

ALEXANDRE SOARES RAMOS

**Previsões de Séries Temporais combinando  
modelos ARMA e Redes Neurais Artificiais**

Recife

2010

ALEXANDRE SOARES RAMOS

**Previsões de Séries Temporais combinando modelos  
ARMA e Redes Neurais Artificiais**

Trabalho apresentado ao Programa de Pós-graduação em Economia da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Doutor em Economia.

Orientador: Francisco de Souza Ramos

Recife  
2010

Catálogo na Fonte  
Bibliotecária Ângela de Fátima Correia Simões, CRB4-773

|       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| R175p | <p>Ramos, Alexandre Soares</p> <p>Previsões de series temporais combinando modelos ARMA e Redes Neurais Artificiais / Alexandre Soares Ramos. - 2010.</p> <p>164 folhas: il. 30 cm.</p> <p>Orientador: Prof. Dr. Francisco de Souza Ramos.</p> <p>Tese (Doutorado em Economia) – Universidade Federal de Pernambuco, CCSA, 2010.</p> <p>Inclui referências e apêndices.</p> <p>1. Redes Neurais (Computação). 2. Mercado financeiro. 3. Ações (Finanças) - Preços. Análise de séries temporais. I. Ramos, Francisco de Souza (Orientador). II. Título.</p> <p>332 CDD (22.ed.)</p> <p>UFPE (CSA 2017 –190)</p> |
|-------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**UNIVERSIDADE FEDERAL DE PERNAMBUCO**  
**CENTRO DE CIÊNCIAS SOCIAIS APLICADAS**  
**DEPARTAMENTO DE ECONOMIA**  
**PIMES/PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

**PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE TESE DO DOUTORADO  
EM ECONOMIA DE**

**ALEXANDRE SOARES RAMOS**

A Comissão Examinadora composta pelos professores abaixo, sob a presidência do primeiro, considera o Candidato Alexandre Soares Ramos **APROVADO**.

Recife, 15/03/2010

---

**Prof. Dr. Francisco de Souza Ramos**  
**Orientador**

---

**Prof. Dr. Ricardo Chaves Lima**  
**Examinador Interno**

---

**Prof. Dr. José Lamartine Távora Júnior**  
**Examinador Interno**

---

**Prof. Dr. Paulo Glício da Rocha**  
**Examinador Externo/FIR-CHESF**

---

**Prof. Dr. Marcos Roberto Góis de Oliveira**  
**Examinador Externo/DCA-UFPE**

# Agradecimentos

A meu orientador, Prof. Francisco Ramos, por sua orientação amiga e pela paciência.

Ao Prof. Ricardo Chaves por ter despertado em mim a vontade de trabalhar com Econometria e com Redes Neurais.

Ao Prof. Ernesto Raposo pela amizade e incentivo.

Ao Prof. Alexandre Stamford pela sua valiosa contribuição.

A minha esposa Priscila Andrade pelo companheirismo e estímulo de sempre. Aos meus pais Hermino e Terezinha pela eterna dedicação. A meu irmão André por tudo.

Aos meus sogros Socorro e Erinaldo pela atenção especial.

Aos meus grandes amigos Renato, Enildo e Agnaldo pelas arejadas discussões. Aos meus colegas de curso pelos excelentes momentos de descontração. Aos funcionários do PIMES que de um modo ou de outro me ajudaram nesta caminhada, em especial, a Patrícia.

Ao PIMES por ter fornecido as condições necessárias para a realização de um bom trabalho. Ao CNPq pelo financiamento de parte deste trabalho.

*“Man survives earthquakes, epidemics, the horrors of disease, and agonies of the soul, but all the time his most tormenting tragedy has been, is, and will always be, the tragedy of the bedroom.” - Leo Tolstoy*

# Resumo

A tomada de decisão em um ambiente envolvendo incerteza é um problema que data dos primórdios da civilização. Atualmente, uma das áreas mais desenvolvidas em termos de metodologia de análise do comportamento do mercado sob incerteza é o setor financeiro. A análise de evolução dos preços de ações tem demandado o uso de um instrumental analítico fortemente fundamentado e que envolve a utilização de instrumental quantitativo bastante avançado. Os avanços recentes no estado da arte na área da econometria e as controvérsias com os resultados teóricos da Economia, demandou e ainda demanda, uma abordagem multidisciplinar para a análise de dados fazendo surgir a criação de novos modelos, em particular, para séries financeiras. Como os modelos tradicionais usados isoladamente, mesmo os não lineares, não apresentam resultados satisfatórios em todos os períodos analisados, uma forma sugerida em muitos casos é combinar modelos para aproveitar as melhores capacidades de previsão de cada um deles, pretendendo com isso, captar os principais parâmetros determinantes da dinâmica das séries temporais.

Nesta tese é proposto a utilização de um sistema híbrido, baseado em Redes Neurais Artificiais (RNA) e Modelos econométricos não-lineares, para realizar previsões com o intuito de se conseguir um melhor resultado com esses modelos quando comparados aos modelos isolados. O objetivo do trabalho é investigar quais modelos fornecem a melhor qualidade de previsão, as limitações desses modelos e se os mesmos têm aplicações práticas quando aplicados a previsões de séries temporais. Como resultado, foi possível concluir que o modelo combinado se ajusta melhor aos dados reais do que os modelos individuais, fato este comprovado através da comparação das diferentes medidas de performance. Por exemplo, no caso da série de preços das ações da Petrobras, o erro quadrático médio (MSE) do modelo combinado é reduzido em 40% quando comparado ao MSE do melhor dos modelos estudados. No conjunto das séries não-financeiras, o MSE do modelo combinado para a série de emissão de CO<sub>2</sub> apresentou uma redução de 65% quando comparado ao MSE do modelo ARIMA que para a presente série foi o modelo individual com melhor performance. Usou-se o MSE como exemplo pois o MSE é a medida mais utilizada pela sua facilidade de cálculo e por ser uma métrica encontrada em praticamente todos os pacotes computacionais que trabalham com estatística e RNAs, facilitando assim a comparação dos resultados. De uma forma geral temos sete parâmetros de comparação entre os modelos em sete séries de teste. Assim temos 49 resultados dos parâmetros de comparação. Em 35 resultados o modelo combinado é o melhor dos três, ou seja, para as séries em teste, em 71,4% das vezes o modelo combinado é o melhor. Em 42 resultados o modelo combinado aparece como o melhor resultado, ou seja, para as séries em teste, em 85,7% das vezes o modelo

combinado é melhor ou tão bom quanto o das RNAs. Enquanto isso, o modelo RNA aparece apenas 7 vezes (14,3%) como o melhor resultado e 13 vezes (26,5% ) das vezes como melhor ou tão bom quanto o modelo Combinado. O Modelo de Box&Jenkins só apresenta um resultado como sendo o melhor na série de CO<sub>2</sub>, na medida BIC. Assim, verifica-se que entre os modelos e dentre essas séries, o modelo Combinado é o mais adequado a se utilizar. Logo, podemos afirmar que, em algumas situações, esses modelos híbridos fornecem melhores resultados quando comparados aos modelos econométricos tradicionais ou mesmo aos modelos puros de Redes Neurais geralmente utilizados para se fazer previsões.

**Palavras-chaves:** Redes Neurais Artificiais. Mercado Financeiro. Modelos Complexos. Previsão de preço de ações. Séries Temporais

# Abstract

Decision making under uncertainty is a problem that has existed since the dawn of civilization. Currently, the financial sector is one of the most advanced in terms of methodological analysis of market behavior under uncertainty. The analysis of evolution of stock prices has required the use of a quantitative analytical tools strongly based. Recent advances in the state of the art in the area of econometrics and controversies with the theoretical results of the Economy have demanded and still demand a multidisciplinary approach to data analysis, creating the creation of new models, particularly for financial series. As the traditional models used alone, even non-linear ones, do not present satisfactory results in all the analyzed periods, a suggested way in many cases is to combine models to take advantage of the best predictive capacities of each one of them, intending to capture the main parameters that determine the dynamics of the time series. This thesis proposes to use a hybrid system based on Dynamic Artificial Neural Networks (DANN) and non-linear econometric models to make predictions, in order to achieve a better result. The objective of this study is to investigate which models provide the highest quality prediction, the limitations of these models and whether they have practical applications when applied to forecasts financial series.

As a result, it was concluded that, in some cases, the combined model fits the data better than the actual individual models, a fact confirmed by comparing the different measures of performance. For example, in the series of stock prices of Petrobras, the mean square error

(MSE) of the combined model is reduced by 40% compared to the MSE of the best studied individual model. In the group of non-financial series, the MSE of the combined model for the CO<sub>2</sub> emission has fallen by 65% compared to the MSE of the ARIMA model for which this series was the individual model with better performance. The MSE was used as an example because the MSE is the most used measure for its ease of calculation and for being a metric found in almost all the computational packages that work with statistics and RNAs, thus facilitating the comparison of the results. In general we have seven parameters of comparison between the models in seven test series. Thus we have 49 results of the comparison parameters. In 35 results the combined model is the best of the three, that is, for the series under test, at 71.4 % of the time the combined model is the best. In 42 results the combined model appears as the best result, that is, for the series under test, at 85.7 % of the time the combined model is better or as good as the RNAs. Meanwhile, the RNA model appears only 7 times (14.3 %) as the best result and 13 times (26.5 %) of the times as better or as good as the Combined model. The Box&Jenkins Model only shows one result as being the best in the CO<sub>2</sub> series, in

BIC measurement. Thus, it is verified that among the models and among these series, the Combined model is the most suitable to be used. Thus, we can say that, in some situations, these hybrid models provide better results when compared to traditional econometric models or even the pure models of neural networks generally used to make predictions.

**Keywords:** Artificial neural network. Financial markets. Complex models. Stock prediction. Time series

# Lista de ilustrações

|                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Imagens do Sol . . . . .                                                                                               | 32 |
| Figura 2 – Série de Manchas Solares (sunspot) - medidas anuais. . . . .                                                           | 32 |
| Figura 3 – Níveis de emissão de dióxido de carbono ( $\text{CO}_2$ ) no Canadá. . . . .                                           | 34 |
| Figura 4 – Série do índice do Ibovespa. . . . .                                                                                   | 36 |
| Figura 5 – Série do preço diário das ações do Bradesco. . . . .                                                                   | 37 |
| Figura 6 – Gráfico da série ações da Petrobras. . . . .                                                                           | 39 |
| Figura 7 – Gráfico da série do índice <i>Dow Jones Industrial Average</i> . . . . .                                               | 40 |
| Figura 8 – Gráfico da série do preço das ações da Itausa. . . . .                                                                 | 42 |
| Figura 9 – Representação básica de um neurônio biológico . . . . .                                                                | 63 |
| Figura 10 – Diagrama esquemático de um neurônio artificial . . . . .                                                              | 64 |
| Figura 11 – Exemplo de uma RNA não recorrente . . . . .                                                                           | 66 |
| Figura 12 – Exemplo de uma RNA recorrente . . . . .                                                                               | 67 |
| Figura 13 – RNA multi-camada . . . . .                                                                                            | 69 |
| Figura 14 – Comparativo entre os valores reais e as previsões geradas pelo modelo $\text{ARIMA}(9,0,0)$ . . . . .                 | 82 |
| Figura 15 – Comparativo entre os valores reais e as previsões geradas pelo modelo $\text{ARIMA}(9,0,1)$ . . . . .                 | 83 |
| Figura 16 – Níveis mensais de dióxido de carbono com símbolos. . . . .                                                            | 83 |
| Figura 17 – ACF amostral para os níveis de $\text{CO}_2$ . . . . .                                                                | 84 |
| Figura 18 – Gráfico da série temporal da primeira diferença dos níveis de $\text{CO}_2$ . . . . .                                 | 85 |
| Figura 19 – ACF amostral da primeira diferença dos níveis de $\text{CO}_2$ . . . . .                                              | 85 |
| Figura 20 – Gráfico da série temporal da primeira diferença e diferença sazonal dos níveis de $\text{CO}_2$ . . . . .             | 86 |
| Figura 21 – ACF amostral da primeira diferença e diferença sazonal dos níveis de $\text{CO}_2$ . . . . .                          | 87 |
| Figura 22 – Resíduos para o modelo $\text{ARIMA}(0,1,1) \times (0,1,1)[12]$ . . . . .                                             | 87 |
| Figura 23 – ACF para o modelo $\text{ARIMA}(0,1,1) \times (0,1,1)[12]$ . . . . .                                                  | 88 |
| Figura 24 – Resíduos para o modelo $\text{ARIMA}(0,1,1) \times (0,1,1)[12]$ . . . . .                                             | 90 |
| Figura 25 – Resíduos para o modelo $\text{ARIMA}(0,1,1) \times (0,1,1)[12]$ . . . . .                                             | 90 |
| Figura 26 – Grupo de Teste (preto) e valores previstos (vermelho) com o modelo $\text{ARIMA}(0,1,1) \times (0,1,1)[12]$ . . . . . | 91 |
| Figura 27 – Gráfico dos retornos do Ibovespa. . . . .                                                                             | 92 |
| Figura 28 – Função de autocorrelação dos retornos. . . . .                                                                        | 94 |
| Figura 29 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição $t$ -Student. . . . .                          | 95 |

|                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 30 – Previsão obtida pelo modelo ARMA-GARCH para os log-retornos de Ibovespa. . . . .                                                        | 96  |
| Figura 31 – Previsão da volatilidade. . . . .                                                                                                       | 97  |
| Figura 32 – Intervalo de confiança. . . . .                                                                                                         | 97  |
| Figura 33 – Gráfico dos retornos da série do Bradesco. . . . .                                                                                      | 98  |
| Figura 34 – Função de autocorrelação dos retornos da série Bradesco. . . . .                                                                        | 99  |
| Figura 35 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição <i>t</i> -Student. . . . .                                       | 101 |
| Figura 36 – Previsão obtida pelo modelo ARMA-GARCH para os log-retornos de Bradesco. . . . .                                                        | 102 |
| Figura 37 – Previsão da volatilidade. . . . .                                                                                                       | 103 |
| Figura 38 – Intervalo de confiança. . . . .                                                                                                         | 103 |
| Figura 39 – Gráfico dos retornos da série do Itausa. . . . .                                                                                        | 104 |
| Figura 40 – Função de autocorrelação dos retornos da série Itausa. . . . .                                                                          | 105 |
| Figura 41 – Gráfico quantil quantil dos resíduos sobre a distribuição <i>t</i> -Student . . . . .                                                   | 107 |
| Figura 42 – Previsão obtida pelo modelo ARMA-TGARCH para os log-retornos da série Itausa. . . . .                                                   | 108 |
| Figura 43 – Previsão da volatilidade. . . . .                                                                                                       | 109 |
| Figura 44 – Intervalo de confiança. . . . .                                                                                                         | 109 |
| Figura 45 – Gráfico dos retornos da série do índice Dow Jones. . . . .                                                                              | 110 |
| Figura 46 – Função de autocorrelação dos retornos dos índices Dow Jones. . . . .                                                                    | 111 |
| Figura 47 – Gráfico dos retornos da Petrobras. . . . .                                                                                              | 113 |
| Figura 48 – Função de autocorrelação dos retornos da série Petrobras. . . . .                                                                       | 114 |
| Figura 49 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição <i>t</i> -Student. <i>Fonte: elaborada pelo autor.</i> . . . . . | 116 |
| Figura 50 – Previsão obtida pelo modelo ARMA-TGARCH para os retornos da Petrobras. . . . .                                                          | 117 |
| Figura 51 – Previsão da volatilidade. . . . .                                                                                                       | 117 |
| Figura 52 – Intervalo de confiança. . . . .                                                                                                         | 118 |
| Figura 53 – Topologia de um Rede Neural com Dinâmica de entrada . . . . .                                                                           | 118 |
| Figura 54 – Comparativo entre os valores reais e as previsões - Manchas Solares. . . . .                                                            | 121 |
| Figura 55 – Comparativo entre os valores reais e as previsões - Bradesco. . . . .                                                                   | 122 |
| Figura 56 – Comparativo entre os valores reais e as previsões - Petrobras. . . . .                                                                  | 123 |
| Figura 57 – Comparativo entre os valores reais e as previsões - Itausa. . . . .                                                                     | 124 |
| Figura 58 – Comparativo entre os valores reais e as previsões - Ibovespa. . . . .                                                                   | 126 |
| Figura 59 – Comparativo entre os valores reais e as previsões -CO <sub>2</sub> . . . . .                                                            | 127 |
| Figura 60 – Comparativo entre os valores reais e as previsões - Dow Jones. . . . .                                                                  | 128 |

|                                                                                                                                                                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 61 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box & Jenkins (linha tracejada) para o conjunto de teste da série do Valor de Fechamento das Ações do Itaúsa. . . . .                    | 130 |
| Figura 62 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações da Petrobras. . . . .                  | 132 |
| Figura 63 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box & Jenkins (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Bradesco. . . . .                   | 133 |
| Figura 64 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série do Índice de preços do Ibovespa. . .                                   | 134 |
| Figura 65 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série do Índice Dow Jones Industrial Average. . . . .                        | 135 |
| Figura 66 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Itaúsa. . . . .    | 136 |
| Figura 67 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão geradas pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações da Petrobras. . . . . | 137 |
| Figura 68 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão geradas pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Bradesco. . . . .  | 138 |
| Figura 69 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série do Índice de preços do Ibovespa. . . . .              | 139 |
| Figura 70 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série do Índice Dow Jones Industrial Average. . . . .       | 140 |

|                                                                                                                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 71 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Manchas Solares. . . . .      | 142 |
| Figura 72 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de CO <sub>2</sub> . . . . .     | 143 |
| Figura 73 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índice do Itaúsa. . . . .     | 144 |
| Figura 74 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índice da Petrobras. . . . .  | 145 |
| Figura 75 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices do Bradesco. . . . .  | 146 |
| Figura 76 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices do Ibovespa. . . . .  | 147 |
| Figura 77 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices da Dow Jones. . . . . | 148 |

# Sumário

|            |                                                            |           |
|------------|------------------------------------------------------------|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>                                          | <b>18</b> |
| <b>1.1</b> | <b>Introdução</b>                                          | <b>18</b> |
| 1.1.1      | Motivação, Justificativa e Objetivos                       | 18        |
| 1.1.2      | A importância das bolsas de valores                        | 19        |
| 1.1.3      | Estrutura da tese                                          | 23        |
| <b>2</b>   | <b>PREVISÃO DE SÉRIES FINANCEIRAS</b>                      | <b>25</b> |
| <b>2.1</b> | <b>Previsão de Séries Temporais</b>                        | <b>25</b> |
| 2.1.1      | Previsão de Séries Financeiras                             | 26        |
| <b>2.2</b> | <b>Caracterização das Séries Temporais Utilizadas</b>      | <b>31</b> |
| 2.2.1      | Série das Manchas Solares ( <i>sunspot</i> )               | 31        |
| 2.2.2      | Série de emissão de dióxido de carbono - CO <sub>2</sub>   | 33        |
| 2.2.3      | Série do índice do Ibovespa                                | 34        |
| 2.2.4      | Série do Bradesco                                          | 37        |
| 2.2.5      | Série do valor de fechamento das ações da Petrobras        | 37        |
| 2.2.6      | Série do índice <i>Dow Jones Industrial Average</i> (DJIA) | 40        |
| 2.2.7      | Série do Itaúsa                                            | 40        |
| <b>3</b>   | <b>MÉTODOS TRADICIONAIS</b>                                | <b>43</b> |
| <b>3.1</b> | <b>Modelos Econométricos</b>                               | <b>43</b> |
| 3.1.0.1    | Objetivos                                                  | 44        |
| 3.1.1      | Modelos probabilísticos                                    | 44        |
| 3.1.1.1    | Processos estacionários                                    | 45        |
| 3.1.1.2    | Função de autocorrelação                                   | 46        |
| 3.1.1.3    | Alguns processos estocásticos                              | 47        |
| 3.1.1.3.1  | Sequência aleatória                                        | 47        |
| 3.1.1.3.2  | Passeio Aleatório                                          | 48        |
| 3.1.1.4    | Filtragem                                                  | 49        |
| 3.1.1.5    | Diferenciação                                              | 49        |
| 3.1.1.6    | Correlograma                                               | 50        |
| 3.1.1.7    | Teste de Box Pierce                                        | 51        |
| 3.1.1.8    | Teste da raiz unitária                                     | 51        |
| 3.1.2      | Modelagem de Box-Jenkins                                   | 53        |
| 3.1.2.1    | Modelos auto-regressivos                                   | 54        |
| 3.1.2.2    | Modelos de médias móveis                                   | 54        |
| 3.1.2.3    | Modelos auto-regressivos e de médias móveis                | 55        |

|            |                                                                                     |           |
|------------|-------------------------------------------------------------------------------------|-----------|
| 3.1.2.4    | Modelos ARIMA e SARIMA . . . . .                                                    | 56        |
| 3.1.2.5    | Cr terios de identifica  o de $p$ e $q$ . . . . .                                   | 56        |
| 3.1.3      | Heterocedasticidade condicional . . . . .                                           | 57        |
| 3.1.4      | Modelos ARCH . . . . .                                                              | 58        |
| 3.1.5      | Modelos GARCH . . . . .                                                             | 59        |
| 3.1.6      | Avalia  o de previs  o . . . . .                                                    | 60        |
| <b>3.2</b> | <b>Modelos de Redes Neurais Artificiais . . . . .</b>                               | <b>62</b> |
| 3.2.1      | O Neur nio Artificial . . . . .                                                     | 63        |
| 3.2.1.1    | Fun  es de Ativa  o . . . . .                                                       | 65        |
| 3.2.1.2    | Topologia das Redes Neurais Artificiais . . . . .                                   | 66        |
| 3.2.1.3    | Tipos de Treinamento . . . . .                                                      | 67        |
| 3.2.2      | Redes Neurais Artificiais Multi-Camadas . . . . .                                   | 68        |
| 3.2.2.1    | Algoritmo de Retropropaga  o do Erro . . . . .                                      | 68        |
| 3.2.3      | Previs  o atrav s de RNAs . . . . .                                                 | 70        |
| 3.2.3.1    | Previs  o <i>Single</i> e <i>Multi-Step</i> . . . . .                               | 72        |
| 3.2.3.2    | Aplica  es . . . . .                                                                | 72        |
| 3.2.4      | Modelagem das RNA's . . . . .                                                       | 74        |
| 3.2.4.1    | Fatores Determinantes no Desempenho das RNAs . . . . .                              | 74        |
| 3.2.4.2    | Controle dos Fatores Determinantes no Desempenho das RNAs . . . . .                 | 75        |
| 3.2.4.2.1  | Padroniza  o . . . . .                                                              | 75        |
| 3.2.4.2.2  | Fun  es de Ativa  o . . . . .                                                       | 76        |
| 3.2.4.2.3  | Fun  es Erro . . . . .                                                              | 77        |
| 3.2.4.2.4  | Par metros de Controle do Gradiente Descendente e do Tempo de Treinamento . . . . . | 78        |
| 3.2.4.2.5  | Arquitetura das RNAs . . . . .                                                      | 78        |
| <b>3.3</b> | <b>Resultados experimentais com m todos tradicionais . . . . .</b>                  | <b>80</b> |
| 3.3.1      | An lise atrav s da Metodologia Box-Jenkins e modelos ARMA-GARCH . . . . .           | 80        |
| 3.3.1.1    | Ajuste de modelos para s rie das Manchas Solares . . . . .                          | 80        |
| 3.3.1.2    | An lise da s rie CO <sub>2</sub> . . . . .                                          | 82        |
| 3.3.1.3    | An lise da s rie do Ibovespa . . . . .                                              | 92        |
| 3.3.1.4    | An lise da s rie do Bradesco . . . . .                                              | 97        |
| 3.3.1.5    | An lise da s rie do Itausa . . . . .                                                | 104       |
| 3.3.1.6    | An lise da s rie <i>Dow Jones</i> . . . . .                                         | 109       |
| 3.3.1.7    | An lise da s rie da Petrobras . . . . .                                             | 113       |
| 3.3.2      | An lise atrav s de Redes Neurais Artificiais Din micas . . . . .                    | 118       |
| 3.3.2.1    | Rede Neural Artificial Din mica (RNAD) . . . . .                                    | 118       |
| 3.3.2.2    | S ries e treinamento . . . . .                                                      | 120       |
| 3.3.2.3    | Resultados Experimentais . . . . .                                                  | 121       |
| 3.3.2.3.1  | S rie Manchas Solares . . . . .                                                     | 121       |
| 3.3.2.3.2  | S rie Bradesco . . . . .                                                            | 122       |

|            |                                                                                                        |            |
|------------|--------------------------------------------------------------------------------------------------------|------------|
| 3.3.2.3.3  | Série Petrobras . . . . .                                                                              | 123        |
| 3.3.2.3.4  | Série Itausa . . . . .                                                                                 | 124        |
| 3.3.2.3.5  | Série Ibovespa . . . . .                                                                               | 126        |
| 3.3.2.3.6  | Série CO <sub>2</sub> . . . . .                                                                        | 127        |
| 3.3.2.3.7  | Série Dow Jones . . . . .                                                                              | 128        |
| 3.3.2.4    | Conclusões . . . . .                                                                                   | 129        |
| <b>4</b>   | <b>MODELO COMBINADO-ESTIMAÇÃO E RESULTADOS . . . . .</b>                                               | <b>130</b> |
| <b>4.1</b> | <b>Experimentos com os modelos de Box e Jenkins . . . . .</b>                                          | <b>130</b> |
| 4.1.1      | Série do valor de fechamento das ações do Itausa . . . . .                                             | 130        |
| 4.1.2      | Série do valor de fechamento das ações da Petrobras . . . . .                                          | 131        |
| 4.1.3      | Série do valor de fechamento das ações do Bradesco . . . . .                                           | 132        |
| 4.1.4      | Série do índice de preços do Ibovespa . . . . .                                                        | 133        |
| 4.1.5      | Série do índice Dow Jones Industrial Average (DJIA) . . . . .                                          | 133        |
| <b>4.2</b> | <b>Resultados experimentais com a aplicação de Redes Neurais Artifi-<br/>ciais Dinâmicas . . . . .</b> | <b>135</b> |
| 4.2.1      | Série do valor de fechamento das ações do Itausa . . . . .                                             | 135        |
| 4.2.2      | Série do valor de fechamento das ações da Petrobras . . . . .                                          | 135        |
| 4.2.3      | Série do valor de fechamento das ações do Bradesco . . . . .                                           | 137        |
| 4.2.4      | Série do índice de preços do Ibovespa . . . . .                                                        | 139        |
| 4.2.5      | Série do índice Dow Jones Industrial Average (DJIA) . . . . .                                          | 139        |
| <b>4.3</b> | <b>Modelo Combinado . . . . .</b>                                                                      | <b>140</b> |
| 4.3.1      | Resultados experimentais com a aplicação do Modelo Combinado . . . . .                                 | 141        |
| 4.3.1.1    | Série de Manchas Solares . . . . .                                                                     | 141        |
| 4.3.1.2    | Série de CO <sub>2</sub> . . . . .                                                                     | 141        |
| 4.3.1.3    | Série do valor de fechamento das ações do Itausa . . . . .                                             | 143        |
| 4.3.1.4    | Série do valor de fechamento das ações da Petrobras . . . . .                                          | 145        |
| 4.3.1.5    | Série do valor de fechamento das ações do Bradesco . . . . .                                           | 146        |
| 4.3.1.6    | Série do índice de preços do Ibovespa . . . . .                                                        | 147        |
| 4.3.1.7    | Série do índice Dow Jones Industrial Average (DJIA) . . . . .                                          | 148        |
| <b>4.4</b> | <b>Comparação entre modelos . . . . .</b>                                                              | <b>149</b> |
| <b>4.5</b> | <b>Conclusões . . . . .</b>                                                                            | <b>151</b> |
| <b>5</b>   | <b>CONCLUSÕES . . . . .</b>                                                                            | <b>153</b> |
|            | <b>Referências . . . . .</b>                                                                           | <b>156</b> |

# 1 Introdução

## 1.1 Introdução

Nessa seção é apresentada a Motivação, Justificativa e Objetivos da tese, em seguida, dada a importância das séries financeiras, e dado que essa tese trata de 5 séries relacionadas a bolsa de valores, uma seção específica sobre a bolsa de valores é apresentada antes de se expor a estrutura da tese.

### 1.1.1 Motivação, Justificativa e Objetivos

A ideia de se prever acontecimentos futuros a partir de informações sobre o passado sempre demandou bastante esforço por parte dos cientistas. O ramo da ciência que trabalha com a elaboração de modelos para se realizar previsões é a Estatística e a cada dia novos modelos estatísticos são propostos na tentativa de se melhorar a qualidade das previsões. Dentre os modelos mais estudados encontram-se os modelos de Box & Jenkins (1976), também conhecidos como modelos ARIMA e bastante utilizados em aplicações econômicas e financeiras.

Uma das limitações dos modelos de Box & Jenkins é o fato de serem modelos lineares e, quando aplicados a situações onde as relações são não-lineares, podem gerar resultados não satisfatórios (Hamilton, 1994). Dessa forma, vários modelos estatísticos não-lineares foram propostos. Entre eles podemos citar os modelos de heterocedasticidade condicional auto-regressivos (ARCH), os modelos de heterocedasticidade condicional auto-regressivos generalizados (GARCH), os modelos auto-regressivos exponenciais (*exponential autoregressive models*), além de vários outros.

Esses modelos apresentam vantagens e desvantagens dependendo das situações nas quais são aplicados. Apesar de se ter como principal objetivo da inclusão da não-linearidade o aumento da capacidade de previsão, os resultados obtidos de maneira geral ainda são insatisfatórios podendo inclusive serem comparados aos resultados dos modelos lineares (De Gooijer e Kumar, 1992). Vários trabalhos mostram que essa deficiência dos modelos não-lineares em realizar previsões melhores do que os modelos lineares ainda persiste nos dias de hoje (Clements *et al*, 2004). É esperado que a sofisticação da estimação, da seleção e dos testes de modelos não-lineares aperfeiçoem o desempenho das previsões.

Neste cenário, temos como alternativa as Redes Neurais Artificiais (RNAs), uma das muitas técnicas de Inteligência Artificial aplicáveis ao problema de previsão. No entanto,

ainda não se tem um veredicto sobre sua performance quando comparadas as técnicas econométricas apesar de alguns artigos como o de Haider *et al* (2009) sugerirem melhores resultados de previsão alcançados com RNAs. Assim, muito esforço ainda necessita ser feito no desenvolvimento de modelos que possam apresentar melhores resultados de previsões. Aliado a essa necessidade, a integração de diferentes modelagens compondo um modelo combinado dessas diferentes técnicas se apresenta como uma área ainda pouco explorada do ponto de vista científico gerando um interessante campo a ser estudado. No artigo de Ömer Faruk (2009) é sugerido um modelo híbrido composto de ARIMA e RNAs para prever o nível de qualidade de água que apresenta melhores resultados de previsão do que os modelos individuais. A idéia dessas combinações é que o ponto forte de uma técnica compense o ponto fraco da outra gerando como resultado da combinação modelos com melhores capacidades de previsão.

Pelo exposto, esta Tese tem como objetivo:

Comparar previsões sobre séries temporais utilizando modelos econométricos do tipo ARMA e Redes Neurais Artificiais Dinâmicas, para o caso dos log-retornos, comparando as performances das diversas previsões, bem como construir um modelo combinado baseado na metodologia de Zou e Yang (2004) para o caso de previsões das séries .

A escolha por este modelo seu deu devido ao modelo desenvolvido por tais autores reduzir a instabilidade da previsão e, portanto, melhorar a precisão da predição, como apresentado por eles em algumas simulações e exemplos de dados reais que indicaram vantagem potencial sobre a seleção do modelo para tais casos.

### 1.1.2 A importância das bolsas de valores

O aumento da capacidade de produção, através de investimentos em capital e recursos humanos, interfere diretamente sobre o desenvolvimento econômico sustentável. Isto significa que quando existe a acumulação do capital produtivo e o aumento da produtividade, os investimentos produzem resultados positivos. Ainda, quando esses investimentos são direcionados para as alternativas que apresentam maiores retornos econômicos e sociais, percebe-se uma aceleração no crescimento. Dessa forma, os investimentos e a forma como são realizados estão na base do processo de desenvolvimento econômico e social.

Taxas elevadas de poupança ou até mesmo alta eficiência na sua intermediação são observadas em países desenvolvidos ou em rápido processo de desenvolvimento. Como viabiliza os investimentos, a eficiência na utilização da poupança é fundamental, pois irá determinar o curso do investimento. Esses fatores, quando associados, irão gerar o círculo virtuoso do desenvolvimento, permitindo à sociedade desfrutar de

níveis crescentes de bem-estar e de redução da pobreza. Isto pode caracterizar que o crescimento econômico está associado a elementos incentivadores da formação de poupança e de sua intermediação eficiente.

Uma continuada eficiência na intermediação de poupanças impulsionou o desenvolvimento da Europa e dos Estados Unidos a partir do final do século passado, através de um sofisticado mercado de capitais. É a existência de um novo mercado, caracterizado pelo vigor e flexibilidade para financiar a nova economia, que explica o maior período de prosperidade da economia americana.

Segundo Nobrega *et al.* (2000, p.6), pode-se observar três caminhos como forma de associar a poupança ao investimento: (1) o autofinanciamento; (2) o governo e (3) o financiamento via mercado de capitais. O autofinanciamento, em que as empresas geram internamente os próprios recursos, é o mais simples e primitivo. O governo, quando financia certas atividades usando a arrecadação de tributos ou a imposição de mecanismos compulsórios de poupança, é limitado e sujeito a distorções. De acordo com Pinheiro (2005, p.155), o financiamento via mercado financeiro e de capitais, é o mais eficiente.

Os investidores institucionais, como as seguradoras, os fundos de pensão e os fundos mútuos, gerenciam grandes recursos. Esses recursos financiam a compra de ações e de papéis de dívida emitidos pelos indivíduos e pelas empresas. Percebe-se que estes investidores influenciam transformações sociais, melhorando a situação econômica das famílias, que passam a ter maiores excedentes de consumo e a demandar alternativas confiáveis para aplicar suas poupanças. É o caso de alternativas que garantam a aposentadoria e a proteção do patrimônio, proporcionando liquidez e oferecendo retornos razoáveis. Dessa forma, percebe-se que nenhum mercado de capitais cresce de forma considerável sem a institucionalização da poupança.

Os mercados são responsáveis por viabilizar o aproveitamento das diversas oportunidades em toda a economia, uma vez que aproxima os dois agentes: o poupador e o tomador. Enquanto um tem excesso de recurso e não tem oportunidade de investi-lo em atividades produtivas, o outro está na situação contrária. Dessa forma, percebe-se que os mercados promovem um aumento geral da produtividade, da eficiência e do bem-estar da sociedade.

Além disso, a diluição do risco de novos investimentos pode ser alcançada através do mercado acionário, o qual constitui uma fonte de incentivo ao desenvolvimento econômico, uma vez que promove a socialização do capital, pulverizando a propriedade das empresas entre pequenos poupadores. Em contrapartida, as empresas tendem a procurar maneiras que proporcionem melhores retornos para os seus acionistas, tendo como objetivo a busca de melhores resultados econômicos, e não apenas interesses específicos.

A desintermediação, uma vez que reduz o número de participantes nas transações,

pode ser considerada como outra característica que favorece o sistema financeiro. Dessa forma, o processo em que um banco capta recurso e empresta a seus clientes, em um empréstimo bancário clássico, não tem tanto valor. Os investidores passam a comprar as ações diretamente das empresas, sem a presença do banco. Este processo reduz custos e aumenta a rapidez nas operações, proporcionando maior liquidez.

A partir dos anos de 1980, em função da tecnologia da informação, o sistema financeiro teve um crescimento considerável. O desenvolvimento das telecomunicações e da informática possibilitou a quebra das barreiras geográficas, levando a um fenômeno que se convencionou chamar de globalização. Esse fenômeno leva a uma execução rápida e segura da transferência e consequente gestão dos recursos.

Nos Estados Unidos, o mercado de capitais se tornou o centro do sistema financeiro, uma vez que possibilita o financiamento de segmentos importantes, como a agricultura e a construção civil. Recursos equivalentes a dois terços do PIB são movimentados pelo mercado hipotecário. Além disso, os mecanismos para abertura e aumento de capital das empresas cresceram nos anos mais recentes. Os processos de criação, financiamento e gestão de empresas foram revolucionados pelos *private equity funds* e *venture capitalists*.

No Brasil, dois aspectos são observados quando é feita uma análise do mercado de capitais: (1) o seu desenvolvimento é importante para o crescimento econômico; (2) os investidores institucionais assumiram papel relevante na última década.

Carvalho (2000) afirma que, apesar de haver muitas críticas em relação à deficiência do sistema bancário brasileiro, os bancos buscam maximizar seus retornos, enquanto instituição privada. Dado o cenário em que não se apresenta o *tradeoff* entre risco e retorno, as instituições não têm incentivos para diversificar suas carteiras, apresentando elevada concentração em financiamentos ao setor público. Segundo Kupfer e Hasenclever (2002), essa disfunção do sistema bancário, que leva o crédito a escassear, representa uma barreira ao investimento e à concorrência na economia.

O crescimento do volume negociado na Bolsa de Valores de São Paulo (BOVESPA) foi propiciado pela integração do Brasil com o mercado internacional. No entanto, descontando o efeito da estabilização, o crescimento parece pequeno diante do potencial de mercado. O valor de mercado de 380 empresas listadas somou R\$ 2,27 trilhões em fevereiro de 2010. Quando comparamos esse número com o volume negociado na Nasdaq observamos que o valor negociado na bolsa brasileira é bastante modesto representando apenas 1% do montante negociado na bolsa americana. No entanto, apesar da grande diferença de recursos, a taxa acumulada de crescimento ao longo dos últimos anos tem sido aproximadamente a mesma para ambas as bolsas.

Nos dados da World Federation of Exchanges, os números brasileiros confirmam a necessidade de ampliação do volume pois mesmo com todo o avanço ainda não passamos dos 40% do PIB. Tal realidade está distante de países com maior tradição no mercado de capitais, como os EUA, cujo volume já ultrapassou 110% do PIB, ou mesmo

o Chile, cujo percentual é duas vezes maior do que para o caso brasileiro.

Além disso, o mercado acionário brasileiro necessita de outros elementos para que possa participar do mercado global de forma mais efetiva. Segundo estudo da consultoria Tendências (2000), o mercado de capitais requer três itens básicos para seu desenvolvimento: liquidez, transparência e facilidade de acesso.

Além desses três itens, existe a necessidade de um trabalho de divulgação e de fortalecimento da imagem do sistema financeiro perante a população. Algumas ações têm sido feitas nesta esfera. Um exemplo é a Bovespa, que busca atrair o público em campanhas e divulgação em eventos em geral. Dados positivos têm sido observados em relação a estas ações, uma vez que a participação de pessoas físicas no volume negociado na Bovespa chegou a 31% do volume médio em janeiro de 2010. Os novos meios de transação que visam reduzir o custo bem como a estabilização da inflação, contribuíram positivamente para esta marca. Se considerarmos apenas o público feminino, em cinco anos, o número de investidoras aumentou em mais de sete vezes (647,74%), passando de 15.030, em 2002, para 112.386, ao final de 2007.

É válido ressaltar que durante a elaboração dos planos de ampliação para o mercado brasileiro de títulos, o baixo nível de renda e escolaridade da população brasileira devem ser observados. Da mesma forma, investimentos para o desenvolvimento de setores-chave da economia nacional, como é o caso da habitação, agricultura e infraestrutura, precisam ser feitos. Muitos desses setores estão incorporados à iniciativa privada pelo processo de privatização, contudo captam internacionalmente recursos via *funding* do mercado de capitais. Esta ação aumenta a importância e o desafio de ampliar e fortalecer as bases desse mercado no Brasil, para que possa assumir o papel que lhe cabe no desenvolvimento econômico de país.

Para que haja o crescimento do investimento e da economia brasileira é preciso mais do que apenas facilitar o acesso estrangeiro e controlar o processo inflacionário. Ações como estas não são suficientes para transformar o mercado de capitais em elemento fundamental para o crescimento da economia. As mudanças institucionais e legais são importantes. É preciso que haja um fortalecimento na integração com o sistema bancário e todo o setor produtivo de forma que este seja dinamizado e os custos transacionais sejam reduzidos.

A busca contínua da integração com outros mercados bem como uma melhoria nas condições de acesso e transparência farão com que os custos transacionais sejam reduzidos, proporcionando o aumento da liquidez. Isso atrairá novos agentes, aplicadores e até mesmo as empresas menores, hoje excluídas do mercado.

Diante do exposto, pode-se observar que as dificuldades para o desenvolvimento da bolsa brasileira têm natureza econômica, política e social. O mercado das ações reflete o comportamento inconstante da economia brasileira, caracterizado por problemas estruturais associados a crises conjunturais. Dessa forma, um dos grandes desafios nesse

cenário de incerteza é realizar previsões sobre preços de ações no mercado acionário brasileiro, pois é algo de fundamental importância para os agentes econômicos que nele operam uma vez que os auxiliam nas tomadas de decisões. No entanto, realizar essas previsões tem se tornado uma tarefa menos árdua nos últimos anos graças a estabilidade monetária alcançada após o ano de 1994 com a implementação do Plano Real que levou a uma estabilidade do nível geral de preços.

### 1.1.3 Estrutura da tese

A estrutura desta Tese é composta dos seguintes capítulos:

Capítulo 1 - Introdução: neste primeiro capítulo é apresentada a motivação, a justificativa e o objetivo da tese; então é feita uma discussão sobre a importância das bolsas de valores e sobre a necessidade de se realizar previsões de séries temporais.

Capítulo 2 - Previsão de Séries Financeiras: neste capítulo é realizada uma breve introdução ao problema de previsão de séries temporais financeiras apresentando uma discussão sobre a utilização de técnicas de Redes Neurais Artificiais e Modelos Econométricos não-lineares, assim como são apresentadas as séries temporais utilizadas como estudo de caso. Para cada uma das séries são discutidas as características e estruturas de correlações, gerando informação para a determinação dos retardos temporais relevantes a serem utilizados no processo de previsão;

Capítulo 3 - Métodos Tradicionais: neste capítulo é definido o problema de previsão de séries temporais. A partir desta definição são expostos os modelos estatísticos de Box & Jenkins (modelos lineares) e os modelos estatísticos não lineares mais utilizados na previsão de séries temporais; Depois são apresentados Modelos de Redes Neurais Artificiais: sendo introduzida a técnica de Redes Neurais Artificiais (RNAs), expondo os fundamentos básicos de forma a gerar condições suficientes para o entendimento da utilização desta técnica para aplicações de previsão; Por fim os Resultados experimentais com métodos tradicionais é apresentado. Nesta seção apresentamos uma análise dos resultados para as previsões realizadas pelos modelos econométricos e pelas redes neurais artificiais para as séries financeiras e também para as não-financeiras. Uma discussão sobre a limitação de uma modelagem combinada é apresentada.

Capítulo 4 - Modelo Combinado - Estimção e Resultados: neste capítulo é proposto a aplicação de uma metodologia formulada por Zou e Yang (2004) para o problema de previsão de séries temporais. Apresentamos a metodologia do modelo combinado assim como discutimos a sua implementação. Por se tratar de séries temporais foi utilizado um modelo combinado de Redes Neurais Artificiais e Modelos Econométricos lineares dado que os não-lineares têm resultados similares como apresentado na introdução. Resultados experimentais para esta metodologia são apresentados e comparados aos demais modelos;

Capítulo 5 - Conclusões: neste capítulo são apresentadas as conclusões dos procedimentos, técnicas e experimentos realizados. Uma discussão a respeito da metodologia proposta é realizada assim como são relacionadas e discutidas as limitações deste trabalho. Por fim são sugeridas propostas de novos trabalhos na mesma linha com possíveis ramificações do modelo implementado.

## 2 Previsão de Séries Financeiras

### 2.1 Previsão de Séries Temporais

Desde os primeiros trabalhos até 1920, a previsão de séries temporais ocorria via extrapolação de dados. Coube a Yule (1926), *apud* Ferson e Tucker (2002) pesquisando as manchas solares, propor a técnica autoregressiva e ser o primeiro a abordar o problema das regressões sem sentido e espúrias. A técnica autoregressiva era puramente linear e consistia em utilizar a soma ponderada das observações anteriores para determinar o valor previsto. Durante aproximadamente cinquenta anos, exceto pela aplicação do filtro adaptativo linear de Widrow e Hoff (1960), em estudos de previsão climática, o modelo baseado em um filtro autoregressivo acrescido do ruído foi praticamente o único a ser utilizado nesta área.

No final dos anos de 1960, George E. P. Box e G. M. Jenkins, publicaram vários trabalhos sobre a teoria de controle e de análise de séries temporais. Em 1970 publicaram o livro *Time series analysis, forecasting and control* (Box e Jenkins, 1970) apresentando uma metodologia para a análise de séries temporais, e em 1976 foi lançada a versão revisada desse livro (Box e Jenkins, 1976) que normalmente é a mais mencionada. Esse trabalho foi importante por reunir as técnicas existentes numa metodologia para construir modelos ARMA que descreviam com uma certa precisão e de forma parcimoniosa o processo gerador da série temporal.

Nas décadas de 1960 e 1970, os modelos de equações simultâneas foram mais utilizados dado que modelos refinados da economia americana baseados em equações simultâneas dominaram a previsão econômica. Entretanto, Sims (1980) sugeriu que a decisão sobre a escolha das variáveis era muito subjetiva. Ele achava que se há uma verdadeira simultaneidade entre um conjunto de variáveis, todas elas devem ser tratadas igualmente; não deve haver distinção a priori entre variáveis de entrada e saída. Foi com este espírito que Sims (1980) apresentou a classe de modelos lineares VAR (*vectors autoregressives*). Os conceitos base associados ao VAR são: dependência temporal; impacto dinâmico de um distúrbio aleatório; métodos de seleção de modelos (Akaike, Schwarz e outros).

