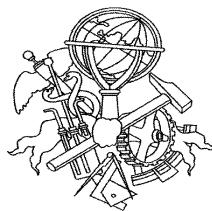


# ANÁLISE DINÂMICA DE MERCADOS FINANCEIROS

Diogo Oliveira Pontes



Mestrado em Engenharia Electrotécnica e de Computadores

Área de Especialização de Automação e Sistemas

Departamento de Engenharia Electrotécnica

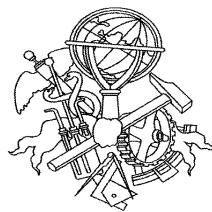
Instituto Superior de Engenharia do Porto

2010



Este relatório satisfaz, parcialmente, os requisitos que constam da Ficha de Disciplina de Tese/Dissertação, do 2º ano, do Mestrado em Engenharia Electrotécnica e de Computadores

Candidato: Diogo Oliveira Pontes, N° 1010598, 101058@isep.ipp.pt  
Orientação científica: Professor José Tenreiro Machado



Mestrado em Engenharia Electrotécnica e de Computadores  
Área de Especialização de Automação e Sistemas  
Departamento de Engenharia Electrotécnica  
Instituto Superior de Engenharia do Porto

12 de julho de 2010



## *Agradecimentos*

Este espaço é dedicado àqueles que deram a sua contribuição para que esta dissertação fosse realizada. A todos eles deixo aqui o meu agradecimento sincero.

Em primeiro lugar agradeço ao Prof. Luís Afonso a forma como orientou inicialmente o meu trabalho, criando as bases que tornaram possíveis a sua evolução.

Em segundo lugar, gostaria de agradecer aos meus pais, José Ernesto Pontes e Berta dos Santos Pontes pela paciência demonstrada.

Finalmente, deixo um agradecimento muito especial a Ana Sofia da Cruz, pelo apoio incondicional e palavras motivadoras dadas ao longo do desenvolvimento deste trabalho.



## *Resumo*

Na sociedade actual, é cada vez mais difícil desassociar o ambiente financeiro do ambiente social, tendo o primeiro influência directa ou indirecta em praticamente todos os aspectos da sociedade.

A esta influência está associada a vasta quantidade de informação e serviços financeiros que possibilitam uma melhor compreensão do ambiente socioeconómico actual, permitindo também o estudo das evoluções e das dinâmicas dos mercados financeiros.

Este trabalho refere-se ao estudo e comparação de algumas ferramentas disponíveis para a análise dinâmica e tentativa de previsão de alguns índices de bolsa escolhidos. Tais métodos a estudar são modelos clássicos como o Autoregressivo, Média Móvel e o Modelo Misto apresentado por Box e Jenkins.

São também propostos dois métodos que tentam distanciar-se dos métodos tradicionais por apenas considerarem para a sua previsão os momentos semelhantes ao momento actual que se tenta prever, ao invés de considerar todo o espectro dos dados disponíveis, tal como os métodos clássicos referidos anteriormente.

### *Palavras-Chave*

Análise Dinâmica, Séries Temporais, Média Móvel, Autoregressivo



## *Abstract*

In today's society, it's becoming more difficult to disassociate the financial environment from the social environment, having the former direct or indirect influence in almost all aspects of society.

It's associated to this influence the vast quantity of information and financial services that allow a better understanding of today's socioeconomic environment, also allowing the possibility to study financial market's evolution and dynamics.

This thesis refers to the study and comparison of some of the tools available for a dynamic analysis and attempts to forecast some chosen stock market indexes. Such tools to be studied are classical models such as the Autoregressive, Moving Average and the Mixed Model proposed by Box & Jenkins.

Two models are also proposed that try to distance themselves from the classical models by just considering for the forecast the moments that are similar to the current moment which they are trying to forecast, instead of considering the full spectrum of the available data, such as the classical models previously referred.

### ***Key Words***

Dynamic Analysis, Time Series, Moving Average, Autoregressive



# Índice

<i>Resumo</i> .....	7
<i>Abstract</i> .....	9
<i>Índice</i> .....	11
<i>Índice de Figuras</i> .....	15
<i>Índice de Tabelas</i> .....	21
<i>Acrónimos</i> .....	25
<i>Simbologia</i> .....	27
<b>1. INTRODUÇÃO</b> .....	29
<b>2. APRESENTAÇÃO DOS DADOS USADOS</b> .....	33
<b>3. TRANSFORMAÇÃO DE DADOS NOS MERCADOS FINANCEIROS</b> .....	81
3.1. DIFERENÇA DE PREÇOS .....	82
3.1.1. DIFERENÇA DE PREÇOS DESCONTADA .....	83
3.1.2. SUCESSIVAS DIFERENÇAS DO LOGARITMO NATURAL DO PREÇO .....	84
3.2. RETORNOS .....	85
3.2.1. RETORNO BRUTO DE PERÍODO ÚNICO .....	85
3.2.2. RETORNO LOGARÍTMICO .....	86
3.3. ESCALA DE TEMPO EM MERCADOS FINANCEIROS .....	87
<b>4. MODELIZAÇÃO DE SUCESSÕES CRONOLÓGICAS</b> .....	91
4.1. IDENTIFICAÇÃO .....	95
4.2. ESTIMAÇÃO .....	104
4.3. ANÁLISE DO DIAGNÓSTICO .....	105
4.3.1. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: SIGNIFICÂNCIA ESTATÍSTICA DOS PARÂMETROS UTILIZADOS .....	105

4.3.2. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: ESTACIONARIDADE E INVERTIBILIDADE DO MODELO ESTIMADO .....	106
4.3.3. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: ESTABILIDADE DO MODELO .....	106
4.3.4. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: REDUNDÂNCIA DE ESTIMATIVAS .....	107
4.3.5. AVALIAÇÃO DA QUALIDADE DO AJUSTAMENTO .....	107
4.4. CRITÉRIO DE SELECÇÃO DE MODELOS .....	109
4.4.1. CRITÉRIO DE AKAIKE .....	109
4.4.2. CRITÉRIO BIC .....	112
<b>5. LINGUAGEM DE PROGRAMAÇÃO R .....</b>	<b>113</b>
5.1. INTRODUÇÃO AO R .....	113
5.2. O AMBIENTE R .....	115
<b>6. METODOLOGIAS CLÁSSICAS DE MODELIZAÇÃO LINEAR DE SÉRIES TEMPORAIS .....</b>	<b>117</b>
6.1. ERRO MÉDIO ABSOLUTO .....	118
6.2. LIMITE DE FUTILIDADE .....	119
6.3. PROCESSO AUTOREGRESSIVO .....	120
6.3.1. ESTIMAÇÃO DOS PARÂMETROS NO MODELO AUTOREGRESSIVO .....	121
6.3.2. RESULTADOS .....	123
6.3.3. OBSERVAÇÕES .....	127
6.4. PROCESSO DE MÉDIA MÓVEL .....	130
6.4.1. MÉDIA MÓVEL SIMPLES .....	130
6.4.2. MÉDIA MÓVEL PONDERADA .....	131
6.4.3. ESTIMAÇÃO DOS PARÂMETROS NO MODELO DE MÉDIA MÓVEL .....	132
6.4.4. RESULTADOS .....	133
6.4.5. OBSERVAÇÕES .....	137
6.5. PROCESSO AUTOREGRESSIVO DE MÉDIA MÓVEL (MODELO MISTO) .....	140
6.5.1. ESTIMAÇÃO DOS PARÂMETROS NO MODELO MISTO .....	141

6.5.2. <i>RESULTADOS</i> .....	142
6.5.3. <i>OBSERVAÇÕES</i> .....	145
6.6. COMPARAÇÃO ENTRE OS MODELOS CLÁSSICOS .....	147
<b>7. PROPOSTA DE DUAS METODOLOGIAS DE ANÁLISE E PREVISÃO</b> .....	<b>149</b>
7.1. APRESENTAÇÃO GERAL DOS MODELOS PROPOSTOS .....	150
7.2. PREVISÃO .....	151
7.3. ESTIMAÇÃO .....	152
7.3.1. <i>TAMANHO DA SEQUÊNCIA</i> .....	152
7.3.2. <i>NÚMERO DE CLASSES</i> .....	152
7.3.3. <i>EPSON</i> .....	153
7.4. OPTIMIZAÇÃO DE PARÂMETROS .....	153
7.4.1. <i>ERRO MÉDIA ABSOLUTO</i> .....	154
7.4.2. <i>PERDA DE INFORMAÇÃO</i> .....	155
7.5. MODELO DE CLASSES .....	156
7.5.1. <i>RESULTADOS</i> .....	157
7.5.2. <i>OBSERVAÇÕES</i> .....	162
7.6. MODELO EPSON .....	165
7.6.1. <i>RESULTADOS</i> .....	166
7.6.2. <i>OBSERVAÇÕES</i> .....	170
<b>8. COMPARAÇÃO DAS METODOLOGIAS ESTUDADAS</b> .....	<b>175</b>
<b>9. CONCLUSÃO</b> .....	<b>179</b>
<i>Bibliografia</i> .....	185



## *Índice de Figuras*

<b>Figura 1</b> – Representação temporal do valor real de fecho de AAPL .....	38
<b>Figura 2</b> – Representação temporal do valor às diferenças de fecho de AAPL .....	38
<b>Figura 3</b> – Representação temporal do valor de Retorno Bruto de fecho de AAPL .....	38
<b>Figura 4</b> – Representação temporal do valor de Retorno Logarítmico de fecho de AAPL .....	38
<b>Figura 5</b> – Representação histográfica do valor real de fecho de AAPL .....	39
<b>Figura 6</b> – Representação histográfica do valor às diferenças de fecho de AAPL .....	39
<b>Figura 7</b> – Representação histográfica do valor de Retorno Bruto de fecho de AAPL .....	39
<b>Figura 8</b> – Representação histográfica do valor de Retorno Logarítmico de AAPL .....	39
<b>Figura 9</b> – Representação temporal do valor real de fecho de BA .....	40
<b>Figura 10</b> – Representação temporal do valor às diferenças de fecho de BA .....	40
<b>Figura 11</b> – Representação temporal do valor de Retorno Bruto de fecho de BA .....	40
<b>Figura 12</b> – Representação temporal do valor de Retorno Logarítmico de fecho de BA .....	40
<b>Figura 13</b> – Representação histográfica do valor real de fecho de BA .....	41
<b>Figura 14</b> – Representação histográfica do valor às diferenças de fecho de BA .....	41
<b>Figura 15</b> – Representação histográfica do valor de Retorno Bruto de fecho de BA .....	41
<b>Figura 16</b> – Representação histográfica do valor de Retorno Logarítmico de BA .....	41
<b>Figura 17</b> – Representação temporal do valor real de fecho de BP .....	42
<b>Figura 18</b> – Representação temporal do valor às diferenças de fecho de BP .....	42
<b>Figura 19</b> – Representação temporal do valor de Retorno Bruto de fecho de BP .....	42
<b>Figura 20</b> – Representação temporal do valor de Retorno Logarítmico de fecho de BP .....	42
<b>Figura 21</b> – Representação histográfica do valor real de fecho de BP .....	43
<b>Figura 22</b> – Representação histográfica do valor às diferenças de fecho de BP .....	43
<b>Figura 23</b> – Representação histográfica do valor de Retorno Bruto de fecho de BP .....	43
<b>Figura 24</b> – Representação histográfica do valor de Retorno Logarítmico de BP .....	43
<b>Figura 25</b> – Representação temporal do valor real de fecho de CVX .....	44
<b>Figura 26</b> – Representação temporal do valor às diferenças de fecho de CVX .....	44
<b>Figura 27</b> – Representação temporal do valor de Retorno Bruto de fecho de CVX .....	44
<b>Figura 28</b> – Representação temporal do valor de Retorno Logarítmico de fecho de CVX .....	44
<b>Figura 29</b> – Representação histográfica do valor real de fecho de CVX .....	45
<b>Figura 30</b> – Representação histográfica do valor às diferenças de fecho de CVX .....	45
<b>Figura 31</b> – Representação histográfica do valor de Retorno Bruto de fecho de CVX .....	45

<b>Figura 32</b> – Representação histográfica do valor de Retorno Logarítmico de CVX .....	45
<b>Figura 33</b> – Representação temporal do valor real de fecho de C .....	46
<b>Figura 34</b> – Representação temporal do valor às diferenças de fecho de C .....	46
<b>Figura 35</b> – Representação temporal do valor de Retorno Bruto de fecho de C .....	46
<b>Figura 36</b> – Representação temporal do valor de Retorno Logarítmico de fecho de C .....	46
<b>Figura 37</b> – Representação histográfica do valor real de fecho de C .....	47
<b>Figura 38</b> – Representação histográfica do valor às diferenças de fecho de C .....	47
<b>Figura 39</b> – Representação histográfica do valor de Retorno Bruto de fecho de C .....	47
<b>Figura 40</b> – Representação histográfica do valor de Retorno Logarítmico de C .....	47
<b>Figura 41</b> – Representação temporal do valor real de fecho de DELL .....	48
<b>Figura 42</b> – Representação temporal do valor às diferenças de fecho de DELL .....	48
<b>Figura 43</b> – Representação temporal do valor de Retorno Bruto de fecho de DELL .....	48
<b>Figura 44</b> – Representação temporal do valor de Retorno Logarítmico de fecho de DELL ....	48
<b>Figura 45</b> – Representação histográfica do valor real de fecho de DELL .....	49
<b>Figura 46</b> – Representação histográfica do valor às diferenças de fecho de DELL .....	49
<b>Figura 47</b> – Representação histográfica do valor de Retorno Bruto de fecho de DELL .....	49
<b>Figura 48</b> – Representação histográfica do valor de Retorno Logarítmico de DELL .....	49
<b>Figura 49</b> – Representação temporal do valor real de fecho de XOM .....	50
<b>Figura 50</b> – Representação temporal do valor às diferenças de fecho de XOM .....	50
<b>Figura 51</b> – Representação temporal do valor de Retorno Bruto de fecho de XOM .....	50
<b>Figura 52</b> – Representação temporal do valor de Retorno Logarítmico de fecho de XOM ....	50
<b>Figura 53</b> – Representação histográfica do valor real de fecho de XOM .....	51
<b>Figura 54</b> – Representação histográfica do valor às diferenças de fecho de XOM .....	51
<b>Figura 55</b> – Representação histográfica do valor de Retorno Bruto de fecho de XOM .....	51
<b>Figura 56</b> – Representação histográfica do valor de Retorno Logarítmico de XOM .....	51
<b>Figura 57</b> – Representação temporal do valor real de fecho de F .....	52
<b>Figura 58</b> – Representação temporal do valor às diferenças de fecho de F .....	52
<b>Figura 59</b> – Representação temporal do valor de Retorno Bruto de fecho de F .....	52
<b>Figura 60</b> – Representação temporal do valor de Retorno Logarítmico de fecho de F .....	52
<b>Figura 61</b> – Representação histográfica do valor real de fecho de F .....	53
<b>Figura 62</b> – Representação histográfica do valor às diferenças de fecho de F .....	53
<b>Figura 63</b> – Representação histográfica do valor de Retorno Bruto de fecho de F .....	53
<b>Figura 64</b> – Representação histográfica do valor de Retorno Logarítmico de F .....	53
<b>Figura 65</b> – Representação temporal do valor real de fecho de GE .....	54
<b>Figura 66</b> – Representação temporal do valor às diferenças de fecho de GE .....	54
<b>Figura 67</b> – Representação temporal do valor de Retorno Bruto de fecho de GE .....	54

<b>Figura 68</b> – Representação temporal do valor de Retorno Logarítmico de fecho de GE .....	54
<b>Figura 69</b> – Representação histográfica do valor real de fecho de GE .....	55
<b>Figura 70</b> – Representação histográfica do valor às diferenças de fecho de GE .....	55
<b>Figura 71</b> – Representação histográfica do valor de Retorno Bruto de fecho de GE .....	55
<b>Figura 72</b> – Representação histográfica do valor de Retorno Logarítmico de GE .....	55
<b>Figura 73</b> – Representação temporal do valor real de fecho de GM .....	56
<b>Figura 74</b> – Representação temporal do valor às diferenças de fecho de GM .....	56
<b>Figura 75</b> – Representação temporal do valor de Retorno Bruto de fecho de GM .....	56
<b>Figura 76</b> – Representação temporal do valor de Retorno Logarítmico de fecho de GM .....	56
<b>Figura 77</b> – Representação histográfica do valor real de fecho de GM .....	57
<b>Figura 78</b> – Representação histográfica do valor às diferenças de fecho de GM .....	57
<b>Figura 79</b> – Representação histográfica do valor de Retorno Bruto de fecho de GM .....	57
<b>Figura 80</b> – Representação histográfica do valor de Retorno Logarítmico de GM .....	57
<b>Figura 81</b> – Representação temporal do valor real de fecho de HPQ .....	58
<b>Figura 82</b> – Representação temporal do valor às diferenças de fecho de HPQ .....	58
<b>Figura 83</b> – Representação temporal do valor de Retorno Bruto de fecho de HPQ .....	58
<b>Figura 84</b> – Representação temporal do valor de Retorno Logarítmico de fecho de HPQ .....	58
<b>Figura 85</b> – Representação histográfica do valor real de fecho de HPQ .....	59
<b>Figura 86</b> – Representação histográfica do valor às diferenças de fecho de HPQ .....	59
<b>Figura 87</b> – Representação histográfica do valor de Retorno Bruto de fecho de HPQ .....	59
<b>Figura 88</b> – Representação histográfica do valor de Retorno Logarítmico de HPQ .....	59
<b>Figura 89</b> – Representação temporal do valor real de fecho de IBM .....	60
<b>Figura 90</b> – Representação temporal do valor às diferenças de fecho de IBM .....	60
<b>Figura 91</b> – Representação temporal do valor de Retorno Bruto de fecho de IBM .....	60
<b>Figura 92</b> – Representação temporal do valor de Retorno Logarítmico de fecho de IBM .....	60
<b>Figura 93</b> – Representação histográfica do valor real de fecho de IBM .....	61
<b>Figura 94</b> – Representação histográfica do valor às diferenças de fecho de IBM .....	61
<b>Figura 95</b> – Representação histográfica do valor de Retorno Bruto de fecho de IBM .....	61
<b>Figura 96</b> – Representação histográfica do valor de Retorno Logarítmico de IBM .....	61
<b>Figura 97</b> – Representação temporal do valor real de fecho de JPM .....	62
<b>Figura 98</b> – Representação temporal do valor às diferenças de fecho de JPM .....	62
<b>Figura 99</b> – Representação temporal do valor de Retorno Bruto de fecho de JPM .....	62
<b>Figura 100</b> – Representação temporal do valor de Retorno Logarítmico de fecho de JPM .....	62
<b>Figura 101</b> – Representação histográfica do valor real de fecho de JPM .....	63
<b>Figura 102</b> – Representação histográfica do valor às diferenças de fecho de JPM .....	63
<b>Figura 103</b> – Representação histográfica do valor de Retorno Bruto de fecho de JPM .....	63

<b>Figura 104</b> – Representação histográfica do valor de Retorno Logarítmico de JPM .....	63
<b>Figura 105</b> – Representação temporal do valor real de fecho de MSFT .....	64
<b>Figura 106</b> – Representação temporal do valor às diferenças de fecho de MSFT .....	64
<b>Figura 107</b> – Representação temporal do valor de Retorno Bruto de fecho de MSFT .....	64
<b>Figura 108</b> – Representação temporal do valor de Retorno Logarítmico de fecho de MSFT ..	64
<b>Figura 109</b> – Representação histográfica do valor real de fecho de MSFT .....	65
<b>Figura 110</b> – Representação histográfica do valor às diferenças de fecho de MSFT .....	65
<b>Figura 111</b> – Representação histográfica do valor de Retorno Bruto de fecho de MSFT .....	65
<b>Figura 112</b> – Representação histográfica do valor de Retorno Logarítmico de MSFT .....	65
<b>Figura 113</b> – Representação temporal do valor real de fecho de SNE .....	66
<b>Figura 114</b> – Representação temporal do valor às diferenças de fecho de SNE .....	66
<b>Figura 115</b> – Representação temporal do valor de Retorno Bruto de fecho de SNE .....	66
<b>Figura 116</b> – Representação temporal do valor de Retorno Logarítmico de fecho de SNE .....	66
<b>Figura 117</b> – Representação histográfica do valor real de fecho de SNE .....	67
<b>Figura 118</b> – Representação histográfica do valor às diferenças de fecho de SNE .....	67
<b>Figura 119</b> – Representação histográfica do valor de Retorno Bruto de fecho de SNE .....	67
<b>Figura 120</b> – Representação histográfica do valor de Retorno Logarítmico de SNE .....	67
<b>Figura 121</b> – Representação temporal do valor real de fecho de FTSE100 .....	68
<b>Figura 122</b> – Representação temporal do valor às diferenças de fecho de FTSE100 .....	68
<b>Figura 123</b> – Representação temporal do valor de Retorno Bruto de fecho de FTSE100 .....	68
<b>Figura 124</b> – Representação temporal do valor de Retorno Logarítmico de FTSE100 .....	68
<b>Figura 125</b> – Representação histográfica do valor real de fecho de FTSE100 .....	69
<b>Figura 126</b> – Representação histográfica do valor às diferenças de fecho de FTSE100 .....	69
<b>Figura 127</b> – Representação histográfica do valor de Retorno Bruto de fecho de FTSE100 ...	69
<b>Figura 128</b> – Representação histográfica do valor de Retorno Logarítmico de FTSE100 .....	69
<b>Figura 129</b> – Representação temporal do valor real de fecho de NASDAQ .....	70
<b>Figura 130</b> – Representação temporal do valor às diferenças de fecho de NASDAQ .....	70
<b>Figura 131</b> – Representação temporal do valor de Retorno Bruto de fecho de NASDAQ .....	70
<b>Figura 132</b> – Representação temporal do valor de Retorno Logarítmico de NASDAQ .....	70
<b>Figura 133</b> – Representação histográfica do valor real de fecho de NASDAQ .....	71
<b>Figura 134</b> – Representação histográfica do valor às diferenças de fecho de NASDAQ .....	71
<b>Figura 135</b> – Representação histográfica do valor de Retorno Bruto de fecho de NASDAQ ..	71
<b>Figura 136</b> – Representação histográfica do valor de Retorno Logarítmico de NASDAQ .....	71
<b>Figura 137</b> – Representação temporal do valor real de fecho de NYSE .....	72
<b>Figura 138</b> – Representação temporal do valor às diferenças de fecho de NYSE .....	72
<b>Figura 139</b> – Representação temporal do valor de Retorno Bruto de fecho de NYSE .....	72

<b>Figura 140</b> – Representação temporal do valor de Retorno Logarítmico de NYSE .....	72
<b>Figura 141</b> – Representação histográfica do valor real de fecho de NYSE .....	73
<b>Figura 142</b> – Representação histográfica do valor às diferenças de fecho de NYSE .....	73
<b>Figura 143</b> – Representação histográfica do valor de Retorno Bruto de fecho de NYSE .....	73
<b>Figura 144</b> – Representação histográfica do valor de Retorno Logarítmico de NYSE .....	73
<b>Figura 145</b> – Representação temporal do valor real de fecho de NIKKEI225 .....	74
<b>Figura 146</b> – Representação temporal do valor às diferenças de fecho de NIKKEI225 .....	74
<b>Figura 147</b> – Representação temporal do valor de Retorno Bruto de fecho de NIKKEI225 ...	74
<b>Figura 148</b> – Representação temporal do valor de Retorno Logarítmico de NIKKEI225 .....	74
<b>Figura 149</b> – Representação histográfica do valor real de fecho de NIKKEI225 .....	75
<b>Figura 150</b> – Representação histográfica do valor às diferenças de fecho de NIKKEI225 .....	75
<b>Figura 151</b> – Representação histográfica do valor de Retorno Bruto de fecho de NIKKEI225	75
<b>Figura 152</b> – Representação histográfica do valor de Retorno Logarítmico de NIKKEI225 ..	75
<b>Figura 153</b> – Representação temporal do valor real de fecho de PSI20 .....	76
<b>Figura 154</b> – Representação temporal do valor às diferenças de fecho de PSI20 .....	76
<b>Figura 155</b> – Representação temporal do valor de Retorno Bruto de fecho de PSI20 .....	76
<b>Figura 156</b> – Representação temporal do valor de Retorno Logarítmico de fecho de PSI20 ..	76
<b>Figura 157</b> – Representação histográfica do valor real de fecho de PSI20 .....	77
<b>Figura 158</b> – Representação histográfica do valor às diferenças de fecho de PSI20 .....	77
<b>Figura 159</b> – Representação histográfica do valor de Retorno Bruto de fecho de PSI20 .....	77
<b>Figura 160</b> – Representação histográfica do valor de Retorno Logarítmico de PSI20 .....	77
<b>Figura 161</b> – Metodologia Box – Jenkins .....	93
<b>Figura 162</b> – Representação da FAC do índice BP para os dados em retorno bruto .....	97
<b>Figura 163</b> – Representação da FAC do índice BP para os dados às diferenças .....	97
<b>Figura 164</b> – Representação da FACP do índice BP para os dados em retorno bruto .....	98
<b>Figura 165</b> – Representação da FACP do índice BP para os dados às diferenças .....	98
<b>Figura 166</b> – Representação da FAC do índice IBM para os dados em retorno bruto .....	99
<b>Figura 167</b> – Representação da FAC do índice IBM para os dados às diferenças .....	99
<b>Figura 168</b> – Representação da FACP do índice IBM para os dados em retorno bruto .....	100
<b>Figura 169</b> – Representação da FACP do índice IBM para os dados às diferenças .....	100
<b>Figura 170</b> – Representação da FAC do índice JPM para os dados em retorno bruto .....	101
<b>Figura 171</b> – Representação da FAC do índice JPM para os dados às diferenças .....	101
<b>Figura 172</b> – Representação da FACP do índice JPM para os dados em retorno bruto .....	102
<b>Figura 172</b> – Representação da FACP do índice JPM para os dados às diferenças .....	102
<b>Figura 174</b> – Ambiente gráfico do R no sistema operativo Mac OS X .....	116

<b>Figura 175</b> – Evolução dos valores reais e de previsão do índice AAPL para Retorno Bruto com $AR_1$ .....	127
<b>Figura 176</b> – Evolução dos valores reais e de previsão do índice XOM para Retorno Bruto com $MA_1$ .....	137
<b>Figura 177</b> – Evolução dos valores reais e de previsão do índice XOM para Retorno Bruto com $ARMA_1$ .....	145
<b>Figura 178</b> – Distribuição dos valores do parâmetro $n_{classes}$ para o modelo de Classes .....	163
<b>Figura 179</b> – Distribuição dos valores do parâmetro $tam_{seq}$ para o modelo de Epson .....	171
<b>Figura 180</b> – Distribuição dos valores do parâmetro $E$ para o modelo de Epson .....	172

## *Índice de Tabelas*

<b>Tabela 1</b> – Características da FAC e FACP nos modelos clássicos .....	96
<b>Tabela 2</b> – Valores do EMA para o método AR dos valores de Retorno Absoluto .....	124
<b>Tabela 3</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método AR de valores de Retorno Absoluto .....	124
<b>Tabela 4</b> – Valores do EMA para o método AR dos valores de Retorno Logarítmico .....	125
<b>Tabela 5</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método AR de valores de Retorno Logarítmico .....	125
<b>Tabela 6</b> – Valores do EMA para o método AR dos valores às Diferenças .....	126
<b>Tabela 7</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método AR de valores às Diferenças .....	126
<b>Tabela 8</b> – Resumo de resultados do modelo AR .....	128
<b>Tabela 9</b> – Valores do EMA para o método MA dos valores de Retorno Absoluto .....	134
<b>Tabela 10</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método MA de valores de Retorno Absoluto .....	134
<b>Tabela 11</b> – Valores do EMA para o método MA dos valores de Retorno Logarítmico .....	135
<b>Tabela 12</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método MA de valores de Retorno Logarítmico .....	135
<b>Tabela 13</b> – Valores do EMA para o método MA dos valores às Diferenças .....	136
<b>Tabela 14</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método MA de valores às Diferenças .....	136
<b>Tabela 15</b> – Resumo de resultados do modelo MA .....	138
<b>Tabela 16</b> – Valores do EMA para o método ARMA dos valores de Retorno Absoluto .....	143
<b>Tabela 17</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método ARMA de valores de Retorno Absoluto .....	143
<b>Tabela 18</b> – Valores do EMA para o método ARMA dos valores de Retorno Logarítmico ..	144
<b>Tabela 19</b> – Valores comparativos de melhoria relativamente ao $L_F$ do método ARMA de valores de Retorno Logarítmico .....	144
<b>Tabela 20</b> – Resumo de resultados do modelo MA .....	146
<b>Tabela 21</b> – Apresentação dos melhores resultados dos modelos clássicos .....	147

<b>Tabela 22</b> – Valores do EMA do modelo de Classes para o índice BA com dados de Retorno Bruto .....	157
<b>Tabela 23</b> – Valores da Perda de Informação do modelo de Classes para o índice BA com dados de Retorno Bruto .....	157
<b>Tabela 24</b> – Resultados do Modelo de Classes para o índice AAPL .....	159
<b>Tabela 25</b> – Resultados do Modelo de Classes para o índice BA .....	159
<b>Tabela 26</b> – Resultados do Modelo de Classes para o índice BP .....	159
<b>Tabela 27</b> – Resultados do Modelo de Classes para o índice CVX .....	159
<b>Tabela 28</b> – Resultados do Modelo de Classes para o índice C .....	159
<b>Tabela 29</b> – Resultados do Modelo de Classes para o índice DELL .....	159
<b>Tabela 30</b> – Resultados do Modelo de Classes para o índice XOM .....	159
<b>Tabela 31</b> – Resultados do Modelo de Classes para o índice F .....	160
<b>Tabela 32</b> – Resultados do Modelo de Classes para o índice GE .....	160
<b>Tabela 33</b> – Resultados do Modelo de Classes para o índice GM .....	160
<b>Tabela 34</b> – Resultados do Modelo de Classes para o índice HPQ .....	160
<b>Tabela 35</b> – Resultados do Modelo de Classes para o índice IBM .....	160
<b>Tabela 36</b> – Resultados do Modelo de Classes para o índice JPM .....	160
<b>Tabela 37</b> – Resultados do Modelo de Classes para o índice MSFT .....	160
<b>Tabela 38</b> – Resultados do Modelo de Classes para o índice SNE .....	161
<b>Tabela 39</b> – Resultados do Modelo de Classes para o índice FTSE .....	161
<b>Tabela 40</b> – Resultados do Modelo de Classes para o índice NASDAQ .....	161
<b>Tabela 41</b> – Resultados do Modelo de Classes para o índice NYSE .....	161
<b>Tabela 42</b> – Resultados do Modelo de Classes para o índice NASDAQ .....	161
<b>Tabela 43</b> – Resultados do Modelo de Classes para o índice PSI20 .....	161
<b>Tabela 44</b> – Valores de Erro Médio Absoluto do modelo Epsilon para o índice DELL com dados às Diferenças .....	166
<b>Tabela 45</b> – Valores de Perda de Informação correspondentes aos valores de EMA do índice DELL com dados às Diferenças .....	166
<b>Tabela 46</b> – Resultados do Modelo Epsilon para o índice AAPL .....	167
<b>Tabela 47</b> – Resultados do Modelo Epsilon para o índice BA .....	167
<b>Tabela 48</b> – Resultados do Modelo Epsilon para o índice BP .....	167
<b>Tabela 49</b> – Resultados do Modelo Epsilon para o índice CVX .....	167
<b>Tabela 50</b> – Resultados do Modelo Epsilon para o índice C .....	167
<b>Tabela 51</b> – Resultados do Modelo Epsilon para o índice DELL .....	167

<b>Tabela 52</b> – Resultados do Modelo Epsilon para o índice XOM .....	168
<b>Tabela 53</b> – Resultados do Modelo Epsilon para o índice F .....	168
<b>Tabela 54</b> – Resultados do Modelo Epsilon para o índice GE .....	168
<b>Tabela 55</b> – Resultados do Modelo Epsilon para o índice GM .....	168
<b>Tabela 56</b> – Resultados do Modelo Epsilon para o índice HPQ .....	168
<b>Tabela 57</b> – Resultados do Modelo Epsilon para o índice IBM .....	168
<b>Tabela 58</b> – Resultados do Modelo Epsilon para o índice JPM .....	169
<b>Tabela 59</b> – Resultados do Modelo Epsilon para o índice MSFT .....	169
<b>Tabela 60</b> – Resultados do Modelo Epsilon para o índice SNE .....	169
<b>Tabela 61</b> – Resultados do Modelo Epsilon para o índice FTSE100 .....	169
<b>Tabela 62</b> – Resultados do Modelo Epsilon para o índice NASDAQ .....	169
<b>Tabela 63</b> – Resultados do Modelo Epsilon para o índice NYSE .....	169
<b>Tabela 64</b> – Resultados do Modelo Epsilon para o índice NIKKEI225 .....	169
<b>Tabela 65</b> – Resultados do Modelo Epsilon para o índice PSI20 .....	169
<b>Tabela 66</b> – Comparação dos modelos estudados com base no valor de $M$ correspondente ..	176



## *Acrónimos*

**AAPL** – Apple Inc.

**AR** – Autoregressive (Autoregressivo)

**ARMA** – Autoregressive Moving Average (Autoregressivo de Média Móvel)

**BA** – Boeing Co.

**BP** – British Petroleum (BP plc)

**CVX** – Chevron Corp.

**C** – Citigroup Inc.

**CSV** – Comma Separated Value

**DELL** – Dell Inc.

**EMA** – Erro Média Absoluto

**F** – Ford Motor Co.

**FAC** – Função de Autocorrelação

**FACP** – Função de Autocorrelação Parcial

**GE** – General Electric Co.

**GM** – General Motors Corporation

**HPQ** – Hewlett-Packard Company

**IBM** – International Business Machines Corp.

**JPM** – JPMorgan Chase & Co.

**L<sub>F</sub>** – Limite de Futilidade

**MA** – Moving Average (Media Móvel)

**MSFT** – Microsoft Corporation

**PI** – Perda de Informação

**SNE** – Sony Corporation

**FTSE 100** – Financial Times Stock Exchange

**NASDAQ** – National Association of Securities Dealers Automated Quotations

**NYSE** – New York Stock Exchange

**PSI20** – Portuguese Stock Index

**XOM** – Exxon Mobil Corp.



## *Simbologia*

$X_t$  – Preço de um bem financeiro no momento  $t$ .

$Y_t$  – Valor transformado de um bem financeiro no momento  $t$ .

$D_t$  – Factor de deflação no momento  $t$ .

$k$  – Atraso.

$p$  – Ordem do modelo Autoregressivo.

$q$  – Ordem do Modelo de Média Móvel.

$\sigma^2$  – Variância.

$\varepsilon$  – Processo de ruído branco.

$S^*$  – Soma dos quadrados.

$\lambda$  – Tamanho da amostra da série temporal.

$w_i$  – Peso  $i$  do processo de Média Móvel Ponderada



# 1. INTRODUÇÃO

O sistema financeiro tem importância fundamental nas economias nacionais e internacionais. A sua influência é abrangente a praticamente todos os sectores da sociedade, tornando-se imperativa a análise e compreensão dos mercados.

Neste âmbito, a aplicação de modelos matemáticos e tecnologias computacionais que suportem uma análise da evolução dos mercados torna-se imprescindível. O sistema financeiro depende de uma grande variedade destas ferramentas assim como da disponibilidade, volume e significado de informação que, graças às evoluções tecnológicas das últimas décadas, permitiu o desenvolvimento de várias soluções, para um estudo suportado do ambiente financeiro.

Não obstante o vasto conhecimento em várias áreas por parte dos intervenientes nos mercados, seja teoria financeira, económica ou política, a utilização de modelos matemáticos e computacionais permite a criação de uma base de conhecimento sólida sobre a qual se poderá construir a estratégia de evolução mais apropriada para cada modificação do ambiente financeiro.

Esta base pode estabelecer-se sobre o estudo das séries temporais financeiras, nomeadamente com a teoria e a prática da análise dos activos ao longo do tempo. As

séries temporais financeiras devem lidar com um ambiente económico-financeiro em permanente mutação, e com o facto de a sua volatilidade não ser directamente observável, resultando em um grau de incerteza elevado. Como resultado desta incerteza adicional, os métodos estatísticos possuem um papel importante na análise de séries temporais financeiras.

Ferramentas matemáticas avançadas podem identificar e utilizar a correlação entre activos de mercado e outros factores envolventes, como alterações climáticas, a variação das taxas de juro ou de desemprego, ou até mesmo alterações de agenda política (nacional ou internacional) para realizar previsões acerca de um determinado activo financeiro.

No entanto, este trabalho irá incidir apenas na evolução de cada índice em particular, não realizando uma correlação entre os activos financeiros em estudo. Contudo, alguns destes foram escolhidos propositadamente por pertencerem ao mesmo ramo ou indústria, de modo a permitir um futuro estudo de tal correlação entre os vários activos.

Existem várias teorias relevantes acerca da funcionalidade e dinâmica dos mercados que se disponibilizam a uma análise formal. Algumas métodos são aqui apresentados, assim como duas propostas de modelos de previsão alternativos.

Os modelos propostos tentaram distanciar-se das metodologias clássicas ao analisar apenas os momentos semelhantes do espectro total ao momento que se está a tentar prever. Serão chamados de Modelo Epsom e Modelos de Classes. A diferença entre estes métodos incide sobre o modo como são transformados os dados que utilizam. Posteriormente à transformação dos dados, o seu método de funcionamento é idêntico.

Pretende-se com estes métodos procurar uma alternativa melhor aos métodos tradicionais. A metodologia de comparação e conclusão acerca de qual o modelo mais adequado irá incidir sobre a média do erro absoluto entre as previsões efectuadas com os vários métodos, e o valor real dos índices.

A organização deste trabalho é a seguir apresentada.

No capítulo 2 expõe-se as características dos dados a serem usados no trabalho, com uma exposição gráfica destes ao longo do tempo e respectivos histogramas, dos seus valores reais, de retorno (bruto e logarítmico) e às diferenças. Serão demonstradas e justificadas as escolhas feitas relativamente ao tipo de transformação de dados utilizada.

Seguidamente, no capítulo 3, fala-se acerca da transformação dos dados emergentes nos mercados financeiros, sendo demonstradas as técnicas de transformação de dados. Também são discutidas as escalas de tempo usadas habitualmente em mercados financeiros e é referida a escolha feita para a escala de tempo adoptada no decorrer do trabalho.

O capítulo 4 revê conceitos fundamentais da análise de séries temporais lineares tal como a estacionariedade, a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP). É também abordada a metodologia de três etapas na escolha dos modelos, em que são analisadas as fases de identificação, estimação e avaliação do diagnóstico. Esta metodologia é seguida neste trabalho como modo de escolha dos modelos de previsão.

No capítulo 5 é feita uma introdução à linguagem computacional R utilizada neste trabalho para a realização dos cálculos de análise e tentativa de previsão dos modelos utilizados. A escolha desta linguagem recaiu sobre o facto de que as funções necessárias para cálculo e previsão das metodologias tradicionais já estão presentes nas bibliotecas do referido software.

No capítulo 6 são expostas e aplicadas algumas metodologias clássicas de modelização linear de séries temporais, como os processos autoregressivos, as médias móveis e a metodologia mista de processos autoregressivos de média móvel, também conhecida como metodologia de Box e Jenkins.

O capítulo 7 apresenta as duas metodologias propostas como alternativa aos modelos clássicos discutidos no capítulo 6. A principal diferença entre as

metodologias propostas e as clássicas discutidas anteriormente reside na quantidade de dados utilizados por cada uma das metodologias. Enquanto que as metodologias clássicas utilizam todos os dados disponíveis do espectro, as metodologias propostas apenas utilizam os dados dos momentos semelhantes ao momento que se pretende prever.

No capítulo 8 é realizada uma avaliação dos vários métodos estudados, comparando os seus resultados de modo a tentar determinar qual será o método mais indicado para o tipo de dados utilizados. Sugestões de melhoramento dos métodos propostos serão incluídas neste capítulo.

No capítulo 9 serão desenvolvidas as várias conclusões obtidas ao longo do trabalho.

## 2. APRESENTAÇÃO DOS DADOS USADOS

A possibilidade de estudar as séries temporais dos mercados financeiros utilizando ferramentas matemáticas é um factor essencial para uma melhor compreensão da sua natureza e para uma possível previsão dos seus futuros cenários. Neste capítulo serão demonstradas graficamente as características dos dados a serem usados neste trabalho, ao longo do tempo e respectivos histogramas, dos seus valores reais, de retorno bruto, de retorno logarítmico e às diferenças.

Este estudo é actualmente facilitado pela vasta quantidade de informação digital disponível dos valores dos activos financeiros. Com o objectivo de fornecer o acesso à informação dos activos financeiros, algumas entidades, como a Google<sup>1</sup> e a Yahoo<sup>2</sup> disponibilizam serviços de informação como os valores de vários activos presentes em algumas bolsas financeiras, assim como notícias relevantes ao ambiente económico.

No entanto, estes serviços gratuitos não possuem as mesmas características que os serviços pagos fornecidos por outras entidades ou até mesmo as versões pagas da

---

<sup>1</sup> <http://www.google.com/finance>

<sup>2</sup> <http://finance.yahoo.com>

Google ou Yahoo. Por exemplo, enquanto que as versões gratuitas actualizam as informações a cada 10 segundos mas com 15 minutos de atraso, as versões pagas actualizam informação ao segundo, frequência exigida pela maioria dos profissionais que trabalham na compra e venda de activos financeiros, por necessitarem de agir imediatamente perante uma alteração importante no cenário económico-social.

Neste trabalho serão apenas utilizados os dados fornecidos pela versão gratuita da Google e da Yahoo, por não haver a necessidade de informação com a frequência fornecida pelos serviços pagos. Um factor adicional que obrigou à utilização da Yahoo Finance é o facto de que alguns índices utilizados neste trabalho não estavam disponíveis na versão gratuita da Google.

Com os serviços gratuitos da Google e Yahoo, é possível obter em ficheiro CSV<sup>3</sup> os valores diários, semanais ou mensais, de cada índice de bolsa num dado período de tempo escolhido. Estes serviços, apesar de gratuitos, disponibilizam os valores dos índices desde a sua entrada em bolsa. No ficheiro disponibilizado são fornecidos os valores de abertura, máximo, mínimo, de fecho e volume de transacções de cada intervalo escolhido do índice.

Para o estudo aqui desenvolvido, decidiu-se utilizar os valores de fecho diário de cada índice para um período que se expande entre Outubro de 1988 e Setembro de 2008, obtendo-se mais de 5000 valores diários. Considera-se que esta amostra seja suficiente para a realização deste estudo.

Foram escolhidos 20 índices de várias indústrias, em que 15 são de empresas individuais cotadas em bolsa e 5 são índices de bolsas financeiras. Os critérios de escolha foram:

- Serem de conhecimento público;
- Estarem cotados na bolsa há 20 anos ou mais (excepção feita ao PSI20);
- A sua informação estar disponibilizada gratuitamente;
- Não possuírem uma sazonalidade evidente;
- Estarem representados em dólares americanos (US\$);

---

<sup>3</sup> Comma Separated Value

Com os critérios referidos em mente, os índices escolhidos foram:

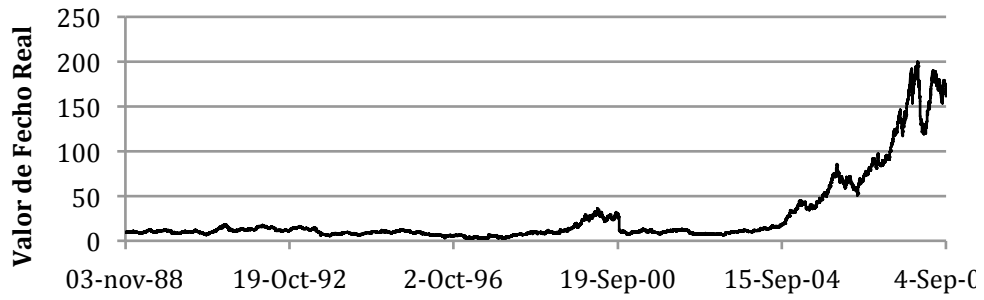
- **Apple Inc. (AAPL)** – A Apple Inc., juntamente com as suas subsidiárias, projecta, produz e vende computadores pessoais, aparelhos de comunicação móvel, e reprodutores de música e vídeo portáteis, assim como vende vários tipos de software relacionado, serviços, periféricos, e soluções de rede. Sector: Tecnologia. Indústria: Computadores Pessoais.
- **Boeing Co. (BA)** – A Boeing Company projecta, desenvolve, fabrica, vende e dá suporte a aviões comerciais e militares, satélites, defesa anti-míssil, voo espacial, e lança sistemas e serviços a nível mundial. Sector: Bens Industriais. Indústria: Aeroespacial/Defesa.
- **BP plc (BP)** – A BP plc fornece combustível para transporte, energia para aquecimento e luz, serviços de retalho, e produtos petroquímicos. Sector: Matéria Básica. Indústria: Petróleo e Gás Integrado.
- **Chevron Corp. (CVX)** – A Chevron Corporation opera como uma companhia integrada de energia a nível mundial. As operações de petróleo da companhia incluem exploração, desenvolvimento, produção e marketing de óleo não refinado e gás natural. Sector: Matéria Básica. Indústria: Petróleo e Gás Integrado.
- **Citigroup, Inc. (C)** – A Citigroup Inc. é uma companhia de serviços financeiros globais, fornece a consumidores, governos, corporações e instituições uma vasta gama de produtos financeiros e serviços. Sector: Financeiro. Indústria: Bancária.
- **Dell Inc. (DELL)** – A Dell Inc. projecta, desenvolve, fabrica, vende e dá suporte a sistemas de computador, assim como providencia serviços relacionados a nível mundial. Sector: Tecnologia. Indústria: Computadores Pessoais.
- **Exxon Mobil Corp. (XOM)** – A Exxon Mobil Corporation envolve-se na exploração, produção, transporte e venda de óleo não refinado e gás

natural. Também se envolve na produção, transporte e venda e produtos derivados do petróleo. Sector: Matéria Básica. Indústria: Petróleo e Gás Integrado.

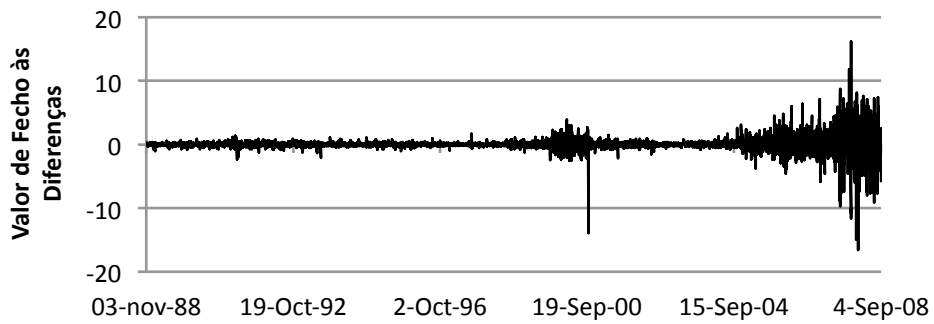
- **Ford Motor Co. (F)** – A Ford Motor Company projecta, desenvolve, produz e dá assistência a automóveis e camiões a nível mundial. Sector: Bens de Consumo. Indústria: Fabricante de Automóveis – Principal.
- **General Electric Co. (GE)** – A General Electric Company funciona como uma companhia mundial de serviços de tecnologia, média e financeiros. Sector: Conglomerados. Indústria: Conglomerados.
- **General Motors Corporation (GM)** – A General Motors Corporation foi a segunda maior construtora automóvel do mundo, cujas principal área de negócio é a produção de automóveis. De momento, devido ao processo de pedido de protecção contra falência, a “nova” GM, a General Motors Company, não tem activos financeiros em troca nos mercados cambiais. Sector: Bens de Consumo. Indústria: Fabricante de Automóveis – Principal.
- **Hewlett-Packard Company (HPQ)** – A Hewlett-Packard Company oferece variados produtos, tecnologias, software, soluções e serviços mundialmente. Sector: Tecnologia. Indústria: Sistemas Computacionais Diversificados.
- **International Business Machines Corp. (IBM)** – A International Business Machines Corp. desenvolve e fabrica produtos de tecnologias de informação (TI) e serviços à escala mundial. Sector: Tecnologia. Indústria: Sistemas Computacionais Diversificados.
- **JPMorgan Chase & Co. (JPM)** – A JPMorgan Chase & Co., uma companhia de *holdings* financeira, fornece variados serviços mundialmente. Sector: Financeiro. Indústria: Bancária.

- **Microsoft Corporation (MSFT)** – A Microsoft Corporation fornece soluções de software e hardware mundialmente. Sector: Tecnologia. Indústria: Software de Aplicação.
- **Sony Corporation (SNE)** – A Sony Corporation, juntamente com as suas subsidiárias, desenvolve, projecta, produz e vende equipamento electrónico, instrumentos, e aparelhos para mercados de consumo e industriais em todo o Mundo. Sector: Bens de Consumo. Indústria: Equipamento Electrónico.
- **Financial Times Stock Exchange (FTSE 100)** – Índice de bolsa de valores das 100 companhias do Reino Unido mais capitalizadas listadas na Bolsa de Valores de Londres.
- **National Association of Securities Dealers Automated Quotations (NASDAQ)** – Bolsa de valores Electrónica, constituída por um conjunto de correctores conectados por um sistema informático.
- **New York Stock Exchange (NYSE)** – Bolsa de Valores de Nova Iorque.
- **NYKKEI 225** – A Média de Acções NIKKEI 225 é o peso médio ponderado das 225 primeiras companhias japonesas listadas na Primeira Secção da Bolsa de Valores de Tokyo.
- **Portuguese Stock Index (PSI20)** – O PSI20 é o principal índice da Euronext Lisboa. É composto pelas 20 maiores empresas portuguesas no mercado de capitais.

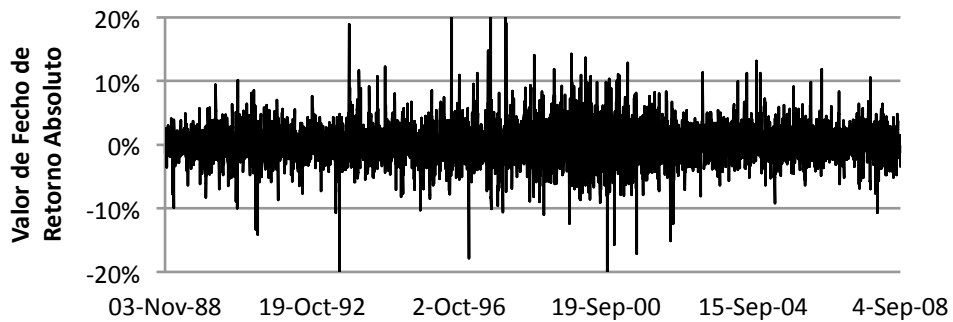
De seguida (Figuras 1–160) são demonstradas as representações gráficas dos índices apresentados, ao longo do tempo e o correspondente histográfico dos seus valores reais e transformados.



