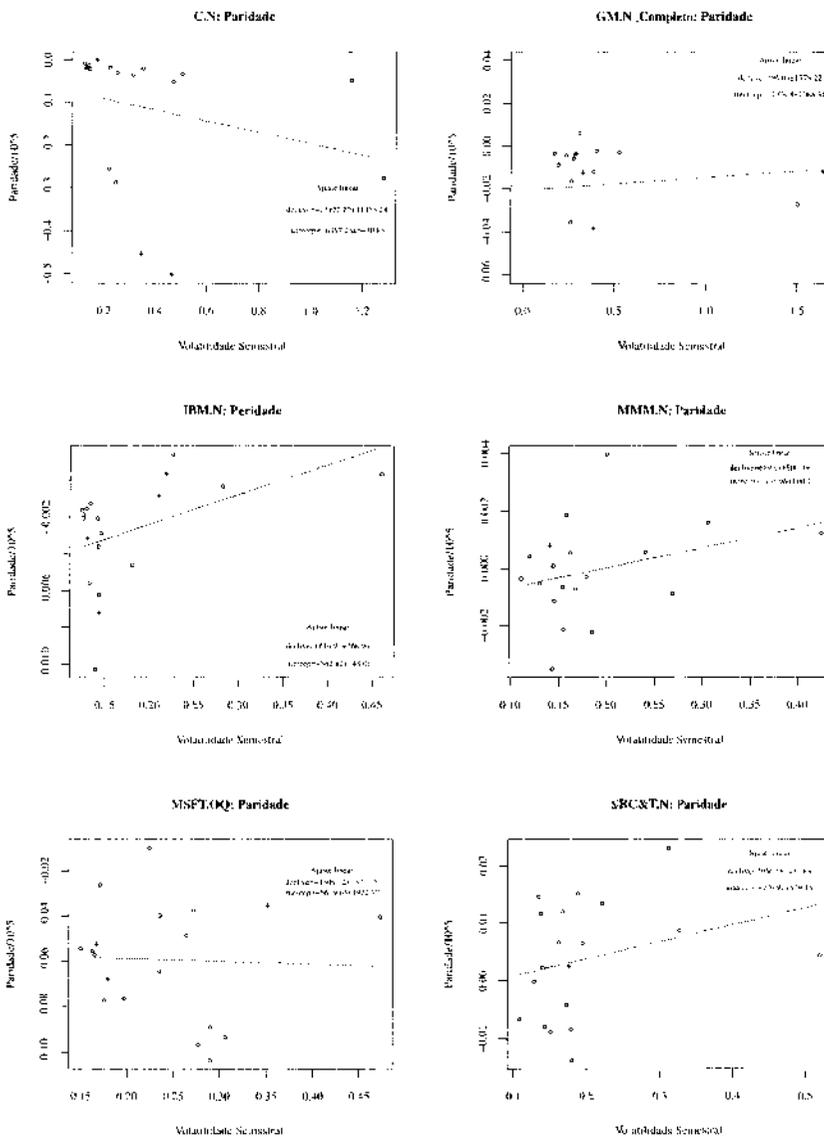


Figura 3.10: Paridade do volume transacionado para cada semestre em função do coeficiente de volatilidade para cada semestre calculado através da Eq. (2.2). A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na volatilidade.

no sistema. Este desacordo pode ser visto para o caso da concavidade com a qual foi possível encontrar uma relação com a média volume ao passo que no caso da volatilidade não existe relação perceptível.

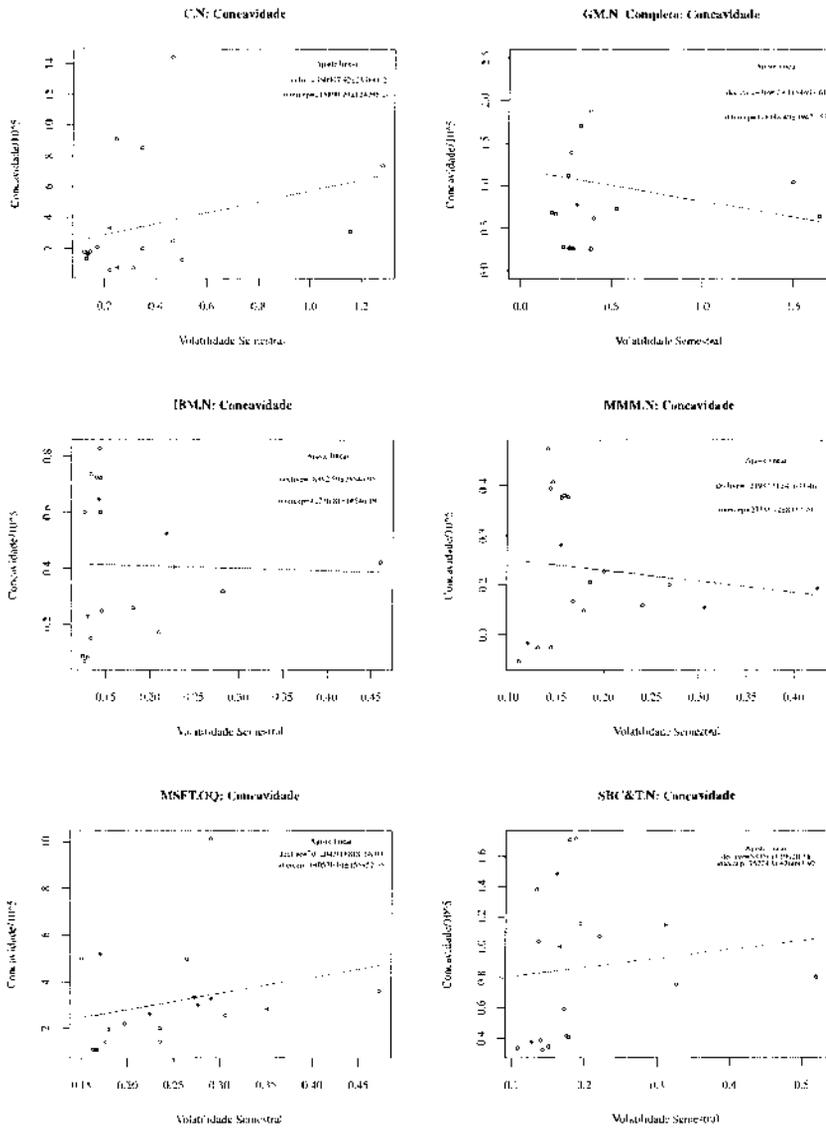


Figura 3.11: Concavidade do volume transacionado para cada semestre em função do coeficiente de volatilidade para cada semestre calculado através da Eq. (2.2). A linha vermelha tracejada corresponde ao ajuste linear. Note que os erros dos declives são da mesma ordem do valor, apontando para independência na volatilidade.

### 3.2 Variância do Volume

A variância do volume negociado representa a dispersão quanto ao seu valor esperado em cada instante. Assim, para que fosse analisada como a dispersão está relacionada à média do volume

foram feitos gráficos, para cada uma das empresas, e para cada semestre, da variância por horário em função da média do volume por horário.

As curvas dos gráficos apresentaram uma concentração maior dos pontos entre os valores 0 e  $4 \times 10^4$  para média da variância, como pode ser verificado na Figura 3.12. Neste intervalo, as curvas são crescentes em todos os semestres apresentando uma dispersão maior nos PP2007 e SS2007, SS2009, PS2010 e PS2011. Olhando para a curva no intervalo integral, temos um padrão de curva crescente porém com uma inclinação pequena, e, ao passar dos semestres, essa curva vai diminuindo de comprimento consequente de uma menor dispersão dos pontos ao longo do semestres.

Com o objetivo de evidenciar as diferenças entre os períodos matutinos e vespertinos sem uma descrição muito acurada da diferença, foram feitos ajustes parabólicos relativo ao período intra-diário total, ao período intra-diário matutino e ao período intra-diário vespertino.

Ao se analisar o gráfico do coeficiente  $t$  do ajuste linear da média da variância ao longo dos semestres,

$$\mathcal{V}_i = t_i * \bar{v}_i^2,$$

onde  $\bar{v}_i^2$  corresponde ao quadrado da média do volume, Figura 3.12, foi possível observar que três semestres apresentaram valores bem altos: SS2004, SS2009 e PS2010, enquanto os valores mais baixos pertencem ao SS2011 e o PS2012.

Uma observação curiosa é que houveram três variações significativas de valores de  $t$  dentro

de um período de uma ano, que aconteceram em 2004 no qual o SS2004 tem um valor de  $t$  que é 3.07 vezes maior do que o valor de  $t$  para o PS2004, em 2008 no qual o SS2008 tem um valor de  $t$  que é 2.44 vezes maior que o do para o PS2008 e no ano de 2011 onde o SS2011 tem um valor de  $t$  correspondente à 0.19 do valor de  $t$  do PS2011. Todos esses anos foram anos conturbados: em 2004 houve um aumento muito grande da inflação em consequência da política adotada pelo governo Bush quando em primeiro de julho de 2004 a taxa FED sobre para 1.25 em razão do rombo nas contas do Estado; em 2008 houve a crise nos Estados Unidos por conta da bolha imobiliária e no Segundo Semestre de 2011 teve-se a crise do limite de dívida dos Estados Unidos de 2011 [para mais informações vide Apêndice B].

Essas mesmas discrepâncias podem ser vistas no gráfico do coeficientes  $m$  e  $v$  do ajuste da média da variância intra-diária da Figura 3.12 porém com proporções um pouco diferentes: o SS2004 tem um valor de  $m$  do ajuste

$$\mathcal{V}_i = m_i * \tilde{v}_i^2,$$

isto é, 3.7 vezes maior que o valor de  $m$  para o PS2004, o SS2008 tem um valor de  $m$  3.14 vezes maior que o do PS2008 e o SS2011 tem um valor de  $m$  correspondente à 0,262 do valor de  $m$  para o PS2011. No segundo semestre de 2004 a FED começa a diminuir a quantidade de dinheiro injetado no sistema bancários e os juros começam a subir, no segundo semestre de 2008 eclode a bolha imobiliária e no final do Primeiro Semestre de 2008 e começo do Segundo Semestre de 2011 os Estados Unidos se encontra na crise de limite de dívida pública.

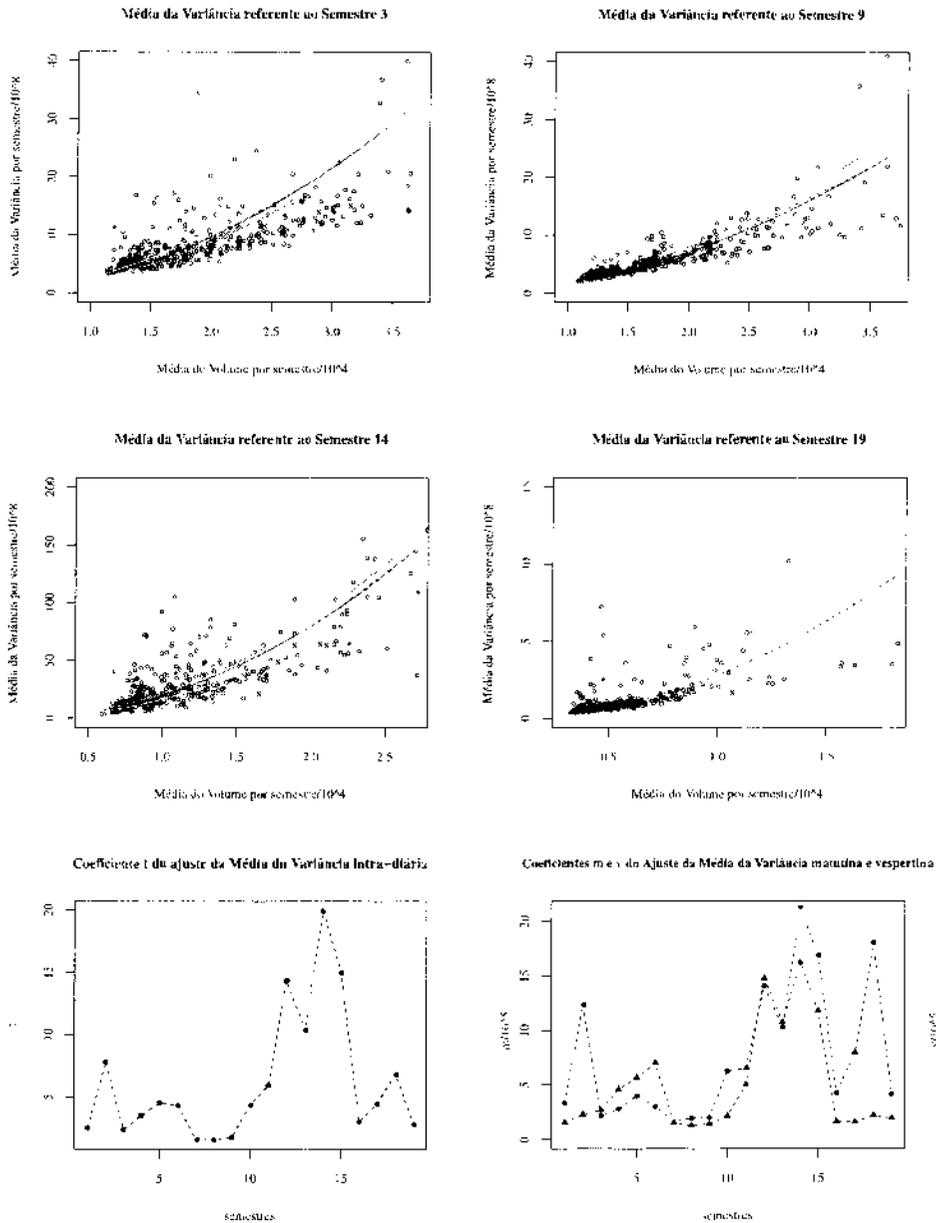


Figura 3.12: Média da Variância em função da Média do volume transacionado por horário intra-diário sobre todas as empresas. A reta tracejada em vermelho corresponde ao ajuste total dos pontos dado por  $\mathcal{V}_i = t_i * \tilde{v}^2_i$ , a reta laranja tracejada representa o ajuste dos pontos pertencente ao período matutino dado por  $\mathcal{V}_i = m_i * \tilde{v}^2_i$ , enquanto a reta tracejada rosa representa o ajuste dos pontos pertencentes ao período vespertino dado por  $\mathcal{V}_i = v_i * \tilde{v}^2_i$

Já para o ajuste dos dados vespertinos representa o ajuste dos pontos pertencentes ao período vespertino dado por

$$\mathcal{V}_i = v_i * \tilde{v}_i^2,$$

que pode ser observado no gráfico do coeficientes  $m$  e  $v$  do ajuste da média da variância intra-diária da Figura 3.12, há uma variação muito grande entre os semestres dos anos de 2009 no qual o valor de  $v$  para o SS2009 é 2.97 vezes maior que o valor de  $v$  para o PS2009 e 2011 no qual o valor de  $v$  para o SS2011 corresponde à 0.14 do valor de  $v$  para o PS2011 . Isto demonstra que os dados matutinos estão em melhor sincronia com os dados intra-diários total do que os dados vespertinos estão com os dados intra-diários total.

Em seguida foram feitos gráficos, para cada semestre, com os valores da Variância para todas as empresas juntas em função do horário para que fosse possível analisar como a dispersão estatística evolui ao longo do tempo, Figura 3.13. O resultado foi um conjunto de pontos que descreveu uma curva em formato U mostrando uma não-estacionariedade.

Os picos das curvas médias, em vermelho, nos gráficos das variâncias do volume referentes ao semestre 1, ao semestre 9 e ao semestre 19 da Figura 3.13 podem ser suavizadas usando a mediana dos valores da variância em função da média, mas, como aqui foi visada a uniformidade de tratamento, usou-se a média dos valores da variância.

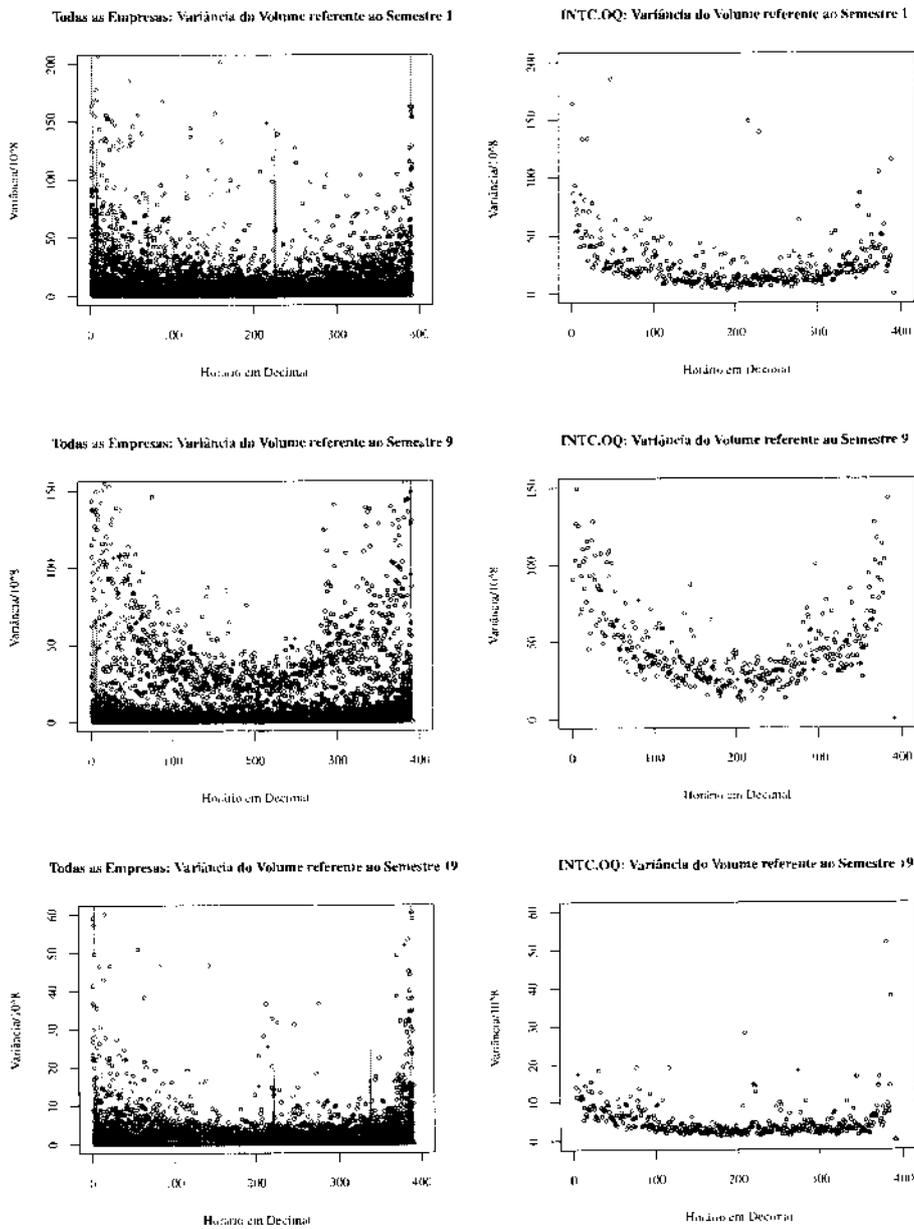


Figura 3.13: Variância do volume transacionado por horário intra-diário de todas as empresas e somente da empresa Intel Corp..

### 3.3 Assimetria

Para se verificar como o grau de afastamento da condição de simetria da distribuição da média do volume se relacionada com a média do volume durante o período de transação intra-diário

foram feitos, para cada semestre, gráficos da assimetria média sobre todas as empresas em função da média do volume por horário como pode ser verificado na Figura 3.14 e pode ser visto para a empresa Citibank Inc. e Microsoft Corp. no Apêndice 3.

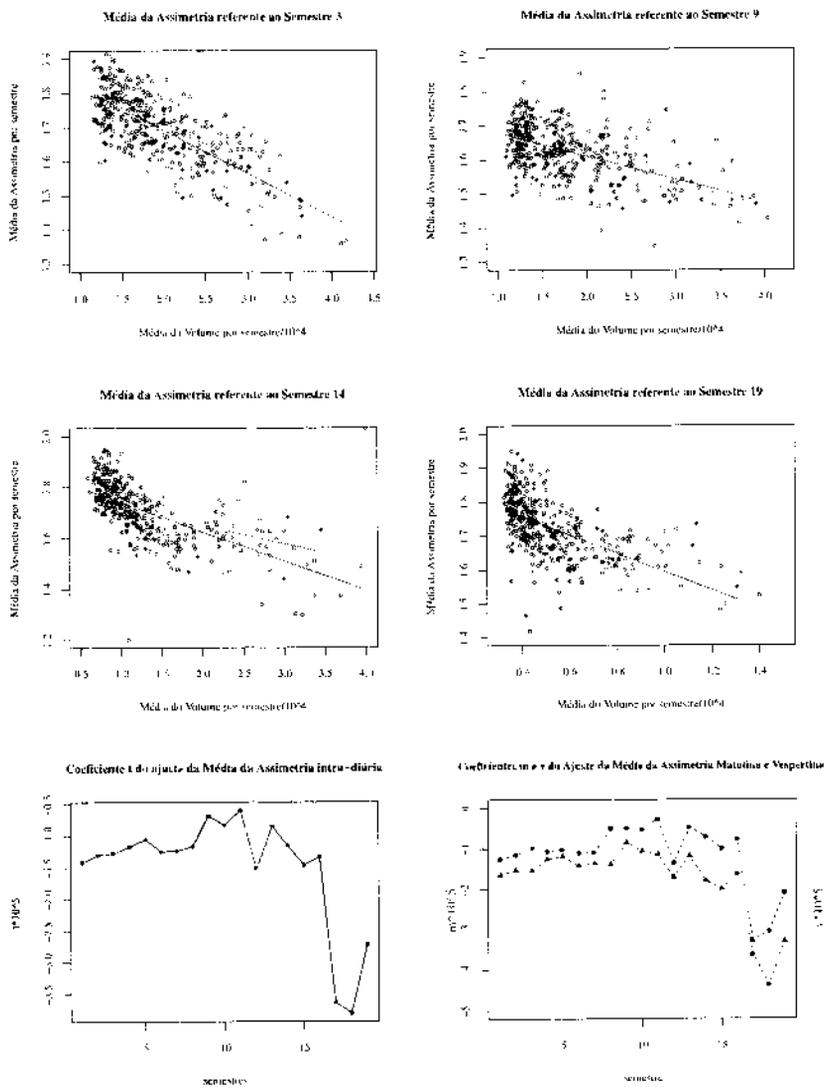


Figura 3.14: Média da Assimetria em função da Média do volume transacionado por horário intra-diário sobre todas as empresas. A reta tracejada em vermelho corresponde ao ajuste total dos pontos dado por  $A_i = t_i \tilde{v}_1 + i_i$ , a reta laranja tracejada representa o ajuste dos pontos pertencente ao período matutino dado por  $A_i = m_i * \tilde{v}_{const} + const'_1$  enquanto a reta tracejada rosa representa o ajuste dos pontos pertencentes ao período vespertino dado por  $A_i = v_i * \tilde{v}_1 + const''_1$ .

Assim como nos gráficos da variância, os pontos concentram-se no intervalo de 0 a  $4.10^4$  para o valor do volume médio. Neste intervalo, a curva é decrescente mas não apresenta um padrão quanto ao ângulo de inclinação. Ao olhar para as curvas no intervalo integral, o resultado foram curvas com um padrão vertical onde os pontos iam ficando mais condensados ao longo dos semestres, assim, o PS2013 apresentou uma curva vertical mais fina e com pontos menos dispersos.

Embora as curvas de assimetria das flutuações de preço tenha um padrão vertical [13], a forma da curva integral não se assemelha em nada como a da assimetria encontrada para o volume.

Ao se analisar os gráficos dos coeficientes de ajuste da Figura 3.14, é possível notar que o maior valor do coeficiente de ajuste  $t$  e  $m$  do ajuste linear total

$$\mathcal{A}_i = t_i \tilde{v}_i + \text{const}_i,$$

$$\mathcal{A}_i = m_i \tilde{v}_i + \text{const}'_i,$$

$$\mathcal{A}_i = v_i \tilde{v}_i + \text{const}''_i,$$

o maior valor pertence ao primeiro semestre de 2008. Porém, para o menor valor tem-se que o menor valor de  $t$  e de  $v$  pertencem ao segundo semestre de 2012 enquanto que o menor valor de  $m$  pertence ao primeiro Semestre de 2011.

Outra característica bem evidente é que a partir do segundo semestre de 2011 há uma queda dos valores bem significativa para os três coeficientes. No segundo semestre de 2011 houve a

crise do limite de dívida dos Estados Unidos de 2011 a qual atingindo todos os países no qual a moeda se baseia no dólar e houve também a crise do euro que, adicionalmente, aumentou o risco de uma crise mundial.

Para os três coeficientes, as variações mais evidentes são: *a)* para os semestres de 2009 com diminuição de 2.55 vezes no valor de  $t$ , diminuição de 4.84 vezes para o valor de  $m$  e diminuição de 1.51 vezes para o valor de  $v$ ; *b)* entre o SS2011 e o PS2012 com diminuição de 2.75 para o valor de  $t$ , diminuição de 4.19 para o valor de  $m$  e diminuição de 2.19 para o valor de  $v$  e *c)* entre o SS2012 e o PS2013 com um aumento de 0.71 vezes para o valor de  $t$ , aumento de 0.68 para o valor de  $m$  e de 0.75 para o valor de  $v$ . Ao analisar essas diferenças pode se constatar que, apesar de as três curvas serem visualmente muito semelhantes, os valores do período vespertino descrevem de uma maneira mais eficaz o período intra-diário.

Em seguida, para se verificar como o grau de afastamento da condição de simetria da distribuição da média do volume evolui ao longo do período de transação intra-diário, foram feitos gráficos para cada semestre da assimetria de todas as empresas em função do tempo, Figura 3.15, nos quais foi possível observar que a partir do SS2007, um pico com amplitude negativa, ou seja, um ponto de mínimo, apareceu por volta do horário 13:10, ou instante 220, e continuou aparecendo em todos os semestres pares dos anos seguintes: SS2008, SS2009, SS2010, SS2011, SS2012 e também no PS2013.