No artigo de Granger e Newbold (1974) foi apresentada uma criteriosa análise sobre regressões espúrias. A regressão de uma variável sobre uma ou mais variáveis muitas vezes pode fornecer resultados sem sentido ou espúrios e uma maneira de se prevenir é testar se as séries temporais são cointegradas. Basicamente, cointegração significa que a combinação de duas ou mais séries individualmente não estacionárias pode resultar em uma série estacionária. O teste apresentado em Engle e Granger (1987) pode ser

utilizado para verificar se duas ou mais séries são cointegradas, ou seja, sugere se há ou não uma relação entre elas a longo prazo (equilíbrio). Neste mesmo trabalho foi proposto um mecanismo de correção de erro para conciliar o comportamento a curto prazo de uma variável com seu comportamento a longo prazo, surgindo a classe de modelos lineares VEC (*vector error correction*).

Em Johansen e Juselius (1990) foi sugerida uma complementação ao teste de Granger. Quando não existe cointegração, pode-se utilizar um modelo linear VAR. Caso contrário, aplica-se um modelo linear VEC. No artigo de Sims *et al.*, (1990) foi analisada a escolha do tamanho do retardo nos testes de raízes unitárias, em modelos VAR e testes de cointegração. Os modelos lineares ARMA, VAR e VEC são muito utilizados nas áreas financeira e econômica. No início dos anos de 1980, a comunidade estatística propôs alternativas aos modelos lineares de memória curta existentes (ARMA e VAR):

1. Modelos não-lineares na média condicional: em Tong e Lim (1980) foram propostos os modelos autoregressivos com regimes determinados por limiares (threshold VAR, TVAR) ou por uma função de transição suave (smooth transition VAR, STVAR) que têm os regimes definidos por uma variável observada e por uma função de transição e combina dois ou mais modelos lineares de uma forma não-linear; no artigo de Subba e Gabr (1984) foram apresentados os modelos bilineares; em Ozaki (1980) foram sugeridos os modelos EXPAR (exponential autoregressive).
2. Modelos não-lineares na variância condicional: Engle (1982) e Bollerslev (1986) apresentaram modelos não-lineares na variância (ARCH - *autoregressive conditional heteroskedasticity* e GARCH - *generalized ARCH*). O objetivo de Engle era descrever o comportamento persistente da volatilidade da série de retornos de um ativo. Nos modelos ARCH a variância condicional muda com o tempo enquanto a variância não condicional permanece constante. O objetivo de Bollerslev (1986) era a generalização dos modelos ARCH como o próprio nome sugere. Posteriormente surgiram variações destes modelos, como o IGARCH (*Integrated GARCH*) e o EGARCH (*Exponential GARCH*).

### 2.1.1 Previsão de Séries Financeiras

Um considerável número de estudos empíricos chamam a atenção para o fato de séries financeiras não verificarem com frequência as hipóteses básicas da econometria (MANDELBROT, 1963; FAMA, 1965; AKGIRAY E BOOTH, 1988; TUCKER E POND, 1988; Cavalcanti, 2003).

A existência da não-normalidade altera de maneira significativa os resultados obtidos uma vez que tratar essas séries como se apresentassem normalidade leva a truncar a série em questão, mantendo-se apenas os pontos que possuem as “propriedades desejáveis” enquanto que os valores extremos são descartados.

Uma grande parcela de estudos empíricos levam a considerar essas séries como sendo geralmente não-estacionárias, o que implica que essas séries podem ter uma raiz unitária de primeira ou de segunda ordem e que conseqüentemente o seu comportamento dinâmico pode ser representado pelo modelo de passeio aleatório. Logo, qualquer desenvolvimento de modelagem econométrica que não considerar a não-estacionariedade da série estudada corre o risco de chegar a resultados pouco confiáveis. Por outro lado, os problemas ligados à escolha de um portfólio ótimo de títulos levam a considerar com atenção a existência eventual de heteroscedasticidade e de autocorrelação dos resíduos nas séries de ativos financeiros. A presença de resíduos heteroscedásticos, isto é, de resíduos cuja variância condicional depende do tempo, conduz a estimadores não eficientes dos parâmetros e da matriz de covariância (HAMILTON, 1994), esta última sendo o indicador de risco nos modelos de escolha de portfólio.

Analogamente, a autocorrelação pode levar a problemas de estimação que têm repercussão sobre a matriz de covariância. Uma vez que os trabalhos empíricos identificaram a existência desses dois últimos fenômenos nas séries financeiras, diversos estudos teóricos foram desenvolvidos com o objetivo de propor modelos estocásticos para levá-los em conta (modelos do tipo ARCH a partir dos trabalhos de ENGLE (1982) e BOLLERSLEV (1986)).

Os modelos EARCH ou ARCH exponencial propostos por Nelson (1991) são uma extensão dos modelos ARCH e surgem como consequência da observação de que os rendimentos das ações são negativamente correlacionados com as variações da volatilidade desses rendimentos. Isso significa que a volatilidade tende a aumentar em reação à divulgação de más notícias e a diminuir no caso contrário. O modelo ARCH-M é outro modelo a ser citado devido a média condicional ser uma função da variância condicional.

A grande questão é saber se os modelos ARCH são capazes de tratar a *leptokurtosis* apresentada nas séries financeiras. Nesse particular, os trabalhos de Milhoj (1985) e Bollerslev (1986) mostram que a distribuição incondicional dos resíduos  $\epsilon_t$ , obtida a partir do modelo GARCH(p,q) com resíduos condicionais normais, apresenta uma *kurtosis* excessiva, em outros termos, essa distribuição incondicional é leptokúrtica (Cavalcanti, 2003).

O modelo do movimento browniano desenvolvido em 1900 pelo matemático Louis Bachelier na sua tese de doutorado é um caso particular dos processos L-estáveis, também denominados “Levy-estáveis”, “ $\alpha$ -estáveis” ou “Pareto-estáveis”. Deve-se enfatizar o caráter pioneiro do trabalho de Bachelier. Seu modelo, na realidade, foi redescoberto em 1905 por Einstein no campo da Física e em 1920 por Norbert Wiener no campo da Matemática. Esse modelo deve, no entanto, ser criticado em sua aplicação aos mercados financeiros, uma vez que ele supõe a normalidade de sua distribuição de probabilidade.

Numa distribuição normal o decréscimo dos pontos situados longe do centro acontece de forma rápida significando que as grandes variações são raras para esse tipo de distribuição. No entanto, a observação do comportamento dos mercados financeiros mostra que as grandes flutuações têm uma grande probabilidade de ocorrência.

A maioria dos estudos empíricos rejeita a hipótese de normalidade das séries financeiras - o que se deve sobretudo à *leptokurtosis* - particularmente no caso em que os intervalos de tempo são pequenos (BOUTILLIER, 1992). Do ponto de vista teórico, a crítica mais importante à utilização da distribuição normal do modelo do movimento browniano desenvolvido por Bachelier foi feita por Mandelbrot (1963), que propõe um modelo alternativo baseado na hipótese de que as mudanças nos preços dos ativos financeiros são L-estáveis e independentes (Cavalcanti, 2003).

A característica de *leptokurtosis* observada nas séries financeiras pode ser explicada através dos processos L-estáveis não-gaussianos. Essa explicação se dá no contexto dos fenômenos de “descontinuidade” e “concentração”, identificados por Mandelbrot, que os tratou separadamente em um primeiro momento. Assim, o modelo proposto em seu artigo de 1963 estuda a descontinuidade e o apresentado em Mandelbrot (1965) analisa a concentração. Posteriormente, os dois fenômenos foram unificados em um terceiro modelo, cuja primeira versão foi apresentada em Mandelbrot (1972) e a segunda em Mandelbrot (1997b). O principal problema com relação aos processos L-estáveis não-gaussianos reside na sua utilização empírica, em decorrência da característica de variância infinita. De um lado, essa característica torna impossível a realização de cálculos econométricos através de métodos que utilizam momentos finitos (como é o caso do método de “mínimos quadrados”). De outro lado, alguns estudos empíricos mostraram uma relação inversa entre periodicidade e *kurtosis*, implicando que a agregação temporal leva a aceitar mais facilmente a hipótese de normalidade, o que é igualmente incompatível com a característica de variância infinita (Cavalcanti, 2003).

A descontinuidade e concentração, eliminam a possibilidade de analisar essas séries através de modelos determinísticos lineares, uma vez que os teoremas básicos do cálculo das probabilidades não se aplicam a essa situação (a não verificação dos teoremas impede a aplicação do procedimento tradicional para lidar com a irregularidade e que consiste em primeiro identificar uma tendência e em seguida, a ela superpor as flutuações).

Essas questões levam às seguintes conclusões: por um lado até agora os modelos ARCH não foram capazes de explicar toda a *leptokurtosis*, e por outro, o modelo do movimento browniano por ser um processo gaussiano foi objeto de uma importante crítica, levando à proposta, por Mandelbrot (1963), de um modelo mais geral, o dos processos L-estáveis, para o tratamento dessa não-normalidade.

Como observou De Gooijer e Kumar (1992), no seu artigo de revisão sobre modelos não-lineares, não se conseguiu evidências claras em favor dos modelos não-lineares,

quando comparados aos modelos lineares, em termos de desempenho de previsão. Tal situação não mudou de forma significativa até os dias atuais como relata Clements *et al.* (2004).

Desta forma, dentre os modelos desenvolvidos até agora para o tratamento econométrico das séries financeiras, não se encontra um modelo capaz de lidar simultaneamente com todas as características estatísticas acima mencionadas (não-normalidade, não-estacionaridade, autocorrelação dos resíduos e heteroscedasticidade), particularmente com a não-normalidade e a heteroscedasticidade.

Logo, na perspectiva de previsão, existe uma grande motivação para a pesquisa de modelos não-lineares como destaca a edição especial do International Journal of Forecasting de abril/junho de 2004 – *Forecasting Economic and Financial Time Series Using Nonlinear Methods*, onde se espera que a sofisticação dos testes, seleção e estimação de modelos não-lineares venha a aperfeiçoar o desempenho na previsão.

Neste cenário, a Inteligência Artificial (IA) é vista como uma alternativa para a previsão de séries temporais, com métodos e procedimentos geralmente não-lineares, e que em termos práticos atuais não apresentam tantas dificuldades técnicas quando comparada aos modelos estatísticos. Muitas técnicas de IA podem ser aplicadas ao problema de previsão, destacando na literatura as técnicas de redes neurais artificiais, algoritmos genéticos e programação genética entre outras.

Muitos trabalhos atuais também exploram a combinação destas técnicas para a geração de soluções otimizadas com a criação de sistemas híbridos inteligentes (GOO-NATILAKE, 1995), compostos pela combinação de técnicas de IA de tal forma que o ponto forte de uma técnica compense o ponto fraco da outra.

A classe dos modelos não-lineares ficou mais popular nos últimos anos, seja porque os dados exibem não linearidades inequívocas, seja pela disponibilidade de classes de modelos não-lineares bem especificados. Por exemplo, os modelos não lineares GARCH e volatilidade estocástica (VE) já são utilizados com sucesso para estimar a volatilidade condicional (risco) em séries financeiras. Como no mercado financeiro retorno e risco estão fortemente relacionados, é importante investigar métodos para estimá-los que incluam não-linearidades. As redes neurais artificiais (RNAs) são candidatas naturais para realizar esta tarefa.

Entretanto, as RNAs só podem ser devidamente exploradas por meio de procedimentos refinados de análise e síntese, ou seja, os recursos de processamento devem ser aplicados na medida certa e na situação apropriada. Tem-se que avaliar os ganhos de desempenho na presença de incrementos de complexidade. A complexidade da implementação de um modelo via RNAs pode aumentar e tornar difícil encontrar a solução global ótima. Entretanto, nas competições patrocinadas pelo Santa Fé Institute, a classe de modelos neurais apresentou os melhores resultados para a previsão de séries temporais multivariadas não-lineares e variantes no tempo (WEIGEND e

GERSHENFELD, 1994).

O foco da questão passou então a incluir a possibilidade da obtenção de informações por meio de técnicas de sistemas dinâmicos não-lineares para auxiliar as metodologias estatísticas consagradas de previsão. Este fato veio a fortalecer a idéia de que deve prevalecer a *performance* de previsão em que, caso fenômenos dinâmicos estejam envolvidos, aspectos temporais estarão presentes.

Em seguida, com o crescimento da capacidade de processamento e de memória dos computadores, viabilizaram-se os estudos de séries temporais com grandes conjuntos de dados a partir de modelos mais complexos. Em Lapedes e Farber (1986) foi utilizado um perceptron multi-camadas (PMC) na predição de problemas populares na comunidade que estudava sistemas caóticos. Em Weigend *et al.* (1990) foi apresentada uma técnica para penalizar parâmetros extras de um PMC, resultando em um modelo mais parcimonioso e mais eficiente que um modelo econométrico correspondente. Utans e Moody (1991) desenvolveram uma metodologia que incluía uma parcela de erro para penalizar o número efetivo de parâmetros de um modelo não-linear qualquer. Refenes *et al.* (1993) propuseram um método para adicionar unidades de neurônios mostrando que este método superou um modelo ARMA equivalente na predição de taxa de câmbio.

Hoje em dia, com o incremento vertiginoso da velocidade de processamento dos computadores permitindo trabalhar Redes Neurais com um enorme número de neurônios o que permite em tese uma melhor capacidade de aprendizado, aliado ao desenvolvimento de *softwares* mais robustos para realizar as estimações dos modelos econométricos, muito se avançou no que diz respeito à qualidade das previsões feitas por esses métodos. No entanto, apesar desse avanço e dos inúmeros trabalhos sobre o tema, não existe ainda um entendimento de qual técnica seria mais apropriada para realizar previsões de séries financeiras.

Com base nestas informações e valendo-se, portanto, da importância que a Bolsa de Valores tem para o desenvolvimento sustentável da economia, a tese proposta se concentrará, assim, na utilização de um modelo combinado proposto por Zou e Yang (2004) para a previsão de séries temporais, usando para tal uma combinação de redes neurais artificiais e modelos econométricos. A idéia é que o ponto forte de uma técnica compense o ponto fraco da outra e então que um modelo formado pela combinação convexa destas consiga realizar melhores previsões.

Tem-se a hipótese de que através da análise da correlação, é possível reunir elementos suficientes para que se possa afirmar a possibilidade de se fazer previsão dos valores de ações através do relacionamento dela com seus valores do passado.

## 2.2 Caracterização das Séries Temporais Utilizadas

No presente texto temos como objetivo avaliar o desempenho das previsões, para as séries adotadas, através da modelagem via modelos lineares (modelos Box-Jenkins, por exemplo) e não-lineares (ARCH ou GARCH, por exemplo, usados na análise de séries financeiras). Em seguida tratamos as mesmas séries utilizando o enfoque de redes neurais. Iremos trabalhar com séries reais que apresentam características peculiares e relevantes no que tange a complexidade, representatividade, variabilidade e aplicação prática. Todas as séries foram coletadas do endereço [www.ime.usp.br/pam](http://www.ime.usp.br/pam) com exceção da série de emissão de  $\text{CO}_2$  coletada do endereço [www.co2now.org](http://www.co2now.org). Listamos, abaixo, as séries que serão analisadas.

- Série das Manchas Solares (sunspot);
- Série com a emissão de gás carbônico ( $\text{CO}_2$ );
- Série do índice Ibovespa;
- Série do preço de fechamento das ações do Bradesco;
- Série do preço de fechamento das ações do Itaúsa;
- Série do índice *Dow Jones Industrial Average*;
- Série do preço de fechamento das ações da Petrobras;

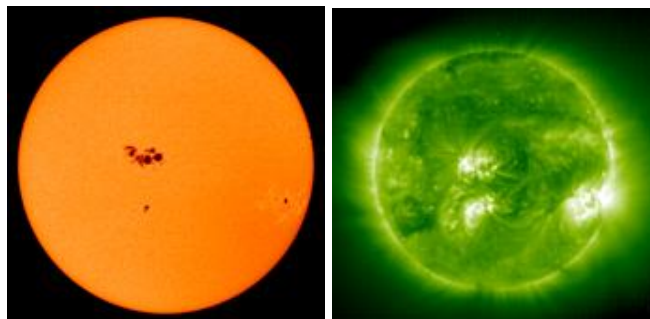
### 2.2.1 Série das Manchas Solares (*sunspot*)

A série de observações das manchas solares remontam desde a antiguidade. Há indícios de que os chineses foram os primeiros a observá-las, possivelmente feitas a olho nu, quando o Sol encontrava-se próximo ao horizonte ou em ocasiões de névoa. A partir do século XVII, quando surgiu a luneta, Galileu observou o Sol, e passou a tomar registros periódicos das manchas solares. As manchas solares são regiões escuras na superfície do Sol, como podem ser visto na Figura 1.

Um aspecto importante é que no centro dessas manchas escuras a temperatura chega a cair 38.09% em comparação com a temperatura da superfície do Sol, que é em torno de 5700 °K. Diversos trabalhos com a análise da série de Manchas Solares podem ser encontrados na literatura, entre os quais podemos destacar Lam *et al.* (2003); Leung *et al.* (2003) e Sfetsos e Siriopoulos (2004).

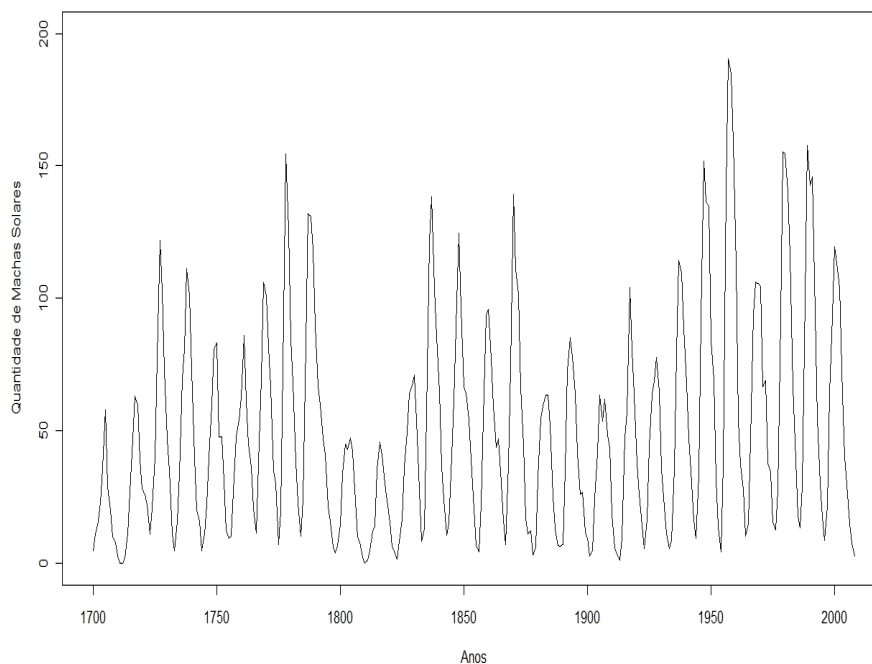
A série de manchas solares (sunspot), analisada nesta tese, refere-se à medidas anuais do período compreendidas de 1700 à 2008, totalizando 309 observações. A Figura 2 mostra o gráfico da série, onde podemos notar um comportamento não linear. A Tabela 1 apresenta medidas descritivas que caracterizam a série.

Figura 1 – Imagens do Sol



(Esquerda) Imagem do Michelson Doppler Imager (MDI) instrumento da nave espacial SOHO's ( *Solar and Heliospheric Observatory* ). (Direita) - Visão completa do Sol com SOHO's pelo telescópio EIT (*Extreme ultraviolet Imaging Telescope*). NASA

Figura 2 – Série de Manchas Solares (sunspot) - medidas anuais.



Fonte: elaborada pelo autor

Tabela 1 – Estatísticas descritivas da série de Manchas Solares (Sunspots). *Fonte: elaborada pelo autor.*

| Estatísticas      | Valores |
|-------------------|---------|
| Nº de observações | 309     |
| Máximo            | 190.20  |
| Mínimo            | 0.00    |
| Média             | 49.75   |
| Variância         | 1636.41 |
| Desvio Padrão     | 40.45   |

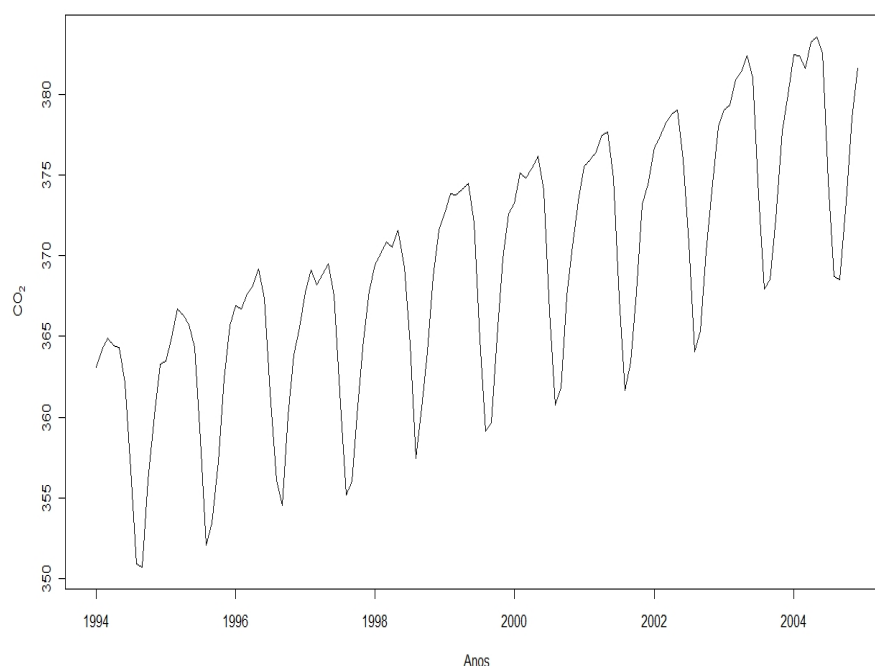
### 2.2.2 Série de emissão de dióxido de carbono - CO<sub>2</sub>

O dióxido de carbono é um composto químico constituído por dois átomos de oxigênio e um de carbono, cuja representação química é CO<sub>2</sub>. O nome comercial do CO<sub>2</sub> é gás carbono, cujo nome deve-se ao seguinte fato: a estrutura do dióxido de carbono é constituída de moléculas de geometria linear e de carácter apolar, dessa forma as atrações intermoleculares são bastante fracas, tornando-o, nas condições ambientais, um gás. Este gás é encontrado nos oceanos, rochas sedimentares e está presente também nos combustíveis fósseis. Muitos organismos no ecossistema e nos oceanos, como as plantas, absorvem o carbono encontrado na atmosfera na forma de dióxido de carbono através do processo de fotossíntese. Por outro lado, os organismos, como plantas e animais, liberam dióxido de carbono para atmosfera através do processo de respiração. Existe ainda o intercâmbio de dióxido de carbono entre os oceanos e a atmosfera por meio da difusão. A concentração de dióxido de carbono na atmosfera tem uma ligeira variação com a mudança das estações do ano, sofrendo influência principalmente do crescimento sazonal de plantas no Hemisfério Norte.

Nos últimos 10 anos o crescimento anual do CO<sub>2</sub> foi aproximadamente de 0.4%. Esse aumento foi impulsionado pela utilização de petróleo, gás, carvão e a destruição das florestas tropicais. A concentração de outros gases como metano e os clorofluorcarbonetos também vem aumentando rapidamente e desta forma contribuindo para o Efeito de Estufa. O efeito conjunto desses gases pode vir a causar um aumento na temperatura global estimado entre 2 a 6°C nos próximos 100 anos. Como consequências, teremos alterações do clima em nível mundial como também aumento do nível médio das águas do mar.

Nesta seção apresentaremos a série referente às medidas mensais relativas aos níveis de emissão de CO<sub>2</sub> no Canadá, no período de janeiro de 1994 à dezembro de 2004, com um total de 262 observações.

Na Tabela 2 mostramos as medidas que sumarizam informações referentes à série. A Figura 3 exhibe o comportamento dos níveis de CO<sub>2</sub> emitidos no Canadá, em que é perceptível características peculiares da série, que serão discutidas no próximo capítulo.

Figura 3 – Níveis de emissão de dióxido de carbono (CO<sub>2</sub>) no Canadá.

Fonte: elaborada pelo autor.

Tabela 2 – Estatísticas descritivas da série dos níveis de emissão de CO<sub>2</sub> no Canadá.

Fonte: elaborada pelo autor.

| Estatísticas      | Valores |
|-------------------|---------|
| Nº de observações | 132     |
| Máximo            | 383.58  |
| Mínimo            | 350.69  |
| Média             | 369.05  |
| Variância         | 59.51   |
| Desvio Padrão     | 7.71    |

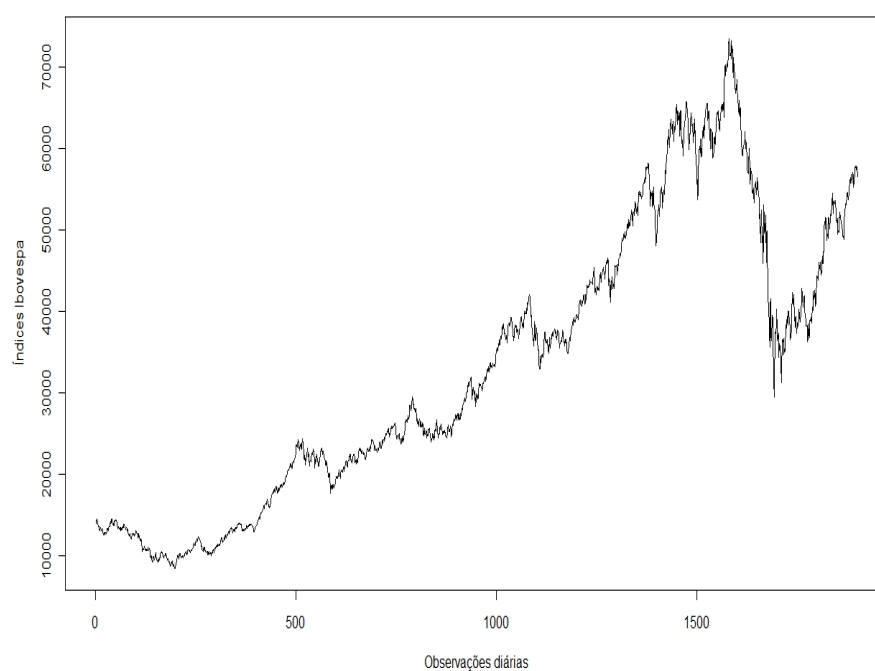
### 2.2.3 Série do índice do Ibovespa

O mais importante indicador do desempenho médio das cotações das ações negociadas na Bolsa de Valores de São Paulo é o Índice Bovespa (Ibovespa). A carteira teórica do índice Ibovespa é integrada pelas ações que, em conjunto, representaram 80% do volume transacionado à vista nos 12 meses anteriores à formação da carteira.

A série do índice Ibovespa, aqui utilizada, é constituída por cotações diárias, no período compreendido entre 02/01/2002 até 31/08/2009, totalizando 1902 observações. A análise para este conjunto de dados será dada de forma semelhante às séries apresentadas anteriormente, no sentido de que fizemos a divisão em dois grupos, Treinamento e Teste, com um total de 1712 e 190 observações cada. A Figura 4 mostra o comportamento

da série e na Tabela 3 apresentamos algumas estatísticas básicas do conjunto de dados em estudo. Por se tratar de uma série financeira a modelagem apresentada no capítulo seguinte será feita analisando-se os retornos.

Figura 4 – Série do índice do Ibovespa.



*Fonte: elaborada pelo autor.*

Tabela 3 – Estatísticas descritivas da série do índice Ibovespa. *Fonte: elaborada pelo autor.*

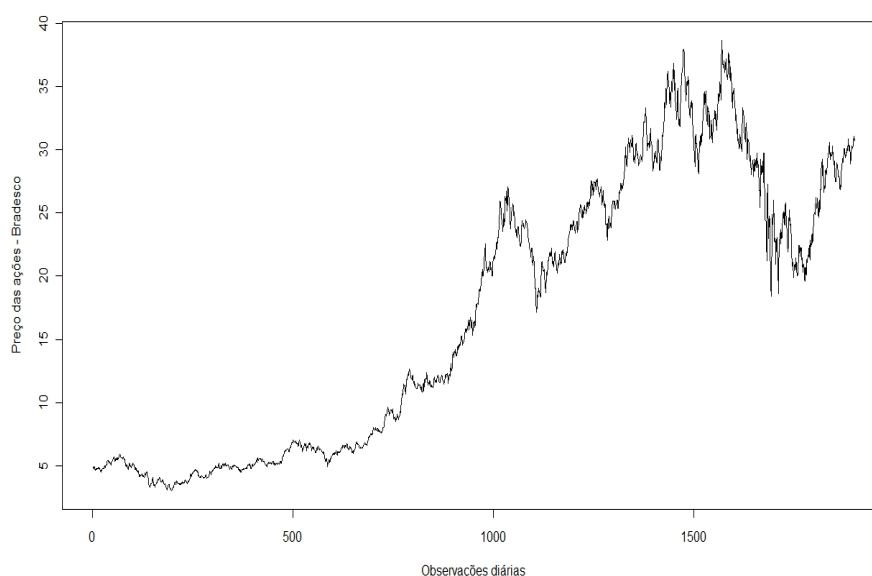
| Estatísticas      | Valores      |
|-------------------|--------------|
| Nº de observações | 1902         |
| Máximo            | 73516.00     |
| Mínimo            | 8370.00      |
| Média             | 33008.34     |
| Variância         | 294628981.91 |
| Desvio Padrão     | 17164.76     |

### 2.2.4 Série do Bradesco

O banco Bradesco foi criado em 1943, na cidade de Marília - São Paulo. Inicialmente foi denominado de Casa Bancária Almeida e depois Banco Brasileiro de Descontos S.A., cuja sigla é “Bradesco”. Tinha como filosofia inicial atrair o pequeno comerciante, funcionário público, diferentemente dos bancos da época, que eram voltados para os grandes proprietários de terras. Foi um dos pioneiros na estimulação do uso de cheques por parte dos correntistas. É o primeiro banco brasileiro a aparecer no ranking das cem mais valiosas marcas do mundo.

No presente texto iremos trabalhar a série do preço diário das ações do Bradesco, em que temos um total de 1902 observações, obtidas no período de 02/01/2002 à 31/08/2009. A Tabela 4 apresenta valores típicos da série, como também alguma estatísticas descritivas do conjunto de dados dos preços das ações do Bradesco. A Figura 5 exibe o comportamento dos dados.

Figura 5 – Série do preço diário das ações do Bradesco.



*Fonte: elaborada pelo autor.*

### 2.2.5 Série do valor de fechamento das ações da Petrobras

A Petrobras é uma empresa brasileira de prospecção, extração e refino de petróleo de grande destaque em nível nacional e internacional, principalmente no que tange a tecnologia de extração de petróleo em águas de alta profundidade, e por isso merece destaque no mercado de ações. Foi fundada no governo Getúlio Vargas, em 1953. Suas atividades foram iniciadas com o acervo recebido do antigo Conselho Nacional do Petróleo (CNP), que manteve sua função fiscalizadora sobre o setor. As ações da

Tabela 4 – Estatísticas descritivas da série do preço das ações do Bradesco. *Fonte: elaborada pelo autor.*

| Estatísticas      | Valores |
|-------------------|---------|
| Nº de observações | 1902    |
| Máximo            | 38.69   |
| Mínimo            | 3.00    |
| Média             | 16.87   |
| Variância         | 117.22  |
| Desvio Padrão     | 10.83   |

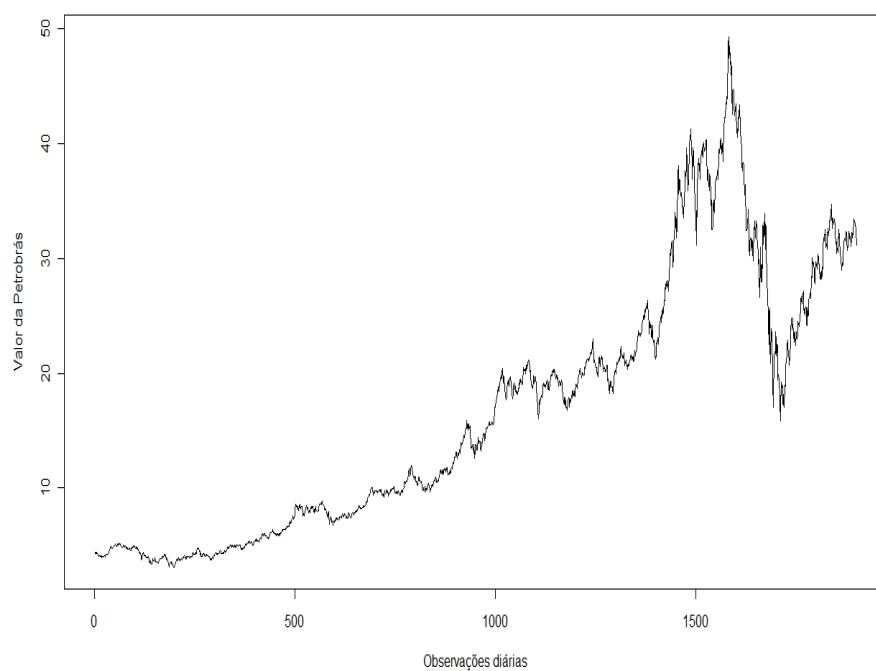
Petrobras têm ganhado grande valorização, nos últimos anos, sendo uma das empresas estatais de capital aberto mais lucrativa da América Latina. Recentemente, com a descoberta da camada Pré-Sal, houve valorização das ações relacionadas a Petrobras, entretanto a concretização dessas valorizações presentes e futuras dependem, em certa medida, dos avanços em termos de eficiência na extração do petróleo.

A série da Petrobras é constituída por observações diárias referentes ao valor de fechamento das ações da Petrobras (PetrobrásON - Código da Bovespa PETR3), no período de 2 de janeiro de 2002 à 31 de agosto de 2009, com um total de 1902 observações. Na Tabela 5 são mostrados os valores de algumas estatísticas básicas da série dos valores de fechamento das ações da Petrobras, já a Figura 6 mostra o gráfico da série.

Tabela 5 – Estatísticas descritivas da série do valor de fechamento das ações da Petrobras. *Fonte: elaborada pelo autor.*

| Estatísticas      | Valores |
|-------------------|---------|
| Nº de observações | 1902    |
| Máximo            | 49.34   |
| Mínimo            | 3.09    |
| Média             | 16.53   |
| Variância         | 124.46  |
| Desvio Padrão     | 11.16   |

Figura 6 – Gráfico da série ações da Petrobras.



*Fonte: elaborada pelo autor.*

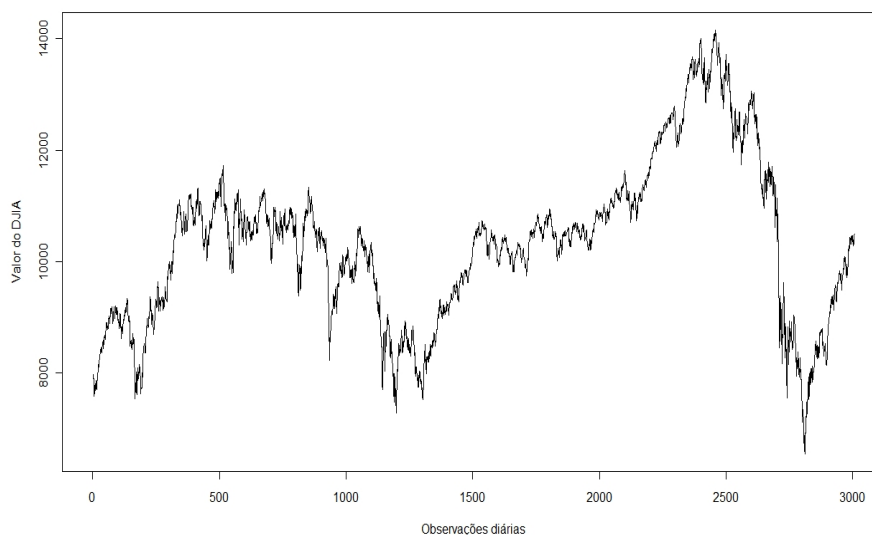
### 2.2.6 Série do índice *Dow Jones Industrial Average* (DJIA)

O índice *Dow Jones Industrial Average* (DJIA) criado por Charles H. Dow, em 1896, representava o valor médio das ações de 12 grandes empresas do setor industrial americano. Tinha como objetivo inicial dar uma ideia do valor médio da evolução e comportamento do mercado de ações, visto que esse índice ainda não estava consolidado. O índice DJIA é ao lado do NASDAQ Composite e do Standard & Poor's 500 um dos principais indicadores do mercado financeiro americano.

O DJIA é composto atualmente por trinta ações que representam todos os importantes setores industriais, formando, assim, um índice robusto capaz de refletir o comportamento das ações relativas ao parque industrial, não só americano, como também os inerentes às grandes multinacionais, seguindo a uma tendência de economia globalizada.

A série do índice *Dow Jones* utilizada na tese é constituída por observações diárias, compreendido no período de 01/01/1998 à 14/12/2009, totalizando 3007 observações. Na Figura 7 apresentamos o comportamento dos dados, já na Tabela 6 mostramos as estatísticas que sumarizam a série dos índices DJIA.

Figura 7 – Gráfico da série do índice *Dow Jones Industrial Average*



*Fonte: elaborada pelo autor.*

### 2.2.7 Série do Itaúsa

O Banco Itaú iniciou sua história em 1944, na cidade de Itaú de Minas na época município de Pratápolis, Minas Gerais. A primeira agência do banco foi aberta em janeiro de 1945.

Tabela 6 – Estatísticas descritivas da série do índice *Dow Jones Industrial Average*. Fonte: elaborada pelo autor.

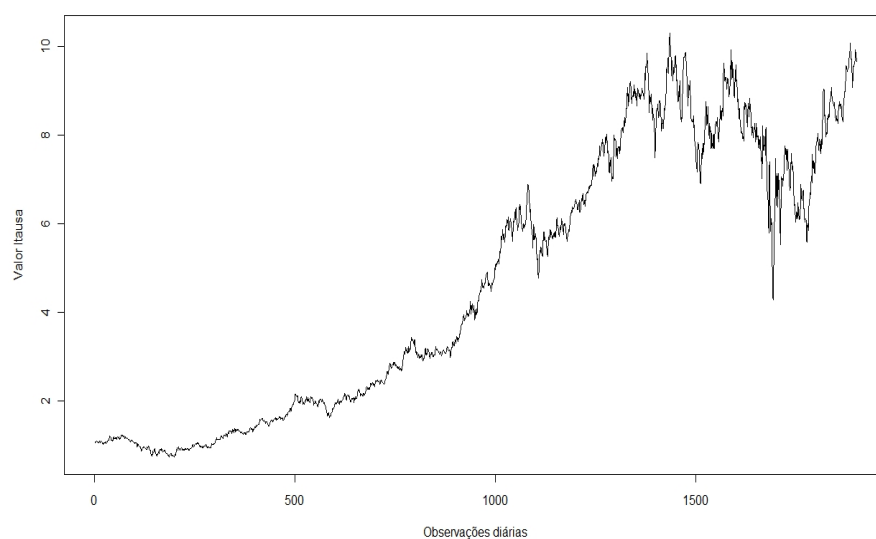
| Estatísticas      | Valores    |
|-------------------|------------|
| Nº de observações | 3007       |
| Máximo            | 14164.53   |
| Mínimo            | 6547.05    |
| Média             | 10319.37   |
| Variância         | 2074675.36 |
| Desvio Padrão     | 1440.37    |

Em 2008, com a fusão de suas operações financeiras com o Unibanco, tornou-se a maior instituição bancária privada do Brasil e do hemisfério sul, figurando ainda entre as vinte maiores do mundo. Na Tabela 7 são apresentadas as estatísticas descritivas e na Figura 8 é apresentado o gráfico da série dos preços das ações do Itaúsa no período de janeiro de 2002 a agosto de 2009. No capítulo seguinte apresentamos a modelagem para esse conjunto de dados.

Tabela 7 – Estatísticas descritivas da série do preço das ações do Itaúsa. Fonte: elaborada pelo autor.

| Estatísticas      | Valores |
|-------------------|---------|
| Nº de observações | 1902    |
| Máximo            | 10.32   |
| Mínimo            | 0.73    |
| Média             | 4.6386  |
| Variância         | 9.12    |
| Desvio Padrão     | 3.02    |

Figura 8 – Gráfico da série do preço das ações da Itausa.



*Fonte: elaborada pelo autor.*

## 3 Métodos Tradicionais

“Na econometria, quem modela muitas vezes se depara com dados provenientes de observações em oposição aos dados experimentais. Isto tem duas implicações importantes para a modelagem empírica. Primeiro, quem modela deve dominar habilidades muito diferentes das necessárias à análise de dados experimentais. . . . Segundo, a separação entre quem coleta dados e quem os analisa exige que quem modela esteja profundamente familiarizado com a natureza e estrutura dos dados em questão.” (Spanos, 1999)

Conforme a citação de Spanos, a econometria de séries temporais não pode ser feita mecanicamente, precisa de compreensão, intuição e habilidade. O objetivo deste capítulo é introduzir os conceitos básicos que estão diretamente associados às séries temporais utilizadas nesta tese, bem como fazer uma breve digressão acerca de algumas metodologias de modelagem.

### 3.1 Modelos Econométricos<sup>1</sup>

Uma série temporal é um conjunto de observações geradas sequencialmente no tempo, (Box e Jenkins, 1976). A característica mais importante deste tipo de dado é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Enquanto em modelos de regressão a ordem das observações não tem muita importância para a análise, em séries temporais a ordem dos dados é de extrema relevância. Dados de séries temporais surgem em vários campos do conhecimento, como na Economia (preços diários de ações, taxa mensal de desemprego e produção industrial), Medicina (eletrocardiograma e eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (precipitação pluviométrica, temperatura diária e velocidade do vento), dentre outros. O estudo de séries temporais requer o uso de notações e técnicas específicas.

As séries temporais podem ser contínuas ou discretas. É dita contínua quando as observações são feitas continuamente no tempo (por exemplo, registro de marés no porto de Suape). Neste caso, definindo o conjunto  $\Pi = \{t : t_1 < t < t_2\}$ , a série será denotada por  $\{y(t) : t \in \Pi\}$ . Contrariamente, uma série temporal é discreta quando as observações são realizadas em tempos específicos (por exemplo, índices diários da Bolsa de Valores de São Paulo). Definindo o conjunto  $\Pi = \{t_1, \dots, t_T\}$ , a série temporal discreta será denotada por  $\{y_t : t \in \Pi\}$ . Por simplicidade, podemos fazer  $\Pi = \{1, 2, \dots, T\}$ . Note que estes termos não se referem à variável observada  $y$ . Esta pode assumir valores

<sup>1</sup> As seções deste capítulo estão baseadas fundamentalmente em Morettin, 2006, a ideia é que a tese fique autocontida.

discretos ou contínuos. Na prática, muitos dados de séries temporais econômicas e financeiras são coletadas em períodos de tempo discretos. Dessa forma, utiliza-se séries temporais discretas ao invés de contínuas, sendo que  $y_t$  representará uma variável aleatória observável da série temporal num dado instante  $t$ .

Uma série temporal também pode ser multivariada. Neste caso, se  $k$  variáveis são observadas a cada tempo, denotamos a série por  $\{y_{1t}, \dots, y_{kt}, t \in \Pi\}$ . Assim, várias séries correlacionadas devem ser analisadas conjuntamente, ou seja, tem-se um vetor de observações para “cada unidade de tempo”.

### 3.1.0.1 Objetivos

De um modo geral, os principais objetivos numa análise de séries temporais são (Franses, 1998):

- Descrição: descrever propriedades da série, como por exemplo, o padrão de tendência, existência de variação sazonal ou cíclica, observações discrepantes (*outliers*), uma variância que se altera no tempo - heterocedasticidade condicional e não-linearidade. Tipicamente, uma série temporal econômica apresenta pelo menos duas ou três dessas características;
- Explicação: usar a variação em uma série para explicar a variação em outra série;
- Predição: prever valores futuros com base em valores passados. Aqui, assume-se que o futuro envolve incerteza, ou seja, as previsões não são perfeitas, embora tenhamos por objetivo reduzir os erros desta previsão;
- Controle: se os valores da série temporal medem a qualidade de um processo de manufatura, o objetivo, neste caso, poderia ser o controle do processo.

### 3.1.1 Modelos probabilísticos

Nesta seção, iremos descrever diversos modelos que se adequam aos dados de séries temporais. Referidos modelos são denominados de processos estocásticos. Matematicamente um processo estocástico pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de pontos  $\Pi$ , que pode ser contínuo ou discreto. Iremos denotar a variável aleatória no tempo  $t$  por  $y(t)$  no caso contínuo (usualmente  $-\infty < t < \infty$ ) e por  $y_t$  no caso discreto (usualmente  $t = 0, \pm 1, \pm 2, \dots$ ). O conjunto de possíveis valores do processo é denotado de espaço de estados e pode ser discreto ou contínuo.

Na análise de séries temporais a situação é bem diferente da maioria dos problemas estatísticos. Embora seja possível variar o tamanho da série observada, usualmente seria impossível fazer mais do que uma observação em cada tempo. Assim, tem-se apenas uma realização do processo estocástico e uma única observação da variável aleatória no tempo  $t$ . Uma maneira de descrever um processo estocástico é através da distribuição de probabilidade conjunta de  $y(t_1), \dots, y(t_k)$  para qualquer conjunto de tempos  $t_1, \dots, t_k$  e qualquer valor de  $k$ . Esta é uma tarefa extremamente complicada e na prática costuma-se descrever um processo estocástico através das funções média, variância e autocovariância. Estas funções são definidas a seguir para o caso contínuo, sendo que definições similares se aplicam ao caso discreto.