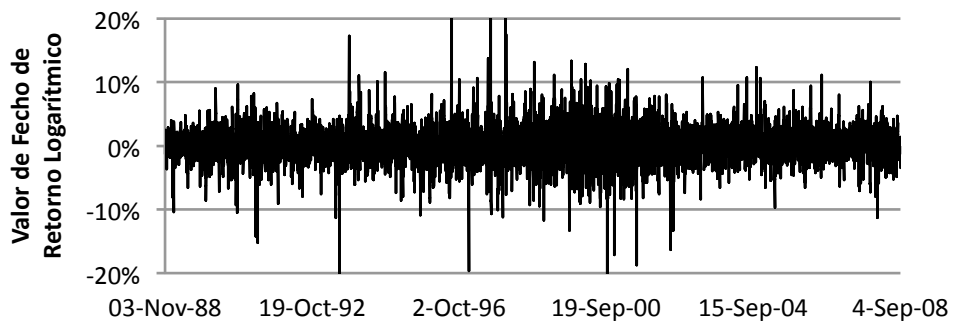
**Figura 1** – Representação temporal do valor real de fecho de AAPL



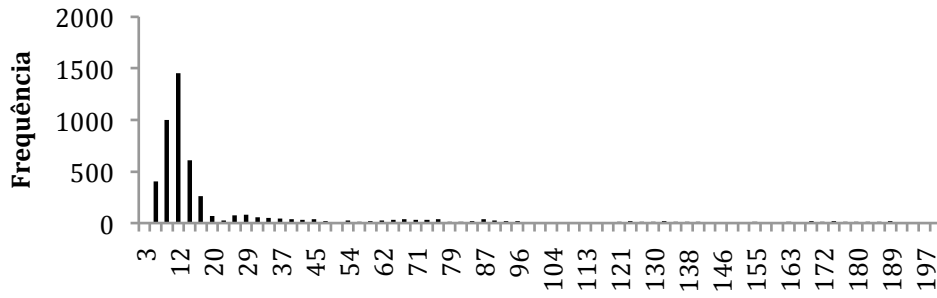
**Figura 2** – Representação temporal do valor às diferenças de fecho de AAPL



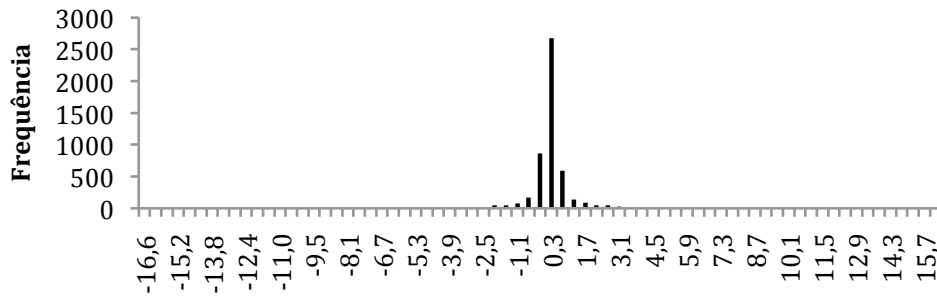
**Figura 3** – Representação temporal do valor de Retorno Bruto de fecho de AAPL



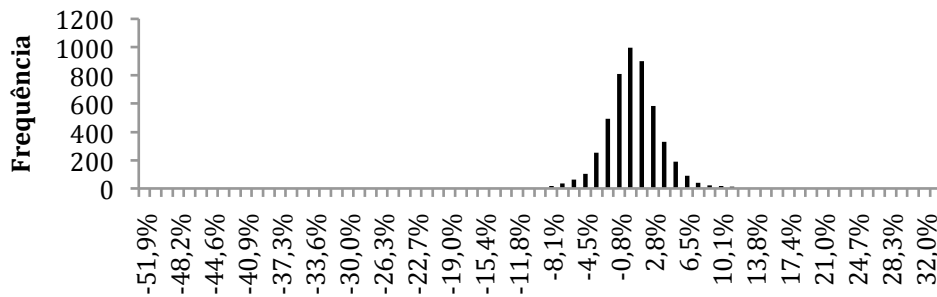
**Figura 4** – Representação temporal do valor de Retorno Logarítmico de fecho de AAPL



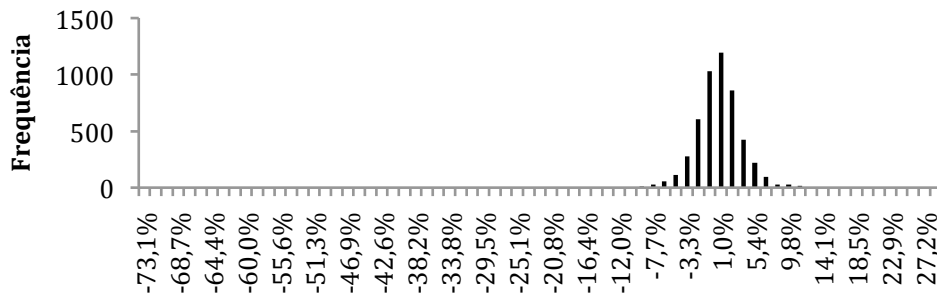
**Figura 5** – Representação histográfica do valor real de fecho de AAPL



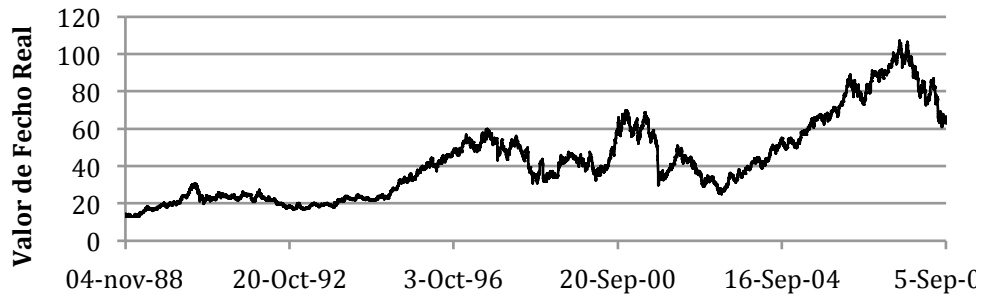
**Figura 6** – Representação histográfica do valor às diferenças de fecho de AAPL



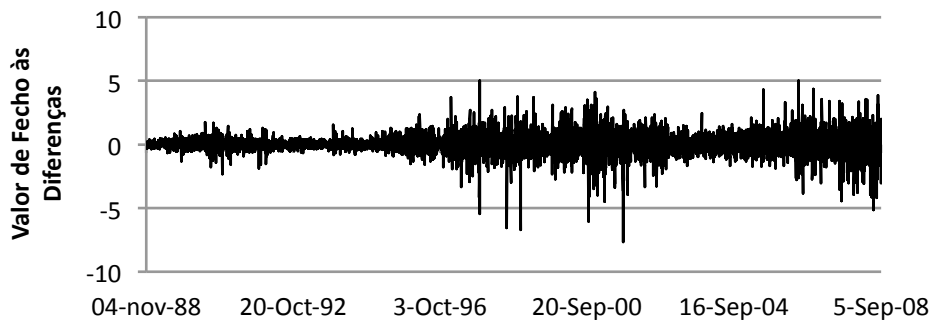
**Figura 7** – Representação histográfica do valor de Retorno Bruto de fecho de AAPL



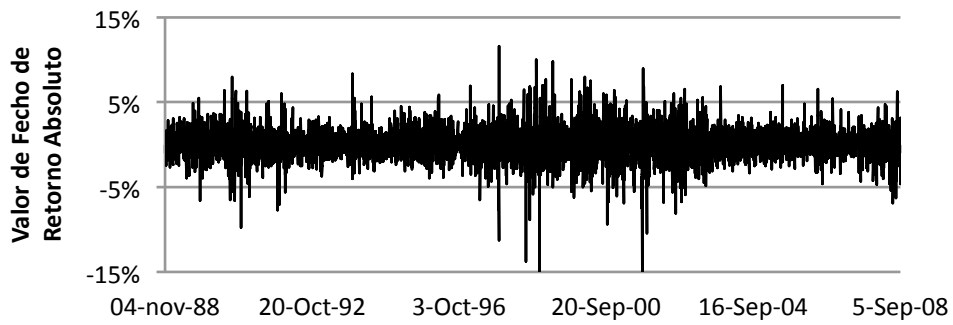
**Figura 8** – Representação histográfica do valor de Retorno Logarítmico de AAPL



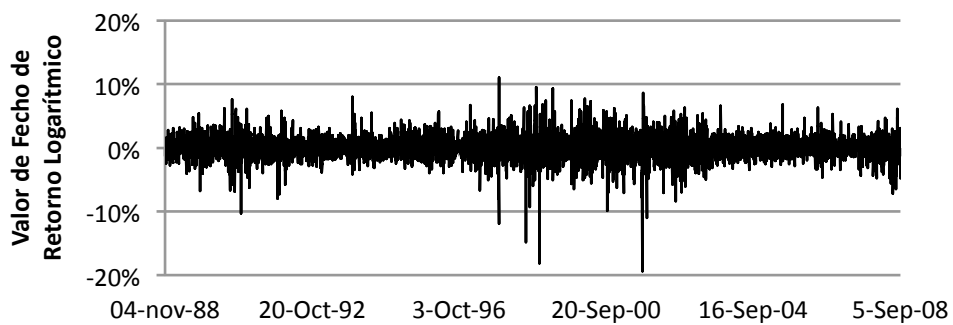
**Figura 9** – Representação temporal do valor real de fecho de BA



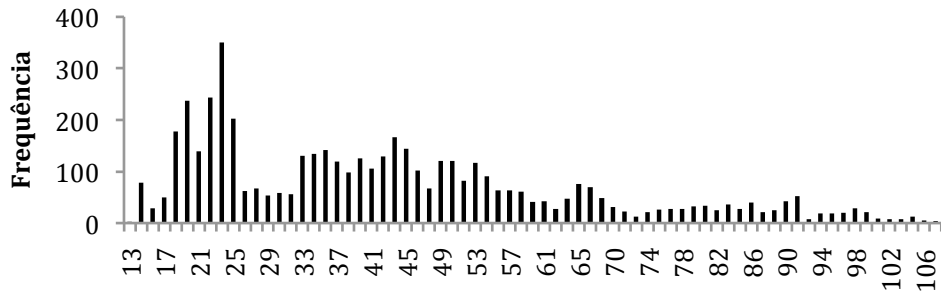
**Figura 10** – Representação temporal do valor às diferenças de fecho de BA



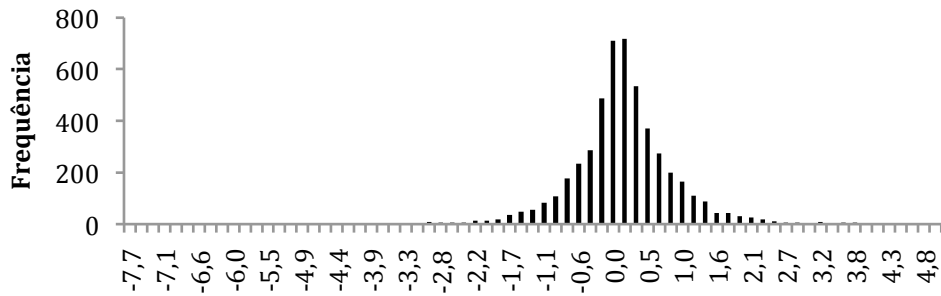
**Figura 11** – Representação temporal do valor de Retorno Bruto de fecho de BA



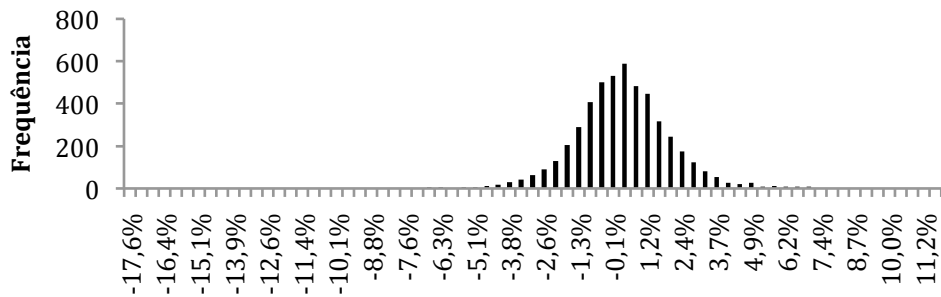
**Figura 12** – Representação temporal do valor de Retorno Logarítmico de BA



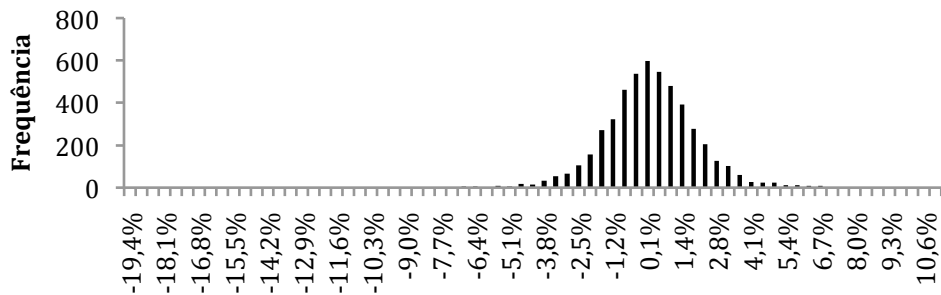
**Figura 13** – Representação histográfica do valor real de fecho de BA



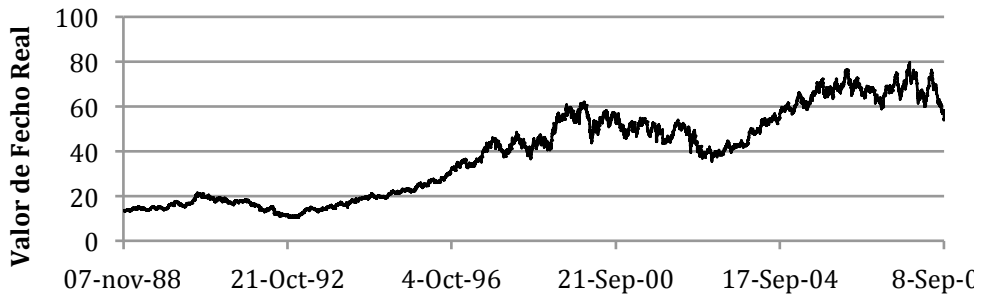
**Figura 14** – Representação histográfica do valor às diferenças de fecho de BA



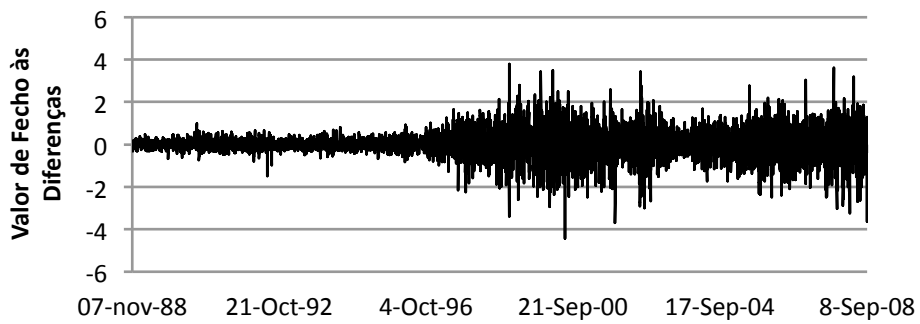
**Figura 15** – Representação histográfica do valor de Retorno Bruto de fecho de BA



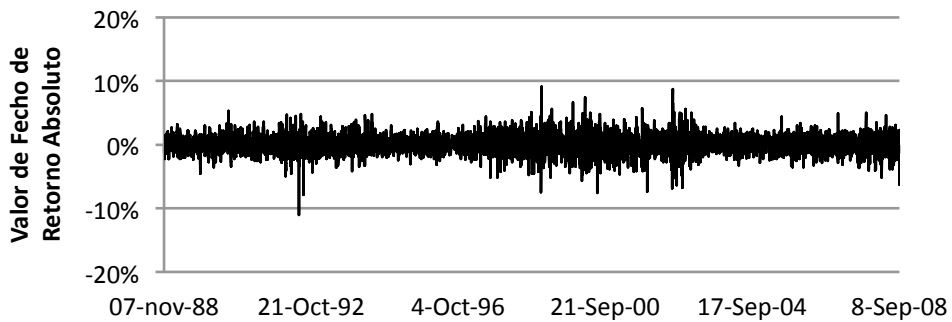
**Figura 16** – Representação histográfica do valor de Retorno Logarítmico de fecho de BA



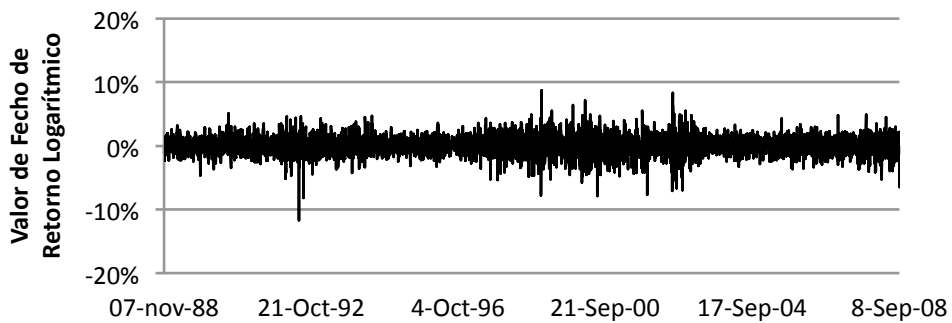
**Figura 17** – Representação temporal do valor real de fecho de BP



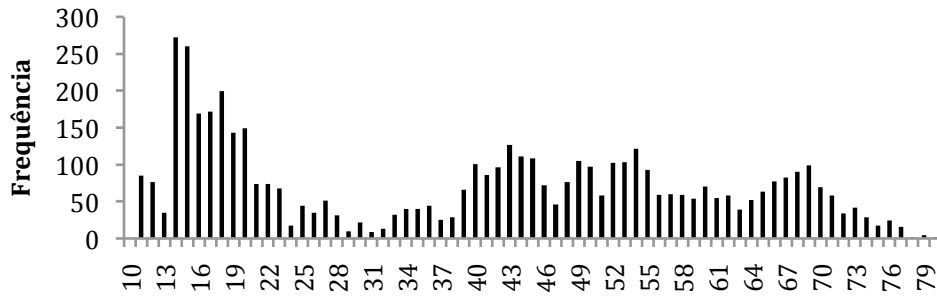
**Figura 18** – Representação temporal do valor às diferenças de fecho de BP



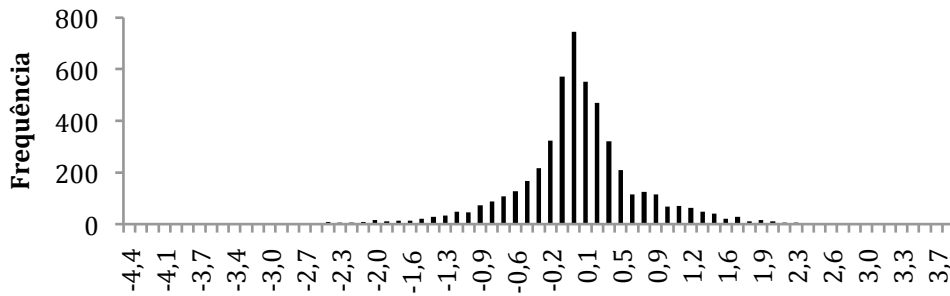
**Figura 19** – Representação temporal do valor de Retorno Bruto de fecho de BP



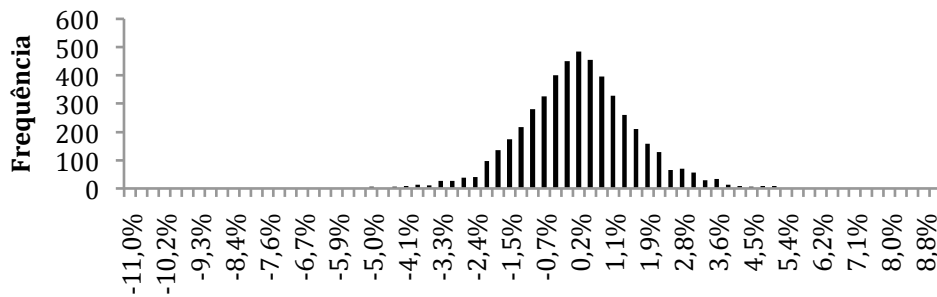
**Figura 20** – Representação temporal do valor de Retorno Logarítmico de fecho de BP



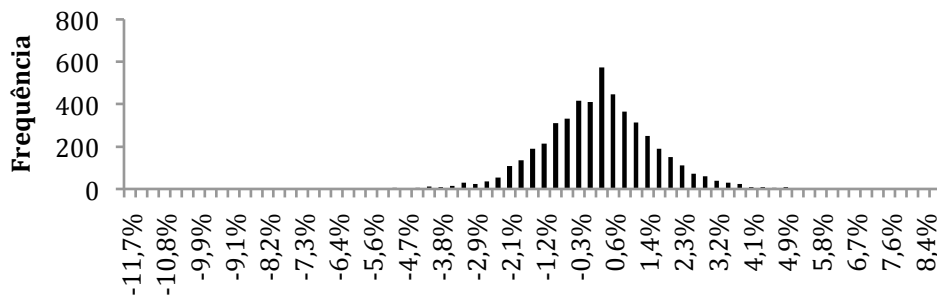
**Figura 21** – Representação histográfica do valor real de fecho de BP



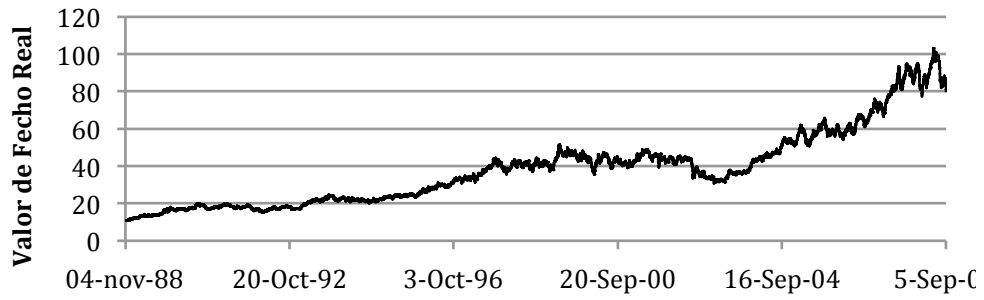
**Figura 22** – Representação histográfica do valor às diferenças de fecho de BP



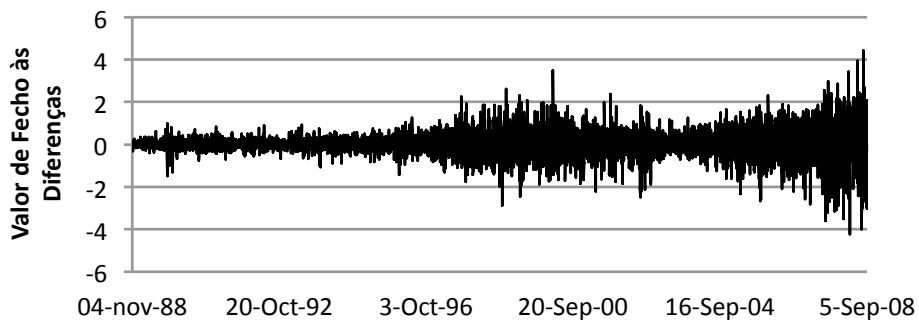
**Figura 23** – Representação histográfica do valor de Retorno Bruto de fecho de BP



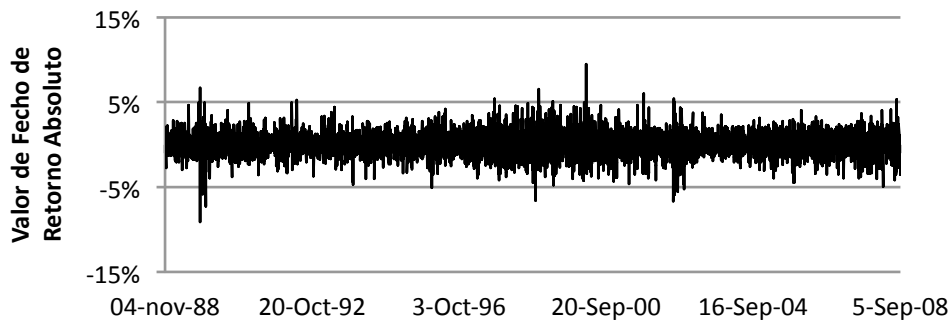
**Figura 24** – Representação histográfica do valor de Retorno Logarítmico de fecho de BP



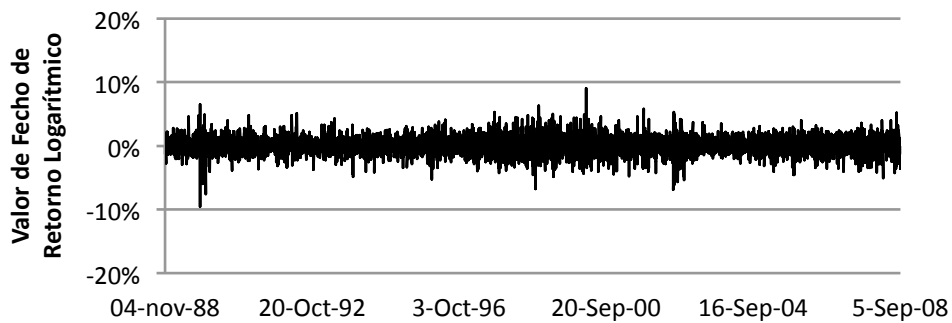
**Figura 25** – Representação temporal do valor real de fecho de CVX



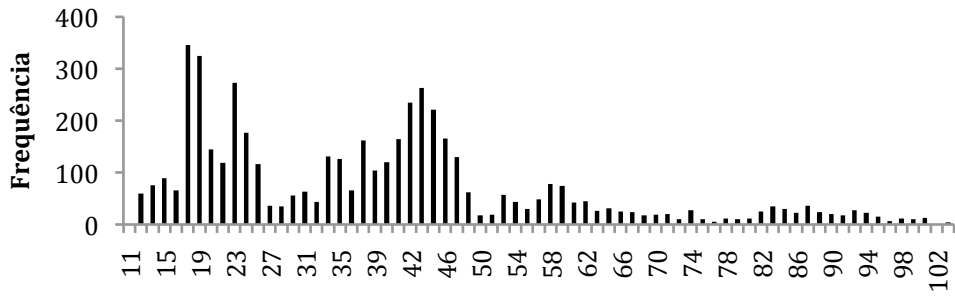
**Figura 26** – Representação temporal do valor às diferenças de fecho de CVX



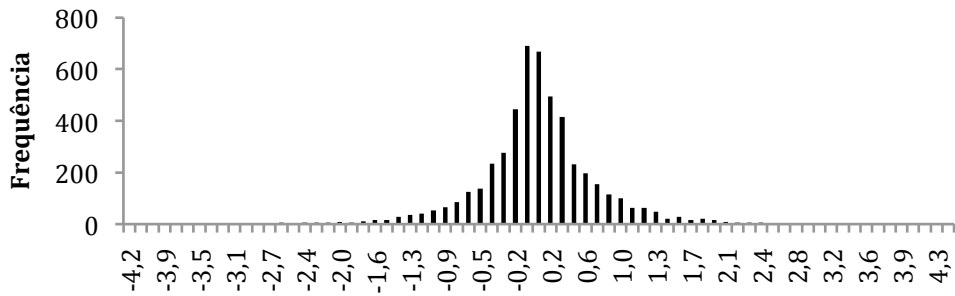
**Figura 27** – Representação temporal do valor de Retorno Bruto de fecho de CVX



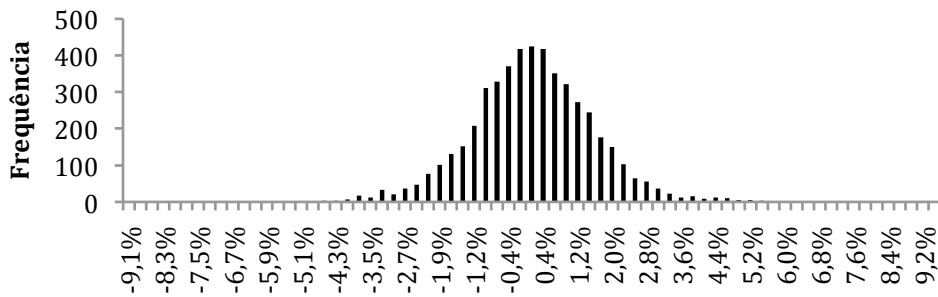
**Figura 28** – Representação temporal do valor de Retorno Logarítmico de CVX



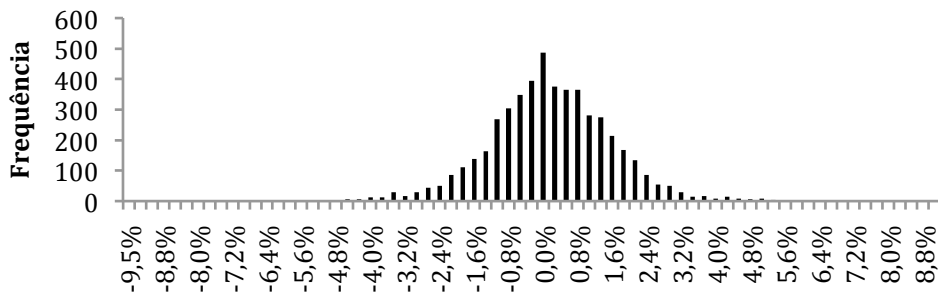
**Figura 29** – Representação histográfica do valor real de fecho de CVX



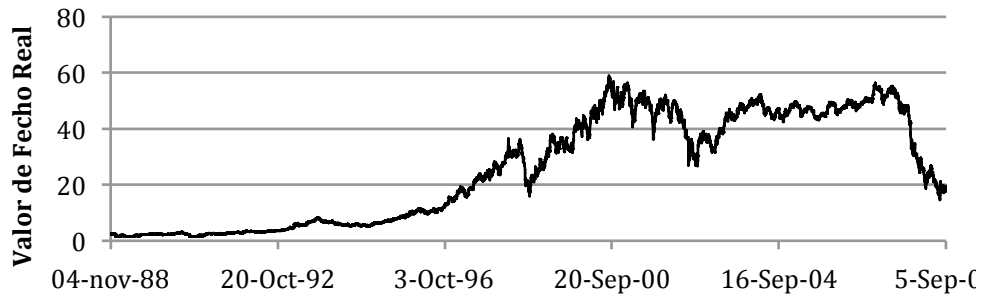
**Figura 30** – Representação histográfica do valor às diferenças de fecho de CVX



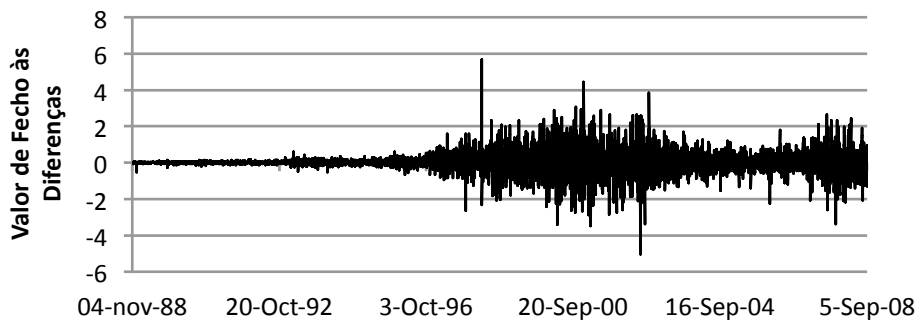
**Figura 31** – Representação histográfica do valor de Retorno Bruto de fecho de CVX



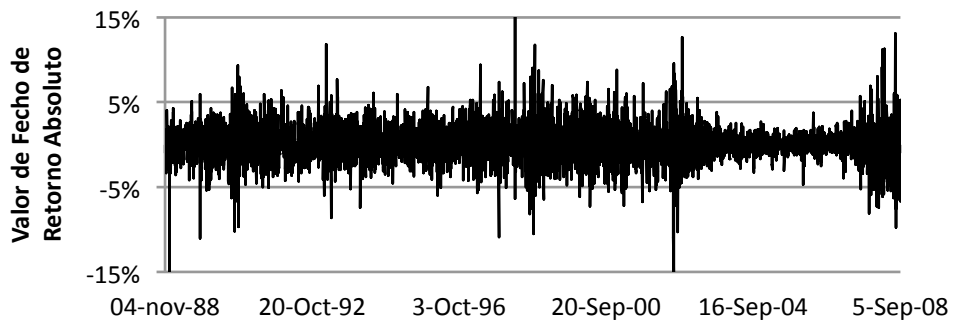
**Figura 32** – Representação histográfica do valor de Retorno Logaritmico de fecho de CVX



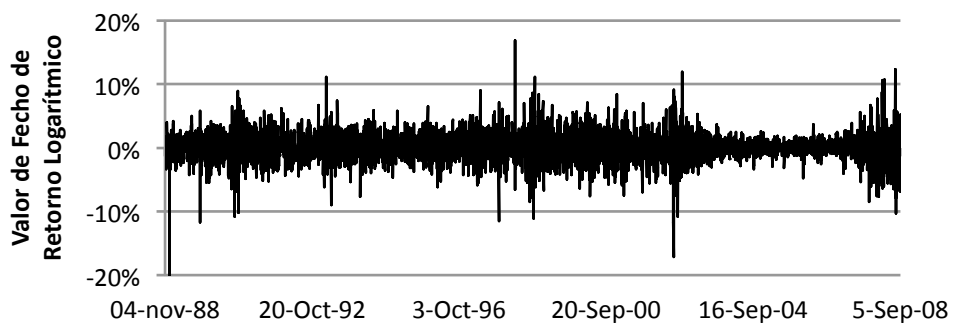
**Figura 33** – Representação temporal do valor real de fecho de C



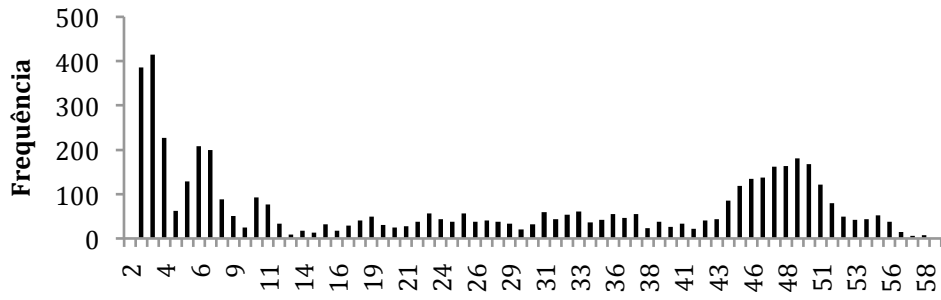
**Figura 34** – Representação temporal do valor às diferenças de fecho de C



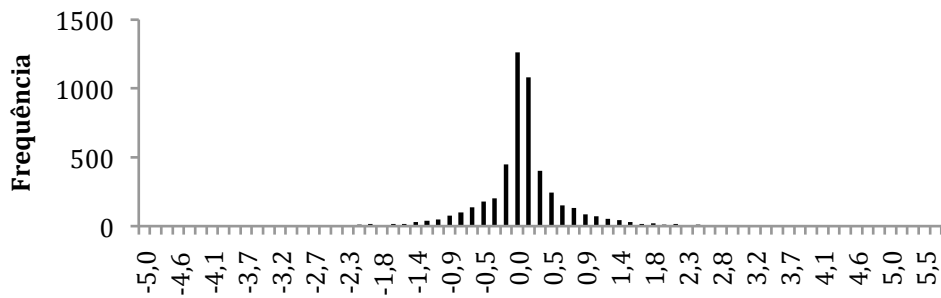
**Figura 35** – Representação temporal do valor de Retorno Bruto de fecho de C



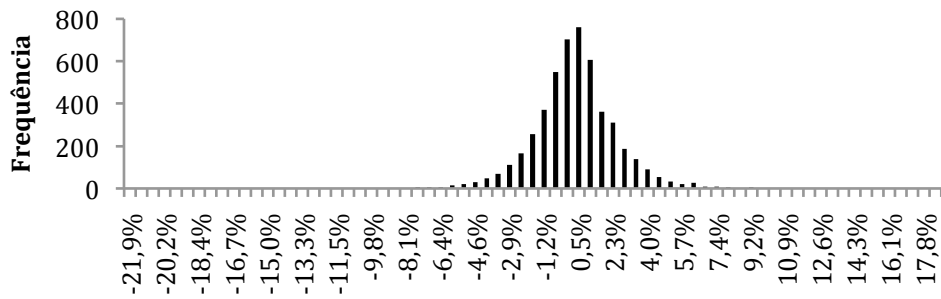
**Figura 36** – Representação temporal do valor de Retorno Logarítmico de fecho de C



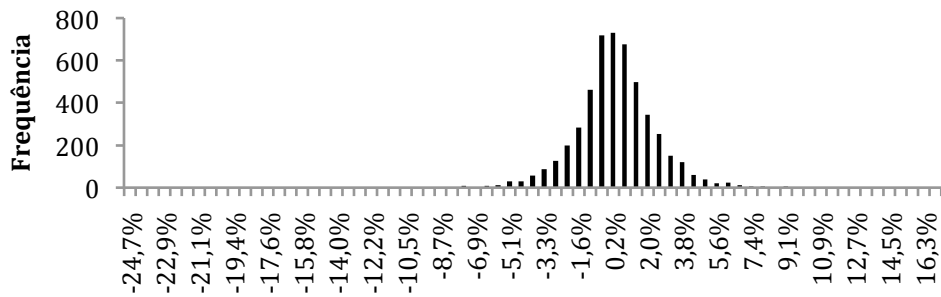
**Figura 37** – Representação histográfica do valor real de fecho de C



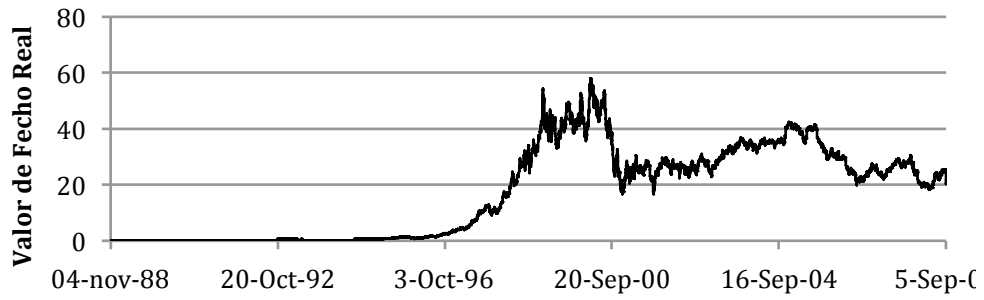
**Figura 38** – Representação histográfica do valor às diferenças de fecho de C



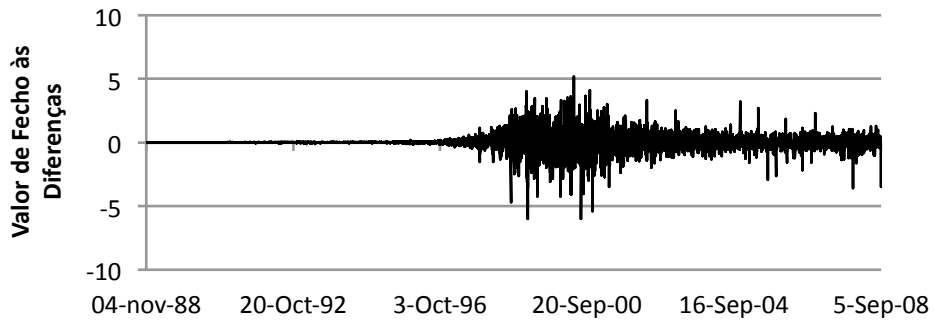
**Figura 39** – Representação histográfica do valor de Retorno Bruto de fecho de C



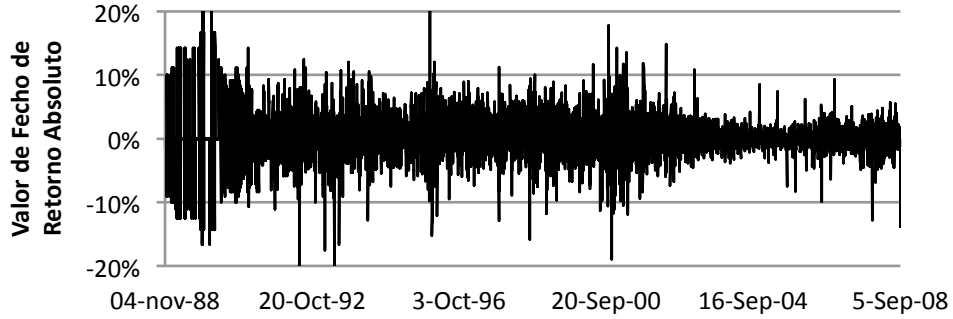
**Figura 40** – Representação histográfica do valor de Retorno Logarítmico de fecho de C



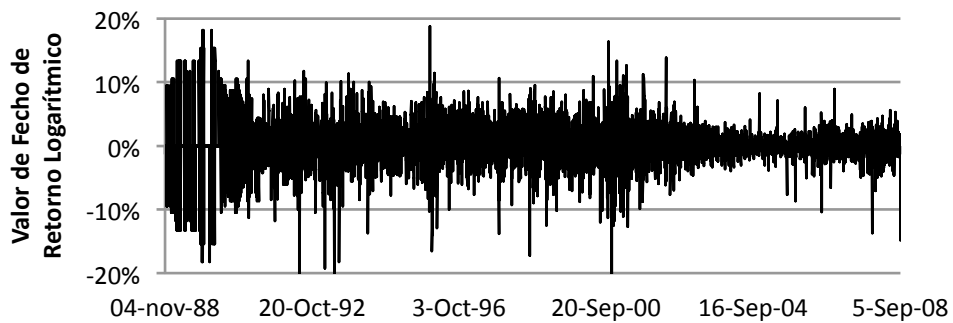
**Figura 41** – Representação temporal do valor real de fecho de DELL



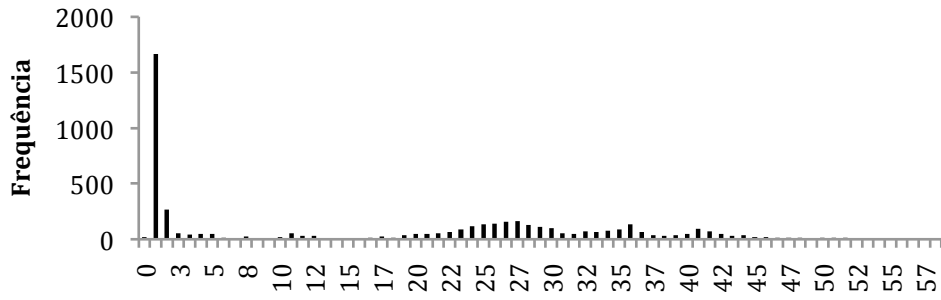
**Figura 42** – Representação temporal do valor às diferenças de fecho de DELL



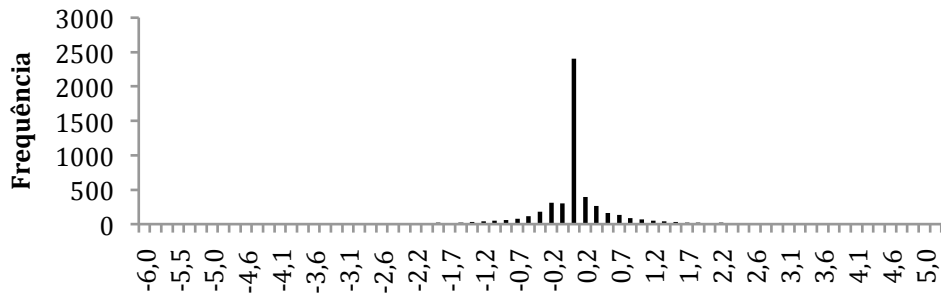
**Figura 43** – Representação temporal do valor de Retorno Bruto de fecho de DELL



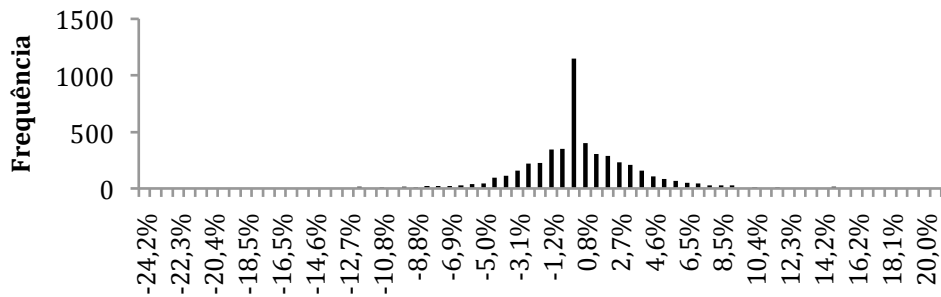
**Figura 44** – Representação temporal do valor de Retorno Logarítmico de fecho de DELL



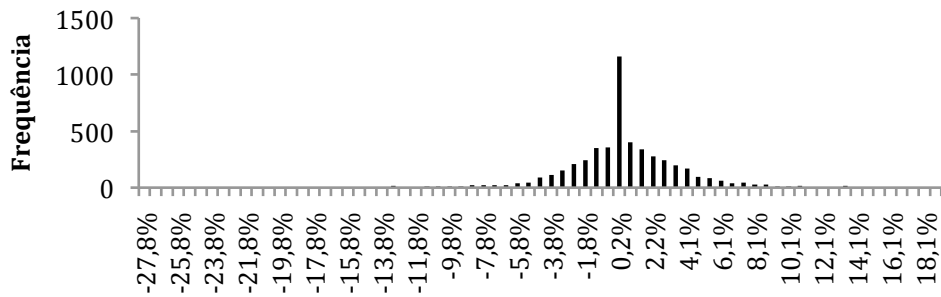
**Figura 45** – Representação histográfica do valor real de fecho de DELL



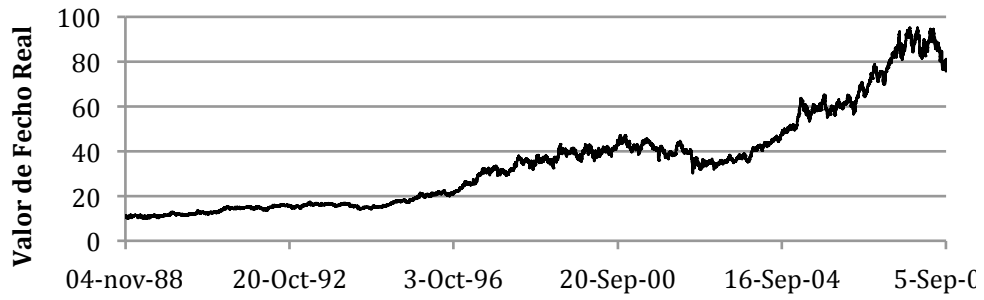
**Figura 46** – Representação histográfica do valor às diferenças de fecho de DELL



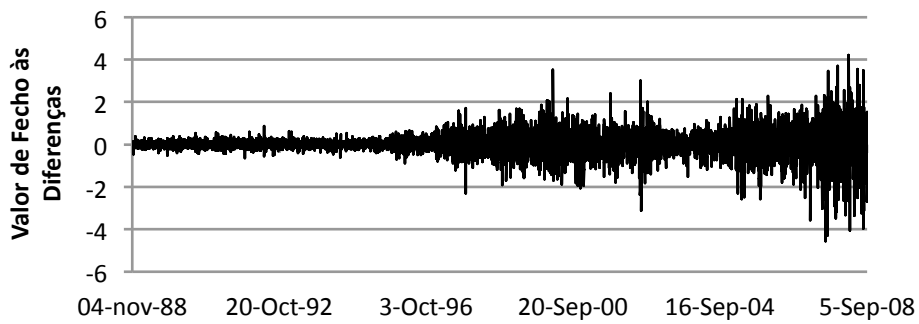
**Figura 47** – Representação histográfica do valor de Retorno Bruto de fecho de DELL



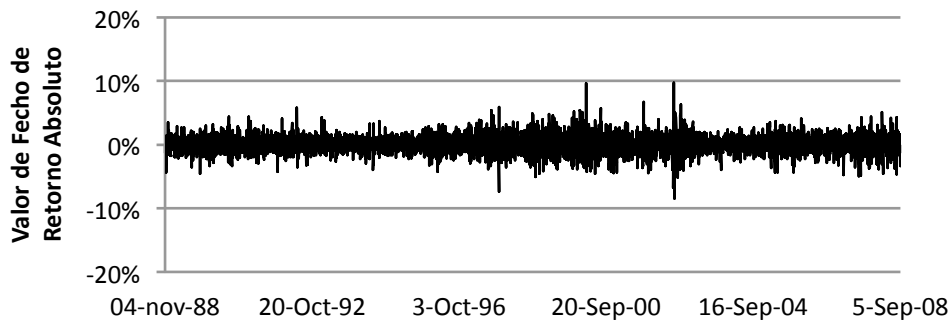
**Figura 48** – Representação histográfica do valor de Retorno Logarítmico de fecho de DELL



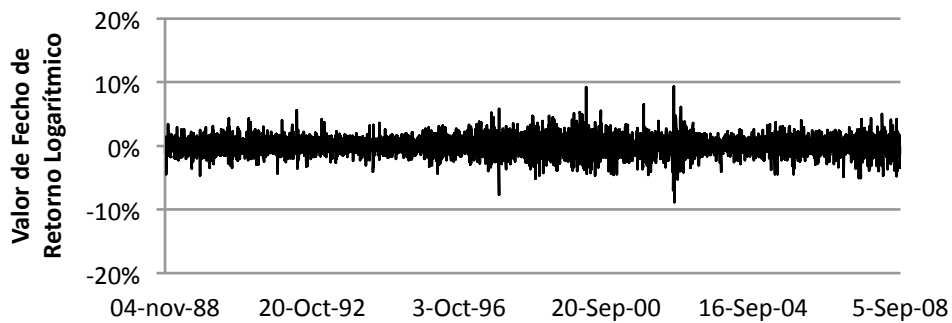
**Figura 49** – Representação temporal do valor real de fecho de XOM



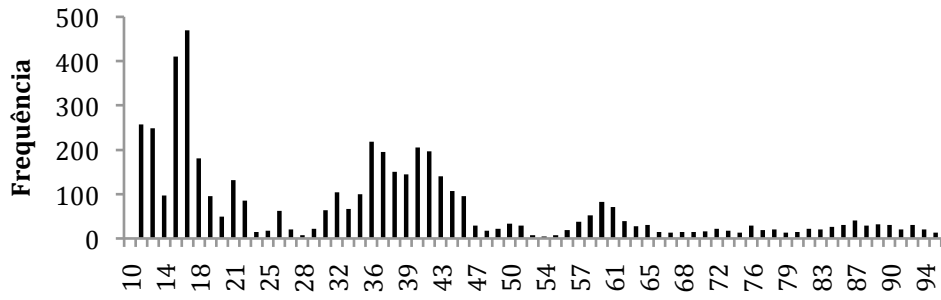
**Figura 50** – Representação temporal do valor às diferenças de fecho de XOM



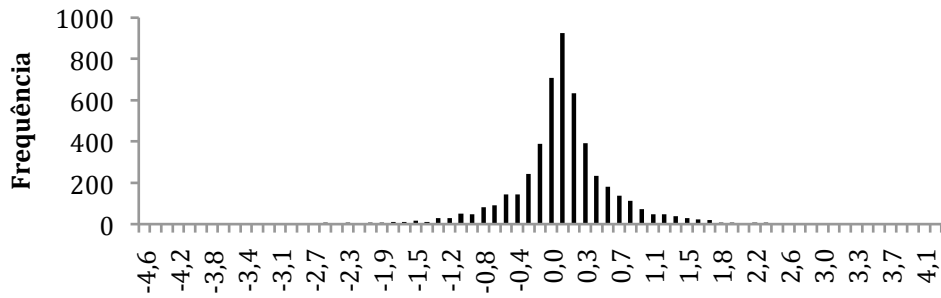
**Figura 51** – Representação temporal do valor de Retorno Bruto de fecho de XOM