Como, para todos os semestres, temos assimetrias positivas, a distribuição do volume ao

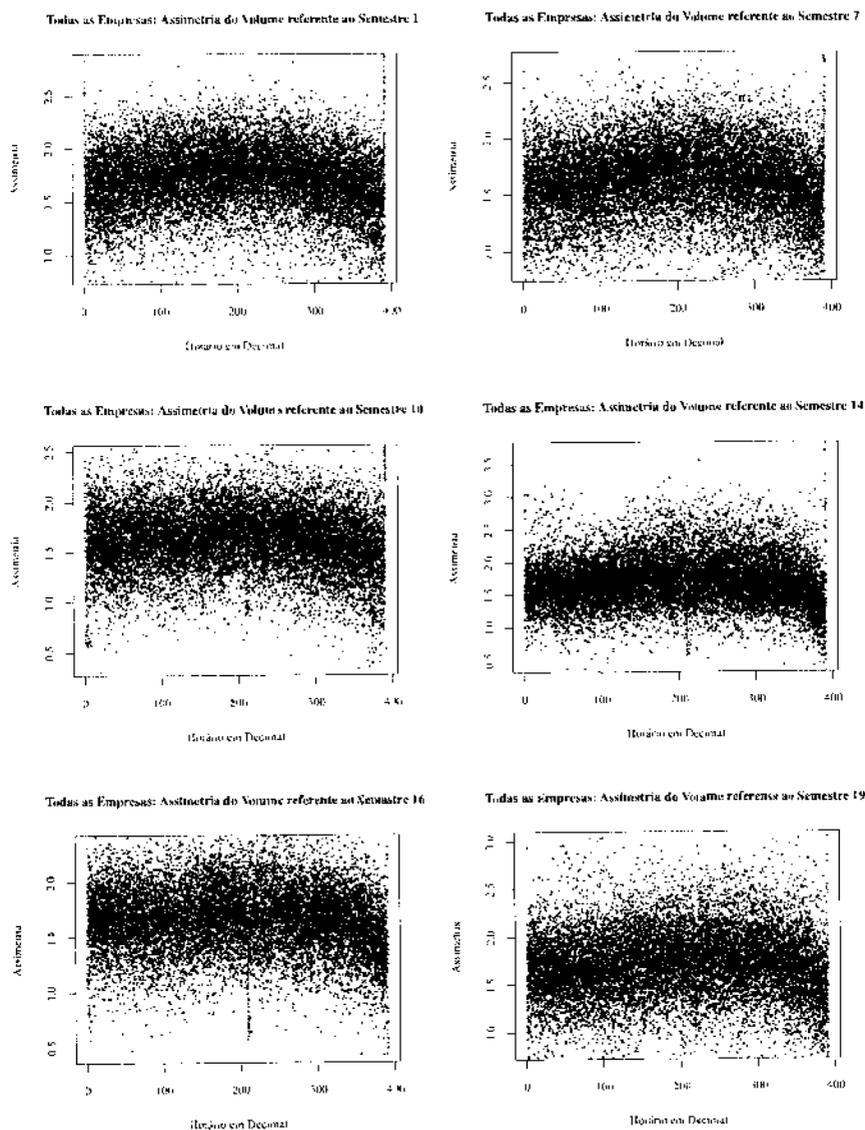


Figura 3.15: Assimetria do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário.

longo do tempo deve ter uma cauda maior para a direita, o que é possível verificar nos gráficos do volume médio por horário em função do horário onde o formato U, tem um ramo maior do lado direito. Isto foi comprovado ao verificar que a média é maior que a mediana.

Observou-se a partir do gráfico do coeficiente  $c$  da Figura 3.16, que o coeficiente  $c_1$  do ajuste

$\mathcal{A}_{ci} = c_i * h_i + cte_i$  no qual  $\mathcal{A}_{ci}$  é a Assimetria de um dado horário  $i$ ,  $h_i$  é um dado horário  $i$  e  $cte_i$  é a constante do ajuste para um dado horário  $i$ , começa a decrescer a partir do PS2007 e diminui bastante nos SS2007 e PS2008 e permanece baixo no SS2008 e no PS2009. Esse é justamente o período no qual os Estado Unidos iniciou e permaneceu em crise.

A curva tracejada em vermelho, representa a média dos pontos que pertencem à uma região dada como região normal. Os pontos considerados para o cálculo desta média foram os do PS2004 até o PS2007, do SS2009 até o PS2011 e o SS2012.

Pode se notar, que pontos não pertencentes a esta região, são pontos correspondentes aos períodos bem conturbados, como o período no qual a bolha imobiliária estourou (2008), assim como o período que logo precede (segundo semestre de 2007 a consciência da existência da bolha já vai além do mundo financeiro), e o período logo após ao estouro (primeiro semestre de 2009). O segundo semestre de 2011 também é turbulento pois, até os Estados Unidos decidirem aumentar o teto da dívida, existiam rumores de calote.

E a partir do gráfico do coeficiente  $d$  da Figura 3.16 foi possível observar uma flutuação muito grande dos valores do coeficiente  $d_i$  do ajuste  $\mathcal{A}_{di} = d_i * h_i + cte'_i$  entre o PS2007 e o SS2007, entre o SS2007 e o PS2008, entre o PS2009 e o SS2009 e entre o SS2012 e o PS2013.

Quando se compara as duas curvas do coeficientes de ajuste  $c_i$  e  $d_i$  como pode ser visto na Figura 3.16, tem-se uma semelhança "invertida" entre o comportamento das curvas a partir do Segundo Semestre de 2009 até o Primeiro Semestre de 2012: a forma é bem similar porém

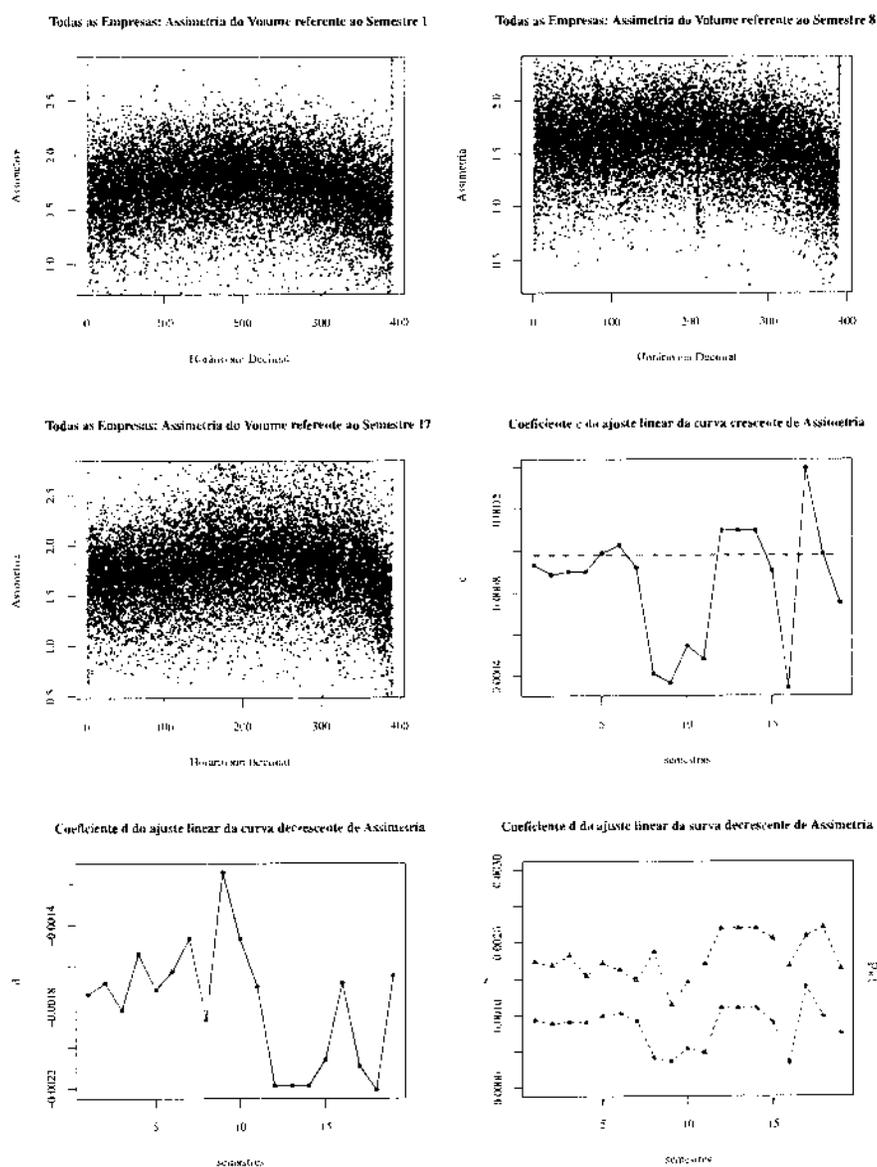


Figura 3.16: Assimetria do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Assimetria sobre todas as empresas para cada horário. As linhas azuis corresponde ao ajuste  $A_{c,i} = c_i * h_i + cte_i$  e  $A_{d,i} = d_i * h_i + cte'_i$ , sendo  $c_i$  e  $cte_i$  os coeficientes de ajuste da reta crescente e  $d_i$  e  $cte'_i$  os coeficiente de ajuste da reta decrescente para um dado semestre  $i$ . A linha vermelha tracejada define a região de normalidade. A curva azul clara representa a curva do coeficiente de ajuste  $c$  e a curva azul escura representa a curva do coeficiente de ajuste  $d$ .

enquanto o coeficiente  $c$  cresce, o coeficiente  $d$  decrescente, quando o coeficiente  $c$  decresce o coeficiente  $d$  cresce. O ponto referente ao Primeiro Semestre de 2012, descontinua o compor-

tamento das duas curvas, o que pode ser reflexo do abalo na economia no SS2011 decorrente da “crise do euro”, da “crise do dólar” e do aumento do risco de uma crise mundial. Outro ponto assim, ocorre para o SS2007, que foi o período onde os economistas passaram a ter uma consciência maior de que se estava a acontecer: o estouro da bolha.

### 3.4 Curtose

Para analisar como o grau de aguçamento da distribuição da Média do volume se relaciona com a própria Média do volume durante o período de transação intra-diário, para cada semestre, foram feitos gráficos da curtose média de todas as empresas em função da média do volume por horário como pode ser visto na Figura 3.17 e da curtose para cada empresa em função da Média do volume por horário como pode ser visto como exemplo nas figuras no Apêndice C.

Teve-se um comportamento semelhante ao comportamento das curva dos gráficos da assimetria em função da média do volume por horário: ao longo dos semestres, a curva vertical foi ficando mais condensada e fina, apresentando menos dispersão entre os pontos. Porém, para o caso da curtose, todas as curvas de todos os semestres apresentaram uma “ramificação”: a curva começa com uma curtose alta, vai diminuindo ficando num intervalo entre 6 e 7 e por volta do horário 12:40 ela começa a apresentar valores mais altos de curtose num intervalo entre 7 e 8, atingindo seus valores máximos por volta do horário 13:10, voltando a diminuir por volta do horário 13:30.

Ao observar as curvas de ajustes dos gráficos da Figura 3.17, é possível verificar uma diferença de comportamento bem evidente entre os pontos do período matutino e os pontos do período vespertino. Essa diferença está presente para todos os 19 semestres analisados.

Ao analisar os gráficos dos coeficientes de ajuste da Figura 3.17 foi possível observar que no gráfico dos coeficientes  $a2^m$  do ajuste polinomial da curtose matutina  $C_i^m = a2_i^m * \tilde{v}_i^2 + a1_i^m * \tilde{v}_i + a0_i^m$  e  $a2^v$  do ajuste polinomial da curtose vespertina  $C_i^v = a2_i^v * \tilde{v}_i^2 + a1_i^v * \tilde{v}_i + a0_i^v$  tem-se uma variação muito grande do comportamento das curvas a partir do SS2006 no qual a curva para o ajuste vespertino passa a apresentar uma certa linearidade aproximando-se a uma linha reta enquanto o a curva para o ajuste vespertino picos de máximos e mínimos bem proeminentes. Além disso, pode se notar que só existe um valor de  $a2^m$  que é menor que  $a2^v$ , pertencente ao SS2012, turbulento em conta da crise das dividas publicas vivida pela Europa e o pelos Estados Unidos.

No gráfico do coeficiente  $a1$  da Figura 3.17 é possível verificar mais uma vez um comportamento bem diferente para o Segundo Semestre de 2012, que além de passar a ser positivo, o valor tomado é bem maior que os demais. Outra característica que pode ser observada é que entre o PS2004 até o PS2006, as curvas de  $a2^m$  e  $a2^v$  apresentam um comportamento muito similar, porém os valores de  $a2^v$  são menores que os valores de  $a2^m$ . O gráfico dos coeficientes de ajuste da curtose matutina  $a1^m$  e vespertina  $a1^v$  expõe curvas de formato totalmente distintos porém as duas tratam-se de curvas, que em média, são crescentes ao longo dos semestre.

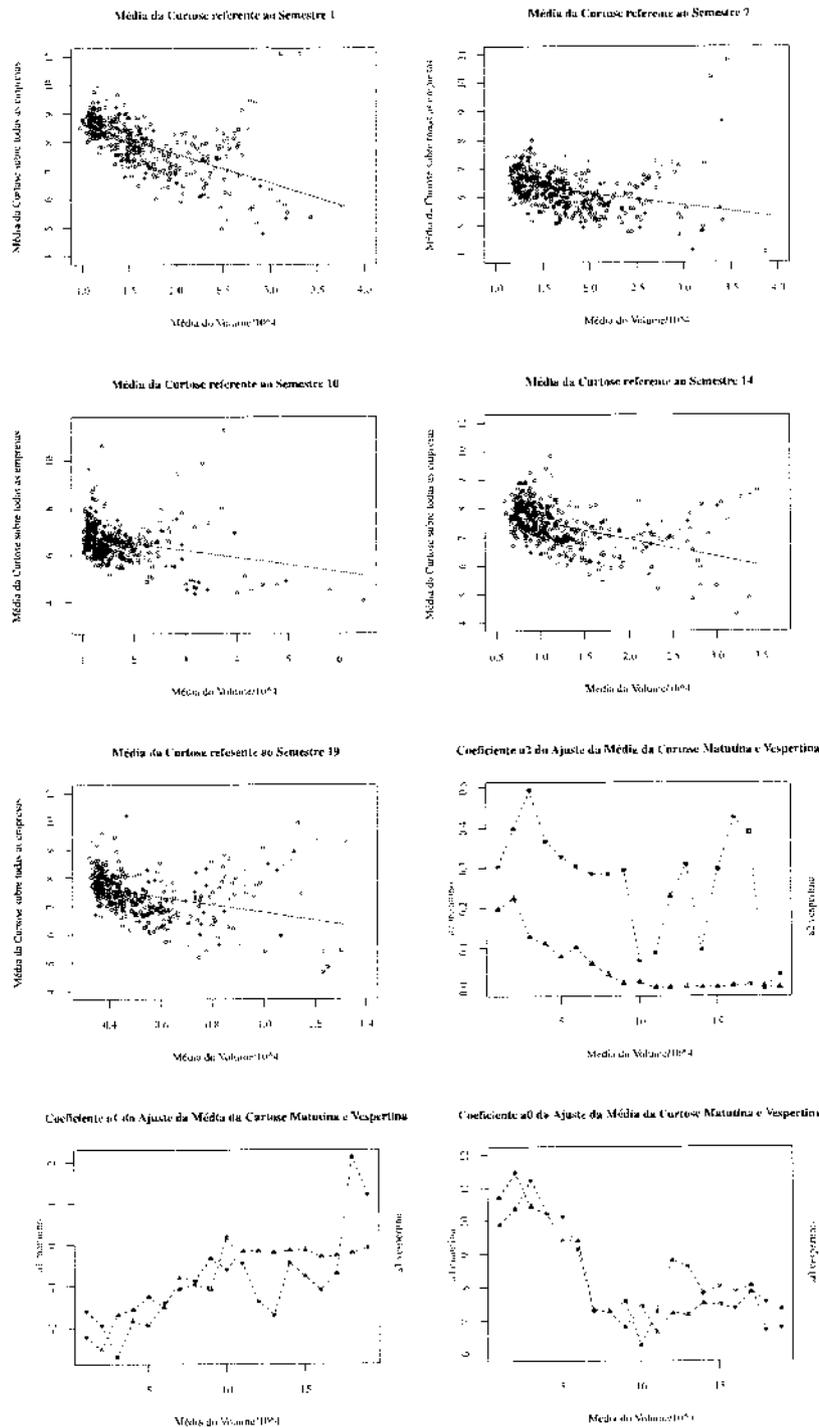


Figura 3.17: Média da curtose em função da média do volume transacionado. A reta tracejada em vermelho é o ajuste linear de todos os pontos  $C_i = t_i * \tilde{v}_i + const_i$ , a reta laranja tracejada(rosa) o ajuste dos pontos do período matutino(vespertino):  $C_{i(m,vc)} = a2_{i(m,vc)}^* \tilde{v}_i^2 + a1_{i(m,vc)}^* \tilde{v}_i + a0_{i(m,vc)}$ .

Além disso, tem-se seis semestres para os quais o coeficiente de ajuste matutino é maior que o vespertino: PS2004 e SS2004, SS2006, SS2008, SS2012 e PS2013; enquanto para os outros o coeficiente de ajuste vespertino é maior que o matutino. Ao se analisar o gráfico dos coeficientes de ajuste  $a0^m$  para a curtose matutina e  $a0^v$  para a curtose vespertina é possível notar uma semelhança entre o comportamento das curvas entre o período do PS2004 até SS2007.

No PS2008, no SS2008 e no PS2009, as curvas apresentam uma total inversão de uma quanto a outra, lembrando-se que se trata do período pré, de e pós estouro da bola imobiliária dos Estados Unidos. Depois desse período há um misto de comportamento semelhante e antagônico entre as curvas.

Para que esta "ramificação", ou melhor, a não injetividade da curva da curtose em função da Média fosse melhor analisada, foram feitos gráficos para cada empresa separada da curtose em função do volume médio e estas não apresentaram tal "ramificação" e nem o comportamento descrito acima.

Então, observou-se os gráficos, para cada semestre, da curtose média sobre todas as empresas em função do horário e a média do volume médio sobre todas as empresas em função do horário. Como pode ser visto na Figura 3.18, o primeiro trata-se de uma curva quase que antissimétrica enquanto o segundo é quase simétrica, fazendo com que, ao se fazer um em função do outro de maneira a parametrizar-se, surge a "ramificação".

Em seguida, para analisar como o grau de aguçamento da distribuição da média do volume

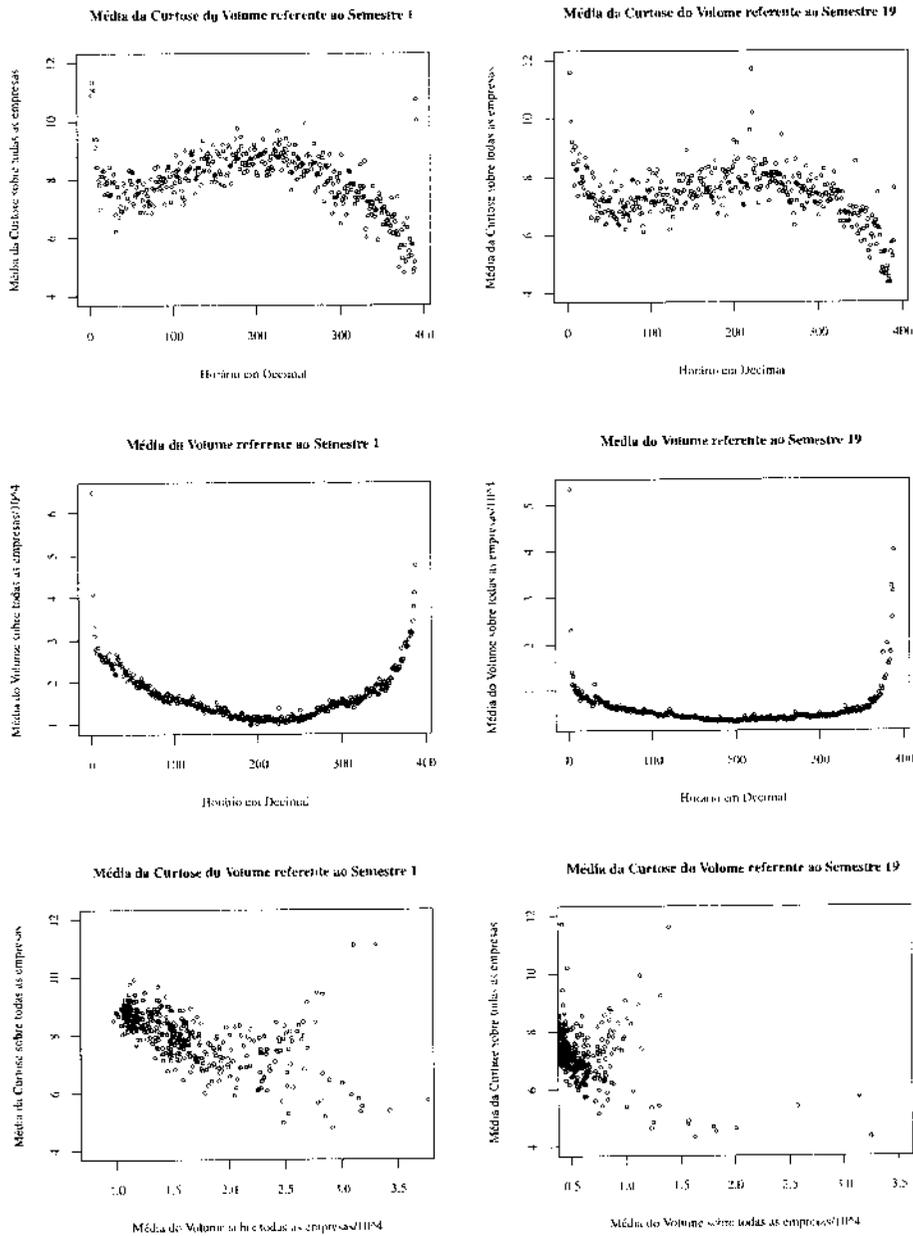


Figura 3.18: Média da Curtose do volume transacionado sobre todas as empresas em função da Média do Volume transacionado (calculadas para o mesmo horário).

evolui ao longo do período intra-diário, foram feitos gráficos para cada semestre da curtose de todas as empresas em função do tempo, Figura 3.19, e o resultado foi uma forma padrão apresentada por todas as empresas, durante todo os semestres: um formato de  $\sim$  com uma

grande abertura e, se, desconsiderados os pontos finais que parecem ser outliers, a curva pode ser ajustada por um polinômio de quinta ordem.

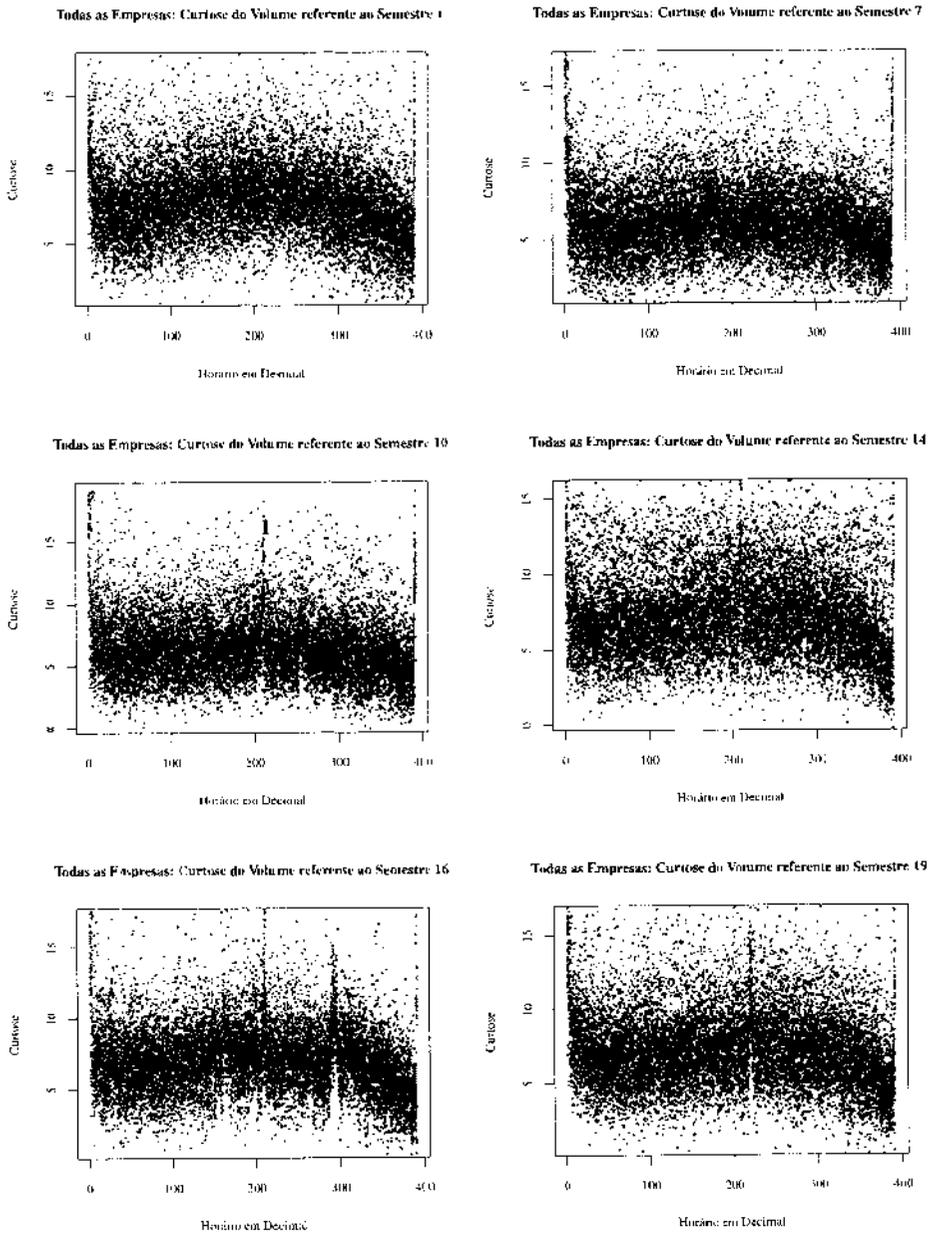


Figura 3.19: Curtose do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário.

Para o caso descrito aqui, a curtose do volume com seu padrão  $\sim$ , é possível verificar que

a curva apresenta um máximo em sua abertura e fechamento e outro máximo bem menos pronunciado por volta das 13:10 horas, indicando que esses períodos são dominados por efeitos idiossincráticos ou efeitos setoriais específicos com pequenas surpresas.

Para os semestres pares de cada ano, foi possível observar a formação de um pico (pico do meio do  $\sim$ ) em torno das 13 horas e 10 minutos, ou seja, no instante 220, a partir do SS2006 e este pico teve sua amplitude aumentada no SS2008 e continuou presente em todos os semestres pares de cada ano seguinte: SS2008, SS2009, SS2010, SS2011, SS2012 e também no PS2013.

Este pico concorre com o chamado “efeito almoço” [2, 33, 34, 8] caracterizado por uma alta volatilidade e tempos curtos de negociações. Mostramos então que para além alteração na volatilidade existem um aumento da não-Gaussianidade, o que naturalmente corresponde a uma alteração funcional da curva. Além disso, tem-se o fato de que o segundo semestre de cada ano apresenta um ritmo de atividade maior, e por este motivo, é mais aparente o padrão sazonal.

Este comportamento sazonal é visto como uma anomalia em decorrência da ineficiência do mercado financeiro por operar em sincronia com as atividades humanas [35], porém até agora não há uma compreensão clara e por isto continua sendo debatida entre economistas, econométricos e estudiosos da economia [36]. Entendemos no entanto que tal sazonalidade está acima de tudo relacionada com a planificação da maioria das empresas que costumam ter segundos semestres mais fortes. Tal é verificável por exemplo com o acompanhamento da arrecadação fiscal que em países economicamente evoluídos apresenta variações da ordem de

2% do Produto Interno Bruto.

