

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
PRODUÇÃO E SISTEMAS**

**EDGAR LEITE DOS SANTOS FILHO**

**PREVISÃO DOS RETORNOS DO ÍNDICE BOVESPA  
USANDO REDES NEURAIS ARTIFICIAIS**

**CURITIBA**

**2008**

**EDGAR LEITE DOS SANTOS FILHO**

**PREVISÃO DOS RETORNOS DO ÍNDICE BOVESPA  
USANDO REDES NEURAS ARTIFICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

Orientador: Prof. Dr. Leandro dos Santos Coelho

**CURITIBA**

**2008**

**EDGAR LEITE DOS SANTOS FILHO**

**PREVISÃO DOS RETORNOS DO ÍNDICE BOVESPA  
USANDO REDES NEURAIAS ARTIFICIAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Sistemas da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Engenharia de Produção e Sistemas.

**CURITIBA**

**2008**

## **AGRADECIMENTOS**

- Em primeiro lugar agradeço a Deus, que me proporcionou a graça de viver, aprender e crescer.
- À minha mãe, Clea Maria Bandeira dos Santos, que sempre me dedicou amor e com quem aprendi a ter orgulho e continuar acreditando.
- À minha esposa, Beatriz de Fatima Storch dos Santos e filha, Bruna Fernanda Storch dos Santos, pela paciência, apoio e carinho, para que juntos pudéssemos provar o sabor dessa conquista.
- Ao Prof. Dr. Leandro dos Santos Coelho, meu orientador, agradeço a grande e positiva influência que efetivamente exerceu em minha formação acadêmica e pela brilhante contribuição neste trabalho.
- Ao Prof. Dr. Wesley Vieira da Silva, pela imensurável ajuda em todas as etapas do trabalho.
- E são tantas as pessoas que, das mais diferentes formas, participaram desta história. A elas, o meu mais sincero agradecimento.

## LISTA DE FIGURAS

- Figura 1.1 – Comportamento dos preços do IBOVESPA à vista entre Jun/2004 – Maio/2008
- Figura 2.1 – a) rede (*feedforward*) acíclica b) rede (recorrente de Elman) cíclica
- Figura 2.2 – Rede parcialmente recorrente de Elman
- Figura 2.3 – Ciclo típico da evolução das conexões dos pesos
- Figura 2.4 – Rede neural *feedforward* com uma camada escondida
- Figura 2.5 – Estrutura geral de um algoritmo evolutivo
- Figura 2.6 – Processo de geração do vetor tentativa  $\underline{v}$  no esquema ED1.
- Figura 2.7 – Ilustração do processo de crossover para  $D=7$ ,  $n=2$  e  $L=3$
- Figura 2.8 – Processo de geração do vetor tentativa  $\underline{v}$  no esquema ED2.
- Figura 4.1 – Série de preços do IBOVESPA à vista entre Jun/2004 – Mai/2008
- Figura 4.2 – Série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Figura 4.3 – Histograma da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Figura 4.4 – Série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Figura 4.5 – Função de autocorrelação dos log-retornos quadrados do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Figura 4.6 – Função de autocorrelação parcial dos log-retornos quadrados do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Figura 4.7 – Série de preços do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.8 – Série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.9 – Histograma da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.10 – Série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.11 – Função de autocorrelação dos log-retornos quadrados do

IBOVESPA futuro entre Jun/2004 e Mai/2008

- Figura 4.12 – Função de autocorrelação parcial dos log-retornos quadrados do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.13 – Visão geral do processo de previsão um passo a frente do IBOVESPA
- Figura 4.14 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método GARCH(1,1) em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.15 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método GARCH(1,1) em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008
- Figura 4.16 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.17 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS em relação à série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Figura 4.18 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.19 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.20 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.21 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008
- Figura 4.22 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *rand/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008

- Figura 4.23 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *rand/1/bin* em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008
- Figura 4.24 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *best/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.25 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *best/1/bin* em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008
- Figura 4.26 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.27 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008
- Figura 4.28 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Figura 4.29 – Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008

## **LISTA DE TABELAS**

- Tabela 4.1 – Estatísticas da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Tabela 4.2 – Estatísticas da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008
- Tabela 4.3 – Estatísticas da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Tabela 4.4 – Estatísticas da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008
- Tabela 4.5 – Medidas de desempenho da série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008
- Tabela 4.6 – Medidas de desempenho da série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008



## LISTA DE ABREVIATURAS

ARFIMA	– Auto-Regressivo Fracionalmente Integrado de Médias Móveis, do inglês, <i>AutoRegressive Fractionally Integrated Moving Average</i>
AR	– Auto-Regressivo, do inglês, <i>AutoRegressive</i>
ARMA	– Auto-Regressivo de Média Móvel, do inglês, <i>AutoRegressive Moving Average</i>
ARIMA	– Auto-Regressivo Integrado de Média Móvel, do inglês, <i>AutoRegressive Integrated Moving Average</i>
ARMA-GARCH	– Modelo ARMA considerando Heterocedasticidade Auto-Regressiva, do inglês, <i>ARMA- Autoregressive Conditional Heteroskedasticity</i>
ARCH	– Heterocedasticidade Condicional Auto-Regressiva, do inglês, <i>Autoregressive Conditional Heteroskedasticity</i>
Quase-Newton	– Método Quase Newton com aproximação da Hessiana proposta por
BFGS	Broyden, Fletcher, Powell e Goldfarb
BOVESPA	– Bolsa de Valores de São Paulo
EGARCH	– GARCH Exponencial, do inglês, <i>Exponencial GARCH</i>
ED	– Evolução Diferencial
EQM	– Erro Quadrático Médio, do inglês, <i>Mean Square Error</i>
FIGARCH	– ARCH Fracionalmente Integrados, do inglês, <i>Fractionally Integrated ARCH</i>
GARCH	– Heterocedasticidade Condicional Auto-Regressiva Generalizada, do inglês, <i>Generalized Autoregressive Conditional Heteroskedasticity</i>
IBOVESPA	– Índice BOVESPA
LOGIT	– Modelagem Logística
MAPE	– Erro Médio Percentual Absoluto, do inglês, <i>Mean Absolute Percentage Error</i>
MAE	– Erro Médio Absoluto, do inglês, <i>Mean Absolute Error</i>
NASDAQ	– <i>North American Securities Dealers Automated Quotation</i>
REQM	– Raiz Quadrada do Erro Quadrático Médio, do inglês, <i>Root Mean Square Error</i>
RNA	– Rede Neural Artificial
RNAE	– Rede Neural Artificial Evolutiva

# SUMÁRIO

<b>AGRADECIMENTOS .....</b>	<b>iv</b>
<b>LISTA DE FIGURAS.....</b>	<b>v</b>
<b>LISTA DE TABELAS .....</b>	<b>viii</b>
<b>LISTA DE ABREVIATURAS.....</b>	<b>ix</b>
<b>SUMÁRIO .....</b>	<b>x</b>
<b>RESUMO.....</b>	<b>xii</b>
<b>ABSTRACT .....</b>	<b>xiii</b>
<b>1. INTRODUÇÃO .....</b>	<b>1</b>
1.1 APRESENTAÇÃO DO CONTEXTO .....	1
1.2 PROBLEMA DE PESQUISA .....	3
1.3 OBJETIVOS DA PESQUISA .....	6
1.3.1 Objetivo geral .....	6
1.3.2 Objetivos específicos.....	6
1.4 JUSTIFICATIVAS DA PESQUISA .....	7
1.4.1 Justificativa teórica .....	7
1.4.2 Justificativa prática .....	13
1.5 ESTRUTURA DO TRABALHO .....	16
<b>2. FUNDAMENTAÇÃO TEÓRICO-EMPÍRICA .....</b>	<b>17</b>
2.1 EFICIÊNCIA DE MERCADO.....	17
2.1.1 Testes de previsibilidade .....	19
2.1.2 Testes da forma semiforte.....	19
2.1.3 Testes da forma forte .....	20
2.2 MODELO DE HOLT-WINTERS .....	20
2.3 REDES NEURAIAS ARTIFICIAIS EVOLUTIVAS.....	22
2.3.1 Fundamentos das redes neurais artificiais .....	23
2.3.2 Redes neurais parcialmente recorrentes .....	25
2.3.3 Evolução dos pesos das conexões .....	28
2.4 MÉTODOS DE OTIMIZAÇÃO BASEADOS EM GRADIENTE .....	29
2.4.1 Algoritmo de retropropagação do erro .....	29
2.4.2 Método Quase-Newton.....	34

2.5 EVOLUÇÃO DIFERENCIAL.....	36
2.5.1 Tamanho da população.....	38
2.5.2 Seleção de ancestrais.....	39
2.5.3 Operador de recombinação.....	39
2.5.4 Operador de mutação.....	39
2.5.5 Estratégias de ED.....	40
2.6 MODELOS DA FAMÍLIA ARCH.....	43
2.6.1 Presença do efeito ARCH.....	45
2.6.2 Modelo ARCH.....	46
2.6.3 Modelo GARCH.....	47
2.7 MEDIDAS DE DESEMPENHO.....	48
<b>3. METODOLOGIA DE PESQUISA.....</b>	<b>50</b>
3.1 CARACTERIZAÇÃO DA PESQUISA.....	50
3.2 COLETA E TRATAMENTO DOS DADOS.....	51
3.3 DEFINIÇÃO DAS VARIÁVEIS A SEREM ESTUDADAS.....	51
3.4 HIPÓTESE DE PESQUISA.....	55
<b>4. APRESENTAÇÃO E ANÁLISE DOS DADOS.....</b>	<b>56</b>
4.1 AS SÉRIES TEMPORAIS DO IBOVESPA.....	56
4.1.1 Análise da série de dados do IBOVESPA à vista.....	56
4.1.2 Análise da série de dados do IBOVESPA futuro.....	62
4.2 COMPARAÇÃO DE DESEMPENHO ENTRE MÉTODOS DE PREVISÃO.....	67
<b>5. CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....</b>	<b>88</b>
5.1 CONCLUSÕES.....	88
5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	90
<b>REFERÊNCIAS.....</b>	<b>91</b>

## RESUMO

Os mercados financeiros e de capitais, especialmente os mercados de ações, são considerados investimentos de alto risco, dominados por incertezas e volatilidades. A previsão no mercado de ações pode ser útil para lidar com esta incerteza e, conseqüentemente, o risco. Como os mercados de ações são influenciados por fatores econômicos, políticos e até mesmo psicológicos, é difícil prever o movimento futuro dos valores. Uma vez que as métricas estatísticas e/ou econométricas tradicionais, usadas como suporte para a análise técnica e fundamentalista, não são capazes de absorver as não linearidades dos conjuntos de dados, torna-se necessária a utilização de procedimentos de previsão avançados. As previsões financeiras por meio das técnicas de análise da volatilidade juntamente com as redes neurais artificiais (RNAs) foram apresentadas como métodos promissores para tais previsões. O projeto de uma RNA tem inspiração em aspectos de aprendizado do cérebro humano. As RNAs têm o potencial de aprender o mecanismo dos mercados de ações, isto é, capturar a complexa dinâmica e as não linearidades das séries temporais do mercado de ações. Neste trabalho, alguns conceitos e teorias sobre séries temporais financeiras ligadas aos mercados de ações, bem como as RNAs e os sistemas baseados em técnicas híbridas utilizadas para solucionar problemas de previsão de interesse dos mercados financeiros, são revisados. Nesta dissertação são validados métodos de previsão baseados em RNAs de Elman, treinadas pelo método Quase Newton do tipo BFGS (método Quase Newton com aproximação da Hessiana proposta por Broyden, Fletcher, Powell e Goldfarb) e com o algoritmo da computação evolutiva denominado de evolução diferencial. Esses métodos são utilizados para a previsão, um passo à frente, dos valores do índice BOVESPA (IBOVESPA), com base em suas cotações históricas tomadas diariamente e sua capacidade preditiva comparada com base no coeficiente de desigualdade U-Theil.

Os métodos baseados em RNAs Elman combinados com o quase Newton do tipo BFGS e a evolução diferencial obtiveram melhores resultados, ainda que não significativamente, em relação ao coeficiente de desigualdade U-Theil, do que o método GARCH(1,1). A combinação Elman e ED *best/1/bin* foi o método com o melhor desempenho entre os validados.

**PALAVRAS-CHAVE:** Redes Neurais Artificiais, Séries Temporais, Evolução Diferencial, GARCH, Previsão.

## ABSTRACT

Financial and capital markets, especially stock markets, are considered high risk investment fields, dominated by uncertainty and volatility. Stock market prediction tries to reduce this uncertainty and consequently the risk. As stock markets are influenced by many economical, political and psychological factors, it is difficult to forecast the movement of future values. Since traditional statistical and econometrics metrics, used for technical and fundamental analysis, are unable to deal with the non-linearity in the dataset, it becomes necessary the utilization of advanced forecasting procedures. Financial forecasting based on volatility analysis and artificial neural networks have been proposed as one of the most promising methods for such predictions. ANN (artificial neural network) simulates the learning capability of the human brain. They have a potential of learning the underlying mechanics of stock markets and capture the complex dynamics and non-linearity of the stock market time series. In this work, some financial time series concepts and theories linked to stock markets will be studied, as well as neural networks based on systems and hybrid techniques used to solve forecasting problems concerning financial markets. This research validates some financial time series forecasting methods based on Elman neural network, trained with BFGS (Quasi-Newton method with Broyden, Fletcher, Powell and Goldfarb hessian approximation) and an evolutionary algorithm named differential evolution. These methods will be used to predict one-step-ahead index values of IBOVESPA, based on daily historical prices and their forecasting power are compared using U-Theil's coefficient.

The methods based on Elman ANN and combined with quasi Newton BFGS and differential evolution obtained better results, even they aren't significant when compared to U-Theil inequality coefficient of GARCH method. Elman ANN combined with *best/1/bin* differential evolution obtained the best performance between validated methods.

**KEYWORDS:** Artificial Neural Networks, Time Series, Differential Evolution, GARCH, Forecasting.

## 1. INTRODUÇÃO

Este capítulo procura fornecer uma visão do que é apresentado ao longo do trabalho, sendo este estruturado em cinco seções. A seção 1.1 refere-se à apresentação do contexto desta dissertação. A seção 1.2 trata do problema de pesquisa abordado na dissertação. A seção 1.3 refere-se aos objetivos geral e específicos. A seção 1.4 justifica esta pesquisa, enquanto estudo científico, e a seção 1.5 é destinada à apresentação da organização do trabalho.

### 1.1 APRESENTAÇÃO DO CONTEXTO

Em um mercado competitivo é necessário se antecipar aos acontecimentos de forma que sejam tomadas ações inovadoras que se transformem em vantagem competitiva. Partindo-se do pressuposto que eventos passados podem explicar o futuro, é necessária a procura de métodos que capturem características do comportamento passado a fim de prever o próximo acontecimento com certo grau de confiança (BÓDIS, 2004).

De acordo com a Hipótese de Mercados Eficientes (FAMA, 1970), o preço das ações possui uma dinâmica aleatória, ou seja, não segue nenhum tipo de ordem ou padrão. Essa questão é fundamental para o debate relacionado com a previsibilidade do mercado, pois, caso a evolução do preço das ações realmente não apresente nenhuma tendência sistemática, ou seja, o mercado seja eficiente na sua formação de preço, é impossível prevê-lo (SANTOS, 2005). O conceito de mercado eficiente é usualmente associado à hipótese de *martingale* ou jogo justo (CAMPBELL *et al.*, 1997), cuja condição é dada por

$$E[P_{t+1} / P_t, P_{t-1}, \dots] = 0, \quad (1.1)$$

onde  $E$  é a esperança matemática e  $P_i$  é o preço do ativo na data  $i$ .

Tomando-se  $P_t$  como o preço do ativo na data  $t$ , a hipótese de *martingale* assegura que o preço de amanhã é esperado ser igual ao preço de hoje, dada a série histórica dos preços do ativo. Alternativamente, a expectativa de variação do preço é zero, quando condicionada à história dos preços; dessa forma, o preço pode tanto cair quanto subir. Vista de uma

perspectiva de previsão, a hipótese de *martingale* implica que a “melhor” previsão do preço de amanhã é simplesmente o preço de hoje. Outro aspecto da hipótese é que as mudanças de preços são não-correlacionadas em todas as defasagens, implicando a ineficácia das regras lineares de previsão das mudanças futuras de preços baseadas apenas no histórico dos preços.

A hipótese de *martingale* levou ao desenvolvimento de uma hipótese correlata que se tornou fundamental nos estudos sobre eficiência dos mercados: a hipótese do caminho aleatório (*random walk hypothesis*). A hipótese (ou modelo) do caminho aleatório pode ser representada pela equação:

$$P_t = \mu + P_{t-1} + \varepsilon_t \quad (1.2)$$

onde  $\mu$  é a média e  $\varepsilon_t$  é um erro independente e identicamente distribuído.

Este modelo diz que o preço do ativo no período  $t+1$  é dado pelo seu preço no período imediatamente anterior, mais um termo de média esperada  $\mu$  conhecido como *drift*, mais um componente de erro não-previsível. O comportamento do termo  $\varepsilon_t$  é importante, e as suas restrições acerca do comportamento deste termo motivaram Campbell *et al.* (1997) a produzir três versões do modelo de caminho aleatório:

*Caminho aleatório 1*: Nesta versão, também denominada de versão “forte” do modelo *random walk*, assume-se que o termo  $\varepsilon_t$  é identicamente distribuído e independente. A independência implica que não só os incrementos são não-correlacionados, mas também qualquer função não-linear dos incrementos será não-correlacionada.

*Caminho aleatório 2*: Esta versão considera que os incrementos do termo  $\varepsilon_t$  são independentes, mas não necessariamente identicamente distribuídos. O relaxamento dessa hipótese mantém a característica da não-previsibilidade linear além de admitir a presença de fatos estilizados em séries financeiras, com mudanças na volatilidade condicional.

*Caminho aleatório 3*: A versão mais geral e com significativo número de testes empíricos substitui a hipótese de incrementos independentes pela hipótese de incrementos dependentes, porém, não-correlacionados. Isto significa, por exemplo, satisfazer a condição de que a

covariância  $Cov[\varepsilon_t^2, \varepsilon_{t-k}^2] \neq 0$ , para todo  $k \neq 0$ . Tal processo indica que os incrementos não são correlacionados, mas, são dependentes, desde que o quadrado dos incrementos seja correlacionado.

Segundo Santos (2005), a pesquisa científica sugeriu, por muito tempo, a impossibilidade de se refutar o modelo do caminho aleatório nas variações de preços dos ativos financeiros. Entretanto, como notam Campbell *et al.* (1997), o avanço computacional e o surgimento de métricas econométricas robustas, passaram a sugerir algum nível de previsibilidade nos retornos dos ativos financeiros, indicando que as variações de preços não podem ser totalmente explicadas pelo método do caminho aleatório e que estratégias de negociação que visem identificar essas tendências podem superar estratégias do tipo comprar e esperar.

## 1.2 PROBLEMA DE PESQUISA

Segundo Kostela (2003), uma série temporal consiste de medidas ou observações de um processo tomadas em períodos regulares de tempo. A principal motivação na pesquisa de séries temporais é o desejo de entender os processos e fenômenos presentes nos sistemas em estudo. Os objetivos principais da análise de séries temporais incluem identificação, reconstrução ou mesmo a previsão de eventos futuros.

Ao se analisar uma ou mais séries temporais, a representação gráfica dos dados sequencialmente ao longo do tempo é fundamental e pode revelar padrões de comportamento. As tendências de crescimento (ou decrescimentos), padrões cíclicos, alterações estruturais, observações fora do comum, entre outras, são facilmente identificadas.

Segundo a abordagem de decomposição clássica, as séries temporais podem ser representadas como uma combinação dos componentes de tendência, ciclicidade e sazonalidade, conforme a equação 1.3 (ABELÉM, 1994), tal que:

$$X_t = T_t + C_t + R_t, \quad (1.3)$$



onde  $T_t$  é o componente de tendência,  $C_t$  é o componente cíclico ou sazonal e  $R_t$  é o componente aleatório.

A tendência é um movimento ascendente ou descendente de longa duração que pode ser persistente ou percebida quando se observa a série temporal de forma mais abrangente. Na Figura 1.1, pode-se observar uma tendência de alta do IBOVESPA à vista, no período de Junho de 2004 até Maio de 2008.

A sazonalidade e ciclicidade são caracterizadas por tendências de curta duração que geralmente obedecem a um padrão de ocorrência dentro de um período de tempo pré-determinando. A sazonalidade é freqüentemente observada em períodos que compreendem 12 meses e a ciclicidade usualmente é notada em períodos maiores que um ano. Na Figura 1.1 pode-se observar um movimento de alta sazonal acontecendo entre os pontos (1) e (2) (jun/2005 a jun/2006), e novamente entre os pontos (2) e (3), respectivamente, de jun/2006 a jun/2007.

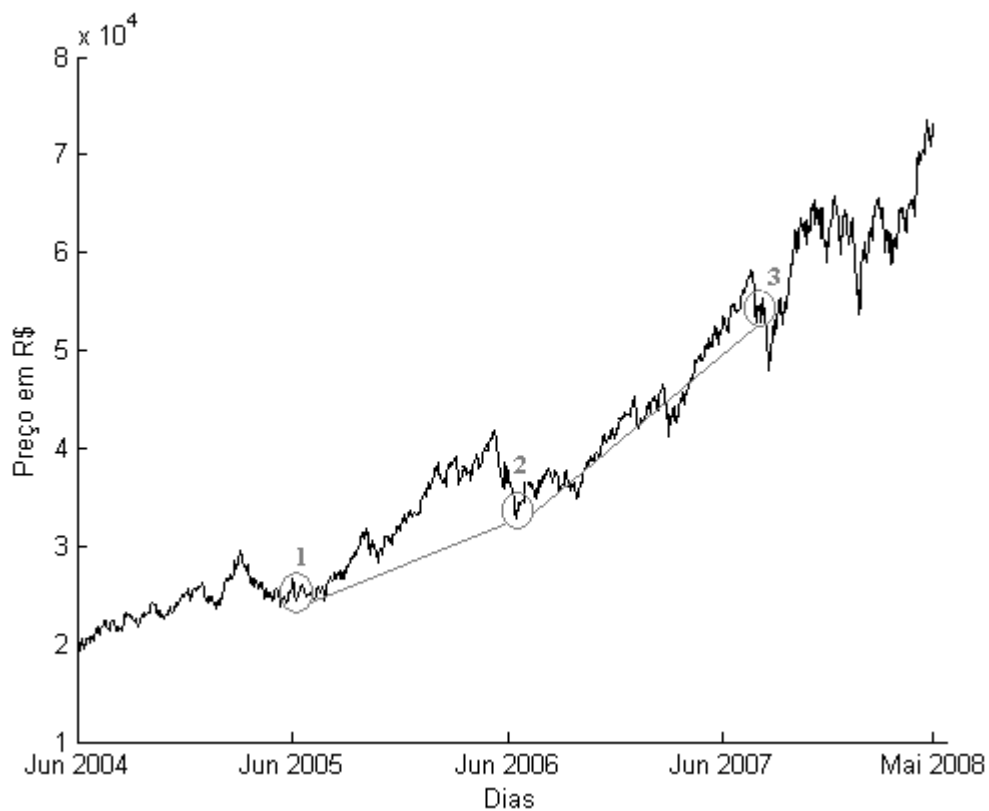


Figura 1.1 - Comportamento dos preços do IBOVESPA à vista entre Jun/2004 e Maio/2008.

A partir da análise no domínio do tempo e/ou análise no domínio de frequência, o analista tem como objetivo construir modelos para, entre outros, investigar o mecanismo gerador da série temporal, descrever o comportamento da série, procurar periodicidades relevantes nos dados e fazer previsões de valores futuros da série. A construção dos modelos deve ser fundamentada nos princípios da simplicidade, parcimônia e facilidade de utilização.

Nas áreas de conhecimento das ciências, tecnologia e economia, existem processos e fenômenos cuja previsão é de grande utilidade. Esses incluem diferentes processos industriais que podem ser modelados, preditos e controlados com base em dados sensoriais. Preços de ações, índices de ação, taxa de câmbios e juros são exemplos de indicadores cuja previsão é de grande utilidade na área econômica (SANTOS, 2005).

Em geral, as observações das séries temporais apresentam autocorrelação e, por causa dessa dependência, faz sentido e é possível prever o futuro das séries temporais. Não importando se é linear ou não-linear em sua natureza, a dependência pode ser estimada pela utilização de métodos estatísticos. Contudo, as observações de um processo frequentemente possuem ruídos que não podem ser cancelados e podem ser dependentes de variáveis desconhecidas que afetam o processo e não podem ser estimadas. Além disso, existem questões a respeito de processos que são, ou não, passíveis de previsão. Mas, neste caso, considera-se que um modelo estatístico pode ser ajustado aos dados disponíveis e com base em algum índice de desempenho, por exemplo, EQM (Erro Médio Quadrático), REQM (Raiz do Erro Médio Quadrático),  $R^2$  (coeficiente de explicação) e coeficiente de desigualdade U-Theil, entre outros, avaliar a utilidade do método.

O problema então se resume em construir modelos capazes de capturar as características do processo usando medidas disponíveis. O modelo que mais precisamente se ajusta à série temporal é considerado como o que melhor caracteriza o sistema. De uma forma geral, o modelo recebe informações de valores passados da série temporal e retorna a previsão.

Na previsão um passo à frente, o modelo fornece a previsão para o próximo valor da série temporal. Na previsão  $n$  passos à frente, o modelo gera  $n$  previsões à frente. Por meio da previsão um passo à frente é possível construir um procedimento iterativo e gerar previsões  $n$

passos à frente, mas, esse procedimento pode não atingir uma precisão aceitável uma vez que acumula o erro.

O problema da seleção das variáveis que servem de entrada para os modelos pode tornar-se complexo, uma vez que essas variáveis são dependentes da natureza do modelo utilizado na previsão. Para os modelos lineares estáticos, a autocorrelação entre os valores consecutivos pode ser estimada e usada como guia para decidir o número de variáveis de entrada.

### 1.3 OBJETIVOS DA PESQUISA

Os objetivos desta pesquisa foram divididos em geral e específicos, os quais são sumarizados nas subseções a seguir.

#### 1.3.1 Objetivo geral

Avaliar a previsão, um passo à frente, dos retornos do Índice Bovespa, à vista e futuro, usando RNA parcialmente recorrente de Elman, treinadas por métodos híbridos compostos por algoritmos de descida de encosta (*descent gradient*) e evolução diferencial.

#### 1.3.2 Objetivos específicos

Visando alcançar o objetivo geral, os seguintes objetivos específicos foram traçados:

- Projetar uma RNA de Elman treinada com algoritmo quase-Newton do tipo BFGS;
- Projetar uma RNA de Elman treinada com algoritmo evolução diferencial;

- Projetar uma RNA de Elman que terá suas conexões ajustadas por um procedimento híbrido utilizando os algoritmos evolução diferencial e quase-Newton do tipo BFGS;
- Realizar a previsão, um passo à frente, utilizando as redes projetadas, e comparar os valores obtidos com o modelo GARCH (Heterocedasticidade Condicional Auto-Regressiva Generalizada, do inglês, *Generalized Autoregressive Conditional Heteroskedasticity*).

## 1.4 JUSTIFICATIVAS DA PESQUISA

As justificativas teóricas e práticas desta dissertação encontram-se estruturadas nas subseções apresentadas a seguir.

### 1.4.1 Justificativa teórica

Uma vez que pesquisas evidenciaram a autocorrelação e a possibilidade de previsão dos retornos de séries temporais financeiras, é necessário o estudo de modelos visando capturar a correlação dos retornos, extrair e modelar as características não-lineares, a fim de fazer previsões de valores futuros da série temporal (FANG, 2003).

Dentre as técnicas de modelagem linear, o alisamento exponencial é uma técnica de previsão linear que busca a modelagem do comportamento geral da série temporal ignorando as observações discrepantes, ou seja, aquelas mais distantes da média. O alisamento exponencial tem suas bases nos estudos realizados entre os anos 1950 e 1960, por Brown (1959), Holt (1957) e Winters (1960), mas, na época não recebeu apoio dos estatísticos e permaneceu sem desenvolver a fundamentação estatística necessária até que Gradner (1985) publicou uma revisão do que havia sido pesquisado até aquela data. Os trabalhos publicados posteriormente por Makridakis e Hibon (1991) focaram nas propriedades empíricas dos métodos, nas propostas de novos métodos de estimação e configuração dos parâmetros iniciais, e também no método de decaimento multiplicativo de Taylor (2003).

Outra técnica linear com significativa quantidade de pesquisas na previsão de séries temporais é a metodologia para construção de modelos ARIMA (Auto-Regressivo Integrado de Média Móvel, do inglês, *Autoregressive Integrated Moving Average*) que foi apresentada pelos pesquisadores Box e Jenkins (1970), e que influenciou significativamente a teoria e prática moderna de análise e previsão de séries temporais.

Quando se trata de modelagem ARIMA, a previsão das séries temporais pode ser realizada a partir de modelos univariados ou modelos de função de transferência. Existem vários métodos para estimação de parâmetros de um modelo ARIMA (Auto-Regressivo de Média Móvel, do inglês, *Autoregressive Integrated Moving Average*), mas, apesar de serem assintoticamente equivalentes no sentido de que tendem para uma distribuição normal, geram resultados significativamente diferentes e, conseqüentemente, influenciam na previsão. Após alguns estudos conduzidos por Newbold (1994), recomenda-se o uso do método da máxima verossimilhança para a estimação de parâmetros desse tipo de modelagem.

Moshiri e Cameron (2000) compararam a previsão da taxa mensal de inflação canadense, utilizando MAE (Erro Médio Absoluto, do inglês, *Mean Absolute Error*) como medida de acurácia dos modelos matemáticos. A previsão foi realizada por modelos ARIMA, redes neurais de múltiplas camadas *feedforward* e modelos estruturais. Os modelos ARIMA foram superiores, mas, na previsão de doze passos (meses), as redes neurais conseguiram pequena vantagem preditiva.

Quando se considera os métodos de análise e previsão de séries temporais não-lineares, atribui-se a Volterra (1930) os primeiros estudos mostrando que qualquer função não-linear contínua em  $t$  poderia ser aproximada por uma série finita. Wiener (1958) se interessou pela representação funcional das séries e desenvolveu o assunto posteriormente. Mesmo depois de terem suas propriedades probabilísticas estudadas extensivamente, os problemas de estimação de parâmetros, ajuste de modelo e previsão, foram negligenciados por muito tempo, o que pode ser atribuído à complexidade do modelo de Wiener.

Embora a pressuposição de linearidade seja útil, no fim dos anos 1970 e início dos anos 1980, pesquisas mostraram, de forma gradativa, que os modelos lineares não geram bons resultados em muitas aplicações reais. Por razões práticas, alguns modelos foram desenvolvidos no mesmo período. De Gooijer e Kumar (1992) produziram um resumo do

desenvolvimento de modelos não-lineares até o início dos anos 1990 e argumentaram sobre a hipótese de que os modelos não-lineares tenham maior poder de predição quando comparados com os lineares.

A família de modelos ARCH (Heterocedasticidade Condicional Auto-Regressiva, do inglês, *Autoregressive Conditional Heteroskedasticity*), introduzida por Engle (1982), é reconhecida atualmente como uma técnica de referência para a modelagem e previsão da volatilidade do retorno de um ativo financeiro, uma vez que procura modelar a variância da variável ao longo do tempo, que é uma característica das séries temporais financeiras. Posteriormente, Bollerslev (1986), motivado pela dificuldade da estimação de um número grande de parâmetros necessários para os modelos ARCH, propôs uma generalização ao modelo, o qual ficou conhecido como GARCH (Heterocedasticidade Condicional Auto-Regressiva Generalizada, do inglês, *Generalized Autoregressive Conditional Heteroskedasticity*).