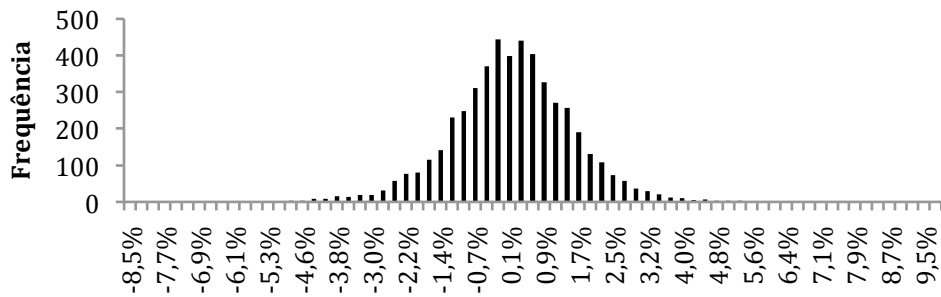
**Figura 52** – Representação temporal do valor de Retorno Logarítmico de XOM



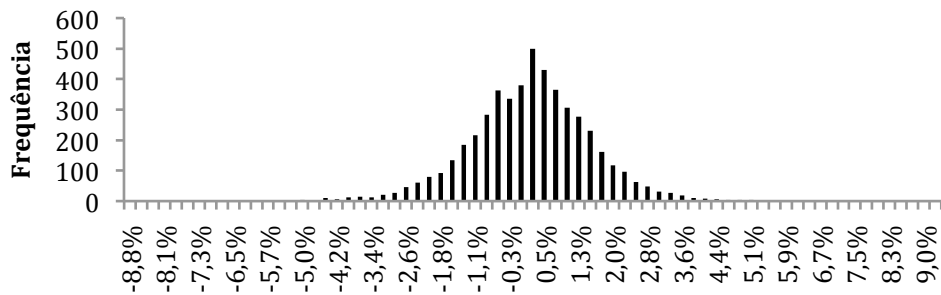
**Figura 53** – Representação histográfica do valor real de fecho de XOM



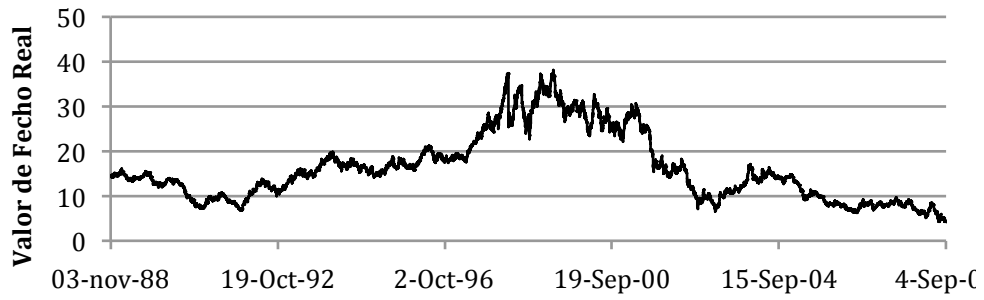
**Figura 54** – Representação histográfica do valor às diferenças de fecho de XOM



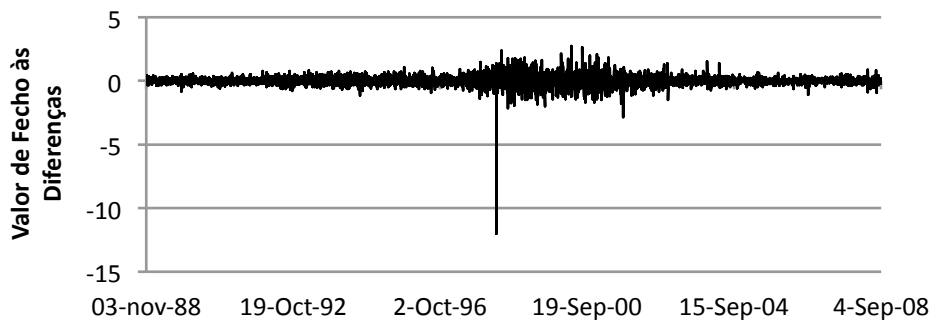
**Figura 55** – Representação histográfica do valor de Retorno Bruto de fecho de XOM



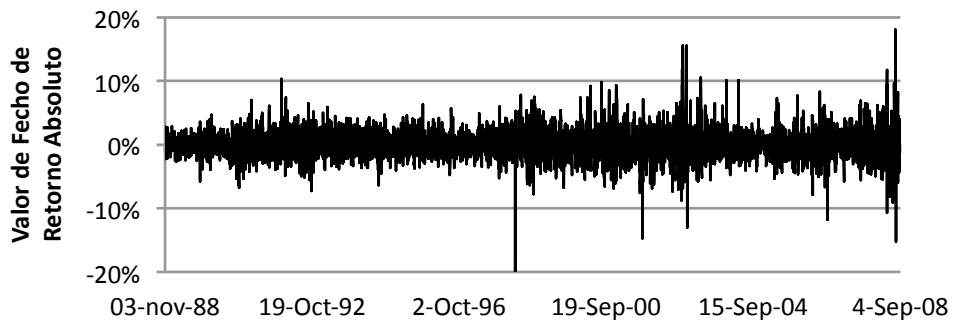
**Figura 56** – Representação histográfica do valor de Retorno Logarítmico de fecho de XOM



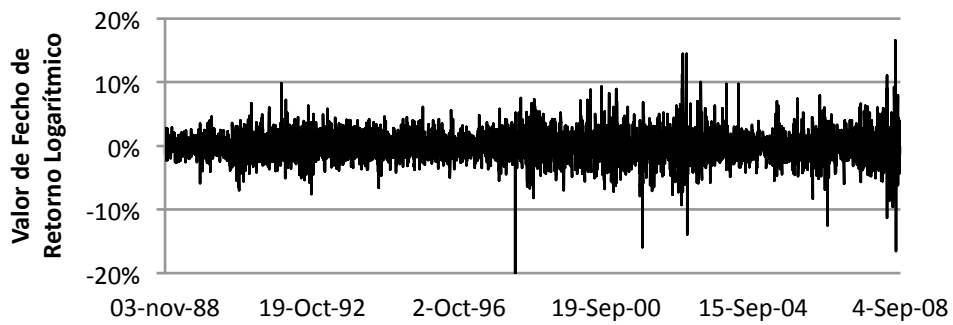
**Figura 57** – Representação temporal do valor real de fecho de F



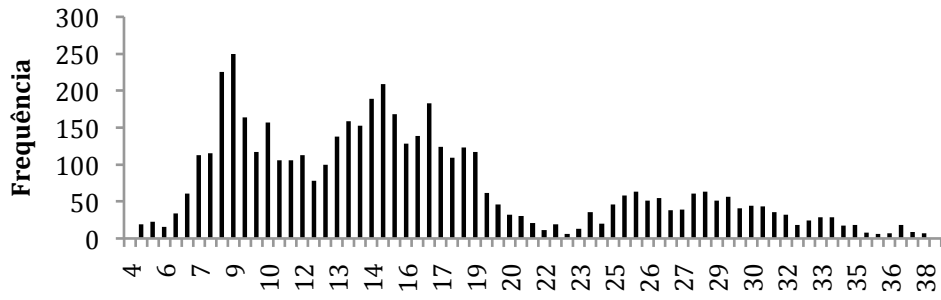
**Figura 58** – Representação temporal do valor às diferenças de fecho de F



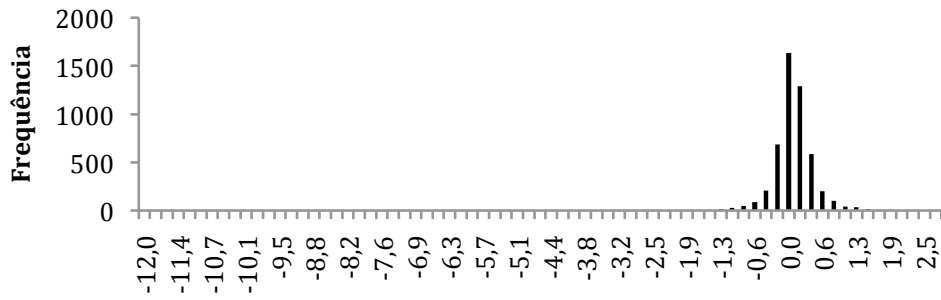
**Figura 59** – Representação temporal do valor de Retorno Bruto de fecho de F



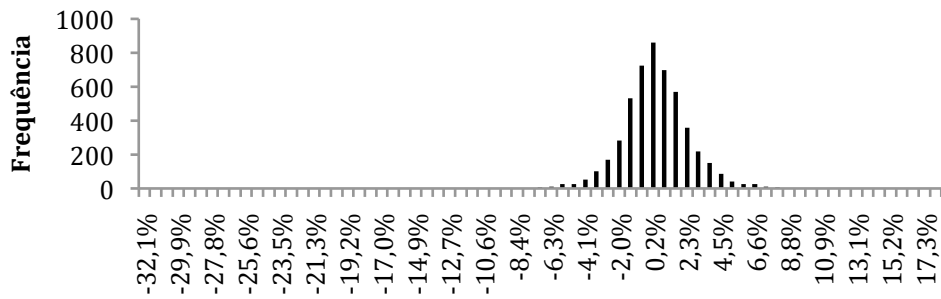
**Figura 60** – Representação temporal do valor de Retorno Logarítmico de fecho de F



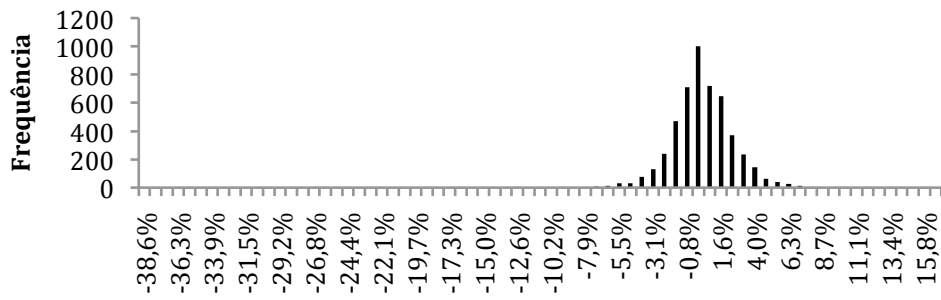
**Figura 61** – Representação histográfica do valor real de fecho de F



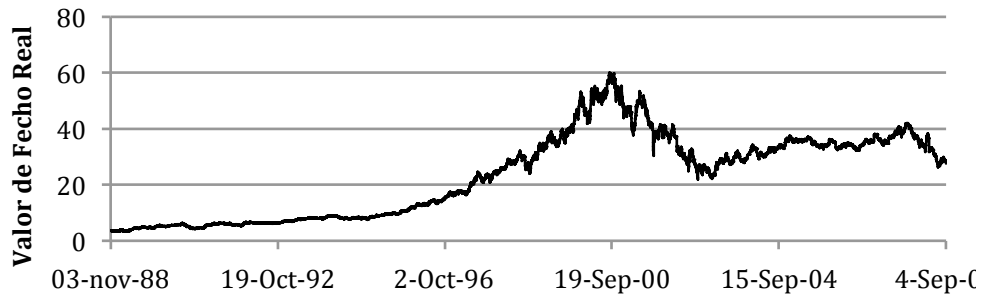
**Figura 62** – Representação histográfica do valor às diferenças de fecho de F



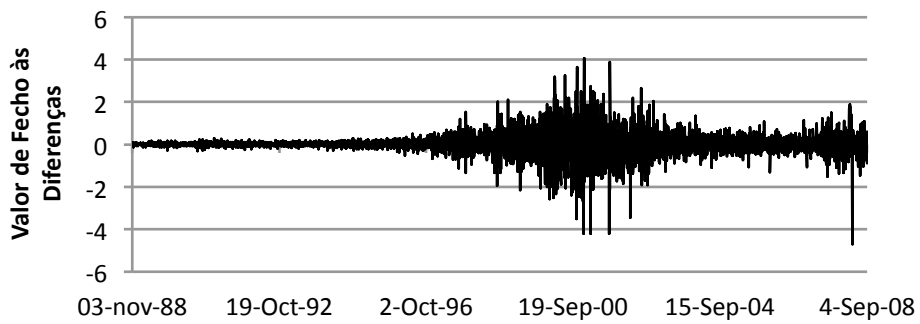
**Figura 63** – Representação histográfica do valor de Retorno Bruto de fecho de F



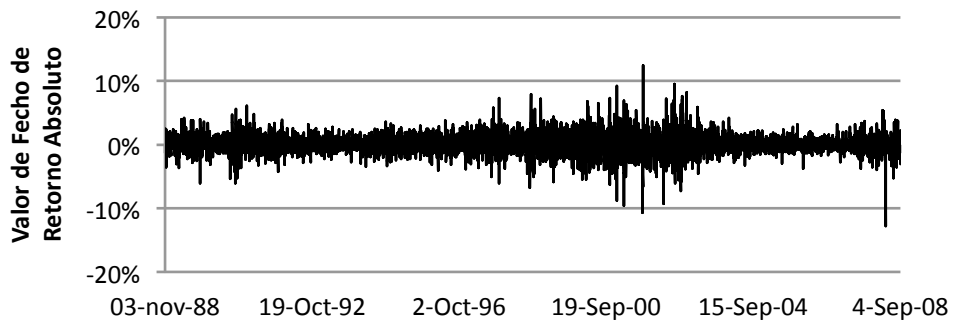
**Figura 64** – Representação histográfica do valor de Retorno Logarítmico de fecho de F



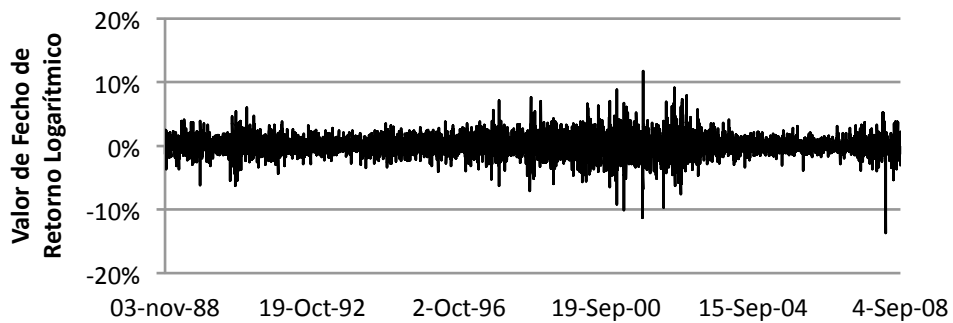
**Figura 65** – Representação temporal do valor real de fecho de GE



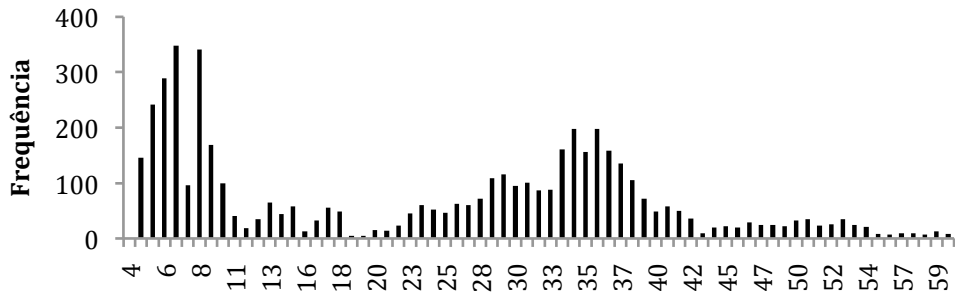
**Figura 66** – Representação temporal do valor às diferenças de fecho de GE



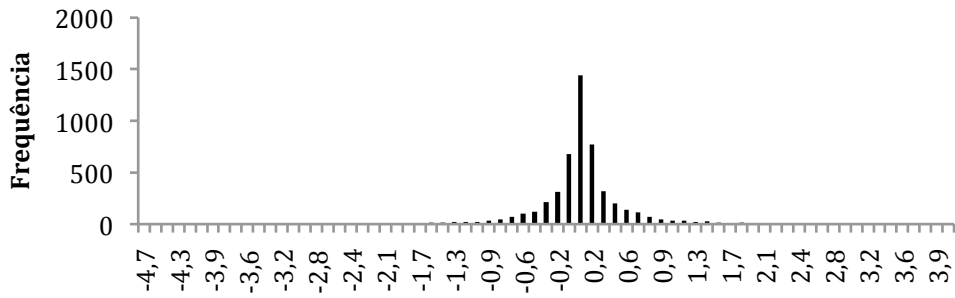
**Figura 67** – Representação temporal do valor de Retorno Bruto de fecho de GE



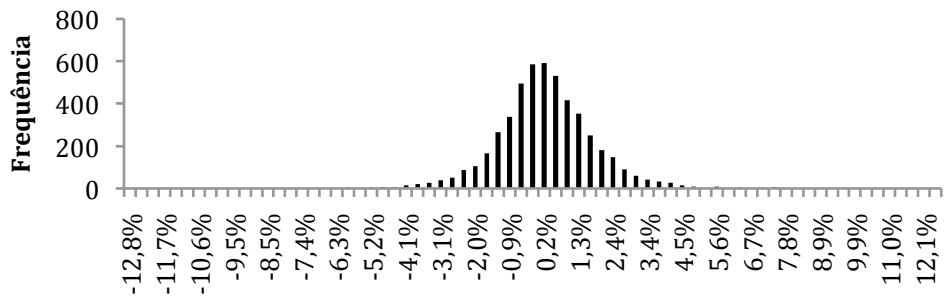
**Figura 68** – Representação temporal do valor de Retorno Logarítmico de GE



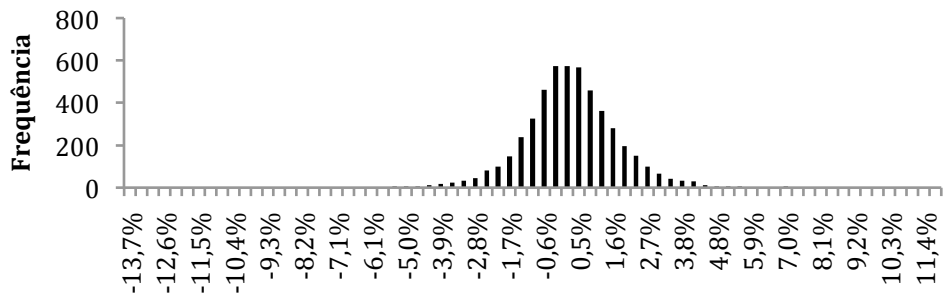
**Figura 69** – Representação histográfica do valor real de fecho de GE



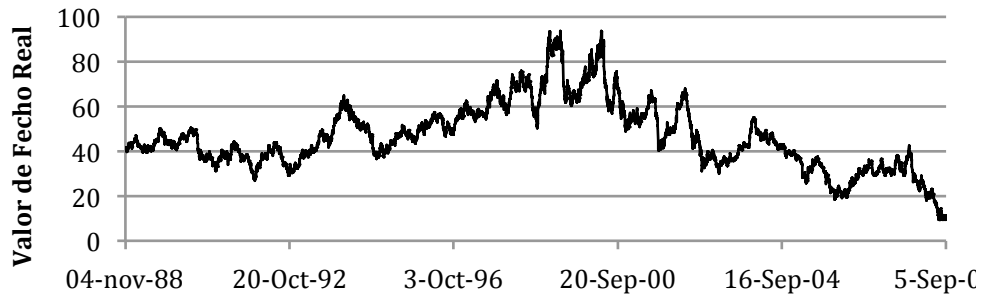
**Figura 70** – Representação histográfica do valor às diferenças de fecho de GE



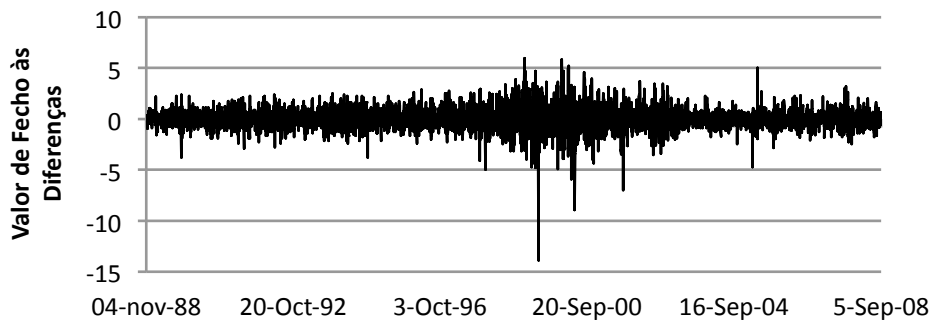
**Figura 71** – Representação histográfica do valor de Retorno Bruto de fecho de GE



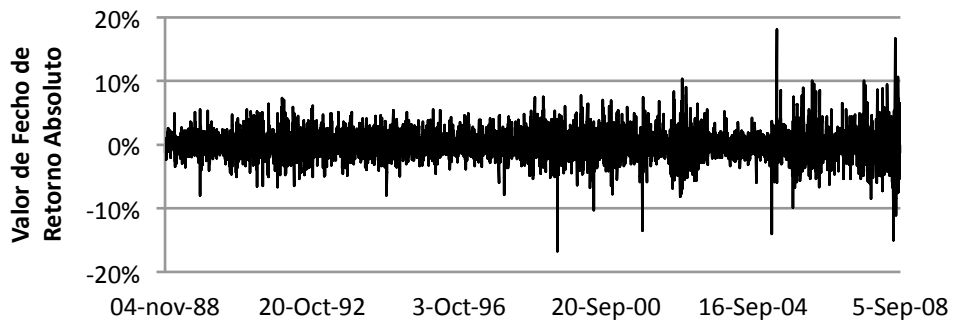
**Figura 72** – Representação histográfica do valor de Retorno Logarítmico de fecho de GE



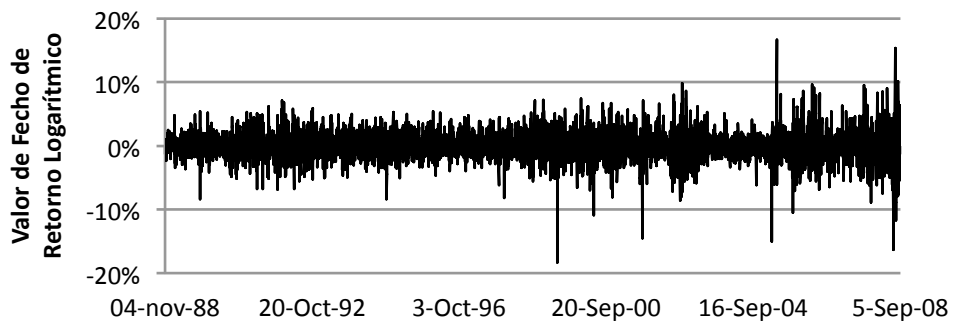
**Figura 73** – Representação temporal do valor real de fecho de GM



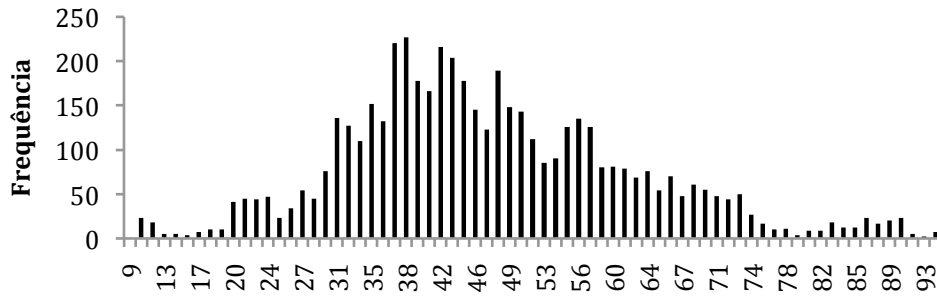
**Figura 74** – Representação temporal do valor às diferenças de fecho de GM



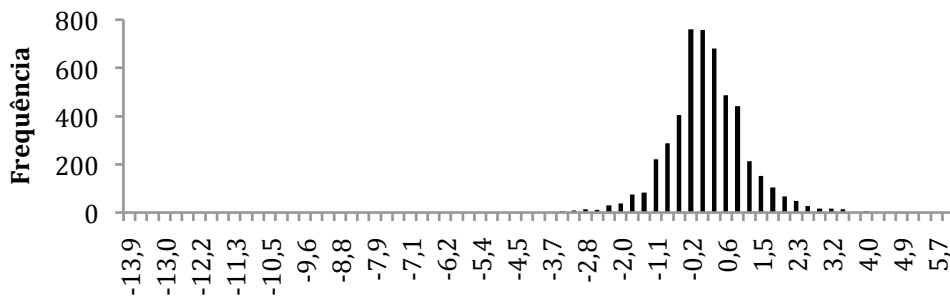
**Figura 75** – Representação temporal do valor de Retorno Bruto de fecho de GM



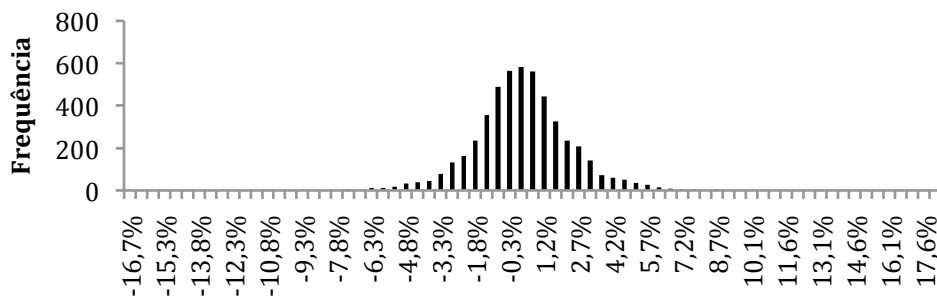
**Figura 76** – Representação temporal do valor de Retorno Logarítmico de fecho de GM



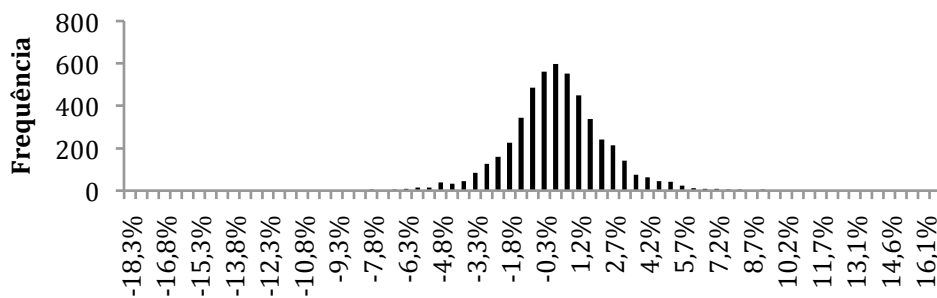
**Figura 77** – Representação histográfica do valor real de fecho de GM



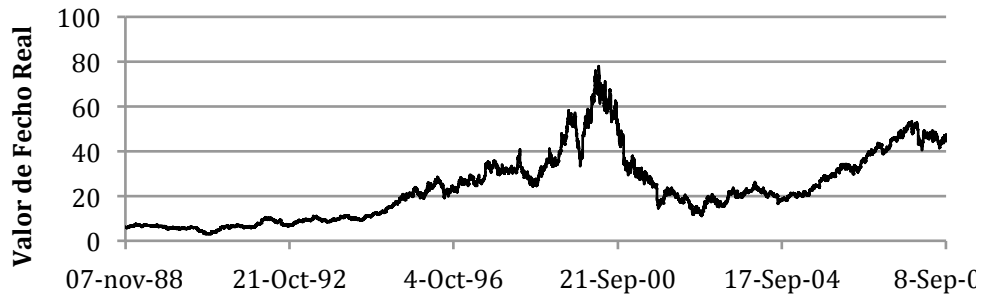
**Figura 78** – Representação histográfica do valor às diferenças de fecho de GM



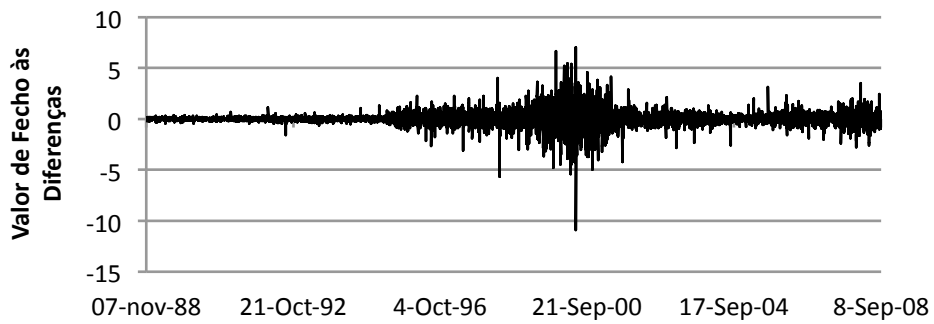
**Figura 79** – Representação histográfica do valor de Retorno Bruto de fecho de GM



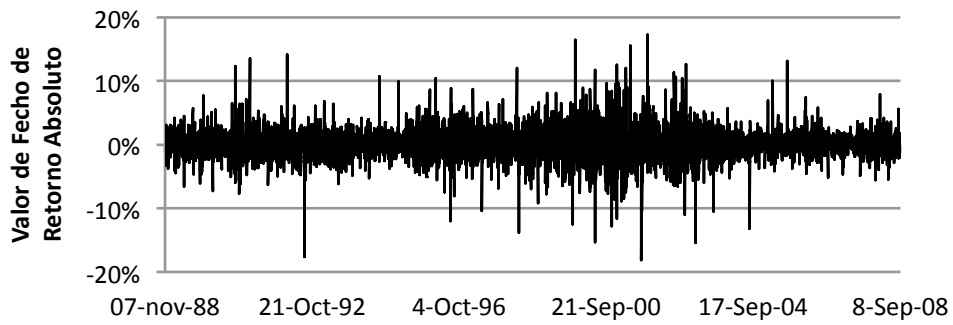
**Figura 80** – Representação histográfica do valor de Retorno Logarítmico de fecho de GM



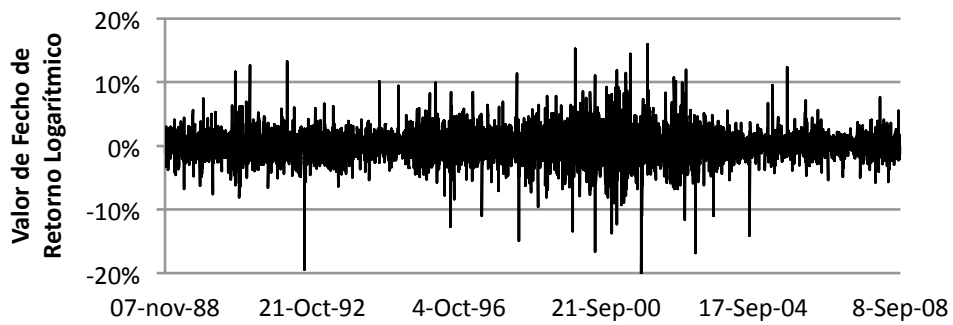
**Figura 81** – Representação temporal do valor real de fecho de HPQ



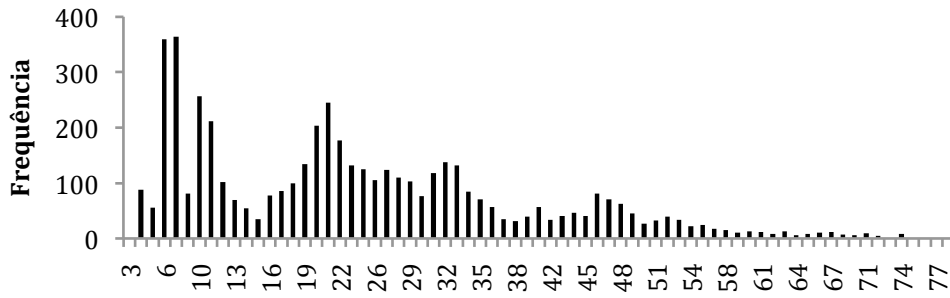
**Figura 82** – Representação temporal do valor às diferenças de fecho de HPQ



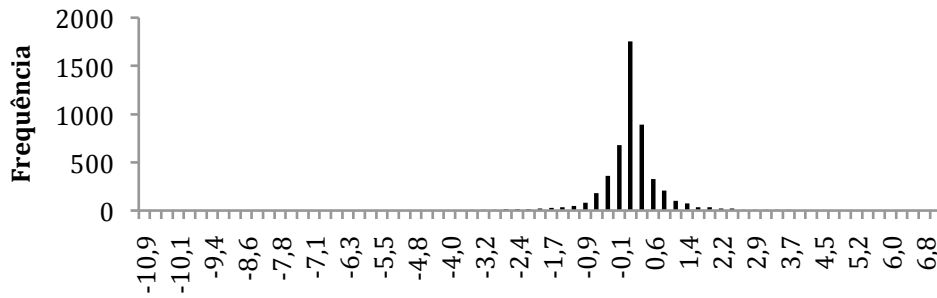
**Figura 83** – Representação temporal do valor de Retorno Bruto de fecho de HPQ



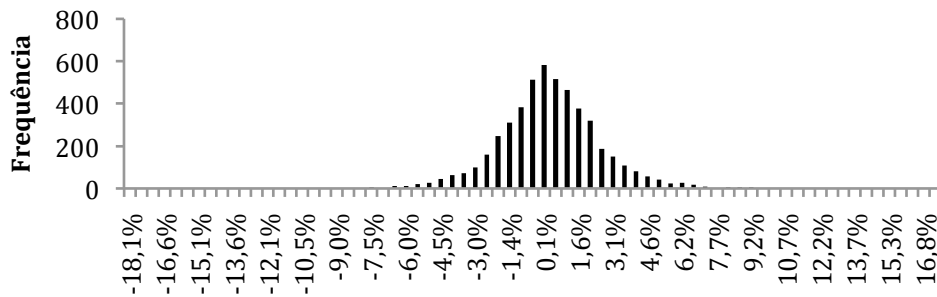
**Figura 84** – Representação temporal do valor de Retorno Logarítmico de HPQ



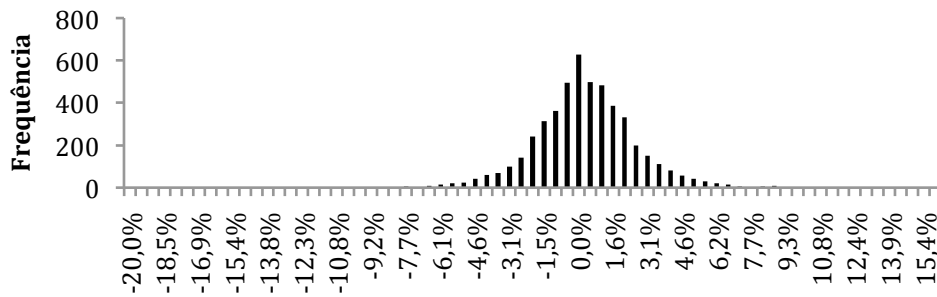
**Figura 85** – Representação histográfica do valor real de fecho de HPQ



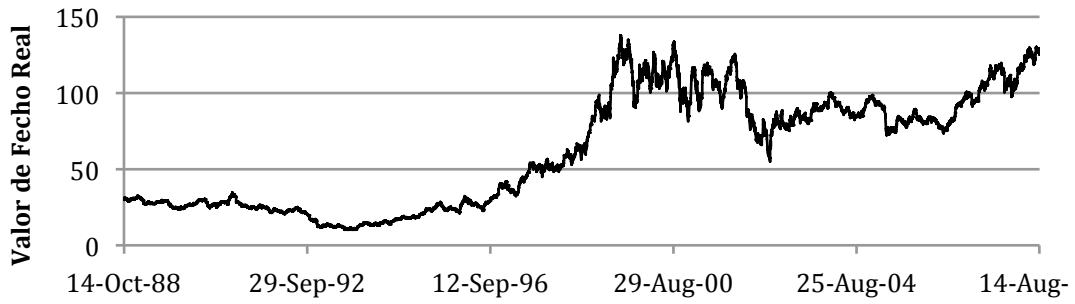
**Figura 86** – Representação histográfica do valor às diferenças de fecho de HPQ



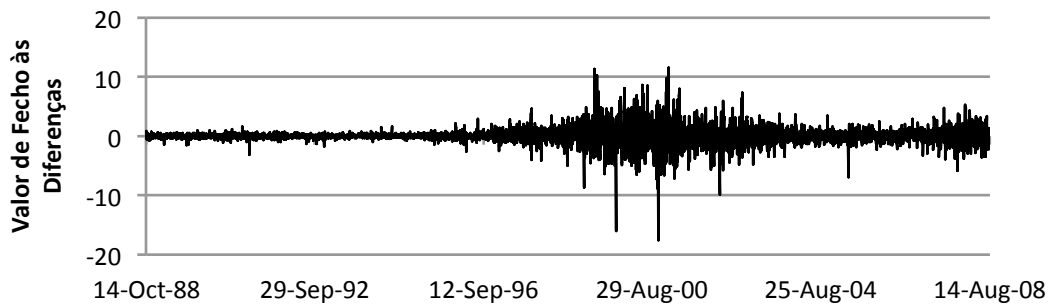
**Figura 87** – Representação histográfica do valor de Retorno Bruto de fecho de HPQ



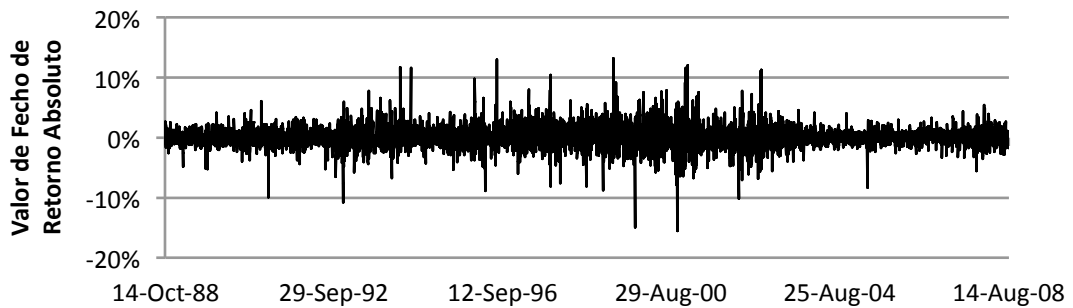
**Figura 88** – Representação histográfica do valor de Retorno Logarítmico de fecho de HPQ



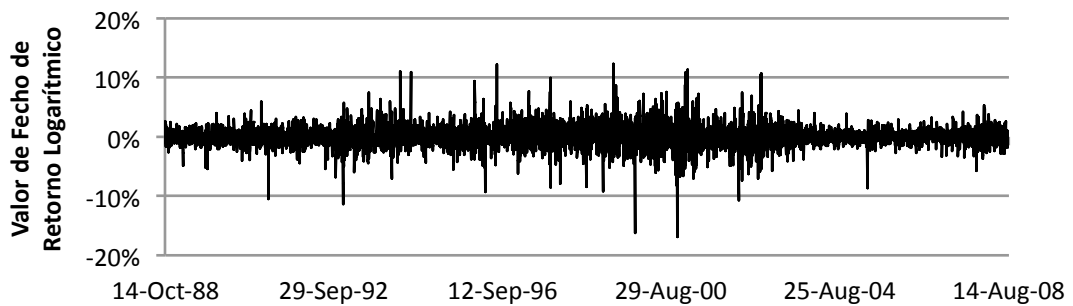
**Figura 89** – Representação temporal do valor real de fecho de IBM



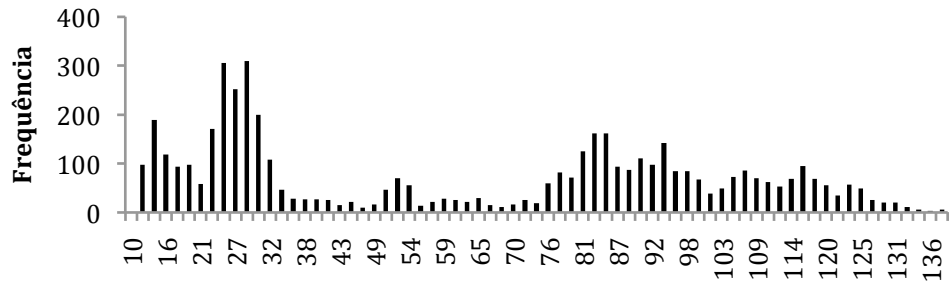
**Figura 90** – Representação temporal do valor às diferenças de fecho de IBM



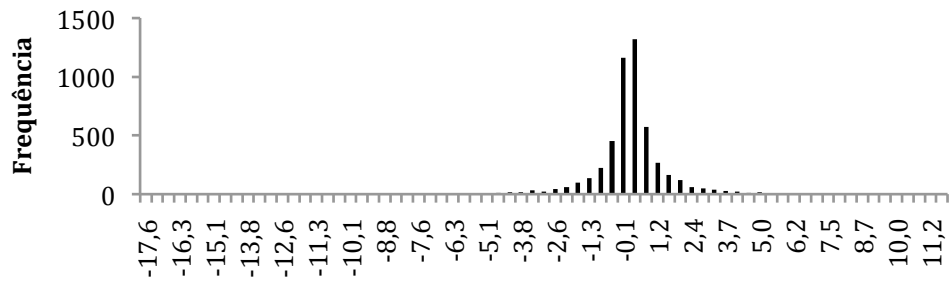
**Figura 91** – Representação temporal do valor de Retorno Bruto de fecho de IBM



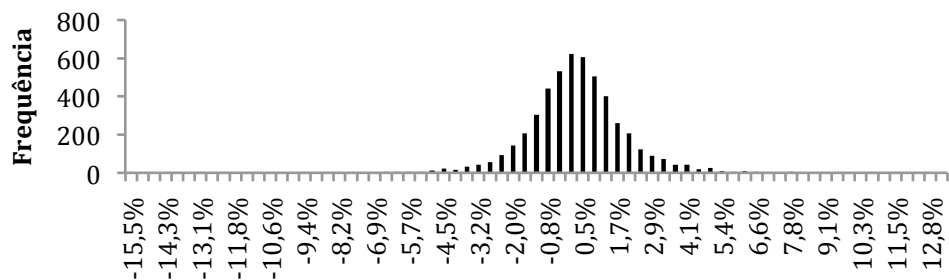
**Figura 92** – Representação temporal do valor de Retorno Logarítmico de fecho de IBM



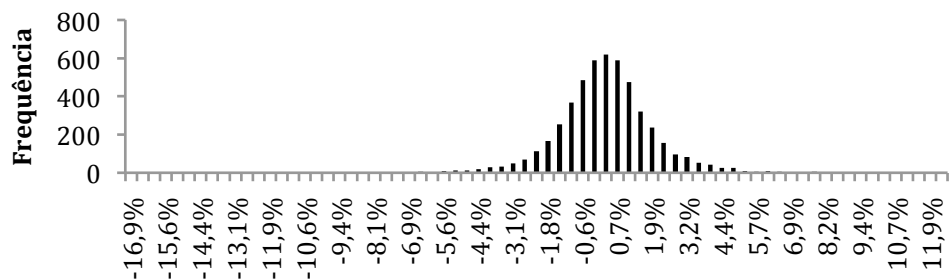
**Figura 93** – Representação histográfica do valor real de fecho de IBM



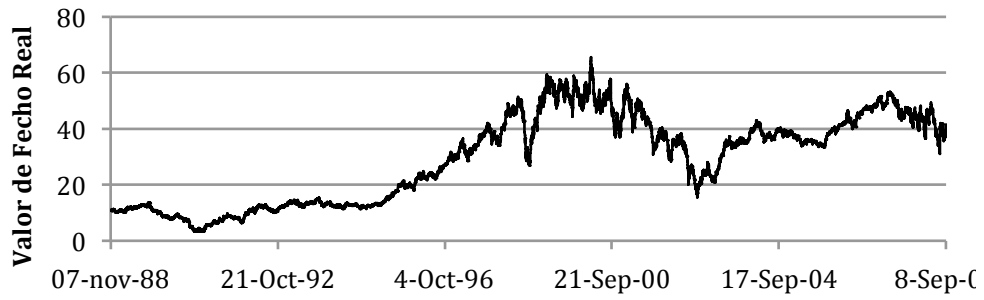
**Figura 94** – Representação histográfica do valor às diferenças de fecho de IBM



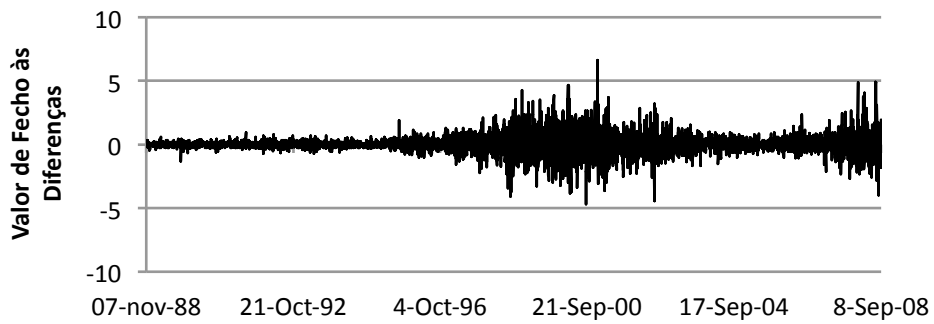
**Figura 95** – Representação histográfica do valor de Retorno Bruto de fecho de IBM



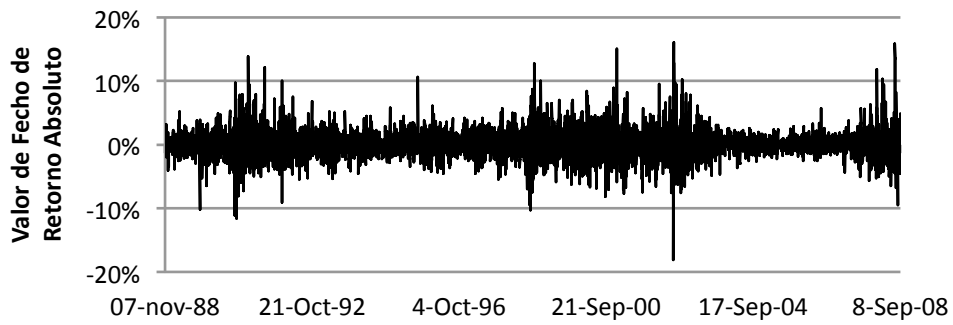
**Figura 96** – Representação histográfica do valor de Retorno Logarítmico de fecho de IBM



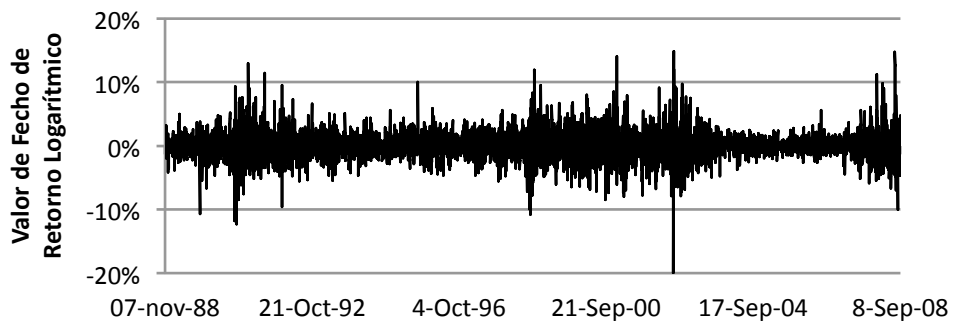
**Figura 97** – Representação temporal do valor real de fecho de JPM



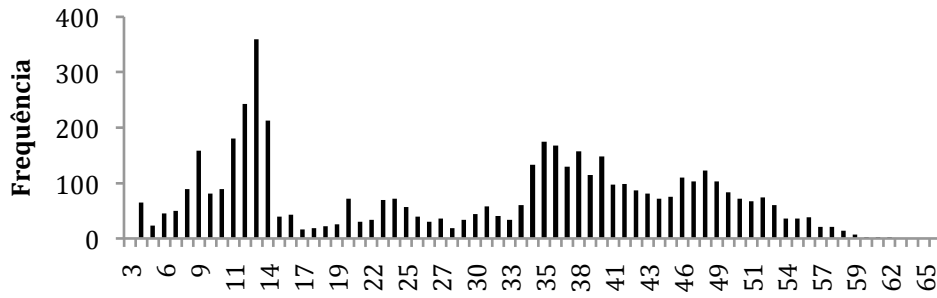
**Figura 98** – Representação temporal do valor às diferenças de fecho de JPM



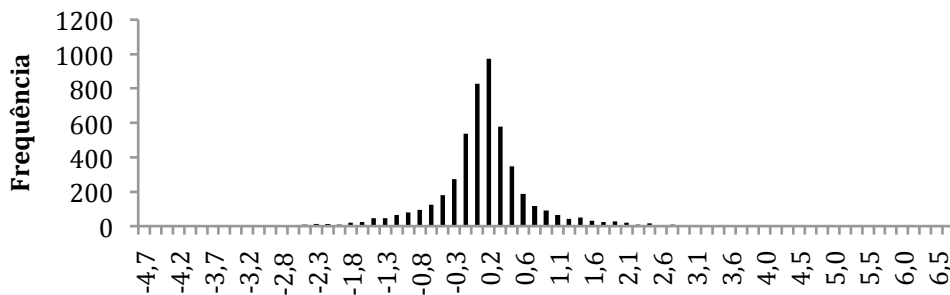
**Figura 99** – Representação temporal do valor de Retorno Bruto de fecho de JPM



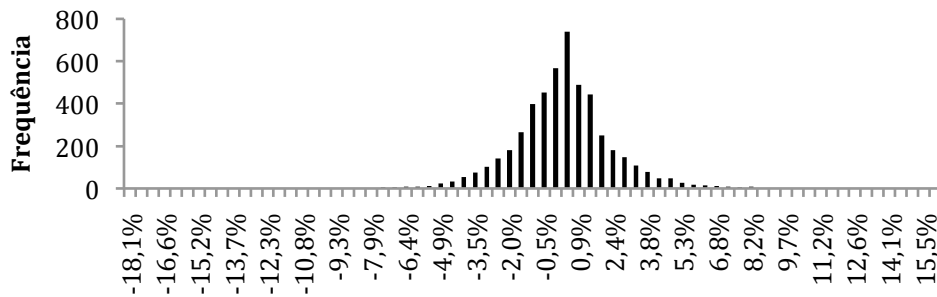
**Figura 100** – Representação temporal do valor de Retorno Logarítmico de JPM



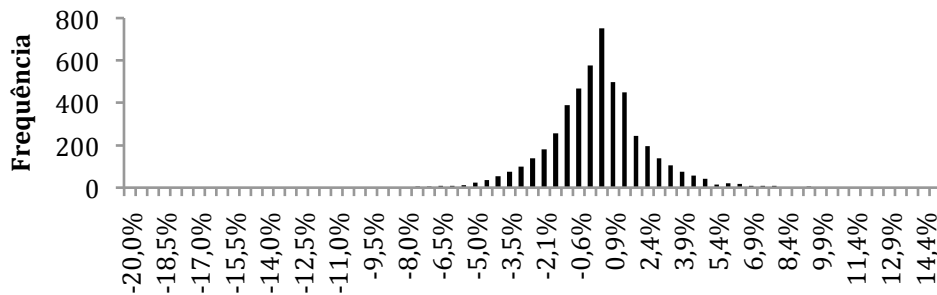
**Figura 101** – Representação histográfica do valor real de fecho de JPM