Nos períodos onde não há picos; pode se dizer períodos de grandes oscilações, os quais começa de manhã logo após a abertura do mercado, diminui perto da hora do almoço, volta a crescer durante a tarde e depois diminui novamente até o fechamento; a componente idios-sincrática da dinâmica torna-se subdominante fazendo com que efeitos externos e secundários prevaleçam.

Ao observar o gráfico do coeficiente  $ai$  na Figura 3.20, percebeu-se que o coeficiente de ajuste inicial  $ai$ , dos ajustes em lei de potência  $C_i = h_i^{-ai} + constante$  do período que começa às 09:32 horas e termina às 10:23 horas (período considerado de relaxação pós choque), sendo  $C_i$  a curtose de um dado horário  $i$  e  $h_i$  um dado horário  $i$ , do PS 2007, SS2007 e PS2008 apresentam um valor bem maior quando considerado o período que começa no PS2004 e termina no SS2010.

Este período é justamente o período de crise vivida pelos Estados Unidos. A partir do PS2011 até o PS2013 obteve-se uma curva crescente. Quando comparado aos coeficientes de decaimento para a média do volume, Figura 3.3, é possível notar que a magnitude do grau de relaxação da Média do Volume transacionado é maior que a do grau de relaxação da média da curtose.

Os coeficientes de ajuste final  $af_i$  do ajuste

$$C_i = C_{\max_i} - (h_i - h_{\max_i})^{af_i},$$

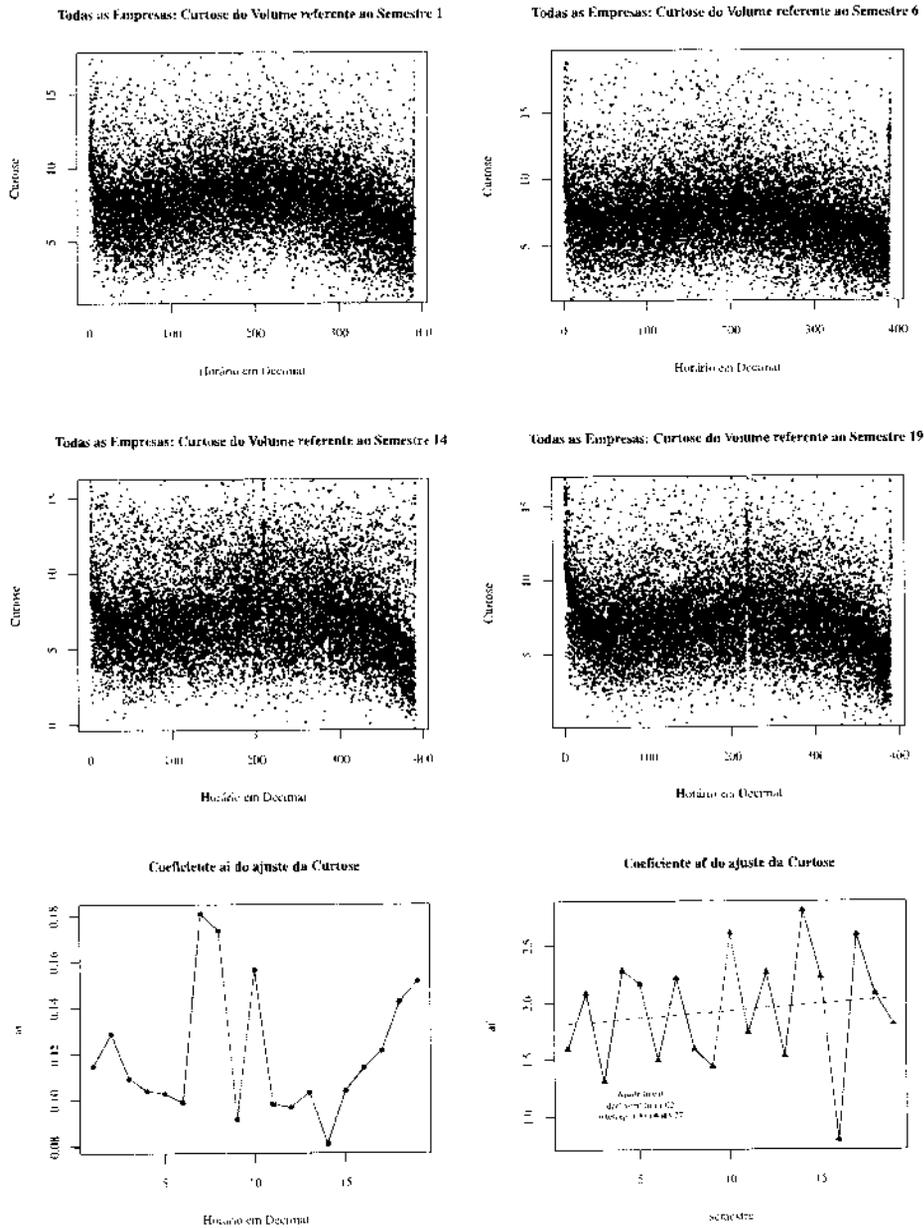


Figura 3.20: Gráfico da Curtose do volume transacionado por horário intra-diário de todas as empresas. A linha vermelha corresponde a média da Curtose sobre todas as empresas para cada horário. A linha azul corresponde ao ajuste  $C_i = h_i^{-a_i} + constante$  e a linha azul escura corresponde ao ajuste  $C_i = C_{max_i} - (h_j - h_{max_i})^{a_i}$  para cada semestre  $i$ .

onde  $C_{max_i}$  é o valor máximo da curtose no intervalo de 270 à 290 do horário decimal e  $h_{max_i}$  é o horário correspondente ao valor de  $C_{max_i}$ , como pode ser observado no gráfico do coeficiente

*af* na Figura 3.20, apresenta uma característica de curva com comportamento crescente em sua totalidade, porém o ajuste linear apresenta um erro  $10^2$  vezes maior do que o valor do ajuste para o declive e  $10^4$  vezes maior do que o valor do ponto de intercepção da ordenada de forma que não se pode classifica-la então como uma curva crescente. Adicionalmente, existe um ponto mínimo bem proeminente pertencente ao SS2011. No segundo semestre de 2011 o mundo se vê com um alto risco de entrar numa crise global, enquanto que os EUA e países da Europa já se encontravam em crise econômica.

# Capítulo 4

## Sazonalidades coletivas

Com o objetivo de analisar como a dinâmica de uma empresa  $i$  depende linearmente da dinâmica de uma outra empresa  $j$ , foram montadas matrizes de dimensão 30 x 30 onde cada entrada contém o valor de correlação entre duas empresas

$$c^{ij}(t; l) \equiv \frac{\frac{1}{D} \sum_{d=d_s}^{d_e} v_i(d, t; l)v_j(d, t; l) - \tilde{v}_i(t; l)\tilde{v}_j(t; l)}{\tilde{v}_i^2(t; l)\tilde{v}_j^2(t; l)}. \quad (4.1)$$

As empresas foram ordenadas em ordem alfabética, assim, a empresa AA corresponde à linha/coluna 1, a empresa AIG corresponde a linha/coluna 2 sucessivamente até a empresa XOM que corresponde à linha/coluna 30.

A matriz de correlação  $C^{(v)}(t)$  é definida simétrica  $c^{ij}(t; l) = c^{ji}(t; l)$  e nos seus elementos da diagonal  $i = j$  têm as auto-correlações de cada empresa resultando em  $\text{Tr} C^{(v)}(t; l) = 30$  sendo por isso uma matriz definida positiva.

Porém, existem algumas dificuldades em quantificar correlações entre duas empresas pois,

diferentemente da maioria dos sistemas físicos, não existe uma lei de primeiros princípios para calcular a “força de interação” entre duas empresas  $i$  e  $j$ , sendo desconhecida a natureza exata da interação direta ou indireta entre duas empresas.

## 4.1 Análise da Componente Principal

A Análise da Componente Principal (PCA) teve como objetivo inicial a redução da dimensão de um conjunto de dados para as direções onde existe uma maior variância, ou seja, descobrir os autovetores correspondentes aos autovalores mais significativos de uma matriz de covariância, e projetar este conjunto de dados nestas direções. Esta análise visa transformar as variáveis originais em um novo conjunto de variáveis que são uma combinação linear de variáveis originais, que sejam não-correlacionadas entre si e que sejam ordenadas de acordo com a magnitude da variação das variáveis originais que elas explicam.

Existem dois métodos gerais para a implementação da PCA quando não há valores faltantes: a decomposição espectral e a decomposição do valor singular.

Aqui foi usado o método de análise de decomposição espectral pois esta analisa a covariância e correlação entre variáveis enquanto o método de decomposição do valor singular analisa a covariância e correlação entre amostras.

O método da decomposição espectral para matrizes quadradas, que são os do tipos das matrizes de correlação tratadas aqui, é descrito como segue.

Para uma dada matriz simétrica e real  $A$  de dimensão  $m \times m$ , tem-se sua decomposição espectral na forma de

$$A = U\Lambda U^T \quad (4.2)$$

sendo  $U$  uma matriz ortogonal de vetores unitários tais que  $UU^T = I$  ou  $\sum_k U_{ki}U_{kj} = \delta_{ij}$  e  $\Lambda$  uma matriz diagonal.

As colunas de  $U$  correspondem aos autovetores de  $A$  e a matriz diagonal  $\Lambda$  tem como seus elementos os autovalores de  $A$ . Se  $A$  é definida positiva, os autovalores serão todos positivos. Se multiplicarmos a Eq. (4.2) por  $U$  tem-se que:

$$AU = U\Lambda U^T U = U\Lambda \quad (4.3)$$

que permite escrever a equação de autovalores definindo  $u_i$  como as colunas de autovetores de  $A$  e  $\Lambda_{ii} = \lambda_i$  os autovalores de  $A$ :

$$Au_i = \lambda_i u_i. \quad (4.4)$$

A partir das informações acima é possível descrever como este método foi usado nas matrizes de correlação aqui calculadas:

Com os dados em mão, foquei-me primeiramente nos cálculos das matrizes de correlação para cada horário do período intra-diário de negociação. As medidas foram realizadas sobre os dias de cada semestre.

Montada a matriz de correlação de dimensão  $30 \times 30$  para cada hora fixa, foi aplicado o método de Rebonato [29], o mesmo método mencionado anteriormente e que pode ser visto com mais detalhes no Apêndice A, para a correção das mesmas pois, dada a finitude das séries é natural que surjam autovalores negativos e estes não fazem sentido pois a variância é uma grandeza não-negativa. Depois de corrigida, foram calculados os autovalores e seus autovetores correspondentes.

Os autovetores são unitários, ou seja, tem comprimento 1, são ortogonais entre si e representam os padrões de comportamento do conjunto estudado. O autovetor correspondente ao autovalor mais significativo descreve o padrão mais importante do comportamento da correlação entre as empresas para uma dada hora fixa mostrando-se um modo de mercado, como pode ser visto adiante.

#### **4.1.1 Matrizes Completas e Matrizes Condicionadas**

Com o objetivo de se avaliar a quantidade mínima de volume transacionado que conseguiria descrever a dinâmica do processo assim como qual a grandeza de volume necessária para isso, foram calculadas matrizes de correlação condicionadas a dois valores de volume de corte.

Os volumes de corte foram estipulados para cada empresa e para cada horário da seguinte forma: dados os conjuntos de volumes  $v$  cada um pertencente a um horário  $t$  para uma dada empresa  $e$ , cada conjunto deste era colocado em ordem decrescente e então 2 volumes de corte

eram estabelecidos sendo o primeiro restringia os dados a 40% do maior valor  $v_{C_{et}}^{40\%}$  (chamadas de Matrizes de Correlação 40%) e o segundo restringia os volumes aos 20% maiores valores  $v_{C_{et}}^{20\%}$  (chamadas de Matrizes de Correlação 20%).

Para que essas matrizes de correlação pudessem ser analisadas visualmente, para cada matriz de correlação e de correlação condicionada foram montados mapas de calor onde a graduação de cores representava a magnitude do valor da correlação entre um par de empresas.

A partir da Figura 4.1 é possível observar que, no começo do período intra-diário, as Matrizes de Correlação apresentam uma coloração predominantemente castanha e rosa indicando que há mais valores de correlação positiva, ou seja, as empresas estão mais correlacionadas entre si.

Com exceção dos horários nos quais as matrizes de correlação apresentam uma mudança bem evidente de comportamento (que será discutido a seguir), ao longo do período intra-diário as Matrizes de Correlação vão se tornando mais amareladas indicando que os valores de correlação são menores em valor absoluto, em sua maioria, bem dentro do intervalo  $-0.5$  e  $0.5$ , de tal forma a ter-se empresas não muito correlacionadas e não muito anti-correlacionadas.

Quando se aproxima do final do período intra-diário, as Matrizes de Correlação passam a ter uma coloração mais amarelada e esverdeada indicando uma maior quantidade de não correlação e de anti-correlações entre as empresas. Com isso é possível argumentar que as empresas, em sua totalidade, começam o período intra-diário mais correlacionadas e terminam o período intra-diário mais anti-correlacionadas.

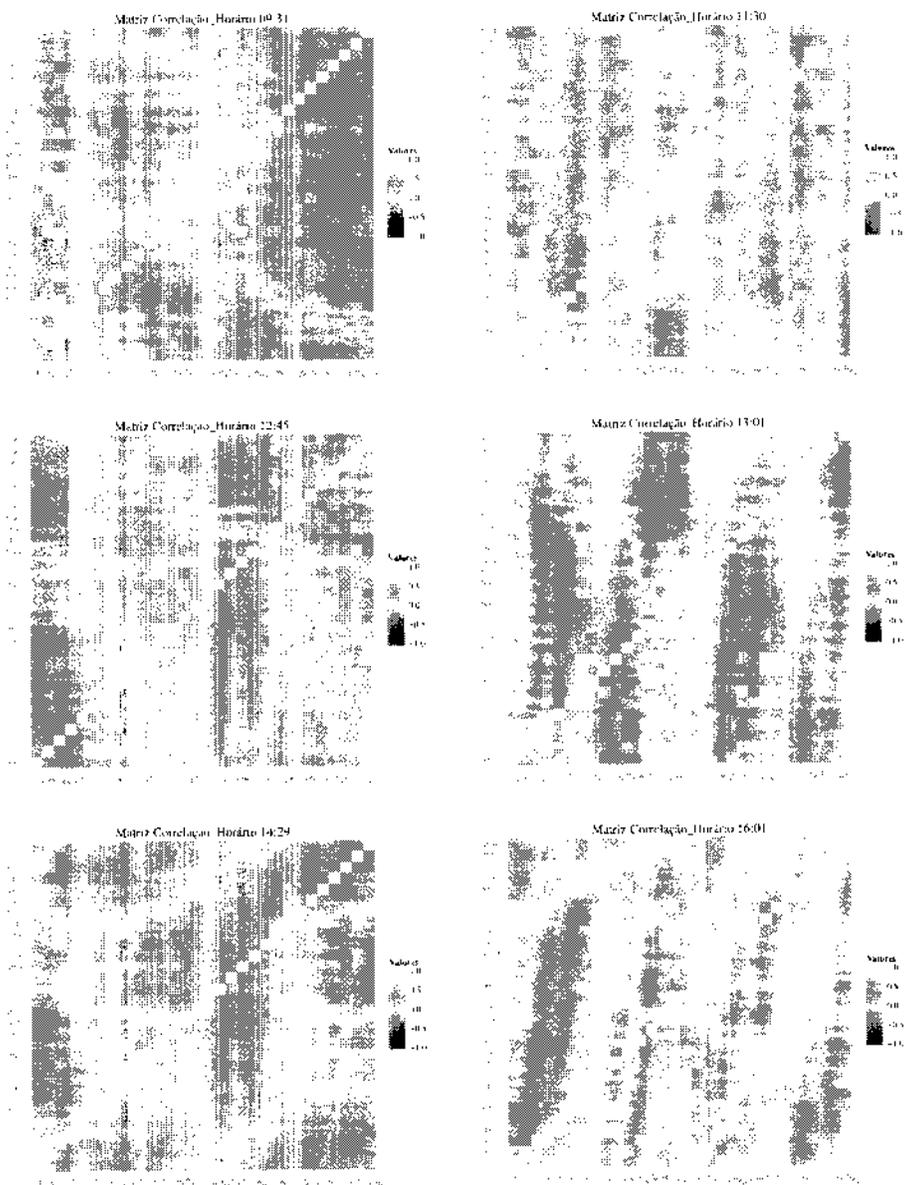


Figura 4.1: Mapa de Calor das Matrizes de Correlação para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:29 horas e 16:01 horas.

A partir da Figura 4.2 pode se observar que as Matrizes de Correlação 40% apresentam uma coloração de fundo que transita entre o verde claro e o amarelo, mostrando que existem muitas empresas que estão medianamente anti-correlacionadas, e há pontos em destaques na cor rosa e castanho escuro expondo que existem empresas que estão fortemente correlacionadas entre

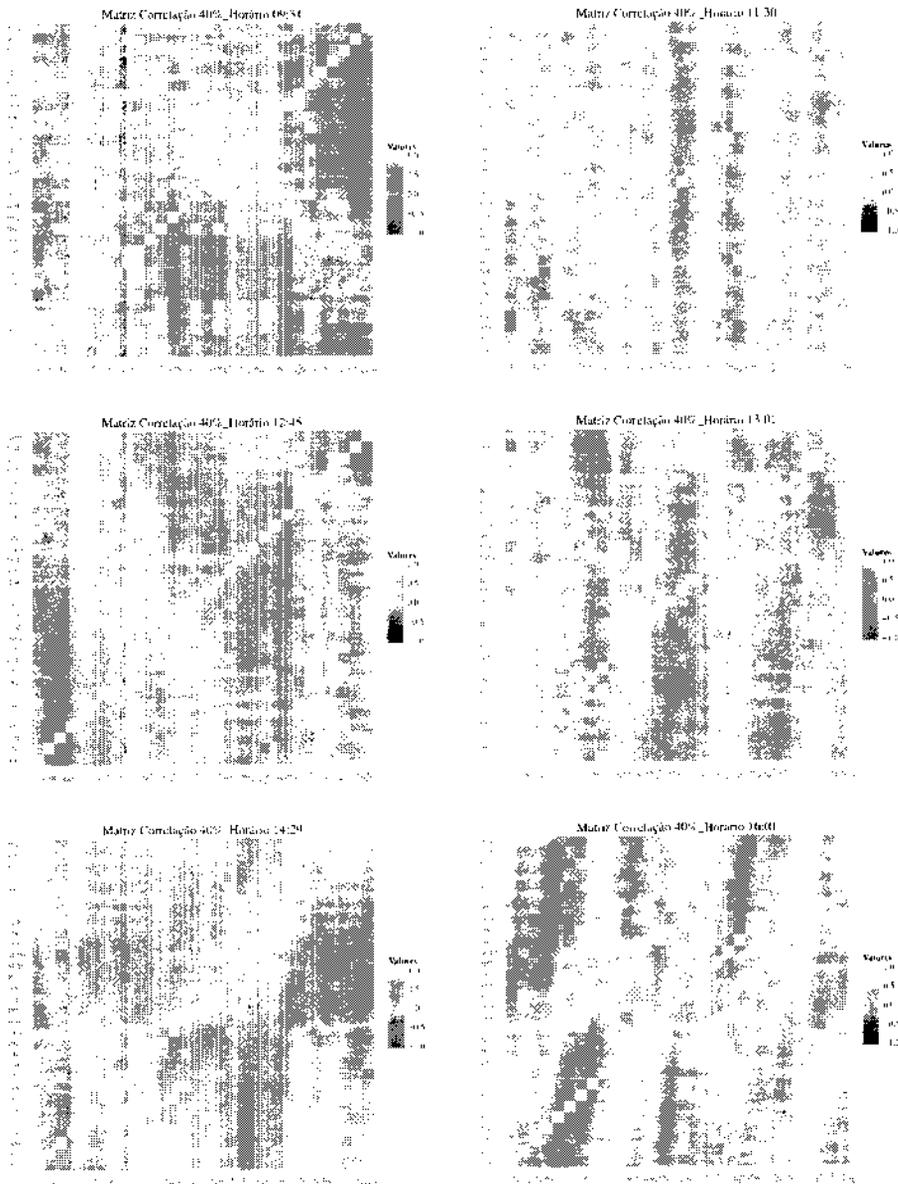


Figura 4.2: Mapa de Calor das Matrizes de Correlação 40% para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:00 horas e 16:01 horas.

si. Com exceção dos horários nos quais as matrizes de correlação apresentam uma mudança bem evidente de comportamento, ao longo do período intra-diário a Matriz de Correlação 40% vai tomando uma coloração entre o castanho claro e o verde claro indicando que as empresas passam a ter um menor valor de correlação em valor absoluto. No final do período intra-diário,

a Matriz de Correlação 40% torna-se mais esverdeadas mas apresenta pontos castanhos indicando uma maior quantidade de empresas não correlacionadas e de empresas que apresentam uma anti-correlação entre  $-0.1$  e  $-0.5$  e uma menor quantidade de empresas que apresentam correlação entre 0 e 0.5.

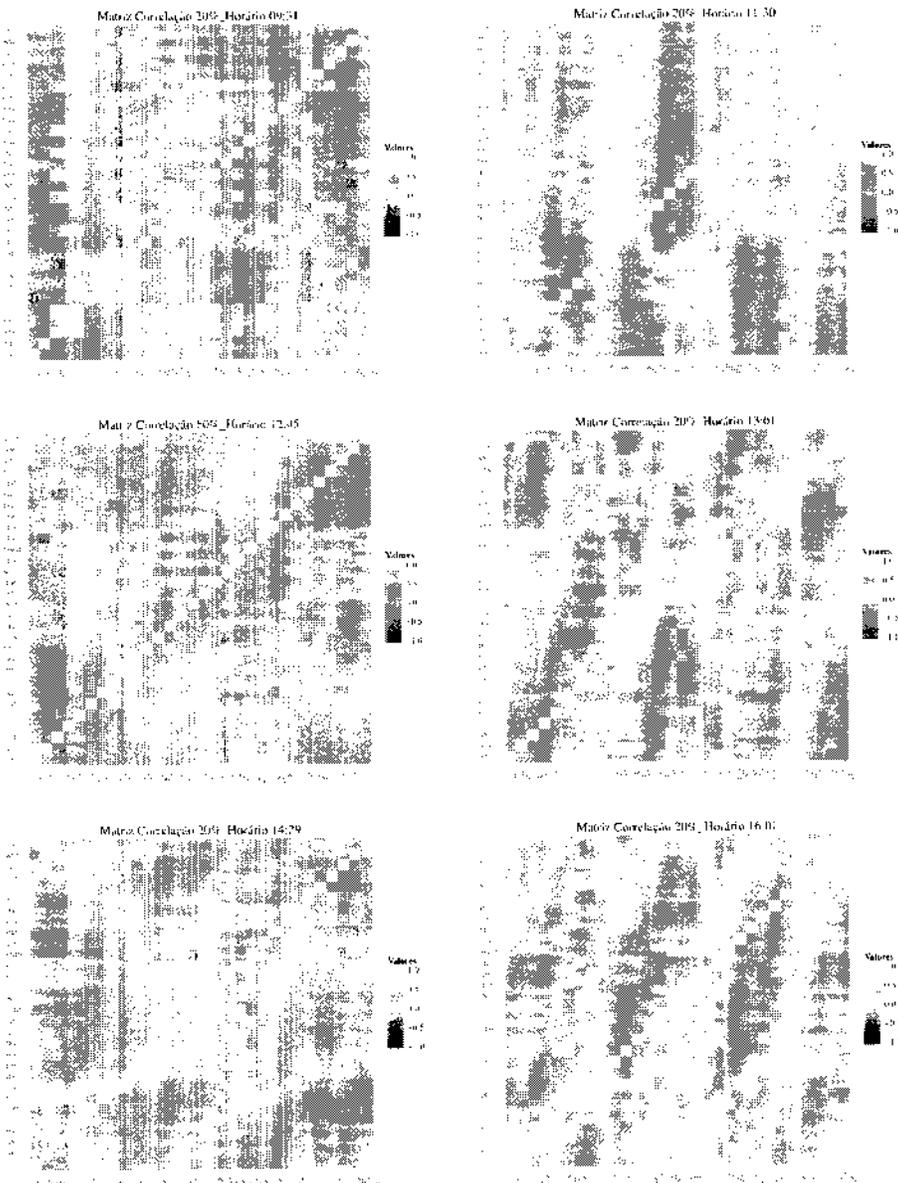


Figura 4.3: Mapa de Calor das Matrizes de Correlação 20% para os horários 09:31 horas, 11:30 horas, 12:45 horas, 13:01 horas, 14:00 horas e 16:01 horas.

Para as Matrizes de Correlação 20%, como pode ser visto na Figura 4.3, tem-se uma mistura de cores porém tem mais valores pertencentes às colorações castanho claro, amarelo e verde claro indicando que há mais empresas que apresentam uma correlação entre 0 e 0.5 e uma anti-correlação entre  $-0.5$  e 0 mas também existem mais empresas muito correlacionadas do que empresa muito anti-correlacionadas.

Então, pode se tomar a Matriz de Correlação 20% do início do período intra-diário como uma matriz de correlação positiva de modo que as empresas começam mais correlacionadas entre si. Com exceção dos horários nos quais as matrizes de correlação apresentam uma mudança bem evidente de comportamento, ao longo do período intra-diário é possível observar que as Matrizes de Correlação 20% vão se tornando mais amareladas e esverdeadas indicando uma maior quantidade de empresas anti-correlacionadas e de empresas não-correlacionadas .

Existem poucas empresas correlacionadas entre si com o valor de correlação predominantemente entre 0 e 0.5. No final do período, esta coloração amarelada e esverdeada persiste mas a quantidade de empresas correlacionadas entre si diminui porém o valor da correlação aumenta, pertencendo ao intervalo entre 0.25 e 0.75.

Outro fato interessante é que, ao olhar os mapas de calor das Matrizes de Correlação, das Matrizes de Correlação 40% e das matrizes de correlação 20% foi possível observar que, por volta de 12:50, as matrizes vão apresentando uma mudança de comportamento voltando a apresentar mais valores de correlação positiva; para 13:01 há formações de regiões bem delimitadas

de correlações fortemente positivas e para 13:02 essas regiões se expandem de modo a ter-se uma Matriz de Correlação quase que totalmente de valores de correlação positivos e a se ter uma Matriz de Correlação 20% e uma Matriz de Correlação 40% de valores de correlação equilibrados apresentando quase a mesma quantidade de correlações, anti-correlações e não correlações. Esse comportamento volta a parecer e se torna recorrente no intervalo entre às 13:12 até as 14:02 porém com mais valores negativos de correlação.

À partir das 15:02, as matrizes começam a apresentar mais anti-correlações e não correlações e esse comportamento segue até as 16:01. De uma maneira geral, pode se dizer que no início do período intra-diário as empresas estão mais correlacionadas entre si, por volta do horário de almoço as empresas apresentam um misto de correlação, não correlação e anti-correlação ou regiões bem delineadas de fortes correlações e quanto mais se aproxima do término do período intra-diário as matrizes vão apresentando maiores quantidades de não correlação ou anti-correlação.

Esses comportamentos foram mais facilmente observados no conjuntos de Matrizes de Correlação 20% por apresentar uma riqueza maior de detalhes já que a diferença entre os valores de correlação são mais facilmente visíveis em consequência de que os maiores valores de correlações e anti-correlações tornam-se mais baixos. Isto indica que houve uma melhoria na definição e caracterização da dinâmica pois a flutuação é menor e demonstra que a redução de dimensionalidade feita evidencia os padrões encontrados sem perder as informações mais importantes.