Lux (2007) investigou a previsibilidade da volatilidade e volume de uma amostra de ações japonesas pelos métodos ARFIMA (Auto-regressivo Fracionalmente Integrado de Média Móvel, do inglês, *Autoregressive Fractionally Integrated Moving Average*) e FIGARCH (GARCH fracionalmente integrado, do inglês, *Fractionally Integrated GARCH*), e a hipótese da superioridade desses métodos quando têm seus resultados comparados com modelos GARCH e ARMA. De uma forma geral, os modelos FIGARCH e ARFIMA apresentaram maior capacidade de previsão quando comparados pelos critérios do MAE (Erro Médio Absoluto, do inglês, Mean Absolute Error) e REQM (Raiz Quadrada do Erro Quadrático Médio, do inglês, Root Mean Square Error).

A previsão de séries temporais, por métodos não-lineares, ainda carece de uma análise mais detalhada e desenvolvimentos. Neste contexto, a proposta deste trabalho se concentra na combinação de redes neurais e métodos de otimização para o seu treinamento.

Chatfield (1995) questionou a possibilidade ter havido uma valorização exagerada das RNAs no que se refere à sua capacidade de previsão. Gorr e Hill (1994) sugeriram que pesquisas futuras deveriam investigar melhor os limites de quando as RNAs superam as técnicas tradicionais.

Segundo Darbellay e Slama (2000), uma RNA pode ser útil para a previsão de sistemas não-lineares que tem relacionamento funcional desconhecido e, por conseqüência, possui ajustamento difícil. A idéia principal das RNAs é que as entradas são filtradas a partir de uma ou mais camadas escondidas, cada uma das quais compostas por unidades escondidas, ou neurônios, antes de produzir o valor para a variável de saída.

Enquadrando a RNA no contexto de métodos estatísticos clássicos, pode-se classificá-la como uma técnica de estimação não-paramétrica, não-linear, multivariada e múltipla, ou seja, é uma função dos dados, sem um modelo de descrição definido. Ser multivariada indica que a entrada da RNA contém duas ou mais variáveis para prever o comportamento futuro de uma ou mais variáveis intertemporais. Não-paramétrica e sem modelo específico são conseqüências de algumas suposições referentes às relações entre as variáveis de entrada e as extrapolações dentro de um horizonte de tempo futuro (LIMA, 2004).

Gao (1996) propôs uma arquitetura de RNA de Elman modificada, onde foram adicionadas conexões entre os resultados da camada de contexto para a camada de saída, e um novo algoritmo para o seu treinamento. Os testes realizados por Gao (1996) verificaram uma melhora no tempo de convergência da RNA de Elman para a identificação de sistemas lineares e não-lineares. Posteriormente, Cheng (2002) provou que a RNA de Elman modificada, possui melhores propriedades dinâmicas quando comparadas com a RNA de Elman, conforme proposta original de Elman (1990).

Pham (1998) verificou que as RNAs de Elman e Jordan geram melhores resultados, na identificação de sistemas lineares e não-lineares, quando treinadas com algoritmos genéticos, quando comparadas com o treinamento clássico com algoritmo de retropropagação do erro. Ge (2007) conseguiu melhor convergência de resultados em relação ao controle e identificação de parâmetros de um motor ultrasônico, com uma RNA de Elman otimizada por um algoritmo de enxame de partículas modificado, em relação a uma RNA de Elman treinada por um algoritmo baseado em algoritmo descida de encosta.

Ilonen (2003) obteve desempenho similar, em relação aos EQMs de previsões com séries temporais de colesterol, câncer e da função seno, em RNAs *feedforward*, treinadas com algoritmo de descida de encosta e evolução diferencial. A verificação demonstrou não haver vantagem na taxa de aprendizado, nem na qualidade da solução, na utilização do algoritmo de

evolução diferencial e recomendaram o uso desta técnica para validação de um resultado alcançado, desenvolvimento de termos de regularização e para o caso de funções de transferência não-convencionais que não possam fornecer informações de gradiente.

Segundo Santos (2005), partindo-se dos resultados de uma série de testes estatísticos que identificam não-linearidades em séries temporais, as séries financeiras possuem uma presença significativa de não-linearidades. Com base nessa constatação é possível afirmar que modelos lineares não são capazes de reproduzir o comportamento das séries de forma apropriada. O autor verificou que os modelos não-lineares com base em RNAs *feedforward*, RNAs-FBR e sistemas nebulosos fornecem melhores previsões quando comparados com modelos lineares (ARIMA-GARCH, AR, MA e ARMA), no curto prazo, sendo que as RNAs *feedforward* obtiveram os melhores resultados na previsão da taxa de câmbio real/dólar.

O'Connor (2005) obteve um retorno sobre o investimento de 23,5% em um período que o DJIA (*Dow Jones Industrial Average*) cresceu 13,03%, prevendo os movimentos do DJIA a partir de uma RNA treinada com informações de preço de *commodities* e taxas de troca de moeda, além dos valores diários de abertura e fechamento do DJIA.

Cheng (2006) testou uma RNA função de base radial, uma função logística na camada escondida e modificou o algoritmo de treinamento para aproveitar a capacidade de predição das RNAs e os resultados em forma de probabilidades gerados a partir da análise logística, que, por consequência, são mais explicativos se comparados às respostas das RNAs de função de base radial. A combinação de RNA e modelagem logística gerou melhores resultados se comparadas com as previsões de cada um dos métodos separados.

Andreeski (2006) comparou o ajustamento das previsões de preços de ações da IBM, conforme a série B apresentada no livro *Time Series Analysis: Forecasting and Control* de Box e Jenkins, gerados por modelagem ARIMA e RNAs, e verificou que mesmo obtendo um ajustamento superior como os modelos de RNAs, a suas previsões freqüentemente apresentam resultados não satisfatórios em termos do coeficiente de explicação  $R^2$ .

Hassan (2006) obteve resultados comparáveis aos modelos ARIMA mesmo que não fazendo tratamentos de sazonalidade, ciclicidade ou regime sobre a série, por meio da



combinação do modelo escondido de Markov, RNAs e algoritmos genéticos em um único modelo, para previsão do preço de ações.

Chang (2006) conseguiu melhores previsões de séries temporais, medidos pelo teste estatístico Ljung-Box *Q-test*, com um sistema neuro-nebuloso adaptativo de inferência, utilizando um modelo GARCH não-linear para minimizar o efeito da variância condicional dos resíduos da previsão.

Brunelli (2006) obteve melhores resultados na previsão da concentração de dióxido de enxofre, em construções próximas a áreas industriais, projetando uma RNA de Elman e comparando suas previsões com outras obtidas com uma RNA *feedforward* e um modelo estocástico.

Roh (2007) obteve melhores previsões da volatilidade do índice KOSPI 200 (*Korea Composite Price Index*), utilizando um método híbrido baseado em RNAs e variação do modelo GARCH, quando comparado com métodos de RNAs sem nenhuma hibridização.

Santos (2007) por sua vez, obteve melhor desempenho de predição utilizando modelos não-lineares, rede neural artificial (RNA) *feedforward*, RNA de função de base radial e sistemas nebulosos Takagi-Sugeno, quando comparados com modelos lineares ARMA, ARMA-GARCH, mas, mencionou a questão da consistência dos resultados que é afetada pela frequência dos dados e os efeitos das não-linearidades das séries de cotação de moedas.

Du (2006), Babu (2007) e Basturk (2008) constataram bons resultados com o algoritmo ED para o treinamento de diversos tipos de RNAs aplicadas em problemas de reconhecimento de contornos de imagens e controle no aquecimento de tubulações.

Assim, em termos teóricos, essa pesquisa procura utilizar a literatura pertinente sobre o assunto abordado, como um arcabouço para a referida pesquisa, visando reduzir o grau de incerteza dos ativos financeiros pesquisados, a partir da análise dos modelos de previsão que possibilitem compreender a maneira pela qual uma série temporal univariada seja prevista, de forma a auxiliar o investidor sobre a alocação de recursos escassos necessários para o planejamento de uma organização.

#### 1.4.2 Justificativa prática

A intenção em estender a pesquisa sobre previsão de séries temporais financeiras, direcionada ao IBOVESPA é, antes de qualquer coisa, fornecer aos analistas de investimentos ferramentas consistentes para a realização de seus planejamentos de estratégias de investimentos.

Quando procura informações sobre como selecionar ações para investimento, o investidor ou mesmo o especulador encontra nas análises, fundamentalista e técnica, estes instrumentos que direcionam a tomada de decisão.

A partir da avaliação de tópicos que se relacionam com o desenvolvimento da empresa, a análise fundamentalista possibilita previsões sobre as perspectivas da mesma. Essa análise utiliza fundamentos econômico-financeiros para a determinação do valor da empresa. Pode-se conceituar a análise fundamentalista como o estudo de toda a informação disponível no mercado sobre determinada empresa, com a finalidade de obter seu verdadeiro valor, e assim formular uma recomendação de investimento. O analista resume e analisa a informação, parte do passado e trata de prever o futuro, para dar sua opinião.

Segundo Pinheiro (2007), o objetivo principal dessa análise é avaliar o comportamento da empresa visando à determinação do seu valor. A análise fundamentalista parte do princípio de que as ações têm valor intrínseco, que está associado com o desempenho da empresa emissora e com a situação geral da economia. Portanto, estuda os fatores que explicam o valor intrínseco de uma empresa, setor ou mercado, colocando em segundo plano os fatores de mercado, tais como preço e volume. Esses fatores são denominados valores fundamentalistas.

A justificativa para o uso desse tipo de análise é antecipar o comportamento futuro de um determinado papel que reflete o valor da empresa no mercado. Isto é, adiantar-se ao mercado. Para que isso seja possível, têm-se que partir de uma hipótese básica: o mercado não é eficiente no curto prazo, ainda que seja no longo prazo. Se não fosse assim, não seria possível adiantar-se ao mercado. Hoje, o preço de uma ação não reflete o verdadeiro valor da empresa, mas, existe uma tendência de que isso ocorra em um futuro próximo. O analista fundamentalista trata de descobrir supervalorizações ou subvalorizações, com base em determinada informação ainda não negociada pelo mercado.

O processo decisório dessa escola de análise de investimentos envolve o cálculo do valor hipotético da empresa, que corresponderia ao seu “preço justo” em determinado momento, e, pela comparação desse com seu preço de mercado, seleciona para investimento aquela com menor relação valor intrínseco/preço de mercado.

Para estimar o valor de uma ação em determinado instante, a análise fundamentalista baseia-se em três aspectos:

i) as demonstrações financeiras passadas, bem como outras informações relevantes, que permitam a projeção mais provável possível da evolução futura da empresa emissora das ações;

ii) a situação atual da economia e estimativas futuras de variáveis com indiscutível influência sobre o preço e o valor de todas as ações negociadas; e

iii) o grau de confiança do investidor em relação ao grau de intervenção governamental.

Além de se basear nestes aspectos, a determinação do valor da empresa pode ser definida também como função de alguns fatores, como, por exemplo:

- lucro esperado em exercícios futuros;
- investimentos realizados e a realizar; e
- fontes de financiamentos utilizadas.

É difícil encontrar no mercado analistas de investimentos que se utilizem exclusivamente dos pilares da escola de análise fundamentalista. Em geral, a análise fundamentalista é utilizada simultaneamente com a análise técnica, que tem como principal objetivo a determinação do momento mais favorável para o investimento. O importante é saber conjugar os fatores de ambas as análises, visto que, em geral, ambas se apóiam em avaliações subjetivas.

Pode-se conceituar a análise técnica como um estudo dos movimentos passados dos preços e dos volumes de negociação de ativos financeiros, com o objetivo de fazer previsões sobre o comportamento futuro dos preços.

A escola técnica ou gráfica baseia-se na tese de que os preços das negociações futuras são fortemente dependentes dos preços das negociações anteriores, sendo possível, então, prever tendências de preços valendo-se da observação de movimentos passados.

Segundo essa análise, as variações dos preços de mercado não são independentes de variações passadas, o mesmo ocorrendo com as taxas de retorno sobre o investimento em qualquer ativo. Segundo a escola técnica, as séries de preços de mercado apresentam padrões identificáveis, e cabe ao analista desenvolver a arte de identificar tendências e saber interpretá-las.

Para essa análise não são relevantes as informações sobre lucros, dividendos, participações no mercado, grau de endividamento ou liquidez da empresa, como são consideradas pela análise fundamentalista. O que importa são os fatores de procura e oferta internos ao mercado, sendo crucial entender a “psicologia” do mercado. Isso porque o mercado é “arbitrado”, ou seja, o impacto dos fatores externos já está embutido nos preços. Portanto, essa análise é o estudo de como os preços movimentam-se, não se preocupando com o motivo de suas movimentações.

A análise mencionada parte de três princípios:

- 1) a ação do mercado reflete todos os fatores envolvidos neste;
- 2) os preços se movimentam em tendências; e
- 3) o futuro reflete o passado.

Assim, o método de previsão proposto neste trabalho pode ser utilizado em conjunto com outros indicadores da análise técnica para a previsão dos próximos movimentos do mercado, confirmando ou não suas indicações e apoiando, como um todo, o processo de tomada de decisão quanto a investimentos.

## 1.5 ESTRUTURA DO TRABALHO

No primeiro capítulo, faz-se uma revisão sobre séries temporais e técnicas de previsão, uma revisão bibliográfica para cada técnica abordada e, por fim, uma justificativa para a utilização de RNAs. No segundo capítulo, é apresentada a fundamentação teórico-empírica focando na hipótese de mercados eficientes, alisamento exponencial por Holt-Winters, redes neurais artificiais, evolução diferencial e modelos da família ARCH. No terceiro capítulo, é realizada uma descrição das variáveis e forma de coleta de dados. No quarto capítulo, são propostas RNAs de Elman treinadas ou otimizadas de forma híbrida com algoritmos evolucionários e algoritmos de descida de encosta. Para essas RNAs de Elman, o algoritmo de evolução diferencial e o método quase-Newton do tipo BFGS serão os algoritmos utilizados nas buscas, globais e locais, combinados e separados, com o objetivo de comparar os coeficientes U-Theil dessas técnicas com o obtido utilizando um modelo GARCH.

## 2. FUNDAMENTAÇÃO TEÓRICO-EMPÍRICA

Este capítulo trata da fundamentação teórico-empírica elaborada com o objetivo de dar suporte aos resultados obtidos a partir dos dados coletados, sendo sumarizado como segue: a seção 2.1 refere-se à eficiência de mercado. A seção 2.2 aborda os fundamentos do modelo de Holt-Winters. A seção 2.3 trata das redes neurais combinadas a algoritmos evolutivos. A seção 2.4 mostra métodos de otimização baseados em gradiente; enquanto a seção 2.5 apresenta a evolução diferencial. Finalizando o capítulo, a seção 2.6 trata dos modelos da família ARCH.

### 2.1 EFICIÊNCIA DE MERCADO

Segundo Fama (1970), quando se fala em mercados eficientes pressupõe-se que os preços dos títulos refletem integralmente todas as informações disponíveis. Esta hipótese é forte uma vez que a condição necessária para que os investidores tenham um incentivo para negociar até que os preços reflitam integralmente todas as informações é a de que o custo da aquisição de informação e o custo de negociação sejam iguais à zero. Observando-se que se os custos são positivos, uma definição mais realista é a de que os preços refletem informação até que os custos marginais de obtenção de informação e negociação não superem o benefício marginal.

A hipótese de mercado eficiente tem sido subdividida em três categorias, cada uma delas relacionada a um tipo distinto de informação. Os testes da forma fraca são usados para verificação se a informação contida em preços passados está refletida nos preços correntes. Os testes da forma semiforte da hipótese de mercado eficiente procuram verificar se as informações publicamente disponíveis estão integralmente refletidas nos preços correntes dos títulos. Finalmente, os testes utilizados para avaliar a forma forte da hipótese de mercado eficiente visam verificar se toda informação, publicamente disponível ou não, está integralmente refletida nos preços dos títulos, e se qualquer tipo de investidor consegue obter lucros extraordinários.

Fama (1991) abordou a eficiência do mercado novamente com o propósito de considerar a evolução dos estudos empíricos na área e o conseqüente aprendizado sobre o tema, propondo uma reformulação e mesmo ampliação da abrangência das três categorias de testes de eficiência estabelecidos.

Ao invés de testar a forma fraca, relacionados como poder de previsão dos retornos passados, Fama propõe a nova denominação de testes de previsibilidade dos retornos. De forma mais abrangente, essa categoria de testes de eficiência inclui também os dividendos e os juros como variáveis de previsão de retornos. Como em testes de eficiência, avalia-se a hipótese conjunta de eficiência do mercado e a robustez do modelo de determinação do preço de equilíbrio, utilizado no teste. Fama (1991) inclui também, nessa categoria, as anomalias detectadas, em testes do tipo *cross-section*, tais como o efeito tamanho e o efeito janeiro.

Recentemente, Tóth (2006) verificou um crescimento da eficiência de mercado quando analisou os retornos da *NYSE (New York Stock Exchange)*. Em se tratando de eficiência de mercado em países emergentes, Brizio (2007) verificou aumento de eficiência no mercado de ações mexicano.

A hipótese de mercado eficiente se preocupa com as condições sob as quais um investidor pode obter retornos extraordinários com um determinado título. Considerando-se um cenário onde, após um anúncio, o preço suba gradativamente durante a semana, em resposta ao anúncio, os investidores que estivessem examinando a seqüência de preços observariam que o preço estaria afastando-se do nível ao qual a ação havia sido negociada anteriormente.

Cajueiro (2008) verificou uma dependência de volatilidade de longo prazo em 41 índices de ações no mercado mundial, mesmo não encontrando evidência de dependência nos retornos dos mesmos. Os testes de previsibilidade de retornos visam verificar se esse tipo de tática de negociação pode conduzir a retornos extraordinários.

Em um segundo cenário, onde um investidor ouve um anúncio de mudança de perspectiva e compra imediatamente ações de uma empresa na expectativa da alta dos preços, pode ser avaliado por um teste da forma semiforte da hipótese de mercado eficiente. Estes testes possibilitam que a estratégia leve a retornos extraordinários. Na forma forte de

eficiência, considerando o cenário analisado anteriormente, o investidor pode avaliar precisamente o impacto do anúncio sobre o valor da empresa e decide comprar ou vender seus títulos.

A forma forte da eficiência de mercado diz que não há qualquer investidor com esta capacidade superior. Como é impossível determinar exatamente como os investidores poderiam usar o anúncio para reavaliar a empresa, os testes da forma forte são exames da capacidade de obtenção de retornos extraordinários por um investidor ou grupo de investidores. Por falta de dados a respeito dos tipos de investidores, o grupo mais frequentemente testado é formado por administradores de fundos de investimento.

#### 2.1.1 Testes de previsibilidade

Em relação aos testes de previsibilidade de retornos, alguns estudos têm relatado padrões temporais onde os retornos são sistematicamente mais altos ou mais baixos em função da hora-do-dia, do dia-da-semana e do mês do ano. Um padrão que tem sido examinado é a diferença de retornos entre os vários dias da semana. Segundo Costa (2000), os resultados obtidos por um estudo sobre o efeito dia-da-semana, na América Latina, permitem concluir que, no Peru e na Venezuela, existe o efeito dia-da-semana, que torna baixa ou negativa a variação nos índices de preços na segunda-feira e positiva, na sexta-feira, no período analisado. Essa conclusão é sustentada pelos testes paramétrico e não-paramétrico. Especificamente no Brasil, não se confirmam os resultados obtidos em estudo anterior, que apresentava clara distinção entre os dias perto do início e do fim de semana. Outro exemplo de padrão temporal é a pesquisa de padrões mensais que constata que os retornos em janeiro são significativamente maiores que os retornos em outros meses, principalmente entre as empresas de pequeno porte.

#### 2.1.2 Testes da forma semiforte

Para testar a forma semiforte da eficiência de mercado, a metodologia de estudos de eventos é padronizada e funciona da seguinte maneira:



- Coleta-se uma amostra de empresas para as quais houve um anúncio inesperado (o evento);
- Determina-se o dia exato do anúncio e chama-se esse dia de data zero;
- Defini-se  $n$  períodos a serem estudados;
- Para cada uma das empresas da amostra, calcula-se o retorno em cada um dos dias estudados;
- Calcula-se o retorno “anormal” para cada um dos dias estudados para cada empresa da amostra;
- Calcula-se para cada dia do período do evento, o retorno anormal médio envolvendo todas as empresas da amostra;
- Soma o retorno anormal de cada dia para calcular o retorno anormal acumulado desde o início do período;
- Examinar e discutir os resultados.

### 2.1.3 Testes da forma forte

Nos testes da forma forte, todos os investidores que detém uma porcentagem das ações de uma empresa acima de um nível pré-determinado, ou estão em uma posição executiva privilegiada, são consideradas *insiders*. Nos Estados Unidos da América, os *insiders* devem informar suas compras e vendas à *SEC (U.S. Securities and Exchange Commission)*.

Neste trabalho, assume-se que o mercado não seja eficiente em sua forma fraca e que pressupõe ser possível prever retornos futuros utilizando-se do histórico de retornos passados de um determinado ativo.

## 2.2 MODELO DE HOLT-WINTERS

Segundo Cavalheiro (2003), o método de Holt-Winters é enquadrado na categoria de alisamento exponencial e utilizado para a previsão de séries temporais que possuem características de tendência e sazonalidade, tanto aditiva quanto multiplicativa. Este método costuma ser eficaz nas previsões de curto prazo (CORTEZ, 1997).

O alisamento exponencial é um procedimento que revisa continuamente a previsão com base na observação mais recente. O alisamento exponencial atribui pesos exponenciais decrescentes à medida que as observações da série temporal são atrasadas no tempo. Em outras palavras, as observações mais recentes são mais significativas para a previsão do que as observações passadas.

As séries temporais, freqüentemente, apresentam comportamento sazonal. A sazonalidade é definida como a tendência dos dados das séries temporais em exibir um comportamento que se repete a cada  $L$  períodos. O termo sazonal refere-se ao período de tempo anterior à repetição do comportamento e  $L$  é o tamanho sazonal em períodos.

No modelo multiplicativo, a amplitude do ciclo sazonal varia em função do tempo, ou seja, a diferença entre o maior e o menor valor da série temporal aumenta ou diminui dependendo das suas características ao longo do tempo. No modelo aditivo, verifica-se que a diferença é constante.

Quando observadas em gráficos, a característica que diferencia esses dois tipos de sazonalidade é que, no caso aditivo, as séries mostram flutuações sazonais estacionárias, independente do nível da série; e no caso multiplicativo, os tamanhos das flutuações sazonais variam conforme o nível da série.

Segundo Makridakis (1998), as expressões para o modelo multiplicativo são:

$$L_t = \alpha \frac{Z_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} - b_{t-1}), \quad (2.1)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \quad (2.2)$$

$$S_t = \gamma \frac{Z_t}{L_t} + (1 - \gamma)S_{t-s}, \quad (2.3)$$

$$F_{t+m} = (L_t - b_t m)S_{t-s+m}, \quad (2.4)$$

e para o modelo aditivo, tem-se:

$$L_t = \alpha(Z_t - S_{t-s}) + (1 - \alpha)(L_{t-1} - b_{t-1}), \quad (2.5)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}, \quad (2.6)$$

$$S_t = \gamma(Z_t - L_t) + (1 - \gamma)S_{t-s}, \quad (2.7)$$

$$F_{t+m} = (L_t + b_t m) + S_{t-s+m}, \quad (2.8)$$

onde  $L_t$  representa a estimativa de nível da série média no momento  $t$ ,  $b_t$  representa a estimativa da inclinação da série no momento  $t$ ,  $S_t$  a estimativa de sazonalidade da série no momento  $t$ ,  $F_{m+t}$  representa a previsão para  $m$  períodos futuros e  $\alpha$ ,  $\beta$ ,  $\gamma$  são as constantes de suavização a serem calculadas, e  $Z_t$  é o valor atual observado na série temporal.

Como o objetivo é encontrar um conjunto de parâmetros que minimize os erros entre a previsão estimada e a observação real, a utilização de um método numérico de otimização pode ser empregado.

### 2.3 REDES NEURAIAS ARTIFICIAIS EVOLUTIVAS

As RNAs evolutivas (RNAEs) referem-se a uma classe especial de RNAs na qual a evolução é uma forma de adaptação em conjunto com o aprendizado. Os algoritmos evolutivos (AE) são usados para executar várias tarefas, tais como treinamento de pesos de conexão, projeto de arquitetura, adaptação de regras de aprendizado, seleção de características de entrada, geração dos pesos iniciais das conexões, extração de regra, entre outros. Uma característica das RNAEs é sua capacidade de adaptação a ambientes dinâmicos. Em outras palavras, as RNAEs podem adaptar-se, efetiva e eficientemente, através da evolução e do aprendizado, a um ambiente bem como às suas mudanças. As RNAEs podem ser consideradas sistemas capazes de trocar suas arquiteturas e aprender regras sem intervenção do projetista.

### 2.3.1 Fundamentos das redes neurais artificiais

Uma RNA consiste de um conjunto de elementos de processamento, também conhecidos como neurônios ou nós, os quais são interconectados. A RNA pode ser descrita como um grafo no qual cada nó  $i$  executa uma função de transferência  $f_i$  na forma de

$$y_i = f_i \left( \sum_{j=1}^n w_{ij} x_j - \theta_i \right), \quad (2.9)$$

onde  $y_i$  é a saída do nó  $i$ ,  $x_j$  é a  $j$ -ésima entrada para o nó, e  $w_{i,j}$  é o peso da conexão entre os nós  $i$  e  $j$ .  $\theta_i$  é o *bias* do nó. Usualmente,  $f_i$  é uma função não-linear, tais como função degrau, sigmóide ou Gaussiana.

As RNAs podem ser separadas em duas classes, *feedforward* e recorrente tal como encontra-se na Figura 2.1, de acordo com sua conectividade. Uma RNA é *feedforward* se as saídas dos neurônios de uma determinada camada forem propagadas para uma camada posterior. Um RNA é recorrente se as saídas dos neurônios em uma determinada camada, realimentarem uma camada anterior.

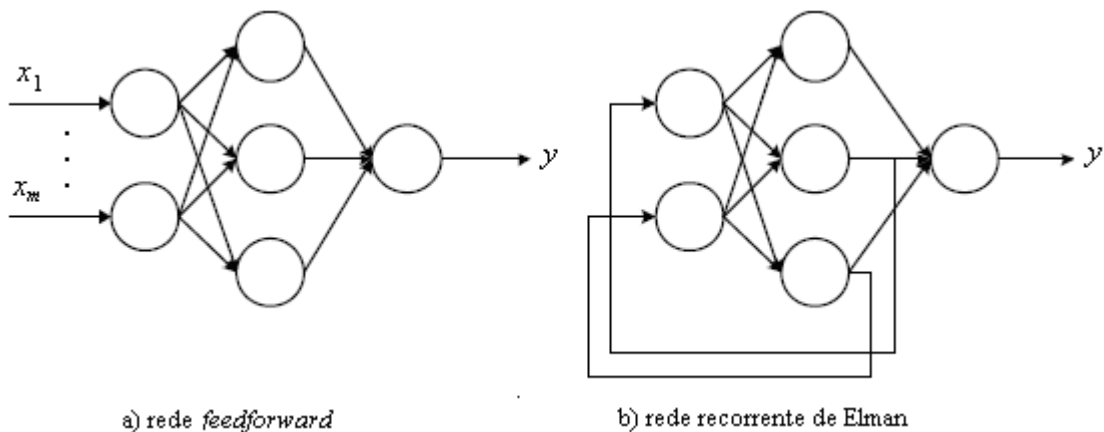


Figura 2.1 a) rede (*feedforward*) acíclica b) rede (recorrente de Elman) cíclica.

Na equação (2.9), cada termo da somatória está relacionado com somente uma entrada  $x_j$ . RNAs de ordem mais altas são aquelas que contém nós nas quais mais de uma entrada são envolvidos na soma dos termos de uma somatória. Por exemplo, um nó de segunda ordem pode ser descrito como,

$$y_i = f_i \left( \sum_{j=1}^n w_{ijk} x_j x_k - \theta_i \right), \quad (2.10)$$

onde todos os símbolos tem definições similares àqueles da equação (2.9).

A arquitetura de uma RNA é determinada por suas estrutura topológica, isto é, a conectividade em geral e a função de transferência de cada nó na rede.

O aprendizado das RNAs é tipicamente alcançado a partir da apresentação de exemplos. Esse processo é chamado de treinamento porque o aprendizado é alcançado pelo ajuste dos pesos das conexões da RNA que, uma vez treinadas, podem executar certas tarefas. O treinamento das RNAs pode ser dividido em supervisionado, não-supervisionado e aprendizado por reforço. O aprendizado supervisionado é baseado na comparação direta entre a saída atual da rede e a saída desejada, também conhecida como saída alvo. Esse tipo de aprendizado é frequentemente formulado como a minimização de uma função de erro tal como o erro quadrático médio total entre a saída da rede e a saída desejada, somados com base em todos os dados disponíveis. Uma otimização baseada em um algoritmo de descida de encosta, tal como o *backpropagation*, pode ser usada para ajustar iterativamente os pesos das conexões de uma RNA a fim de minimizar o erro. O aprendizado por reforço é um caso particular de aprendizado supervisionado onde a saída desejada não é conhecida. Nesse caso, a saída atual é pode ser simplesmente classificada como correta ou não correta. O aprendizado não-supervisionado é baseado nas correlações dos dados de entrada e nenhuma informação correta sobre a saída está disponível para o aprendizado.

A essência de um algoritmo de aprendizado é a regra de atualização dos pesos a qual determina como as conexões são modificadas. Exemplos de regras incluem a regra delta, regra de aprendizado Hebbiana e a regra de aprendizado competitivo (YAO, 1999).

### 2.3.2 Redes neurais parcialmente recorrentes

As redes neurais de múltiplas camadas *feedforward* geralmente utilizam uma janela de tempo como artifício para o processamento temporal, o que consiste em ter uma entrada composta por um conjunto de dados representando um intervalo temporal como se eles formassem um padrão estático. Esta solução não é satisfatória para o processamento temporal e com base nesta deficiência é necessário estender a estrutura das redes neurais multicamadas de forma que seja possível oferecer um comportamento dinâmico à rede.

Segundo Soto (1999), as redes neurais recorrentes possuem uma ou mais conexões de realimentação que, por sua vez, podem ser classificadas como locais ou globais. A realimentação local acontece em nível de um único neurônio e a realimentação global envolve alguma(s) camada(s) completa(s). Outro critério de classificação refere-se à direção da comunicação entre os neurônios e assim rede recorrente pode ser enquadrada como totalmente ou parcialmente recorrente.

As redes totalmente recorrentes, também conhecidas como simétricas, são caracterizadas por terem seus neurônios se comunicando em ambos os sentidos. As redes parcialmente recorrentes surgiram a partir de estudos que utilizam ligações recorrentes em apenas alguns neurônios pela necessidade de resolver problemas de instabilidade e complexidade de treinamento das redes totalmente recorrentes (SHULLI, 2005).

Segundo Cheng (2002), a RNA de Elman é uma rede neural recorrente que pode ser enquadrada entre uma RNA *feedforward* e uma rede recorrente pura. Pela existência de nós de contexto e de conexões locais recorrentes entre a camada de contexto e a camada escondida, tem vantagens dinâmicas quando comparadas às RNAs estáticas, tais como a RNA *feedforward* e a RNA-FBR.