**Média**  $\mu(t) = E[y(t)],$

**Variância**  $\sigma^2(t) = Var[y(t)],$

**Autocovariância**  $\gamma(t_1, t_2) = E[y(t_1) - \mu(t_1)][y(t_2) - \mu(t_2)].$

Note que a função de variância é um caso especial da função de autocovariância quando  $t_1 = t_2$ . Momentos de ordem mais alta do processo também podem ser definidos mas raramente são utilizados na prática. Em geral, as funções  $\mu(t)$  e  $\gamma(t_1, t_2)$  são suficientes.

#### 3.1.1.1 Processos estacionários

Uma importante classe de processos estocásticos são os chamados processos estacionários. Uma série temporal é dita estritamente (fortemente) estacionária se a distribuição de probabilidade conjunta de  $y(t_1), \dots, y(t_k)$  é a mesma de  $y(t_1 + h), \dots, y(t_k + h)$ , ou seja, o deslocamento da origem dos tempos por uma quantidade  $h$  não tem efeito na distribuição conjunta, dependendo apenas dos intervalos entre  $t_1, \dots, t_k$ . Em particular, para  $k = 1$ , a estacionariedade estrita implica que a distribuição de  $y(t)$  é a mesma para todo  $t$ . Desta forma, se os dois primeiros momentos forem finitos, temos que

$$\mu(t) = \mu \quad \text{e} \quad \sigma^2(t) = \sigma^2$$

são constantes. Para  $k = 2$ , a distribuição conjunta de  $y(t_1)$  e  $y(t_2)$  depende apenas da distância  $t_2 - t_1$ , denominada de defasagem. A função de autocovariância  $\gamma(t_1, t_2)$  também depende apenas de  $t_2 - t_1$  e pode ser escrita como  $\gamma(h)$ , onde

$$\gamma(h) = E[y(t) - \mu][y(t+h) - \mu] = Cov[y(t), y(t+h)]$$

é chamado de coeficiente de autocovariância defasagem  $h$ . Note que o tamanho de  $\gamma(h)$  depende da escala em que  $y(t)$  está sendo mensurada. Por isto, padroniza-se a função de autocovariância dando origem a função de autocorrelação

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)},$$

que mede a correlação entre  $y(t)$  e  $y(t + h)$ .

Na prática, é muito difícil utilizar a definição de estacionariedade estrita e costuma-se definir estacionariedade de uma forma menos restrita. Um processo estocástico  $\{y(t), t \in \Pi\}$  é dito ser estacionário de segunda ordem ou fracamente estacionário se a sua média é constante e sua função de autocovariância depende apenas da defasagem, isto é

$$E[y(t)] = \mu \quad \text{e} \quad \text{Cov}[y(t), y(t + h)] = \gamma(h).$$

Nenhum outro tipo de suposição é feita a respeito dos momentos de ordem mais alta. Além disso, fazendo  $h = 0$ , temos  $\text{Var}[y(t)] = \gamma(0)$ , ou seja, a variância do processo. Importante destacar que em processos normais a distribuição fica completamente caracterizada pelo primeiro e segundo momentos, isto é, por  $\mu(t)$  e  $\gamma(t_1, t_2)$ . Neste caso, estacionariedade fraca implica em estacionariedade estrita. Além disso, estacionariedade forte implica em estacionariedade fraca desde que os dois primeiros momentos sejam finitos.

### 3.1.1.2 Função de autocorrelação

A função de autocorrelação parcial (acf) de um processo estocástico estacionário é uma ferramenta importante para analisar as propriedades de uma determinada série. Se um processo estocástico estacionário  $y(t)$  tem média  $\mu$  e variância  $\sigma^2$ , então

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\sigma^2},$$

e, portanto,  $\rho(0) = 1$ . Seguem ainda as seguintes propriedades facilmente verificáveis:

1. A correlação entre  $y(t)$  e  $y(t + h)$  é a mesma que entre  $y(t)$  e  $y(t - h)$ , ou seja  $\rho(h) = \rho(-h)$ ;
2.  $-1 < \rho(h) < 1$ ;
3. Embora um processo estocástico tenha uma estrutura de autocovariância única, em geral, o contrário não é verdadeiro. É possível encontrar vários processos com a mesma função de autocorrelação, o que dificulta ainda mais a interpretação do correlograma.

### 3.1.1.3 Alguns processos estocásticos

#### 3.1.1.3.1 Sequência aleatória

Um processo em tempo discreto é dito puramente aleatório se consiste em uma sequência de variáveis aleatórias  $u_t$  independentes e identicamente distribuídas (*iid*). Isto implica nas propriedades abaixo listadas:

1.  $E(u_t) = E(u_t | u_{t-1}, u_{t-2}, \dots) = \mu;$
2.  $Var(u_t) = Var(u_t | u_{t-1}, u_{t-2}, \dots) = \sigma_u^2;$
3.  $\gamma(h) = Cov(u_t, u_{t+h}) = 0, h = \pm 1, \pm 2, \dots$

Como a média e a função de autocovarianância não dependem do tempo, o processo é estacionário em segunda ordem. A função de autocorrelação é simplesmente

$$\rho(h) = \begin{cases} 1 & \text{se } h = 0 \\ 0 & \text{se } h = \pm 1, \pm 2, \dots \end{cases}$$

Um processo puramente aleatório pode ser útil, por exemplo, na construção de processos mais complicados. As propriedades acima podem ser entendidas como ausência de correlação serial e homoscedasticidade condicional (variância condicional constante).

## 3.1.1.3.2 Passeio Aleatório

Seja  $u_t$  um ruído branco com média  $\mu$  e variância  $\sigma_u^2$ . Um processo  $y_t$  é denominado de passeio aleatório se

$$y_t = y_{t-1} + u_t,$$

fazendo substituições sucessivas obtemos

$$\begin{aligned} y_t &= y_{t-2} + u_{t-1} + u_t \\ &= y_{t-3} + u_{t-2} + u_{t-1} + u_t \\ &\vdots \\ &= y_0 + \sum_{j=1}^t u_j, \end{aligned}$$

se iniciarmos o processo em  $y_0 = 0$ , temos que

$$\begin{aligned} E(y_t) &= \sum_{j=1}^t E(u_j) = t\mu, \\ \text{Var}(y_t) &= \sum_{j=1}^t \text{Var}(u_j) = t\sigma_u^2, \end{aligned}$$

além disso, a função de autocovariância é dada por

$$\text{Cov}(y_t, y_{t-j}) = \text{Cov}(u_1 + u_2 + \cdots + u_{t-j} + \cdots + u_t, u_1 + u_2 + \cdots + u_t) = (t-j)\sigma_u^2$$

,  
e, portanto, a função de autocorrelação fica

$$\rho_t(h) = \frac{t-j}{t}.$$

Como a média, a variância e as autocovariâncias dependem de  $t$ , este processo é não-estacionário. No entanto, é interessante notar que a primeira diferença de um passeio aleatório é estacionária, uma vez que

$$\Delta y_t = y_t - y_{t-1} = u_t.$$

### 3.1.1.4 Filtragem

Um procedimento para analisar séries com tendência é utilizar filtros lineares. Um filtro linear converte uma série  $y_t$  em outra  $x_t$  através da operação linear

$$x_t = \sum_{j=-q}^s a_j y_{t+j},$$

onde  $a_j$  é um conjunto de pesos. Além disso, como queremos estimar a média local, os pesos devem ser  $\sum_{j=-q}^s a_j = 1$ , garantindo que  $\min\{y_t\} < x_t < \max\{y_t\}$ . Neste caso, a operação é chamada média móvel. Estas, em geral, são simétricas com  $s = q$  e  $a_{-r} = a_r$ . Por exemplo, se  $s = q = 2$ , temos que

$$x_t = a_1 y_{t-1} + a_2 y_{t-2} + a_0 y_t + a_1 y_{t+1} + a_2 y_{t+2}.$$

O caso mais simples é quando todos os pesos  $a_j$  têm o mesmo valor e, devido à restrição de soma 1, segue que  $a_j = \frac{1}{2q+1}$ , para  $j = -q, \dots, q$ . Desta forma, o valor suavizado de  $x_t$  é dado por

$$x_t = \frac{1}{2q+1} \sum_{j=-q}^s y_{t+j}.$$

Qualquer que seja o filtro utilizado,  $x_t$  é uma estimativa da tendência no tempo  $t$  e  $y_t - x_t$  é uma série livre de tendência.

### 3.1.1.5 Diferenciação

Um tipo especial de filtro muito útil para remover uma componente de tendência polinomial consiste em diferenciar a série até que ela se torne estacionária. Para dados não sazonais, a primeira diferença é, em geral, suficiente para induzir estacionariedade aproximada. A nova série  $x_1, \dots, x_T$  é formada a partir da série original  $y_2, \dots, y_T$ , onde

$$x_t = y_t - y_{t-1} = \Delta y_t.$$

Note que isto nada mais é do que um filtro (assimétrico) com coeficientes 1 e -1. Diferenciação de primeira ordem é a mais utilizada, sendo que ocasionalmente uma diferenciação de segunda ordem pode ser requerida. Neste último caso, teremos

$$x_t = \Delta^2 y_t = \Delta(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}.$$

Além disso, a diferenciação pode ser útil como ferramenta exploratória. Observações discrepantes, por exemplo, podem ter um efeito distoante na série diferenciada e uma representação gráfica pode auxiliar na identificação destes pontos.

## 3.1.1.6 Correlograma

Uma das formas de checar se o modelo estimado fornece uma representação adequada dos dados é através de uma análise das autocorrelações residuais. Segundo Box e Pierce (1970) e McLeod (1978), as distribuições assintóticas das autocorrelações amostrais dos resíduos de um modelo ARMA  $(p, q)$  possuem distribuição  $N(0, 1)$ . Portanto, se as autocorrelações amostrais estiverem dentro do intervalo

$$IC_{aprox}(\rho, 95\%) = [-2/\sqrt{T}; 2/\sqrt{T}],$$

há indícios de que o modelo esteja corretamente especificado. O coeficiente de autocorrelação amostral de ordem  $j$  é definido como

$$r_j = \frac{\sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}}{\sum_{t=1}^T \hat{u}_t^2}.$$

Se alguma dessas autocorrelações estiver fora do intervalo de confiança, existem evidências amostrais contra a adequação do modelo.

## 3.1.1.7 Teste de Box Pierce

Ao invés de verificarmos as autocorrelações residuais individualmente, podemos testar se um grupo de autocorrelações é significativamente diferente de zero através das chamadas estatísticas  $Q$ . Para modelos ARMA, Box & Jenkins (1970) sugeriram o uso do teste de Box-Pierce para as hipóteses

$$\begin{cases} H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \\ H_1: \rho_j \neq 0 \text{ para algum } j \in \{1, 2, \dots, m\}, \end{cases}$$

sendo a estatística de teste dada por

$$Q = T \sum_{j=1}^m r_j^2.$$

Na prática, o número  $m$  de autocorrelações amostrais é tipicamente escolhido entre 10 e 30. Se o modelo ajustado for apropriado, então  $Q$  terá distribuição aproximadamente qui-quadrado com  $m - p - q$  graus de liberdade. Assim, valores grandes de  $Q$  fornecem indicação contra a hipótese de que as autocorrelações são todas nulas, em favor da hipótese de que ao menos uma delas é diferente de zero. O teste de Box-Pierce não tem bom desempenho em amostras pequenas, no sentido de que a distribuição se afasta da qui-quadrado. Vários testes alternativos foram sugeridos na literatura e o mais conhecido é o teste de Ljung-Box, onde a estatística de teste é dada por

$$Q_{LB} = T(T+2) \sum_{j=1}^m \frac{r_j^2}{T-j}.$$

Sua distribuição amostral também é aproximadamente qui-quadrado com  $m - p - q$  graus de liberdade.

## 3.1.1.8 Teste da raiz unitária

Um teste alternativo de estacionariedade que se tornou bastante popular é o teste da raiz unitária. Considere o seguinte modelo:

$$y_t = y_{t-1} + u_t, \quad (3.1)$$

em que  $u_t$  é um ruído branco e (3.1) é um auto-regressivo de primeira ordem (AR(1)). Se o coeficiente de  $y_{t-1}$  for de fato igual a 1, defrontamo-nos com o que é conhecido como problema da raiz unitária, isto é, o processo não é estacionário. Portanto, se estimamos a regressão

$$y_t = \rho y_{t-1} + u_t \quad (3.2)$$

e de fato verificarmos que  $\rho = 1$ , então, a variável estocástica  $y$  tem uma raiz unitária. A equação (3.2) é frequentemente expressa de uma forma alternativa, a saber:

$$\begin{aligned}\Delta y_t &= (\rho - 1)y_{t-1} + u_t \\ &= \delta y_{t-1} + u_t\end{aligned}\tag{3.3}$$

em que  $\delta = (\rho - 1)$ . Portanto, testar a hipótese  $\rho = 1$  é equivalente a testar se  $\delta = 0$ . Sob a hipótese nula, a estatística  $t$  é calculada de modo convencional e é conhecida como estatística  $\tau$ , cujos valores críticos foram tabulados por Dickey e Fuller com base em simulações de Monte Carlo. Na literatura especializada, o teste  $\tau$  é conhecido como teste Dickey-Fuller (DF).

Se o valor calculado da estatística  $\tau$  exceder o valor crítico  $\tau$  de  $DF$ , então, não rejeitamos a hipótese de que a série temporal seja estacionária. Caso contrário, a série temporal é não-estacionária. Por razões teóricas e práticas, o teste DF é aplicado substituindo a equação (3.1) pelas seguintes equações:

$$\begin{aligned}\Delta y_t &= \beta_1 + \delta y_{t-1} + u_t, \\ \Delta y_t &= \beta_1 + \beta_2 t + \delta y_{t-1} + u_t,\end{aligned}\tag{3.4}$$

em que  $t$  é a variável tempo ou tendência. Em cada caso, a hipótese nula é  $\delta = 0$ , ou seja, há raiz unitária. A diferença entre (3.1) e as duas outras regressões reside na inclusão da constante (intercepto) e do termo de tendência. Se o termo de erro  $u_t$  é autocorrelacionado, modificamos (3.4) como segue:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + u_t.\tag{3.5}$$

Quando o teste DF é aplicado em modelos do tipo (3.5), é designado de teste aumentado de Dickey-Fuller (ADF). A estatística de teste ADF tem a mesma distribuição assintótica que a estatística DF, de modo que podem ser usados os mesmo valores críticos.

Entretanto, o teste ADF sofre de problemas de baixo poder estatístico e distorções de tamanho de amostra. Nesse contexto Elliot, Rottemberg e Stock (1996) e Ng e Perron (2001) propõem o teste MADF, o qual é uma versão modificada do teste ADF e supera os problemas citados anteriormente. Essas modificações baseiam-se em dois aspectos centrais: (i) a extração de tendência em séries de tempo usando mínimos quadrados ordinários (OLS) é ineficiente, e (ii) uma adequada seleção para a ordem de defasagem do termo aumentado é imperativo de maneira a obter uma melhor aproximação para o verdadeiro processo gerador de dados.

No primeiro aspecto, (i), Elliot, Rottemberg e Stock (1996) extraem a tendência estocástica da série usando os mínimos quadrados generalizados (GLS). No que concerne ao segundo aspecto, (ii), Ng e Perron (2001) mostram que os critérios de informação tradicionais de seleção, i.e. Akaike (AIC) e Schwarz (SIC), tendem a selecionar baixos valores para a defasagem  $m$ , quando se tem uma grande raiz negativa no polinômio de médias móveis da séries, conduzindo os testes de raízes unitárias a sérias distorções. Desse modo, desenvolveu-se o critério modificado de informação de Akaike (MAIC) para a seleção de defasagem auto-regressiva, objetivando minimizar as distorções provocadas pela seleção inadequada da defasagem.

### 3.1.2 Modelagem de Box-Jenkins

Uma metodologia bastante utilizada na análise de modelos paramétricos é conhecida como abordagem de Box & Jenkins (1970). Esta metodologia consiste em ajustar modelos auto-regressivos integrados de média móveis,  $ARIMA(p, d, q)$ , e suas variações a um conjunto de dados. A ênfase está não na construção de modelos com equações únicas ou simultâneas, mas na análise das propriedades probabilísticas, ou estocásticas, das séries temporais econômicas segundo a filosofia de *deixar que os dados falem por si*. Diferentemente dos modelos de regressão nos quais  $y_t$  é explicado por  $k$  regressores  $x_1, x_2, x_3, \dots, x_k$ , os modelos do tipo Box-Jenkins permitem que  $y_t$  seja explicado por valores passados, ou defasados, do próprio  $y$  e dos termos de erro estocásticos.

É importante destacar que, para empregar o método Box-Jenkins, precisamos ter em mãos uma série temporal estacionária ou uma que possa se tornar estacionária após uma ou mais diferenciações. O motivo para se exigir a estacionariedade foi assim apresentado por Pokorny (1987):

“O objetivo do método Box-Jenkins é identificar e estimar um modelo estatístico que possa ser interpretado como tendo sido gerado pelos dados amostrais. Se esse modelo estimado for usado para fazer previsões, devemos pressupor que as características desse modelo são constantes ao longo do tempo e especialmente nos períodos futuros. Assim, a razão para exigir dados estacionários é que qualquer modelo que venha a ser inferido desses dados possa ele próprio ser interpretado como estacionário ou estável, oferecendo, assim, uma base válida para a previsão.”

A estratégia para construção do modelo é baseada em um ciclo iterativo e utiliza os próprios dados para a escolha da estrutura deste modelo. Inicialmente, é proposta uma classe de modelos, procede-se a identificação de um modelo particular com base na análise de autocorrelações, autocorrelações parciais e outros critérios; em seguida os parâmetros são estimados, e os resíduos do modelo ajustado são então avaliados. Caso o modelo selecionado não seja adequado, o ciclo é reiniciado. Em geral, os modelos mais convenientes são parcimoniosos, ou seja, contêm poucos parâmetros e mesmo

assim geram previsões precisas. Segundo Gujarati (2006), a modelagem ARIMA pelo método Box-Jenkins é mais arte que ciência; é necessário considerável habilidade para escolher o modelo ARIMA certo.

### 3.1.2.1 Modelos auto-regressivos

O modelo auto-regressivo de ordem  $p$ , denotado  $AR(p)$ , é definido como

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t,$$

onde  $\epsilon_t \sim RB(0, \sigma^2)$ ,  $\phi_1, \phi_2, \dots, \phi_p$  são parâmetros auto-regressivos e  $c$  é um parâmetro incluído para permitir que  $y_t$  tenha média diferente de zero. Assim, processos AR podem ser usados como modelos se for razoável assumir que o valor atual de uma série temporal depende do seu passado imediato mais um erro aleatório.

O modelo  $AR(p)$  pode ser reescrito de forma mais compacta usando o operador de defasagem, denotado por  $B$  e definido como  $B^n y_t = y_{t-n}$ , da seguinte maneira

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) y_t = c + \epsilon_t,$$

ou

$$\phi(B) y_t = c + \epsilon_t,$$

onde  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$  é um polinômio auto-regressivo. Se todas as raízes da equação  $\phi(z) = 0$  estiverem fora do círculo unitário, o processo  $AR(p)$  será estacionário. Neste caso, a média de  $y_t$  é

$$E(y_t) = \mu = \frac{c}{1 - \phi_1 - \phi_2 - \cdots - \phi_p},$$

e a  $j$ -ésima autocovariância é dada por

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}, j > 0. \quad (3.6)$$

Dividindo-se (3.6) por  $\gamma_0 = Var(y_t)$ , chega-se às “equações de Yule-Walker”:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \cdots + \phi_p \rho_{j-p}, j > 0. \quad (3.7)$$

Em um processo  $AR(p)$  não existe correlação direta entre  $y_t$  e  $y_{t-p-1}$ ,  $y_{t-p-2}$ , .... Consequentemente, todos os coeficientes de correlação parcial são nulos para  $j > p$ .

### 3.1.2.2 Modelos de médias móveis

Dizemos que  $y_t$  segue um processo de médias móveis de ordem  $q$ , denotado por  $MA(q)$ , se

$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

onde  $\epsilon_t \sim RB(0, \sigma^2)$ . Usando o operador de defasagem  $B$ , este processo pode ser reescrito como  $y_t = \mu + \theta(B)\epsilon_t$ , onde  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$  é um polinômio de médias móveis. A média da variável  $y_t$  é  $E(y_t) = \mu$  e sua variância é dada por

$$\text{Var}(y_t) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2).$$

As autocovariâncias são

$$\gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_q\theta_{q-j})\sigma^2, & j = 1, 2, \dots, q \\ 0, & \text{para } j = q+1, q+2, \dots \end{cases}$$

Como a média e a variância são constantes e  $\gamma_j$  não depende de  $t$  o processo é estacionário independente dos valores dos parâmetros. Contudo, nem sempre é único. A fim de se ter unicidade, imporemos uma condição adicional de invertibilidade que garante que existe um único processo  $MA(q)$  para uma dada função de autocorrelação. Um processo  $MA(q)$  é invertível quando todas as raízes da equação  $\theta(B) = 0$  estão fora do círculo unitário.

### 3.1.2.3 Modelos auto-regressivos e de médias móveis

Combinando os modelos AR e MA pode-se obter uma representação adequada com um número menor de parâmetros. Processos auto-regressivos médias móveis (ARMA) formam um classe de modelos muito úteis e parcimoniosos para descrever dados de séries temporais. O modelo  $ARMA(p, q)$  é dado por

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q},$$

onde  $\epsilon_t \sim RB(0, \sigma^2)$ . Note que, modelos AR ou MA podem ser obtidos como casos especiais quando  $p = 0$  ou  $q = 0$ , respectivamente. Usando o operador de defasagem, o modelo pode ser reescrito como

$$\phi(B)y_t = \theta(B)\epsilon_t,$$

onde  $\phi(B)$  e  $\theta(B)$  são os polinômios AR e MA usuais. A condição de estacionariedade para um processo  $ARMA(p, q)$  é a mesma de um processo  $AR(p)$  e a condição de invertibilidade é a mesma de um processo  $MA(q)$ .

Nota-se que as funções de autocorrelação e autocorrelação parcial ficam consideravelmente mais complicadas em processos ARMA. De um modo geral, para um processo  $ARMA(p, q)$  estacionário a função de autocorrelação tem um decaimento exponencial ou oscilatório após a defasagem  $q$  enquanto que a função de autocorrelação parcial tem o mesmo comportamento após a defasagem  $p$ .

### 3.1.2.4 Modelos ARIMA e SARIMA

Os modelos discutidos até agora são apropriados para descrever séries temporais estacionárias. Assim, para ajustar estes modelos à uma série temporal observada é necessário remover as fontes de variação não estacionárias. Por exemplo, se a série observada for não estacionária na média pode-se tentar remover a tendência tomando-se uma ou mais diferenças. Um modelo ARMA no qual  $y_t$  é substituído pela sua  $d$ -ésima diferença  $\Delta^d y_t$  é capaz de descrever alguns tipos de séries não estacionárias. O processo auto-regressivo integrado médias móveis ARIMA( $p, d, q$ ) é dado por

$$\phi(B)(1 - B)^d(y_t - \mu)\theta(B)\epsilon_t. \quad (3.8)$$

O modelo (3.8) supõe que a  $d$ -ésima diferença da série  $y_t$  pode ser representada por um modelo ARMA, estacionário e invertível. Um processo que se torna estacionário após  $d$  diferenças é dito ser não estacionário homogêneo, ou integrado de ordem  $d$ ,  $I(d)$ . Na prática, valores pequenos são em geral especificados para  $d$ , sendo  $d = 1$  o valor mais utilizado.

Com o objetivo de levar em consideração padrões de flutuações sazonal, ampliou-se a classe de modelos ARIMA. Muitas vezes não é possível transformar  $y_t$  de forma a remover a sazonalidade, ou seja, a própria sazonalidade pode apresentar um padrão dinâmico. Isto significa que há necessidade de se considerar uma sazonalidade estocástica e ajustar à série original um modelo ARIMA sazonal, que passar a ser denominado de SARIMA. Seja  $y_t$  a série de interesse observada com período de sazonalidade  $s$ . A classe de modelos sazonais multiplicativos  $(p, d, q) \times (P, D, Q)$  é dada por

$$\phi(B)\Phi(B^s) \left[ (1 - B)^d(1 - B^s)^D - \mu \right] y_t = \theta(B)\Theta(B^s)\epsilon_t,$$

onde  $\theta(B^s) = 1 + \theta_1 B^s + \theta_2 B^{2s} + \dots + \theta_Q B^{Qs}$ ,  $\Phi(B^s) = 1 + \Phi_1 B^s + \Phi_2 B^{2s} + \dots + \Phi_P B^{Ps}$  e  $D$  indica o número de “diferenças sazonais”.

### 3.1.2.5 Critérios de identificação de $p$ e $q$

A forma mais utilizada para a identificação dos valores de  $p$  e  $q$  em um processo ARMA( $p, q$ ), são os critérios de informação apresentados a seguir:

1. Para modelar processos AR puros:

$$FPE = \frac{T + p}{T - p} \tilde{\sigma}^2;$$

$$\text{onde } \tilde{\sigma}^2 = \sum_{j=1}^T \frac{e_j^2}{T}.$$

## 2. Critério de Informação de Akaike(AIC)

$$AIC = -2\log(\hat{L}) + 2(p + q);$$

onde  $\hat{L}$  é a verossimilhança estimada.

## 3. Critério de Akaike Corrigido(AICc)

$$AICc = -2\log(\hat{L}) + \frac{2(p + q)}{T - p - q - 1};$$

## 4. Critério Bayesiano(BIC)

$$BIC = -2\log(\hat{L}) + (p + q) \log(T).$$

Destaca-se que o  $AIC$  não é consistente e que, mesmo assintoticamente, superestima a dimensão do modelo. Por outro lado, o  $BIC$  é consistente, isto é

$$\text{plim}(\hat{p}) = p \quad \text{e} \quad \text{plim}(\hat{q}) = q.$$

Para  $T > 8$ , o  $BIC$  penaliza a inclusão de novos termos no modelo mais pesadamente que o  $AIC$ . No que diz respeito ao  $AICc$ , este é uma versão corrigida do  $AIC$  e assintoticamente os dois critérios são equivalentes.

## 3.1.3 Heterocedasticidade condicional

Campbell, Lo, e MacKinlay (1997) argumentam que “é logicamente inconsistente como estatisticamente ineficiente utilizar medidas de volatilidade que são baseadas na suposição de volatilidade constante ao longo do tempo”. No caso de dados financeiros, por exemplo, grandes e pequenos erros tendem a ocorrer em clusters (agrupamentos), ou seja, grandes retornos são seguidos por grandes retornos e pequenos retornos por pequenos retornos. Assim, temos que os retornos são serialmente correlacionados. Em modelos de séries temporais para séries financeiras devemos fazer uso de modelos não-lineares, tais como ARCH, GARCH, EGARCH, TAR, contudo no presente trabalho abordaremos apenas os modelos ARCH e GARCH. A seguir, fazemos a distinção entre séries temporais lineares e não-lineares

- (a) Séries Temporais Lineares: os choques são assumidos como sendo não correlacionados mas não independentes e identicamente distribuídos (*iid*).
- (b) Séries Temporais Não Lineares: choques são assumidos *iid*, mas existe uma função não linear relacionando a série temporal observada e os choques subjacentes.

### 3.1.4 Modelos ARCH

O modelo ARCH (Autoregressive Conditional Heteroskedasticity), ou seja, modelo autoregressivo com heterocedasticidade condicional foi proposto inicialmente por Engle (1982). Desde então, têm aparecido inúmeros estudos com refinamentos e modificações do modelo ARCH básico e também com aplicações empíricas em taxa de câmbio, mercado de ações e muitos outros tipos de ativos Bollerslev et al. (1994). Os modelos ARCH permitem que os erros sejam não correlacionados, porém dependentes, ou seja,  $\text{Cov}(r_t^2, r_{t-1}^2)$  pode ser não nula.

**Definição 1.1** Um modelo ARCH( $m$ ) é definido por

$$r_t = \sqrt{h_t} \varepsilon_t \quad (3.9)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2, \quad (3.10)$$

em que  $\varepsilon_t$  são *iid* com média zero,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, m-1$ ,  $\alpha_m > 0$ . As distribuições que geralmente são usadas para  $\varepsilon_t$ , são  $\varepsilon_t \sim \mathcal{N}(0,1)$ ,  $\varepsilon_t \sim t_v$ , vide Morettin (2008). Contudo em séries financeiras é razoável o uso de distribuições com caudas mais pesadas em comparação com a distribuição normal. Apresentaremos a seguir algumas propriedades dos modelos ARCH, considerando-se o caso de  $m = 1$ , assim temos que o modelo é dado

$$r_t = \sqrt{h_t} \varepsilon_t \quad (3.11)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2, \quad (3.12)$$

em que  $\alpha_0 > 0$  e  $\alpha_1 > 0$  e  $\varepsilon_t$  é um ruído branco. Dinardo et al. apresentam as seguintes propriedades para os modelos ARCH:

- (i)  $\mathbb{E}(r_t) = 0$ , ou seja,  $r_t$  tem média zero;

*Prova*

$$r_t = \varepsilon_t \sqrt{\alpha_0 + \alpha_1 r_{t-1}^2}$$

$$\mathbb{E}_{t-1}(r_t) = \underbrace{\mathbb{E}_{t-1}(\varepsilon_t)}_0 \sqrt{\alpha_0 + \alpha_1 r_{t-1}^2}$$

$$\mathbb{E}_{t-1}(r_t) = 0$$

$$\mathbb{E}_{t-2} \mathbb{E}_{t-1}(r_t) = 0$$

$$(\dots)$$

$$\mathbb{E}(r_t) = 0$$

- (ii)  $r_t$  tem variância condicional dada por  $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$ ;

*Prova*

$$\begin{aligned}
r_t^2 &= \varepsilon_t^2 [\alpha_0 + \alpha_1 r_{t-1}^2] \\
\mathbb{E}_{t-1}(r_t^2) &= \sigma_\varepsilon^2 [\alpha_0 + \alpha_1 r_{t-1}^2] \\
\mathbb{E}_{t-1}(r_t^2) &= 1 [\alpha_0 + \alpha_1 r_{t-1}^2] \\
\mathbb{E}_{t-1}(r_t^2) &= \sigma_t^2
\end{aligned}$$

(iii)  $r_t$  tem variância incondicional dada por  $\sigma^2 = \frac{\alpha_0}{1-\alpha_1}$ ;

*Prova*

$$\begin{aligned}
r_t^2 &= \varepsilon_t^2 [\alpha_0 + \alpha_1 r_{t-1}^2] \\
\mathbb{E}_{t-2}\mathbb{E}_{t-1}(r_t^2) &= \mathbb{E}_{t-2}[\alpha_0 + \alpha_1 r_{t-1}^2] = \alpha_0 + \alpha_1 \mathbb{E}_{t-2}(r_{t-1}^2) = \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 r_{t-2}^2 \\
\mathbb{E}_{t-3}\mathbb{E}_{t-2}\mathbb{E}_{t-1}(r_t^2) &= \mathbb{E}_{t-3}[\alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 r_{t-2}^2] = \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 \mathbb{E}_{t-3} r_{t-2}^2 \\
\mathbb{E}_{t-3}\mathbb{E}_{t-2}\mathbb{E}_{t-1}(r_t^2) &= \alpha_0 + \alpha_0 \alpha_1 + \alpha_0 \alpha_1^2 + \alpha_1^3 r_{t-3}^2 \\
(\dots) \\
\mathbb{E}_0\mathbb{E}_1\mathbb{E}_2, \dots, \mathbb{E}_{t-2}\mathbb{E}_{t-1}(r_t^2) &= \alpha_0 [1 + \alpha_1 + \alpha_1^2 + \dots + \alpha_1^{t-1}] + \alpha_1^t r_0^2 = \frac{\alpha_0}{1-\alpha_1} = \sigma^2. \text{ Logo,} \\
&\text{n\~ao-condicionalmente o processo \u00e9 homoced\u00e1stico.}
\end{aligned}$$

(iv)  $r_t$  tem auto-covari\u00e2ncias zero;

*Prova*

$$\mathbb{E}_{t-1}(r_t r_{t-1}) = r_{t-1} \mathbb{E}_{t-1}(r_t) = 0$$

No que tange a curtose, que ser\u00e1 chamada de  $K$ , Higgins et al. (1993) e Morettin (2008) mostram que o processo tem uma cauda mais pesada do que a distribui\u00e7\u00e3o Normal, sendo que:

$$K = 3 \left( \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} \right) > 3$$

### 3.1.5 Modelos GARCH

Bollerslev (1986) sugeriu uma generaliza\u00e7\u00e3o dos modelos ARCH, que foi denominada de GARCH, “generalized ARCH”. De forma semelhante aos modelos ARMA, temos que os modelos GARCH \u00e9 mais parcimonioso no sentido de apresentar menos par\u00e2metros em compara\u00e7\u00e3o com o modelo ARCH quando usado para descrever a volatilidade.

**Defini\u00e7\u00e3o 1.2** Um modelo GARCH(m,n) \u00e9 definido por

$$r_t = \sqrt{h_t} \varepsilon_t \quad (3.13)$$

$$h_t = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + \sum_{j=1}^n \beta_j h_{t-j}^2, \quad (3.14)$$

em que  $\varepsilon$  são variáveis aleatórias *iid*, com média zero,  $\alpha > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, m-1$ ,  $\beta_j \geq 0$ ,  $j = 1, \dots, n-1$ ,  $\alpha_m > 0$ ,  $\beta_n > 0$ ,  $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$ ,  $q = \max(m, n)$ . Como no caso dos modelos ARCH, as distribuições geralmente usadas para  $\varepsilon_t$  são as distribuições normal e *t*-Student.

### 3.1.6 Avaliação de previsão

Em séries temporais, um dos principais objetivos é fazer previsões de valores futuros. Para tanto, é necessário escolher o melhor método de previsão e avaliar a precisão das previsões futuras provenientes do método selecionado. Podemos citar alguns trabalhos escritos sobre o desempenho de previsões, tais como: Diebold e Mariano (1995), Anders (2006), Barucci e Reno (2002). As medidas de performance de previsões são usualmente definidas em termos dos erros de previsão, ao qual podemos definir da seguinte maneira,

$$e_j = y_j - \hat{y}_j \quad (3.15)$$

onde  $\hat{y}_j$  denota a previsão de  $y_j$ . A capacidade de previsão de um modelo estimado pode ser avaliada através da utilização das medidas descritas a seguir:

A partir da equação (3.15) obtemos a medida mais utilizada na avaliação do desempenho de previsão,

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2, \quad (3.16)$$

em que  $n$  denota a quantidade de passo que previstos para a série temporal.

Avaliar previsões a partir do MSE (Mean Square Error) pode trazer alguns transtornos, no sentido de que esta medida não exhibe, de forma clara, o comportamento do modelo de previsão, ver Clements e Hendry (1993). Nesse contexto, para uma melhor compreensão e avaliação dos modelos de previsão outras medidas de desempenho também serão utilizadas.

Iremos adotar o Erro Percentual Médio Absoluto (Mean e Percentual Error - *MAPE*),

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{y_t} \right|, \quad (3.17)$$

em que  $e_t$  é o erro de previsão dado pela equação (3.15) e  $y_t$  é o valor da série no instante  $t$ .

Uma outra medida de desempenho de previsão é a estatística U de Theil, ver Steurer e Hann (1996), que toma como base o *MSE* do preditor normalizado pelo erro de

previsão de um modelo do tipo random walk. Um modelo random walk assume que o melhor valor de previsão para o tempo  $t + 1$  é o valor obtido no tempo  $t$ . A estatística U de Theil é dada por,

$$Theil = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=2}^n (\hat{y}_t - \hat{y}_{t-1})^2}, \quad (3.18)$$

assim, temos que se  $Theil = 1$  o modelo tem um desempenho igual a um modelo do tipo random walk; se  $Theil > 1$  o desempenho é inferior a um modelo do tipo random walk e se  $Theil < 1$  o preditor terá um desempenho superior a um modelo do tipo random walk. Em um modelo ideal a estatística U de Theil deve tender a zero.

A medida POCID (Prediction on Change in Direction) é usada para avaliar se há alteração na direção da previsão, ou seja, se o valor futuro irá aumentar ou diminuir em relação ao valor atual. Em suma, podemos dizer que essa medida fornece o percentual de acerto quanto à tendência da série, é definida por,

$$POCID = 100 \times \frac{\sum_{t=1}^n D_t}{n}, \quad (3.19)$$

em que,

Uma outra medida de desempenho relativo é o ARV (Average Relative Variance), que é dada por,

$$ARV = \frac{1}{n} \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{\sum_{t=1}^n (\hat{y}_t - \bar{y}_t)^2}, \quad (3.20)$$

em que  $n$  é a quantidade de pontos preditos,  $y_t$  é o valor real da série no instante  $t$ ,  $\hat{y}_t$  é o valor estimado sob o modelo adotado e  $\bar{y}_t$  é a média da série temporal. Através de  $ARV$  verificamos o ganho de desempenho de previsão do modelo relativo a uma previsão da média da série, onde se  $ARV = 1$ , temos que o modelo é equivalente a se tomarmos a média da série e usarmos esta como previsão. Se  $ARV > 1$ , o modelo é inferior a previsão realizada pela média da série e se  $ARV < 1$ , então o modelo tem um desempenho de previsão superior quando comparado com a previsão através da média da série. Em um modelo ideal  $ARV$  deve tender a zero.

Conforme Spanos, o processo de modelagem requer cuidado redobrado por parte de quem ajusta os modelos e, por mais esforço que os econométristas façam, os problemas relacionados às estimações dos modelos continuam levando a se buscar outras formas de realizar previsões.

## 3.2 Modelos de Redes Neurais Artificiais

No início da década de 40 surgiram os primeiros conceitos de máquinas baseadas no comportamento das células nervosas biológicas. Em 1943, Warren McCulloch e Walter Pitts apresentaram a primeira discussão sofisticada sobre *neuro-logical network*.

Em 1947, publicaram: “How we know universals”, seu segundo estudo sobre o tema. A partir destas idéias, surgiu o movimento intelectual chamado “Cybernetics”, o qual tentou combinar conceitos de biologia, psicologia, engenharia e matemática. A década terminou com a publicação do livro de Donald Hebb “The Organization of Behavior”.

A era “Cybernetics” abriu a perspectiva de se construir máquinas baseadas no conhecimento e no início da década de 50, os primeiros protótipos neste campo, com arquiteturas específicas para desempenhar tarefas determinadas começaram a acontecer. Contudo, instigados pelo fato dos animais aprenderem a fazer muitas coisas para as quais não tinham sido “construídos”, passou-se a ter como meta a construção de máquinas capazes de aprender. Com isso, a década de 50 presenciou o surgimento de muitos sistemas que aprendiam baseados em reforço e em formas simples de aprendizado e, o primeiro deles, foi provavelmente feito por Minsky em 1951. No final dos anos 50, o campo de pesquisa em redes neurais estagnou-se, sem nenhuma descoberta de peso na área, enquanto que importantes avanços tinham sido alcançados em sistemas especialistas.

Em 1962, com a publicação do livro de Frank Rosenblatt “Principles of Neurodynamics”, ressurgiu o interesse pela RNAs.

Rosenblatt definiu máquinas denominadas “perceptrons”, e um rol de teorias sobre as mesmas foi provado. Entusiasmados com as idéias, muitos pesquisadores passaram a crer que os “perceptrons” seriam a base para uma inteligência artificial. Contudo constatou-se que essas redes falhavam na solução de alguns problemas simples, similares a outros nos quais elas tinham sucesso. O rápido entusiasmo foi interrompido em 1969 com o livro de Marvin Minsky e Seymour Papert chamado “Perceptrons”, onde eles provaram que as redes neurais de uma única camada de Rosenblatt eram incapazes de resolver alguns problemas elementares como o do *ou-exclusivo*. Como Minsky possuía grande prestígio, seu livro alcançou credibilidade tal que à quase totalidade da comunidade científica abandonou as pesquisas na área por um longo período.

Um silêncio quase geral na área marca a década de 70 e o início da de 80. Poucos pesquisadores como Grossberg, Kohonen, Anderson, Hopfield, Marr, von der Malsburg e Copper continuaram estudando e desenvolveram trabalhos interessantes, principalmente nas áreas de memória associativa endereçável pelo conteúdo, sistema visual e reformulações do problema de aprendizado das redes. Em 1982, Hopfield deu uma importante contribuição com a introdução da função energia e pela ênfase da noção de memórias como atratores dinamicamente estáveis. Entretanto, talvez a maior influência desta época ainda tenha vindo dos “perceptrons” de Rosenblatt pois, baseados em suas

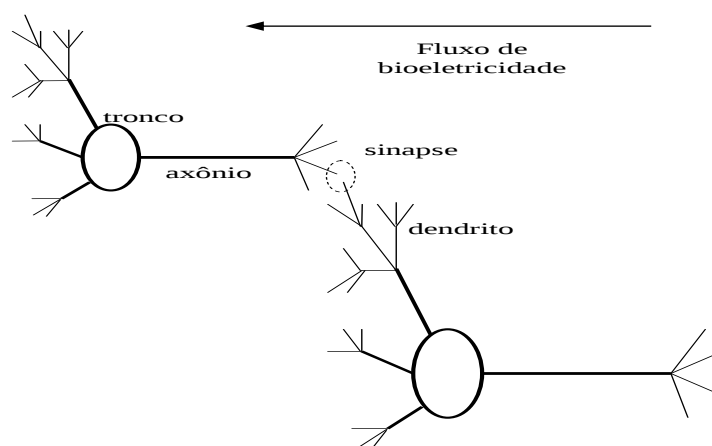
teorias, vários pesquisadores foram desenvolvendo algoritmos para o ajustes dos pesos das redes neurais multi-camada, culminando com o surgimento do algoritmo de “Backpropagation” (retropropagação). O “Backpropagation” parece ter sido primeiramente proposto por Werbos em 1974 e de forma independente redescoberto por volta de 1985 por Rumelhart, Hinton e Williams, tendo Parker e Le Cun também propostos algoritmos semelhantes.

Com o “Backpropagation” volta a crescer o interesse na área, mais maduro e consciente. Uma parte massiva da teoria de redes neurais tem sido transformada em aplicações, com o surgimento de novas corporações dedicadas à comercialização desta técnica. O crescimento nas atividades de pesquisa tem sido muito grande e novas áreas de atuação das redes neurais têm surgido, tais como: compressão de imagem e voz, reconhecimento de caracteres manuscritos, diagnósticos médicos, conservação de energia e previsões de séries temporais, em especial séries econômicas.

### 3.2.1 O Neurônio Artificial<sup>2</sup>

O cérebro humano é composto por mais ou menos  $10^{11}$  neurônios de diversos tipos diferentes. A figura (9) mostra o esquema de um neurônio comum.

Figura 9 – Representação básica de um neurônio biológico



O núcleo da célula está localizado no corpo, ou soma, da mesma. Conectados ao corpo da célula estão fibras nervosas com estruturas similares a raízes, chamadas dendritos e uma única ramificação nervosa mais grossa chamada axônio, da qual surgem sub ramificações. No fim dela estão os pontos de transmissão para os outros neurônios, chamados de junções sinápticas ou sinapses. As sinapses podem ocorrer

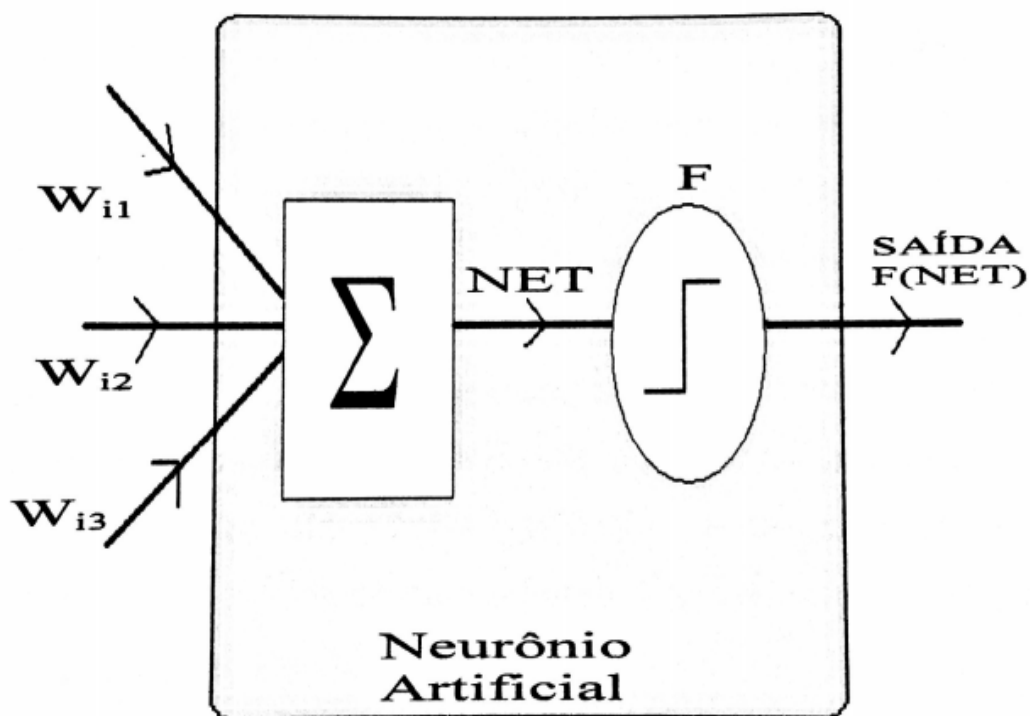
<sup>2</sup> Esta seção está baseada no livro de Braga *et al.*, 2007

tanto entre os dendritos como entre dendritos e o corpo de neurônios diferentes. O axônio de um neurônio típico possui algumas centenas de sinapses que irão se conectar com outros neurônios.

A transmissão do sinal de uma célula para outra é regida por um complexo processo químico, no qual substâncias neurotransmissoras são liberadas pelo neurônio. O efeito é um aumento ou uma queda no potencial elétrico no corpo da célula receptora. Se este potencial alcançar o limite de ativação da célula, um pulso ou uma ação potencial de potência e duração fixa é enviada através do axônio. Diz-se então que o neurônio está ativo.