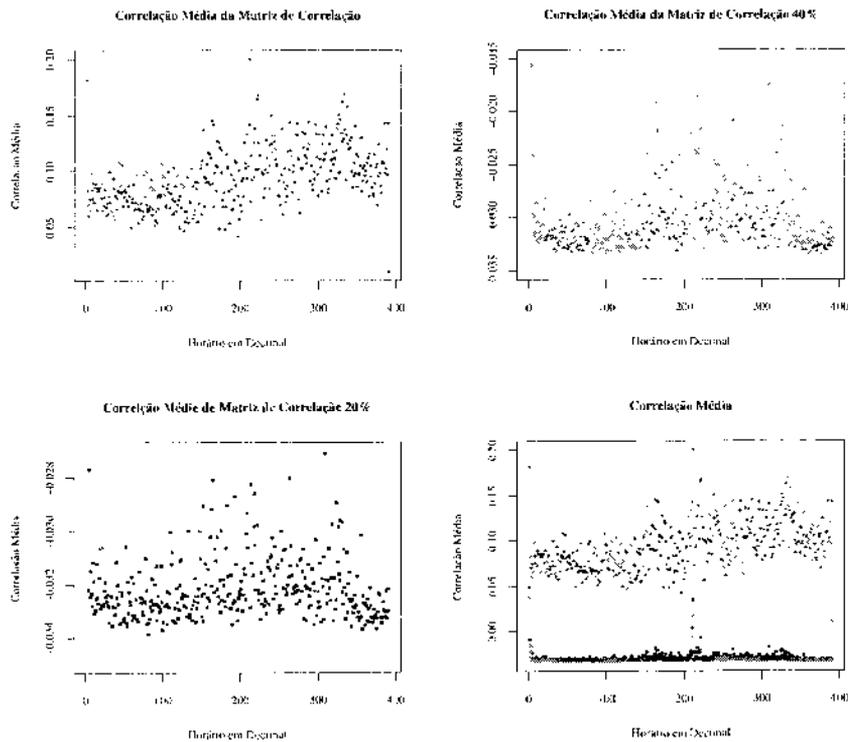


Figura 4.4: Curva da Média do Valor de Correlação para as Matriz de Correlação, Matriz Correlação 20% e Matriz Correlação 40%.

Por fim, foi possível notar que a diferença entre a matriz de correlação de 20% e a de 40% não é tão significativa, mostrando que os autovalores que somados correspondem a 20% do autovalor total são os, de fato, mais significativos e, por tanto, seus autovetores correspondentes descrevem os padrões que mais caracterizam a dinâmica.

Para que essas características fossem quantificadas, foi calculada a média das correlações da região acima do traço das matrizes. Assim, para  $i < j$ , os resultados obtidos para a média de  $c_{ij}$  para as Matrizes de Correlação, Matrizes de Correlação 20% e Matrizes de Correlação 40% para cada horário intra-diário foram dispostos na Figura 4.4.

Note-se que os os períodos em que surgem correlações mais relevantes na matriz de correlação geral são aqueles em que também existem correlações fortes nos valores extremos, de onde podemos deduzir que as correlações são dominadas pelas correlações entre valores “extremos”.

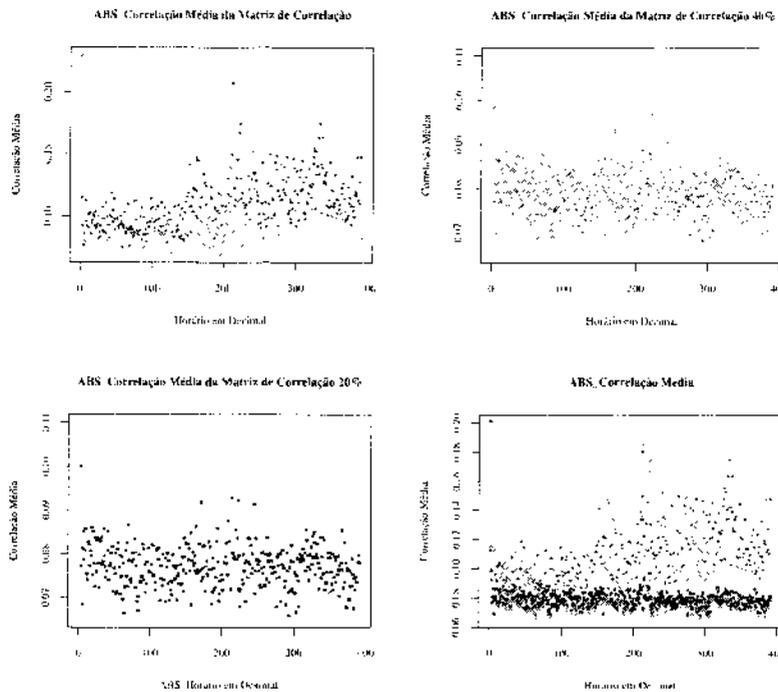


Figura 4.5: Curva da Média do Valor Absoluto de Correlação para as Matriz de Correlação, Matriz Correlação 20% e Matriz Correlação 40%.

### 4.1.2 Matrizes Completas e Matrizes Condicionadas em Valor Absoluto

Para se observar melhor as matrizes de correlação quanto às empresas estarem ou não correlacionadas ou anti-correlacionadas entre si, foram feitos mapas de calor com os valores absolutos de correlação.

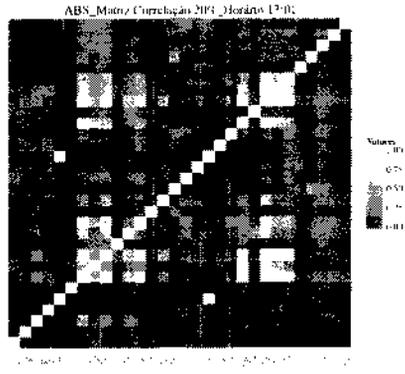
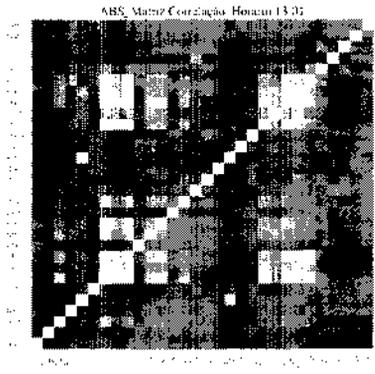
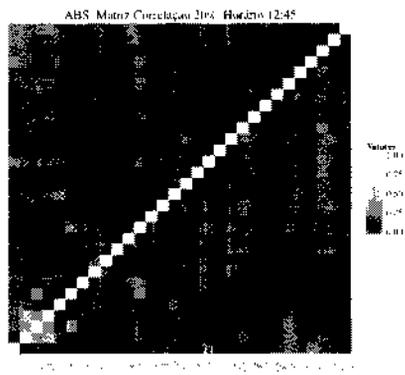
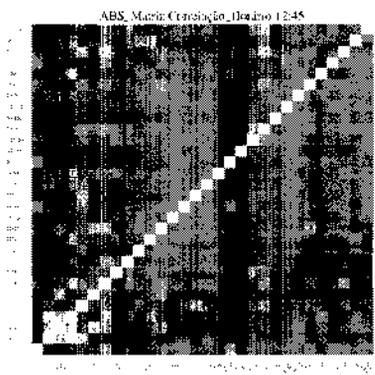
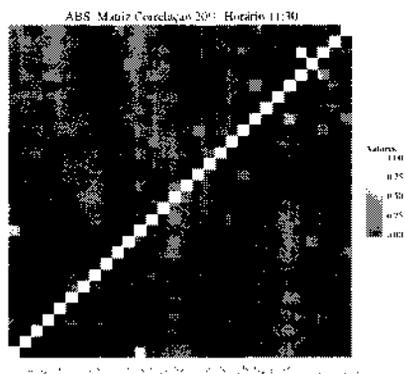
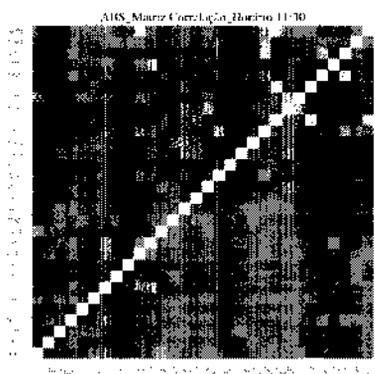
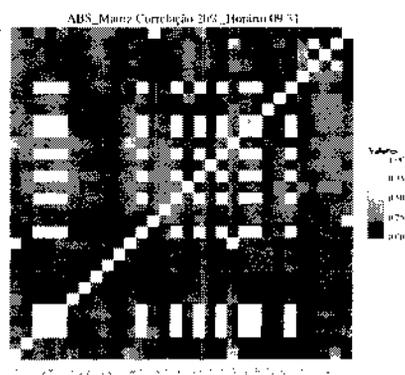
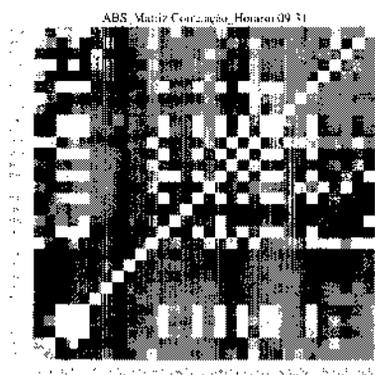
Observando a Figura 4.6 é possível confirmar o que se havia observado nas Figura 4.1, Figura

4.2 e Figura 4.3: no início do período intra-diário existe um maior número de correlações ou anti-correlações entre as empresas e, com o passar do dia, tem-se um número muito grande de não correlações e de correlações com valor absoluto pequeno, até 0.25, e poucas empresas com valor absoluto de correlação entre 0.25 e 0.50 e quase não há correlações com valores absolutos maiores que 0.5.

O comportamento visto nos horários de abertura, 09:31 e nos horários perto do almoço, por volta das 13:00 horas, ainda apresentam o comportamento observados nos mapas de calor de correlações relativas porém em valor absoluto é possível delinear melhor as áreas de altos valores de correlações. Relembre-se que para a análise da curtose foi encontrado um máximo na região das 13:00 horas.

Adicionalmente, para todos os horários é possível verificar que os pontos mais claros correspondentes aos valores absolutos mais altos de correlação são coincidentes para as Matrizes de Correlação e as Matrizes de Correlação 20% permitindo afirmar que os valores maiores de volume são os grandes responsáveis pela correlação entre os volumes das empresas.

De maneira a quantificar tais características, foram calculados a correlação média da região superior ao traço das Matrizes de Correlação em valor absoluto, das Matrizes de Correlação 20% em valor absoluto e das Matrizes de Correlação 40% em valor absoluto, como pode ser visto na Figura 4.5.



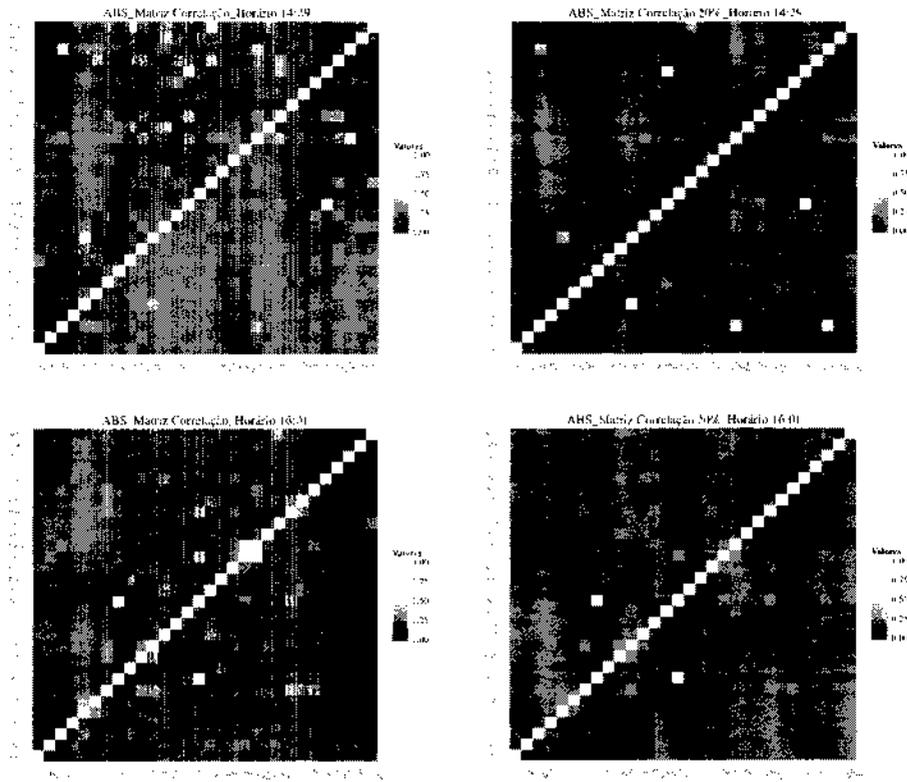
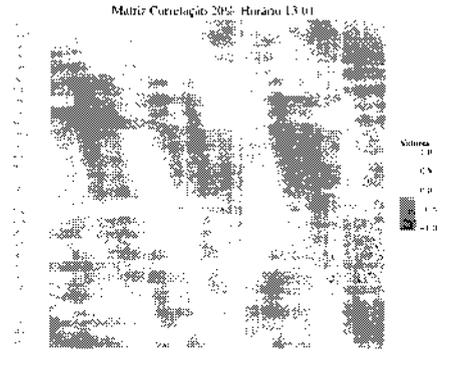
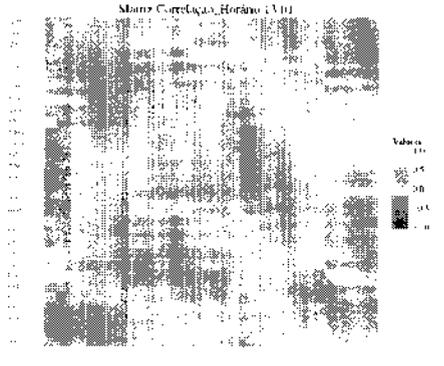
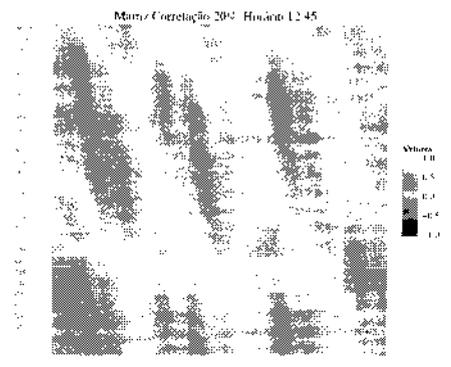
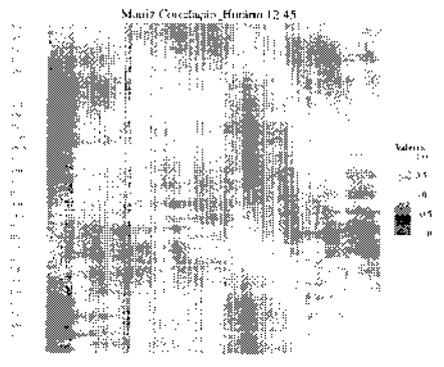
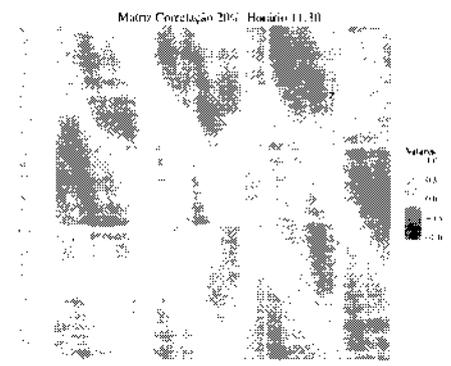
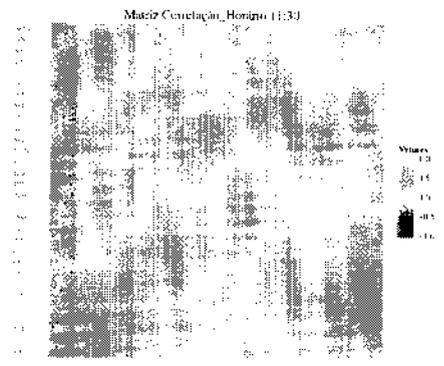
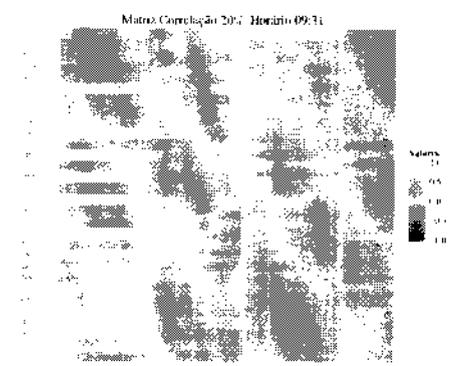
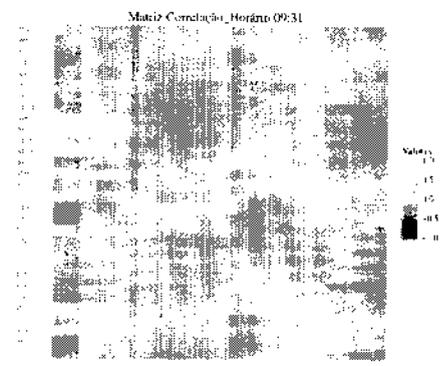


Figura 4.6: Mapa de Calor das Matrizes de Correlação e Matrizes de Correlação 20% em Valor Absoluto para os horários 09:31 horas, 13:01 horas e 16:01 horas.

### Ordenação das Matrizes Completas e Condicionadas

Para que as regiões de maior correlação fossem analisadas, assim como a própria distribuição de cores ( diferentes valores de correlação ) que predominavam nos gráficos de calor, os elementos de cada linha de cada matriz de correlação foram reorganizada de modo decrescente de valor de correlação. Esta organização destrói as propriedades matemáticas de  $C$ , não podendo por isso se falar de uma matriz no sentido puro.

Os gráficos de calor estão dispostos nas Figura 4.7 e Figura 4.8 e, a partir deles, é possível confirmar o que já tínhamos observado: as Matrizes de Correlação 20% apresentam uma to-



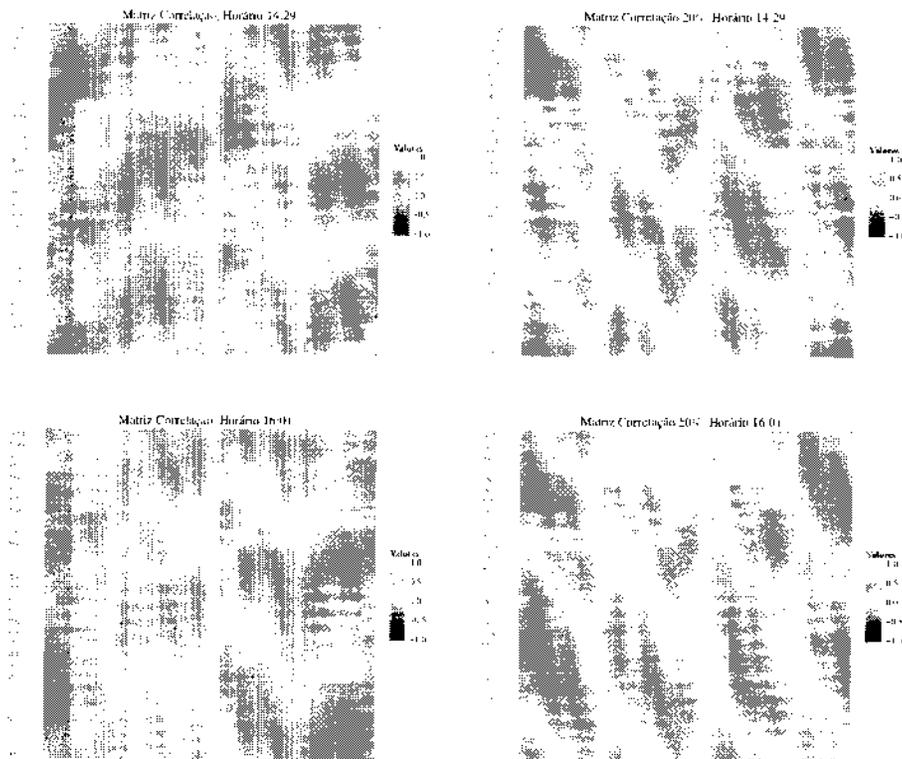


Figura 4.7: Mapa de Calor das Matrizes de Correlação e Matrizes de Correlação 20% em ordem decrescente para os horários indicados.

nalidade mais esverdeada em consequência do aumento de anti-correlações entre as empresas.

Essas anti-correlações são expostas quando usamos o volume de corte a fim de se ter apenas as transações de grandes volumes que podem ser tomadas como transações de peso.

Além disso, o fato de que as Matrizes de Correlação 20% possuem áreas de alta correlação mais bem definidas também se confirma.

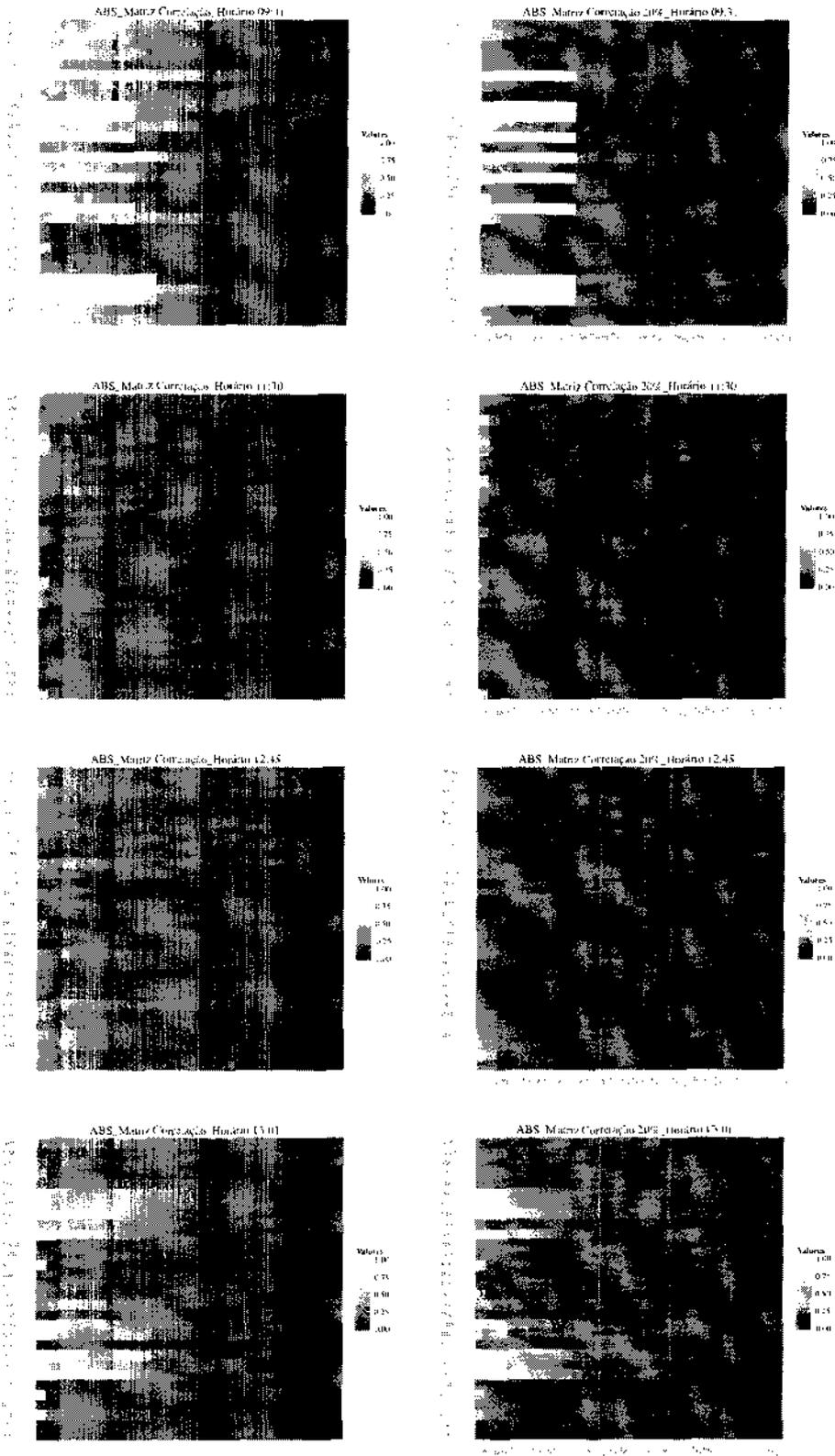
Quanto ao período intra-diário, assim como foi observado nas Figura 4.1, Figura 4.2, Figura 4.3 e Figura 4.6, o horário da abertura e do almoço mostram um padrão de comportamento bem diferentes das demais, certamente por o momento de abertura ser considerado um instante

de choque, como já foi referido anteriormente, e pelo horário de almoço ser uma característica sazonal bem influente. Tanto para às 11:30 horas como para às 12:45 horas, tem-se gráficos de calor com pouquíssimos pontos de valor de correlação maior que 0.5, certamente decorrente do aumento do número de negociações.

Esse comportamento volta a acontecer às 14:29 horas e no horário de fechamento porém com uma pequena maior quantidade de correlações acima de 0.5 e uma menor quantidade de correlações entre 0 e 0.5, sendo resultado de que algumas negociações já foram concretizadas para o dia de negociações mas ainda há um bom trânsito delas.

Já o comportamento do horário de almoço, 13:01 horas, e do horário de abertura diferem bastante dos demais: tem-se muito mais valores de correlações altamente positivas e altamente negativas. Isso deve-se ao fato de que, antes do horário de abertura, o mercado não estava a negociar e durante o horário do almoço há uma diminuição das negociações, porém as especulações e informações fazem que no período anterior à abertura e durante o almoço existam tentativas de manipulação de agentes nos mercados para tirar vantagens de forma a criar as perturbações que ocorrem no mercado no período do almoço.

Este tipo de tentativa de manipulação pode ser visto no mercado de câmbio e na “guerra do almoço” existente entre Tóquio e Hong Kong que, por começarem mais cedo, eles têm a capacidade de liderar.