A estrutura básica de uma RNA de Elman é apresentada na Figura 2.2. A RNA de Elman possui uma estrutura de quatro camadas: camada de entrada, camada escondida, camada de contexto e camada de saída. Existem pesos de conexões ajustáveis para cada duas camadas vizinhas. A RNA de Elman é geralmente considerada um tipo especial de RNA *feedforward* com neurônios adicionais de memória com realimentação local. A realimentação

das conexões da camada de contexto para a camada de entrada faz a RNA de Elman ser sensível às entradas históricas o que é importante na modelagem de sistemas dinâmicos.

A notação usada utilizada nesta seção é dada a seguir:

$w1_{i,j}$ : O peso que conecta o nó  $i$  da camada de entrada com o nó  $j$  na camada escondida;

$w2_{i,j}$ : O peso que conecta o nó  $i$  da camada escondida com o nó  $j$  na camada de saída;

$w3_{i,j}$ : O peso que conecta o nó  $i$  da camada de contexto com o nó  $j$  na camada escondida;

$m, n, r$ : O número de nós na camada de entrada, saída e camadas escondidas respectivamente;

$u_{i,(k)}, y_{j,(k)}$ : Entradas e saídas da RNA de Elman, onde  $i=1,2,\dots,m$ , e  $j=1,2,\dots,n$ ;

$x_{i,(k)}$ : Saída do nó escondido  $i$ , onde  $i=1,2,\dots,r$ ;

$c_{i,(k)}$ : Saída do nó de contexto  $i$ , isto é, a saída do nó escondido  $i$  da última iteração;

$Z^{-1}$ : Uma unidade de atraso.

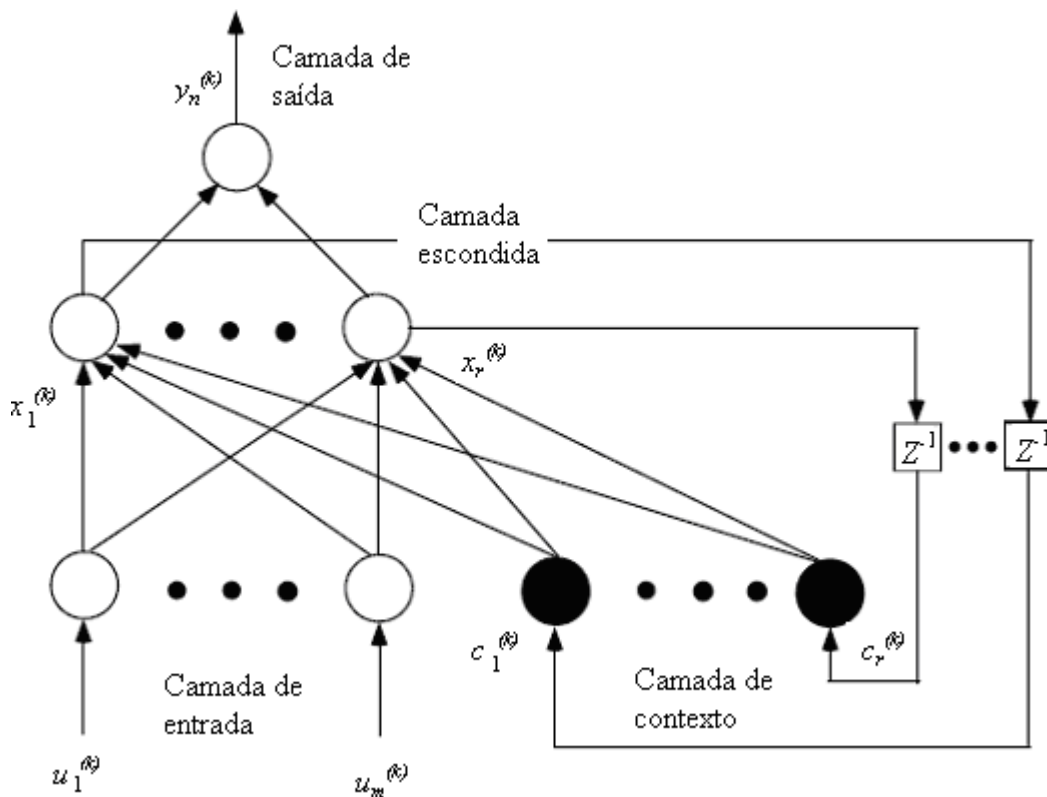


Figura 2.2. Rede parcialmente recorrente de Elman.

Para cada unidade na camada escondida, uma unidade denominada unidade de contexto é adicionada à RNA de Elman. A unidade de contexto é totalmente conectada com todas as unidades escondidas de maneira que existe um peso de toda unidade de contexto para toda unidade na camada escondida. Existem também conexões recorrentes das unidades escondidas de volta para as unidades de contexto. Mas, cada unidade de contexto está conectada somente à sua unidade de contexto associada, conforme apresentado na Figura 2.3.

Os pesos das conexões recorrentes são fixos e os pesos das conexões para as camadas posteriores são treinados com o algoritmo *backpropagation*. Na fase *forward*, as unidades de contexto comportam-se como unidades de entrada. Os valores das unidades escondidas e das unidades de saída são calculados da mesma forma que as RNAs *feedforward*. Após calcular as saídas das unidades escondidas, os valores obtidos são copiados para as unidades de contexto correspondentes através das conexões recorrentes (unidade de atraso). Na primeira iteração do treinamento, os valores das unidades recorrentes devem ser iniciados. Durante a fase de retropropagação, os valores dos pesos são ajustados pelo algoritmo de retropropagação do erro.

As entradas da rede são:  $u_{i(k)}$ ,  $y_{j(k)}$ ,  $x_{i(k)}$  e as saídas de camada podem ser dadas por

$$x_j(k) = f\left(\sum_{i=1}^m w_{2,i,j} u_i(k) + \sum_{i=1}^r w_{1,i,j} c_i(k)\right), \quad (2.11)$$

$$x_i(k) = x_i(k-1), \quad (2.12)$$

$$y_j(k) = g\left(\sum_{i=1}^r w_{3,i,j} x_i(k)\right), \quad (2.13)$$

onde  $f(\cdot)$  e  $g(\cdot)$  são funções com saídas lineares ou não-lineares da camada escondida e da camada de saída, respectivamente.

Por causa dessa característica dinâmica da RNA de Elman provida por meio das conexões internas, ela não precisa representar o estado como entrada ou sinal de treinamento. Está é a vantagem da RNA de Elman em relação às RNA *feedforward* estáticas.



### 2.3.3 Evolução dos pesos das conexões

Segundo Yao (1999), a evolução foi introduzida nas RNAs em três níveis diferentes: pesos de conexão, arquiteturas, e regras de aprendizado. A evolução dos pesos das conexões introduz uma abordagem de treinamento global e adaptativa, especialmente no aprendizado por reforço e no aprendizado recorrente da rede, onde o aprendizado baseado em descida de encosta frequentemente encontra dificuldades. A evolução das arquiteturas habilita a RNA a modificar sua estrutura, para realizar tarefas diferentes e assim provê uma abordagem automática para a evolução do projeto e adaptação de pesos de conexões da RNA.

O treinamento dos pesos de uma RNA é formulado, usualmente, como a minimização de uma função de erro, por exemplo, o REQM entre a saída desejada e a saída atual da rede e ajustando iterativamente as conexões dos pesos. Muitos algoritmos de treinamento, tal como o *error backpropagation* e o algoritmo de gradiente conjugado, são baseados na descida de encosta. Esses algoritmos frequentemente ficam presos em um mínimo local e são incapazes de encontrar um mínimo global se a função de erro não for diferenciável e/ou unimodal.

Uma forma de contornar os problemas dos algoritmos baseados em gradiente é a adoção de uma RNAE, por exemplo, formular um processo de treinamento como a evolução dos pesos das conexões em um ambiente determinado pela arquitetura e tarefa de aprendizado. Os AEs podem ser utilizados efetivamente na evolução global, para encontrar um conjunto de pesos das conexões próximos ao ótimo, e sem cálculo do gradiente. O valor do erro, objetivo de uma RNA, pode ser definido com base nas necessidades específicas da tarefa a ser executada. Dois fatores frequentemente considerados na formulação da função de erro são: o erro entre a saída deseje e a saída atual, e a complexidade da RNA.

A abordagem evolutiva de treinamento dos pesos de uma RNA consiste na escolha da representação das conexões dos pesos e no processo de evolução juntamente com seu esquema de representação e operadores de mutação e cruzamento. Um ciclo típico da evolução das conexões dos pesos é mostrado na Figura 2.3, onde a evolução termina quando o erro atinge um valor maior ou igual a um valor predefinido ou a população tenha convergido (YAO, 1999).

1. Decodifica cada indivíduo (genótipo) da geração atual em um conjunto de pesos de conexões e construa uma RNA com os pesos.
2. Avalia a RNA computando o REQM total entre as saídas desejadas e obtidas. O *fitness* de um indivíduo pode ser determinado pelo erro. Quando mais alto o erro, menor o *fitness*. O mapeamento ótimo do erro é dependente do problema. Um termo de regularização pode ser incluído na função de *fitness* para penalizar pesos grandes.
3. Seleciona os pais para reprodução baseados em seus valores de *fitness*.
4. Aplica os operadores de busca (cruzamento e mutação) aos pais, para gerar os descendentes, para compor a próxima geração.

Figura 2.3. Ciclo típico da evolução das conexões dos pesos.

Segundo Yao (1999), embora os AEs sejam bons nas buscas globais, eles são ineficientes para uma busca local. A eficiência do treinamento evolucionário pode ser significativamente melhorada pela incorporação de um procedimento de busca local, por exemplo, uma combinação da habilidade de busca global dos AEs para encontrar uma boa região e uma busca local por algoritmo da retropropagação do erro.

Lee (1996) usou algoritmos genéticos (AG) para uma busca de um conjunto de pesos de conexões próximo do resultado desejado e usou *backpropagation* para a busca local a partir dos pesos iniciais. Os resultados mostraram o treinamento híbrido como mais eficiente quando comparados com os resultados dos treinamentos realizados com o *backpropagation* ou AG sozinhos. Se for considerado que o *backpropagation* tem que ser executado várias vezes, até encontrar uma solução satisfatória, porque é sensível aos pesos iniciais, o treinamento híbrido torna-se competitivo.

## 2.4 MÉTODOS DE OTIMIZAÇÃO BASEADOS EM GRADIENTE

### 2.4.1 Algoritmo de retropropagação do erro

A rede *feedforward*, também conhecida como rede de retropropagação ou *perceptron* de múltiplas camadas, é uma classe importante de redes neurais por sua topologia simples e capacidade de aproximação. É de conhecimento que o *perceptron* só pode resolver problemas

separáveis linearmente ou linearmente independentes (HAYKIN, 2001). Problemas não linearmente separáveis requerem que as redes neurais tenham uma representação intermediária apropriada dos padrões de entrada pela introdução de camadas escondidas não-lineares. A construção de uma rede *feedforward* é inspirada nessa idéia. Uma rede *feedforward* de uma única camada escondida é apresentada na Figura 2.4.

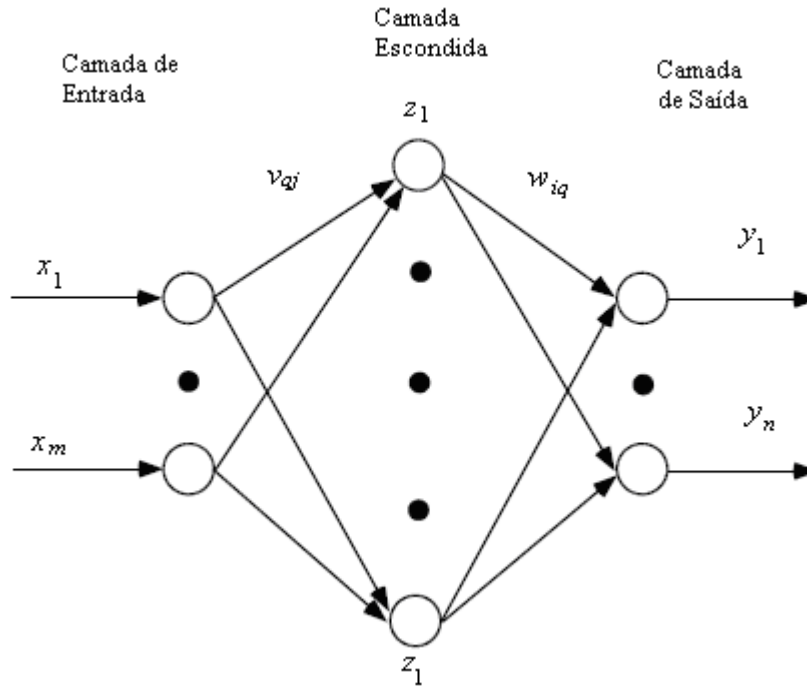


Figura 2.4. Rede neural *feedforward* com uma camada escondida.

Segundo Haykin (2001), a invenção do algoritmo de aprendizado por retropropagação para redes neurais *feedforward* é um marco no desenvolvimento histórico das redes neurais. Somente depois da introdução do algoritmo de aprendizado, as propriedades das redes neurais foram reconhecidas, e as pesquisas sobre redes neurais conseguiram atenção de pesquisadores de diferentes áreas de conhecimento. Sem perda de generalidade, o algoritmo de retropropagação é deduzido para uma rede *feedforward* com uma única camada escondida.

Na Figura 2.4,  $v_{qj}$  é o peso conectando o nó  $j$  na camada de entrada com o nó  $q$  na camada escondida, e  $w_{iq}$  o peso que conecta o nó  $q$  da camada escondida com o nó  $i$  na camada de saída. Existem  $m$ ,  $l$  e  $n$  nós na camada de entrada, camada escondida, e camada de saída, respectivamente. Assuma que  $(\mathbf{x}, \mathbf{d})$  representa uma par de exemplo de treinamento;  $\mathbf{x} = \{x_1, x_2, x_m\}$  é o padrão de entrada, e  $\mathbf{d} = \{d_1, d_2, d_n\}$  é a saída desejada. O nó  $q$  na camada escondida recebe a entrada e obtém a saída transformada conforme calculado por

$$net_q = \sum_{j=1}^m v_{qj} x_j. \quad (2.14)$$

Neste caso:

$$z_q = a(net_q) = a\left(\sum_{j=1}^m v_{qj} x_j\right), \quad (2.15)$$

onde  $a(\cdot)$  é a função de ativação de um nó na rede neural *feedforward*. O cálculo da entrada do nó  $i$  na camada de saída é dado por:

$$net_i = \sum_{q=1}^l w_{iq} z_q = \sum_{q=1}^l w_{iq} a\left(\sum_{j=1}^m v_{qj} x_j\right). \quad (2.16)$$

Portanto, a saída correspondente é:

$$y_i = a(net_i) = a\left(\sum_{q=1}^l w_{iq} z_q\right) = a\left[\sum_{q=1}^l w_{iq} a\left(\sum_{j=1}^m v_{qj} x_j\right)\right]. \quad (2.17)$$

As equações (2.15) a (2.17) mostram o procedimento de propagação desde a entrada até a saída da rede neural *feedforward*. O algoritmo básico de retropropagação é baseado nessas equações e no seguinte procedimento de retropropagação de erro. A seguir a função de custo  $E(W)$  para a rede *feedforward* minimizar é definida como:

$$E(W) = \frac{1}{2} \sum_{i=1}^n (d_i - y_i)^2 = \frac{1}{2} \sum_{i=1}^n [d_i - a(net_i)]^2 = \frac{1}{2} \sum_{i=1}^n \left[ d_i - a\left(\sum_{q=1}^l w_{iq} z_q\right) \right]^2 \quad (2.18)$$

onde  $W$  representa os pesos das conexões  $v_{qj}$  e  $w_{iq}$ . Os pesos  $w_{iq}$  são atualizados pelo princípio da descida de encosta, tal que:

$$\Delta w_{iq} = -\eta \frac{\partial E}{\partial w_{iq}}, \quad (2.19)$$

onde  $\eta$  é a taxa de aprendizado da rede neural *feedforward*. Aplicando a regra da cadeia para  $\frac{\partial E}{\partial w_{iq}}$ , obtém-se:

$$\Delta w_{iq} = -\eta \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial net_i} \frac{\partial net_i}{\partial w_{iq}} = \eta (d_i - y_i) a'(net_i) z_q = \eta \delta_{oi} z_q, \quad (2.20)$$

onde  $a'(\cdot)$  é a derivada de primeira ordem da função  $a(\cdot)$ . Com base na equação (2.19), a definição do termo geral de retropropagação  $\delta_{oi}$  para a camada de saída é dada por:

$$\delta_{oi} = -\frac{\partial E}{\partial net_i} = -\frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial net_i} = (d_i - y_i) a'(net_i). \quad (2.21)$$

Para atualizar os pesos conectando a camada de entrada com a camada de saída, aplica-se a descida íngreme e novamente a regra da cadeia:

$$\Delta v_{aj} = -\eta \frac{\partial E}{\partial y_{aj}} = -\eta \frac{\partial E}{\partial net_q} \frac{\partial net_q}{\partial v_{aj}} = -\eta \frac{\partial E}{\partial z_q} \frac{\partial z_q}{\partial net_q} \frac{\partial net_q}{\partial v_{aj}}. \quad (2.22)$$

Uma vez que  $z_q$  contribui com cada termo de erro da função custo, gera-se:

$$\Delta v_{aj} = \eta \sum_{i=1}^n [(d_i - y_i) a'(net_i) w_{iq}] a'(net_q) x_j. \quad (2.23)$$

Substituindo (2.20) em (2.22), tem-se:

$$\Delta v_{aj} = \eta \sum_{i=1}^n (\delta_{oi} w_{iq}) a'(net_q) x_j = \eta \delta_{hq} x_j. \quad (2.24)$$

Similar a  $\delta_{oi}$ , o erro geral de retropropagação para a camada escondida é definido como:

$$\delta_{hq} = -\frac{\partial E}{\partial net_q} = -\frac{\partial E}{\partial z_q} \frac{\partial z_q}{\partial net_q} = a(net_q) \sum_{i=1}^n \delta_{oi} w_{iq} . \quad (2.25)$$

Comparando-se (2.24) com (2.20), conclui-se que  $\delta_{hq}$  é determinado por  $\delta_{oi}$ , que é obtido a partir da camada de saída para a camada escondida. Portanto, fica aparente que esses erros de retropropagação,  $\delta_{hq}$  e  $\delta_{oi}$ , são propagados através das camadas para a implementação do algoritmo de aprendizado. O procedimento pode ser generalizado e aplicado a redes com mais de uma camada escondida. Até o momento foi apresentada a versão do algoritmo de treinamento por retropropagação baseado em padrões, o que significa que a atualização dos pesos da rede neural é realizada imediatamente após a apresentação de um par de treinamentos. Entretanto, uma outra opção é o treinamento em lote, onde os pesos só são atualizados após a apresentação de todo o conjunto de padrões. O treinamento por padrões é classificado como incremental e pode ser utilizado para aplicações que precisam de atualizações em tempo real, enquanto o treinamento em lote, frequentemente, apresenta melhor convergência.

Segundo Kostela (2003), a rede *feedforward* é considerada um modelo semi-paramétrico, o que significa que o número efetivo de parâmetros pode ser menor do que o número de parâmetros disponíveis. O número efetivo de parâmetros determina a complexidade do modelo. Para pesos pequenos, o mapeamento da rede é quase linear e possui baixa complexidade efetiva uma vez que a região central da função de ativação sigmóide pode ser aproximada por uma transformação linear.

Um problema prático das RNAs é a escolha correta da quantidade de neurônios na camada oculta ou a correta regularização dos parâmetros. Frequentemente o modelo que minimiza os erros para os dados de treinamento não generaliza bem novos dados. Isso acontece porque o modelo começa a aprender o ruído dos dados de treinamento. Uma vez que a rede *feedforward* é um modelo flexível, e algoritmos eficientes são aplicados na busca de parâmetros ótimos, é provável que ocorra excesso de treinamento e como consequência é necessária a aplicação de métodos de regularização para que a capacidade de generalização não seja prejudicada. Tradicionalmente a complexidade da rede *feedforward* é controlada pela parada antecipada ou métodos de decrescimento de pesos (KOSTELA, 2003).

### 2.4.2 Método Quase-Newton

Segundo Haykin (2001), os métodos de quase-Newton são basicamente métodos de gradiente descritos pela seguinte equação de atualização:

$$w(n+1) = w(n) + \eta(s)n s(n), \quad (2.26)$$

onde o vetor de direção  $s(n)$  é definido em termos do vetor gradiente  $g(n)$  por

$$s(n) = -S(n)g(n). \quad (2.27)$$

A matriz  $S(n)$  é uma matriz definida positiva que é ajustada de uma iteração para a seguinte. Isto é feito de modo que o vetor de direção  $s(n)$  aproxime a *direção de Newton*, ou seja,

$$-\left(\frac{\partial^2 \mathcal{E}_{med}}{\partial w^2}\right)^{-1} * \left(\frac{\partial \mathcal{E}_{med}}{\partial w}\right). \quad (2.28)$$

Os métodos quase-Newton utilizam informações de segunda ordem acerca da superfície de erro, sem realmente requerer o conhecimento da matriz hessiana  $H$ . O processo utiliza dois vetores iterativos sucessivos  $w(n)$  e  $w(n+1)$ , juntamente com os respectivos vetores de gradiente  $g(n)$  e  $g(n+1)$ . Considerando-se que

$$q(n) = g(n+1) - g(n) \quad (2.29)$$

e

$$\Delta w(n) = w(n+1) - w(n), \quad (2.30)$$

pode-se então derivar a informação de curvatura usando a fórmula aproximada:

$$q(n) \cong \left(\frac{\partial}{\partial w} g(n)\right) \Delta w(n). \quad (2.31)$$

Em particular, dado  $w$  incrementos de peso linearmente independentes  $\Delta w(0), \Delta w(1), \dots, \Delta w(w-1)$  e os respectivos incrementos de gradiente  $q(0), q(1), \dots, q(w-1)$ , pode-se aproximar a matriz hessiana  $H$  por

$$H \cong [q(0), q(1), \dots, q(w-1)][\Delta w(0), \Delta w(1), \dots, \Delta w(w-1)]^{-1}, \quad (2.32)$$

e a matriz Hessiana inversa é dada por:

$$H^{-1} \cong [\Delta w(0), \Delta w(1), \dots, \Delta w(w-1)][q(0), q(1), \dots, q(w-1)]^{-1}. \quad (2.33)$$

Quando a função custo  $\mathcal{E}_{med}(w)$  é quadrática, as equações (2.29) e (2.30) são exatas.

Na classe mais popular de métodos quase-Newton, a matriz  $S(n+1)$  é obtida a partir do seu valor prévio  $S(n)$  e dos vetores  $\Delta w(n)$  e  $q(n)$ , utilizando a recursão (FLETCHER, 1987; BERTSEKAS, 1995), tal que:

$$S(n+1) = S(n) + \frac{\Delta w(n)\Delta w^T(n)}{q^T(n)q(n)} - \frac{S(n)q(n)q^T(n)S(n)}{q^T(n)S(n)q(n)} + \xi(n)[q^T(n)S(n)q(n)]v(n)v^T(n), \quad (2.34)$$

onde

$$v(n) = \frac{\Delta w(n)}{\Delta w^T(n)\Delta w(n)} - \frac{S(n)q(n)}{q^T(n)S(n)q(n)} \quad (2.35)$$

e  $0 \leq \xi(n) \leq 1, \forall n$ .

O algoritmo é iniciado com uma matriz definida positiva arbitrária  $S(0)$ . A forma particular do método quase-Newton é parametrizada de acordo com a definição de  $\eta(n)$ , como indicado a seguir (FLETCHER, 1987):

- para  $\xi(n) = 0$  para todo  $n$ , obtém-se o *algoritmo de Davison-Fletcher-Powell* (DFP), que é historicamente o primeiro algoritmo quase Newton;
- para  $\xi(n) = 1$  para todo  $n$ , obtém-se o *algoritmo Broyden-Fletcher-Goldfarb-Shanno* (BFGS), que é considerado a melhor forma de método quase-Newton, conhecida atualmente.



## 2.5 EVOLUÇÃO DIFERENCIAL

Os algoritmos evolutivos referem-se à classe de algoritmos de busca, estocásticos, baseados em população, que são desenvolvidos a partir de idéias e princípios da evolução natural. Seguindo essas idéias, estão incluídas as estratégias evolutivas, programação evolutiva, algoritmos genéticos e evolução diferencial. Uma característica de todos esses algoritmos é a busca baseada em população, onde os indivíduos de uma população competem e trocam informação entre si a fim de executar alguma tarefa. Um estrutura geral de AEs pode ser descrito pela Figura 2.5.

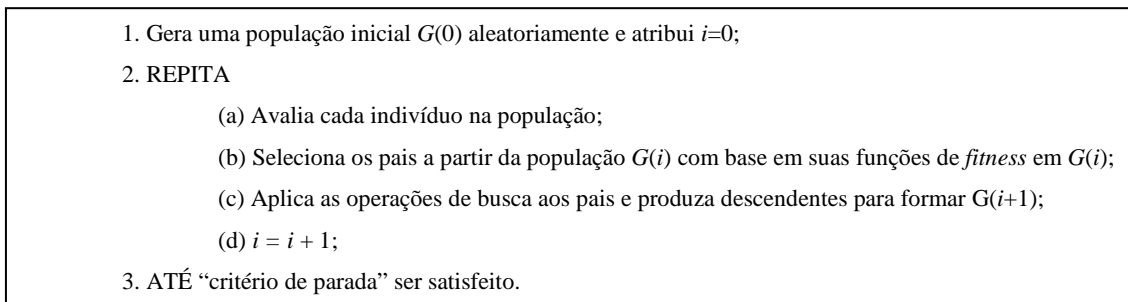


Figura 2.5 Estrutura geral de um algoritmo evolutivo.

Segundo Yao (1999), os AEs são úteis para a resolução de problemas complexos que possuam vários ótimos locais. Eles são menos susceptíveis a ficar presos em mínimos locais do que os algoritmos de busca baseados em informação do gradiente (*backpropagation*, Quase Newton BFGS, etc.). Eles podem ser aplicados a problemas onde uma função objetivo não exista ou não esteja explicitamente definida e, como não dependem da informação do gradiente, são adequados a problemas onde tal informação não está disponível ou é obtida ou estimada de forma custosa.

O algoritmo ED (Evolução Diferencial) é uma técnica de otimização global, que é um método heurístico capaz de lidar com funções não-lineares e não-diferenciáveis (STORN; PRICE, 1997). Da mesma forma que o algoritmo genético, utiliza-se de operações bio-inspiradas de cruzamento, mutação e seleção como mecanismos de evolução. O algoritmo também emprega um cruzamento não-uniforme que privilegia um determinado vetor ancestral em detrimento de outros ancestrais utilizados no cruzamento. A principal diferença entre um

algoritmo genético e ED é a operação de mutação. Nos algoritmos genéticos, a mutação é causada por pequenas alterações nos genes e no algoritmo de ED é realizada com base na combinação de indivíduos (DU, 2007). O cruzamento não-uniforme embaralha informações das combinações com melhor *fitness* e faz com que a busca seja executada no espaço de soluções. O algoritmo ED introduz uma operação de mutação baseada em diferenças entre pares de soluções selecionadas de forma aleatória na população.

Como todos os algoritmos genéticos, a ED manipula uma população de soluções. A população  $p$  de uma geração  $g$  tem  $NP$  vetores (indivíduos  $X^g$ ). Cada indivíduo representa uma solução potencial por:

$$p = \{X_i^g\}, i = 1, 2, \dots, NP. \quad (2.36)$$

Por sua vez, a solução  $X^g$  contém  $D$  variáveis, tal que:

$$X_i^g = x_{ij}^g, j = 1, 2, \dots, D. \quad (2.37)$$

Para cada geração, os indivíduos são atualizados num esquema de reprodução. Para produzir um novo indivíduo, as operações de diferenciação e recombinação são aplicadas ao conjunto de soluções gerado.

Um conjunto de indivíduos é extraído aleatoriamente para a operação de diferenciação. Um vetor de diferença  $\delta$  e um vetor base  $\beta$  são projetados com base nestes indivíduos extraídos. Assim, o resultado da operação de diferenciação é

$$\omega = \beta + F\delta, \quad (2.38)$$

onde  $F > 0$  é a constante de diferenciação.

A próxima operação de combinação representa um caso típico de troca de partes de um indivíduo. Um indivíduo tentativa  $\omega$  herda cromossomos com alguma probabilidade. Assim,

$$\omega = \begin{cases} \omega_j, & \text{se } rand < CR, \\ x_{i,j}^g, & \text{qualquer outro caso,} \end{cases} \quad (2.39)$$

onde  $j = 1, 2, \dots, D$  e  $CR \in [0, 1]$  é a constante de recombinação.

Os parâmetros  $F$  e  $CR$  afetam a velocidade de convergência e a robustez do processo de busca. Os valores ótimos dependem das características da função objetivo e do tamanho da população  $NP$ , e assim, a seleção dos valores de parâmetros ótimos depende da aplicação, onde

$$X_i^{g+1} = \begin{cases} \omega & \text{se } f(\omega) < f(X_i^g), \\ x_i^g & \text{qualquer outro caso.} \end{cases} \quad (2.40)$$

Para todos os algoritmos evolutivos os parâmetros operacionais controlam o nível entre *exploitation* (usar o material existente na busca de uma solução melhor) e *exploration* (procurar por genes melhores). As seções a seguir descrevem os principais parâmetros e valores recomendados para o algoritmo ED.

### 2.5.1 Tamanho da população

Price e Storn (1997) recomendaram o tamanho da população de 5 a 20 vezes a dimensionalidade do problema, com multiplicadores de 2 e 100 representando os valores, mínimo e máximo. Estes valores têm funcionado na prática, demonstrando que a quantidade de material genético contida nas populações é suficiente. Entretanto, outras pesquisas com algoritmos evolutivos produziram melhores resultados com multiplicadores entre 1,5 e 2 (MAYER, 2002).

### 2.5.2 Seleção de ancestrais

Para seleção de ancestrais, uma grande quantidade de métodos de seleção foi utilizada no passado, incluindo roleta (a escolha tradicional para algoritmo genético), seleção completa, truncamento e torneio (onde o tamanho 2 mostra-se como o padrão atual). O algoritmo ED utiliza a seleção completa (cada ancestral é considerada em cada geração).

### 2.5.3 Operador de recombinação

Em relação à taxa de recombinação, vários estudos sobre algoritmos evolutivos mostraram que recombinação e mutação têm um efeito sinérgico (MICHALEWICZ e FOGEL, 2000). A ED incorpora ambos em uma só operação usando uma forma de cruzamento uniforme. Em ED, a recombinação é controlada por um *CR* especificado pelo usuário e para cada parâmetro, um alelo modificado ou seu ancestral é utilizado. Storn e Price (1997) sugeriram valores de *CR* entre 0 e 1 para uma otimização detalhada (apesar de lenta), 1 e 0 para convergência mais rápida (apesar de arriscada), e 0,5 como valor recomendado.

### 2.5.4 Operador de mutação

ED não tem uma taxa de mutação definida e ao invés disso obtém esse parâmetro como um efeito do *CR*. Estudos anteriores mostraram que tanto taxas baixas (0,01) quanto altas (1,0) são efetivas (MAYER, 2002). Um *CR* de 0,5 pode ser utilizado como valor inicial para taxa de mutação, mas estudos mostraram que a forma de mutação aplicada não é importante contanto que exista uma para guiar o processo de exploração.

De forma única, que permite auto-adaptação da quantidade de mutação ao longo do progresso da otimização, a *ED* adiciona uma diferença entre dois ancestrais para um terceiro ancestral. Esta é uma forma aritmética aplicada para cada valor real que poder ser intermediário ou extrapolativo, dependendo de um fator de escala (*F*). Storn e Price (1997) recomendam um valor de *F* entre 0,4 e 1, tendo 0,5 como escolha inicial.