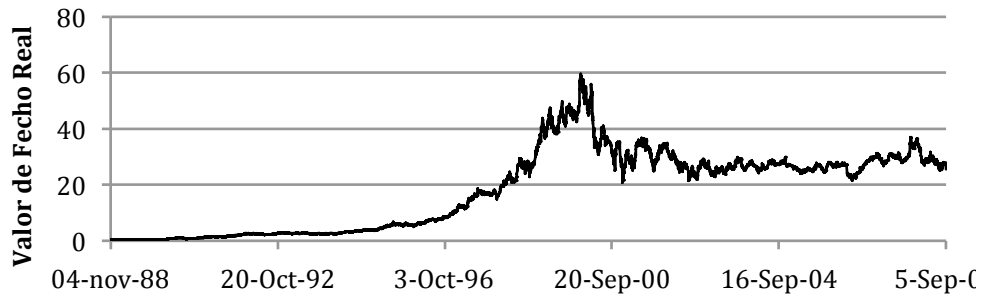
**Figura 102** – Representação histográfica do valor às diferenças de fecho de JPM



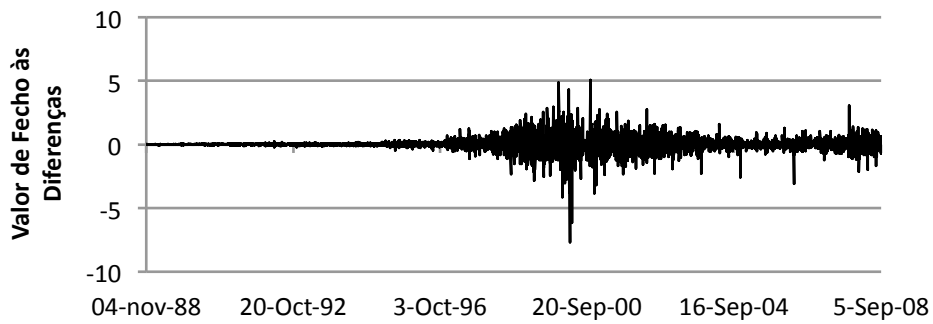
**Figura 103** – Representação histográfica do valor de Retorno Bruto de fecho de JPM



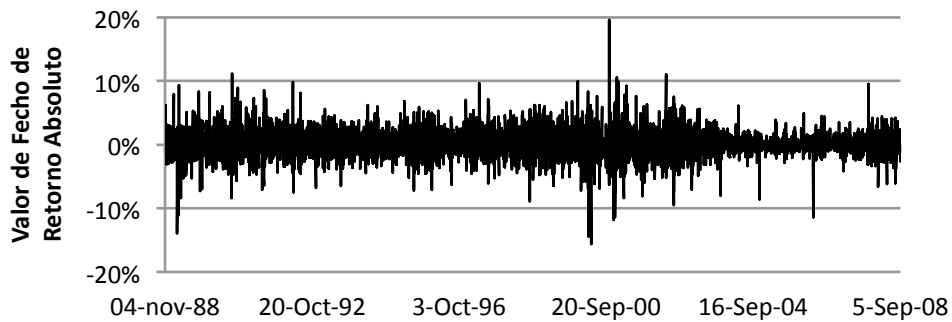
**Figura 104** – Representação histográfica do valor de Retorno Logarítmico de fecho de JPM



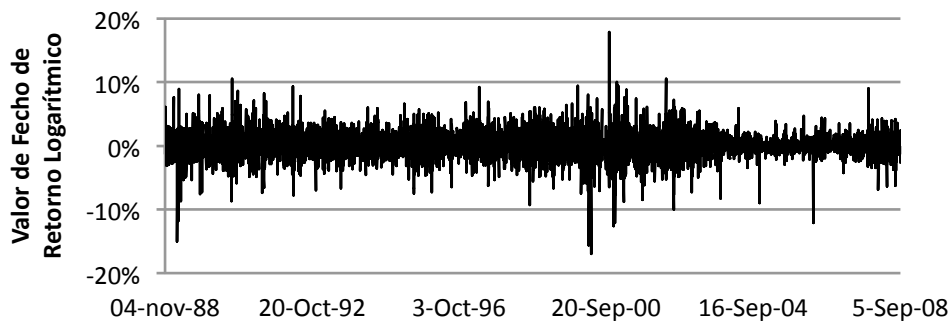
**Figura 105** – Representação temporal do valor real de fecho de MSFT



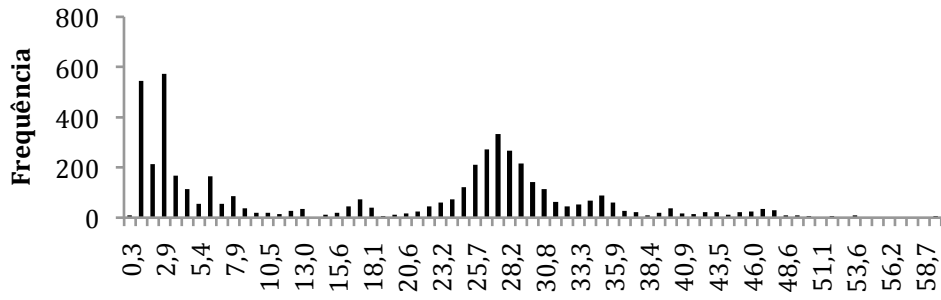
**Figura 106** – Representação temporal do valor às diferenças de fecho de MSFT



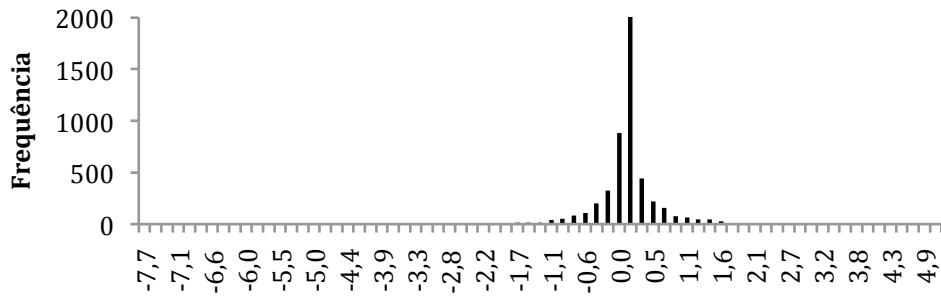
**Figura 107** – Representação temporal do valor de Retorno Bruto de fecho de MSFT



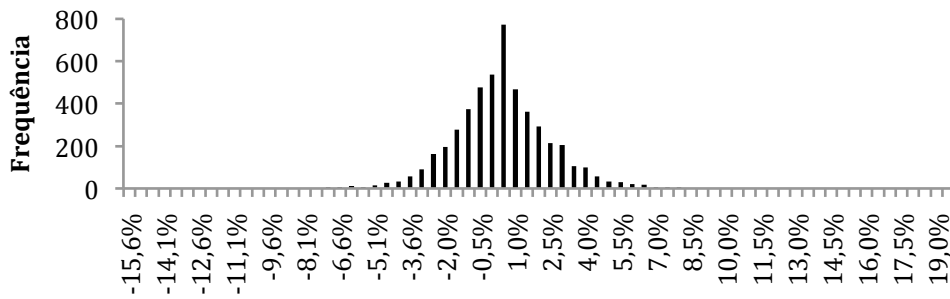
**Figura 108** – Representação temporal do valor de Retorno Logarítmico de MSFT



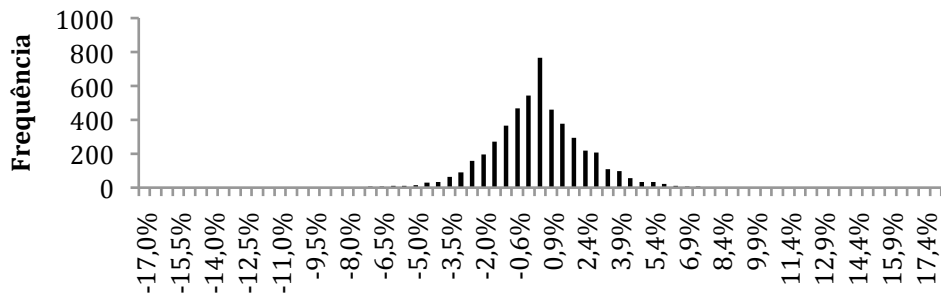
**Figura 109** – Representação histográfica do valor real de fecho de MSFT



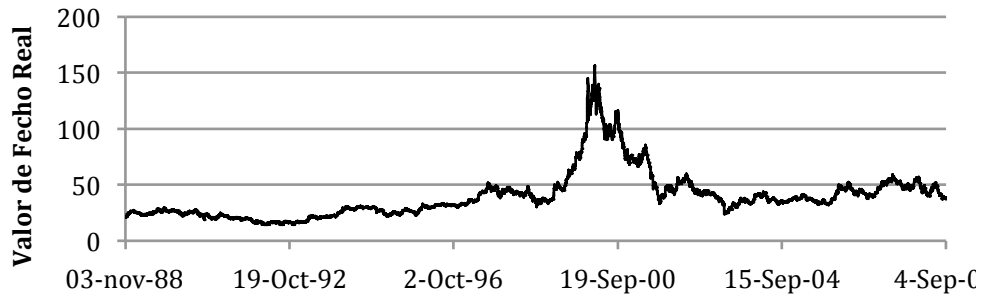
**Figura 110** – Representação histográfica do valor às diferenças de fecho de MSFT



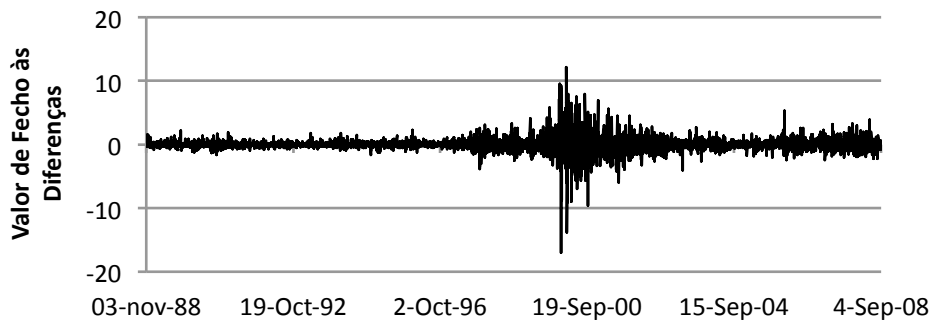
**Figura 111** – Representação histográfica do valor de Retorno Bruto de fecho de MSFT



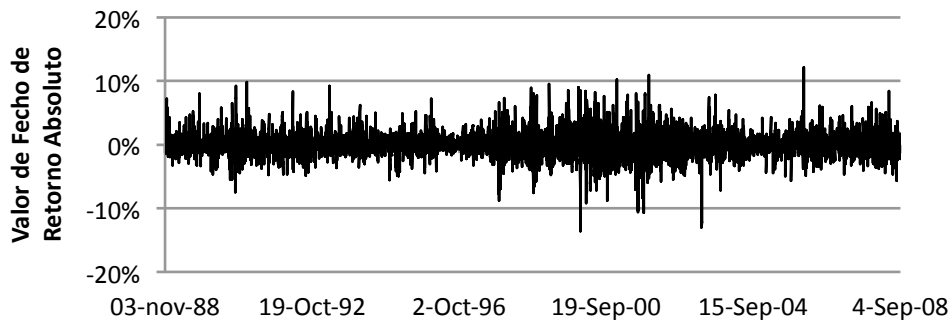
**Figura 112** – Representação histográfica do valor de Retorno Logarítmico de fecho de MSFT



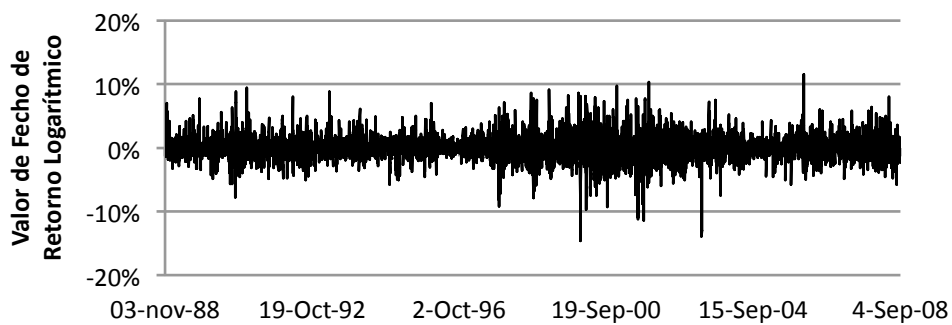
**Figura 113** – Representação temporal do valor real de fecho de SNE



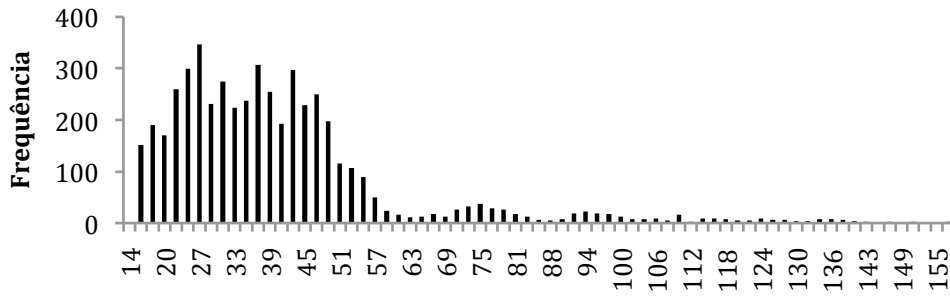
**Figura 114** – Representação temporal do valor às diferenças de fecho de SNE



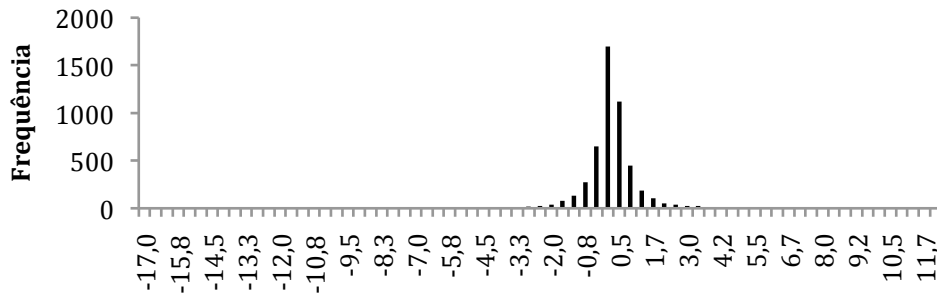
**Figura 115** – Representação temporal do valor de Retorno Bruto de fecho de SNE



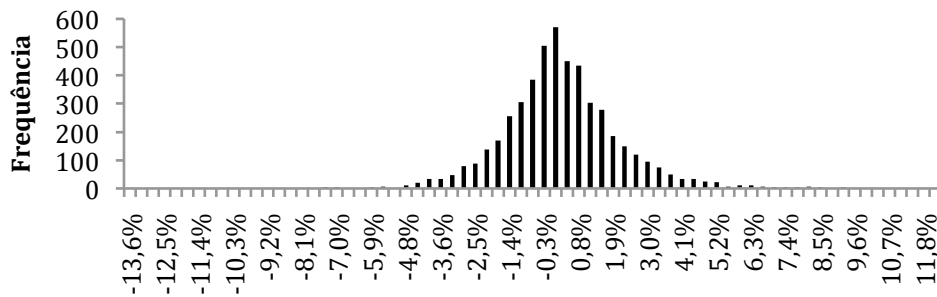
**Figura 116** – Representação temporal do valor de Retorno Logarítmico de SNE



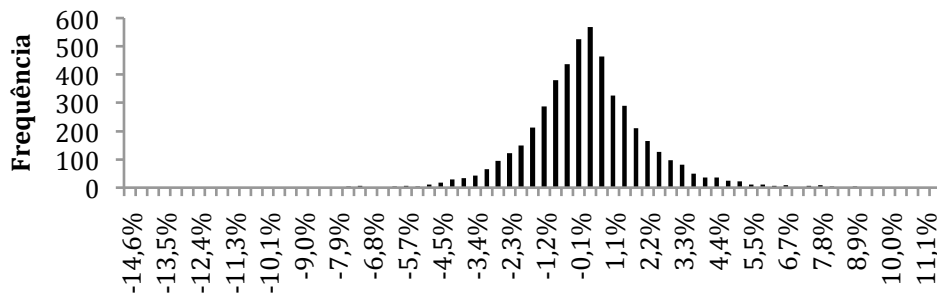
**Figura 117** – Representação histográfica do valor real de fecho de SNE



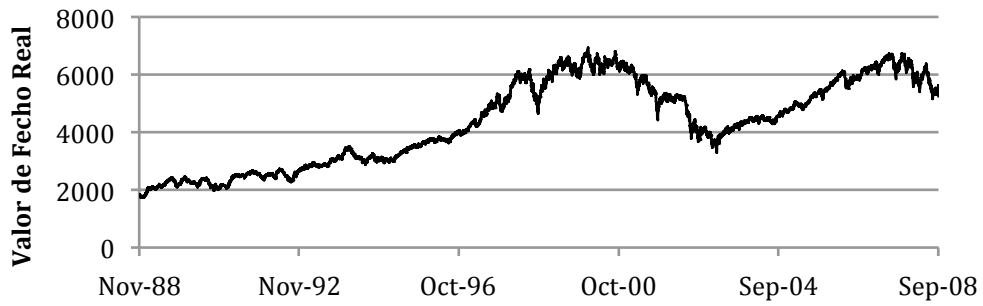
**Figura 118** – Representação histográfica do valor às diferenças de fecho de SNE



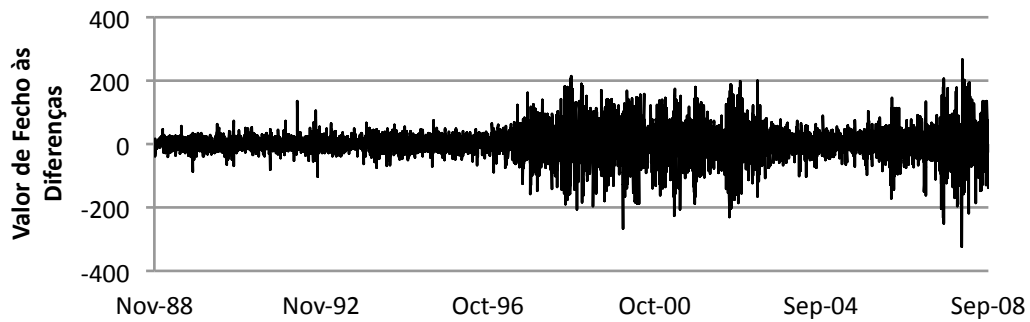
**Figura 119** – Representação histográfica do valor de Retorno Bruto de fecho de SNE



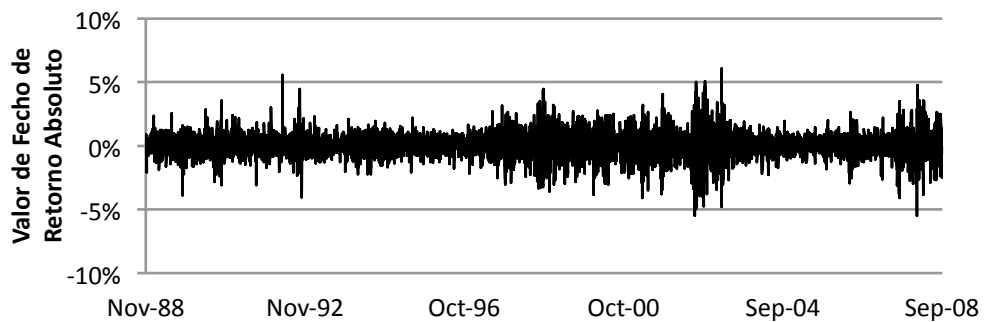
**Figura 120** – Representação histográfica do valor de Retorno Logarítmico de fecho de SNE



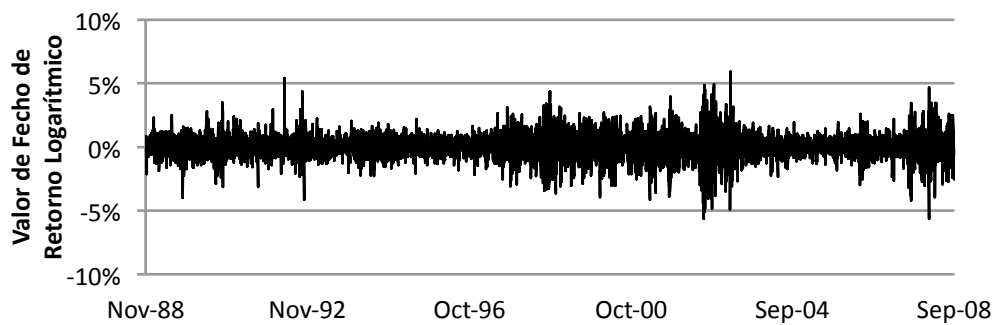
**Figura 121** – Representação temporal do valor real de fecho de FTSE 100



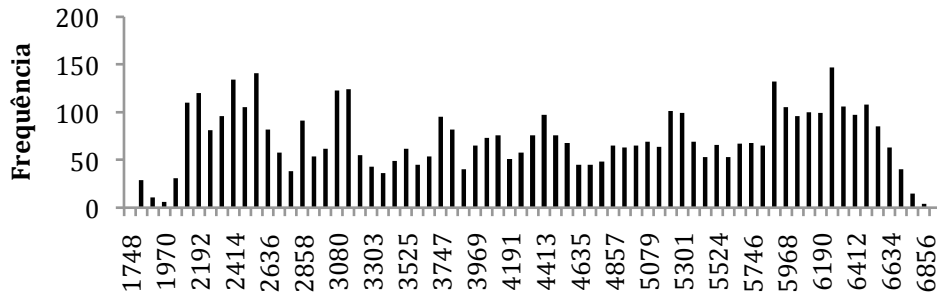
**Figura 122** – Representação temporal do valor às diferenças de fecho de FTSE 100



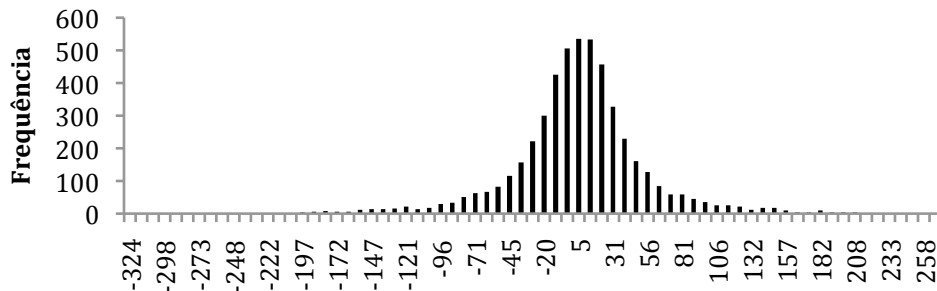
**Figura 123** – Representação temporal do valor de Retorno Bruto de fecho de FTSE 100



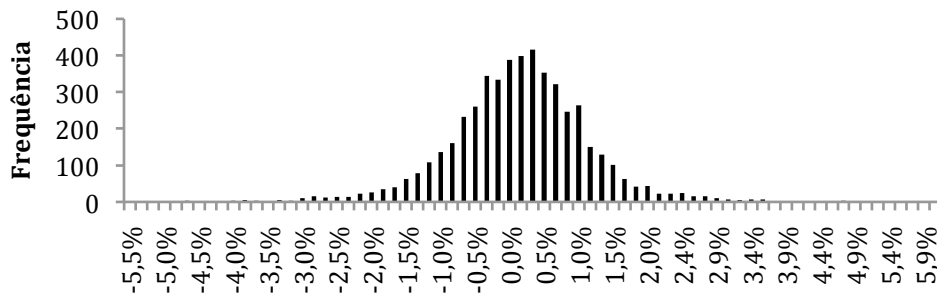
**Figura 124** – Representação temporal do valor de Retorno Logarítmico de FTSE 100



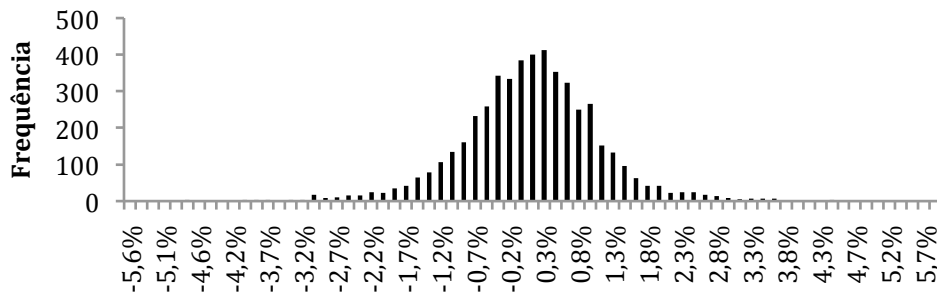
**Figura 125** – Representação histográfica do valor real de fecho de FTSE 100



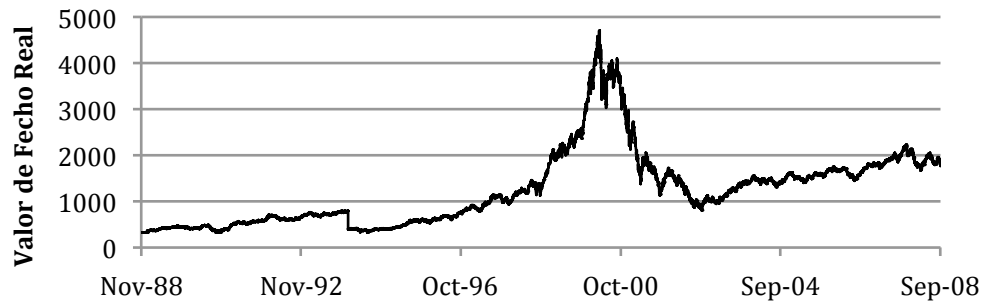
**Figura 126** – Representação histográfica do valor às diferenças de fecho de FTSE 100



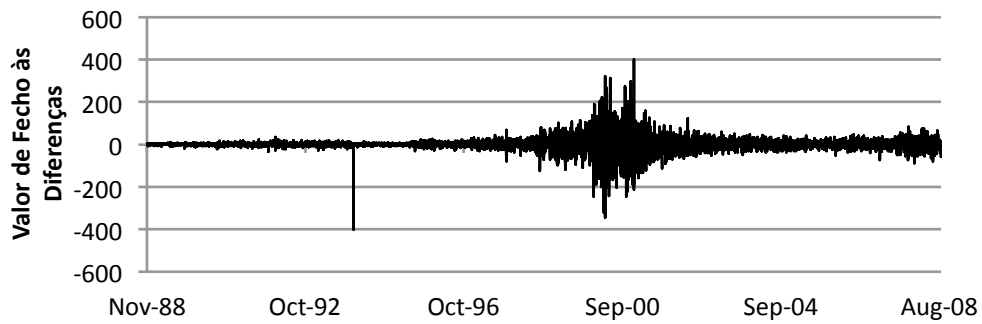
**Figura 127** – Representação histográfica do valor de Retorno Bruto de fecho de FTSE 100



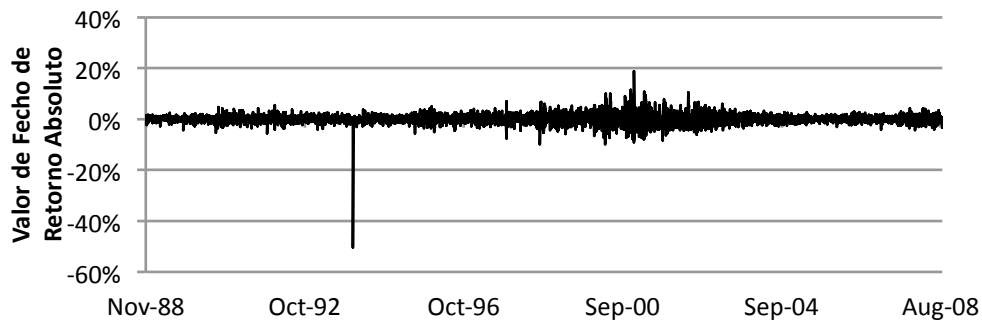
**Figura 128** – Representação histográfica do valor de Retorno Logarítmico de FTSE 100



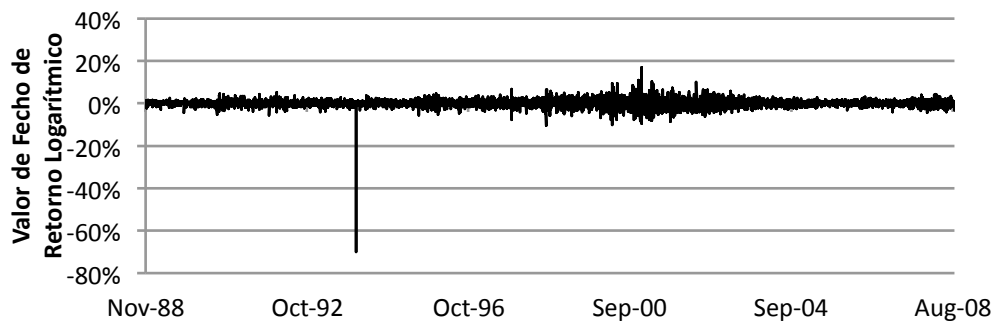
**Figura 129** – Representação temporal do valor real de fecho de NASDAQ



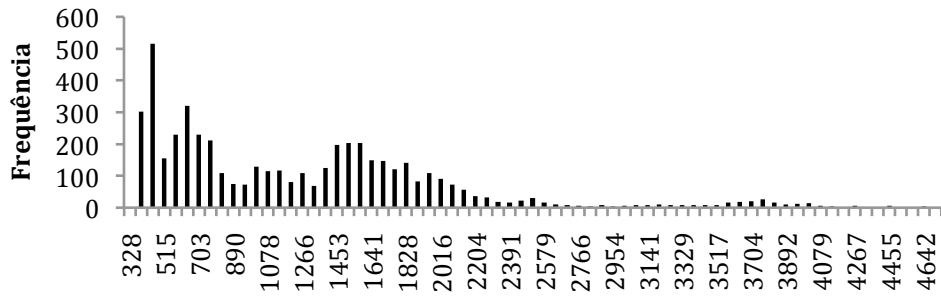
**Figura 130** – Representação temporal do valor às diferenças de fecho de NASDAQ



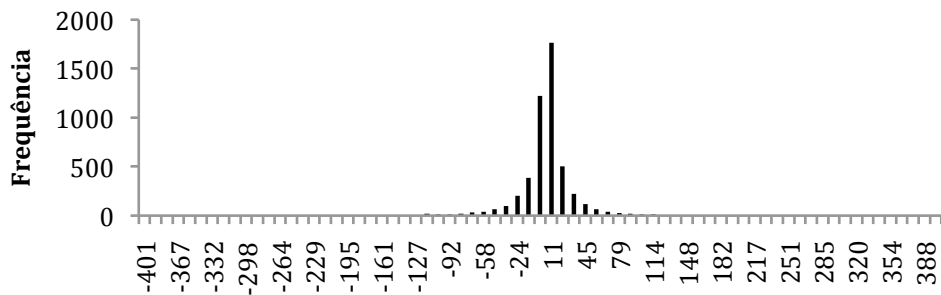
**Figura 131** – Representação temporal do valor de Retorno Bruto de fecho de NASDAQ



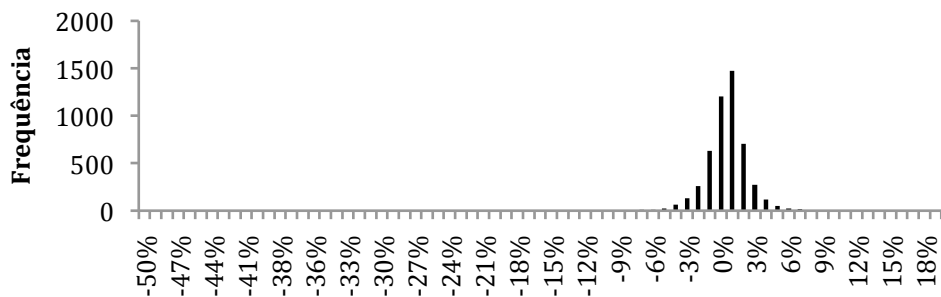
**Figura 132** – Representação temporal do valor de Retorno Logarítmico de fecho NASDAQ



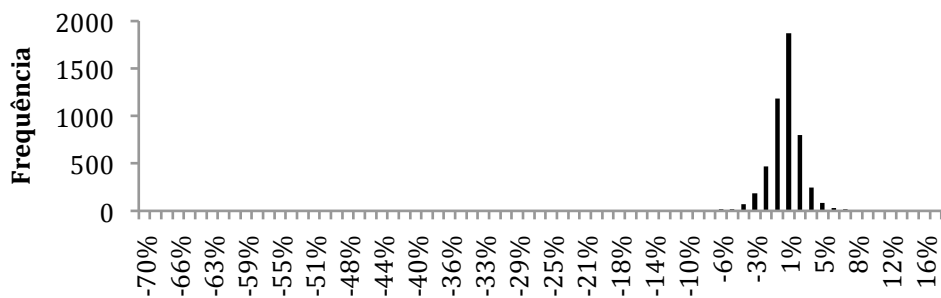
**Figura 133** – Representação histográfica do valor real de fecho de NASDAQ



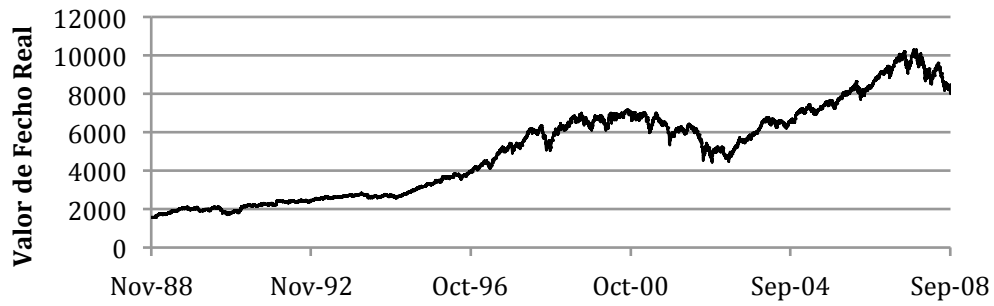
**Figura 134** – Representação histográfica do valor às diferenças de fecho de NASDAQ



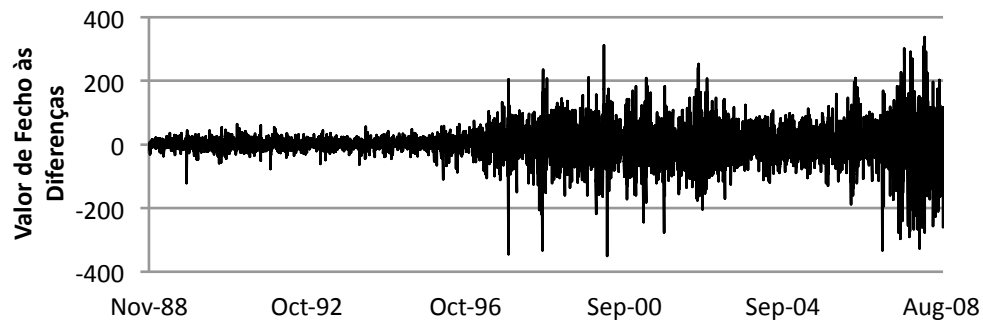
**Figura 135** – Representação histográfica do valor de Retorno Bruto de fecho de NASDAQ



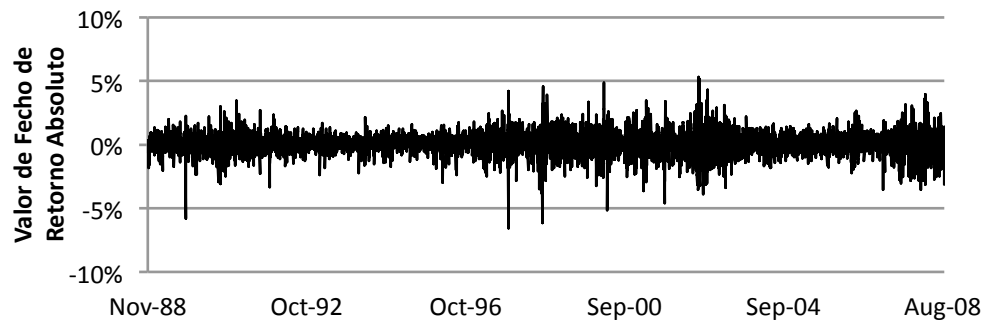
**Figura 136** – Representação histográfica do valor de Retorno Logarítmico de NASDAQ



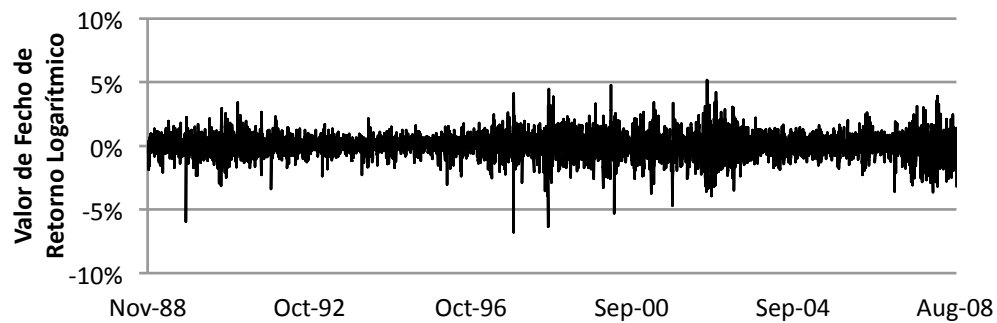
**Figura 137** – Representação temporal do valor real de fecho de NYSE



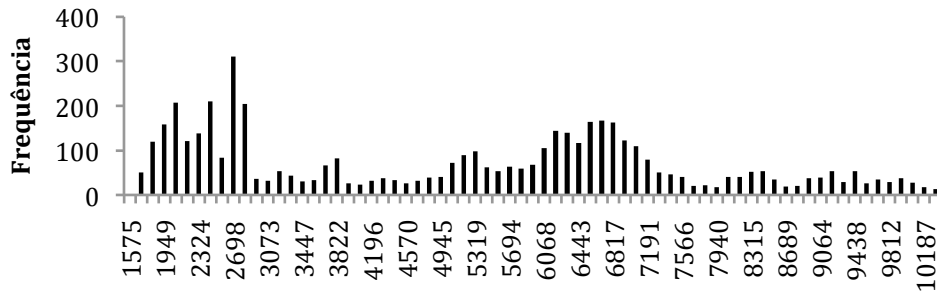
**Figura 138** – Representação temporal do valor às diferenças de fecho de NYSE



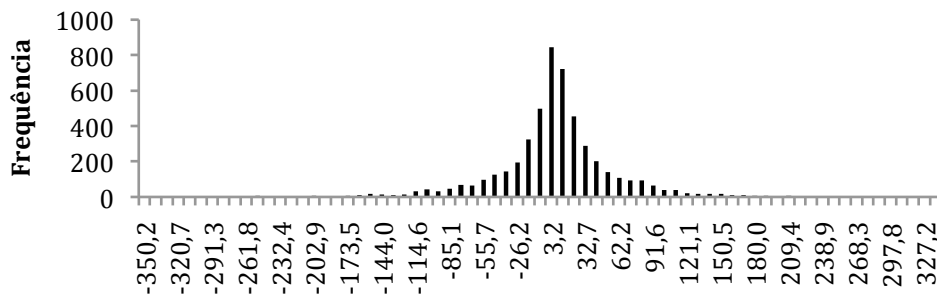
**Figura 139** – Representação temporal do valor de Retorno Bruto de fecho de NYSE



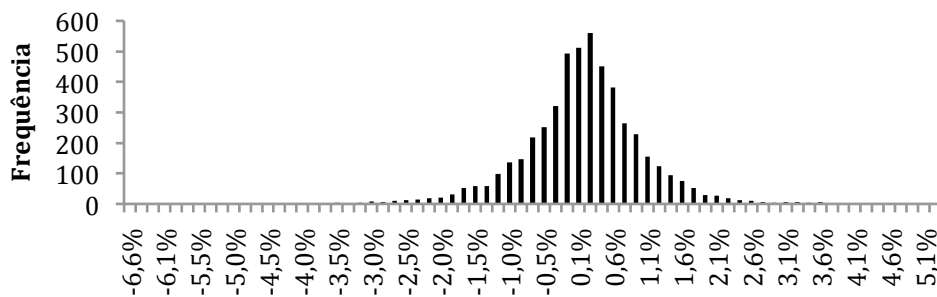
**Figura 140** – Representação temporal do valor de Retorno Logarítmico de fecho de NYSE



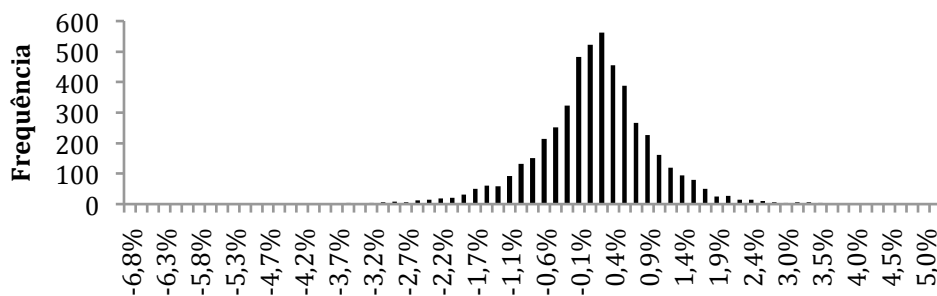
**Figura 141** – Representação histográfica do valor real de fecho de NYSE



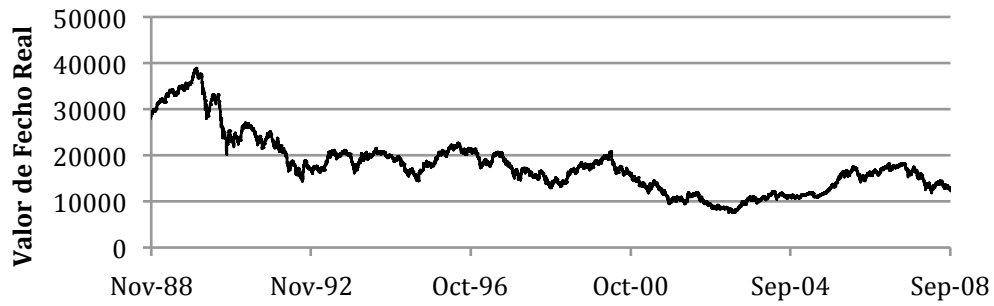
**Figura 142** – Representação histográfica do valor às diferenças de fecho de NYSE



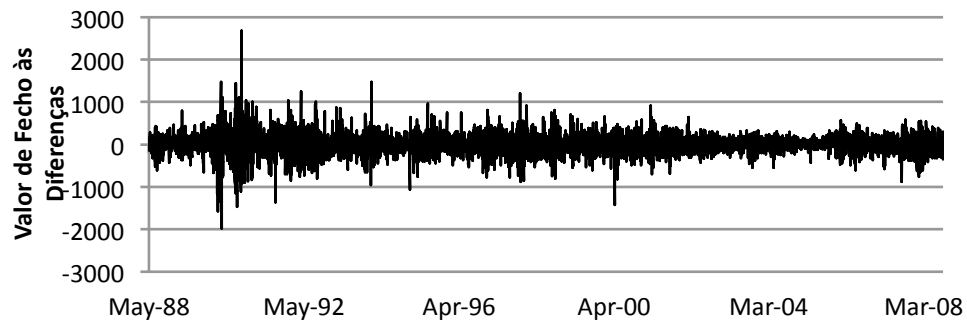
**Figura 143** – Representação histográfica do valor de Retorno Bruto de fecho de NYSE



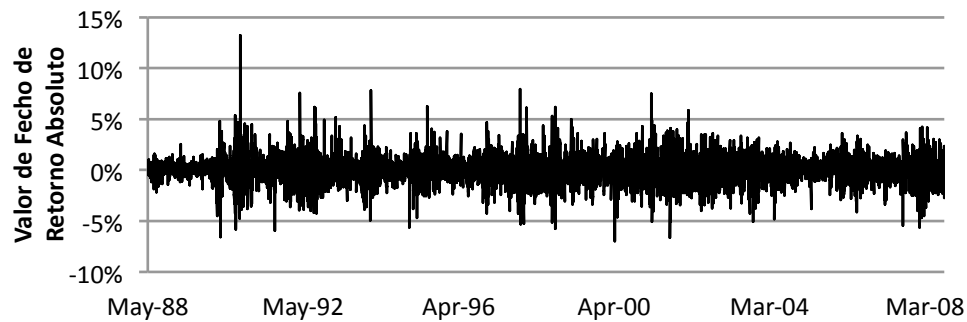
**Figura 144** – Representação histográfica do valor de Retorno Logarítmico de fecho de NYSE



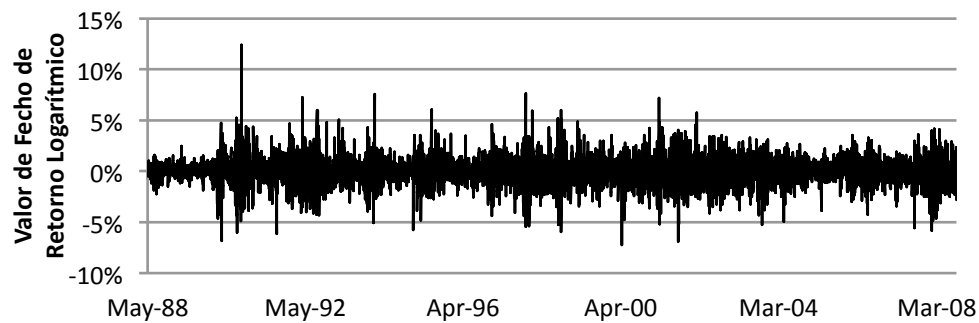
**Figura 145** – Representação temporal do valor real de fecho de NYKKEI 225



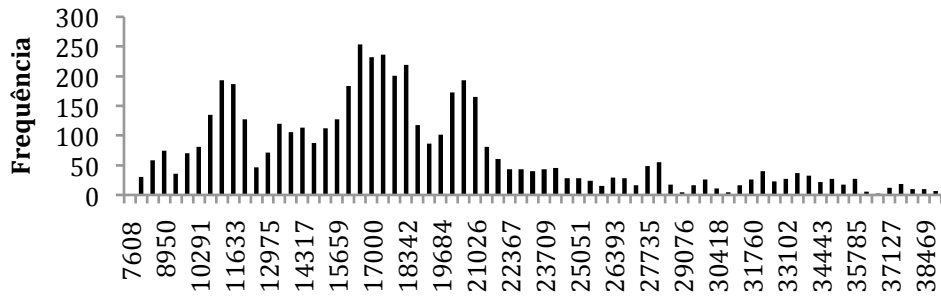
**Figura 146** – Representação temporal do valor às diferenças de fecho de NYKKEI 225



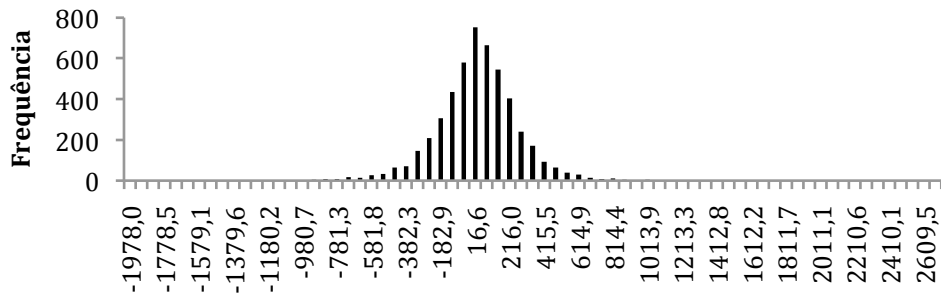
**Figura 147** – Representação temporal do valor de Retorno Bruto de fecho de NYKKEI 225



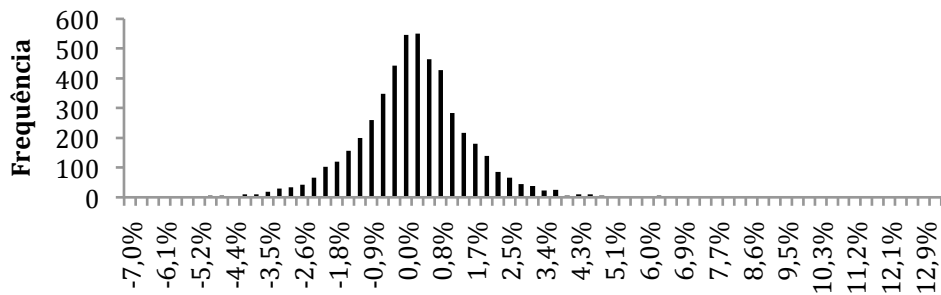
**Figura 148** – Representação temporal do valor de Retorno Logarítmico de NYKKEI 225



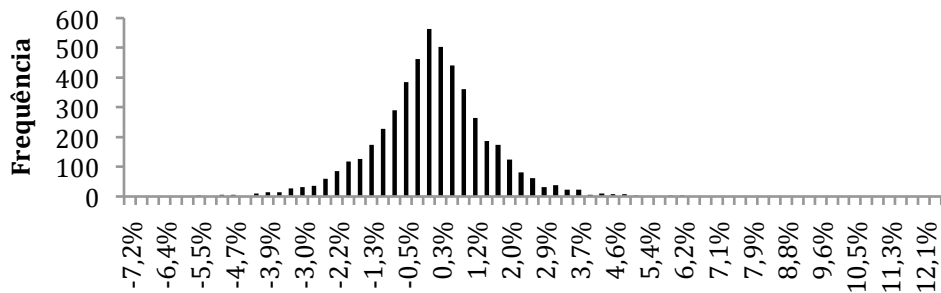
**Figura 149** – Representação histográfica do valor real de fecho de NYKKEI 225



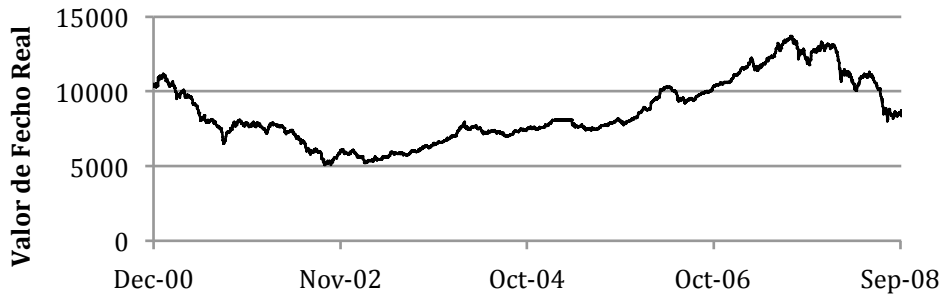
**Figura 150** – Representação histográfica do valor às diferenças de fecho de NYKKEI 225



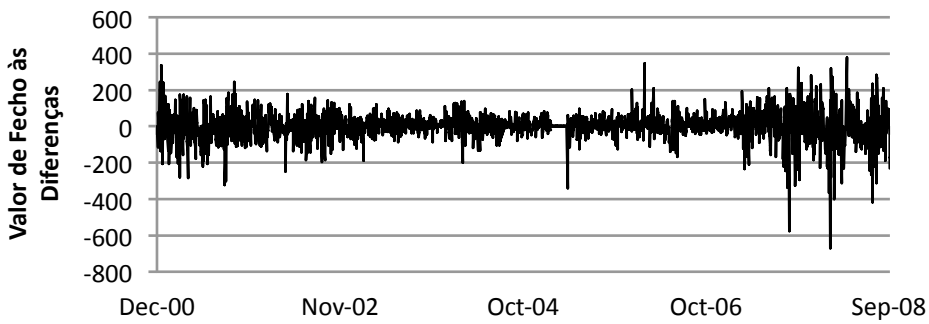
**Figura 151** – Representação histográfica do valor de Retorno Bruto de fecho NYKKEI 225



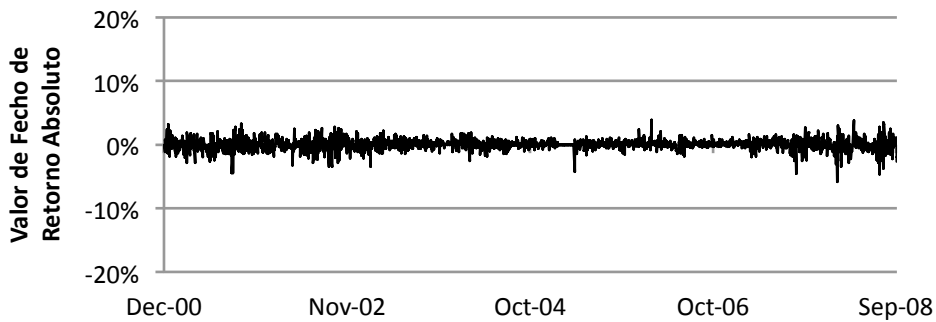
**Figura 152** – Representação histográfica do valor de Retorno Logarítmico de NYKKEI 225



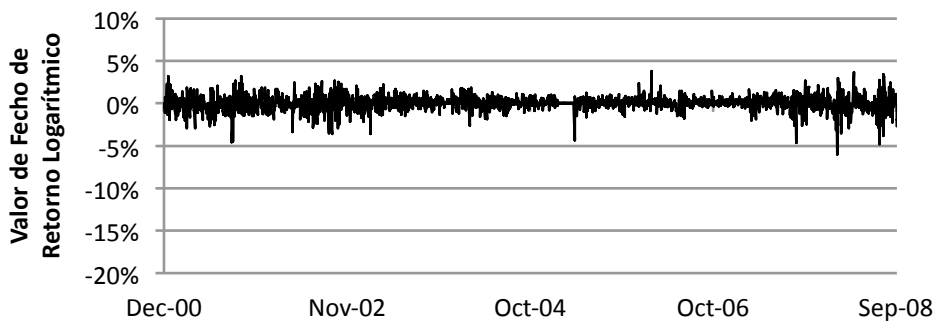
**Figura 153** – Representação temporal do valor real de fecho de PSI20



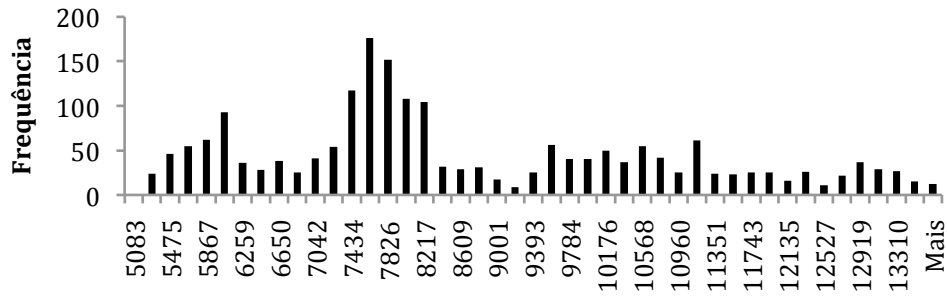
**Figura 154** – Representação temporal do valor às diferenças de fecho de PSI20



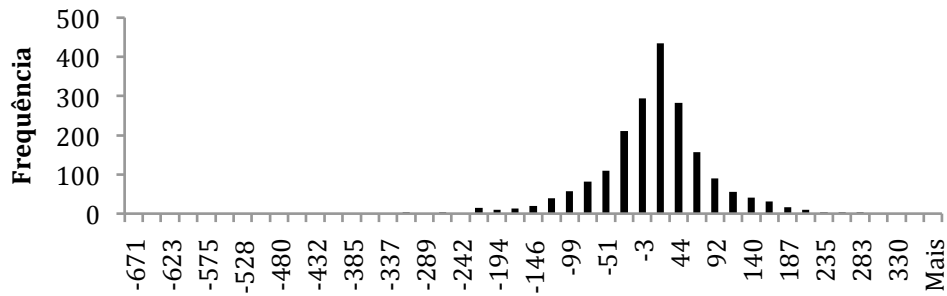
**Figura 155** – Representação temporal do valor de Retorno Bruto de fecho de PSI20



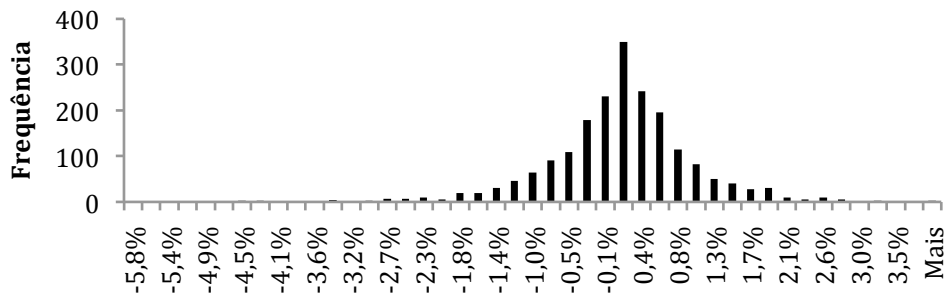
**Figura 156** – Representação temporal do valor de Retorno Logarítmico de PSI20



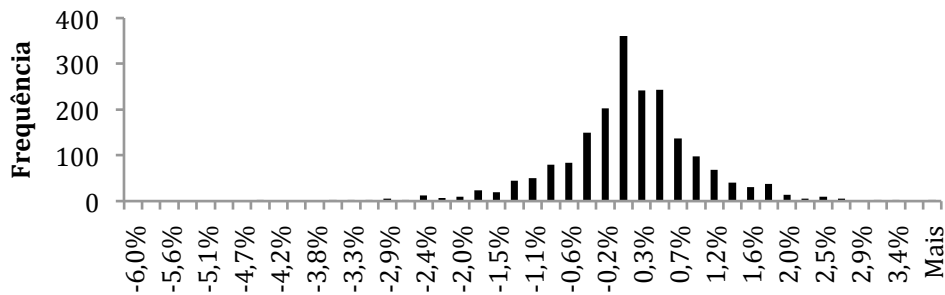
**Figura 157** – Representação histográfica do valor real de fecho de PSI20



**Figura 158** – Representação histográfica do valor às diferenças de fecho de PSI20



**Figura 159** – Representação histográfica do valor de Retorno Bruto de fecho de PSI20



**Figura 160** – Representação histográfica do valor de Retorno Logarítmico de fecho de PSI20

É possível verificar que nos valores reais, isto é, em valores não transformados, há uma tendência crescente do seu valor ao longo do tempo. Este é o efeito da inflação e pagamento de juros nos preços dos activos das bolsas financeiras. Apesar das variações afectas aos activos provocados por variadas influências, o efeito de inflação parece sobrepor-se, num espaço de tempo suficientemente grande, a todos os outros factores.