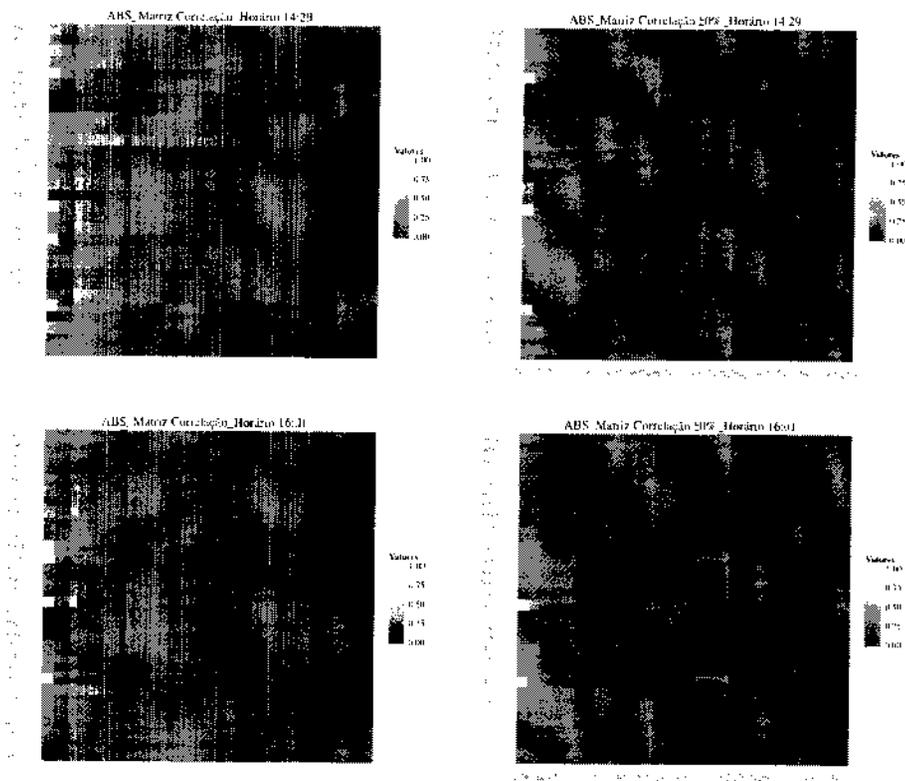


Figura 4.8: Mapa de Calor das Matrizes de Correlação, e Matrizes de Correlação 20% em ordem decrescente do valor absoluto das correlações para os horários indicados.

## 4.2 Análise à luz da Teoria da Matriz Randômica

Os maiores e menores autovalores obtidos de cada Matriz de Correlação, Matriz de Correlação 20% e Matriz de Correlação 40% relativa a cada horário intra-diário calculadas na Seção 3.1 foram comparados com o maior e o menor autovalor obtido com os limites teóricos da Teoria da Matriz Randômica calculados através da equação

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \quad (4.5)$$

que pode ser visto com mais detalhes na seção 2.3. O resultado obtido foi um autovalor máximo teórico de 2.203 e um autovalor mínimo teórico de 0.266. Porém, o segundo valor já se

encontra bastante próximo do valor teórico máximo o que aponta para a não-relevância como modo diretor a partir desta ordem. Estes resultados são bastante surpreendentes, pois para variáveis claramente menos correlacionadas como é o caso das flutuações de preço e considerando uma frequência de amostragem mais baixa foi encontrado um número elevado de valores com relevância de modo, mais precisamente os sete primeiros autovalores [13].

Na Figura 4.9 é possível visualizar a porcentagem dos autovalores máximos em relação ao autovalor máximo teórico para cada horário e a porcentagem dos autovalores mínimos em relação ao autovalor mínimo teórico para cada horário das Matrizes de Correlação (representados pelos pontos de cor turquesa), das Matrizes de Correlação 40% (representados pelos pontos de cor rosa) e das Matrizes de Correlação 20% (representados pelos pontos de cor laranja). Como os autovalores mínimos das Matrizes de Correlação 20% e das Matrizes de Correlação 40% são da ordem de  $10^{-16}$  enquanto o teórico é de 0.266, temos duas retas sobrepostas no gráfico do autovalor mínimo em relação ao teórico.

Para a Matriz de Correlação tem-se que o valor médio dos autovalores máximos é 4.19 sendo 1.9 vezes maior que o valor do autovalor máximo teórico e a média dos autovalores mínimos é de 0.22 sendo 0.82 vezes o valor do autovalor mínimo teórico. O valor médio do autovalor máximo da Matriz de Correlação 40% é de 2.29 que corresponde à 1.04 vezes o valor teórico enquanto o valor médio do autovalor mínimo é de  $1.87 \times 10^{-16}$  sendo  $7.03 \times 10^{-16}$  vezes o valor do autovalor mínimo teórico. Por fim, o valor médio do autovalor máximo da Matriz de Correlação 20% é de 2.26 correspondendo à 1.03 vezes o valor do autovalor máximo teórico e

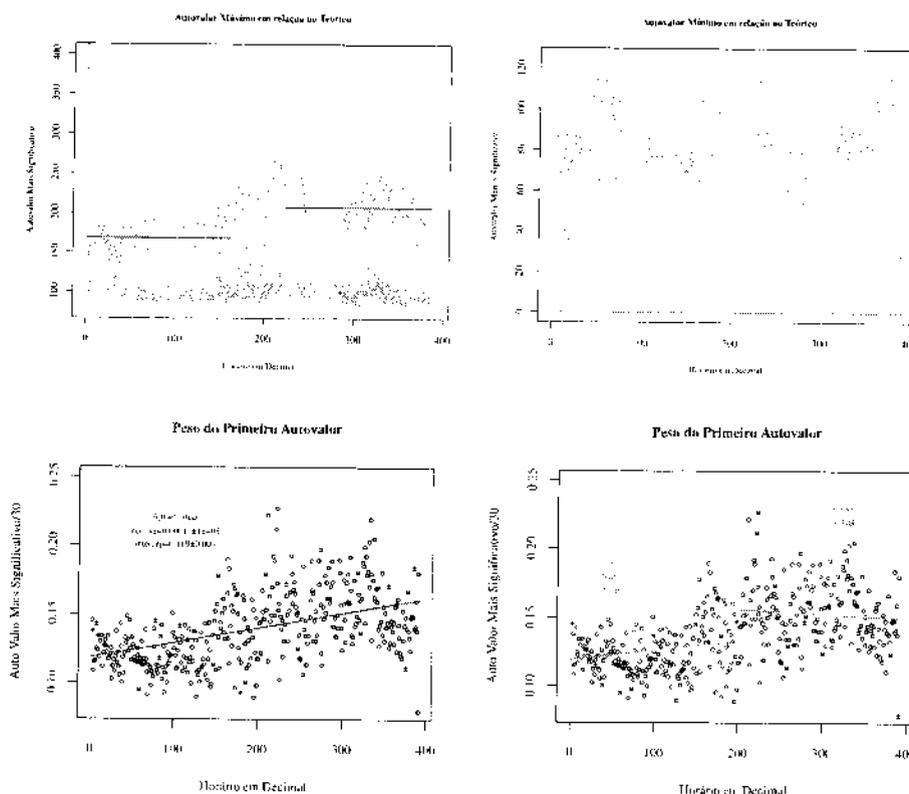


Figura 4.9: Porcentagem dos autovalores máximos(mínimos) em relação ao autovalor máximo(mínimo) das Matrizes de Correlação (turquesa), Matrizes de Correlação 40%(rosa) e Matrizes de Correlação 20% (laranja). Na direita baixa: Peso do Autovalor mais significativo para cada horário. A linha tracejada vermelha corresponde ao ajuste linear de todo o período intra-diário enquanto a linha tracejada laranja corresponde ao ajuste linear do período matutino e a linha tracejada rosa corresponde ao ajuste linear do período vespertino. Os erros dos declives são da mesma ordem do seu valor apontando para existência de dois patamares.

o valor médio do autovalor mínimo é  $3.08 \times 10^{-16}$  correspondendo à  $1.16 \times 10^{-15}$  vezes o valor teórico do autovalor mínimo.

Além disso, foi possível observar que em média o autovalor máximo das Matrizes de Correlação é de 13.97% da soma de todos os autovalores que é de 30, em média o autovalor máximo das Matrizes de Correlação 40% é de 7.63% da soma de todos os autovalores e o maior auto-

valor das Matrizes de Correlação 20% correspondem em média à 7.54% do valor da soma de todos os autovalores.

Adicionalmente, a porcentagem do autovalor máximo das Matrizes de Correlação, representados pelos pontos em turquesa, parecem ter dois subconjuntos de dados diferentes, um pertencente ao intervalo de horário entre às 09:31 horas até as 12:44 horas e o outro pertencente ao intervalo entre às 13:14 até às 15:57 horas, como pode ser visto pelas retas vermelhas no gráfico dos autovalores máximos em relação ao teórico na Figura 4.9 que representam a média de cada subconjunto.

Para que essa diferença fosse analisada, foi aplicado o teste t-Student num intervalo de confiança de 95%. O resultado foi um  $t_{calculado} = -12.09$  enquanto o valor t tabelado para 120 graus de liberdade, um intervalo de 95% de confiança é de  $t_{teórico} = 1.654$  e um  $p - value$  de  $2.2 \times 10^{-16}$ .

Como  $p - value$  é muito menor que o intervalo de confiança de 5%, não podemos aceitar a hipótese nula de igualdade das médias. Isto é confirmado pelo fato de que o  $|t_{calculado}| > |t_{teórico}|$ . Desta maneira, pode se considerar que se tratam de dois subconjuntos de comportamentos diferentes.

Com esse resultado, novamente, pode se inferir que o autovalor máximo trata-se de um modo de mercado e como visto, o maior autovalor é fortemente não-Gaussiano, tendendo à uniformidade, sugerindo então que todas as empresas participam. Isto pode ser quantificado

com precisão pelo conceito de rácios de participação inversa que vem da teoria de localização onde se encontra o fato que todos os componentes participam aproximadamente igual ao valor próprio [15]. Isto implica que cada empresa está conectada com todas as outras empresas, indicando que todo o mercado se move junto e a presença de correlações que permeiam todo o sistema.

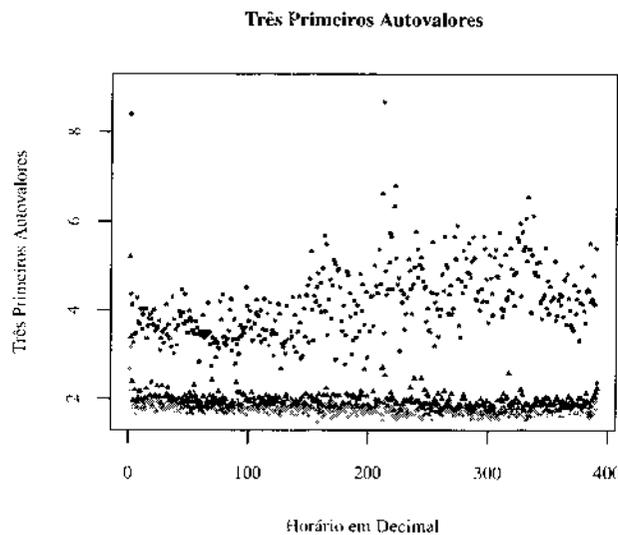


Figura 4.10: Primeiro autovalor mais significativo, representado pela curva vermelha, do segundo autovalor mais significativo, representado pela curva azul, e do terceiro autovalor mais significativo, representado pela curva verde, em função dos horários intra-diários.

Nos gráfico da Figura 4.10 foram colocadas as três curvas dos três primeiros autovalores mais significativos. É fácil notar que até o horário decimal 200 que corresponde às 12:50 horas, tem-se um comportamento parecido entre as duas curvas. O valor do primeiro autovalor é bastante mais elevado que o valor máximo teórico.

Adicionalmente é possível notar que com exceção do primeiro minuto que tem um compor-

tamento muito particular o que é devido ao momento de choque já comentado anteriormente, os primeiros 150 minutos intra-diários apresentam um modo de mercado basicamente independente do horário. Entre as 12:00 horas e as 15:00 horas, se percebe uma dependência crescente do modo de mercado com o horário e daí em diante este modo decresce.

Para examinar a estabilidade no tempo dos autovalores e dos autovetores correspondentes que se desviam do limite da RMT, tomou-se o período total analisado no capítulo sobre sazonalidades o qual é composto de 19 semestres. Para cada um dos semestres, foram calculadas as matrizes de correlação para cada horário que ia de 9:31 até 16:01, num intervalo de 1 minuto, e em seguida seus valores próprios e vetores próprios.

Para cada semestre, foram identificados os três primeiros autovalores e seus autovetores correspondentes. Sendo  $1S2004i$  com  $i=1,2,3$  os autovetores correspondentes aos 3 maiores autovalores do primeiro semestre de 2004 até o  $1S2013i$  os autovetores correspondentes aos 3 primeiros autovalores do primeiro semestre de 2013, a estabilidade foi medida através do produto interno entre o autovetor do primeiro período de tempo, isto é do primeiro semestre de 2004, com os auto vetores dos próximos períodos  $1S2004i \times (2S2004i, \dots, 1S2013i)$ . Para uma análise mais visual, foram montados os gráficos destes produtos internos e foi observado que este pode quantificar o quão divergente é um autovalor dos demais e que este desvio representa um grau de estabilidade temporal diferente.

Ou seja, se faz a seguinte análise: para o primeiro autovetor que corresponde ao maior

autovalor, por exemplo, faz-se o produto interno entre o primeiro autovetor do primeiro semestre de 2004 para uma dada hora fixa com o primeiro autovetor para esta mesma dada hora de todos os outros semestres. Isso foi feito a cada 15 minutos no intervalo intra-diário de 09:31 até 16:01.

Os resultados foram que, a medida que o autovetor é menos significativo, menor é a divergência entre os pontos e que, para os três autovetores, o produto interno entre eles ao longo do dia não apresentou um padrão que descrevesse de modo claro alguma evolução temporal porém é visível, como pode ser visto na Figura 4.11, que o comportamento dos produtos internos no horário final do período intra-diário e no horário de almoço diferem daqueles encontrados nos outros horários e os pontos referentes ao produto interno entre os primeiros autovalores mais significativos, pontos em vermelho, do Segundo Semestre de 2009 e ao Primeiro Semestre de 2010 se destacam dos demais já que nestes semestres não houve negociação de ações da General Motors Corp..

Novamente, quanto ao período do almoço, é fácil observar no gráfico das 12:45 horas na Figura 4.11 que no Segundo Semestre de 2009 e no Primeiro Semestre de 2010 existe uma aproximação entre os autovalores, onde o valor do produto interno entre os primeiros autovalores e entre os terceiros autovalores mais significativos diminui enquanto os valores do produto interno entre os segundos autovalores mais significativos aumenta.

Adicionalmente, ao observar o gráfico do desvio padrão da Figura 4.11, é possível verificar que os valores de desvio padrão em unidades de média para o produto interno entre os segundos

autovetores mais significativos e para o produto interno entre os terceiros autovetores mais significativos são bem semelhantes enquanto que para o produto interno entre os primeiros autovetores mais significativos o desvio padrão em unidade de média apresenta valores de maior magnitude.

Porém, a diferença entre os pontos da curva do desvio padrão em unidades de média do produto interno entre os primeiros autovetores mais significativos, entre os segundos autovetores mais significativos e entre os terceiros autovetores mais significativos apresentam magnitudes semelhantes sugerindo que, apesar de que para os primeiros autovalores há uma variabilidade de valores maiores, essa variabilidade acontece para todos os horários.

Há também uma semelhança curiosa entre o comportamento das três curvas para os primeiros três horários e os dois últimos horários, sendo que esses dois últimos apresentam uma diferença significativa entre os seus valores. Certamente em consequência que o mercado está à ponto de encerrar e então há uma variação muito grande do comportamento das negociações.

Semelhantemente, porém com um valor de diferença menor, é o que acontece na abertura no mercado: a diferença grande entre os três últimos pontos é consequente do período pós colapso que acontece no instante após a abertura do mercado.

Para verificar o comportamento dos produtos internos entre os três primeiros autovetores mais significativos, tratou-se os semestres em forma de sobreposição de modo a ter-se 109 semestres totais em sobreposição.

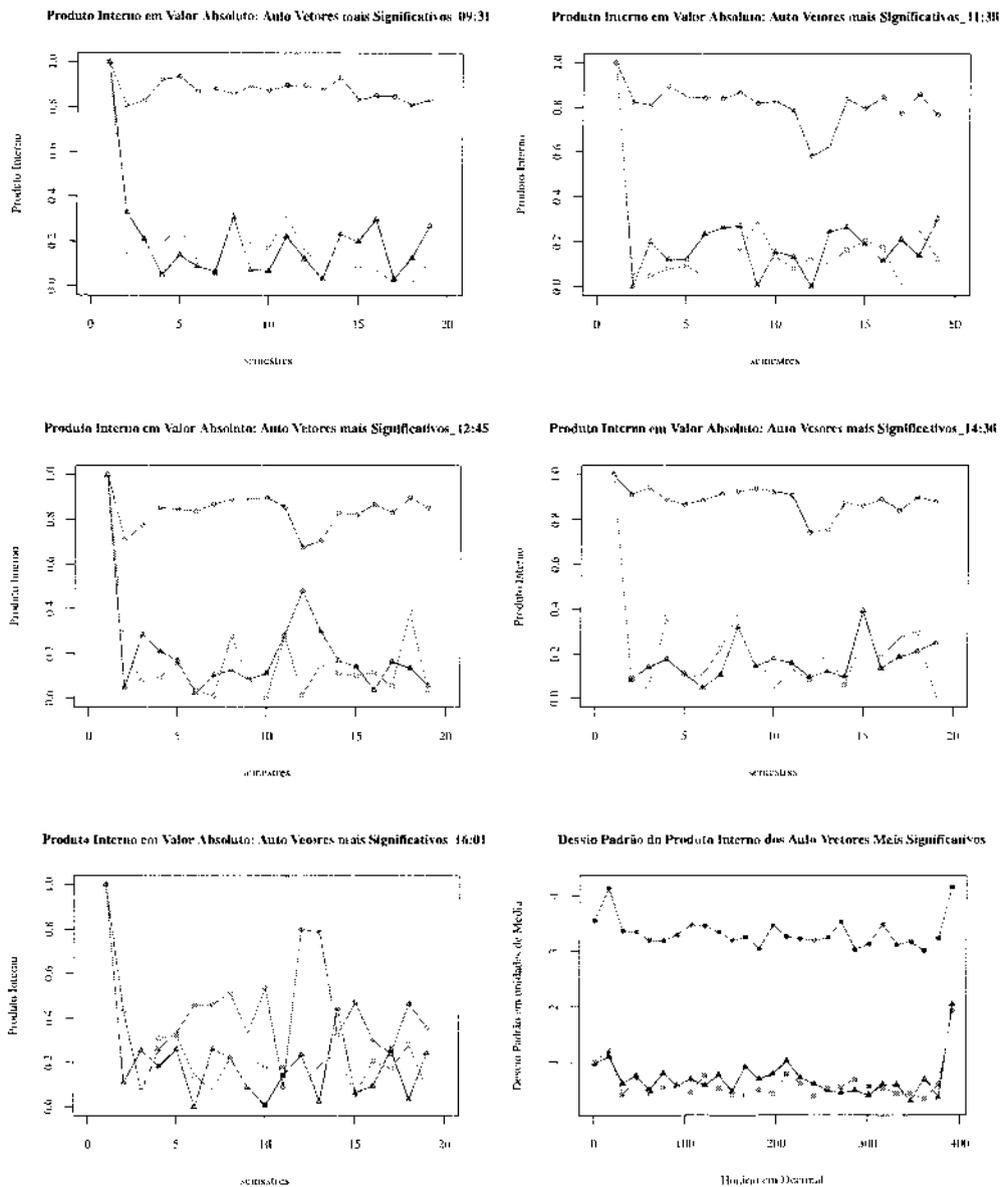


Figura 4.11: Produto Interno dos três primeiros autovetores mais significativos e Desvio Padrão em unidades de Média. Os pontos vermelhos/azuis/verde representam o produto interno entre os auto vetores associados aos primeiros/segundos/terceiros autovetores mais significativos.

O resultado que pode ser visto na Figura 4.12, evidencia que o produto interno dos primeiros autovalores mais significativos dos grupos de semestres em sobreposição que contém os dados do segundo semestre de 2009 e do primeiro semestre de 2010 se destacam dos demais pelo fato

de não existir valores para a empresa General Motors Corp. e, por tanto para esses semestres as matrizes de correlação são formada apenas das 29 demais empresas.

Além disso, os comportamentos do produto interno entre os primeiros autovetores mais significativo para às 16:01 horas e para às 12:45 horas diferem dos encontrados nos demais horários. Esses resultados são convergentes com aqueles encontrados para os semestres em modo contíguo.

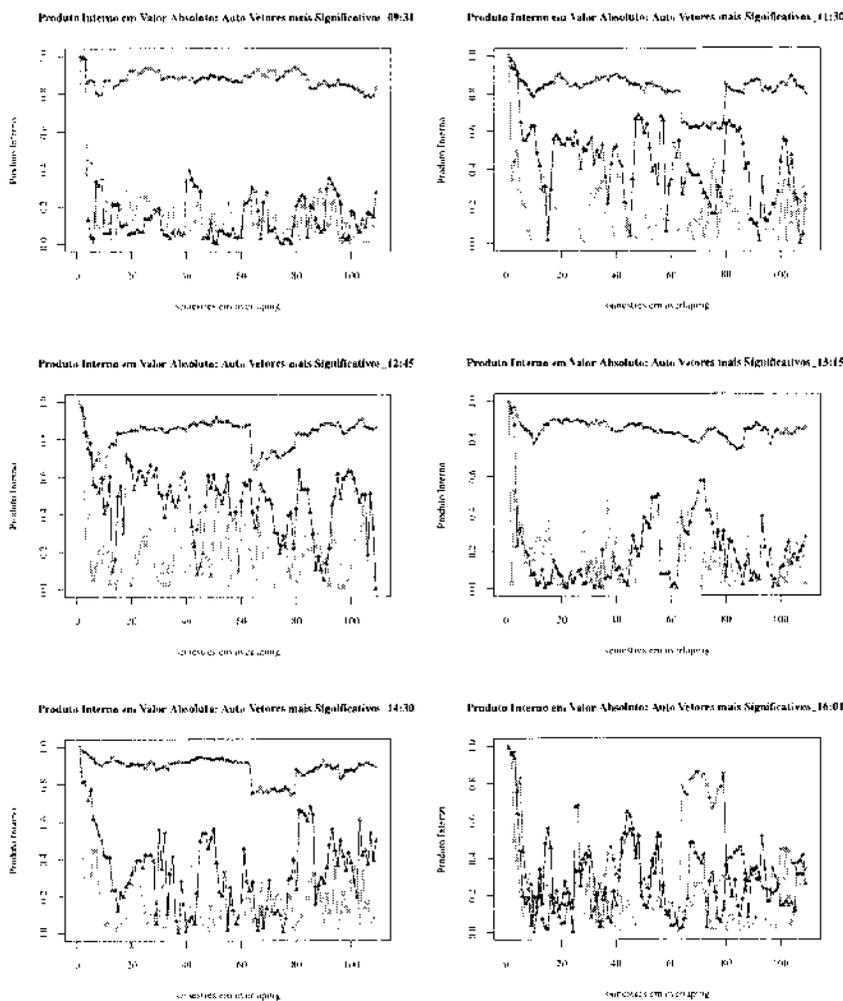


Figura 4.12: Os pontos vermelhos/azuis/verdes representam o produto interno entre os autovetores associados aos primeiros/segundos/terceiros autovalores mais significativos.

De um modo um pouco diferente, defini para cada semestre vetores formados pelos autovalores em cada horário de determinada ordem

$$\vec{A}_i(\text{semestre}) \equiv \frac{1}{\sqrt{\sum_{j=1}^{391} \lambda_i^2(j)}} \{\lambda_i(1), \dots, \lambda_i(391)\}. \quad (4.6)$$

Em seguida se realizou o produto interno dos vetores ( $i=1,2,3$ ) do Primeiro Semestre de 2004 com os restantes semestres.

Como pode ser visto na Figura 4.13, à medida que baixamos a ordem do autovalor a coerência de  $\vec{A}_i$  aumenta, resultado que é remanescente dos resultados da teoria da matriz randômica [30] que apontam para uma perda de estabilidade temporal dos autovetores à medida que o autovalor correspondente se aproxima do limite superior previsto pela teoria. Mais especificamente, é possível observar que do primeiro semestre em sobreposição até o semestre 47 em sobreposição, que engloba desde o primeiro semestre de 2004 até o começo do Primeiro Semestre de 2008, tem-se um comportamento decrescente para a curva do produto interno entre os primeiros autovalores. Depois deste período, a curva muda seu comportamento para os semestres a partir do primeiro semestre de 2008, lembrando que 2008 foi o colapso da bolha imobiliária, e volta a mudar de comportamento no semestre 65 em sobreposição e novamente no semestre 90 em sobreposição. O semestre 65 em sobreposição corresponde ao interlace entre o meio do primeiro semestre de 2009 e o meio do segundo semestre de 2009, período no mercado financeiro vivia uma turbulência pelo medo de alguns países, nomeadamente da zona do Euro não cumprirem os seus compromissos de dívida pública, e o semestre 90 em sobreposição corresponde o período de interlace entre o final do Primeiro Semestre de 2011 e o final

do Segundo semestre de 2011, período no qual ocorreu a crise do limite da dívida pública dos EUA.

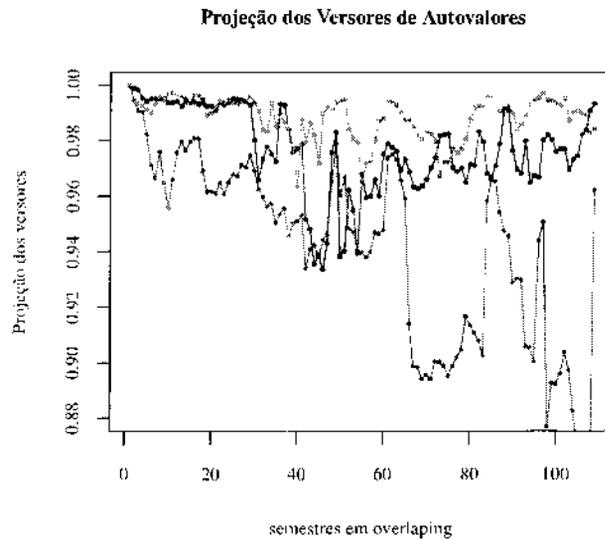


Figura 4.13: Projeção dos primeiros três versores de autovalores  $\vec{\lambda}_i$  ao longo dos semestres em sobreposição. Os pontos vermelhos/azuis/verdes representam o produto interno entre os versores associados ao primeiro/segundo/terceiro autovalor mais significativo. Quatro pontos referentes ao primeiro autovalor foram omitidos para que se tivesse uma ampliação do gráfico. Esses pontos correspondem aos semestres 105, 106, 107 e 108 e valem 0.867, 0.858, 0.862 e 0.843 respectivamente.

Uma análise intra-diária também foi feita e para isso foram usadas as matrizes de correlação calculadas com os dados do Segundo Semestre de 2004. O resultado foram 391 matrizes que iam desde 09:31 até 16:01.

Com o objetivo de se verificar como o peso do autovalor  $\lambda_1(t)/N$  evolui dentro do período intra-diário foram calculados todos os primeiros maiores autovalores, um para cada horário, e todos eles foram divididos pelo número de empresas  $N=30$ . Em seguida esses pesos foram dispostos em função do horário num gráfico.

Como pode ser visto na Figura 4.9, no gráfico do peso do primeiro autovalor, o resultado foi uma curva com concavidade negativa entre 09:31 até aproximadamente 12:25 com um ponto de mínimo em aproximadamente 10:50 e depois a curva passa a ter uma concavidade positiva até às 16:01 com um ponto de máximo em aproximadamente 15:10 horas. Esse formato se assemelha com  $\smile$ , o mesmo visto para a curva da Curtose em função do horário intra-diário.