### 2.5.5 Estratégias de ED

Na primeira variante de ED (ED1), para cada vetor  $X_{i,G}$ ,  $i = 0, 1, 2, \dots, NP-1$ , um vetor tentativa  $\underline{V}$  é gerado conforme

$$\underline{V} = \underline{X}_{r_1,G} + F(\underline{X}_{r_2,G} - \underline{X}_{r_3,G}), \quad (2.41)$$

com  $r_1, r_2, r_3 \in [0, NP-1]$ , inteiro e diferentes, e  $F > 0$ .

Os inteiros  $r_1$ ,  $r_2$  e  $r_3$  são escolhidos aleatoriamente de um intervalo  $[0, NP-1]$  e são diferentes do índice corrente  $i$ .  $F$  é uma fator real e constante que controla a amplificação da variação diferencial  $(\underline{X}_{r_2,G} - \underline{X}_{r_3,G})$ . A Figura 2.6 mostra um exemplo que ilustra os diferentes vetores utilizados no esquema ED1. O vetor  $\underline{X}_{r_1}$  é selecionado aleatoriamente. Os vetores  $\underline{X}_{r_2}$  e  $\underline{X}_{r_3}$  são selecionados aleatoriamente, executada a operação de diferença e seu resultado multiplicado pelo fator de amplificação  $F$ . Em seguida, é executada a operação de adição para a geração do vetor  $\underline{V}$ .

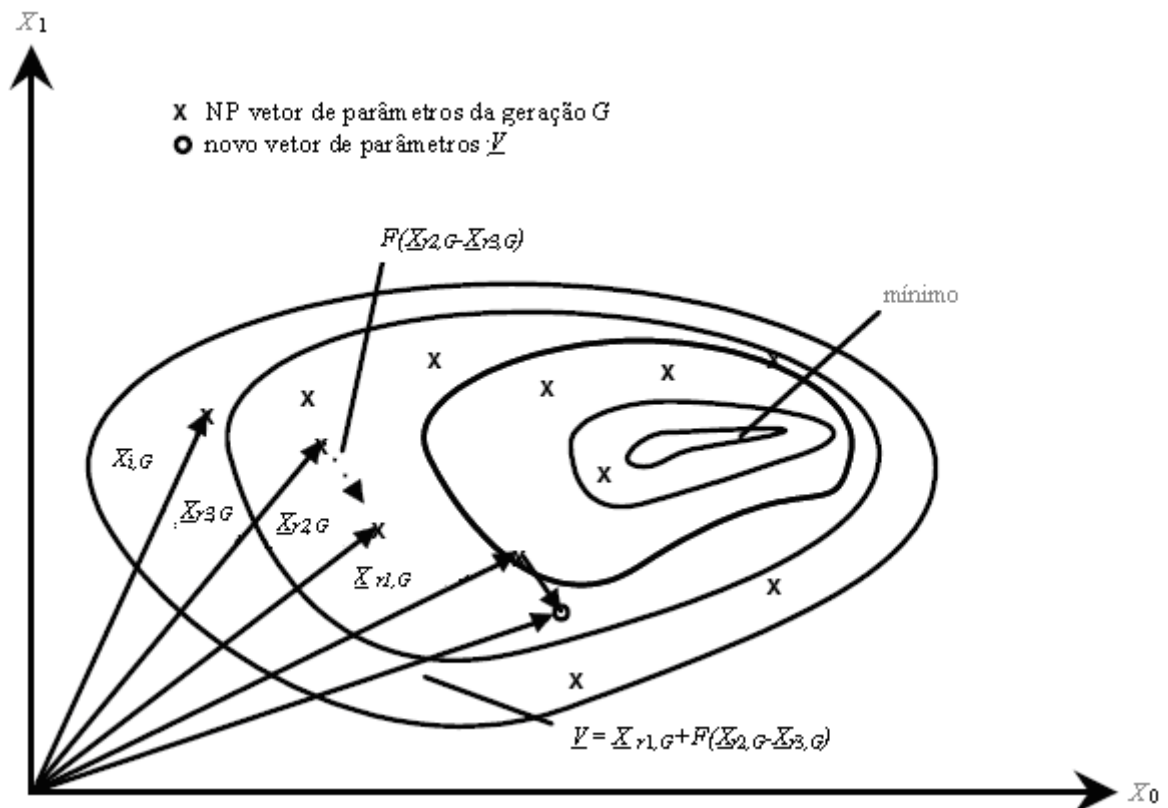


Figura 2.6. Processo de geração do vetor tentativa  $\underline{V}$  no esquema ED1.

Para aumentar a diversidade dos vetores de parâmetro, é gerado um vetor  $u$ , definido nas equações (2.42) e (2.43), com base no vetor tentativa e valores modificados segundo a probabilidade definida em CR.

$$u = (u_0, u_1, \dots, u_{D-1})^T \quad (2.42)$$

com

$$u_j = \begin{cases} v_j, & \text{para } j = \langle n \rangle_D, \langle n+1 \rangle_D, \dots, \langle n+L-1 \rangle_D \\ (X_{i,G})_j, & \text{para todos os outros } j \in [0, D-1], \end{cases} \quad (2.43)$$

onde  $\langle \cdot \rangle_D$  é a função módulo com módulo  $D$ .

A equação (2.42) produz uma seqüência de vetores  $u$  com tamanho idêntico a  $v$ , onde alguns elementos do vetor recebem os valores originais de  $v$  e outros, com base em uma probabilidade determinada pelo parâmetro CR são modificados a título de cruzamento. Escolher um subgrupo de parâmetros para mutação é similar ao processo conhecido como *crossover* na teorias de algoritmos genéticos e estratégias evolutivas. A idéia é ilustrada na Figura 2.7 para  $D=7$ ,  $n=2$  e  $L=3$ . O índice  $n$  iniciando em 2 foi escolhido aleatoriamente dentro do intervalo inteiro  $[0, D-1]$ . O inteiro  $L$  denota quantidade de parâmetros que devem ser trocados a partir de elemento  $n$ . O algoritmo que determina  $L$  funciona conforme a linhas de pseudocódigo a seguir, onde  $rand()$  gera números aleatórios pertencentes ao intervalo  $[0,1)$ :

$L = 0$ ;

faça {

$L = L + 1$ ;

} enquanto ( $rand < CR$ ) e ( $L < D$ ));

Como resultado, a probabilidade  $\Pr(L \geq v) = (CR)^{v-1}$ ,  $\forall v > 0$ .  $CR \in [0,1]$  é a probabilidade de cruzamento e constitui uma variável de controle para o esquema ED1. As decisões aleatórias tanto para  $n$  quanto para  $L$  são realizadas para cada vetor de tentativa  $\underline{V}$ .

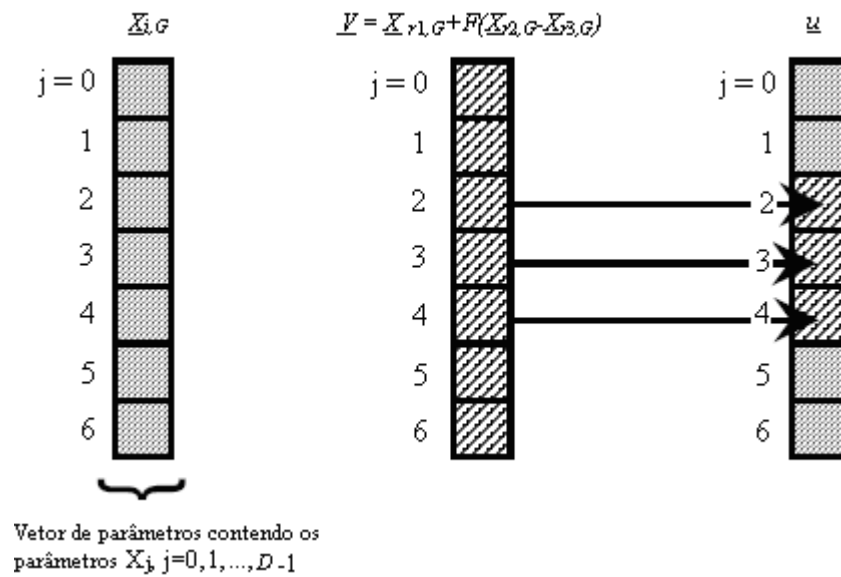


Figura 2.7. Ilustração do processo de *crossover* para  $D=7$ ,  $n=2$  e  $L=3$ .

Para decidir quando o vetor  $u$  deve ou não se transformar em um membro da população  $G+1$ , usa-se comparação com  $X_{i,G}$ . Se a função objetivo retornar um valor para o vetor  $u$  que seja menor do que o valor gerado para  $X_{i,G}$ ,  $X_{i,G+1}$  recebe  $u$ , do contrário o valor antigo de  $X_{i,G}$  é retido.

O esquema evolução diferencial com duas diferenças (ED2) funciona da mesma forma que ED1, mas gera o vetor  $\underline{V}$  de acordo com a seguinte equação:

$$\underline{V} = \underline{X}_{i,G} + \lambda * (\underline{X}_{best,G} - \underline{X}_{i,G}) + F * (\underline{X}_{r2,G} - \underline{X}_{r3,G}), \quad (2.44)$$

onde  $\lambda$  uma variável de controle adicional. A idéia associada a  $\lambda$  é dar significado à “fome” do esquema incorporando o melhor vetor de solução em um determinado momento  $\underline{X}_{best,G}$ . Esta característica pode ser útil para funções objetivo nas quais o mínimo global é relativamente fácil de ser encontrado. A Figura 2.8 mostra o processo de geração do vetor definido na equação (2.44). A construção de  $u$  a partir de  $\underline{V}$  e  $\underline{X}_{i,G}$  bem como a decisão do processo funcionam da mesma forma que em ED1.

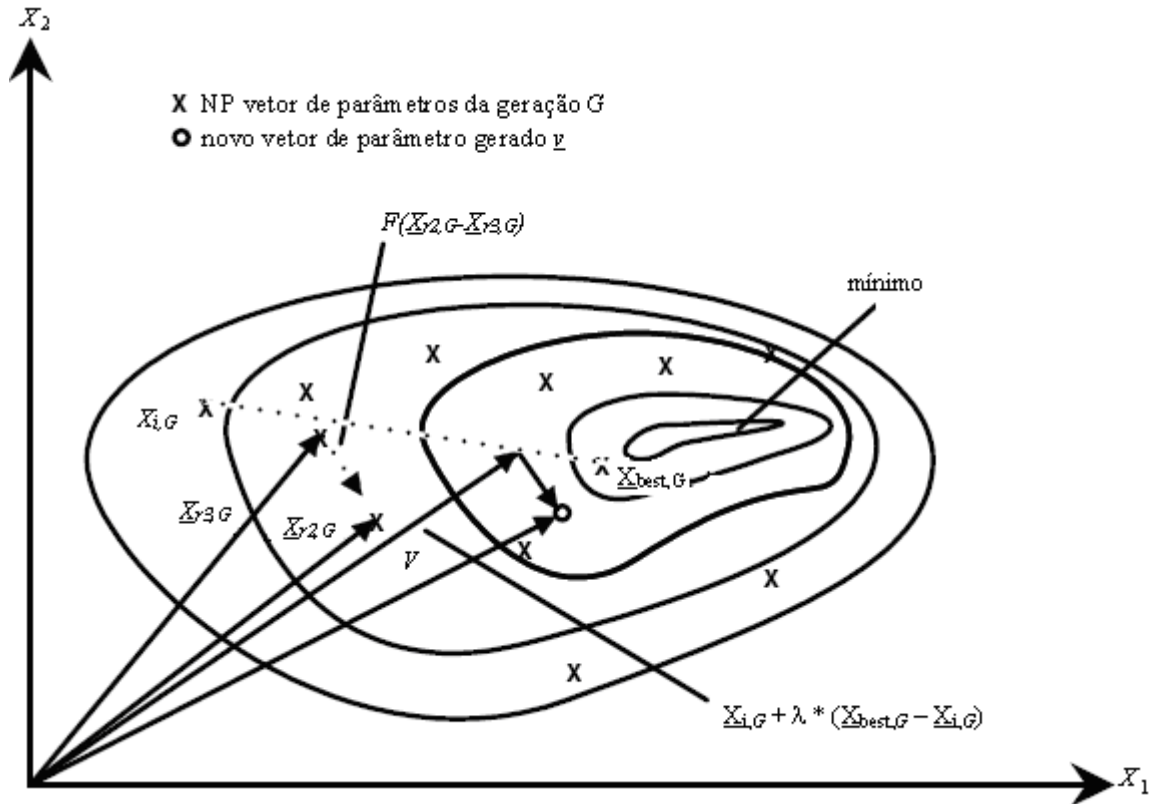


Figura 2.8 Processo de geração do vetor tentativa  $V$  no esquema ED2.

## 2.6 MODELOS DA FAMÍLIA ARCH

Os modelos ARCH (Heterocedasticidade Condicional Auto-Regressiva, do inglês, *Autoregressive Conditional Heteroskedasticity*) são usados para modelar a volatilidade do retorno de um ativo e são exemplos de modelos de heterocedasticidade condicional. A volatilidade tem sua utilidade em várias aplicações financeiras, como por exemplo, em precificação de opções e na alocação de recursos em estruturas de variância mínima.

Segundo Tsay (2005), a volatilidade não pode ser observada diretamente dos dados de retorno uma vez que existe somente uma observação em um dia de pregão. Se forem considerados dados intradiários (retorno de 15 minutos ou frequências maiores) é possível o cálculo da volatilidade diária, mas, será necessária uma análise cuidadosa dos dados uma vez dados de alta frequência possuem pouca informação sobre a volatilidade diária.



Embora não seja diretamente observável, a volatilidade apresenta os seguintes comportamentos comuns em séries de retornos de ativos:

- agrupamento de volatilidade – pode ser alta para alguns períodos e baixa para outros;
- evolui de maneira contínua ao longo do tempo – é raro acontecer saltos;
- não diverge para o infinito – as variações ficam restritas a uma faixa de valores (significa que a volatilidade é estatisticamente estacionária); e
- tem comportamento diferenciado na reação a um retorno significativamente positivo e um retorno significativamente negativo (comportamento conhecido como efeito de alavancagem).

Essas características são motivos de pesquisa dos modelos de previsão de volatilidade e a diferença de um modelo está na capacidade de capturar um determinado comportamento, por exemplo, modelo EGARCH (GARCH exponencial, do inglês, *Exponencial GARCH*) para capturar o efeito de alavancagem.

As alterações no comportamento da variância dos erros de uma série de dados é característica frequentemente encontrada em séries temporais financeiras e é denominada heterocedasticidade. Neste tipo de série, a heterocedasticidade apresenta-se a partir de grandes (pequenos) retornos absolutos seguidos por outros grandes (pequenos) retornos. Isto significa que existem períodos que mostram agrupamentos de alta (baixa) volatilidade e estes podem ser observados em séries diárias e semanais, de ações, *commodities* e câmbio.

Alguns coeficientes de correlação e o conjunto dos quatro primeiros momentos (média, variância, assimetria e curtose) da Estatística podem ajudar a identificar o modelo GARCH que melhor captura o comportamento da série.

### 2.6.1 Presença do efeito ARCH

Segundo Alexander (2005), por consequência do agrupamento, observa-se uma forte autocorrelação dos retornos ao quadrado, o que pode ser verificado calculando-se o coeficiente de correlação de primeira ordem dos retornos ao quadrado como mostrado na equação (2.45).

$$ccpo = \frac{\sum_{t=2}^T r_t^2 r_{t-1}^2}{\sum_{t=2}^T r_t^4}, \quad (2.45)$$

onde  $r$  é o retorno e  $T$  é o tamanho da amostra.

A significância do coeficiente de correlação de primeira ordem pode ser medida a partir do teste de autocorrelação de Box-Pierce. O teste de Box-Pierce é uma forma de teste do multiplicador de Lagrange. É uma variável que segue assintoticamente uma distribuição qui-quadrado com 1 grau de liberdade, para o caso de correlação de primeira ordem, e, por consequência, seu valor crítico a 1% é de 6,635. A equação com teste Box-Pierce para ordem  $p$ , e dada por:

$$Q = T \sum_{n=1}^p \varphi(n)^2, \quad (2.46)$$

onde  $T$  é o tamanho da amostra e  $\varphi(n)^2$ , a autocorrelação de ordem  $n$ .

O teste de Box-Pierce é simples em relação à programação computacional, mas não robusto em termos de resultado. É necessário considerar outras medidas estatísticas no sentido de confirmar o resultado do teste. Em uma situação hipotética onde se verifica assimetria negativa e extremo excesso de curtose, é provável que o resultado do teste esteja sendo influenciado por *outliers* extremamente negativos que induzem uma baixa autocorrelação.

Outro aspecto a ser considerado na identificação do modelo está relacionado com a assimetria presente nas séries e essa característica pode ser verificada pelo coeficiente de autocorrelação de primeira ordem entre os retornos defasados e os retornos correntes ao quadrado, ou seja:

$$capo = \frac{\sum_{t=2}^T r_t^2 r_{t-1}}{\sqrt{\sum_{t=2}^T r_t^4 \sum_{t=2}^T r_{t-1}^2}}. \quad (2.47)$$

onde  $r$  é o retorno.

Caso se obtenha um resultado negativo para o coeficiente e o teste de Box-Pierce for estatisticamente diferente de zero, conclui-se que existe assimetria no agrupamento, a qual não será capturada por modelos GARCH simétricos.

### 2.6.2 Modelo ARCH

O modelo ARCH, de Engle (1982), modela a volatilidade e apóia-se na idéia de que o resíduo  $\alpha_t$  do retorno de um ativo não é serialmente correlacionado, mas é dependente de  $\alpha_t$  e esta dependência pode ser representada por uma função quadrática de seus valores defasados. Um modelo ARCH( $m$ ) pode ser representado da seguinte forma:

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2, \quad (2.48)$$

onde  $\{\varepsilon_t\}$  é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância 1,  $\alpha_0 > 0$ , e  $\alpha_i \geq 0$  para todo  $i > 0$ . Na prática, pressupõe-se que  $\varepsilon_t$  possui distribuição normal, t-Student ou distribuição generalizada de erro. Pela estrutura do modelo percebe-se que grandes variações do passado implicam em grande variância. Isto significa que grandes variações tendem a ser seguidas por outras grandes variações.

É necessário considerar as seguintes desvantagens dos modelos ARCH:

- a) o modelo assume que *shocks* positivos e negativos tem o mesmo efeito sobre a volatilidade uma vez que ele depende do quadrado do *shock* anterior, mas, na prática é sabido que as séries financeiras respondem de forma diferente ante a *shocks* negativos e positivos;
- b) o modelo ARCH é restritivo uma vez que  $\alpha_1^2$  deve assumir valores no intervalo  $[0, 1/3]$ , se a série tem curtose finita. Na prática, essa restrição dificulta a modelagem com muitos parâmetros e limita a habilidade dos modelos ARCH em capturar excesso de curtose;
- c) o modelo ARCH não fornece informações sobre o que gerou as variações nas séries financeiras. É simplesmente uma forma de descrever o comportamento da variância condicional;
- d) é possível que o modelo superestime a volatilidade porque responde lentamente a grandes *shocks* da série de retornos.

### 2.6.3 Modelo GARCH

Com a intenção de resolver o problema de excesso de parâmetros do modelo ARCH, Bollerslev (1986) propôs uma extensão conhecida como modelo ARCH generalizado (GARCH). Ao modelo foi adicionada uma variável para capturar o comportamento da variância condicional em separado. Para uma série de logaritmos de retornos  $r_t$ , onde  $a_t = r_t - \mu_t$  é o resíduo no tempo  $t$ ,  $a_t$  segue um modelo GARCH( $m, s$ ) se

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2, \quad (2.49)$$

onde  $\{\varepsilon_t\}$  é uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com média zero e variância 1,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$  e  $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$ . É importante explicitar  $\alpha_i = 0$  para  $i > m$ ,  $\beta_j = 0$  para  $j > s$  e  $\alpha_i + \beta_i$  implica que a variância incondicional de  $\alpha_t$  é finita e a variância incondicional evolui ao longo do tempo.

Os modelos GARCH sofrem da mesma deficiência que os modelos ARCH uma vez que não diferencia uma variação negativa de uma variação positiva e conforme observado por Tsay (2005), o mercado reage de forma diferenciada a esses eventos.

## 2.7 MEDIDAS DE DESEMPENHO

Nesta dissertação, foram adotadas as medidas de desempenho detalhadas a seguir.

- Raiz do Erro Quadrático Médio (REQM): Avalia o tamanho do erro de previsão e é calculada através da média de erros de cada previsão em relação ao valor real, elevados ao quadrado:

$$REQM = \sqrt{\frac{1}{N} \sum_{t=1}^N (x_t - \hat{x}_t)^2} \quad (2.50)$$

onde  $x_t$  e  $\hat{x}_t$  representam o valor previsto e o real e  $N$  é o tamanho do conjunto de dados. Quando mais próximo de zero é o REQM, melhor é o modelo de previsão.

- Índice de Desigualdade de Theil (U-Theil): Avalia o ajustamento da série prevista à série original. Quanto mais próximo a zero, maior é o ajustamento da série prevista em relação a série original. Em contrapartida, valores mais próximos da unidade indicam que o modelo não conseguiu fazer boas previsões. O índice de U-Theil é definido por

$$U - Theil = \frac{\sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2}}{\sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t)^2 + \frac{1}{N} \sum_{t=1}^N (y_t)^2}}, \quad (2.51)$$

onde  $\hat{y}_t$  é o valor previsto para o período  $t$ ,  $y_t$  é o valor observado e  $N$  é o número de observações.

### 3. METODOLOGIA DE PESQUISA

Neste capítulo, são apresentados os aspectos relacionados à metodologia aplicada nesta dissertação. Na seção 3.1, encontra-se a caracterização da pesquisa onde este trabalho é classificado segundo sua natureza, seus objetivos e procedimento técnicos. Na seção 3.2, a forma da coleta e o tratamento dos dados são descritos brevemente. A seção 3.3 refere-se à definição das variáveis a serem estudadas, onde será descrito todo o processo de formação e atualização da carteira IBOVESPA e a seção 3.4 traz a hipótese de pesquisa a ser testada no capítulo 4.

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

Para caracterizar o presente trabalho é utilizada a forma clássica de classificação de pesquisas científicas definida por Silva e Menezes (2005), que considera a pesquisa em relação à sua natureza, a abordagem do problema utilizada, do ponto de vista de seus objetivos, e do ponto de vista de seus procedimentos técnicos:

- Natureza da Pesquisa: este trabalho pode ser caracterizado como uma pesquisa aplicada, pois tem como objetivo a geração de conhecimentos para utilização prática, conhecimentos estes que são dirigidos à solução de problemas específicos. A pesquisa aplicada envolve dogmas e interesses específicos (SILVA; MENEZES, 2005);
- Abordagem do Problema: este trabalho pode ser definido também como uma pesquisa predominantemente quantitativa por utilizar-se, em sua maioria, métodos matemáticos para explicar os resultados pesquisados. Nesta abordagem, valoriza-se o entendimento dos resultados matemáticos, bem como sua interpretação por parte do pesquisador;

- Em relação aos objetivos: o presente trabalho pode ser caracterizado como uma pesquisa explicativa, uma vez que visa identificar os fatores que determinam ou contribuem para a ocorrência dos fenômenos;
- Procedimentos técnicos: a classificação do presente trabalho, em relação aos procedimentos técnicos utilizados, se enquadra como uma pesquisa com delineamento *ex-post facto*, uma vez que é baseada em experimentos realizados após a ocorrência dos fatos, e sendo assim, o pesquisador não tem controle sobre as variáveis analisadas.

### 3.2 COLETA E TRATAMENTO DOS DADOS

Para o tratamento dos dados coletados são utilizadas métricas predominantemente quantitativas, a fim de que se tenham resultados matemáticos que corroborem com a teoria utilizada na previsão de séries temporais.

Os dados da série temporal do índice BOVESPA à vista, foram conseguidos no sítio [www.finance.yahoo.com](http://www.finance.yahoo.com); a série histórica de índice BOVESPA futuro foi obtida na BMF (Bolsa de Mercadorias e Futuros), em [www.bmf.com.br](http://www.bmf.com.br). O período de dados considerado neste trabalho tem início em 01/06/2004 e término em 30/05/2008.

### 3.3 DEFINIÇÃO DAS VARIÁVEIS A SEREM ESTUDADAS

O Índice Bovespa é o valor atual, em moeda corrente, de uma carteira teórica de ações constituída em 02/01/1968, a partir de uma aplicação hipotética. Supõe-se não ter sido efetuado nenhum investimento adicional, desde então, considerando-se somente os ajustes efetuados em decorrência da distribuição de proventos pelas empresas emissoras (tais como reinversão de dividendos recebidos e do valor apurado com a venda de direitos de subscrição, e manutenção em carteira das ações recebidas em bonificação). Dessa forma, o índice reflete não apenas as variações dos preços das ações, mas também o impacto da distribuição dos



proventos, sendo considerado um indicador que avalia o retorno total de suas ações componentes.

O valor absoluto do índice não tem grande relevância. O importante é sua variação entre duas datas que definem determinado período de tempo. Mas, de qualquer forma, é essencial saber que o valor absoluto do índice Bovespa corresponde ao valor (monetário) de mercado de uma carteira selecionada de ações negociadas na Bolsa de Valores de São Paulo em determinado instante. A variação deste índice é, portanto, uma taxa de lucratividade (retorno) que seria auferida por quem mantivesse esta carteira durante o período de análise.

A finalidade básica do índice é a de servir como indicador médio do comportamento do mercado. Para tanto, sua composição procura aproximar-se o mais possível da real configuração das negociações à vista (lote-padrão) na Bovespa. Em termos de liquidez, as ações integrantes da carteira teórica do índice Bovespa respondem por mais de 80% do número de negócios e do volume financeiro verificados no mercado à vista (lote-padrão) da Bovespa.

O índice de liquidez da ação é calculado conforme a equação a seguir:

$$L = \sqrt{(N * G) * P}, \quad (3.2)$$

onde

*L*: índice de liquidez da ação;

*N*: participação da ação no mercado, em termos de volume (R\$), no período considerado;

*G*: número relativo de negócios, ou seja, participação percentual da ação no mercado, em termos de número de negócios, no período considerado;

*P*: presença relativa da ação no mercado, ou seja, número de pregões em que a ação apresenta pelo menos um negócio em relação ao número total de pregões do período considerado.

O índice de liquidez das ações consiste, portanto, numa média geométrica das participações percentuais da ação no mercado em termos de volume monetário e número de

negócios, ponderada pela presença relativa. A média geométrica de  $N$  e  $G$  é apropriada para medir a liquidez, pois, à medida que os dois indicadores se afastam, há uma queda do resultado. Para que o índice de liquidez seja alto é necessário que a ação participe de todos os pregões do período com, simultaneamente, alto volume de negociações e elevado número de negócios.

A negociabilidade é um conceito central na metodologia de construção e manutenção do índice Bovespa. O índice de negociabilidade é definido por

$$IN = \sqrt{\frac{n_i}{N} * \frac{v_i}{V}}, \quad (3.3)$$

onde:

$IN$ : índice de negociabilidade;

$n_i$ : número de negócios com a ação  $i$  no mercado a vista (lote-padrão);

$N$ : número total de negócios no mercado à vista da BOVESPA (lote-padrão);

$v_i$ : volume financeiro gerado pelos negócios com a ação  $i$  no mercado à vista (lote-padrão);

$V$ : volume financeiro total do mercado à vista da BOVESPA (lote-padrão).

Com uma metodologia de fácil acompanhamento pelo mercado, o índice Bovespa representa fielmente o comportamento médio das principais ações transacionadas, e o perfil das negociações à vista observadas nos pregões da Bovespa. O índice Bovespa é o somatório dos pesos (quantidade teórica da ação multiplicada pelo seu último preço) das ações integrantes de sua carteira teórica. Assim sendo, pode ser apurado por meio da seguinte equação:

$$Ibovespa_t = \sum_{i=1}^n P_{i,t} * Q_{i,t} \quad (3.4)$$

onde

$Ibovespa_t$ : índice bovespa no instante  $t$ ;

$n$ : número total de ações componentes da carteira teórica;

$P$ : último preço da ação  $i$  no instante  $t$ ;

$Q$ : quantidade teórica da ação  $i$  na carteira no instante  $t$ .

Os seguintes critérios metodológicos são considerados na construção e manutenção da carteira teórica do índice Bovespa:

- Participam da carteira teórica do índice Bovespa as ações de maior negociabilidade nos últimos 12 meses e que, em conjunto, representam pelo menos 80% da soma dos índices de negociabilidade apurados para todas as ações negociadas à vista nos pregões do período analisado. Adicionalmente, exige-se que cada ação selecionada por este critério tenha participado, em pelo menos, 80% dos pregões do período e que seu volume de negócios no mercado à vista (em R\$) corresponda a mais do que 0,1% do volume total (R\$) de negociações no mesmo período;
- A constituição da carteira teórica do índice é refeita a cada quatro meses no primeiro dia útil de cada quadrimestre do ano (janeiro/abril; maio/agosto e setembro/dezembro), com a finalidade de atualizar a representatividade da amostra. Na abertura do pregão do dia da mudança, evidentemente, as carteiras “antiga” e “nova” têm o mesmo valor;
- As ações componentes da carteira do índice têm suas participações estabelecidas pelos respectivos índices de negociabilidade. A carteira do índice é ponderada pela negociabilidade das ações componentes. A consequência mais imediata deste critério é que o índice passa a espelhar eventuais contrações de mercado que se manifestam em prazos relativamente longos (últimos 12 meses);
- Não recebendo investimentos adicionais, a carteira segue com seu valor estabelecido pelas cotações das ações componentes a cada momento.

A Bovespa calcula seu índice em tempo real, considerando os preços dos últimos negócios efetuados no mercado à vista (lote-padrão), com ações componentes de sua carteira. Sua divulgação é realizada pela rede de difusão da Bovespa e também retransmitida por uma série de assinantes, sendo possível, dessa forma, acompanhar *on-line* seu comportamento em qualquer parte do Brasil ou do mundo.

### 3.4 HIPÓTESE DE PESQUISA

Uma hipótese metodológica pode ser vista como um enunciado ou afirmação que se faz com o objetivo de posterior comprovação ou não.

Para Luna (1997), a formulação de hipótese é quase inevitável para quem é estudioso da área de pesquisa. Geralmente, com base em análises do conhecimento disponível, o pesquisador acaba concentrando-se naquilo que pode surgir como resultado da pesquisa.

Assim, após verificar a formulação do problema de pesquisa, tem-se a seguinte hipótese de pesquisa:

H<sub>1</sub>: O desempenho dos modelos de RNA de Elman, treinados com os algoritmos ED e/ou quase-Newton do tipo BFGS são superiores ao modelo GARCH(1,1).

A hipótese de pesquisa será corroborada ou não com base no coeficiente de desigualdade U-Theil. Caso o U-Theil dos modelos de RNA de Elman, treinados com os algoritmos ED e/ou quase-Newton do tipo BFGS, seja inferior, ao U-Theil do modelo GARCH(1,1), a hipótese será confirmada. Caso contrário, refutada.

## 4. APRESENTAÇÃO E ANÁLISE DOS DADOS

Neste capítulo tem-se os resultados auferidos a partir dos dados coletados. O capítulo encontra-se estruturado em duas seções que tratam: das séries temporais do IBOVESPA à vista e futuro, bem como, tratam da comparação entre os métodos de previsão propostos.