No entanto, este efeito da inflação dos preços é eliminado quando se opera sobre valores transformados, como os de retorno (bruto ou logarítmico) e às diferenças. Aqui é considerada apenas a variação entre cada dois valores em cada observação. A eliminação do efeito da inflação provoca uma normalização dos preços dos activos, aproximando-os, como é possível verificar nas representações histográficas, de um gráfico com forma de sino centrado na vizinhança de zero. Tal evidencia uma não aleatoriedade no que toca aos valores transformados dos activos, pois variam principalmente em torno de um valor médio, indiciando assim a possibilidade da sua previsão.

Outro fenómeno observável nas representações ao longo do tempo para alguns dos índices de bolsa representados é o de, ao possuírem valores baixos e estáveis num período de tempo alargado, verificarem-se baixos valores transformados às diferenças para o mesmo período. Tal pode ser observado nos gráficos de AAPL (figuras 1 e 2), C (figuras 33 e 34), DELL (figuras 41 e 42), GE (figuras 65 e 66) e MSFT (figuras 105 e 106). Isto traduz-se em representações histográficas com altas frequências para valores na vizinhança de zero, e baixas frequências para os restantes valores (figuras 5-8, 37-40, 45-48, 69-72 e 109-112).

Contrariamente aos índices referidos que possuem uma frequência alta para valores  $\approx 0$ , aqueles que apresentam uma maior distribuição das frequências no espectro total, como BA (figura 13), BP (figura 21), CVX (figura 29), F (figura 61), GM (figura 77), IBM (figura 93), FTSE 100 (figura 125) e NIKKEI 225 (figura 149), possuem uma distribuição de valores transformados mais ampla e com uma frequência mais baixa.

Períodos com alta volatilidade relativamente ao restante espectro traduz-se numa grande variação dos valores às diferenças. No entanto, esta variação não é notável nos valores transformados de retorno (bruto e logarítmico). Este fenómeno é observável para AAPL (figuras 1-4), DELL (figuras 41-44), GE (figuras 65-68), HPQ (figuras 81-84), IBM (figuras 89-92), MSFT (figuras 105-108), SNE (figuras 113-116) e NASDAQ (figuras 129-132).

É também digno de nota a pequena diferença entre as representações dos valores transformados de retorno bruto e retorno logarítmico. Esta diferença situa-se nas centésimas de ponto percentual (0.01%). No entanto, serão mesmo assim estudadas ambas as transformações, na tentativa de verificar se esta diferença se torna significativa nas previsões efectuadas.

Como referido anteriormente, os dados transformados dos preços dos índices em estudo aparentam possuir alguma normalização dos seus valores. Este trabalho assentar-se-á nesta característica na tentativa de prever valores futuros dos índices em estudo.



# 3. TRANSFORMAÇÃO DE DADOS DOS MERCADOS FINANCEIROS

A análise estatística de dados do mercado é essencial, tanto pela razão fundamental de compreender a dinâmica do mercado, como por razões relacionadas com os problemas chave de fixar o preço opcional e gestão de portfolio. Neste capítulo é referida a transformação dos dados emergentes nos mercados financeiros. São também discutidas as escalas de tempo usadas habitualmente em mercados financeiros e é referida a escolha feita para a escala de tempo usada no decorrer do trabalho.

Na análise de mercados financeiros as escalas adoptadas são usualmente dadas em unidades (moedas de câmbio) que flutuam no seu valor ao longo do tempo e em transacções que ocorrem em momentos e com intensidade aleatória. Por esta razão, deve ser tomado cuidado na selecção da variável mais apropriada a ser estudada, tomando em consideração as suposições associadas com cada escolha possível (Mantegna & Stanley, 2000).

O preço de bens financeiros é normalmente a moeda do país em que esse bem financeiro em particular está localizado. O valor da moeda não é constante ao longo do tempo. Uma moeda pode mudar o seu valor devido a:

- inflação
- crescimento ou abrandamento económico
- flutuações no mercado de valores global

Defina-se  $X_t$  como o preço de um bem financeiro no momento  $t$ . Qual é a variável estocástica apropriada para se investigar? São possíveis diferentes escolhas e cada uma tem os seus méritos e seus problemas.

### 3.1. DIFERENÇA DE PREÇOS

Em primeiro lugar pode-se investigar a mudança de preços,

$$Y_t = X_{t+\Delta t} - X_t \quad (3.1)$$

em que  $Y_t$  é a diferença entre os preços no momento  $\Delta t$  e  $t$ .

O mérito desta escolha é que não são necessárias transformações não lineares ou estocásticas. Um problema desta definição é que não toma em consideração as mudanças no preço resultantes da inflação ou pagamento de juros, fazendo com que a distribuição dos valores se aproxime de uma curva de aspecto de sino.

### 3.1.1. DIFERENÇA DE PREÇOS DESCONTADA

Alternativamente, pode-se analisar mudanças de preço deflacionadas, ou descontadas,

$$Y_{D_t} = [X_{t+\Delta t}] \cdot D_t \quad (3.2)$$

onde  $D_t$  é um factor de deflação. Os méritos desta técnica são:

- Define as mudanças de preço num sentido absoluto, quando é mais apropriado usar uma variação relativa, uma vez que as variações absolutas estão muito dependentes das unidades/grandezas dos valores e, conseqüentemente, não permitem comparações com outros índices (Dow Zhengdao, 2005).
- Os preços são dados em termos de dinheiro “constante” – os ganhos possíveis com investimentos sem risco são contabilizados pelo factor  $D_t$ .

O problema é que os factores de deflação e de desconto são imprevisíveis ao longo do tempo, e não há uma única escolha para  $D_t$ .

### 3.1.2. SUCESSIVAS DIFERENÇAS DO LOGARITMO NATURAL DO PREÇO

Outra opção ainda consiste em estudar as sucessivas diferenças do logaritmo natural do preço,

$$Y_{S_t} = \ln X_{\Delta t+t} - \ln X_t \quad (3.3)$$

O mérito desta escolha é que a correcção média das mudanças de escala é incorporada sem requerer factores de deflação ou de desconto.

Os problemas são:

- A correcção de uma mudança de escala seria correcta apenas se a taxa de crescimento da economia fosse constante, todavia, a taxa de crescimento geralmente flutua, e essas flutuações não estão incorporadas na definição (3.3)
- É usada uma transformação não linear, e a não linearidade afecta fortemente as propriedades estatísticas de um processo estocástico. Note-se que a informação tomada por  $Y_{S_t}$  mistura características da dinâmica do bem financeiro juntamente com aspectos que envolvem flutuações de indicadores macroeconómicos.

## 3.2. RETORNOS

É também possível analisar-se os retornos, definidos como,

$$Y_t = \frac{X_{\Delta t+t} - X_t}{X_t} \quad (3.4)$$

em que  $Y_t$  é o retorno do preço no momento  $t$ .

A utilização do retorno permite a observação das variações dos preços num sentido relativo, eliminando os efeitos da inflação ou pagamento de juros, referidos anteriormente. Com a eliminação os efeitos da inflação e dos pagamentos de juros, a distribuição dos valores de retorno já se aproxima de uma distribuição em forma de sino, com os valores concentrados em volta de um valor central, neste caso, de um retorno com valor zero.

Considere-se dois modos diferentes de retorno aqui apresentados,

- Retorno bruto  $Y_t = \frac{X_t - X_{t-1}}{X_{t-1}} \quad (3.5)$

- Retorno Logarítmico  $y_t = \ln(1 + Y_t) \quad (3.6)$

### 3.2.1. RETORNO BRUTO DE PERÍODO ÚNICO

Um retorno bruto simples é calculado,

$$1 + Y_t = \frac{X_t}{X_{t+1}} \quad \text{ou} \quad X_t = X_{t-1}(1 + Y_t) \quad (3.7)$$

O retorno simples é,

$$Y_t = \frac{X_t}{X_{t-1}} - 1 = \frac{X_t - X_{t-1}}{X_{t-1}} \quad (3.8)$$

O mérito desta definição é fornecer uma percentagem directa dos ganhos e perdas num dado período de tempo. O problema é a sua sensibilidade para escalar mudanças para longos horizontes de tempo (Pollock, 1993).

Este é o tipo de retorno mais utilizado nos relatórios financeiros, pelo seu carácter intuitivo que permite uma avaliação simples da evolução instantânea do activo financeiro.

### 3.2.2. RETORNO LOGARÍTMICO

O retorno logarítmico é dado por:

$$y_t = \ln(1 + Y_t) = \ln \frac{X_t}{X_{t-1}} = x_t - x_{t-1} \quad (3.9)$$

onde  $x_t = \ln X_t$ .

Retornos de composição contínua  $r_t$  têm vantagens sobre os simples retornos brutos  $R_t$ . Primeiro, considerando retornos de períodos múltiplos, tem-se:

$$y_t[k] = y_t + y_{t+1} + \dots + y_{t-k+1} \quad (3.10)$$

Como tal, o retorno multi-período de composição contínua é simplesmente a soma dos retornos de período único compostos continuamente envolvidos. Em segundo lugar, as características estatísticas são mais fáceis de tratar, pois sendo uma composição relativa e sem dimensões, pode ser directamente usada para comparar activos com diferentes preços.

Para dados com alta frequência, ou dados que são gravados em pequenos períodos de tempo influenciados muito fracamente pela inflação, ambas as medidas normalmente usadas e listadas anteriormente, são aproximadamente equivalentes. No entanto, para investigações ao longo de um largo período de tempo, o retorno logarítmico é a transformação mais usada (Dow Zhengdao, 2005).

Os três tipos de transformação efectuados sobre os dados neste trabalho são Diferenças, Retorno Bruto e Retorno Logarítmico. São realizados testes no sentido de se verificar qual é o tipo de transformação de dados mais apropriado para realizar previsões com este tipo de dados.

### 3.3. ESCALAS DE TEMPO EM MERCADOS FINANCEIROS

Seguidamente considera-se a problemática da escolha da escala de tempo apropriada para usar na análise de dados financeiros. Possíveis candidatos para uma escala de tempo ‘correcta’ incluem:

- O tempo físico
- O tempo de mercado (ou de troca)
- O número de transacções

Uma escolha indiscutível não é clara.

O tempo físico é bem definido, mas a troca de acções fecha durante a noite, aos fins-de-semana e durante feriados. Uma limitação similar está também presente num mercado global tal como o mercado de câmbio estrangeiro. Apesar de este mercado estar activo 24 horas por dia, a organização social de negócio e a presença de ciclo biológicos forçam a actividade do mercado a ter restrições temporais em cada região financeira do mundo. Com a escolha do tempo físico, não se sabe como modelar a dinâmica estocástica de preços e a chegada de informação durante as horas em que o mercado está fechado.

Tempo de mercado está bem definido em trocas de acções: é o tempo que passa durante as horas em que o mercado está aberto. No mercado de câmbio estrangeiro, coincide com o tempo físico. Estudos empíricos tentaram determinar a variância de mudanças de preços logarítmica observada de fecho a fecho em mercados financeiros. Estes estudos mostram que a variância determinada considerando valores de fecho de dias sucessivos é apenas aproximadamente 20% mais baixa que a variância determinada considerando valores de fecho ao longo dos fins-de-semana (Fama, 1965) (French, 1980). Esta prova empírica corrobora a escolha de usar tempo de mercado na modelização da dinâmica de preços. De facto, o tempo de mercado é a escolha mais comum em estudos de investigação e nos estudos realizados na determinação da volatilidade do fixar de preço opcional. No entanto, podem surgir problemas com esta definição, nomeadamente,

- Informação que afecta a dinâmica do preço de um bem financeiro pode ser publicada enquanto o mercado está fechado,
- Em análises de índices financeiros com dados de alta frequência, as mudanças de preços ocorridos durante a noite são tratadas como mudanças de preço de curto prazo,
- A actividade de mercado é considerada implicitamente como uniforme durante as horas de mercado.

Esta última suposição não é verificada por análises empíricas. A actividade não é uniforme durante as horas de mercado, seja em termos de volume seja em números de contractos. Em vez disso, é observado um ciclo diário nos dados de mercado: a volatilidade é mais alta nas horas de abertura e fecho e, usualmente, o valor mais baixo da sessão ocorre durante as horas intermédias.

Se ‘tempo’ é definido em termos do número de transacções, então uma fonte de aleatoriedade observada em mercados financeiros é eliminada, especificamente o tempo passado entre transacções. No entanto, a segunda fonte de aleatoriedade, o volume de transacções, mantém-se.

A escala de tempo a ser usada no desenvolvimento deste trabalho é a do tempo dos mercados, a qual é a considerada a mais utilizada. A razão principal pela

escolha desta escala recai sobre o facto de que é a escala disponibilizada pelos serviços utilizados de informação financeira. A desvantagem desta escala é a de não ser possível considerar o efeito do fim-de-semana nas alterações dos valores dos índices. Em trabalhos posteriores poderá ser aplicado um algoritmo computacional que identifique e considere os limites semanais, incluindo feriados e outras ocasiões em que as bolsas se encontrem fechadas.

A frequência adoptada é a diária, sendo utilizados os valores de fecho do dia. Novamente a escolha da frequência encontra-se limitada pela frequência máxima dos dados fornecidos pelos serviços gratuitos de informação financeira, aos quais se recorre no decorrer deste trabalho.

Em trabalhos futuros poderá ser colocada a hipótese de estudo de dados com frequências inferiores, como semanais ou até mesmo mensais. No entanto, pode-se considerar que os dados com frequência mensal serão mais úteis em estudos que utilizem dados com características sazonais.

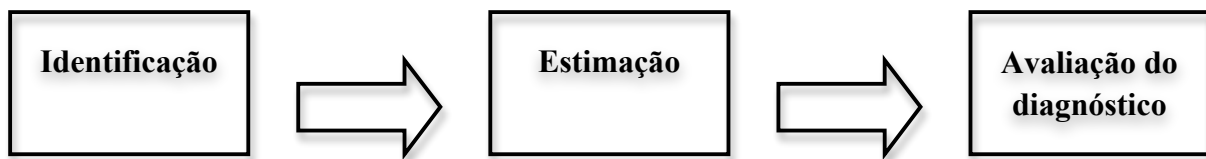


# 4. MODELIZAÇÃO DE SUCESSÕES CRONOLÓGICAS

Neste capítulo são revistos conceitos fundamentais da análise de séries temporais lineares tal como a estacionariedade, a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP). É também abordada a metodologia de três etapas na escolha de modelos, em que são analisadas as fases de identificação, estimação e avaliação do diagnóstico.

Define-se sucessão cronológica como um conjunto de observações associadas a determinado fenómeno aleatório, efectuadas em períodos sucessivos de tempo e relacionadas estatisticamente. Constitui objectivo principal da análise de sucessões cronológicas encontrar um “bom modelo” que tenha em conta as relações existentes entre as observações, permitindo, deste modo, a sua descrição. Pankratz (Pankratz, 1983) admite que “um bom modelo inclui o menor número de parâmetros estimados necessários para adaptar adequadamente os padrões nos dados disponíveis”.

Para alcançar um “bom modelo”, Box e Jenkins (Box & Jenkins, 1970) propuseram uma metodologia (figura 161) de modelização em três etapas:



Esta metodologia é a mais frequente em estudos de modelização.

A primeira etapa, identificação, em que se pretende um modelo ARMA<sup>4</sup> que aparentemente descreva a sucessão cronológica, compreende as seguintes sub-etapas:

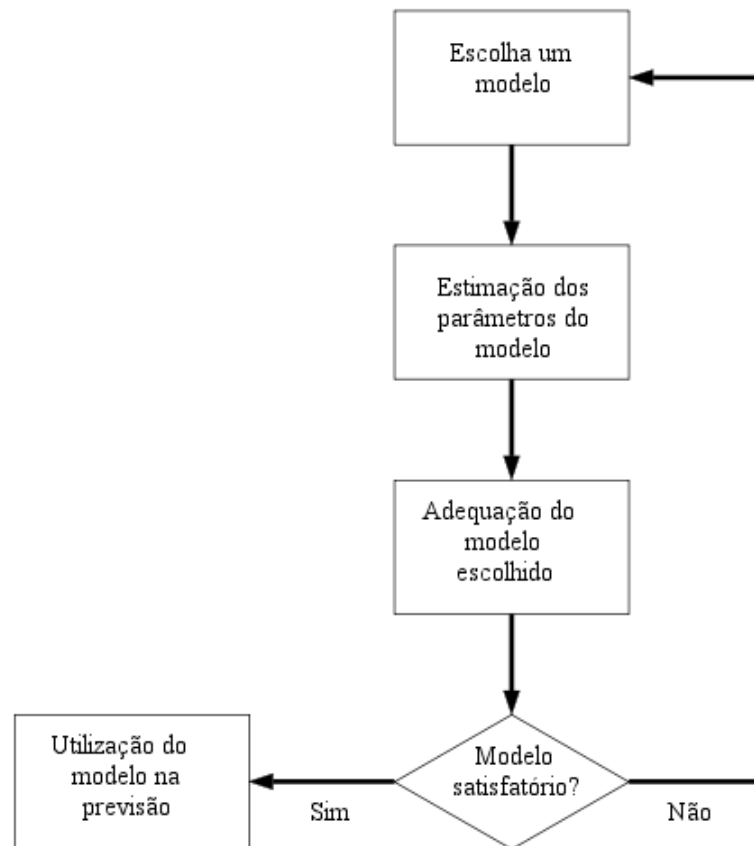
- Estacionarização da sucessão cronológica
- Selecção de um modelo ARMA

Como, em geral, as sucessões cronológicas não se encontram estacionarizadas e os modelos ARMA correspondem a processos estacionários, há que operar sobre os dados originais um certo conjunto de transformações de modo a que a sucessão resultante possa ser descrita por tais modelos. Ou seja, na primeira sub-etapa deve proceder-se à estacionarização da sucessão cronológica original, de modo a estabilizar a variância, a neutralizar a tendência e a eliminar movimentos de carácter estritamente periódico.

Uma vez estacionarizada a sucessão passa-se à sub-etapa seguinte, nomeadamente a selecção de um modelo ARMA, que é realizada através da comparação das FAC e FACP estimadas com as definições teóricas das mesmas funções correspondentes aos diferentes modelos, como apresentados na Tabela 1, apresentada mais à frente.

---

<sup>4</sup> Autoregressive Moving Average



**Figura 161** – Metodologia Box-Jenkins

Na fase do processo de identificação a perícia do analista reveste-se de particular importância.

Na etapa de estimação dos parâmetros do modelo identificado na etapa anterior o analista tem, de uma forma geral e na óptica das aplicações práticas, pouca intervenção, dado que os pacotes informáticos fornecem sem grandes exigências as estimativas dos parâmetros.

Na última etapa, avaliação do diagnóstico, procede-se ao controlo da adequação do modelo identificado e estimado nas etapas anteriores, havendo a distinguir duas sub-etapas:

- Qualidade estatística do modelo
- Qualidade do ajustamento

Na primeira sub-etapa o modelo é analisado à luz de um certo conjunto de critérios referentes à qualidade estatística das estimativas obtidas. Na segunda, com base no estabelecimento de ensaios de hipóteses, estuda-se a adequação do modelo estimado à sucessão cronológica em estudo através dos correspondentes resíduos.

Um modelo que não satisfaça aos critérios aplicados nas duas sub-etapas deve ser rejeitado. Neste caso, as informações recolhidas durante o processo de avaliação podem sugerir indicações que orientem a formulação de um novo modelo e repete-se o ciclo até se encontrar um modelo satisfatório para descrever a sucessão cronológica em causa.

A prática aconselha a procura de modelos alternativos que possam ser considerados igualmente satisfatórios face aos diferentes critérios de avaliação. Em particular, devem avaliar-se os modelos “vizinhos” do primeiro modelo satisfatório encontrado, bem como modelos resultantes da alteração das transformações inicialmente empregues na estacionarização. De facto, em situações concretas, podem existir modelos distintos que descrevem a sucessão em causa com elevado grau de qualidade. Nestes casos o analista é confrontado com a escolha do “melhor” modelo, problema cuja resolução pode eventualmente ser efectuada com base em critérios adicionais, objecto de exposição posterior.

Encontrado o “melhor” modelo à luz dos últimos critérios e aplicado o mesmo no estabelecimento de previsões é natural esperar que conduza a erros de previsão inferiores aos associados com outros modelos. Contudo, a prática tem demonstrado que nem sempre esta situação é verificada, ou seja, nem sempre os modelos que melhor “explicam” uma sucessão em determinado intervalo de tempo são os que fornecem melhores previsões. Assim se o objectivo final da análise é o estabelecimento de previsões, deve proceder-se ao estudo comparativo, através do cálculo dos respectivos erros de previsão para um período pós-amstral ou não utilizado na estimação dos diferentes modelos considerados satisfatórios pelos critérios de avaliação do diagnóstico.

## 4.1. IDENTIFICAÇÃO

As diversas técnicas usadas na identificação não têm grande precisão e muitas delas dependem de métodos gráficos. Conseqüentemente, na etapa de identificação está-se mais interessado em reduzir o número de possíveis modelos do que em escolher o “melhor” modelo.

a) **Estacionarização da sucessão cronológica.** A identificação inicia-se pela estacionarização da sucessão cronológica, através de sucessivas transformações, até se conseguir a estabilização da variância, a neutralização da tendência e a eliminação de movimentos estritamente periódicos. As transformações mais usuais estão definidas no capítulo 4 de (Murteira, Müller, & Turkman, 1993).

A função de autocorrelação (FAC) estimada pode também sugerir algumas transformações. Uma convergência lenta para zero (figura 162) indicia uma não estacionarização em média e uma convergência lenta para zero sobre os atrasos múltiplos de 12 sugere a aplicação de uma diferenciação sazonal de grau um.

b) **Identificação pelas FAC e FACP estimadas.** A sub-etapa respeitante à selecção de um modelo é considerada um momento crucial no processo de modelização, sendo condição indispensável para uma correcta identificação o conhecimento profundo da Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) dos modelos mais usuais.

Existem três grandes classes de processos estacionários – AR<sup>5</sup>, MA<sup>6</sup> e ARMA – com características bem distintas em termos de FAC e FACP. São estas as classes estudadas neste trabalho.

---

<sup>5</sup> Autoregressive – Autoregressivo

<sup>6</sup> Moving Average – Média Móvel

**Tabela 1** – Características da FAC e FACP nos modelos clássicos

<b>Processo</b>	<b>FAC</b>	<b>FACP</b>
<b>AR</b>	Decaimento para zero sob forma exponencial ou sinusoidal amortecida	Decaimento brusco para zero a partir de um certo atraso $k$
<b>MA</b>	Decaimento brusco para zero a partir de um certo atraso $k$	Decaimento para zero sob forma exponencial ou sinusoidal amortecida
<b>ARMA</b>	Decaimento para zero sob forma exponencial ou sinusoidal amortecida	Decaimento para zero sob forma exponencial ou sinusoidal amortecida

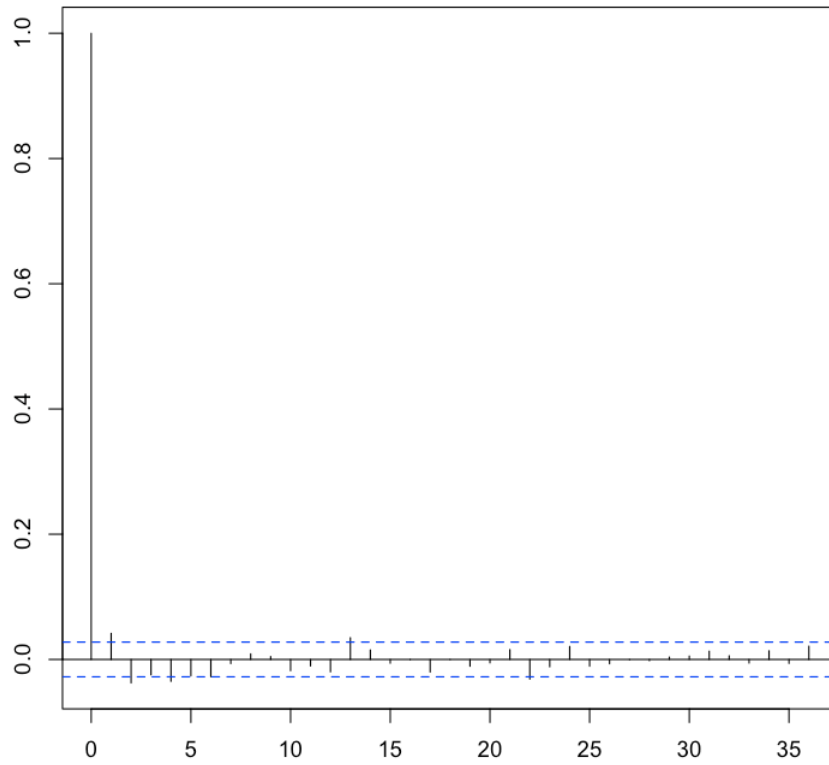
Os modelos multiplicativos dos componentes não sazonal e sazonal podem ser também caracterizados em termos das respectivas FAC e FACP, apesar destas funções possuírem um maior grau de complexidade por combinarem as especificidades de cada uma das componentes.

A identificação dos modelos faz-se, então, a partir das FAC e FACP estimadas para a sucessão estacionarizada, por comparação com os figurinos teóricos que se descreveram nos aspectos essenciais. Em particular, na identificação dos modelos multiplicativos, a componente não sazonal é modelizada através da observação dos primeiros valores das FAC e FACP estimadas.

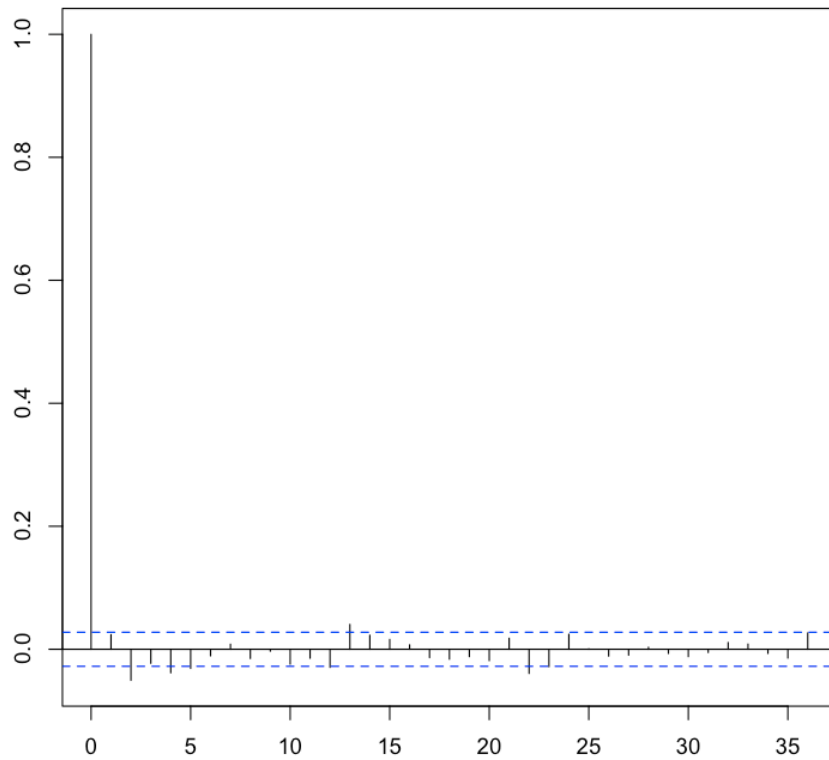
Para a identificação é essencial a presença de uma única autocorrelação e de uma única autocorrelação parcial significativa para  $k = 1$ , já que o padrão do decaimento para zero da FAC e da FACP é difícil de detectar, embora o decaimento da FACP pareça mais brusco.

Tome-se de seguida como exemplo ilustrativo (figuras 162-173) as representações gráficas da estimação da FAC e da FACP para três índices utilizados: BP, IBM e JPM.

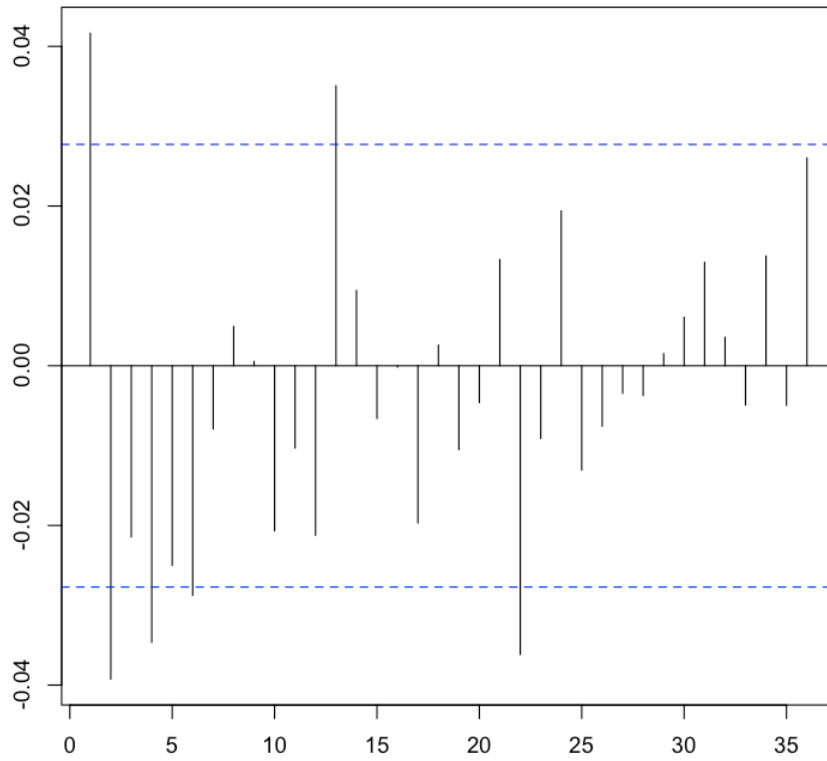
As FAC e FACP foram efectuadas sobre os dados de retorno bruto e às diferenças utilizando funções nativas da linguagem R, que será apresentado no capítulo 5.



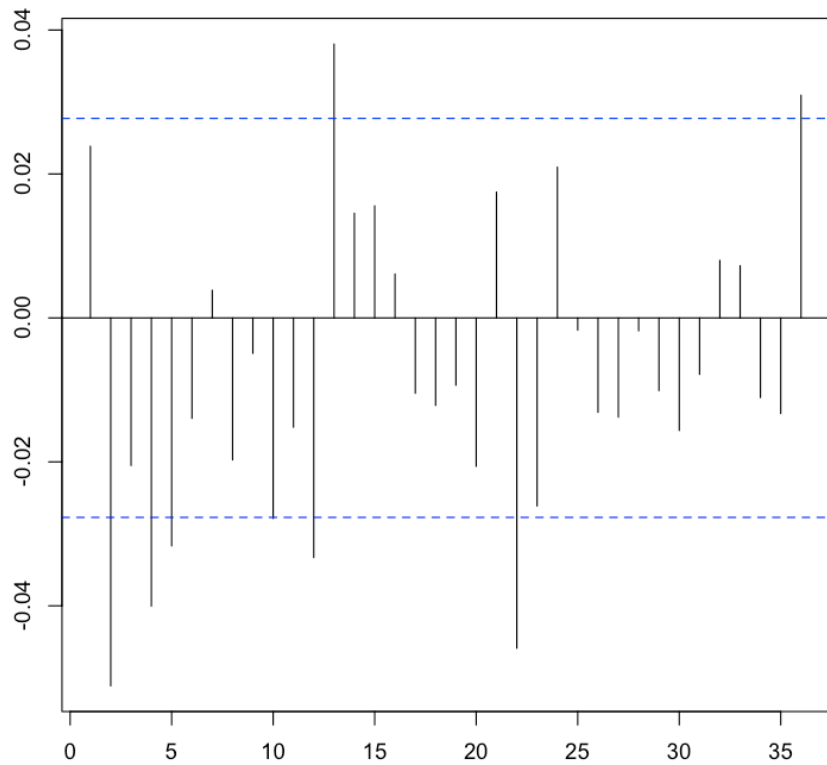
**Figura 162** – Representação da FAC do índice BP para os dados em retorno bruto



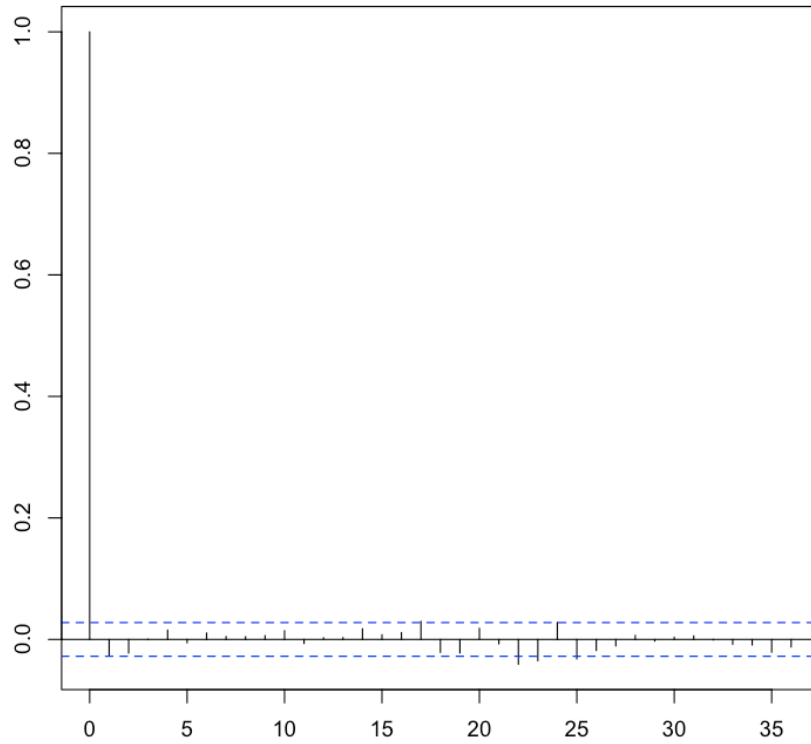
**Figura 163** – Representação da FAC do índice BP para os dados às diferenças



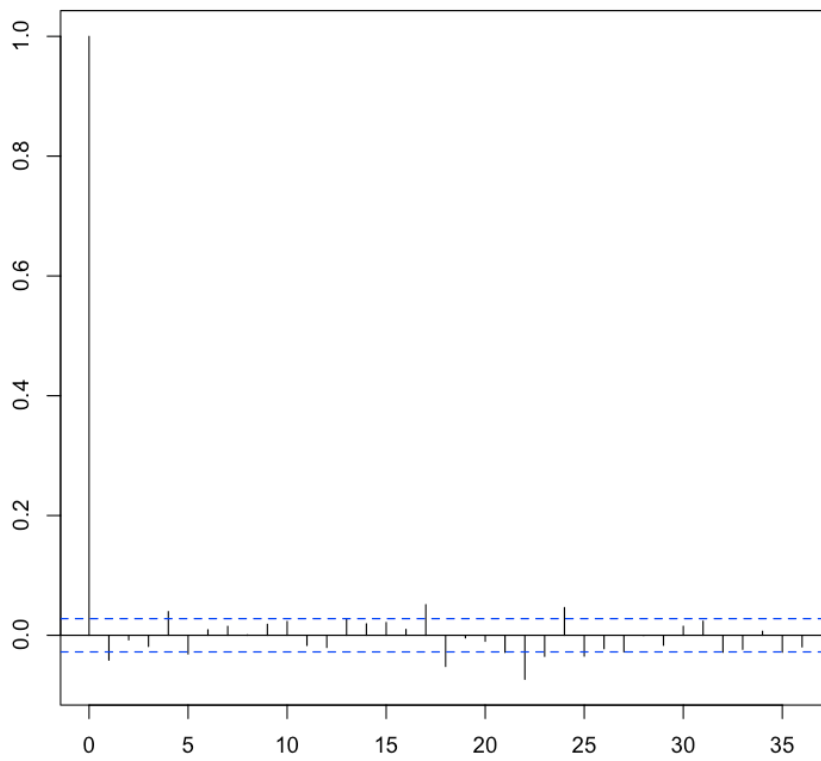
**Figura 164** – Representação da FACP do índice BP para os dados em retorno bruto



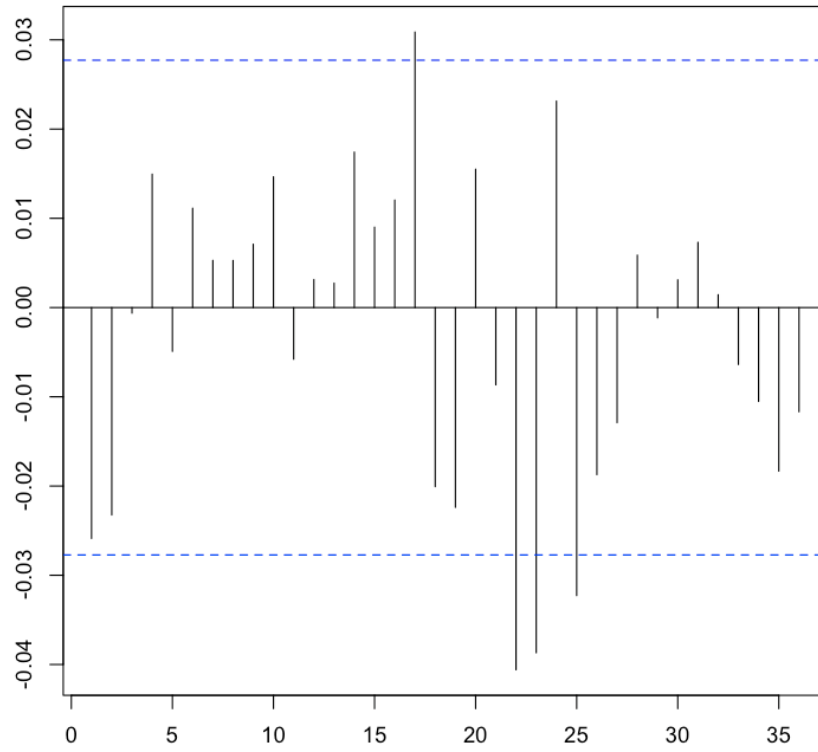
**Figura 165** – Representação da FACP do índice BP para os dados às diferenças



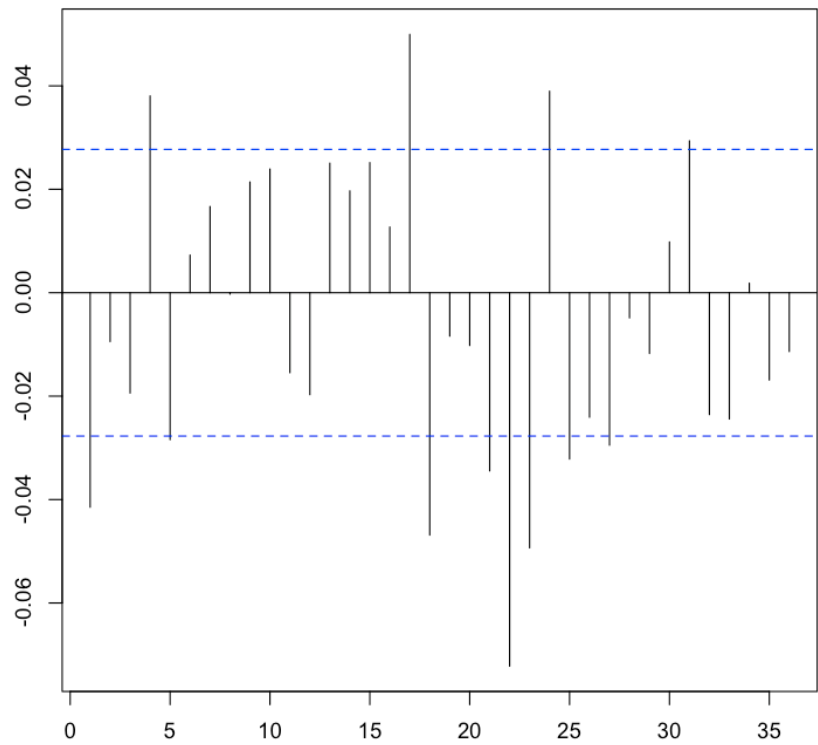
**Figura 166** – Representação da FAC do índice IBM para os dados em retorno bruto



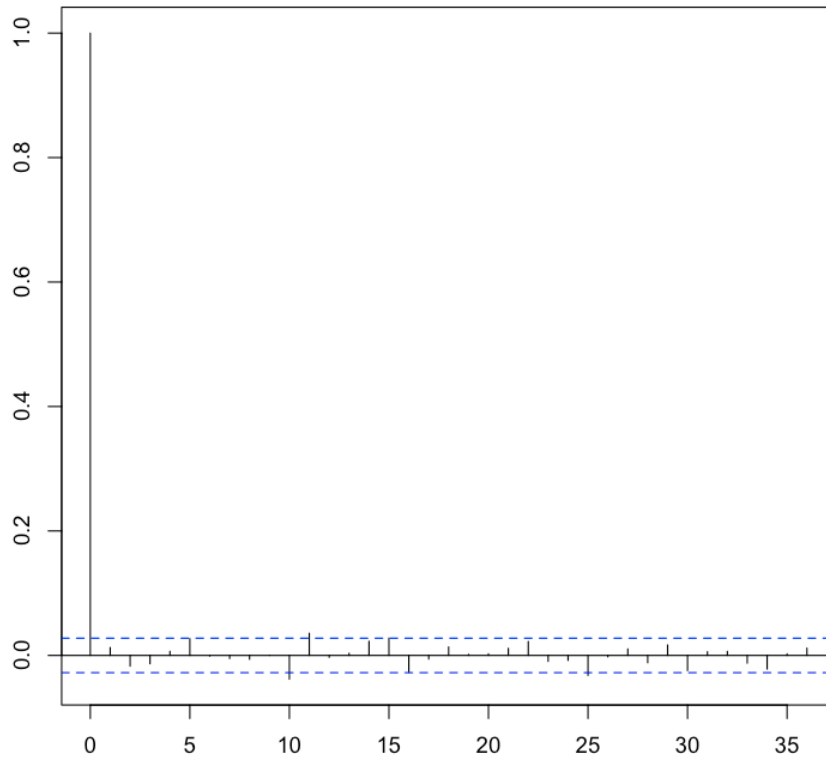
**Figura 167** – Representação da FAC do índice IBM para os dados às diferenças



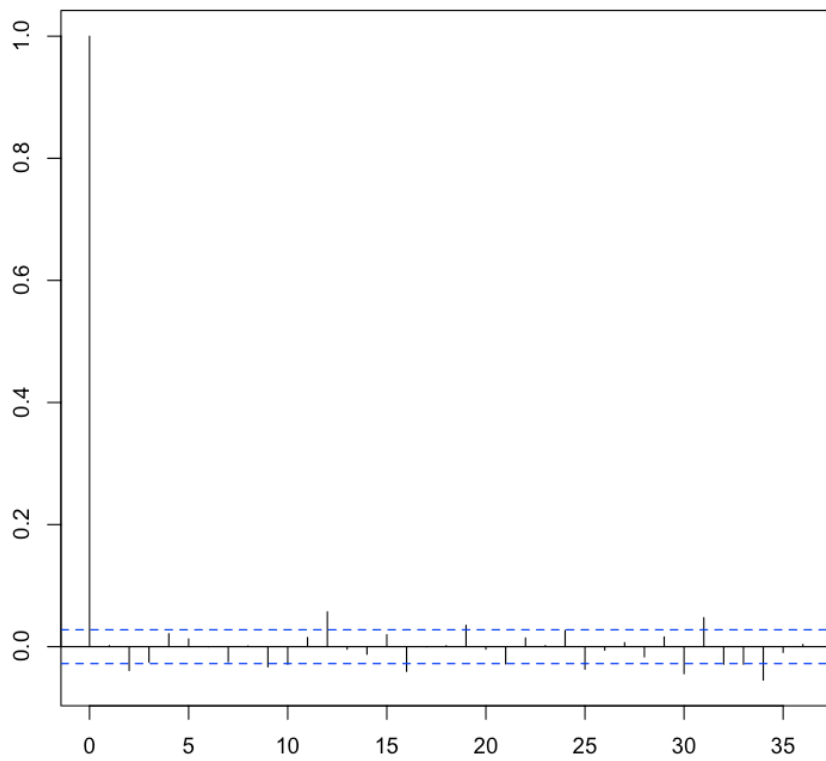
**Figura 168** – Representação da FACP do índice IBM para os dados em retorno bruto



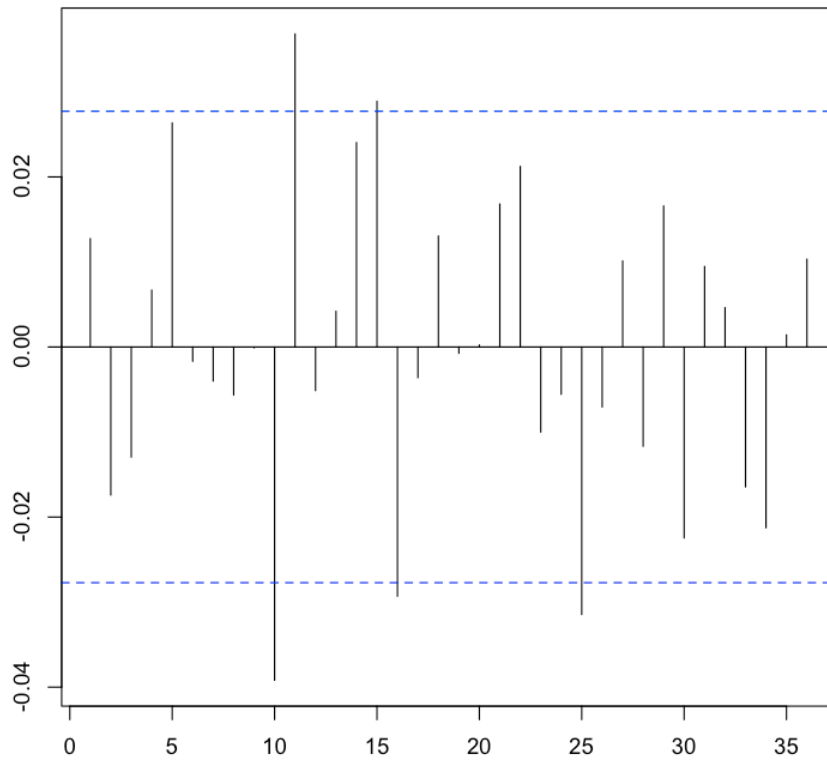
**Figura 169** – Representação da FACP do índice IBM para os dados às diferenças



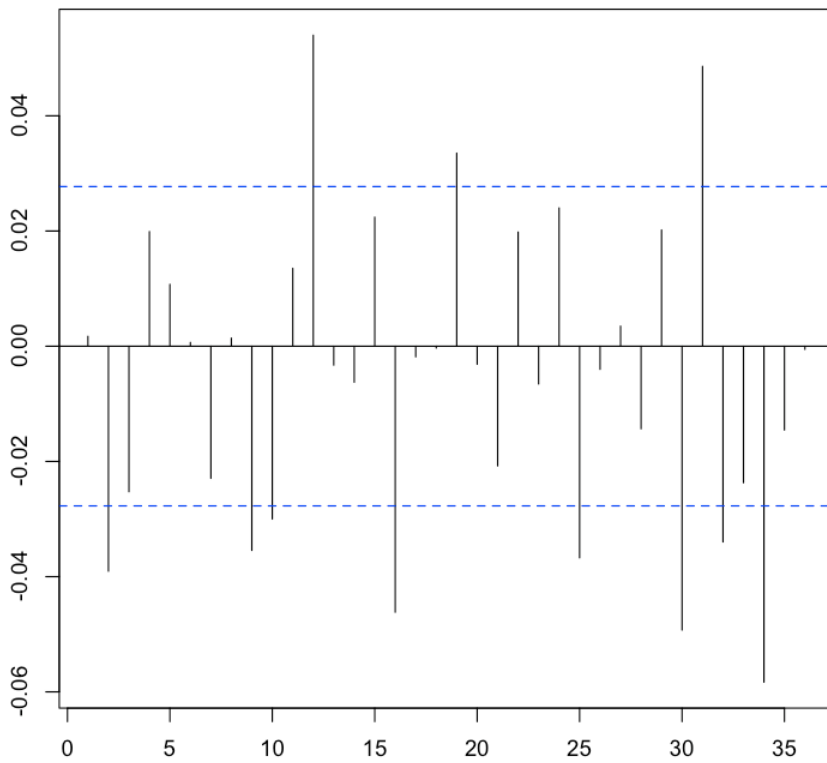
**Figura 170** – Representação da FAC do índice JPM para os dados em retorno bruto



**Figura 171** – Representação da FAC do índice JPM para os dados às diferenças



**Figura 172** – Representação da FACP do índice JPM para os dados em retorno bruto



**Figura 173** – Representação da FACP do índice JPM para os dados às diferenças

Analisando os gráficos das figuras 162-173, e com base na Tabela 1, pode-se considerar que o modelo indicado para a análise dos dados utilizados é o das Médias Móveis (MA), devido à FAC ter um decaimento brusco para zero a partir de um atraso  $k$ , e por a FACP ter um decaimento para zero sob a forma sinusoidal amortecida.