Essa transição da concavidade da curva está dentro do intervalo de almoço, mostrando mais um efeito de sazonalidade devido ao comportamento e hábitos humano. Foi possível observar que de 12:20 até as 16:00, os pesos tem maiores valores do que o intervalo anterior inferindo assim que os modos de mercado são mais fortes e significativos durante este período, ou seja, neste as empresas apresentam uma dinâmica conjunta mais evidenciada.

Quanto ao conjuntos de valores dos pesos do primeiro autovalor mais significativo, teve-se um peso mínimo de 8% , um peso máximo de 28% e uma média de 14% mostrando ser bastante significativo. Ou seja, se todo os autovalores tivessem o mesmo valor, cada um teria um peso de 3.33%. Tomando esta referência é possível, então, observar que em média o primeiro autovalor tem um peso 4.42 vezes maior, sendo mais uma evidência de que ele se trata de um modo de mercado.

No gráfico do peso do autovalor separado por período vespertino e matutino na Figura 4.9 é possível observar que os dados do período vespertino se comportam diferentemente do período matutino. Como foi mencionado acima, temos uma inversão de concavidade. Pelas retas de

ajuste laranja e rosa, pode se também observar que os dados matutinos tem um comportamento crescente enquanto os vespertinos tem um comportamento levemente decrescente porém ainda com um peso maior, em média, que o peso para no período matutino.

Esta propriedade explicita que o modo de mercado se torna mais relevante quanto à direção dinâmica do mercado no período que começa logo após o almoço e vai até o horário de fechamento, com um valor maior para o peso por volta das 15:10 horas.

Quanto maior for o peso do primeiro autovalor, mais ele influenciará no dinamismo do mercado. O fato disso acontecer logo após o horário de almoço é decorrente do fato que a partir deste horário algumas negociações já foram consolidadas e isto influencia as transações que ainda estão sendo feitas, ou seja, há um direcionamento no mercado.

Para que a diferença entre o conjunto de pontos pertencentes ao período matutino e o conjunto de pontos pertencente ao período vespertino, foi aplicado o teste t-Student num intervalo de confiança de 95%. O resultado foi um  $t_{calculado} = -10.54$  enquanto o valor t tabelado para 194 graus de liberdade num intervalo de 95% de confiança é de  $t_{critico} = 1.653$  e um  $p - value$  de  $2.2 \times 10^{-16}$ .

Como  $p - value$  é muito menor que o intervalo de confiança de 5% e como  $|t_{calculado}| > |t_{critico}|$ , não podemos aceitar a hipótese nula de igualdade das médias. Assim, pode se considerar que se tratam se subconjuntos de diferentes comportamentos.

Se todas as empresas tivessem o mesmo peso, teria-se um versor autovetor  $u_1(t)$  correspon-

dente ao autovalor  $\lambda_1(t)$  bem próximo ao versor uniforme

$$\vec{u}_{\text{unif}} = \frac{1}{\sqrt{N}}\{1, \dots, 1\},$$

( $N = 30$ ). A fim de se verificar o comportamento do produto interno  $\vec{u}_1(t) \cdot \vec{u}_{\text{unif}}$  ao longo do período intra-diário, foi traçada uma curva deste em função do tempo.

O resultado, como pode ser visto na Figura 4.14, trata-se de uma curva crescente em sua totalidade considerando o valor absoluto do produto interno, ou seja, até às 16:00 tem-se que o valor absoluto do produto interno apresenta um crescimento ao longo do dia mostrando que durante o período de transação intra-diário considerado a significância do autovetor correspondente ao autovalor mais significativo persiste e se torna mais forte.

Já o gráfico do produto interno entre os autovetores que diferencia o período matutino do vespertino na Figura 4.14 exibe que o comportamento dos dados pertencente ao período vespertino difere daquele encontrado para os dados pertencente ao período matutino: o ajuste dos dados vespertinos exibe uma maior inclinação.

Para analisar como decai a coerência dos primeiros autovetores precisa-se ter cuidado ao realizar-se médias intradiárias já que o volume ao longo deste período não é estacionário,. Assim, deve-se considerar a evolução do produto entre:

1.  $u_1(t_i) \cdot u_1(t_i + \delta)$  no qual  $\delta$  vai de  $t_i$  até  $t_f - 1$  onde  $t_i$  trata-se do horário inicial, ou seja 09:31, e  $t_f$  trata-se do horário final, ou seja 16:01;

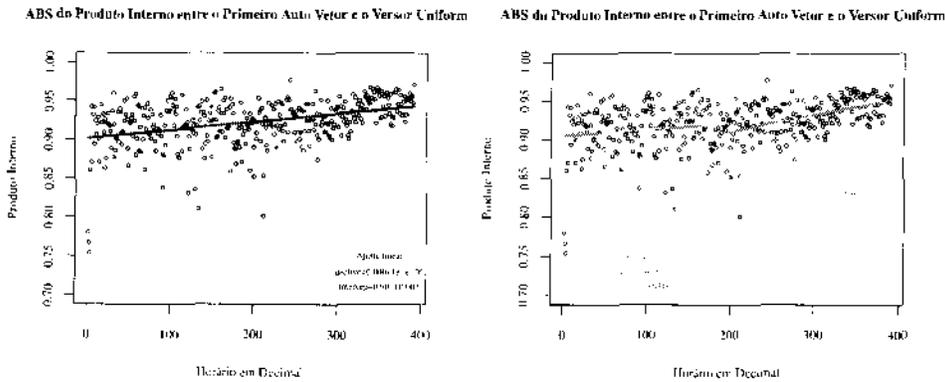


Figura 4.14: Projeção do autovetor mais significativo para cada horário no Versor Uniforme. A linha tracejada vermelha corresponde ao ajuste linear de todo o período intra-diário enquanto a linha tracejada laranja corresponde ao ajuste linear do período matutino e a linha tracejada rosa corresponde ao ajuste linear do período vespertino. Note que no período vespertino o declive e o erro são da mesma ordem de grandeza apontando assim para a independência da projeção nos horários desta parte da sessão.

Como pode ser visto nos dois primeiros gráficos da Figura 4.15, o resultado foi uma curva de concavidade positiva bem suave com média de 0.71 onde os pontos flutuam em torno da média entre um intervalo de valores que apresenta uma forma linear de 0.55 e 0.83, mas a medida que o tempo passa o intervalo de valores nos quais os pontos flutuam vai reduzindo até chegar entre 0.68 e 0.76.

É fácil notar que o comportamento do período matutino difere do vespertino: o ajuste matutino é uma curva decrescente enquanto o ajuste do vespertino trata-se de uma curva crescente e com uma inclinação mais acentuada. Já o ajuste dos dados intra-diários trata-se de uma curva crescente.

2.  $u_1(t_f) \cdot u_1(t_f - \delta)$  no qual  $\delta$  vai de  $t_i$  até  $t_f - 1$ ;

Como pode ser visto nos dois gráficos do meio da Figura 4.15, o resultado foi uma curva

que possui uma concavidade positiva até 12:30 e passa a ter uma concavidade negativa até às 16:01 também suaves, com média em 0.19 onde os pontos flutuam em torno da média entre um intervalo médio de 0.05 até 0.4.

O ajuste dos dados intra-diários é uma reta decrescente, o ajuste dos dados pertencentes ao período matutino é uma reta crescente e já o do período vespertino trata-se de uma reta decrescente.

3.  $u_1(t_e) \cdot u_1(t_e - \delta)$  no qual  $\delta$  vai de  $t_c - 1$  até  $t_e - t_f$  e  $t_e$  trata-se do horário de fechamento dos mercados na Europa, ou seja, às 14:00 horas.

Como pode ser visto nos dois últimos gráficos da Figura 4.15, o resultado foi uma curva com que apresenta um padrão senoidal de amplitude não uniforme, com concavidade negativa até às 12:00 horas, depois uma concavidade positiva até às 13:30 horas e volta a ter uma concavidade negativa até às 16:01, de média 0.87 onde os pontos flutuam num intervalo inicial de aproximadamente 0.66 até 0.9 e este intervalo vai diminuindo até 0.83 até 0.87.

Aqui também é fácil notar que o comportamento do período matutino se assemelha ao do vespertino quanto à direção de crescimento e a inclinação. Além disso, estes ajustes se assemelham com o ajuste dos dados intra-diários.

O resultado acima, Figura 4.9, para o peso do primeiro autovalor mais significativo, como já comentado, foi de 16% em média, mostrando ser significativo porém não o suficientes para

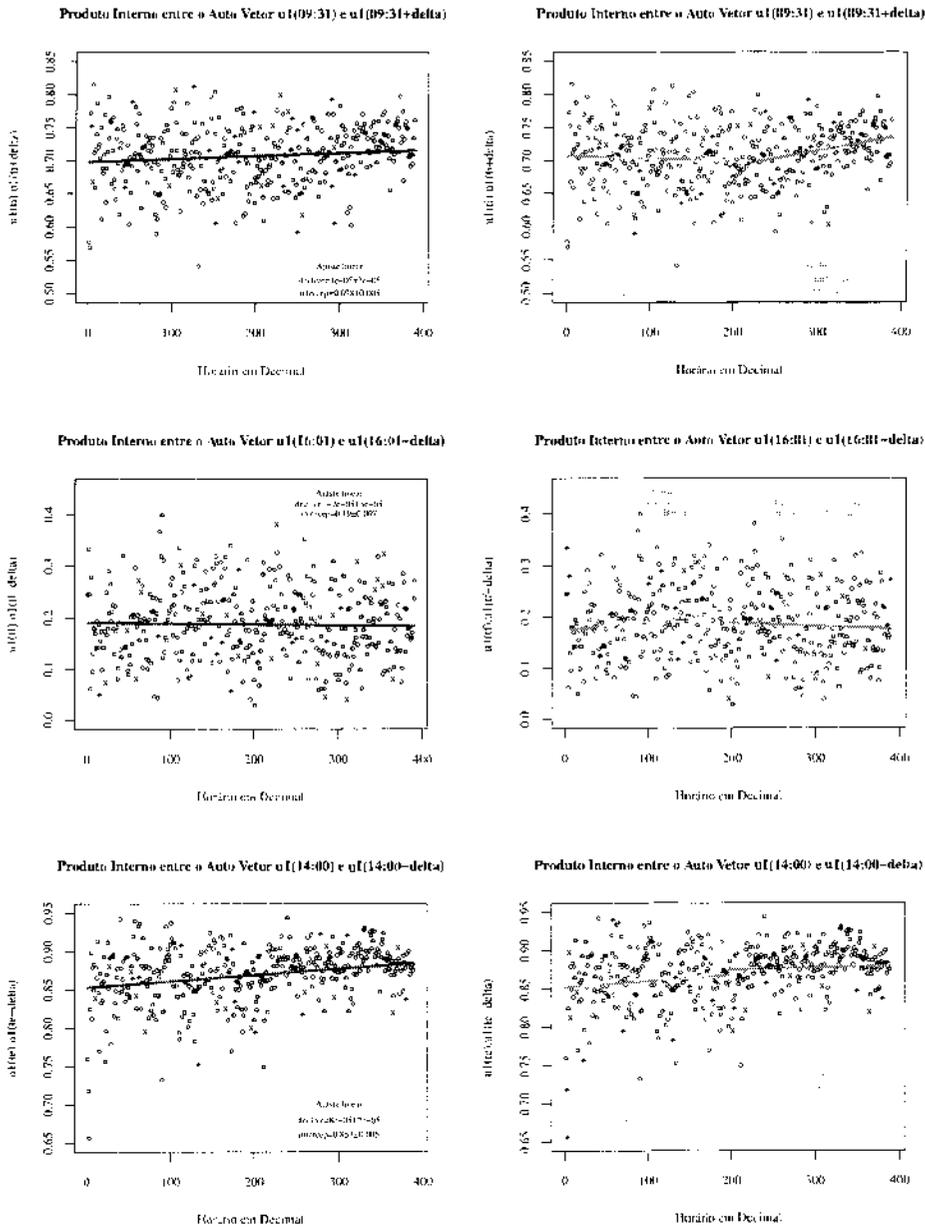


Figura 4.15: Produto Interno entre os três primeiros autovetores mais significativos. Os pontos vermelhos representam o produto interno entre os primeiros autovetores mais significativos, os pontos em azuis representam o produto interno entre os segundos autovetores mais significativos e os pontos em verde representam o produto interno entre os terceiros autovetores mais significativos.

para representar o comportamento do sistema.

A fim de se verificar qual são os pesos dos cinco primeiros autovalores somados  $\frac{1}{N} \sum_{i=1}^5 \lambda_i(t)$

e compará-lo com o primeiro, foi montada outra curva de em função do horário.

Como pode ser visto nos gráficos do peso dos cinco primeiros autovalores da Figura 4.16, a curva obtida também foi uma curva que apresentava mudança de concavidade em 12:30, a curva com concavidade positiva apresenta um ponto de mínimo em 11:20 horas e a curva de concavidade negativa apresenta uma concavidade um pouco mais plana com um máximo em 14:10 horas.

Quanto aos valores dos pesos dos cinco primeiros autovalores mais significativos, tem-se um máximo de 68% e um mínimo de 31% e uma média de 37% em relação ao total. Estes valores ainda são baixos para estipular que somente os 5 primeiros autovetores apresentam todos os padrões suficientes para não perder nenhuma informação importante da dinâmica estudada.

O próximo passo, foi buscar a quantidade mínima de autovalores mais significativos  $n_\lambda = \min n \in [1, N]$  necessárias para se ter em média  $\frac{1}{N} \sum_{i=1}^{n_\lambda} \lambda_i(t) \geq 0.8$ . Isso é satisfeito quando considerado os 17 primeiros autovalores, tendo-se um peso mínimo de 76%, um peso máximo de 99% e um peso médio de 80%.

Para se verificar como evolui esse menor subconjunto suficiente de autovalores os quais seus autovetores correspondentes podem descrever o sistema como um todo através dos principais e mais importantes padrões, foi feito a curva desse peso dos 17 primeiros autovalores para cada horário como pode ser visto nos gráficos para o peso dos dezessete primeiros autovalores da Figura 4.16.

O resultado foi que a curva de concavidade negativa apresentou uma planificação ainda maior que as curvas dos pesos do primeiro autovalor mais significativo e dos cinco primeiros autovalores mais significativos. O ponto de mínimo da curva com concavidade positiva permaneceu em 11:20 horas e o ponto de máximo da curva com concavidade negativa também permaneceu em 14:10.

Quando comparados as curvas dos pesos do primeiro autovalor mais significativo, do peso dos cinco primeiros autovalores mais significativos e do peso dos dezessete primeiros autovalores mais significativos, ver Figura 4.9, Figura 4.16 e Figura 4.16 referente à todo o período intra-diário, é possível notar que o ângulo de inclinação da reta do ajuste vai diminuindo se tornando quase que linear para a curva do peso dos dezessete primeiros autovalores mais significativos.

Se comparados os gráficos que distinguem os pontos pertencentes a cada período matutino e vespertino da Figura 4.9, da Figura 4.16 e da Figura 4.16, é possível verificar que o ajuste do período matutino para o peso do primeiro autovalor é uma reta crescente, enquanto o ajuste do período matutino para o peso dos cinco primeiros autovalores trata-se de uma reta decrescente e a reta do ajuste do período matutino do peso dos dezessete autovalores também é decrescente e apresenta uma inclinação um pouco maior.

Já para o período vespertino, temos uma reta de ajuste decrescente para todas as três curvas e as inclinações também são bem semelhantes. Isto leva-se a concluir que a diferença do com-

portamento total das curvas dos pesos é influenciada mais fortemente pelos dados do período matutino.

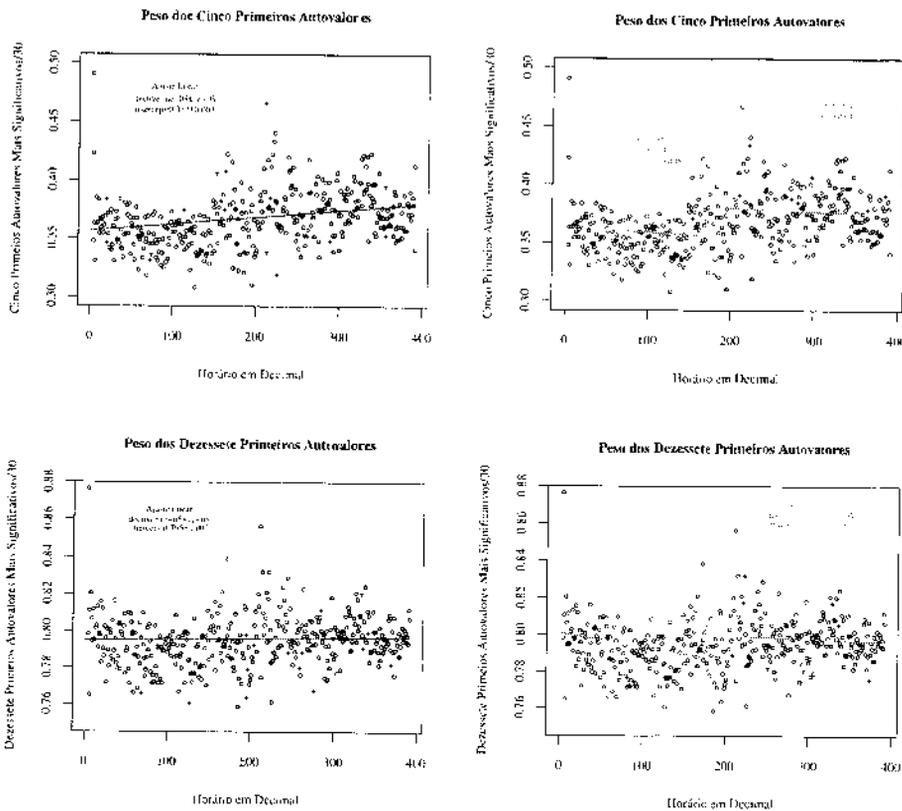


Figura 4.16: Em cima: Peso dos cinco primeiros Autovalores mais Significativos; em baixo: Peso dos dezessete primeiros Autovalores mais Significativos. A reta vermelha representa o ajuste linear intra-diário, a reta laranja representa o ajuste linear dos dados matutinos, a reta rosa representa o ajuste linear dos dados vespertinos.

Uma diferença também visível entre os três gráficos, do peso do primeiro autovalor, do peso dos cinco primeiros autovalores e do peso dos 17 primeiros autovalores, foi que, à medida que foram contabilizados mais autovalores na soma a curva com concavidade negativa foi ficando mais plana e a diferença dos valores dos pesos entre a primeira curva com concavidade positiva e a segunda curva com concavidade negativa foi diminuindo.

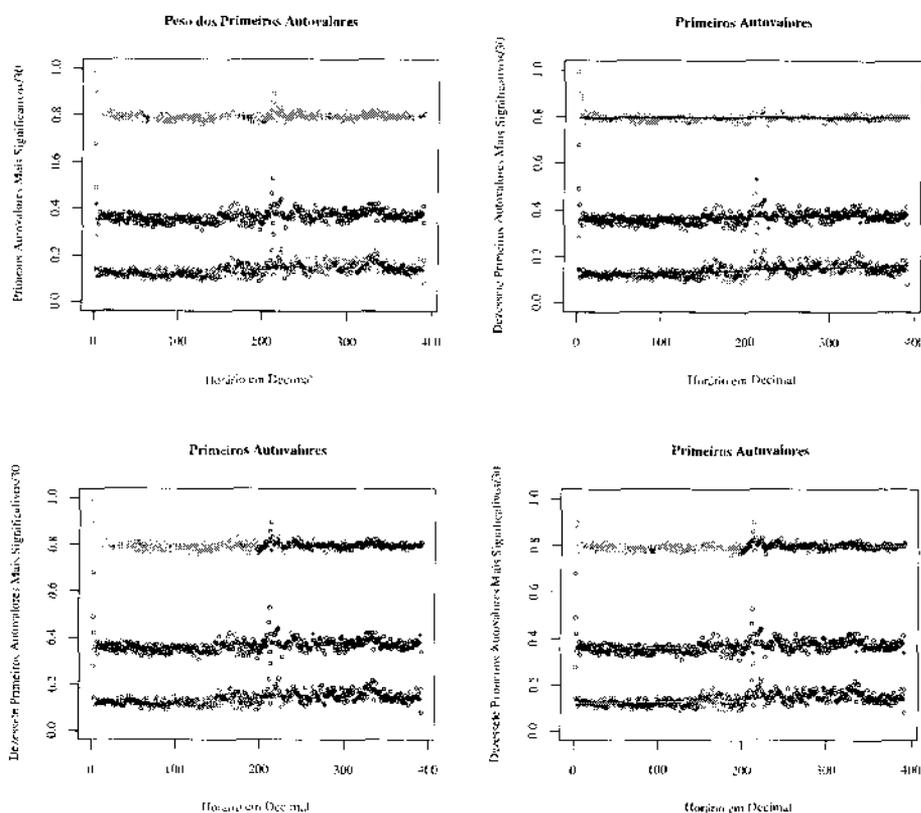


Figura 4.17: Pesos das somas dos dezessete primeiros (em verde e em verde escuro) e dos cinco primeiros (em azul e em azul escuro) Autovalores mais Significativos e do Peso do primeiro Autovalor mais Significativo (em vermelho e em vermelho escuro). As cores claras representam o período matutino enquanto as cores escuras representam o período vespertino.

Porém, há duas características em comum: todas elas apresentam um valor máximo entre a abertura e o fechamento do período intra-diário no horário de 13:02 horas e o comportamento da curva dos valores pertencentes ao período matutino é diferente do comportamento da curva do período vespertino.

## Capítulo 5

### Notas finais e Perspetivas Futuras

Nesta dissertação levei em consideração o estudo sobre a natureza não-estacionária do volume negociado de ações de referência — também chamadas de *blue chip* — que compuseram o famoso índice Dow Jones Industrial Average durante o período que ia do Primeiro Semestre de 2004 até o Primeiro Semestre de 2013. Essa quantidade é, a par das flutuações de preço, uma quantidade crucial para a descrição de um mercado financeiro — exemplo típico de sistema complexo — e desde há muito tomada como um indicador indireto de informação no mercado. Numa primeira fase, a análise se focou na evolução das propriedades estatísticas individuais ao longo do dia de negociação e como um eventual perfil diário tem evoluído ao longo do tempo, fruto dos ciclos econômicos, alteração nas abordagens à negociação entre outros fatores. Em seguida, procurou se entender a dinâmica coletiva de negociação analisando o comportamento das matrizes de correlação. Como conhecido há muitos anos, existe uma maior atividade no início e no fim das sessões de negócios na bolsa, fenômeno conhecido como

o “U-shape” dos mercados financeiros. O trabalho realizado procurou explorar as consequências dessa dinâmica de atividade e avaliar o quanto essa dinâmica é uma propriedade robusta dos mercados financeiros. Para tal, avaliei como as quantidades obtidas variam entre semestres contíguos.

Considerando a abertura de um mercado, assumi que esta pode ser vista como uma introdução de uma perturbação no sistema (relembre-se que existe um período de pré-mercado). Usando como análogo o comportamento pós-choque de vários sistemas complexos analisei a relaxação pós-abertura e verifiquei que esta pode ser descrita por uma lei de potência.

Este relaxamento é diferente do relaxamento exponencial apresentado por modelos de retorno à média apresentado por processos como de dinâmica diferencial estocástica do tipo Feller.

Ao analisar como esta relaxação se tem comportado ao longo dos anos verifiquei que até ao primeiro semestre de 2008 o expoente é 0.29 (compatível com o valor encontrado para pós-choques de volatilidade encontrados para o mesmo período [45]) a partir do segundo semestre do mesmo ano dá-se uma alteração estatisticamente significativa do valor do expoente que se eleva para 0.37. Esta alteração pode ser entendida como uma indicação do impacto da crise financeira de 2008 no comportamento dos agentes no mercado financeiro. Como consequência da modificação do valor do expoente tem-se que o tempo para o volume negociado seja metade daquele negociado no primeiro minuto, passou de dez para seis minutos, o que se entenderia como

uma indicação de que o mercado passou a funcionar mais rapidamente. Complementarmente, esta relaxação mais rápida deixa a entender que existem alterações na forma do U. Fazendo uso de funções simples, parti para o estudo de propriedades que designei de concavidade e a paridade em relação ao meio da sessão do perfil de volume médio negociado.

No que diz respeito à paridade, não foi possível encontrar um comportamento característico para além de se verificar que o seu valor médio calculado sobre as companhias é positivo com um pico pronunciado na crise de 2008. Ou seja, nessa altura houve por parte dos agentes uma opção clara de primeiramente esperar para ver e não deixar decisões para o dia seguinte onde notícias e medidas tomadas no pós-mercado pudessem afectar o portfólio. Consequentemente é viável pensar-se que a paridade pode ser um uma forma indirecta de se avaliar o sentimento de confiança ou de aversão ao risco por parte dos agentes de mercado. Algo que pode ser estudado com maior profundidade em trabalhos futuros, *e.g.*, condicionando as médias dos perfis (para cada empresa) a intervalos de variação de preço.

Já para a concavidade, que calculei como uma média da segunda derivada, verifiquei que as indicações deixadas pela análise do relaxamento do valor inicial se comprovam e que a concavidade do perfil intra-diário tem vindo a diminuir. Dito de uma forma gráfica pode se afirmar que cada vez mais o formato-U se aproxima de um formato-□.

Não obstante o resultado da concavidade parecer mais apelativo à primeira vista, a existência de uma paridade sistematicamente diferente de zero sugere que o comportamento dos

agentes nos períodos matutino e vespertino são diferentes, algo que se aprofundará em análises a momentos de ordem superior.

Desta forma assim como se estudou a reação à abertura, valerá a pena estudar a reação ao fechamento através de uma função semelhante e verificar o comportamento do expoente ao longo dos semestres, nomeadamente se existiu uma mudança pronunciada. Pelo comportamento da paridade e da concavidade, acredito que tal mudança tenha ocorrido com significância estatística.

Feita esta análise da relação do perfil intra-diário com as diferentes escalas de tempo, procurei entender a relação deste perfil com quantidades financeiras como o volume médio diário negociado, a volatilidade (relembrando as hipótese MDH e SAIH) e — seguindo o lema dos mercados financeiros de que é preciso volume para que os preços se modifiquem — com as flutuações de preço. Para os dois últimos casos não foi possível encontrar qualquer relação, porém no caso do volume médio diário verifiquei uma relação de proporcionalidade direta entre esta quantidade e a concavidade. Este fato tem uma consequência relevante: o aumento do volume diário negociado significa mais do que um eventual deslocamento do valor de base, corresponde a um aumento da concavidade, *i.e.*, na forma específica do perfil intra-diário.