### 4.1 AS SÉRIES TEMPORAIS DO IBOVESPA

A amostra utilizada compreende o período de Junho de 2004 a Maio de 2008, totalizando 986 cotações de fechamento do IBOVESPA à vista e 987 cotações do IBOVESPA futuro. Para o processo de estimação dos parâmetros do modelo GARCH e treinamento das redes neurais, são utilizadas as primeiras 789 observações do IBOVESPA à vista e 790 do IBOVESPA futuro, e reservadas as observações restantes das amostras para a fase de avaliação do modelo onde o conjunto de validação é estimado com base nos parâmetros identificados.

A importância de prever a volatilidade de uma série temporal financeira reside no fato do potencial investidor conhecer o risco assumido nos períodos ex-ante, podendo o mesmo, traçar estratégias que maximizem os retornos esperados de um ativo ou carteira.

#### 4.1.1 Análise da série de dados do IBOVESPA à vista

Na análise visual da série do IBOVESPA à vista disposta na Figura 4.1, é possível perceber a presença de tendência ascendente, sazonalidade e valores discrepantes provenientes de movimentos abruptos de subida ou descida, próximos a outubro de 2007 e março de 2008, confirmando as quedas referentes à crise de hipotecas *subprime* que atingiu os Estados Unidos da América e refletiu no mercado de ações brasileiro.

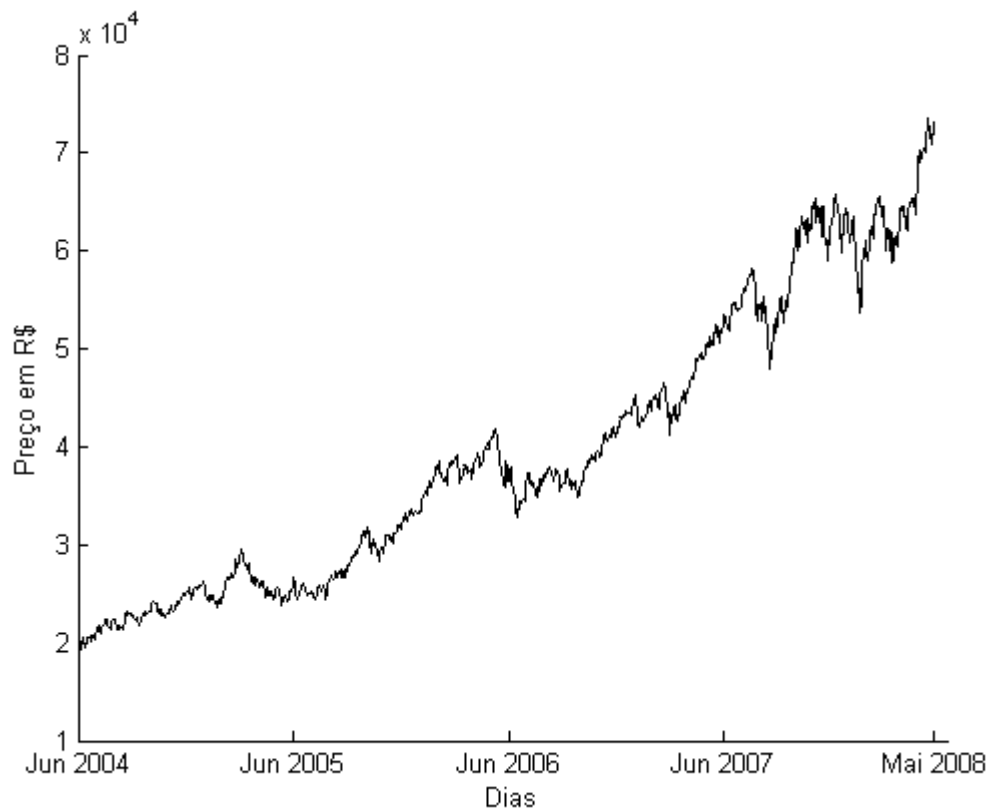


Figura 4.1. Série de preços do IBOVESPA à vista entre Jun/2004 e Mai/2008.

A Figura 4.2 mostra a série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008. As medidas de estatística descritiva obtidas para a série são apresentadas a seguir na Tabela 4.1.

Tabela 4.1. Estatísticas da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008.

Média	Desvio Padrão	Assimetria	Curtose
0,0013	0,0163	-0,2780	3,9245

A série apresenta agrupamento de volatilidade, destacado na Figura 4.2 pelos círculos. É possível identificar momentos de alta volatilidade com reações de euforia, quando se observa grandes retornos positivos, e pânico, representados por alguns retornos extremamente negativos em respostas a notícias ruins e inesperadas. O agrupamento de volatilidade implica a existência de uma forte autocorrelação entre os retornos quadrados (ALEXANDER, 2005). A assimetria negativa e o excesso de curtose indicam que existe pelo menos um retorno extremamente negativo e, como consequência, uma falsa medida de não-correlação.

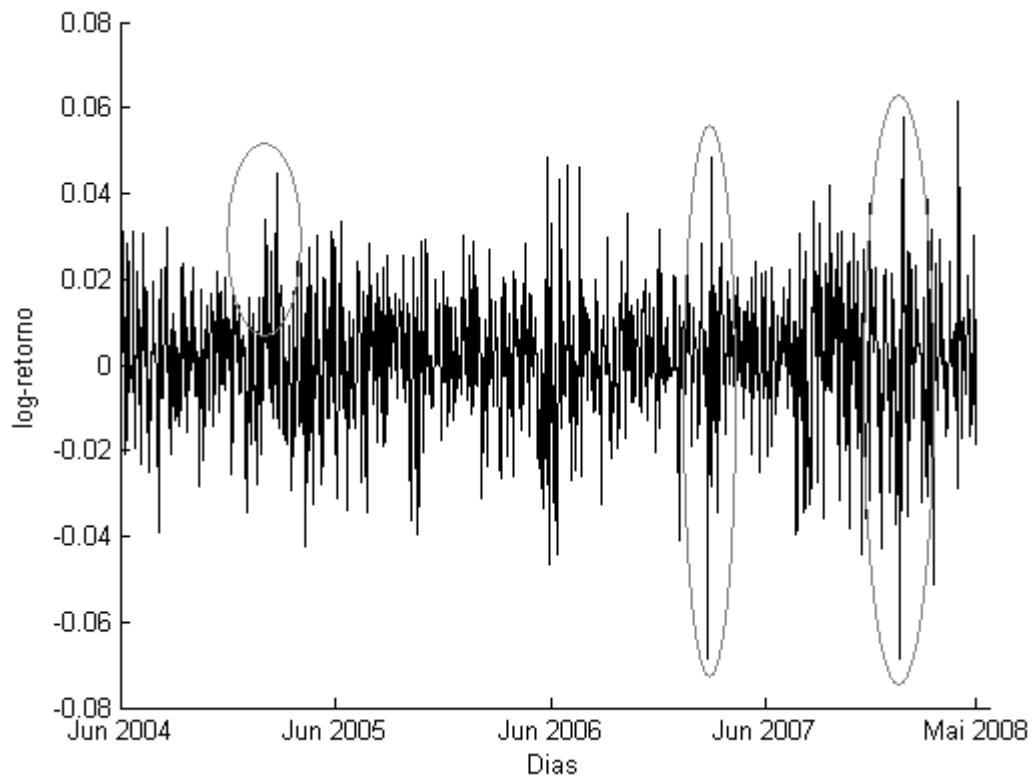


Figura 4.2. Série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008.

A Figura 4.3 apresenta o histograma da série de retornos log-normal do IBOVESPA à vista entre Jun/2004 e Mai/2008. Visualmente é possível perceber que os dados não seguem uma distribuição normal, e sim uma distribuição próxima da normal. Para a confirmação da não-normalidade foi executado o teste de Jarque-Bera (JARQUE e BERA, 1987) com nível de significância 0,05 e os resultados obtidos são apresentados na Tabela 4.2.

Tabela 4.2. Estatísticas da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008.

H0 – Hipótese nula (normalidade)	Probabilidade	Estatística
H1 – Hipótese alternativa (não-normalidade)		assintoticamente equivalente a $\chi^2$ (2 graus de liberdade)
Não-normalidade	0,00000	47,12271

Percebeu-se à luz da tabela 4.2 que o teste Jarque-Bera rejeitou a hipótese de normalidade, dado que apresentou uma estatística qui-quadrado igual a 47,12271 contra um valor tabelado igual a 5,991.

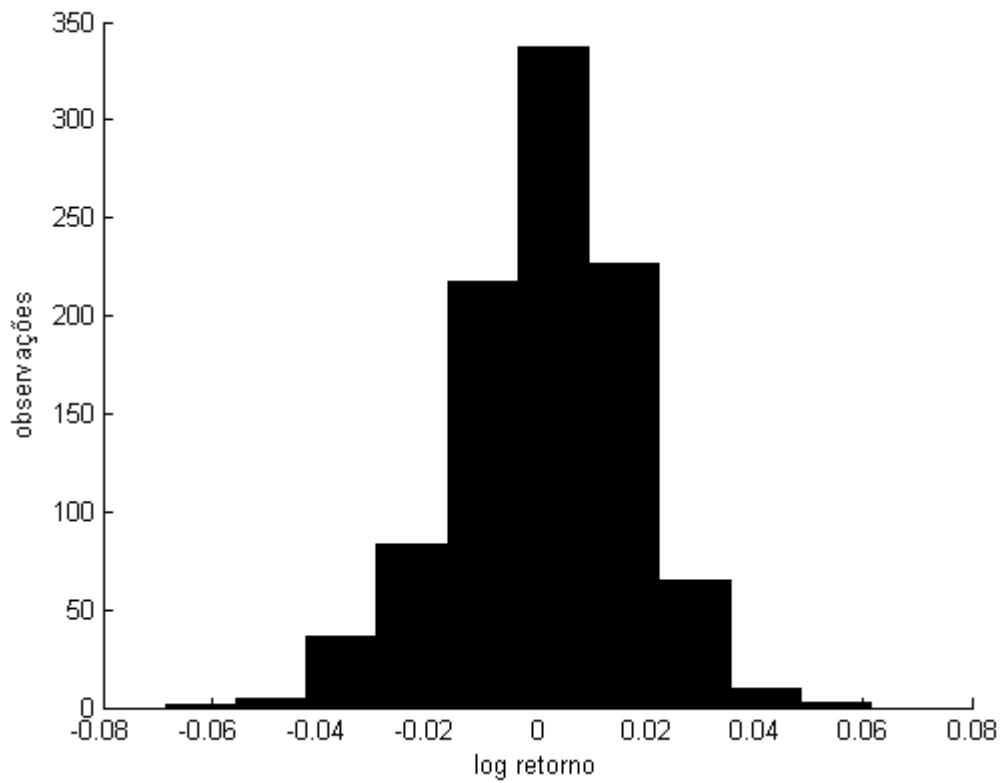


Figura 4.3. Histograma da série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008.

A Figura 4.4 apresenta a série de log-retornos quadrados do IBOVESPA à vista entre Jun/2004 e Mai/2008, mostrando elevada volatilidade nos períodos próximos de Mai/2007, Set/2007 e Mar/2008. A modelagem da heterocedasticidade pelo modelo GARCH é realizada com base nos retornos quadrados.



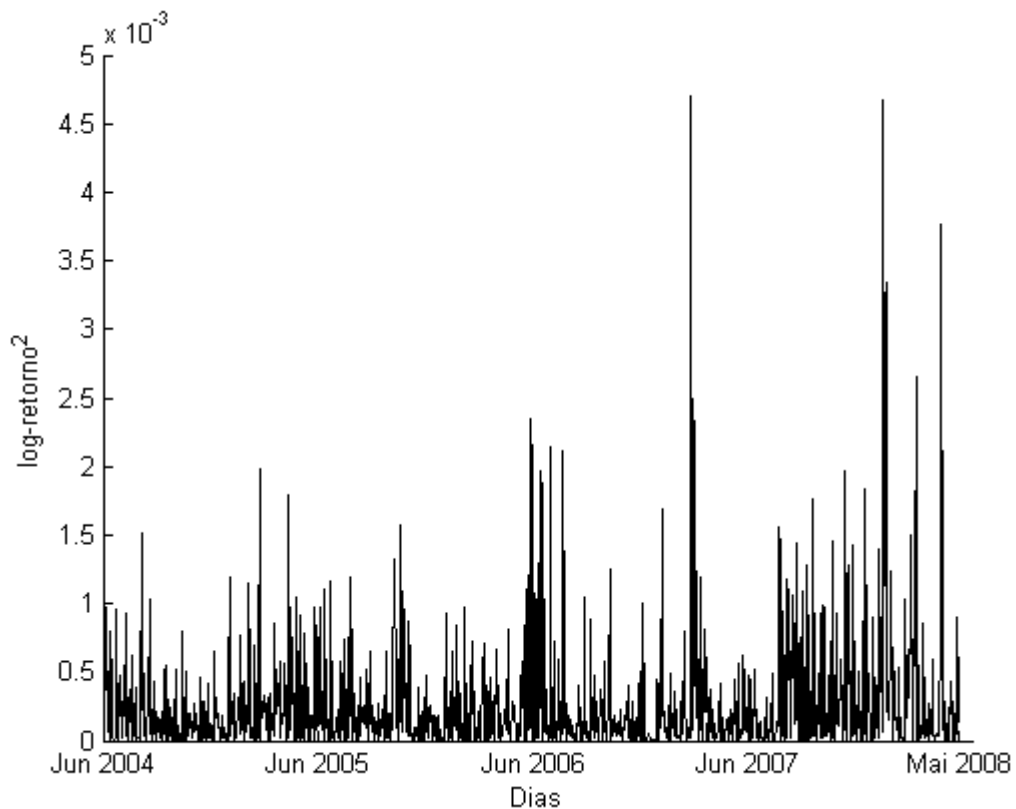


Figura 4.4. Série de log-retornos do IBOVESPA à vista entre Jun/2004 e Mai/2008.

As funções de autocorrelação dos log-retornos do IBOVESPA à vista mostram que não existe autocorrelação na série e uma equação de média condicional baseada em uma constante que capture a média de longo prazo é suficiente.

As Figuras 4.5 e 4.6, respectivamente, mostram as funções de autocorrelação amostral e autocorrelação parcial dos retornos quadrados. Observando a Figura 4.5, nota-se a existência de autocorrelação entre os retornos quadrados, uma vez que a autocorrelação de alguns atrasos se apresenta fora do intervalo de confiança. Na Figura 4.6, observa-se que seis defasagens é o número máximo de atrasos de um modelo GARCH, visto que até sexta defasagem a autocorrelação está fora do intervalo de confiança. Entretanto, neste trabalho, é utilizado um modelo GARCH(1,1) por este ser o ajustamento de referência na área financeira (ALEXANDER, 2005).

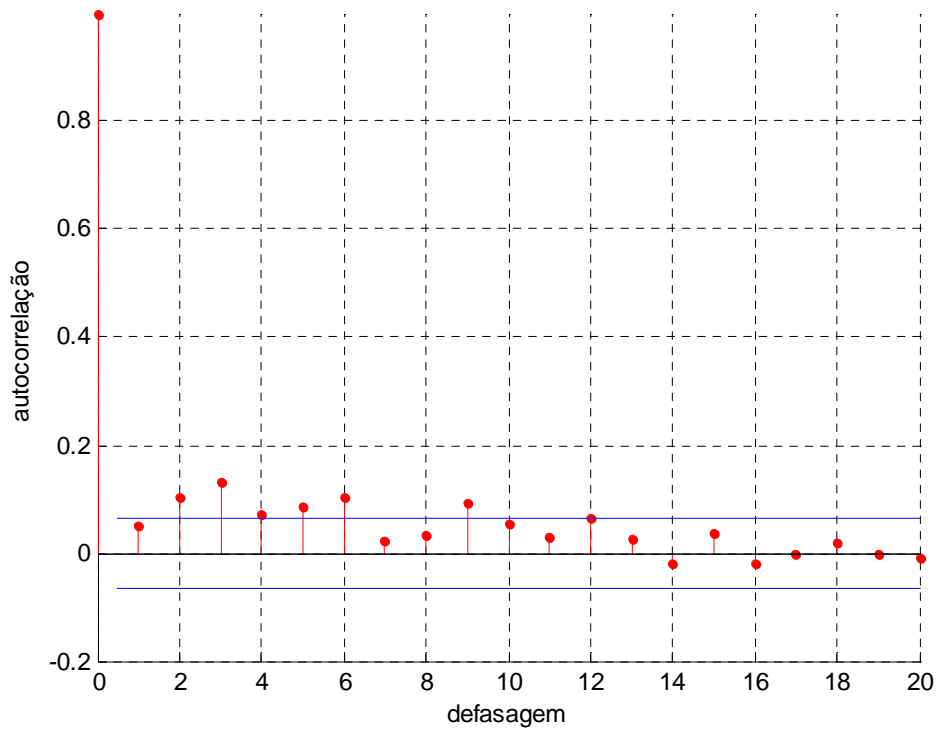


Figura 4.5. Função de autocorrelação dos log-retornos quadrados do IBOVESPA à vista entre Jun/2004 e Mai/2008.

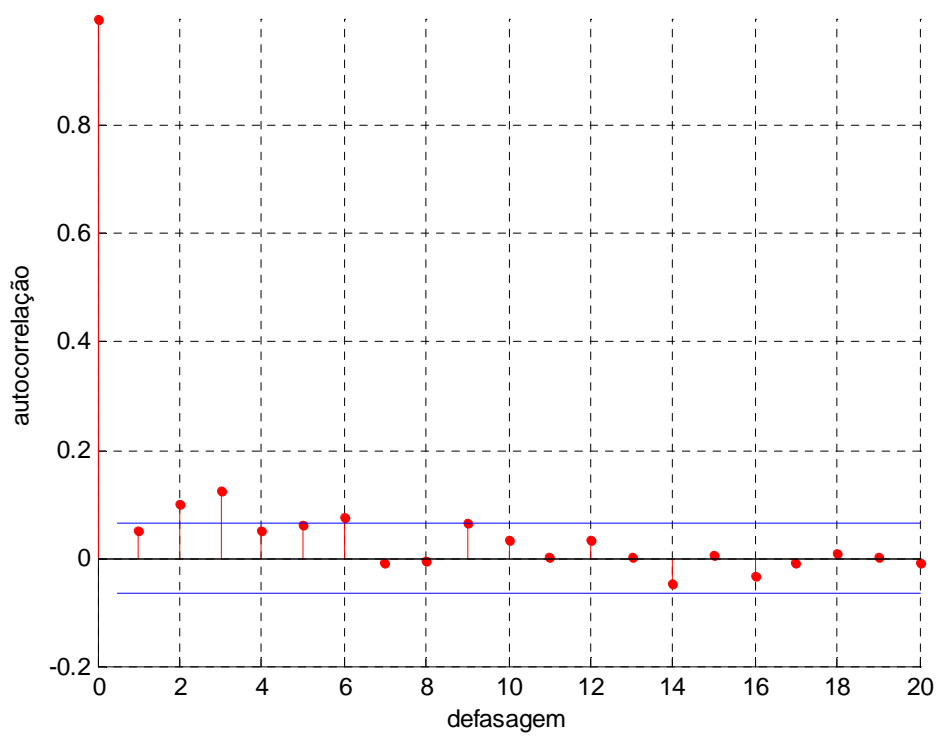


Figura 4.6. Função de autocorrelação parcial dos log-retornos quadrados do IBOVESPA à vista entre Jun/2004 e Mai/2008.

#### 4.1.2 Análise da série de dados do IBOVESPA futuro

Na análise visual da série do IBOVESPA futuro visto na Figura 4.7, é possível perceber a presença de tendência ascendente e sazonalidade nas quedas de junho de 2006 e junho de 2008. Os valores discrepantes provenientes de movimentos abruptos, próximos a outubro de 2007 e março de 2008, confirmam as quedas referentes à crise de hipotecas *subprime*, a falta de liquidez financeira no mercado bancário e pressões inflacionárias que atingiram os Estados Unidos da América e refletiram no mercado acionário brasileiro.

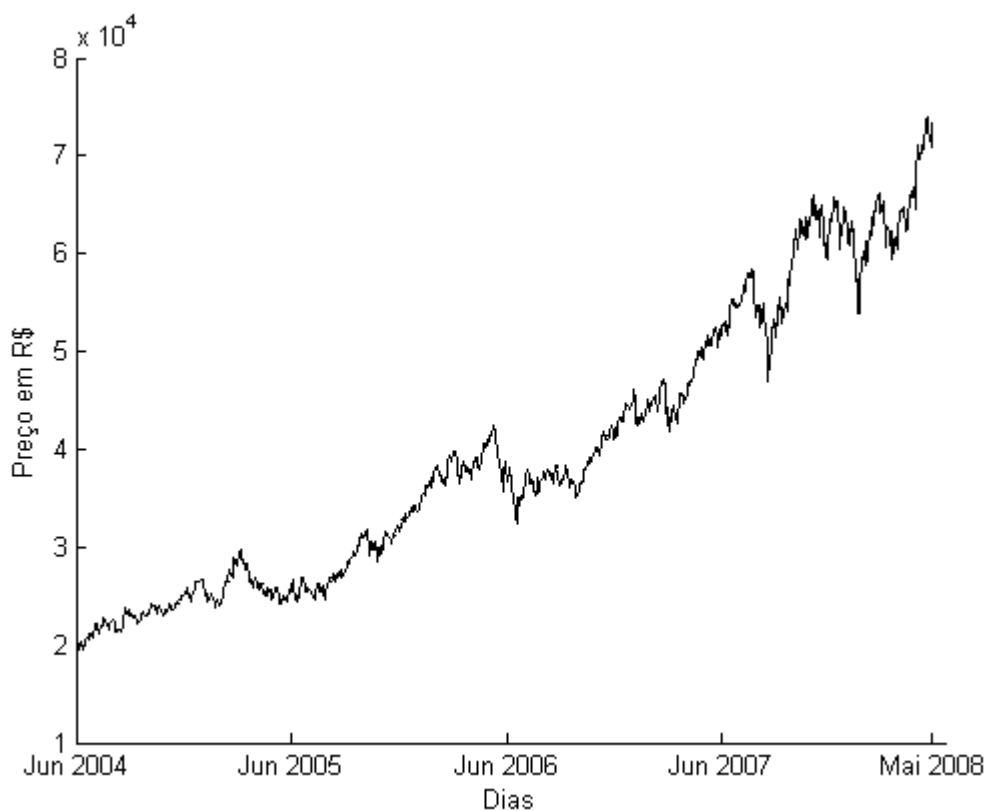


Figura 4.7. Série de preços do IBOVESPA futuro entre Jun/2004 e Mai/2008.

A Figura 4.8 mostra a série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008. As medidas de estatística descritiva obtidas para a série são apresentadas na tabela 4.3.

Tabela 4.3. Estatísticas da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008.

Média	Desvio Padrão	Assimetria	Curtose
0,0013	0,0171	-0,1201	4,1781

A série temporal apresenta um agrupamento de volatilidade. É possível identificar momentos de alta volatilidade com reações de euforia, quando se observa grandes retornos positivos, e pânico, representados por alguns retornos extremamente negativos em resposta a notícias ruins e inesperadas. O agrupamento de volatilidade implica a existência de uma forte autocorrelação entre os retornos quadrados (ALEXANDER, 2005). A assimetria negativa e o excesso de curtose indicam que existe pelo menos um retorno extremamente negativo e como consequência uma falsa medida de não-correlação.

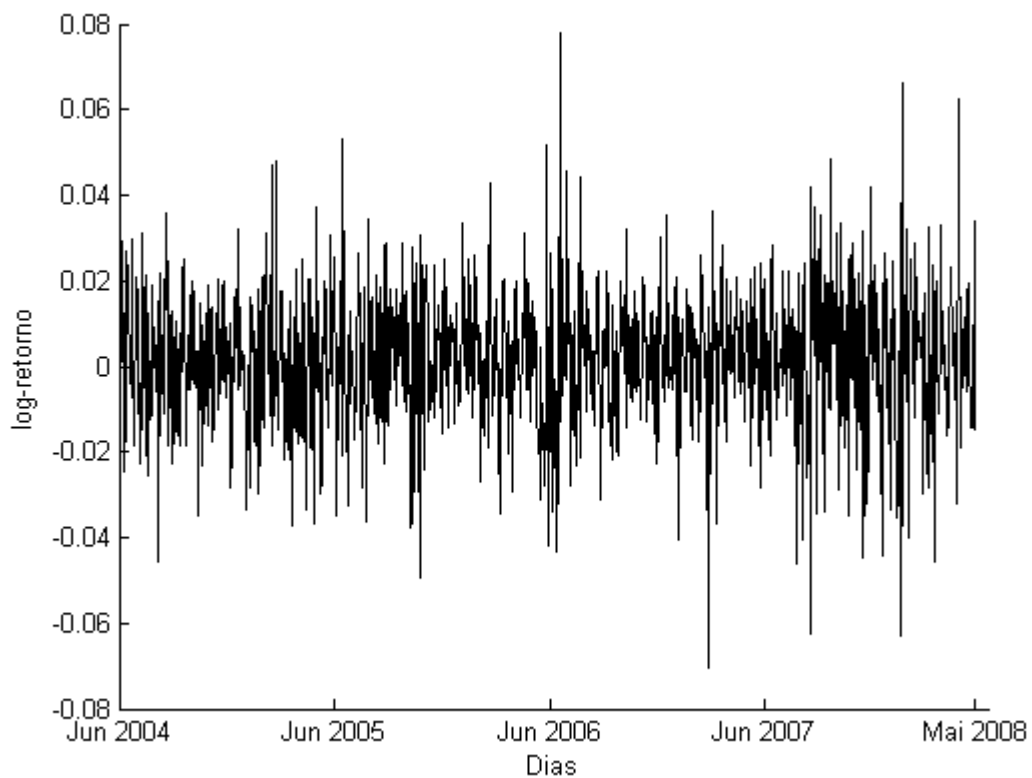


Figura 4.8. Série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008.

A Figura 4.9 apresenta o histograma da série de retornos log-normal do IBOVESPA futuro entre Jun/2004 e Mai/2008. Visualmente é possível perceber que os dados seguem uma distribuição próxima da normal. O teste de Jarque-Bera (JARQUE; BERA, 1987) com nível de significância 0,05 confirma a suposição visual de que os retornos não seguem distribuição normal e os resultados obtidos são apresentados na Tabela 4.4.

Tabela 4.4. Estatísticas da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008.

H0 – Hipótese nula (normalidade) H1 – Hipótese alternativa (não-normalidade)	Probabilidade	Estatística assintoticamente equivalente a qui- quadrado (2 graus de liberdade)
não-normalidade	0,00000	58,56982

Como o resultado do teste segue uma distribuição qui-quadrado com dois graus de liberdade, para ser significativo a 5% sua estatística deve estar maior que 5,99 e confirmar o resultado em relação a não-normalidade.

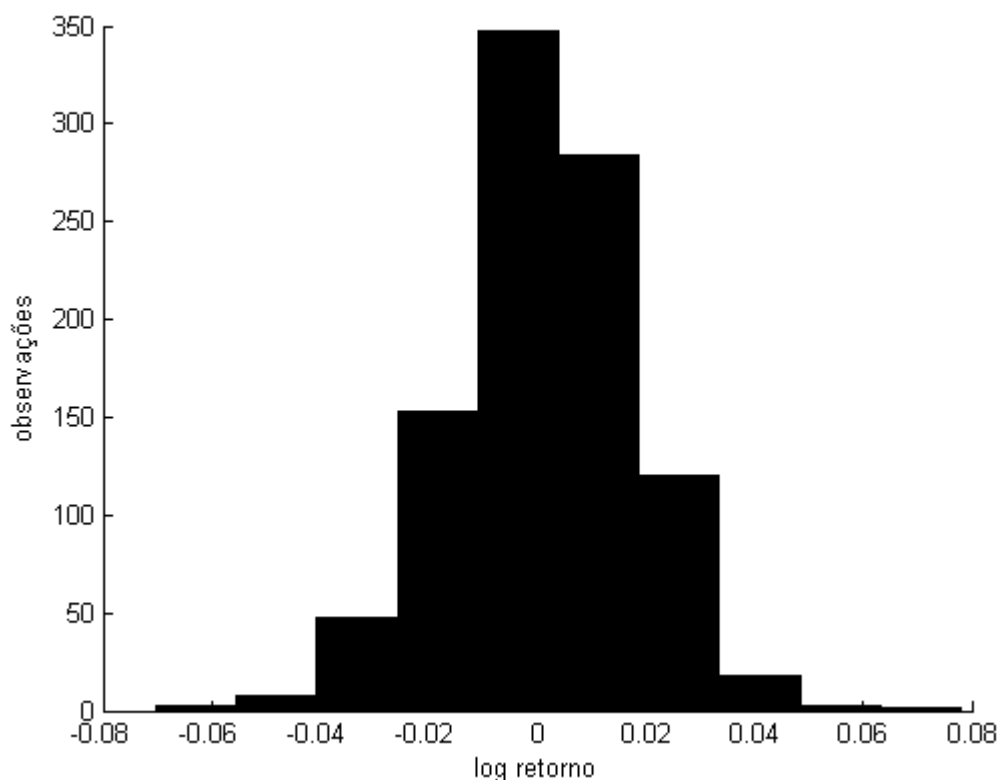


Figura 4.9. Histograma da série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008.

A Figura 4.10 apresenta a série de retornos quadrados log-normal do IBOVESPA futuro entre Jun/2004 e Mai/2008. Nesta pesquisa, a modelagem da heterocedasticidade pelo modelo GARCH é realizada com base nos retornos quadrados.

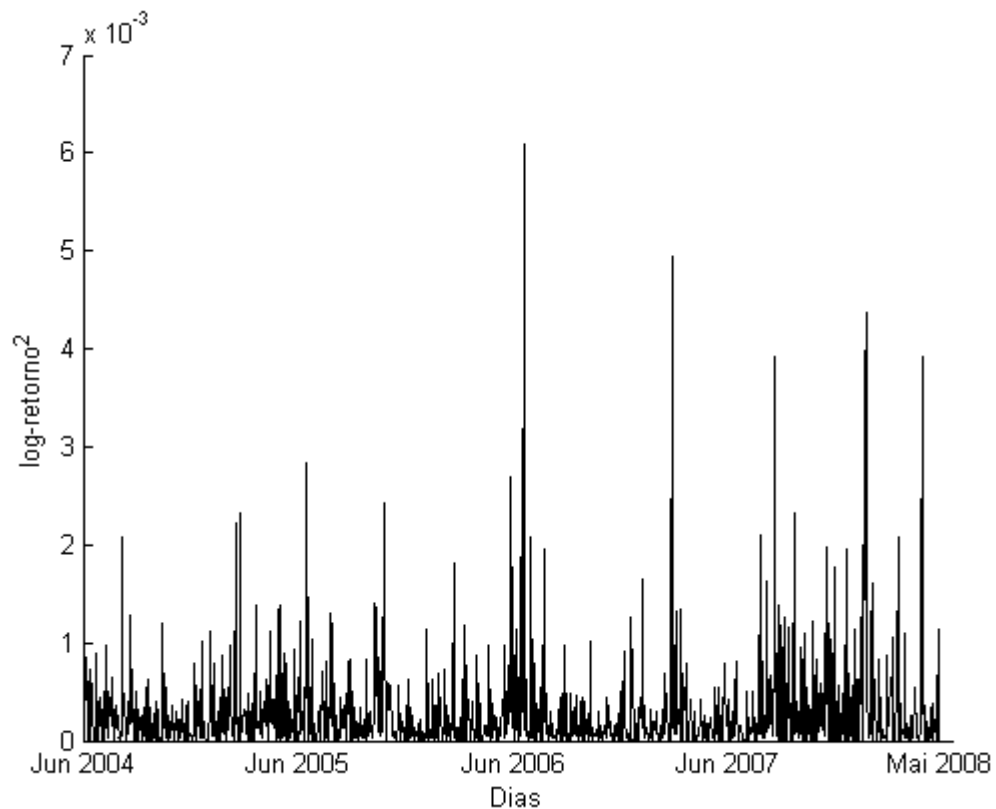


Figura 4.10. Série de log-retornos quadrados do IBOVESPA futuro entre Jun/2004 e Mai/2008.