Projetado para imitar - em primeira ordem - um neurônio biológico, um neurônio artificial é, essencialmente, um conjunto de entradas e uma saída com um centro de processamento, que obedece a uma regra ou função matemática - função de ativação. Cada entrada é multiplicada por um peso correspondente ( $W_{ij}$ ), gerando entradas ponderadas, de forma análoga à força das sinapses. Em seguida todas estas entradas ponderadas são somadas, obtendo-se um valor  $NET$  que será comparado com o valor limite para ativação do neurônio ( $F$ ). Caso este valor alcance o valor limite de ativação do neurônio, ele se ativará, caso contrário ele ficará inativo. A Figura (10) mostra o modelo que implementa esta idéia.

Figura 10 – Diagrama esquemático de um neurônio artificial



### 3.2.1.1 Funções de Ativação

A função de ativação  $F$  é a que processa o sinal  $NET(N)$  para produzir a saída final do neurônio,  $SAÍDA(S)$ . Esta função pode ter várias formas: uma simples função linear; uma função limiar (função degrau); ou ainda uma função que simule mais precisamente as características não lineares do neurônio biológico. No caso de  $F$  ser uma função linear, a saída do neurônio é dada por:

$$S(N) = k \cdot N \quad (3.21)$$

onde  $k$  é uma constante.

As chamadas funções “squashing”, são utilizadas quando se quer imitar com maior resolução os neurônios biológicos. Elas caracterizam-se por não deixarem  $S$  exceder certos valores limites, os quais, em geral, são menores que os valores de  $N$ . A função “squashing” mais usada é a função logística ou sigmóide, a qual é representada matematicamente por

$$S(N) = \frac{1}{1 + \exp^{-N}} \quad (3.22)$$

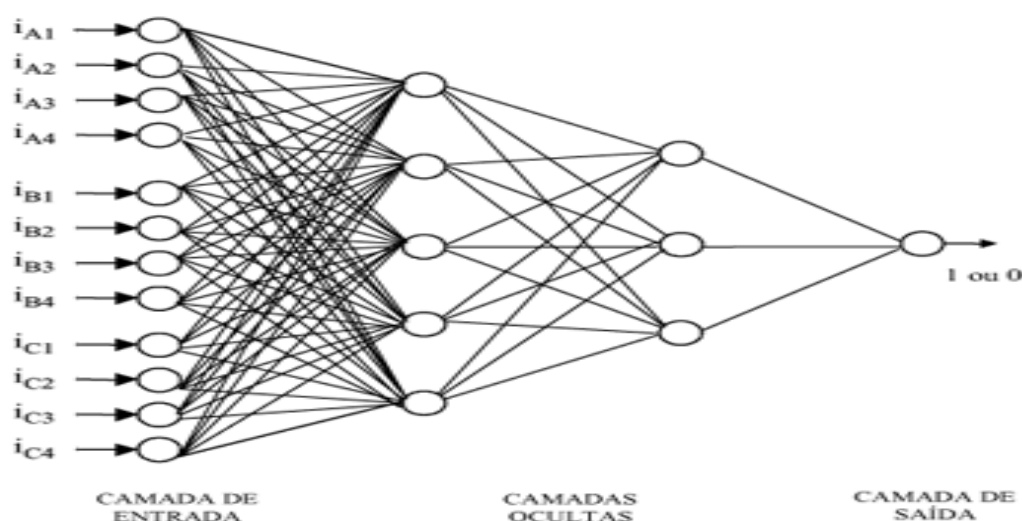
Definindo-se o ganho como a razão entre a variação da saída  $SAÍDA$  e a variação da entrada  $NET$ , pode-se obter ganhos de um valor baixo para grandes excitações negativa, passando por valores altos para excitação zero, até voltar a cair quando a excitação se torna grande e positiva. Em 1973 Grossberg descobriu que esta característica de ganho não-linear resolvia o problema do dilema da saturação com ruído. Neste dilema, Grossberg questionava como a mesma rede poderia manipular tanto sinais pequenos como grandes. Pequenos sinais necessitariam de um alto ganho através da rede para produzir saídas usáveis, exigindo um grande número de estágios de alto ganho. Contudo, estes estágios poderiam saturar a saída, amplificando o ruído (variáveis randômicas) que está presente em qualquer rede criada. Para sinais de entrada com valores altos, os estágios de alto ganho também iriam saturar a saída, eliminando qualquer saída usável. A região central de alto ganho da função sigmóide resolve este problema do processamento de pequenos sinais, enquanto que suas regiões de ganho decrescente nos extremos negativo e positivo são adequadas a grandes excitações.

Mesmo ignorando diversas características do neurônio natural, tais como a não consideração dos atrasos de tempo que afetam a dinâmica do sistema - as entradas produzem saídas imediatas - e a não inclusão dos efeitos de sincronismo ou de modulação de frequência - característica que alguns pesquisadores acham de fundamental importância, as RNAs formadas por simples neurônios artificiais possuem atributos semelhantes aos do sistema biológico, como a capacidade de aprendizado e generalização, podendo-se dizer que a essência do funcionamento do neurônio natural foi absorvida.

### 3.2.1.2 Topologia das Redes Neurais Artificiais

As várias topologias de redes neurais (estratégias de como os neurônios estão conectados entre si) estão divididas, basicamente, em duas classes: não-recorrentes e recorrentes. As RNAs não-recorrentes são aquelas que não realimentação suas entradas com as suas saídas e por isso são também ditas “sem memória”. A estrutura das RNAs não-recorrentes é em camadas, podendo estas RNAs serem formadas por uma (RNA de camada única) ou mais camadas (RNA multi-camada). Redes neurais multi-camadas contêm um conjunto de neurônios de entrada, uma camada de saída e uma ou mais camadas escondidas. A entrada não é considerada uma camada da rede, pelo fato de apenas distribuir os padrões. A camada com os neurônios que fornecem a saída da rede é chamada camada de saída. As camadas que não possuem ligações diretas nem com a entrada, nem com a saída são chamadas de camadas escondidas. Não existem conexões ligando um neurônio de uma camada a outro de uma camada prévia, nem a um neurônio da mesma camada (Figura (11)). As RNAs de uma camada, também chamadas de “perceptrons”, por possuírem um espectro de representações limitado, não serão tratadas nesta tese.

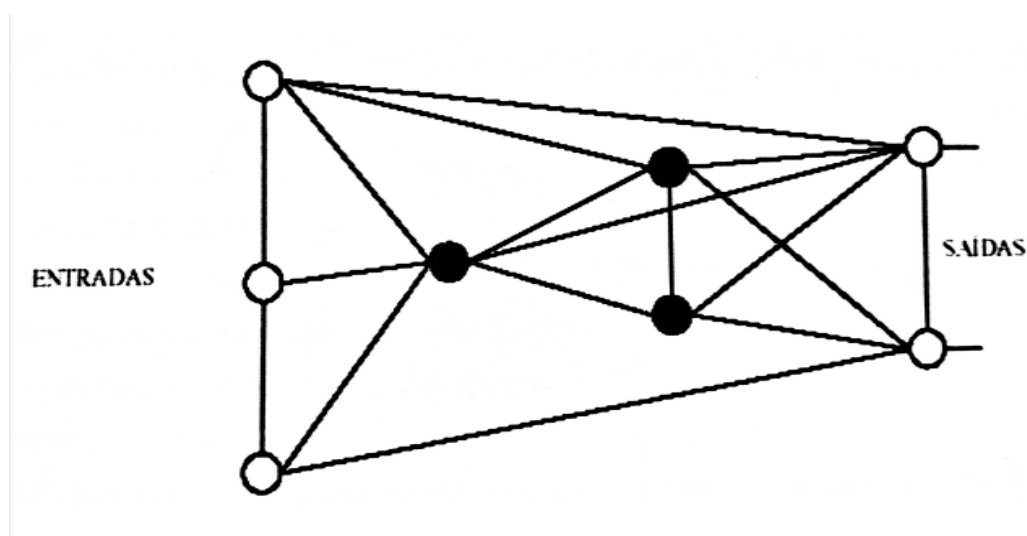
Figura 11 – Exemplo de uma RNA não recorrente



As RNAs recorrentes são que contêm realimentação das entradas com as saídas, sendo suas saídas determinadas pelas entradas atuais e pelas saídas anteriores. Além disso, não são obrigatoriamente organizadas em camadas e quando são, estas redes podem possuir interligações entre neurônios da mesma camada e entre camadas não consecutivas, gerando interconexões bem mais complexas que as RNAs não-recorrentes (Figura (3.2.1.2)).

As RNAs recorrentes, também chamadas de RNAs com memória por possuírem realimentação, respondem a estímulos dinamicamente, isto é, após aplicar uma nova entrada, a saída é calculada e então realimentada para modificar a entrada. Para as redes estáveis, este processo é repetido várias vezes, produzindo pequenas mudanças

Figura 12 – Exemplo de uma RNA recorrente



nas saídas, até estas ficarem constantes. Todavia, como todo sistema retroalimentado, as RNAs recorrentes nem sempre são estáveis, mesmo com entradas constantes. O fato de não se conseguir prever quais redes seriam estáveis foi um problema que preocupou os pesquisadores até o início da década de 80, quando Cohen e Grossberg provaram um poderoso teorema para definir quando as RNAs eram estáveis. Este teorema determina que para as RNAs recorrentes alcançarem um estado estável é necessário que possuam conexões simétricas, melhor dizendo,  $W_{ij} = W_{ji}$ , sendo  $W_{ii} = 0$ . Contribuições importantes também foram dadas por John Hopfield, tanto para a teoria como para a aplicação, sendo que algumas configurações passaram a ser chamadas de redes de Hopfield em sua homenagem. Hinton e Sejnowski também introduziram regras gerais de treinamento para redes, denominadas por eles de máquina de Boltzmann devido ao cálculo das saídas dos neurônios artificiais ser efetuado através de probabilidade com distribuição de Boltzmann.

#### 3.2.1.3 Tipos de Treinamento

Uma descrição detalhada dos tipos de treinamento pode ser encontrada em Bocanegra (2002).

Existe uma grande variedade de algoritmos tanto para o treinamento supervisionado, como para o não supervisionado. Muitos deles baseiam-se no algoritmo proposto por D. O. Hebb em 1961 entretanto, os algoritmos mais eficientes foram desenvolvidos nos últimos trinta anos, baseados nos trabalhos de Rosenblatt (1962), Windrow (1959) e Windrow e Hoff (1960). Nesta tese, será utilizado o mais difundido algoritmo de treinamento: o de retropropagação (“Backpropagation”).

### 3.2.2 Redes Neurais Artificiais Multi-Camadas

O entendimento do problema da separabilidade linear, na década de 60, superou a capacidade limitada de representação das RNAs com uma única camada. Contudo, apesar de se conhecer como aumentar a capacidade de representação da rede com o acréscimo de novas camadas, faltava ainda uma forma de treiná-la. A descoberta veio em meados da década de 70, com a criação de vários algoritmos, entre eles o de retropropagação (“Backpropagation”). O “Backpropagation” se tornou um dos mais populares algoritmos para o treinamento, sendo em parte responsável pelo ressurgimento do interesse na área de Redes Neurais Artificiais.

#### 3.2.2.1 Algoritmo de Retropropagação do Erro

Basicamente, o algoritmo de retropropagação consiste em fornecer variações controladas nos pesos da RNA, objetivando minimizar o erro obtido na saída através do aprendizado do vetor de treinamento (entrada-saída). Para isto, o algoritmo baseia-se no método do gradiente descendente, o qual dada uma medida de erro  $E[W]$  procura melhorar o conjunto de pesos  $W_{ij}$  deslizando para baixo na superfície definida no espaço de  $W$ . Em outras palavras, o algoritmo do gradiente descendente sugere mudanças nos pesos  $W_{ij}$  por uma quantidade  $\Delta W_{ij}$  proporcional ao gradiente de  $E$ :

$$\Delta W_{ij} = -\eta \cdot \left( \frac{\partial E}{\partial W_{ij}} \right) \quad (3.23)$$

onde  $\eta$  é conhecido como taxa de aprendizado,  $i$  refere-se ao neurônio  $i$  da camada posterior e  $j$  ao neurônio  $j$  da camada anterior. Executando-se a derivada e fazendo-se a mudança para cada padrão de entrada individualmente, chega-se a:

$$\Delta W_{ij} = -\eta \delta_i X_j \quad (3.24)$$

onde,

$$\delta_i = (T - O)F'(N) \quad (3.25)$$

A equação (3.25) é conhecida como regra delta.  $F'(N)$  refere-se a derivada da função de ativação,  $T$  é o vetor alvo e  $O$  é a saída fornecida pela RNA.

Como existe um valor alvo, o ajuste dos pesos da camada de saída é facilmente executado utilizando-se as equações (3.24) e (3.25). O que não se verifica com relação à atualização dos pesos da camada escondida, mais complicada pois, não existem valores alvo para a comparação. Assim, o cálculo do delta ( $\delta$ ) deve ser feito substituindo-se a diferença entre o vetor alvo e a saída fornecida pela RNA, na equação (3.25) pelo somatório dos produtos entre os pesos que ligam o neurônio em questão com os

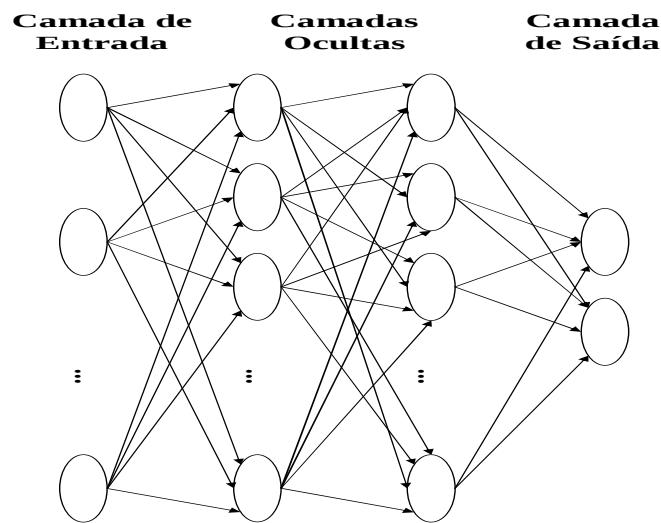
neurônios da camada de saída e os deltas dos respectivos neurônios da camada de saída:

$$\delta_i = \left( \sum_j \delta_j W_{ij} \right) F'(N). \quad (3.26)$$

Após calculado o  $\delta$  para o neurônio  $i$ , os pesos ligados a este neurônio podem ser ajustados através da equação (3.24) e assim sucessivamente para todos os neurônios da camada escondida. Caso haja mais de uma camada escondida o procedimento é o mesmo, os  $\delta$ 's devem ser calculados e todos os pesos associados àquela camada devem ser ajustados, sendo que agora a camada escondida posterior é tida como referência para o cálculo do novo delta.

A Figura (13) mostra uma rede multicamada apropriada para o treinamento com o algoritmo de retropropagação.

Figura 13 – RNA multi-camada



Como fora mencionado anteriormente, o treinamento da rede visa ajustar os pesos, tal que a aplicação de um conjunto de entradas produza um conjunto de saídas desejadas. Antes de iniciar o processo de treinamento, todos os pesos devem ser inicializados randomicamente com valores pequenos, garantindo desta forma que a rede não ficará saturada com grandes valores de pesos e prevenindo-se contra certas patologias de treinamento.

O treinamento pode ser dividido em passos da seguinte forma:

1. Selecionar o próximo par do conjunto de treinamento e aplicar o vetor de entrada a rede.
2. Calcular a saída da rede.

3. Calcular o erro entre a saída da rede e a saída do alvo.
4. Ajustar os pesos da rede de maneira a minimizar o erro.
5. Repetir o passo 1 até o passo 4 para cada vetor do conjunto de treinamento, até o erro se tornar aceitavelmente baixo para o conjunto inteiro.

Os passos 1 e 2 constituem a etapa de propagação para frente, onde o sinal de entrada é transmitido através da rede da entrada até a saída. Os passos 3 e 4, por sua vez, constituem-se na etapa de retropropagação, onde o erro calculado é propagado de volta através da rede para ajustar os pesos.

Apesar do grande sucesso das aplicações do algoritmo de retropropagação e da sua enorme popularidade, muitos problemas ainda o acompanham. Entre eles, o que mais se destacam são:

- O longo período de treinamento: ocorre principalmente para problemas complexos e não há garantias que depois deste tempo o treinamento tenha obtido êxito;
- Mínimos locais: como a superfície do erro geralmente é cheia de vales e desníveis e o algoritmo emprega um tipo de gradiente descendente, existe sempre a possibilidade de ficar preso em um mínimo local;
- Paralisia da rede: durante o treinamento os pesos podem ser ajustados para valores muito grandes, os quais vão levar a derivada da função de ativação a zero (no caso de funções “squashing”), impedindo a rede de aprender o conjunto de treinamento.

Por conta de suas limitações, o algoritmo de retropropagação sofreu inúmeras variações nos últimos anos. Uma alternativa para diminuir o tempo de treinamento, assegurando um maior grau de estabilidade para rede, chamada “momentum”, foi proposta por Rumelhart, Hinton e Williams em 1986 e se mostrou bastante eficiente para alguns problemas. Outras alternativas também foram propostas por Parker em 1987, bem como por Stornetta e Huberman. Até agora, todavia, nenhuma tem solucionado de forma definitiva e confiável os problemas do “Backpropagation”.

### 3.2.3 Previsão através de RNAs

Como observado nas seções anteriores, para se realizar uma análise, ou mesmo uma previsão de uma série temporal utilizando procedimentos padrões (estatísticos), é necessário, além de todas as considerações e transformações que possam vir a ser feitas para tornar a série adequada, um modelamento matemático *a priori*. Aliado a isto, tem-se o fato de que as técnicas estatísticas para predição e tomada de decisão não têm

bons resultados para muitas aplicações que possuem tamanho de amostras restrito e com alguma não linearidade no conjunto de dados.

Em função deste desempenho, considerado apenas moderado principalmente na área financeira, dos métodos estatísticos convencionais para previsão de séries temporais, os analistas passaram a procurar formas alternativas de previsão. Dentre as muitas técnicas que surgiram, as RNAs ganharam especial atenção pela sua habilidade de aprendizado e sua capacidade de generalização, associação e busca paralela. Estas qualidades as tornam capazes de identificar e assimilar as características mais marcantes das séries, tais como sazonalidade, periodicidade, tendências, entre outras, na maioria das vezes camufladas por ruídos, sem necessitar do laborioso passo da formulação teórica, imprescindível para os procedimentos estatísticos.

Além disso, as RNAs também têm se destacado pelos seguintes aspectos:

- Possibilidade de considerar o comportamento não-linear dos fenômenos responsáveis pela geração dos dados de entrada;
- Habilidade de aproximar qualquer mapeamento entrada/saída de natureza contínua;
- Necessidade de pouco conhecimento estatístico sobre o ambiente no qual a rede está inserida;
- Capacidade de aprendizado, a qual é atingida através de uma sessão de treinamento com exemplos entrada/saída que sejam representativos do ambiente;
- Capacidade de generalização, a qual permite à rede ter um desempenho satisfatório em resposta a dados não pertencentes ao conjunto de treino;
- Tolerância a falhas, o que permite à rede continuar a apresentar resultados aceitáveis no caso de falha de alguns neurônios - unidades computacionais básicas das redes neurais artificiais;

Afora as vantagens, as RNAs possuem problemas não negligenciáveis. Um dos principais é a falta de procedimentos para definir com precisão o número de camadas escondidas ou o número de neurônios em cada uma destas camadas, em outras palavras, a topologia mais apropriada para a aplicação.

Outro problema é a incapacidade de explicação dos resultados obtidos. Estando as informações das RNAs intrínsecas aos pesos, não existem meios explícitos (ou imediatos) de se justificar os resultados fornecidos pelas redes neurais o que, em certas aplicações, como na concessão de empréstimos ou concessão de cartões de crédito, é fundamental.

Um terceiro problema, desta feita mais específico das RNAs que utilizam o algoritmo de retropropagação para o treinamento, é a possibilidade de não convergência das redes

neurais devido a mínimos locais. Todavia, a introdução de fatores no algoritmo como o termo de momentum tem minimizado este problema para certas aplicações.

### 3.2.3.1 Previsão *Single* e *Multi-Step*

Como foi visto, existem vários procedimentos de previsão que vão desde os simples e intuitivos até os mais quantitativos e complexos. As RNAs, quando usadas para previsão de séries temporais, normalmente utilizam a filosofia do procedimento de análise de séries temporais, o qual baseia-se na observação dos dados históricos do sistema a fim de descobrir regularidades em fenômenos aparentemente randômicos, frequentemente encobertos por ruídos. Neste contexto, as previsões podem ser classificadas em: previsões “multi-step” e previsões “single-step”.

As previsões “multi-step” caracterizam-se por possuir realimentação das saídas das RNAs para as entradas das mesmas. Neste previsão, o sistema neural usa um conjunto de valores correntes da série para prever os valores futuros desta série por um período fixo. Em seguida, esta previsão é realimentada na entrada do sistema para prever o próximo período. Estas previsões são muito usadas para identificar tendências e pontos de mudanças preponderantes nas séries. Devido ao erro que é inserido a cada nova previsão, as previsões “multi-step” são usadas para prever apenas alguns passos a frente, em média de 15 a 20 passos, quando o erro ainda é aceitável.

Por serem retroalimentadas, as previsões “multi-step” ficam restritas a previsões univariantes, já que no caso das multivariantes, nem todas as séries fornecidas nas entradas são previstas pelas RNAs para poderem ser realimentadas.

Nas previsões “single-step” não existe realimentação. As RNAs prevêem um passo à frente utilizando apenas os valores anteriores da série, podendo este passo tanto ser para previsões de curto prazo como para previsões de médio e longo prazo, bastando que se tenha dados suficientes para treinar a rede. A previsão “single-step” também serve para avaliar a adaptabilidade e a robustez do sistema, mostrando que mesmo quando as RNAs fazem previsões erradas, elas são capazes de se auto corrigirem e fazer as próximas previsões corretamente. Além disso, as previsões “single-step” podem ser usadas como alarme, permitindo ao investir, no caso de séries financeiras, comprar ou vender antes de um aumento ou de uma queda nos preços.

Como não utilizam realimentação, as previsões “single-step” estão habilitadas a fazer tanto previsões univariantes com multivariantes, tornando-se neste aspecto muito mais poderosa que as previsões “multi-step”.

### 3.2.3.2 Aplicações

As RNAs vêm se tornando uma poderosa ferramenta para processamento de informações financeiras devido a sua capacidade em manipular dados imprecisos e ruidosos muitas vezes com vantagens sobre vários métodos estatísticos, essencialmente lineares.

Muitas aplicações têm sido desenvolvidas nos últimos anos na área financeira, sendo que a maioria destas aplicações têm mostrado bons resultados quando comparados com procedimentos estatísticos. Um bom exemplo desta superioridade está no estudo feito por Chakraborty *et al.*, (1992), onde os autores utilizaram RNAs para fazer previsões univariadas e multivariadas no preço dos grãos em três cidades americanas: Buffalo, Minneapolis e Kansas. O algoritmo de treinamento foi o de retropropagação, usando o erro quadrático médio (MSE) como função objetivo para o treinamento. Os autores utilizaram dados do período de 1972 a 1980 e os resultados das RNAs foram comparados com o modelo estatístico de Tiao e Tsay comprovando a eficiência das RNAs.

Poddig, (1993) apresenta um outro bom exemplo bem sucedido de aplicação das RNAs na previsão de séries financeiras, onde se utilizou RNAs para prever as mudanças na taxa de câmbio USD/DM, um, três e cinco dias à frente. O autor baseou-se em estudos da SIEMENS Corporate Research e utilizou 44 séries de indicadores fundamentais e técnicos como commodities, metais preciosos, indicadores do mercado financeiro e moedas estrangeiras, entre outros, para alimentar as RNAs. As análises foram divididas em duas fases. Na primeira fase ele utilizou redes com uma camada e redes MLP (“Multiple Layer Perceptrons”) com duas camadas (uma escondida), que foram treinadas com o algoritmo de retropropagação. O autor fez comparações entre os resultados das duas redes e entre elas e métodos estatísticos baseados em análise de fatores discriminantes, tendo as RNAs mostrado melhores resultados.

Na segunda fase de análises, o autor usou redes RCP (“ReCurrent Perceptron”) em busca de resultados com menor flutuação e maior estabilidade. Desta feita, Poddig comparou os resultados das RNAs com métodos de regressão multivariadas, obtendo novamente melhores resultados para as RNAs.

Refenes *et al.*, (1992) aplicaram RNAs na previsão de taxa de câmbio. Previsões “single” e “multi-step” utilizando o algoritmo de retropropagação foram feitas para a taxa de câmbio US/DM. Os autores analisaram os principais fatores no projetos de RNAs, como arquitetura, função erro, taxa de aprendizado e função de ativação. Os autores discutem também o efeito de diferentes funções de ativação na convergência das RNAs, mostrando que, para o caso em questão, as funções simétricas (*tanh*) convergem mais rápido que as funções assimétricas (*sigmoide*). Por fim, os autores compararam os resultados com modelos estatísticos de alisamento exponencial e Box-Jenkins puramente autorregressivo, obtendo resultados bastante satisfatórios em uma carteira de investimentos, com no mínimo 22% de lucro contra 2-5% dos métodos convencionais.

Inúmeras aplicações bem sucedidas foram realizadas na previsão de séries temporais, na área financeira e demais áreas. Entre elas pode-se destacar séries como: “sunspots”, consumo de energia, entre outras, mostrando que o interesse pela área tem crescido vertiginosamente.

### 3.2.4 Modelagem das RNAs

Infelizmente, o pequeno entendimento das complexas interrelações entre os parâmetros que influenciam o desempenho da rede (como função de ativação, função erro, parâmetros de controle do gradiente descendente e do tempo de treinamento, arquitetura da rede), permite um grau de liberdade adicional no projeto de RNAs. Este grau de liberdade na escolha dos parâmetros adequados, torna as aplicações altamente dependente da expertise do projetista e da experimentação manual para que se obtenham sucesso.

#### 3.2.4.1 Fatores Determinantes no Desempenho das RNAs

O desempenho das RNAs em geral é avaliado em função de três fatores: convergência, generalização e escalabilidade.

A convergência é um fator importantíssimo para o bom desempenho das RNAs, pois está relacionada à capacidade das RNAs de aprender. A convergência é o que indica quando o procedimento de treinamento é capaz de aprender uma classificação definida nos padrões apresentados. RNAs com topologias fixas que utilizam o algoritmo de retropropagação provam a convergência mostrando que em um tempo limite, quando o tempo de treinamento tende a infinito, o erro minimizado pelo método do gradiente descendente tende a zero. Outros métodos (sub-ótimos) provam a convergência mostrando que, na condição de tempo de treinamento tendendo a infinito, o método pode classificar o número máximo de possíveis mapeamentos com uma probabilidade arbitrariamente grande  $p \leq 1$ . A utilização da convergência como fator determinante para o desempenho da rede é feita por motivos óbvios, já que se uma determinada RNA não consegue aprender o conjunto de treinamento que lhe foi apresentado (não convergente), dificilmente conseguirá oferecer bons resultados no momento dos testes.

A habilidade das RNAs de reconhecer padrões fora do conjunto de treinamento - generalização - é a principal propriedade que deve ser exigida de um sistema neural. Ela determina a quantidade de dados necessário para treinar o sistema a fim de que este responda corretamente para padrões fora do conjunto de treinamento. Apesar das várias aplicações bem sucedidas de RNAs descritas na literatura, as condições para que boas generalizações sejam encontradas ainda não são totalmente conhecidas. Estudos teóricos têm mostrado que a probabilidade de generalizações corretas depende do tamanho do espaço de hipóteses (i.e. número total de redes consideradas), do tamanho do espaço de soluções (i.e. conjunto de RNAs que fornece boas soluções) e do número de exemplos do treinamento. Se o espaço de hipóteses é muito grande e/ou o número de exemplos do treinamento é muito pequeno, então existirá um vasto número de RNAs, as quais são consistentes para o conjunto de treinamento. Porém, somente uma pequena porção deste conjunto se encontra no espaço de soluções corretas, fazendo com

que seja provável uma generalização pobre.

De forma geral, a diminuição dos parâmetros livres (número de conexões) da rede aumenta a probabilidade de uma boa generalização, muito embora, isto deve ser feito sem reduzir demais o tamanho da rede, já que esta redução pode implicar na incapacidade da rede em computar a função desejada. Escolher um conjunto de treinamento bastante representativo é outro ponto importante, apresentando as relações fundamentais que devem ser aprendidas pelas RNAs. A grande maioria dos pesquisadores experientes concordam que uma boa representação dos dados de entrada é um ponto de fundamental importância para o sucesso de uma aplicação.

A escalabilidade está relacionada a ambos, convergência e generalização, e é relativa ao procedimento de ajustar os parâmetros das RNAs para que elas consigam convergir e generalizar bem. Sabe-se que para uma RNA ter um bom desempenho, ela deve ser grande o suficiente para aprender o problema (convergir), mas também pequena o bastante para generalizar bem.

#### 3.2.4.2 Controle dos Fatores Determinantes no Desempenho das RNAs

O controle dos fatores determinantes de desempenho se dá pela escolha adequada dos parâmetros da rede como: função de ativação, função erro, parâmetros de controle do gradiente descendente e do tempo de treinamento e arquitetura da rede, os quais, como dito anteriormente, variam de aplicação para aplicação e não possuem regras de escolha muito bem definidas.

##### 3.2.4.2.1 Padronização

Mesmo não controlando diretamente os fatores determinantes de desempenho das RNAs, a padronização pode, frequentemente, melhorar a performance das RNAs através da remoção de características insignificantes do conjunto de dados. As informações relevantes das séries são, geralmente, independentes de valores de compensação (“offset”) e de desvios padrão. Estas características podem obscurecer a tarefa das RNAs. A padronização, entretanto, admite várias formas e exige cuidados, já que é muito fácil remover, inadvertidamente, informações preciosas dos dados. Por outro lado, a padronização correta dos dados pode melhorar consideravelmente o desempenho de uma rede neural.

A remoção da média é um dos tipos mais usados de padronização e, em muitos casos, a informação significativa de uma série está contida em pequenas variações que se encontram adicionadas a um grande valor de compensação. Este valor de compensação pode ser retirado se é sabido que ele é irrelevante para o sistema e está inibindo o treinamento. Para isto, basta que se retire de todos os pontos da série a média da mesma.

Melhor dizendo, para cada componente dos vetores da série, tem-se:

$$\Omega_{sm} = \Omega - \bar{\Omega} \quad (3.27)$$

A diferenciação dos dados é uma padronização utilizada principalmente quando se tem valores com pequenas variações absolutas, ficando a informação significativa na variação percentual dos mesmos. As pequenas variações em termos absolutos comumente escondem a informação importante, piorando o desempenho da rede. Assim, para cada componente dos vetores da série, tem-se:

$$\Omega_d = \frac{\Omega_t - \Omega_{t-1}}{\Delta t} \quad (3.28)$$

Grandes variações na magnitude dos componentes do vetor é outro fator que pode confundir a rede. Nestes casos, faz-se a padronização nas magnitudes do vetor, dividindo cada componente do vetor pelo desvio padrão calculado sobre o total de componentes. Com isto, para cada componente dos vetores da série, tem-se:

$$\Omega_{sdp} = \frac{\Omega}{\sigma_{\Omega}} \quad (3.29)$$

A padronizações não lineares dos dados, como o logaritmo, são usadas principalmente em casos onde a magnitude dos dados varia de muitas ordens de grandeza, mas a informação significativa esta contida na variação percentual. Por exemplo, quando os maiores valores dos dados excedem a região da rede, as variações percentuais nos pequenos valores provavelmente serão perdidas. Neste caso, o logaritmo dos dados irá ajustar as faixas, tal que os maiores valores serão comprimidos mais que os pequenos, alocando uma faixa constante para os desvios percentuais, independente do valor dos dados.

Para essas padronizações, quando a característica está sobre todos os componentes dos vetores, faz-se uma padronização total. Caso os componentes dos vetores representem diferentes tipos de dados, faz-se uma padronização vertical para cada tipo. Já no caso em que componentes adjacentes do vetor representam amostras consecutivas de um processo contínuo, usa-se a padronização horizontal. Por fim, em casos onde os dados possuem uma distribuição multi modal, com os vários modos tendo diferentes médias e desvios padrões, usa-se a padronização local.

#### 3.2.4.2.2 Funções de Ativação

A primeira importante consideração no projeto de RNAs é a escolha da função de ativação neuronal. Existem vários tipos de funções utilizadas, como funções lineares, funções de limiar ou funções não lineares, que simulam mais precisamente as características dos neurônios biológicos. A escolha da função de ativação varia muito em função da aplicação e do algoritmo de treinamento utilizado. Entre as funções não lineares,

as funções “squashing” são as mais usadas para todos os tipos de treinamento. Elas são mais complexas, diferenciáveis e comportam-se bem para a maioria das aplicações. Pode-se classificar as funções “squashing” em : assimétricas e simétricas. Uma típica função “squashing” assimétrica é a função logística, também chamada de sigmoide. No caso das funções simétricas, sua representante típica é a tangente hiperbólica. A família das funções “squashing” é definida por:

$$f(x) = k + \frac{c}{1 + \exp^{Tx}} \quad (3.30)$$

onde com a escolha apropriada para os valores de  $k$ ,  $c$  e  $T$  são obtidas as funções sigmoide e tangente hiperbólica típicas. Os parâmetros  $k$  e  $c$  influenciam na amplitude da função, enquanto que o parâmetro  $T$  determina a inclinação da curva na origem. Em aplicações que possuem os valores dos padrões de treinamento muito altos, utiliza-se a função “squashing” somente nos neurônios das camadas escondidas, ficando os neurônios da camada de saída com funções lineares. Este artifício tem o objetivo de evitar a paralisia da rede, pois como os valores de entradas são altos, mesmo pequenos erros podem ser grandes em valores absolutos, forçando os neurônios a trabalharem em regiões onde a derivada das funções “squashing” é praticamente nula. Como o erro que é retro propagado para o treinamento é diretamente proporcional a esta derivada, o processo de treinamento fica paralisado e a rede não converge.

Escolher funções “squashing” assimétricas ou simétricas depende muito da aplicação. Em geral, quando números negativos são utilizados, opta-se por funções simétricas como a  $\tanh$ , já que estas são bipolares.

#### 3.2.4.2.3 Funções Erro

A função erro informa o grau de convergência das RNAs. Através dela, é possível saber o quanto as RNAs estão aprendendo sobre os padrões apresentados durante o treinamento. A função erro mais comumente usada é a MSE (Mean Square Error) dada por:

$$MSE = \frac{1}{n} \sum_{t=1}^n (T_t - O_t)^2 \quad (3.31)$$

onde  $n$  é o número de padrões apresentados.

Apesar de nem sempre ser a função mais indicada, devido aos problemas causados por trabalhar com médias, a MSE é a mais utilizada pela sua facilidade de cálculo e por ser uma métrica encontrada em praticamente todos os pacotes computacionais que trabalham com estatística e RNAs, facilitando assim a comparação dos resultados.

Uma outra opção de função erro é a MAD (Mean Absolute Deviation), a qual é dada pela fórmula:

$$MSE = \frac{1}{n} \sum_{t=1}^n |T_t - O_t|^2 \quad (3.32)$$

A diferença básica entre o MAD e o MSE é que o segundo eleva o erro ao quadrado, penalizando erros maiores com muito mais rigor.

#### 3.2.4.2.4 Parâmetros de Controle do Gradiente Descendente e do Tempo de Treinamento

O terceiro importante mecanismo para controlar o desempenho do aprendizado das RNAs é a escolha dos parâmetros de controle do gradiente descendente e do tempo de treinamento. Pesquisadores têm experimentado termos extras na fórmula do gradiente descendente como o termo de momentum, segunda derivada, entre outros. O termo de *momentum*, é utilizado para diminuir o tempo de treinamento e o risco de mínimos locais, sem que o risco de oscilações das RNAs aumente.

Reduzir o tempo de treinamento em fatores da ordem de 100, poder ser alcançado utilizando-se o artifício da segunda derivada. Nesta técnica, ao invés de utilizar-se a inclinação da superfície do erro (primeira derivada) como no algoritmo de retropropagação, utiliza-se a taxa de variação da inclinação (segunda derivada) desta superfície.

A taxa de aprendizado influencia fortemente no tempo de treinamento e no controle do gradiente descendente, cujo o papel essencial sobre a magnitude das mudanças dos pesos. Encontrar uma taxa de aprendizado apropriada para cada aplicação é uma tarefa difícil que depende basicamente de experimentações. Uma taxa de aprendizado pequena implica em pequenas variações, mesmo quando grandes mudanças são necessárias, tornando o treinamento lento e aumentando as chances de paradas em mínimos locais. Por outro lado, altas taxas de aprendizado provocam grandes mudanças nos pesos, diminuindo o tempo de treinamento e evitando mínimos locais, mas podendo levar a RNA a saturação ou mesmo à oscilação, comprometendo todo o processo de aprendizado. Uma estratégia que tem se mostrado bastante eficiente é o uso de taxas de aprendizado dinâmicas, as quais começam com valores altos e vão diminuindo a medida que o treinamento vai evoluindo. A sugestão feita por Hertz e adotada nesta tese, é a de usar uma taxa de aprendizado adaptativa que checa, a cada apresentação do conjunto de treinamento, se a função erro diminuiu. Em caso negativo, o passo é desfeito, a taxa de aprendizado é diminuída e os padrões são reapresentados. Este procedimento é repetido até que o erro passe a diminuir. Quando a função erro diminui, o passo é validado e o treinamento continua. Caso a função erro diminua por um número  $k$  de vezes consecutivas (em nosso caso  $k = 10$ ), a taxa de aprendizado é aumentada para acelerar o treinamento.

#### 3.2.4.2.5 Arquitetura das RNAs

A arquitetura da rede é muito importante para as capacidades de convergência e generalização das RNAs. A escolha de uma arquitetura adequada, nem tão grande e nem tão pequena, é a principal responsável pelo sucesso de uma aplicação. Para se

obter uma boa habilidade de generalização, deve-se projetar a rede colocando dentro dela o máximo conhecimento possível sobre o problema (e.g. a topologia do espaço de entradas) e limitar o número de conexões adequadamente.

Existem várias técnicas que tentam encontrar a arquitetura ideal através da diminuição de neurônios redundantes e do compartilhamento de pesos. Pode-se citar Hertz, (1991) e Refenes (1991), entre outras, porém, nenhuma conseguiu resultados convincentes para todas as aplicações. Existe também a possibilidade de fazer uma busca no espaço de todas as arquiteturas possíveis. Para isso, deve-se treinar cada uma das arquiteturas separadamente e então avaliá-las com uma função erro apropriada que incorpore o desempenho e o número de unidades. A busca também pode ser feita através de um algoritmo genético, tal que as boas arquiteturas encontradas em uma triagem sejam mantidas para serem combinadas com boas arquiteturas de outras triagens. Contudo, estes tipos de busca se mostram pouco práticos para aplicações que utilizam grandes RNAs, onde cada treinamento exige um alto esforço computacional. A verdade é que a arquitetura ideal das RNAs ainda continua sendo determinada principalmente pela experimentação manual e pela experiência e sensibilidade do projetista.

Como vimos neste capítulo, a rede neural artificial é uma importante técnica de IA para se realizar previsões. Os modelos de previsão de preço no mercado de ações seguem basicamente duas linhas: uma de caráter mais técnico, utiliza-se de dados históricos (volume de transações, preços passados, volatilidade etc.) para determinar os preços futuros das ações e, outra baseada em fundamentos econômicos, utilizada para fazer previsões taxas de juros, preços e retornos de outras ações e outras variáveis micro ou macroeconômicas. A ala acadêmica mais conservadora contesta essa análise técnica argumentando que esse comportamento é irracional dado a hipótese de eficiência dos mercados. Segundo esta hipótese, o preço atual de uma ação já reflete toda a informação dos preços passados. O argumento para isto é que uma oportunidade de lucro será aproveitada imediatamente e conseqüentemente desaparecerá. Ou seja, o mercado é tão eficiente que impossível comprar ou vender tão rapidamente de forma a se beneficiar continuamente. Como consequência dessa hipótese, portanto, os preços das ações seguem um passeio aleatório (*random walk*) não havendo uma forma de se prever valores futuros baseado apenas em valores passados.

Desacreditando na hipótese de eficiência de mercado, um grande número de investidores têm tomado suas decisões de investimento com vistas em previsões utilizando dados históricos como nos modelos GARCH vistos nas seções anteriores. Essas decisões são tomadas sob a premissa de que existe um padrão dos dados e que, este padrão, pode ser usado para prever valores futuros. Na existência desse padrão, pode-se, a princípio, aplicar técnicas automatizadas de reconhecimento de padrões como por exemplo a técnica de redes neurais artificiais (RNA). No entanto, muito se tem para avançar na utilização dessas redes para se realizar as previsões, em particular na escolha dos

melhores parâmetros da rede, de forma que se consiga resultados de previsões mais robustos.

### 3.3 Resultados experimentais com métodos tradicionais

#### 3.3.1 Análise através da Metodologia Box-Jenkins e modelos ARMA-GARCH

Nesta seção trataremos as séries escolhidas com a utilização da metodologia Box-Jenkins apresentando as sete medidas de desempenho de modelos discutidas que serão utilizadas mais adiante para comparação dos modelos testados nesta tese.

##### 3.3.1.1 Ajuste de modelos para série das Manchas Solares

Exibimos na Seção 2.2.1 uma breve descrição sobre o conjunto de dados das Manchas Solares. Apresentamos, a seguir, os modelos ajustados para esses dados. Através da Figura 2, podemos observar que a série não apresenta indícios de tendência, e isso fica comprovado pelo teste de não estacionariedade, proposto por Kwiatkowski et al. (1992), em que observamos um  $p$ -valor de 0.09, ou seja, 9%, indicando a não rejeição da hipótese nula de não tendência.

Diversos modelos foram propostos para a série de Manchas Solares, tendo em vista que foi uma das primeiras séries a serem estudadas devido às características peculiares, tais como a não-linearidade e a difícil previsão desta. Morettin e Tolo (2004), sugerem um modelo AR(2), já Ferreira (2006) apresenta previsões com um modelo ARIMA(9,0,1), entre outros. No presente trabalho fizemos a escolha do modelo através dos critérios de avaliação AIC - *Akaike Information Criterion* e/ou BIC - *Bayesian Information Criterion*, detalhes em Akaike (1974) e Schwarz (1978), respectivamente.

As Tabelas 8 e 9 apresentam os valores dos critérios para modelos ARIMA( $p,0,0$ ) e ARIMA( $p,0,1$ ), onde fizemos  $p \leq 10$ . Pelo critério AIC os modelos escolhidos são ARIMA(9,0,0) e ARIMA(9,0,1).

As Figuras 14-15 apresentam os valores reais do *Grupo de Teste* e os valores preditos para cada modelo. A Tabela 10 apresenta os valores das estatísticas da avaliação das previsões. Notamos que ambos os modelos apresentam valores relativamente próximos para as previsões do *Grupo de Teste*, dentre as medidas de avaliação de previsão podemos destacar POICD e U-Theil, em que para o POICD tem-se resultados significativos de 79.31%, muito superiores aos 50% obtidos por uma previsão do tipo sucesso ou fracasso. No que tange à medida U-Theil obtivemos o valor 0.7415 e 0.7403, para os modelos ARIMA(9,0,0) e ARIMA(9,0,1), respectivamente.

Tabela 8 – Valores obtidos através dos critérios de seleção para os modelos da forma  $ARIMA(p,0,0)$ . Fonte: elaborada pelo autor.

| ARIMA(p,0,0) | AIC      | pesos AIC | BIC      | pesos BIC |
|--------------|----------|-----------|----------|-----------|
| 1            | 1758.633 | 0.000     | 1762.268 | 0.000     |
| 2            | 1640.557 | 0.000     | 1647.827 | 0.000     |
| 3            | 1636.640 | 0.000     | 1647.544 | 0.000     |
| 4            | 1617.070 | 0.000     | 1631.609 | 0.000     |
| 5            | 1613.792 | 0.000     | 1631.966 | 0.000     |
| 6            | 1590.526 | 0.000     | 1612.335 | 0.000     |
| 7            | 1568.459 | 0.000     | 1593.902 | 0.000     |
| 8            | 1548.789 | 0.001     | 1577.867 | 0.006     |
| 9            | 1535.036 | 0.663     | 1567.749 | 0.919     |
| 10           | 1536.397 | 0.336     | 1572.745 | 0.076     |

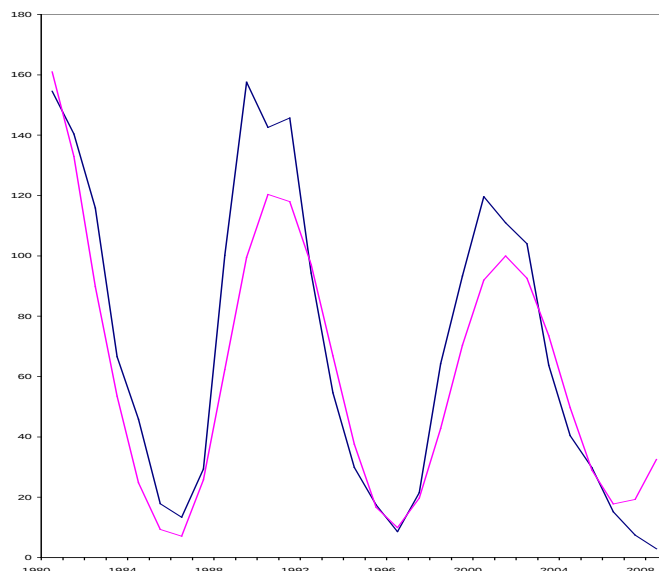
Tabela 9 – Valores obtidos através dos critérios de seleção para os modelos da forma  $ARIMA(p,0,1)$ . Fonte: elaborada pelo autor.

| ARIMA(p,0,1) | AIC      | pesos AIC | BIC      | pesos BIC |
|--------------|----------|-----------|----------|-----------|
| 1            | 1646.512 | 0.000     | 1650.147 | 0.000     |
| 2            | 1565.322 | 0.000     | 1572.591 | 0.000     |
| 3            | 1564.500 | 0.000     | 1575.404 | 0.000     |
| 4            | 1566.470 | 0.000     | 1581.009 | 0.000     |
| 5            | 1568.221 | 0.000     | 1586.395 | 0.000     |
| 6            | 1535.733 | 0.016     | 1557.542 | 0.362     |
| 7            | 1531.846 | 0.112     | 1557.289 | 0.411     |
| 8            | 1529.837 | 0.306     | 1558.915 | 0.182     |
| 9            | 1529.172 | 0.427     | 1561.885 | 0.041     |
| 10           | 1531.419 | 0.139     | 1567.767 | 0.002     |

Tabela 10 – Medidas de desempenho de previsão para os modelos analisados. Fonte: elaborada pelo autor.

| Medidas | $\hat{y}_t$  |              |
|---------|--------------|--------------|
|         | ARIMA(9,0,0) | ARIMA(9,0,1) |
| MSE     | 380.3000     | 382.6400     |
| MAPE    | 59.85%       | 59.08%       |
| U Theil | 0.7415       | 0.7403       |
| POICD   | 79.31%       | 79.31%       |
| ARV     | 0.0070       | 0.007        |
| AIC     | 192.2879     | 192.4660     |
| BIC     | 214.6975     | 241.9370     |

Figura 14 – Comparativo entre os valores reais e as previsões geradas pelo modelo ARIMA(9,0,0).