O trabalho realizado para a análise foi estendido para os outros momentos de onde se retiram as seguintes relações:

- A variância apresenta um perfil em forma de U. Neste caso, não realizei análise relativa à

concavidade e paridade e respetiva evolução, mas tendo em conta a análise da paridade do perfil do valor médio avaliou-se a relação em a variância e o valor médio do volume em cada horário de negociação. Encontrei uma relação de proporcionalidade direta e procurei explorar a eventual diferença entre os períodos matutino e vespertino. A análise mostrou que a diferença da relação variância-média do volume entre os dois períodos acentua-se em situações de crise tornando-se a proporcionalidade matutina significativamente maiores do que a vespertina;

- A assimetria (*skewness*) apresenta um perfil  $\cap$  bastante achatado e equivalente ao obtido para o perfil da curtose das flutuações de preço. Esta mudança de perfil implica numa relação de proporcionalidade inversa entre esta quantidade e o valor médio. Novamente, recorrendo a uma análise de primeira ordem, podemos verificar que o comportamento da relação é relativamente uniforme à exceção dos últimos 3 semestres onde se verifica uma alteração significativa em que a proporcionalidade aumenta em valor absoluto. Já no que diz respeito à diferença entre os período matutino e vespertino verifica-se diferença que contudo parece manter-se insensível ao passar dos semestres e consequente alteração do cenário macroeconómico. Essa diferença fica mais clarividente numa aproximação de primeira ordem a  $\cap$  onde se verifica uma relação de um fator 2 entre os dois declives;
- A curtose tem um comportamento intra-diário que faz lembrar a forma  $\sim$ , *i.e.*, existe um decréscimo de curtose durante a primeira hora (aproximadamente), depois um crescimento até o meio da sessão para depois terminar significativamente abaixo e cerca de

3 vezes menor do que o valor inicial. Temos assim dois períodos de relaxação da curtose, o primeiro referente à abertura de mercado e a segunda relacionada com o pico pós-almoço. Notei que existe bastante oscilação de ambos os valores entre os semestres; contudo é perceptível que os valores médios são muito diferentes, na ordem de um fator de 10, favorável ao relaxamento no fecho. Todavia, é na representação direta da relação entre curtose e média que é possível verificar a diferença entre os períodos matutino e vespertino. Especificamente, ao passo que para o caso vespertino tem-se um comportamento monotonamente decrescente, no período matutino tem-se um claro comportamento parabólico. *Grosso modo*, dado a curtose medir eminentemente o afastamento em relação à Gaussiana — também chamada de distribuição Normal — acaba por fornecer uma medida da anormalidade contida na distribuição, ou em outras palavras, da surpresa que esta carrega. Por conseguinte, podemos entender que a negociação é marcada por dois picos de surpresa, a abertura e o pós-almoço sendo para o fim da sessão, apesar de existir um aumento do valor médio do volume e da sua variância, o comportamento do volume vai diminuindo a sua surpresa intrínseca.

Toda a análise anterior tratou o volume das empresas como uma variável independente e distribuída de forma semelhante. Sabendo que o mercado financeiro é um exemplo comum de complexidade, devemos pensar como se comporta a dinâmica do volume negociado ao longo de um dia estudando a sua matriz de correlação.

A esse respeito a análise que apresentei permite dizer que:

- Quando decomposta em autovalores, o maior autovalor é sistematicamente muito maior do que o valor dado pela teoria das matrizes randômicas indicando que se trata de um modo de mercado;
- Verifiquei que, no que diz respeito ao comportamento coletivo do volume negociado, também foi possível descortinar um perfil intra-diário marcado por um perfil apresentado pelo maior autovalor que se destaca claramente dos demais. Curiosamente, este modo de mercado é bastante mais fraco do que aquele encontrado para as flutuações de preço, dado os pesos dos maiores autovalores, no determinante de cada uma das matrizes de correlação, apresenta um fator de relação da ordem de 2. Utilizo o advérbio curiosamente, pois é sabido que as autocorrelações de flutuações de preço caem rapidamente para o nível de ruído ao passo que o volume negociado apresenta autocorrelações que decaem lentamente;
- A inspeção visual das matrizes de correlação mostra duas alturas da sessão de negociação em que existe um maior número de pares de empresas com valores elevados de correlação: o início da sessão — e por inerência o início do período matutino — e as 13 horas que marcam o início do período vespertino. Por análise estatística verificou-se que esta diferença de comportamentos é significativa. Deve recordar-se que esses períodos são aqueles em que a curtose individual também atinge maiores valores. Entendo a conjuga-

ção dos fenômenos da seguinte maneira: no início da manhã e no início da tarde são as alturas em que surgem valores mais elevados (ou coloquialmente, mais surpreendentes); no entanto, esses valores elevados ocorrem com grau relevante de coordenação.

- Ainda nesta inspeção visual é de notar a substancial diferença de comportamentos entre a abertura de mercado — com fortes correlações — e o fim de mercado, em que existem muito poucas correlações significativas. Novamente, pode se ligar este comportamento ao comportamento da curtose; no final da sessão existe uma diminuição da curtose, ou seja, uma maior independência entre os valores negociados conduz a uma maior “normalidade” dos valores de  $v$ .
- Os demais autovalores mais significativos (segundo e terceiro) se mostram relativamente constantes ao longo do período de negociação à exceção de aumentos relevantes no início e no fim da sessão assim como o pico de almoço para o segundo maior valor; este comportamento difere daquele verificado para as flutuações de preço para as quais se verifica um decaimento pronunciado desses autovalores.
- Analisando os autovetores, pode se concluir que o primeiro autovetor é relativamente próximo da uniformidade relativamente ao peso das empresas durante a maior parte do tempo. Mais especificamente, o produto interno entre o maior autovalor cresce entre 0.9 e 0.95. Nos primeiros minutos apresenta valores entre 0.75 e 0.8 e um no início da tarde no pós-almoço em que cai novamente para um valor da ordem de 0.8;

- Assim, genericamente podemos entender que ao longo do dia as empresas se vão tornando mais independentes umas às outras. As posições no início são tomadas tendo em conta a informação existente para as empresas e tendo em conta diversas relações de interdependência. Ao longo da manhã o impacto da informação sobre uma empresa na outra vai relaxando voltando a haver um ajuste no início da tarde para novamente se dissipar à medida que a sessão se aproxima do fim.
- Sobre a não-estacionariedade, ao longo dos semestres verifiquei uma relativa constância do primeiro autovetor com flutuações sensíveis nos três semestres após a crise de 2008 à exceção do fim do período intra-diário de negociação onde a coerência para com o primeiro semestre cai significativamente para níveis próximos de ruído.
- Procurando descrever o comportamento ao longo do dia temos que a tomada de posições e constituição de carteira no início da sessão é regida por algumas empresas cuja influência se vai diluindo ao longo do dia até que no fim do dia a situação se tornou muito próxima da uniformidade, com a maioria das flutuações centradas no nível do ruído.
- Comparando os conjuntos diários de segundos e terceiros autovalores ao longo dos semestres foi possível anotar bastante consistência dos mesmos. Já no que se refere ao modo de mercado notou-se uma alteração substancial após a crise financeira de 2008.

Este conjunto de resultados indicam assim a relevância da crise de 2008 e o seu papel na alteração da dinâmica do mercado. Apesar de não ser possível afirmar categoricamente se essas

alterações se prendem com uma nova atitude dos agentes tomada por livre vontade ou fruto de nova legislação, a percepção é que a primeira se sobrepõe dado não serem conhecidas mudanças radicais nas regras de mercado.

## 5.1 Questões futuras

O trabalho aqui realizado fornece um vasto número de indicações sobre a validade do estudo de propriedades intra-diárias quer a nível empírico quer a nível teórico. Listamos aqui alguns casos que poderão ser de interesse:

Sob o ponto de vista académico, pode logo se pensar no desenvolvimentos ou aprimoramento de modelos de agentes que permitam a reprodução destas propriedades. Tal poderá passar pela introdução de um relógio que afete o comportamento dos agentes, nomeadamente a sua propensão para atividade de negociação.

Pensando em questões práticas estes resultados podem ser úteis para a uma melhor gestão de carteiras de ativos. Deve se notar que os diferentes comportamentos entre os perfis intra-diários das flutuações de preços e de volume negociado apontam para a existência de um perfil intra-diário para quantidades importantes em negociação, nomeadamente para os provedores de liquidez, tal como acontece com a profundidade de mercado da ação, *i.e.*, a quantidade de ações que é necessário negociar para que o preço se mova de uma unidade.

No espírito da problemática entre as abordagens MDH e SAIH pode se estudar:

- o comportamento intra-diário e não estacionário da volatilidade e a sua dependência na definição utilizada;
- as correlações intra-diárias entre volatilidade e volume negociado.

Adicionalmente vale a pena olhar para o chamado efeito de alavancagem e comparar a estrutura diária da correlação entre flutuações de preço e volatilidade assim como de flutuações de preço com o volume negociado. Esta análise deverá ser estendida a diferentes mercados com o intuito de se avaliar se existe um comportamento universal, o eventual impacto da capitalização e da liquidez ou se estas propriedades intra-diárias — bem como a sua não estacionariedade — pode servir para classificar mercados quanto à sua liquidez e estado de desenvolvimento.

Ainda na questão de matrizes aleatórias se sabe que o volume negociado é no longo termo bem descrito por uma distribuição genérica da distribuição  $F$ , também conhecida como distribuição  $q$ -Gamma. Seria relevante fazer um estudo no contexto de matrizes aleatórias que trata-se este tudo de variáveis na linha do que já foi feito com variáveis que seguem distribuições  $q$ -Gaussianas e log-normal.

Para concluir, pode se falar que a literatura se tem centrado no estudo das correlações para descrever o comportamento dinâmico dos sistemas financeiros. No entanto, as correlações são de fato o efeito da existência de interações. Deste modo é colocar-se a seguinte questão: que tipo de interações — ou mais humildemente, que intensidade de determinada interação — existem entre as diferentes companhias para que as flutuações de preço (volatilidade/volume ne-

gocido) apresentem aquelas propriedades estatísticas? Um método tentador para a obtenção desta resposta corresponde à aplicação de métodos de máxima entropia associando os valores da correlação a multiplicadores de Lagrange.

Contudo, as dependências não se esgotam na matriz de correlação. É sabido que o fator de correlação de Pearson usado no cálculo das matrizes de correlação mede sobretudo dependências lineares e que para o cálculo de dependências de ordem superior a abordagem mais adequada implica a utilização de formas entrópicas (ou derivadas). Um dos mais fortes candidatos é a criação de matrizes  $Q^{op}$  cujas entradas são os índices entrópicos ótimos calculados a partir de entropia mútua generalizada-normalizada. Este trabalho implica o desenvolvimento físico-matemático das propriedades deste tipo de matrizes, que podem envolver a sua relação com teoria de cópulas multidimensionais, bem como o desenvolvimento de estimadores de entropia mútua generalizada-normalizada que mitiguem os efeitos de tamanho finito das séries.

# Apêndice A

## Descrição da fórmula da volatilidade

Usando o movimento aleatório para descrever a dinâmica dos preços das ações, Parkinson [37] comparou a constante de difusão da Física Estatística com a variância dos movimentos dos preços das ações supondo que uma partícula pontual se move unidimensionalmente e é submetida a um passeio aleatório com uma constante de difusão  $D$ .

Então, a probabilidade de encontrar tal partícula num intervalo  $(x, x + dx)$  em um instante  $t$ , sabendo que no instante inicial  $t = 0$  a partícula estava em  $x_0$  é

$$P(x, t) = \frac{dx}{\sqrt{2\pi Dt}} \exp\left[-\frac{(x - x_0)^2}{2Dt}\right].$$

Ao se comparar com uma distribuição normal, é possível constatar que  $D$  trata-se da variância do deslocamento da partícula no intervalo  $x - x_0$  no primeiro instante de tempo após o inicial.

Sendo  $d_i$  o deslocamento da partícula durante o intervalo de tempo  $i$ , ou seja,  $d_i = x(i) -$

$x(i-1)$ , com  $i = 1, 2, \dots, n$ , tem-se a estimativa para a constante de difusão dada por

$$D_x = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$$

com uma média de deslocamento dado por

$$\bar{d} = \frac{1}{n-1} \sum_{m=1}^n d_m.$$

Perante esta abordagem, para a transformada do logaritmo do preço, mudanças sobre qualquer intervalo de tempo segue uma distribuição normal [41] de média zero e variância proporcional ao comprimento do intervalo exibindo integrais de caminho contínuas. Porém, não é assumido que essas integrais podem ser observadas em qualquer ponto devido as restrições as quais as negociações só ocorrem, frequentemente, em pontos discretos no tempo e os mercados de câmbios fecham por certo períodos de tempo.

Sendo  $(S_1, S_2, \dots, S_{n+1})$  a serie de preços de ações cotadas em intervalos iguais de unidade de tempo, ao igualarmos

$$r_i = \ln \frac{S_{i+1}}{S_i}$$

tem-se uma taxa de retorno nula e um número de dias de transações anuais de 252 e uma taxa  $n$  de retornos num intervalo de tempo  $i$ . Assim, o estimador anual Close-Close é simplesmente o desvio padrão clássico da distribuição

$$\sigma_{cc} = \sqrt{252 \cdot \frac{1}{n} \cdot \sum_{i=1}^n r_i^2}$$

que também passa ser a raiz quadrada da definição da constante de difusão  $D$ .

Com base no estimador de Parkinson, e usando o pressuposto que as ações seguem um movimento browniano geométrico, Garman e Klass [41] modelou um estimador de volatilidade que levou em conta os preços de abertura e fechamento das ações além dos preços máximos e mínimos levados em consideração por Parkinson. Este estimador levou o nome de Estimador de Volatilidade High-Low-Open-Close e tem a forma de

$$\sigma_{HI,OC} = \sqrt{\frac{252}{n} \sum_{i=1}^n \left[ \frac{1}{2} \left( \ln \frac{H_i}{L_i} \right)^2 - (2 \ln 2 - 1) \left( \ln \frac{C_i}{O_i} \right)^2 \right]}.$$

Este é muito mais eficiente que o Estimador Close-Close, sendo necessário sete vezes menos observações para se obter o mesmo resultado [41] além de a distribuição amostral do coeficiente de volatilidade obtida através deste ser mais apertada.

Deve se enfatizar que hoje em dia é bem conhecido que os preços não realizam um caminho Browniano, mas executam processos difusivos que apresentam flutuações não-Gaussianas. Desta forma um objeto de trabalho interessante no futuro é a determinação de uma fórmula para a volatilidade que tenha em consideração este dado empírico.

## **Apêndice B**

### **Cenário Econômico entre 2004 até 2013**

Conhecida como crise sub-prime, a crise financeira caracterizada pela excessiva especulação sobre ativos de alto risco financiados por empréstimos bancários teve seu início com a transação de títulos hipotecários residenciais nos Estados Unidos e acabou se transformando numa crise financeira de grande proporção para toda a economia americana estendendo-se também à economia mundial devido aos laços econômicos entre os EUA e os demais países.

A gravidade da crise econômica americana foi consequente de uma sucessão de elementos inéditos, até então, para a economia mundial.

Até o início da década de 1970, os preços dos imóveis permaneciam praticamente constantes. Durante a década de 1970, esses preços duplicaram em efeito da alta inflação monetária ocorrida naquela década. Em 1980 houve uma mini-bolha que eclodiu no início da década de 1990 e a partir de 1993, no início do governo Clinton, os preços voltam a subir continuamente, acelerando vertiginosamente a partir de 2001 até entrar em colapso em 2008.

Na década de 1990, foram intensificadas as duas políticas governamentais voltadas exclusivamente para o setor imobiliário com o objetivo de aumentar o número de proprietários de imóveis.

De um lado havia duas empresas privadas criadas pelo Congresso americano, a Federal National Mortgage e a Federal Home Loan Mortgage Corporation, que se voltavam exclusivamente para o mercado imobiliário para garantir a liquidez ao mercado de hipotecas através da compra dos empréstimos feitos pelos bancos para compradores de imóveis.

Sempre que um banco concede um empréstimo o valor do seu ativo (valor do empréstimo) aumenta e, conseqüentemente, o valor do seu passivo (dinheiro fornecido pelo banco) também. Por uma questão de regulamentação bancária do Banco Central americano e do Banco da Basileia, existe um limite para o crescimento destes ativos de tal forma que os ativos devem manter uma proporção máxima em relação ao patrimônio líquido do banco.

Porém, essas duas empresas compravam esses empréstimos, ou títulos hipotecários, permitindo que os bancos continuarem a fazer novos empréstimos. Porém, como a relação passava a ser entre essas duas empresas e os tomadores de empréstimos, se estes não pagassem as empresas não teriam fluxo de caixa, remarcariam os títulos imobiliários para um valor zero afetando severamente seus patrimônios líquidos enquanto que os bancos não eram afetados.

Essas duas empresas então tinham duas opções: manter em suas carteiras os empréstimos que compravam do banco ( de forma a aferir as receitas) ou vendiam esses empréstimos para

investidores ao redor do mundo, chamados de títulos lastrados em hipotecas. Este processo acabou gerando mais empréstimos do que o normal de modo que este direcionamento de recursos para o mercado imobiliário inflasse os preços dos imóveis.

Para atrair um contínuo financiamento de investidores, Freddie Mac e Fannie Mae tiveram uma linha especial de crédito junto ao Tesouro Americano no valor de 2.25 bilhões de dólares, pois deste modo, caso os empréstimos não fossem pagos elas seriam socorridas pelo governo americano.

Do outro lado havia a política governamental de ação afirmativa para empréstimos. Além da Fannie e Freddie, agencia governamentais de vários tipos eram utilizadas para reduzir os padrões de empréstimos pressionando os bancos a fazerem empréstimos mais arriscados em nome da "igualdade racial" e seriam processados por discriminação e racismo caso recusassem.

Após um estudo feito pela Federal Reserve de Boston ter afirmado que existiam claras evidências de que a minorias recebiam menos empréstimos que os brancos, a lei Community Reinvestment Act, criada na década de 1970, é amplamente revigorada no governo de Clinton deixando os bancos destinados à processos por discriminação caso esses não emprestassem dinheiro para um volume alto da minoria.

A partir de 1995, o governo americano passou a pressionar os bancos a fazerem empréstimos arriscados sem verificarem o histórico de crédito, poupança do tomador de empréstimo ou até mesmo sem comparar a magnitude do pagamento da hipoteca frente à sua renda. Em soma,

o Banco central americano afirmou aos bancos que a simples participação destes tomadores em programas de aconselhamento de crédito seria suficiente para provar a capacidade de um tomador de empréstimo de baixa renda honrar suas dívidas.

Na década de 2000, o aumento na demanda dos imóveis estimulado pelo fácil acesso ao financiamentos, imposto pelo governo, gerou um contínuo aumento nos preços dos imóveis de forma a atrair especuladores para o mercado imobiliário: tornou-se comum um indivíduo tomar um empréstimo para comprar uma casa, e depois de um curto período de tempo revende-la por um preço muito maior ou, aqueles que não compravam com intenção de revender, utilizavam os imóveis como um caixa eletrônico de modo que sempre que o imóvel valorizava, o indivíduo ia ao banco e negociava um novo empréstimo para gastar com bens de consumo usando o novo valor da sua casa como colateral.

Em 2008, a Fannie e a Freddie detinham justas quase metade das hipotecas do país e praticamente 75% das hipotecas recém-concedidas. O que tornou possível essa contínua especulação e a compra e revenda dos títulos lastrados em hipotecas foram os fundos concedidos pelos bancos de investimentos que compravam pacotes de ativos das Fannie e da Freddie mesmo sabendo que haviam ativos bons, de mutuários com bom histórico de crédito, e ativos ruins (mutuários sem nenhum histórico de crédito, porque as agências de classificação de risco deram classificação máxima AAA para estes ativos.

Com todo o novo dinheiro criado pelo Banco Central americano (Fed) que foi multiplicado

por meio do processo de reservas fracionárias pelo sistema bancário junto com a facilidade da concessão de empréstimos dado por critérios gerados por políticas governamentais e a classificação máxima dada pelas as agências de classificação de risco aos títulos imobiliários, fizeram com que grandes bancos americanos e estrangeiros comprassem esses títulos em grande volume permitindo que Fannie e Freddie continuassem a dar liquidez ao mercado imobiliário de forma a perpetuar a bolha.

Porém foi o Fed que tornou possível todo o sentimento artificial de prosperidade do setor imobiliário pois foi o dinheiro por ele criado quem proveu o principal estímulo ao aumento exponencial de preços dos imóveis. Todo esse processo começou no final dos anos 2000 após o estouro da bolha das empresas de tecnologia. A Fed, temendo a recessão, aumentou a injeção de dinheiro no sistema bancário para gerar uma redução de juros fazendo que essas caíssem de 6,5% para 1% em 2001, ficando assim até o segundo semestre de 2004.

O resultado deste aumento na base monetária foi bancos cheios de dinheiro para ser emprestado, aumentando o crédito total em 100% de 2000 à 2008. A partir do segundo semestre de 2004, a economia americana já dá sinal de ter se recuperado da recessão de 2001 e, assim, a FED começa a diminuir a quantidade de dinheiro injetado no sistema bancário fazendo com que os juros comecem a subir. Os juros das hipotecas com taxas ajustáveis vai de 3.5% no início de 2004 para 6% no segundo semestre de 2006 e os juros vão de 5,5% à 7% neste mesmo período.

Em consequência ao aumento de juros, a demanda por imóveis cai porém como havia um

acentuado aumento na oferta de imóveis, os preços dos mesmos começam a cair.

A percepção de que os preços não iriam mais aumentar fez com que todas as pessoas que tinham comprado imóveis para especular parassem de pagar suas hipotecas porque, como elas não precisaram dar nenhum pagamento de entrada, abandonando os imóveis não perderiam nada.

Já para aqueles não especuladores, o motivo de se ter deixado de pagar as hipotecas era que os juros haviam se tornando tão altos que era impossível pagar as prestações.

Desta maneira a inadimplência começou em 2006, disparou em 2007 e explodiu em 2008.

O aumento dos incumprimentos fez com que os bancos de investimentos que haviam comprado títulos lastrados em hipotecas não mais auferissem essa receita, fazendo com que os ativos reduzissem a zero. A redução nos ativos sem uma redução conjunta nos passivos fez com que esses bancos sofressem uma redução enorme do seu patrimônio líquido fazendo com que esses parassem de conceder novos empréstimos gerando o congelamento no mercado de crédito. A crise passa a se tornar mundialmente visível em 2007 quando o banco francês BNP Paribas anunciou que havia suspenso saques em dois dos seus fundos que haviam investido volumosamente em títulos lastrados em hipotecas americanas. A partir daí foram vários bancos e seguradoras informando que não tinham mais dinheiro passando a pedir intervenções da FED e do Banco Central da Inglaterra, por exemplo.

A FED então, depois de emprestar 125 bilhões de dólares à seguradora AIG em troca de

80% da empresa, passa a comprar todos os títulos hipotecários em posse dos bancos limpando o balancete dos bancos e fazendo com que a base monetária explodisse transformando-se em reservas de dinheiro mantidas pelos bancos depositadas junto ao Fed. Esta nova política adotada pelo FED resultou em um subsídio para o sistema bancário de forma que seus lucros fossem mantidos e os prejuízos fossem socializados.

Uma segunda crise, conseqüente da crise do sub-prime, foi a crise das dívidas soberanas na qual o grande acúmulo da dívida governamental fez estourar a capacidade de endividamento das nações desenvolvidas causando uma grande turbulência financeira provocada pelo medo de essas nações decretarem o incumprimento da dívida.

A crise americana expôs as políticas fiscais insustentáveis dos países na Europa e no mundo.

Em conseqüência disso, a Grécia, que foi um dos países que não conseguiu reformas fiscais, decretou no final de 2009 que não tinha mais condições de pagar suas contas. A partir disso, os investidores passaram a exigir mais rentabilidade sobre os títulos da Grécia e de outros países endividados da região. Em 9 de maio de 2010, os países membros da zona do euro criaram o Fundo Europeu de Estabilidade Financeira para facilitar o fornecimento de empréstimos de emergência aos países em dificuldade financeira.

Em 2010, o mundo encontrava-se numa guerra cambial. A desvalorização de moedas é uma medida adotada por vários países para tentar resolver seus problemas internos. No final de 2010, o Banco Central dos Estados Unidos injetam 600 bilhões de dólares na economia.

Concomitantemente, a Europa vivia a crise do euro e de insolvência na Grécia, Irlanda e Portugal e, associado à possibilidade de que o Reino Unido, a Espanha e a Itália não conseguissem honrar seus compromissos, o mercado financeiro sobre um grande abalo.

Em 2011, os Estados Unidos enfrenta um problema subsequente da tentativa de evitar uma crise econômica profunda em 2008: a enorme ampliação da dívida pública. Este ficou conhecido como crise do limite da dívida pública dos EUA de 2011 e foi um longo período de negociação e discussão no Congresso dos Estados Unidos sobre se o país deveria aumentar o limite da dívida pública e quais políticas fiscais deveriam ser aplicadas após este aumento. Enquanto isso o país também tentava negociar um aumento nos impostos e um corte nos gastos.

Esta crise do limite de dívidas dos EUA foi o estopim da segunda crise e, durante o período de negociação, a especulação internacional sobre a solvência americana cresceu de maneira a sensibilizar o mercado financeiro. Em 31 de agosto de 2011, a crise mais imediata dos EUA terminou quando um acordo entre os legisladores do Partido Republicado, que eram opostos a quaisquer aumentos dos impostos, e os legisladores do Partido Democrata, que defendiam que o aumento dos impostos era tão necessário quanto um corte nos gastos, conseguiu elevar o limite de gastos.

Porém, o mercado não reagiu positivamente e nos dias que se sucederam, a maior parte das bolsas de valores do mundo todo fecharam em forte queda. Diante disso, a agência de classificação de riscos Standard & Poor's pela primeira vez rebaixou a nota de dívida pública

dos Estados Unidos de AAA para AA+ em consequência à crescente dívida e ao grande déficit de orçamento. Em resposta, as bolsas de valores mundiais calcularam perdas altíssimas. Além disso, as economias da zona do euro, que já estava em profunda recessão, em agosto de 2011 cresceram menos do que o previsto. Parte dos mercados acionários mundiais e de ações de bancos recuperaram-se após o rebaixamento da nota da dívida dos EUA .

Em dezembro de 2011, o Banco Central Europeu disponibilizou uma linha de crédito com taxas baixíssimas à disposição dos bancos em dificuldade na zona do euro. Este programa, nomeado de Operação de Refinanciamento de Longo Prazo, teve como objetivo manter a capacidade dos bancos europeus de conceder empréstimos.

Em fevereiro de 2012, mais recursos são disponibilizados para este programa pelo Banco Central Europeu.

No Primeiro Semestre de 2013, os países europeus em crise como a Grécia, Itália Espanha e Portugal, experimentaram um crescimento quase nulo e taxas altíssimas de desemprego.

Nos EUA, teve-se melhora do desemprego mas tem como problema central da crise por ele vivida ainda o elevado déficit fiscal, apesar de ter havido uma pequena redução.

Além disso, a economia da China entra numa fase de transição onde ela passa a investir menos e a consumir mais produtos internos fazendo que seu PIB crescesse de forma menos acentuada.

## **Apêndice C**

# **Gráficos típicos para as empresas individuais (Citigroup Inc. e Microsoft Corp.)**

Apresentam-se em seguida alguns gráficos para empresas individuais de análise intra-diária de momentos estatísticos cujo comportamento médio é apresentado no Capítulo 3.

A Microsoft Corp. e o Citigroup Inc. foram escolhidos pois suas ações são negociadas nos dois mercados de onde usamos dados e por serem aquelas que apresentam comportamentos relevantes.

Refira-se que em grande parte do período analisado a Microsoft Corp. foi a empresa com maior capitalização bolsista do mundo, sendo que em nosso semestre de referência esse posto foi ocupado pelo Citigroup Inc..