As Figuras 4.11 e 4.12, respectivamente, mostram as funções de autocorrelação amostral e autocorrelação parcial dos retornos quadrados. Pode-se observar a existência de autocorrelação entre os retornos quadrados, o que permite a modelagem GARCH. Diferente da série do IBOVESPA à vista, o atraso máximo do modelo GARCH pode alcançar a décima defasagem. Entretanto, neste trabalho, é utilizado o GARCH(1,1) para a comparação entre os modelos.

Da mesma forma que na série temporal do IBOVESPA à vista, a série do IBOVESPA futuro também não apresenta autocorrelação no retorno e tem sua média condicional modelada a partir de uma constante para capturar a média de longo prazo dos preços.

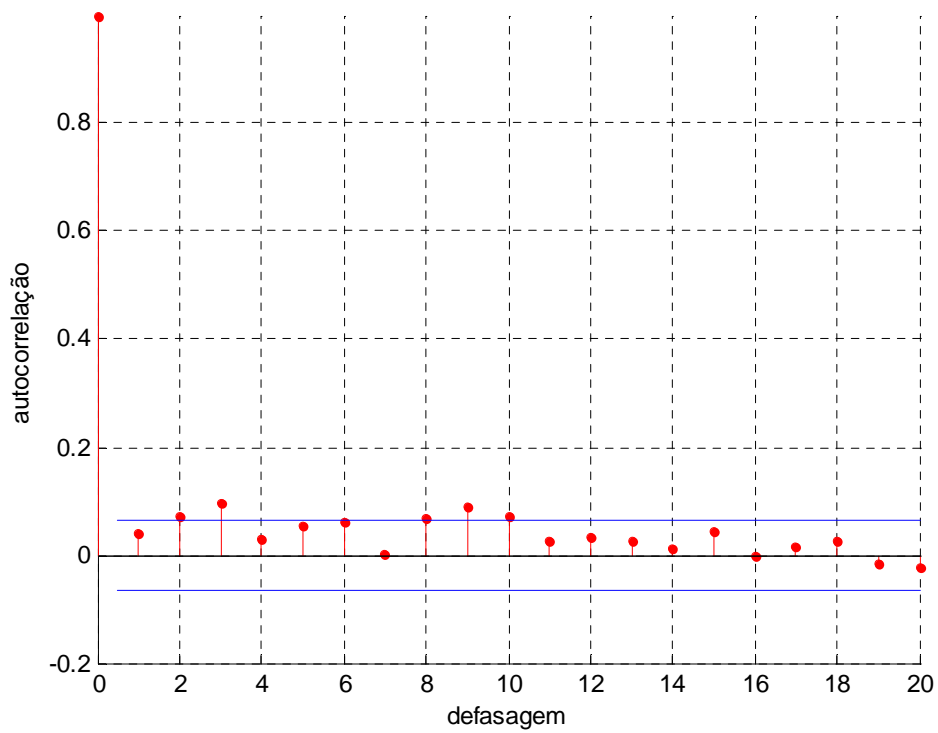


Figura 4.11. Função de autocorrelação dos log-retornos quadrados do IBOVESPA futuro entre Jun/2004 e Mai/2008.

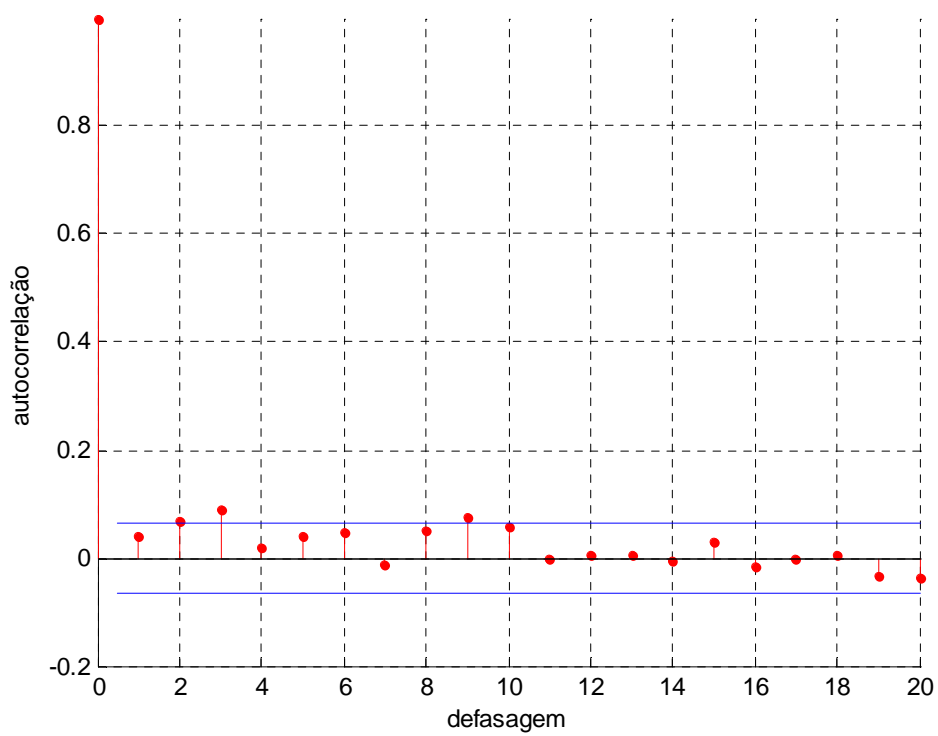


Figura 4.12. Função de autocorrelação parcial dos log-retornos quadrados do IBOVESPA futuro entre Jun/2004 e Mai/2008.

## 4.2 COMPARAÇÃO DE DESEMPENHO ENTRE MÉTODOS DE PREVISÃO

Para a modelagem das séries IBOVESPA utilizando a técnica GARCH, o requisito de que os dados sejam estacionários deve ser levado em consideração. Na intenção de obter um conjunto de dados estacionários e dessa forma eliminar componentes de tendência e sazonalidade, é realizada uma transformação logarítmica na série de preços e tomada a sua primeira diferença. Após as transformações, uma série com distribuição supostamente log-normal e não correlacionada, em relação à sua média, é obtida. Uma série não correlacionada em relação à média pode apresentar agrupamentos de volatilidade, o que é mostrado por meio do gráfico de autocorrelação dos retornos quadrados das séries do IBOVESPA, e permite a aplicação de modelos GARCH para a sua modelagem.

A arquitetura das RNAs de Elman é composta por três neurônios na camada de entrada, cinco neurônios na camada escondida e um neurônio na camada de saída que representa a previsão um passo à frente em relação ao valor da camada de entrada. Como o objetivo das RNAs é a previsão um passo à frente dos retornos da série do IBOVESPA, à vista e futuro, a escolha de um único neurônio de saída foi obtida de forma direta.

A quantidade de entradas, inicialmente, foi ajustada para 1 retorno defasado e variações da quantidade de entradas no intervalo de 1 até 15 foram testadas a fim de obter o melhor ajustamento, em relação ao coeficiente U-Theil, com o menor número de entradas defasadas.

A decisão da quantidade de neurônios na camada escondida foi baseada no ajustamento do coeficiente U-Theil e a quantidade de tempo gasto para execução total do método. A característica da RNA de Elman de ter unidades de contexto para armazenarem os resultados da camada escondida, faz com que a quantidade conexões entre os neurônios cresça rapidamente quando são adicionados neurônios a essa camada e isso exige mais tempo para o treinamento das RNAs. Uma vez que se procurava manter a quantidade de tempo, necessário para o treinamento das RNAs, compatível com o método GARCH, a arquitetura contendo apenas 3 entradas defasadas e 5 neurônios na camada escondida foi a que obteve o melhor ajustamento U-Theil e possui tempo de execução similar ao método GARCH.



Para fins de critério de parada, o máximo de 2000 chamadas à função objetivo são executadas. Neste caso, a função objetivo é o cálculo do REQM com base nas previsões obtidas pela RNA utilizando-se o conjunto de pesos atuais. Se for atingido o critério de parada no qual o erro é menor ou igual a  $1 \times 10^{-6}$ , o treinamento também é encerrado. Para as RNAs treinadas por algoritmos híbridos, são executadas 1000 avaliações da função objetivo com o primeiro algoritmo (descida de encosta ou ED) e com base na melhor solução é iniciado o treinamento com o segundo algoritmo (descida de encosta ou ED).

Na evolução diferencial é adotado o algoritmo ED auto-adaptável (SALMAN *et al.*, 2007) e as variações da equação de diferenças utilizadas são, respectivamente, *best/1/bin* e *rand/1/bin*. Inicialmente, é atribuído ao parâmetro *F* o valor 0,5 e ao parâmetro *CR*, o valor de 0,9 sendo que o valor de *F* será modificado conforme as regras do ED auto-adaptável. O tamanho da população é de 20 indivíduos. *best/1/bin* é uma variante do esquema de algoritmo de ED que leva o melhor indivíduo de uma geração considerada como atual, para a geração posterior. No caso da variante *rand/1/bin*, um indivíduo aleatório, sem garantias de ser o melhor, é escolhido para a próxima geração. Em ambos os casos, o valor 1 significa que a equação de recombinação (2.41) utiliza somente uma diferença e o *bin* faz com que o algoritmo considere uma distribuição binomial. Na Figura 4.13, ED(2) refere-se à variante *best/1/bin* e ED(1) refere-se à variante *rand/1/bin*.

Na fase de validação, para todos os modelos, os erros (resíduos) gerados, quando da comparação entre os valores previstos e realizados, são submetidos a um modelo de regressão linear e os valores ajustados, com base nos coeficientes estimados pelo modelo de regressão, são adicionados à previsão original. Esse modelo de regressão linear dos erros leva em consideração o valor de um erro defasado e uma constante representando a média do erro de longo prazo do modelo. A intenção deste procedimento é melhorar a previsão por meio de um componente similar a um filtro de média móvel, MA(1). Modelos de regressão de erros de ordens superiores foram testados, mas, a utilização de somente 1 defasagem para série de erros foi o modelo que mais contribuiu para a melhora do desempenho dos métodos testados neste trabalho.

A Figura 4.13 apresenta uma visão geral de como são conduzidos os procedimentos de entrada de dados, previsão de um passo à frente dos valores do IBOVESPA, à vista e futuro, e a geração das medidas de desempenho para a comparação dos modelos.

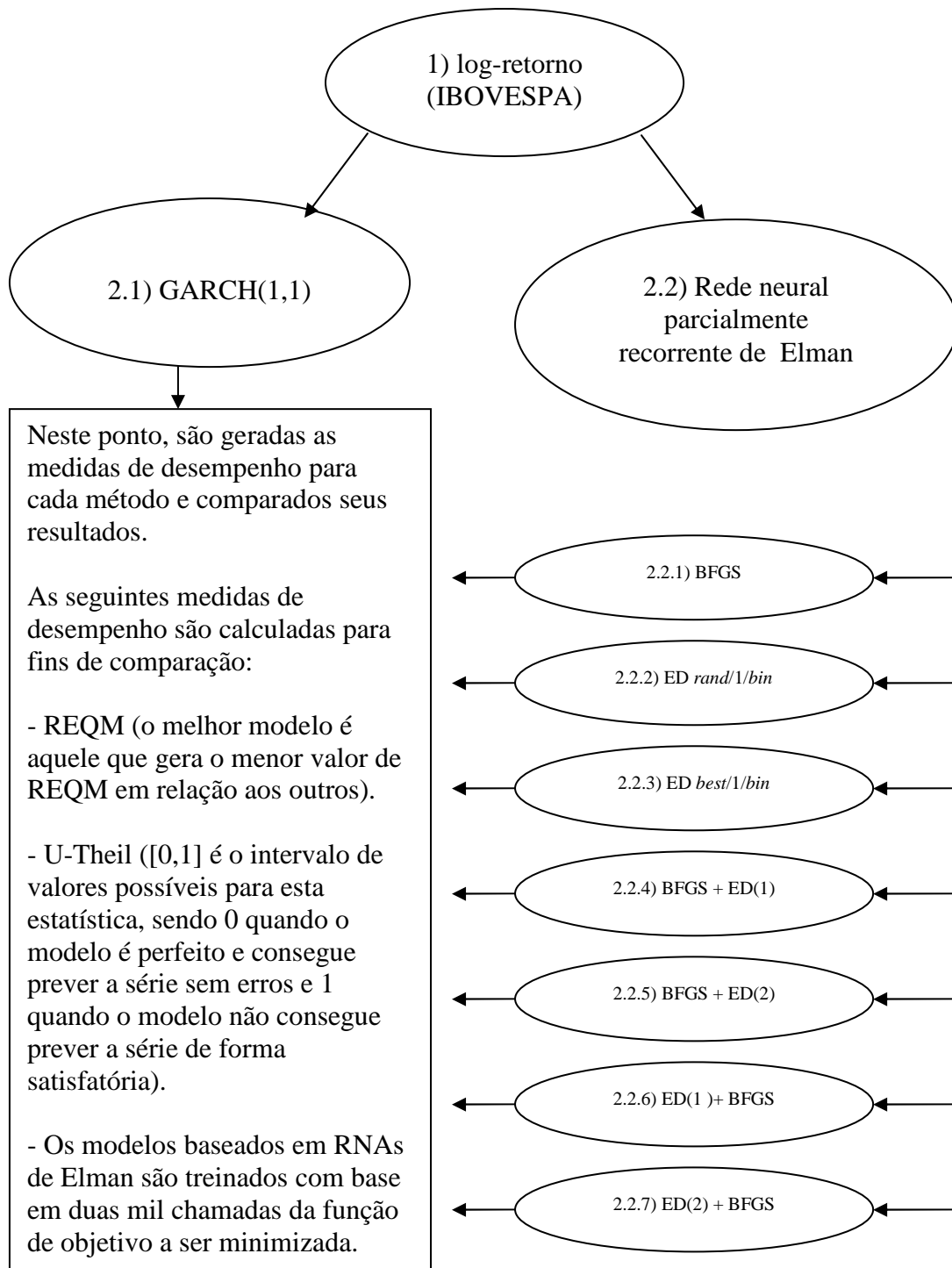


Figura 4.13. Visão geral do processo de previsão um passo a frente do IBOVESPA.

A partir da Figura 4.14 até a Figura 4.29, respectivamente, são apresentados os preços realizados e previstos bem como o erro quadrado das previsões do IBOVESPA, à vista e futuro, entre Ago/2007 e Mai/2008. Como indicados na legenda, a linha tracejada refere-se ao retorno realizado enquanto que a linha sólida, a previsão obtida pelos métodos.

As Figuras 4.14 e 4.15 mostram os gráficos de previsão obtidos pelo método GARCH(1,1) para a séries do IBOVESPA, à vista e futuro. Neste caso, uma vez que não foram observadas autocorrelação dos log-retornos do IBOVESPA, à vista e futuro, ajustou-se o seguinte modelo para a modelagem da média e variância condicionais, respectivamente:

$$\begin{aligned} y_t &= c + u_t, \\ u_t &= \sigma_t \varepsilon_t, \\ \sigma^2 &= \omega + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{4.1}$$

onde  $y_t$  é a média condicional,  $c$  é uma constante que representa a média de longo prazo e  $u_t$ , o erro do modelo que será modelado pela equação da variância condicional.  $\sigma^2$  é previsão da variância,  $\omega, \alpha_1, \beta_1$  são os parâmetros da equação da variância,  $u_{t-1}^2$ , o erro anterior e  $\sigma_{t-1}^2$ , a variância anterior.

Para o IBOVESPA à vista, o método GARCH(1,1) obteve REQM=0,0200, U-Theil=0,8661, com a presença de 71,43% de acertos dos movimentos ascendentes e 63,37% dos movimentos descendentes após a aplicação do procedimento de regressão dos resíduos.

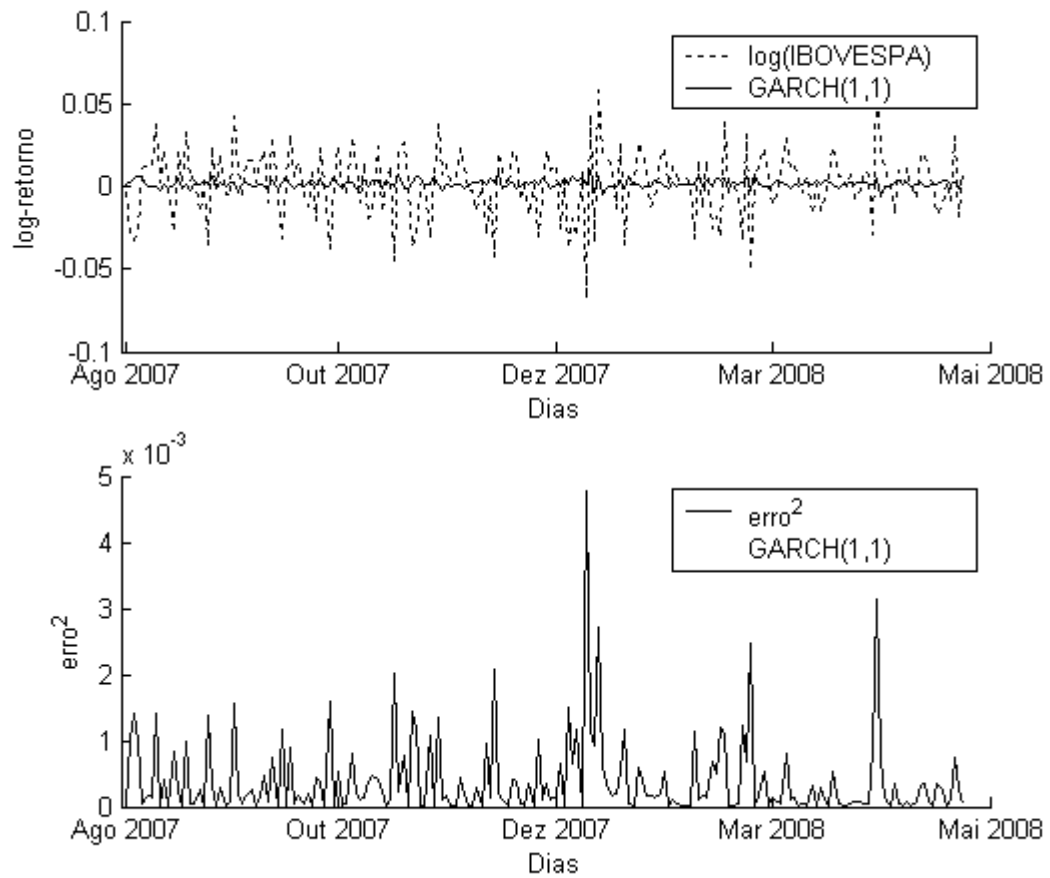


Figura 4.14. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método GARCH(1,1) em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, o método GARCH(1,1) obteve  $REQM=0,0206$ ,  $U\text{-Theil}=0,8479$ , com 74,47% de acertos dos movimentos ascendentes e 70,71% dos movimentos descendentes.

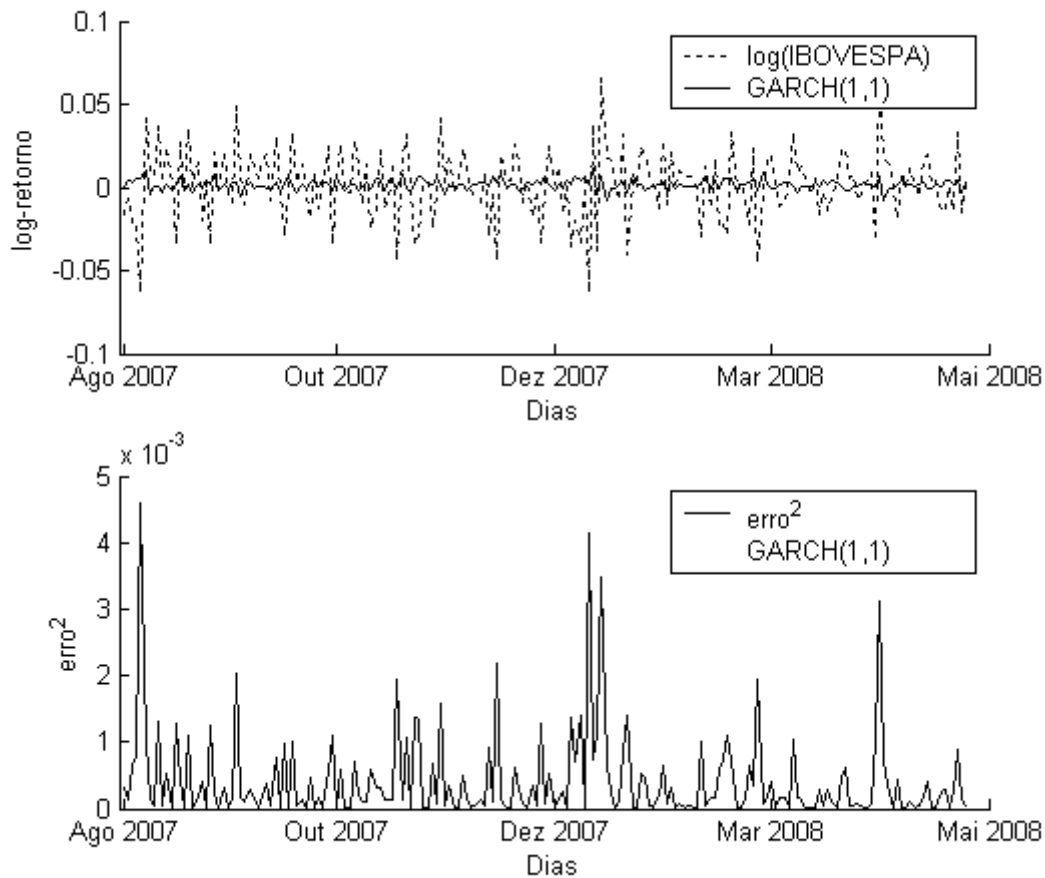


Figura 4.15. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método GARCH(1,1) em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

As Figuras 4.16 e 4.17, respectivamente, mostram os gráficos da previsão obtidos pela RNA de Elman treinada com o método quase-Newton to tipo BFGS. Na série do IBOVESPA à vista, obteve-se  $REQM=0,0203$ ,  $U\text{-Theil}=0,8566$ , com 79,12% de previsão correta no movimento ascendente, 69,31% de previsões corretas do movimento descendente dos retornos.

É possível observar que o método tende a penalizar a previsão posterior a um retorno negativo e, mesmo na presença de retorno posterior positivo, percebe-se uma previsão negativa. A amplitude da previsão acompanha o retorno positivo, mas, é realizada no sentido contrário. Esse efeito é muito evidente próximo ao início do mês de Mai/2008 quando o Brasil recebeu classificação de grau de investimento e a previsão do retorno apresenta uma amplitude significativa, mas na direção negativa.

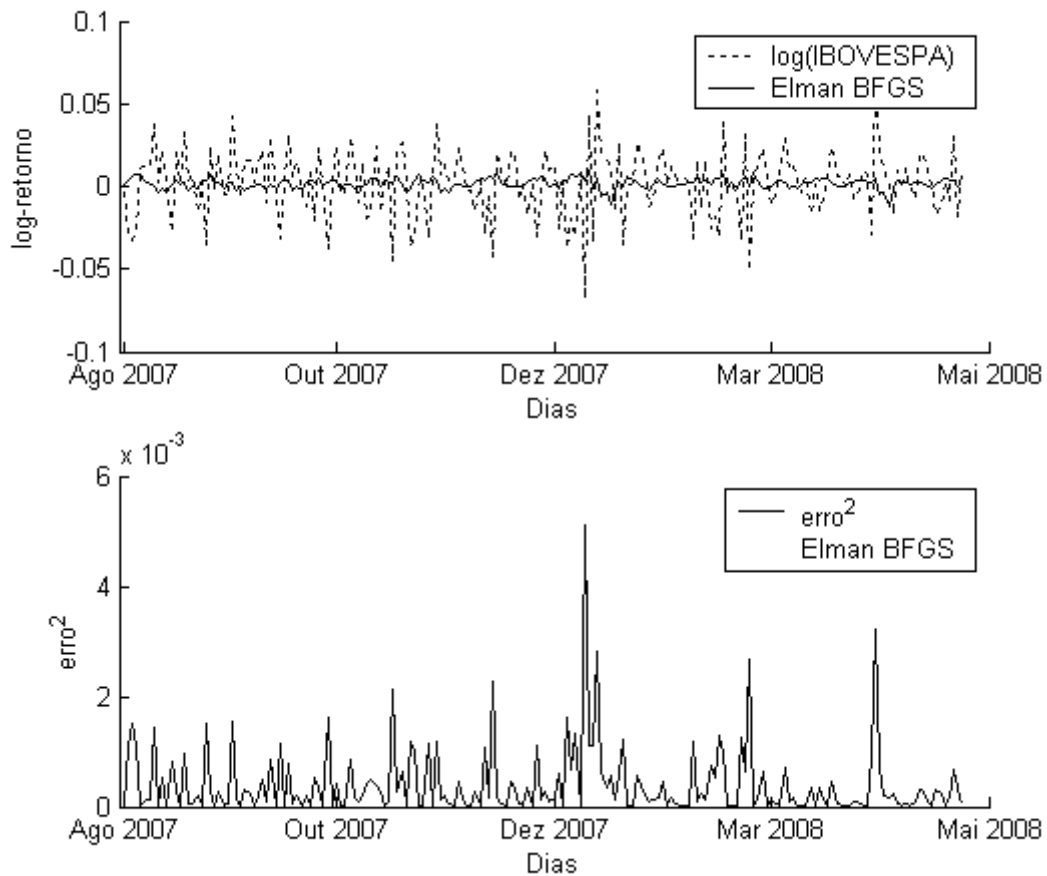


Figura 4.16. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Na série temporal do IBOVESPA futuro, a RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS obteve  $REQM=0,0206$  e estatística  $U\text{-Theil}=0,8398$ . Na previsão da direção o método obteve desempenho de 72,34% nas previsões de movimentos ascendentes e 71,72%, nos movimentos descendentes. Neste caso, apesar de ter um valor  $U\text{-Theil}$  menor, o percentual de previsões corretas, do movimento ascendente, foi inferior ao método  $GARCH(1,1)$ .

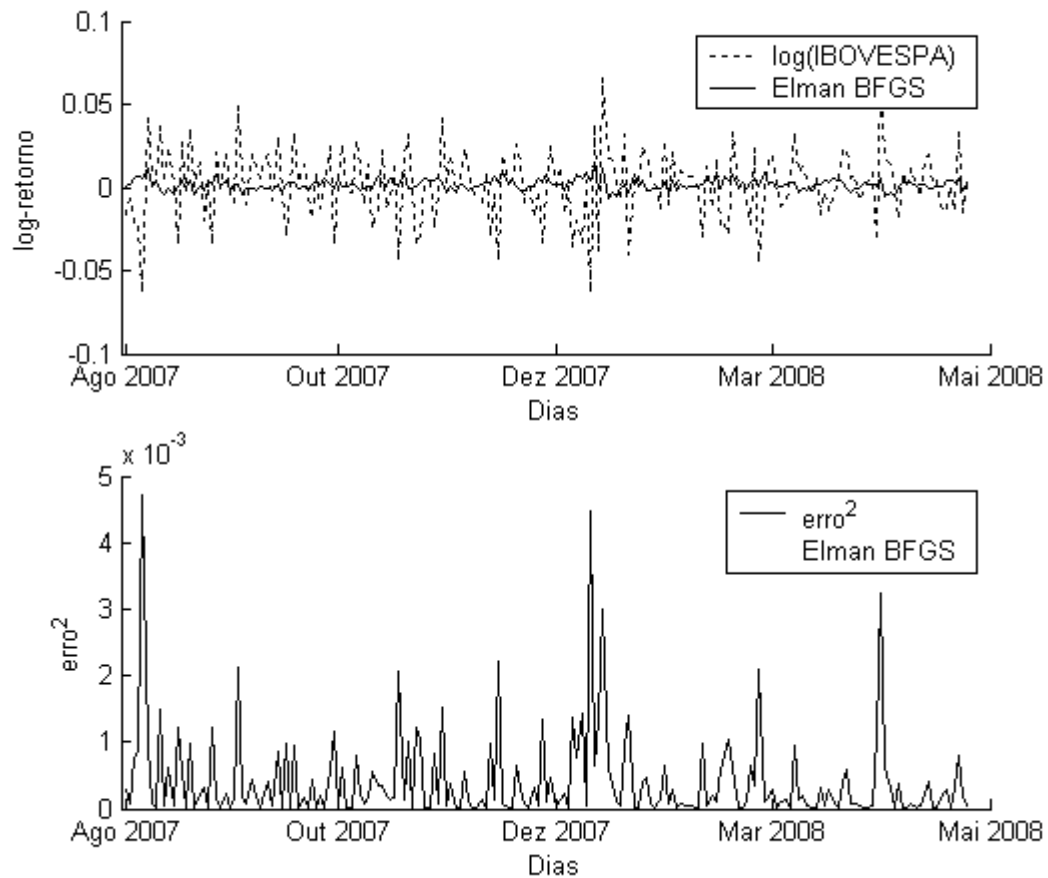


Figura 4.17. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS em relação à série de log-retornos do IBOVESPA futuro entre Jun/2004 e Mai/2008.

As Figuras 4.18 e 4.19 mostram os gráficos da previsão obtidos pela RNA de Elman, treinada com o algoritmo de ED *rand/1/bin*. Para a série temporal do IBOVESPA à vista, o REQM obtido foi de 0,0202 e coeficiente U-Theil=0,8694. Na previsão da direção, o método obteve desempenho 75,82% nas previsões de movimentos ascendentes e 69,31%, nos movimentos descendentes.

Quando comparado ao método GARCH(1,1), a RNA de Elman treinada com o algoritmo ED *rand/1/bin* obteve valor de REQM superior. Em relação ao coeficiente U-Theil, o método GARCH(1,1) obteve valor inferior a RNA de Elman treinada como algoritmo ED *rand/1/bin*. Com base nas afirmações anteriores, é possível concluir que a RNA de Elman treinada com o algoritmo ED *rand/1/bin* não possui capacidade de previsão superior ao método GARCH(1,1).