No entanto, as conclusões obtidas neste trabalho incidem principalmente nos resultados obtidos através dos cálculos de previsão efectuados, sendo possível realizar uma comparação entre as conclusões obtidas a partir da aplicação das FAC e FACP e os resultados obtidos experimentalmente.

**c) Identificação pela função de autocorrelação estendida.** O decaimento brusco para zero das FAC nos modelos MA e das FACP no modelo AR, apresentado no capítulo 6, permite identificar a ordem  $q$  dos primeiros e a ordem  $p$  dos segundos, com  $p, q \in \mathbb{N}$ . No entanto, nos modelos ARMA aquelas funções têm um comportamento amortecido para zero e não apresentam cortes bruscos, dificultando, deste modo, a identificação das ordens  $p$  e  $q$  dos operadores autoregressivos e de médias móveis.

Tsiao e Tsay (Tiao & Tsay, 1984) sugerem uma metodologia de identificação das ordens dos processos ARMA baseada em regressões iterativas e no conceito de função de autocorrelação estendida daí decorrente. No entanto, não sendo de especial interesse para este trabalho, esta metodologia não é aqui desenvolvida.

## 4.2. ESTIMAÇÃO

A etapa que se segue à da identificação é a da estimação dos parâmetros desconhecidos,

$$(\varphi_1, \dots, \varphi_p), (\theta_1, \dots, \theta_q), \sigma_\varepsilon^2 = \nu\{\varepsilon_t\} \quad (4.1)$$

do modelo ARMA(p,q).

No modelo das diversas metodologias existentes constata-se que apenas nos modelos mais simples é possível definir formas explícitas para os estimadores dos parâmetros. Em geral, os estimadores surgem como soluções de sistemas de equações não lineares, que necessitam de técnicas de análise numérica e de cálculo computacional para o estabelecimento das correspondentes estimativas.

Na óptica das aplicações esta etapa não requer, em geral, grande intervenção por parte do analista, dada a existência de pacotes informáticos que fornecem, uma vez identificado o modelo, as estimativas dos respectivos parâmetros.

No capítulo 6, dedicado aos modelos clássicos Autoregressivo (AR), Médias Móveis (MA) e Misto (ARMA) e é apresentada com detalhe a metodologia de estimação de parâmetros destes modelos.

### 4.3. ANÁLISE DO DIAGNÓSTICO

Tendo-se identificado um modelo ARMA e estimados os respectivos parâmetros, passa-se à etapa em que se procura avaliar, quer a qualidade estatística do modelo, quer a qualidade do ajustamento. A avaliação do diagnóstico reveste-se de especial importância e está intimamente relacionada com a de identificação. Se o modelo seleccionado for rejeitado durante o processo de avaliação, deve-se voltar à fase de identificação numa nova tentativa de escolher outros modelos com a informação recolhida no decorrer daquele processo.

#### 4.3.1. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: SIGNIFICÂNCIA ESTATÍSTICA DOS PARÂMETROS ESTIMADOS

Tendo presente o princípio da parcimónia<sup>7</sup>, devem analisar-se as estimativas obtidas para os parâmetros com o objectivo de eliminar aqueles que sejam desnecessários, ou seja, os parâmetros que não possam considerar-se significativamente diferentes de zero. A aplicação deste princípio tem grande importância pois a prática tem demonstrado que, em geral, modelos parcimoniosos produzem melhores previsões. Tem, por isso, interesse proceder, para cada um dos parâmetros do modelo identificado, ao ensaio da hipótese da nulidade. Considerando a coincidência no limite de *Student* com a Normal, a hipótese da nulidade pode rejeitar-se, ao nível aproximado de 5%, quando sai,

$$\left| \frac{\sqrt{N}\bar{\alpha}_i}{v_{ij}(\bar{\alpha})} \right| \geq 1,96 \quad (4.2)$$

caso em que o parâmetro se considera significativo e deve manter-se o modelo.

---

<sup>7</sup> O princípio da parcimónia é um ideal filosófico aplicado à ciência que indica que, no caso de se encontrar mais do que uma explicação para uma dada observação, deve-se adoptar a mais simples.

#### 4.3.2. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: ESTACIONARIDADE E INVERTIBILIDADE DO MODELO ESTIMADO

Os parâmetros estimados podem utilizar-se para averiguar a estacionaridade e invertibilidade do modelo, sobretudo nos casos em que os seus valores se encontram próximos da fronteira da região em que se verificam tais propriedades.

#### 4.3.3. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: ESTABILIDADE DO MODELO ESTIMADO

A análise da matriz de correlação entre os estimadores dos parâmetros deve incluir-se na etapa de avaliação do diagnóstico. Como é evidente, existe sempre uma certa correlação entre os estimadores obtidos, mas a presença de correlações elevadas sugere, em geral, a má qualidade dos mesmos. Quando dois estimadores são fortemente correlacionados, uma variação na estimativa do primeiro arrasta uma variação na estimativa do primeiro arrasta uma variação na estimativa do segundo sem que isso implique grandes alterações na soma dos quadrados dos resíduos (minimizada no processo de estimação). Este facto tem como consequência que uma pequena alteração na sucessão cronológica pode facilmente produzir estimativas significativamente diferentes. Nestas condições o modelo estimado não é o mais apropriado para descrever o comportamento futuro da sucessão cronológica, a não ser que este se afaste pouco do comportamento histórico. Está-se na presença de um modelo instável.

Como regra prática, consideram-se os estimadores altamente correlacionados quando o correspondente coeficiente de correlação for em módulo superior a 0,7. Assim, todas as vezes que ocorrer uma situação deste tipo devem procurar-se modelos alternativos.

#### 4.3.4. AVALIAÇÃO DA QUALIDADE ESTATÍSTICA: REDUNDÂNCIA ENTRE ESTIMATIVAS

Multiplicando ambos os operadores autoregressivos e de médias móveis do modelo misto,

$$\varphi_B X_t = \theta_B \varepsilon_t \quad (4.3)$$

por um mesmo factor  $(1 - \alpha\beta)$ , obtém-se um modelo idêntico ao modelo autoregressivo,

$$(1 - \alpha\beta) \cdot \varphi_B X_t = (1 - \alpha\beta) \cdot \theta_B \varepsilon_t \quad (4.4)$$

Sendo, portanto, aquele factor redundante.

Modelos que contenham factores redundantes ou quase redundantes devem ser evitados, não só por respeito ao princípio da parcimónia, como pela má qualidade das estimativas obtidas.

#### 4.3.5. AVALIAÇÃO DA QUALIDADE DO AJUSTAMENTO

A qualidade do ajustamento de um modelo estimado pode medir-se através da análise dos correspondentes resíduos. Assim, se o modelo ARMA(p,q),

$$\varphi_B X_t = \theta_B \varepsilon_t \quad (4.5)$$

foi identificado e se  $\bar{\varphi} = (\bar{\varphi}_1, \dots, \bar{\varphi}_p)$ ,  $\bar{\theta} = (\bar{\theta}_1, \dots, \bar{\theta}_q)$  representam os estimadores da máxima verosimilhança, os resíduos  $\varepsilon_t$  são definidos por,

$$\bar{\varepsilon}_t = \bar{\theta}_B^{-1} \cdot \bar{\varphi}_B X_t, \quad t = 1, 2, \dots, N \quad (4.6)$$

onde,

$$\bar{\varphi}_B = (1 - \bar{\varphi}_1 B - \dots - \bar{\varphi}_p B^p), \bar{\theta}_B = (1 - \bar{\theta}_1 B - \dots - \bar{\theta}_q B^q) \quad (4.7)$$

Box e Jenkins (Box & Jenkins, 1970) referem que nestas condições, se o modelo for adequado para descrever a sucessão cronológica, tem-se,

$$\bar{\varepsilon}_t = \varepsilon_t + O\left(\frac{1}{\sqrt{N}}\right) \quad (4.8)$$

o que significa que à medida que o número de observações aumenta, os resíduos aproximam-se do ruído branco,  $\varepsilon_t$ . Deste modo pode pensar-se que a análise dos resíduos pode dar indicações sobre a adequação do modelo estimado. Isto é, se os resíduos tiverem um “comportamento” análogo a um ruído branco, pode admitir-se que o modelo estimado descreve bem a sucessão em estudo.

Um ruído branco fica caracterizado pela forma peculiar das respectivas FAC e FACP, pelo que o comportamento dos resíduos costuma avaliar-se em termos dos correspondentes valores estimados destas funções.

Designando  $\bar{\rho}_{\bar{\varepsilon}}$  e  $\bar{\varphi}_{kk}(\bar{\varepsilon})$ , respectivamente, as FAC e FACP estimadas a partir da sucessão dos resíduos, vai definir-se um conjunto de testes estatísticos que permitem avaliar a qualidade do ajustamento. Estes testes, não sendo demonstrados neste trabalho, são desenvolvidos e exemplificados em (Murteira, Müller, & Turkman, 1993). Estes testes designam-se por:

- Teste de Kendall e Stuart (1968)
- Teste de Bartlett (1946)
- Teste de Box-Pierce (1970)
- Teste de Jenkins (1956) e Daniels (1956)

## 4.4. CRITÉRIO DE SELECÇÃO DE MODELOS

Na análise de uma sucessão cronológica podem existir vários modelos que a podem descrever de forma satisfatória, ou seja, é possível, por vezes, encontrar mais do que um modelo que passe o crivo dos diferentes critérios de avaliação do diagnóstico anteriormente estudados.

Em alguns casos a escolha do melhor modelo é tarefa fácil, noutras torna-se mais difícil e exige que se recorra a critérios de selecção de modelos. Estes critérios são, em geral, baseados em estatísticas construídas a partir da sucessão residual correspondente a um determinado modelo ajustado ou baseados nos erros de previsão produzidos por esse mesmo modelo. De seguida apresentam-se os mais importantes (Murteira, Müller, & Turkman, 1993).

### 4.4.1. CRITÉRIO DE AKAIKE

Considere-se um modelo com  $m$  parâmetros, estimados de forma independente, que foi ajustado a uma sucessão cronológica. Akaike (Akaike, A new look at the statistical model identification, 1974), utilizando conceitos da teoria da informação, estabeleceu um critério para medir a qualidade do ajustamento, definindo a grandeza AIC,

$$AIC_m = -2 \cdot \ln[\text{máximaverosimilhança}] + 2m \quad (4.9)$$

De acordo com este critério a ordem óptima de um modelo correspondente ao valor de  $m$  que minimiza o correspondente  $AIC_m$ .

O critério de Akaike para optar por um de vários modelos concorrentes pode justificar-se pela relação que tem com a medida de divergência de Kulkback-Leiber (Turkman, 1985). Contudo é também susceptível de interpretação directa: é lógico considerar um modelo relativamente melhor do que outro ou outros quando tem

verossimilhança mais elevada; contudo, modelos com maior número de parâmetros têm provavelmente verossimilhança mais elevada. Assim, se pretende encontrar-se um modelo parsimonioso, os modelos com elevado número de parâmetros devem ser penalizados e é por isso que o termo  $2m$  entra em (4.9); o que não é muito claro é a razão que leva a dar esta forma de penalização.

Tendo em conta as deduções da secção 4.3 referentes à estimação das três classes de modelos (AR, MA e ARMA), pode estabelecer-se genericamente que o logaritmo da função de verossimilhança para  $n$  observações efectivas é dado, a menos de uma constante aditiva, por,

$$L_{\alpha} = -\frac{n}{2} \cdot \ln(\sigma_{\varepsilon}^2) - \frac{1}{2\sigma_{\varepsilon}^2} \cdot S^*(\alpha) \quad (4.10)$$

onde  $S^*$  representa a soma dos quadrados de uma quantidade previamente definida como apropriada para cada uma das classes de modelos e  $\alpha$  o conjunto dos  $m$  parâmetros do modelo.

Maximizando esta verossimilhança em relação a  $\alpha$  e a  $\sigma_{\varepsilon}^2$  obtém-se como estimador deste último parâmetro,

$$\sigma_{\varepsilon}^2 = \frac{1}{n} \cdot S^*(\bar{\alpha}) \quad (4.11)$$

onde  $\bar{\alpha}$  representa o estimador da máxima verossimilhança de  $\alpha$ .

Assim, o máximo de  $L_{\bar{\alpha}}$  vem,

$$L_{\bar{\alpha}} = -\frac{n}{2} \cdot \ln(\bar{\sigma}_{\varepsilon}^2) - \frac{n}{2} \quad (4.12)$$

ignorando o segundo termo desta quantidade por ser independente de  $\alpha$  e  $m$ , a grandeza AIC anteriormente definida é equivalente a,

$$AIC_m = n \cdot \ln(\bar{\sigma}_\varepsilon^2) + 2m \quad (4.13)$$

O critério de Akaike, além de permitir pesquisar a ordem óptima de um modelo ajustado, é também utilizado como critério de selecção de modelos, resolvendo o problema da discriminação entre os que foram aceites na fase de avaliação do diagnóstico. À luz de tal critério deve seleccionar-se como “melhor” o modelo para o qual o correspondente AIC tenha o menor valor. Pela forma como o critério é definido verifica-se que são penalizados os modelos com maior número de parâmetros e com maior variância residual. Quer dizer, um modelo menos parcimonioso só é seleccionado se os seus resíduos tiverem uma variância significativamente mais pequena (melhor ajustamento) do que os modelos concorrentes mais parcimoniosos.

Por último, nota-se que, de acordo com a teoria exposta, o critério de Akaike apenas pode aplicar-se aos modelos que resultam de ajustamentos à mesma sucessão cronológica estacionarizada. Se a uma determinada sucessão correspondem, por exemplo, duas sucessões estacionárias (se as transformações aplicadas forem diferentes), os modelos ajustados a cada uma das sucessões não podem comparar-se por este critério.

#### 4.4.2. CRITÉRIO BIC

Mais recentemente, Akaike (Akaike, A Bayesian extension of the minimum AIC procedure, 1978) desenvolveu uma extensão Bayesiana do critério anterior, definindo para um modelo com  $m$  parâmetros e  $n$  observações efectivas a grandeza,

$$BIC_m = n \cdot \ln(\bar{\sigma}_\varepsilon^2) - (n - m) \cdot \ln(1) - \frac{m}{n} + m \cdot \ln(n) + m \cdot \ln\left[\frac{1}{m} \cdot \frac{\bar{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} - 1\right] \quad (4.14)$$

onde  $\bar{\sigma}_\varepsilon^2$  é o estimador da máxima verosimilhança de  $\sigma_\varepsilon^2$ .

Quando  $m$  é pequeno relativamente a  $n$  pode utilizar-se a aproximação,

$$\left[-(n - m) \cdot \ln(1) - \frac{m}{n}\right] = m \quad (4.15)$$

obtendo-se a expressão ligeiramente mais simples,

$$BIC_m = n \cdot \ln(\bar{\sigma}_\varepsilon^2) + m(1 + \ln(n)) + m \cdot \ln\left[\frac{1}{m} \cdot \frac{\bar{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} - 1\right] \quad (4.14)$$

Deste modo a diferença básica entre as grandezas AIC e BIC reside no segundo termo da última,  $2m$ , que foi substituído por  $m(1 + \ln(n))$ . Este facto tem como efeito um aumento do peso associado ao termo penalizador da falta de parcimónia e, conseqüentemente, a minimização do BIC leva, em geral, à selecção de modelos de menor ordem do que os obtidos pela minimização do AIC.

# 5. LINGUAGEM DE PROGRAMAÇÃO *R*

Neste capítulo é feita uma introdução à linguagem computacional *R* utilizada neste trabalho para a realização dos cálculos de análise e tentativa de previsão dos modelos utilizados. A escolha desta linguagem recaiu sobre o facto de que as funções necessárias para cálculo e previsão das metodologias tradicionais – Autoregressiva, Média Móvel e Modelo Misto – que já se encontram nas bibliotecas do *R*.

## 5.1. INTRODUÇÃO A *R*

O *R*<sup>8</sup> é uma linguagem e ambiente para computação gráfica e estatística. Trata-se de um projecto GNU<sup>9</sup> similar à linguagem e ambiente S que foi desenvolvida nos Laboratórios Bell (anteriormente AT&T, agora Lucent Technologies) por John Chambers e colegas. A linguagem *R* pode ser considerada como uma implementação diferente de *S*. Há algumas diferenças importantes, mas muito do código de *S* corre inalterado em *R*.

---

<sup>8</sup> <http://www.r-project.org>

<sup>9</sup> <http://www.gnu.org>

O *R* foi inicialmente desenvolvido por Ihaka e Gentleman, ambos da Universidade de Auckland, na Nova Zelândia. O actual desenvolvimento de *R* é levado a cabo por uma equipa constituída por uma dúzia de pessoas de diferentes instituições à volta do mundo. O desenvolvimento de *R* toma como vantagem o crescimento de uma comunidade que coopera no seu desenvolvimento devido à sua filosofia de open source. De facto, o código fonte de todos os componentes de *R* está disponível gratuitamente para inspecção e/ou adaptação na Internet.

O *R* fornece uma grande variedade de técnicas estatísticas (modelização linear e não linear, testes estatísticos clássicos, análises de séries temporais, classificação, ...) e gráficas, e é altamente expansível. A linguagem  $S^{10}$  é normalmente o veículo de escolha para investigação em metodologia estatística, e *R* fornece um caminho Open Source para a participação nessa actividade.

Um dos pontos fortes do *R* consiste na facilidade com que se consegue produzir representações gráficas com qualidade de publicação, incluindo símbolos matemáticos e fórmulas quando necessário. Foi tido um grande cuidado sobre as escolhas de design básico dos gráficos, mas o utilizador retém o controlo total.

O *R* está disponível como Software Livre sob os termos da GNU General Public License<sup>11</sup> da Free Software Foundation no formato de código fonte. Compila e corre numa variedade de plataformas UNIX e sistemas similares (incluindo FreeBSD e Linux), Windows e MacOS.

---

<sup>10</sup> A linguagem *S* é uma linguagem de programação estatística desenvolvida inicialmente por J. Chambers, R. Becker e A. Wilks nos *Bell Laboratories*.

<sup>11</sup> [www.r-project.org/COPYING](http://www.r-project.org/COPYING)

## 5.2. O AMBIENTE R

O ambiente do *R* é uma suite integrada de soluções de software para manipulação de dados, cálculos e demonstração gráfica, e inclui:

- uma solução eficiente de manipulação e gravação de dados
- uma suite de operadores para cálculo em vectores, em especial matrizes
- uma colecção vasta, coerente e integrada de ferramentas intermédias para análises de dados
- soluções gráficas para análises e representação de dados, seja no ecrã ou gravada em memória
- uma linguagem de programação bem desenvolvida, simples e eficiente que inclui condições, ciclos, funções recursivas definidas pelo utilizador e soluções de input e output

O termo “ambiente” é usado intencionalmente para caracteriza-lo como um sistema coerente e planeado, em vez de uma adição incremental de ferramentas específicas e inflexíveis, como é frequentemente o caso com outros softwares de análise de dados. Na figura 174, observa-se um típico ambiente gráfico de *R*, com consola, editor de dados, representação gráfica, editor de código e gestor de pacotes.

O *R*, tal como o *S*, permite aos utilizadores acrescentar funcionalidades adicionais definindo novas funções. Muito do sistema em si é escrito no dialecto *R* de *S*, o que o torna fácil para os utilizadores seguirem as escolhas algorítmicas feitas. Para tarefas computacionalmente intensivas, código em C, C++ e Fortran pode ser adicionada e chamado ao correr o programa. Utilizadores avançados podem escrever código C de modo a manipular directamente objectos em *R*.

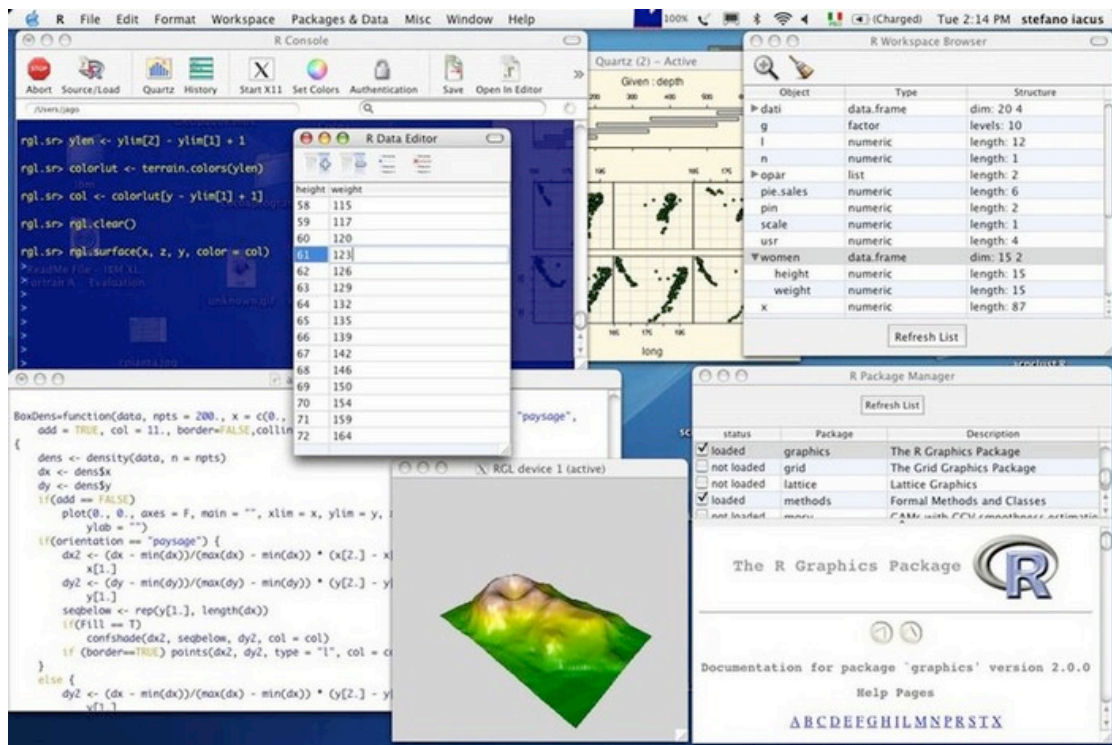


Figura 174 – Ambiente gráfico do R no sistema operativo Mac OS X

O R é considerado como um ambiente dentro do qual técnicas estatísticas são implementadas. Pode ser facilmente expandido através de pacotes. Há cerca de oito pacotes fornecidos com a distribuição de R, e muitos mais disponibilizados através da família CRAN de páginas de internet que cobrem um vasto leque de estatísticas modernas.

O R tem a sua própria documentação em formato do tipo LaTeX, que é usada para fornecer documentação abrangente.

Os pacotes utilizados neste trabalho, para facilitar o cálculo da previsão dos modelos AR, MA e ARMA são:

- `ar{stats}` – Adapta o modelo autoregressivo de séries temporais aos dados, seleccionando por defeito a complexidade pela AIC
- `arima{stats}` – Adapta um modelo ARIMA a uma série temporal univariada.

# 6. METODOLOGIAS CLÁSSICAS DE MODELIZAÇÃO LINEAR DE SÉRIES TEMPORAIS

Neste capítulo é dada uma breve descrição de alguns modelos matemáticos lineares mais utilizados na descrição, análise e previsão de índices financeiros. Começa-se por descrever os processos mais simples, tais como o Autoregressivo (AR), o processo de Médias Móveis (MA), e subsequentes generalizações com modelos Mistos (ARMA).

No final de cada secção são também demonstrados os resultados da comparação entre as várias metodologias. O valor apresentado corresponde ao Erro Médio Absoluto (EMA), para cada índice e para cada ordem do modelo. Este valor possibilita uma comparação com o valor do Limite de Futilidade ( $L_F$ ), de modo a verificar a adequação do modelo e respectiva ordem para a previsão de dado índice.

## 6.1. ERRO MÉDIO ABSOLUTO

Pretende-se o estabelecimento de uma medida de comparação entre os vários modelos de modelização linear de séries temporais, de modo a obter um processo de escolha do modelo mais adequado para determinada série temporal. Decidiu-se que o valor utilizado para essa comparação será o valor de Erro Médio Absoluto (EMA).

Considere-se  $Y_t$  como o conjunto de valores transformados (retorno bruto, retorno logarítmico ou às diferenças) de um dado índice financeiro em estudo. Neste trabalho,  $Y_t$  possui um tamanho de 5000 amostras, como referido no capítulo 2. Desta amostra, será usada uma amostra dos últimos  $\lambda$  ( $\lambda = 500$ ) valores conhecidos de  $Y_t$ . Denomina-se esta amostra como  $Y_C$ .

Seguidamente é criada a amostra, denominada de  $Y_P$ , com tamanho  $\lambda$  com os valores de previsão calculados para o intervalo da amostra de  $Y_t$  correspondente a  $Y_C$ , ou seja, para os últimos  $\lambda$  valores. Para tal, o método realiza  $\lambda$  iterações, para cada momento de  $Y_{t-k+i}$ , com  $i = \{1, \dots, \lambda\}$ .

Com a amostra dos valores calculados para a previsão,  $Y_P$ , pode-se calcular o Erro Médio Absoluto com a equação (6.1),

$$EMA = \frac{\sum_{i=0}^{\lambda} |Y_{C_i} - Y_{P_i}|}{\lambda} \quad (6.1)$$

## 6.2. LIMITE DE FUTILIDADE

Considere-se que, na situação de total ausência de informação acerca da série temporal em estudo, a melhor previsão para o valor não transformado do momento seguinte ao momento actual,  $X_{t+1}$ , é a do valor actual,  $X_t$ . Ou seja, tem-se,

$$X_{t+1} = X_t \quad (6.2)$$

No caso dos dados transformados,  $Y_t$ , isto é, para os valores de retorno bruto, retorno logarítmico e às diferenças, isto traduz-se em valores de zero, visto que a diferença entre valores iguais é nula.

$$Y_{t+1} = Y_t = 0 \quad (6.3)$$

Neste trabalho é utilizada esta heurística como linha de base para a verificação da adequação da capacidade de previsão de um modelo perante uma série temporal. Por outras palavras, se a média do erro de previsão do modelo for superior ao valor da variável do limite de futilidade, pode-se considerar que o modelo não estará adequado a ser utilizado como método de previsão para a dada série temporal, tornando assim a sua utilização inconsequente.

O valor da variável do limite de futilidade é calculado como a média da diferença absoluta entre os valores reais e o valor da previsão considerando a ausência de informação. No caso dos valores transformados, o valor da previsão considerando a ausência de informação é zero, como referido anteriormente. Assim, temos que o valor da variável do limite de futilidade,  $L_F$ , é dado por:

$$L_F = \frac{\sum_{i=0}^{\lambda} |Y_{C_i} - 0|}{\lambda} = \frac{\sum_{i=0}^{\lambda} |Y_{C_i}|}{\lambda} \quad (6.4)$$

onde  $\lambda$  é o tamanho da amostra da série temporal em estudo.

### 6.3. PROCESSOS AUTOREGRESSIVOS

O modelo autoregressivo insere-se no grupo de modelos de previsão lineares que tentam prever a saída de um sistema baseado em entradas e saídas anteriores do sistema. Estes modelos são modelos não causais, que, contrariamente aos modelos econométricos, usam apenas os valores anteriores do sinal para realizar a modelização e análise deste.

Um modelo que depende apenas das saídas anteriores do sistema é chamado de modelo autoregressivo (Autoregressive – AR). Processos autoregressivos, tal como o nome implica, regredem sobre os processos anteriores.

Este modelo definido por

$$Y_t = \alpha_1 \cdot Y_{t-1} + \dots + \alpha_p \cdot Y_{t-p} + \varepsilon_t \quad (6.5)$$

indica que os últimos  $p$  valores de  $Y_{t-i}$  para  $i = 1, \dots, p$  determinam conjuntamente a expectativa prevista de  $Y_t$  quando fornecidos dados dos períodos anteriores (Tsay, 2006).  $Y_t$  é uma combinação linear dos  $p$  valores mais recentes, adicionado de um termo de erro  $\varepsilon_t$ , que incorpora tudo o que é novo na série no momento  $t$  que não é explicado pelos valores passados. Isto é como um modelo de regressões múltiplas que não regrediu sobre variáveis independentes, mas sobre valores passados; assim, o termo “autoregressivo” é utilizado (Arsham, 1994).

Tratam-se de processos invertíveis. Em (Murteira, Müller, & Turkman, 1993), mostra-se que são também estacionárias quando as raízes do polinómio autoregressivo de ordem  $p$  estão fora do círculo unitário.

### 6.3.1. ESTIMAÇÃO DOS PARÂMETROS NOS MODELOS AUTOREGRESSIVOS

Sob a hipótese da normalidade do ruído branco, isto é, admitindo que os  $\varepsilon_t$  formam uma sucessão de variáveis aleatórias, o processo  $AR_p$  é Gaussiano. Supondo, sem perda de generalidade, que  $\mu=0$ , pode provar-se (Priestley, 2001) que o logaritmo da verosimilhança é dado, a menos de uma constante aditiva, pela expressão,

$$L(\varphi_1, \dots, \varphi_p) = \frac{-N}{2} \cdot \ln(\sigma_\varepsilon^2) + \frac{1}{2} \ln|V_p| = \frac{S^*(\varphi_1, \dots, \varphi_p)}{2\sigma_\varepsilon^2} \quad (6.6)$$

e,

$$S^*(\varphi_1, \dots, \varphi_p) = \sum_{i=1}^p \sum_{j=1}^p v_{ij} X_i X_j + \sum_{i=p+1}^N (X_i - \varphi_1 X_{i-1} - \dots - \varphi_p X_{i-p})^2 \quad (6.7)$$

com  $v_{ij}$  elemento genérico de  $V_p$ .

Uma vez que a matriz  $V_p$  depende dos parâmetros  $(\varphi_1, \dots, \varphi_p)$ , as derivadas de  $\ln|V_p|$  são difíceis de obter, resultando que os estimadores de máxima verosimilhança são solução de um sistema de equações não lineares. Contudo, demonstra-se em (Box & Jenkins, 1970) que para um  $N$  suficientemente grande, o termo  $\ln|V_p|$  é dominado por  $S^*(\varphi_1, \dots, \varphi_p)$ , podendo desprezar-se na expressão (6.6). Deste modo, obtêm-se aproximações aos estimadores da máxima verosimilhança minimizando  $S^*(\varphi_1, \dots, \varphi_p)$ . Uma segunda aproximação pode efectuar-se quando  $p$  é pequeno comparado com  $N$  e consiste em tomar,

$$L(\varphi_1, \dots, \varphi_p) = \text{constante} - \frac{S^*(\varphi_1, \dots, \varphi_p)}{2\sigma_\varepsilon^2} \quad (6.8)$$

com,

$$S(\varphi_1, \dots, \varphi_p) = \sum_{i=p+1}^N (X_i - \varphi_1 X_{i-1} - \dots - \varphi_p X_{i-p})^2 \quad (6.9)$$

saindo os estimadores da minimização de (6.8). Esta segunda aproximação corresponde ao método da máxima verosimilhança condicional em que os valores de  $(X_1, \dots, X_p)$  se consideram dados e em que, portanto, não se considera a respectiva distribuição conjunta.

### 6.3.2. RESULTADOS

De seguida são apresentados os valores resultantes dos cálculos efectuados pelo código desenvolvido em R.

Decidiu-se realizar experiências com a ordem  $p$  no intervalo  $p = \{1, \dots, 5\}$ , para cada uma transformações realizadas nos dados das séries financeiras em estudo.

São assim apresentados (Tabelas 2 a 5) os resultados para os dados transformados em retorno absoluto, seguido dos retornos logarítmicos e finalmente para valores às diferenças.

Para cada um dos índices, e em cada tipo de transformação de dados, é também indicado o correspondente valor do Limite de Futilidade ( $L_F$ ), de modo a ser possível uma comparação imediata dos resultados obtidos com o  $L_F$ . Com este valor de  $L_F$ , é criado uma tabela com valores comparativos, de modo a verificar-se a percentagem de melhoria  $M$  de cada ordem  $p$  de AR para cada índice.  $M$  permite comparar os vários tipos de transformação efectuada nos valores do sinal. Um valor positivo de percentagem indica um método melhor que a simples utilização do  $L_F$ . Deste modo, também é possível comparar as diferenças em utilizar os dados com as transformações em retorno bruto ou o retorno logarítmico. Para facilitar a análise, os valores de melhoria positivos estão indicados a cinza nas tabelas.

**Tabela 2** – Valores do EMA para o método AR dos valores de Retorno Absoluto

	Ordem de AR					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6860	1,6847	1,6867	1,6899	1,6896	1,6866
BA	0,9839	0,9841	0,9840	0,9868	0,9866	0,9855
BP	0,9585	0,9586	0,9592	0,9589	0,9584	0,9577
CVX	1,0555	1,0571	1,0570	1,0554	1,0550	1,0585
C	0,6743	0,6746	0,6741	0,6746	0,6743	0,6704
DELL	1,4987	1,4997	1,4989	1,4984	1,4905	1,4858
XOM	1,1245	1,1313	1,1319	1,1368	1,1366	1,1333
F	1,5173	1,5173	1,5169	1,5159	1,5159	1,5051
GE	0,6356	0,6348	0,6347	0,6352	0,6348	0,6319
GM	1,7925	1,7911	1,7910	1,7905	1,7894	1,7881
HPQ	1,0754	1,0720	1,0733	1,0712	1,0716	1,0786
IBM	0,7296	0,7315	0,7317	0,7311	0,7311	0,7264
JPM	0,7564	0,7570	0,7568	0,7565	0,7564	0,7571
MSFT	1,0912	1,0899	1,0948	1,0950	1,0946	1,0871
SNE	1,1637	1,1640	1,1639	1,1652	1,1652	1,1634
FTSE100	0,8444	0,8454	0,8469	0,8486	0,8487	0,8442
NASDAQ	0,9563	0,9551	0,9552	0,9553	0,9551	0,9626
NYSE	0,7671	0,7680	0,7680	0,7675	0,7656	0,7665
NIKKEI225	1,0195	1,0227	1,0240	1,0244	1,0245	1,0216
PSI20	0,4838	0,4869	0,4844	0,4834	0,4840	0,4827

**Tabela 3** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método AR de valores de Retorno Absoluto

	Ordem de AR				
	1	2	3	4	5
AAPL	0,04%	0,11%	-0,01%	-0,20%	-0,18%
BA	0,16%	0,14%	0,15%	-0,13%	-0,11%
BP	-0,08%	-0,09%	-0,16%	-0,13%	-0,07%
CVX	0,28%	0,13%	0,14%	0,29%	0,33%
C	-0,58%	-0,63%	-0,55%	-0,63%	-0,58%
DELL	-0,87%	-0,94%	-0,88%	-0,85%	-0,32%
XOM	0,78%	0,18%	0,12%	-0,31%	-0,29%
F	-0,81%	-0,81%	-0,78%	-0,72%	-0,72%
GE	-0,59%	-0,46%	-0,44%	-0,52%	-0,46%
GM	-0,25%	-0,17%	-0,16%	-0,13%	-0,07%
HPQ	0,30%	0,61%	0,49%	0,69%	0,65%
IBM	-0,44%	-0,70%	-0,73%	-0,65%	-0,65%
JPM	0,09%	0,01%	0,04%	0,08%	0,09%
MSFT	-0,38%	-0,26%	-0,71%	-0,73%	-0,69%
SNE	-0,03%	-0,05%	-0,04%	-0,15%	-0,15%
FTSE100	-0,02%	-0,14%	-0,32%	-0,52%	-0,53%
NASDAQ	0,65%	0,78%	0,77%	0,76%	0,78%
NYSE	-0,08%	-0,20%	-0,20%	-0,13%	0,12%
NIKKEI225	0,21%	-0,11%	-0,23%	-0,27%	-0,28%
PSI20	-0,23%	-0,87%	-0,35%	-0,15%	-0,27%

**Tabela 4** – Valores do EMA para o método AR dos valores de Retorno Logarítmico

	Ordem de AR					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6780	1,6771	1,6788	1,6825	1,6816	1,6785
BA	0,9838	0,9839	0,9839	0,9867	0,9866	0,9844
BP	0,9583	0,9583	0,9588	0,9587	0,9580	0,9576
CVX	1,0555	1,0571	1,0571	1,0554	1,0552	1,0581
C	0,6735	0,6736	0,6733	0,6737	0,6734	0,6701
DELL	1,4959	1,4963	1,4957	1,4956	1,4885	1,4878
XOM	1,1257	1,1323	1,1332	1,1382	1,1379	1,1335
F	1,5178	1,5178	1,5174	1,5166	1,5166	1,5065
GE	0,6342	0,6332	0,6333	0,6337	0,6333	0,6316
GM	1,7891	1,7877	1,7877	1,7872	1,7861	1,7858
HPQ	1,0731	1,0701	1,0714	1,0692	1,0698	1,0756
IBM	0,7291	0,7309	0,7310	0,7302	0,7302	0,7270
JPM	0,7565	0,7571	0,7571	0,7568	0,7569	0,7566
MSFT	1,0881	1,0868	1,0916	1,0918	1,0914	1,0861
SNE	1,1617	1,1620	1,1618	1,1631	1,1631	1,1617
FTSE100	0,8445	0,8445	0,8471	0,8488	0,8489	0,8444
NASDAQ	0,9585	0,9569	0,9571	0,9572	0,9569	0,9628
NYSE	0,7679	0,7686	0,7687	0,7682	0,7662	0,7670
NIKKEI225	1,0209	1,0238	1,0255	1,0259	1,0261	1,0229
PSI20	0,4843	0,4874	0,4848	0,4838	0,4843	0,4827

**Tabela 5** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método AR de valores de Retorno Logarítmico

	Ordem de AR				
	1	2	3	4	5
AAPL	0,03%	0,08%	-0,02%	-0,24%	-0,18%
BA	0,06%	0,05%	0,05%	-0,23%	-0,22%
BP	-0,07%	-0,07%	-0,13%	-0,11%	-0,04%
CVX	0,25%	0,09%	0,09%	0,26%	0,27%
C	-0,51%	-0,52%	-0,48%	-0,54%	-0,49%
DELL	-0,54%	-0,57%	-0,53%	-0,52%	-0,05%
XOM	0,69%	0,11%	0,03%	-0,41%	-0,39%
F	-0,75%	-0,75%	-0,72%	-0,67%	-0,67%
GE	-0,41%	-0,25%	-0,27%	-0,33%	-0,27%
GM	-0,18%	-0,11%	-0,11%	-0,08%	-0,02%
HPQ	0,23%	0,51%	0,39%	0,60%	0,54%
IBM	-0,29%	-0,54%	-0,55%	-0,44%	-0,44%
JPM	0,01%	-0,07%	-0,07%	-0,03%	-0,04%
MSFT	-0,18%	-0,06%	-0,51%	-0,52%	-0,49%
SNE	0,00%	-0,03%	-0,01%	-0,12%	-0,12%
FTSE100	-0,01%	-0,01%	-0,32%	-0,52%	-0,53%
NASDAQ	0,45%	0,61%	0,59%	0,58%	0,61%
NYSE	-0,12%	-0,21%	-0,22%	-0,16%	0,10%
NIKKEI225	0,20%	-0,09%	-0,25%	-0,29%	-0,31%
PSI20	-0,33%	-0,97%	-0,44%	-0,23%	-0,33%

**Tabela 6** – Valores do EMA para o método AR dos valores às Diferenças

	Ordem de AR					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,2239	1,2275	1,2279	1,2301	1,2300	1,2225
BA	0,7788	0,7783	0,7787	0,7813	0,7811	0,7810
BP	0,6476	0,6482	0,6488	0,6489	0,6484	0,6476
CVX	0,6877	0,6893	0,6894	0,6893	0,6889	0,6887
C	0,3314	0,3313	0,3312	0,3314	0,3309	0,3315
DELL	0,3553	0,3554	0,3553	0,3554	0,3553	0,3553
XOM	0,9282	0,9324	0,9333	0,9362	0,9367	0,9326
F	0,1238	0,1238	0,1238	0,1238	0,1238	0,1220
GE	0,2209	0,2204	0,2205	0,2205	0,2203	0,2203
GM	0,4947	0,4947	0,4945	0,4943	0,4940	0,4929
HPQ	0,3637	0,3630	0,3633	0,3623	0,3623	0,3648
IBM	0,6051	0,6058	0,6072	0,6056	0,6059	0,6028
JPM	0,3294	0,3301	0,3302	0,3297	0,3297	0,3297
MSFT	0,3223	0,3231	0,3251	0,3255	0,3254	0,3224
SNE	0,4715	0,4718	0,4738	0,4740	0,4740	0,4720
FTSE100	51,4776	51,6082	51,6990	51,7435	51,7597	51,5324
NASDAQ	18,1116	18,1001	18,1127	18,1187	18,1012	18,1374
NYSE	71,0219	71,1548	71,1690	71,1775	70,9537	71,1076
NIKKEI225	157,4088	158,0049	158,2614	158,2771	158,0795	157,4307
PSI20	53,0623	53,5141	53,1653	53,1193	53,4691	52,8405

**Tabela 7** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método AR de valores às Diferenças

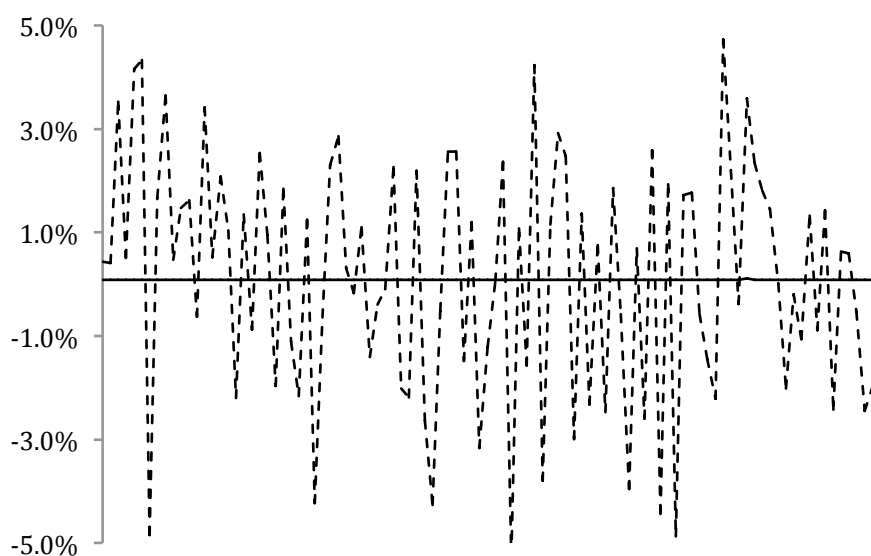
	Ordem de AR				
	1	2	3	4	5
AAPL	-0,11%	-0,41%	-0,44%	-0,62%	-0,61%
BA	0,28%	0,35%	0,29%	-0,04%	-0,01%
BP	0,00%	-0,09%	-0,19%	-0,20%	-0,12%
CVX	0,15%	-0,09%	-0,10%	-0,09%	-0,03%
C	0,03%	0,06%	0,09%	0,03%	0,18%
DELL	0,00%	-0,03%	0,00%	-0,03%	0,00%
XOM	0,47%	0,02%	-0,08%	-0,39%	-0,44%
F	-1,48%	-1,48%	-1,48%	-1,48%	-1,48%
GE	-0,27%	-0,05%	-0,09%	-0,09%	0,00%
GM	-0,37%	-0,37%	-0,32%	-0,28%	-0,22%
HPQ	0,30%	0,49%	0,41%	0,69%	0,69%
IBM	-0,38%	-0,50%	-0,73%	-0,46%	-0,51%
JPM	0,09%	-0,12%	-0,15%	0,00%	0,00%
MSFT	0,03%	-0,22%	-0,84%	-0,96%	-0,93%
SNE	0,11%	0,04%	-0,38%	-0,42%	-0,42%
FTSE100	0,11%	-0,15%	-0,32%	-0,41%	-0,44%
NASDAQ	0,14%	0,21%	0,14%	0,10%	0,20%
NYSE	0,12%	-0,07%	-0,09%	-0,10%	0,22%
NIKKEI225	0,01%	-0,36%	-0,53%	-0,54%	-0,41%
PSI20	-0,42%	-1,27%	-0,61%	-0,53%	-1,19%

### 6.3.3. OBSERVAÇÕES

Analisa-se de seguida os resultados para o Erro Médio Absoluto (EMA) obtidos com o modelo Autoregressivo para os índices em estudo.

Uma observação quase imediata é a grande aproximação dos valores de previsão obtidos aos valores do Limite de Futilidade,  $L_F$ . Este fenómeno é explicado tendo em consideração o modo como é obtido o  $L_F$ , com a equação (6.3), e os valores calculados pelo modelo Autoregressivo, observáveis na Figura 175. Verifica-se que as previsões efectuadas pelo modelo Autoregressivo aproximam-se muito do 0.01%, colocando-se apenas ligeiramente acima deste valor. Também é verificável que nunca é realizada uma previsão negativa.

Como tal, é de esperar que os valores do EMA sejam muito próximos do valor de  $L_F$ , para qualquer um dos índices, e para qualquer uma das ordens  $p$  de AR.



**Figura 175** – Evolução dos valores reais e de previsão do índice AAPL para Retorno Bruto com AR<sub>1</sub>

Em termos de comparação entre os vários tipos de transformação dos dados, verifica-se que os dados transformados às diferenças possuem um maior número de

índices que apresentam uma melhoria das previsões comparativamente ao  $L_F$ , com 12 dos 20 índices a apresentar um valor positivo de melhoria para uma ou mais ordens  $p$  de AR. Para os valores de Retorno Bruto e Retorno Logarítmico, apenas 9 dos 20 índices apresentam uma ou mais ordens  $p$  de AR com um valor positivo de melhoria.

Comparando de seguida os resultados obtidos entre as transformações de dados de retorno bruto e retorno logarítmico, observa-se que os valores de EMA obtidos com estas transformações, apresentam valores de melhoria positivo para as mesmas ordens  $p$  de AR para cada um dos índices. Logo, a diferença de utilização entre ambas as transformações infere sobre a percentagem de melhoria relativamente ao  $L_F$  correspondente. Neste caso, nota-se percentagens de melhoria superiores para os dados que sofreram uma transformação para retorno bruto, do que para os transformados em retorno logarítmico.

Apresenta-se de seguida a Tabela 8, que resume os melhores resultados obtidos com o modelo AR. Os índices estudados que não estão presentes na tabela são os índices para os quais não foi possível encontrar uma melhoria relativamente ao valor de  $L_F$ .

**Tabela 8** – Resumo de resultados do modelo AR

Índice	Ordem $p$ de AR	Transformação	M
AAPL	2	Retorno Bruto	0,11%
BA	2	Diferenças	0,35%
CVX	5	Retorno Bruto	0,33%
C	5	Diferenças	0,18%
XOM	1	Retorno Bruto	0,78%
HPQ	4 e 5	Diferenças	0,69%
JPM	1 e 5	Retorno Bruto	0,09%
MSFT	1	Diferenças	0,03%
SNE	1	Diferenças	0,11%
FTSE100	1	Diferenças	0,11%
NASDAQ	5	Retorno Bruto	0,78%
NYSE	5	Diferenças	0,22%
NIKKEI225	1	Retorno Bruto	0,21%

Observando a tabela 8, pode-se inferir que, para as séries dos índices estudados, com o modelo Autoregressivo, os valores transformados em Retorno Logarítmico têm sempre resultados piores em termos de melhoria comparativamente ao valor de  $L_F$  do que os valores transformados em retorno bruto.