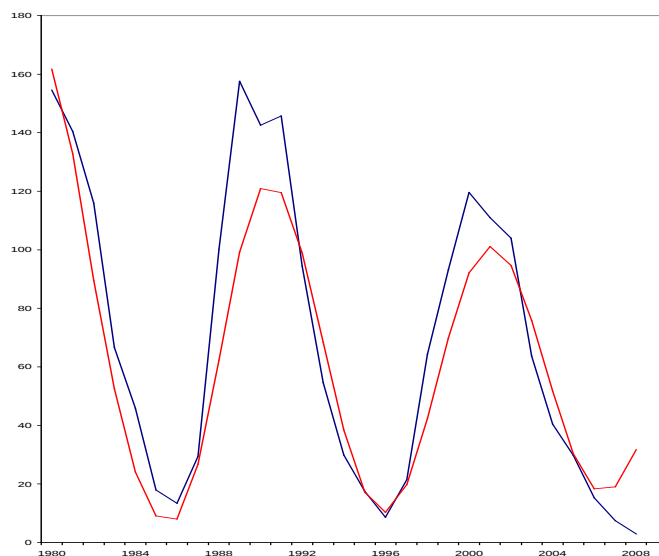
*Fonte: elaborada pelo autor.*

### 3.3.1.2 Análise da série CO<sub>2</sub>

A série dos níveis mensais de emissão de dióxido de carbono (CO<sub>2</sub>) é apresentada na figura 3. Analisando a série observamos que existe uma tendência e também um comportamento sazonal que pode ser melhor observado na Figura 16. Adicionalmente, é sabido que os níveis de dióxido de carbono são altos durante o período do inverno e mais alto ainda durante os meses do verão.

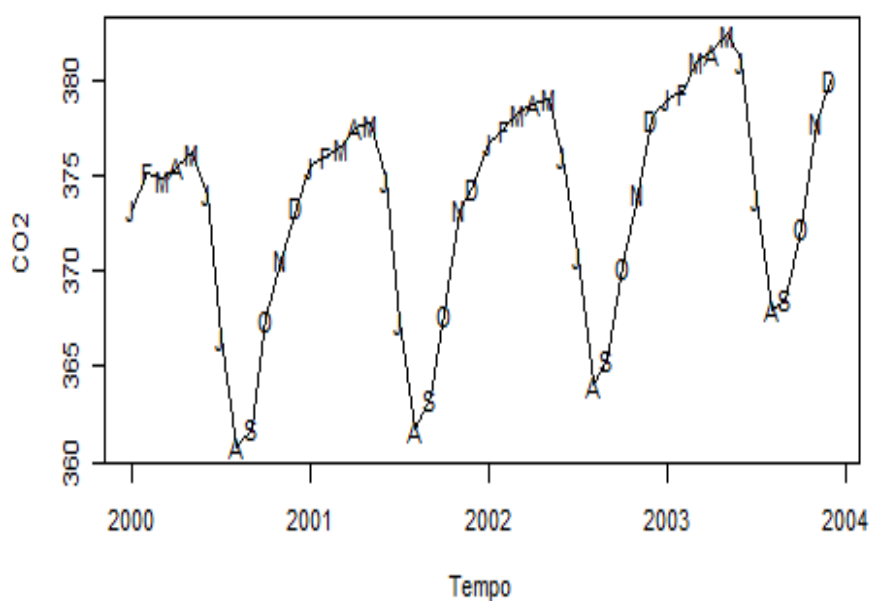
A tendência ascendente da série de CO<sub>2</sub> nos dá indício de que a mesma é não estacionária. Na Figura 17, temos a função de autocorrelação para esta série. Notamos que as relações de autocorrelação sazonal apresentam bastante destaque observando a forte correlação nas defasagens 12, 24, 36, e assim por diante.

Figura 15 – Comparativo entre os valores reais e as previsões geradas pelo modelo ARIMA(9,0,1).

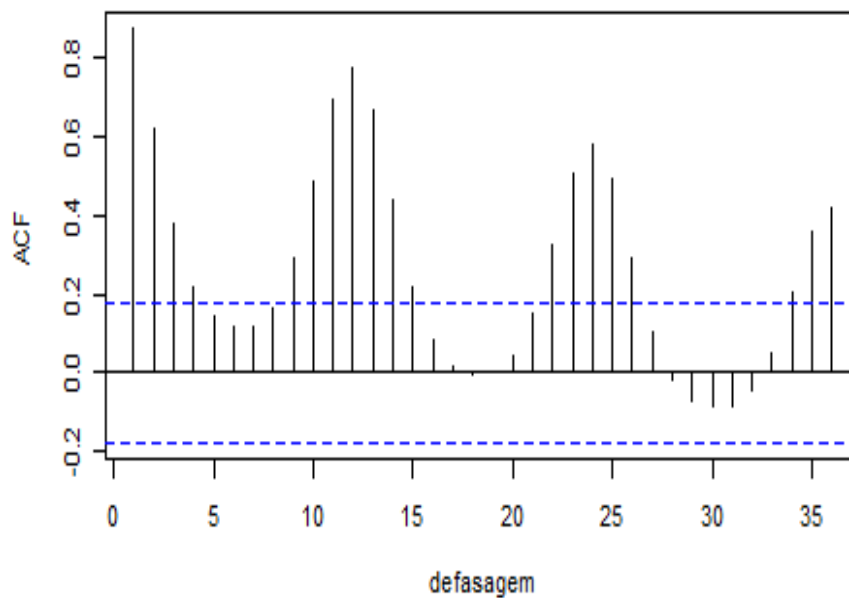


*Fonte: elaborada pelo autor.*

Figura 16 – Níveis mensais de dióxido de carbono com símbolos.

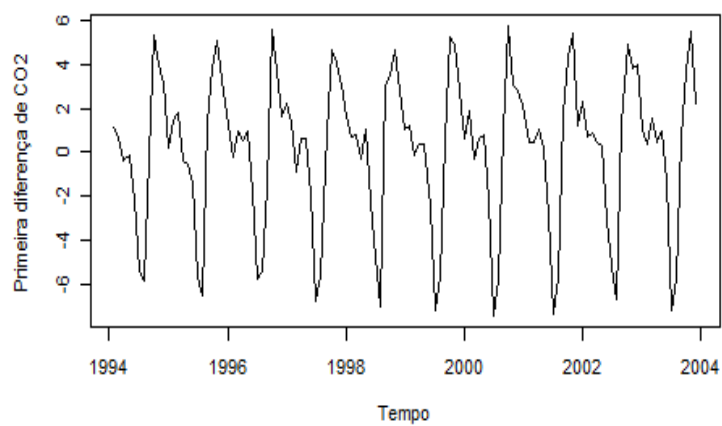


*Fonte: elaborada pelo autor.*

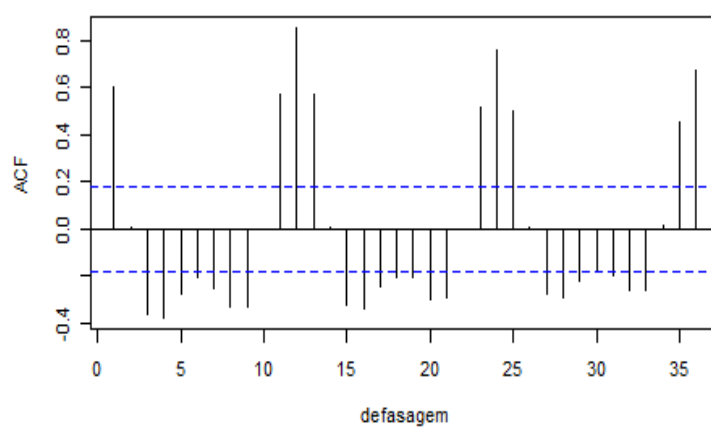
Figura 17 – ACF amostral para os níveis de CO<sub>2</sub>.

*Fonte: elaborada pelo autor.*

Diferenciando a série observamos que não há mais tendência, porém a sazonalidade persiste, como pode ser observado nas Figura 18 e 19. Para contornarmos este problema, iremos aplicar uma diferença sazonal na série.

Figura 18 – Gráfico da série temporal da primeira diferença dos níveis de CO<sub>2</sub>.

Fonte: elaborada pelo autor.

Figura 19 – ACF amostral da primeira diferença dos níveis de CO<sub>2</sub>.

Fonte: elaborada pelo autor.

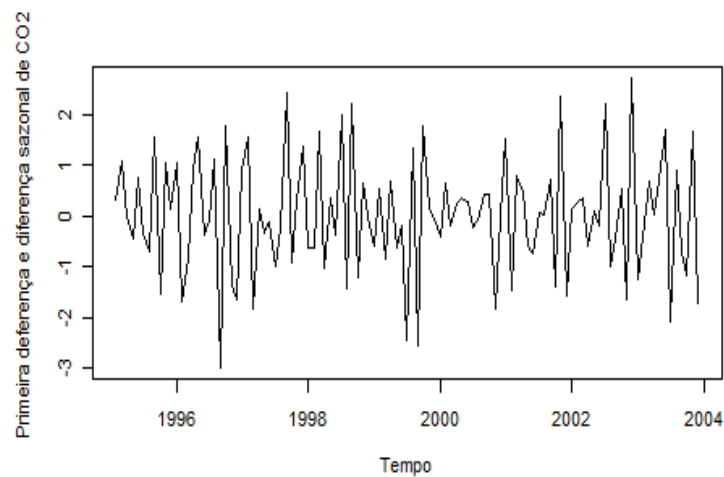
Tabela 11 – Estimativas dos parâmetros para o modelo CO<sub>2</sub>. *Fonte: elaborada pelo autor.*

| coeficientes                                                   | $\theta$ | $\Theta$ |
|----------------------------------------------------------------|----------|----------|
| Estimativas                                                    | -0.5618  | -0.9118  |
| Erro padrão                                                    | 0.0838   | 0.2381   |
| $\sigma_e^2 = 0.4958$ : log likelihood = -124.11, AIC = 252.23 |          |          |

Na Figura 20, apresentamos o resultado da série após aplicarmos uma diferença simples e uma diferença sazonal. Notamos que a série aparentemente não apresenta mais nenhum comportamento sazonal. A Figura 21, sugere que um modelo simples que incorpore as autocorrelações de defasagens 1 e 12 seja adequado. Consideramos inicialmente o modelo multiplicativo, sazonal ARIMA(0,1,1)x(0,1,1)[12]

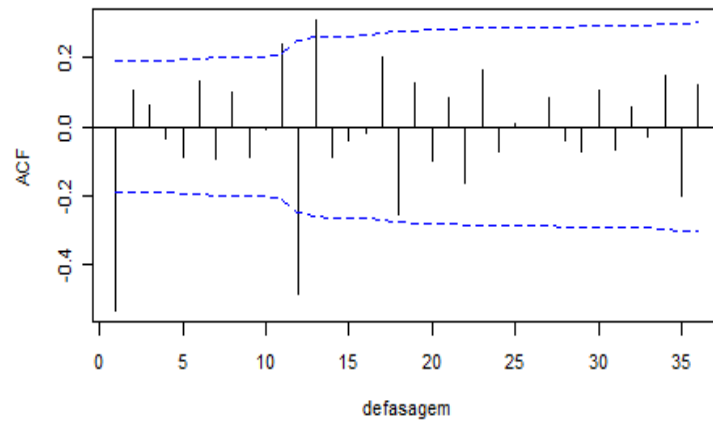
$$\nabla_{12}\nabla Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13},$$

que incorpora a maioria desses requisitos. Como de costume, todos os modelos são preliminares e sujeitos a revisão na fase de diagnóstico da construção do modelo.

Figura 20 – Gráfico da série temporal da primeira diferença e diferença sazonal dos níveis de CO<sub>2</sub>.

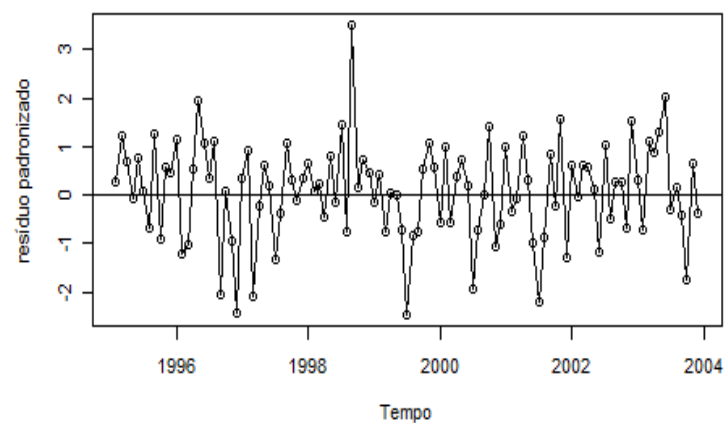
*Fonte: elaborada pelo autor.*

Com a especificação do modelo, agora serão estimados os coeficientes. Na Tabela 11, são apresentadas as estimativas de máxima verossimilhança e os erros-padrão para o modelo ARIMA(0,1,1)x(0,1,1)[12] para a série de CO<sub>2</sub>. As estimativas do modelo são altamente significativas.

Figura 21 – ACF amostral da primeira diferença e diferença sazonal dos níveis de CO<sub>2</sub>.

Fonte: elaborada pelo autor.

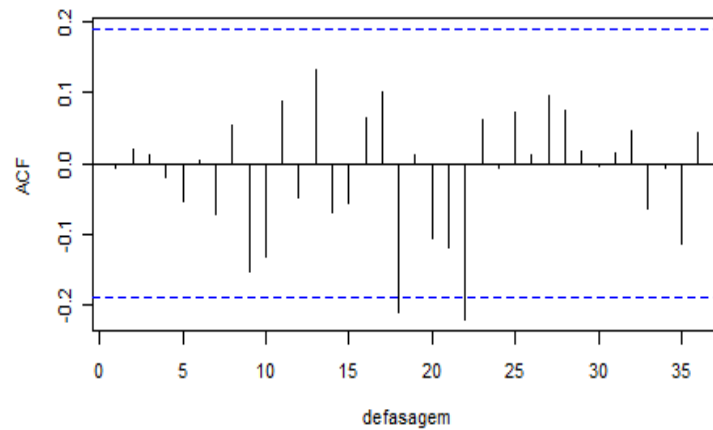
Apresentamos uma análise de diagnóstico do modelo estimado,  $ARIMA(0,1,1) \times (0,1,1)$  [12], em que a Figura 22 exibe o comportamento dos resíduos padronizados. Podemos notar que não há indícios de violação das pressuposições. Contudo, devemos investigar a existência de pontos influentes e/ou aberrantes, como por exemplo o valor observado em setembro de 1998.

Figura 22 – Resíduos para o modelo  $ARIMA(0,1,1) \times (0,1,1)$  [12].

Fonte: elaborada pelo autor.

Para obtermos uma interpretação melhor da qualidade do ajuste do modelo, foi feito o gráfico ACF para os resíduos padronizados da série, (ver Figura 23). Nela observamos que apenas a correlação de defasagem 22 apresenta um valor significativo com correlação de  $-0.19$ . Além disso, não é possível pensar em uma interpretação razoável para a dependência na defasagem 22. Por fim, não devemos nos preocupar com o fato da autocorrelação da defasagem 36 apresentar valor significativo pois isso pode ser facilmente resultado do acaso. Com exceção da defasagem 22, o modelo parece ter capturado a essência da dependência temporal da série.

Figura 23 – ACF para o modelo  $ARIMA(0,1,1) \times (0,1,1)[12]$



*Fonte: elaborada pelo autor.*

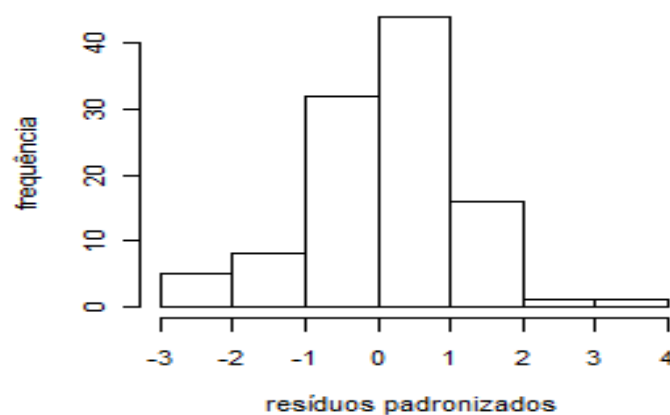
O teste Ljung-Box, apresentado na Tabela 12, mostra a estatística  $\chi^2$  com valor igual a 7.45, com 10 graus de liberdade, e um  $p$ -valor igual a 0.68, indicando que o modelo capturou o padrão apresentado pela série. Para investigarmos a questão da normalidade dos resíduos, apresentamos na Figura 24 o histograma dos resíduos.

Tabela 12 – Teste Ljung-Box. *Fonte: elaborada pelo autor.*

| $\chi^2$ | gl | p-value |
|----------|----|---------|
| 7.45     | 10 | 0.68    |

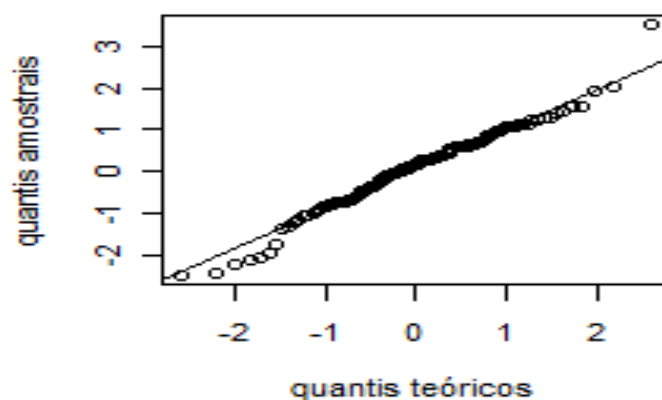
A Figura 24, apresenta uma *forma de sino*, porém aparentemente não é o ideal. Talvez analisando o gráfico quantil-quantil tenhamos mais informações sobre a forma da distribuição dos resíduos. Nesta gráfico, que apresentamos na Figura 25, observamos a presença de *outliers* na cauda superior, mas ao utilizarmos o teste de normalidade Shapiro-Wilks obtemos uma estatística  $W = 0.98$ , com um p-valor igual a 0.08. Desta forma, a normalidade dos resíduo não é rejeitada em qualquer nível de significância menor que 8%.

Figura 24 – Resíduos para o modelo  $ARIMA(0,1,1) \times (0,1,1)$ [12].



Fonte: elaborada pelo autor.

Figura 25 – Resíduos para o modelo  $ARIMA(0,1,1) \times (0,1,1)$ [12].

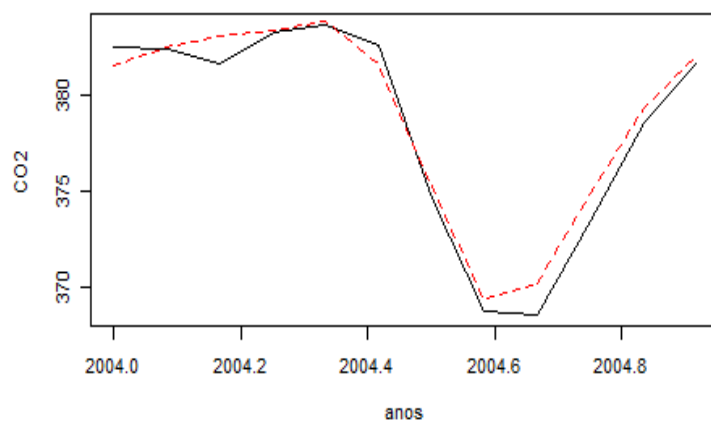


Fonte: elaborada pelo autor.

Após a realização da análise de diagnóstico do modelo  $ARIMA(0,1,1)(0,1,1)$ [12], chegamos a conclusão que ele mostrou-se adequado para o estudo da série mensal de

emissão de CO<sub>2</sub>. O passo seguinte é a realizar previsões para o ano de 2004 e em seguida compará-las aos resultados reais do *Grupo de Teste*. Adicionalmente, serão calculadas medidas de qualidade de previsão.

Figura 26 – Grupo de Teste (preto) e valores previstos (vermelho) com o modelo  $ARIMA(0,1,1) \times (0,1,1)$ [12].

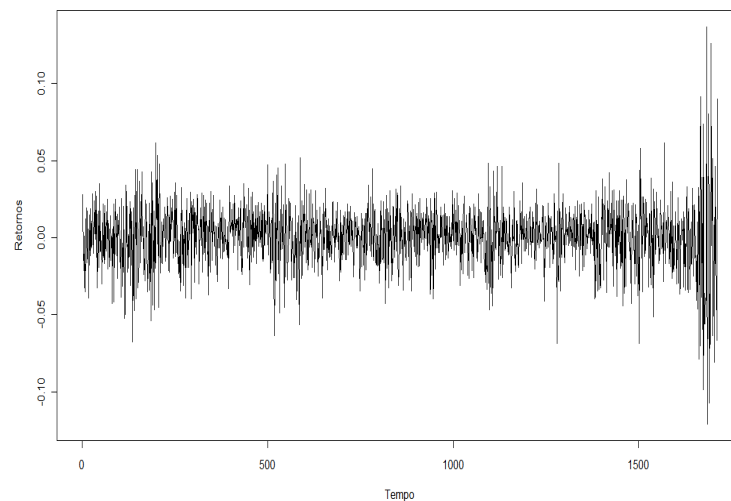


Fonte: elaborada pelo autor.

### 3.3.1.3 Análise da série do Ibovespa

Nas Figuras 4 e 27 são apresentados respectivamente os gráficos da série Ibovespa e a da série dos seus log-retornos (ou retornos). Da Figura 4 nota-se que a série apresenta uma forte tendência, ou seja, não deve apresentar estacionariedade, indicando que a média e possivelmente a variância não são constantes no período analisado (Janeiro/02 à Novembro/08). Nota-se ainda uma quebra estrutural ocorrida no final de 2008 (próximo à observação 1600 na figura 4) . Tal quebra é reflexo da crise mundial iniciada nos Estados Unidos, a qual teve reflexo em praticamente todos os países do mundo. Já na Figura 27, é exibido os log-retornos, em que podemos observar que esta série apresenta-se de forma estacionária, contudo podemos notar que os últimos log-retornos apresentam uma variância elevada.

Figura 27 – Gráfico dos retornos do Ibovespa.



Fonte: elaborada pelo autor.

A Tabela 13 apresenta as estatísticas básicas da série dos log-retornos de Ibovespa. Observa-se que a série apresenta uma possível assimetria à esquerda, pois o seu coeficiente de assimetria é menor do que zero e também apresenta curtose maior do que 3 aparentando que a distribuição dos log-retornos possui um cauda mais pesada do que a distribuição normal. Na Tabela 14 apresentamos os resultados do teste de aderência (Kolmogorov-Smirnov) para testar se podemos considerar para esta série a distribuição normal. Observando os resultados, existe evidências ao nível de 5% de que a séries dos log-retornos não segue uma distribuição normal.

Tabela 13 – Estatística básicas. Fonte: elaborada pelo autor.

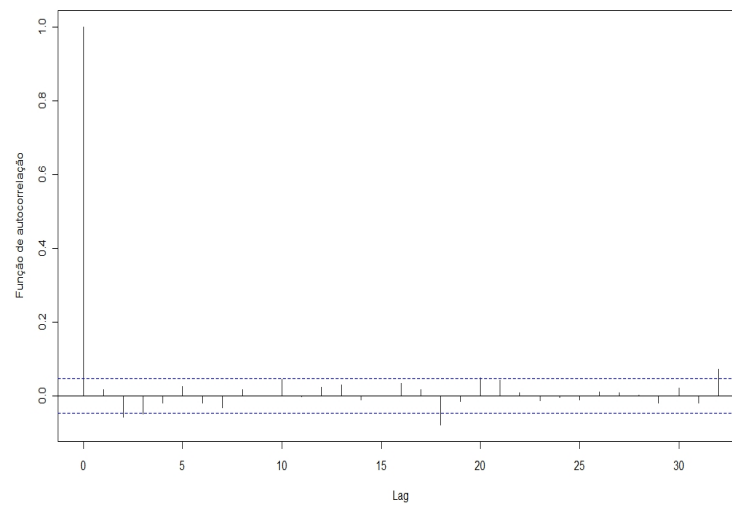
| Estatísticas           | Valores    |
|------------------------|------------|
| Mínimo                 | −0.1210000 |
| 1 <sup>o</sup> Quartil | −0.0099730 |
| Mediana                | 0.0013680  |
| Média                  | 0.0004747  |
| 3 <sup>o</sup> Quartil | 0.0102200  |
| Máximo                 | 0.1368000  |
| Assimetria             | −0.2067545 |
| Curtose                | 7.6159980  |

Tabela 14 – Teste de Kolmogorov-Smirnof para série dos log-retornos do Ibovespa. Fonte: elaborada pelo autor.

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0542                                  | 0.01           |

Iremos ajustar um modelo ARMA-GARCH na série dos log-retornos com distribuição  $t$ -Student, visto que essa distribuição apresenta uma cauda mais pesada, atendendo às evidências apresentadas. Na Figura 28 é apresentada a função de autocorrelação (acf) dos log-retornos. Analisando o gráfico da acf, existem *lags*, acima do intervalo de confiança indicando que a série não é ruído branco. Dessa forma, faremos o ajuste do modelo ARMA(2,2)-GARCH(1,1), com distribuição  $t$ -Student.

Figura 28 – Função de autocorrelação dos retornos.



Fonte: elaborada pelo autor.

Tabela 15 – Estimativas do Modelo *Fonte: elaborada pelo autor.*

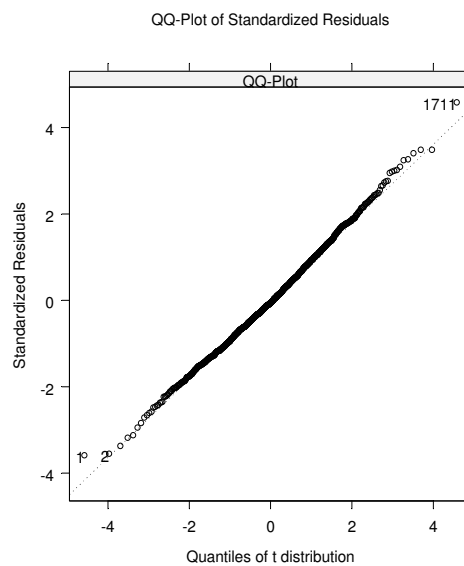
| Parâmetros | Estimativa  | Erro-Padrão | $t$ valor | $p$ -valor (%) |
|------------|-------------|-------------|-----------|----------------|
| $\phi_0$   | 0.00041396  | 0.0001971   | 2.1       | 1.792          |
| $\phi_1$   | 1.39670665  | 0.2748      | 5.083     | 2.064          |
| $\phi_2$   | -0.67507427 | 0.2293      | -2.944    | 0.16           |
| $\theta_1$ | -1.39990724 | 0.2845      | -4.921    | 0              |
| $\theta_2$ | 0.65139065  | 0.2448      | 2.661     | 0.0397         |
| $\alpha_0$ | 0.00000787  | 0.0000004   | 2.333     | 0.0987         |
| $\alpha_1$ | 0.07259116  | 0.01466     | 4.951     | 0              |
| $\beta_1$  | 0.90659293  | 0.02096     | 43.250    | 0              |

Os parâmetros do modelos são descritos na Tabela 15. Após o ajuste do modelo fizemos uma análise dos resíduos para verificar se as pressuposições são violadas. A Tabela 16 mostra o resultado do teste Ljung-Box para os resíduos e os quadrados do resíduos do modelo. De acordo com os resultados não rejeitamos a hipótese de que os resíduos e os quadrados do resíduos do modelo são ruídos branco.

Tabela 16 – Teste Ljung-Box. Fonte: elaborada pelo autor.

| r                      | Estimativa | $p$ -valor (%) |
|------------------------|------------|----------------|
| Resíduos               | 6.518      | 88.78          |
| Quadrados dos Resíduos | 8.85       | 71.57          |

Na Figura 29 temos o qq-plot dos resíduos do modelo sobre a distribuição  $t$ -Student com 13.19546 graus de liberdade (valor estimado). Podemos observar uma possível aproximação com a distribuição, apesar do primeiro ponto amostral se distanciar um pouco desta. Realizamos um teste de aderência (Kolmogorov-Smirnov) para testar se os resíduos seguem a distribuição  $t$ -Student. O resultado do teste Kolmogorov-Smirnov é apresentado na Tabela 17.

Figura 29 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição  $t$ -Student.

Fonte: elaborada pelo autor.

Tabela 17 – Teste do Kolmogorov-Smirnov. Fonte: elaborada pelo autor.

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0191                                  | 56.36          |

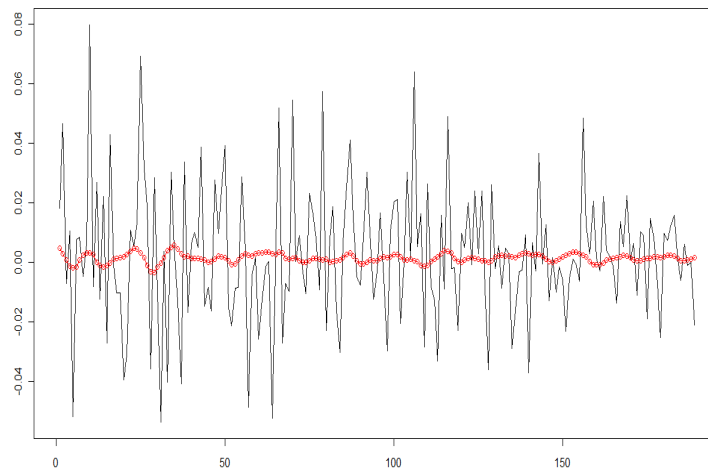
Os teste anteriores indicam que o modelo ajustado explica bem o comportamento da série dos log-retornos da Ibovespa. Assim, o modelo estimado é:

$$\begin{aligned} X_t &= 0.00041396 + 1.39670665X_{t-1} - 0.675087427X_{t-2} + a_t - 1.3990724a_{t-1} \\ &= +0.65139065a_{t-2} \\ a_t &= \sigma_t e_t \\ \sigma_t^2 &= 0.00000787 + 0.07259116a_t^2 + 0.90659293\sigma_{t-1}^2 \end{aligned}$$

onde  $e_t$  segue um distribuição  $t$ -Student padronizada com 13.19546 graus de liberdade.

As previsões realizadas utilizando o modelo ARMA-GARCH são apresentados na Figura 30. A linha contínua indica o valor da série dos log-retornos dos índices do Ibovespa e a linha central marcada com círculo indica as previsões.

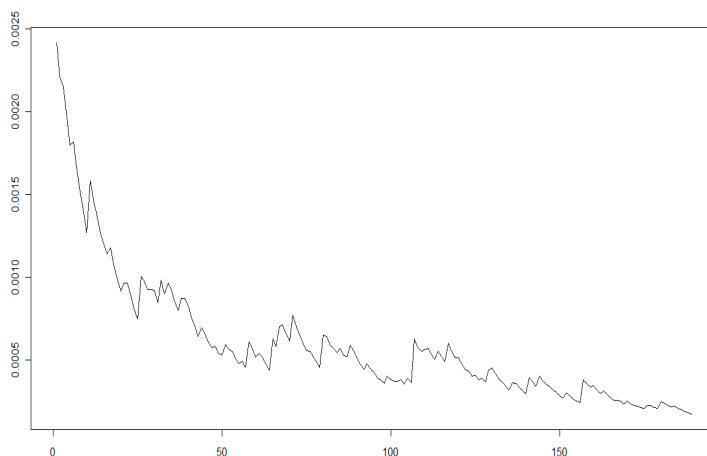
Figura 30 – Previsão obtida pelo modelo ARMA-GARCH para os log-retornos de Ibovespa.



Fonte: elaborada pelo autor.

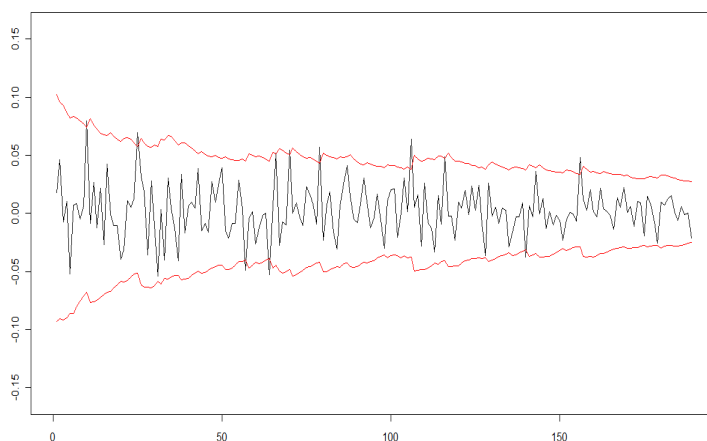
Na Figura 31 é mostrada a previsão da volatilidade, observa-se que durante esse período a volatilidade decai exponencialmente, devido a recuperação do mercado pós-crise. Na Figura 32 é apresentado o intervalo de confiança de 95% e a série original. Percebe-se que os valores da série estão na sua grande maioria dentro do intervalo de confiança, sendo portanto o modelo ajustado um bom modelo.

Figura 31 – Previsão da volatilidade.



Fonte: elaborada pelo autor.

Figura 32 – Intervalo de confiança.



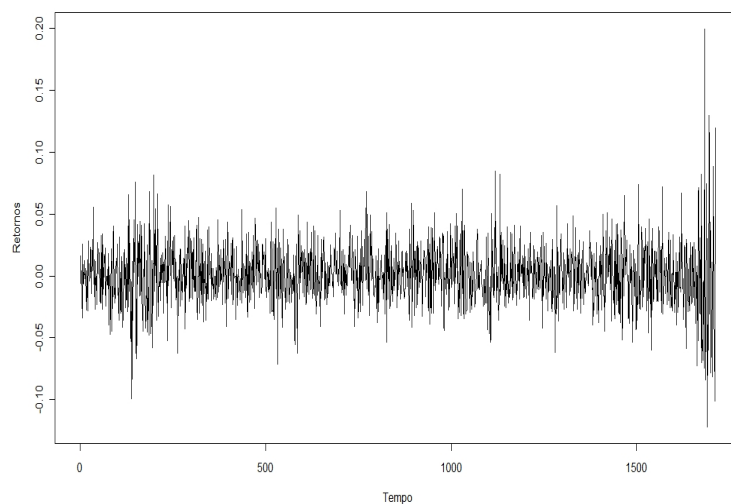
Fonte: elaborada pelo autor.

#### 3.3.1.4 Análise da série do Bradesco

As figuras 5 e 33 representam os gráficos da série original e dos log-retornos para a série de preços do Bradesco. Analisando os gráficos, podemos observar que a série é não-

estacionária e a não apresenta variância constante. Notamos também que a variância parece ser maior no último período da série.

Figura 33 – Gráfico dos retornos da série do Bradesco.



Fonte: elaborada pelo autor.

Tabela 18 – Estatísticas básicas. *Fonte: elaborada pelo autor.*

| Estatísticas           | Valores   |
|------------------------|-----------|
| Mínimo                 | -0.1222   |
| 1 <sup>o</sup> Quartil | -0.01328  |
| Mediana                | 0         |
| Média                  | 0.000792  |
| 3 <sup>o</sup> Quartil | 0.01464   |
| Máximo                 | 0.1998    |
| Assimetria             | 0.3566356 |
| Curtose                | 7.387346  |

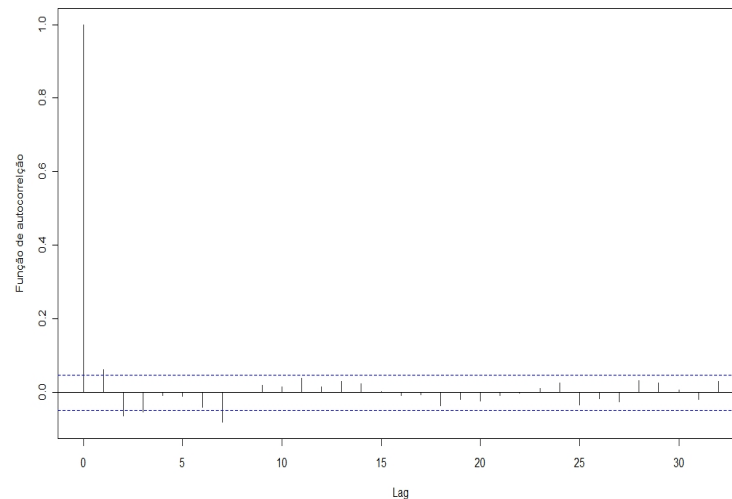
Apresentamos na Tabela 18 as estatísticas básicas da série dos log-retornos de Bradesco observa-se que a série apresenta uma possível assimetria à direita, pois o seu coeficiente de assimetria é maior do que zero e também apresenta curtose maior do 3 aparentando que a distribuição dos log-retornos possui um cauda mais pesada do que a distribuição normal. A Tabela 19 exibe o resultado do teste de aderência para testar se a série possui uma distribuição normal, portanto existe evidências ao nível de 5% de que a séries dos log-retornos não segue uma distribuição normal.

Tabela 19 – Teste do Kolmogorov-Smirnof para série dos log-retornos do Bradesco.  
Fonte: elaborada pelo autor.

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0391                                  | 1.06           |

Diante disso vamos ajustar um modelo ARMA-GARCH na série dos log-retornos com distribuição  $t$ -Student. A Figura 3.26 mostra a função de autocorrelação (a.c.f) dos log-retornos. Analisando o gráfico da a.c.f, nota-se que existem *lags* acima do intervalo de confiança portanto a série, indicando não se tratar de um ruído branco.

Figura 34 – Função de autocorrelação dos retornos da série Bradesco.



Fonte: elaborada pelo autor.

Inicialmente ajustamos um modelo ARMA(9,0)-GARCH(1,1) com distribuição  $t$ -Student para a série dos log-retornos do Bradesco. As estimativas dos parâmetros são mostrados na Tabela 20. Observa-se que existem parâmetros não significativos ao nível de 5%, portanto retiramos os parâmetros não significativos do modelo, os quais são apresentado na Tabela 21.

Tabela 20 – Estimativas do Modelo. *Fonte: elaborada pelo autor.*

| Parâmetros | Estimativa  | Erro-Padrão | $t$ valor | $p$ -valor (%) |
|------------|-------------|-------------|-----------|----------------|
| $\phi_0$   | 0.00120826  | 0.0005096   | 2.370     | 0.8937         |
| $\phi_1$   | 0.05721067  | 0.02575     | 2.222     | 1.320          |
| $\phi_2$   | -0.03291184 | 0.02581     | -1.4837   | 6.903*         |
| $\phi_3$   | -0.03773725 | 0.02417     | -1.5611   | 5.935*         |
| $\phi_4$   | -0.00683171 | 0.02523     | -0.2708   | 39.33*         |
| $\phi_5$   | -0.02049693 | 0.02502     | -0.8175   | 20.69*         |
| $\phi_6$   | -0.01402319 | 0.02439     | -0.575    | 28.27*         |
| $\phi_7$   | -0.03078086 | 0.0247      | -1.2460   | 10.65*         |
| $\phi_8$   | -0.00643094 | 0.0256      | -0.2512   | 40.09*         |
| $\phi_9$   | 0.04191817  | 0.02393     | 1.752     | 3.997          |
| $\alpha_0$ | 0.00001955  | 0.0000078   | 2.4981    | 0.6289         |
| $\alpha_1$ | 0.07991532  | 0.011726    | 4.6316    | 0.000002       |
| $\beta_1$  | 0.88516083  | 0.02776     | 31.8848   | 0              |

Tabela 21 – Estimativas do Modelo Final. *Fonte: elaborada pelo autor.*

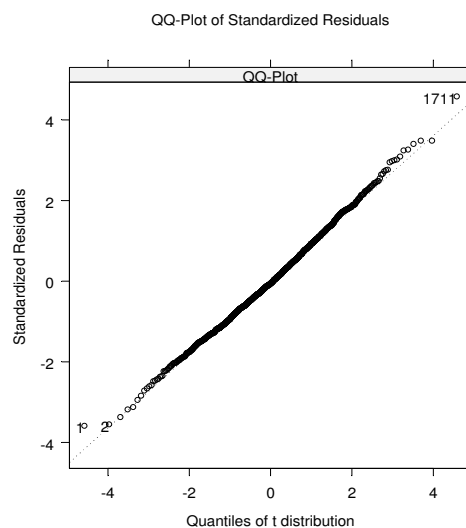
| Parâmetros | Estimativa | Erro-Padrão | $t$ valor | $p$ -valor (%) |
|------------|------------|-------------|-----------|----------------|
| $\phi_0$   | 0.00097983 | 0.0005028   | 2.370     | 2.574          |
| $\phi_1$   | 0.0595478  | 0.02551     | 2.234     | 0.9855         |
| $\phi_9$   | 0.04545008 | 0.0238      | 1.909     | 2.819          |
| $\alpha_0$ | 0.00001955 | 0.0000078   | 2.4981    | 0.6289         |
| $\alpha_1$ | 0.07900865 | 0.01718     | 4.6000    | 0.000002       |
| $\beta_1$  | 0.88602139 | 0.02787     | 31.796    | 0              |

Após o ajuste do modelo, fizemos uma análise dos resíduos, no qual vamos testar as pressuposições do modelo. A Tabela 22 mostra o resultado do teste Ljung-box para os resíduos e os quadrados dos resíduos do modelo, portanto não rejeitamos a hipótese de que os resíduos e os quadrados dos resíduos do modelo sejam ruídos branco.

Tabela 22 – Teste Ljung-Box. *Fonte: elaborada pelo autor.*

| r                      | Estimativas | p-valor (%) |
|------------------------|-------------|-------------|
| Resíduos               | 9.988       | 61.70       |
| Quadrados dos Resíduos | 12.86       | 37.98       |

Na Figura 3.27 temos o gráfico quantil quantil dos resíduos do modelo sobre a distribuição  $t$ -Student com 12.40431 graus de liberdade, podemos observar aparentemente uma boa aproximação com a distribuição. Fazendo o teste de aderência, Kolmogorov-Smirnof, para testar se os resíduos seguem a distribuição  $t$ -Student. O resultado do teste Kolmogorov-Smirnof é mostrado na Tabela 23, portanto não existem evidência ao nível de significância de 5% de que a distribuição dos resíduos do modelo ajustado segue uma distribuição  $t$ -Student com 12.40431 graus de liberdade.

Figura 35 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição  $t$ -Student.

*Fonte: elaborada pelo autor.*

Tabela 23 – Teste do Kolmogorov-Smirnof. *Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | p-valor (%) |
|-----------------------------------------|-------------|
| 0.0275                                  | 15.12       |

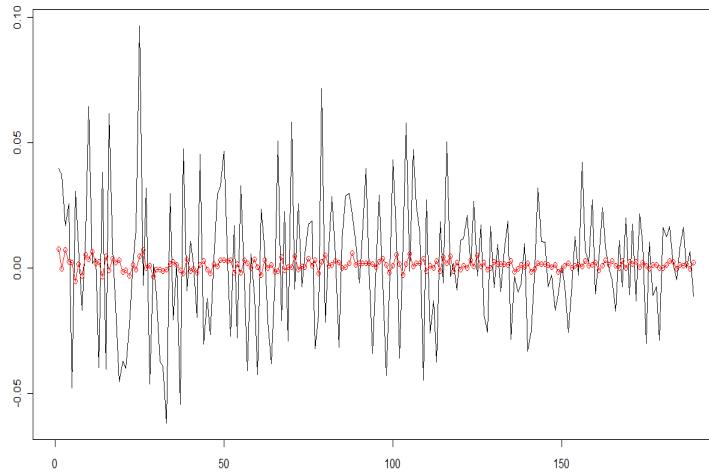
Os testes anteriores indicam que o modelo ajustado explicam bem o comportamento da série dos log-retornos da série de preço das ações do Bradesco. O modelo estimado é o seguinte:

$$\begin{aligned} X_t &= 0.00097983 + 0.0595478X_{t-1} + 0.04545008X_{t-9} + a_t \\ a_t &= \sigma_t e_t \\ \sigma_t^2 &= 0.00001955 + 0.07900865a_t^2 + 0.88602139\sigma_{t-1}^2, \end{aligned}$$

em que  $e_t$  segue um distribuição  $t$ -Student padronizada com 12.40431 graus de liberdade.