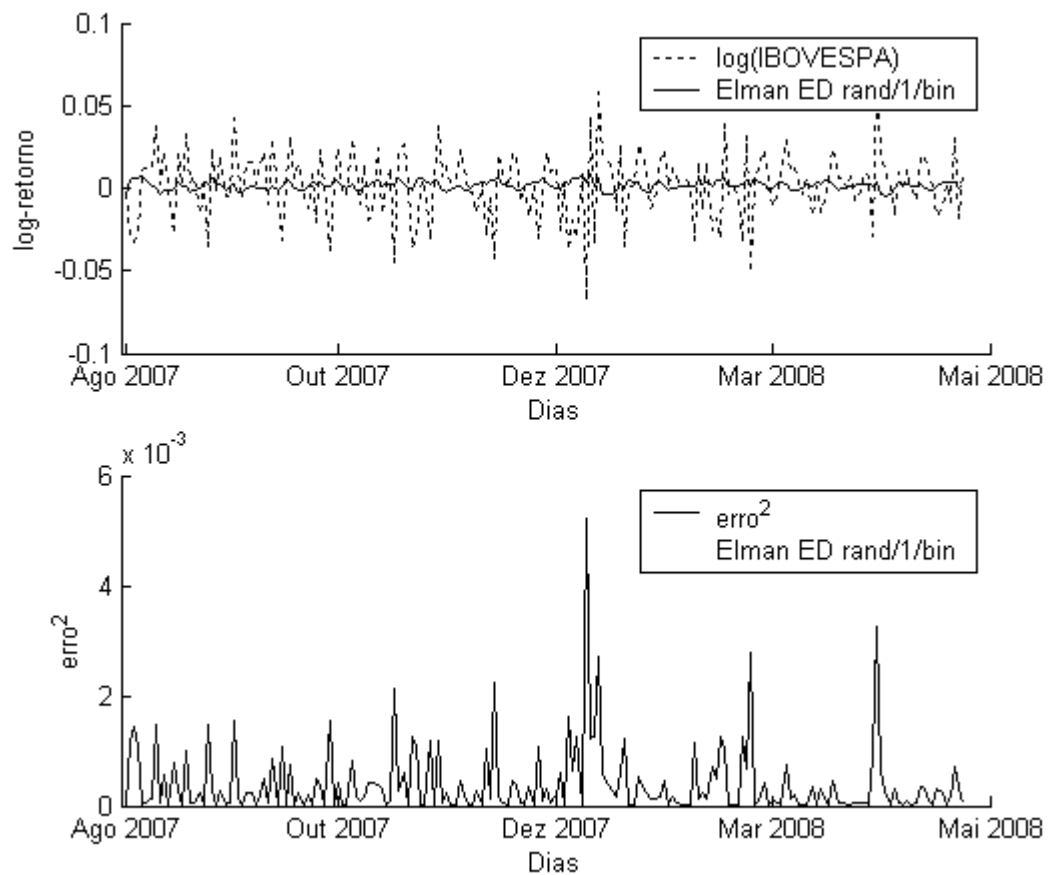


Figura 4.18. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED rand/1/bin em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro o método obteve  $REQM=0,0206$ ,  $U\text{-Theil}=0,8465$ , com percentual de 72,34% em acerto de previsão dos movimentos ascendentes e 72,73% de acertos nos movimentos descendentes.

O desempenho do coeficiente  $U\text{-Theil}$  é similar ao método  $GARCH(1,1)$ . Na comparação dos resultados do coeficiente  $U\text{-Theil}$ , para a série temporal do IBOVESPA futuro, com o método  $GARCH(1,1)$ , é possível perceber uma capacidade de previsão superior, mesmo que não significativa, da RNA de Elman treinada com o algoritmo *rand/1/bin*.



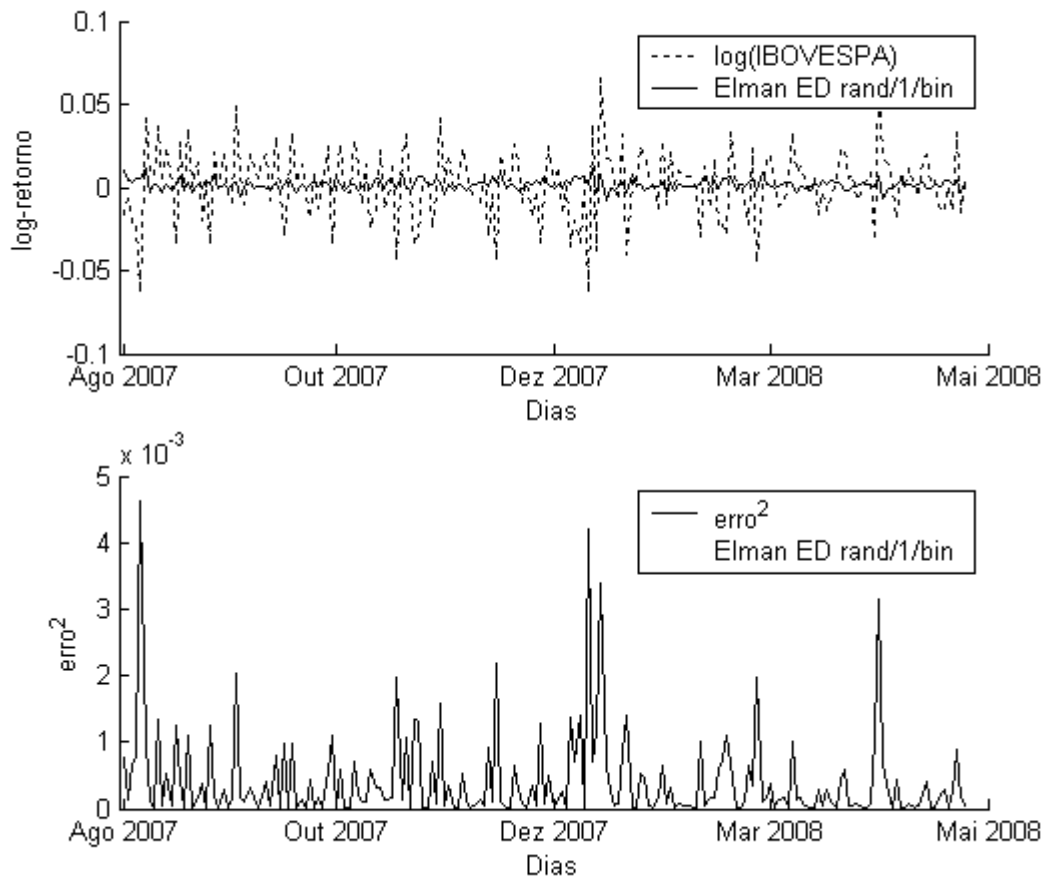


Figura 4.19. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado ED rand/1/bin em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

As Figura 4.20 e 4.21 mostram os gráficos da previsão obtidos pela RNA de Elman, treinada com o algoritmo de ED *best/1/bin*. Para IBOVESPA à vista, o REQM obteve o desempenho de 0,0200 e estatística U-Theil=0,8432. Na previsão da direção, o método obteve desempenho de 78,02% nas previsões de movimentos ascendentes e 62,38% nos movimentos descendentes.

Os valores de REQM e U-Theil que a RNA de Elman treinada com o algoritmo *best/1/bin* obteve para a série do IBOVESPA à vista, foram os menores, portanto melhores, quando comparados aos obtidos com o método GARCH(1,1), a RNA treinada com algoritmo quase-Newton do tipo BFGS e a RNA treinada com o algoritmo *rand/1/bin*.

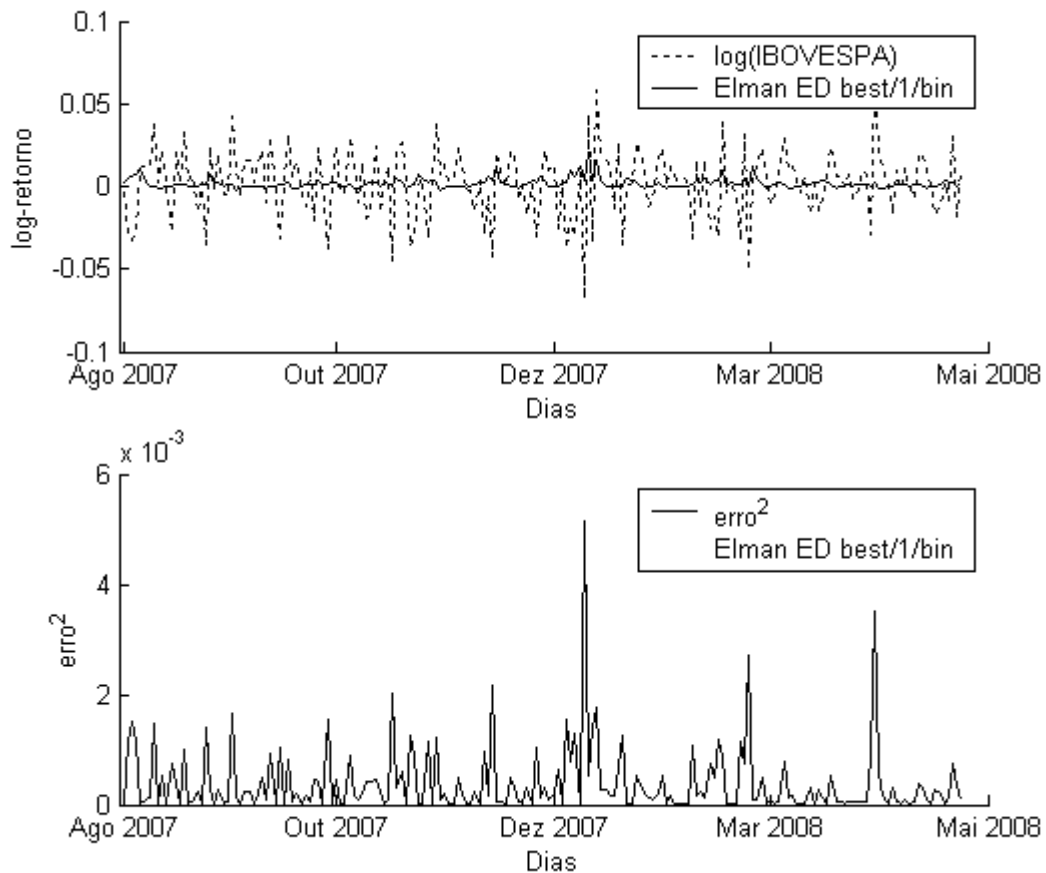


Figura 4.20. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED best/1/bin em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, o método obteve  $REQM=0,0201$ ,  $U\text{-Theil}=0,7759$ , percentual de 65,96% em acerto de previsão dos movimentos ascendentes e 68,69% de acertos nos movimentos descendentes. O desempenho do coeficiente  $U\text{-Theil}$  e o  $REQM$  são os melhores obtidos na comparação com os outros métodos verificados neste trabalho.

Ainda considerando a análise da série temporal do IBOVESPA futuro, a RNA de Elman treinada com o algoritmo *best/1/bin* obteve os menores percentuais de acertos, tanto nos movimentos ascendentes quanto nos descendentes, quando comparados com os obtidos pelos outros métodos pesquisados. Entretanto, o coeficiente  $U\text{-Theil}$ , que é a estatística utilizada neste trabalho para comparação dos métodos, é o menor obtido ente os métodos comparados e por isso, a RNA de Elman treinada com o algoritmo *best/1/bin*, tem a maior capacidade de previsão para série temporal do IBOVESPA futuro.

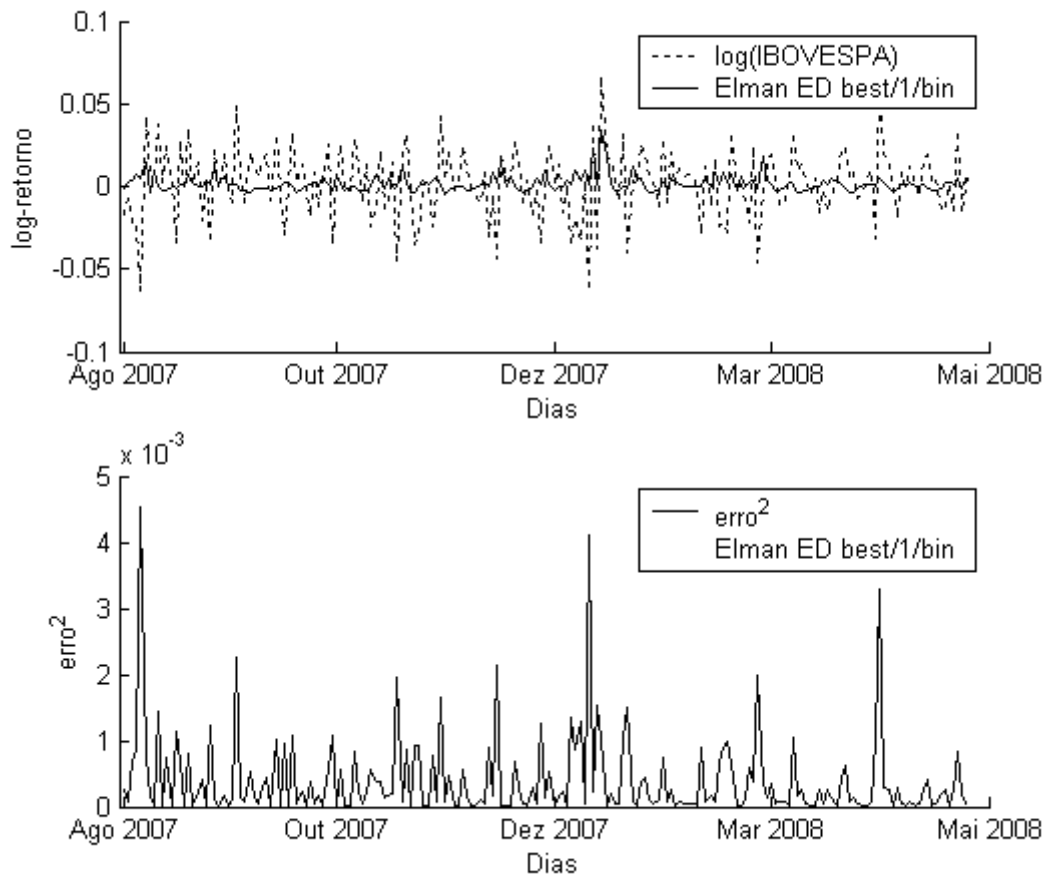


Figura 4.21. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED best/1/bin em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

A Figura 4.22 e 4.23, respectivamente, mostram os gráficos da previsão obtidos pela RNA de Elman, treinada com o algoritmo quase-Newton do tipo BFGS e ED *rand/1/bin*. Para IBOVESPA à vista, o REQM obteve o desempenho de 0,0200 e estatística U-Theil de 0,8469. Na previsão da direção, o método obteve desempenho de 75,82% nas previsões de movimentos ascendentes e 68,32% nos movimentos descendentes.

As previsões do IBOVESPA à vista obtidas pela RNA de Elman treinada com a combinação do algoritmo quase-Newton do tipo BFGS e ED *rand/1/bin*, apresenta coeficiente U-Theil inferior aos coeficientes obtidos pelas RNAs de Elman treinadas com os mesmos métodos utilizados de forma separada. Portanto, verificou-se um melhor ajustamento na utilização deste método híbrido se comparado com as RNAs treinadas com os algoritmos de forma separada.

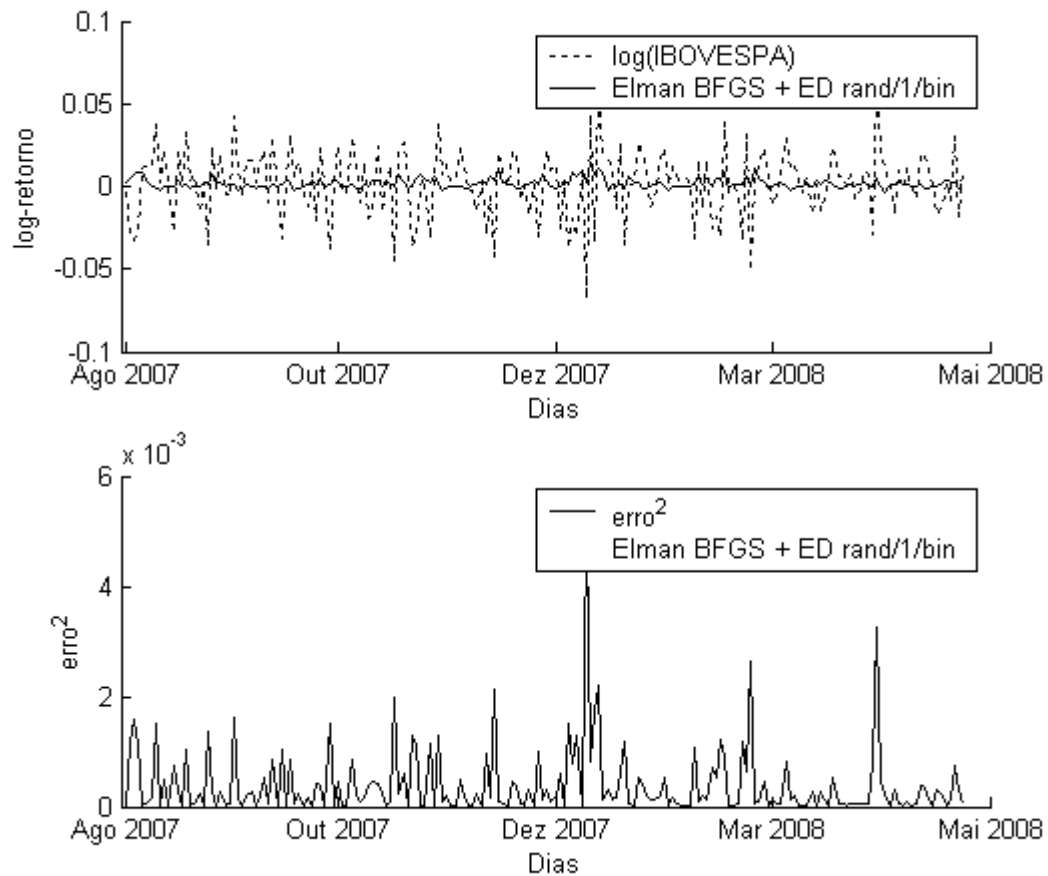


Figura 4.22. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *rand/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, a RNA de Elman combinando o algoritmo quase-Newton do tipo BFGS e ED *rand/1/bin* obteve REQM de 0,0208, U-Theil de 0,8449, com percentual de 75,53% em acertos de previsão dos movimentos ascendentes e 70,71% de acertos nos movimentos descendentes. O REQM de 0,0208 é o maior obtido para a série do IBOVESPA futuro, portanto mostra-se como o de menor desempenho em relação a esta medida. Entretanto, comparando o coeficiente de U-Theil obtido pela RNA de Elman combinando o algoritmo quase-Newton do tipo BFGS e ED *rand/1/bin*, medida de referência para esta pesquisa, verificou-se um ajustamento similar aos obtidos pelos outros métodos.

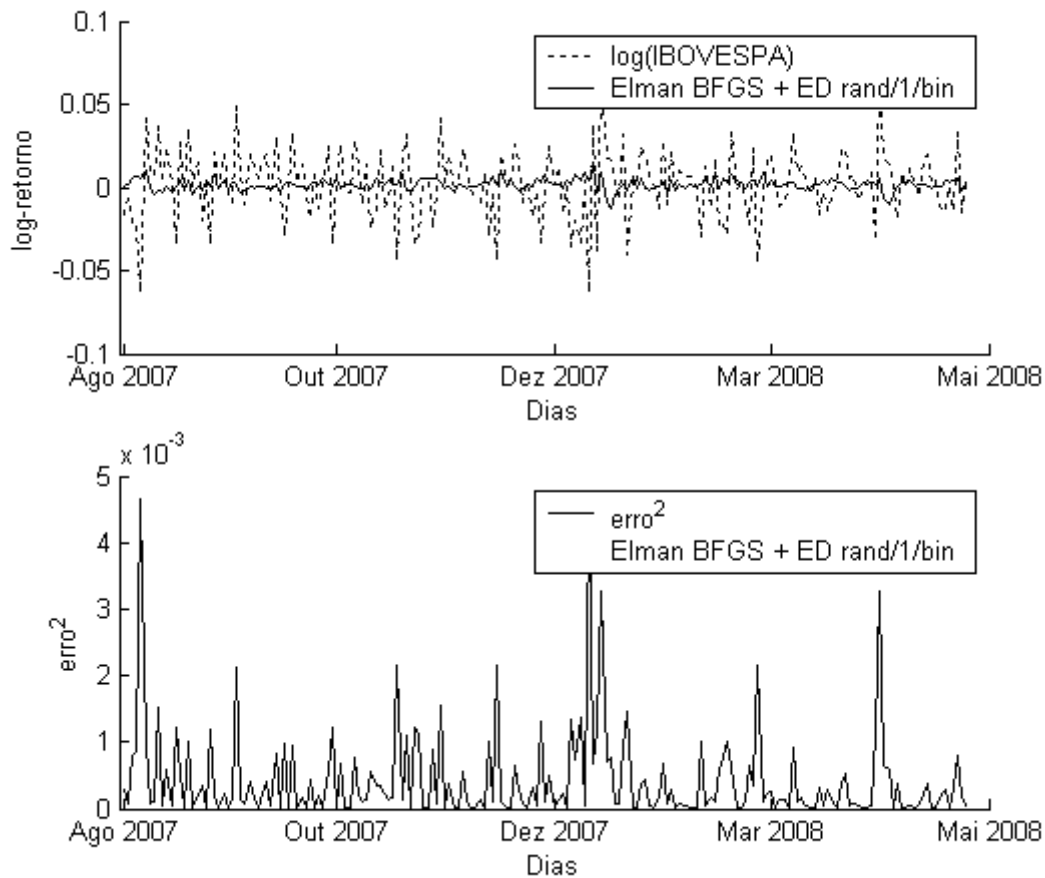


Figura 4.23. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED rand/1/bin em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

As Figuras 4.24 e 4.25 mostram os gráficos da previsão obtidos pela RNA de Elman, treinada com o algoritmo de quase-Newton do tipo BFGS e ED *best/1/bin*. Para IBOVESPA à vista, o REQM obteve o desempenho de 0,0203 e estatística U-Theil 0,8474. Na previsão da direção, o método obteve desempenho de 71,43% nas previsões de movimentos ascendentes e 63,37% nos movimentos descendentes.

Apesar de ter superado o resultado do coeficiente U-Theil do método GARCH(1,1) e da RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS, a RNA de Elman combinando o algoritmo quase-Newton do tipo BFGS e ED *best/1/bin* não supera a RNA de Elman treinada com o algoritmo *best/1/bin* utilizado de forma separada.

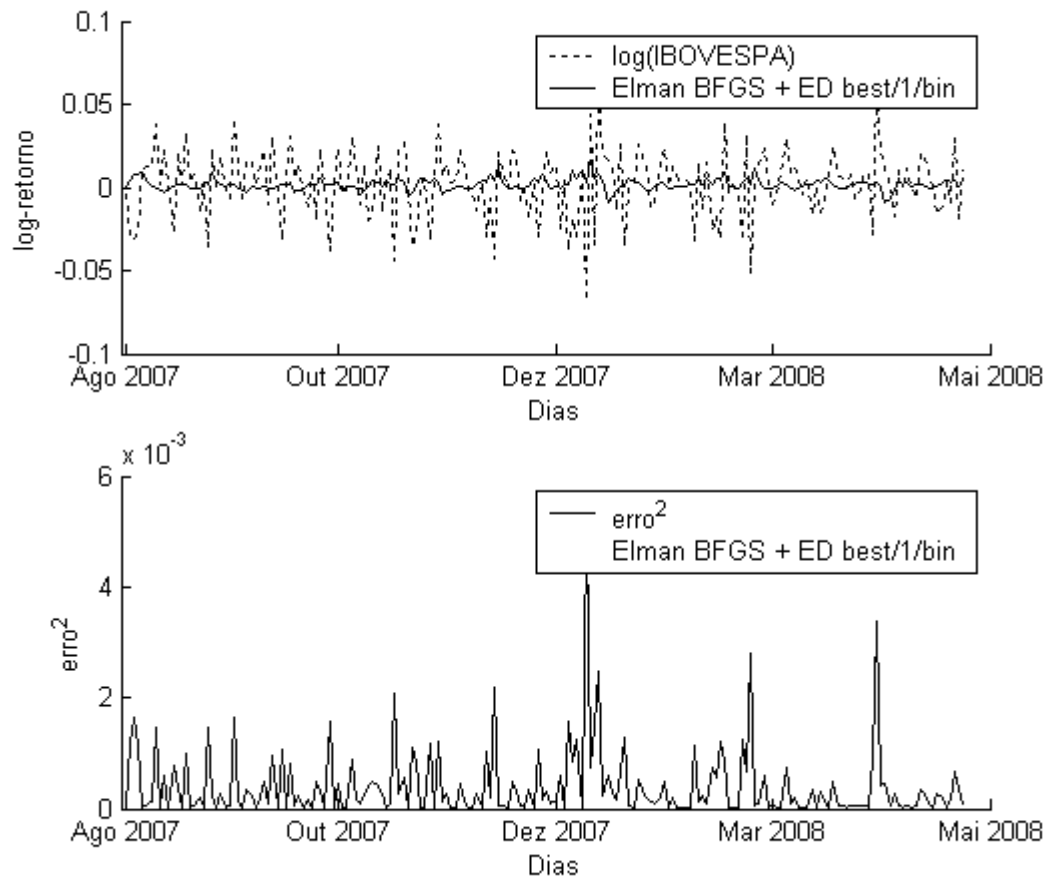


Figura 4.24. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED *best/1/bin* em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, o método obteve  $REQM=0,0205$ ,  $U\text{-Theil}=0,8342$ , percentual de 73,40% em acerto de previsão dos movimentos ascendentes e 69,70% de acertos nos movimentos descendentes. Comparando o resultado do coeficiente de U-Theil obtido pela RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS e ED *best/1/bin*, verificou-se sua superioridade em relação ao método GARCH(1,1) e a RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS.

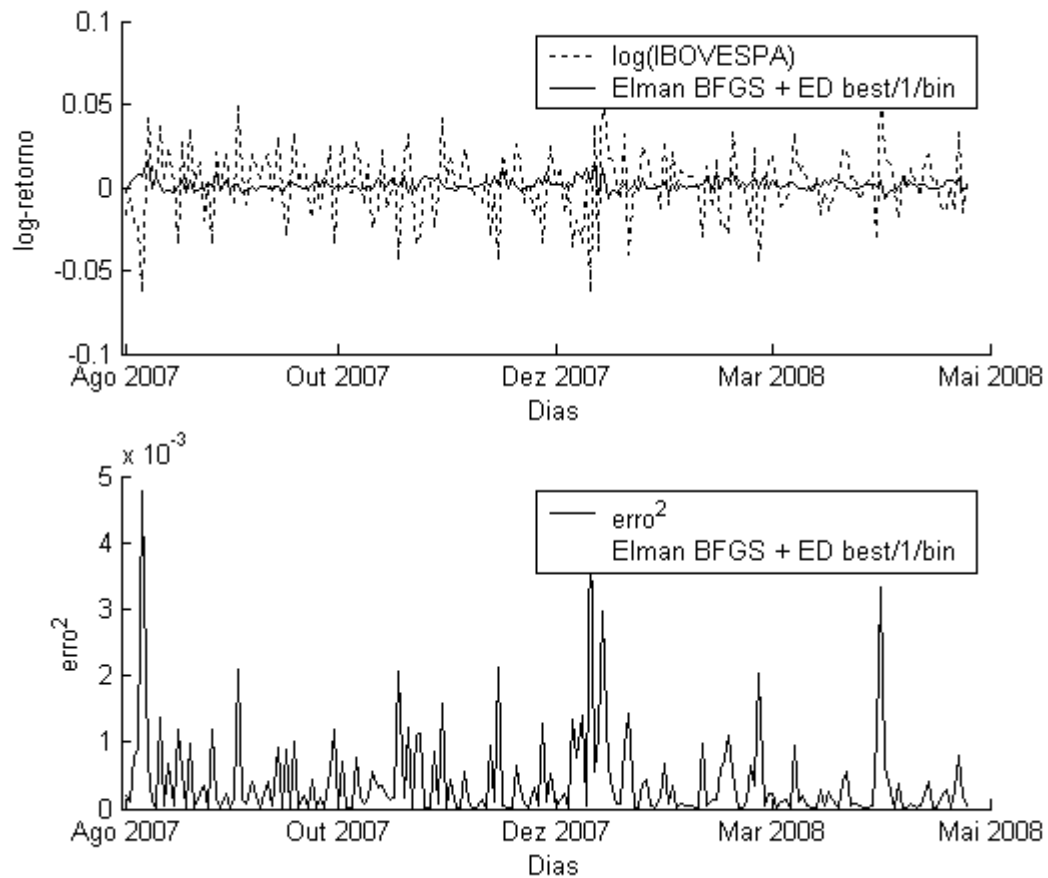


Figura 4.25. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a BFGS e ED best/1/bin em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

As Figuras 4.26 e 4.27, respectivamente, mostram os gráficos da previsão obtidos pela RNA de Elman treinada com o algoritmo ED *rand/1/bin* e quase-Newton do tipo BFGS. Para IBOVESPA à vista, o REQM obteve o desempenho de 0,0201 e estatística U-Theil 0,8628. Na previsão da direção, o método obteve desempenho 78,02% nas previsões de movimentos ascendentes e 69,31% nos movimentos descendentes.

A RNA de Elman treinada com o algoritmo ED *rand/1/bin* e quase-Newton do tipo BFGS supera o coeficiente de U-Theil obtido pela RNA de Elman treinada pelo algoritmo *rand/1/bin*, mas, não supera a RNA de Elman treinada pelo algoritmo quase-Newton do tipo BFGS. Em relação ao GARCH(1,1), o método obteve desempenho similar na medida do coeficiente U-Theil mas se mostrou superior no percentual de acertos de movimento ascendentes e descendentes.

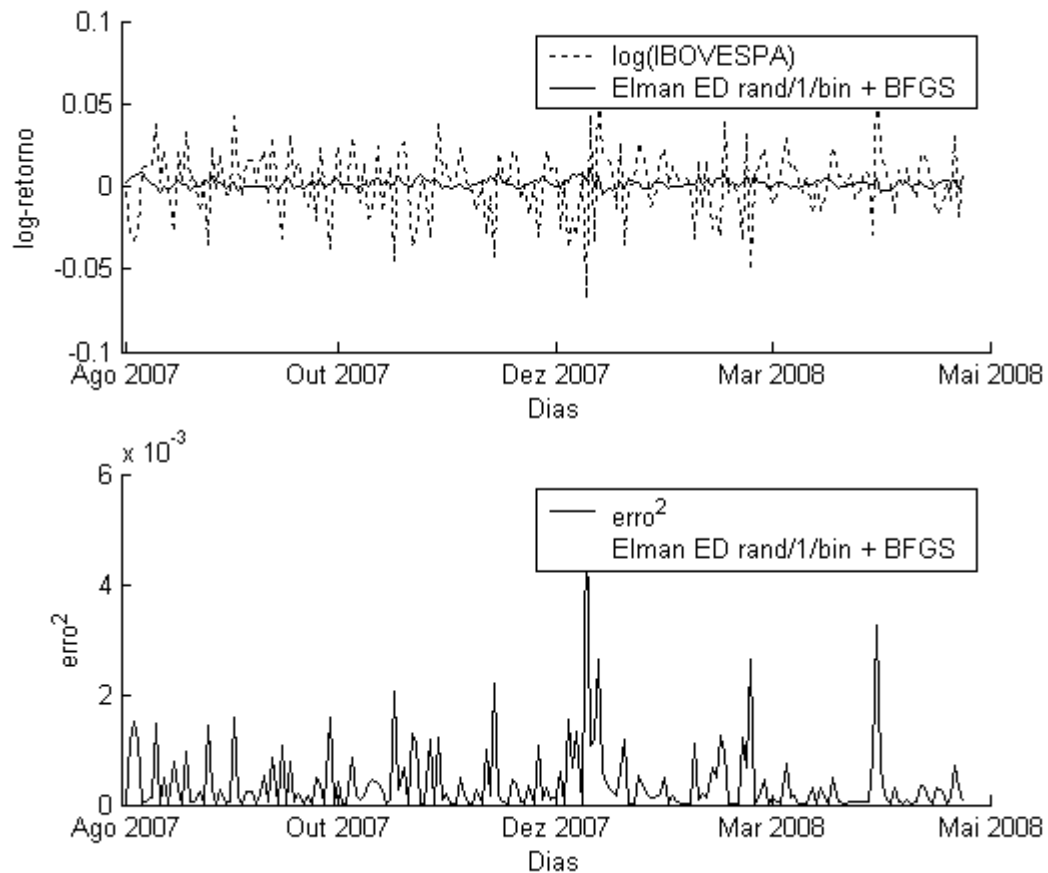


Figura 4.26. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, a RNA de Elman treinada com o algoritmo ED *rand/1/bin* e quase-Newton do tipo BFGS obteve REQM de 0,0206, U-Theil de 0,8455, percentual de 75,53% em acerto de previsão dos movimentos ascendentes e 67,68% de acertos nos movimentos descendentes. Comparando o resultado do coeficiente de U-Theil obtido, verificou-se a superioridade, não significativa, do método em relação ao GARCH(1,1) mas inferior quando comparado a RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS.