Dos 13 índices que se obteve valores positivos de  $M$ , em 6 foi a transformação em Retorno Bruto que obteve o melhor resultado, e nos restantes 7 foi a transformação às diferenças.

Também se verifica que em 6 índices, foi a ordem  $p = 5$  que obteve melhores resultados. Possivelmente, aumentando o valor da ordem  $p$  de AR seriam obtidos valores melhores de EMA.

## 6.4. PROCESSOS DE MÉDIAS MÓVEIS

Há várias maneiras de introduzir os modelos de Médias Móveis (Moving Average – MA). Um método é tratar o modelo como uma simples extensão de séries de ruído branco. Outro método, aqui explorado, é tratar o modelo como um modelo autoregressivo de ordem infinita com algumas restrições de parâmetros (Tsay, 2006).

### 6.4.1 MÉDIA MÓVEL SIMPLES

Um dos métodos mais intuitivo de previsão é o das médias móveis, já que este simplesmente toma um certo número de períodos passados e soma-os, dividindo a seguir pelo número de períodos (Arsham, 1994). A média móvel simples é uma técnica eficaz e eficiente desde que a série temporal seja estacionária tanto na sua média como na variância. A seguinte fórmula é usada para encontrar a média móvel, MA, para o período  $t + I$ ,

$$MA_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-n+1}}{n} \quad (6.10)$$

onde  $n$  é o número de observações usadas no cálculo e  $Y_t$  é o último valor observado.

A previsão para o momento  $t + I$  é a previsão para todos os momentos temporais futuros. No entanto, esta previsão apenas é revista quando dados novos se tornam disponíveis.

A fórmula geral de uma média móvel pode ser expressa como

$$Y_t = \mu_0 \cdot \varepsilon_t + \mu_1 \cdot \varepsilon_{t-1} + \dots + \mu_q \cdot \varepsilon_{t-q} \quad (6.11)$$

onde  $\varepsilon_t$  é um processo de ruído branco. Este modelo pode ser comparado com um modelo autoregressivo do tipo demonstrado em (Arsham, 1994).

Deste modo, enquanto que um processo autoregressivo depende de uma combinação linear da função  $Y_t$  com os seus próprios valores desfasados do tempo, o processo de média móvel depende de uma combinação similar da função  $\varepsilon_t$  com os seus valores desfasados. A afinidade dos dois tipos de processo é ainda confirmado quando é reconhecido que um processo autoregressivo de ordem finita é equivalente a um processo de média móvel de ordem infinita e que, inversamente, um processo de média móvel de ordem finita é apenas um processo autoregressivo de ordem infinita (Pollock, 1993), como ilustrado na seguinte função (6.11).

$$\begin{aligned} AR_p &\Rightarrow MA_\infty \\ MA_q &\Rightarrow AR_\infty \end{aligned} \quad (6.11)$$

O modelo de médias móveis implica que os efeitos produzidos pelas inovações só perduram por um curto período de tempo o que contrasta com o que se passa com os processos autoregressivos em que os efeitos persistem por um período de tempo mais alargado (Murteira, Müller, & Turkman, 1993).

#### 6.4.2. MÉDIA MÓVEL PONDERADA

Este método é bastante poderoso e económico em termos computacionais. Sendo frequentemente usado quando são necessárias previsões repetidas. Utiliza métodos tal como a soma dos dígitos e métodos de ajustamentos às tendências. Como exemplo, uma média móvel ponderada (Weighed Moving Average – WMA) de ordem 3 é dado por:

$$WMA_3 = w_1 \cdot Y_t + w_2 \cdot Y_{t-1} + w_3 \cdot Y_{t-2} \quad (6.12)$$

onde os pesos  $w_i$ ,  $i = 1, 2, 3$ , são quaisquer números positivos tais que  $w_1 + w_2 + w_3 = 1$ .

### 6.4.3. ESTIMAÇÃO DOS PARÂMETROS NOS MODELOS DE MÉDIAS MÓVEIS

O método dos momentos, aplicado aos processos de médias móveis, permite, através das correspondentes equações de Yule-Walker, determinar estimadores para os respectivos parâmetros. Por exemplo, para um processo de média móvel de ordem 1,  $MA_1$ , têm-se uma equação teórica,

$$\rho_1 = \frac{-\theta}{1 + \theta^2} \quad (6.13)$$

de onde sai o estimador,

$$\bar{\theta} = \frac{-1 \pm \sqrt{1 - \bar{\rho}_1^2}}{2\bar{\rho}_1} \quad (6.14)$$

que deve tomar-se com o sinal que garanta a invertibilidade. No entanto, apesar da viabilidade do método prova-se que conduz, em geral, a estimadores não eficientes, sendo por isso apenas utilizado para fornecer valores preliminares na inicialização de algoritmos de pesquisa de estimadores da máxima verosimilhança.

Mantendo a hipótese anterior sobre o ruído branco, Box e Jenkins (Box & Jenkins, 1970) desenvolveram uma metodologia de máxima verosimilhança condicional para os processos de média móvel de ordem  $q$ ,  $MA_q$ ,

$$Y_t = \varepsilon_t - \varphi_1 \cdot \varepsilon_{t-1} - \dots - \varphi_q \cdot \varepsilon_{t-q} \quad (6.15)$$

Que consiste em fixar os valores iniciais,

$$\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \dots = \varepsilon_{-(q-1)} = 0 \quad (6.16)$$

e considerar a seguinte transformação,

$$Y_t = \varepsilon_t - \varphi_1 \cdot \varepsilon_{t-1} - \dots - \varphi_q \cdot \varepsilon_{t-q}, \quad q < t \leq N \quad (6.17)$$

Uma vez que o Jacobiano desta transformação é igual à unidade, o logaritmo da verosimilhança assume a forma,

$$L(\varphi_1, \dots, \varphi_q) = \text{constante} - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^N \varepsilon_i^2 \quad (6.18)$$

Ensaando valores de  $(\varphi_1, \dots, \varphi_q)$ , compatíveis com a invertibilidade do processo  $MA_q$ , podem pesquisar-se numericamente os valores que maximizam a verosimilhança (6.17) ou, equivalentemente, minimizam,

$$S(\varphi_1, \dots, \varphi_q) = \sum_{i=1}^N \varepsilon_i^2 \quad (6.19)$$

Assim, com o auxílio de meios informáticos, obtêm-se os chamados estimadores da máxima verosimilhança condicional que são também estimadores dos mínimos quadrados condicionais.

#### 6.4.4. RESULTADOS

Tal como para o modelo Autoregressivo, são de seguida apresentados (Tabelas 9-14) os resultados do EMA para o modelo de Média Móvel de ordem  $q$ , com  $q = \{1, \dots, 5\}$ , para os dados transformados em Retorno Bruto, Retorno Logarítmico e às Diferenças.

O cálculo do EMA foi realizado com a ajuda das funções presentes no compilador de R, tal como para o cálculo do EMA do modelo Autoregressivo. O algoritmo encontra-se para observação em anexo.

A disposição dos dados e a subsequente análise é igual à realizada para o modelo anterior.

**Tabela 9** – Valores do EMA para o método MA dos valores de Retorno Absoluto

	Ordem de MA					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6860	1,6848	1,6868	1,6894	1,6888	1,6866
BA	0,9840	0,9841	0,9840	0,9864	0,9864	0,9855
BP	0,9586	0,9585	0,9591	0,9591	0,9585	0,9577
CVX	1,0554	1,0570	1,0568	1,0556	1,0556	1,0585
C	0,6746	0,6746	0,6740	0,6744	0,6742	0,6704
DELL	1,4995	1,5003	1,4983	1,4979	1,4902	1,4858
XOM	1,1246	1,1337	1,1349	1,1388	1,1381	1,1333
F	1,5173	1,5173	1,5169	1,5161	1,5158	1,5051
GE	0,6356	0,6348	0,6347	0,6352	0,6347	0,6319
GM	1,7927	1,7912	1,7909	1,7907	1,7894	1,7881
HPQ	1,0752	1,0716	1,0727	1,0709	1,0712	1,0786
IBM	0,7297	0,7316	0,7316	0,7312	0,7311	0,7264
JPM	0,7564	0,7570	0,7568	0,7565	0,7566	0,7571
MSFT	1,0911	1,0898	1,0949	1,0949	1,0943	1,0871
SNE	1,1637	1,1639	1,1638	1,1651	1,1652	1,1634
FTSE100	0,8444	0,8450	0,8474	0,8485	0,8487	0,8442
NASDAQ	0,9562	0,9551	0,9551	0,9552	0,9552	0,9626
NYSE	0,7673	0,7678	0,7681	0,7675	0,7655	0,7665
NIKKEI225	1,0193	1,0228	1,0245	1,0249	1,0247	1,0216
PSI20	0,4841	0,4872	0,4847	0,4833	0,4836	0,4827

**Tabela 10** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método MA de valores de Retorno Absoluto

	Ordem de MA				
	1	2	3	4	5
AAPL	0,04%	0,11%	-0,01%	-0,17%	-0,13%
BA	0,15%	0,14%	0,15%	-0,09%	-0,09%
BP	-0,09%	-0,08%	-0,15%	-0,15%	-0,08%
CVX	0,29%	0,14%	0,16%	0,27%	0,27%
C	-0,63%	-0,63%	-0,54%	-0,60%	-0,57%
DELL	-0,92%	-0,98%	-0,84%	-0,81%	-0,30%
XOM	0,77%	-0,04%	-0,14%	-0,49%	-0,42%
F	-0,81%	-0,81%	-0,78%	-0,73%	-0,71%
GE	-0,59%	-0,46%	-0,44%	-0,52%	-0,44%
GM	-0,26%	-0,17%	-0,16%	-0,15%	-0,07%
HPQ	0,32%	0,65%	0,55%	0,71%	0,69%
IBM	-0,45%	-0,72%	-0,72%	-0,66%	-0,65%
JPM	0,09%	0,01%	0,04%	0,08%	0,07%
MSFT	-0,37%	-0,25%	-0,72%	-0,72%	-0,66%
SNE	-0,03%	-0,04%	-0,03%	-0,15%	-0,15%
FTSE100	-0,02%	-0,09%	-0,38%	-0,51%	-0,53%
NASDAQ	0,66%	0,78%	0,78%	0,77%	0,77%
NYSE	-0,10%	-0,17%	-0,21%	-0,13%	0,13%
NIKKEI225	0,23%	-0,12%	-0,28%	-0,32%	-0,30%
PSI20	-0,29%	-0,93%	-0,41%	-0,12%	-0,19%

**Tabela 11** – Valores do EMA para o método MA dos valores de Retorno Logarítmico

	Ordem de MA					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6780	1,6772	1,6789	1,6818	1,6806	1,6785
BA	0,9839	0,9840	0,9839	0,9863	0,9864	0,9844
BP	0,9584	0,9582	0,9588	0,9588	0,9582	0,9576
CVX	1,0554	1,0571	1,0569	1,0557	1,0557	1,0581
C	0,6738	0,6736	0,6731	0,6735	0,6734	0,6701
DELL	1,4965	1,4968	1,4947	1,4950	1,4884	1,4878
XOM	1,1258	1,1348	1,1363	1,1402	1,1395	1,1335
F	1,5178	1,5178	1,5174	1,5167	1,5165	1,5065
GE	0,6343	0,6333	0,6332	0,6337	0,6332	0,6316
GM	1,7893	1,7877	1,7875	1,7874	1,7862	1,7858
HPQ	1,0730	1,0698	1,0710	1,0690	1,0695	1,0756
IBM	0,7292	0,7310	0,7309	0,7302	0,7302	0,7270
JPM	0,7565	0,7571	0,7571	0,7568	0,7571	0,7566
MSFT	1,0880	1,0868	1,0916	1,0917	1,0911	1,0861
SNE	1,1617	1,1619	1,1618	1,1630	1,1630	1,1617
FTSE100	0,8445	0,8452	0,8475	0,8487	0,8488	0,8444
NASDAQ	0,9585	0,9569	0,9571	0,9572	0,9570	0,9628
NYSE	0,7680	0,7685	0,7688	0,7682	0,7662	0,7670
NIKKEI225	1,0207	1,0240	1,0261	1,0264	1,0263	1,0229
PSI20	0,4847	0,4878	0,4851	0,4837	0,4840	0,4827

**Tabela 12** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método MA de valores de Retorno Logarítmico

	Ordem de MA				
	1	2	3	4	5
AAPL	0,03%	0,08%	-0,02%	-0,20%	-0,13%
BA	0,05%	0,04%	0,05%	-0,19%	-0,20%
BP	-0,08%	-0,06%	-0,13%	-0,13%	-0,06%
CVX	0,26%	0,09%	0,11%	0,23%	0,23%
C	-0,55%	-0,52%	-0,45%	-0,51%	-0,49%
DELL	-0,58%	-0,60%	-0,46%	-0,48%	-0,04%
XOM	0,68%	-0,11%	-0,25%	-0,59%	-0,53%
F	-0,75%	-0,75%	-0,72%	-0,68%	-0,66%
GE	-0,43%	-0,27%	-0,25%	-0,33%	-0,25%
GM	-0,20%	-0,11%	-0,10%	-0,09%	-0,02%
HPQ	0,24%	0,54%	0,43%	0,61%	0,57%
IBM	-0,30%	-0,55%	-0,54%	-0,44%	-0,44%
JPM	0,01%	-0,07%	-0,07%	-0,03%	-0,07%
MSFT	-0,17%	-0,06%	-0,51%	-0,52%	-0,46%
SNE	0,00%	-0,02%	-0,01%	-0,11%	-0,11%
FTSE100	-0,01%	-0,09%	-0,37%	-0,51%	-0,52%
NASDAQ	0,45%	0,61%	0,59%	0,58%	0,60%
NYSE	-0,13%	-0,20%	-0,23%	-0,16%	0,10%
NIKKEI225	0,22%	-0,11%	-0,31%	-0,34%	-0,33%
PSI20	-0,41%	-1,06%	-0,50%	-0,21%	-0,27%

**Tabela 13** – Valores do EMA para o método MA dos valores às Diferenças

	Ordem de MA					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,2237	1,2271	1,2278	1,2300	1,2299	1,2225
BA	0,7788	0,7783	0,7787	0,7812	0,7809	0,7810
BP	0,6476	0,6483	0,6491	0,6493	0,6495	0,6476
CVX	0,6877	0,6891	0,6893	0,6892	0,6891	0,6887
C	0,3314	0,3312	0,3311	0,3312	0,3309	0,3315
DELL	0,3553	0,3554	0,3553	0,3554	0,3553	0,3553
XOM	0,9284	0,9333	0,9345	0,9369	0,9372	0,9326
F	0,1238	0,1238	0,1238	0,1238	0,1238	0,1220
GE	0,2209	0,2204	0,2204	0,2205	0,2202	0,2203
GM	0,4948	0,4946	0,4945	0,4944	0,4941	0,4929
HPQ	0,3636	0,3630	0,3634	0,3621	0,3619	0,3648
IBM	0,6053	0,6060	0,6070	0,6059	0,6057	0,6028
JPM	0,3294	0,3300	0,3302	0,3297	0,3297	0,3297
MSFT	0,3223	0,3231	0,3254	0,3257	0,3252	0,3224
SNE	0,4716	0,4722	0,4731	0,4741	0,4741	0,4720
FTSE100	51,4737	51,5765	51,7070	51,7434	51,7662	51,5324
NASDAQ	18,1182	18,1046	18,1090	18,1165	18,1090	18,1374
NYSE	71,0287	71,1477	71,1804	71,1799	70,9275	71,1076
NIKKEI225	157,4055	158,0549	158,2829	158,3004	158,0798	157,4307
PSI20	53,0873	53,5484	53,3188	53,2288	53,4062	52,8405

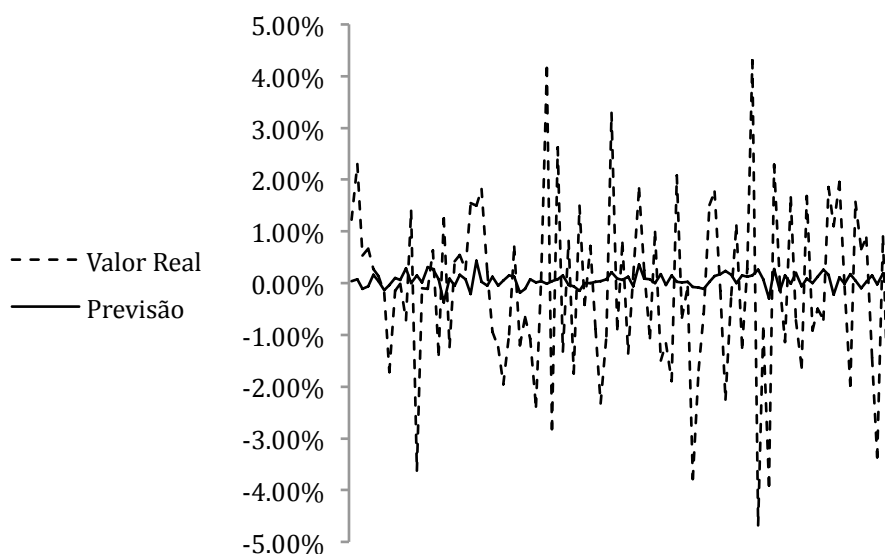
**Tabela 14** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método MA de valores às Diferenças

	Ordem de MA				
	1	2	3	4	5
AAPL	-0,10%	-0,38%	-0,43%	-0,61%	-0,61%
BA	0,28%	0,35%	0,29%	-0,03%	0,01%
BP	0,00%	-0,11%	-0,23%	-0,26%	-0,29%
CVX	0,15%	-0,06%	-0,09%	-0,07%	-0,06%
C	0,03%	0,09%	0,12%	0,09%	0,18%
DELL	0,00%	-0,03%	0,00%	-0,03%	0,00%
XOM	0,45%	-0,08%	-0,20%	-0,46%	-0,49%
F	-1,48%	-1,48%	-1,48%	-1,48%	-1,48%
GE	-0,27%	-0,05%	-0,05%	-0,09%	0,05%
GM	-0,39%	-0,34%	-0,32%	-0,30%	-0,24%
HPQ	0,33%	0,49%	0,38%	0,74%	0,79%
IBM	-0,41%	-0,53%	-0,70%	-0,51%	-0,48%
JPM	0,09%	-0,09%	-0,15%	0,00%	0,00%
MSFT	0,03%	-0,22%	-0,93%	-1,02%	-0,87%
SNE	0,08%	-0,04%	-0,23%	-0,44%	-0,44%
FTSE100	0,11%	-0,09%	-0,34%	-0,41%	-0,45%
NASDAQ	0,11%	0,18%	0,16%	0,12%	0,16%
NYSE	0,11%	-0,06%	-0,10%	-0,10%	0,25%
NIKKEI225	0,02%	-0,40%	-0,54%	-0,55%	-0,41%
PSI20	-0,47%	-1,34%	-0,91%	-0,73%	-1,07%

### 6.4.5. OBSERVAÇÕES

São aqui demonstradas as observações possíveis com os resultados obtidos através do cálculo do modelo de Médias Móveis (MA), de ordem  $q = \{1, \dots, 5\}$ .

Assim como nos resultados do modelo anterior, os valores obtidos para o EMA dos vários índices, e para as várias ordens  $q$  do modelo MA, aproximam-se bastante dos valores de  $L_F$  correspondentes. No entanto, é possível observar, na Figura 176 que os valores de previsão já apresentam um comportamento ligeiramente diferente daquele presente nas previsões do modelo AR, representados na figura 175. Nomeadamente, este modelo a previsão apresenta valores negativos e mais dispersos. Todavia, estas previsões localizam-se, tal como as previsões para o modelo AR, numa zona próxima do valor de 0.00%. Assim, é de esperar que os valores obtidos de EMA também se aproximem do valor correspondente de  $L_F$ .



**Figura 176** - Evolução dos valores reais e de previsão do índice XOM para Retorno Bruto com MA<sub>1</sub>

Os resultados demonstram também a diferença entre os tipos de transformação de dados para o cálculo de previsão. Assim como para o modelo anterior, há mais índices para os quais é possível obter uma melhoria positiva com os dados transformados às diferenças, comparativamente aos dados transformados em retornos.

Nomeadamente, para os valores transformados às diferenças, obteve-se 13 dos 20 índices com melhorias positivas, contrariamente aos apenas 9 para valores de retorno.

Comparando os resultados do EMA entre os obtidos para valores de retorno bruto e para retorno logarítmico, observa-se novamente que os resultados para o primeiro são melhores em todas as instâncias em que se nota uma melhoria relativamente ao valor de  $L_F$ .

Apresenta-se de seguida a Tabela 15, que resume os melhores resultados obtidos com o modelo MA. Os índices estudados que não estão presentes na tabela são os índices para os quais não foi possível encontrar uma melhoria relativamente ao valor de  $L_F$ .

**Tabela 15** – Resumo de resultados do modelo MA

Índice	Ordem $q$ de MA	Transformação	Melhoria
AAPL	2	Retorno Bruto	0,11%
BA	2	Diferenças	0,35%
CVX	1	Retorno Bruto	0,29%
C	5	Diferenças	0,18%
XOM	1	Retorno Bruto	0,77%
GE	5	Diferenças	0,05%
HPQ	5	Diferenças	0,79%
JPM	1	Retorno Bruto	0,09%
MSFT	1	Diferenças	0,03%
SNE	1	Diferenças	0,08%
FTSE100	1	Diferenças	0,11%
NASDAQ	2 e 3	Retorno Bruto	0,78%
NYSE	5	Diferenças	0,25%
NIKKEI225	1	Retorno Bruto	0,23%

Analisando a Tabela 15, novamente se verifica um maior número de índices de séries temporais nos quais a transformação de dados às Diferenças inferiu numa maior melhoria do que nos resultados de Retorno Bruto.

Também, e assim como nos resultados do modelo Autoregressivo, os dados transformados em Retorno Bruto têm sempre um resultado melhor que o correspondente valor para Retorno Logarítmico.

Finalmente, a ordem  $q$  do modelo MA com melhores resultados tende a ser de valor baixo, com  $q = 1$  ou  $q = 2$ .

## 6.5. PROCESSO AUTOREGRESSIVO DE MÉDIA MÓVEL (MODELO MISTO)

Em algumas aplicações, os modelos AR e MA discutidos nas secções anteriores tornam-se pesados computacionalmente pois um poderá necessitar de um modelo de alta ordem com muitos parâmetros para adequadamente descrever a estrutura dinâmica dos dados. De modo a ultrapassar esta dificuldade, os modelos Autoregressivos de Média Móvel (Autoregressive Moving Average – ARMA) são introduzidos. Um modelo ARMA combina as ideias dos modelos AR e MA numa forma compacta de modo a que o número de parâmetros usados seja mantido pequeno. Para as séries de retorno em finanças, a possibilidade de usar modelos ARMA é baixa. No entanto, o conceito dos modelos ARMA é altamente relevante em modelização volátil (séries que apresentam grande volatilidade (Tsay, 2006)).

O processo  $ARMA_{pq}$ , definido por

$$\alpha_0 Y_t + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} = \mu_0 \cdot \varepsilon_t + \mu_1 \cdot \varepsilon_{t-1} + \dots + \mu_q \cdot \varepsilon_{t-q} \quad (6.20)$$

normalizado para  $\alpha_0 = 1$ , fazendo  $\mu_0 = 1$  ou  $\sigma_\varepsilon^2 = 1$ . Também pode ser escrita em notação sumariada como  $\mu_L \cdot \varepsilon_t = \alpha_L \cdot Y_t$ . Se as raízes da equação  $\alpha_L = 0$  estiverem fora do círculo unitário, o processo pode ser representado por,

$$Y_t = \alpha_L^{-1} \cdot \mu_L \cdot \varepsilon_t \quad (6.21)$$

que é o processo de média móvel de ordem infinita.

### 6.5.1. ESTIMAÇÃO DOS PARÂMETROS NOS MODELOS MISTOS

Continuando a manter-se a hipótese anterior sobre o ruído branco, Box e Jenkins (Box & Jenkins, 1970) desenvolveram uma metodologia semelhante à descrita na secção anterior, fixando como condições iniciais o seguinte conjunto de valores,

$$\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_p = \varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{q-p-1} = 0, \quad q > p+1 \quad (6.22)$$

o que permite escrever,

$$\begin{aligned} \varepsilon_{p+q} &= X_{p+q} - \varphi_1 \cdot X_{p+q-1} - \dots - \varphi_p \cdot X_q + \theta_1 \cdot \varepsilon_{p+q-1} + \dots + \theta_{q-1} \cdot \varepsilon_{p-1}, \\ \varepsilon_1 &= X_1 \cdot \varphi_1 \cdot X_{t-1} - \dots - \varphi_p \cdot X_{t-p} + \theta_1 \cdot \varepsilon_{t-1} + \dots + \theta_q \cdot \varepsilon_{t-q}, \end{aligned} \quad p+q < t \leq N \quad (6.23)$$

Calculando a soma dos quadrados,

$$S(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q) = \sum_{i=p+1}^N \varepsilon_i^2 \quad (6.24)$$

e procedendo, por cálculo numérico, à respectiva minimização, obtêm-se os estimadores da máxima verosimilhança condicional.

O estimador do parâmetro  $\sigma_\varepsilon^2$  assume agora a forma,

$$\sigma_\varepsilon^2 = \frac{1}{N-2p-q} \cdot S(\bar{\varphi}_1, \dots, \bar{\varphi}_p, \bar{\theta}_1, \dots, \bar{\theta}_q) \quad (6.25)$$

onde o denominador resulta de se terem usado apenas  $N-p$  observações e estimado  $p+q$  parâmetros.

## 6.5.2. RESULTADOS

De seguida são apresentados (Tabelas 16 - 19) os resultados do EMA para o modelo Misto de ordem  $pq$ , com  $p = q = \{1, \dots, 5\}$ , para os dados transformados em Retorno Bruto e Retorno Logarítmico. Os resultados dos dados transformados às Diferenças não foram calculados para este modelo por problemas de convergência na execução do código de *R*.

Novamente, o cálculo do EMA foi realizado com a ajuda das funções presentes no compilador de *R*, encontrando-se o algoritmo para observação em anexo.

A disposição dos dados e a subsequente análise é igual à realizada para ambos os modelos anteriores, a exceção feita para os resultados dos dados às Diferenças. Deste modo será facilitada a realização de uma análise final comparativa dos modelos aqui apresentados.

**Tabela 16** – Valores do EMA para o método ARMA dos valores de Retorno Absoluto

	Ordem de ARMA					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6860	1,6844	1,6921	1,6871	1,6879	1,6866
BA	0,9857	0,9842	0,9825	0,9815	0,9831	0,9855
BP	0,9583	0,9586	0,9580	0,9587	0,9585	0,9577
CVX	1,0558	1,0569	1,0573	1,0573	1,0580	1,0585
C	0,6753	0,6738	0,6738	0,6758	0,6777	0,6704
DELL	1,4979	1,4977	1,4928	1,5198	1,5086	1,4858
XOM	1,1367	1,1379	1,1377	1,1386	1,1373	1,1333
F	1,5175	1,5171	1,5153	1,5180	1,5142	1,5051
GE	0,6352	0,6346	0,6345	0,6331	0,6351	0,6319
GM	1,7918	1,7903	1,7896	1,7883	1,7905	1,7881
HPQ	1,0713	1,0703	1,0708	1,0738	1,0736	1,0786
IBM	0,7311	0,7312	0,7315	0,7326	0,7315	0,7264
JPM	0,7578	0,7563	0,7590	0,7594	0,7626	0,7571
MSFT	1,0918	1,0909	1,0967	1,0958	1,1066	1,0871
SNE	1,1637	1,1661	1,1632	1,1690	1,1671	1,1634
FTSE100	0,8436	0,8486	0,8498	0,8482	0,8479	0,8442
NASDAQ	0,9554	0,9549	0,9554	0,9557	0,9535	0,9626
NYSE	0,7675	0,7682	0,7679	0,7658	0,7670	0,7665
NIKKEI225	1,0178	1,0256	1,0253	1,0256	1,0233	1,0216
PSI20	0,4870	0,4823	0,4919	0,4897	0,4910	0,4827

**Tabela 17** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método ARMA de valores de Retorno Absoluto

	Ordem de ARMA				
	1	2	3	4	5
AAPL	0,04%	0,13%	-0,33%	-0,03%	-0,08%
BA	-0,02%	0,13%	0,30%	0,41%	0,24%
BP	-0,06%	-0,09%	-0,03%	-0,10%	-0,08%
CVX	0,26%	0,15%	0,11%	0,11%	0,05%
C	-0,73%	-0,51%	-0,51%	-0,81%	-1,09%
DELL	-0,81%	-0,80%	-0,47%	-2,29%	-1,53%
XOM	-0,30%	-0,41%	-0,39%	-0,47%	-0,35%
F	-0,82%	-0,80%	-0,68%	-0,86%	-0,60%
GE	-0,52%	-0,43%	-0,41%	-0,19%	-0,51%
GM	-0,21%	-0,12%	-0,08%	-0,01%	-0,13%
HPQ	0,68%	0,77%	0,72%	0,45%	0,46%
IBM	-0,65%	-0,66%	-0,70%	-0,85%	-0,70%
JPM	-0,09%	0,11%	-0,25%	-0,30%	-0,73%
MSFT	-0,43%	-0,35%	-0,88%	-0,80%	-1,79%
SNE	-0,03%	-0,23%	0,02%	-0,48%	-0,32%
FTSE100	0,07%	-0,52%	-0,66%	-0,47%	-0,44%
NASDAQ	0,75%	0,80%	0,75%	0,72%	0,95%
NYSE	-0,13%	-0,22%	-0,18%	0,09%	-0,07%
NIKKEI225	0,37%	-0,39%	-0,36%	-0,39%	-0,17%
PSI20	-0,89%	0,08%	-1,91%	-1,45%	-1,72%

**Tabela 18** – Valores do EMA para o método ARMA dos valores de Retorno Logarítmico

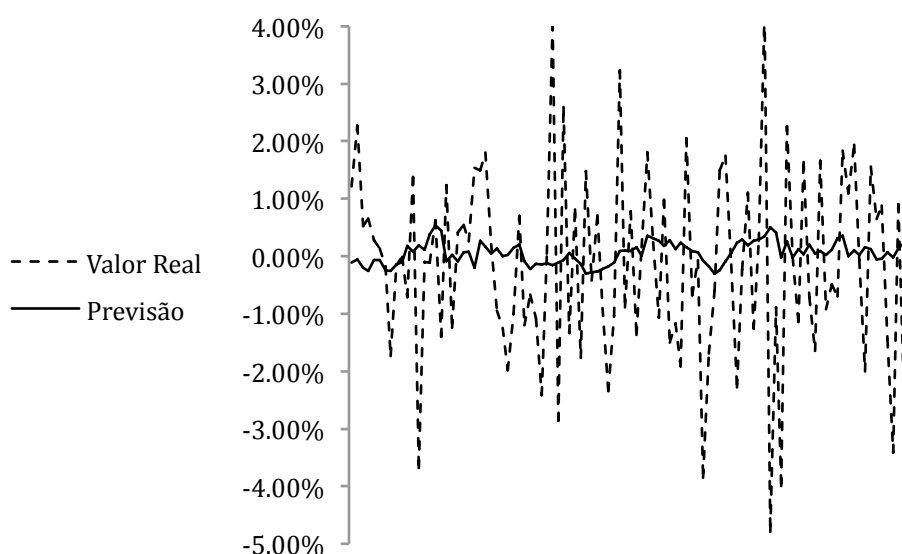
	Ordem de ARMA					L <sub>F</sub>
	1	2	3	4	5	
AAPL	1,6780	1,6768	1,6847	1,6796	1,6818	1,6785
BA	0,9856	0,9841	0,9821	0,9821	0,9811	0,9844
BP	0,9580	0,9582	0,9578	0,9587	0,9586	0,9576
CVX	1,0559	1,0570	1,0570	1,0575	1,0543	1,0581
C	0,6744	0,6729	0,6730	0,6777	0,6774	0,6701
DELL	1,4941	1,4929	1,4902	1,5111	1,5110	1,4878
XOM	1,1381	1,1393	1,1390	1,1396	1,1407	1,1335
F	1,5180	1,5174	1,5173	1,5169	1,5145	1,5065
GE	0,6337	0,6330	0,6329	0,6318	0,6357	0,6316
GM	1,7886	1,7869	1,7876	1,7868	1,7872	1,7858
HPQ	1,0698	1,0685	1,0715	1,0755	1,0729	1,0756
IBM	0,7305	0,7304	0,7308	0,7319	0,7284	0,7270
JPM	0,7661	0,7566	0,7572	0,7588	0,7632	0,7566
MSFT	1,0890	1,0880	1,0932	1,0926	1,1042	1,0861
SNE	1,1617	1,1644	1,1614	1,1678	1,1635	1,1617
FTSE100	0,8439	0,8496	0,8508	0,8483	0,8478	0,8444
NASDAQ	0,9576	0,9570	0,9575	0,9579	0,9560	0,9628
NYSE	0,7683	0,7694	0,7673	0,7670	0,7681	0,7670
NIKKEI225	1,0192	1,0271	1,0269	1,0280	1,0269	1,0229
PSI20	0,4875	0,4811	0,4926	0,4880	0,4895	0,4827

**Tabela 19** – Valores comparativos de melhoria relativamente ao L<sub>F</sub> do método ARMA de valores de Retorno Logarítmico

	Ordem de ARMA				
	1	2	3	4	5
AAPL	0,03%	0,10%	-0,37%	-0,07%	-0,20%
BA	-0,12%	0,03%	0,23%	0,23%	0,34%
BP	-0,04%	-0,06%	-0,02%	-0,11%	-0,10%
CVX	0,21%	0,10%	0,10%	0,06%	0,36%
C	-0,64%	-0,42%	-0,43%	-1,13%	-1,09%
DELL	-0,42%	-0,34%	-0,16%	-1,57%	-1,56%
XOM	-0,41%	-0,51%	-0,49%	-0,54%	-0,64%
F	-0,76%	-0,72%	-0,72%	-0,69%	-0,53%
GE	-0,33%	-0,22%	-0,21%	-0,03%	-0,65%
GM	-0,16%	-0,06%	-0,10%	-0,06%	-0,08%
HPQ	0,54%	0,66%	0,38%	0,01%	0,25%
IBM	-0,48%	-0,47%	-0,52%	-0,67%	-0,19%
JPM	-1,26%	0,00%	-0,08%	-0,29%	-0,87%
MSFT	-0,27%	-0,17%	-0,65%	-0,60%	-1,67%
SNE	0,00%	-0,23%	0,03%	-0,53%	-0,15%
FTSE100	0,06%	-0,62%	-0,76%	-0,46%	-0,40%
NASDAQ	0,54%	0,60%	0,55%	0,51%	0,71%
NYSE	-0,17%	-0,31%	-0,04%	0,00%	-0,14%
NIKKEI225	0,36%	-0,41%	-0,39%	-0,50%	-0,39%
PSI20	-0,99%	0,33%	-2,05%	-1,10%	-1,41%

### 6.5.3. OBSERVAÇÕES

Comparativamente com os resultados obtidos em ambos os modelos anteriores, verifica-se que aqueles aqui obtidos com o modelo ARMA não diferem muito dos obtidos nesses modelos anteriores. Novamente a razão apontada é a baixa variação dos valores da previsão obtida pelo método, verificável na Figura 177. Tal como para o modelo de Média Móvel, as previsões mantêm-se muito próximas do valor de 0.00% para o Retorno Bruto.



**Figura 177** – Evolução dos valores reais e de previsão do índice XOM para Retorno Bruto com ARMA<sub>1</sub>

Verificam-se novamente melhores resultados para os cálculos do EMA para os dados transformados em Retorno Bruto do que para os resultados obtidos do cálculo que utiliza os dados transformados em Retorno Logarítmico.

Na tabela 20, é realizada a conclusão dos resultados obtidos com o modelo ARMA aqui apresentado. Novamente, os índices não presentes são aqueles para os quais não foram encontradas melhorias relativamente ao valor do Limite de Futilidade,  $L_F$ .

**Tabela 20** – Resumo de resultados do modelo MA

Índice	Ordem $pq$ de ARMA	Transformação	Melhoria
AAPL	2	Retorno Bruto	0,13%
BA	4	Retorno Bruto	0,41%
CVX	5	Retorno Logarítmico	0,36%
HPQ	2	Retorno Bruto	0,77%
JPM	2	Retorno Bruto	0,11%
SNE	3	Retorno Logarítmico	0,03%
FTSE100	1	Retorno Bruto	0,07%
NASDAQ	5	Retorno Bruto	0,95%
NYSE	4	Retorno Bruto	0,09%
NIKKEI225	1	Retorno Bruto	0,37%
PSI20	2	Retorno Logarítmico	0,33%

Como indicado anteriormente, para o cálculo dos EMA para o modelo Misto, apenas foram utilizados os dados transformados em ambas as vertentes dos Retornos.

Contrariamente a ambos os modelos anteriores, no modelo Misto obteve-se melhorias superiores com os dados transformados em Retorno Logarítmico, para 3 dos 11 índices estudados com uma melhoria relativamente ao valor do Limite de Futilidade correspondente.

Com o objectivo de fazer uma comparação entre os três modelos clássicos apresentados anteriormente, a próxima secção é dedicada à exposição e comparação dos melhores resultados.

## 6.6. COMPARAÇÃO ENTRE OS MODELOS CLÁSSICOS

Nesta secção é apresentada na Tabela 21 uma comparação final entre o melhor resultado obtido a partir da previsão realizada pelos três modelos estudados, para cada uma das séries temporais em estudo.

Pretende-se encontrar o modelo clássico mais adequado ao tipo de séries temporais aqui em estudo, comparando o valor do cociente de melhoria  $M$  para cada um dos índices. Os valores mais altos de  $M$  estão indicados a cinzento na tabela, de modo a facilitar a análise dos resultados.

Nos índices para os quais um modelo não retornou uma melhoria positiva, é indicado com o símbolo (-).

**Tabela 21** – Apresentação dos melhores resultados dos modelos clássicos

Índice	AR	Transformação	MA	Transformação	ARMA	Transformação
AAPL	0,11%	Ret. Bruto	0,11%	Ret. Bruto	0,13%	Ret. Bruto
BA	0,35%	Diferenças	0,35%	Diferenças	0,34%	Ret. Logarítmico
BP	-	-	-	-	-	-
CVX	0,33%	Ret. Bruto	0,29%	Ret. Bruto	0,36%	Ret. Logarítmico
C	0,18%	Diferenças	0,18%	Diferenças	-	-
DELL	-	-	-	-	-	-
XOM	0,78%	Ret. Bruto	0,77%	Ret. Bruto	-	-
F	-	-	-	-	-	-
GE	-	-	0,05%	Diferenças	-	-
GM	-	-	-	-	-	-
HPQ	0,69%	Diferenças	0,79%	Diferenças	0,77%	Ret. Bruto
IBM	-	-	-	-	-	-
JPM	0,09%	Ret. Bruto	0,09%	Ret. Bruto	0,11%	Ret. Bruto
MSFT	0,03%	Diferenças	0,03%	Diferenças	-	-
SNE	0,11%	Diferenças	0,11%	Diferenças	0,03%	Ret. Logarítmico
FTSE100	0,11%	Diferenças	0,11%	Diferenças	0,07%	Ret. Bruto
NASDAQ	0,78%	Ret. Bruto	0,78%	Ret. Bruto	0,95%	Ret. Bruto
NYSE	0,22%	Diferenças	0,25%	Diferenças	-	-
NIKKEI225	0,21%	Ret. Bruto	0,23%	Ret. Bruto	0,37%	Ret. Bruto
PSI20	-	-	-	-	0,33%	Ret. Logarítmico

Analisando a tabela 21, uma conclusão imediata é de que não há um modelo claramente superior aos outros, visto os melhores resultados estarem distribuídos pelos três modelos clássicos, com o modelo Autoregressivo a ter 6 resultados superiores, o modelo de Médias Móveis com 7 resultados superiores, e o modelo Misto com 6 resultados superiores.

Na secção 4.1, foi indicado que, através da análise da Função de Autocorrelação e da Função de Autocorrelação Parcial, seria o modelo de Média Móvel o mais indicado na utilização para previsões nas séries temporais estudadas. No entanto, apesar de este modelo apresentar melhores percentagens de melhoria para sete índices, em cinco desses índices obteve o mesmo valor obtido no modelo Autoregressivo.

Como tal, os resultados indicam possivelmente que a análise gráfica das FAC e FACP das séries temporais de cada índice não é suficiente para saber qual o modelo indicado para previsão de dada série temporal.

No entanto, deve-se tomar em conta que para o modelo Misto não foi possível realizar os testes para os dados transformados às Diferenças, factor que poderá ter pesado no facto de o modelo de Média Móvel ter obtido melhores resultados que o modelo Misto.

# 7. PROPOSTA DE DUAS METODOLOGIAS DE ANÁLISE E PREVISÃO

Os métodos clássicos estudados apresentam, como limitação, a necessidade de a série temporal analisada ser considerada na sua totalidade, tomando em consideração momentos que não são relevantes para a previsão desejada. Na tentativa de ultrapassar esta questão e conseguir uma previsão mais acertada, foram desenvolvidos dois métodos baseados na repetição de estados que, teoricamente, se adaptam melhor à série temporal.

Este capítulo apresenta as duas metodologias propostas como alternativa aos modelos clássicos discutidos no capítulo 6. Estas metodologias têm como nome Modelo de Classes e Modelo Epson. A principal diferença entre as metodologias propostas e as clássicas discutidas anteriormente reside na quantidade de dados utilizados por cada uma das metodologias. Enquanto que os modelos clássicos utilizam todos os dados disponíveis, estes modelos apenas utilizam os dados dos momentos semelhantes ao momento que se pretende prever.

## 7.1. APRESENTAÇÃO GERAL DOS MODELOS PROPOSTOS

Considere-se  $X_t$  como a série de valores em estudo com  $t$  a representar o último valor observado da série. Ambos os modelos aqui apresentados, desenvolvidos nas secções 7.5 e 7.6, baseiam-se na procura de momentos – sequências de  $l$  valores – iguais ao momento em estudo, do qual se pretende tentar prever o valor seguinte,  $X_{t+l}$ . O conhecimento de valores passados que ocorreram após uma sequência semelhante àquela em estudo, permite obter uma possível previsão, realizando uma média ponderada dos resultados obtidos.

No entanto, esta procura torna-se muito difícil quando se tenta encontrar sequências iguais em séries temporais cujos valores pertencem a  $\mathfrak{R}$ , fazendo com que as buscas retornem poucos ou nenhum valores. De modo a ultrapassar este problema, são realizadas transformações nos dados das amostras. Estas transformações, que são o que distingue ambos os modelos propostos, serão descritas em detalhes nas secções 7.5 e 7.6.

Com a transformação dos valores das séries, torna-se mais fácil para o modelo encontrar momentos iguais ou semelhantes ao longo desta.

A fase de previsão do modelo, explicada de seguida na secção 7.2, consiste em percorrer toda a amostra da série disponível e comparar a última sequência da série com a sequência no momento actual da busca. Se todos os valores forem iguais, então o valor seguinte a essa sequência é guardado. Após toda a série ser percorrida, é realizada uma média dos valores obtidos.

Na secção 7.3 é explicado o método de optimização e validação dos modelos.

## 7.2. PREVISÃO

Considere-se a série de valores transformados,  $Y_t$ , de tamanho  $N$ , com e com o momento  $t_N$  a ser o último valor observado. Pretende-se saber o valor de  $Y_{tN+1}$ , pelo que para tal, serão guardados em uma sequência,  $s_a$ , os últimos  $l$  valores conhecidos de  $Y_t$ ,

$$s_a = \{Y_{t_a-l-1}, \dots, Y_{t_a-1}, Y_{t_a}\} \quad (7.1)$$

É percorrida toda a série  $Y_t$ , com  $t = \{1, 2, \dots, N\}$  sendo guardado em cada  $t$ , os  $l$  valores na sequência  $s_b$ ,

$$s_b = \{Y_t, Y_{t+1}, \dots, Y_{t+l-1}\} \quad (7.2)$$

e, a cada  $t$ , se a expressão (7.3) for verdadeira,

$$s_a = s_b \quad (7.3)$$

o valor de  $Y_{t+l}$  é guardado como possível valor de previsão. Também é guardado, no caso do modelo Epsom, o número de vezes que a expressão (7.3) é verdadeira,  $nc$ .

Quando  $t = N - l$ , é feita uma média destes valores, de modo a ser extraída uma previsão,  $P$ .

## 7.3. ESTIMAÇÃO

A implementação dos métodos em estudo requer a determinação do valor dos parâmetros a serem usados na previsão dos valores da série temporal. Estes parâmetros são o Tamanho da Sequência ( $tam_{seq}$ ), o Número de Classes ( $n_{classes}$ ) e o Epsilon ( $E$ ).

### 7.3.1. TAMANHO DA SEQUÊNCIA

Utilizado em ambos os modelos,  $tam_{seq}$  indica o número de valores do momento em estudo para determinado  $t$ . Pode-se considerar que um  $tam_{seq}$  muito pequeno ( $< 3$ ) não fornecerá uma representação válida do momento, e um  $tam_{seq}$  muito grande poderá incorrer numa situação em que seja muito difícil, ou até mesmo impossível, encontrar um momento semelhante. Através de experiências iniciais durante o desenvolvimento computacional dos modelos, verificou-se que seria mais indicado usar um  $tam_{seq} = \{1; 1 = 3,4,5,6,7\}$  para o modelo Epsilon e  $tam_{seq} = \{1; 1 = 3,4,5\}$  no modelo de Classes.

### 7.3.2. NÚMERO DE CLASSES

Utilizado no modelo de Classes,  $n_{classes}$  indica em quantos grupos (classes) irá ser dividido o espectro de valores presentes numa dada série temporal  $X_t$ . Cada classe é caracterizada por um determinado intervalo  $\Delta$ , dependente do valor máximo e mínimo de  $X_t$  e de  $n_{classes}$ . O valor atribuído a cada classe será o seu valor médio. Durante o desenvolvimento computacional, verificou-se que seria indicado começar com  $n_{classes} = \{10,15,20,25,30,35,40,45,50\}$ .

### 7.3.3. EPSON

Utilizado no modelo Epson,  $E$  representa o intervalo em volta do valor de  $Y_t$  dentro do qual todos os valores encontrados serão considerados equivalentes a  $Y_t$ . Os valores experimentais iniciais usados serão no intervalo  $[0, 1]$ , com incrementos de  $E = 0,1$ .

Pode-se logicamente deduzir que valores muito grandes de  $E$  poderão conduzir a um cenário de grande correspondência entre  $s_a$  e  $s_b$ , mas com valores que na realidade não são semelhantes, produzindo resultados de previsão não desejáveis. Por outro lado, a elevada resolução de  $E$  ( $E = 0$ ), resulta num cenário em que poucas ou nenhuma correspondências entre  $s_a$  e  $s_b$  ocorrem.

## 7.4. OPTIMIZAÇÃO DE PARÂMETROS

A técnica utilizada na otimização dos parâmetros dos métodos em estudo incide na tentativa de previsão para um período conhecido da série temporal, tendo comprimento de  $\lambda$  amostras, utilizando todas as combinações dos parâmetros possíveis.

Para este período, o modelo irá realizar  $\lambda$  iterações, produzindo  $\lambda$  valores de previsão para cada combinação possível dos parâmetros,  $X_{P_t}$ .

Contrariamente aos métodos tradicionais, que conseguem realizar uma previsão válida para todas as iterações requeridas para o intervalo de teste, os modelos propostos por vezes não são capazes de encontrar uma previsão que corresponda à combinação de parâmetros escolhidos para essa iteração. Como tal, verificou-se a necessidade de escolher uma combinação de parâmetros que, para além de possuírem um Erro Médio Absoluto (EMA) baixo, também tivessem uma Percentagem de Perda de Informação (PI) baixa.

Após a realização de alguns testes, considerou-se que poderia ser válida a heurística que indica que a combinação de parâmetros escolhida será aquela que apresentar o EMA mais baixo no grupo de combinações de parâmetros que possuem uma PI inferior a 10%.

#### 7.4.1. ERRO MÉDIO ABSOLUTO

O Erro Médio Absoluto (EMA) foi apresentado na secção 6.1. Recorda-se que representa a média da diferença entre o valor real da série temporal e o valor de previsão calculado pelo modelo utilizado. Assim, o EMA é calculado através da fórmula,

$$EMA = \frac{\sum_{i=0}^{\lambda} |Y_{C_i} - Y_{P_i}|}{\lambda} \quad (6.1)$$

#### 7.4.2. PERCENTAGEM DE PERDA DE INFORMAÇÃO

Durante o desenvolvimento e teste dos modelos, foi verificado que, para certas combinações de parâmetros dos modelos propostos, estes não retornavam qualquer valor de previsão calculado. Este fenômeno ocorre devido às características numéricas dos parâmetros usados, que tornavam praticamente impossível o encontrar de ocorrências iguais.

Nestes casos, o modelo retorna um valor correspondente ao Limite de Futilidade. Este método, como explicado anteriormente, assume que a melhor previsão para o momento  $t + 1$ , no caso de ausência de informação adicional, será o valor do momento  $t$ . No caso da utilização de valores transformados, o valor retornado será zero.

Considera-se que, quanto maior for o número de iterações nas quais o método não consegue fazer qualquer previsão, maior é a Percentagem de Perda de Informação (PI). Resultados de combinações de parâmetros que obtenham perdas de informação elevadas não devem ser tomados em consideração, pois o valor de previsão dado por esta iteração é obtido, não por efeito do método, mas por falta de informação. Estes resultados tendem a aproximar-se do valor médio de Limite de Futilidade.

## 7.5. MODELO DE CLASSES

Como referido anteriormente, a probabilidade de a condição,

$$s_a = s_b \quad (7.6)$$

em que  $s_a = \{Y_{t_a-l-1}, \dots, Y_{t_a-1}, Y_{t_a}\}$  e  $s_b = \{Y_t, Y_{t+1}, \dots, Y_{t+l-1}\}$ , ser verdadeira é próxima de zero, para valores não adaptados. Assim há a necessidade de fazer uma transformação nos dados da série temporal. O modelo das Classes aplica essa transformação antes de realizar o cálculo da previsão.

Considere-se, novamente,  $Y_t$  como a série temporal de valores transformados de  $X_t$ . Os valores de  $Y_t$  apresentam-se como retornos (absolutos e logarítmicos) e às diferenças. Considere-se também  $Z_t$  como a série dos valores modelizados a partir de  $Y_t$ , ou seja, valores que foram obtidos através da aplicação das transformações usadas em cada modelo.

De modo a realizar a transformação  $Y_t \Rightarrow Z_t$ , é necessário o cálculo de três parâmetros presentes na série  $Y_t$ . Estes parâmetros são  $n_{classes}$  (já apresentado na secção 7.3.2),  $max_{Y_t}$  e  $min_{Y_t}$ .

Os parâmetros  $max_{Y_t}$  e  $min_{Y_t}$  representam, respectivamente, os valores máximos e mínimos da série  $Y_t$ . Estes parâmetros são utilizados na determinação do tamanho do intervalo de classes,  $\Delta$ , que é calculado como

$$\Delta = \frac{\max_{Y_t} - \min_{Y_t}}{n_{classes}} \quad (7.7)$$

Após a determinação do valor de  $\Delta$ , é procurada a classe  $n$  correspondente a cada valor de  $Y_t$ , sendo

$$n = \{1, \dots, n_{classes}\} \quad (7.8)$$

E o seu valor determinado através da dedução da equação para  $\Delta$ ,

$$\min_{Y_t} + (n + 1) \cdot \Delta \leq Y_t < n \cdot \Delta \quad (7.9)$$

### 7.5.1. RESULTADOS

De seguida são apresentados os resultados obtidos para o modelo de Classes. Com o objectivo de se compreender melhor o modo como os valores foram obtidos, são expostos todos os resultados obtidos para a série temporal de Retorno Bruto do índice BA, para os vários número de classes, e vários tamanhos da sequência. O valor do Limite de Futilidade deste índice para dados em Retorno Bruto é  $L_F = 0,9855$ .