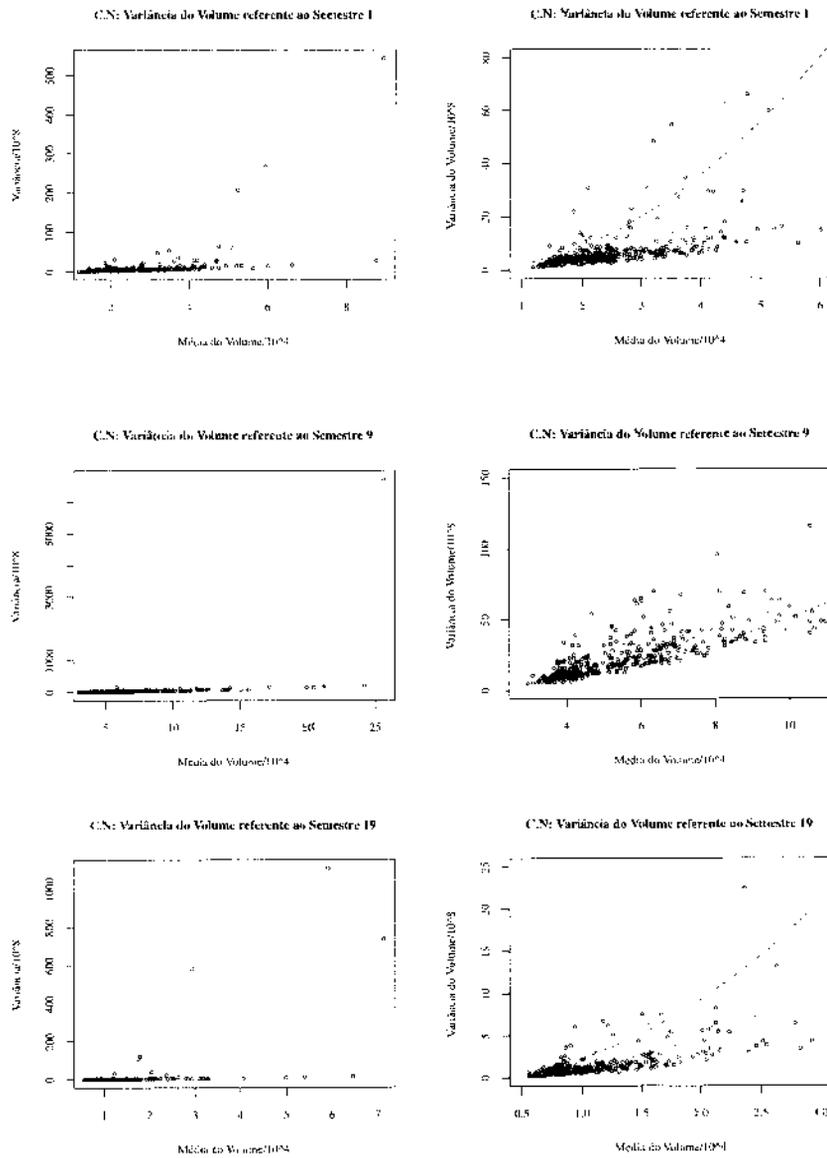


Figura C.1: Variância do volume em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos.

## C.1 Assimetria

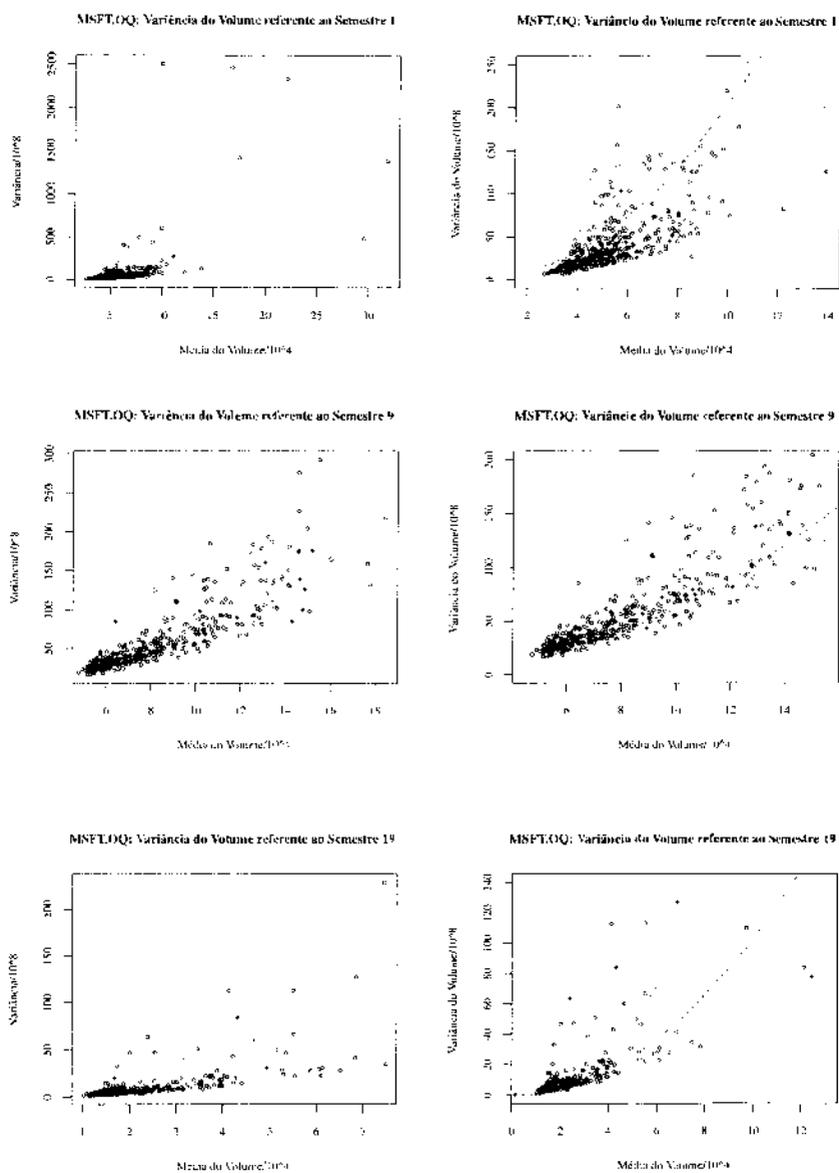


Figura C.2: Variância do volume em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos.

## C.2 Curtose

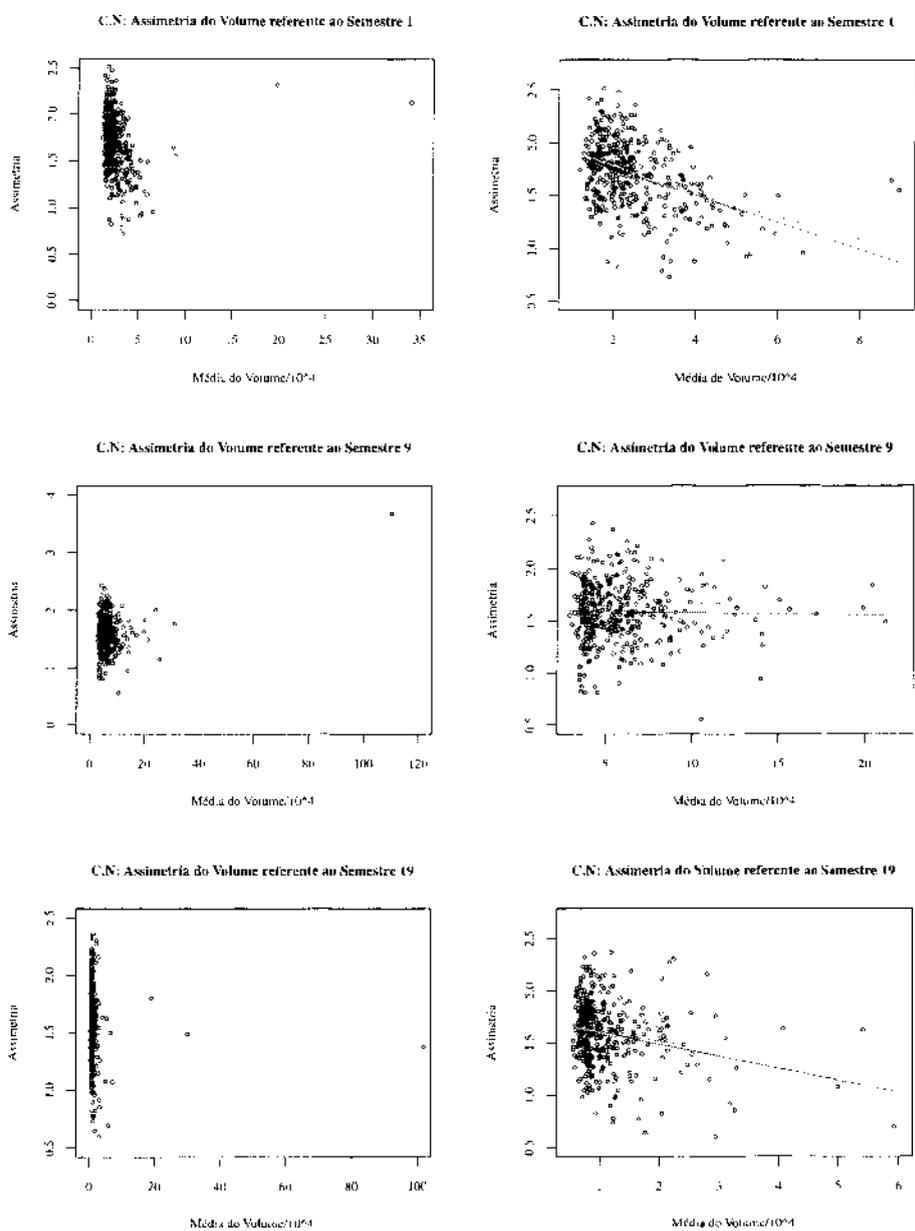


Figura C.3: Assimetria do volume transacionado em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos.

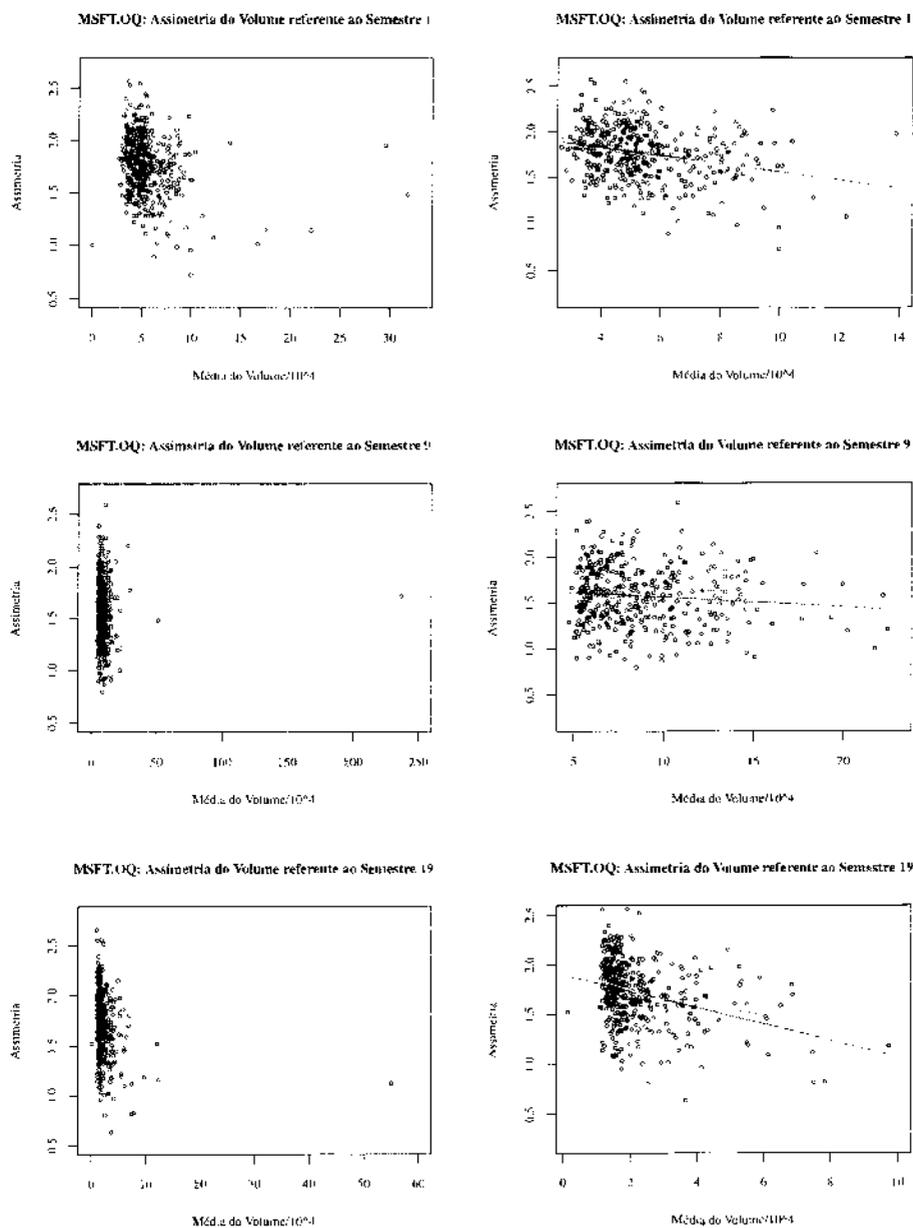


Figura C.4: Assimetria do volume transacionado em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos.

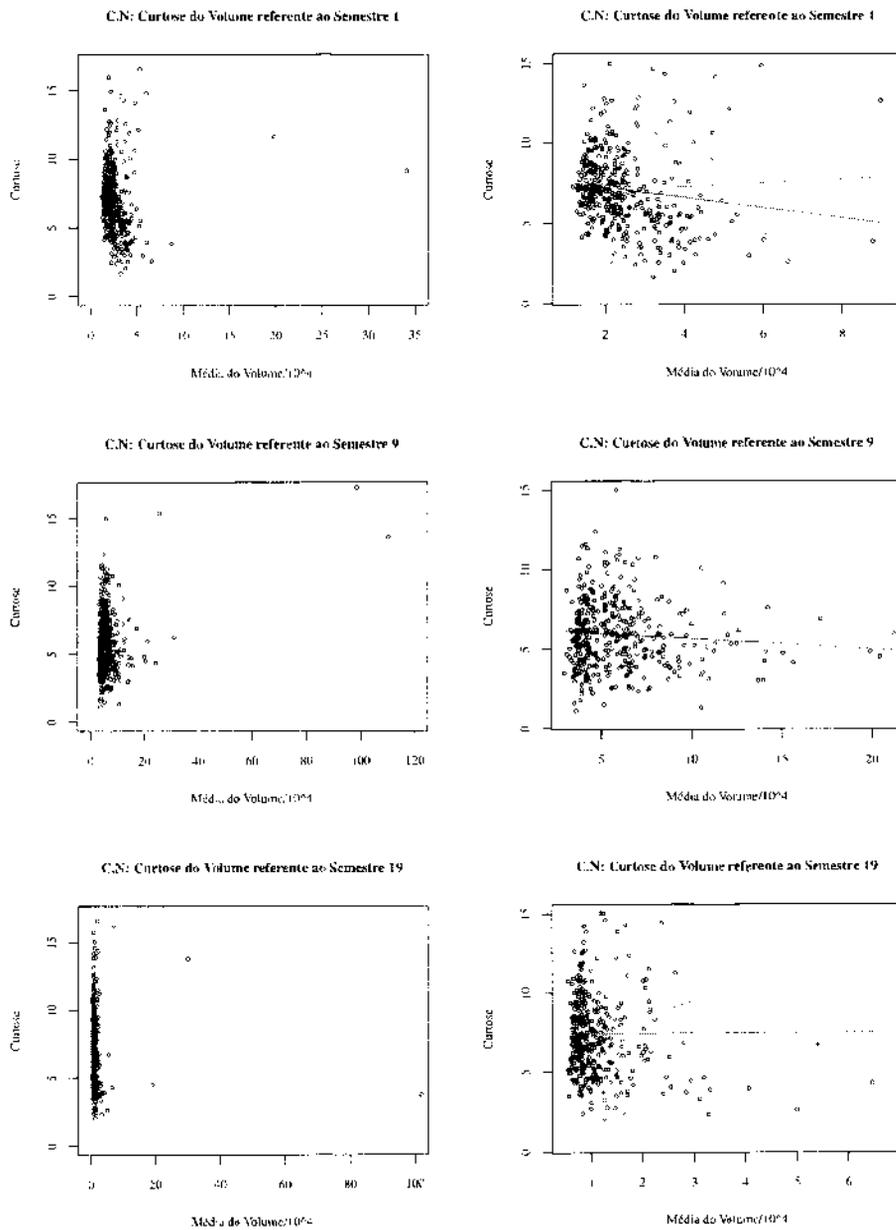


Figura C.5: Curtose do volume transacionado em função da média para o Citigroup. As retas tracejadas representam ajuste a lineares aos dois períodos.

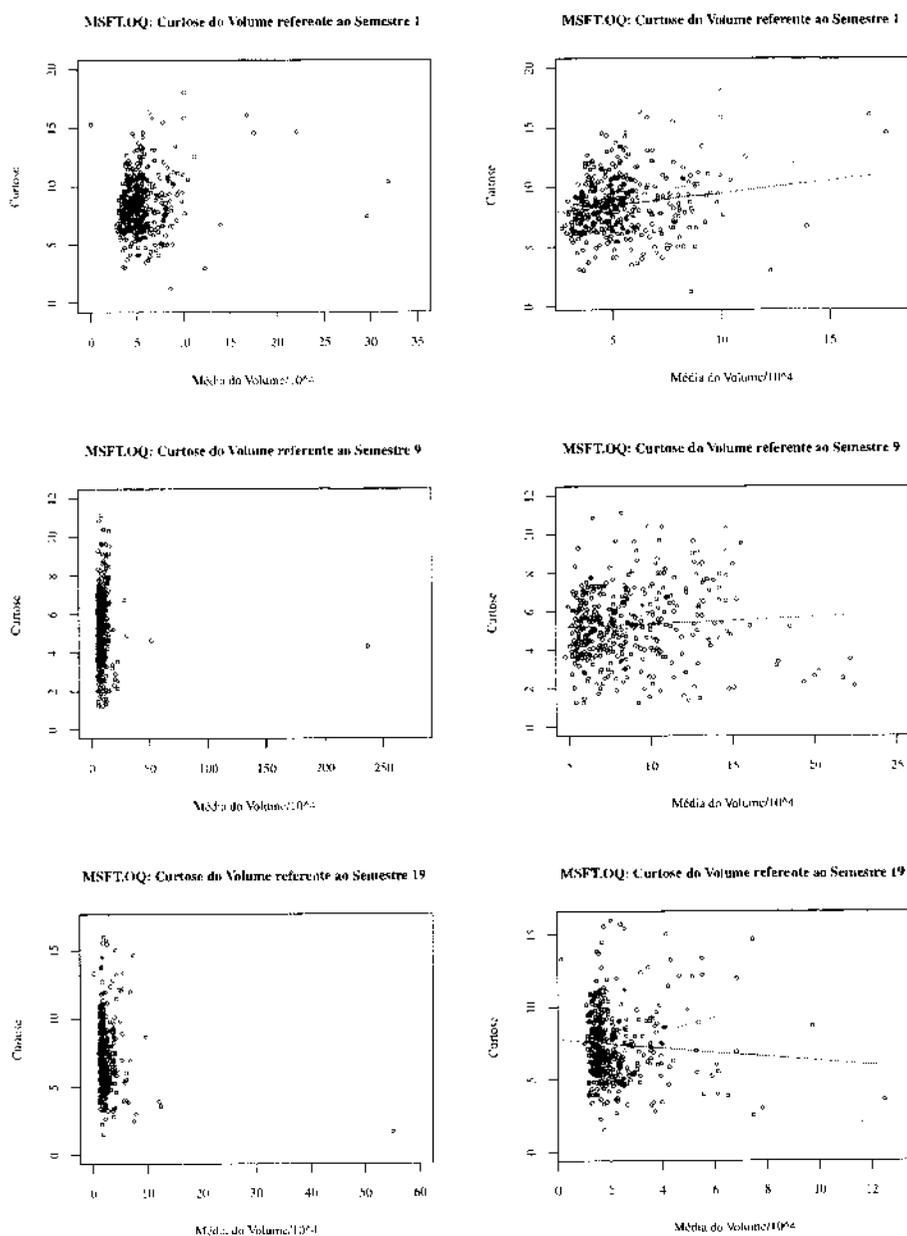


Figura C.6: Curtose do volume transacionado em função da média para a Microsoft. As retas tracejadas representam ajuste a lineares aos dois períodos.

# Bibliografia

- [1] F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, Singapore, 1985)
- [2] A. Admati and P. Pfleiderer, *Rev. Financ. Stud.* **13**, 40 (1988).
- [3] P. K. Clark, *Econometrica*, **41**, 135 (1973).
- [4] H.J. Jensen, *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems* (Cambridge University Press, Cambridge, 1988);  
P. Bak, *How Nature Works: The Science of Self-organized Criticality* (Copernicus New York, 1997);  
M.E.J. Newman, *Proc. R. Soc. London B*, **263**, 1605 (1996).
- [5] T. Copeland, *J. Financ.* **31**, 1149 (1976).
- [6] S.M.D. Queirós, *Trading volume in financial markets: an introductory review*, 2016.
- [7] R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999);  
J.P. Bouchaud and M. Potters, *Theory of Financial Risks: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, 2000);  
*Empirical Science of Financial Fluctuations: The Advent of Econophysics*, edited by H. Takayasu (Springer-Verlag, Berlin, 2002);  
J. Voit, *The Statistical Mechanics of Financial Markets* (Springer-Verlag, Berlin, 2003);  
S.M. Duarte Queirós, C. Anteneodo and C. Tsallis, Power-law distributions in economics: a nonextensive statistical approach, in: *Noise and Fluctuations in Econophysics and Fi-*

- nance, edited by D. Abbot, J.P. Bouchaud, X. Gabaix and J.L. McCauley, Proc. of SPIE **5848**, 151 (2005).
- [8] M. Dacorogna, R. Gençay, U. Müller, R. Olsen, and O. Pictet, *An Introduction to High-Frequency Finance* (Academic Press, London, 2001);
- [9] S. Camargo, S.M. Duarte Queirós and C. Anteneodo, Eur. Phys. J. B **86**, 159 (2013).
- [10] P. Rocha, F. Raischel, J.P. Boto and P.G. Lind, Uncovering the evolution of non-stationary stochastic variables: the example of asset volume-price fluctuations (2015) Preprint. arXiv:1510.07280 [q-fin.ST].
- [11] R. A. Wood, T. H. McInish and J. K. Ord, J. Finance **40**, 723 (1985).
- [12] P. C. Jain and G.-H. Joh, J. Finan. Quant. Anal. **23**, 269 (1988).
- [13] R. Allez and J.-P. Bouchaud, New J. Phys. **13**, 025010 (2011).
- [14] S. Camargo, S.M. Duarte Queirós and C. Anteneodo, Phys. Rev. E **84**, 046702 (2011).
- [15] T. Andersen and T. Bollerslev, J. Empir. Financ. **4**, 115 (1997).
- [16] P.I. Good, *Resampling Methods: A Practical Guide to Data Analysis* (Springer Science+Business Media, New York, 1999)
- [17] T.-H. Kim and H.L. White, Finance Res. Lett. **1**, 56 (2004).
- [18] E.P. Wigner, Ann. Math. **53**, 36 (1951).
- [19] E.P. Wigner, Proc. Cambridge Philos. Soc. **47**, 479 (1951).
- [20] E.P. Wigner, *Conference on Neutron Physics by Time-of-Flight* (Gatlinburg, Tennessee, 1956).
- [21] F.J. Dyson, J. Math. Phys. **3**, 1191 (1962).
- [22] F.J. Dyson, J. Math. Phys. **4**, 713 (1963).
- [23] M.L. Mehta, *Random Matrices* (Academic Press, Boston, 1991).

- [24] T. Guhr, A. Muller-Groeling e H.A. Weidenmuller, Phys. Rep. **299**, (1998).
- [25] A. Tulino e S. Verdú, Foundations and Trends® in Communications and Information Theory, **1**, 1 (2004).
- [26] J. Baik, G. Ben Arous, and S. Peche. Ann. Probab. **33** (2005).
- [27] G. Biroli, J.-P. Bouchaud e M. Potters, Eurohys. Lett. **78**, 10001 (2007).
- [28] J. Baik, G. Ben Arous, and S. Péché, Ann. Probab. **33**, 1643 (2005).
- [29] P. Jäckel and R. Rebonato, *The most general methodology to create a valid correlation matrix for risk management and option pricing purposes*, Quantitative Research Centre of the NatWest Group, 1999.
- [30] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley, Physica A **287**, 374 (2000).
- [31] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, T. Guhr, and H.E. Stanley, Physica Review E **65**, 066126 (2002).
- [32] P. Embrechts, F. Lindskog and A. McNeil, Modelling Dependence with Copulas and Applications to Risk Management, *Handbook of Heavy Tailed Distributions in Finance*, edited by S.T. Rachev (Elsevier/North-Holland, Amsterdam, 2001).
- [33] J. Hasbrouck, *Empirical market microstructure: The institutions, economics and econometrics of securities trading* (New York: Oxford University Press, 2007).
- [34] K.C. Chan, W.G. Christie, and P.H. Schultz, J. Business **68**, 35 (1995).
- [35] T. Schneeweis, J.R. Woolridge, J. Finan. Quant. Anal. **14**, 939 (1979).
- [36] A. Chakraborti, M. Patriarca, and M.S. Santhanam, *Financial time-series analysis: A brief overview*, edited by A. Chatterjee and B.K. Chakrabarti, Econophysics of Markets and Business Networks (Springer-Verlag, Milan, 2007).
- [37] M. Parkinson M., J. Business **53**, 62 (1980).

- [38] A. Joseph, I. Vodenska, H. Eugene Stanley, and G. Chen, *Neteconomics: Novel Forecasting Techniques from the Combination of Big Data*, Network Science and Economics, Março 2014.
- [39] J.M. Karpoff, J. Finan. Quant. Anal. **22** 109 1997.
- [40] P. Gopikrishnan, V. Plerou, X. Gabaix, and H.E. Stanley, Phys. Rev. E **62**, R4493 (2000).
- [41] M.B. Garman and M.J. Klass, J. Business **53**, 67 (1980).
- [42] D. Féral e S. Péché, *The largest eigenvalue of rank one deformation of large Wigner matrices*, Comm. Math. Phys. **272**, 185 (2007);  
S. Péché, Probab. Theory Relat. Fields **134**, 127 (2006).
- [43] Z. Burda, A. Görlich, A. Jarosz and J. Jurkiewicz, Physica A **343**, 295 (2004).
- [44] J. W. Silverstein and Z. D. Bai, J. Multivar. Anal., **54**, 175 (1995).
- [45] F. Lillo and R.N. Mantegna, Phys. Rev. E **68**, 016119 (2006);  
D. Sornette, Y. Malevergne and J.F. Muzy, Risk **16**, 67 (2003).