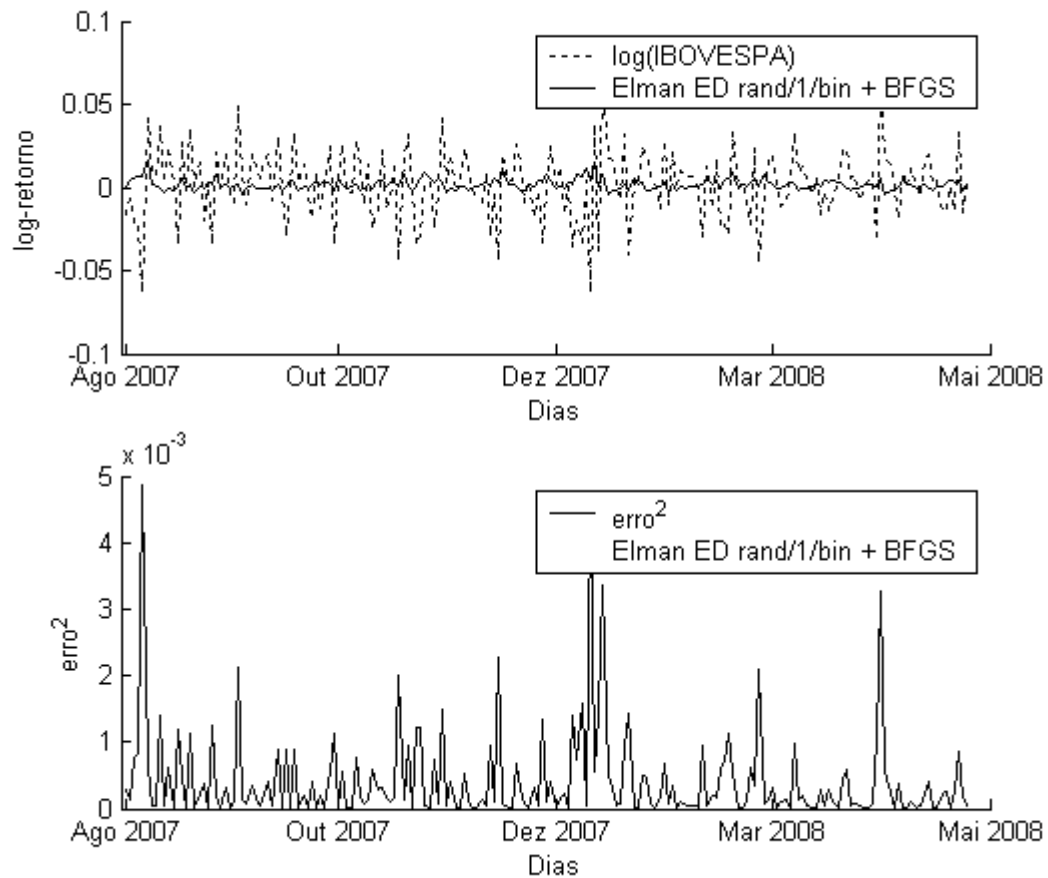


Figura 4.27. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *rand/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

As Figuras 4.28 e 4.29 mostram os gráficos da previsão obtidos pela RNA de Elman, treinada com o algoritmo ED *best/1/bin* e quase-Newton do tipo BFGS. Para a séries de retornos do IBOVESPA à vista, o REQM obteve o desempenho de 0,0201 e estatística U-Theil=0,8617. Na previsão da direção, o método obteve desempenho de 79,12% nas previsões de movimentos ascendentes e 63,37% nos movimentos descendentes.

A RNA de Elman combinando ED *rand/1/bin* e quase-Newton do tipo BFGS supera o coeficiente de U-Theil obtido pela RNA de Elman treinada com o algoritmo ED *rand/1/bin*, mas, não supera a RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS quando utilizados individualmente no treinamento das RNAs. Em relação ao GARCH(1,1), o método obteve desempenho similar na medida do coeficiente U-Theil mas se mostrou superior no percentual de acertos de movimento ascendentes e descendentes.

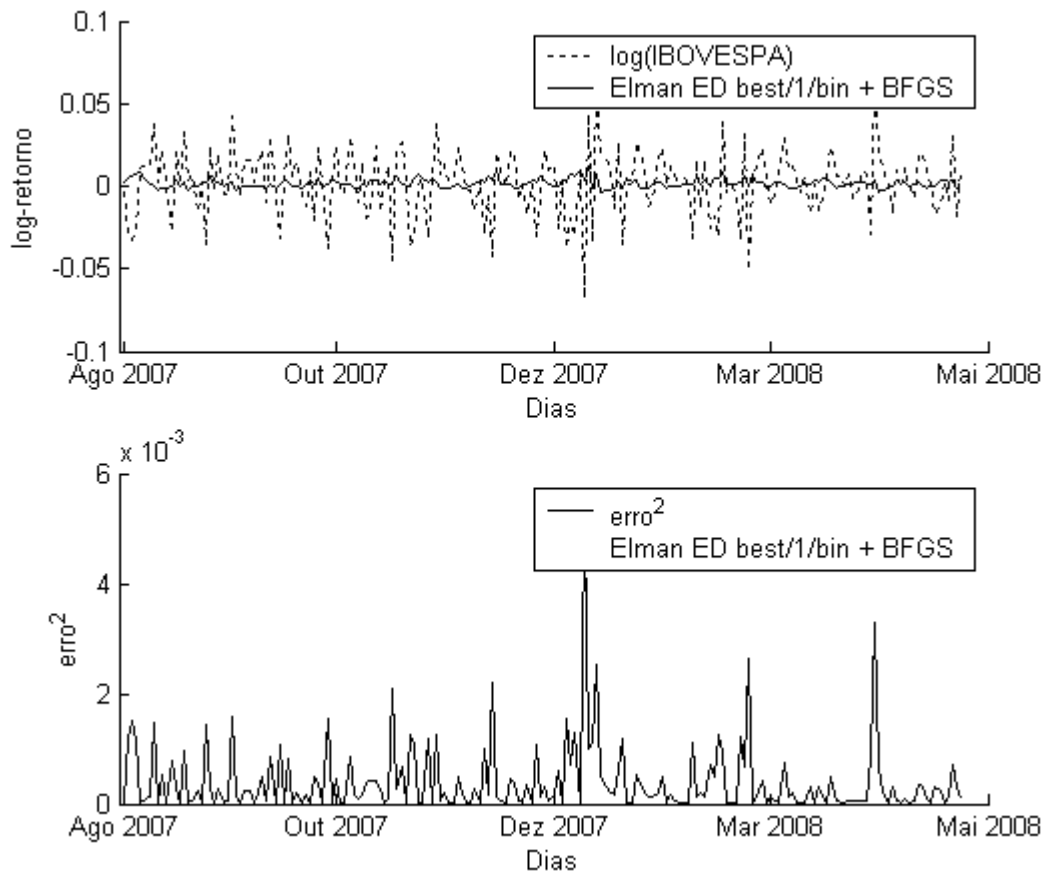


Figura 4.28. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008.

Para o IBOVESPA futuro, a RNA de Elman combinando ED *best/1/bin* e quase-Newton do tipo BFGS obteve REQM de 0,0206, U-Theil de 0,8369, com percentual de 75,53% em acerto de previsão dos movimentos ascendentes e 71,72% de acertos nos movimentos descendentes. Comparando o resultado do coeficiente U-Theil obtido, verificou-se a superioridade, não significativa, deste método em relação ao GARCH(1,1) e a RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS, mas inferior quando comparado a RNA de Elman treinada com o algoritmo ED *best/1/bin*.

O REQM do GARCH(1,1) é similar àqueles métodos que combinam RNA de Elman com quase-Newton do tipo BFGS, ED *rand/1/bin*, ED *rand/1/bin* com quase-Newton do tipo BFGS, e ED *best/1/bin* com quase-Newton do tipo BFGS. As exceções são os métodos ED

*best/1/bin*, e quase-Newton do tipo BFGS combinado ao ED *best/1/bin*, que obtiveram um REQM menor do que o método GARCH(1,1).

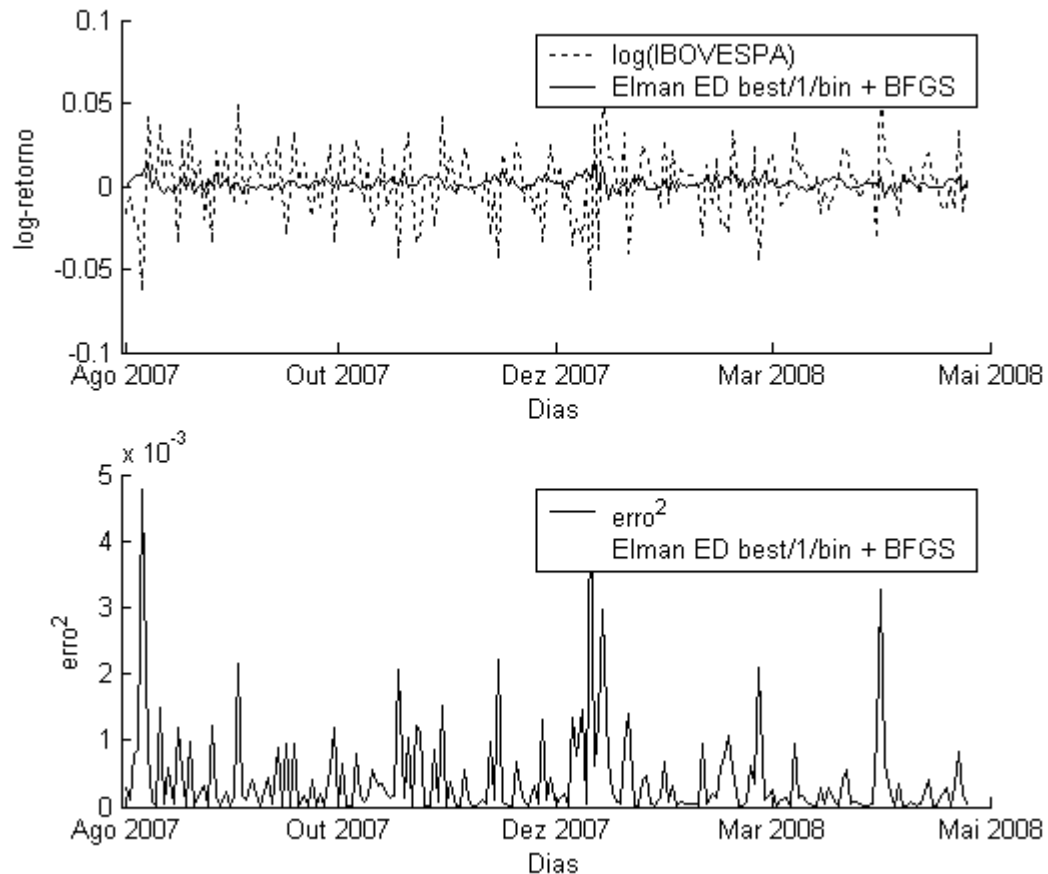


Figura 4.29. Previsão uma passo à frente e sinal de erro, ao quadrado, gerados pelo método Elman combinado a ED *best/1/bin* e BFGS em relação à série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008.

A Tabela 4.5 sumariza as medidas de desempenho obtidas com base nas previsões do IBOVESPA à vista, um passo à frente, de cada um dos modelos apresentados na Figura 4.13.

De uma forma geral, os REQMs estão nos limites de 0,0200 e 0,0203, o que mostra um desempenho similar entre todos os métodos analisados. A confirmação desta similaridade é observada no coeficiente U-Theil, que também tem uma faixa de valores entre 0,8432 e 0,8661. O método GARCH(1,1) só não foi superado pela RNA treinada com o método ED *rand/1/bin* quando comparados os coeficientes U-Theil.

A RNA de Elman treinada com o algoritmo quase-Newton do tipo BFGS e o método híbrido combinando o algoritmo ED *best/1/bin* e quase-Newton do tipo BFGS obtiveram o maior número de acertos dos movimentos ascendentes. As RNAs treinadas com o algoritmo quase-Newton do tipo BFGS, ED *rand/1/bin* e o método híbrido combinando o algoritmo ED *rand/1/bin* e quase-Newton do tipo BFGS, obtiveram mais acertos no descendentes.

Tabela 4.5. Medidas de desempenho da série de log-retornos do IBOVESPA à vista entre Ago/2007 e Mai/2008

Modelo	REQM E.	REQM V.	U-Theil	Mov. Asc.	Mov. Desc.	Média Erro
GARCH(1,1)	0,0151	0,0200	0,8661	71,43%	63,37%	0,0001
Elman BFGS	0,0151	0,0203	0,8566	79,12%	69,31%	0,0008
Elman ED(1)	0,0151	0,0202	0,8694	75,82%	69,31%	0,0079
Elman ED(2)	0,0151	0,0200	0,8432	78,02%	62,38%	-0,0001
Elman BFGS + ED(1)	0,0151	0,0200	0,8469	75,82%	68,32%	-0,0000
Elman BFGS + ED(2)	0,0150	0,0203	0,8474	71,43%	63,37%	0,0004
Elman ED(1) + BFGS	0,0151	0,0201	0,8628	78,02%	69,31%	0,0004
Elman ED(2) + BFGS	0,0151	0,0201	0,8617	79,12%	63,37%	0,0004

A Tabela 4.6 apresenta as medidas de desempenho obtidas com base nas previsões do IBOVESPA futuro, um passo à frente, de cada um dos modelos apresentados na Figura 4.13.

Para o IBOVESPA futuro, a faixa de REQM está entre 0,0201 e 0,0206, o que mostra um equilíbrio entre os métodos de previsão no que diz respeito a sua capacidade de previsão. A RNA treinada com o algoritmo ED *best/1/bin* obteve o melhor desempenho quanto ao REQM e o coeficiente U-Theil. Os métodos híbridos também obtiveram melhor desempenho em relação ao coeficiente U-Theil, quando comparados ao modelo GARCH(1,1).

Tabela 4.6. Medidas de desempenho da série de log-retornos do IBOVESPA futuro entre Ago/2007 e Mai/2008

Modelo	REQM E.	REQM V.	U-Theil	Mov. Asc.	Mov. Desc.	Média Erro
GARCH(1,1)	0,0161	0,0206	0,8479	74,47%	70,71%	0,0000
Elman BFGS	0,0160	0,0206	0,8398	72,34%	71,72%	0,0004
Elman ED(1)	0,0160	0,0206	0,8465	72,34%	72,73%	-0,0098
Elman ED(2)	0,0159	0,0201	0,7759	65,96%	68,69%	-0,0003
Elman BFGS + ED(1)	0,0160	0,0208	0,8449	75,53%	70,71%	0,0006
Elman BFGS + ED(2)	0,0160	0,0205	0,8342	73,40%	69,70%	0,0004
Elman ED(1) + BFGS	0,0160	0,0206	0,8455	75,53%	67,68%	0,0002
Elman ED(2) + BFGS	0,0160	0,0206	0,8369	75,53%	71,72%	0,0003

## 5. CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Neste capítulo, serão apresentadas as conclusões deste trabalho, analisando o desempenho das RNAs para a previsão, um passo à frente, do IBOVESPA, à vista e futuro, nos testes executados. Apresentam-se também os possíveis trabalhos futuros desta dissertação.

### 5.1 CONCLUSÕES

No mercado de ações é necessário se antecipar aos acontecimentos para que seja possível se obter lucros nas transações. Partindo-se do pressuposto que eventos passados podem explicar o futuro, é necessária a procura de métodos que capturem características do comportamento passado a fim de prever o próximo acontecimento.

O objetivo deste trabalho foi a comparação entre métodos baseados em RNAs de Elman treinadas com o algoritmo ED *rand/1/bin*, ED *best/1/bin*, quase-Newton do tipo BFGS e o treinamento híbrido combinando quase-Newton do tipo BFGS e os métodos ED. Os métodos híbridos foram combinados de forma a executar buscas globais com os algoritmos ED seguidos de uma busca local por quase-Newton do tipo BFGS ou busca inicial por quase-Newton do tipo BFGS seguida por uma busca com um dos métodos ED. Dessa forma foi possível verificar se existe vantagem na busca quando da utilização de um método híbrido em relação à aplicação do método puro e também medir o desempenho dos métodos baseados em RNA de Elman em relação ao modelo GARCH(1,1).

Nesta pesquisa, o algoritmo de evolução diferencial adaptativo (SALMAN *et al.*, 2007) foi utilizado, sozinho e combinado, com o algoritmo de descida de encosta, quase-Newton do tipo BFGS, para o treinamento de RNAs de Elman, com o objetivo de previsão dos retornos, um passo à frente, do IBOVESPA, à vista e futuro, no período de Ago/2007 e Mai/2008.

Um modelo GARCH(1,1) foi ajustado para servir como referência clássica uma vez que esta técnica é amplamente utilizada no mercado financeiro para a modelagem da volatilidade e previsão de retornos. O coeficiente U-Theil foi a medida de desempenho utilizada para a comparação dos modelos, visto que esta procura avaliar o ajustamento da série prevista em relação a série realizada. O treinamento da RNA de Elman com o algoritmo ED *best/1/bin* obteve o menor valor para o coeficiente U-Theil e foi validado como o de melhor desempenho quando comparado aos outros métodos de treinamento de RNA testados e ao modelo GARCH(1,1).

Na comparação entre o menor e o maior coeficiente de U-Theil, foi observado um pequeno intervalo de valores, o que caracteriza o equilíbrio e similar capacidade de previsão das técnicas verificadas. O treinamento das RNAs com os métodos híbridos, combinando algoritmo ED e quase-Newton do tipo BFGS, não supera significativamente os resultados obtidos com os métodos utilizados separadamente. Os métodos híbridos mostraram-se melhores na previsão da direção dos retornos e, de forma consistente, obtiveram percentuais de acertos de direção, melhores do que os demais.

O modelo GARCH(1,1) obteve melhores percentuais de acertos após ter os resíduos de previsão ajustados por uma modelo de regressão linear considerando um atraso. Visto isso, é possível afirmar que, se a informação do gradiente estiver disponível, ela deve ser utilizada. De forma correspondente, se a superfície de erro for imperfeita e a informação do gradiente consistentemente levar o processo de busca a ficar preso a um mínimo local, as abordagens estocásticas e os métodos de otimização global podem se tornar a única alternativa. A comparação do REQM das previsões obtidas com as RNAs de Elman, considerando aquelas treinadas com algoritmo de descida de encosta, algoritmo ED e pelos métodos híbridos, e o modelo GARCH, mostrou uma pequena vantagem na utilização das RNAs.

A avaliação da previsão, um passo a frente, dos retornos do IBOVESPA, à vista e futuro, usando rede neural parcialmente recorrente de Elman, treinadas por métodos híbridos compostos por algoritmos de descida de encosta e evolução diferencial foi realizada e os objetivos desta dissertação foram alcançados. A hipótese de pesquisa foi aceita uma vez que os métodos de previsão baseados em RNAs, de uma forma geral, obtiveram melhores resultados, em relação ao coeficiente U-Theil, do que a previsão realizada pelo método

GARCH(1,1). O fato de prever os retornos do IBOVESPA, à vista ou futuro, leva a conclusão de que o mercado não é eficiente em sua forma fraca.

## 5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Uma pesquisa futura pode considerar a minimização da função de máxima verossimilhança, do modelo GARCH(1,1), por um algoritmo ED bem como uma RNA treinada pelo mesmo algoritmo para a modelagem dos resíduos da previsão. Outra possibilidade é a adição de outros parâmetros de entrada na RNA de Elman. Uma vez que a informação do IBOVESPA futuro está disponível, ela pode ser utilizada como parâmetro de entrada da RNA ou mesmo um coeficiente que represente uma relação entre os índices pode servir de informação adicional para a RNA. Uma terceira opção a ser validada é a utilização da saída atual da RNA como parâmetro de entrada da avaliação posterior. Outra possibilidade é a utilização da informação do erro anterior como entrada da RNA.

## REFERÊNCIAS

ABELÉM, A. J. G. **Redes Neurais Artificiais na Previsão de Séries Temporais**. Dissertação de Mestrado – Pontifícia Universidade Católica, Rio de Janeiro/RJ, Departamento de Engenharia Elétrica, Setembro 1994.

ALEXANDER, C. **Modelos de Mercados: um Guia para Análise de Informações Financeiras**. Tradução José Carlos de Souza Santos - São Paulo/SP: Bolsa de Mercadorias e Futuros, 2005.

ANDREESKI, C. J.; VASANT, P. M.; STANKOVSKI, M. J.; DIMIROVSKI, G. M. **Elman NN and Time Series in Forecasting Models for Decision Making**. World Automation Congress (WAC), Budapest, Hungary, p. 1-11, 2006.

BABU, B. V.; RAKESH, A. **Modified Differential Evolution (MDE) for Optimization of Non-Linear Chemical Processes**. Computers and Chemical Engineering, vol. 30, no. 6-7, p. 989-1002, 2006.

BABU, B. V.; MUNAWAR, S. A. **Differential Evolution Strategies for Optimal Design of Shell-and-Tube Heat Exchangers**. Chemical Engineering Science, vol. 62, no. 14, p. 3720-3739, 2007.

BARBOSA, A. **Análise de Demanda do Álcool utilizando os Métodos de Suavização Exponencial**. Monografia Bacharelado em Estatística, Universidade Estadual de Maringá, Maringá/PR, 2005.

BARRETO, A. M. S. **Algoritmo Genético dos Mínimos Quadrados Ortogonal para o Treinamento de Redes RFB**. Dissertação (Mestrado), Programa de Pós Graduação em Engenharia Civil, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia/Universidade Federal do Rio de Janeiro - Rio de Janeiro/RJ, 2003.

BÓDIS, L. **Financial Times Series Forecasting using Artificial Neural Networks**. Master Thesis – “Babes-Bolyai” University, Faculty of Mathematics and Computer Science – Department of Computer Science, Cluj-Napoca, Romania, 2004.

BRUNELLI, U.; PIAZZA, V.; PIGANTO, L.; SORBELLO, F.; VITABILE, S. **Three Hours ahead Prevision of SO<sub>2</sub> Pollutant Concentration using an Elman Neural based Forecaster**. Building and Environment, vol. 43, no. 3, p. 304-314, 2006.



BASTURK, A.; GUNAY, E. **Efficient Edge Detection in Digital Images Using a Cellular Neural Network Optimized by Differential Evolution Algorithm**. Expert Systems with Applications, In Press, doi:10.1016/j.eswa.2008.01.082, 2008.

CARVALHO, J. V. **Modelagem Temporal das Medidas de Vazão de Drenos na Barragem de Funil (RJ) utilizando Redes Neurais e Métodos Estatísticos**. Tese de Doutorado Pontifícia Universidade Católica / Rio de Janeiro – Departamento de Engenharia Civil – Rio de Janeiro/RJ, 2005.

CAVALHEIRO, D. **Método de Previsão de Demanda Aplicada ao Planejamento da Produção de Indústrias de Alimentos**. Dissertação de Mestrado, Universidade Federal de Santa Catarina – Departamento de Engenharia Mecânica – Florianópolis/SC, 2003.

CHANG, B. R. **Applying Nonlinear Generalized Autoregressive Conditional Heteroscedasticity to Compensate ANFIS Outputs Tuned by Adaptive Support Vector Regression**. Fuzzy Sets and Systems, vol. 157, no. 13, p. 1832-1850, 2006.

CHENG, C.; CHEN, C.; FU C. **Financial Distress Prediction by a Radial Basis Function Network with Logit Analysis Learning**. Computers and Mathematics with Applications, vol. 51, no. 3-4, p. 579-588, 2006.

CHENG, Y.; QI, W.; CAI, W. **Dynamic properties of Elman and modified Elman neural network**. Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, China, 2002.

CIARLINI, P.; MANISCALCO, U. **Wavelets and Elman Neural Networks for Monitoring Environmental Variables**. Journal of Computational and Applied Mathematics, doi: 10.1016/j.cam.2007.10.040, 2008.

CORTEZ, P. A. R. **Algoritmos Genéticos e Redes Neurais na Previsão de Séries Temporais**. Dissertação de Mestrado– Universidade do Minho, Braga, Portugal, 1997.

DU, J.; HUANG, D.; WANG, X.; GU, X. **Shape Recognition Based on Neural Networks Trained by Differential Evolution Algorithm**. Neurocomputing, vol. 70, no. 4-6, p. 896-903, 2007.

EHLERS, R. S. **Análise de Séries Temporais**, 3ª ed, Universidade Federal do Paraná, Curitiba/PR, 2005.

ELMAN, J. L. **Finding Structure in Time.** Cognitive Science, vol. 14, p.179-211, 1990.

ENGLE, R. **GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics.** Journal of Economics Perspective, vol. 15, no. 4, p. 157–168, 2001.

FANG, Y.; XU, D. **The Predictability of Asset Returns: An Approach Combining Technical Analysis and Time Series Forecasts.** International Journal of Forecasting, vol. 19, no. 13, p. 369-385, 2003.

FAMA, E. **Efficient Capital Markets: A Review of Theory and Empirical Work.** Journal of Finance, vol. 25, no. 2, p. 383-417, 1970.

GAO, X. Z.; GAO, X. M.; OVASKA, S. J. **A Modified Elman Neural Network Model with Application to Dynamical System Identification.** Proceedings of IEEE International Conference, Systems, Man, and Cybernetics, 1996, Beijing, China, vol. 2, p. 1376-1381, 1996.

GE, H.; LIANG, Y.; MARCHESE, M. **A Modified Particle Swarm Optimization-based Dynamic Recurrent Neural Network for Identifying and Controlling Nonlinear Systems.** Computers and Structures, vol. 85, no. 21-22, p. 1611-1622, 2007.

GONÇALVES, A. **Prognosticação Quantitativa Utilizando Técnicas de Alisamento Baseadas em Séries Temporais** – Monografia de Graduação em Engenharia de Produção – Universidade Federal de Ouro Preto – Departamento de Engenharia de Produção, Ouro Preto/MG, 2003.

GOOIJER, J. G. De; HYNDMAN, R. J. **25 Years of IIF Time Series Forecasting: A Selective Review** – Tinbergen Institute Discussion Paper, University of Amsterdam, Holanda – TI 068/04, 2005.

GROTHMANN, R. **Multi-Agent Market Modeling Based on Neural Networks.** Thesis Doctor of Economics – Faculty of Economics – University of Bremen, Munich, Germany, 2003.

HASSAN, Md. R.; BAIJUNTH, N.; KIRLEY, M. **A Fusion Model of HMM, ANN and GA for Stock Marketing Forecasting.** Expert Systems with Applications, vol. 33, no. 1, p.171-180, 2006.

HAYKIN, S. **Redes Neurais: Princípios e Prática.** 2ª ed., Porto Alegre/RS, Editora Bookman, 2001.

JARQUE C. M.; BERA, A. K. **A Test for Normality of Observations and Regression Residuals.** International Statistical Review, vol. 55, no. 2, p. 163-172, 1987.

KOSTELA, T. **Neural Network Methods In Analysing and Modelling Time Varying Processes.** Doctoral Thesis – Helsinki University of Technology – Department of Electrical and Communications Engineering, Finland, 2003.

LEE, S. W. **Off-line Recognition of Totally Unconstrained Handwritten Numerals using Multilayer Cluster Neural Network.** IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 6, p. 648-652, 1996.

LIMA, F. G. **Um Método de Análise e Previsão de Sucessões Cronológicas Unidimensionais Lineares e Não-Lineares.** Tese de Doutorado - Universidade de São Paulo, São Paulo/SP, 2004.

LIN, C.; HUANG, J.; GEN, M.; TZENG, G. **Recurrent Neural Network for Dynamic Portfolio Selection.** Applied Mathematics and Computation, vol. 175, no. 2, p. 1139-1146, 2006.

LUNA, S. V. **Planejamento de Pesquisa: Uma Introdução.** São Paulo: Educ, 1997.

LUX, T.; KAIZOJI, T. **Forecasting Volatility and Volume in the Tokyo Stock Market: Long Memory, Fractality and Regime Switching.** Journal of Economic Dynamics & Control, vol. 31, no. 6, p. 1808-1843, 2007.

MARTIN, C. **Prognóstico de Demanda de Potência Elétrica para Planejamento e Operação de Sistemas Elétricos.** Dissertação de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis/SC, 2005.

MAYER, D. G.; KINGHORN, B. P.; ARCHER, A. A. **Differential Evolution An Easy and Efficient Evolutionary Algorithm for Model Optimisation.** Agricultural Systems, vol. 83, no. 3, p. 315-328, 2005.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais, 2ª ed.,** São Paulo/SP: Edgard Blücher, 2006.

O'CONNOR, N.; MADDEN, M. G. **A Neural Network Approach to Predicting Stock Exchange Movements using External Factors.** Knowledge-Based Systems, vol. 19, no. 5, p. 371-378, 2006.

PHAM, D. T.; KARABOGA, D. **Training Elman and Jordan Networks for System Identification using Genetic Algorithms.** Artificial Intelligence in Engineering, vol. 13, no. 2, p. 107-117, 1998.

PINHEIRO, J. L. **Mercado de Capitais – Fundamentos e Técnicas.** 4<sup>a</sup> ed, São Paulo/SP, Atlas, 2007.

REN, S.; GAO, L. **Resolve of Overlapping Voltammetric Signals in using a Wavelet Packet Transform based Elman Recurrent Neural Network.** Journal of Eletroanalytical Chemistry, vol. 586, no. 1, p. 23-30, 2005.

ROH, T. H. **Forecasting the Volatility of Stock Price Index.** Expert Systems with Applications, vol. 33, no. 4, p. 916-922, 2007.

SALMAN, A. ENGELBRECHT, A. P.; OMRAN, M. G. H.; **Empirical Analysis of Self-Adaptive Differential Evolution.** European Journal of Operational Research, vol. 183, no. 2, p. 785-804, 2007.

SANTOS, A. A. P. **Previsão Não-linear da Taxa de Câmbio Real/Dólar utilizando Redes Neurais e Sistemas Nebulosos.** Dissertação de Mestrado em Economia, Universidade Federal de Santa Catarina, Florianópolis/SC, 2005.

SANTOS, A. A. P.; COSTA, N. C. A.; COELHO, L. S. **Computational Intelligence Approaches and Linear Models in Case Studies of Forecasting Exchange Rates.** Expert Systems with Applications, vol. 33, no. 4, p. 816-833, 2007.

SCHULLI, E. S., **Reconhecimento de Gestos de Maestro Utilizando Redes Neurais Artificiais Parcialmente Recorrentes.** Dissertação de Mestrado em Engenharia Elétrica, Universidade Federal do Paraná, Curitiba/PR, 2005.

SILVA, E. L.; MENEZES, E. M. **Metodologia da Pesquisa e Elaboração de Dissertação.** – 4<sup>a</sup> ed. rev. atual. Laboratório de Ensino a Distância da UFSC, Florianópolis/SC, 2005.

SOTO, C. P. **Redes Neurais Temporais para o Tratamento de Sistemas Variantes no Tempo.** Dissertação de Mestrado em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro – Rio de Janeiro/RJ, 1999.

STORN, R.; PRICE K. **Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces**, *Jornal of Global Optimization*, vol. 11, no. 4, p. 341-359, 1997.

TSAY, R. S. **Analysis of Financial Time Series**, 2 ed., New Jersey: Wiley Series In Probability and Statistics, New Jersey, USA, 2005.

YAO, X. **Evolving Artificial Neural Networks**, *Proceedings of IEEE*, vol. 87, no. 9, p. 1423-1439, 1999.