As previsões realizadas utilizando o modelo ARMA-GARCH são apresentados na Figura 36. A linha contínua indica o valor da série dos log-retornos do preço das ações do Bradesco e a linha central com marcado de círculo indica as previsões.

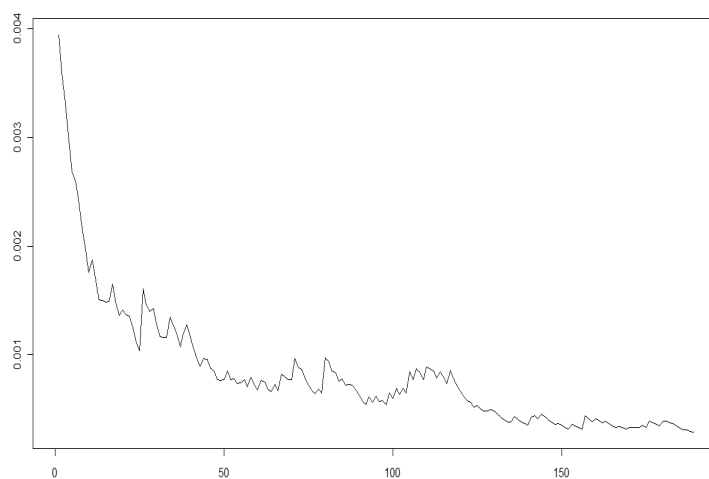
Figura 36 – Previsão obtida pelo modelo ARMA-GARCH para os log-retornos de Bradesco.



*Fonte: elaborada pelo autor.*

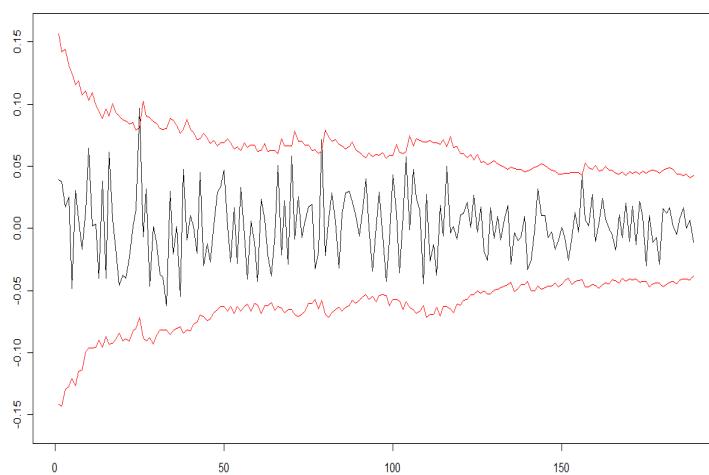
Na Figura 37 é mostrada a previsão da volatilidade, observa-se que durante esse período a volatilidade decai exponencialmente, devido a recuperação do mercado pós-crise. Na Figura 38 é apresentado o intervalo de confiança de 95% e a série original, percebe-se que os valores da série estão na sua grande maioria dentro do intervalo de confiança, portanto modelo ajustado é um bom modelo.

Figura 37 – Previsão da volatilidade.



*Fonte: elaborada pelo autor.*

Figura 38 – Intervalo de confiança.

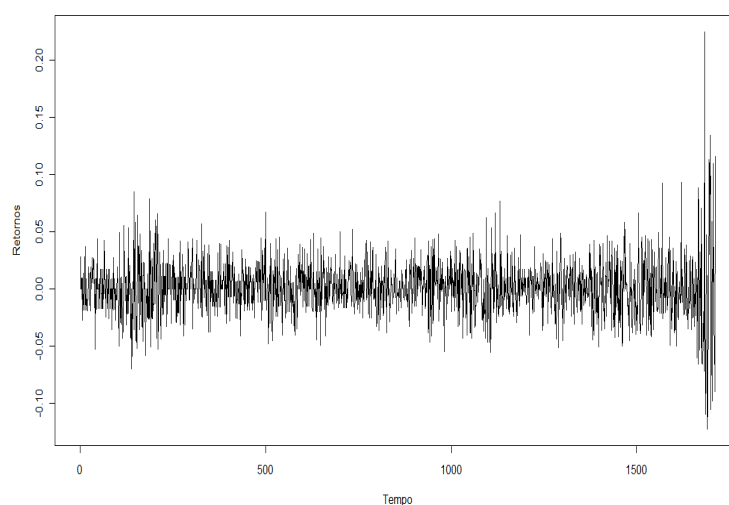


*Fonte: elaborada pelo autor.*

### 3.3.1.5 Análise da série do Itaú

Nas Figuras 8 e 39 são apresentados respectivamente, os gráficos da série Itaú e os seus log-retornos. Note que as séries Itaú e Bradesco são bem similares, uma vez que fazem parte do mesmo segmento e portanto os seus desempenhos seguem trajetórias comuns. Analisando os gráficos, percebemos que a série Itaú parece ser não-estacionária enquanto que a dos log-retornos aparentam apresentar estacionariedade.

Figura 39 – Gráfico dos retornos da série do Itaú.



*Fonte: elaborada pelo autor.*

Tabela 24 – Estatísticas básicas. *Fonte: elaborada pelo autor.*

| Estatísticas           | Valores   |
|------------------------|-----------|
| Mínimo                 | -0.12280  |
| 1 <sup>o</sup> Quartil | -0.01325  |
| Mediana                | 0         |
| Média                  | 0.0009721 |
| 3 <sup>o</sup> Quartil | 0.01432   |
| Máximo                 | 0.2249    |
| Assimetria             | 0.5240198 |
| Curtose                | 9.106308  |

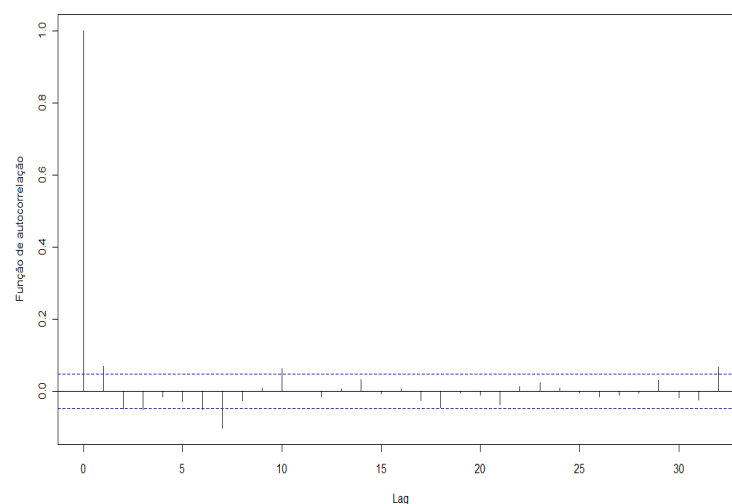
A Tabela 24 apresenta as estatísticas básicas da série dos log-retornos dos preços das ações do Itaúsa. Observa-se que esta série apresenta uma possível assimetria à direita, pois o seu coeficiente de assimetria é maior do que zero e também apresenta curtose maior do 3 aparentando que a distribuição dos log-retornos possui uma cauda mais pesada do que a da distribuição normal. Na Tabela 25 apresentamos o teste de aderência para testar se a série possui uma distribuição normal. A partir dos resultados obtidos com o teste Kolmogorov-Smirnov, chegamos a conclusão que existem evidências estatísticas ao nível de 5% de que a série dos log-retornos não segue uma distribuição normal.

Tabela 25 – Teste do Kolmogorov-Smirnov para série dos log-retornos de Itaúsa. *Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0514                                  | 0.02           |

Iremos ajustar um modelo ARMA-TGARCH na série dos log-retornos com distribuição  $t$ , pois a série apresenta uma assimetria relevante. Na Figura 3.32, é apresentada a função de autocorrelação (acf) dos log-retornos. Analisando este gráfico, observamos que existem defasagens, acima do intervalo de confiança, portanto, a série não é um ruído branco.

Figura 40 – Função de autocorrelação dos retornos da série Itaúsa.



*Fonte: elaborada pelo autor.*

Inicialmente, será ajustado um modelo ARMA(10,0)-TGARCH(1,1) com distribuição  $t$ -Student para a série dos log-retornos de Itaúsa. As estimativas dos parâmetros são mostrados na Tabela 26. Observa-se que existem parâmetros não significativos ao nível de 5%. Retirando os parâmetros não significativos obtemos os resultados apresentados na Tabela 27.

Tabela 26 – Estimativas dos coeficientes do modelo. *Fonte: elaborada pelo autor.*

| Parâmetros  | Estimativa  | Erro-Padrão | t valor | p-valor (%) |
|-------------|-------------|-------------|---------|-------------|
| $\phi_0$    | 0.00134951  | 0.0005294   | 2.5491  | 0.54        |
| $\phi_1$    | 0.01342077  | 0.02526     | 0.5314  | 29.76*      |
| $\phi_2$    | -0.02254002 | 0.02451     | -0.9195 | 17.90*      |
| $\phi_3$    | -0.04987605 | 0.02495     | -1.9987 | 2.29        |
| $\phi_4$    | -0.02453101 | 0.02516     | -0.9749 | 16.49*      |
| $\phi_5$    | -0.0393679  | 0.02524     | -1.56   | 5.947       |
| $\phi_6$    | -0.03094266 | 0.02514     | -1.2307 | 10.93*      |
| $\phi_7$    | -0.04962835 | 0.02447     | -2.0284 | 2.13        |
| $\phi_8$    | -0.00662033 | 0.02365     | -0.2799 | 38.98*      |
| $\phi_9$    | 0.01115974  | 0.02394     | 0.4661  | 32.06*      |
| $\phi_{10}$ | 0.05329267  | 0.02405     | 2.2164  | 1.34        |
| $\alpha_0$  | 0.00001051  | 0.000004    | 2.5360  | 0.56        |
| $\alpha_1$  | 0.02627581  | 0.01271     | 2.0678  | 1.94        |
| $\beta_1$   | 0.91925717  | 0.01695     | 54.2441 | 0           |
| $\gamma_1$  | 0.07547384  | 0.02162     | 3.4904  | 0.02        |

Tabela 27 – Estimativas do Modelo Final. *Fonte: elaborada pelo autor.*

| Parâmetros  | Estimativa  | Erro-Padrão | t valor | p-valor (%) |
|-------------|-------------|-------------|---------|-------------|
| $\phi_0$    | 0.00112662  | 0.0005092   | 2.211   | 1.36        |
| $\phi_3$    | -0.04813252 | 0.02487     | -1,935  | 2.65        |
| $\phi_7$    | -0.04852181 | 0.02432     | -1.995  | 2.31        |
| $\phi_{10}$ | 0.05813204  | 0.02409     | 2.413   | 0.79        |
| $\alpha_0$  | 0.00001037  | 0.000004    | 2.517   | 0.59        |
| $\alpha_1$  | 0.02335499  | 0.01253     | 1.864   | 3.12        |
| $\beta_1$   | 0.92097809  | 0.01670     | 55.145  | 0           |
| $\gamma_1$  | 0.07967913  | 0.02201     | 3.621   | 0.02        |

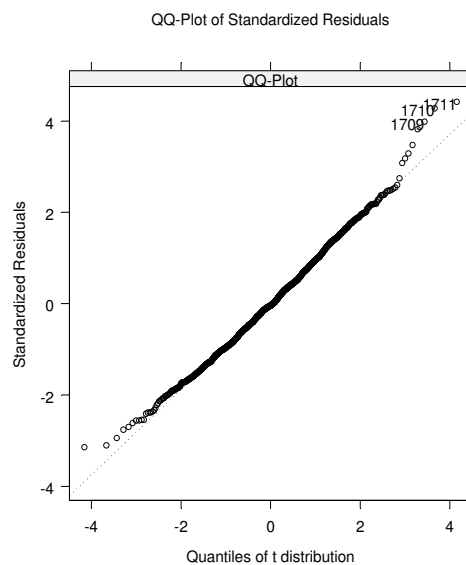
Após o ajuste do modelo será feita uma análise dos resíduos do modelo, no qual serão testadas as suas pressuposições. A Tabela 28 mostra o resultado do teste Ljung-box para os resíduos e os quadrados dos resíduos do modelo. De acordo com os resultados não rejeitamos a hipótese de que os resíduos e os quadrados dos resíduos do modelo sejam ruído branco.

Tabela 28 – Teste Ljung-Box. *Fonte: elaborada pelo autor.*

| r                      | Estimativa | p-valor (%) |
|------------------------|------------|-------------|
| Resíduos               | 5.989      | 91.67       |
| Quadrados dos Resíduos | 8.527      | 74.27       |

Na Figura 3.33 temos o gráfico quantil-quantil dos resíduos do modelo sobre a distribuição  $t$ -Student com 18.09087 graus de liberdade. Podemos observar aparentemente uma boa aproximação com a distribuição, porém a distribuição dos resíduos do modelo aparenta possuir um cauda direita mais pesada do que a  $t$ . Realizando um teste de aderência (Kolmogorov-Smirnof) para testar se os resíduos seguem a distribuição  $t$ -Student podemos concluir que não existe evidência ao nível de significância de 5% de que a distribuição dos resíduos do modelo ajustado siga uma distribuição  $t$ -Student com 12.40431 grau de liberdade. O resultado do teste Kolmogorov-Smirnof é mostrado na Tabela 29.

Figura 41 – Gráfico quantil quantil dos resíduos sobre a distribuição  $t$ -Student



*Fonte: elaborada pelo autor.*

Tabela 29 – Teste do Kolmogorov-Smirnov. *Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0246                                  | 25.32          |

Com os resultados obtidos dos testes anteriores temos uma indicação de que o modelo ajustado explica bem o comportamento da série dos log-retornos da Itausa. Assim, temos como modelo estimado:

$$X_t = 0.00112662 - 0.044813252X_{t-3} - 0.04852181X_{t-7} + 0.05813204X_{t-10} + a_t + a_{t-1}$$

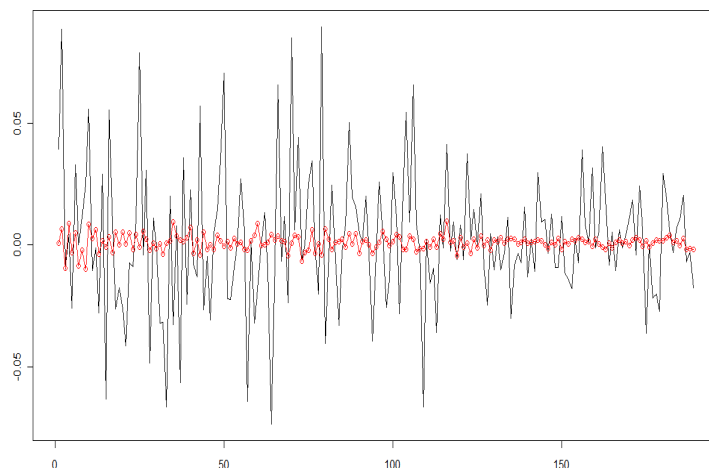
$$a_t = \sigma_t e_t$$

$$\sigma_t^2 = 0.00001037 + 0.02335499a_t^2 + 0.92097809\sigma_{t-1}^2 + 0.07967913a^2 d_t,$$

onde  $e_t$  segue uma distribuição  $t$ -Student com 18.09087 graus de liberdade e  $d$  é uma variável dummy que possui valor 0 se  $e_t \geq 0$  e 1 se  $e_t < 0$ .

As previsões realizadas utilizando o modelo ARMA-TGARCH são apresentadas na Figura 42. A linha contínua indica o valor da série dos log-retornos de Itausa e a linha central com marcado de círculo representa as previsões.

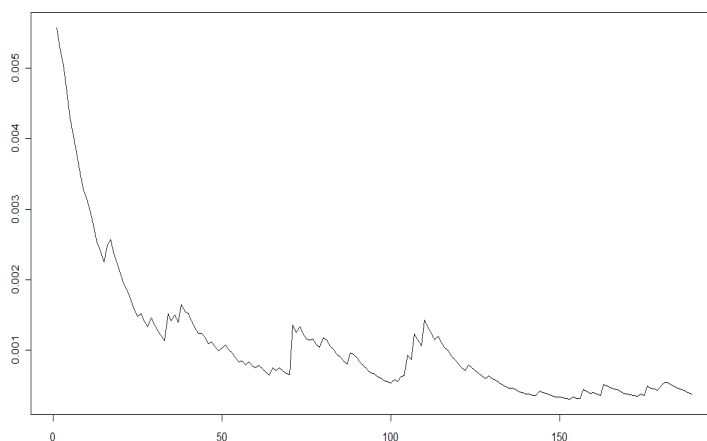
Figura 42 – Previsão obtida pelo modelo ARMA-TGARCH para os log-retornos da série Itausa.



*Fonte: elaborada pelo autor.*

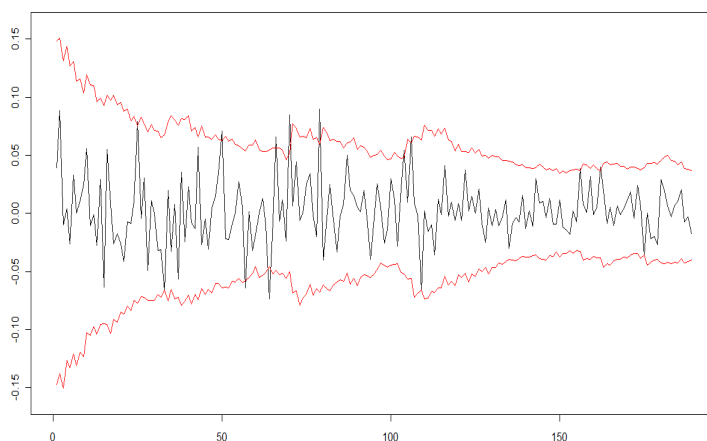
Na Figura 43 é mostrada uma previsão da volatilidade. Observa-se que durante esse período a volatilidade decai exponencialmente, devido a recuperação do mercado pós-crise. Na Figura 44 é apresentado o intervalo de confiança de 95% e a série original. Podemos perceber que os valores da série estão na sua grande maioria dentro do intervalo de confiança, sugerindo que o modelo ajustado parece ser adequado.

Figura 43 – Previsão da volatilidade.



*Fonte: elaborada pelo autor.*

Figura 44 – Intervalo de confiança.



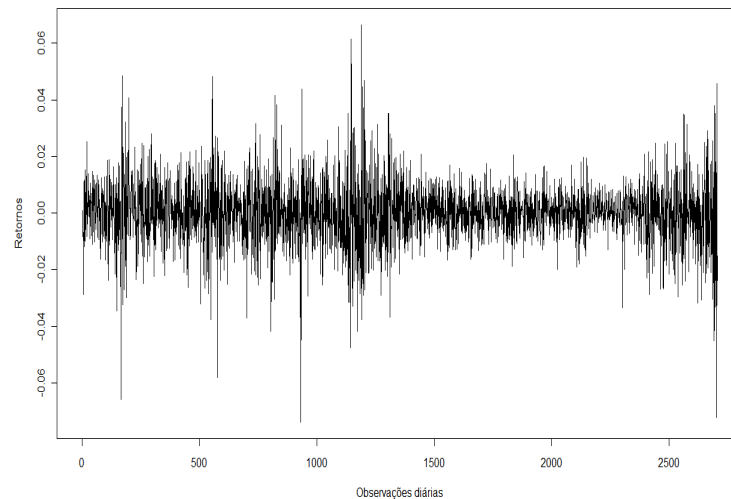
*Fonte: elaborada pelo autor.*

#### 3.3.1.6 Análise da série Dow Jones.

Os gráficos para a série do índice Dow Jones e os seus log-retornos, respectivamente podem ser vistos nas figuras 7 e 45. Na figura 7 notamos que a série apresenta tendência

para o período analisado (Janeiro/98 à Dezembro/05). Nota-se também quebras estruturais, observadas no final dos anos 2003 e 2008, em que esta última foi ocasionada por uma crise no setor imobiliário americano.

Figura 45 – Gráfico dos retornos da série do índice Dow Jones.



*Fonte: elaborada pelo autor.*

Tabela 30 – Estatísticas básicas. *Fonte: elaborada pelo autor.*

| Estatísticas           | Valores |
|------------------------|---------|
| Mínimo                 | -0.0740 |
| 1 <sup>o</sup> Quartil | -0.0057 |
| Mediana                | 0.0004  |
| Média                  | 0.0001  |
| 3 <sup>o</sup> Quartil | 0.0060  |
| Máximo                 | 0.0666  |
| Assimetria             | -0.1559 |
| Curtose                | 3.9596  |

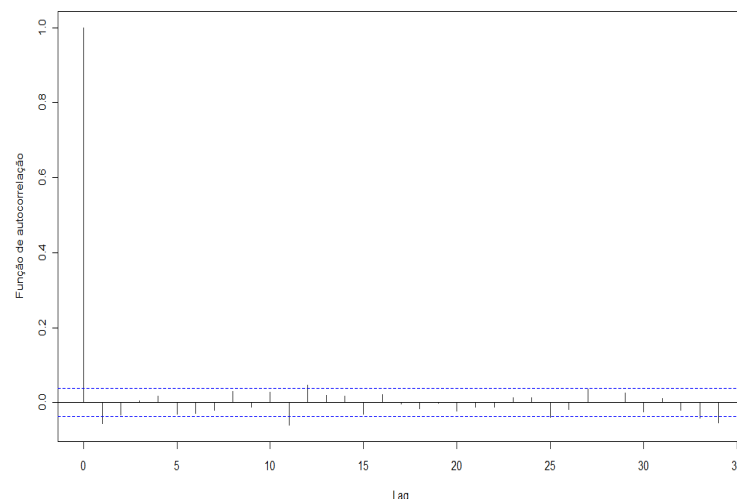
A Tabela 30 apresenta as estatísticas descritivas da série dos retornos do índice Dow Jones. Observa-se que esta série apresenta uma possível assimetria à esquerda, dado que seu coeficiente de assimetria é menor do que zero e também apresenta curtose maior do 3 indicando que a distribuição dos log-retornos possui uma cauda mais pesada do que a da distribuição normal. Na Tabela 31 apresentamos o teste de aderência.. A partir dos resultados obtidos com o teste Kolmogorov-Smirnov, chegamos a conclusão que existem evidências estatísticas ao nível de 5% que a séries dos retornos não segue uma distribuição normal.

Tabela 31 – Teste do Kolmogorov-Smirnof para série dos log-retornos de Dow Jones.  
*Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0574                                  | 0.0            |

O modelo ajustado à série foi um ARMA-TGARCH na série dos retornos com distribuição  $t$ , visto que a série apresenta assimetria. Na Figura 3.38, é apresentada a função de autocorrelação (acf) dos retornos. Analisando o gráfico, notamos que existem defasagens, acima do intervalo de confiança, levando a desconsiderar a hipótese de ruído branco.

Figura 46 – Função de autocorrelação dos retornos dos índices Dow Jones.



*Fonte: elaborada pelo autor.*

Propomos como modelo inicial um ARMA(12,0)-TGARCH(1,1) com distribuição  $t$ -Student para a série dos retornos. As estimativas dos parâmetros são mostradas na Tabela 32. Observa-se que existem parâmetros não significativos ao nível de 5%, levando-os conseqüentemente a serem excluídos. O resultado após a retirada pode ser visto na tabela 33.

Tabela 32 – Estimativas dos coeficientes do modelo. *Fonte: elaborada pelo autor.*

| Parâmetros  | Estimativa | Erro-Padrão | $t$ valor | $p$ -valor (%) |
|-------------|------------|-------------|-----------|----------------|
| $\phi_0$    | 0.000267   | 0.000167    | 1.5940    | 11.12*         |
| $\phi_1$    | -0.004209  | 0.02082     | -2.0215   | 4.33           |
| $\phi_2$    | -0.017370  | 0.01960     | -0.8864   | 37.55*         |
| $\phi_3$    | -0.002071  | 0.01980     | -0.1046   | 91.66*         |
| $\phi_4$    | 0.002071   | 0.01874     | 0.1105    | 91.19*         |
| $\phi_5$    | -0.031510  | 0.01897     | -1.6604   | 9.69*          |
| $\phi_6$    | -0.012870  | 0.01903     | -0.6762   | 49.89*         |
| $\phi_7$    | -0.007757  | 0.01871     | -0.4145   | 67.85*         |
| $\phi_8$    | 0.004816   | 0.01816     | 0.2651    | 79.09*         |
| $\phi_9$    | -0.010180  | 0.01894     | -0.5373   | 59.11*         |
| $\phi_{10}$ | 0.026690   | 0.01884     | 1.4168    | 15.67*         |
| $\phi_{11}$ | -0.01666   | 0.01842     | -0.904    | 36.59*         |
| $\phi_{12}$ | 0.050720   | 0.01866     | 2.7182    | 0.66           |
| $\alpha_0$  | 0.0000007  | 0.00000001  | 3.6923    | 0.023          |
| $\alpha_1$  | -0.004917  | -0.008799   | -0.5589   | 57.63*         |
| $\beta_1$   | 0.9148     | 0.007690    | 118.3187  | 0              |
| $\gamma_1$  | 0.1141     | 0.01364     | 8.3656    | 0              |

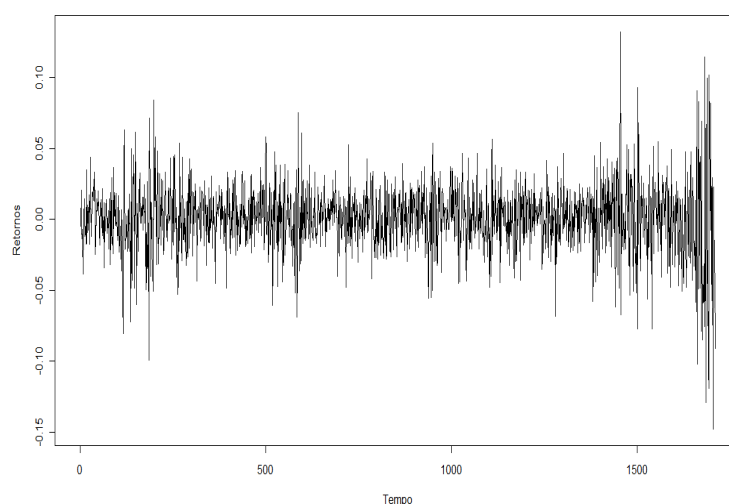
Tabela 33 – Estimativas do Modelo Final. *Fonte: elaborada pelo autor.*

| Parâmetros  | Estimativa   | Erro-Padrão | $t$ valor  | $p$ -valor (%) |
|-------------|--------------|-------------|------------|----------------|
| $\phi_0$    | 0.0002667    | 0.0001673   | 1.594      | 1.11           |
| $\phi_1$    | -0.04209     | 0.0208199   | -2.021     | 0.433          |
| $\phi_{12}$ | 0.050720     | 0.0195997   | 2.5880     | 0.09713        |
| $\alpha_0$  | 0.0000007151 | 0.0197961   | 0.00003612 | 1.00           |
| $\alpha_1$  | -0.004917    | 0.0187371   | -0.2624    | 7.93           |
| $\beta_1$   | 0.9418       | 0.0189743   | 49.640     | 0              |
| $\gamma_1$  | 0.1141       | 0.0190279   | 5.9980     | 0              |

### 3.3.1.7 Análise da série da Petrobras

Na figura 6 podemos observar a série de preços das ações da Petrobras. Como a série aparenta ser não-estacionária e também por se tratar de uma série financeira, devemos analisar os log-retornos como nas séries analisadas anteriormente. A Figura 47 apresenta a série dos retornos dos preços das ações da Petrobras. O comportamento dos retornos é semelhantes das séries Ibovespa, Bradesco e Itausa, no sentido de que a variabilidade parece aumentar no último período da série, ao que tudo indica por conta da crise ocorrida no final de 2008.

Figura 47 – Gráfico dos retornos da Petrobras.



*Fonte: elaborada pelo autor.*

Tabela 34 – Estatísticas descritivas. *Fonte: elaborada pelo autor.*

| Estatísticas           | Valores   |
|------------------------|-----------|
| Mínimo                 | -0.147939 |
| 1 <sup>o</sup> Quartil | -0.011969 |
| Mediana                | 0.001459  |
| Média                  | 0.000759  |
| 3 <sup>o</sup> Quartil | 0.014502  |
| Máximo                 | 0.132311  |
| Assimetria             | -0.320335 |
| Curtose                | 3.911537  |

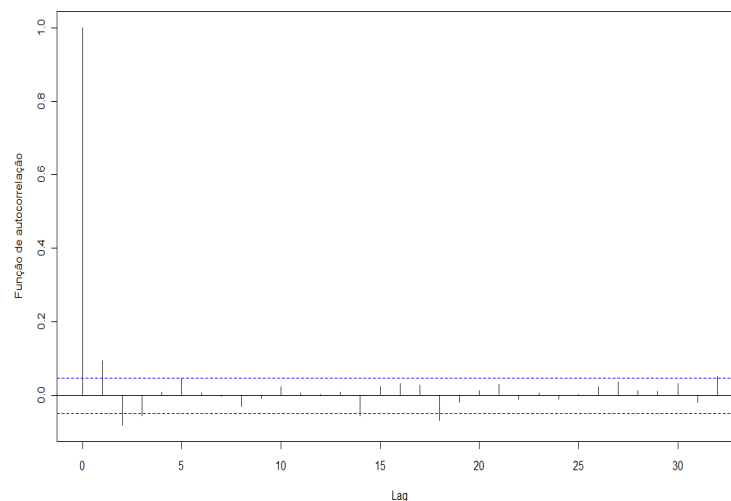
A Tabela 34 apresenta as estatísticas descritivas da série dos retornos dos preços das ações da Petrobras. De acordo com os resultados, esta série apresenta uma possível assimetria à esquerda já que o seu coeficiente de assimetria é menor do que zero. A curtose maior do três indica que a distribuição dos retornos possui cauda mais pesada do que a da distribuição normal. Na Tabela 35 apresentamos o resultados do teste de aderência. A partir dos resultados obtidos com o teste Kolmogorov-Smirnov, chegamos a conclusão de que existem evidências estatísticas ao nível de 5% de que a série dos retornos não segue uma distribuição normal.

Tabela 35 – Teste do Kolmogorov-Smirnof para série dos retornos da Petrobras. *Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0572                                  | 0              |

Ajustamos um modelo ARMA-TGARCH à série dos retornos com distribuição  $t$ -Student. Na Figura 3.40 mostramos a função de autocorrelação dos retornos. Analisando o gráfico da acf, notamos que existem *lags* acima do intervalo de confiança, indicando não se tratar da função de autocorrelação de um ruído branco.

Figura 48 – Função de autocorrelação dos retornos da série Petrobras.



*Fonte: elaborada pelo autor.*

O modelo ajustado para os dados foi um ARMA(2,0)-TGARCH(1,1) com distribuição  $t$ -Student para a série dos retornos. As estimativas dos parâmetros são mostrados na Tabela 36. Observa-se que todos os parâmetros são significativos ao nível de 5%.

Tabela 36 – Estimativas do Modelo. *Fonte: elaborada pelo autor.*

| Parâmetros | Estimativa  | Erro-Padrão | $t$ valor | $p$ -valor (%) |
|------------|-------------|-------------|-----------|----------------|
| $\phi_0$   | 0.00146732  | 0.0004700   | 3.122     | 0.182599       |
| $\phi_1$   | 0.09123592  | 0.02581     | 3.535     | 0.041811       |
| $\phi_2$   | -0.06073933 | 0.02456     | -2.473    | 1.349182       |
| $\alpha_0$ | 0.00002534  | 0.00006072  | 4.174     | 0.003148       |
| $\alpha_1$ | 0.02491858  | 0.01679     | 1.484     | 13.78          |
| $\beta_1$  | 0.87339364  | 0.02409     | 35.313    | 0              |
| $\gamma_1$ | 0.13890979  | 0.03142     | 4.422     | 0              |

Após o ajuste do modelo, fizemos uma análise dos resíduos, no qual vamos testar os pressupostos do modelo. A Tabela 37 mostra o resultado do teste Ljung-box para os resíduos e os quadrados dos resíduos do modelo. De acordo com o resultado não podemos rejeitar a hipótese de que os resíduos e os quadrados dos resíduos do modelo sejam ruídos branco.

Tabela 37 – Teste Ljung-Box. *Fonte: elaborada pelo autor.*

| r                      | Estimativas | $p$ -valor (%) |
|------------------------|-------------|----------------|
| Resíduos               | 7.434       | 82.77          |
| Quadrados dos Resíduos | 5.981       | 91.71          |

Na Figura 3.41 temos o gráfico quantil quantil dos resíduos do modelo sobre a distribuição  $t$ -Student com 10.13615 graus de liberdade. Podemos observar aparentemente uma boa aproximação com a distribuição.

O resultado do teste Kolmogorov-Smirnov é mostrado na Tabela 38. Não existe evidência ao nível de significância de 5% de que a distribuição dos resíduos do modelo ajustado segue uma distribuição  $t$ -Student com 10.13615 graus de liberdade.

Tabela 38 – Teste do Kolmogorov-Smirnov. *Fonte: elaborada pelo autor.*

| Estatística do teste Kolmogorov-Smirnov | $p$ -valor (%) |
|-----------------------------------------|----------------|
| 0.0122                                  | 96.0           |

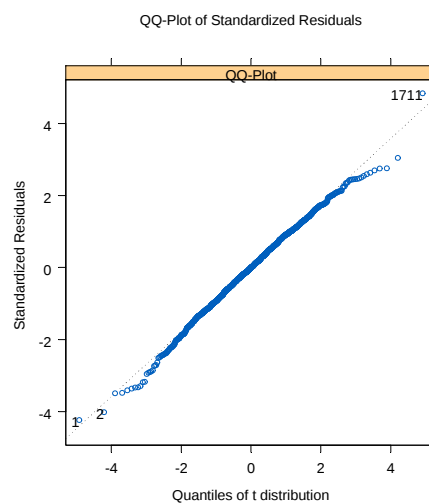


Figura 49 – Gráfico quantil quantil dos resíduos padronizados sobre a distribuição  $t$ -Student. *Fonte: elaborada pelo autor.*

Os testes anteriores indicam que o modelo ajustado explicam bem o comportamento da série dos retornos da Petrobras. O modelo estimado é:

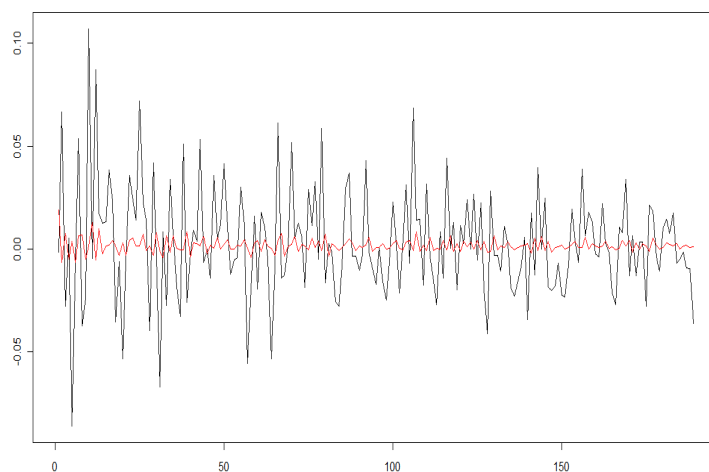
$$\begin{aligned}
 X_t &= 0.00146732 + 0.09123592X_{t-1} - 0.06073933X_{t-2} + a_t \\
 a_t &= \sigma_t e_t \\
 \sigma_t^2 &= 0.00002534 + 0.02491858a_t^2 + 0.85068561\sigma_{t-1}^2 + 0.13890979a^2 d_t,
 \end{aligned}$$

em que  $e_t$  segue uma distribuição  $t$ -Student padronizada com 10.13615 graus de liberdade e  $d$  é uma variável dummy que apresenta valor 0 se  $e_t > 0$  e 1 se  $e_t < 0$ .

As previsões realizadas utilizando o modelo ARMA-TGARCH são apresentadas na Figura 50. A linha de cor preta indica o valor da série dos retornos do preço das ações da Petrobras e a linha de cor vermelha indica as previsões obtidas a partir do modelo ajustado.

Na Figura 51 é mostrada a previsão da volatilidade. Na Figura 52 é apresentado o intervalo de confiança de 95% e a série original. Podemos observar que os valores da

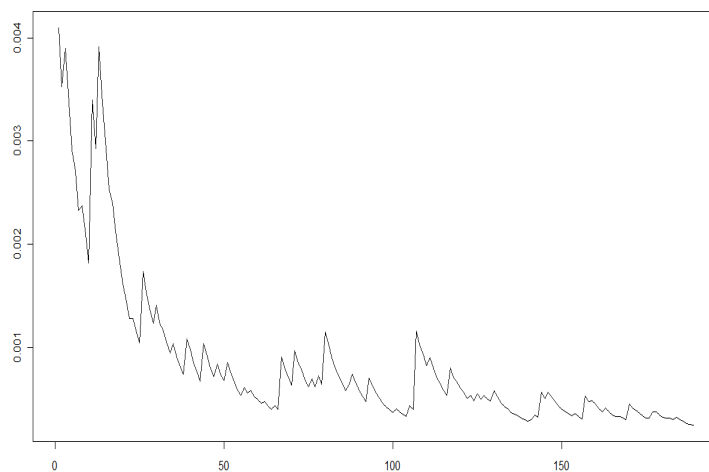
Figura 50 – Previsão obtida pelo modelo ARMA-TGARCH para os retornos da Petrobras.



*Fonte: elaborada pelo autor.*

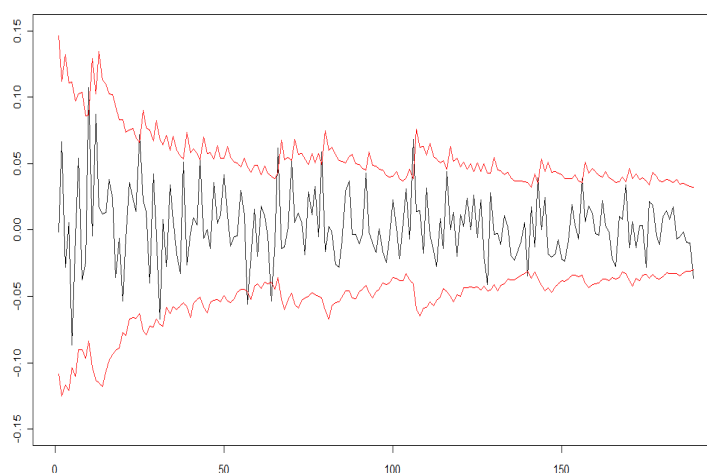
série estão na sua grande maioria dentro do intervalo de confiança indicando um bom ajuste do modelo.

Figura 51 – Previsão da volatilidade.



*Fonte: elaborada pelo autor.*

Figura 52 – Intervalo de confiança.



*Fonte: elaborada pelo autor.*

### 3.3.2 Análise através de Redes Neurais Artificiais Dinâmicas

#### 3.3.2.1 Rede Neural Artificial Dinâmica (RNAD)

As séries matemáticas e estatísticas, como o próprio nome diz, apresentam características ou padrões que modificam de acordo com o passar do tempo. Sendo que a forma em que os valores variam no tempo não caracteriza uma única função e sim um conjunto delas sobrepostas. Devido a estas características procura-se utilizar redes neurais com características dinâmicas para acompanhar a variabilidade dos valores de acordo com o tempo De Azevedo (1994).

Devido às características dinâmicas das séries matemáticas, neste trabalho propõem-se a utilização de Redes Neurais Artificiais Dinâmicas (RNAD), mais especificamente uma RNAD com dinâmica de entrada. Neste tipo de RNA existe uma malha de atraso na entrada da rede com a finalidade de criar o efeito de memória na rede, ou seja, fazer com que a saída da RNA dependa do elemento anterior ( $t - 1$ ). A topologia desse tipo de RNA pode ser visualizada através da figura 53.

Figura 53 – Topologia de uma Rede Neural com Dinâmica de entrada

figure=rnad.pdf, angle=90, scale=0.30

As características dinâmicas deste tipo de rede estão nos elementos de entrada da RNA. A malha de atraso da entrada faz com que todos os pontos da série passem em cada um dos neurônios de entrada da rede, sendo que, o último ponto é eliminado após ser computado o seu valor pela RNA. Este procedimento se repete até o final da série ou do padrão a ser analisado. A principal característica desta RNA é a de verificar a

maneira com que os pontos variam no decorrer do tempo e não a disposição dos pontos no plano, como é o caso da RNA estática, facilitando desta forma, a identificação de sinais variantes no tempo, como é o caso das séries aleatórias. Portanto, pode-se dizer que a RNA estática é indicada para sinais estáticos (onde o padrão se encontra estável ou não variante no tempo); em contrapartida, a RNAD é indicada para sinais variantes no tempo (Rodrigues *et al.*, 1999 e Rodrigues, 2003).

A entrada da RNAD, utilizada nesta tese, é constituída de quatro elementos de atraso (quatro valores da série) e, desta forma, quatro neurônios de entrada, interligados a cada elemento de atraso. A Camada intermediária foi composta de 50 neurônios, todos interconectados com os elementos da camada de entrada. Essa quantidade de neurônios foi escolhida com base no conjunto de treinamento a ser utilizado pela RNAD, em torno de 1700 elementos. Não existe uma regra para a definição da quantidade de neurônios na camada intermediária, mas é comprovado que a matriz de pesos da entrada, composta da quantidade de neurônios de entrada e de neurônios da camada intermediária é responsável pela capacidade de armazenamento de informações durante o treinamento. Quanto menor essa matriz de pesos, menor será a capacidade da RNA se adaptar a uma solução para um determinado conjunto de treinamento. Porém, o aumento em excessivo dessa matriz com a utilização de um conjunto de treinamento pequeno, pode fazer com que a rede “memorize” alguns desses valores, fazendo com que a RNA perca o poder de generalização. Neste caso muitos pesos sinápticos ficam sem utilização, apenas caminhos por alguns pesos são especializados para gerar a saída desejada. Portanto, existe um comprometimento entre o conjunto de treinamento e a quantidade de neurônios de cada camada da RNA. A quantidade de neurônio deve ser dimensionada segundo a complexidade do problema e com a quantidade de elementos a serem treinados. Pois, se a quantidade de neurônios for pequena e conjunto de treinamento muito grande, a RNA não conseguira apresentar um valor de saída entre os parâmetros mínimos de erro desejados, isto é, vai apresentar um erro maior que o desejado (Rodrigues *et al.*, 2001).

A saída da RNA é composta de apenas um neurônio, representando um valor de saída para cada quatro elementos da série de entrada. Portanto, a RNA gera um vetor de saída, durante a fase de treinamento, menor em quatro unidades que o vetor treinado.

Todos os neurônios utilizados para compor a RNAD deste trabalho possuem a função de ativação do tipo bipolar (função tangente hiperbólica) e o algoritmo que foi utilizado foi o backpropagation. A implementação do algoritmo foi realizada com o auxílio do software MATLAB, com o auxílio do toolbox Neural Networks The Mathworks (1996).

## 3.3.2.2 Séries e treinamento

Para o treinamento foi utilizado uma parte da série original, a outra parte foi utilizada para a validação da RNAD. Já para a criação das RNAs foi definido neste trabalho um valor fixo de elementos a serem considerados na determinação do próximo valor da série pela RNA, adotou-se quatro elementos em instantes de tempo anteriores ( $t - 4, t - 3, t - 2, t - 1$ ) para a entrada da RNAD, que geram um elemento de saída da RNA (instante  $t$ ). Assim o número de neurônios da entrada das RNAs é quatro, igual ao número de elementos de retardo. Logo, foram utilizados quatro pontos da série em sequência como valor de entrada para o treinamento e um quinto elemento da sequência como padrão de saída, isto é, o padrão desejado para sequência de entrada. O treinamento ocorreu sempre incrementando um ponto da série, porém mantendo os outros três deslocados no tempo. Conforme mostra o exemplo na Tabela 39.

Tabela 39 – Exemplo de valores de entrada e saída para as RNAD's. *Fonte: elaborada pelo autor.*

| Entrada 1 da RNAD | Entrada 2 da RNAD | Entrada 3 da RNAD | Entrada 4 da RNAD | Saída da RNA |
|-------------------|-------------------|-------------------|-------------------|--------------|
| Valor 1           | Valor 2           | Valor 3           | Valor 4           | Valor 5      |
| Valor 2           | Valor 3           | Valor 4           | Valor 5           | Valor 6      |
| Valor 3           | Valor 4           | Valor 5           | Valor 6           | Valor 7      |

A atualização dos pesos da rede é realizada a cada ponto, de acordo com a entrada dos pontos pela malha de atraso. O padrão alvo para saída da rede apresentado na Tabela 39 somente é alterado após a rede ter realizado o treinamento com a sequência de entrada.

Para cada série treinada não foi estipulado um número de épocas máximo para o treinamento. O treinamento das redes neurais foi realizado até que o critério do erro médio quadrático (MSE) fosse satisfeito, utilizando-se uma taxa de aprendizado fixa da RNAD (larning rate) igual a 0.01. Foram treinadas uma RNAD para cada série, devido as suas peculiaridades.

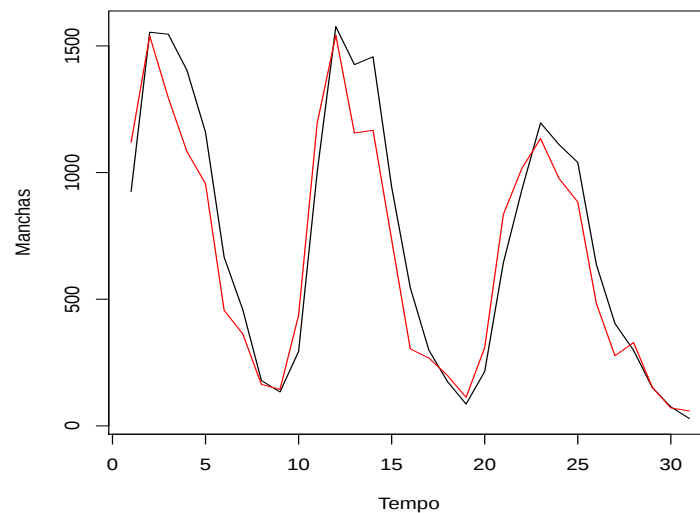
A etapa de validação da RNAD foi realizada, da mesma forma que no treinamento, porém são apresentados 4 pontos para entrada da rede, esperando a previsão do quinto ponto pela RNAD.