Como referido anteriormente, nos cálculos efectuados foram feitos testes para todas as combinações entre o tamanho da sequência,  $tam_{seq} = \{3, 4, 5\}$  e número de classes,  $n_{classes} = \{10, 15, 20, 25, 30, 35, 40, 45, 50\}$ .

**Tabela 22** – Valores do EMA do modelo de Classes para o índice BA com dados de Retorno Bruto

		$n_{classes}$								
		10	15	20	25	30	35	40	45	50
$tam_{seq}$	3	1,0087	1,0104	1,0570	1,0866	1,0745	1,1177	1,1107	1,1265	1,2060
	4	1,0228	1,0749	1,1399	1,1429	1,1215	1,1236	1,0565	1,0613	1,0181
	5	1,0460	1,1587	1,1467	1,0817	1,0594	1,0318	1,0143	1,0062	0,9874

**Tabela 23** – Valores da PI do modelo de Classes para o índice BA com dados de Retorno Bruto

		$n_{classes}$								
		10	15	20	25	30	35	40	45	50
$tam_{seq}$	3	1,8%	6,4%	8,6%	11,8%	16,4%	21,8%	27,8%	35,0%	38,6%
	4	4,2%	11,0%	19,4%	29,4%	44,4%	60,4%	67,4%	79,0%	86,4%
	5	8,8%	23,2%	42,6%	62,8%	78,2%	88,8%	91,6%	95,0%	97,4%

A importância dada aos resultados da Perda de Informação prende-se com o facto de que, ao tomar-se apenas em consideração os valores do EMA obtidos com o

método, incorre-se na possibilidade de escolher uma combinação de parâmetros que não traduz a capacidade de previsão do modelo. Como exemplo, considere-se os valores obtidos no índice BA, apresentados nas Tabelas 22 e 23. Analisando apenas os valores do EMA na Tabela 22, poder-se-ia deduzir que o melhor resultado seria aquele obtido com a combinação de  $n_{classes} = 50$  e  $tam_{seq} = 5$ , por este ter o Erro Médio Absoluto mais baixo, com  $EMA = 0,9874$ . No entanto, adicionando a informação demonstrada na Tabela 24, verifica-se que esta combinação resulta de uma alta Perda de Informação, no valor de 97,4%. Significa isto que, para 97,4% das iterações que procuram uma previsão e não retornaram nenhum valor, o valor retornado de previsão dada é 0%, seguindo a ideologia utilizada no cálculo do Limite de Futilidade.

Assim, é fácil deduzir-se que para uma combinação de parâmetros com uma perda de informação muito alta, o valor de EMA obtido irá aproximar-se do Limite de Futilidade, por ser calculado aproximadamente com os mesmos dados. Tal valor não é de interesse para o objectivo do modelo.

Deste modo, com o exemplo referido, a combinação escolhida para a série de dados é  $n_{classes} = 10$  e  $tam_{seq} = 3$ , com um resultado de 1,0087. Isto resulta numa quociente de melhoria ( $M$ ) de:

$$M = \frac{L_F - EMA}{EMA} = \frac{0,9855 - 1,0087}{0,9855} = -2,35\% \quad (7.10)$$

De seguida, são apresentados nas Tabelas 24 – 43 os resultados obtidos para o modelo das Classes, com a indicação dos parâmetros  $n_{classes}$  e  $tam_{seq}$  que tiveram o melhor resultado tendo em consideração a heurística em uso. Também é indicado o valor de  $L_F$  correspondente, assim como o quociente da melhoria ( $M$ ) realizada pelo método, de modo a facilitar a comparação entre modelos, e o valor da percentagem de Perda de Informação ( $PI$ ).

São apresentados os resultados para os três tipos de transformação de dados – Retorno Bruto, Retorno Logarítmico e às Diferenças – para ser possível uma comparação dos resultados obtidos por cada método. O tipo de transformação que tiver o maior quociente de melhoria  $M$  é indicado a cinzento nas tabelas.

**Tabela 24** – Resultados do Modelo de Classes para o índice AAPL

AAPL	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,6866	3	15	1,7218	1,0%	-2,09%
Ret. Logarítmico	1,6785	3	35	1,7374	5,6%	-3,51%
Diferenças	1,2225	3	10	1,3065	12,0%	-6,87%

**Tabela 25** – Resultados do Modelo de Classes para o índice BA

BA	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,9855	3	10	1,0087	1,8%	-2,35%
Ret. Logarítmico	0,9844	3	10	0,9929	2,8%	-0,86%
Diferenças	0,7810	3	10	0,8093	10,4%	-3,62%

**Tabela 26** – Resultados do Modelo de Classes para o índice BP

BP	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,9577	3	10	1,0120	2,2%	-5,67%
Ret. Logarítmico	0,9576	3	10	1,0024	2,4%	-4,68%
Diferenças	0,6476	3	10	0,6944	11,8%	-7,23%

**Tabela 27** – Resultados do Modelo de Classes para o índice CVX

CVX	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,0585	4	10	1,0957	1,6%	-3,51%
Ret. Logarítmico	1,0581	3	15	1,1032	3,8%	-4,26%
Diferenças	0,6887	3	10	0,7848	20,8%	-13,95%

**Tabela 28** – Resultados do Modelo de Classes para o índice C

C	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,6704	3	20	0,6875	2,8%	-2,55%
Ret. Logarítmico	0,6701	3	20	0,6937	2,4%	-3,52%
Diferenças	0,3315	3	15	0,3411	6,8%	-2,90%

**Tabela 29** – Resultados do Modelo de Classes para o índice DELL

DELL	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,4858	3	10	1,5352	3,4%	-3,32%
Ret. Logarítmico	1,4878	3	15	1,5416	6,8%	-3,62%
Diferenças	0,3553	3	10	0,3734	2,2%	-5,09%

**Tabela 30** – Resultados do Modelo de Classes para o índice XOM

XOM	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,1333	3	10	1,1934	6,8%	-5,30%
Ret. Logarítmico	1,1335	3	10	1,2013	8,4%	-5,98%
Diferenças	0,9326	3	10	1,0280	24,8%	-10,23%

**Tabela 31** – Resultados do Modelo de Classes para o índice F

F	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,5051	3	15	1,5367	6,0%	-2,10%
Ret. Logarítmico	1,5065	3	15	1,5646	3,8%	-3,86%
Diferenças	0,1220	3	35	0,1270	1,6%	-4,10%

**Tabela 32** – Resultados do Modelo de Classes para o índice GE

GE	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,6319	3	20	0,6464	3,4%	-2,29%
Ret. Logarítmico	0,6316	3	15	0,6522	3,0%	-3,26%
Diferenças	0,2203	4	10	0,2236	4,2%	-1,50%

**Tabela 33** – Resultados do Modelo de Classes para o índice GM

GM	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,7881	3	10	1,9414	8,2%	-8,57%
Ret. Logarítmico	1,7858	3	10	1,8915	7,4%	-5,92%
Diferenças	0,4929	3	10	0,5077	1,6%	-3,00%

**Tabela 34** – Resultados do Modelo de Classes para o índice HPQ

HPQ	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,0786	3	10	1,1078	2,6%	-2,71%
Ret. Logarítmico	1,0756	3	10	1,1212	2,8%	-4,24%
Diferenças	0,3648	4	15	0,3665	7,8%	-0,47%

**Tabela 35** – Resultados do Modelo de Classes para o índice IBM

IBM	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,7264	3	20	0,7544	2,6%	-3,85%
Ret. Logarítmico	0,7270	3	20	0,7519	3,6%	-3,43%
Diferenças	0,6028	3	15	0,6124	1,0%	-1,59%

**Tabela 36** – Resultados do Modelo de Classes para o índice JPM

JPM	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	0,7571	3	15	0,7873	3,2%	-3,99%
Ret. Logarítmico	0,7566	3	25	0,7918	4,8%	-4,65%
Diferenças	0,3297	3	10	0,3441	3,6%	-4,37%

**Tabela 37** – Resultados do Modelo de Classes para o índice MSFT

MSFT	L <sub>F</sub>	tam <sub>seq</sub>	n <sub>classes</sub>	EMA	PI	M
Ret. Bruto	1,0871	3	15	1,1406	7,2%	-4,92%
Ret. Logarítmico	1,0861	3	10	1,1333	3,4%	-4,35%
Diferenças	0,3224	4	10	0,3305	7,4%	-2,51%

**Tabela 38** – Resultados do Modelo de Classes para o índice SNE

SNE	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	1,1634	3	10	1,2204	3,0%	-4,90%
Ret. Logarítmico	1,1617	3	15	1,2239	5,6%	-5,35%
Diferenças	0,4720	3	25	0,4959	4,8%	-5,06%

**Tabela 39** – Resultados do Modelo de Classes para o índice FTSE

FTSE	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	0,8442	3	10	0,9135	16,2%	-8,21%
Ret. Logarítmico	0,8444	3	10	0,9134	13,8%	-8,17%
Diferenças	51,5324	3	10	55,2152	21,4%	-7,15%

**Tabela 40** – Resultados do Modelo de Classes para o índice NASDAQ

NASDAQ	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	0,9626	3	30	0,9657	5,2%	-0,32%
Ret. Logarítmico	0,9628	3	25	0,9940	2,8%	-3,24%
Diferenças	18,1374	4	10	18,5600	4,6%	-2,33%

**Tabela 41** – Resultados do Modelo de Classes para o índice NYSE

NYSE	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	0,7665	3	10	0,8296	12,4%	-8,23%
Ret. Logarítmico	0,7670	3	10	0,8073	13,0%	-5,25%
Diferenças	71,1076	3	10	76,4512	27,8%	-7,51%

**Tabela 42** – Resultados do Modelo de Classes para o índice NASDAQ

NIKKEI225	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	1,0216	3	10	1,0704	4,8%	-4,78%
Ret. Logarítmico	1,0229	3	15	1,0837	7,4%	-5,94%
Diferenças	157,4307	3	15	163,4277	2,8%	-3,81%

**Tabela 43** – Resultados do Modelo de Classes para o índice PSI20

PSI20	$L_F$	$tam_{seq}$	$n_{classes}$	EMA	PI	M
Ret. Bruto	0,4827	3	10	0,5231	11,2%	-8,37%
Ret. Logarítmico	0,4827	3	10	0,5181	10,8%	-7,33%
Diferenças	52,8405	3	10	54,6020	9,4%	-3,33%

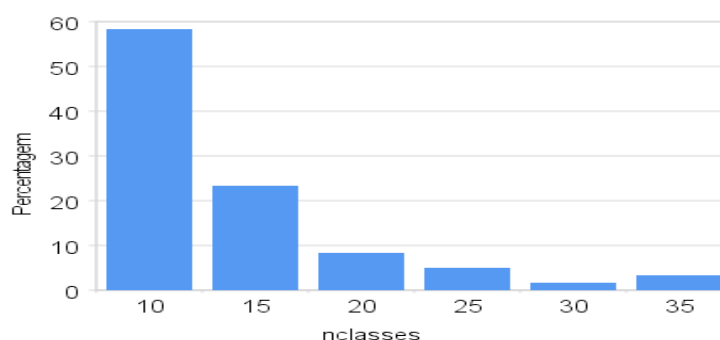
### 7.5.2. OBSERVAÇÕES

Analisando as tabelas 24 – 43 dos resultados do modelo das Classes, nota-se inicialmente que nenhum dos resultados de EMA obtidos pelos cálculos deste modelo obteve uma melhoria positiva relativamente ao  $L_F$  correspondente. Tal demonstra a inaptidão dos parâmetros iniciais deste modelo para a previsão das séries temporais em estudo.

Tome-se também em consideração nas tabelas 24, 25, 26, 27, 30, 39, 41 e 43, que para alguns tipos de transformação de dados, principalmente para os transformados às Diferenças, que o valor da Perda de Informação (PI) é superior ao limite estipulado de 10%. Nestes casos, foi escolhido como valor de EMA o menor valor obtido.

Relativamente ao parâmetro  $tam_{seq}$ , verifica-se que na maioria dos resultados – 56 num total de 60 – esta variável tem o valor  $tam_{seq} = 3$ . Apenas nos restantes 4 resultados é que se obtêm um valor  $tam_{seq} = 4$ , e nenhum resultado para  $tam_{seq} = 5$ . Com estes resultados se pode concluir que para uma sequência maior, é mais difícil para o modelo proposto encontrar uma sequência semelhante, por ser aumentado o número de valores para os quais deve encontrar um semelhante na mesma ordem. Para iterações futuras, não será necessário realizar testes com  $tam_{seq} = 5$  ou superior.

Observando a distribuição dos valores do parâmetro  $n_{classes}$  na Figura 178, verifica-se que a maior percentagem de resultados foi obtida para  $n_{classes} = 10$ , seguido de  $n_{classes} = 15$ . Para  $n_{classes}$  acima destes valores a percentagem de ocorrência decresce, até ao ponto de não haver ocorrências para valores acima de  $n_{classes} = 35$ . Este fenómeno pode ser explicado pelo facto de que à medida que aumenta o valor de  $n_{classes}$ , o correspondente valor de  $\Delta$  diminui, tornando-se mais difícil para o modelo conseguir encontrar valores semelhantes ao pretendido. Derivado disto, a percentagem de Perda de Informação aumenta, tornando a combinação de parâmetros não aceitável perante a heurística escolhida. Para iterações futuras, não será necessário realizar testes com  $n_{classes} = 40$  ou superior.



**Figura 178** – Distribuição dos valores do parâmetro  $n_{classes}$  para o modelo das Classes

Uma primeira possibilidade na busca de melhores resultados para este modelo seria a procura de uma melhor otimização dos valores dos parâmetros utilizados. Ou seja, tomando a combinação de parâmetros que obteve o melhor resultado para um dado índice, pode-se realizar os cálculos para valores vizinhos desses parâmetros. Por exemplo, observando a tabela 40, do índice NASDAQ, verifica-se que a combinação de parâmetros que retorna o melhor resultado é a de  $tam_{seq} = 3$  e  $n_{classes} = 30$ , para os dados com transformação em Retorno Bruto. Pode-se fazer a otimização realizando os cálculos para  $n_{classes} = \{29, 31\}$ , ou seja, para os valores inteiros na vizinhança do valor de  $n_{classes}$  original. Na eventualidade de se obter um resultado com um  $M$  maior do que o obtido pela combinação original de parâmetros, poderia repetir-se o cálculo para o seguinte valor mais próximo, por exemplo,  $n_{classes} = 28$ , caso tivesse sido o valor de  $n_{classes} = 29$  a ter melhor resultado. Se este novo cálculo também retornasse um valor de  $M$  maior, repetia-se o processo, até se encontrar o melhor valor para  $M$ . Uma vez encontrado este valor, a combinação de parâmetros que lhe deu origem seria a indicada a fornecer o melhor resultado para uma previsão para a dada série.

No entanto, este método de otimização poderá, para algumas séries temporais estudadas, trazer piores resultados de previsões. Tome-se como exemplo a tabela 33, do índice GM. A combinação de parâmetros que retornam o melhor resultado para os três tipos de transformação de dados é de  $\{n_{classes} = 10; tam_{seq} = 3\}$ . Esta combinação de parâmetros demonstra ser a mais frequente nos resultados obtidos, como observado anteriormente na Figura 177. Se, ao realizar cálculos para valores de  $n_{classes}$  menores que 10, houver a possibilidade de aumentar demasiado o valor de  $\Delta$  correspondente, ocorre que todos os valores verificados pelas iterações do

modelo sejam aceites como momentos semelhantes ao momento que se tenta prever. Este fenómeno provoca assim um distanciamento do intuito inicial do modelo – ter em consideração apenas momentos semelhantes ao momento que se tenta prever – e uma aproximação dos modelos clássicos, como o das Médias Móveis.

Seria como tal recomendável a optimização do parâmetro  $n_{classes}$  apenas para valores superiores ao valor mínimo inicial.

Relativamente ao tipo de transformação de dados, verifica-se um melhor desempenho com os valores transformados em Retorno Bruto do que os valores obtidos em Retorno Logarítmico, a par do que foi verificado nos modelos clássicos. Obteve-se melhores resultados em 9 dos 20 índices com Retorno Bruto, comparativamente melhor do os 3 em 20 obtidos com dados transformados em Retorno Logarítmico.

A par dos resultados para Retorno Bruto, os resultados obtidos com dados transformados às Diferenças obtiveram o melhor desempenho em 8 dos 20 índices. Tal como para os modelos clássicos, apenas se pode concluir que, na maioria dos casos, os dados transformados em Retorno Bruto têm melhor desempenho que os cálculos usando dados transformados em Retorno Logarítmico. No entanto, não é possível fazer tal distinção entre o desempenho para valores transformados em Retorno Bruto e para valores transformados às Diferenças.

Para resumir, verificou-se que o modelo proposto não consegue retornar melhores valores do que os valores do Limite de Futilidade,  $L_F$ . No entanto, será possível optimizar os parâmetros  $n_{classes}$  e  $tam_{seq}$  de modo tentar-se obter resultados melhores.

Também se pode diminuir o tempo de computação ao limitar-se os valores de  $n_{classes}$  no intervalo  $[10; 40[$  e os de  $tam_{seq}$  no intervalo  $[3; 4]$ .

## 7.6. MODELO EPSON

O modelo Epsom fornece uma solução diferente para a possibilidade da condição (7.11) ser verdadeira, em que  $s_a = \{Y_{t_a-l-1}, \dots, Y_{t_a-1}, Y_{t_a}\}$  e  $s_b = \{Y_t, Y_{t+1}, \dots, Y_{t+l-1}\}$

$$s_a = s_b \quad (7.11)$$

Neste modelo, e contrariamente ao modelo das Classes, não é efectuada uma transformação sobre os dados da série temporal em estudo. Tome-se novamente  $Y_t$  como a série de valores transformados de  $X_t$ , e  $s_a$  como um intervalo de tamanho  $tam_{seq}$  que representa os últimos  $tam_{seq}$  valores da série  $Y_t$ .

Como indicado anteriormente, o modo de previsão dos modelos propostos infere sobre a busca de momentos semelhantes ao momento actual,  $s_a$ . Assim, considere-se  $s_b$  como um intervalo de tamanho  $tam_{seq}$  que representa uma sequência de valores num qualquer ponto da série  $Y_t$ .

Para verificar se o momento  $s_b$  é semelhante a  $s_a$ , é dada a variável  $E$  aos limites de  $s_a$ . Assim é colocada uma margem em volta dos valores de  $s_a$  para ser considerada uma semelhança entre os valores correspondentes de  $s_a$  e  $s_b$  caso estes últimos se encontrem dentro dos limites impostos pela variável  $E$ . Por outras palavras, se a condição (7.12) for verdadeira,

$$(s_a - E/2) \leq s_b \leq (s_a + E/2) \quad (7.12)$$

o valor seguinte na série  $Y_t$  é considerado como previsão válida para o momento  $s_a$ . O cálculo da previsão final é feito como uma média de todos os valores de previsão obtidos ao longo das iterações realizadas pelo modelo na série  $Y_t$ .

### 7.6.1. RESULTADOS

São aqui demonstrados os resultados obtidos para o modelo Epsom. De modo a compreender-se melhor a forma de obtenção dos resultados, são expostos nas Tabelas 44 e 45 todos os resultados de Erro Médio Absoluto e respectiva Perda de Informação obtidos para o índice DELL com os dados transformados às Diferenças. Para o modelo Epsom, são utilizados como parâmetros o  $tam_{seq}$ , (também utilizado no modelo das Classes) com valores  $tam_{seq} = \{3, 4, 5, 6, 7\}$  e o Epsom, com valores  $E = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ . Para o índice DELL, com dados transformados às Diferenças, tem-se  $L_F = 0,3553$ .

**Tabela 44** – Valores de EMA do modelo Epsom para o índice DELL com dados às Diferenças

		Epsom										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
tam <sub>seq</sub>	3	0,3553	0,4226	0,4705	0,4748	0,4450	0,4198	0,4085	0,3971	0,3858	0,3823	0,3763
	4	0,3553	0,3921	0,4356	0,4506	0,4354	0,4133	0,4084	0,3950	0,3879	0,3806	0,3773
	5	0,3553	0,3661	0,4112	0,4307	0,4267	0,4122	0,4034	0,3965	0,3900	0,3834	0,3759
	6	0,3553	0,3587	0,3908	0,4175	0,4294	0,4184	0,4028	0,3949	3,3888	0,3793	0,3719
	7	0,3553	0,3558	0,3670	0,3983	0,4157	0,4123	0,3979	0,3867	0,3767	0,3769	0,3706

**Tabela 45** – Valores da PI correspondentes aos valores de EMA do índice DELL com dados às Diferenças

		Epsom										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
tam <sub>seq</sub>	3	99,2%	20,8%	4,4%	1,6%	0,8%	1,0%	0,4%	0,4%	0,6%	0,2%	0,2%
	4	100,0%	55,2%	19,4%	6,0%	2,4%	1,8%	1,0%	0,8%	0,4%	0,2%	0,0%
	5	100,0%	87,4%	38,4%	15,0%	6,0%	2,8%	1,6%	0,6%	0,6%	0,2%	0,2%
	6	100,0%	96,4%	61,8%	26,2%	10,4%	5,6%	2,6%	1,4%	1,4%	0,6%	0,8%
	7	100,0%	99,2%	81,4%	41,0%	17,8%	9,6%	4,8%	2,4%	2,2%	1,4%	1,0%

Verifica-se que, assim como para o modelo de Classes, os valores de PI diminuem à medida que aumentam os valores de Epsom. A justificação para este fenómeno é a mesma que para o modelo de Classes, ou seja, à medida que aumenta o intervalo à volta de  $s_a$ , aumenta o número de valores da série  $Y_t$  que são considerados semelhantes a  $s_a$ .

São apresentados nas seguintes tabelas 46 – 65 os resultados obtidos com o modelo Epsilon. Demonstra-se os correspondentes valores de  $L_F$ ,  $tam_{seq}$ ,  $E$ ,  $PI$ ,  $EMA$  e  $M$  para possibilitar a comparação entre os diferentes tipos de transformação de dados em estudo – Retorno Bruto, Retorno Logarítmico e às Diferenças.

**Tabela 46** – Resultados do Modelo Epsilon para o índice AAPL

AAPL	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,6866	4	1	2,2761	6,6%	-34,95%
Ret. Logarítmico	1,6785	4	0,9	2,2515	10,0%	-34,14%
Diferenças	1,2225	3	1	1,4754	9,0%	-20,69%

**Tabela 47** – Resultados do Modelo Epsilon para o índice BA

BA	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,9855	6	1	1,2682	7,8%	-28,69%
Ret. Logarítmico	0,9844	4	1	1,3432	1,8%	-36,45%
Diferenças	0,7810	5	1	0,9159	8,4%	-17,27%

**Tabela 48** – Resultados do Modelo Epsilon para o índice BP

BP	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,9577	6	1	1,1904	6,0%	-24,30%
Ret. Logarítmico	0,9576	6	1	1,1883	6,2%	-24,09%
Diferenças	0,6476	7	1	0,7196	4,2%	-11,12%

**Tabela 49** – Resultados do Modelo Epsilon para o índice CVX

CVX	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,0585	6	1	1,3084	8,0%	-23,61%
Ret. Logarítmico	1,0581	6	1	1,3085	6,2%	-23,67%
Diferenças	0,6887	7	1	0,7740	4,4%	-12,39%

**Tabela 50** – Resultados do Modelo Epsilon para o índice C

C	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,6704	7	1	0,8457	7,2%	-26,15%
Ret. Logarítmico	0,6701	7	1	0,8446	7,0%	-26,04%
Diferenças	0,3315	5	1	0,3548	0,0%	-7,03%

**Tabela 51** – Resultados do Modelo Epsilon para o índice DELL

DELL	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,4858	3	0,5	1,9332	8,2%	-30,11%
Ret. Logarítmico	1,4878	4	1	1,9408	6,4%	-30,45%
Diferenças	0,3553	7	1	0,3706	1,0%	-4,31%

**Tabela 52** – Resultados do Modelo Epsilon para o índice XOM

XOM	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	1,1333	4	0,9	1,4474	6,2%	-27,72%
Ret. Logarítmico	1,1335	4	0,9	1,4424	6,4%	-27,25%
Diferenças	0,9326	5	1	1,0364	9,4%	-11,13%

**Tabela 53** – Resultados do Modelo Epsilon para o índice F

F	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	1,5051	4	0,9	1,8643	10,0%	-23,87%
Ret. Logarítmico	1,5065	4	1	1,8830	7,6%	-24,99%
Diferenças	0,1220	7	1	0,1215	0,4%	0,41%

**Tabela 54** – Resultados do Modelo Epsilon para o índice GE

GE	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	0,6319	7	1	0,7835	0,0%	-23,99%
Ret. Logarítmico	0,6316	7	1	0,7840	1,8%	-24,13%
Diferenças	0,2203	7	1	0,2215	0,0%	-0,54%

**Tabela 55** – Resultados do Modelo Epsilon para o índice GM

GM	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	1,7881	3	0,9	2,2148	9,6%	-23,86%
Ret. Logarítmico	1,7858	3	0,9	2,2174	9,8%	-24,17%
Diferenças	0,4929	7	1	0,5377	0,0%	-9,09%

**Tabela 56** – Resultados do Modelo Epsilon para o índice HPQ

HPQ	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	1,0786	5	1	1,3827	0,8%	-28,19%
Ret. Logarítmico	1,0756	5	1	1,3902	6,6%	-29,25%
Diferenças	0,3650	7	1	0,3810	3,4%	-4,38%

**Tabela 57** – Resultados do Modelo Epsilon para o índice IBM

IBM	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	0,7264	6	0,8	0,8732	7,4%	-20,21%
Ret. Logarítmico	0,7270	7	1	0,8766	4,6%	-20,58%
Diferenças	0,6033	6	1	0,6612	2,0%	-9,60%

**Tabela 58** – Resultados do Modelo Epsilon para o índice JPM

JPM	L <sub>F</sub>	tam <sub>seq</sub>	E	EMA	PI	M
Ret. Bruto	0,7571	7	1	0,9624	7,6%	-27,12%
Ret. Logarítmico	0,7566	5	1	1,4181	7,2%	-87,43%
Diferenças	0,3297	4	1	0,3529	0,0%	-7,04%

**Tabela 59** – Resultados do Modelo Epsilon para o índice MSFT

MSFT	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,0871	5	1	1,4068	7,0%	-29,41%
Ret. Logarítmico	1,0861	5	1	1,4181	7,2%	-30,57%
Diferenças	0,3224	6	1	0,3293	0,4%	-2,14%

**Tabela 60** – Resultados do Modelo Epsilon para o índice SNE

SNE	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,1634	4	0,9	1,4887	6,2%	-27,96%
Ret. Logarítmico	1,1617	5	1	1,5172	3,6%	-30,60%
Diferenças	0,4720	7	1	0,5117	2,6%	-8,41%

**Tabela 61** – Resultados do Modelo Epsilon para o índice FTSE100

FTSE100	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,8442	5	1	1,0025	6,6%	-18,75%
Ret. Logarítmico	0,8444	5	1	1,0038	6,4%	-18,88%

**Tabela 62** – Resultados do Modelo Epsilon para o índice NASDAQ

NASDAQ	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,9626	5	0,9	1,2319	8,0%	-27,98%
Ret. Logarítmico	0,9628	5	0,9	1,2320	7,8%	-27,96%

**Tabela 63** – Resultados do Modelo Epsilon para o índice NYSE

NYSE	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,7665	5	0,9	0,9377	8,0%	-22,34%
Ret. Logarítmico	0,7670	5	0,9	0,9408	7,8%	-22,66%

**Tabela 64** – Resultados do Modelo Epsilon para o índice NIKKEI225

NIKKEI225	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	1,0216	5	1	1,2542	9,2%	-22,77%
Ret. Logarítmico	1,0229	5	1	1,2557	9,0%	-22,76%

**Tabela 65** – Resultados do Modelo Epsilon para o índice PSI20

PSI20	$L_F$	$tam_{seq}$	E	EMA	PI	M
Ret. Bruto	0,4827	7	1	0,5165	4,8%	-7,00%
Ret. Logarítmico	0,4827	7	1	0,5175	4,6%	-7,21%

## 7.6.2. OBSERVAÇÕES

De seguida efectua-se uma análise dos resultados obtidos para o modelo Epon. São analisados inicialmente os valores obtidos para o quociente de melhoria  $M$ , que permite uma comparação entre os restantes modelos de previsão utilizados neste trabalho.

Seguidamente, analisam-se as características inerentes dos resultados, sendo considerada a heurística presente, que consiste em escolher a combinação de parâmetros com o menor EMA dentro do grupo com  $PI < 10\%$ . São analisadas as frequências obtidas para os parâmetros utilizados,  $tam_{seq}$  e  $E$ . Igualmente importante é a análise aos tipos de transformação de dados realizadas – Retorno Bruto, Retorno Logarítmico e às Diferenças – e estudar qual destes tipos obtém os melhores resultados.

Finalmente, analisa-se a adequação do modelo Epon às séries temporais estudadas, e estimam-se possíveis melhorias e optimizações ao modelo.

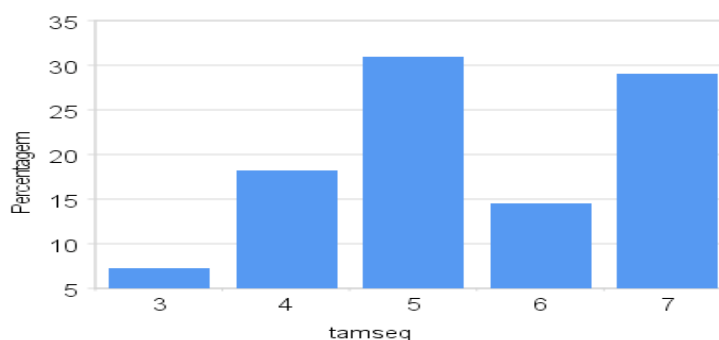
Observa-se que, tal como no modelo das Classes, os resultados obtidos para o cociente de melhoria  $M$  são na maioria valores negativos, excepção feita ao índice F, com os valores transformados às diferenças (Tabela 53). Verifica-se também, e apesar de este ser um tema mais aprofundado no capítulo 8, que estes valores demonstram ser mais baixos em melhoria do que os obtidos para o modelo das Classes.

Outro aspecto importante a notar é a ausência de valores de resultados para as Diferenças nos índices FTSE100, NASDAQ, NYSE, NIKKEI225 e PSI20, verificável respectivamente nas tabelas 61 a 65. A razão desta ausência assenta no incapacidade de utilização do modelo Epon nestas séries, devido aos valores da variável  $E$  utilizados. Recorde-se que é utilizado  $E = [0; 1]$  com intervalos de  $E = 0,1$ . Observe-se também nas tabelas 39 a 43 os valores de  $L_F$  para os referidos índices com dados às Diferenças. Verifica-se que estes valores para  $L_F$  encontram-se numa escala mais elevada relativamente aos valores de  $L_F$  dos restantes índices. Esta diferença de escala

provoca um aumento do valor da Perda de Informação, pois torna-se praticamente impossível encontrar semelhanças entre valores com diferenças até  $E = 1$ , quando estas séries apresentam possíveis variações muito superiores, na escala das unidades, ou até mesmo na escala das dezenas de unidades.

Foi tentado experimentalmente uma alteração aos valores de Epsom, multiplicando por um factor heurístico de 10, para possibilitar uma maior adequação do modelo às séries em estudo. No entanto, os resultados obtidos não foram animadores, obtendo-se valores de Perda de Informação superiores a 30%, inviabilizando, assim, a possibilidade de utilizar este modelo com os dados transformados às Diferenças, para os índices indicados anteriormente. Assim decidiu-se não incluir estes resultados nas tabelas anteriores.

Outro aspecto de análise importante são os valores obtidos para os parâmetros inerentes ao modelo:  $tam_{seq}$  e  $E$ . Observe-se a distribuição dos valores de  $tam_{seq}$  para os resultados obtidos nas tabelas 46 a 65 na Figura 179.



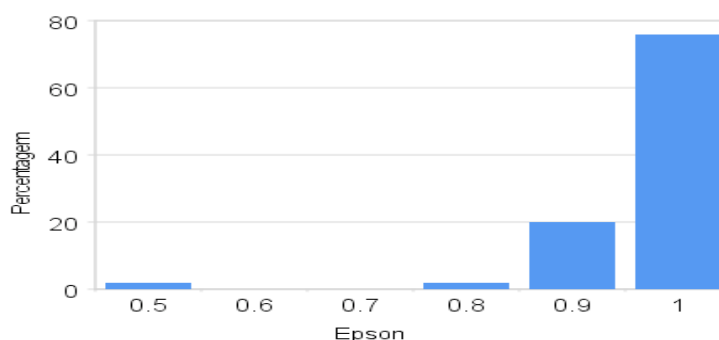
**Figura 179** – Distribuição dos valores do parâmetro  $tam_{seq}$  para o modelo de Epsom

Verifica-se que, contrariamente ao deduzido no modelo das Classes, há uma maior incidência para valores de  $tam_{seq} = 5$  (30,91%) e  $tam_{seq} = 7$  (29,09%). Tratam-se estes resultados de dissidência cognitiva, pois seria de esperar obter melhores resultados com sequências de tamanhos mais pequenos, como verificado para o modelo das Classes. No entanto, tenha-se em consideração a Tabela 64, que demonstra que à medida que aumenta o valor de  $tam_{seq}$ , também aumenta o valor da Perda de Informação. Assim, enquanto que os cálculos de previsão do modelo afastam o valor de EMA para longe de  $L_F$ , este aumento de perda de informação

aproxima novamente os valores para mais perto de  $L_F$ . Como tal, haverá uma diminuição do EMA para valores altos de  $tam_{seq}$ .

Face a estas observações, poderia concluir-se que, para obter melhores resultados com este modelo, bastaria aumentar o valor de  $tam_{seq}$ . No entanto, isto levaria a resultados que, apesar de se aproximarem do valor de  $L_F$ , nunca seriam melhores do que esse valor, ou seja, o correspondente cociente de melhoria nunca teria valores positivos.

O parâmetro  $E$  é outro que se deve analisar de modo a ser possível uma optimização do modelo e, assim, conduzia a uma melhoria dos resultados. Observando a Figura 180, verifica-se que a maioria dos melhores resultados obtidos para cada índice foram obtidos utilizando  $E = 0,9$  e  $E = 1,0$ . Apenas para uma pequena percentagem de resultados foram utilizados valores de  $E$  diferentes, podendo esta percentagem ser considerada não significativa.



**Figura 180** – Distribuição dos valores do parâmetro  $E$  para o modelo de Epson

Assim como para o modelo das Classes, pode-se considerar uma optimização dos valores de  $E$ . Inicialmente realizam-se testes para valores  $E > 1,0$ . No entanto, na realização de cálculos de previsão para estes valores, deve-se ter em consideração dois aspectos. O primeiro é que inicialmente será visível a diminuição do valor da Perda de Informação. É fácil deduzir tal resultado, visível na Tabela 45, pois à medida que se aumenta o intervalo, nos quais os valores encontrados são aceites como equivalentes aos dos segmento  $s_a$ , há maior probabilidade de se produzir uma previsão, logo diminuindo a Perda de Informação. O segundo aspecto a ter em consideração é semelhante ao verificado no modelo das Classes, em que um aumento

do intervalo  $\Delta$  – por diminuição do valor de  $n_{classes}$  – leva a uma aproximação ao valor de  $L_F$ . Neste modelo, o aumento do valor de  $E$  pode levar a este fenómeno. Nesta ordem de ideias, esta variação deve ser realizada com cuidado.

A optimização dos valores de  $E$  pode ser realizada como referido no modelo de Classes. Ao verificar-se qual a combinação dos parâmetros usados no modelo com o melhor resultado, realiza-se testes idênticos para os valores de  $E$  na vizinhança do valor escolhido. Na situação de existir um valor de  $E$  com melhor valor do que o obtido com a primeira realização de testes, realizam-se novamente testes na vizinhança do novo valor de  $E$ , aumentando a resolução do valor de  $E$  aplicado. Estas iterações podem ser realizadas até se obter o melhor resultado possível de EMA.

De seguida passa-se a analisar por fim os resultados relativamente ao tipo de transformação realizada nos dados das séries – Retorno Bruto, Retorno Logarítmico, Diferenças. Contrariamente aos restantes modelos estudados, o modelo Epsilon obteve sempre melhores resultados com a transformação às Diferenças, desde que os dados estivessem disponíveis nesse formato. Tal apenas não se verificou nos últimos 5 índices, por razões já discutidas.

Isto demonstra uma melhor adaptação do modelo aos dados transformados em Diferenças. Relativamente à diferença entre os resultados obtidos com os valores em Retorno Bruto e em Retorno Logarítmico, não se verifica uma variação significativa entre estes.

Concluindo, verificou-se que este modelo apresenta resultados piores do que aqueles obtidos no modelo de Classes, no que trata a resultados obtidos com os dados em Retorno Bruto ou Logarítmico. No caso de se utilizar valores às Diferenças, os resultados são melhores. Uma comparação geral no próximo capítulo irá colocar alguma luz sobre as variações de resultados obtidas relativamente a cada modelo estudado. Também foi verificado que os valores dos parâmetros utilizados podem ser optimizados, utilizando outros intervalos de dados para  $tam_{seq}$  e  $E$ . No entanto, e como foi indicado anteriormente, deve-se realizar estas optimizações com cuidado.



# 8. COMPARAÇÃO DAS METODOLOGIAS ESTUDADAS

Neste capítulo é realizada uma comparação dos modelos estudados, incidindo-se sobre os resultados obtidos nos cálculos de previsão efectuados. Para tal, é apresentado uma tabela (Tabela 66) contendo os melhores resultados obtido a partir de cada modelo, para cada índice.

Pretende-se, assim, analisar qual será o modelo que melhor se adapta a cada um dos índices e, se possível, concluir acerca do melhor modelo a utilizar de um modo geral para este tipo de séries temporais.

De modo a obter uma comparação adimensional entre os vários modelos, é apresentado o valor do cociente de melhoria  $M$  possibilitando, assim, a análise pretendida.

No caso em que não foi obtido em modelo algum um valor positivo de  $M$  para determinado índice, será indicado na mesma o melhor resultado para cada modelo, mesmo sendo este negativo.

**Tabela 66** – Comparação dos modelos estudados com base no valor de  $M$  correspondente

	AR	MA	ARMA	Classes	Epson
AAPL	0,11%	0,11%	0,13%	- 2,09%	- 20,69%
BA	0,35%	0,35%	0,41%	- 0,86%	- 17,27%
BP	0,00%	0,00%	- 0,02%	- 4,68%	- 11,12%
CVX	0,33%	0,29%	0,36%	- 3,51%	- 12,39%
C	0,18%	0,18%	- 0,42%	- 2,55%	- 7,03%
DELL	0,00%	0,00%	- 0,16%	- 3,32%	- 4,31%
XOM	0,78%	0,77%	- 0,30%	- 5,30%	- 11,13%
F	- 0,67%	- 0,66%	- 0,53%	- 2,10%	0,41%
GE	0,00%	0,05%	- 0,03%	- 1,50%	- 0,54%
GM	-0,02%	- 0,02%	- 0,01%	- 3,00%	- 9,09%
HPQ	0,69%	0,79%	0,77%	- 0,47%	- 4,38%
IBM	- 0,29%	- 0,30%	- 0,19%	- 1,59%	- 9,60%
JPM	0,09%	0,09%	0,11%	- 3,99%	- 7,04%
MSFT	0,03%	0,03%	- 0,17%	- 2,51%	- 2,14%
SNE	0,11%	0,08%	0,03%	- 4,90%	- 8,41%
FTSE100	0,11%	0,11%	0,07%*	- 7,15%	- 18,75%*
NASDAQ	0,78%	0,78%	0,95%*	- 0,32%	- 27,96%*
NYSE	0,22%	0,25%	0,00%*	- 5,25%	- 22,34%*
NIKKEI225	0,21%	0,23%	0,37%*	- 3,81%	- 22,76%*
PSI20	- 0,15%	- 0,12%	0,33%*	- 3,33%	- 7,00%*

A comparação entre os modelos clássicos – AR, MA e ARMA – já foi discutida na secção 6.6. Denota-se uma melhor adaptação do modelo MA para um maior número de índices, considerando apenas os índices com resultado  $M$  positivo. Para além do ao demonstrado na secção 6.6, ao considerar-se também os valores negativos de  $M$ , têm-se que o modelo ARMA obtém o mesmo número de índices com melhor resultado, a par do modelo MA, considerando o resultado do índice IBM, com  $M = -0,19\%$ .

Estes resultados demonstram a inadequação dos modelos propostos para a realização de previsões nas séries temporais estudadas. Verifica-se que os resultados obtidos por estes modelos são significativamente piores do que os obtidos pelos modelos clássicos. Excepção feita para o resultado positivo do modelo Epson para o índice F, com  $M = 0,41\%$ .

---

\* Melhor resultado obtido apenas entre Retorno Bruto e Retorno Logarítmico, por ausência de resultado para dados às Diferenças

No entanto, não se encontra uma justificação óbvia para este resultado, mesmo tendo em consideração as conclusões efectuadas no capítulo 2, relativamente à distribuição dos valores transformados dos vários índices.

Apesar de o melhor resultado para os modelos propostos partir do modelo Epson, este apresenta, de um modo geral, piores resultados que os correspondentes do modelo das Classes.

Foi referido nas secções 7.5.2 e 7.6.2 que os modelos propostos podem ser optimizados, sendo possível obter melhores resultados do que aqueles apresentados. No entanto, não é credível que se obtenha resultados melhores do que os modelos clássicos, sem uma alteração profunda de como é realizada a previsão e o cálculo do EMA.



## 9. CONCLUSÃO

Os sistemas financeiros dependem de uma grande variedade de ferramentas matemáticas e tecnologias computacionais que possibilitam a análise do ambiente financeiro. Neste trabalho analisou-se algumas dessas ferramentas computacionais, e foram propostas duas metodologias que se apresentam como uma alternativa aos modelos clássicos aqui estudados.

A facilitar os estudos computacionais, é disponibilizada uma vasta quantidade de informação digital que, com as ferramentas adequadas, permite uma análise de um ou vários aspectos do ambiente financeiro em que está inserida essa informação. Aqui mostrou-se que tal informação pode ser obtida gratuitamente através de determinados serviços, como a Google Finance e a Yahoo Finance. Apesar das limitações impostas pela versão gratuita destes serviços, é possível obter os dados necessários para a realização das análises efectuadas neste trabalho.

A utilização de dados com a frequência diária mostrou ser adequada aos testes realizados. Em trabalhos futuros, a disponibilidade de dados com frequência mais elevada - por exemplo, ao minuto - poderá permitir outro tipo de estudo, tendo em conta outras considerações, tais como a evolução de um activo financeiro ao longo do dia, cujo maior volume de transacções se processa no início e no final da sessão. Outra possibilidade será utilizar dados com frequência menor - semanal ou mensal.

No entanto, e como referido anteriormente, será de maior interesse utilizar uma frequência mensal para activos financeiros fortemente sujeitos a uma sazonalidade, algo não presente nos activos escolhidos para este trabalho.

De modo a utilizar-se as ferramentas matemáticas propostas, dá-se a necessidade de realizar transformações nos dados utilizados. Como os activos utilizados são fornecidos com os valores reais na moeda do país em que este está inserido, tais valores ficam sujeitos a efeitos da inflação e pagamentos de taxas de juro, o que provoca o aumento do seu valor ao longo do tempo. Realiza-se, assim, as transformações das séries em dados de Retorno Bruto, Retorno Logarítmico e às Diferenças. As exposições gráficas dos dados transformados demonstraram uma normalização das séries, com as suas representações histográficas a aproximarem-se de curvas tipo sino centradas num valor médio, normalmente próximo do valor zero. Tal normalização poderá permitir a previsão dos valores dos activos ao longo do tempo.

A metodologia de modelização das sucessões cronológicas utilizada neste trabalho baseia-se na metodologia apresentada por Box e Jenkins. Esta metodologia infere na escolha inicial de um modelo, estimando de seguida os respectivos parâmetros e posterior verificação da adequação do modelo escolhido. Se o modelo escolhido for adequado, é utilizado na previsão. Contrariamente, realiza-se novamente o ciclo utilizando um novo modelo.

O modo de escolha inicial do modelo recai sobre a análise dos figurinos gráficos da Função de Autocorrelação e da Função de Autocorrelação Parcial das séries temporais. A evolução gráfica ao longo dos atrasos  $k$  permite inferir, baseando-se nas características teóricas dos modelos clássicos utilizados - Autoregressivo, Média Móvel e Modelo Misto - qual o modelo mais adequado a realizar a previsão da série temporal.

Ao realizar a Função de Autocorrelação e Função de Autocorrelação Parcial nas séries em estudo, verificou-se nas representações gráficas que existe um decaimento brusco para zero a partir de um certo atraso  $k$ , e um decaimento para zero sob forma sinusoidal amortecida, respectivamente. Isto possibilitou concluir que o

modelo mais adequado a ser utilizado na previsão das séries seria o modelo de Média Móvel.

Com o intuito de verificar a validade desta conclusão, foram realizados testes para todos os modelos estudados neste trabalho permitindo, assim, uma comparação entre estes e, se possível, inferir acerca do melhor modelo para utilizar na previsão das séries em estudo. Introduziu-se como parâmetro de comparação entre modelos o Erro Médio Absoluto, que representa a média da diferença entre o valor real do índice e o valor da previsão calculado pelo modelo, para um determinado intervalo do índice. Considera-se como o modelo mais adequado aquele que apresentar o menor Erro Médio Absoluto. Também foi introduzido o Limite de Futilidade, que permite verificar se o modelo utilizado é, para além do mais adequado, capaz de produzir um resultado de previsão melhor do que se não fosse utilizado nenhum modelo. Podem-se considerar não adequados os modelos que produzam resultados de Erro Médio Absoluto mais altos do que o respectivo Limite de Futilidade.

Numa primeira parte, realizou-se um estudo dos modelos clássicos – Autoregressivo, Média Móvel e Modelo Misto. O intuito de tal estudo é tentar concluir qual destes modelos será o mais adequado para as séries temporais. No entanto, foi verificado que os resultados obtidos não forneceram um modelo claramente superior aos outros, em termos de melhoria relativamente ao Limite de Futilidade. O número de índices para os quais foram obtidos melhorias relativamente ao Limite de Futilidade estão distribuídos entre os três modelos em estudo. O modelo de Média Móvel apresenta valores mais altos de melhoria para 7 dos 20 índices, com o modelo Misto a ter 6 índices. No entanto, dos 7 índices, 5 são partilhados com o modelo Autoregressivo, por terem o mesmo valor de melhoria.

Esta observação vai contra a conclusão obtida com os figurinos teóricos da Função de Autocorrelação e Função de Autocorrelação Parcial, que demonstravam que o modelo de Média Móvel seria o mais indicado. Tal indica que não basta uma análise das FAC e FACP para saber qual o modelo mais adequado, mas que tal análise apenas servirá de ponto de partida.

Como alternativa aos modelos clássicos, foram propostos duas metodologias de análise e previsão de séries temporais, o modelo de Classes e o modelo Epsom. A principal diferença entre os modelos propostos e as metodologias clássicas está na quantidade de informação utilizada para a realização da previsão por parte de cada modelo. Enquanto que os modelos clássicos utilizam todo o espectro de informação disponível, os modelos propostos apenas procuram utilizar os valores de momentos semelhantes ao momento que estão a tentar prever. Tenta-se assim criar uma metodologia que melhor se adapte às séries temporais e que obtenha melhores resultados.

A metodologia das Classes coloca cada valor do activo dentro da classe que possui o intervalo de valores ao qual pertence esse valor. Deste modo, consegue traduzir os valores que se inserem dentro de tal intervalo em um único valor, facilitando assim o processo de busca de momento semelhantes ao que se pretende prever. Os parâmetros utilizados por este modelo para a sua optimização são  $n_{classes}$  e  $tam_{seq}$ .

A realização de testes com o modelo de Classes retornou sempre resultados de melhoria negativa, indicando que este modelo, e o modo como está aplicado, não será o mais indicado para a análise e previsão das séries temporais. Há a possibilidade de optimizar este modelo com valores dos parâmetros  $n_{classes}$  e  $tam_{seq}$  diferentes daqueles que são utilizados, mas tal optimização deve ser realizada com cuidado, e os resultados actuais não fazem prever melhorias positivas.

O modelo Epsom difere do modelo de Classes na maneira como considera dois valores semelhantes, aquando da busca por momentos semelhantes ao que tenta prever. Quando compara dois valores, se estes se encontrarem dentro de um intervalo de determinado valor  $E$ , tais valores serão considerados semelhantes. Não se pode, à partida, inferir se este método é melhor do que o método utilizado pelo modelo das Classes, pois só uma análise e comparação dos resultados obtidos é que poderá fornecer tal informação.

Assim como no modelo de Classes, também não foram obtidos valores positivos de melhoria, à excepção de um índice. De um modo geral, os resultados

obtidos pelo modelo de Epsom demonstraram ser piores do que aqueles obtidos pelo modelo de Classes, e ainda piores do que os valores obtidos pelos modelos clássicos. É possível, tal como para o modelo de Classes, uma optimização dos valores dos parâmetros utilizados,  $E$  e  $tam_{seq}$ . No entanto, tal optimização deve ser realizada com cuidado, e à imagem do modelo de Classes, os resultados apresentados não auferem uma boa imagem da possibilidade de obter valores do EMA que resultem em melhorias positivas.

No entanto, acredita-se ser possível utilizar a metodologia inerente aos modelos propostos para a criação de um modelo que se adapte melhor às séries temporais, obtendo-se assim melhores resultados do que os obtidos com os modelos clássicos.



## *Bibliografia*

- Akaike, H. (1978). A Bayesian extension of the minimum AIC procedure. 9-14.
- Akaike, H. (1974). A new look at the statistical model identification.
- Arsham, H. (1994). Time-Critical Decision Making for Business Administration. *Obtido de <http://home.ubalt.edu/ntsbarsh/stat-data/Forecast.htm>*
- Box, G. E., & Jenkins, G. M. (1970). Time Series Analysis, Forecasting and Control. *San Francisco: Holden-Day.*
- Dow Zhengdao, W. (2005). Microsimulation of Financial Markets. *Universiteit van Amsterdam.*
- Fama, E. F. (1965). The Behaviour of Stock Market Prices. *Journal of Business* , 34-105.
- French, K. R. (1980). Stock Returns and the Weekend Effect. *Journal of Financial Economics* , 55-69.
- Mantegna, R., & Stanley, H. E. (2000). An Introduction to Econophysics. Cambridge, *United Kingdom: Press Syndicate of the University of Cambridge.*
- Murteira, B. J., Müller, D. A., & Turkman, K. F. (1993). Análise de Sucessões Cronológicas. *McGraw Hill.*
- Pankratz, A. (1983). Forecasting with Univariate Box-Jenkins Models: Concept and Cases. *New York: Wiley.*
- Pollock, D. S. (1993). A Short Course of Time Series Analysis and Forecasting. *Vienna.*
- Tiao, G. C., & Tsay, R. S. (1984). Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and non-stationary ARIMA models.
- Tsay, R. (2006). Analysis of Financial Time Series.
- Turkman, K. F. (1985). The Choice of extremal models by Akaike's information criteria. *J. of Hidrology* , 307-315.