### 3.3.2.3 Resultados Experimentais

#### 3.3.2.3.1 Série Manchas Solares

Na Figura 54 mostramos um gráfico comparativo entre a série real (linha preta) e a série prevista (linha vermelha) pela RNA, para o grupo de teste. Observamos que as RNADs capturam o comportamento do fenômeno gerador da série, apresentando previsões bastante próximas dos valores reais.

Figura 54 – Comparativo entre os valores reais e as previsões - Manchas Solares.



Fonte: elaborada pelo autor.

Na Tabela 40 apresentamos um comparativo entre os modelos selecionados pelas RNADs e pela metodologia Box e Jenkins. Nesta Tabela, observa-se que o modelo escolhido pelo método de RNAs apresenta desempenho superior ao modelo escolhido pela metodologia Box e Jenkins para todas as medidas de performance.

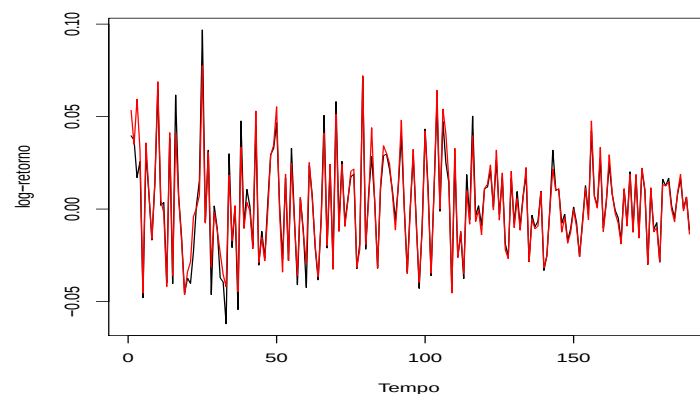
Tabela 40 – Medidas de desempenho de previsão para os modelo analisado - Manchas Solares. *Fonte: elaborada pelo autor.*

| Medidas | RNAD     | Box & Jenkins | Melhor desempenho |
|---------|----------|---------------|-------------------|
| MSE     | 248.6455 | 380.3000      | RNAD              |
| MAPE    | 22.27%   | 59.85%        | RNAD              |
| U Theil | 0.3116   | 0.7415        | RNAD              |
| POICD   | 93.10%   | 79.31%        | RNAD              |
| ARV     | 0.00434  | 0.0070        | RNAD              |
| AIC     | 171.9600 | 192.2879      | RNAD              |
| BIC     | 199.0600 | 214.6975      | RNAD              |

### 3.3.2.3.2 Série Bradesco

Na Figura 55 apresentamos o gráfico para o grupo de teste da série Bradesco e suas respectivas previsões usando um modelo de RNADs. Nesta Figura é possível notar que as previsões estão bem próximas dos valores reais. Novamente observamos que o modelo de RNADs conseguiu capturar bem o comportamento apresentado pela série de retornos do Bradesco.

Figura 55 – Comparativo entre os valores reais e as previsões - Bradesco.



*Fonte: elaborada pelo autor.*

Na Tabela 41 são apresentadas as medidas de desempenho dos modelos escolhidos pela metodologia Box e Jenkins e RNAs. Nessa Tabela, assim como no caso da série das Manchas Solares, observa-se que o modelo escolhido pela RNAs apresenta desempenho superior ao modelo escolhido pela metodologia Box e Jenkins para todas as medidas de performance.

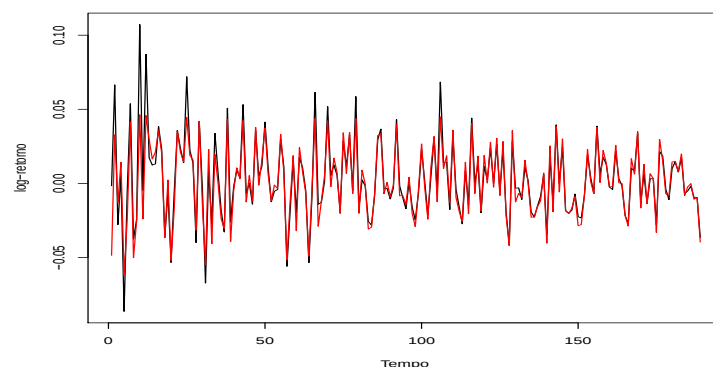
Tabela 41 – Medidas de desempenho de previsão para os modelo analisado - Bradesco.  
Fonte: elaborada pelo autor.

| Medidas | RNAD     | Box & Jenkins | Melhor desempenho |
|---------|----------|---------------|-------------------|
| MSE     | 0.000038 | 0.0007        | RNAD              |
| MAPE    | 30.29%   | 105.8400%     | RNAD              |
| U Theil | 0.0268   | 74.1866       | RNAD              |
| POICD   | 96.30%   | 31.22%        | RNAD              |
| ARV     | 0.000294 | 0.7263        | RNAD              |
| AIC     | -1913.70 | -1361.1524    | RNAD              |
| BIC     | -1980.83 | -1410.7446    | RNAD              |

### 3.3.2.3.3 Série Petrobras

Assim como nos demais casos, podemos ver no gráfico da figura 56 as previsões feitas pelo modelo de RNADs para a série de retornos dos preços das ações da Petrobras e compará-las com os valores reais. Percebe-se mais uma vez que as previsões (linha vermelha) acompanham os valores reais caracterizando uma boa adequabilidade do modelo.

Figura 56 – Comparativo entre os valores reais e as previsões - Petrobras.



Fonte: elaborada pelo autor.

Na Tabela 42 apresentamos um comparativo entre os modelos selecionados pelas RNAs e pela metodologia Box e Jenkins. Podemos notar que o modelo de RNAD apresenta um desempenho superior à metodologia Box e Jenkins para todas as medidas de performance. Observamos que medidas importantes como o MSE com valores inferiores ao do caso dos modelos econométricos ratificam a superioridade das RNADs.

Tabela 42 – Medidas de desempenho de previsão para os modelo analisado - Petrobras.

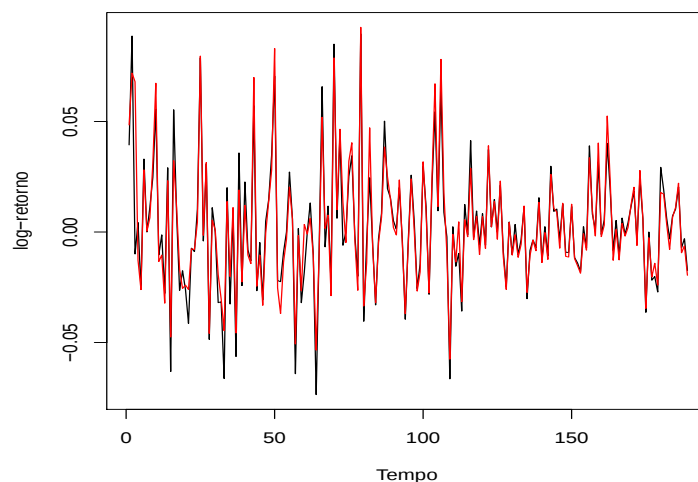
*Fonte: elaborada pelo autor.*

| Medidas | RNAD         | Box & Jenkins | Melhor desempenho |
|---------|--------------|---------------|-------------------|
| MSE     | 0.00009      | 0.0007        | RNAD              |
| MAPE    | 77.10%       | 120.14%       | RNAD              |
| U Theil | 0.0662       | 23.7510       | Box & Jenkins     |
| POICD   | 95.24%       | 36.51%        | RNAD              |
| ARV     | 0.000784     | 0.3128        | RNAD              |
| AIC     | -1749.0246   | 11.9791       | RNAD              |
| BIC     | -1810.930233 | -37.2887      | RNAD              |

#### 3.3.2.3.4 Série Itausa

Na Figura 57 mostramos um gráfico comparativo entre a série real (linha preta) e a série prevista (linha vermelha) pela RNAD, para a série dos retornos do Itausa. Nota-se que os valores previstos sobrepõem à série real, indicando uma grande proximidade entre os valores reais e previstos.

Figura 57 – Comparativo entre os valores reais e as previsões - Itausa.



*Fonte: elaborada pelo autor.*

Na Tabela 43 apresentamos um comparativo entre os modelos selecionados pelas RNAs e pela metodologia Box & Jenkins. Tal qual os demais casos, o método de RNADs apresenta desempenho superior ao modelo escolhido pela metodologia Box & Jenkins para todas as medidas de performance.

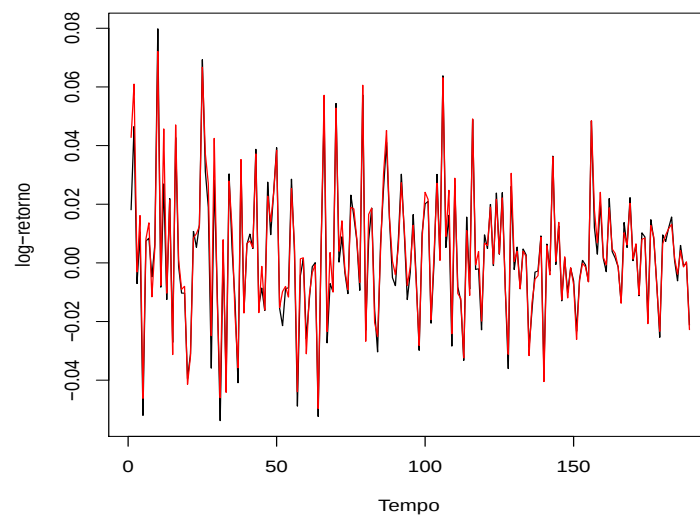
Tabela 43 – Medidas de desempenho de previsão para os modelo analisado - Itausa.  
*Fonte: elaborada pelo autor.*

| Medidas | RNAD       | Box & Jenkins | Melhor desempenho |
|---------|------------|---------------|-------------------|
| MSE     | 0.0001     | 0.0008        | RNAD              |
| MAPE    | 47.8000%   | 99.5300%      | RNAD              |
| U Theil | 0.0569     | 33.7562       | RNAD              |
| POICD   | 95.7700%   | 53.9700%      | RNAD              |
| ARV     | 0.000574   | 0.3634        | RNAD              |
| AIC     | -1770.5800 | -1325.8735    | RNAD              |
| BIC     | -1833.1600 | -1448.4825    | RNAD              |

### 3.3.2.3.5 Série Ibovespa

Na Figura 58 pode ser visto um gráfico comparativo entre a série real (linha preta) e a série prevista pela RNAD (linha vermelha) para o grupo de teste da série dos retornos do Ibovespa. Aqui também há uma proximidade acentuada entre os valores reais e previstos, demonstrando a uma superioridade dos modelos de redes neurais para realizar previsões desta série.

Figura 58 – Comparativo entre os valores reais e as previsões - Ibovespa.



*Fonte: elaborada pelo autor.*

Na Tabela 3.37 apresentamos um comparativo entre os modelos selecionados pelas RNADs e pela metodologia Box & Jenkins. Mais uma vez o método de RNADs apresenta desempenho superior ao modelo escolhido pela metodologia Box & Jenkins.

Tabela 44 – Medidas de desempenho de previsão para os modelo analisado - Ibovespa.

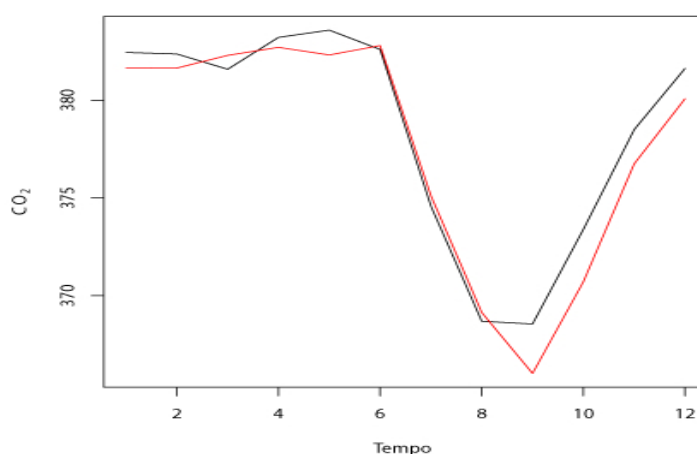
*Fonte: elaborada pelo autor.*

| Medidas | RNAD       | Box & Jenkins | Melhor desempenho |
|---------|------------|---------------|-------------------|
| MSE     | 0.000019   | 0.0005        | RNAD              |
| MAPE    | 82.5100%   | 150.8200%     | RNAD              |
| U Theil | 0.0174     | 703.6880      | RNAD              |
| POICD   | 96.3000%   | 49.7400%      | RNAD              |
| ARV     | 0.0002     | 0.6777        | RNAD              |
| AIC     | -2046.3600 | -1421.4898    | RNAD              |
| BIC     | -2117.7100 | -1490.3359    | RNAD              |

### 3.3.2.3.6 Série CO<sub>2</sub>

Na Figura 59 mostramos um gráfico comparativo entre a série real (linha preta) e a série prevista (linha vermelha) pela RNAD, para o grupo de teste. Note que nesta que as RNAs capturam o comportamento do fenômeno gerador da série, apresentando previsões bastante próximas dos valores reais.

Figura 59 – Comparativo entre os valores reais e as previsões -CO<sub>2</sub>.



*Fonte: elaborada pelo autor.*

Na Tabela 45 são apresentadas as medidas de desempenho dos modelos escolhidos pela metodologia Box e Jenkins e RNADs. Note que aqui o modelo escolhido pela metodologia Box e Jenkins apresenta desempenho superior à aquele apresentado pelo método das RNADs (exceto POICD). Além de capturar de forma superior o comportamento do fenômeno gerador da série, o modelo escolhido pela metodologia Box e Jenkins forneceu medidas ainda mais precisas, uma vez que aquelas obtidas pelas RNADs já foram relativamente precisas.

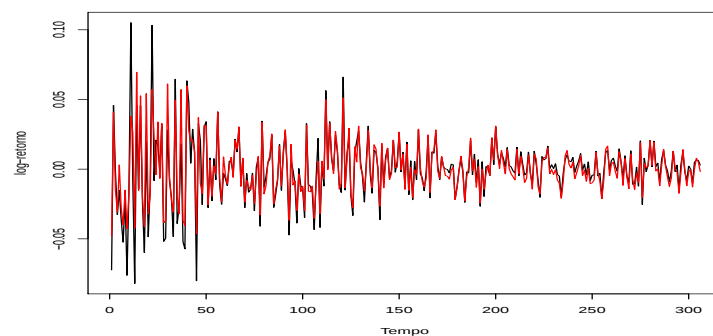
Tabela 45 – Medidas de desempenho de previsão para os modelo analisado -  $CO_2$ . Fonte: elaborada pelo autor.

| Medidas | RNAD     | Box & Jenkins | Melhor desempenho |
|---------|----------|---------------|-------------------|
| MSE     | 1.9243   | 0.8506        | Box & Jenkins     |
| MAPE    | 0.3000%  | 0.2000%       | Box & Jenkins     |
| U Theil | 0.1321   | 0.0761        | Box & Jenkins     |
| POICD   | 66.6700% | 58.3300%      | RNAD              |
| ARV     | 0.0048   | 0.0029        | Box & Jenkins     |
| AIC     | 17.8546  | 8.05578       | Box & Jenkins     |
| BIC     | 16.1273  | 2.2485        | Box & Jenkins     |

### 3.3.2.3.7 Série Dow Jones

Na Figura 60, as previsões feitas pelo modelo de RNADs acompanham extritamente os valores reais mostrando um bom ajuste do modelo.

Figura 60 – Comparativo entre os valores reais e as previsões - Dow Jones.



Fonte: elaborada pelo autor.

Na Tabela 46 apresentamos um comparativo entre os modelos selecionados pelas RNADs e pela metodologia Box e Jenkins. Assim como na maioria das séries anteriores, o método de RNAs apresenta desempenho superior ao modelo escolhido pela metodologia Box e Jenkins.

Tabela 46 – Medidas de desempenho de previsão para os modelo analisado - Dow Jones.  
*Fonte: elaborada pelo autor.*

| Medidas | RNAD       | Box & Jenkins | Melhor desempenho |
|---------|------------|---------------|-------------------|
| MSE     | 0.00001    | 0.0005        | RNAD              |
| MAPE    | 102.82%    | 105.9300%     | RNAD              |
| U Theil | 0.10158    | 138.1880      | RNAD              |
| POICD   | 92.33%     | 45.1800%      | RNAD              |
| ARV     | 0.0007918  | 1.0035        | RNAD              |
| AIC     | -2749.1021 | -2267.7226    | RNAD              |
| BIC     | -2820.5745 | -2327.7859    | RNAD              |

#### 3.3.2.4 Conclusões

Neste capítulo modelamos os log-retornos das séries descritas no capítulo 2 utilizando uma técnica de Inteligência Artificial denominada Redes Neurais Artificiais, em particular as RNAs dinâmicas, assim como modelos econométricos do tipo ARMA-GARCH. Os resultados alcançados revelam valores de MSEs muito menores para as RNADs quando comparadas aos MSEs dos modelos ARMA-GARCH para o caso das séries financeiras. Como as RNADs se ajustam incomparavelmente melhor do que os modelos econométricos, a metodologia de combinação de modelos através de funções peso calculadas de acordo com as performances de cada modelo, não se aplica para essas séries de log-retornos pois o resultado de uma combinação daria  $p_i = 1$  para a previsão feita pelo melhor modelo, onde  $p_i$  é o peso atribuído à previsão do modelo  $i$ , e  $p_i = 0$  para as previsões dos demais modelos. Desta forma, o resultado da previsão é exatamente igual para o melhor modelo individual e para o modelo combinado.

Então, para estudarmos se o método combinado pode de fato melhorar os resultados das previsões, decidimos realizar as análises das séries sem nenhum tipo de transformação nos dados, relaxando assim algumas suposições desse modelos. Ou seja, trabalhamos com as séries reais de preços apresentadas no capítulo 2. Sendo assim, foram selecionados modelos ARIMAs pela metodologia Box e Jenkins e posteriormente calculadas suas previsões para cada uma das série financeiras aqui estudadas. Em seguida, foram calculadas as previsões pelas RNADs e por fim aplicado o modelo combinado dessas duas metodologias para o cálculo das previsões. O resultados obtidos serão apresentados no capítulo 5.

## 4 Modelo Combinado-Estimação e Resultados

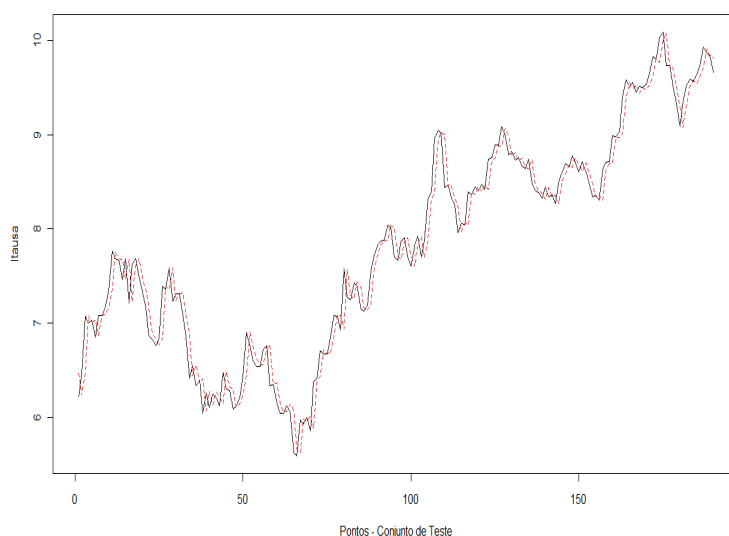
### 4.1 Experimentos com os modelos de Box e Jenkins

Para realizar as previsões com modelos estatísticos das séries de preços das diferentes ações e índices, utilizamos os modelos de Box e Jenkins (ARIMA( $p,d,q$ )). Para cada série ajustamos um modelo ARIMA adequado e, em seguida, realizamos previsões para o conjunto de teste, assim como calculamos as medidas de performance utilizadas nesta tese.

#### 4.1.1 Série do valor de fechamento das ações do Itausa

O modelo eleito nos experimentos de Box e Jenkins para a série do valor de fechamento das ações do Itausa foi o ARIMA(1,0,1), onde os coeficientes auto-regressivos e de médias móveis são:  $\phi_1 = 0.9892$  e  $\theta_1 = -0.0176$ . Um gráfico comparativo entre os valores da série (linha sólida) e os valores de previsão obtidos à partir do modelo ARIMA(1,0,1) (linha tracejada) é mostrado na Figura 6.4.

Figura 61 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box & Jenkins (linha tracejada) para o conjunto de teste da série do Valor de Fechamento das Ações do Itausa.



*Fonte: elaborada pelo autor.*

As medidas de desempenho, apresentadas na Tabela 47, mostram para o conjunto

de teste um erro MSE de aproximadamente 0.0415 e um erro MAPE de 2.01%. No que tange à medida U de Theil, tem-se o valor de 1.0543. A medida POICD foi de 42.86% para o conjunto de teste e a medida ARV apresentou um valor aproximado de 0.0002. Por fim, o resultado dos critérios de informação AIC e BIC foi de -589.57 e -614.67, respectivamente.

Tabela 47 – Medidas de desempenho de previsão para o modelo ARIMA(1,0,1) aplicado à série do Valor de Fechamento das Ações do Itausa. *Fonte: elaborada pelo autor.*

| Medidas | ARIMA(1,0,1) |
|---------|--------------|
| MSE     | 0.0415       |
| MAPE    | 2.010%       |
| U Theil | 1.0543       |
| POICD   | 42.8600%     |
| ARV     | 0.0002       |
| AIC     | -589.5742    |
| BIC     | -614.6718    |

#### 4.1.2 Série do valor de fechamento das ações da Petrobras

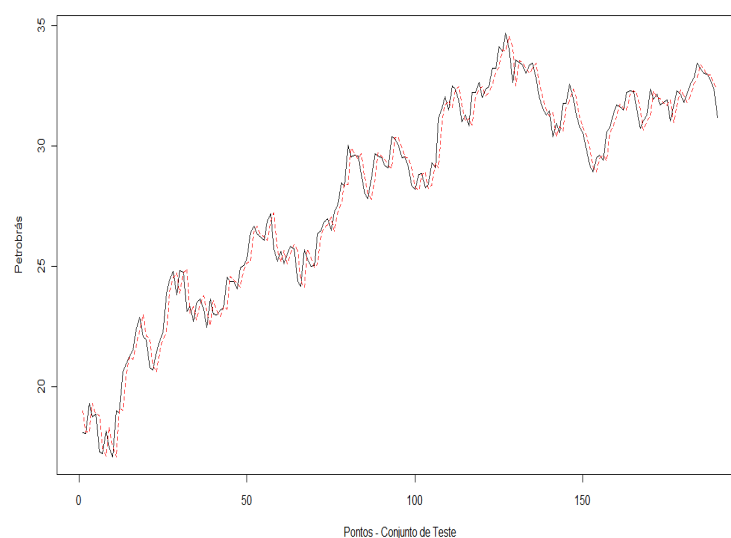
Para a série do valor de fechamento das ações da Petrobras foi escolhido o ARIMA(3,0,2), com coeficientes auto-regressivos e de médias móveis iguais a:  $\phi_1 = -0.1547$ ,  $\phi_2 = 0.8034$ ,  $\phi_3 = 0.3337$ ,  $\theta_1 = 1.1362$  e  $\theta_2 = 0.3802$ , respectivamente. A Figura 4.2 apresenta um gráfico comparativo entre os valores reais (linha sólida) e os valores previstos através do modelo em estudo (linha tracejada).

As medidas de performance de previsão para o conjunto de teste são apresentadas na Tabela 48.

Tabela 48 – Medidas de desempenho de previsão para o modelo ARIMA(3,0,2) aplicado à série do Valor de Fechamento das Ações da Petrobras. *Fonte: elaborada pelo autor.*

| Medidas | ARIMA(3,0,2) |
|---------|--------------|
| MSE     | 0.44986      |
| MAPE    | 1.92%        |
| U Theil | 1.0208       |
| POICD   | 49.21%       |
| ARV     | 0.000121     |
| AIC     | -138.97682   |
| BIC     | -149.76974   |

Figura 62 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações da Petrobras.



*Fonte: elaborada pelo autor.*

#### 4.1.3 Série do valor de fechamento das ações do Bradesco

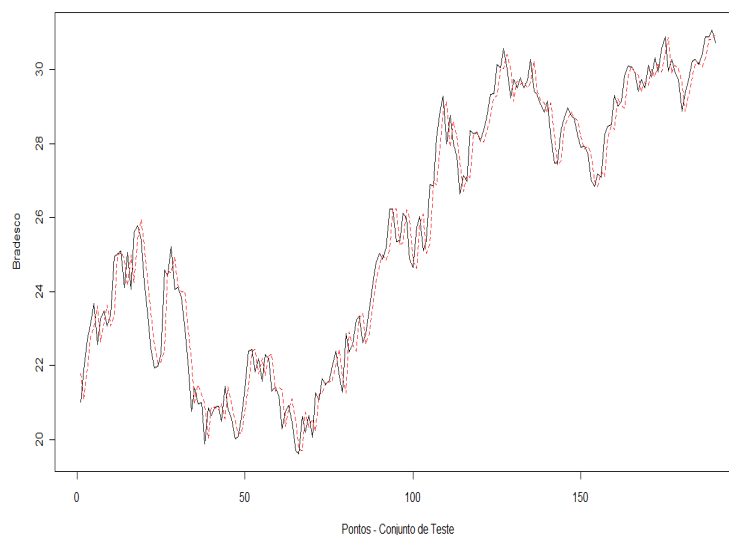
Para a série do valor de fechamento das ações do Bradesco o modelo Box e Jenkins obtido foi o ARIMA(1,0,3), em que os coeficientes auto-regressivo e médias móveis estimados são:  $\phi_1 = 0.9897$ ,  $\theta_1 = -0.0327$ ,  $\theta_2 = 0.0476$  e  $\theta_3 = -0.1206$ , respectivamente. Um gráfico comparativo entre a série real do valor de fechamento das ações do Bradesco (linha sólida) e a previsão (linha tracejada) gerada a partir do modelo ARIMA(1,0,3) é exibido na Figura 4.3.

As medidas de desempenho podem ser vistas na Tabela 49.

Tabela 49 – Medidas de desempenho de previsão para o modelo ARIMA(1,0,3) aplicado à série do Valor de Fechamento das Ações do Bradesco. *Fonte: elaborada pelo autor.*

| Medidas | ARIMA(1,0,3) |
|---------|--------------|
| MSE     | 0.40263      |
| MAPE    | 2.0400%      |
| U Theil | 1.0722       |
| POICD   | 44.4400%     |
| ARV     | 0.0002       |
| AIC     | -159.9361    |
| BIC     | -171.3944    |

Figura 63 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box & Jenkins (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Bradesco.



*Fonte: elaborada pelo autor.*

#### 4.1.4 Série do índice de preços do Ibovespa

O Modelo eleito através da análise de Box e Jenkins para a série do Índice de preços do Ibovespa foi o modelo ARIMA(2,0,0), onde os coeficientes do modelo são,  $\phi_1 = 0.9430$  e  $\phi_2 = 0.0519$ . A Figura 4.4 mostra a comparação entre os valores da série (linha sólida) e as previsões obtidas pelo modelo ajustado, ARIMA(2,0,0), (linha tracejada).

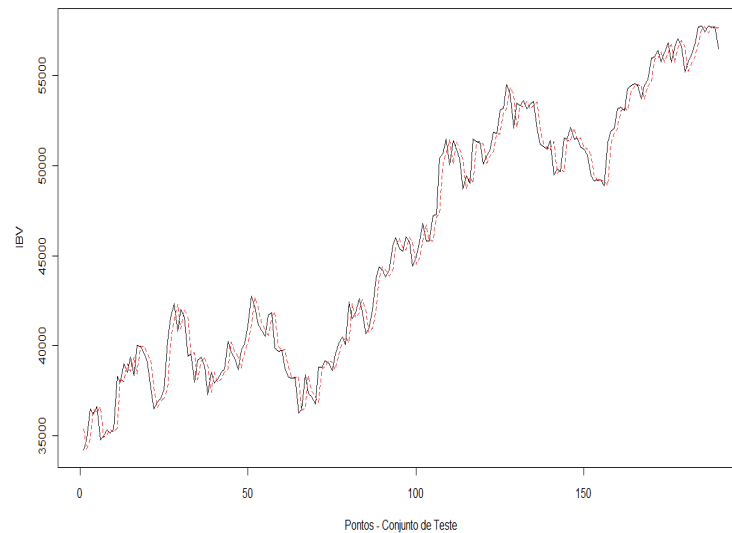
Tabela 50 – Medidas de desempenho de previsão para o modelo ARIMA(2,0,0) aplicado à série do Índice de preços do Ibovespa. *Fonte: elaborada pelo autor.*

| Medidas | ARIMA(2,0,0) |
|---------|--------------|
| MSE     | 936433.15113 |
| MAPE    | 1.65%        |
| U Theil | 1.1208       |
| POICD   | 48.15%       |
| ARV     | 0.000104     |
| AIC     | 2610.71852   |
| BIC     | 2687.21751   |

#### 4.1.5 Série do índice Dow Jones Industrial Average (DJIA)

Para a série do Índice Dow Jones Industrial Average (DJIA) o modelo de Box e Jenkins selecionado foi o ARIMA(1,0,2), em que os coeficientes auto-regressivo e médias

Figura 64 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série do Índice de preços do Ibovespa.



*Fonte: elaborada pelo autor.*

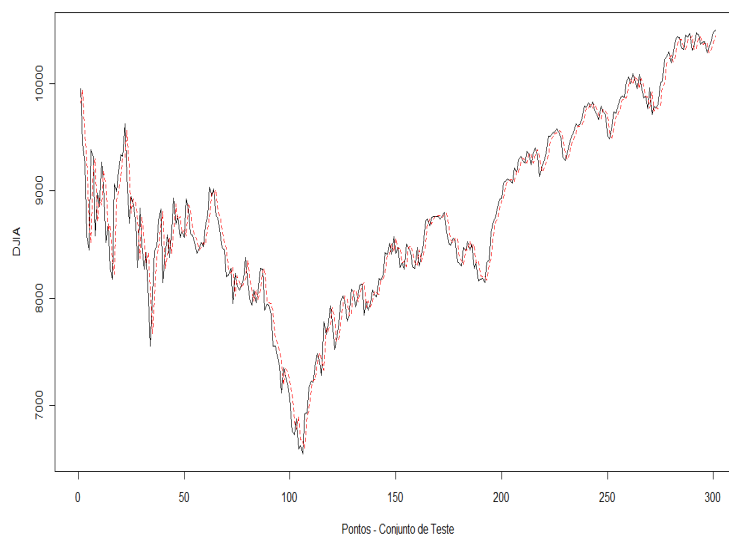
móveis estimados são:  $\phi_1 = 0.9974$ ,  $\theta_1 = -0.0614$  e  $\theta_2 = -0.0321$ . A Figura 4.5 exibe o comportamento dos valores reais da série (linha sólida) e as previsões geradas a partir do modelo ARIMA(1,0,2), (linha tracejada).

Tabela 51 – Medidas de desempenho de previsão para o modelo ARIMA(1,0,2) aplicado à série do Índice de Dow Jones Industrial Average. *Fonte: elaborada pelo autor.*

| Medidas | ARIMA(1,0,2) |
|---------|--------------|
| MSE     | 35693.79     |
| MAPE    | 1.53%        |
| U Theil | 1.51849      |
| POICD   | 47.51%       |
| ARV     | 0.00016      |
| AIC     | 3169.30232   |
| BIC     | 3235.68144   |

As medidas de desempenho da previsão para o modelo ajustado são apresentadas na Tabela 51.

Figura 65 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo modelo de Box e Jenkins (linha tracejada) para o conjunto de teste da série do Índice Dow Jones Industrial Average.



*Fonte: elaborada pelo autor.*

## 4.2 Resultados experimentais com a aplicação de Redes Neurais Artificiais Dinâmicas

### 4.2.1 Série do valor de fechamento das ações do Itausa

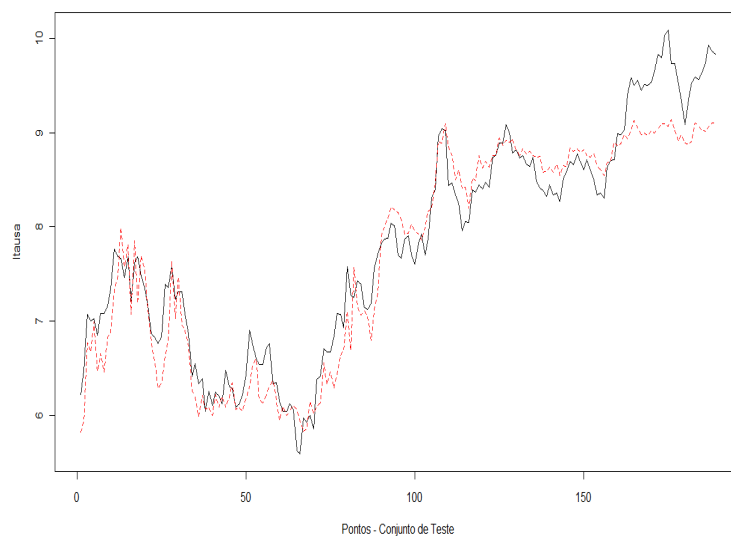
Apresentamos nesta subseção os resultados das previsões para a série do valor de fechamento das ações do Itausa realizadas através dos modelos de Redes Neurais Artificiais Dinâmicas (RNADs). A Figura 4.6 mostra o comportamento dos valores reais (linha cheia) e os previstos pelas RNADs (linha tracejada). As medidas de desempenho são apresentadas na Tabela 52.

As medidas de desempenho, apresentadas na Tabela 52, mostram para o conjunto de teste um erro MSE de aproximadamente 0.04396 e um erro MAPE de 2.13%. No que tange à medida U de Theil, tem-se o valor de 1.6833. A medida POCID foi de 58.20% para o conjunto de teste e a medida ARV apresentou um valor aproximado de 0.000187. Por fim, o resultado dos critérios de informação AIC e BIC foi de -578.51534.57 e -603.26186, respectivamente.

### 4.2.2 Série do valor de fechamento das ações da Petrobras

Através das RNADs a previsão para a série do valor de fechamento das ações da Petrobras pode ser vista na Figura 4.7 onde temos um gráfico comparativo entre os

Figura 66 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Itaúsa.



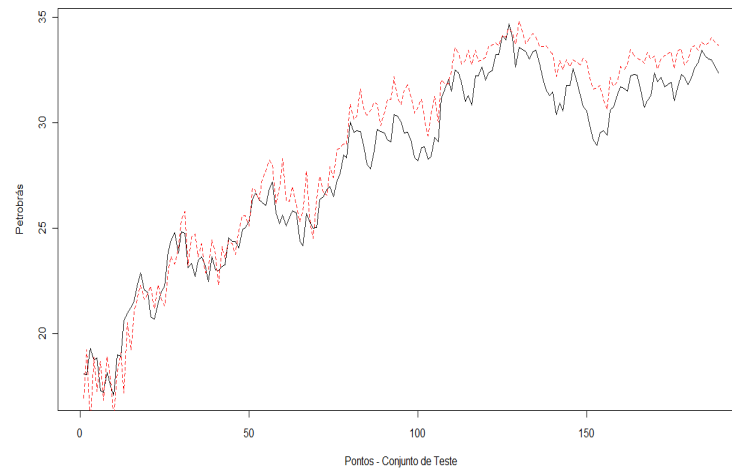
*Fonte: elaborada pelo autor.*

Tabela 52 – Medidas de desempenho de previsão para a rede neural ajustada. *Fonte: elaborada pelo autor.*

| Medidas | Valores    |
|---------|------------|
| MSE     | 0.04396    |
| MAPE    | 2.13%      |
| U Theil | 1.6833     |
| POICD   | 58.20%     |
| ARV     | 0.000187   |
| AIC     | -578.51534 |
| BIC     | -603.26186 |

valores reais (linha sólida) e os valores previstos (linha tracejada).

Figura 67 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão geradas pelo metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações da Petrobras.



*Fonte: elaborada pelo autor.*

As medidas de performance de previsão podem ser visualizadas na Tabela 53

Tabela 53 – Medidas de desempenho de previsão para redes neurais. *Fonte: elaborada pelo autor.*

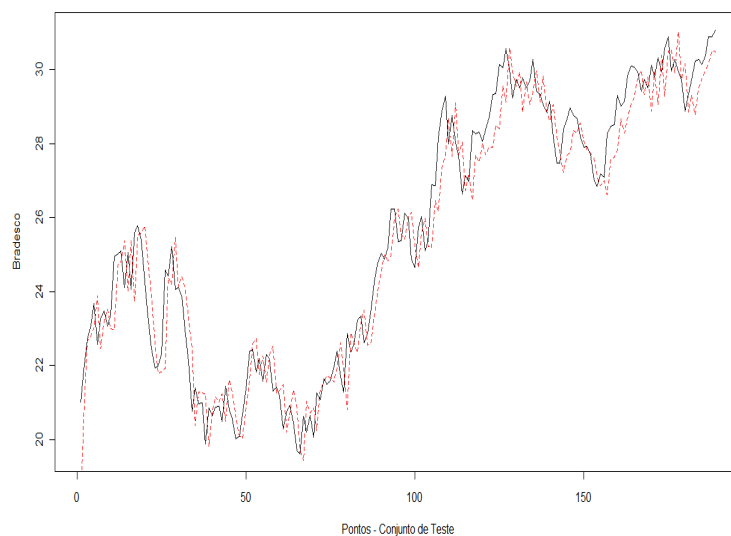
| Medidas | Valores    |
|---------|------------|
| MSE     | 0.45995    |
| MAPE    | 2.07%      |
| U Theil | 1.7063     |
| POICD   | 71.96%     |
| ARV     | 0.000113   |
| AIC     | -134.78369 |
| BIC     | -145.44349 |

#### 4.2.3 Série do valor de fechamento das ações do Bradesco

A figura 4.8 mostra um gráfico comparativo entre a série real do valor de fechamento das ações do Bradesco (linha sólida) e a previsão obtida a partir de Redes Neurais Artificiais (linha tracejada).

Podemos observar o resultados das diferentes medidas de performance na Tabela 54.

Figura 68 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão geradas pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série o Valor de Fechamento das Ações do Bradesco.



*Fonte: elaborada pelo autor.*

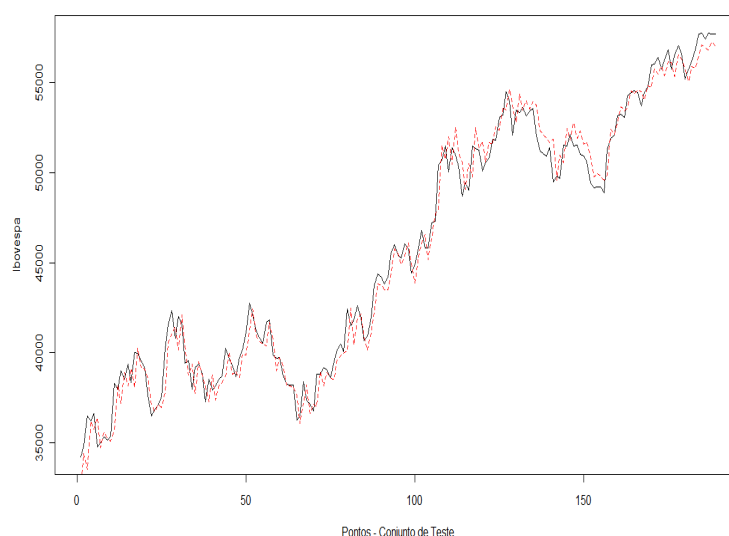
Tabela 54 – Medidas de desempenho de previsão obtidas através de redes neurais. *Fonte: elaborada pelo autor.*

| Medidas | Valores    |
|---------|------------|
| MSE     | 0.214471   |
| MAPE    | 1.49%      |
| U Theil | 0.2841     |
| POICD   | 77.35%     |
| ARV     | 0.000098   |
| AIC     | -278.98059 |
| BIC     | -294.21807 |

#### 4.2.4 Série do índice de preços do Ibovespa

A Figura 4.9 mostra a comparação entre os valores reais (linha sólida) e as previsões obtidas através das Redes Neurais Artificiais (linha tracejada).

Figura 69 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pelo metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série do Índice de preços do Ibovespa.



*Fonte: elaborada pelo autor.*

As medidas de desempenho da previsão para o modelo ajustado são apresentadas na Tabela 55.

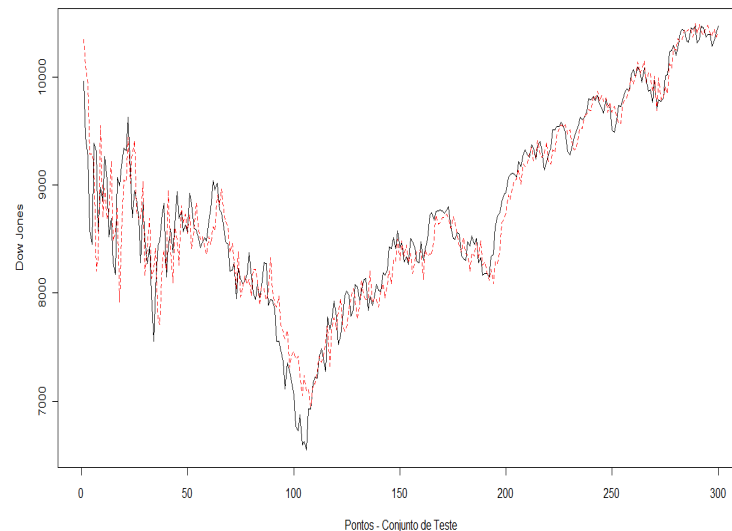
Tabela 55 – Medidas de desempenho de previsão para redes neurais. *Fonte: elaborada pelo autor.*

| Medidas | Valores      |
|---------|--------------|
| MSE     | 939062.03175 |
| MAPE    | 1.73%        |
| U Theil | 0.7917       |
| POICD   | 50.79%       |
| ARV     | 0.000097     |
| AIC     | 2611.248359  |
| BIC     | 2687.76179   |

#### 4.2.5 Série do índice Dow Jones Industrial Average (DJIA)

Na Figura 4.10 podemos observar o comportamento das previsões feitas pelas RNADs para a série do Índice Dow Jones Industrial Average (DJIA), assim como o comportamento da série real.

Figura 70 – Gráfico comparativo entre os valores reais (linha sólida) e os valores de previsão gerados pela metodologia de Redes Neurais Artificiais (linha tracejada) para o conjunto de teste da série do Índice Dow Jones Industrial Average.



*Fonte: elaborada pelo autor.*

As medidas de performance da previsão para o modelo de RNADs são apresentadas na Tabela 56.

Tabela 56 – Medidas de desempenho de previsão obtidas aplicando Redes Neurais à série dos Índices Dow Jones Industrial Average. *Fonte: elaborada pelo autor.*

| Medidas | Valores     |
|---------|-------------|
| MSE     | 25878.4408  |
| MAPE    | 1.33%       |
| U Theil | 1.02753     |
| POICD   | 80.40%      |
| ARV     | 0.000012    |
| AIC     | 3072.510815 |
| BIC     | 3136.638974 |

### 4.3 Modelo Combinado

Na prática, é muito difícil encontrar um modelo que tenha alta performance em todos os períodos. Para incrementar a qualidade das previsões individuais foram criados os modelos combinados. Utilizaremos o método de combinação utilizado em Zou e Yang

(2004).<sup>1</sup> Esta metodologia de combinação é representada pelas equações (4.1) e (4.2).

$$\hat{y}(n+1) = \sum_k p^k(n+1) \hat{y}_k(n+1) \quad (4.1)$$

em que  $\hat{y}_k$  é o valor estimado de  $y$  pelo modelo  $k$  e o peso  $p^k(n+1)$  é calculado por meio de

$$p^k(n+1) = \frac{\exp[-\lambda(n) \sum_{j=1}^n (y(j) - \hat{y}_k(j))^2]}{\sum_{j=1}^n \exp[-\lambda(n) \sum_{j=1}^n (y(j) - \hat{y}_k(j))^2]} \quad (4.2)$$

em que  $\lambda(n) = 1/(2S^2)$  e  $S^2$  é a variância amostral. Intuitivamente, numa janela  $n$ , o modelo que tiver performance melhor terá pesos maiores, já o modelo que tiver performance fraca terá pesos menores.

Estas classes de modelos serão ajustadas para implementar a previsão das séries temporais a serem estudadas. Especificamente serão implementados os modelos formados pela combinação de Redes Neurais Artificiais Dinâmicas e modelos ARIMA. Apresentaremos os gráficos comparativos dos resultados das previsões com os valores reais para cada série assim como calcularemos as medidas de performance para podermos comparar os resultados obtidos com as diferentes técnicas de previsão. Para o cálculo do modelo Combinado foi desenvolvido um programa em linguagem C.

### 4.3.1 Resultados experimentais com a aplicação do Modelo Combinado

#### 4.3.1.1 Série de Manchas Solares

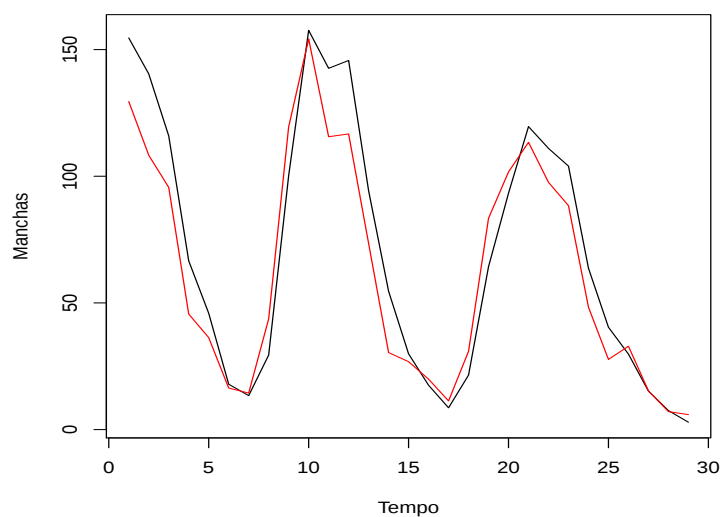
A Figura 71, apresenta um comparativo entre os valores reais da série e os valores previstos para o modelo combinado gerado pela combinação da respectiva RNAD com o ARIMA (9,0,0) para o caso da previsão da série de Manchas Solares. Analisando a figura, observamos que os resultados obtidos simulam o comportamento da série real para os pontos do conjunto de teste. A Tabela 57 apresenta as medidas de performance para as previsões do modelo combinado. O resultado para o conjunto teste do MSE foi de aproximadamente 248.610 e um erro MAPE de 22.27%. No que tange à medida U de Theil, tem-se o valor de 0.310. A medida POCID foi de 93.100% e a medida ARV apresentou um valor aproximado de 0.004. Por fim, o resultado dos critérios de informação AIC e BIC foi de 171.960 e 199.055, respectivamente.

#### 4.3.1.2 Série de CO<sub>2</sub>

Podemos observar na Figura 72 os valores previstos comparados aos valores reais para os pontos pertencentes ao grupo de teste da série de emissão de CO<sub>2</sub>. As previsões

<sup>1</sup> Zou e Yang (2004) utilizaram como base de dados para a aplicação do modelo combinado as seguintes séries: média diária de defeitos por caminhão no final da linha de montagem, nível de água do lago Huron, produção mensal de tijolos na Austrália e vendas mensais de casas novas nos Estados Unidos.

Figura 71 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Manchas Solares.



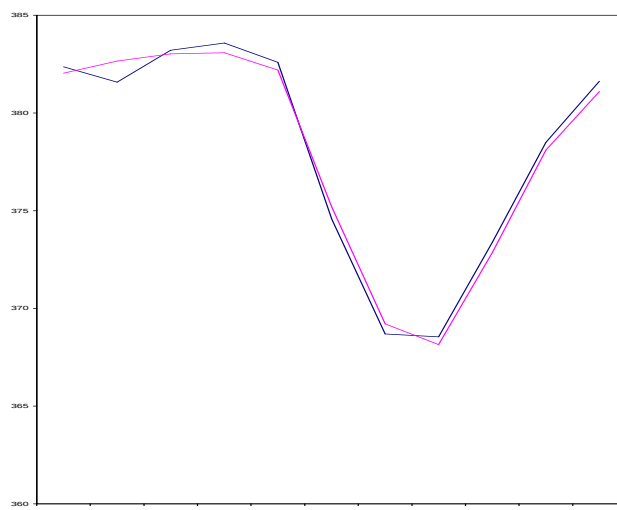
*Fonte: elaborada pelo autor.*

Tabela 57 – Medidas de desempenho de previsão obtidas aplicando o método combinado à série de Manchas Solares. *Fonte: elaborada pelo autor.*

| Medidas | Valores |
|---------|---------|
| MSE     | 248.610 |
| MAPE    | 22.27%  |
| U Theil | 0.310   |
| POICD   | 93.100% |
| ARV     | 0.004   |
| AIC     | 171.960 |
| BIC     | 199.055 |

foram realizadas com a combinação de RNADs e o modelo ARIMA (0,1,1)×(0,1,1)[12]. Na Tabela 58 calculamos as medidas a serem utilizadas para verificar as qualidades das previsões apresentadas pelo modelo combinado.

Figura 72 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de CO<sub>2</sub>.



*Fonte: elaborada pelo autor.*

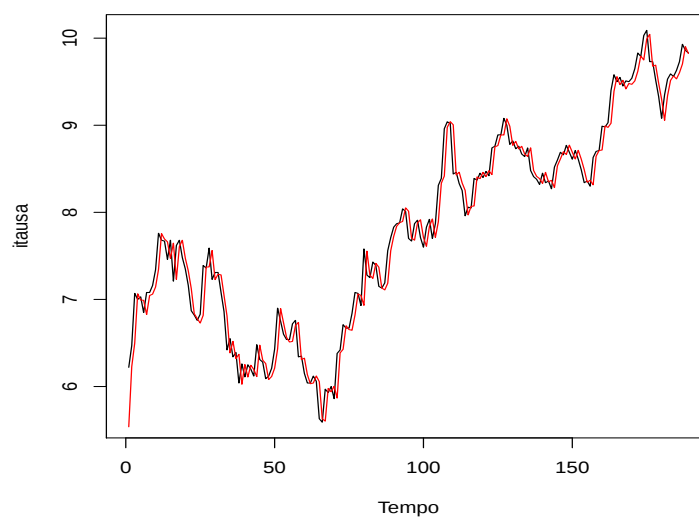
Tabela 58 – Medidas de desempenho de previsão obtidas aplicando o método combinado à série de CO<sub>2</sub>. *Fonte: elaborada pelo autor.*

| Medidas | Valores  |
|---------|----------|
| MSE     | 0.292817 |
| MAPE    | 0.130%   |
| U Theil | 0.0221   |
| POICD   | 81.82%   |
| ARV     | 0.000901 |
| AIC     | 3.51     |
| BIC     | 14.6153  |

#### 4.3.1.3 Série do valor de fechamento das ações do Itaúsa

O gráfico apresentado na Figura 73 apresenta as previsões obtidas através do modelo combinado e os valores reais para o grupo de teste da série Itaúsa. O modelo combinado foi gerado a partir de combinações de RNADs com o modelo ARIMA (1,0,1). Já na Tabela 59 apresentamos algumas medidas utilizadas para quantificar a qualidade das previsões obtidas através do modelo combinado.

Figura 73 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índice do Itausa.



*Fonte: elaborada pelo autor.*

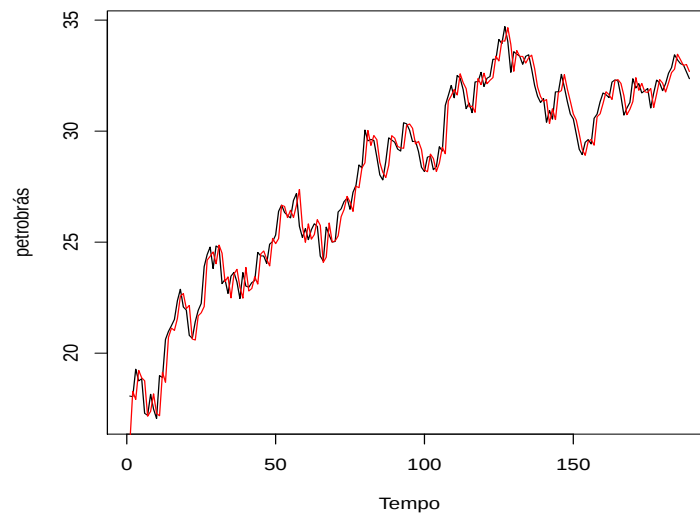
Tabela 59 – Medidas de desempenho de previsão para o método combinado a série do valor de fechamento das ações do Itausa. *Fonte: elaborada pelo autor.*

| Medidas | Valores      |
|---------|--------------|
| MSE     | 0.037699116  |
| MAPE    | 1.94%        |
| U Theil | 1.32926345   |
| POICD   | 51.85%       |
| ARV     | 0.000154879  |
| AIC     | -607.5644225 |
| BIC     | -633.2331343 |

## 4.3.1.4 Série do valor de fechamento das ações da Petrobras

O gráfico apresentado na Figura 74 apresenta as previsões obtidas através do modelo combinado e os valores reais para o grupo de teste da série de preços das ações da Petrobras. O modelo combinado foi gerado a partir de combinações de RNADs com o modelo ARIMA (3,0,2). A Tabela 60 apresenta algumas medidas utilizadas para quantificar a qualidade das previsões obtidas através do modelo combinado.

Figura 74 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índice da Petrobras.



*Fonte: elaborada pelo autor.*

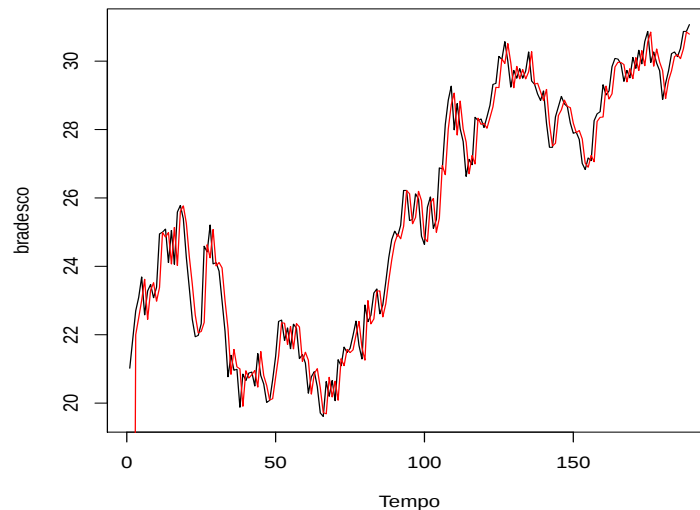
Tabela 60 – Medidas de desempenho de previsão pelo método comninado para a série do valor de fechamento das ações da Petrobras. *Fonte: elaborada pelo autor.*

| Medidas | Valores   |
|---------|-----------|
| MSE     | 0.371917  |
| MAPE    | 1.82%     |
| U Theil | 1.0556    |
| POICD   | 51.32%    |
| ARV     | 0.000095  |
| AIC     | -174.9371 |
| BIC     | -186.872  |

## 4.3.1.5 Série do valor de fechamento das ações do Bradesco

Na Figura 75 temos o gráfico comparativo entre os valores reais e os valores previstos para o grupo de teste da série das ações do Bradesco. O modelo combinado foi gerado a partir de combinações de RNADs com o modelo ARIMA (1,0,3). Foram também calculadas algumas medidas para verificarmos a qualidade das previsões. Tais medidas, estão apresentadas na Tabela 61.

Figura 75 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices do Bradesco.



*Fonte: elaborada pelo autor.*

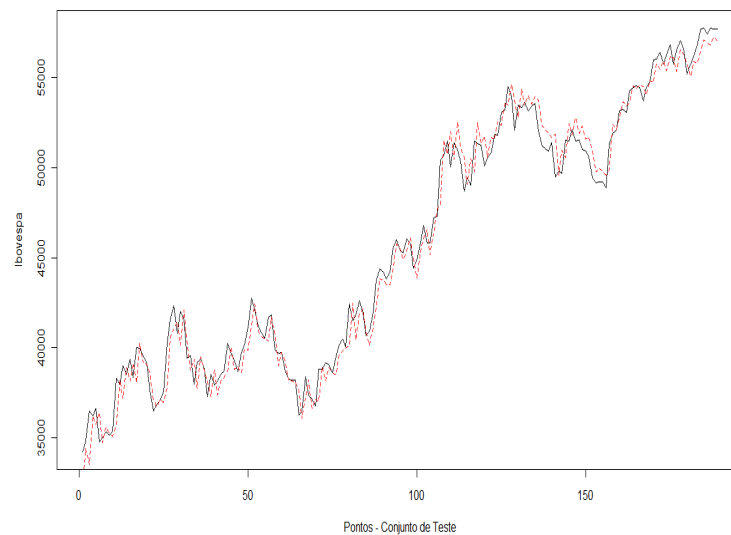
Tabela 61 – Medidas de desempenho de previsão obtidas pelo método combinado para a série do valor de fechamento das ações do Bradesco. *Fonte: elaborada pelo autor.*

| Medidas | Valores   |
|---------|-----------|
| MSE     | 0.20778   |
| MAPE    | 1.45%     |
| U Theil | 0.89070   |
| POICD   | 64.55%    |
| ARV     | 0.00001   |
| AIC     | -284.9661 |
| BIC     | -300.3936 |

## 4.3.1.6 Série do índice de preços do Ibovespa

Na Figura 76, temos o gráfico que apresenta um comparativo entre os valores previstos obtidos modelo combinado e o valores reais do grupo de teste. O modelo combinado foi gerado a partir de combinações de RNADs com o modelo ARIMA (2,0,0). Temos ainda a Tabela 62 que apresenta as medidas de performance para quantificar a qualidade das previsões.

Figura 76 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices do Ibovespa.



*Fonte: elaborada pelo autor.*

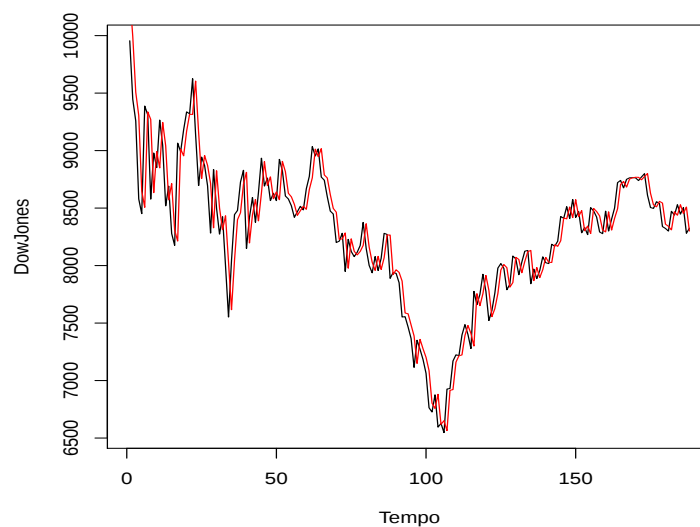
Tabela 62 – Medidas de desempenho de previsão obtidos pelo método combinado para a série do índice de preços do Ibovespa. *Fonte: elaborada pelo autor.*

| Medidas | Valores      |
|---------|--------------|
| MSE     | 908432.55978 |
| MAPE    | 1.63%        |
| U Theil | 0.9707       |
| POICD   | 51.85%       |
| ARV     | 0.000098     |
| AIC     | 2604.980951  |
| BIC     | 2681.297807  |

#### 4.3.1.7 Série do índice Dow Jones Industrial Average (DJIA)

Na Figura 77 temos um gráfico comparativo entre os valores reais e os valores previstos obtidos através do modelo combinado. O modelo combinado foi gerado a partir de combinações de RNADs com o modelo ARIMA (1,0,2). Na Tabela 63 apresentamos algumas medidas de desempenho para as previsões deste modelo.

Figura 77 – Gráfico comparativo entre os valores reais (linha preta) e os valores de previsão gerados pelo método combinado (linha vermelha) para o conjunto de teste da série de Índices da Dow Jones.



*Fonte: elaborada pelo autor.*

Tabela 63 – Medidas de desempenho de previsão obtidas aplicando o método combinado à série dos Índices Dow Jones Industrial Average. *Fonte: elaborada pelo autor.*

| Medidas | Valores     |
|---------|-------------|
| MSE     | 21324.32957 |
| MAPE    | 1.08%       |
| U Theil | 1.632400167 |
| POICD   | 51.00%      |
| ARV     | 0.00019     |
| AIC     | 3002.28118  |
| BIC     | 3056.086804 |

## 4.4 Comparação entre modelos

Na tabelas a seguir, apresentamos um comparativo entre as medidas de performance dos modelos de redes neurais dinâmicas, modelos ARIMAs e modelos combinados. Observando a tabela 64, conclui-se que para o caso da série de manchas solares, o modelo combinado apresenta desempenho apenas levemente superior aos modelos de RNADs. Para este caso em particular, a utilização do modelo não apresenta maiores vantagens. Observamos que as medidas de AIC e BIC são praticamente idênticas para o modelo combinado e as RNAs.

Tabela 64 – Medidas de desempenho de previsão para os modelos analisados - Manchas Solares. *Fonte: elaborada pelo autor.*

| Medidas | RNAD     | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|----------|---------------|--------------------|-------------------|
| MSE     | 248.6455 | 380.3000      | 248.6100           | Comb              |
| MAPE    | 22.27%   | 59.85%        | 22.27%             | Comb-RNAD         |
| U Theil | 0.312    | 0.7415        | 0.310              | Comb              |
| POICD   | 93.10%   | 79.31%        | 93.100%            | Comb-RNAD         |
| ARV     | 0.004    | 0.0070        | 0.004              | Comb-RNA          |
| AIC     | 171.9600 | 192.2879      | 171.960            | Comb-RNAD         |
| BIC     | 199.0600 | 214.6975      | 199.050            | Comb              |

Na Tabela 65 apresentamos os resultados das medidas de performance para a série de CO<sub>2</sub>. Observa-se que o modelo combinado apresenta desempenho bem superior aos modelos escolhidos pela metodologia de Box e Jenkins e aos escolhidos pelas RNADs. Analisando o MSE concluímos que o modelo combinado reduz o erro em mais de 60% quando comparado ao MSE do modelo individual com melhor performance. Com exceção do BIC, todas as outras medidas favoreceram a modelagem combinada.

Tabela 65 – Medidas de desempenho de previsão para os modelos analisados - CO<sub>2</sub>. *Fonte: elaborada pelo autor.*

| Medidas | RNAD     | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|----------|---------------|--------------------|-------------------|
| MSE     | 1.9243   | 0.8506        | 0.292817           | Comb              |
| MAPE    | 0.3000%  | 0.2000%       | 0.130%             | Comb              |
| U Theil | 0.1321   | 0.0761        | 0.0221             | Comb              |
| POICD   | 66.6700% | 58.3300%      | 81.82%             | Comb              |
| ARV     | 0.0048   | 0.0029        | 0.000901           | Comb              |
| AIC     | 17.8546  | 8.05578       | 3.51               | Comb              |
| BIC     | 16.1273  | 2.2485        | 14.6153            | Box & Jenkins     |

Na Tabela 66 sintetizamos os resultados para as medidas de performance da série do Itausa. Novamente o modelo combinado apresenta as melhores medidas de performance. Com exceção da medida POICD que sugere melhores resultados com RNADs,

todas as outras medidas apontam para uma melhor capacidade de previsão do modelo combinado.

Tabela 66 – Medidas de desempenho de previsão para os modelos analisados - Itausa.  
Fonte: elaborada pelo autor.

| Medidas | RNAD      | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|-----------|---------------|--------------------|-------------------|
| MSE     | 0.04396   | 0.041464      | 0.037699116        | Comb              |
| MAPE    | 2.13%     | 2.01%         | 1.94%              | Comb              |
| U Theil | 1.833     | 1.0543        | 1.32926345         | Comb              |
| POICD   | 58.20%    | 42.86%        | 51.85%             | RNAD              |
| ARV     | 0.000187  | 0.000171      | 0.000154           | Comb              |
| AIC     | -578.5153 | -589.5742     | -607.5644          | Comb              |
| BIC     | -603.2619 | -614.6718     | -633.2331          | Comb              |

A Tabela 67 apresenta os resultados das medidas de desempenho para a série do Bradesco. Comparando as medidas de AIC e BIC, observamos que o modelo combinado apresenta os menores valores dessas medidas sugerindo uma maior capacidade de se realizar previsões com esses modelos.

Tabela 67 – Medidas de desempenho de previsão para os modelos analisados - Bradesco.  
Fonte: elaborada pelo autor.

| Medidas | RNAD       | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|------------|---------------|--------------------|-------------------|
| MSE     | 0.214471   | 0.40263       | 0.207785           | Comb              |
| MAPE    | 1.49%      | 2.04%         | 1.45%              | Comb              |
| U Theil | 0.2841     | 1.0722        | 0.8907             | RNAD              |
| POICD   | 77.35%     | 44.44%        | 64.55%             | RNAD              |
| ARV     | 0.000001   | 0.000181      | 0.000001           | RNAD - Comb       |
| AIC     | -278.98059 | -159.93615    | -284.96614         | Comb              |
| BIC     | -294.21807 | -171.39445    | -300.39364         | Comb              |

Os resultados das medidas de performance para a série da Petrobras pode ser visto na tabela 68. Comparando os erros médios quadráticos (MSE) dos diferentes modelos, observamos uma redução de 17,32% no MSE quando utilizamos a combinação das modelagens. Pela tabela, observa-se que o modelo combinado apresenta superioridade quando comparado aos modelos individuais. Nota-se ainda que apenas a medida POICD não sugere o modelo combinado como melhor método de previsão para a referida série.

Na Tabela 69 apresentamos um comparativo entre os modelos selecionados pelas RNAs, pela metodologia Box e Jenkins e pelo modelo combinado para a série do Ibovespa. Nessa tabela, observa-se que o modelo combinado apresenta desempenho

Tabela 68 – Medidas de desempenho de previsão para os modelos analisados - Petrobras.  
*Fonte: elaborada pelo autor.*

| Medidas | RNAD       | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|------------|---------------|--------------------|-------------------|
| MSE     | 0.45995    | 0.44986       | 0.371917           | Comb              |
| MAPE    | 2.07%      | 1.92%         | 1.82%              | Comb              |
| U Theil | 1.7063     | 1.0208        | 1.0556             | Comb              |
| POICD   | 71.96%     | 49.21%        | 51.32%             | RNAD              |
| ARV     | 0.0001     | 0.000121      | 0.000095           | Comb              |
| AIC     | -134.78364 | -138.97682    | -174.9371          | Comb              |
| BIC     | -145.44349 | -149.76974    | -186.872           | Comb              |

superior aos demais modelos em praticamente todas as medidas de qualidade de previsão.

Tabela 69 – Medidas de desempenho de previsão para os modelos analisados - Ibovespa.  
*Fonte: elaborada pelo autor.*

| Medidas | RNAD         | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|--------------|---------------|--------------------|-------------------|
| MSE     | 939062.03175 | 936433.15113  | 908432.559786      | Comb              |
| MAPE    | 1.73%        | 1.65%         | 1.63%              | Comb              |
| U Theil | 0.7917       | 1.1208        | 0.9707             | Comb              |
| POICD   | 50.79%       | 48.15%        | 51.85%             | Comb              |
| ARV     | 0.000097     | 0.000104      | 0.000098           | Comb-RNAD         |
| AIC     | 2611.248359  | 2610.71852    | 2604.980951        | Comb              |
| BIC     | 2687.76179   | 2687.21751    | 2681.297807        | Comb              |

Por fim, a tabela 70 apresenta a comparação das medidas de desempenho para a série do Dow Jones. O MSE do modelo combinado apresenta uma redução de aproximadamente 20% quando comparado aos demais MSEs. A medida MAPE também sugere um melhor ajuste da metodologia de modelos combinados uma vez que O MAPE da combinação é de 1,08% enquanto que para as RNADs essa medida é de 1,33% e para o ARIMA 1,53%.

## 4.5 Conclusões

Neste capítulo realizamos uma aplicação do modelo combinado, proposto por Zou e Yang (2004), para o caso de séries financeiras além de duas outras séries não-financeiras porém de difícil previsão. Foi possível demonstrar que, de um modo geral, a modelagem composta apresentou superioridade quando comparada aos modelos individuais utilizados na combinação. Com isto, podemos aumentar a capacidade de previsão de diferentes modelos fazendo suas combinações com modelos que venham a

Tabela 70 – Medidas de desempenho de previsão para os modelos analisados - Dow Jones. Fonte: elaborada pelo autor.

| Medidas | RNAD        | Box & Jenkins | Modelos Combinados | Melhor desempenho |
|---------|-------------|---------------|--------------------|-------------------|
| MSE     | 25878.4408  | 35693.79      | 21324.32957        | Comb              |
| MAPE    | 1.33%       | 1.53%         | 1.08%              | Comb              |
| U Theil | 1.02753     | 1.51849       | 1.632400167        | RNAD              |
| POICD   | 80.40%      | 47.51%        | 51.00%             | RNAD              |
| ARV     | 0.000012    | 0.00016       | 0.00019            | RNAD              |
| AIC     | 3072.510815 | 3169.30232    | 3002.28118         | Comb              |
| BIC     | 3136.638974 | 3235.68144    | 3056.086804        | Comb              |

compensar as deficiências apresentadas pelos primeiros. Em particular, para as séries da Petrobras, Dow Jones, Ibovespa, Itausa, Bradesco e CO<sub>2</sub>, as medidas de performance indicaram claramente a preferência pelos modelos combinados para a realização de previsões. Por exemplo, o modelo combinado aplicado à série da Petrobras reduz o MSE em 17,32% quando comparado ao melhor modelo individual. Já para a série do Dow Jones essa redução é de 17,60%. Para a série do Itausa a redução do MSE é de 9,09% e, no caso da série de CO<sub>2</sub>, a redução chega a 65,57%. Usou-se o MSE como exemplo pois o MSE é a medida mais utilizada pela sua facilidade de cálculo e por ser uma métrica encontrada em praticamente todos os pacotes computacionais que trabalham com estatística e RNAs, facilitando assim a comparação dos resultados. Como estamos interessados em realizar previsões de séries temporais nas quais qualquer incremento da capacidade de previsão pode trazer enormes economias, as reduções dos erros alcançadas com o uso do modelo combinado no presente capítulo, mais do que justificam a sua utilização.

## 5 Conclusões

Neste capítulo apresentamos as conclusões dos estudos e investigações experimentais realizadas para a resolução do problema de previsão de séries temporais. Discussões sobre a adequação da metodologia de Zou e Yang - *Combining time series models for forecasting*, 2004 - são apresentadas, considerando a sua viabilidade para aplicações em séries financeiras. Por fim, também são apresentados possíveis desdobramentos para trabalhos futuros que poderão contribuir para a evolução da metodologia de modelos combinados, bem como para um melhor entendimento do problema de previsão de séries temporais.

Nesta tese foi abordado o problema de previsões de séries temporais com o estudo de técnicas de estatística e de inteligência artificial. Os modelos estatísticos comentados, lineares e não-lineares, são os modelos mais populares encontrados na literatura especializada. Desta forma, esses modelos apresentam-se como um ponto de partida e de formalização dos estudos das leis que governam os fenômenos geradores das séries temporais.

Dentre os modelos de Inteligência Artificial (IA), foram analisados aqueles encontrados de forma mais comum nas aplicações de previsão de séries temporais relatadas na literatura: as Redes Neurais Artificiais (RNA).

Entre as diferentes técnicas de modelagem neural, optamos por trabalhar com as redes neurais artificiais dinâmicas. As características dinâmicas deste tipo de rede estão nos elementos de entrada da RNA. A malha de atraso da entrada faz com que todos os pontos da série passe em cada um dos neurônios de entrada da rede criando um efeito de memória temporal. Esse tipo de RNA apresenta vantagens comparativas com os outros tipos de rede por ser apropriada para processos de aprendizagem nos quais a variável "tempo" se constitui em um importante ingrediente para a "memorização".

Em um primeiro instante, várias séries temporais de relevância na literatura e em problemas do mundo real foram agrupadas para a consolidação de um conjunto de dados (*benchmark*) para a análise de desempenho dos modelos em estudo, formado por séries reais de fenômenos naturais (série do CO<sub>2</sub> e série das manchas solares) e séries econômico-financeiras (série do Ibovespa, série do Dow Jones, série das ações do Itaú, Bradesco e Petrobras).

Para o estabelecimento de um nível de referência para o desempenho de previsão do conjunto de séries temporais, foram realizados experimentos com os modelos econômétricos, visto que estes são totalmente entendidos e bastante utilizados na prática.

Com a referência estabelecida, foram montados vários experimentos que utilizaram unicamente as RNAs do tipo MLP para a previsão das séries temporais. Os ajustes de configurações aplicados às RNAs foram escolhidos por meio de tentativa e erro

através de vários experimentos preliminares com o objetivo de otimizar os modelos. Os resultados alcançados pelas RNADs, de forma geral, superam expressivamente os resultados obtidos pelos modelos econométricos tradicionais para as séries de retornos dos preços das ações. Esses retornos são frequentemente utilizados em estudos de finanças como uma medida do risco já que representam as variações de preços dos ativos.

Diversas classes de modelos econométricos podem ser utilizadas para modelar os retornos das séries financeiras. Entre esses podemos citar os modelos ARIMA, ARCH e GARCH utilizados na presente tese para a modelagem.

Analizando os experimentos, as RNADs apresentaram um desempenho bastante superior aos modelos econométricos. Os MSEs dos modelos de redes neurais chegam a ser até 20 vezes menores que os dos modelos econométricos. Desta forma, não faz sentido estimar um modelo combinado para realizar previsões dos retornos das séries, uma vez que o resultado da previsão seria exatamente igual ao resultado das RNADs. Essa conclusão é um dos resultados da presente tese. Além da grande diferença entre as capacidades de previsão dos modelos individuais, temos que as somas dos quadrados dos resíduos dos modelos individuais ajustados não são da mesma ordem de grandeza de  $2S^2$  fazendo com que os fatores

$$\exp\left[-\left(\frac{1}{2S^2} \sum_{j=1}^n (y(j) - \widehat{y_k(j)})^2\right)\right] \quad (5.1)$$

amplifiquem ainda mais essas diferenças e conduzindo a pesos com valores 0 e 1 na fórmula da equação 4.1. Assim, o resultado da previsão com os modelos combinados é exatamente igual ao obtido com o modelo de melhor performance.

Por outro lado, se ao invés de modelar os riscos resolvemos modelar as séries reais de preços, a utilização de modelos combinados se faz necessário para melhor a capacidade de se realizar previsões. Utilizamos para a modelagem dessas séries novamente modelos de RNADs porém com modelos de Box & Jenkins (ARMA). Para essas séries, ao compararmos a soma dos quadrados dos resíduos para os modelos até a  $n$ -ésima observação com duas vezes a variância amostral da série real até essa observação, obtemos para a expressão entre colchetes na equação 5.1 um valor próximo a 1. Assim, os pesos da equação 4.1 terão valores diferentes de 0 e 1.

Desta forma, foi sugerido a utilização de um modelo combinado formado pela combinação convexa dos modelos de RNAs e modelos econométricos com o intuito de melhorar a capacidade de previsão uma vez que esta combinação faria com que as qualidades de um modelo compensassem as desvantagens do outro e vice-versa.

São sete parâmetros de comparação entre os modelos em sete séries de teste. Assim temos 49 resultados dos parâmetros de comparação. Em 35 resultados o modelo combinado é o melhor dos três, ou seja, para as séries em teste, em 71,4% das vezes o modelo

combinado é o melhor. Em 42 resultados o modelo combinado aparece como o melhor resultado, ou seja, para as séries em teste, em 85,7% das vezes o modelo combinado é melhor ou tão bom quanto o das RNAs. Enquanto isso, o modelo RNA aparece apenas sete vezes (14,3%) como o melhor resultado e treze (26,5%) das vezes como melhor ou tão bom quanto o modelo Combinado. O Modelo de Box & Jenkins só apresenta 1 resultado como sendo o melhor na série de CO<sub>2</sub>, na medida BIC. Assim, verifica-se que entre os modelos e dentre essas séries, o modelo Combinado é o mais adequado a se utilizar. As tabelas apresentadas para cada série com as medidas detalhadas podem ser utilizadas para outras análises mas aqui essa análise parece suficiente para apontar a soberania do modelo sobre os demais.

Com os resultados obtidos, pode-se notar que os modelos combinados podem superar em muito as previsões realizadas por modelos de Box e Jenkins diminuindo os riscos e aumentando os retornos.

Embora os modelos combinados tenham alcançado resultados expressivos, ainda há vários pontos em aberto que necessitam de um estudo cuidadoso e aprofundado.

Um desses pontos a ser considerado é o conjunto de condições necessárias para a utilização dos modelos combinados. Quando a variância das séries reais é pequena, é necessários que os modelos a serem combinados apresentem capacidades de previsão semelhantes pois, caso contrário, o modelo combinado irá gerar valores que colapsarão para os gerados pelo melhor modelo.

Uma extensão imediata do trabalho exposto é a investigação de outros tipos de modelagens que possam vir integrar o modelo combinado melhorando sua capacidade de previsão. Podemos citar os modelos de algoritmos genéticos, os modelos de polinômios autoregressivos (PN) como os utilizados por Hong e Lee (2003) ou mesmo os modelos de coeficientes funcionais (FC) proposto por Cai, Fan e Yao (2000). No presente trabalho utilizamos uma combinação de apenas dois modelos como no trabalho de Zou e Yang (2004), porém podemos inserir o número de modelos que quisermos.

Uma outra perspectiva é procurar modelos alternativos para combinar com as RNADs para modelar os retornos. Com a variância da série real conhecida, os candidatos seriam modelos que tivessem erros quadráticos médios muito próximos do estimado para a RNAD uma vez que os baixos valores da variância amostral para as séries de retornos, amplificam na equação 4.1 qualquer pequena diferença entre os modelos.

Como já existem estudos sobre a computação quântica como em Hirvensalo (2004) e também de redes neurais quânticas como em Kouda *et al.* (2005), uma perspectiva futura seria estudar modelos combinados quânticos aplicados ao problema de previsão de séries temporais financeiras, analisando as correlações em estados complexos da representação de informação.

## Referências

- Akgiray, V., e Booth, G. G. The stable-law model of stock returns. *J. Bus. Econ. Statist.*, 6, 51-57, 1988.
- BOUTILIER, H. La toxicomanie à l'adolescence. In: GRAMMER, C. *et al. L'adolescence, crise familiale: therapie familiale par phases*, Toulouse: Éres, p. 111-130, 1992.
- Barucci, E. and R. Reno. On measuring volatility and the GARCH forecasting performance. *Journal of International Financial Markets, Institutions and Money* 12, 182-200; On measuring volatility of diffusion processes with high frequency data. *Economic Letters* 74, 371-378.
- Binney, J. J., Dowrick, N. J., Fisher, A. J. and Newman, M. E. J. (1992), **The Theory of critical phenomena: An introduction to the renormalization group**. Clarendon Press, Oxford, UK.
- BOCANEGRA, C. W. Procedimentos para tornar mais efetivo o uso das redes neurais artificiais no planejamento de transportes. *Dissertação Escola de Engenharia de São Carlos*, São Paulo, 2002.
- BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* Elsevier, v. 31(3), p. 307-327, 1986.
- BOLLERSLEV, T. and MELVIN, M. Bid-ask spreads and volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics*, v. 36, p. 355-372, 1994.
- BOX, G. and JENKINS, G. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1976.
- BOX, G.E.P. and G.M. JENKINS (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.
- BOX, G. E. P. and PIERCE, D. A. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. *Journal of the American Statistical Association.*, v. 65, p. 1509-1526, 1970.
- Cai, Z., J. Fan, and Q. Yao, "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association* 95:451 (2000), 941-956.
- LO, A. W. and MACHINLAY, A. C. *The econometrics of financial markets*. Princeton University Press, 1997.

- CARVALHO, A.G., Ascensão e Declínio do Mercado de Capitais no Brasil – a Experiência dos Anos 90. *Revista Economia Aplicada*, V. 4, N. 3, 2000.
- CAVALCANTI, A., Instabilidade e Não-Linearidades nos Mercados Financeiros, *Anais do XXXI Encontro Nacional de Economia, ANPEC*, Rio de Janeiro, 2003.
- CHAKRABORTY, K. et. al. Forecasting the behavior of multivariate Time Series Using Neural Networks. *Neural Networks*, E.U.A., n. 5, p. 961-970, 1992.
- CLEMENTS, M.; FRANCES, P.; SWANSON, N. Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, v. 20, p. 169-183, 2004. *Processes*, "Review of Economics and Statistics, November 1988, pp. 638-647.
- CLEMENTS, M. P. and Hendry, D. F. On the limitations of comparing mean square forecast errors. *Journal of Forecasting*, v 12, p. 617-637, 1993.
- DE AZEVEDO, F. M.; BARRETO, J. M. e LIMA, W.C. In: *Identification of arm dynamics using neural networks*, 46a Reunião Anual – SBPC., Vitória, ES, 1994. P. 97.
- DE GOOIJER, J. G. and KUMAR, K. Some recent developments in non-linear time series modelling, testing, and forecasting. *International Journal of Forecasting* Elsevier, v. 8(2), p. 135-156, 1992.
- F.X. Diebold and R. Mariano, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13 (1995), pp. 253–263.
- ELLIOT, G., ROTHENBERG, T. J. and STOCK, J. H. Efficient tests for an autoregressive unit root. *Econometrica*, v. 64, n. 4, p. 813-836, 1996.
- ENGLE, R. F. and GRANGER, C. W. J. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, v. 55, p. 251-276, 1987.
- ENGLE, R. Autoregressive conditional heteroscedasticity with estimates of the variance of u.k inflation. *Econometrica*, v. 50, p. 987-1008, 1982.
- FAMA, E. The behavior of stock-market prices. *Journal of Business*, v. 38, p. 31-105, 1965.
- FARUK, D. O. *A hybrid neural network and ARIMA model for water quality time series prediction*. Engineering Applications of Artificial Intelligence, 2009.
- Ferson, S. and W.T. Tucker. Probability Bounds with Uncertain Input Distributions and Correlations. Proceedings of the 6th World Multiconference on Systemics, Cybernetics, and Informatics, July 14-18, 2002, Orlando, FL.

- FRANSES, P. H. *Time series models for business and economic forecasting*. Cambridge University Press, 1998.
- S. Khebbal & S. Goonatillake *Intelligent Hybrid Systems: Issues, Classifications and Future Directions*, *Intelligent Hybrid Systems*, pp. 1-20, Wiley, London, 1995.
- GRANGER, C. W. J. and NEWBOLD, P. Spurious regressions in econometrics. *Journal of Econometrics* v. 2, p. 111–120, 1974.
- Grossberg, S. Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, L11:213-257, 1973.
- GUJARATI, D.N. *Basic Econometrics*, Nova York: McGraw-Hill, 2006.
- HAMILTON, J. *Time Series Analysis*. Princeton University Press, 1994.
- Hertz, S. R. (1991) Streams, phones, and transitions: toward a phonological and phonetic model of formant timing. *J. of Phonetics* 19, 91-109, 1991
- Higgins, P. G. (1993). Hybrid Human-Computer Production Scheduling. In *Proceedings of ACME'93*, Adelaide, November 1993, pp. 203-207.
- Hirvensalo, M. *Quantum Computing*. Natural Computing Series. Springer, second edition, 2004.
- Yongmiao Hong and Tae-Hwy Lee Inference on Predictability of Foreign Exchange Rates via Generalized Spectrum and Nonlinear Time Series Models *The Review of Economics and Statistics*, Vol. 85, No. 4 (Nov., 2003), pp. 1048-1062
- HOPFIELD, J. J. Neural network and physical systems with emergent collective computation abilities. In: *National Academy of science*, Anais, E.U.A., p. 2554-2558, 1982.
- JOHANSEN, S. and JUSELIUS, K. Maximum likelihood estimation and inference on cointegration-with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* v. 52(2), p. 169-210, 1990.
- N. Kouda, N. Matsui, H. Nishimura e F. Peper. An examination of qubit neural network in controlling an inverted pendulum. *Neural Processing Letter*, 22:227-290, 2005.
- KUPFER, D. e HASENCLEVER, L. *Economia Industrial: Fundamentos Teóricos e Práticos no Brasil*. Rio de Janeiro: Editora Campus, 2002.
- D. Kwiatkowski, P.C.B. Phillips, P. Schmidt and Y. Shin, Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54 (1992), pp. 159–178.

- S.H. Ling, F.H.F. Leung, H.K. Lam, P.K.S. Tam, and Y.S. Lce, "A novel CA-based neural network for shan-term load forecasting," *IEEE Trans. Ind. Eleciron.* vol.. 50, no. 4, pp. 793-799, Aug. 2003.
- LAPEDES, A. and FARBER, R. *Programming a massively parallel, computation universal system: static behavior. A, In neural networks for computing.* New York - American institute of physics, 1986.
- Leung, A.K., Andersen, J.S., Mann, M., Lamond, A.I. (2003) Bioinformatic analysis of the nucleolus *Biochem. J.* . 376, 553–569
- MANDELBROT, B. The variation of certain speculative prices. *The Journal of Business of the University of Chicagov.* 36, p. 394-419, 1963.
- Mandelbrot, B. "Une Classe de Processus Stochastiques Homothétiques à Soi: Application à la Loi Climatologique de H. E. Hurst", *Comptes Rendus*, 260, 3274-7, Paris, 1965.
- MANDELBROT, B. *Fractales, Hasard et Finance.* Paris: Flammarion, 1997.
- McCulloch, W. e Pitts, W., "A Logical Calculus of Ideas Immanent in Nervous Activity", 1943, *Bulletin of Mathematical Biophysics* 5:115-133.
- McCulloch, W. e Pitts, W., "On how we know universals: The perception of auditory and visual forms", *Bulletin of Mathematical Biophysics* 9:127-147, 1947.
- McLeod, A.I. On the distribution and applications of residual autocorrelations in Box-Jenkins modelling, *Journal of the Royal Statistical Society B* 40, 296-302, 1978.
- MILHOJ, A. The moment structure of ARCH processes. *Scandinavian Journal of Statisticsv.* 12, p. 281–292, 1985.
- Morettin, P. A. *Cálculo : funções de uma e várias variáveis* 2008
- NELSON, D. B. Conditional heteroskedasticity in asset returns: A new approach. *Econometricav.* 59, p. 347-370, 1991.
- NG, S. and PERRON, P. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, v. 69, n. 6, 2001.
- NOBREGA, M., LOYOLA, G., GUEDES FILHO, E. M., PASQUAL, D. O mercado de capitais: sua importância para o desenvolvimento e os entraves com que se defronta no Brasil. *São Paulo: Bovespa*, 2000. Disponível em: <<http://www.bovespa.com.br>> Acesso em: 11/11/2003.

- OZAKI, T. Non-linear time series models for non-linear random vibration. *Journal of applied probability*, 17:84-93, 1980.
- PINHEIRO, J. L. *Mercado de capitais: fundamentos e técnicas*. 3º Ed. São Paulo: Atlas, 2005.
- POKORNY, M. *An Introduction to Econometrics*. Nova York: Basil Blackwell, p. 343, 1987.
- Polchinski, J. (1984), Renormalization group and effective lagrangians. **Nuclear Physics B**, 231:269–295.
- Politzer, H. D. (1974), Asymptotic freedom: An approach to strong interactions. **Physics Reports**, 14(4):129–180.
- REFENES A. N., *et al* Currency Exchange rate prediction and Neural Network Design Strategies. *Neural computing & Applications*sv. 1(1), 1993.
- RODRIGUES, M.A.B. e DE AZEVEDO, F. e. M. N. J. e. A. V. Dynamic neural network for recognition of sleep spindles and k complex. *European Medical And Biological Engineering Conference*, 1999.
- RODRIGUES, M.A.B e ANDRIANI, V. M. e. M. N. J. e. D. A. F. Uma rede neural dinâmica para identificador de padrões eletrográficos. *Segundo Congreso Lati-noamericano de Ingeniería Biomédica*, HABANA, 2001.
- RODRIGUES (a), M.B. *Sistema para Detecção e Classificação de Apnéias do sono apartir de Registros Polissonográficos*. Tese (Doutorado em Engenharia Biomédica) Centro Tecnológico, Universidade Federal de Santa Catarina. Florianópolis.2003.
- Rosenblatt, F. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." Spartan Books. 1958
- G. Schwarz, Estimating the dimension of a model, *Annals of Statistics* 6 (1978), pp. 461–464
- SFETSOS, A. and SIRIOULOS, C. Times series in forecasting with a hybrid clustering scheme and parttner recognition. *IEEE Transactions on Systems, Man, And, Cybernetics - Part A: Systems and Humans*sv. 34, p.399-405, 2004.
- SIMS, C. A., STOCK, J. H. and WATSON, M. W. Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*sv. 58(1), p. 113-44, 1990.
- SIMS, C. A. Macroeconomics and Reality. *Econometrica*sv. 48(1), p.1-48, 1980.

- SPANOS, A. *Probability, Theory and Statistical Inference: Econometric Modeling with Observational Data*. Reino Unido: Cambridge University Press, p.21, 1999.
- SUBBA, R. and GABR, M. *An introduction to bispectral analysis and bilinear time series models*. Springer-Verlag, 1984.
- Tucker, A., and Pond, L. "The Probability Distribution Of Foreign Exchange Price Changes: Tests Of Candidate
- O mercado de capitais: Sua importância para o desenvolvimento e os entraves com que se defronta o Brasil, [www.bovespa.com.br/pdf/relatório2.pdf](http://www.bovespa.com.br/pdf/relatório2.pdf)
- The Mathworks, Inc. MATLAB Product Family Manuals. Natick, 1996.
- TONG, H. and LIM, K. S. Threshold Autoregression, Limit Cycles and Cyclical Data (with discussion). *Journal of the Royal Statistical Society Series B*, v. 42, p. 245-292, 1980.
- HANN, T. H., STEURER, E. Much ado about nothing? exchange rate forecasting: Neural networks vs. linear models using monthly and weekly data.] *Neurocomputing*, v. 10, p. 323-339, 1996.
- UTANS, J and MOODY, J. E. Selecting neural network architectures via the prediction risk: application to corporate bond rating prediction. *Anais: Proc. 1st Int. Conference on AI Applications on Wall Street, IEEE Computer Society Press*, 1991.
- WEIGEND, A. S. and GERSHENFELD, N. A. *Time Series Prediction: Forecasting the Future and Understanding the Past*. Perseus Books, 1994.
- WEIGEND, A. S., RUMELHART, D. E., and HUBERMAN, B. A. (1990). Predicting the future - a connectionist approach. *International Journal of Neural System*, 1:193-209.
- WINDROW, B. e HOFF, M. Adaptive Switching Circuits, IRE WESCON Convention Record, New York, N.Y., 1960.
- Yule, G.U. Why do we sometimes get nonsense-correlations between time series?. *Journal of the Royal Statistical Society* 89 (1926), pp. 1-69
- Hui Zou, Yuhong Yang Combining time series models for forecasting *International Journal of Forecasting* 20 (2004) 69- 84